COSS Notes on trading

Carlos Oscar Sorzano
14th December 2015

Contents

Chapter 1

Introduction

1.1 Utility theory

An *utility function* is a twice-differentiable function of wealth $U(w)$ $(w > 0)$ that has the properties of *non-satiation* ($U'(w) > 0$, that is, a little bit of more wealth is always desirable) and *risk aversion* ($U''(w) < 0$) (Norstad, 1999b). These two properties make the function to be monotonically increasing and concave. $U(w) = \sqrt{w}$ is one such a function. These properties stem from the Von Neumann–Morgenstern expected utility theorem. There are four axioms that make an investor rational:

- Completeness: For every investment choice A and B, either $A \succeq B$ or $A \preceq B$, i.e.. given two choices one is preferred to the other or both are indifferent.

- Transitivity: For every A, B and C with $A \succeq B$ and $B \succeq C$, we must have $A \succeq C$, i.e., the investor is consequent with his preferences.

- Independence: Let A, B, and C be three lotteries with $A \succeq B$, and let $t \in (0, 1]$; then $tA + (1 - t)C \succeq tB + (1 - t)C$, i.e., the choices A, B, C keep their preference order independently of the third (this is the weakest assumption).

- Continuity: Let A, B and C be lotteries with $A \succeq B \succeq C$; then there exists a probability p such that B is equally good as $pA + (1 - p)C$, i.e., B is half-way between A and C.

The *Principle of expected utility maximization* states that a rational investor maximizes his expected utility of wealth. In general, the risk-averse investor will always refuse to play a fair game (for instance, betting on tossing a coin), a game with expected return of 0% (this is a consequence of the concavity of the utility function). If the return is greater than 0%, he may choose or not to play the game. Another way of looking at risk aversion is that investors attach greater weight to losses than they do to gains of equal magnitude (the slope of the utilty function is larger at the left than at the right).

Let us consider an investment that brings wealth from a value w_0 to a value $w_1 = \mathrm{E}\{U(w)\}$ (note that we take expectations because the outcome of the investment is a random variable). The *certainty equivalent* is a wealth value w_{eqv} such that

$$U(w_{eqv}) = U(w_1) \tag{1.1}$$

If $w_0 < w_{eqv}$, the investor will consider the investment attractive. Otherwise, he will see it not so attractive. If $w_0 <= w_{eqv}$, the investor will be indifferent between undertaking the investment or doing nothing.

If $U(w)$ is an utility function, any positive affine transformation

$$V(w) = aU(w) + b \tag{1.2}$$

with $a > 0$, is also an utility function. Actually, if two utility functions are related by an affine transformation, they induce exactly the same behaviour in the investor, and both functions are said to be the "same".

The *Iso-elastic utility* function is a parametric family of utility functions defined by a coefficient *Coefficient of risk aversion*, A, as

$$U_A(w) = \begin{cases} \frac{w^{1-A}-1}{1-A} & A > 0, A \neq 1 \\ \log(w) & A = 1 \end{cases} \tag{1.3}$$

When$A$ increases, the investor is more risk-averse. The square-root utility introduced at the beginning belongs to this family (with $A = 0.5$). A property of this family of functions is the iso-elasticity property

$$U_A(kw) = f(k)U_A(w) + g(k) \tag{1.4}$$

It implies that if a given percentage asset allocation is optimal for some current level of wealth, that same percentage asset allocation is also optimal for all other levels of wealth (Norstad, 1999b). In other words, an investor with iso-elastic utility function has a constant attitude towards risk expressed as a percentage of his current wealth.

The *Negative exponential utility* functions are defined as

$$U_B(w) = -e^{-Bw} \tag{1.5}$$

This family of functions is invariant to translations in wealth:

$$U_B(w + w_0) = f(w_0)U_B(w) + g(w_0) \tag{1.6}$$

In general as these investors' wealth increases, they become more conservative. They invest the same amount of money no matter their wealth. They have a constant attitude towards risk expressed in absolute dollar terms. This property is called *constant absolute risk aversion*.

For any utility function we may calculate the *Arrow-Pratt measure of absolute risk-aversion* (ARA)

$$A(w) = -\frac{U''(w)}{U'(w)} \tag{1.7}$$

Two utitlity functions are the "same" iff the ARAs of both functions are identical. We may also define the measure of relative risk-aversion

$$R(w) = wA(w) \tag{1.8}$$

For example

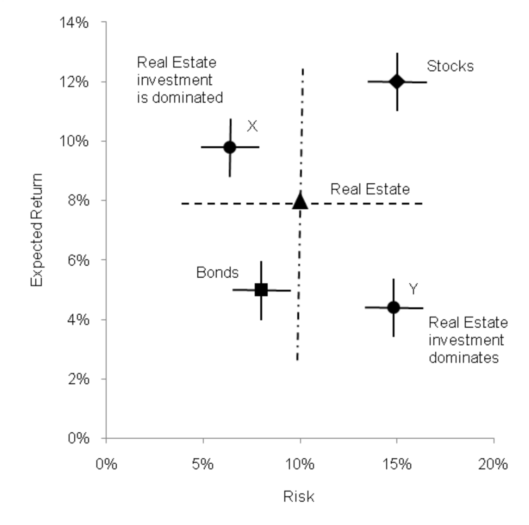


Figure 1.1: Comparison of several investments in a risk (X-axis) and return (Y-axis) plane.

Family	Definition	$A(w)$	$R(w)$
Iso-elastic	$\begin{cases} \frac{w^{1-A}-1}{1-A} & A>0, A\neq 1 \\ \log(w) & A=1 \end{cases}$	$\frac{A}{w}$	A
Linear relative risk-aversion	$\int_0^w x^{-A}e^{-Bx}dx$	$\frac{A}{w}+B$	$A+Bw$
Negative exponential	$-e^{-Bw}$	B	Bw
Quadratic	$w-Bw^2$	$\frac{2B}{1-2Bw}$	$\frac{2Bw}{1-2Bw}$

We see that the linear relative risk-aversion family is a generalization of both (iso-elastic and negative exponential) families.

Given two investments I_1 and I_2 whose ending values at time t are $w_1(t)$ and $w_2(t)$, it is said that I_1 is *more efficient* than I_2 iff (Norstad, 1999a)

$$\mathrm{E}\{U(w_1(t))\} > \mathrm{E}\{U(w_2(t))\} \tag{1.9}$$

for all utility functions U. In the Fig. 1.1, we are comparing investments in real-state versus other possible investments. Real-state investment is dominated by investment X (because X has more return and less standard deviation), but it dominates Y. Real-state is not comparable to bonds or stocks.

Assume that I_1 and I_2 have returns R_1 and R_2 that are normally distributed with distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Then, I_1 is more efficient than I_2 iff $\mu_1 \geq \mu_2$ and $\sigma_1 \leq \sigma_2$ (Norstad, 1999a). However, bear in mind that the normal distribution is not usually a good approximation to the return distribution (a log normal distribution is a better model).

Given a set of feasible investments, I is said to be *efficient* if there is no other investment in the set more efficient than I. Fig. 1.2 shows the efficient frontier of a set of investments. None of the investments in the frontier is dominated by any other investment, while the inefficient portfolios are dominated by the minimum variance portfolio (the leftmost investment).

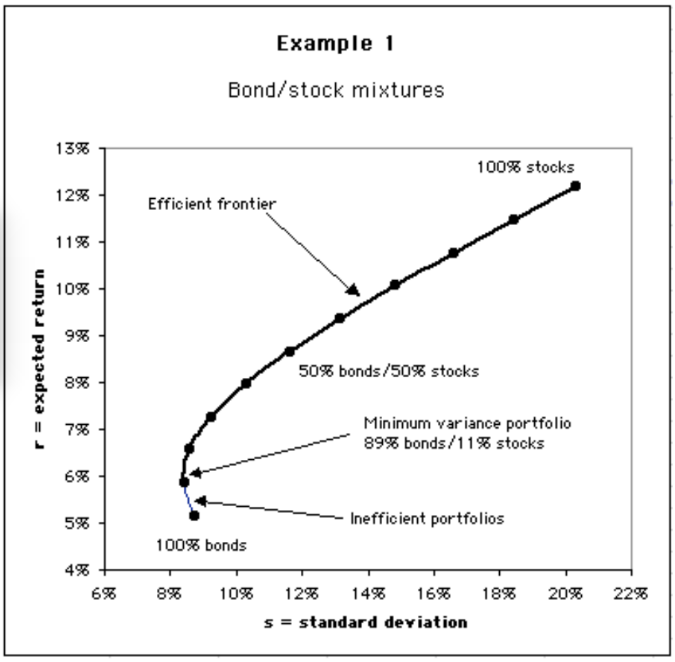


Figure 1.2: Efficient frontier of feasible investments in US bonds and stocks.

The *Risk premium* is the minimum amount of money by which the expected return on a risky asset must exceed the known return on a risk-free asset in order to induce an individual to hold the risky asset rather than the risk-free asset. Formally, the risk premium is π in the equation

$$U(R_f) = \mathrm{E}\{U(R_f + \pi + x\} \tag{1.10}$$

where R_f is the return of a risk-free asset, and x is a zero-mean random variable.

To finalize this introduction, let us quote Norstad (Norstad, 1999b): *Without some kind of evidence there is no reason to believe that any of these particular utility functions which we have examined describes all investors or even any individual investor or the average investor. It is entirely reasonable for an investor's attitudes towards risk to vary with the amount of wealth the investor has accumulated, and it's reasonable for different investors to have different patterns of risk aversion as functions of wealth.*

William Sharpe says the following about the notions of constant relative risk aversion and constant absolute risk aversion in reference: "The assumption of constant relative risk aversion seems much closer to the preferences of most investors than does that of constant absolute risk aversion. Nonetheless, it is by no means guaranteed to reflect every Investor's attitude. Some may wish to take on more risk ... as their wealth increases. Others may wish to take on less. Many analysts counsel a decrease in such risk as one ages. Some strategies are based on acceptance of more or less risk, based on economic conditions. And so on.

For these and other reasons it is important to at least consider strategies in which an Investor's risk tolerance... changes from time to time. However, such changes, if required at all, will likely be far more gradual than those associated with a constant risk tolerance expressed in terms of end-of-period value."

1.2 Stochastic processes

A *Discrete Stochastic Process* is a collection of random variables $\{X_0, X_1, X_2, ...\}$. A *Markov chain* is a discrete stochastic process in which the probability of observing a certain value at time t only depends on the observed value at time $t-1$ (that is, the rest of the history of the random process does not affect the next sample)

$$\Pr\{X_t|X_{t-1}, X_{t-2}, ..., X_0\} = \Pr\{X_t|X_{t-1}\} \tag{1.11}$$

This implies that the slope at the current point of a time series does not affect the next sample in the time series, i.e., the slope is a false illusion.

A *Martingale* is a discrete random process that satisfies

$$\begin{array}{rcl} \mathrm{E}\{|X_t|\} & < & \infty \\ \mathrm{E}\{X_{t+1}|X_t, X_{t-1}, ..., X_0\} & = & X_t \end{array} \tag{1.12}$$

That is the expected value of next observation is the current value of the time series. This is a way of modelling fair games, and it implies that the expected win from one sample to the next is 0. In other words, in the long run there is no hope to win in a martingale.

Given a stochastic process, a *Stopping Time* or *Stopping Rule* is another random variable τ that takes the values $0, 1, 2, ...$ such that $\Pr\{\tau = k\}$ only depends on $\{X_0, X_1, ..., X_k\}$. For example, in a coin tossing game (head: +1\$, tails: -1\$), an example of a stopping time is the first time at which the balance becomes a fixed value (positive or negative). An example of a non-stopping time is the first time that the balance has a peak. This is not a stopping time because for assessing that $\tau = k$ is the peak time, we need to look into the future (specifically, at $k+1$). There is a theorem that states that if τ is a bounded, stopping time for a martingale, then $\mathrm{E}\{X_\tau\} = X_0$. That is, there is no win at the stopping time either.

Given a stochastic process, the random variable defined as $S_N = \sum\limits_{k=0}^{N} X_k$ is a *Random Walk*. The random walk is a Markov chain and a martingale. One of the simplest random walks is the one defined on the random variables $X_k = 1$ (with probability 1/2) or $X_k = -1$ (with probability 1/2). An interesting problem for this random walk is the following: Given that $S_N = k$, what is the probability that the random walk hits the value $S = G$ (G for gain) before hitting the value $S = -L$ (L for loss), let us call $f(k)$ that probability. Note that

$$\begin{array}{rcl} f(k) & = & \frac{1}{2}f(k-1) + \frac{1}{2}f(k+1) \\ f(G) & = & 1 \\ f(-L) & = & 0 \end{array} \tag{1.13}$$

This is a recurrence equation whose solution is

$$f(k) = \frac{L+k}{L+G} \tag{1.14}$$

If the X_k variables are independent, then

$$\begin{aligned} \mathrm{E}\{S_N\} &= N\mathrm{E}\{X_k\} = 0 \\ \mathrm{Var}\{S_N\} &= N\mathrm{Var}\{X_k\} = N\sigma_X^2 \end{aligned} \tag{1.15}$$

That is, the standard deviation of S_N increases with $\sqrt{N}$

$$\sigma_{S_N} = \sqrt{N}\sigma_X \tag{1.16}$$

As can be seen, stock prices do not "indefinitely" grow with time. This is an indication that they are not a "perfect" random walk. However, random walks are still a good model for many of their properties.

Random walks are not stationary processes because their variance increases with time. For a simple random walk with white noise steps we have

$$\begin{aligned} \mathrm{Var}\{S_N\} &= N\sigma_X^2 \\ \mathrm{Cov}\{S_N, S_{N-k}\} &= (N-k)\sigma_X^2 \quad 0 < k < N \\ \mathrm{Corr}\{S_N, S_{N-k}\} &= \frac{\sqrt{N-k}}{\sqrt{N}} = \sqrt{1-\frac{k}{N}} \end{aligned} \tag{1.17}$$

Again, the correlation and covariance depend on time, N, which makes the random walk to be non-stationary.

1.2.1 Lognormal random walks

In this section we will present the relationship between random walks and stochastic differential equations. For doing so, we will start by reasoning on return compounding (Norstad, 2005). Consider an initial amount of money X_0. Let us assume that along the year, we are given N times (dividing the year in N small periods) an interest $\frac{\mu}{N}$. Let us calculate the year compound return:

$$X_1 = X_0\left(1+\frac{\mu}{N}\right)^N = X_0(1+R) \Rightarrow R = \left(1+\frac{\mu}{N}\right)^N - 1 \tag{1.18}$$

When N goes to infinity

$$R = \lim_{N\to\infty}\left(1+\frac{\mu}{N}\right)^N - 1 = e^\mu - 1 \tag{1.19}$$

and after a year the new balance is

$$X_1 = X_0(1+R) = X_0 e^\mu \tag{1.20}$$

After t years, we would have

$$X(t) = X_0 e^{\mu t} \tag{1.21}$$

If we differentiate this function we get

$$dX(t) = \mu X_0 e^{\mu t} dt = \mu X(t) dt \tag{1.22}$$

From which we get a differential equation that the *continuous compounding* must satisfy

$$\frac{dX}{X} = \mu dt \tag{1.23}$$

Let us consider now N investment periods, each one with a random return R_k. At the end of the N periods, we would have:

$$X_N = X_0 \prod_{k=1}^{N} (1 + R_k) \Rightarrow \log \frac{X_n}{X_0} = \sum_{k=1}^{N} \log(1 + R_k) \tag{1.24}$$

That is, $\log \frac{X_n}{X_0}$ is a random walk, and applying the Central Limit Theorem, its asymptotic distribution is normal. Even after 1 day, it is reasonably normal (since in a day we have 25,200 trading seconds). We may now extend the differential equation above to include uncertainty around the instantaneous return rate μ

$$\frac{dX}{X} = \mu(1 + \sigma R(t))dt \tag{1.25}$$

where $R(t) \sim N(0, 1)$. The solution of this equation is

$$\log \frac{X(t)}{X(0)} = \int_0^t \mu(1 + \sigma R(\tau))d\tau = \mu t + \sigma \int_0^t R(\tau)d\tau = \mu t + \sigma S(t) \tag{1.26}$$

where $S(t)$ is a continuous random walk, and $S(t) \sim N(0, t)$. Finally, we have the general expression for the balance

$$X(t) = X_0 e^{\mu t + \sigma S(t)} \tag{1.27}$$

Compare this equation to Eq. 1.21. Gathering the information above, we may state that

$$\log \frac{X(t)}{X(0)} \sim N(\mu t, t\sigma^2) \tag{1.28}$$

that is, the variance of our balance increases with time. We may estimate μ and σ from the annual returns of X as

$$\begin{aligned} \mu &= \mathrm{E}\{\log(1 + R)\} \\ \sigma &= \sqrt{\mathrm{Var}\{\log(1 + R)\}} \end{aligned} \tag{1.29}$$

For the S&P 500 from 1926 to 1994, the estimates are $\mu = 6.63\%$ and $\sigma = 19.65\%$ when t is expressed in years (Norstad, 2002a)[Table 2]. For cash (30 day US Treasure Bills), $\mu = 0.54\%$ and $\sigma = 4.22\%$. Finally, for 20 year US bonds, $\mu = 1.64\%$ and $\sigma = 9.60\%$.

Some financial analysts recommends that investments should span several investment periods (time diversification) as a way to reduce the variance of the average return. Although this is true (the average return is defined as $\bar{R} = \frac{1}{N} \sum_{k=1}^{N} R_k$ and $\mathrm{Var}\{\bar{R}\} = \frac{\sigma_R^2}{N^2}$), what we care is about the value of the investment (X_N), and not the average return ($\bar{R}$), and as explained above the variance of X_N increases with time. This goes against the popular "wisdom" that investing in the long term in volatile assets reduces the risk.

Although the similarity between stock markets and random walks is appealing, there are a number of assumptions of random walks that are violated by stock markets (Norstad, 2005):

1. Independence: Random walks assume that the variables being summed (in our case $\log(1+R_k)$) are independent. But this is not the case in stock markets, if the market goes up one day, the probability that it will also go up the following day is slightly larger than 0.5 (but so little larger that we cannot exploit this property, because of the trading costs, to invest). This is a positive correlation between trading days that creates a short-term momemtum. There is also a negative correlation with respect to the market: if a stock has outperformed the market for a relatively long period, its price will tend, eventually (and the problem is that nobody knows when), to the mean. This is called "reversion to the mean". Markets seem to be more volatile in the short-term than random walks and less volatile than random walks in the long term.

2. Identical distribution: Random walks assume that the variables being summed are identically distributed. However, there is no good reason to think that the expected return of a given stock is constant over time. It depends on the performance of the company, its competitors, the market state, the global state of affairs, ... Standard deviations are not constant either, there are periods of lower volatility and periods of larger volatility.

3. Finite variance: Random walks assume that the variables being summed have finite variance. However, some researchers believe that returns have a fractal nature or stable Paretian distribution, and the variance of some of these distributions do not have finite variance. Return distributions are also fat-tailed compared to the log-normal (very large losses or gains are more likely than in the log-normal).

In any case, although imperfectly, random walks are a good approximation to understand some of the properties of stock markets.

There are rather advanced models for log-return random walks. One of such models is Barndoff-Nielsen and Shephard (BN-S) (Nicolato and Venardos, 2003). The model is based on a random differential equation given by

$$\begin{aligned} dX_t &= (\mu + \beta\sigma_t^2)dt + \sigma_t dW_t + \rho dZ_{\lambda t} \\ d\sigma_t^2 &= -\lambda\sigma_t^2 dt + dZ_{\lambda t} \end{aligned} \tag{1.30}$$

where W_t is a Brownian random process and $Z_{\lambda t}$ is a subordinator (that is a Lévy process with no Gaussian component and positive increments). Alternative models are exponential Lévy processes and jump diffusion models.

1.2.2 The nature of the random walk

Bouchaud et al. (2004) provides a model for the causes of the random walk in stocks. Let us see some excerpts of this article

[...] we show that the random walk nature of traded prices results from a very delicate interplay between two opposite tendencies: long-range correlated market orders that lead to super-diffusion (or persistence), and mean reverting limit orders that lead to sub-diffusion (or anti-persistence). We define and study a model where the price, at any instant, is the result of the impact of all past trades, mediated by a non constant 'propagator' in time that describes the response of the market to a single trade.

[...]

One of the central predictions of EMH is thus that prices should be random walks in time which (to a good approximation) they indeed are. This was interpreted early on as a success of EMH. However, as pointed out by Schiller, the observed volatility of markets is far too high to be compatible with the idea of fully rational pricing. The frantic activity observed in financial markets is another problem: on liquid stocks, there is typically one trade every 5 seconds, whereas the time lag between of relevant news is certainly much larger. More fundamentally, the assumption of rational, perfectly informed agents seems intuitively much too strong, and has been criticized by many. Even the very concept of the fair price of a company appears to be somewhat dubious.

There is a model at the other extreme of the spectrum where prices also follow a pure random walk, but for a totally different reason. Assume that agents, instead of being fully rational, have zero intelligence and take random decisions to buy or to sell, but that their action is interpreted by all the others agents as potentially containing some information.

[...]

Of course, reality should lie somewhere in the middle: clearly, the price cannot wander arbitrarily far from a reasonable value, and trades cannot all be random. The interesting question is to know which of the two pictures is closest to reality.

In this paper, we want to argue, based on a series of detailed empirical results obtained on trade by trade data, that the random walk nature of prices is in fact highly non trivial and results from a fine-tuned competition between two populations of traders, liquidity providers ('market-makers') on the one hand, and liquidity takers (sometimes called 'informed traders'). For reasons that we explain in more details below, liquidity providers act such as to create anti-persistence (or mean reversion) in price changes that would lead to a sub-diffusive behaviour of the price, whereas 'liquidity takers' action leads to long range persistence and super-diffusive behaviour. Both effects very precisely compensate and lead to an overall diffusive behaviour, at least to a first approximation, such that (statistical) arbitrage opportunities are absent, as expected.

They analyzed trading orders accurate to the second. In the following ϵ is a variable that takes the value $+1$ if the price of the following trade is higher than the last price, or -1 if the price is smaller. Let p_n denote the price of the n-th trade, let us define $D(l)$ as

$$D(l) = \left\langle (p_{n+l} - p_n)^2 \right\rangle \tag{1.31}$$

In the absence of linear correlations, $D(l)$ has a strictly diffusive behaviour

$$D(l) = Dl \tag{1.32}$$

for some constant D. The absence of linear correlations in price changes is equivalent to saying that (statistical) arbitrage opportunies are absent, even for high frequency trading. Real markets almost behave in this way in real life (see Fig. 1.3). *The conclusion is that the random walk (diffusive) behaviour of stock prices appears even at the trade by trade level, with a diffusion constant D which is of the order of the typical bid-ask squared. From Fig. 1.3 one indeed sees that* $D(1) \approx 0.01$ *Euros, which is precisely the tick size, and FT has a typical bid-ask spread equal to one or two ticks. This coincidence is interesting. It might*

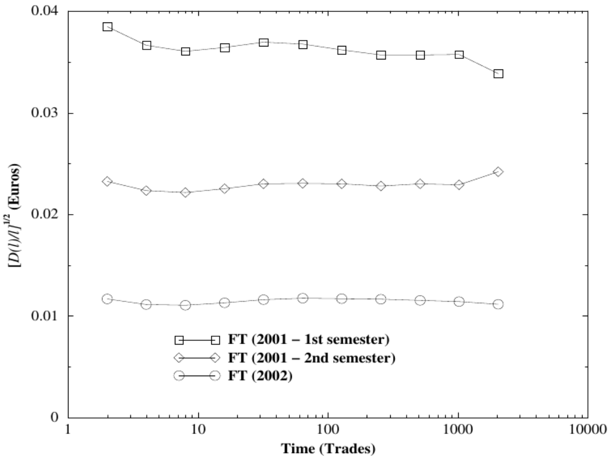


Figure 1.3: $\sqrt{\frac{D(l)}{l}}$ for France Telecom over three different time periods.

suggests that price changes are to a large extent induced by the trading activity itself, independently of real news.

To better understand the impact of trading on the price changes, they study the response function $R(l)$ at a given volume V

$$R(l,V) = \langle (p_{n+l} - p_n)\epsilon_n \rangle|_{V_n=V} \tag{1.33}$$

They propose that this response can be factorized as

$$R(l,V) = R(l)f(V) = \log(V) \tag{1.34}$$

and an average $R(l)$ is shown at Fig. 1.4. The particular amplitude and maximum location depends on the specific stock. However, the shape does not, most stocks have this behaviour. It has also been stablished that approximately

$$D(l) \propto R^2(l) \tag{1.35}$$

Analyzing this data: ... *means that one can hardly detect the statistical presence of informed trades that correctly anticipate the sign of the price change on a short term basis, such as to at least cover their trading costs. This result is consistent with the conclusion of other studies, where it is established that investors 'trade too much', and that the uninformed price pressure is large.*

Long term correlations

All the above results are compatible with a 'zero intelligence' picture of financial markets, where each trade is random in sign and shifts the price permanently, because all other participants update their evaluation of the stock price as a function of the last trade. [...]. This model of a totally random stock market is however qualitatively incorrect for the following reason. Although, as mentioned above, the statistics of price changes reveals very little temporal correlations, the

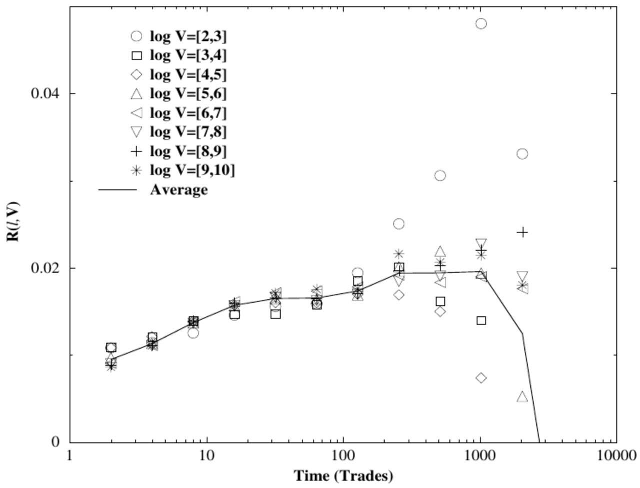


Figure 1.4: $\frac{R(l,V)}{\log(V)} = R(l)$ for different volumes.

correlation function of the sign ϵ_n of the trades, on the other hand, reveals very slowly decaying correlations.

Let C_0 be the correlation function defined as

$$C_0(l) = \langle \epsilon_{n+l}\epsilon_n \rangle - \langle \epsilon_n \rangle^2 \tag{1.36}$$

It has been shown that

$$C_0(l) = \frac{C_0}{l^\gamma} \tag{1.37}$$

with γ taking typical values like 0.2, 0.5, 0.6, ... Always $\gamma < 1$, which implies that the integral of $C_0(l)$ is divergent. This integral can be thought of as the number of correlated succesive trades. For $\gamma = 0.2$, the number of correlated trades is about 50. We may also consider correlation functions including the volume, it is shown that (see Fig. 1.5)

$$\begin{aligned} C_1(l) &= \langle \epsilon_{n+l}\epsilon_n \log(V_n) \rangle \approx C_0(l) \langle \log(V) \rangle \\ C_2(l) &= \langle \epsilon_{n+l} \log(V_{n+l}) \epsilon_n \log(V_n) \rangle \approx C_0(l) \langle \log(V) \rangle^2 \end{aligned} \tag{1.38}$$

Micro-model of price fluctuations

They postulate the following trade superposition model

$$p_n = \sum_{n'<n} G_0(n-n')\epsilon_{n'} \log(V_{n'}) + \sum_{n'<n} \eta_{n'} \tag{1.39}$$

where p_n is the price at time n, G_0 is the bare impact function or propagator (a deterministic function) of a single trade. η_n are random variables independent of ϵ_n and uncorrelated in time. They show that this model explains the behaviour

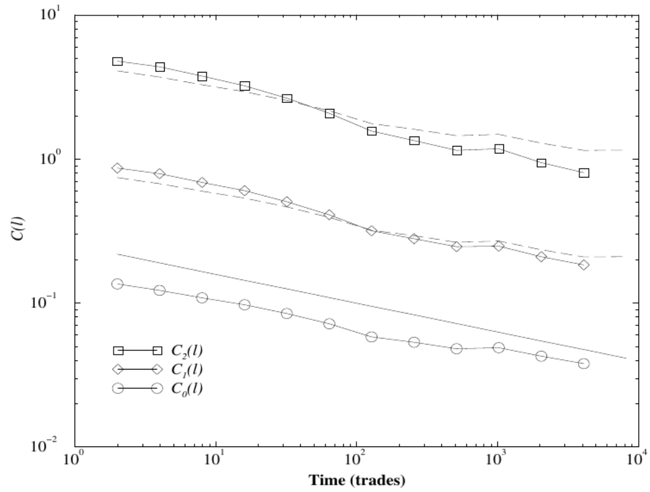


Figure 1.5: $C_0(l), C_1(l), C_2(l)$. The straight line corresponds to $l^{-\gamma}$ with $\gamma = 0.2$.

of $R(l)$ and $D(l)$ as observed in real markets. The propagator is parametrized as (see Fig. 1.6)

$$G_0(l) = \frac{\Gamma_0 l_0^\beta}{(l_0 + l)^\beta} \tag{1.40}$$

There is a critical value $\beta_c = \frac{1-\gamma}{2}$ where γ is the coefficient of the correlation of the ϵ_n (see Eq. 1.37). For $\beta > \beta_c$ the price is sub-diffusive, which means that the price changes show anti-persistence; while for $\beta < \beta_c$ the price is super-diffusive, i.e., the price is persistent.

Critical balance of opposite forces: Market orders vs. limit orders

Although trading occurs for a large variety of reasons, it is useful to recognize that traders organize in two broad categories:

- *One is that of 'liquidity takers', that trigger trades by putting in market orders. The motivation for this category of traders might be to take advantage of some 'information', and make a profit from correctly anticipating future price changes. Information can in fact be of very different nature: fundamental (firm based), macro-economical, political, statistical (based on regularities of price patterns), etc. Unfortunately, information is often hard to interpret correctly, and it is probable that many of these 'information' driven trades are misguided . For example, systematic hedge funds which take decisions based on statistical pattern recognition have a typical success rate of only 52%. There is no compelling reason to believe that the intuition of traders in markets room fares much better than that. Since market orders allows one to be immediately executed, many impatient investors, who want to liquidate their position, or hedge, etc. might be tempted to place market orders, even at the expense of the bid-ask spread.*

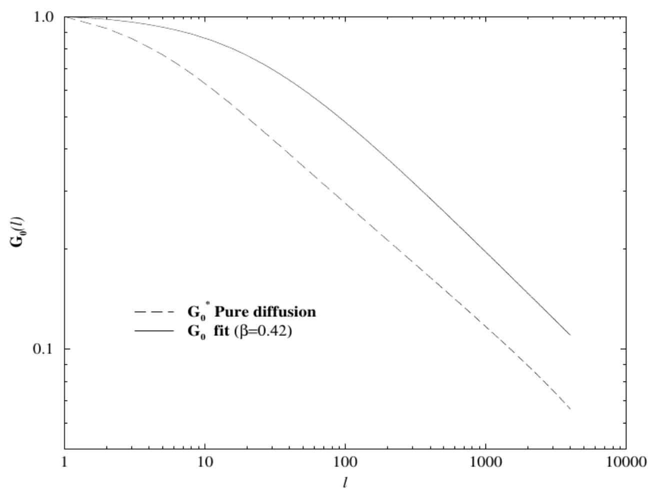


Figure 1.6: G_0 example. $\beta = 0.42$ and $l_0 = 20$.

- *The other category is that of 'liquidity providers' (or 'market makers', although on electronic markets all participants can act as liquidity providers by putting in limit orders), who offer to buy or to sell but avoid taking any bare position on the market. Their profit comes from the bid-ask spread: the sell price is always slightly larger than the buy price, so that each round turn operation leads to a profit equal to the spread, at least if the midpoint has not changed in the mean time (see below).*

This is where the game becomes interesting. Assume that a liquidity taker wants to buy, so that an increased number of buy orders arrive on the market. The liquidity providers is tempted to increase the offer (or ask) price a because the buyer might be informed and really know that the current price is too low and that it will most probably increase in the near future. Should this happen, the liquidity provider, who has to close his position later, might have to buy back at a much higher price and experience a loss. In order not to trigger a sudden increase of a that would make their trade costly, liquidity takers obviously need to put on not too large orders. This is the rationale for dividing one's order in small chunks and disperse these as much as possible over time so as not to appear on the 'radar screens'. Doing so liquidity takers necessarily create some temporal correlations in the sign of the trades. Since these traders probably have a somewhat broad spectrum of volumes to trade, and therefore of trading horizons (from a few minutes to several weeks), this can easily explain the slow, power-law decay of the sign correlation function $C_0(l)$ *reported above.*

Now, if the market orders in fact do not contain useful information but are the result of hedging, noise trading, misguided interpretations, errors, etc., then the price should not move up on the long run, and should eventually mean revert to its previous value. Liquidity providers are obviously the active force behind this mean reversion, again because closing their position will be costly if the price has moved up too far from the initial price.

To summarize: liquidity takers must dilute their orders and create long range correlations in the trade signs, whereas liquidity providers must correctly handle the fact that liquidity takers might either possess useful information (a rare situation, but that can be very costly since the price can jump as a result of some significant news), or might not be informed at all and trade randomly. By slowly mean reverting the price, market makers minimize the probability that they either sell too low, or have to buy back too high. The delicate balance between these conflicting tendencies conspire to put the market at the border between persistence (if mean reversion is too weak, i.e. $\beta < \beta_c$*) or anti-persistence (if mean reversion is too strong, i.e.* $\beta > \beta_c$ *), and therefore eliminate arbitrage opportunities.*

It is actually enlightening to propose a simple model that could explain how market makers enforce this mean reversion. Assume that upon placing limit orders, there is a systematic bias toward some moving average of past prices. If this average is for simplicity taken to be an exponential moving average, the continuous time description of this will read:

$$\begin{aligned} \frac{dp_t}{dt} &= -\Omega(p_t - \bar{p}_t) + \eta_t \\ \frac{d\bar{p}_t}{dt} &= -\kappa(p_t - \bar{p}_t) \end{aligned} \tag{1.41}$$

where η_t *is the random driving force due to trading,* Ω *the inverse time scale for the strength of the mean reversion, and* $\frac{1}{\kappa}$ *the 'memory' time over which the average price* p_t *is computed. The first equation means that liquidity providers tend to mean revert the price toward* $\bar{p}_t$*, while the second describes the update of the exponential moving average* $\bar{p}_t$ *with time.*

Conclusion

The aim of this paper was to study in details the statistics of price changes at the trade by trade level, and to analyze the interplay between the impact of each trade on the price and the volatility. Empirical data shows that (a) the price (midpoint) process is close to being purely diffusive, even at the trade by trade scale (b) the temporal structure of the impact function first increases and reaches a maximum after 100 - 1000 trades, before decreasing back, with a rather limited overall variation (typically a factor 2) and (c) the sign of the trades shows surprisingly long range (power-law) correlations. The paradox is that if the impact of each trade was permanent, the price process should be strongly super-diffusive and the average response function should increase by a large factor as a function of the time-lag.

As a possible resolution of this paradox, we have proposed a micro-model of prices, Eq. 1.39 where the price at any instant is the causal result of all past trades, mediated by what we called a bare impact function, or propagator G0 . All the empirical results can be reconciled if one assumes that this bare propagator also decays as a power-law in time, with an exponent which is precisely tuned to a critical value, ensuring simultaneously that prices are diffusive on long time scales and that the response function is nearly constant. Therefore, the seemingly trivial random walk behaviour of price changes in fact results from a fined-tuned competition between two opposite effects, one leading to super-diffusion (the autocorrelation of trades) and the other leading to sub-diffusion (the decay of the bare impact function). The cancellation is however not exact: the non trivial behaviour of the average response function allows one to detect *small, but systematic deviations from a purely diffusive behaviour, deviations that are hardly detectable on the price fluctuations themselves.*

In financial terms, the competition is between liquidity takers, that create long range correlations by dividing their trading volume in small quantities, and liquidity providers that tend to mean revert the price such as to optimize their gains. The resulting absence of correlations in price changes, and therefore of arbitrage opportunities is often postulated a priori in the economics literature, but the details of the mechanism that removes these arbitrage opportunities are rather obscure. The main message of this paper is that the random walk nature of price changes is not due to the unpredictable nature of incoming news, but appears as a dynamical consequence of the competition between antagonist market forces. In fact, the role of real (and correctly interpreted) news appears to be rather thin: we have defined a model independent indicator of the fraction of 'informed' trades, as the asymmetry of the probability distribution of the signed price variation, where the sign is that of the trade at the initial time. Information triggered trades should reveal in a detectable positive skew of this distribution, in particular in the tails. Consistently with other studies, our empirical results only show very weak asymmetry, barely sufficient to cover trading costs, which means that only a small fraction of trades can a posteriori described as truly informed, whereas most trades can be classified as noise. This result is most probably one of the mechanism needed to explain the excess volatility puzzle first raised by Schiller.

Behavioral finance explanation

Hommes (2006) surveys many Heterogeneous Agent Models proposed in the literature. One of the most interesting is the one of Lux (1995). It is based on Physics master equation. Let us assume there are N trading agents who respond to fundamental and technical analysis. Fundamental traders buy a stock when its market price is below their estimated "true" fundamental price of the asset, p_f. They sell when the market price is above the fundamental price. Chartist traders buy when the price is going up, and they sell when the price is going down. In this model, traders are allowed to switch from optimistic to pessimistic, and from chartist to fundamentalist, and viceversa. Let n_f be the number of fundamental traders and n_c the number of chartist traders. Within chartist traders, let n_c^+ be the number optimistic (bullish) traders and n_c^- the number of pessimistic (bearish) traders. The opinion index is defined as

$$x = \frac{n_c^+ - n_c^-}{2} \in [-1, 1] \tag{1.42}$$

Let us define the proportion of chartist traders as

$$z = \frac{n_c}{N} \in [0, 1] \tag{1.43}$$

A chartist buys (sells) a fixed amount t_c of the asset per period when he is optimistic (pessimistic). The excess demand created by chartist is modelled as

$$ED_c = (n_c^+ - n_c^-)t_c = xzNt_c \tag{1.44}$$

A fundamentalist buys (sells) when the market price, p, is below (above) the fundamental price, p_f. The excess demand created by fundamentalists is

$$ED_f = n_f \gamma(p_f - p) = (1 - z)N\gamma(p_f - p) \tag{1.45}$$

where γ measures the reaction speed of fundamentalists to price deviations. Finally, a market maker adjusts prices according to excess demands

$$\frac{dp}{dt} = \beta(ED_c + ED_f) = \beta(xzNt_c + (1-z)N\gamma(p_f - p)) \tag{1.46}$$

where β denotes the speed of the adjustment. Lux (1995) proposes that the index opinion changes as

$$\frac{dx}{dt} = \frac{\frac{dn_c^+}{dt} - \frac{dn_c^-}{dt}}{n_c} - \frac{n_c^+ + n_c^-}{2n_c^2}\frac{dn_c}{dt} \tag{1.47}$$

Finally, the proportion of chartists is governed by

$$\frac{dz}{dt} = \frac{1}{N}\frac{dn_c}{dt} \tag{1.48}$$

With these equations they were capable of producing price behaviors and correlations like the ones shown in Fig. 1.7.

Economical explanation

Greenwald et al. (2014) explains the quarterly fluctuation of stock markets in terms of three underlying economical variables. These three variables explain up to the 85% of the fluctuations.

We begin by identifying three mutually orthogonal observable economic disturbances that are associated with the vast majority (over 85%) of ‡uctuations in real quarterly stock market wealth since the early 1950s. Econometrically, these shocks are measured as specific orthogonal movements in consumption, labor income, and asset wealth (net worth), identified from a cointegrated vector autoregression (VAR) and extracted using a recursive orthogonalization procedure.

Specifically, the consumption innovation in the empirical VAR would recover a total factor productivity (TFP) shock, the labor income innovation would recover a factors share shock that reallocates the rewards of production without affecting the size of rewards, and the wealth innovation would recover a shock to shareholder risk aversion that moves the stochastic discount factor pricing assets independently of stock market fundamentals or real activity such as consumption and labor earnings. We show that the dynamic responses to these mutually orthogonal VAR innovations produced from model generated data are remarkably similar to those obtained from historical data.

With this theoretical interpretation of the observable disturbances in hand, we turn to the question of how these distinct shocks have affected stock market wealth over time. We find that the vast majority of short- and medium-term stock market fluctuations in historical data are driven by risk aversion shocks, revealed as movements in wealth that are orthogonal to consumption and labor income both contemporaneously (an identifying assumption), and at all subsequent horizons (a result). Although transitory, these shocks are quite persistent and explain 75% of variation in the log difference of stock market wealth on a quarterly basis. These facts are well explained by the model, in which the orthogonal wealth shocks originate from independent shifts in investors' willingness to bear risk. At longer horizons, the relative importance of the shocks changes.

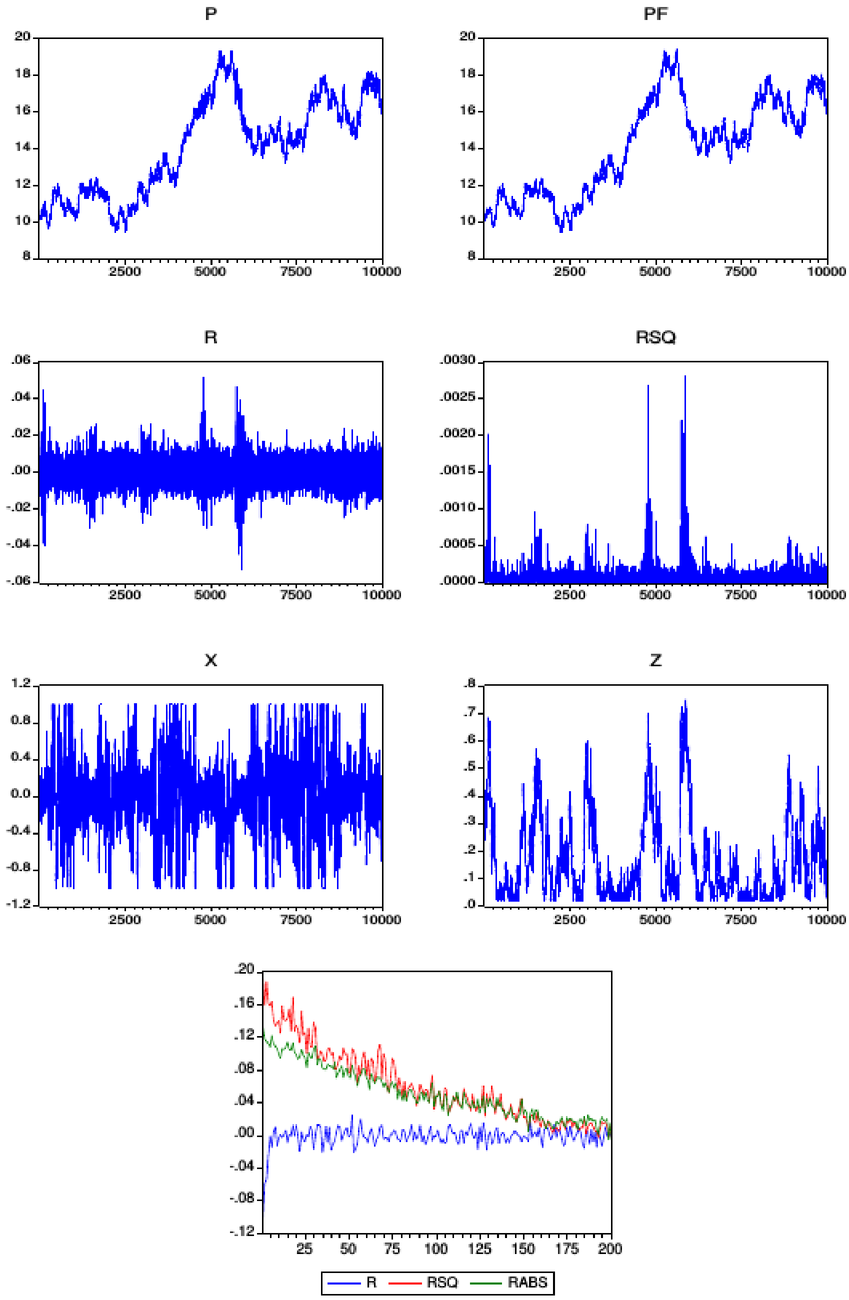


Figure 1.7: Time series of prices (top left), fundamental price (top right), returns (second panel, left), squared returns (second panel, right), opinion index (third panel, left), fraction of chartists (third panel, right) and autocorrelation patterns (bottom panel) of returns, squared returns and absolute returns.

Although the factors share shock explains virtually none of the variation in the real level of the stock market over cycles of a quarter or two, it explains roughly 40% over cycles two to three decades long. These facts are well explained by the model economy, which is subject to small but highly persistent innovations that shift the allocation of rewards between shareholders and workers independently from the magnitude of those rewards. By contrast, consumption shocks, both in the model and in the data, play a small role in the stochastic ‡uctuations of the stock market at all horizons. The crucial aspect of the model that makes it consistent with this finding is its heterogeneous agent specification. This finding contradicts representative agent asset pricing models in which shocks that drive aggregate consumption play a central role in stock market fluctuations.

[...]

Our findings are also informative about the origins of risk premia ‡uctuations. Share-holders in the model are close to risk-neutral most of the time but subject to rare "crises" in their willingness to bear risk, captured in the model by infrequent, large spikes in risk aversion that generate a "flight to safety". Even though these flights are rare and extreme, a time-varying expectation that risk tolerance could crash in the future generates plausible variation in the price-dividend ratio and empirically reasonable predictability in excess stock market returns. Time-variation in the risk premium, both in the model and the data, is revealed by the wealth shocks, which are orthogonal to movements in consumption and labor income. We find that these innovations also bear little relation to other traditional macroeconomic fundamentals such as dividends, earnings, consumption volatility, or broad-based macroeconomic uncertainty, and none of these other variables forecast equity premia. These findings are hard to reconcile with models in which time-varying risk premia arise from habits (which vary with innovations in consumption), stochastic consumption volatility, or consumption uncertainty.

1.2.3 Non-random walks

Lo and MacKinlay (2011) maintains that the price of an asset is not a random walk. It has instead some underlying trend either for fundamental reasons (the performance and quality of a company) or for momentum reasons (many people buying the same stock makes its price raise, which attract more people). They propose that the returns are modelled by

$$\log R[n] = \mu + \epsilon[n] \tag{1.49}$$

where $\epsilon[n]$ is a random innovation (which can be modelled by a GARCH-like model).

Dhar and Kumar (2001) gives some reasons why prices do not follow a random-walk. Here follow some paragraphs of this article.

Investors may trade for a variety of reasons such as portfolio rebalancing, tax-loss selling, liquidity reasons, lifecycle considerations, over-confidence or for purely speculative reasons. Trading may also be driven by changes in investor beliefs about the future stock prices and these beliefs are likely to be influenced by past price trends. Along with the fundamental information about the firm, investors may look at price trends to formulate their trading decisions and they may follow trend-based heuristics ("rules of thumb") such as momentum and *contrarian strategies to decide when to buy and when to sell. Furthermore, investors may differ systematically in their reaction to past trends and the degree to which they follow momentum and contrarian strategies. Clearly, differences in investor trading behavior may be present along dimensions other than price trends and a more complete characterization of investor heterogeneity may help us better understand the link between regularities in investor trading behavior and regularities (mean reversion, low auto-correlation in returns, etc.) in the behavior of asset prices. In this paper we take an initial step in the investigation of investor heterogeneity and analyze the extent to which short-term price trends influence the trading behavior of individual investors. We identify investor segments that systematically trade on trends and characterize the differences in the trading patterns of these identified investor segments.*

The primary reasons for price trend tracking behavior may be psychological. Experimental research in cognitive psychology has shown that people often see patterns in short sequences that may be random (Tversky and Kahneman 1971). In some settings people expect a reversion of trends ("gambler's fallacy"). For instance, in coin tossing experiments, after a sequence of heads, people overestimate the probability of a tail because they expect an equal number of heads and tails even in small sequences. People show a tendency to draw inferences using the law of large numbers in small samples where the law does not apply. In various other situations, people expect a continuation of the past trend ("hot hand"). Gilovich, Vallone, and Tversky (1985) find that people expect continuation of a hot streak in basketball even though it has been found that basketball shots follow a random walk. These experimental results suggest that investors are likely to detect patterns in stock prices even if stock prices follow a random walk and trend tracking behavior may be widespread in the financial markets.

Recent research on investor behavior, both experimental and empirical, suggest that investors tend to look at short-term and long-term price trends in formulating their trading decisions. Experimental evidence of trend-tracking behavior has been documented by Andreassen and Kraus (1990), DeBondt (1991) and DeBondt (1993). Andreassen and Kraus (1990) find that experimental subjects follow a trend chasing strategy, extrapolating price changes, when price trends are dominant. DeBondt (1993) finds that individual investors expect continuation of upward price trends in bullish markets and continuation of downward trends in bearish markets. In contrast, "experts" behave as contrarians, i.e., they expect price-reversals in both bullish and bearish markets (DeBondt 1991).

[...]

We find that both buying and selling decisions of investors in our sample are influenced by short-term (less than 3 months) price trends. By comparing the observed distributions of average trend before buys and average trend before sells with the average trend distributions (obtained using Monte Carlo simulations) when investors trade randomly, the null of non-trend motivated random trading is easily rejected (p-value $<$ *0.002). Using Monte Carlo simulations again, we examine investor heterogeneity in trading based on prior returns and classify investors into (i) momentum buy (MB), (ii) momentum sell (MS), (iii) contrarian buy (CB) or (iv) contrarian sell (CS) category. [...] A comparison of the portfolio characteristics and demographics of the identified investor segments reveal no significant differences. However, the trading characteristics of the segments show systematic differences, particularly in their response to reference points such as monthly high and low prices and in their strategies for selling* *losers. Contrarian buy investors are more likely to buy near monthly low prices while the contrarian sell investors tend to sell near the monthly high prices. The momentum investors do not exhibit such timing behavior. All four investor segments are reluctant to sell losers but the effect is the strongest for contrarian sell investors who expect price reversals and hence show a greater tendency to hold on to the losers. The effect is very weak for momentum sell investors who believe that a downward price trend is likely to continue and hence are more likely to realize their losses.*

[...]

reversals and hence show a greater tendency to hold on to the losers. If the observed systematic differences in trading behavior among investor segments is widespread among other groups of investors in the market, it may have important implications for asset pricing. Asset pricing models incorporating patterns of systematic investor trading behavior have recently been proposed (Barberis, Huang, and Santos 2001), (Barberis and Huang 2001) and results from our paper can provide further input to these theoretical asset pricing models. The behavioral patterns that we have uncovered can also help us understand the properties of asset returns. Various agent-based models of financial markets (Huang and Day 1993, Farmer and Joshi 2001) have shown that markets with momentum and contrarian (or value) investors can generate price dynamics that exhibit characteristics of empirical returns series. These agent-based models that simulate market dynamics can be further refined and made more realistic by utilizing the results from our paper.

Our results also suggest that noise trader risk (DeLong, Shleifer, Summers, and Waldmann 1990) in the market may be limited. Momentum and contrarian investors have diametrically opposite expectations and their presence in the market introduce both destabilizing and restor- ing forces in the market. In the presence of these two opposing forces, asset prices may not diverge significantly from the fundamental value and the amount of noise trader risk in the market may be limited. At any given time, the magnitude of this risk will depend upon the relative proportion of momentum and contrarian investors in the market and so the "popula- tion dynamics" of momentum and contrarian investors may determine the magnitude of noise trader risk.

Finally, the existence of momentum and contrarian investors may also partially explain why there is often high trading volume and large price movements in the market without any significant news events (Cutler, Poterba, and Summers 1989). The "internal dynamics" of momentum and contrarian investors may be partially responsible for such regularities in the market. Papers such as (Goetzmann and Massa 2000b) have already started to provide evidence that internal risk factors (or "market created risk" (Kraus and Smith 1989)), factors that originate from the trading behavior of market participants, may add to the explanatory power of asset pricing models.

Even though psychological factors explain well why investors exhibit trend tracking behav- ior, non-psychological factors such as informational asymmetries (Hirshleifer, Subrahmanyam, and Titman 1994) and differential interpretation of information (Kandel and Pearson 1993) may also induce trend tracking behavior among investors. Non-psychological factors may also be responsible for the observed disposition effect. A recent paper by (Ranguelova 2000) finds that the disposition effect is present primarily in large cap stocks and surprisingly, in the lower decile stocks, the propensity to sell losers is higher than the propensity *to sell winners. Cer tainly, more research is needed to determine whether the psychological or the non-psychological factors are the main determinants of the observed patterns in investor trading behavior.*

Conrad and Kaul (1998) also finds that momemtum and contrarian strategies are the only ones that consistently give profits.

In this article we use a single unifying framework to analyze the sources of profits to a wide spectrum of return-based trading strategies implemented in the literature. We show that less than 50% of the 120 strategies implemented in the article yield statistically significant profits and, unconditionally, momentum and contrarian strategies are equally likely to be successful. However, when we condition on the return horizon (short, medium, or long) of the strategy, or the time period during which it is implemented, two patterns emerge. A momentum strategy is usually profitable at the medium (3- to 12-months) horizon, while a contrarian strategy nets statistically significant profits at long horizons, but only during the 1926–1947 subperiod. More importantly, our results show that the cross-sectional variation in the mean returns of individual securities included in these strategies play an important role in their profitability. The cross-sectional variation can potentially account for the profitability of momentum strategies and it is also responsible for attenuating the profits from price reversals to long-horizon contrarian strategies.

Trading strategies that apparently "beat the market" date back to the inception of trading in financial markets. A number of practitioners and academics in the pre-market-efficiency era (i.e. pre-1960s) believed that predictable patterns in stock returns could lead to "abnormal" profits to trading strategies. [...]

In this interplay between momentum and contrarian strategies, there must be a little bit of overreaction (a kind of second order system). This overreaction has been reported by Bondt and Thaler (1985).

Research in experimental psychology suggests that, in violation of Bayes' rule, most people tend to "overreact" to unexpected and dramatic news events. This study of market efficiency investigates whether such behavior affects stock prices. The empirical evidence, based on CRSP monthly return data, is consistent with the overreaction hypothesis. Substantial weak form market inefficiencies are discovered. The results also shed new light on the January returns earned by prior "winners" and "losers". Portfolios of losers experience exceptionally large January returns as late as five years after portfolio formation.

Brock et al. (1992) shows the profitability of simple trading rules like those based on moving averages (when a fast trend MA goes above 1% of a slow trend MA there is a buy signal, and when it goes below -1%, there is a sell signal, typically 1-200, 2-100, 2-150 MA curves are considered) or those based on resistences (maxima of the last 150 days) and supports (minima of the last 150 days). These profits of these simple rules cannot be explained by AR, GARCH, EGARCH models and random walks.

This paper tests two of the simplest and most popular trading rules—moving average and trading range break—by utilizing the Dow Jones Index from 1897 to 1986. Standard statistical analysis is extended through the use of bootstrap techniques. Overall, our results provide strong support for the technical strategies. The returns obtained from these strategies are not consistent with four popular null models: the random walk, the AR(1), the GARCH-M, and the Exponential GARCH. Buy signals consistently generate higher returns than sell signals, and further, the returns following buy signals are less volatile than returns following *sell signals, and further, the returns following buy signals are less volatile than returns following sell signals. Moreover, returns following sell signals are negative, which is not easily explained by any of the currently existing equilibrium models.*

1.3 Stochastic differential equations

A stochastic differential equation is a differential equation in which one of the participating terms is stochastic. For instance, the Geometric Brownian Motion differential equation is

$$\frac{\frac{dS}{dt}}{S} = \mu + \sigma \frac{dB}{dt} \tag{1.50}$$

or equivalently

$$dS = \mu S dt + \sigma S dB \tag{1.51}$$

where S is the price of an asset over time an B is some Brownian motion driving the stock price. The integral version of this equation is

$$S_t = S_0 + \int_0^t \mu S(\tau) d\tau + \int_0^t \sigma B(\tau) d\tau \tag{1.52}$$

The first integral is well defined in the standard sense, but the second one involves a stochastic signal and it is not so standard.

In the particular case of the Geometric Brownian Motion, the solution is

$$S(t) = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)} \tag{1.53}$$

and it is written $S(t) \sim GBM(\mu, \sigma^2)$.

Another example is given by the assumption that $X(t) = \log S(t)$ is an Ornstein-Uhlenbeck (OU) process, this implies that

$$dX = -\gamma(X - \alpha)dt + \sigma dB \tag{1.54}$$

Its solution is

$$S(t) = \exp\left(\alpha + exp(-\gamma t)(X(0) - \alpha) + \sigma \exp(-\gamma t) \int_0^t \exp(-\gamma\tau) dB(\tau)\right) \tag{1.55}$$

An important result regarding stochastic differential equations is Itô's lemma. Suppose that $X(t)$ is an Itô's process with

$$dX = a(X,t)dt + b(X,t)dB \tag{1.56}$$

Let $Y(t) = F(X,t)$, then

$$dY = \left(\frac{\partial F}{\partial X} a + \frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial X^2} b^2\right) dt + \frac{\partial F}{\partial X} b dB \tag{1.57}$$

For example

$$\begin{array}{rcl} & & dX = \mu dt + \sigma dB \\ Y = X^2 & \Rightarrow & dY = dBdB + dt \\ Y = e^{at+bX} & \Rightarrow & dY = e^{at+bX}\left[\left(a + \frac{1}{2}b^2\right)dt + bdB\right] \\ Y = e^{-\frac{1}{2}\sigma^2 t + \sigma B} & \Rightarrow & dY = \sigma dB \end{array} \quad (1.58)$$

An interesting process of this kind is Cox-Ingersoll-Ross model (CIR model) which is described in the form

$$dX = \alpha(\mu(t) - X)dt + \sigma\sqrt{X}dB \quad (1.59)$$

This does not have analytic solution. But we may simulate any differential equation using an explicit Euler scheme

$$X_n - X_{n-1} = a(X_{n-1}, (n-1)T_s)T_s + b(X_{n-1}, (n-1)T_s)(W_n - W_{n-1}) \quad (1.60)$$

Starting from X_0 we may simulate this process many times, Monte Carlo, and compute the empirical distribution of X_N. From this distribution we may calculate average values, confidence intervals, ...

An important family of stochastic processes are the jump-diffusion models that are defined by the differential equation

$$\frac{dS}{S^-} = \mu dt + \sigma dBdJ \quad (1.61)$$

where S^- is the left limit of $S(t)$ and

$$J(t) = \sum_{i=1}^{N(t)}(Y_i - 1) \quad (1.62)$$

where $N(t)$ is the number of jumps up to time t (a random variable) and Y_j is a sequence of IID random variables. The solution is of the form

$$S(t) = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)} \prod_{i=1}^{N(t)} Y_i \quad (1.63)$$

The following table shows some important differential equations and their solutions (Saito and Mitsui, 1993)

Martingale	$dX = XdW$	$X(t) = e^{-\frac{1}{2}t + W(t)}$
Submartingale	$dX = Xdt + XdW$	$X(t) = e^{\frac{1}{2}t + W(t)}$
Supermartingale	$dX = -Xdt + XdW$	$X(t) = e^{-\frac{3}{2}t + W(t)}$
	$dX = X(1-X)dt + XdW$	$X(t) = \frac{e^{\frac{1}{2}t + W(t)}}{2 + \int_0^t e^{\frac{1}{2}\tau + W(\tau)} d\tau}$

Itô's integral is defined as the inverse of Itô's derivative

$$F = \int f dB + \int g dt \Leftrightarrow dF = f dB + g dt \quad (1.64)$$

Itô's integral can be thought as the limit of Riemmanian sums if the leftmost point of the small intervals are considered in the Riemman integration.

Using Itô's integral we may define Itô's isommetry. It is said that $f(t)$ is adapted to $B(t)$ iff

$$\mathrm{E}\left\{\left(\int_0^t f(s)dB\right)^2\right\} = \mathrm{E}\left\{\int_0^t f^2(s)dB\right\} \tag{1.65}$$

For $f(s) = 1$ we get the quadratic variation of Brownian motion.

1.3.1 Brownian motion

We may think of a Brownian motion (also called a Wiener process) as the limit of a random walk when the time steps are infinitely small. The following theorem guarantees its existence.

Theorem. There exists a probability distribution over the set of real valued functions

$$B : \mathbb{R}^+ \to \mathbb{R} \tag{1.66}$$

such that

- The function $B(t)$ starts at 0. $\Pr\{B(0) = 0\} = 1$
- The function $B(t)$ is stationary. $B(t_F) - B(t_0) \sim N(0, t_F - t_0)$
- Increments are independent. If (t_1, s_1) and (t_2, s_2) define two non-overlapping time intervals, then $B(s_1) - B(t_1)$ is independent from $B(s_2) - B(t_2)$

The following are some properties of a Brownian motion:

- It crosses the time axis infinitely often.
- It does not go too much away from t^2, at any time t_0, we have $B(t_0) \sim N(0, t_0)$.
- It is nowhere differentiable. It means that standard Calculus is not correct, and Itô's Calculus has to be used.
- Let us denote $M(t)$ as the maximum of $B(t)$ up to time t ($M(t) = \max_{s \le t} B(s)$). Then

$$\Pr\{M(t) > a\} = 2\Pr\{B(t) > a\} \quad \forall t, a > 0 \tag{1.67}$$

- The probability of being above or below a certain level is symmetric. Let us denote as t_a the time at which $B(t)$ crosses the line $B(t) = a$ from below (the first time it reaches a). Then

$$\Pr\{B(t) - B(t_a) > 0 | t_a < t\} = \Pr\{B(t) - B(t_a) < 0 | t_a < t\} \tag{1.68}$$

- It fulfills a property called *Quadratic variation.*

$$\lim_{n\to\infty} \sum_{t=1}^{n} \left(B\left(\frac{t}{n}T\right) - B\left(\frac{t-1}{n}T\right)\right)^2 = T \tag{1.69}$$

  A consequence of this property is

$$(dB)^2 = dt \tag{1.70}$$

Let us consider a function of a Brownian motion $x = f(B)$. In standard Calculus we normally have

$$df = f'(B)dB \tag{1.71}$$

But with stochastic processes this is incorrect. In the case of the Brownian motion (which fulfills the Quadratic variation property), we have

$$df = f'(B)dB + \frac{f''(B)}{2}dt \tag{1.72}$$

This latter formula is similar to a Taylor expansion of order 2.

We may now consider a function $x = f(t, B)$, a simplified version of Itô's lemma states

$$df = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial B^2}\right)dt + \frac{\partial f}{\partial t}dB \tag{1.73}$$

Let $B(t) \sim N(0, t)$ be a brownian motion and $X(t) = \int f(t)dB$, then $X(t)$ is normally distributed. For example

$$X(t) = \int \sigma dB \sim N(0, \sigma^2 t) \tag{1.74}$$

The integral of Brownian motions are martingales. Formally,

$$\int f(t, B)dB \tag{1.75}$$

is a martingale as long as f is an L^2 function. As a consequence, the stochastic process defined by the differential equation

$$dX = \mu(t, B)dt + \sigma(t, B)dB \tag{1.76}$$

is a martingale as long as there is no drift ($\mu(t, B) = 0$).

Thanks to Girsanov theorem (which is a consequence of Radon-Nikodym derivative) we may shift from a drifted brownian motion to a martingale by a simple multiplication. Let $B(t)$ be a Brownian motion with drift and $\tilde{B}(t)$ a Brownian motion without drift. Let $P(\omega)$ be the probability of a path ω according to the Brownian motion B and $\tilde{P}(\omega)$ be the probability of a path ω according to the Brownian motion $\tilde{B}$. Then, P and $\tilde{P}$ are equivalent, that is,

$$P(\omega) = Z(\omega)\tilde{P}(\omega) \tag{1.77}$$

with

$$Z(\omega) = \frac{d\tilde{P}}{dP}(\omega) = e^{-\mu\omega(T) - \frac{1}{2}\mu^2 T} \tag{1.78}$$

As a consequence, assume that we have some portfolio whose value $R(t)$ is modelled according to the probability distribution P, then

$$\mathrm{E}_P\{R(t)\} = \mathrm{E}_{\tilde{P}}\{Z(\omega)R(t)\} \tag{1.79}$$

1.3.2 Numerical methods

As for the ODEs, there are numerical methods, which are similar to the numerical methods of the ODEs. Consider a scalar autonomous Itô stochastic differential equation:

$$dX = f(X)dt + g(X)dW \quad X(0) = x_0 \tag{1.80}$$

In the following $\Delta W_n = W_n - W_{n-1}$, and ΔW_n must be a zero-mean random variable, whose variance is h (see Eq. 1.16). It is typical to take $\Delta W_n = \epsilon\sqrt{h}$ being $\epsilon \sim N(0,1)$ a random number.

Here are some of these numerical methods (Saito and Mitsui, 1993):

- Euler-Maruyama: Local order 1 and global order 1

$$X_{n+1} = X_n + f(X_n)h + g(X_n)\Delta W_n \tag{1.81}$$

- Heun: Local order 2 and global order 2

$$X_{n+1} = X_n + \frac{F_1 + F_2}{2}h + \frac{G_1 + G_2}{2}\Delta W_n \tag{1.82}$$

with

$$\begin{aligned} F(x) &= \left(f - \tfrac{1}{2}g'g\right)(x) \\ F_1 &= F(X_n) \\ G_1 &= g(X_n) \\ F_2 &= F(X_n + F_1 h + G_1\Delta W_n) \\ G_2 &= g(X_n + F_1 h + G_1\Delta W_n) \end{aligned} \tag{1.83}$$

- Kloeden: Local order 2 and global order 2, derivative free

$$X_{n+1} = X_n + F_1 h + G_1\Delta W_n + (G_2 - G_1)h^{-\frac{1}{2}}\frac{(\Delta W_n)^2 - h}{2} \tag{1.84}$$

with

$$\begin{aligned} F_1 &= f(X_n) \\ G_1 &= g(X_n) \\ G_2 &= g(X_n + G_1\sqrt{h}) \end{aligned} \tag{1.85}$$

Saito and Mitsui (1993) discusses some more schemes, some of them of order 3.

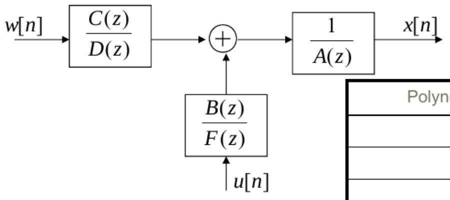
1.4 Statistical time series models

There are a number of statistical models trying to reproduce the same autocorrelation (and, consequently, spectral) characteristics as the input signal. These models may be also used for forecasting. We may distinguish between 3 different issues: model definition, model estimation, and forecasting.

1.4.1 Model definition

In the following let us assume that $X_d[n]$ denotes the samples of a time series of interest, and $W[n]$ is normally some input white noise signal. $U[n]$ is an external, known signal. See Fig. 1.8 for a graphical summary of many of these models.

General model



$$X(z) = \frac{B(z)}{F(z)A(z)}U(z) + \frac{C(z)}{D(z)A(z)}W(z)$$

Polynomials used	Name of the model
A	AR
C	MA
AC	ARMA
ACD	ARIMA
AB	ARX
ABC	ARMAX
ABD	ARARX
ABCD	ARARMAX
BFCD	Box-Jenkins

Figure 1.8: Different kinds of models.

MA(Q) models

Moving average models are defined by

$$X[n] = \sum_{k=0}^{Q} b_k W[n-k] \tag{1.86}$$

The transfer function of this system is

$$H(z) = \frac{X(z)}{W(z)} = \sum_{k=0}^{Q} b_k z^{-k} = B(z) \tag{1.87}$$

If the input $W[n]$ is normally distributed $N(0, \sigma_W^2)$, then

$$X[n] \sim N\left(0, \sigma_W^2 \sum_{k=0}^{Q} b_k^2\right) \tag{1.88}$$

The autocorrelation of the output is

$$\Gamma_X[l] = \begin{cases} \Gamma_X[-l] & l < 0 \\ \sigma_W^2 \sum\limits_{k=0}^{Q-l} b_k b_{k+l} & 0 \leq l \leq Q \\ 0 & l > Q \end{cases} \tag{1.89}$$

MA models can be extended to non-causal systems

$$X[n] = \sum_{k=Q_0}^{Q_F} b_k W[n-k] \tag{1.90}$$

with $Q_0 < 0$. They can also be extended to non-linear models as

$$X[n] = \sum_{k=0}^{Q} b_k f(W[n-k]) \tag{1.91}$$

where f is some non-linear function, or we may use Volterra kernels to express this non-linearity

$$\begin{aligned} X[n] = & \sum_{k=0}^{Q} b_k W[n-k] \\ & + \sum_{k=0}^{Q} \sum_{k'=0}^{Q'} b_{kk'} W[n-k]W[n-k'] \\ & + \sum_{k=0}^{Q} \sum_{k'=0}^{Q'} \sum_{k''=0}^{Q''} b_{kk'k''} W[n-k]W[n-k']W[n-k''] \\ & \ldots \end{aligned} \tag{1.92}$$

AR(P) models

Autoregressive models are defined as

$$X[n] = \sum_{k=1}^{P} a_k X[n-k] + W[n] \tag{1.93}$$

The autocorrelation of this process is given by Yule-Walker equations

$$\Gamma_X[l] = \sigma_W^2 \delta[l] + \sum_{k=1}^{P} a_k \Gamma_X[l-k] \tag{1.94}$$

Its solution is of the form

$$\Gamma_X[l] = \sum_{k=1}^{P} A_k |z_k|^l \tag{1.95}$$

where z_k are the poles of the transfer function

$$H(z) = \frac{X(z)}{W(z)} = \frac{1}{1 - \sum\limits_{k=1}^{P} a_k z^{-k}} = \frac{1}{A(z)} \tag{1.96}$$

Again, we can make non-linear extensions like

- <u>Non-linear AR</u>

$$X[n] = f(X[n-1], X[n-2], ..., X[n-P]) + W[n] \tag{1.97}$$

- <u>Time-varying AR</u>

$$X[n] = \sum_{k=1}^{P} a_k[n] X[n-k] + W[n] \tag{1.98}$$

- Smooth transition AR

$$X[n] = \left(\sum_{k=1}^{P} a_k[n]X[n-k]\right) p[n] + \left(\sum_{k=1}^{P} a'_k[n]X[n-k]\right) (1-p[n]) + W[n] \tag{1.99}$$

- Random coefficients AR

$$X[n] = \sum_{k=1}^{P} (a_k + \epsilon[n])X[n-k] + W[n] \tag{1.100}$$

- Bilinear models

$$X[n] = \sum_{k=1}^{P} a_k X[n-k] + \sum_{k=1}^{Q} b_k X[n-k]W[n-k-M] + W[n] \tag{1.101}$$

- Threshold AR

$$X[n] = \begin{cases} \sum_{k=1}^{P} a_k X[n-k] + W[n] & X[n-d] \leq t \\ \sum_{k=1}^{P} a'_k X[n-k] + W[n] & X[n-d] > t \end{cases} \tag{1.102}$$

- Smooth threshold AR (STAR)

$$X[n] = \left(\sum_{k=1}^{P} a_k[n]X[n-k]\right) + \left(\sum_{k=1}^{P} a'_k[n]X[n-k]\right) S(X[n-d]) + W[n] \tag{1.103}$$

ARMA(P,Q) models

ARMA models are defined as

$$X[n] = \sum_{k=1}^{P} a_k X[n-k] + \sum_{k=0}^{Q} b_k W[n-k] \tag{1.104}$$

The transfer functions is a combination of both AR and MA models

$$H(z) = \frac{B(z)}{A(z)} \tag{1.105}$$

The autocorrelation is

$$\Gamma_X[l] = \sigma_W^2 \sum_{k=0}^{Q} b_k h[k-l] + \sum_{k=1}^{P} a_k \Gamma_X[l-k] \tag{1.106}$$

where $h[k-l]$ are samples of the impulse response.

ARIMA(P,D,Q) models

We may think of an ARIMA model as an ARMA model followed by an integrator. The ARMA model produces an intermediate signal $X_D[n]$

$$\begin{array}{rcl} X_d(z) & = & \frac{B(z)}{A(z)}W(z) \\ X(z) & = & \frac{1}{(1-z^{-1})^D}X_D(z) \end{array} \tag{1.107}$$

Another way of looking at the model is that the D derivative of $X[n]$ follows an ARMA model. If D is rational, then the model is called ARFIMA. The difference equation for $D = 1$ clearly shows this

$$(X[n] - X[n-1]) = \sum_{k=1}^{P} a_k X[n-k] + \sum_{k=0}^{Q} b_k W[n-k] \tag{1.108}$$

Bhardwaj and Swanson (2006) examines several ARIMA models (normally $P = 1$, $Q = 1$ or $Q = 2$ with $D \in [0.25, 0.6]$) for the signals $|R[n]|$, $R^2[n]$, and $\log(R^2[n])$.

SARIMA(P,D,Q)x(p,d,q)s models

The seasonal ARIMA or Box-Jenkins model is a model in which there are two levels of derivatives: one with a seasonal component ($D = 12$) and another one with a standard derivation. We may think of the model as a white noise input, downsampled by a factor s (the seasonal component, e.g., $s = 30$, $s = 12$, ...), then an ARIMA(P,D,Q) model, upsampling by a factor s, and another ARIMA(p,d,q) model.

$$\begin{array}{rcl} X_s(z) & = & H_{ARIMA(P,D,Q)}(z^s)W(z) \\ X(z) & = & H_{ARIMA(p,d,q)}(z)X_s(z) \end{array} \tag{1.109}$$

Here goes an example of SARIMA$(1, 0, 0) \times (0, 1, 1)_{12}$ in time domain

$$(X[n]-X[n-12]) = B_0W[n]+B_1W[n-12]+a_1(X[n-1]-X[n-13]) \tag{1.110}$$

GARCH models

Heterocedastic models assume that the signal is a random walk, with zero mean and whose variance changes over time

$$X[n] = \sigma[n]W[n] \tag{1.111}$$

The ARCH model is an MA model on the variance

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{P} a_k x^2[n-k] \tag{1.112}$$

while the GARCH model proposes an ARMA model

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{P} a_k x^2[n-k] + \sum_{k=1}^{Q} b_k \sigma^2[n-k] \tag{1.113}$$

The model is constrained by $a_k, bk \geq 0$. The model is unique and stationary if

$$\sum_{k=1}^{P} a_k + \sum_{k=1}^{Q} b_k < 1 \tag{1.114}$$

In the long-run the local variance, $\sigma^2[n]$, tends to σ_0^2. The speed at which this reversion to the mean happens is controlled by the sum above which is named the persistence.

It produces signals with zero mean, $\mathrm{E}\{X[n]\} = 0$, and they lack correlation

$$\Gamma_X[l] = \frac{\sigma_0^2}{1 - \sum\limits_{k=1}^{max\{P,Q\}} (a_k + b_k)} \delta[l] \tag{1.115}$$

The exponential GARCH (EGARCH) work with the logarithm of the time series while the GARCH model proposes an ARMA model

$$\log \sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{P} a_k \log x^2[n-k] + \sum_{k=1}^{Q} b_k \log \sigma^2[n-k] \tag{1.116}$$

and the integrated GARCH (IGARCH) relaxes the condition to

$$\sum_{k=1}^{P} a_k + \sum_{k=1}^{Q} b_k = 1 \tag{1.117}$$

The non-linear asymmetric GARCH (NGARCH or NAGARCH) proposes

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{P} a_k (x[n-k] - \theta\sigma[n-k])^2 + \sum_{k=1}^{Q} b_k \sigma^2[n-k] \tag{1.118}$$

The GARCH-M (GARCH in Mean) adds heterocedasticity to the mean equation

$$y[n] = \beta u[n] + \lambda\sigma[n] + x[n] \tag{1.119}$$

where $u[n]$ is an external driving input, and $x[n]$ follows a GARCH model. The QGARCH (Quadratic GARCH) is also capable of representing assymetry as

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{P} a_k x^2[n-k] + \sum_{k=1}^{Q} b_k \sigma^2[n-k] + \sum_{k=1}^{R} c_k x[n-k] \tag{1.120}$$

Another way to have assymetric GARCH is by the use of thresholds (GJR-GARCH, TGARCH)

$$\begin{aligned} \sigma^2[n] &= \sigma_0^2 + \sum_{k=1}^{P} (a_k + a'_k U(x[n])) x^2[n-k] + \sum_{k=1}^{Q} b_k \sigma^2[n-k] \\ \sigma^2[n] &= \sigma_0^2 + \sum_{k=1}^{P} a_k x^2[n-k] + \sum_{k=1}^{Q} (b_k + b'_k U(x[n])) \sigma^2[n-k] \end{aligned} \tag{1.121}$$

where $U(x)$ is Heaviside's step function.

GARCH models can be useful to model the driving input in the rest of models (ARMA, ARIMA, ...) as shown in Baillie et al. (1996).

Multivariate time series

let $\mathbf{x}[n]$ be a multivariate time series whose mean is $\boldsymbol{\mu}_{\mathbf{x}}$ and its covariance n_0 lag matrix is defined as

$$\Gamma_{\mathbf{x}}[n_0] = \mathrm{E}\{(\mathbf{x}[n] - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}[n-n_0] - \boldsymbol{\mu}_{\mathbf{x}})^T\} \tag{1.122}$$

The autocorrelation matrix is $\Gamma[0]$ and it is a symmetric matrix. The correlation matrix is defined as

$$R_{\mathbf{x}}[n_0] = D_{\mathbf{x}}^{-1/2}\Gamma_{\mathbf{x}}[n_0]D_{\mathbf{x}}^{-1/2} \tag{1.123}$$

where $D_{\mathbf{x}} = \mathrm{diag}\{\Gamma_{\mathbf{x}}[0]\}$. The covariance matrix fulfills

$$\Gamma_{\mathbf{x}}[-n_0] = (\Gamma_{\mathbf{x}}[n_0])^T \tag{1.124}$$

If $(\Gamma_{\mathbf{x}}[n_0])_{ij} \neq 0$ for some n_0, then it is said that $x_i[n]$ leads $x_j[n]$. If $x_i[n]$ leads $x_j[n]$ and $x_j[n]$ leads $x_i[n]$, then there is feedback between the two time series.

We may check whether two series have the same variance by the Bartlett, Levene or Brown-Forsyth tests.

The multivariate Wold decomposition states that any time series can be expressed as the sum of a deterministic time series, $\mathbf{v}[n]$ plus a moving average of a white noise. Formally,

$$\mathbf{x}[n] = \mathbf{v}[n] + \sum_{k=0}^{\infty} B_k \mathbf{w}[n-k] \tag{1.125}$$

where $\boldsymbol{\mu}_{\mathbf{w}} = \mathbf{0}$, $\Gamma_{\mathbf{w}}[0] = \Sigma_{\mathbf{w}}$ is a positive-semidefinite matrix, $\Gamma_{\mathbf{w}}[n_0] = 0$ for any lag $n_0 \neq 0$ (the process is uncorrelated to itself), and $\Gamma_{\mathbf{x},\mathbf{w}}[n_0] = 0$ (the process is uncorrelated to the time series). The B_k matrices are such that $B_0 = I$ and $\sum\limits_{k=0}^{\infty} B_k B_k^T$ converges.

We may have vector MA as well as AR processes

$$\begin{aligned} \mathbf{x}[n] &= \sum_{k=0}^{Q} B_k \mathbf{w}[n-k] \\ \mathbf{x}[n] &= \sum_{k=1}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \end{aligned} \tag{1.126}$$

The interesting part of the multivariate models is that we allow the j-th component of the time series at time $n-k$ to have an influence on the value of the i-th component at time n. A VAR(P) process is stationary if all the roots of the polynomial

$$\det\{I - A_1 z^{-1} - A_2 z^{-2} - ... - A_k z^{-k}\} = 0 \tag{1.127}$$

are within the unit complex circle. Alternatively, some time-series have some diverging properties. They are said to have unit roots. To detect unit-roots we may use the Dickey-Fuller test or the Phillips-Perron test.

VAR(P) models are useful to define Granger causality. After fitting the model, we may test whether all the regression coefficients of x_j on x_i are 0

$$H_0 : (A_1)_{ij} = (A_2)_{ij} = ... = (A_P)_{ij} = 0 \tag{1.128}$$

If we reject the null hypothesis, then we say that $x_j[n]$ Granger causes $x_i[n]$. Note that x_j may cause x_i and viceversa, in this case we have feedback.

Cointegration

If we look at the time series of the Dow Jones and DAX, both are rather similar, too much to be independent random walks. Cointegration is a way of analyzing two time series with underlying common random processes. A multivariate time series $\mathbf{x}[n]$ is *Integrated of order* d, $I(d)$, iff its d differenced process, $(1-L)^d\{\mathbf{x}[n]\}$ is stationary. For instance, a random walk is an $I(1)$ process

$$x[n] = x[n-1] + w[n] \tag{1.129}$$

We need to distinguish between cointegration and spurious regression. For instance, let us assume we have two independent random walks, $x_1[n]$ and $x_2[n]$ and we calculate the regression

$$x_2[n] = \alpha + \beta x_1[n] + \epsilon[n] \tag{1.130}$$

Since both signals are independent, we should see with increasing sample sizes that

- The Durbin-Watson statistics is close to 0.
- R^2 is too large.
- $\epsilon[n]$ is $I(1)$.
- The estimates of α and β are inconsistent between different estimations.
- The t_β statistic diverges with $\sqrt{T}$.

But for the regression

$$\Delta y[n] = \beta \Delta x[n] + \epsilon[n] \tag{1.131}$$

we find that

- β has the usual distribution around 0.
- The t_β values are t-distributed.
- $\epsilon[n]$ is white noise.

For two truly cointegrated variables, the regression in Eq. (1.130)

- β is superconsistent, that is it converges with rate T instead of $\sqrt{T}$.
- The t_β statistic is asymptotically normal only if $\epsilon[n]$ is not serially correlated.

If $\mathbf{x}[n]$ has a Vector AR(P) model (VAR(P))

$$A(L)\mathbf{x}[n] = \mathbf{w}[n], \tag{1.132}$$

then $A(L)$ can be factorized as $A(L) = (1-L)^d A'(L)$.

Interestingly, all the components of $\mathbf{x}[n]$ may be $I(1)$ but $\mathbf{x}[n]$ is not $I(1)$. In the same way, linear combinations of the time series (without any differencing) may be stationary. This brings to the definition of cointegration. A time series $\mathbf{x}[n]$ is cointegrated iff all its components are $I(1)$ and there exists a linear combination of its components, $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}\mathbf{x}[n]$ is a stationary process (i.e. $I(0)$).

Some examples are

- Term structure of interest rates: expectations hypothesis.
- Purchase power parity in foreign exchange: cointegration among exchange rate, foreign and domestic prices.
- Money demand: cointegration among money, income, prices and interest rates.
- Covered interest rate parity: cointegration among forward and spot exchange rates.
- Law of one price: cointegration among identical/equivalent assets that must be valued identically to limit arbitrage (e.g. prices of same asset on different trading venues).

Vector Error Correction Models

VECM are a way of estimating cointegrations. They are also called Cointegrated VAR(P) models. Let us analyze the VAR(P) model

$$\mathbf{x}[n] = \sum_{k=1}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \tag{1.133}$$

Let us assume that $\mathbf{x}[n]$ is a $I(1)$ model. We now subtract $\mathbf{x}[n-1]$ on both sides

$$\begin{aligned} \mathbf{x}[n] - \mathbf{x}[n-1] &= (A_1 - I)\mathbf{x}[n-1] + \sum_{k=2}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \\ \Delta\mathbf{x}[n] &= (A_1 - I)\mathbf{x}[n-1] + \sum_{k=2}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \end{aligned} \tag{1.134}$$

We now subtract and add $(A_1 - I)\mathbf{x}[n-2]$ on the right-hand side

$$\begin{aligned} \Delta\mathbf{x}[n] &= (A_1 - I)(\mathbf{x}[n-1] - \mathbf{x}[n-2]) + (A_2 + A_1 - I)\mathbf{x}[n-2] \\ &\quad + \sum_{k=3}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \\ &= (A_1 - I)\Delta\mathbf{x}[n-1] + (A_2 + A_1 - I)\mathbf{x}[n-2] + \sum_{k=3}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \end{aligned} \tag{1.135}$$

We now subtract and add $(A_2 + A_1 - I)\mathbf{x}[n-3]$ on the right-hand side

$$\begin{aligned} \Delta\mathbf{x}[n] &= (A_1 - I)\Delta\mathbf{x}[n-1] + (A_2 + A_1 - I)\Delta\mathbf{x}[n-2] \\ &\quad (A_3 + A_2 + A_1 - I)\Delta\mathbf{x}[n-3] + \sum_{k=4}^{P} A_k \mathbf{x}[n-k] + \mathbf{w}[n] \end{aligned} \tag{1.136}$$

If we keep doing this up to P times, we arrive to

$$\Delta\mathbf{x}[n] = \Pi\mathbf{x}[n-1] + \sum_{k=1}^{P-1} \Gamma_k \Delta\mathbf{x}[n-k] + \mathbf{w}[n] \tag{1.137}$$

where

$$\begin{aligned} \Pi &= A_1 + A_2 + ... + A_P - I \\ \Gamma_k &= -A_{k+1} - A_{k+2} - ... - A_P \end{aligned} \tag{1.138}$$

This is the Error Correction Model. In this model, the differences of the original signal, which are stationary because the original signal was $I(1)$, are expressed as a linear combination of $\mathbf{x}[n-1]$ plus a linear combination of previous differences. The left-hand side is stationary, and all terms in the right-hand side are clearly stationary, so the first one, $\Pi\mathbf{x}[n-1]$, must also be stationary (this term is called the error correction point). This term contains the cointegration of $\mathbf{x}[n]$. $\mathbf{x}[n]$ is $I(1)$, but $\Pi\mathbf{x}[n-1]$ is $I(0)$ (stationary). Consequently Π must be singular. If the rank of Π is 0, then $\Pi = 0$ and there are no cointegrating terms. If the rank of Π is k, then $\mathbf{x}[n]$ cannot be $I(1)$ but $I(0)$, i.e., stationary. But if the rank of Π is r, then it can be factorized as

$$\Pi = CD^T \tag{1.139}$$

being C and D $m \times r$ matrices (m is the number of components of $\mathbf{x}[n]$) of rank r. The columns of D give the cointegration relationships of $\mathbf{x}[n]$ and there are $m - r$ common stochastic trends. These trends are identified through the orthogonal complement of the matrix C. In the long run, the time series $\mathbf{x}[n]$ tends to one of its equilibrium points, which are the solutions of

$$\Pi\mathbf{x} = \mathbf{0} \tag{1.140}$$

To test for cointegration we may use the trace test with increasing values of r (0,1,2,...)

$$\begin{array}{c} H_0 : \text{Rank}\{\Pi\} = r \\ H_A : \text{Rank}\{\Pi\} > r \end{array} \tag{1.141}$$

or the Maximum eigenvalue test

$$\begin{array}{c} H_0 : \text{Rank}\{\Pi\} = r \\ H_A : \text{Rank}\{\Pi\} = r + 1 \end{array} \tag{1.142}$$

Sreedharan (2004) presents an interesting application of VECM to stock market returns. He uses a 4-valued vector (Open, High, Low, Close). The model is

$$\Delta\mathbf{x}[n] = \boldsymbol{\delta}_0 + \sum_{i=1}^{P} \Gamma_i \Delta\mathbf{x}[n-i] + \sum_{i=0}^{1} B_i \boldsymbol{\zeta}[n-i] + \sum_{i=0}^{1} C_i^+ \boldsymbol{\eta}^+[n-i] + \sum_{i=0}^{1} C_i^+ \boldsymbol{\eta}^-[n-i] + \mathbf{w}[n] \tag{1.143}$$

where $\boldsymbol{\zeta}$ is a measure of the current and immediate-past "normal" information and $\boldsymbol{\eta}$ are the currenct and immediate-past positive and negative expectations. They show that the log of each component of $\mathbf{x}(t)$ is non-stationary, although the logarithmic returns are stationary. The cross-correlation between some variables are clearly different from 0. For instance, the log return of open and the log return of close, low or high; the log return of high and the log return of close, low or high; the log return of low and the log return of low or close. They analyze the histogram of the log-returns and they are highly peaked and moderately skewed (to negative values). They build a VAR(10) model on the log returns. But the model does not explain much of the variance of the close log-returns. For this reason, they hypothesize a model misspecification and look for cointegration equations. They find 3 cointegrations:

$$\begin{array}{rcl} C[n] & = & 0.999681O[n] \\ H[n] & = & 1.005055O[n] \\ L[n] & = & 0.994943O[n] \end{array} \tag{1.144}$$

and conclude that the error correction process is a no-arbitrage process. Since there are 3 cointegration equations, there is a single common stochastic trend. The VECM model is still poor in the R^2 of the log returns of close. Then they construct three cointegrating variables that are used as leading exogeneous variables. In this case, they identify these cointegration variables with the $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ vectors required in their model. After identifying these variables, they show that the residuals $\mathbf{w}$ are normally distributed (which in other cases they are not).

They conclude that *The model also supports the view that asset price dynamics comprise of normal and abnormal shocks [see Merton (1976)]:*

1. *The normal shocks can be due to "temporary imbalance between supply and demand", changes in the price of risk or in the economic outlook, or other new information that causes marginal changes in the asset value.*

2. *The abnormal shocks are due to "the arrival of new important information about the asset that has more than a marginal effect on value".*

1.4.2 Model estimation

The determination of the MA model can be done by brute force by solving the non-linear system implied by the autocorrelation function (Eq. 1.89). For AR models, we may recursively solve the Yule-Walker equations. Bhardwaj and Swanson (2006) provides several methods to estimate d in ARIMA models. They normally work in Fourier space. Baillie et al. (1996) shows how to jointly estimate an ARIMA-GARCH model. Model selection is performed through the classical ways: Akaike Information Criterion (with for a model with P parameters and N training samples)

$$AIC(P) = \log \sigma_\epsilon^2 + P\frac{2}{N}, \tag{1.145}$$

the Bayesian Information Criterion

$$BIC(P) = \log \sigma_\epsilon^2 + P\frac{\log N}{N}, \tag{1.146}$$

or the Final Prediction error

$$FPE(P) = \frac{N+P}{N-P}\sigma_\epsilon^2. \tag{1.147}$$

After fitting the model we should check that the residuals are 1) zero-mean, 2) uncorrelated to themselves, 3) normally distributed.

The Chow test checks whether the same model can be fitted to two different regions of the time series. It does so through the analysis of the residuals with the best model fitted to each one of the possible regions and comparing it to the residuals using a single model (see Fig. 1.9).

1.4.3 Forecasting

As shown in Fig. 1.10, forecasting formulas have to be specifically taylored to each model.

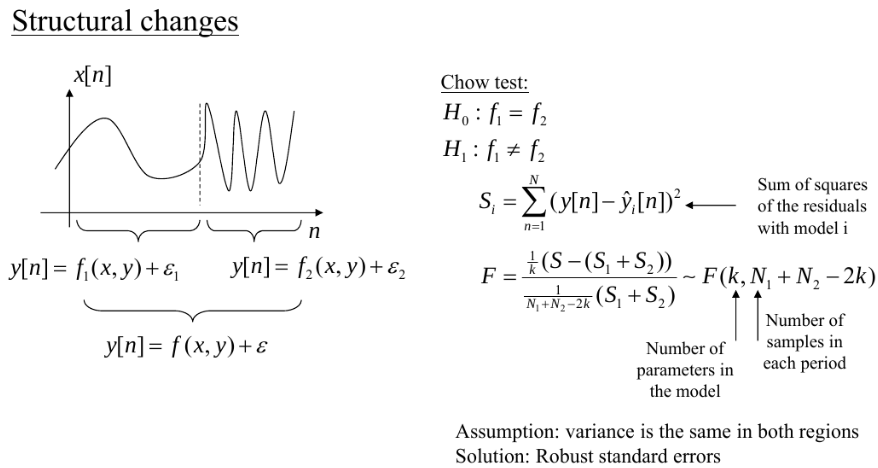


Figure 1.9: Checking for structural heterogeneity

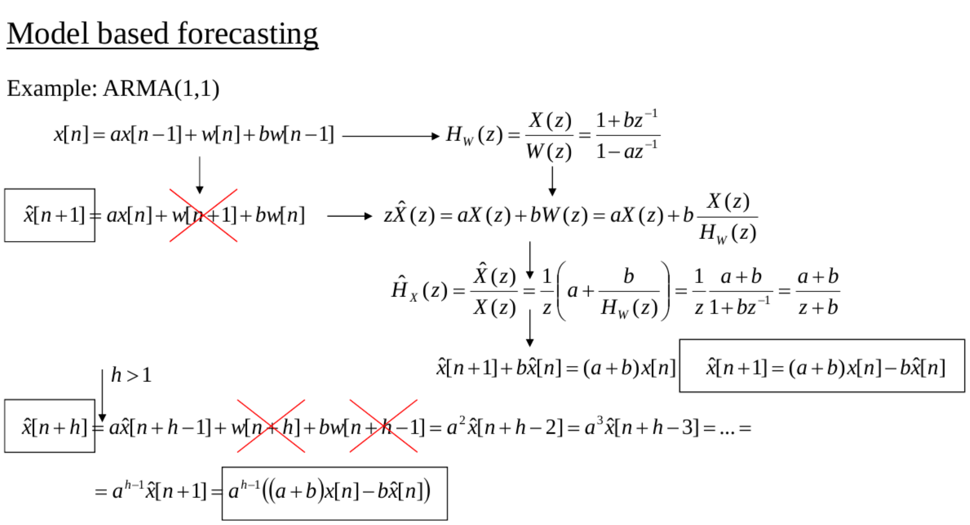


Figure 1.10: Example of ARMA forecasting.

Two possible ways of checking the predictability ability of two competing models is through the Diebold and Mariano and the encompassing tests Bhardwaj and Swanson (2006). These tests check

$$\begin{array}{ll} H_0: & \mathrm{E}\{f(X_0[n+l]) - f(X_1[n+l])\} = 0 \\ H_A: & \mathrm{E}\{f(X_0[n+l]) - f(X_1[n+l])\} \neq 0 \end{array} \quad (1.148)$$

where X_0 and X_1 are the two competing predictions, l is the prediction horizon, f is any loss function. The encompassing alternative hypothesis is $H_A : ... > 0$.

1.4.4 Pricing models

Some interesting models combine several ideas from generic time series models. In the following L is the lag operator (its Z transform is z^{-1}). Some of these follow:

- ARFIMA-GARCH with external inputs: Baillie et al. (1996)

$$\begin{array}{rcl} A(L)(1-L)^D(R[n] - \mu - b_1X_1[n] - \delta\sigma[n]) & = & B(L)\epsilon[n] \\ \alpha(L)\sigma^2[n] & = & \sigma_0^2 + \beta(L)\epsilon^2[n] + b_2X_2[n] \end{array} \quad (1.149)$$

  where X_1 and X_2 are two external inputs.

- Duan's NGARCH model: Hsieh and Ritchken (2005)

$$\begin{array}{rcl} \log R[n] & = & R_f + (\delta + \epsilon[n])\sigma[n] - \frac{1}{2}\sigma^2[n] \\ \sigma^2[n] & = & \sigma_0^2 + \alpha_1\sigma^2[n-1] + \alpha_2(\epsilon[n] - \gamma)^2 \end{array} \quad (1.150)$$

- Heston and Nandi model: Hsieh and Ritchken (2005)

$$\begin{array}{rcl} \log R[n] & = & R_f + \epsilon[n]\sigma[n] + a\sigma^2[n] \\ \sigma^2[n] & = & \sigma_0^2 + \alpha_1\sigma^2[n-1] + \alpha_2(\epsilon[n] - \gamma\sigma[n])^2 \end{array} \quad (1.151)$$

1.5 State-space filtering

1.5.1 Linear state-space filters

Let us assume that we have a dynamic system whose state is $\mathbf{s}[n]$ whose transition equation is

$$\mathbf{s}[n+1] = T[n]\mathbf{s}[n] + R[n]\mathbf{w}[n] \quad (1.152)$$

where $T[n]$ is a matrix, $\mathbf{w}[n]$ is $N(\mathbf{0}, \Sigma_w)$ and Σ_w is an arbitrary positive definite matrix. The observation equation is

$$\mathbf{x}[n] = H[n]\mathbf{s}[n] + \mathbf{u}[n] \quad (1.153)$$

where $H[n]$ is a matrix, $\mathbf{u}[n]$ is $N(\mathbf{0}, \Sigma_u)$ and Σ_u is an arbitrary positive definite matrix. The joint equation is

$$\begin{pmatrix} \mathbf{s}[n+1] \\ \mathbf{x}[n] \end{pmatrix} = \begin{pmatrix} T[n] \\ H[n] \end{pmatrix} \mathbf{s}[n] + \begin{pmatrix} R[n]\mathbf{w}[n] \\ \mathbf{u}[n] \end{pmatrix} = \begin{pmatrix} T[n] \\ H[n] \end{pmatrix} \mathbf{s}[n] + \boldsymbol{\epsilon}[n] \quad (1.154)$$

Note that

$$\Sigma_{\boldsymbol{\epsilon}} = \begin{pmatrix} R[n]\Sigma_{\mathbf{w}}R^T[n] & 0 \\ 0 & \Sigma_u \end{pmatrix} \quad (1.155)$$

Many models seen so far respond to this scheme. For example:

- CAPM: Consider

$$R[n] = \alpha[n] + \beta[n]R_m[n] + \epsilon[n] \tag{1.156}$$

Let the state be formed by

$$\mathbf{s}[n+1] = \begin{pmatrix} \alpha[n+1] \\ \beta[n+1] \end{pmatrix} = \begin{pmatrix} \alpha[n] \\ \beta[n] \end{pmatrix} + \mathbf{w}[n] \tag{1.157}$$

That is, α and β are two random walks. The observation equation is given by

$$R[n] = \begin{pmatrix} 1 & R_m[n] \end{pmatrix} \mathbf{s}[n] + \epsilon[n] \tag{1.158}$$

- Time-varying regression: Consider the time-varying regression

$$R[n] = \beta[n]\mathbf{x}[n] + \epsilon[n] \tag{1.159}$$

Let the state equation be

$$\mathbf{s}[n+1] = \mathbf{x}[n] + \mathbf{w}[n] \tag{1.160}$$

and the observation equation

$$R[n] = \beta[n]\mathbf{x}[n] + \epsilon[n] \tag{1.161}$$

If $\Sigma_{\mathbf{w}} = 0$, then we have the normal linear regression model.

- AR(P) model: Consider the AR model

$$\left(1 - \sum_{k=1}^{P} a_k L^k\right) y[n] = \epsilon[n] \tag{1.162}$$

In this case, the state vector is $\mathbf{s}[n] = (y[n], y[n-1], ..., y[n-P+1])^T$ and the transition equation

$$\mathbf{s}[n+1] = \begin{pmatrix} a_1 & a_2 & ... & a_P \\ 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix} \mathbf{s}[n] + \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{pmatrix} \mathbf{w}[n] \tag{1.163}$$

and the measurement equation

$$y[n] = \begin{pmatrix} 1 & 0 & ... & 0 \end{pmatrix} \mathbf{s}[n] + \begin{pmatrix} 0 & 0 & ... & 0 \end{pmatrix} \mathbf{u}[n] \tag{1.164}$$

- MA(Q) model: Consider the MA model

$$y[n] = \left(\sum_{k=0}^{P} b_k L^k\right) \epsilon[n] \tag{1.165}$$

In this case, the state vector is $\mathbf{s}[n] = (\epsilon[n-1], \epsilon[n-2], ..., \epsilon[n-Q])^T$ and the transition equation

$$\mathbf{s}[n+1] = \begin{pmatrix} 0 & 0 & ... & 0 \\ 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix} \mathbf{s}[n] + \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{pmatrix} \mathbf{w}[n] \quad (1.166)$$

and the measurement equation

$$y[n] = \begin{pmatrix} b_0 & b_1 & ... & b_Q \end{pmatrix} \mathbf{s}[n] + \begin{pmatrix} 0 & 0 & ... & 0 \end{pmatrix} \mathbf{u}[n] \quad (1.167)$$

Kalman filter

The Kalman filter allows the computations of the parameters and states of the linear state-space model given the previous observations. There are two basic steps:

1. Predict state: Predict (*a priori*) the current state and its covariance.

$$\begin{array}{rcl} \hat{\mathbf{s}}_0[n] & = & T[n]\hat{\mathbf{s}}[n-1] \\ \hat{\Sigma}_{\mathbf{s},0}[n] & = & T[n]\hat{\Sigma}_{\mathbf{s}}[n-1]T^T[n] + \Sigma_{\mathbf{w}} \end{array} \quad (1.168)$$

2. Predict observation:

$$\begin{array}{rcl} \hat{\mathbf{x}}[n] & = & H[n]\hat{\mathbf{s}}_0[n] \\ \hat{\mathbf{u}}[n] & = & \mathbf{x}[n] - \hat{\mathbf{x}}[n] \\ \hat{\Sigma}_{\mathbf{x}}[n] & = & H[n]\hat{\Sigma}_{\mathbf{s},0}[n]H^T[n] + \Sigma_{\mathbf{u}} \end{array} \quad (1.169)$$

3. Update state: Compute the Kalman gain matrix

$$G[n] = \hat{\Sigma}_{\mathbf{s},0}[n]H^T[n]\hat{\Sigma}_{\mathbf{x}}^{-1}[n] \quad (1.170)$$

and update the state using the residual information

$$\begin{array}{rcl} \hat{\mathbf{s}}[n] & = & \hat{\mathbf{s}}_0[n] + G[n]\hat{\mathbf{u}}[n] \\ \hat{\Sigma}_{\mathbf{s}}[n] & = & (I - G[n]H[n])\hat{\Sigma}_{\mathbf{x}}[n] \end{array} \quad (1.171)$$

The model parameter estimation $(T, H, \Sigma_{\mathbf{w}}, \Sigma_{\mathbf{u}})$ are estimated through Maximum Likelihood of the observations.

Extended Kalman filter

The Extended Kalman filter (EKF) can be applied to non-linear transition and measurement models

$$\begin{array}{rcl} \mathbf{s}[n+1] & = & \mathbf{t}(\mathbf{s}[n]) + R[n]\mathbf{w}[n] \\ \mathbf{x}[n] & = & \mathbf{h}(\mathbf{s}[n]) + \mathbf{u}[n] \end{array} \quad (1.172)$$

The trick is to linearize these equations by using the Kalman filter with the Jacobian matrices of the transformations $\mathbf{t}$ and $\mathbf{h}$ evaluated at the current state

$$\begin{array}{c} T[n] = \frac{d\mathbf{t}}{d\mathbf{s}}(\mathbf{s}[n]) \\ H[n] = \frac{d\mathbf{h}}{d\mathbf{s}}(\mathbf{s}[n]) \end{array} \quad (1.173)$$

However, we may still use $\mathbf{t}$ and $\mathbf{h}$ at those places where it is easy

1. Predict state: Predict (*a priori*) the current state and its covariance.

$$\begin{aligned} \hat{\mathbf{s}}_0[n] &= \mathbf{t}(\hat{\mathbf{s}}[n-1]) \Longleftarrow \\ \hat{\Sigma}_{\mathbf{s},0}[n] &= T[n]\hat{\Sigma}_{\mathbf{s}}[n-1]T^T[n] + \Sigma_{\mathbf{w}} \end{aligned} \tag{1.174}$$

2. Predict observation:

$$\begin{aligned} \hat{\mathbf{x}}[n] &= \mathbf{h}(\hat{\mathbf{s}}_0[n]) \Longleftarrow \\ \hat{\mathbf{u}}[n] &= \mathbf{x}[n] - \hat{\mathbf{x}}[n] \\ \hat{\Sigma}_{\mathbf{x}}[n] &= H[n]\hat{\Sigma}_{\mathbf{s},0}[n]H^T[n] + \Sigma_{\mathbf{u}} \end{aligned} \tag{1.175}$$

3. Update state: Compute the Kalman gain matrix

$$G[n] = \hat{\Sigma}_{\mathbf{s},0}[n]H^T[n]\hat{\Sigma}_{\mathbf{x}}^{-1}[n] \tag{1.176}$$

and update the state using the residual information

$$\begin{aligned} \hat{\mathbf{s}}[n] &= \hat{\mathbf{s}}_0[n] + G[n]\hat{\mathbf{u}}[n] \\ \hat{\Sigma}_{\mathbf{s}}[n] &= (I - G[n]H[n])\hat{\Sigma}_{\mathbf{x}}[n] \end{aligned} \tag{1.177}$$

Unscented Kalman filter

For non-linear dynamics, instead of linearizing as we have done in the EKF (previous section), we may adopt a different strategy: let us assume that we want to find the best Gaussian approximation to the state distribution. The state distribution is given by the output of a non-linear function $\mathbf{t}$. For doing so, we may get points (they are called sigma points) from a Gaussian distribution, sufficiently separated so as to represent well the Gaussian with very few points, transform them through $\mathbf{t}$ (this is called the unscented transform), and fit a Gaussian distribution to the output points. We need to find $2d+1$ weights w_i and sigma points $\mathbf{s}_i$, being d the number of components of $\mathbf{s}$, such that if $\mathbf{t}$ is the identity transformation, we recover the covariance matrix at the input space

$$\begin{gathered} \sum_i w_i = 1 \\ \sum_i w_i(\mathbf{s}_i - \overline{\mathbf{s}})(\mathbf{s}_i - \overline{\mathbf{s}})^T = \Sigma_{\mathbf{s}} \end{gathered} \tag{1.178}$$

There is no unique solution to this problem. One possible way is to choose

$$\begin{aligned} \mathbf{s}_0 &= \boldsymbol{\mu}_{\mathbf{s}} \\ \mathbf{s}_i &= \boldsymbol{\mu}_{\mathbf{s}} + \mathrm{Col}_i\left\{\sqrt{(d+\lambda)\Sigma_{\mathbf{s}}}\right\} \quad i = 1, ..., d \\ \mathbf{s}_i &= \boldsymbol{\mu}_{\mathbf{s}} - \mathrm{Col}_i\left\{\sqrt{(d+\lambda)\Sigma_{\mathbf{s}}}\right\} \quad i = d+1, ..., 2d \end{aligned} \tag{1.179}$$

where λ is a scaling parameter and $\mathrm{Col}_i\{\cdot\}$ is an operator that extracts the i-th column of a matrix. We will use two different sets of weights: one for computing the mean at the output space, and another one for calculating the covariance. Basically, both are of the form

$$\begin{aligned} w_0 &= \tfrac{\lambda}{\lambda+d} \\ w_i &= \tfrac{1}{2(\lambda+d)} \quad i = 1, ..., 2d \end{aligned} \tag{1.180}$$

For computing the covariance, the weights are as shown above except

$$w_0 = \frac{\lambda}{\lambda + d} + (1 - \alpha^2 + \beta) \tag{1.181}$$

where α and β are two parameters. $\beta = 2$ is optimal for Gaussians. α must be in the interval $(0, 1]$. Finally, it is suggested to use $\lambda = \alpha^2(\kappa + d) - d$, where $\kappa \geq 0$ controls how far the sigma points are from the mean.

We now change the Extended Kalman Filter. The first thing is to change the state prediction by the values proposed by the unscented transform

1. Predict state: Predict (*a priori*) the current state and its covariance. The sigma points have to be generated according to the current estimate of $\hat{\mathbf{s}}[n-1]$ and $\Sigma_{\mathbf{s}}[n-1]$

$$\begin{aligned} \hat{\mathbf{s}}_0[n] &= \sum_{i=0}^{2n} w_i \mathbf{t}(\mathbf{s}_i[n-1]) \Longleftarrow \\ \hat{\Sigma}_{\mathbf{s},0}[n] &= \sum_{i=0}^{2n} w_i (\mathbf{t}(\mathbf{s}_i) - \hat{\mathbf{s}}_0[n])(\mathbf{t}(\mathbf{s}_i) - \hat{\mathbf{s}}_0[n])^T + \Sigma_{\mathbf{w}} \Longleftarrow \end{aligned} \tag{1.182}$$

2. Predict observation: We now generate new sigma points using $\hat{\mathbf{s}}_0[n]$ and $\hat{\Sigma}_{\mathbf{s},0}[n]$ and estimate the value in the space of observations

$$\begin{aligned} \hat{\mathbf{x}}[n] &= \sum_{i=0}^{2n} w_i \mathbf{h}(\mathbf{s}_i[n-1]) \Longleftarrow \\ \hat{\mathbf{u}}[n] &= \mathbf{x}[n] - \hat{\mathbf{x}}[n] \\ \hat{\Sigma}_{\mathbf{x}}[n] &= \sum_{i=0}^{2n} w_i (\mathbf{h}(\mathbf{s}_i) - \hat{\mathbf{x}}[n])(\mathbf{h}(\mathbf{s}_i) - \hat{\mathbf{x}}[n])^T + \Sigma_{\mathbf{u}} \Longleftarrow \end{aligned} \tag{1.183}$$

3. Update state: Finally, with the sigma points of the step above, we compute the Kalman gain matrix

$$G[n] = \left(\sum_{i=0}^{2n} w_i (\mathbf{s}_i - \hat{\mathbf{s}}_0[n])(\mathbf{h}(\mathbf{s}_i) - \hat{\mathbf{x}}[n])^T \right) \hat{\Sigma}_{\mathbf{x}}^{-1}[n] \Longleftarrow \tag{1.184}$$

and update the state using the residual information

$$\begin{aligned} \hat{\mathbf{s}}[n] &= \hat{\mathbf{s}}_0[n] + G[n]\hat{\mathbf{u}}[n] \\ \hat{\Sigma}_{\mathbf{s}}[n] &= (I - G[n]H[n])\hat{\Sigma}_{\mathbf{x}}[n] \end{aligned} \tag{1.185}$$

Fig. 1.11 shows the difference between the Unscented Kalman Filter and the Extended Kalman Filter.

Extended Information Filter

An alternative representation of the multivariate Gaussian $N(\boldsymbol{\mu}, \Sigma)$ is given by the canonical representation of the distribution characterized by the information matrix

$$\Omega = \Sigma^{-1} \tag{1.186}$$

and the information vector

$$\boldsymbol{\xi} = \Sigma^{-1}\boldsymbol{\mu} \tag{1.187}$$

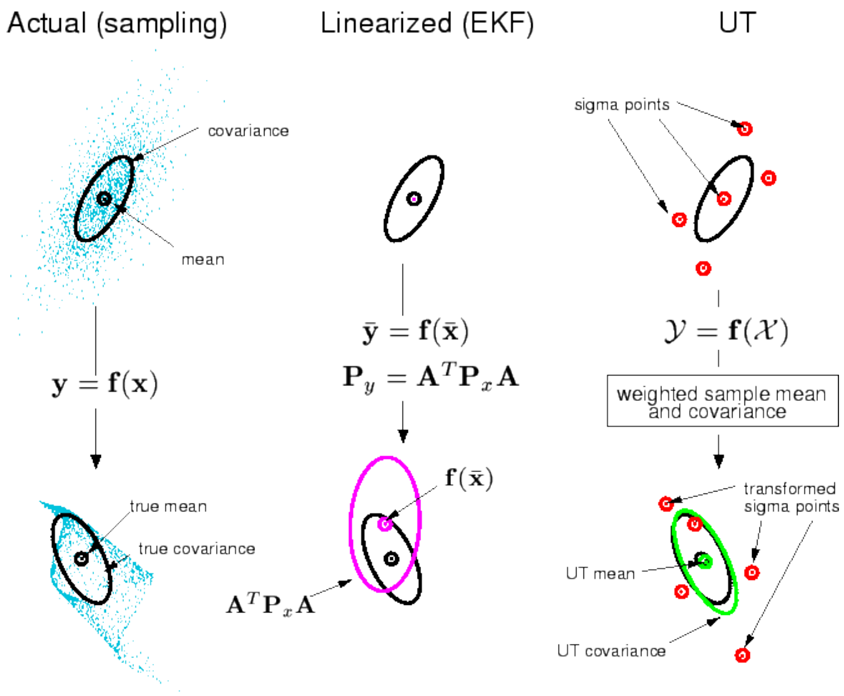


Figure 1.11: Comparison of UKF and EKF.

The probability density function can be expressed:

$$
\begin{aligned}
f(\mathbf{x}) &= \det(2\pi\Sigma)^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \\
&= \frac{\exp\left(-\frac{1}{2}\boldsymbol{\xi}^T\Omega^{-1}\boldsymbol{\xi}\right)}{\det(2\pi\Omega)^{1/2}} \exp\left(-\tfrac{1}{2}\mathbf{x}^T\Omega\mathbf{x} + \mathbf{x}^T\boldsymbol{\xi}\right)
\end{aligned}
\tag{1.188}
$$

The advantage of this new representation is that conditioning is easier (see Fig. 1.12) in the canonical expression while marginalization is easier in the momemtum expression.

The Kalman filter in the information form is called the Information Filter. Now the steps are

1. Predict state:

$$
\begin{aligned}
\hat{\Omega}_{\mathbf{s},0}[n] &= (T[n]\hat{\Omega}_{\mathbf{s}}^{-1}[n-1]T^T[n] + \Sigma_{\mathbf{w}})^{-1} \\
\hat{\boldsymbol{\xi}}_0[n] &= \hat{\Omega}_{\mathbf{s},0}[n]T[n]\hat{\Omega}_{\mathbf{s}}^{-1}[n-1]\boldsymbol{\xi}[n-1]
\end{aligned}
\tag{1.189}
$$

2. Update state:

$$
\begin{aligned}
\hat{\boldsymbol{\xi}}[n] &= \hat{\boldsymbol{\xi}}_0[n] + H^T[n]\Sigma_{\mathbf{u}}^{-1}\mathbf{x}[n] \\
\hat{\Omega}_{\mathbf{s}}[n] &= \hat{\Omega}_{\mathbf{s},0}[n] + H^T[n]\Sigma_{\mathbf{u}}^{-1}H[n]
\end{aligned}
\tag{1.190}
$$

There is no explicit calculation of the Kalman gain matrix or the observation prediction. The KF makes an efficient prediction and a slow correction, while the IF is just the opposite.

We may also extend the Information Filter to non-linear functions

$$p(\boldsymbol{\alpha},\boldsymbol{\beta}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_\alpha\\ \boldsymbol{\mu}_\beta\end{bmatrix}, \begin{bmatrix}\Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta}\\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta}\end{bmatrix}\right) = \mathcal{N}^{-1}\left(\begin{bmatrix}\boldsymbol{\eta}_\alpha\\ \boldsymbol{\eta}_\beta\end{bmatrix}, \begin{bmatrix}\Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta}\\ \Lambda_{\beta\alpha} & \Lambda_{\beta\beta}\end{bmatrix}\right)$$

	MARGINALIZATION $p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha},\boldsymbol{\beta})d\boldsymbol{\beta}$	CONDITIONING $p(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = p(\boldsymbol{\alpha},\boldsymbol{\beta})/p(\boldsymbol{\beta})$
COV. FORM	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$ $\Sigma = \Sigma_{\alpha\alpha}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$
INFO. FORM	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

Figure 1.12: Conditioning and marginalization of a multivariate Gaussian.

1. Predict state:

$$\begin{aligned}\hat{\Omega}_{\mathbf{s},0}[n] &= (T[n]\hat{\Omega}_{\mathbf{s}}^{-1}[n-1]T^T[n] + \Sigma_{\mathbf{w}})^{-1}\\ \hat{\mathbf{x}}_0[n] &= \hat{\Omega}_{\mathbf{s}}^{-1}[n-1]\boldsymbol{\xi}[n-1]\\ \hat{\boldsymbol{\xi}}_0[n] &= \hat{\Omega}_{\mathbf{s},0}[n]\mathbf{t}(\hat{\mathbf{x}}_0[n])\end{aligned} \quad (1.191)$$

2. Update state:

$$\begin{aligned}\hat{\boldsymbol{\xi}}[n] &= \hat{\boldsymbol{\xi}}_0[n] + H^T[n]\Sigma_{\mathbf{u}}^{-1}(\mathbf{x}[n] - (\mathbf{h}(\hat{\mathbf{x}}_0[n]) - H[n]\hat{\mathbf{x}}_0[n]))\\ \hat{\Omega}_{\mathbf{s}}[n] &= \hat{\Omega}_{\mathbf{s},0}[n] + H^T[n]\Sigma_{\mathbf{u}}^{-1}H[n]\end{aligned} \quad (1.192)$$

The EIF has been reported to be more numerically stable than the EKF in some situations, although their expressivity is the same (they can address the same problems). However, IF allows for Sparse IF which may be very powerful.

Particle Filters

The idea is to be capable of dealing with any arbitrary distribution of states. This is done by sampling the current estimate of the state distribution with N_p particles. We have a set of samples and weights. There are many flavours of particle filters (Chen, 2003). The Sequential Importance Sampling has the following steps:

1. Sample from the proposal:

$$\mathbf{s}_{j,0}[n] \sim p(\mathbf{s}[n]|\mathbf{s}_{1,2,\ldots,N_p}[n-1]) \quad (1.193)$$

   That is we use the previous estimates of the states. If we have a transition function $\mathbf{t}(\mathbf{s})$, we may take randomly a state j of iteration $n-1$, $\mathbf{s}_j[n-1]$, take a random sample around a small kernel around it, $\tilde{\mathbf{s}}_j[n-1]$, and predict how the state is updated over time

$$\mathbf{s}_{j,0}[n] = \mathbf{f}(\tilde{\mathbf{s}}_j[n-1]) \quad (1.194)$$

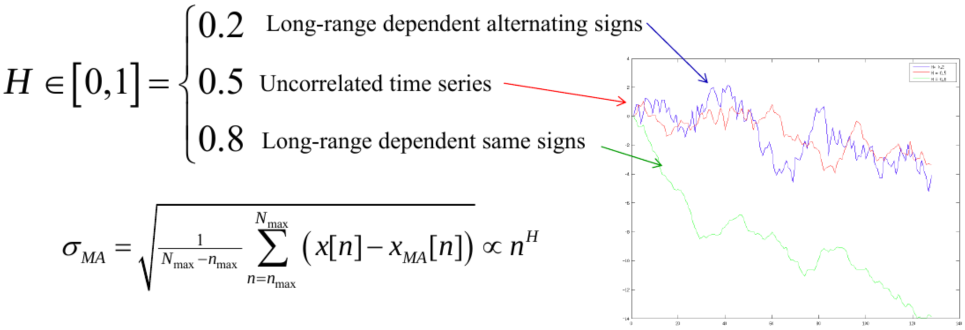


Figure 1.13: Example of Hurst exponents.

2. Compute the importance weights:

$$w_j[n] \sim p(\mathbf{x}[n]|\mathbf{s}_{j,0}[n]) \tag{1.195}$$

   That is, the weight of a particle is proportional to the probability of observing the current measured values if $\mathbf{s}_j[n]$ were the current state of the system. If we have a non-linear function that relates states to measurements we may estimate the likelihood of these observations

$$w_j[n] = f_{\mathbf{u}}(\mathbf{x}[n] - \mathbf{g}(\mathbf{s}_j[n])) \tag{1.196}$$

   The weigts so-computed must be normalized by the sum of weights so that their sum is 1.

3. Resampling: Draw with replacement N_p particles from the set $\mathbf{s}_{j,0}[n]$ according to the importance weights $w_j[n]$. This is a new set of particles that is reused for step $n+1$.

1.6 Self-similar signals

Let D be the fractal dimension of a signal. Then, its power spectral density falls off as

$$S_X(\omega) \propto \omega^{-\alpha} \tag{1.197}$$

where $\alpha = 5 - 2D$. The Hurst exponent is defined as $H = 2 - D$ (see Fig. 1.13).

Recurrence plots are used to distinguish between stochastic and chaotic signals (see Fig. 1.14).

The correlation dimension is a simple way of checking the chaotic nature of a signal (see Fig. 1.15) and another way is through the Lyapunov exponent (see Fig. 1.16).

1.7 Empirical properties of asset returns

The following paragraphs are excerpts of Cont (2001).

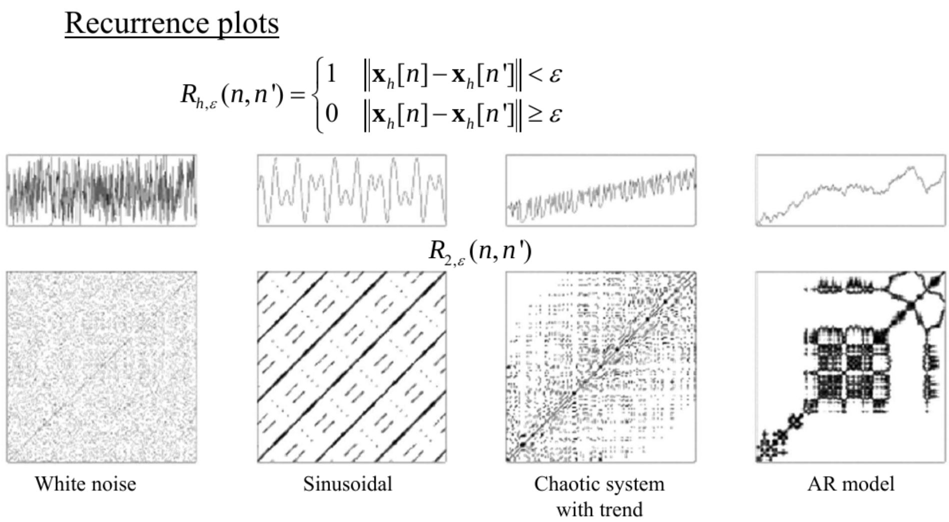


Figure 1.14: Example of recurrence plots.

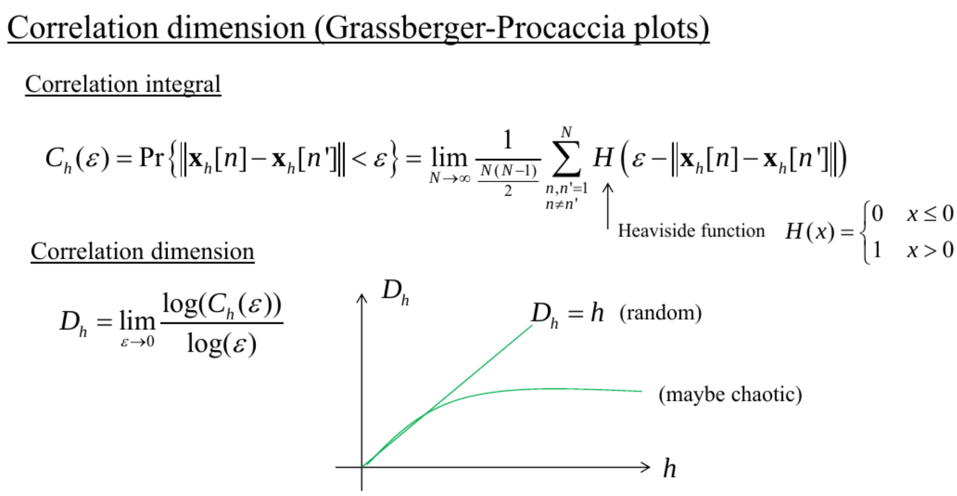


Figure 1.15: Definition of the correlation dimension.

Brock-Dechert-Scheinkman (BDS) Test

H_0: samples are iid $\quad V_{h,\varepsilon} = \dfrac{C_h(\varepsilon) - \left(C_1(\varepsilon)\right)^h}{\sqrt{N}} \sim N(0,1)$

Lyapunov exponent

Consider two time points such that $\quad \left\| \mathbf{x}_h[n] - \mathbf{x}_h[n'] \right\| = \delta_0 \ll 1$

Consider the distance m samples later $\quad \delta_m = \left\| \mathbf{x}_h[n+m] - \mathbf{x}_h[n'+m] \right\|$

The Lyapunov exponent relates these two distances $\quad \delta_m = \delta_0 e^{\lambda m}$

$\lambda > 0 \longrightarrow$ Histories diverge: chaos, cannot be predicted

$\lambda < 0 \longrightarrow$ Histories converge: can be predicted

Maximal Lyapunov exponent

$$\lambda = \lim_{m,\delta_0 \to \infty} \frac{1}{m} \log \frac{\delta_m}{\delta_0}$$

Figure 1.16: Definition of the Lyapunov exponent.

1.7.1 Stylized properties

Let us start by stating a set of stylized statistical facts which are common to a wide set of financial assets.

- Absence of autocorrelations: (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales (20 minutes) for which microstructure effects come into play.
- Heavy tails: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.
- Gain/loss asymmetry: one observes large drawdowns in stock prices and stock index values but not equally large upward movements.
- Aggregational Gaussianity: as one increases the time scale t over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.
- Intermittency: returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility estimators.
- Volatility clustering: different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.

- Conditional heavy tails: even after correcting returns for volatility clustering (e.g. via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.

- Slow decay of autocorrelation in absolute returns: the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\beta \in [0.2, 0.4]$. This is sometimes interpreted as a sign of long-range dependence.

- Leverage effect: most measures of volatility of an asset are negatively correlated with the returns of that asset.

- Volume/volatility correlation: trading volume is correlated with all measures of volatility.

- Asymmetry in time scales: coarse-grained measures of volatility predict fine-scale volatility better than the other way round.

1.7.2 Distributional properties

Distribution family

From the empirical features described above, one can conclude that, in order for a parametric model to successfully reproduce all the above properties of the marginal distributions it must have at least four parameters: a location parameter, a scale (volatility) parameter, a parameter describing the decay of the tails and eventually an asymmetry parameter allowing the left and right tails to have different behaviours. For example, normal inverse Gaussian distributions, generalized hyperbolic distributions and exponentially truncated stable distributions meet these requirements (see Fig. 1.17). The choice among these classes is then a matter of analytical and numerical tractability.

Weighing the tail and extreme values

The tail index k of a distribution may be defined as the order of the highest absolute moment which is finite. The higher the tail index, the thinner the tail; for a Gaussian or exponential tail, $k = \infty$ (all moments are finite), while for a power-law distribution with exponent α, the tail index is equal to α. [...] Measuring the tail index of a distribution gives a measure of how heavy the tail is.

A simple method, suggested by Mandelbrot, is to represent the sample moments (or cumulants) as a function of the sample size n. If the theoretical moment is finite then the sample moment will eventually settle down to a region defined around its theoretical limit and fluctuate around that value. In the case where the true value is infinite the sample moment will either diverge as a function of sample size or exhibit erratic behaviour and large fluctuations. Applying this method to time series of cotton prices, Mandelbrot conjectured that the theoretical variance may be infinite since the sample variance did not converge to a particular value as the sample size increased and continued to fluctuate incessantly.

Normal-inverse Gaussian (NIG)

Parameters	μ location (real) α tail heaviness (real) β asymmetry parameter (real) δ scale parameter (real) $\gamma = \sqrt{\alpha^2 - \beta^2}$
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}\, e^{\delta\gamma+\beta(x-\mu)}$ K_j denotes a modified Bessel function of the third kind[1]
Mean	$\mu + \delta\beta/\gamma$
Variance	$\delta\alpha^2/\gamma^3$
Skewness	$3\beta/(\alpha\sqrt{\delta\gamma})$
Ex. kurtosis	$3(1+4\beta^2/\alpha^2)/(\delta\gamma)$
MGF	$e^{\mu z+\delta(\gamma-\sqrt{\alpha^2-(\beta+z)^2})}$
CF	$e^{i\mu z+\delta(\gamma-\sqrt{\alpha^2-(\beta+iz)^2})}$

Generalised hyperbolic

Parameters	λ (real) α (real) β asymmetry parameter (real) δ scale parameter (real) μ location (real) $\gamma = \sqrt{\alpha^2 - \beta^2}$
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)}\, e^{\beta(x-\mu)} \times \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\sqrt{\delta^2+(x-\mu)^2}/\alpha\right)^{1/2-\lambda}}$
Mean	$\mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)}$
Variance	$\frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \frac{K_{\lambda+1}^2(\delta\gamma)}{K_\lambda^2(\delta\gamma)}\right)$
MGF	$\frac{e^{\mu z}\gamma^\lambda}{(\sqrt{\alpha^2-(\beta+z)^2})^\lambda}\frac{K_\lambda(\delta\sqrt{\alpha^2-(\beta+z)^2})}{K_\lambda(\delta\gamma)}$

Figure 1.17: Distributions with 4 parameters, capable of representing stock returns.

Measuring the tails is important in order to estimate the *Value at risk*, that is define as a high quantile of the loss distribution of a portfolio over a certain time horizon. The Fisher-Tippet theorem relates the distribution of the log-returns to some limit distributions of extreme values.

Fisher-Tippet theorem: Extreme value theorem for IID sequence. Assume the log returns R_t form an IID sequence with distribution F. Let us denote as M_N the maximum return over a period of N samples, If there exist normalizing constants $(\lambda_N\ ,\ \sigma_N)$ and a non-degenerate limit distribution H for the normalized maximum return:

$$\Pr\left\{\frac{M_N - \lambda_N}{\sigma_N} \le x\right\} \underset{x\to\infty}{\longrightarrow} H(x) \tag{1.198}$$

then the limit distribution H is either a Gumbel, Weibull or Fréchet distribution.

Gumbel	$H(x) = \exp(-e^{-x})$
Fréchet	$H(x) = \exp(-x^{-\alpha})\ 1_{x>0}$
Weibull	$H(x) = \exp(-(-x)^{-\alpha})\ 1_{x\leqslant 0} + 1_{x>0}$

The three distributions can be unified in the Cramer-von Mises distribution

$$H_\xi(x) = \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right) \tag{1.199}$$

For $\xi > 0$ we get a Fréchet distribution, $\xi = 0$ Gumbel, $\xi < 0$ Weibull. When applied to daily return stocks, market indices and exchange rates we observe $0.2 < \xi < 0.4$ which means a tail index $2 < \alpha < 5$. These studies seem to validate the power-law nature of log returns with an exponent around 3.

Dependence properties

The following properties are observed in stock markets:

1. Absence of linear autocorrelation. This property implies that traditional tools of signal processing which are based on second-order properties, in the time domain—autocovariance analysis, ARMA modelling—or in the spectral domain— Fourier analysis, linear filtering—cannot distinguish between asset returns and white noise. This points out the need for non-linear measures of dependence in order to characterize the dependence properties of asset returns.

2. Volatility clustering and nonlinear dependence. Returns are not independent (if they were, any non-linear function of the returns would not have any autocorrelation). However, this is not true and simple non-linear functions like absolute value or squared returns exhibit significant positive autocorrelation. This is a quantitative signature of *Volatility clustering*: large price variations are more likely to be followed by large price variations. GARCH models (see Section 2.4) try to capture this dependence. It has been shown that the α-correlation defined as

$$C_\alpha[k] = \mathrm{Corr}\{|R[n]|^\alpha, |R[n+k]|^\alpha\} \tag{1.200}$$

decreases as

$$C_\alpha[k] \approx Ak^{-\beta} \tag{1.201}$$

with $\beta \in [0.2, 0.4]$ for absolute ($\alpha = 1$) or squared ($\alpha = 2$) returns. The slow decay is sometimes interpreted as a signe of long-range dependence (see Fig. 1.18). Recent work on multifractal stochastic volatility models have shown that

$$C_0[k] = \mathrm{Corr}\{\log |R[n]|, \log |R[n+k]|\} \approx A \log \frac{B}{\Delta t + k} \tag{1.202}$$

where Δt is the sampling rate (time difference between two consecutive return samples). Another measure of nonlinear dependence is the *Leverage effect*, i.e., a positive correlation

$$L[k] = \mathrm{Corr}\{R[n], |R[n+k]|^2\} \tag{1.203}$$

However, this effect is assymetric $L[k] \neq L[-k]$ and normally, $L[k] \approx 0$ if $k < 0$. This correlation motivates the decomposition of the returns as

$$R[n] = \sigma[n]\epsilon[n] \tag{1.204}$$

where $\epsilon[n]$ is a white noise signal (see Sections 1.2 and 2.4). A word of caution should be added to all these observed correlations. The fact that returns are heavy tailed implies that estimating the correlation values is more difficult and wider confidence intervals are expected. In other words, the estimates of the correlations are not so reliable.

3. Dependence among assets. The most interesting features of the matrix Σ_R are its eigenvalues λ_i and its eigenvectors $\mathbf{e}_i$, which have been usually interpreted in economic terms as factors of randomness underlying market

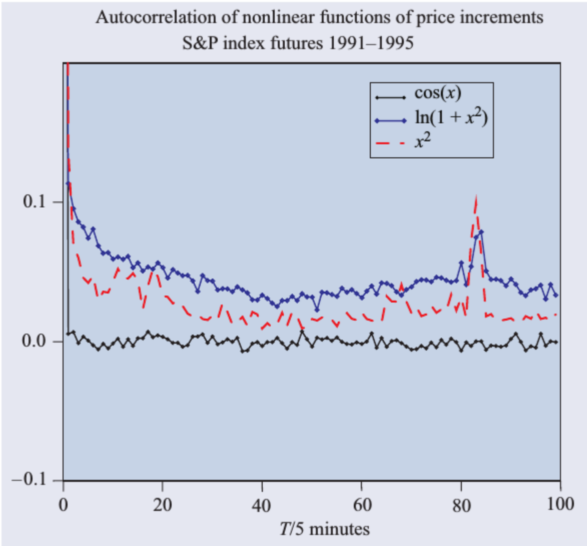


Figure 1.18: Autocorrelation of non-linear function of returns.

movements. In a recent empirical study of the covariance matrix of 406 NYSE assets, Laloux et al showed that among the 406 available eigenvalues and principal components, apart from the highest eigenvalue (whose eigenvector roughly corresponds to the market index) and the next few (ten) highest eigenvalues, the other eigenvectors and eigenvalues do not seem to contain any information: in fact, their marginal distribution closely resembles the spectral distribution of a positive symmetric matrix with random entries whose distribution is the "most random possible"—i.e., entropy maximizing. These results strongly question the validity of the use of the sample covariance matrix as an input for portfolio optimization, as suggested by classical methods such as mean-variance optimization, and support the rationale behind factor models such as the CAPM and APT, where the correlations between a large number of assets are represented through a small number of factors.

4. Dependence among extreme values. A relevant quantity is the conditional probability of a large (negative) return in one stock given a large negative movement in another stock. It is important to remark that two assets may have extremal correlations while their covariance is zero: covariance does not measure the correlation of extremes. Some recent theoretical work has been done in this direction using copulas and multivariate extreme value theory, but a lot remains to be done on empirical grounds.

1.7.3 Pathwise properties

Hölder regularity

In mathematical terms, the regularity of a function may be characterized by its local Hölder exponents. A function f is h-Hölder continuous at point t_0 iff there exists a polynomial of degree $< h$ such that

$$|f(t) - P(t - t_0)| \leq K_{t_0}|t - t_0|^h \tag{1.205}$$

in a neighborhood of t_0, where K_{t_0} is a constant. Let $C^h(t_0)$ be the space of (real-valued) functions which verify the above property at t_0. A function f is said to have local *H ölder exponent* α if for $h < \alpha$, $f \in C^h(t_0)$ and for $h > \alpha$, $f \notin C^h(t_0)$. Let $h_f(t)$ denote the local Hölder exponent of f at point t. If $h_f(t_0) \geq 1$, then f is differentiable at point t_0, whereas a discontinuity of f at t_0 implies $h_f(t_0) = 0$. More generally, the higher the value of $h_f(t_0)$, the greater the local regularity of f at t_0.

In the case of a sample path $X_t(\omega)$ of a stochastic process X_t, $h_{X(\omega)(t)} = h_\omega(t)$ depends on the particular sample path considered, i.e., on ω. There are however some famous exceptions: for example for fractional Brownian motion with self-similarity parameter H, $h_{B(t)} = \frac{1}{H}$ almost everywhere with probability one, i.e., for almost all sample paths. Note however that no such results hold for sample paths of Lévy processes or even stable Lévy motion.

Singularity spectrum

Given that the local Hölder exponent may vary from sample path to sample path in the case of a stochastic process, it is not a robust statistical tool for characterizing signal roughness: the notion of a singularity spectrum of a signal was introduced to give a less detailed but more stable characterization of the local smoothness structure of a function in a “statistical” sense.

Let f be a real-valued function and for each $\alpha > 0$ define the set of points at which f has local Hölder exponent h:

$$\Omega(\alpha) = \{t, h_f(t) = \alpha\} \tag{1.206}$$

The *singularity spectrum* of f is the function $D : \mathbb{R}^+ \to \mathbb{R}$ which associates to each $\alpha > 0$ the Hausdorff–Besicovich dimension of $\Omega(\alpha)$ (a measure of the fractal dimension):

$$D(\alpha) = dim_{HB}(\alpha) \tag{1.207}$$

Using the above definition, one may associate to each sample path $X_t(\omega)$ of a stochastic process X_t its singularity spectrum $d_\omega(\alpha)$. If d_ω is “strongly dependent” on ω then the empirical estimation of the singularity spectrum is not likely to give much information about the properties of the process X_t. Fortunately, this turns out not to be the case: it has been shown that, for large classes of stochastic processes, the singularity spectrum is the same for almost all sample paths.

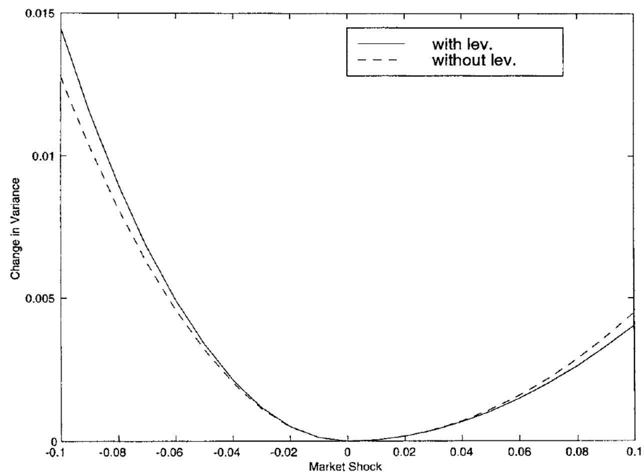


Figure 1.19: Dependence between market shock (regression residual) and the change in variance.

The USD/DEM high-frequency exchange rate data using the structure function method. The spectra have a support ranging from 0.3 to 0.9 (with some variations depending on the data set chosen) with a maximum centred around 0.55–0.6. Note that 0.55–0.6 is the range of values of the "Hurst exponent" reported in many studies of financial time series using the R/S or similar techniques.

1.7.4 Assymetric volatility

We may predict the evolution of an asset return as a CAPM-like regression, in which the market return, a risk-free return and some residual are involved. The residual is called the "market shock". It is known that residuals have an impact on the asset volatility (Bekaert and Wu, 2000) (see Fig. 1.19). An important contribution of this article is that it distinguishes between the market shock on the reference index (all the market is behaving better or worse) and the market shock on a specific stock (the effect of each one is different). It also couples CAPM and GARCH models (the residual in a CAPM-like regression can be modelled as a GARCH process).

Let us see the explanatory dynamics proposed by the article: *We begin by considering news (shocks) at the market level. Bad news at the market level has two effects. First, whereas news is evidence of higher current volatility in the market, investors also likely revise the conditional variance since volatility is persistent. According to the CAPM, this increased conditional volatility at the market level has to be compensated by a higher expected return, leading to an immediate decline in the current value of the market. The price decline will not cease until the expected return is sufficiently high. Hence a negative return shock may generate a significant increase in conditional volatility. Second, the marketwide price decline leads to higher leverage at the market level and*

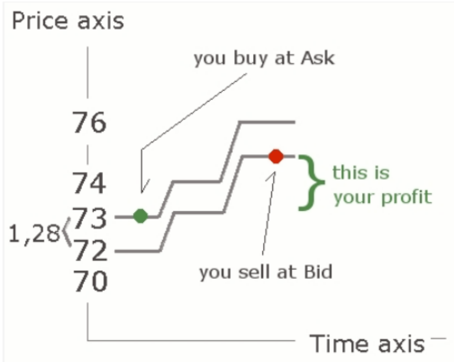


Figure 1.20: Difference between bid and ask prices.

hence higher stock volatility. That is, the leverage effect reinforces the volatility feedback effect.

When good news arrives in the market, there are again two effects. First, news brings about higher current period market volatility and an upward revision of the conditional volatility. When volatility increases, prices decline to induce higher expected returns, offsetting the initial price movement. The volatility feedback effect dampens the original volatility response. Second, the resulting market rally (positive return shock) reduces leverage and decreases conditional volatility at the market level. Hence the net impact on stock return volatility is not clear.

For the initial impact of news at the firm level, the reasoning remains largely the same: bad and good news generate opposing leverage effects which reinforce (offset) the volatility embedded in the bad (good) news event. What is different is the volatility feedback. A necessary condition for volatility feedback to be observed at the firm level is that the covariance of the firm's return increases in response to market shocks. If the shock is completely idiosyncratic, the covariance between the market return and individual firm return should not change, and no change in the required risk premium occurs. Hence idiosyncratic shocks generate volatility asymmetry purely through a leverage effect. Volatility feedback at the firm level occurs when marketwide shocks increase the covariance of the firm's return with the market. Such covariance behavior would be implied by a CAPM model with constant (positive) firm betas and seems generally plausible. The impact on the conditional covariance is likely to be different across firms. For firms with high systematic risk, marketwide shocks may significantly increase their conditional covariance with the market. The resulting higher required return then leads to a volatility feedback effect on the conditional volatility, which would be absent or weaker for firms less sensitive to market level shocks.

1.8 Modern portfolio theory

In a stock market you buy at *ask* price and sell at *bid* price. The difference is the return R (see Fig. 1.20).

When we want to buy or sell an asset, we may face a liquidity risk (this

Figure 1.21: Liquidity risk.

is an inability to easily enter or exit a position; see Fig. 1.21). (From the Investopedia) *There are at least three perspectives on market liquidity. The most popular and crudest measure is the bid-ask spread; this is also called width. A low or narrow bid-ask spread is said to be "tight" and tends to reflect a more liquid market. Depth refers to the ability of the market to absorb the sale (exit) of a position. An individual investor who sells shares of Google, for example, is not likely to impact the share price; on the other hand, an institutional investor selling a large block of shares in a small capitalization company will probably cause the price to fall. Finally, resiliency refers to the market's ability to bounce back from temporarily incorrect prices. To summarize:*

- *The bid-ask spread measures liquidity in the price dimension and it is a feature of the market not the seller (or the seller's position). Financial models that incorporate bid-ask spread adjust for exogenous liquidity and are exogenous liquidity models.*
- *Position size, relative to the market, is a feature of the seller. Models that use this are measuring liquidity in the quantity dimension and are generally known as endogenous liquidity models.*
- *Resiliency measures liquidity in the time dimensions and such models are currently rare.*

At one extreme, high market liquidity would be characterized by an owner of a small position relative to a deep market who is exiting into a tight bid-ask spread and a highly resilient market.

The way of including the bid-ask spread is normally by subtracting half its size from the return, or adding it to the Value-at-Risk. The idea is to consider the worse scenario.

Let R_i $(i = 1, 2, ..., N)$ be random variables representing the return of N assets. A portfolio is defined by the weight of each one of these assets, $\mathbf{p}$, so that the return of the portfolio is defined by another random variable

$$R_{\mathbf{p}} = \sum_{i=1}^{N} p_i R_i = \mathbf{p}^T \mathbf{R} \tag{1.208}$$

where $\mathbf{R}$ is a vector with all the R_i variables. The expected return of the portfolio can be calculated as

$$\mathrm{E}\{R_{\mathbf{p}}\} \triangleq \mu_{\mathbf{p}} = \mathrm{E}\{\mathbf{p}^T \mathbf{R}\} = \mathbf{p}^T \mathrm{E}\{\mathbf{R}\} = \mathbf{p}^T \boldsymbol{\mu}_{\mathbf{R}} \tag{1.209}$$

and its variance

$$\begin{aligned} \mathrm{Var}\{R_{\mathbf{p}}\} \triangleq \sigma_{\mathbf{p}}^2 &= \mathrm{E}\{(R_p - \mu_{\mathbf{p}})^2\} \\ &= \mathbf{p}^T \Sigma_{\mathbf{R}} \mathbf{p} \end{aligned} \tag{1.210}$$

where $\Sigma_{\mathbf{R}}$ is the covariance matrix of the return vector. We may use the correlation matrix $\Gamma_{\mathbf{R}}$ by exploiting the relationship between the covariance and correlation matrices

$$\Gamma_{\mathbf{R}} = \Lambda^{-\frac{1}{2}} \Sigma_{\mathbf{R}} \Lambda^{-\frac{1}{2}} \Leftrightarrow \Sigma_{\mathbf{R}} = \Lambda^{\frac{1}{2}} \Gamma_{\mathbf{R}} \Lambda^{\frac{1}{2}} \tag{1.211}$$

where $\Lambda^{\frac{1}{2}}$ is a diagonal matrix formed by all the standard deviations of the R_i variables, that is, $\Lambda = \mathrm{diag}\{\Sigma\}$. Using the correlation matrix, the variance of the portfolio return becomes

$$\sigma_{\mathbf{p}}^2 = \mathbf{p}^T \Lambda^{\frac{1}{2}} \Gamma_{\mathbf{R}} \Lambda^{\frac{1}{2}} \mathbf{p} \tag{1.212}$$

It is this variance of the portfolio what is normally called the *risk* of the portfolio.

Note that this formulation is completely general and that it allows including a cash position (*e.g.*, if it is variable 1, then $\mu_1 = 0$, $\sigma_1 = 0$, $\rho_{1j} = \rho_{j1} = 0$ for $j \neq 1$) as well as a fixed income asset (*e.g.*, if it is variable 1, then $\mu_1 = r_0$, $\sigma_1 = 0$, $\rho_{1j} = \rho_{j1} = 0$ for $j \neq 1$).

The portfolio management problem amounts to maximizing the expected portfolio return while keeping the risk within limits

$$\begin{aligned} \mathbf{p}^* = \ & \mathrm{argmax} \quad \mathbf{p}^T \boldsymbol{\mu}_{\mathbf{R}} \\ & \text{s.t.} \quad \mathbf{p}^T \Lambda^{\frac{1}{2}} \Gamma_{\mathbf{R}} \Lambda^{\frac{1}{2}} \mathbf{p} \leq \sigma_{max}^2 \end{aligned} \tag{1.213}$$

where σ_{max}^2 is the maximum accepted risk. If we cannot borrow money to do the investment we have to add the constraints

$$\begin{aligned} \mathbf{p}^T \mathbf{1} &= 1 \\ -\mathbf{1} \leq \mathbf{p} &\leq \mathbf{1} \end{aligned} \tag{1.214}$$

that is, all the proportions must add up to 1. If we can only go long, that is, short positions cannot be taken, then we also need to add the constraint

$$\mathbf{p} \geq \mathbf{0} \tag{1.215}$$

We need now some optimization procedures to solve the problems above.

Fig. 1.22 shows the trajectories in the $\mu_{\mathbf{p}}$,$\sigma_{\mathbf{p}}$ plane of different portfolios in which the proportion of Assets 1 and 2 vary. Note the important dependence of the trajectory on the correlation, c in the figure, between the two assets. Fig. 1.23 shows how as the portfolio is diversified, the portfolio risk approaches the market risk.

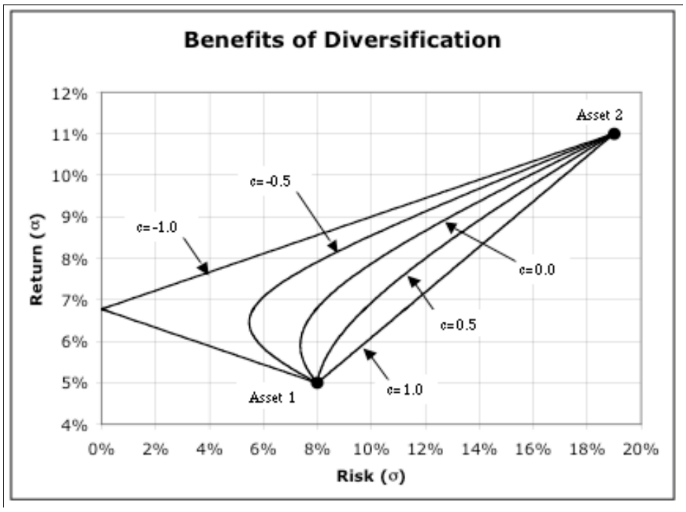


Figure 1.22: Trajectories as a function of the linear combination of two assets, depending on the correlation between both assets.

An interesting feature of diversification is that it provides reasonable returns in a volatile environment. Assume there are two assets A and B. In the first period, A earns 100% and B loses 50%. In the second period, A loses 50% and B earns 100%. On their own, each asset over the two periods stays at the same place (you earn 0%). However, a rebalanced portfolio (50-50%) would earn 25% on the first period (1.25=0.5*2+0.5*0.5), then the portfolio is rebalanced, and on the second period it earns another 25% (1.25=0.5*0.5+0.5*2).

Using leverage takes us away from the efficient frontier as shown in Fig. 1.24. The Risk Parity strategy looks at maximizing the Sharpe Ratio, which for a two asset portfolio it is the slope of the tangent to the efficient frontier when one of the assets is cash or any other known investment. The use of leverage allows us to move along this tangent.

Passive portfolio management uses some simple rule to define **p** like replicating the S&P 500 proportions. Empirically it has been shown that these simple rules may outerform 50% of active portfolios (Elton et al., 2003)[Chap. 26](*Market efficiency is one of the majer paradigms of financial economics. Modern theories of efficiency argue that informed investors in an efficient market will earn just enough to compensate for the cost of obtaining the information. Mutual fund managers are commonly viewed as the prototype of informed investors. [...] (However), mutual fund managers underperfom passive portfolios. Furthermore, funds with higher fees and turnover underperform those with lower fees and turnover.* Elton et al. (1993)). Although simple, some decisions may be taken like skipping some of the small capitalization stocks in order to save transaction costs, or choosing a small subset of assets that mimick the behaviour of the reference index.

This article does much to explain short-term persistence in equity mutual fund returns with common factors in stock returns and investment costs. Buying last year's top-decile mutual funds and selling last year's bottom-decile funds

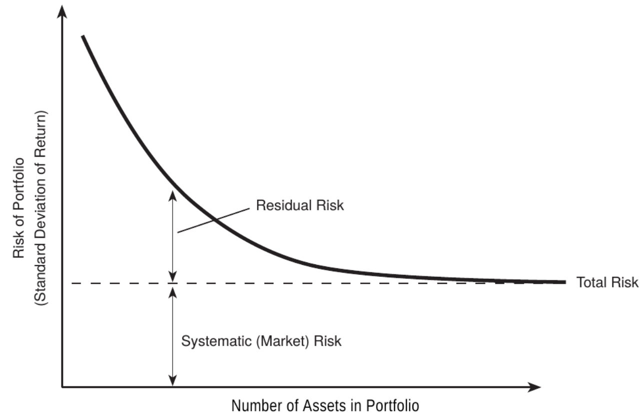


Figure 1.23: Balance between residual risk and market risk as the portfolio is diversified (Barra, 2007)[Fig.2.1].

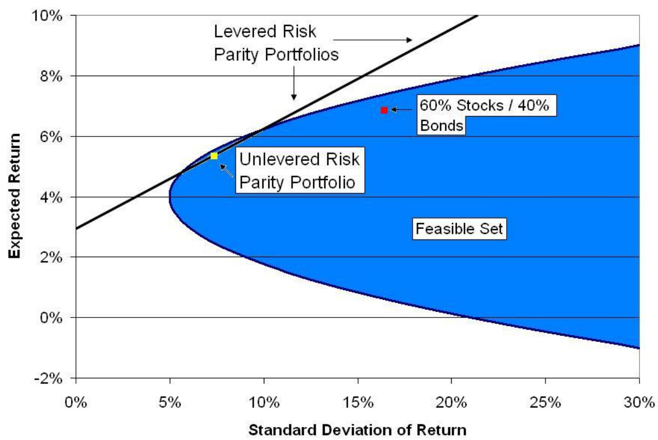


Figure 1.24: Risk parity and levered portfolios.

yields a return of 8 percent per year. Of this spread, differences in the market value and momentum of stocks held explain 4.6 percent, differences in expense ratios explain. 0.7 percent, and differences in transaction costs explain 1 percent. Sorting mutual funds on longer horizons of past returns yields smaller spreads in mean returns, all but about 1 percent of which are attrib- utable to common factors, expense ratios, and transaction costs. Further, the spread in mean return unexplained by common factors and investment costs is concentrated in strong underperformance by the bottom decile relative to the remaining sample. Of the spread in annual return remaining after the 4-factor model, expense ratios, and transaction costs, approximately two-thirds is attributable to the spread between the ninth- and tenth-decile portfolios.

I also find that expense ratios, portfolio turnover, and load fees are significantly and negatively related to performance. Expense ratios appear to reduce performance a little more than one-for-one, Turnover reduces performance about 95 basis points for every buy-and-sell transaction. Differences in costs per transaction account for some of the spread in the best- and worst-performing mutual funds. Surprisingly, load funds substantially underperform no-load funds. After controlling for the correlation between expenses and loads, and removing the worst-performing quintile of funds, the average load fund underperforms the average no-load fund by approximately 80 basis points per year. (Carhart, 1997)

This raises another question. Why do we see any money remain in funds that are predicted to do poorly and in fact do perform poorly? I propose a possible explanation for this phenomenon: The existence of two clienteles, a sophisticated clientele and a second clientele that I will refer to as a disadvantaged clientele.

The sophisticated clientele directs its money to funds based on performance. The disadvantaged clientele consists of three groups:

1. *Unsophisticated investors—a group that directs its money to funds based at least in part on other influences such as advertising and advice from brokers*

2. *Institutionally disadvantaged investors—a group primarily represented by pension accounts that are restricted by the plan they are part of to a set of funds that underperforms the best active funds.*

3. *Tax disadvantaged investors—a group that has held one or more funds for enough time so that capital gains taxes make it inefficient to remove money from these funds. This group can still act as sophisticated investors in placing new money*

All of the evidence in this article is consistent with this hypothesis.

1. *The stock of money underperforms appropriate benchmarks. The stock of money is likely to contain a large percentage of the funds invested by the disadvantaged clientele.*

2. *The flow of money performs better than appropriate benchmarks. Sophisticated investors are likely to constitute a larger percentage of new cash flows into and out of mutual funds. The investor who moves cash into and out of funds earns a positive risk adjusted return and gets the services provided by mutual funds at no net cost.*

3. *The flow of new money into the best performing funds is much larger than the flow of money out of the poorer performing funds. Tax disadvantaged investors and to some extent institutionally disadvantaged investors will not or cannot move money out of bad funds, but tax disadvantaged investors and to some extent institutionally disadvantaged investors can place new money in good performing funds.*

4. *The flow of new money into and out of mutual funds underperforms naive rules for selecting mutual funds as examined in Table II. Again I explain this as due to the presence of both a disadvantaged clientele and a sophisticated clientele. The sophisticated investor earns more than the average positive return on marginal cash flows; the disadvantaged investor earns less.*

5. *Because sophisticated investors can't short sell funds, they cannot eliminate inefficient funds. However, by disinvesting (or not investing) in these funds they eliminate the worst performing funds in the sample over time. This is the reason that 28 percent of the poorer performing funds in my sample merged or changed to noncommon stock funds over the ten year sample period.*

(Gruber, 1996)

Some authors defend the persistent good and bad performance of funds and relate it to its managers (Brown and Goetzmann, 1995), that is, there are funds that are consistently at the top of funds, and funds consistently at the bottom (and eventually disappearing). Hedge funds seem to go ahead of mutual funds, although they do not outperform market indices (Ackermann et al., 1999).

1.8.1 Normal returns

Some authors like Markowitz and Sharpe consider yearly returns to be normally distributed. Although yearly returns are better modelled by a log-normal distribution, using a Gaussian distribution makes some derivations easier.

For a given time horizon, suppose that the return on an investment I is normally distributed with mean R_I and standard deviation σ_I. Let U be the negative exponential utility function with risk aversion coefficient A ($A > 0$), $U_A(w) = -e^{-Aw}$. Let w_0 be the investor's wealth at the beginning of the period and $w(t)$ the wealth at the end of the period t. Then, (Norstad, 1999a)

$$\mathrm{E}\{U_A(w(t))\} = -e^{-Aw_0(1+R_I-\frac{1}{2}Aw_0\sigma_I^2)} \tag{1.216}$$

Consequently, when this investor tries to choose amongst different investment choices, he will select the one maximizing

$$R_I - \frac{1}{2}Aw_0\sigma_I^2 \tag{1.217}$$

As w_0 increases, risk-aversion increases. Some people apply the utility function to the return only (not the wealth), in that case

$$\mathrm{E}\{U_A(r_I)\} = -e^{-A(R_I-\frac{1}{2}A\sigma_I^2)} \tag{1.218}$$

and the investment chosen is the one maximizing

$$R_I - \frac{1}{2}A\sigma_I^2 \tag{1.219}$$

The choice is now independent of the current wealth. This is similar, although not identical, to the iso-elastic utility function.

Fig. 1.25 shows the feasible set of investments and its efficient frontier. It also shows the isoutility curves for two different risk aversion coefficients ($A = 3$, less aversion; and $A = 10$, more aversion). The feasible set and the efficient frontier is the same for all investors. What changes from investor to investor is the utility curve. The optimal investment is the point at which the rightmost isocurve is tangential to the efficient frontier.

1.8.2 Log-normal returns

Let the investment I be log-normally distributed under the random walk model with mean continuously yearly return μ_I and standard devation σ_I^2. Let U be the isoelastic utility function

$$U_A(w) = \frac{w^{1-A} - 1}{1 - A} \tag{1.220}$$

Then, the expected utility at the end of a period of t years is (Norstad, 1999a)

$$\mathrm{E}\{U_A(w(t))\} = \frac{1}{1-A}\left(w_0^{1-A}e^{-(A-1)t(\mu - \frac{1}{2}(A-1)\sigma_I^2)} - 1\right) \tag{1.221}$$

When $A = 1$, the corresponding expected value is

$$\mathrm{E}\{U_1(w(t))\} = \log(w_0) + \mu t \tag{1.222}$$

The investor will choose the investment that maximizes

$$\mu - \frac{1}{2}(A-1)\sigma_I^2 = \left(\mu + \frac{1}{2}\sigma_I^2\right) - \frac{1}{2}A\sigma_I^2 \tag{1.223}$$

The term $\mu + \frac{1}{2}\sigma_I^2$ is the arithmetic mean of the yearly returns, and is normally denoted by α_i. The investment I_1 is more efficient that I_2 iff $\alpha_1 \geq \alpha_2$ and $\sigma_1 \leq \sigma_2$ (Norstad, 1999a).

We may carry over the differential equation formulation of the value of a specific asset (see Eq. 1.26)

$$\log\frac{X(t)}{X(0)} = \mu t + \sigma S(t) \tag{1.224}$$

to a portfolio by simply considering the portfolio weights

$$\log\frac{W(t)}{W(0)} = \mu_{\mathbf{p}} t + \sigma_{\mathbf{p}} S(t) \tag{1.225}$$

where

$$\begin{aligned} \mu_{\mathbf{p}} &= \mathbf{p}^T \boldsymbol{\mu}_{\mathbf{R}} \\ \sigma_{\mathbf{p}} &= \mathbf{p}^T \Sigma_{\mathbf{R}} \mathbf{p} \end{aligned} \tag{1.226}$$

In this model we are assuming that we are continuously rebalancing the portfolio (selling and buying assets) so that the proportion of investment in each one of the assets is kept constant after the end of each dt period. As a consequence, the portfolio value, $W(t)$, is also a log-normal random walk.

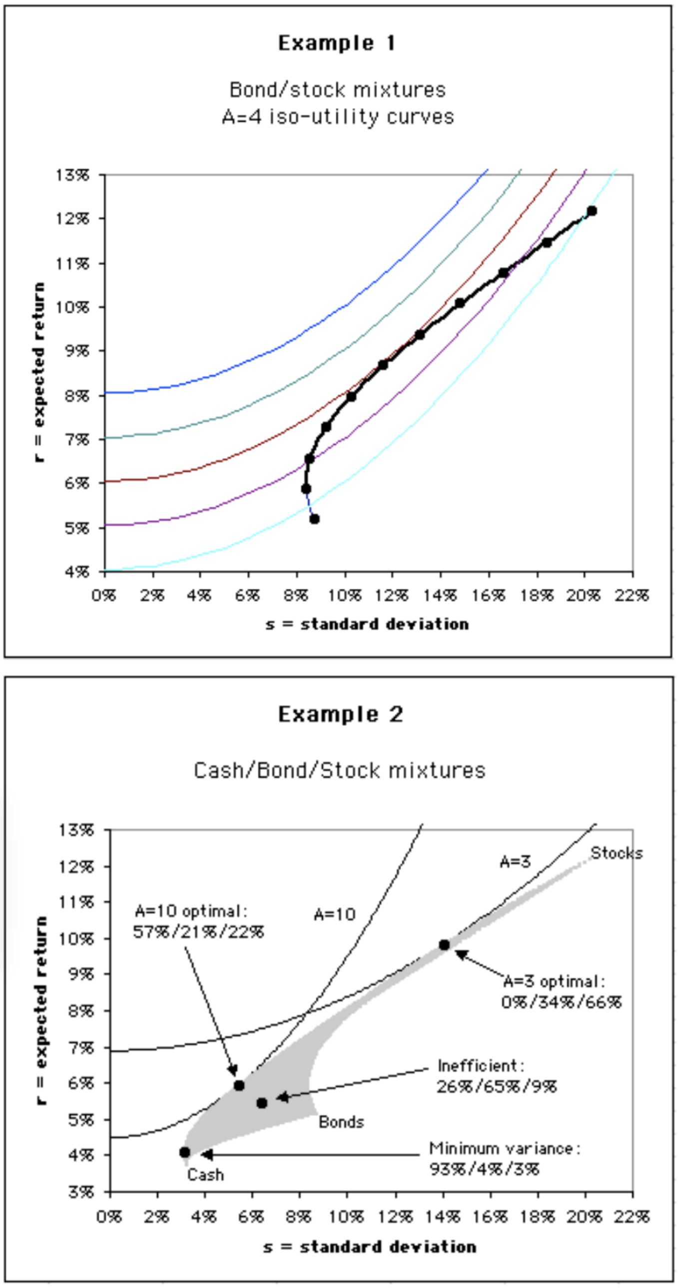


Figure 1.25: Feasible investments in US bonds, stocks (top) and cash (bottom), and its effcient frontier (the leftmost border).

1.8.3 Construction of portfolios

Minimum variance portfolios

Minimum variance portfolios , are those that minimize the portfolio return variance Elton et al. (2003):

$$\begin{aligned}\mathbf{p}^* &= \text{argmin} \quad \mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \\ &\quad\ \text{s.t.} \qquad\ \mathbf{p}^T\mathbf{1} = 1\end{aligned} \tag{1.227}$$

The Lagrangean multipliers method requires finding solutions of the equation system

$$\begin{aligned}2\Sigma_{\mathbf{R}}\mathbf{p}^* + \lambda\mathbf{1} &= \mathbf{0} \\ (\mathbf{p}^*)^T\mathbf{1} &= 1\end{aligned} \tag{1.228}$$

Its solution can be easily calculated through the matrix equation system

$$\begin{pmatrix}2\Sigma_{\mathbf{R}} & \mathbf{1} \\ \mathbf{1}^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{p}^* \\ \lambda\end{pmatrix} = \begin{pmatrix}\mathbf{0} \\ 1\end{pmatrix} \tag{1.229}$$

Maximum utility portfolios

Under the iso-elastic utility function, the Maximum utility portfolios, are those that maximize (Eq. 1.223, Norstad (2002a)):

$$\mathbf{p}^* = \text{argmax} \quad \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} - \tfrac{1}{2}(A-1)\mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \tag{1.230}$$

Calculating the derivative and equating to $\mathbf{0}$, the solution of this problem is given by the linear equation

$$\boldsymbol{\mu}_{\mathbf{R}} - (A-1)\Sigma_{\mathbf{R}}\mathbf{p}^* = \mathbf{0} \Rightarrow \mathbf{p}^* = \frac{1}{A-1}\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}} \tag{1.231}$$

Note that this portfolio is not constrained so that short are allowed ($p_i < 0$) and leverage is allowed ($\mathbf{p}^T\mathbf{1} \neq 1$). The solution above is for $A \neq 1$. For $A = 1$ (risk indifferent investor), the investor maximizes (see Eq. 1.222)

$$\mathbf{p}^* = \text{argmax} \quad \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} = \mathbf{e}_j \tag{1.232}$$

where $j = \text{argmax}\boldsymbol{\mu}_{\mathbf{R}}$, that is, the investor puts all his money of the asset with maximum average return. Actually, the solution $\mathbf{e}_j$ assumes that the investor is not borrowing money.

We may obviously add budget constraints to the problem

$$\begin{aligned}\mathbf{p}^* &= \text{argmax} \quad \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} - \tfrac{1}{2}(A-1)\mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \\ &\quad\ \text{s.t.} \qquad\ \mathbf{p}^T\mathbf{1} = 1\end{aligned} \tag{1.233}$$

The Lagrangean of this problem yields the equation system

$$\begin{pmatrix}(A-1)\Sigma_{\mathbf{R}} & \mathbf{1} \\ \mathbf{1}^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{p}^* \\ \lambda\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}_{\mathbf{R}} \\ 1\end{pmatrix} \tag{1.234}$$

For every risk-aversion coefficient A, we may calculate the optimal portfolio allocation, $\mathbf{p}^*$, and from this we calculate $\mu_{\mathbf{p}^*}$ and $\sigma_{\mathbf{p}^*}$ producing a point in the return-risk plane. The *efficient frontier* is the curve resulting of gathering all these points for all possible $A \in [0, \infty)$. Let us do it for the simple case of the unconstrained problem (see Eq. 1.231).

$$\left.\begin{array}{rcl} \mu_{\mathbf{p}^*} & = & \frac{1}{A-1}\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}} \\ \sigma_{\mathbf{p}^*}^2 & = & \left(\frac{1}{A-1}\right)^2\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}} \end{array}\right\} \Rightarrow \mu_{\mathbf{p}^*} = (A-1)\sigma_{\mathbf{p}^*}^2 \tag{1.235}$$

Although this dependence is only valid in the case of unconstrained portfolios for the iso-elastic utility function, it highlights several important properties of the efficient frontier for any other case:

- The return is an increasing function of the risk. Once the portfolio has been optimally diversified for a given A, the only way of getting more return is by taking more risk.
- The infinite aversion to risk $A \to \infty$ goes to the minimum variance portfolio, and the slope at it is infinite (the tangent is vertical).
- The efficient frontier is a concave function.

In the next section we develop a more involved example

Efficient portfolios

Efficient portfolios are those that maximize the portfolio return while keeping the risk at a fixed value Markowitz (1959),:

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmax} & \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} \\ & & \text{s.t.} & \mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} = \sigma_0^2 \\ & & & \mathbf{p}^T\mathbf{1} = 1 \end{array} \tag{1.236}$$

The Langrangean multipliers method for this problem implies finding solutions for the equation system:

$$\begin{array}{rcl} \boldsymbol{\mu}_{\mathbf{R}} + 2\lambda_1\Sigma_{\mathbf{R}}\mathbf{p}^* + \lambda_2\mathbf{1} & = & \mathbf{0} \\ (\mathbf{p}^*)^T\Sigma_{\mathbf{R}}\mathbf{p}^* & = & \sigma_0^2 \\ (\mathbf{p}^*)^T\mathbf{1} & = & 1 \end{array} \tag{1.237}$$

where the unknows are $\mathbf{p}^*$, λ_1 and λ_2. Although feasible, this equation system is non-linear, and a related but slightly different problem is normally preferred

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmin} & \mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \\ & & \text{s.t.} & \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} = r_0 \\ & & & \mathbf{p}^T\mathbf{1} = 1 \end{array} \tag{1.238}$$

The Lagrangean multipliers method requires finding solutions of the equation system

$$\begin{array}{rcl} 2\Sigma_{\mathbf{R}}\mathbf{p}^* + \lambda_1\boldsymbol{\mu}_{\mathbf{R}} + \lambda_2\mathbf{1} & = & \mathbf{0} \\ (\mathbf{p}^*)^T\boldsymbol{\mu}_{\mathbf{R}} & = & r_0 \\ (\mathbf{p}^*)^T\mathbf{1} & = & 1 \end{array} \tag{1.239}$$

Now, the equation system is linear again and its solution can be easily calculated through the matrix equation system

$$\begin{pmatrix} 2\Sigma_{\mathbf{R}} & \boldsymbol{\mu}_{\mathbf{R}} & \mathbf{1} \\ \boldsymbol{\mu}_{\mathbf{R}}^T & 0 & 0 \\ \mathbf{1}^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p}^* \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ r_0 \\ 1 \end{pmatrix} \tag{1.240}$$

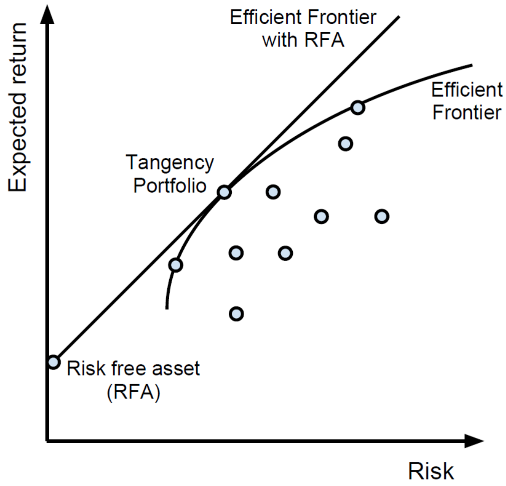


Figure 1.26: Efficient frontier is an hyperbola

The solution of this problem is

$$\begin{array}{rcl}\mathbf{p}^* & = & \frac{1}{\Delta}\left((\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}})\Sigma_{\mathbf{R}}^{-1}\mathbf{1}-(\mathbf{1}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}})\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}}\right)\\ & & +\frac{r_0}{\Delta}\left((\mathbf{1}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{1})\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}}-(\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{1})\Sigma_{\mathbf{R}}^{-1}\mathbf{1}\right)\end{array} \quad (1.241)$$

with

$$\Delta=(\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}})(\mathbf{1}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{1})-(\boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{1})^2>0 \quad (1.242)$$

Now, we will look for the efficient frontier of this problem. Let us define some auxiliary variables to simplify notation

$$\begin{array}{rcl}a & = & \mathbf{1}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{1}\\ b & = & \mathbf{1}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}}\\ c & = & \boldsymbol{\mu}_{\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}}\end{array} \quad (1.243)$$

If we substitute the optimal portfolio, $\mathbf{p}^*$ into the portfolio variance, $\sigma^2_{\mathbf{p}^*}$, we would have

$$\sigma^2_{\mathbf{p}^*}=\frac{1}{a}+\frac{a}{\Delta}\left(r_0-\frac{b}{a}\right)^2=\frac{1}{\Delta}\begin{pmatrix}1 & r_0\end{pmatrix}\begin{pmatrix}c & -b\\ -b & a\end{pmatrix}\begin{pmatrix}1\\ r_0\end{pmatrix} \quad (1.244)$$

Remind that $\mu_{\mathbf{p}^*}=r_0$. This is the equation of a hyperbola in the risk-return plane (see Fig. 1.26). As in our previous example, the efficient frontier is concave, the risk increases with increasing return, and the tangent is vertical at the minimum variance portfolio.

We can also pose an utility version of this problem

$$\begin{array}{rcll}\mathbf{p}^* & = & \text{argmax} & \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}}-\frac{1}{2}A\mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p}\\ & & \text{s.t.} & \mathbf{p}^T\mathbf{1}=1\end{array} \quad (1.245)$$

The solution of this problem can be efficiently performed through the critical line algorithm (Norstad, 2002b).

Mutual fund separation problem

Note that the optimal portfolio above can be written in the form

$$\mathbf{p}^* = \mathbf{p}_A^* + r_0 \mathbf{p}_B^* \tag{1.246}$$

Consider two mutual funds, each one with target returns r_1 and r_2. The optimal portfolios of each one of these mutual funds will be

$$\begin{array}{l} \mathbf{p}_1^* = \mathbf{p}_A^* + r_1 \mathbf{p}_B^* \\ \mathbf{p}_2^* = \mathbf{p}_A^* + r_2 \mathbf{p}_B^* \end{array} \tag{1.247}$$

The mutual fund separation problem states that we may achieve the target portfolio return r_0 by a linear combination of investments in the two funds 1 and 2 (which may save a substantial amount of trading costs)

$$\mathbf{p}^* = \frac{r_0 - r_2}{r_1 - r_2}\mathbf{p}_1^* + \frac{r_1 - r_0}{r_1 - r_2}\mathbf{p}_2^* \tag{1.248}$$

Risk free assets

The formulation above is not valid for risk-free or cash assets, because it involves inverting a covariance matrix with a zero column and row. The problem is then formulated slightly differently to account for the lack of variance of one of the assets. Let us denote as r_f the return of the risk-free asset. The rest of variables $(\mathbf{p}, \boldsymbol{\mu}_\mathbf{R}, \Sigma_\mathbf{R})$ refer to the risky assets.

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmin} & \mathbf{p}^T \Sigma_\mathbf{R} \mathbf{p} \\ & & \text{s.t.} & (1 - \mathbf{p}^T \mathbf{1}) r_f + \mathbf{p}^T \boldsymbol{\mu}_\mathbf{R} = r_0 \end{array} \tag{1.249}$$

The solution is

$$\mathbf{p}^* = \frac{r_0 - r_f}{(\boldsymbol{\mu}_\mathbf{R} - r_f \mathbf{1})^T \Sigma_\mathbf{R}^{-1} (\boldsymbol{\mu}_\mathbf{R} - r_f \mathbf{1})} \Sigma_\mathbf{R}^{-1} (\boldsymbol{\mu}_\mathbf{R} - r_f \mathbf{1}) \tag{1.250}$$

If $r_0 < r_f$, then $\mathbf{p}^* = \mathbf{0}$.

No leverage

If we cannot borrow money for our operations, then the problem above must be rewritten as

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmin} & \mathbf{p}^T \Sigma_\mathbf{R} \mathbf{p} \\ & & \text{s.t.} & \mathbf{p}^T \boldsymbol{\mu}_\mathbf{R} = r_0 \\ & & & \mathbf{p}^T \mathbf{1} = 1 \\ & & & -\mathbf{1} \leq \mathbf{p} \leq \mathbf{1} \end{array} \tag{1.251}$$

or equivalently

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmin} & \mathbf{p}^T \Sigma_\mathbf{R} \mathbf{p} \\ & & \text{s.t.} & \mathbf{p}^T \boldsymbol{\mu}_\mathbf{R} = r_0 \\ & & & \mathbf{p}^T \mathbf{1} = 1 \\ & & & \mathbf{1} - \mathbf{p} \geq \mathbf{0} \\ & & & \mathbf{1} + \mathbf{p} \geq \mathbf{0} \end{array} \tag{1.252}$$

The Karush-Kuhn-Tucker conditions for this problem are

$$\begin{array}{rcll} 2\Sigma_{\mathbf{R}}\mathbf{p}^* + \lambda_1\boldsymbol{\mu}_{\mathbf{R}} + \lambda_2\mathbf{1} - \boldsymbol{\mu}_1\mathbf{1} + \boldsymbol{\mu}_2\mathbf{1} & = & \mathbf{0} & \\ (\mathbf{p}^*)^T\boldsymbol{\mu}_{\mathbf{R}} & = & r_0 & \\ (\mathbf{p}^*)^T\mathbf{1} & = & 1 & \\ \mu_{1i}(1-p_i^*) & = & 0 & \forall i \\ \mu_{2i}(1+p_i^*) & = & 0 & \forall i \\ \mu_{1i} & \geq & 0 & \forall i \\ \mu_{2i} & \geq & 0 & \forall i \end{array} \tag{1.253}$$

This equation system is non-linear (due to the multiplications $\mu_{1i}p_i$ and $\mu_{2i}p_i$).

Empirical validation

DeMiguel et al. (2009) compares 14 portfolio construction models with 7 different datasets. In their words the conclusions are: *From the above discussion, we conclude that of the strategies from the op- timizing models, there is no single strategy that always dominates the 1/N strategy in terms of Sharpe ratio. In general, the 1/N strategy has Sharpe ratios that are higher (or statistically indistinguishable) relative to the constrained policies, which, in turn, have Sharpe rat ios that are higher than those for the unconstrained policies. In terms of CEQ, no strategy from the optimal models is consistently better than the benchmark 1/N strategy. And in terms of turnover, only the "vw" strategy, in which the investor holds the market portfolio and does not trade at all, is better than the 1/N strategy.*

The models are listed in Fig. 1.27. Most of them respond to the optimization

$$\text{argmax} \quad \boldsymbol{\mu}_{\mathbf{R}} - \frac{1}{2}A\mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \tag{1.254}$$

whose solution is

$$\mathbf{p}^* = \frac{\Sigma_{\mathbf{R}}^{-1}\boldsymbol{\mu}_{\mathbf{R}}}{\mathbf{1}^T\Sigma_{\mathbf{R}}\boldsymbol{\mu}_{\mathbf{R}}} \tag{1.255}$$

although they differ in the way they estimate $\boldsymbol{\mu}_{\mathbf{R}}$ and $\Sigma_{\mathbf{R}}$. The classical approach is the one presented in these notes in which the mean and covariance matrices are estimated from historical data. Bayesian estimation presumes an *a priori* distribution of the covariance matrix and refines the historical estimate with this prior. The moment restricted portfolios are those constrained to add up to 1. The metrics used to measure the performance are: the Sharpe-ratio

$$SR = \frac{\mu_{\mathbf{p}}}{\sigma_{\mathbf{p}}}, \tag{1.256}$$

the certainty-equivalent

$$CEQ = \mu_{\mathbf{p}} - \frac{1}{2}A\sigma_{\mathbf{p}}, \tag{1.257}$$

and the turnover (how many trading transactions).

1.8.4 Log optimal portfolios

The following notes have been taken from (Györfi et al., 2012) with some adaptations in the notation. Let $X_0, X_1, ..., X_n$ denote the price of an asset at times

Table 1
List of various asset-allocation models considered

#	Model	Abbreviation
Naive		
0.	$1/N$ with rebalancing *(benchmark strategy)*	ew or $1/N$
Classical approach that ignores estimation error		
1.	Sample-based mean-variance	mv
Bayesian approach to estimation error		
2.	Bayesian diffuse-prior	Not reported
3.	Bayes-Stein	bs
4.	Bayesian Data-and-Model	dm
Moment restrictions		
5.	Minimum-variance	min
6.	Value-weighted market portfolio	vw
7.	MacKinlay and Pastor's (2000) missing-factor model	mp
Portfolio constraints		
8.	Sample-based mean-variance with shortsale constraints	mv-c
9.	Bayes-Stein with shortsale constraints	bs-c
10.	Minimum-variance with shortsale constraints	min-c
11.	Minimum-variance with generalized constraints	g-min-c
Optimal combinations of portfolios		
12.	Kan and Zhou's (2007) "three-fund" model	mv-min
13.	Mixture of minimum-variance and $1/N$	ew-min
14.	Garlappi, Uppal, and Wang's (2007) multi-prior model	Not reported

This table lists the various asset-allocation models we consider. The last column of the table gives the abbreviation used to refer to the strategy in the tables where we compare the performance of the optimal portfolio strategies to that of the $1/N$ strategy. The results for two strategies are not reported. The reason for not reporting the results for the Bayesian diffuse-prior strategy is that for an estimation period that is of the length that we are considering (60 or 120 months), the Bayesian diffuse-prior portfolio is very similar to the sample-based mean-variance portfolio. The reason for not reporting the results for the multi-prior robust portfolio described in Garlappi, Uppal, and Wang (2007) is that they show that the optimal robust portfolio is a weighted average of the mean-variance and minimum-variance portfolios, the results for both of which are already being reported.

Figure 1.27: List of models employed in the portfolio comparison

$0, 1, 2, ..., n$. The price is given by a random variable and in order to normalize let $X_0 = 1$. We may represent the normalized prices as a function of an exponential growth rate W_n as

$$X_n = e^{nW_n} \tag{1.258}$$

from where we can estimate an average exponential growth rate up to time n

$$W_n = \frac{1}{n}\log X_n \tag{1.259}$$

The asymptotic average exponential growth rate is

$$W = \lim_{n\to\infty}\frac{1}{n}\log X_n \tag{1.260}$$

Static portfolio selection

If we have a portfolio with proportions $\mathbf{p}$ (fixed over trading periods, that is why it is called static), and an initial capital S_0, then after 1 trading period, the new capital is

$$S_1 = S_0(\mathbf{p}^T(\mathbf{1} + \mathbf{r}_1)) \tag{1.261}$$

where $\mathbf{r}_1$ is the specific return observed at time 1. After two trading periods

$$S_2 = S_0(\mathbf{p}^T(\mathbf{1} + \mathbf{r}_1))(\mathbf{p}^T(\mathbf{1} + \mathbf{r}_2)) \tag{1.262}$$

The asymptotic growth rate is

$$\begin{aligned} W &= \lim_{n\to\infty}\tfrac{1}{n}\log\left(S_0\prod_{i=1}^{n}\mathbf{p}^T(\mathbf{1}+\mathbf{r}_i)\right) \\ &= \lim_{n\to\infty}\left(\tfrac{1}{n}\log S_0 + \tfrac{1}{n}\sum_{i=1}^{n}\log(\mathbf{p}^T(\mathbf{1}+\mathbf{r}_i))\right) \\ &= \lim_{n\to\infty}\left(\tfrac{1}{n}\sum_{i=1}^{n}\log(\mathbf{p}^T(\mathbf{1}+\mathbf{r}_i))\right) \\ &= \lim_{n\to\infty}\left(\tfrac{1}{n}\sum_{i=1}^{n}\log(1+\mathbf{p}^T\mathbf{r}_i)\right) \end{aligned} \tag{1.263}$$

If the market is memory-less (that is, each return is an independent and identically distributed variable, IID), then the principle of log-optimality that states that (Györfi et al., 2012)

$$\begin{aligned} S_n(\mathbf{p}) &\approx e^{n\mathrm{E}\{\log(1+\mathbf{p}^T\mathbf{r}_1)\}} \\ \mathrm{E}\{S_n(\mathbf{p})\} &= e^{n\log(1+\mathbf{p}^T\mathrm{E}\{\mathbf{r}_1\})} \end{aligned} \tag{1.264}$$

The *log-optimal portfolio* is given by

$$\mathbf{p}^* = \operatorname{argmax}\mathrm{E}\{\log(1+\mathbf{p}^T\mathbf{r}_1)\} \tag{1.265}$$

If the distribution of returns is unknown, we may calculate the log-optimal portfolio (Eq. 1.265) from the historical data

$$\mathbf{p}^* = \operatorname{argmax}\frac{1}{n}\sum_{i=1}^{n}\log(1+\mathbf{p}^T\mathbf{r}_i) \tag{1.266}$$

Note that

$$S_n(\mathbf{p}) \ll \mathrm{E}\{S_n(\mathbf{p})\} \tag{1.267}$$

If we try to maximize $\mathrm{E}\{S_n(\mathbf{p})\}$, this translates in the maximization of

$$\mathbf{p}^* = \operatorname{argmax} \mathbf{p}^T \mathrm{E}\{\mathbf{r}_1\} = \operatorname{argmax} \mathbf{p}^T \boldsymbol{\mu}_{\mathbf{R}} \tag{1.268}$$

The maximization problem above gives $\mathbf{p}^* \mathbf{e}_j$ where j is the index of the maximum value of $\boldsymbol{\mu}_{\mathbf{R}}$. Markowitz was the first in constraining the problem so that the solution diversifes the investment

$$\begin{array}{rcll} \mathbf{p}^* & = & \operatorname{argmax} & \mathbf{p}^T \boldsymbol{\mu}_{\mathbf{R}} \\ & & \text{s.t.} & \mathbf{p}^T \Sigma_{\mathbf{R}} \mathbf{p} \leq \sigma_0^2 \end{array} \tag{1.269}$$

The semi-log optimal portfolio approximates the logarithm by its second-order Taylor expansion ($\log(1+x) \approx x - \frac{1}{2}x^2$)

$$\begin{array}{rcl} \mathbf{p}^* & = & \operatorname{argmax} \mathrm{E}\left\{\log(1+\mathbf{p}^T\mathbf{r}_i)\right\} \\ & \approx & \operatorname{argmax} \mathrm{E}\left\{\mathbf{p}^T\mathbf{r}_i - \frac{1}{2}\left(\mathbf{p}^T\mathbf{r}_i\right)^2\right\} \\ & \approx & \operatorname{argmax} \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}} - \frac{1}{2}\mathbf{p}^T\Sigma_{\mathbf{R}}\mathbf{p} \end{array} \tag{1.270}$$

This equation is formally identical to the one of maximum-utility portfolios (Eq. 1.231) with $A = 2$ and the generic formula summarizing several investment models (Eq. 1.254).

We may add short-selling, no-ruin and leverage constraints to this optimization problem (Horváth and Urbán, 2012). Short-selling positions are represented by negative values in the portfolio vector. We will assume that short-selling is for free. The portfolio return after 1 period of investment is still

$$S_1 = S_0(1 + \mathbf{p}^T\mathbf{r}_1) \tag{1.271}$$

Note that the asset returns may be positive or negative, as well as the portfolio proportion, p_i. Let us assume that the maximum return in absolute value is bounded by a value B (typically $0.3 < B < 0.4$). The maximum loss that may occur is bounded by

$$|\mathbf{p}|^T|\mathbf{r}_1| \leq B|\mathbf{p}|^T\mathbf{1} \tag{1.272}$$

That is, we guarantee no-ruin if

$$B|\mathbf{p}|^T\mathbf{1} < 1 \tag{1.273}$$

This is one of the constraints to add to the optimization problem. This constrain makes the feasible set of solutions to be non-convex so that the problem has to be transformed (by unfolding the portfolio vector into a positive and negative component) in order to become a convex problem (Horváth and Urbán, 2012).

If we add leverage a leverage level, $L_B > 1$, then the portfolio is constrained to be

$$|\mathbf{p}|^T\mathbf{1} = L_B \tag{1.274}$$

Again, the set of feasible portfolios is non-convex, although it can be helped with the vector unfolding technique. No-ruin is guaranteed if

$$L_B < \frac{1}{B} \tag{1.275}$$

In case that one of the assets is cash, which may be used to lend money at an interest r, then L_B is modified to

$$L_{B,r} = \frac{1+r}{B+r} \tag{1.276}$$

Horváth and Urbán (2012) shows the performance of this portfolio construction in a dynamic portfolio selection (see next section) in which a nearest neighbour regression is employed. They report annual growths of 83%.

Dynamic portfolio selection

In practice, the capital invested depends on a portfolio, $\mathbf{p}_i$, that changes dynamically over time (note that to construct $\mathbf{p}_i$ we only use information up to $i-1$), so that

$$S_1 = S_0(\mathbf{p}_1^T(\mathbf{1}+\mathbf{r}_1)) \tag{1.277}$$

and

$$S_2 = S_0(\mathbf{p}_1^T(\mathbf{1}+\mathbf{r}_1))(\mathbf{p}_2^T(\mathbf{1}+\mathbf{r}_2)) \tag{1.278}$$

The asymptotic growth rate is

$$\begin{aligned} W &= \lim_{n\to\infty} \frac{1}{n} \log\left(S_0 \prod_{i=1}^{n} \mathbf{p}_i^T(\mathbf{1}+\mathbf{r}_i)\right) \\ &= \lim_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} \log(1+\mathbf{p}_i^T\mathbf{r}_i)\right) \end{aligned} \tag{1.279}$$

If the returns are stationary, then the log-optimal portfolio problem can still be solved by (Györfi et al., 2012)

$$\mathbf{p}_i^* = \operatorname{argmax} \mathrm{E}\{\log(1+\mathbf{p}^T\mathbf{r}_i)|\mathbf{r}_{i-1}\} \tag{1.280}$$

and all is reduced to a maximization problem in which we need to estimate a regression function of the form $\mathrm{E}\{Y|\mathbf{X}\}$. Because of the stationarity of the returns, the data is a sequence of IID copies of $(\mathbf{X}, Y)$, up to the time $i-1$ we have a data collection

$$D_{i-1} = \{(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), ..., (\mathbf{X}_{i-1}, Y_{i-1})\} \tag{1.281}$$

Due to the complex nature of the observed data non-linear and local regressors are preferred. Amongst these, prominent examples are:

- Partition regression: if we divide the predictor space, $\mathbf{X} \in \mathbb{R}^d$ (that is, there are d assets in the market), into small cells, then we look for the cell of $\mathbf{X}$ and compute the average of the Y values associated to that cell.
- Kernel regression: we may compute a weighted average between all samples giving more weight to the most similar samples

$$Y = \frac{\sum\limits_{k=1}^{i-1} K\left(\frac{\|\mathbf{X}-\mathbf{X}_k\|}{h}\right) Y_k}{\sum\limits_{k=1}^{i-1} K\left(\frac{\|\mathbf{X}-\mathbf{X}_k\|}{h}\right)} \tag{1.282}$$

- kNN regression: alternatively we may compute the k nearest neighbours to sample $\mathbf{X}$ and compute the average of those neighbours.

We may integrate this strategy of selection of portfolios into a more complex scheme. Note that the previous strategy amounts to the construction of a regression function of $\log(1+\mathbf{p}^T\mathbf{r}_i)$ dependent on the previously observed returns $\mathbf{r}_{i-1}$. We may construct many different regressors (for example, changing the information available to each regressor or the regression algorithm). Let us assume we construct M regressors. Each regressor is called an expert and it proposes a portfolio $\mathbf{p}_i^{(m)}$. The combined portfolio is a weighted average of all the proposals

$$\mathbf{p} = \frac{\sum\limits_{m=1}^{M} w_m \mathbf{p}_i^{(m)}}{\sum\limits_{m=1}^{M} w_m} \tag{1.283}$$

The weights can be determined by an *a priori* distribution of the experts (q_m such that $\sum\limits_{m=1}^{M} q_m = 1$) and the past performance of each expert

$$w_m = q_m e^{\eta \log S_{i-1}^{(m)}} = q_m \left(S_{i-1}^{(m)}\right)^{\eta} \tag{1.284}$$

being η a positive scalar parameter and $S_{i-1}^{(m)}$ is the capital obtained using only the m-th expert. The so-called histogram, kernel and nearest neighbour based strategies in Györfi et al. (2012) respond to this scheme.

1.8.5 Information theory results

Information theory results assume that the asset prices at all times, $\mathbf{X}[n]$, are extracted from the same known distribution, $f(\mathbf{X})$. This is clearly no the case in reality because the , but it provides some intuition on the properties of log-optimal portfolios. The following are some results from information theory that apply to log-optimal portfolios (Cover and Thomas, 2012)[Chap. 16]:

- The set of log-optimal portfolios with respect to a given distribution of prices (and, consequently, returns) is convex.
- The wealth following a log-optimal portfolio exceeds the wealth of any other for almost any sequence of returns.
- The increase of wealth due to side information Y is bounded by $I(\mathbf{X};Y)$, that is, any side information Y increases the wealth only if it brings any useful information on the prices $\mathbf{X}$.

In practice, the distribution of prices is not known, but it can be estimated from previous price samples. In this case, the optimal portfolio is called the universal portfolio and it turns out to be the combination of many constantly rebalanced portfolios. The idea is mathematically well grounded and works well if the underlying assets are "random walks". But in practice, funds going with CRPs have desappeared (Mountain View Analytics founded by T. Cover).

1.8.6 Alternative portfolio theories

Geometric mean return

Assume that an asset i has expected returns R_{i1} with probability p_{i1}, R_{i2} with probability p_{i2}, ... The expected geometric mean of the asset is (Elton et al., 2003)[Chap. 11]

$$R_{Gi} = (1 + R_{i1})^{p_{i1}}(1 + R_{i2})^{p_{i2}}...(1 + R_{iN_i})^{p_{iN_i}} - 1 \tag{1.285}$$

We may construct a vector with all these expected values for a set of assets, $\boldsymbol{\mu}$. If we have a portfolio with proportions p_i in each asset, the expected geometric mean return would be

$$\mu_{\mathbf{p}} = (1 + R_{G1})^{p_1}(1 + R_{G2})^{p_2}...(1 + R_{GN})^{p_N} - 1 \tag{1.286}$$

We may maximize $\mu_{\mathbf{p}}$ or $\log(\mu_{\mathbf{p}} + 1)$

$$\mathbf{p}^* = \mathrm{argmax} \sum_{i=1}^{N} p_i \log(1 + R_{Gi}) \tag{1.287}$$

This is similar to the mean-variance efficient portfolio if the geometric returns are log-normally distributed. The geometric mean return portfolio tends to diversify more than the modern portfolio alternatives.

Safety first

Safety first was proposed by Roy and we want to minimize the probability of having a loss larger than a given threshold R_l (Elton et al., 2003):

$$\mathbf{p}^* = \mathrm{argmin} \Pr\{\mu_p < R_l\} \tag{1.288}$$

If the returns are normal, then this amounts to measure the Safety-First Ratio

$$SFR = \frac{\mu_R - R_l}{\sigma_R} \tag{1.289}$$

where μ_R and σ_R are the mean and standard deviation of the asset under study. If returns are log-normally distributed, the ratio would be defined similarly although using the logarithm.

Another safety first criterion is Kataoka's criterion

$$\begin{array}{rcll} \mathbf{p}^* & = & \mathrm{argmax} & \mathbf{p}^T \boldsymbol{\mu}_R \\ & & \text{s.t.} & \Pr\{R < \mathbf{p}^T \boldsymbol{\mu}_R\} \leq \epsilon \end{array} \tag{1.290}$$

Here we are maximizing the minimum return that guarantees that is achieved with probability $1 - \epsilon$.

Telser's criterion is

$$\begin{array}{rcll} \mathbf{p}^* & = & \mathrm{argmax} & \mathbf{p}^T \boldsymbol{\mu}_R \\ & & \text{s.t.} & \Pr\{\mathbf{p}^T \boldsymbol{\mu}_R < R_{min}\} \leq \epsilon \end{array} \tag{1.291}$$

That is, maximize the return such that the probability of achieving a portfolio return less than a given threshold is kept under control.

Stochastic dominance

Stochastic dominance is a way of establishing a partial order (an order is partial if given two choices A and B, none of them is preferred to the other). We may establish different kinds of stochastic dominance:

1. First order dominance: A dominates B iff $F_A(x) \leq F_B(x)$ for all x, with strict inequality at some x. Another definition is A dominates B iff $\Pr\{R_A > x\} \geq \Pr\{R_B > x\}$ for all x, with strict inequality at some x.

2. Second order dominance: A dominates B iff $\int\limits_{-\infty}^{x} (F_B(t) - F_A(t))dt \geq 0$ for all x, with strict inequality at some x. Another definition is A dominates B iff $\mathrm{E}\{u(A)\} \geq \mathrm{E}\{u(A)\}$ for any utility function u. First order dominance implies second order dominance. For second order dominance it is necessary that $\mathrm{E}_A\{x\} \geq \mathrm{E}_B\{x\}$ and $\min\limits_A\{x\} \leq \min\limits_B\{x\}$ (i.e., the left tail of B is thicker than the left tail of A).

3. Second order dominance: A dominates B iff $F_A(x) \leq F_B(x)$ for all x, with strict inequality at some x. Another definition is A dominates B iff $\mathrm{E}\{u(A)\} \geq \mathrm{E}\{u(A)\}$ for any utility function u.

There are higher order dominances. The higher the order, the less relevant is the dominance. According to this approach, investments are ordered by stochastic dominance and the best investments are chosen for the portolio. First order dominance is related to the fact that investors prefer more to less, while second order dominance is related to the fact that investors are risk averse (Elton et al., 2003)[Chap. 11]. If returns are normal, then the second-order dominance leads to a portfolio that is efficient from a return-risk point of view.

Value at risk

Given a confidence level α, the VaR at confidence $1-\alpha$ is the α quantile of the return distribution:

$$VaR_\alpha(R) = \sup\{r | F_R(r) \leq \alpha\} \tag{1.292}$$

If the returns are normal, it is

$$VaR_\alpha(R) = \mu_R + z_\alpha \sigma_R \tag{1.293}$$

If the returns are not normal, then we may look directly at the corresponding quantile or, for portfolios, use a Monte Carlo simulation.

Market neutrality

A strategy is market neutral with respect to a source of risk, if the portfolio return has zero-correlation with that source of risk. An example of such a strategy is pairs trading. Two very much correlated companies, like Coca-Cola and Pepsi, both having similar highs and lows, may be invested in opposite ways (long on Company 1, and short on Company 2). On average, this strategy has zero return (since the gain of one of the investments only compensate the losses of the other; not even so if transaction costs are considered). However, we may try to track the disagreement between both prices, so that we give a little bit more weight to the one lagging behind, hoping that the two prices will close the return gap.

1.8.7 International diversification

When an investment is performed in another currency, we must have into account the variance of the currency conversion. So that an investment of capital S_0 after a period of trading is valued

$$S_1 = S_0(1 + R_a)(1 + R_c) \tag{1.294}$$

where R_a is the return of the asset and R_c the return of the currency over the same period. The overall return is

$$\begin{aligned} R &= (1 + R_a)(1 + R_c) - 1 \\ &= R_a + R_c + R_a R_c \end{aligned} \tag{1.295}$$

The mean and variance of the overall return is

$$\begin{aligned} \mu_R &= \mu_a + \mu_c + (\mu_a \mu_c + \sigma_{a,c}) \\ \sigma_R^2 &= \sigma_{a^2,c^2} + (\sigma_a^2 + \mu_a^2)(\sigma_c^2 + \mu_c^2) - (\sigma_{a,c} + \mu_a \mu_c)^2 \end{aligned} \tag{1.296}$$

where $\sigma_{a,c}$ is the covariance between the asset and the currency, and σ_{a^2,c^2} is the covariance between R_A^2 and R_C^2. The covariance between two investments performed in a foreign market is even more complicated. Overall, the investment problem is much more complex, because we have to add models for the variation of the currency conversion.

1.8.8 Portfolio evaluation

R_f is the return of a risk-free asset. R_i the return of the asset under evaluation. μ_i and σ_i their mean and standard deviation respectively. β_i is the β coefficient of the CAPM model (that is, a value related to the correlation between the asset and the market index; see Section 1.9.1). Let R_{min} be the minimum acceptable return. Let us define the Lower Partial Moment of order n as

$$LPM_{ni} = \frac{1}{N} \sum_{n=0}^{N-1} (\max\{R_{min} - R_i[n], 0\})^n \tag{1.297}$$

This is measuring the mean ($n = 1$) or "variance" ($n = 2$) of the losses ($R_i[n] < R_{min}$). Similarly, we may measure the moments of gains

$$HPM_{ni} = \frac{1}{N} \sum_{n=0}^{N-1} (\max\{R_i[n] - R_{min}, 0\})^n \tag{1.298}$$

Let us denote as MD_{in} the size of the n-th maximum drawdown (MD_{i1} is the worst drawdown, MD_{i2} the second worst, ...; they are negative return quantities)

There are several ways of evaluating the goodness of an investment (Aldridge, 2009)[Chap. 5]:

- Sharpe ratio: This is good if returns are normally distributed.

$$SR_i = \frac{\mu_i - R_f}{\sigma_i} \tag{1.299}$$

If the returns were normally distributed, the SR from T samples is estimated with an error whose mean is 0 and its variance is

$$\hat{\sigma}^2_{SR_i} = \frac{1 + 0.5SR_i^2}{T} \tag{1.300}$$

- Treynor ratio: This is good if returns are normally distributed and the investor wishes to split his holdings between the asset and the market portfolio.

$$TR_i = \frac{\mu_i - R_f}{\beta_i} \tag{1.301}$$

- Jensen's α: Measures the trading return in excess to the return predicted by the CAPM model. This is good if returns are normally distributed and the investor wishes to split his holdings between the asset and the market portfolio. It can be manipulated by leverage.

$$\alpha_i = \mu_i - (R_f + \beta_i(R_m - R_f)) \tag{1.302}$$

- Shadwick's Ω: Measures the trading return in excess to the minimum acceptable and compares it to the average of the losses

$$\Omega_i = \frac{\mu_i - R_{min}}{LPM_{1i}} + 1 \tag{1.303}$$

- Sortino ratio: This is equivalent to the Sharpe ratio but only considering the variance of losses

$$SoR_i = \frac{\mu_i - R_{min}}{\sqrt{LPM_{2i}}} \tag{1.304}$$

- Upside potential ratio: This measure computes the mean gain to the variance of losses

$$UPR_i = \frac{HPM_{1i}}{\sqrt{LPM_{2i}}} \tag{1.305}$$

- Calmar ratio: This measure computes the mean return to the maximum drawdown in the whole history

$$CR_i = \frac{\mu_i - R_f}{-MD_{i1}} \tag{1.306}$$

- Sterling ratio: Since the maximum drawdown can be rather extreme, we average the annual (or any other period) maximum drawdowns as a way of having a better estimate of how drawdowns look like

$$CR_i = \frac{\mu_i - R_f}{-\overline{MD}_{i1}} \tag{1.307}$$

- Excess return to Value at Risk: We may use the Value at Risk as a measure of the drawdown

$$CR_i = \frac{\mu_i - R_f}{-VaR_i} \tag{1.308}$$

1.8.9 The limitations of portfolio optimization

Optimization theory is good to understand how the market works, but it is not useful to perform any practical trading. Here are some reasons (Norstad, 2002b):

"Estimating expected returns is notoriously difficult and controversial. Given the large amount of noise in the prices of risky assets like stocks, it is statistically impossible to accurately estimate expected returns from even long historical time series data with more accuracy than a percent or two. Trying to estimate expected returns based on fundamental considerations is even more difficult, and it is impossible to trust any such estimate with an accuracy of more than a percent or two. The optimization algorithm is simply too sensitive to this kind of inaccuracy to be useful.

[...]

The high level of talent in the financial analysis world combined with the extraordinary high level of competition in that world must cause us to be suspicious of the possibility of earning any consistent abnormal profits by using optimizers. Indeed, the long and detailed historical record does not reveal any serious instances of such success.

[...]

The proper way to think of these theories is that the models they build are a good way to at least start to think about how markets work, but they do not give us any kind of edge in trying to beat those markets.

Put yet another way, portfolio optimizer programs by themselves do not help one become smarter than other investors. It is impossible for such a program by itself to give an investor an advantage, because everyone has access to these programs.

Eugene Fama has been reported to have said that if one feels like wasting some time, it's OK to play around with optimization programs, but it is indeed a waste of time. We largely concur with this sentiment, except that playing around with such programs can often help students visualize the mathematics and underlying economic ideas when they are first becoming exposed to the theories. For practical applications, however, they are useless. "

Another reason why trading and portfolio optimization cannot be done fully automatically is beacause if everybody would use the same trading rules, they would create a kind of "synchronized" movement in the market, which would turn the market very unstable. Actually, its stability seems to be relatively week and to result from the exertion of multiple opposing forces (see Sec. 1.2.2).

1.9 Asset pricing

This task is also called Valuation process, it is the way we predict which will be the expected return for each one of the assets.

1.9.1 Capital Asset Pricing Model

This model tries to determine which is the expected value of a given asset through the formula

$$\mathrm{E}\{R_i\} = R_f + \beta_i(\mathrm{E}\{R_m\} - R_f) \tag{1.309}$$

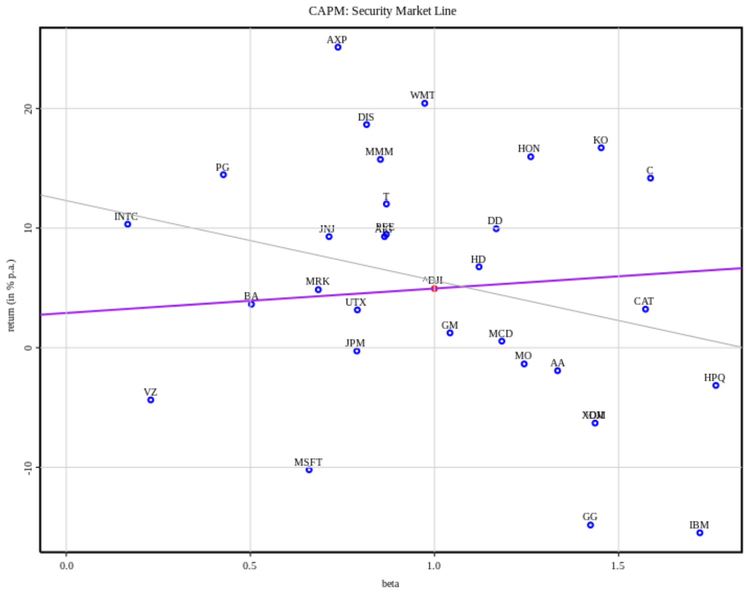


Figure 1.28: Dow jones index and its 30 stocks: the CAPM and the security market line (purple). the grey line is the linear model. mean yield is assumed to be 2.9%.

where R_i is the asset we are interested in, R_f is the return of a risk-free asset, R_m is the return of the market, and β_i is the sensitivity of the asset to the market excess returns. It is calculated as

$$\beta_i = \frac{\text{Cov}\{R_i, R_m\}}{\text{Var}\{R_m\}} \tag{1.310}$$

The Security Market Line is the straight line of $\text{E}\{R_i\}$ as a function of β_i. Given an asset β we can immediately calculate which is the expected return of an asset with that β. Fig. 1.28 shows the CAPM model and the regression line for the DJI components over the 3 years. Note the dispersion about the CAPM and regression lines. In the words of Ivo Welch: *"Unfortunately, in real life, despite its wide use, the evidence in favor of practical use and application of the CAPM is either weak or non-existent. If you use the CAPM, you do so based primarily on a belief that it should work, not based on empirical evidence. Say again: the evidence suggests that, even if the CAPM held, input estimates for corporate cash flows that will occur far in the future are usually so imprecise that they render the CAPM practically useless"* (Welch, 2014)[Chapter 9].

The *Consumption CAPM* proposes a similar model to the one of CAPM, but substituting R_m by an aggregate measure of consumption per capita.

One of the problems of the estimation of β_i is that the Ordinary Least Squares estimate is not too stable. Eisenbeiß et al. (2007) proposes a regularized estimation in which

$$\text{E}\{R_i\} = R_f + (\beta_i + \tilde{\beta}_i(t))(\text{E}\{R_m\} - R_f) \tag{1.311}$$

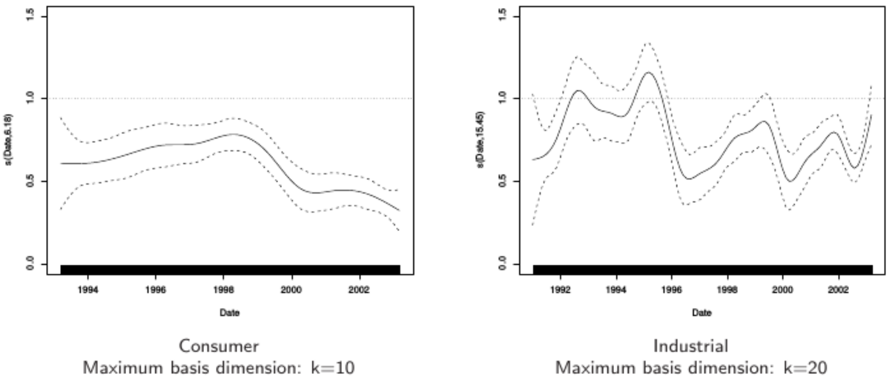


Figure 1.29: Smooth β_i curves for a couple of sector indices.

with the constraint that $\int \tilde{\beta}_i(t))dt = 0$, that is, the β_i coefficient dependence on time is explicitly considered through a B-spline model, which is regularized to have small second derivative. The regression is weighted by the local variance, $\frac{1}{\sigma_i^2(t)}$ as to accommodate heterocedasticity. The β_i fitting is now much smoother over time as shown in Fig. 1.29. In Kauermann et al. (2011), they show that this spline smoothing compares favourably with respect to Hodrick-Prescott filter (a spline smoothing with white noise, instead of correlated noise, and regularized on the spline second derivative), bandpass and highpass filters. A alternative to this smooth fitting is the Kalman Filter.

1.9.2 Arbitrage Pricing Theory

We may presume that the returns observed in the market can be explained as linear combinations of a few underlying factors (Barra, 2007). In this way,

$$\mathbf{R} = \mathbf{R}_0 + X\mathbf{F} + \boldsymbol{\epsilon} \tag{1.312}$$

The assumptions are

$$\begin{aligned} \mathrm{E}\{\boldsymbol{\epsilon}\} &= \mathbf{0} \\ \mathrm{E}\{\mathbf{F}\} &= \mathbf{0} \\ \mathrm{Cov}\{\mathbf{F}\} &= I \\ \mathrm{Cov}\{\boldsymbol{\epsilon}\} &= \Sigma_{\boldsymbol{\epsilon}} \\ \mathrm{Cov}\{\boldsymbol{\epsilon}, \mathbf{F}\} &= 0 \end{aligned} \tag{1.313}$$

The covariance matrix of $\mathbf{R}$ can be computed from the covariance matrices of $\mathbf{F}$ and $\boldsymbol{\epsilon}$ (the residuals are supposed to be independent of the factors and independent amongst themselves)

$$\Sigma_{\mathbf{R}} = X\Sigma_{\mathbf{F}}X^T + \Sigma_{\boldsymbol{\epsilon}} \tag{1.314}$$

We may distinguish between the unconditional moments

$$\begin{aligned} \mathrm{E}\{\mathbf{R}\} &= \boldsymbol{\mu}_{\mathbf{R}} = \mathbf{R}_0 + X\boldsymbol{\mu}_{\mathbf{F}} \\ \mathrm{Cov}\{\mathbf{R}\} &= \Sigma_{\mathbf{R}} = X\Sigma_{\mathbf{F}}X^T + \Sigma_{\boldsymbol{\epsilon}} \end{aligned} \tag{1.315}$$

and the conditional ones

$$\begin{array}{rcl} \mathrm{E}\{\mathbf{R}|\mathbf{F}[n]\} & = & \mathbf{R}_0 + X\mathbf{F}[n] \\ \mathrm{Cov}\{\mathbf{R}|\mathbf{F}[n]\} & = & \Sigma_{\boldsymbol{\epsilon}} \end{array} \tag{1.316}$$

$\mathbf{R}_0$ and X must be determined by linear regression.

If the factors are specified by the user (like the overall market return, the return of an industry index, the return of the U.S. Treasure Bond, ...), then the only unknown of this problem is the regression parameters (X and $\mathbf{R}_0$). This would be a regression problem. If both, the parameters and the factors are unknown, then this is a Factor Analysis problem.

The relationship between $\mathbf{F}$ and $\mathbf{R}$ may be non-linear:

$$\mathbf{R} = \mathbf{h}(\mathbf{F}) + \boldsymbol{\epsilon} \tag{1.317}$$

However, if the deviations of $\mathbf{F}$ from $\boldsymbol{\mu}_{\mathbf{F}}$ are small so that the Taylor expansion is valid

$$\begin{array}{rcl} \mathbf{h}(\mathbf{F}) & \approx & \mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}) + D_{\mathbf{F}}\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}})(\mathbf{F} - \boldsymbol{\mu}_{\mathbf{F}}) \\ & = & (\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}) - D_{\mathbf{F}}\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}})\boldsymbol{\mu}_{\mathbf{F}}) + (D_{\mathbf{F}}\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}))\mathbf{F} \end{array} \tag{1.318}$$

See the Mathematical Tools Appendix for the definition of the derivative of a multivariate function. Then, we would be back to the linear case above, in which

$$\begin{array}{rcl} \mathbf{R}_0 & = & \mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}) - (D_{\mathbf{F}}\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}))\boldsymbol{\mu}_{\mathbf{F}} \\ X & = & D_{\mathbf{F}}\mathbf{h}(\boldsymbol{\mu}_{\mathbf{F}}) \end{array} \tag{1.319}$$

Under the linear factor model, the mean and variance of the portfolio can be easily calculated in terms of the factors

$$\begin{array}{rcl} \mathrm{E}\{R_{\mathbf{p}}\} & = & \mathbf{p}^T(\mathbf{R}_0 + X\boldsymbol{\mu}_{\mathbf{F}}) \\ \mathrm{Var}\{R_{\mathbf{p}}\} & = & \mathbf{p}^T(X\Sigma_{\mathbf{F}}X^T + \Sigma_{\boldsymbol{\epsilon}})\mathbf{p} \end{array} \tag{1.320}$$

Under the non-linear factor model, we have

$$\begin{array}{rcl} \mathrm{E}\{R_{\mathbf{p}}\} & = & \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{h}(\mathbf{F})} \\ \mathrm{Var}\{R_{\mathbf{p}}\} & = & \mathbf{p}^T(\Sigma_{\mathbf{h}(\mathbf{F})} + \Sigma_{\boldsymbol{\epsilon}})\mathbf{p} \end{array} \tag{1.321}$$

where

$$\begin{array}{rcl} \boldsymbol{\mu}_{\mathbf{h}(\mathbf{F})} & = & \mathrm{E}_{\mathbf{F}}\{\mathbf{h}(\mathbf{F})\} \\ \Sigma_{\mathbf{h}(\mathbf{F})} & = & \mathrm{E}_{\mathbf{F}}\{(\mathbf{h}(\mathbf{F}) - \boldsymbol{\mu}_{\mathbf{h}(\mathbf{F})})(\mathbf{h}(\mathbf{F}) - \boldsymbol{\mu}_{\mathbf{h}(\mathbf{F})})^T\} \end{array} \tag{1.322}$$

The *financial elasticity*, β, of an equity with respect to a factor (normally, the market return) is the regression coefficient of the return when that single factor is used as predictor

$$R_i = R_{0i} + \beta_{ij}F_j + \epsilon_i \tag{1.323}$$

Then

$$\beta_{ij} = \frac{\mathrm{Cov}\{R_i, F_j\}}{\mathrm{Var}\{F_j\}} = \frac{\sigma_{ij}}{\sigma_j^2} = \rho_{ij}\frac{\sigma_i}{\sigma_j} \tag{1.324}$$

where σ_i^2 is the variance of R_i, σ_j^2 is the variance of F_j, σ_{ij} is the covariance between the equity return and the factor, and ρij is the correlation index between these two variables.

An heuristic way of constructing portfolios (Elton et al., 2003)[Chap. 9] is by choosing equities by descending order of excess return to β

$$\frac{\mathrm{E}\{R_i\} - R_{\mathrm{free}}}{\beta_{1m}} \tag{1.325}$$

where $\mathrm{E}\{R_i\}$ is the expected return for the i-th equity, R_{free} is the return of a risk-free asset, and β_{im} is the β of the stock when regressed to the market index. (Elton et al., 2003)[Chap. 9] provides formulas for the calculation of $\mathbf{p}$ under this heuristic and some simplifications (like a model with a single index).

We may also calculate the financial elasticity between a portfolio and a factor

$$\begin{aligned} \mathrm{Cov}\{R_p, F_j\} &= \mathrm{E}\{(\mathbf{p}^T\mathbf{R} - \mathbf{p}^T\boldsymbol{\mu}_{\mathbf{R}})(\mathbf{e}_j^T\mathbf{F} - \mathbf{e}_j^T\boldsymbol{\mu}_{\mathbf{F}})\} \\ &= \mathbf{p}^T\mathrm{E}\{(\mathbf{R} - \boldsymbol{\mu}_{\mathbf{R}})(\mathbf{F} - \boldsymbol{\mu}_{\mathbf{F}})^T\}\mathbf{e}_j \\ &= \mathbf{p}^T\mathrm{E}\{(\mathbf{R}_0 + X\mathbf{F} + \boldsymbol{\epsilon} - \mathbf{R}_0 - X\boldsymbol{\mu}_{\mathbf{F}})(\mathbf{F} - \boldsymbol{\mu}_{\mathbf{F}})^T\}\mathbf{e}_j \\ &= \mathbf{p}^T X\Sigma_{\mathbf{F}}\mathbf{e}_j \end{aligned} \tag{1.326}$$

Note that the factors in $\mathbf{F}$ do not need to be orthogonal to each other. By applying a Principal Component Analysis we may find a set of orthogonal factors $\mathbf{f}$ such that

$$\mathbf{F} = \boldsymbol{\mu}_{\mathbf{F}} + B\mathbf{f} + \mathbf{u} \tag{1.327}$$

Under this decomposition we would have

$$\Sigma_{\mathbf{F}} = B\Sigma_{\mathbf{f}}B^T + \Sigma_{\mathbf{u}} \tag{1.328}$$

being $\Sigma_{\mathbf{f}}$ and $\Sigma_{\mathbf{u}}$ diagonal matrices.

Some models (Barra, 2007)[Chap. 3] try to model also the variance $\boldsymbol{\epsilon}$ at Eq. 1.312. At each time point n, the forecasted standard deviation for a particular asset i is of the form

$$\hat{\sigma}_i[n] = \kappa_i[n](1 + \hat{R}_i[n])\hat{R}_G[n] \tag{1.329}$$

where $\hat{R}_i[n]$ is a prediction of the return of asset i at time n, $\hat{R}_G[n]$ is a prediction of the global return of the market at time n and $\kappa_i[n]$ is a scaling factor. On their turn, the global market return is predicted through an AR model

$$\hat{R}_G[n] = R_{0G} + \sum_{k=1}^{P} a_k R_G[n-k] \tag{1.330}$$

and a factor model

$$\hat{\mathbf{R}}[n] = Y[n]\mathbf{g}[n] \tag{1.331}$$

where $Y[n]$ is a time-varying matrix and $\mathbf{g}[n]$ the factors used to predict the returns of each asset.

We may use this linear model in a different way. Let us assume now that we want to discover factors. Then, we fix the matrix X (for instance by using indicator variables, e.g., which industry this asset belongs to, ...)

$$\mathbf{R} = X\mathbf{F} + \boldsymbol{\epsilon} \tag{1.332}$$

The constraint is that X^TX is an invertible matrix (normally it is chosen to be diagonal). Then, we can look for the factors solving for $\mathbf{F}$ in a LS sense

$$\mathbf{F} = (X^TX)^{-1}X^T\mathbf{R} \tag{1.333}$$

In this way, industry factor models are built.

A more general approach is provided by Factor Analysis (FA) and Principal Component Analysis (PCA). In factor analysis, it is assumed that there are a set of latent, normally distributed, independent factor $\mathbf{F} \sim N(0, I)$ such that they generate the observed returns

$$\mathbf{R} = \mathbf{R}_0 + X\mathbf{F} + \boldsymbol{\epsilon} \tag{1.334}$$

The goal of FA is to determine X. This is done through a Maximum Likelihood approach, in which all the processes involved are assumed to be Gaussian. Similarly, PCA uses the same model, but it constructs X with the first eigenvectors of the matrix $\Sigma_{\mathbf{R}}$ (the eigenvectors associated to the largest eigenvalues).

1.9.3 Discounted cash flow

This model states that the value of a share of a stock is equal to the present value of the cash flow that the stockholder expects to receive from it. Let X_t be the price at time t and D_t the dividend received at time t. Let k be the appropriate discount rate (like inflation). Then the current value of a dividend received at time $t+1$ and the price at the same time is given by (Elton et al., 2003)[Chap. 18]

$$X_t = \frac{D_{t+1}}{1+k} + \frac{X_{t+1}}{1+k} \tag{1.335}$$

This is a recursive equation whose solution is

$$X_t = \frac{D_{t+1}}{1+k} + \frac{D_{t+2}}{(1+k)^2} + \frac{D_{t+3}}{(1+k)^3} + ... \tag{1.336}$$

If the dividends grow at a constant rate, $D_{t+n} = D_{t+n-1}(1+g)$, then we have a geometric series, whose sum is

$$X_t = \frac{D_{t+1}}{k-g} \tag{1.337}$$

g may be estimated from the expected return of the firm on its investments, r, and the fraction of earnings retained by the firm as $g = rb$. The constant growth model assumes that the firm will maintain a stable dividen policy (keep its retention rate constant), and earn a stable return on new investments over time.

1.9.4 Cross-sectional regression

The difference between cross-sectional regression and time-series regression is that cross-sectional regression performs a regression between $\mathbf{x}_t$ and y, i.e., a number of predictor variables like (earnings growth rate, dividend payout rate, standard deviation in growth rate, inflation, ...) measured at time t; while time-series regression performs the regression between $..., \mathbf{x}_{t-2}, \mathbf{x}_{t-1}, \mathbf{x}_t$ and y, i.e., all the historical knowledge is employed.

Cross-sectional regression has been employed to predict the Price/Earnings Ratio (PER) as a function of variables like earnings growth rate, dividend payout rate, standard deviation in growth rate, oil price, inflation, etc. (Elton et al., 2003)[Chap. 18]. It is important to note that the regression coefficients may vary depending on the market trend, so that different regressions may be performed.

1.10 Machine learning

1.10.1 Distance learning

Do et al. (2015) proposes an interesting approach to learning distances that maximize a classification of time series. Let us summarize here their approach. There are two types of time series distances: 1) those that care about the specific values $d_V(\mathbf{x},\mathbf{y})$ (like the Euclidean, Minkowski and Mahalanobis distances) and 2) those that care about the shape $d_S(\mathbf{x},\mathbf{y})$ (like the correlation or the temporal correlation). The temporal correlation is defined as

$$Cort_r(\mathbf{x},\mathbf{y}) = \frac{\sum\limits_{|n-n'|\leq r}(x[n]-x[n'])(y[n]-y[n'])}{\sqrt{\sum\limits_{|n-n'|\leq r}(x[n]-x[n'])^2}\sqrt{\sum\limits_{|n-n'|\leq r}(y[n]-y[n'])^2}} \quad (1.338)$$

Pearson correlation is a particular case of this measure when $r = N-1$ (being N the number of samples in the time series). We may turn the temporal correlation into a distance by

$$d_S(\mathbf{x},\mathbf{y}) = \frac{1-Cort_r(\mathbf{x},\mathbf{y})}{2} \quad (1.339)$$

We may also combine values and shape into a single distance based on a parameter α with any of the formulas below

$$\begin{array}{rcl} d_\alpha(\mathbf{x},\mathbf{y}) & = & \frac{2d_V(\mathbf{x},\mathbf{y})}{1+\exp(1-2d_S(\mathbf{x},\mathbf{y}))} \\ d_\alpha(\mathbf{x},\mathbf{y}) & = & \alpha d_V(\mathbf{x},\mathbf{y})+(1-\alpha)d_S(\mathbf{x},\mathbf{y}) \\ d_\alpha(\mathbf{x},\mathbf{y}) & = & d_V^\alpha(\mathbf{x},\mathbf{y})d_S^{1-\alpha}(\mathbf{x},\mathbf{y}) \end{array} \quad (1.340)$$

We may also combine several metrics

$$d_\mathbf{w}(\mathbf{x},\mathbf{y}) = \sum_i w_i d_i(\mathbf{x},\mathbf{y}) \quad (1.341)$$

Let us define a training set $\{\mathbf{x}_i, y_i\}$ that is going to be classified with a kNN classifier. The goal is to learn the metrics $d_\mathbf{w}$ that minimizes the classification eror. This can be done by pulling similar samples towards the central sample and by pushing disimilar samples outside of the neighbourhoods. Let $N_k(\mathbf{x}_i)$ be the k neighbours of the time series $\mathbf{x}_i$. Let $S_k(\mathbf{x}_i)$ be the subset of the neighbourhood with the same label as the sample i, and $\bar{S}_k(\mathbf{x}_i)$ the set of samples with disimilar label. Then the goal is to minimize

$$\min_\mathbf{w} \sum_{j\in S_k(\mathbf{x}_i)} d_\mathbf{w}(\mathbf{x}_i,\mathbf{x}_j) + C\sum_{j\in \bar{S}_k(\mathbf{x}_i)} d_\mathbf{w}(\mathbf{x}_i,\mathbf{x}_j) \quad (1.342)$$

(Their optimization problem is slightly different, but this one shows better the philosophy). They use 1NN classifiers for their experiments and decrease the classification error between 1 and 3%. Surprisingly, for many datasets, the d_V or d_S alone provide the best classifiers.

1.10.2 Strangeness measures

Given a vector $\mathbf{x}$ with label y, we may try to characterize how strange it is with respect to a training set. For doing so we have different measures (Ho and Wechsler, 2010):

- Using kNN: We may compute the distance of $\mathbf{x}$ to the k closest samples with label y, and to the k closest samples with a label different from y and compare both distances. The smaller the ratio, the less strange the new sample is.

$$s = \frac{\sum\limits_{j=1}^{k} d_j^y}{\sum\limits_{j=1}^{k} d_j^{\neg y}} \tag{1.343}$$

- Using SVM: Let us assume we have trained a SVM with N samples, and we want to determine the strangeness of a new sample. Assume $y \in \{-1, 1\}$. We may look for its Lagrange multiplier by minimizing the dual problem

$$Q(\boldsymbol{\lambda}) = -\frac{1}{2}\sum_{i=1}^{N+1}\sum_{j=1}^{N+1}\lambda_i\lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{N+1}\lambda_i \tag{1.344}$$

subject to $\sum\limits_{i=1}^{N+1}\lambda_i y_i = 0$ and $0 \leq \lambda_i \leq C$. The idea is that well classified samples will have 0 lagrange multiplier, and samples within the margin will have a Lagrange multiplier different from 0. A Gaussian kernel has been found to perform well (Ho and Wechsler, 2010) and it is recommended some kind of online SVM.

$$s = \lambda_{N+1} \tag{1.345}$$

- Using clustering: We may cluster the training set, including the new sample and compute the distance of the new sample to its centroid.

$$s = d(\mathbf{x}, \mathbf{c}) \tag{1.346}$$

- Using regression: We may calculate a regression, $y = f(\mathbf{x})$ and estimate an error function g (like $g(\mathbf{x}_i) = \log|y_i - f(\mathbf{x}_i)|$). Then, the strangeness is defined as

$$s = \frac{|y - f(\mathbf{x})|}{\exp(g(\mathbf{x}))} \tag{1.347}$$

1.11 High-frequency trading

High-frequency trading exploits the intraday variability in order to increase profits. The following table shows the potential maximum gain in the EUR/USD exchange ratio (Aldridge, 2009)[Chap. 7]

Statistic	10 s	1 min	10 min	1 h	1 day
Max. Gain (%)	319	90	18.5	6.5	0.6
Avg. Range per period (%)	0.04	0.06	0.13	0.27	0.6
Number of intraday periods	8640	1440	144	24	1

The Sharpe ratio has to be adjusted by a factor that accounts for the number of periods

$$SR_i = \frac{\mu_i - R_f}{\sigma_i}\sqrt{PD} \tag{1.348}$$

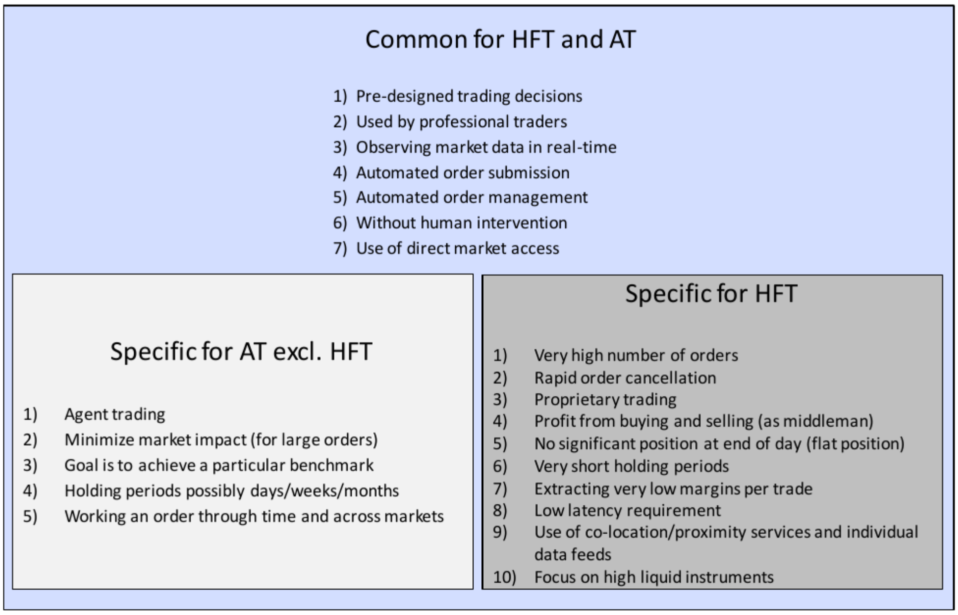


Figure 1.30: Similarities and dissimilarities between Algorithmic Trading and High-Frequency Trading.

where P is the number of intraday periods, and D is the number of trading days in the year.

High-frequency trading has been shown to increase the efficiency of the market and to improve its quality. However, regulatory issues must be undertaken to avoid situations like the "flash crash" (May 6th, 2010) which seemed to be caused by a combination of factors (Chlistalla et al., 2011). The following paragraphs are extracted from Gomber et al. (2011) and gives an overview of the families of High-Frequency Trading algorithms. The paper distinguishes between Algorithmic Trading (AT) and High-Frequency Trading (HFT) (see Fig. 1.30).

1.11.1 Algorithmic trading

The Scope of Algorithmic Trading Strategies

HFT is mostly defined as a subset of AT strategies. However, not all algorithmic strategies are necessarily high frequent. Most non-HFT algorithmic strategies aim at minimizing the market impact of (large) orders. They slice the order into several smaller child orders and spread these child orders out across time (and/or venues) according to a pre-set benchmark. The following subsections describe some of the more common non-HFT algorithmic strategies.

The classification into four generations is based on Almgren (2009) and includes information from Johnson (2010). First generation algorithms focus solely on benchmarks that are based on market generated data (e.g. VWAP) *and are independent from the actual order and the order book situation at order arrival, while the second generation tries to define the benchmark based on the individual order and to handle the trade-off between market impact and timing risk. Third generation algorithms are furthermore able to adapt to their own performance during executions. A fourth generation – that is not included in the Almgren (2009) classification – consists of so called newsreader algorithms.*

First Generation Execution Algorithms

Participation Rate Algorithms

Participation rate algorithms are relatively simple. They are geared to participate in the market up to a predefined volume. Such an algorithm could for example try to participate by trading 5% of the volume in the target instrument(s) until it has built or liquidated a target position. Since these algorithms target traded volume, they reflect the current market volume in their orders. Variants of these algorithms add execution periods during which orders are submitted to the market or maximum volumes or prices. Furthermore, randomized participation rates are used to make the algorithm harder to detect for other market participants.

Time Weighted Average Price (TWAP) Algorithms

TWAP algorithms divide a large order into slices that are sent to the market in equally distributed time intervals. Before the execution begins, the size of the slices as well as the execution period is defined. For example, the algorithm could be set to buy 12,000 shares within one hour in blocks of 2,000 shares, resulting in 6 orders for 2,000 shares which are sent to the market every 10 minutes. TWAP algorithms can vary their order sizes and time intervals to prevent detection by other market participants.

Volume Weighted Average Price (VWAP) Algorithms

VWAP algorithms try to match or beat the volume weighted average price (their benchmark) over a specified period of time. VWAP can be calculated applying the following formula for n trades, each with an execution price pn and size vn (Johnson 2010):

$$VWAP = \frac{\sum_n V_n P_n}{\sum_n V_n} \tag{1.349}$$

Since trades are being weighted according to their size, large trades have a greater impact on the VWAP than small ones. VWAP algorithms are based on historical volume profiles of the respective equity in the relevant market to estimate the intraday/target period volume patterns.

Second Generation Execution Algorithms

The most prominent second generation algorithms try to minimize implementation shortfall. The current price/midpoint at the time of arrival of an order serves as a benchmark, which shall be met or outperformed (order based benchmark). Implementation shortfall algorithms try to minimize the market impact of a large order taking into account potential negative price movements during the execution process (timing risk). To hedge against an adverse price trend, these algorithms predetermine an execution plan based on historical data, and split an order into as many as necessary but as few as possible sub orders. In

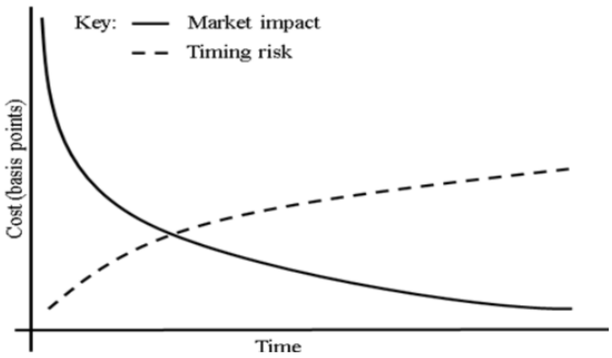


Figure 1.31: Market impact vs. timing risk.

contrast to TWAP or VWAP, these orders will be scattered over a period which is just long enough to dampen the market impact of the overall order (Johnson 2010). Figure 1.31 shows the trade-off between minimizing market impact and timing risk.

Third Generation Execution Algorithms

Adaptive algorithms form the third generation in Almgren's classification (Almgren 2009). These algorithms follow a more sophisticated approach than implementation shortfall algorithms. Instead of determining a pre-set schedule, these algorithms re-evaluate and adapt their execution schedule during the execution period, making them adaptive to changing market conditions and reflecting gains/losses in the execution period by a more/less aggressive execution schedule.

Newsreader algorithms

Investors have been relying on news to make their investment decisions ever since the first stock market opened its gates. Since then, traders who possess valuable information have been using it to generate profits. However, there is a limit to the quantity of data a human trader can analyze, and maybe even more important, the human nature of an investor/trader limits the speed with which he/she can read incoming news. This has led to the development of newsreader algorithms.

These automated newsreaders employ statistical methods as well as text-mining techniques to discern the likely impact of news announcements on the market. Newsreader algorithms rely on high-speed market data. Exchanges and news agencies have developed low latency news feeds, which provide algorithmic traders with electronically processable news.

1.11.2 High-Frequency Trading

While consolidated information on the major players in HFT is still scarce, the community of market participants leveraging HFT technologies to implement

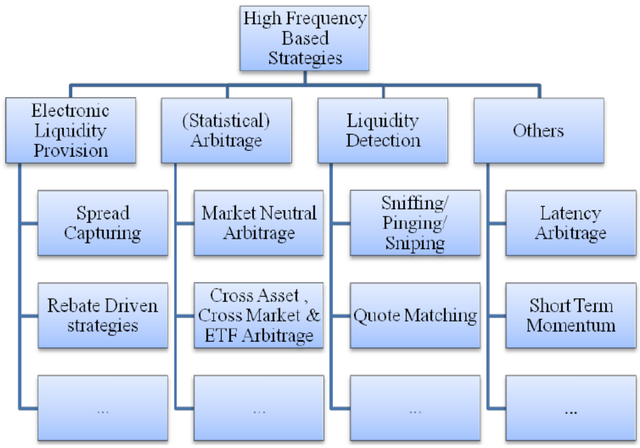


Figure 1.32: Common HFT techniques.

their trading strategies is highly diverse. Its members range from broker-dealer operated proprietary trading firms and broker-dealer market making operations to specialized HFT boutiques to quantitative hedge funds leveraging HFT technology in order to increase the profits from their investment and trading strategies (see Easthope and Lee 2009). There is (i) a multitude of different institutions with different business models that use HFT and (ii) there are many hybrid forms, e.g. broker-dealers which run their proprietary trading books applying HFT techniques. Therefore, in the assessment of HFT it is very important to take a functional rather than an institutional perspective. In order to achieve a level playing field, all institutions that apply HFT based trading strategies have to be taken into consideration independent of whether HFT is their core or an add-on technology to implement trading strategies.

The scope of HFT strategies

While the universe of HFT strategies is to diverse and opaque to name them all, some of these strategies are well known and not necessarily new to the markets. The notion of HFT often relates to traditional trading strategies that use the possibilities provided by state-of- the-art IT. HFT is a means to employ specific trading strategies rather than a trading strategy in itself. Therefore, instead of trying to assess HFT as such, it is necessary to have a close look at the individual strategies that use HFT technologies (see Figure 1.32). The following subsections shed light on some of the best known and probably most prominent HFT based strategies.

Electronic Liquidity Provision

One of the most common HFT strategies is to act as a liquidity provider. While many HFTs provide the market with liquidity like registered market makers, they frequently do not face formal obligations to quote in the markets in which they

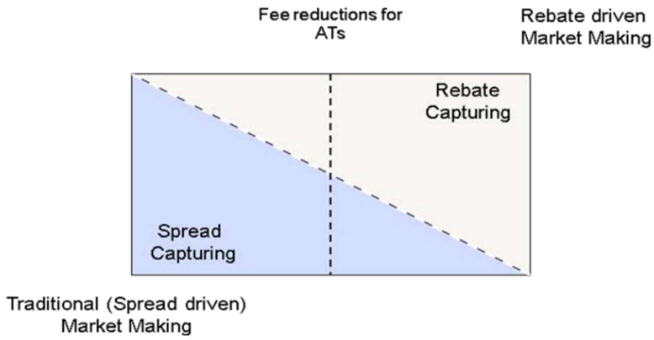


Figure 1.33: Revenue sources for HFT based on liquidity provision.

are active. HFT liquidity providers have two basic sources of revenues: (i) They provide markets with liquidity and earn the spread between bid and ask limits and (ii) trading venues incentivize these liquidity provides by granting rebates or reduced transaction fees in order to increase market quality and attractiveness. Figure 1.33 depicts these different revenue sources for HFT electronic liquidity provision strategies.

Spread Capturing

A HFT strategy, which closely resembles its traditional counterpart, i.e. market making, is spread capturing. These liquidity providers profit from the spread between bid and ask prices by continuously buying and selling securities (ASIC 2010a). With each trade, these liquidity providers reap the spread between the (higher) price at which market participants can buy securities and the (lower) one at which they can sell securities.

Rebate Driven Strategies

Other liquidity provision strategies are built around particular incentive schemes of some markets. In order to attract liquidity providers and react to increasing competition among markets, some trading venues have adopted asymmetric pricing: members removing liquidity from the market (taker; aggressive trading) are charged a higher fee while traders who submit liquidity to the market (maker; passive trading) are charged a lower fee or are even provided a rebate. An asymmetric fee structure is supposed to incentivize liquidity provision. A market operator's rationale for applying maker-taker pricing is given by the following: traders supplying liquidity on both sides (buy and sell) of the order book earn their profits from the market spread. Fee reductions or even rebates for makers shall stimulate a market's liquidity by firstly attracting more traders to post passive order flow in form of limit orders. Secondly, those traders submitting limit orders shall be incentivized and enabled to quote more aggressively, thus narrowing the spread. The respective loss of profits from doing so is supposed to be compensated by a rebate. If this holds true, those markets appear favorable over their rivals and market orders are attracted enhancing the probability for the makers to have their orders executed (Lutat 2010).

(Statistical) Arbitrage

Opportunities to conduct arbitrage strategies frequently exist only for very brief periods (fractions of a second). Since computers are able to scan the markets for such short-lived possibilities, arbitrage has become a major strategy applied by HFTs. These HFTs conduct arbitrage in the same way as their traditional counterparts; they leverage state of the art technology to profit from small and short-lived discrepancies between securities. The following types of arbitrage are not limited to HFT, but are conducted by non-automated market participants as well. Since arbitrageurs react on existing inefficiencies, they are mainly takers of liquidity.

Market Neutral Arbitrage

This form of statistical arbitrage aims to be "market neutral". Arbitrageurs try to hold instruments while simultaneously shorting other instruments. Since the instruments are closely correlated, gains and losses due to movements of the general market will (mostly) offset each other. However, in order to gain from this strategy, arbitrageurs sell an instrument which they deem to have a relatively lower intrinsic value, while simultaneously buying an instrument, which reacts very similar (ideally identical) to changes in the market environment and which they deem to have a relatively higher intrinsic value. If the respective valuation of these instruments "normalizes" into the expected direction, the arbitrageur liquidates its market neutral position. Gains from this strategy result from the difference between the individual valuation of the assets at the time the position is opened and their "normalized" prices at the time the position is liquidated. Since this strategy offers protection against market movements, it is highly attractive for HFTs and traditional arbitrageurs alike. (Aldridge 2010)

Cross Asset, Cross Market & Exchange Traded Fund (ETF) Arbitrage

An established arbitrage strategy is to trade instruments across markets or to trade related instruments and to profit from pricing inefficiencies across markets: if an asset shows differing prices across marketplaces, arbitrageurs generate profits by selling the asset on the market where it is valued higher and simultaneously buying it on another market where it is valued lower. Cross market arbitrage strategies have profited from the increased market fragmentation in Europe as described in section two. A higher number of markets increases the probability that an instrument has different prices across these markets. Similarly, arbitrageurs can profit from inefficiencies across assets: if, e.g. an option is priced too high relative to its underlying; arbitrageurs can earn profits by selling the option and simultaneously buying the underlying. In a similar way, ETF arbitrageurs trade ETFs against their underlying and profit from respective pricing inefficiencies. Since such inefficiencies exist only shortly on modern securities markets, HFTs leverage their speed advantage to trade against them (see Aldridge 2010 for more information).

Liquidity Detection

Another category of HFT strategies is liquidity detection. These HFTs try to discern the patterns other market participants leave in the markets and adjust their actions accordingly. Liquidity detectors focus their attention on large orders and employ various strategies to detect sliced orders, hidden orders, orders *being submitted by execution algorithms or to gain further information about electronic limit order books (ASIC 2010a). Liquidity detectors gathering information about algorithmic traders are frequently referred to as "sniffing out" other algorithms. Other detectors "ping" or "snipe" in order books or dark pools to retrieve information from them (see e.g. ASIC 2010a).*

Another possible way to use HFT technology would be a high speed version of the "quote matching" strategy described by Harris (2003). Using this strategy, a trader who has detected a large order within the order book places his own order ahead of the large order. If he has detected for example a large buy order, he places his own buy order at a slightly higher limit. Should prices now move upwards, he profits from the rise. However, should prices fall, the large order resting in the book serves as an option/hedge against which the trader can sell his own shares, thereby limiting his possible losses as long as the large limit order rests within the book.

Other High-Frequency Trading Strategies

Latency Arbitrage

Some market participants accuse HFTs of conducting a form of arbitrage which is purely based on their faster access to market data. This modern form of arbitrage, where HFTs are said to be able to see (and interpret) new market information before many market participants even receive it, is frequently referred to as latency arbitrage. These latency arbitrageurs leverage direct data feeds and co-located infrastructure to minimize their reaction times. Especially in the U.S., where many market participants rely on the "national best bid and offer" (NBBO), latency arbitrageurs are said to be able to profit from their speed advantage in comparison to the NBBO (see e.g. Gaffen 2009). Since actions of these market participants are said to impair the prices at which other traders (e.g. buy side execution algorithms) are able to trade, they are often called "predatory". While it is not possible for the authors to assess the actual effect of latency arbitrage on securities markets or the magnitude at which this strategy is conducted, it seems that the discussion described above is currently limited to the U.S. and its NBBO. Therefore, at least those forms which are built around this distinctive feature of the U.S. market system are not applicable in European markets, where no (statutory) NBBO exists.

Short-Term Momentum Strategies

Market participants leveraging HFT technologies to conduct short-term momentum strategies are a modern equivalent to classical day traders. In contrast to many other HFT based strategies they are neither focused on providing the market with liquidity, nor are they targeting market inefficiencies. They usually trade aggressively (taking liquidity) and aim at earning profits from market movements/trends. Their trading decisions can be based on events influencing securities markets and/or the movements of the markets themselves. Momentum based trading strategies are not new and have been implemented by traditional traders for a long time.

1.11.3 Conclusions

HFT is a technical means to implement established trading strategies. *HFT is not a trading strategy as such but it applies latest technological advances* *in market access, market data access and order routing to maximize the returns of established and well known trading strategies. Therefore, the assessment and the regulatory discussion about HFT should focus on underlying strategies rather than on HFT as such.*

HFT is a natural evolution of the securities markets instead of a completely new phenomenon*. Since the advent of electronic markets, market participants tried to minimize trading costs and to maximize their profits from electronic executions. While "HFT" is a relatively new term, the underlying concept is not new at all. From the first quote machines to direct market access tools to smart order routing systems, there is a clear evolutionary process in market participants' adoption of new technologies in changing market environments, triggered by competition, innovation and regulation. Like all other technologies, HFT enables sophisticated market participants to achieve legitimate rewards on their investments – especially in technology – and compensation for their market, counterparty and operational risk exposures.*

A lot of problems related to HFT are rooted in the U.S. market structure*. Both the flash crash on May 6, 2010 and the discussions about flash orders relate to the U.S. equity market structure and the NMS. Some market observers argued that HFT is the key problem of the flash crash and around flash orders. However, the U.S. trade-through rule and a circuit breaker regime that neither targeted at individual equities nor sufficiently aligned among U.S. trading venues are relevant causes for both problems. In Europe, where a more flexible best execution regime without re-routing obligations has been implemented by MiFID and a share-by-share volatility safeguard regime has been in existence for two decades, no market quality problems related to HFT have been documented so far. Therefore, a European approach to the subject matter is required and Europe should be cautious in addressing and fixing a problem that exists in a different market structure, hence potentially creating risks for European market efficiency and market quality.*

The majority of HFT strategies contributes to market liquidity (market-making strategies) or to price discovery and market efficiency (arbitrage strategies)*. Preventing these strategies by inadequate regulation or by impairing underlying business models through excessive burdens may trigger counterproductive and unforeseen effects to market efficiency and quality. Any arguments that try to associate or equate HFT based strategies with market abuse miss the point; there is no ground for treating entities that are applying HFT differently from other market participants in this respect. However, any approach that uses the new possibilities of sophisticated IT to run abusive strategies against market integrity or in order to deliberately exercise disruptive or confusing effects on other market participants must be effectively combated by supervisory authorities.*

Academic literature mostly shows positive effects of AT-/HFT based strategies on market quality*. Six out of eight recently published or publicly accessible papers, focusing on HFT, do not find evidence for negative effects of HFT on market quality. On the contrary, the majority argues that HFT generally contributes to market quality and price formation. In this regard, most studies find positive effects on liquidity and short term volatility. Only one paper, in its theoretical part, critically points out that under certain circumstances HFT might increase an adverse selection problem. The issue of HFT behavior under market stress has not been in the focus of many analyses so far, but in* *case of the flash crash one study documents that HFT exacerbated volatility. It should be noted that empirical research is restricted by a lack of easily accessible and reliable data on HFT activities and market sizing. As of today, it is nearly impossible for researchers (and regulators) to identify exactly on an order-by-order basis whether the respective action can be allocated to HFT operations. Here, further research – ideally in cooperation with HFT entities – is highly desirable.*

In contrast to internalization or dark pool trading, HFT market making strategies face relevant adverse selection costs as they provide liquidity on lit markets without knowing their counterparties. *HFT market makers face the traditional problems of market makers concerning adverse selection costs and have to manage, minimize and compensate their losses of trading against informed order flow. In contrast to internalization systems or dark venues in the OTC space, where banks and brokers in their role as market access intermediaries know the identity of their counterparty and are able to "cream skim" uninformed order flow, HFTs are not informed on the toxicity of their counterparts. Therefore, HFT market makers provide an important function for market quality in supervised and regulated trading venues. Highly automated trading strategies carried out in the OTC space create potential issues for fairness and price discovery.*

Any assessment of HFT based strategies has to take a functional rather than an institutional approach. *HFT is applied by different groups of sophisticated market players from top-tier investment banks to specialized proprietary trading boutiques. Any regulatory approach focusing on specialized players alone risks(i) to undermine a level playing field and (ii) exclude a relevant part of HFT based strategies. In order to manage systemic risk adequately, supervisory authorities have to consider all market participants using automated trading techniques. In this context, it has to be taken into account that the separation of investment banks' proprietary trading operations increases the share of entities that would not be subject to registration as an investment firm, due to the current MiFID Article 2.1 (d) exemptions.*

The high penetration of HFT based strategies underscores the dependency of players in today's financial markets on reliable and thoroughly supervised technology. *Therefore: (i) entities running HFT based strategies need to establish sophisticated risk management tools and operational safeguards and have to be able to demonstrate that they are in full control of their algorithms at any time, e.g. by logging and recording algorithms input and output parameters for supervisory investigations and back testing. (ii) Market operators as well as clearing & settlement organizations have to be able to handle peak volumes and have to be capable of protecting themselves against technical failures in members' algorithms, e.g. by requiring that a human trader responsible for the algorithm is always available during trading hours. (iii) Regulators need a full picture of potential systemic risks triggered by HFT, require people with specific skills and regulatory tools to assess trading algorithms and their functionality, e.g. to be enabled for near-time reactions and rapid investigations in case of market stress.*

Any regulatory interventions in Europe should try to preserve the benefits of HFT while mitigating the risks as far as possible. *The fragmentation of liquidity triggered by MiFID has led to a structural break and prepared the ground for HFT strategies that were not profitable in the pre-MiFID* *environment. However, these changes have reduced both explicit and implicit trading costs and improved market quality in European lit equity markets. Regulatory interventions should attempt to improve overall market quality, resilience as well as robustness in the given, technology-driven environment by assuring that ...*

1. ***... a diversity of trading strategies prevails and that artificial systemic risks are prevented****. Based on a clear functional approach in the assessment of HFT, any undue regulatory burdens for smaller players should be avoided. Furthermore, it is key to prevent any systemic risks by an "equalization" of algorithms that is triggered by a need for HFTs to create largely similar algorithms in order to be compliant with regulatory requirements. It is vital for our financial markets that multiple players of different size, with diverse business models and with different strategies are able to compete.*

2. ***... economic rationale rather than obligations drive the willingness of traders to act as liquidity providers****. HFT quoting obligations are in sharp contrast to the business model of HFTs that relies on minimizing risk, keeping positions for shortest periods and staying mostly flat. Therefore, a quoting obligation and the resulting shut-down of HFT strategies would likely reduce market liquidity instead of improving it. The history of financial markets shows that in times of extreme market stress even designated liquidity providers prioritize sanctions for not fulfilling their obligations over bankruptcy. The key challenge both for regulators and market operators is the design of the right economic incentives rather than imposing obligations/fines that drive liquidity providers temporarily or completely out of markets. The incentives should be based on the respective contribution to market liquidity of market makers independent of whether they are designated or voluntary liquidity providers.*

3. ***... co-location and proximity services are implemented on a level playing field****. In contrast to floor trading (where physical presence and physical strength influences access to deals) or to remote access (where the distance to the location of an electronic trading venue's backend influences the round trip latencies in order execution), HFT, set up in a fair and non-discriminatory co- location environment (especially concerning pricing), assures equality in the access to market data feeds and to the main matching engine. This fairness also has to relate to the provision of co-location or proximity services as such: all market participants have to demonstrate that their providers are able to warrant the physical and operational integrity of their engines independent of the nature of the hosting entity, i.e. Regulated Market, MTF or Network Services Provider.*

4. ***... volatility safeguards are aligned among European trading venues, reflect the HFT reality and ensure that all investors are able to react in times of market stress****. Although the flash crash is a U.S. phenomenon, market operators in Europe have to rethink their safeguards in a fragmented high-speed environment. Extreme market movements should trigger aligned pan-European circuit breakers that enable even retail investors to react and to consider how to position them-*

selves with new orders during a general market halt. This study provides some ideas how this can be put into operation.

The market relevance of HFT requires supervision but also transparency and open communication to assure confidence and trust in securities markets. *With a market share of HFT undoubtedly above one third of trading volume in major markets, it is necessary to enable regulators to assess the robustness and reliability of HFT systems and risk management operations. Given the public sensitivity to innovations in the financial sector after the crisis, it is furthermore the responsibility of entities applying HFT to proactively communicate on their internal safeguards and risk management mechanisms. One has to accept that HFTs cannot publicly release their intellectual property rights and the core of their business models, i.e. the mechanisms of their algorithms and operations. But the observable unwillingness by lot of entities to interact with the public, the media or with other market participants as well as an appearance and behavior shrouded in mystery is not a means to generate trust – especially in the aftermath of the flash crash on May 6, 2010. HFT entities act in their own interest by contributing to an environment where objectivity rather than perception dominates the center of the debate: they have to actively draw attention to the fact that they are an evolution of modern securities markets, supply substantial liquidity and contribute to price discovery for the benefit of all market participants.*

1.11.4 Testing for local randomness

The following ideas are taken from Aldridge (2009)[Chap. 7]

- Non-parametric runs test: Let N_+ denote the number of periods with a positive gain, and N_- the number of periods with negative gain. Let N denote the number of runs (e.g., +++——++- has 4 runs). The Wald–Wolfowitz runs test states that if the sample is really random we should expect a number of runs whose mean and variance are

$$\begin{aligned} \mu_0 &= \tfrac{2N_+N_-}{N_++N_-}+1 \\ \sigma_0^2 &= \tfrac{(\mu_0-1)(\mu_0-2)}{N_++N_--1} \end{aligned} \tag{1.350}$$

  We may calculate the Z-score corresponding to our observation

$$Z = \frac{|N-\mu_0|-0.5}{\sigma_0} \tag{1.351}$$

  and use the Gaussian tables to determine the p-value that the actual number of runs is really random.

- Random-walk test: The idea of this test by Lo and MacKinlay is to check whether the returns follow

$$\log R[n] = \mu + w[n] \tag{1.352}$$

  The idea is that if prices measured with a sampling T_s are random, then measured with $2T_s$ should also be random. Let us consider $2N$ samples.

The test estimates

$$\begin{aligned}\hat{\mu} &= \frac{1}{2N}\sum_{n=0}^{2N-1}\log R[n] \\ \hat{\sigma}_a^2 &= \frac{1}{2N}\sum_{n=0}^{2N-1}(\log R[n]-\hat{\mu})^2 \\ \hat{\sigma}_b^2 &= \frac{1}{2N}\sum_{k=0}^{N-1}(\log R[2k+1]-\log R[2k]-2\hat{\mu})^2\end{aligned} \quad (1.353)$$

The statistic of the test is

$$J_r = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2} - 1 \quad (1.354)$$

Under the null hypothesis

$$\sqrt{N}J_r \sim N(0,1) \quad (1.355)$$

- Autoregression tests: If the market is efficient, then there cannot be any dependence of the current return on previous returns. Two models are proposed: the unrestricted one

$$R_i[n] = R_0 + \sum_{k=1}^{P} a_k R_i[n-k] + w[n] \quad (1.356)$$

and the restricted

$$R_i[n] = R_0 + w[n] \quad (1.357)$$

The Market inefficiency is defined as a function of the R^2 of both fits

$$\eta = 1 - \frac{R^2_{restricted}}{R^2_{unrestricted}} \quad (1.358)$$

- Martingale tests: If the market is efficient and information is automatically incorporated into the price, then prices should behave as a martingale. Martingale tests are rather involved (Escanciano and Lobato, 2009). There are three families: 1) tests based on the finite past of the time series (e.g. Box-Pierce Portmanteau Q_p test for linear and nonlinear relationships of the time series with itself; the Variance Ratio, although this test is sensitive to sign compensation in the autocorrelation function); 2) tests based on the infinite past of the time series (these tests are performed in Fourier space and they use calculations on the Periodogram as statistics, e.g., Kolmogorov-Smirnov, Cramer von Mises or Lobato and Velasco); 3) tests based on nonlinear measures of dependence. The evidence found by Escanciano and Lobato (2009) for currency exchange data favours the hypothesis that exchange rates behave as a martingale at the daily and weekly level.

1.11.5 Testing for local changes

Let us assume that a data stream is generated according to some model 0, and that at some point, it switches to be generated from a different model 1. The change detection algorithms aim at detecting such changes. Ho and

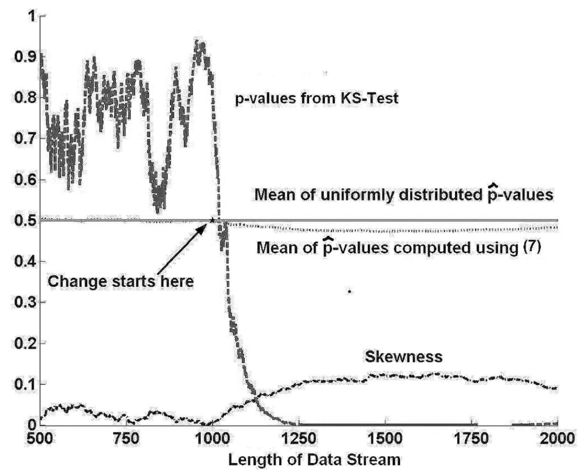


Figure 1.34: Kolmogorov-Smirnov test on the uniform distribution of $\hat{p}[i]$.

Wechsler (2010) proposes one of such algorithms based on a martingale and the assumption that within the data generated by a single model, data points are interchangeable. Given any sequence of incoming vectors, $\mathbf{x}_i$, we may construct a martingale based on them. For doing so, we calculate the strangeness of each vector, s_i, according to some criterion (see Sec. 1.10.2). The randomized power martingale, of parameter $\epsilon \in [0, 1]$, with n samples is constructed as

$$M_\epsilon[n] = \prod_{i=1}^{n} (\epsilon \hat{p}^{\epsilon-1}[i]) \tag{1.359}$$

where $\hat{p}[i]$ is the percentile of the strangeness of the sample i with respect to all its previous samples. The exchangeability property states that within the samples generated by the same model, the order of observation can be arbitrarily permuted, and consequently, $\hat{p}[i]$ behaves as sequence of independent, and identically distributed (IID) observations uniformly distributed in the $[0, 1]$ interval. We may analyze the p-value of a Kolmogorov-Smirnov test checking the uniform distribution of $\hat{p}[i]$ (see Fig. 1.34). This test takes between 100-200 samples to detect the change.

Alternatively, we may use a test based on the maximum value of the martingale. If there is no change in the model, it should be $M_\epsilon[n] < \lambda$. λ is chosen through Doob's maximal inequality:

$$\Pr\left\{\max_{k \leq n} M_\epsilon[k] \geq \lambda\right\} \leq \frac{1}{\lambda} \tag{1.360}$$

This gives an upper bound of the Type I error. If one wants $\alpha = \frac{1}{\lambda} = 0.05$, then we would reject the hypothesis that the data is generated from the same model all the time if $M_\epsilon[n] \geq 20$. Normally, $M_\epsilon[n]$ has very small values. However, they grow when the model changes as shown in Fig. 1.35. They choose λ between 6 and 20 for the detection with strangeness based on Gaussian SVMs.

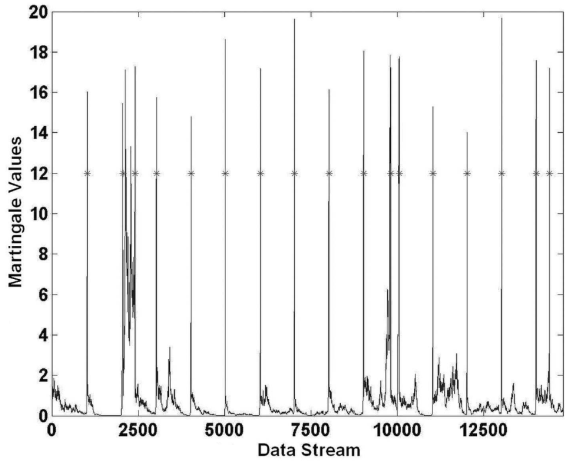


Figure 1.35: Example values of the martingale $M_\epsilon[n]$.

1.11.6 Trading on market microstructure

Inventory models: Liquidity provision

One way of exploiting the microstructure is through the bid-ask spread. This is what liquidity providers do. The size of the spread can be understood through the Gambler's Ruin Problem. The probability of ruin in the long run is (Aldridge, 2009)[Chap. 10]

$$\Pr\{Ruin\} = \left(\frac{\Pr\{Loss\}Loss}{\Pr\{Gain\}Gain}\right)^{InitialWealth} \tag{1.361}$$

The liquidity provider buys or sells 1 share, so $Gain = Price_{ask}$ and $Loss = Price_{bid}$. The probability of lossing or gaining a share is the probability of a buyer or a seller arriving to the market, λ_{buyer} and λ_{seller} (assumed to be Poisson processes). The initial wealth is the initial number of shares. The probability that the market maker runs out of cash is

$$\Pr\{Ruin\} = \left(\frac{\lambda_{seller}Price_{bid}}{\lambda_{buyer}Price_{ask}}\right)^{InitialWealth} \tag{1.362}$$

To remain in business, he needs that

$$\lambda_{seller}Price_{bid} < \lambda_{buyer}Price_{ask} \tag{1.363}$$

The bid-ask spread is larger in illiquid assets, it is also larger if the rate of buyers and sellers are balanced.

Information models

This strategy exploits the information assymetry between well-informed and uninformed traders. This assymetry can be measured from the bid-ask spread (Aldridge, 2009)[Chap. 11], the agressiveness of orders, the order flow. An interesting approach to the problem is the following. Assume that α is the probabilty of occurrance of an event having an impact on the price of an equity. Well-informed traders know that the impact will be positive on the price with pobability δ and negative on the price with probability $1-\delta$. So they place buy ($\delta > 0.5$) or sell ($\delta < 0.5$) trades according to their knowledge at a rate μ. Uninformed traders continue to place trade orders on both sides (buy and sell) at a rate ω. The probability that informed trading is taking place is

$$\Pr\{InformedTrading\} = \frac{\alpha\mu}{\alpha\mu + 2\omega} \tag{1.364}$$

Event arbitrage

This strategy places orders according to the expected market reaction to an event (global or national economy, industry specific, company specific, ...) (Aldridge, 2009)[Chap. 12]. For this strategy we need to analyze historical data with similar events. Once the previous events are identified we may perform:

- Directional analysis: in the periods after the event, does the price consistently move up (or down)? This can be analyzed through a test on the proportion of periods with a positive value, and check if it is significantly different from 0.5.
- Point analysis: what is the price of equilibrium after the event? This is done through a regression similar to the CAPM model.

Statistical arbitrage

We may identify pairs of assets whose historical difference $S_i[n] - S_j[n]$ is rather stable and trade in the compesatory direction when their difference is significantly away from the nominal value. A similar idea can be performed with highly correlated assets. Here go some generic strategies (Aldridge, 2009)[Chap. 13]:

- Equity classes of the same issuer: Since it is the same company behind, both equities will be rather correlated.
- Market neutral arbitrage: We may find two equities whose β of the CAPM model are rather similar. Then, we can be long on one (the one with larger α) and short on the other, on average the losses of one will be compensated by the gains of the other. The difference in α between the two equities must be significant so that it covers the trading costs and results in some profit.
- Liquidity arbitrage: Small liquidity implies higher returns. We may identify equities that are more sensitive to this effect. We may do so by fitting

  $$R_i^e[n+1] = \theta + \beta R_i[n] + \gamma \mathrm{sign}(R_i^e[n]) V_i[n] + \epsilon[n+1] \tag{1.365}$$

  $R_i^e[n]$ is the excess return ($R_i^e[n] = R_i[n] - R_M[n]$). The assets with larger γ are more sensitive to volume variations.
- Large-to-small information spillovers: The market of small companies (<1B$) is less efficient than that of medium (<10B$) or large companies (>10B$). The reason is that they are not attractive to large investors because they

would significantly affect the price with their trades. Small investors, with low-technology, are the ones investing in this market. News propagate much more slowly in this market.

1.11.7 Trading pairs

Elliott et al. (2005) and Chen et al. (2012) present two different strategies for trading on pairs of assets. The first one is based on regression while the second is based on Kalman filters (see Sec. 1.5.1). In their backtesting of the strategies they have an annualized return of 9%, volatility of 7.6%, Sharpe ratio of 1.14, and maximum drawdown of 5.6%.

The spread between the returns of the two assets is assumed to be a mean reverting Ornstein-Uhlenbeck process characterized by

$$dX = \theta(\mu - X)dt + \sigma dW \tag{1.366}$$

μ is the "equilibrium" spread, and θ determines the speed at which the process tends to equilibrium. The mean and variance of this process are

$$\begin{aligned} \mathrm{E}\{X\} &= \mu \\ \mathrm{Var}\{X\} &= \frac{\sigma^2}{2\theta} \end{aligned} \tag{1.367}$$

Calibration of the spread parameters

We need to determine the parameters of the Ornstein-Uhlenbeck process. This may be rather complicated (Dietz and Kutoyants, 2003; Fasen, 2013), although some easy starting point is the following. A discretization of the differential equation above gives

$$X[n+1] = X[n]e^{-\theta T_s} + \mu(1 - e^{-\theta T_s}) + \sigma\sqrt{\frac{1 - e^{-2\theta T_s}}{2\theta}}w[n] \tag{1.368}$$

This equation is of the form

$$X[n+1] = aX[n] + b + w[n] \tag{1.369}$$

If we fit (robust fitting is recommended) this line on the observed data, then we may recover the spread parameters as

$$\begin{aligned} \theta &= -\frac{\log a}{T_s} \\ \mu &= \frac{b}{1-a} \\ \sigma &= \hat{\sigma}_w\sqrt{\frac{-2\log a}{T_s(1-a^2)}} \end{aligned} \tag{1.370}$$

Linear regression strategy

Let the daily returns of two financial products satisfy the stochastic differential equation:

$$\frac{\frac{dS_1}{dt}}{S_1} - \beta\frac{\frac{dS_2}{dt}}{S_2} = dX \tag{1.371}$$

where X is the random spread. The weights of the portfolio are:

$$\begin{aligned} w_1 &= \frac{1}{1+\beta} \\ w_2 &= \frac{\beta}{1+\beta} \end{aligned} \tag{1.372}$$

We will go long on one of the assets and short on the other. Due to the rebalancing costs, it is better not to change β too frequently (at least, keep it fixed for 5-10 periods) so that the trasaction costs do not grow too much.

Let us define the returns as

$$R_i[n] = \frac{S_i[n] - S_i[n-1]}{S_i[n-1]} \tag{1.373}$$

Then, β is estimated from the regression

$$R_1[n] = \beta R_2[n] + \epsilon[n] \tag{1.374}$$

Once β is fitted (in a moving window or all the data available), then we estimate the spread time series as

$$X[n] = \sum_{k=1}^{n} \epsilon[k] \tag{1.375}$$

Then, we can use the model estimation of the Ornstein-Uhlenbeck process in the previous section. Finally, we compute the Z-score of the spread time series

$$z[n] = \frac{X[n] - \mu_X}{\sigma_X} \tag{1.376}$$

The trading rules are

- Open long if $z[n] < -z_1$
- Open short if $z[n] > z_2$
- Close long if $z[n] > -z_3$
- Close short if $z[n] < z_4$

Kalman filter strategy

In this second approach, the spread is assumed to be the state of a Kalman filter (Eq. 1.369) and then it is observed

$$\begin{array}{rcl} s[n+1] &=& as[n] + b + w[n] \\ x[n] &=& s[n] + u[n] \end{array} \tag{1.377}$$

The system parameters are $a, b, \sigma_w^2, \sigma_u^2$ which are determined by Expectation Maximization (see Chen et al. (2012) for details).

The trading rules are the following

- Open long and short $x[n] > s[n] + \delta_1$, we expect that the current spread will shrink.
- Open long and short $x[n] < s[n] - \delta_1$, we expect that the current spread will expand.
- Close position if $|x[n] - s[n]| \leq \delta_2$, for some predetermined value δ_2.

δ_1 is a value that includes transaction costs and our profit.

Javaheri et al. (2003) gives several examples of Kalman filters and particle filters applied to the prediction of commodity prices and volatility. They show that particle filters perform better than the Kalman Filter and the Extended Kalman filter. Platania and Rogers (2004) shows how to use particle filters to model the bid-ask intraday prices.

Chapter 2

Mathematical tools

2.1 Lagrangean multipliers method

The Lagrangean method can be applied to constrained optimization problems in which the constraints imply equalities:

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmax} & f(\mathbf{p}) \\ & & \text{s.t.} & \mathbf{g}(\mathbf{p}) = \mathbf{0} \end{array} \tag{2.1}$$

The Lagrangean function related to the problem is

$$L(\mathbf{p}) = f(\mathbf{p}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{p}) = f(\mathbf{p}) + \sum_{i=1}^{k} \lambda_i g_i(\mathbf{p}) \tag{2.2}$$

The Lagrange theorem states that any solution of the problem above must satisfy the equation system

$$\begin{array}{rcl} D_{\mathbf{p}} L & = & D_{\mathbf{p}} f(\mathbf{p}) + \boldsymbol{\lambda}^T (D_{\mathbf{p}} \mathbf{g}(\mathbf{p})) = \mathbf{0} \\ D_{\boldsymbol{\lambda}} L & = & \mathbf{g}(\mathbf{p}) = \mathbf{0} \end{array} \tag{2.3}$$

The converse is not true, any solution of the equation system above does not need to be a solution of the optimization problem. Consequently, we must verify that the solutions of the equation system are actually solutions of the optimization problem. The equation system only has a solution if

$$\text{rank}\{D_{\mathbf{p}} \mathbf{g}(\mathbf{p})\} = k \tag{2.4}$$

where $\mathbf{g}(\mathbf{p})$ is a vector function with k components ($\mathbf{g}(\mathbf{p}) : \mathbb{R}^n \to \mathbb{R}^k$). Remind that the vector derivative of a vector-valued function is

$$D_{\mathbf{p}} \mathbf{g}(\mathbf{p}) = \begin{pmatrix} \frac{\partial g_1}{\partial p_1} & \frac{\partial g_1}{\partial p_2} & \dots & \frac{\partial g_1}{\partial p_n} \\ \frac{\partial g_2}{\partial p_1} & \frac{\partial g_2}{\partial p_2} & \dots & \frac{\partial g_2}{\partial p_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_k}{\partial p_1} & \frac{\partial g_k}{\partial p_2} & \dots & \frac{\partial g_k}{\partial p_n} \end{pmatrix} (\mathbf{p}) \tag{2.5}$$

To verify if it is a maximum, let us define the set $\mathcal{Z}(\mathbf{p}^*)$ as

$$\mathcal{Z}(\mathbf{p}^*) = \{\mathbf{z} \in \mathbb{R}^n | (D_{\mathbf{p}} \mathbf{g}(\mathbf{p}^*))\mathbf{z} = \mathbf{0}\} \tag{2.6}$$

If f has a maximum at $\mathbf{p}^*$, then $\mathbf{z}^T(D^2_{\mathbf{p}}L(\mathbf{p}^*))\mathbf{z} \leq 0$ for all $\mathbf{z} \in \mathcal{Z}(\mathbf{p}^*)$. The Hessian of the Lagrangean function is given by

$$D^2_{\mathbf{p}}L(\mathbf{p}) = D^2_{\mathbf{p}}f(\mathbf{p}) + \sum_{i=1}^{k} \lambda_i D^2_{\mathbf{p}} g_i(\mathbf{p}) \tag{2.7}$$

Remind that the Hessian of a scalar function is given by

$$D^2_{\mathbf{p}}f(\mathbf{p}) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 p_1} & \frac{\partial^2 f}{\partial p_1 \partial p_2} & \cdots & \frac{\partial^2 f}{\partial p_1 \partial p_n} \\ \frac{\partial^2 f}{\partial p_2 \partial p_1} & \frac{\partial^2 f}{\partial^2 p_2} & \cdots & \frac{\partial^2 f}{\partial p_2 \partial p_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial p_n \partial p_1} & \frac{\partial^2 f}{\partial p_n \partial p_2} & \cdots & \frac{\partial^2 f}{\partial^2 p_n} \end{pmatrix} (\mathbf{p}) \tag{2.8}$$

2.2 Karush-Kuhn-Tucker conditions

Karush-Kuhn-Tucker conditions are the extension of Lagrange multipliers method to constraints with inequalities

$$\begin{array}{rcll} \mathbf{p}^* & = & \text{argmax} & f(\mathbf{p}) \\ & & \text{s.t.} & \mathbf{h}(\mathbf{p}) \geq \mathbf{0} \end{array} \tag{2.9}$$

A condition is said to be effective at a point $\mathbf{p}$ if at that point the equality holds ($h_i(\mathbf{p}) = 0$). Presume $\mathbf{p}^*$ is a local maximum of the function f satisfying the $\mathbf{h}$ conditions (f and $\mathbf{h}$ are assumed to be continuous functions). Let us denote by E the number of effective conditions at $\mathbf{p}^*$ and by $\mathbf{h}_E$ the corresponding vector valued function. Assume that

$$\text{rank}\{D_{\mathbf{p}}\mathbf{h}_E(\mathbf{p}^*)\} = E \tag{2.10}$$

Then, the Karush-Kuhn-Tucker theorem states that there exists a vector $\boldsymbol{\mu}$ such that

$$\begin{array}{c} \forall i \quad \mu_i \geq 0 \text{ and } \mu_i h_i(\mathbf{p}^*) = 0 \\ D_{\mathbf{p}}f(\mathbf{p}^*) + \boldsymbol{\mu}^T(D_{\mathbf{p}}\mathbf{h}(\mathbf{p}^*)) = \mathbf{0} \end{array} \tag{2.11}$$

2.3 Estimation of covariance matrices

The covariance matrix between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\text{Cov}\{X, Y\} = \text{E}\{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})^T\} \tag{2.12}$$

With historical data we may estimate the ij-th entry of this matrix as

$$\hat{\sigma}_{ij} = \frac{1}{N-1} \sum_{n=1}^{N} (X_i[n] - \overline{X_i})(Y_j[n] - \overline{Y_j}) \tag{2.13}$$

This estimation gives the same weight to all samples, disregarding how old they are. Instead, we may give more weight to newer samples (Barra, 2007)[Chap. 3]

$$\hat{\sigma}_{ij} = \frac{1}{\sum\limits_{n=1}^{N} w[n]} \sum_{n=1}^{N} w[n](X_i[n] - \overline{X_i})(Y_j[n] - \overline{Y_j}) \tag{2.14}$$

with

$$w[n] = \lambda^{\frac{N-n}{N_{1/2}}} \tag{2.15}$$

where $N_{1/2}$ is called the half-life period.

2.4 Covariance matrix scaling

We may estimate the covariance matrix to learn the structure of the assets. But on a daily operation we may adapt the covariance matrix to the current market conditions. This can be done by locally estimating the variance of each asset and scaling the covariance matrix accordingly. One possibility is the time weighted estimate above

$$\hat{\sigma}_i^2 = \frac{1}{\sum\limits_{n=1}^{N} w[n]} \sum_{n=1}^{N} w[n](X_i[n] - \overline{X_i})^2 \tag{2.16}$$

The DEWIV model (Daily Exponentially Weighted Index Volatility) belongs to this family of models. Another possibility is to use a GARCH model (Generalized Autoregressive conditional heteroskedasticity). The GARCH(Q,P) model states that the residuals $\epsilon[n]$ are generated as

$$\epsilon[n] = \sigma[n]z[n] \tag{2.17}$$

where $z[n]$ is a white noise signal and

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{Q} a_k \epsilon^2[n-k] + \sum_{k=1}^{P} b_k \sigma^2[n-k] \tag{2.18}$$

Pure GARCH models fail to capture relatively higher volatility following periods of below-normal returns (Barra, 2007)[Chap. 3]. This can be solved by extending the GARCH model with terms depending on $\epsilon[n]$

$$\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^{Q} a_k \epsilon^2[n-k] + \sum_{k=1}^{P} b_k \sigma^2[n-k] + \sum_{k=1}^{R} a_k \epsilon[n-k] \tag{2.19}$$

To scale the covariance matrix, we can multiply a diagonal matrix whose entries are the ratio between the new and old standard deviations of each variable

$$S = \begin{pmatrix} \frac{\sigma_1^{new}}{\sigma_1^{old}} & 0 & ... & 0 \\ 0 & \frac{\sigma_2^{new}}{\sigma_2^{old}} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \frac{\sigma_M^{new}}{\sigma_M^{old}} \end{pmatrix} \tag{2.20}$$

being M the number of variables, and

$$\Sigma_R^{new} = S\Sigma_R^{old} S \tag{2.21}$$

Bibliography

Ackermann, C., McEnally, R., Ravenscraft, D., 1999. The performance of hedge funds: Risk, return, and incentives. Journal of Finance , 833–874.

Aldridge, I., 2009. High-frequency trading: a practical guide to algorithmic strategies and trading systems. Vol. 459. John Wiley and Sons.

Baillie, R. T., Chung, C.-F., Tieslau, M. A., 1996. Analysing inflation by the fractionally integrated arfima-garch model. Journal of applied econometrics 11 (1), 23–40.

Barra, M., 2007. Barra Risk Model.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13 (1), 1–42.

Bhardwaj, G., Swanson, N. R., 2006. An empirical investigation of the usefulness of arfima models for predicting macroeconomic and financial time series. Journal of Econometrics 131 (1), 539–578.

Bondt, W. F., Thaler, R., 1985. Does the stock market overreact? The Journal of finance 40 (3), 793–805.

Bouchaud, J.-P., Gefen, Y., Potters, M., Wyart, M., 2004. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. Quantitative Finance 4 (2), 176–190.

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of finance , 1731–1764.

Brown, S. J., Goetzmann, W. N., 1995. Performance persistence. Journal of finance , 679–698.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of finance 52 (1), 57–82.

Chen, Y., Ren, W., Lu, X., 2012. Machine learning in pairs trading strategies. Tech. rep., Stanford University.

Chen, Z., 2003. Bayesian filtering: From kalman filters to particle filters, and beyond. Statistics 182 (1), 1–69.

Chlistalla, M., Speyer, B., Kaiser, S., Mayer, T., 2011. High-frequency trading. Deutsche Bank Research 7.

Conrad, J., Kaul, G., 1998. An anatomy of trading strategies. Review of Financial studies 11 (3), 489–519.

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

Cover, T. M., Thomas, J. A., 2012. Elements of information theory. John Wiley & Sons.

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

Dhar, R., Kumar, A., 2001. A non-random walk down the main street: Impact of price trends on trading decisions of individual investors. Yale International Center for Finance. Working Paper .

Dietz, H. M., Kutoyants, Y. A., 2003. Parameter estimation for some non-recurrent solutions of sde. Statistics & Decisions/International mathematical Journal for stochastic methods and models 21 (1/2003), 29–46.

Do, C.-T., Douzal-Chouakria, A., Marié, S., Rombaut, M., 2015. Multiple metric learning for large margin knn classification of time series. In: EUSIPCO 2015.

Eisenbeiß, M., Kauermann, G., Semmler, W., 2007. Estimating beta-coefficients of german stock data: a non-parametric approach. The European Journal of Finance 13 (6), 503–522.

Elliott, R. J., Van Der Hoek, J., Malcolm, W. P., 2005. Pairs trading. Quantitative Finance 5 (3), 271–276.

Elton, E. J., Gruber, M. J., Brown, S. J., Goetzmann, W. N., 2003. Modern portfolio theory and investment analysis. John Wiley and Sons, Inc.

Elton, E. J., Gruber, M. J., Das, S., Hlavka, M., 1993. Efficiency with costly information: A reinterpretation of evidence from managed portfolios. Review of Financial studies 6 (1), 1–22.

Escanciano, J. C., Lobato, I. N., 2009. Testing the martingale hypothesis. Palgrave Hand-book of Econometrics, Palgrave, MacMillan .

Fasen, V., 2013. Statistical estimation of multivariate ornstein–uhlenbeck processes and applications to co-integration. Journal of Econometrics 172 (2), 325–337.

Gomber, P., Arndt, B., Lutat, M., Uhle, T., 2011. High-frequency trading. Available at SSRN 1858626 .

Greenwald, D. L., Lettau, M., Ludvigson, S. C., 2014. Origins of stock market fluctuations. Tech. rep., National Bureau of Economic Research.

Gruber, M. J., 1996. Another puzzle: The growth in actively managed mutual funds. The journal of finance 51 (3), 783–810.

Györfi, L., Ottucsák, G., Urbán, A., 2012. Machine Learning for Financial Engineering. World scientific, Ch. Empirical log-optimal portfolio selections: a survey, pp. 81–118.

Ho, S.-S., Wechsler, H., 2010. A martingale framework for detecting changes in data streams by testing exchangeability. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (12), 2113–2127.

Hommes, C. H., 2006. Heterogeneous agent models in economics and finance. Handbook of computational economics 2, 1109–1186.

Horváth, M., Urbán, A., 2012. Machine Learning for Financial Engineering. World scientific, Ch. Growth-optimal portfolio selection with short selling and leverage, pp. 153–178.

Hsieh, K., Ritchken, P., 2005. An empirical comparison of garch option pricing models. Review of derivatives research 8 (3), 129–150.

Javaheri, A., Lautier, D., Galli, A., 2003. Filtering in finance. Wilmott 2003, 67–83.

Kauermann, G., Krivobokova, T., Semmler, W., 2011. Filtering time series with penalized splines. Studies in Nonlinear Dynamics & Econometrics 15 (2).

Lo, A. W., MacKinlay, A. C., 2011. A non-random walk down Wall Street. Princeton University Press.

Lux, T., 1995. Herd behaviour, bubbles and crashes. The economic journal , 881–896.

Markowitz, H., 1959. Portfolio Selection: Efficient Diversification of Investments. Wiley.

Nicolato, E., Venardos, E., 2003. Option pricing in stochastic volatility models of the ornstein-uhlenbeck type. Mathematical finance 13 (4), 445–466.

Norstad, J., 1999a. An introduction to portfolio theory.

Norstad, J., 1999b. An introduction to utility theory.

Norstad, J., 2002a. Portfolio optimization. part 1-unconstrained portfolios.

Norstad, J., 2002b. Portfolio optimization. part 2-constrained portfolios.

Norstad, J., 2005. Random walks.

Platania, A., Rogers, L., 2004. Particle filtering in high frequency data. preprint .

Saito, Y., Mitsui, T., 1993. Simulation of stochastic differential equations. Annals of the Institute of Statistical Mathematics 45 (3), 419–432.

Sreedharan, N. J., 2004. A Vector Error Correction Model (VECM) of Stockmarket Returns. University of Tasmania.

Welch, I., 2014. Corporate finance. Welch.

Index