Applications and Uses of Digital Filters in Finance

by
Johan Boissard

A thesis submitted to the
Department of Management, Technology and Economics (d-mtec)
of the Swiss Federal Institute of Technology Zurich (ethz)
in partial fulfillment of the requirements for the degree of

Master of Science in Management, Technology and Economics

Jointly supervised by

Dr. Donnacha Daly — swissQuant Group AG
Dr. Peter Cauwels — Chair of Entrepreunerial Risks, ETH Zurich
Prof. Didier Sornette — Chair of Entrepreunerial Risks, ETH Zurich

December 2012

Abstract

In finance, moving averages are commonly used to reveal trends from time-series. From those trends, information about the market or forecasts are often deduced. However, when the term moving average is used, most of the time it implies the use of the simple moving average (similar weight for all the observations) or the exponentially weighted moving average (weights decrease exponentially).

From a signal processing point of view, moving averages are a subset of digital filters and many were already designed to solve problems in a broad range of fields (e.g. telecommunication, image processing, acoustic, medicine, ...) . Hence, in this study, we investigate the substitution of the ubiquitously used simple and exponentially weighted moving averages (SMA and EWMA) with new moving averages that are borrowed from the signal processing field and try to beat the performance of the regularly used filters.

We first research investing strategies where moving averages are used to extract trends from time-series. After being coupled with a trading rule, those trends recommend a short or long position. We are interested in replacing the typically used EWMA with filters that have a less abrupt decay in weights; rather than have an impulse function that is only convex, an inflection point is introduced so that the filters have desirable characteristics both from the SMA and EWMA. In addition, we also test the use of a filter that has oscillations (positive and negative weights). We measure the performances of the filters with standard indicators and then compare the different filters for different types of markets. Tests are conducted on a portfolio of over 100 securities traded daily including equities, forex and commodities. Results showed that in general the filter are significantly different and all show similar performances. However, under certain circumstances, half windows produce better views than the EWMA.

In a second part, we investigate the computation of the Value-at-Risk (VaR) within the RiskMetrics^{TM} framework. This method was first introduced in 1994 by JP Morgan and is widely used among practitioners. In order to compute the VaR, the method uses two steps: variance forecasting and then finding the quantile from the variance. The variance forecasting is done by applying the EWMA on the squared returns of the time-series of interest. This process is simple since the EWMA needs only one parameter. We test the substitution of the EWMA with other one-parameter filters and assess the difference in performance over a whole portfolio to see if notable improvements can be achieved. In this part, the EWMA turned out to be the best performing filter, even though the other ones are not much worse.

Contents

1. Introduction **6**
- 1.1. Moving Averages . . . 6
- 1.2. Trading Environment . . . 9
- 1.3. Value-at-Risk and the RiskMetrics^{TM} Framework . . . 10

2. Moving averages and digital filters **11**
- 2.1. LTI Systems . . . 11
  - 2.1.1. Factorization of an LTI System . . . 12
  - 2.1.2. Impulse Response of a Filter . . . 12
  - 2.1.3. Finite and Infinite Impulse Response . . . 13
- 2.2. Z-transform . . . 13
  - 2.2.1. LTI System Representation in the Z-plane . . . 13
  - 2.2.2. Normalizing Moving Averages . . . 15
  - 2.2.3. Energy of a filter and System Stability . . . 16
- 2.3. The Discrete Time Fourier Transform - DTFT . . . 17
  - 2.3.1. System Representation . . . 17
- 2.4. Physical Meaning of Filtering Financial Time-series . . . 18
- 2.5. The SMA as a Finite Impulse Response (FIR) filter . . . 19
  - 2.5.1. Parameter of the SMA . . . 19
- 2.6. Infinite Impulse Response (IIR) filters . . . 19
  - 2.6.1. The Exponentially Weighted Moving Average as an IIR filter . . . 22
  - 2.6.2. Cumulative Weights . . . 22
  - 2.6.3. Parameter λ . . . 23
  - 2.6.4. Double Exponentially Weighted Moving Average . . . 25
  - 2.6.5. Triple Exponentially Weighted Moving Average . . . 27
  - 2.6.6. The Resonator - 2 poles IIR filter . . . 27
- 2.7. Windows . . . 31
  - 2.7.1. Rectangular Window . . . 31
  - 2.7.2. Using Windows as Smoothing Moving Averages . . . 31
  - 2.7.3. Half-windows . . . 33
  - 2.7.4. Comparison of half-window filters with the SMA . . . 34

3. Trading Strategies using Moving Averages **37**
- 3.1. Trading Strategies . . . 38
  - 3.1.1. Relative Strength Index (RSI) . . . 39
  - 3.1.2. Trading Rules . . . 40
- 3.2. Measuring Performance . . . 42
  - 3.2.1. Return computation . . . 42
  - 3.2.2. Computing the profit and loss . . . 42
  - 3.2.3. Sharpe Ratio . . . 42
  - 3.2.4. Directional quality . . . 43

3.2.5. Random Strategies . . . 44
3.2.6. Statistical Tests . . . 45
3.3. Backtesting Methodology . . . 45
3.3.1. Chosen Trading Rules . . . 45
3.3.2. Trading Costs . . . 45
3.3.3. Looking Ahead and In-sample optimization . . . 46
3.3.4. Parameter Sweep . . . 47
3.3.5. Parameter Space and Overfitting . . . 48
3.4. Simulation Results . . . 48
3.4.1. Overview . . . 48
3.4.2. Exploring the Sharpe ratios versus directional Qualities . . . 49
3.4.3. Results obtained with Q_d percentile performance measures . . . 52
3.4.4. Confirming Graphical Observations with Statistical Tests . . . 55
3.4.5. Strategy #5 without transaction cost . . . 56
3.4.6. Mean Trade Per Return and the Problem of Transaction Cost . . . 56
3.4.7. Conclusion . . . 56

4. Substitution of the EWMA with other Moving Averages for the Computation of the Value at Risk 62
4.1. Value at Risk . . . 62
4.2. Risk Metrics Framework . . . 63
4.2.1. Assumptions of the RiskMetrics^{TM} model . . . 63
4.2.2. Variance forecasting . . . 63
4.2.3. New Moving averages investigated . . . 64
4.2.4. Comparison of filters . . . 64
4.2.5. Inferring VaR from the forecasted variance . . . 64
4.2.6. Autocorrelation . . . 67
4.2.7. Comparison with GARCH models . . . 68
4.3. Backtesting Methodology . . . 69
4.4. Results . . . 70

5. Conclusion 78

A. Appendix 80
A.1. Digital Filter Theory: complement . . . 80
A.1.1. Short reference for complex numbers . . . 80
A.1.2. The inverse Z-transform . . . 80
A.1.3. Direct Computation of the Discrete Time Fourier Transform . . . 80
A.1.4. The Discrete Fourier Transform (DFT) . . . 82
A.1.5. Parseval's Theorem . . . 82
A.1.6. Main Types of Filters . . . 82
A.2. Additional filters and derivations . . . 84
A.2.1. Triangular Moving Average (TMA) . . . 84
A.2.2. Generalization of the order of the SMA - the n-SMA . . . 85
A.2.3. Linear Moving Average . . . 85
A.2.4. Generalization of the EWMA to the n-EWMA . . . 86
A.3. Backtesting Framework Structure . . . 87
A.3.1. Settings . . . 87

A.3.2. Tickers Selection 88
A.3.3. Fetch Data and Formatting 88
A.4. Database Structure and Content 88
A.4.1. Portfolio 90
A.4.2. Results 92

1. Introduction

In finance, moving averages are probably one of the most common technical analysis indicator and has been widely used in the last decades. Their major application is to detect trends in signals or compute signal statistics. From those trends, information about the market or forecasts are often deduced. In this thesis, we investigate the substitution of the ubiquitously used SMA and EWMA in finance with other moving averages.

1.1. Moving Averages

Theoretically, there is an infinity of possible moving averages, however practitioners almost exclusively use the simple moving average (SMA) and the exponentially moving average (EWMA). The SMA is the conventional form of a moving average and is defined as

$$y_n = \frac{1}{N}(x_n + x_{n-1} + ... + x_{n-N+1}) = \frac{1}{N}\sum_{k=0}^{N} x_{n-k}. \tag{1.1}$$

where x is the input time-series and y the output time-series. This process is depicted in Figure 1.1 where one can see that the input (The Dow Jones Industrial Average Index settlement prices, DJIA) is largely smoothed by the moving averages and thus reveals the trends of the input signal. Moving averages also always lag the signal. This lag is roughly half the length of the moving average considered and we see in Figure 1.1 that since the length is $N = 100$, the moving average lags the signal by about 50 trading days.

Since the SMA weights all observations equally, it does not take in account the assumption that some past observations are more relevant than others. This leads to weighted moving averages where every considered observation is given a specific weight and are defined as

$$y_n = h_0 x_n + h_1 x_{n-1} + ...h_N x_{n-N+1} = \sum_{k=0}^{N-1} h_k x_{n-k} \tag{1.2}$$

where h_k is the weight associated with the kth observation.

In order to be able to compare the time-series with the the result of the moving average, it is important that the sum of h_k equals one ($\sum_{k=0}^{N} h_k = 1$) and that h_k are non-negative. Moreover, note that N can be taken in the interval $[1, \infty]$. As we will se more formally later, h completely describes a moving average and is called the impulse response. For the SMA, as all weights are constant, we have an impulse response that is $h_k = \frac{1}{N}$ for $k \in [0, N-1]$.

When it comes to do forecasting, a common property of the weighted moving averages is that weights are decreasing as it is often assumed that recent observations are more relevant than older ones to forecast near future values, i.e. h_k are decreasing so that $h_k \geq h_{k+1}$. A famous and widely used example of the latter is the exponentially weighted moving average that is defined as a weighted moving average with exponentially decreasing weights and is written as follows

$$y_n = \lambda y_{n-1} + (1 + \lambda)x_n. \tag{1.3}$$

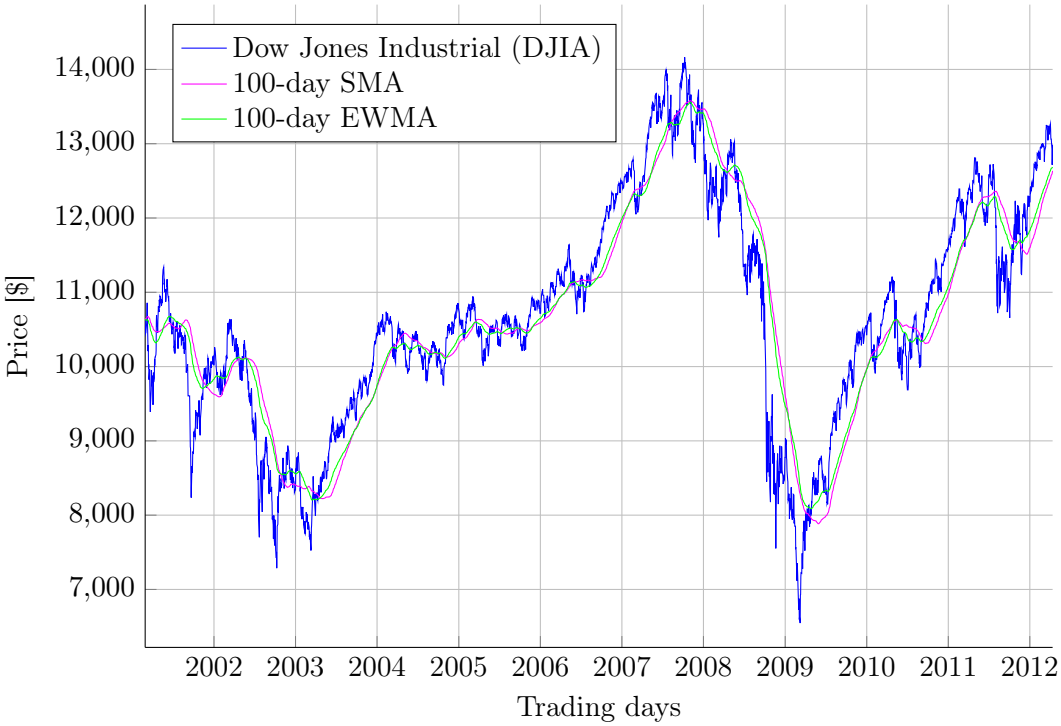


Figure 1.1.: Illustration of the Dow Jones Industrial Index Average (DJIA) settlement prices compared with a 100-day simple moving average

Here λ is the decay rate and can be comprised between 0 and 1; note that a $\lambda > 1$ would lead to a diverging process. For a fair comparison with the SMA, the relation $N = (1+\lambda)/(1-\lambda)$ is used to convert the λ parameter to days (see Section 2.6.1 for the derivation). Unlike the SMA, the EWMA comprises an autoregressive part and when expanded, it can also be written as

$$y_n = (1-\lambda)\sum_{k=0}^{\infty} \lambda^k x_{n-k}. \tag{1.4}$$

From Equation 1.4, we see that the weight associated with the $n-k$th observation is $h_k = (1-\lambda)\lambda^k$. The fact that weights decrease with observations allows the moving average to react quicker than the SMA. This can be seen in Figure 1.1, where the EWMA outputs a signal that is more sensible to local changes.

Nevertheless, the EWMA decays rapidly and is strictly convex; weights in the tail become quickly insignificant in comparison with the first weights; this is illustrated in Figure 1.2. Hence, in this thesis, we investigate new unconventional moving averages that have impulse responses that are constantly decreasing but with the inclusion of an inflection point (see Figure 1.3). These moving averages thus have an impulse response that is somehow flat in the beginning (like the SMA) but still decreasing and then, after the inflection point, decaying faster, like the EWMA. We also tested the so-called double and triple EWMA. Those moving averages are EWMA filters in serial (two or three times). Finally, we also investigated the use of the resonator which is a filter oscillating but nonetheless exponentially decaying.

Hence, in this thesis, we are interested to see if by substituting the commonly used EWMA or SMA with other moving averages, their performances can be beaten in different financial applications. In order to do this, in Chapter 2 we first briefly present some results of the digital

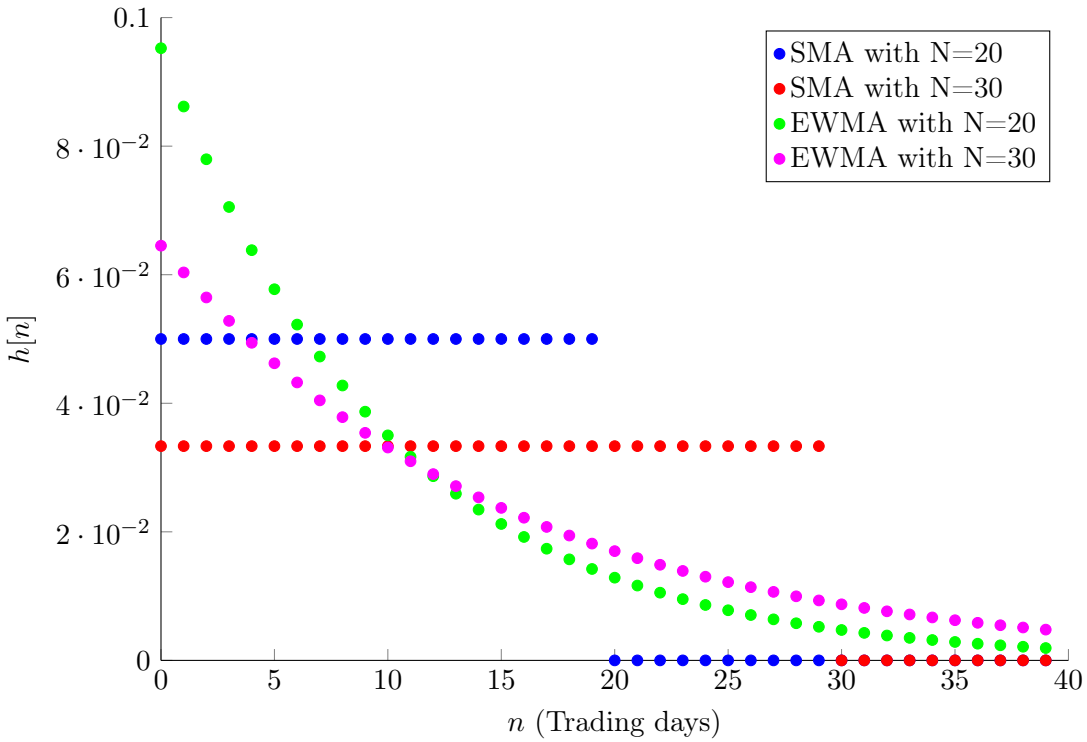


Figure 1.2.: Difference between the SMA and EWMA weights. The SMA has constant weights whereas the EWMA has exponentially decreasing weights. The parameter N is given in day-equivalent (explained in Section 2.6.3).

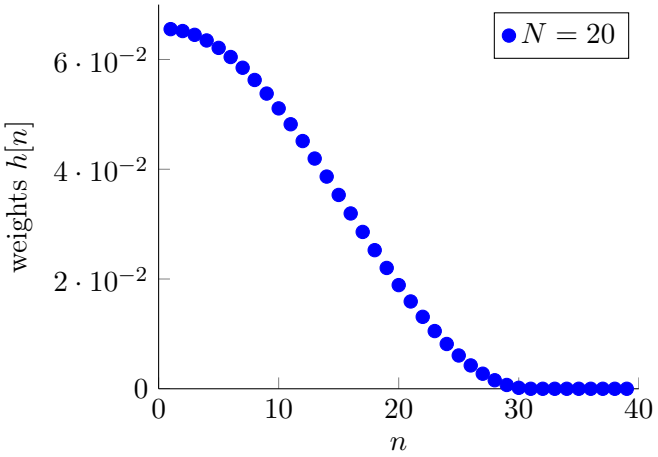


Figure 1.3.: Typical example of a moving average with weights that are constantly decreasing but with the inclusion of an inflection point (reverted logistically shaped). The parameter is $N = 20$ day equivalent.

signal processing theory and introduce the new moving averages.

In order to be able to quantify the performances of the different moving averages, we backtested them in two different financial environments where moving averages have been extensively used in the last 30 years. In Chapter 3, we tested them in a simple trading framework and in Chapter 4, the EWMA is substituted with other moving averages to forecast the daily volatility and then to compute the value-at-risk.

1.2. Trading Environment

In Chapter 3, we use moving averages to extract trends from time-series of daily traded securities. We do this on close prices for the last 10 years (2002 to 2012). After being coupled with a trading rule, those trends recommend a short or long position that form a trading strategy. The performances of the trading strategies are then measured using standard metrics and compared to the performance of the EWMA and SMA. It is important to note that this kind of trading strategy assumes that the market is not totally efficient and that, by filtering out high frequency noise, real market information can be retrieved. This goes against the Efficient Market Hypothesis that states that achieving constant excess returns based on historical information is not possible. However, [Carhart, 1997] introduced the momentum factor implying that markets have cycles as an intrinsic mechanism. Moreover [Clare et al., 2012] and [Hurst et al., 2012] showed that trend-following strategies have better results than traditional buy and hold positions.

Substituting the EWMA or the SMA with other weighted moving averages did not seem very well researched. Beside [Tillson] who used the same kind of approach but only backtested its findings on the NASDAQ, no articles on the matter were found.

It seems, that in this field, most of the research literature rather focuses on the use of adaptive filters such as the Kalman filter [Wesen et al., Nair et al., 2010, Beitollah Akbari Moghaddam and Hassan Haleh] or machine learning techniques including neural networks, support vector machine or the genetic algorithm [Yao and Tan, 2000, Wesen et al., DeBrunner and Charpentier, 2002, Saito, 2007]. Additionally, some scholars also investigate the use of GARCH models to forecast stock prices [Liu, 1999].

1.3. Value-at-Risk and the RiskMetrics™ Framework

In Chapter 4, we investigate the computation of the Value-at-Risk (VaR) within the RiskMetrics framework ™[J.P. Morgan and Reuters, 1996]. In order to compute the VaR, the method uses two steps: variance forecasting and then finding the quantile from the variance. The variance forecasting is done by applying the EWMA on the squared returns of the time-series of interest. Here again, we test the substitution of the EWMA with other one-parameter filters and assess the difference in performance in terms of violations over a whole portfolio to see if notable improvements can be achieved in the accuracy of the VaR value. We haven't found any paper treating the same subject. Scholars rather seem to focus on more advanced variance forecasting methods such as GARCH [Engle, 1982] or multi-fractals [Boisbourdain, 2008].

2. Moving averages and digital filters

It turns out that moving averages are a special case of digital filters, for which reasons we outline in this chapter the main results of the digital signal processing theory. For seminal treatment on the subject, the reader is invited to consult [Oppenheimer and Schafer, 1999].

A digital filter is a system that takes a discrete-time signal input $x[n]$, applies different linear mathematical operations to it and returns another discrete-time signal output $y[n]$, see Figure 2.1 for a typical graphical representation. Most of the time, filters are designed in order to extract specific aspects of the input (e.g. the low-frequency component of the input), to reduce noise or unwanted spectral components, or compress the input signal. In this study, as we are working with financial time-series, we will mostly be interested in extracting persistent features of the input (e.g.strong market cycle signals) and reducing unwanted dynamics (e.g. noise) in order to generate forecasting views.

2.1. LTI Systems

The action of smoothing time-series with moving averages , which weights are stationary, can be seen as a special case of Linear Time Invariant Systems (LTI systems). *Linear* means that the output behaves linearly with respect to the input, or mathematically if the system is denoted by $S\{\cdot\}$ and the input and output discrete-time signals as $x[n]$ and $y[n]$:

$$y[n] = S\{a_1 x_1[n] + a_2 x_2[n]\} = a_1 S\{x_1[n]\} + a_2 S\{x_2[n]\} \tag{2.1}$$

Time invariant means that the system's behavior does not change with time or

$$y[n] = S\{x[n]\} \Rightarrow y[n-k] = S\{x[n-k]\}. \tag{2.2}$$

Discrete LTI systems can always be described in terms of difference equations (weighted sum of delayed outputs equaling a weighted sum of delayed inputs):

$$a_0 y[n] + a_1 y[n-1] + ... + a_N y[n-N+1] = b_0 x[n] + b_1 x[n-1] + ... + b_M x[n-M+1]. \tag{2.3}$$

From Equation 2.3, one can see that (normalizing with $a_0 = 1$) the output $y[n]$ can always be

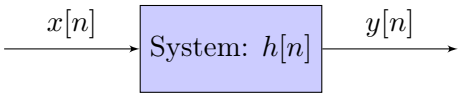


Figure 2.1.: Typical representation of a LTI system where the system $S\{\cdot\}$ is fed with the input $x[n]$ and the output is $y[n]$. The system is fully described by its impulse response $h[n]$.

expressed as

$$y[n] = \underbrace{-\sum_{k=1}^{N-1} a_k y[n-k]}_{\text{AR}} + \underbrace{\sum_{k=0}^{M-1} b_k x[n-k]}_{\text{MA}} \tag{2.4}$$

This is the most straightforward way to compute the output of a filter. In this case the filter is completely described by the coefficients a_k and b_k. The Matlab® function `filter` uses this algorithm to compute the output of a digital filter.

2.1.1. Factorization of an LTI System

Let the Kronecker's delta be defined as

$$\delta[n-k] := \delta_{nk} := \begin{cases} 1 & \text{if } n = k \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

and $*$ be the discrete convolution operator so that

$$x[n] * y[n] = y[n] * x[n] := \sum_{k=-\infty}^{\infty} y[k]x[k-n]. \tag{2.6}$$

Using Equation 2.5 and 2.6, in Equation 2.3, the input $x[n]$ and output $y[n]$ can be factorized so that

$$y[n] * \left(\sum_{k=0}^{N-1} a_k \delta[n-k] \right) = x[n] * \left(\sum_{k=0}^{M-1} b_k \delta[n-k] \right). \tag{2.7}$$

This notation will turn out to be useful when used with the Z-transform which will be further discussed in Section 2.2.

2.1.2. Impulse Response of a Filter

Every LTI digital filter has an associated impulse response $h[n]$ that completely represents the system and obeys the following equation

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k] \tag{2.8}$$

where N can be taken between 1 and ∞. At this point it is interesting to note that Equation 2.8 is actually the definition of a moving average if in addition $\sum_{k=0}^{N-1} h[k] = 1$ and $h[k] \geq 0$ are fulfilled.

Hence, moving averages are a subset of digital filters and the impulse response $h[n]$ represents the weights applied to the observations of the time-series considered. This also means that the SMA and EWMA can be analyzed as digital filters and new moving averages can be designed using this theory.

The impulse response of any digital filter can be derived from Equation 2.7 and the Z-transform that we introduce later.

2.1.3. Finite and Infinite Impulse Response

In Equation 2.4, the coefficients b_k represent the moving average (MA) part and the a_k coefficients the autoregressive part (AR) of the filter. As we already saw, the SMA only has an MA part and thus its impulse response is simply represented by the coefficients b_k. On the other hand, the EWMA has an AR component but by expanding the AR term, one can rewrite it only with MA coefficients but needs an infinite number of b_k to recreate the AR effect as mentioned in Chapter 1. In fact, all filters having an AR part can be rewritten only with an MA part but in order to recreate the exact process an infinite number of MA coefficients is needed.

In signal processing, practitioners call filters that are made exclusively of an MA part *Finite Impulse Response (FIR)* filters because their impulse response length is finite and filters with an AR part *Infinite Impulse Response (IIR)* filters because their impulse response length is infinite.

2.2. Z-transform

In order to better understand the smoothing qualities of a digital filter it is useful to look at it in the frequency domain. The Z-transform is conceptually equivalent to the Laplace transform in discrete-time. It transforms a time-domain signal into a complex frequency-domain representation. Let the operator $\mathcal{Z}\{\cdot\}$ denote the Z-transform and thus

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n\in\mathbb{Z}} x[n]z^{-n}. \tag{2.9}$$

Using Equation 2.9, one can derive some interesting properties of the Z-transform. A short list of useful ones is given below.

$$\begin{aligned} \mathcal{Z}\{a_1x_1[n] + a_2x_2[n]\} &= a_1X_1(z) + a_2X_2(z) && (2.10)\\ \mathcal{Z}\{\delta[n-k]\} &= z^{-k} && (2.11)\\ \mathcal{Z}\{x[-n]\} &= X(z^{-1}). && (2.12) \end{aligned}$$

Among all the Z-transform's properties, the most interesting one is probably the one that allows to transform a convolution into a multiplication such that

$$\mathcal{Z}\{f[n] * g[n]\} = F(z)G(z). \tag{2.13}$$

2.2.1. LTI System Representation in the Z-plane

Using the previous properties (Equation 2.13), one can easily write the Z-transform of Equation 2.7 as:

$$Y(z)\prod_{k=0}^{N} a_k z^{-k} = X(z)\prod_{k=0}^{M} b_k z^{-k} \tag{2.14}$$

which can in turn be rewritten

$$\frac{Y(z)}{X(z)} = H(z) = \frac{\prod_{k=0}^{M} b_k z^{-k}}{\prod_{k=0}^{N} a_k z^{-k}} = \frac{B(z)}{A(z)}. \tag{2.15}$$

In the Equation 2.15 $H(z)$ is introduced. This quantity is called the *transfer function* because it describes the mapping in transferring the filter input to its output. It is the Z-transform of the impulse response. Note that in the time-domain the output of a filter is given by a

convolution between the input $x[n]$ and the impulse response $h[n]$ (Equation 2.8), whereas, in the Z-domain, it is given by a multiplication between the Z-transform of the input $X(z)$ and the transfer function $H(z)$ so that $Y(z) = H(z)X(z)$. This is a straightforward application of Equation 2.13.

It is often convenient to represent $H(z)$ in terms of its zeros and poles. Indeed, in Equation 2.15, the function $A(z)$ and $B(z)$ can be factorized as polynomials and thus $H(z)$ can be fully described by its zeros $\{z_i|B(z_i) = 0\}$, poles $\{p_i|A(p_i) = 0\}$ and a multiplicative constant. Note that FIR filters have in addition a trivial pole at $z = 0$, so that $1/B(z) = 0$. As such, it is common to represent a transfer function using a so-called zero-pole diagram where zeros are commonly shown as "o" and poles as "x".

To illustrate this and the following concepts, we pick the example of a simple moving average with $N = 3$. In this case, since $b_0 = b_1 = b_2 = \frac{1}{3}$ and $a_0 = 1$ (FIR filter, no autoregressive part), the impulse response is $h[n] = \frac{1}{3}(\delta[n] + \delta[n-1] + \delta[n-2])$. Using the properties of the Z-transform, it is easy to find the transfer function $H(z) = \frac{1}{3}(1 + z^{-1} + z^{-2})$ which can be factorized in $H(z) = \frac{1}{3}(1 - e^{j2/3\pi}z^{-1})(1 - e^{-j2/3\pi}z^{-1}) = B(z)$ and thus its two zeros are $z_1 = e^{j2/3\pi}$ and $z_2 = z_1^*$. The complete function $H(z)$ is illustrated in Figure 2.2 where the effect of the pole and zeros can be clearly seen; $H(z_1) = H(z_2) = 0$ and the function grows towards infinity at the trivial pole $H(p)\big|_{p=0}$. A more typical and understandable illustration of $H(z)$ is the illustration in Figure 2.3 where only the location of the zeros and poles are shown. As we will see in more details in Section 2.3, in this case, the zeros dampen the frequency response for high frequencies and thus the filter is a low-pass filter.

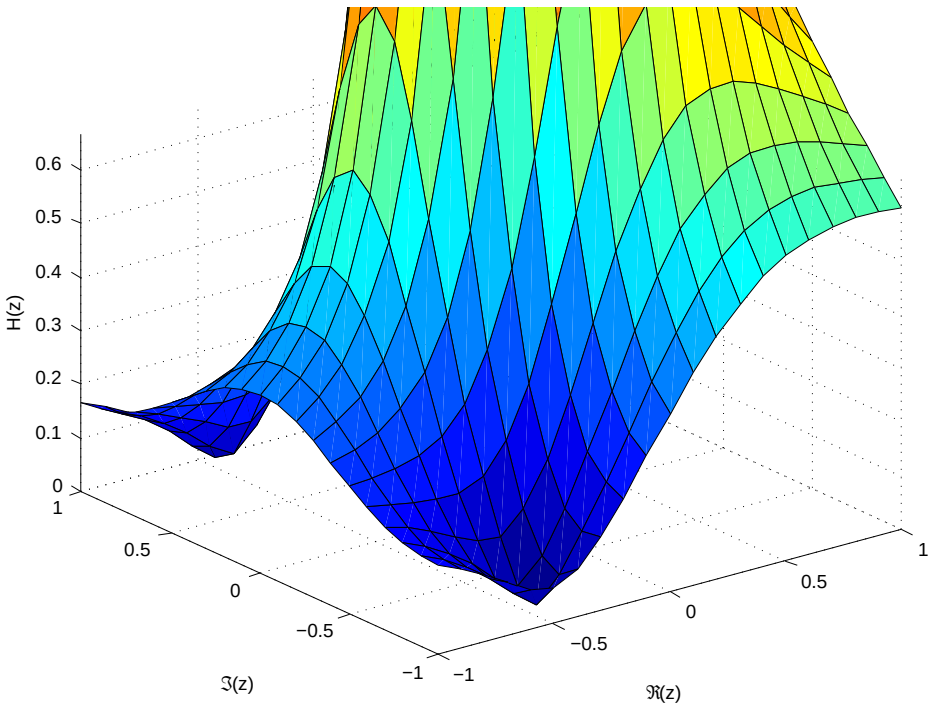


Figure 2.2.: Illustration of $H(z) = \frac{1}{3}(1 + z^{-1} + z^{-2})$ in the Z-plane.

In order to retrieve the impulse response $h[n]$ from the transfer response $H(z)$, one can either

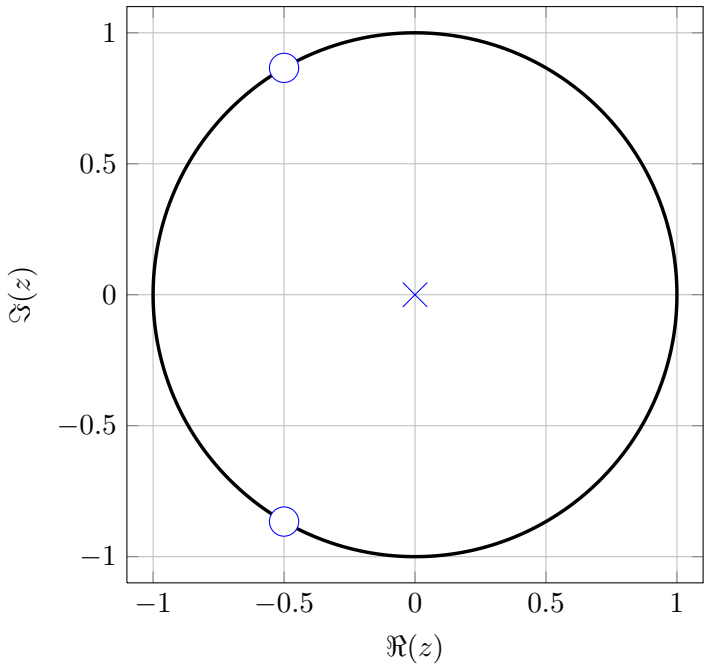


Figure 2.3.: Example of a filter representation in the Z-plane with one trivial pole at $p_1 = 0$ and zeros at $z_1 = e^{j2/3\pi} = -\frac{1}{2} + j\frac{\sqrt{3}}{2}$ and $z_2 = z_1^*$

use the inverse Z-transform (Out of the scope of this thesis, see Equation A.11) or look it up in a table of pre-computed Z-transforms (see Table A.1) and use them along with the Z-transform properties.

2.2.2. Normalizing Moving Averages

To conclude with this section, we recall that one of the two properties moving averages must fulfill is that $\sum_{k=0}^{N-1} h[k] = 1$. That is the sum of the weights must equal one. This property is sometimes tricky to test in the time domain but when converted to the Z-domain, it is considerably simplified since it is equivalent to evaluate the Z-transform at $z = 1$, $H(z)\big|_{z=1}$.

The proof can be obtained by simply writing the equation of the Z-transform and then evaluating at 1.

$$H(z) = \sum_{k\in\mathbb{Z}} h[k]z^{-k} \tag{2.16}$$

and

$$H(1) = \sum_{k\in\mathbb{Z}} h[k]. \tag{2.17}$$

Training Period and Cumulative Weights

The cumulative weight is a measure that indicates which fraction of weight is contained in the N first points of the filter. We have

$$CW_N = \frac{\sum_{k=0}^{N-1} h_k}{\underbrace{\sum_{k=0}^{\infty} h_k}_{1}}. \tag{2.18}$$

This relation is especially important for IIR filters as they theoretically require an infinite number of points to have $CW_N = 1$ in all the other cases the sum of the weights will never equal one. In [J.P. Morgan and Reuters, 1996], the authors introduce a tolerance measure Ω that measure how much weight is being neglected when computing the average that is defined as

$$\Omega = 1 - CW_N = \sum_{n=N}^{\infty} h_n. \tag{2.19}$$

In the FIR case, $\Omega = 0$ when the number of observations is at least the length of the filter (length of the impulse response). Results become meaningful only after the Mth observation as all the preceding outputs will have $\Omega < 1$. However, for IIR filters, $\Omega = 0$ can never be achieved in practice but a reasonable threshold can be set (typically below 2%). Moreover, the exact relation between N and Ω is different for every IIR filter. However, one can also plot the cumulative impulse response of the filter and decide on a threshold from the plot.

Hence for any filter, the first output points are meaningless because it takes a certain amount of points to reach a reasonable Ω. We call this period the *training period* or *transient*. A concrete example will be given for the EWMA in Section 2.6.2.

2.2.3. Energy of a filter and System Stability

In signal processing, the energy of a filter is defined as the squared sum of the coefficients of the impulse response[1] or

$$E = \sum_{k\in\mathbb{Z}}^{N-1} h[k]^2. \tag{2.20}$$

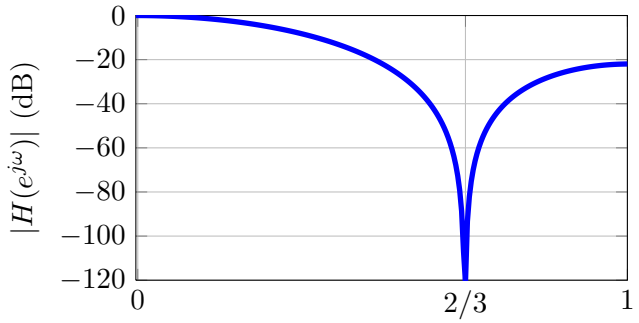
One can show with the help of Parseval's theorem (Section A.1.5) that the energy is the same in the time-domain and the frequency domain. Hence, we will use this relation to ensure that the comparison of two different filters is fair by tuning their parameters so that their energy is the same.

Moreover, the concept of energy is also used to describe the stability of a system. If the filter fulfills $E < \infty$, the system is said to be BIBO stable. BIBO stable stands for Bounded-Input Bounded-Output and is generally a desirable characteristic for any filter and assures it does not diverge to infinity. In the Z-domain, the necessary and sufficient condition for BIBO stability is that all poles lie within the unit circle on the Z-plane.

[1] Formally the energy is defined as $\sum_{k=0}^{N-1} |h[k]|^2$ but since here we consider only real valued impulse responses, the absolute value condition is relaxed.

2.3. The Discrete Time Fourier Transform - DTFT

The Discrete Time Fourier Transform (DTFT) allows to visualize the frequency response of a filter. It is the Z-transform evaluated on the unit circle, i.e. the associated DTFT of $H(z)$ is $H(e^{j\omega})$. From this it follows that the DTFT is always 2π-periodic, i.e. $H(e^{j\omega}) = H(e^{j\omega+2\pi})$. This is called the Discrete Time Fourier Transform (DTFT). In the same way that the Fourier transform is a special case of the Laplace transform in the continuous world, the DTFT is a special case of the Z-transform. Since the DTFT is a complex function, usually the amplitude $|H(e^{j\omega})|$ and the phase $\angle H(e^{j\omega})$ are shown to describe the filter. See Figure 2.4 for an illustration of the frequency response of the SMA example. The effect of the zeros shown in Figure 2.3 is clearly marked by a strong decrease in frequency at the zero position. A pole on the unit circle would do the opposite effect. The shape of the frequency response can be estimated from any pole-zero diagram by traveling along the unit circle and looking at the surrounding poles and zeros. If a zero is close to the current position, the value of $|H(e^{j\omega})|$ will go down whereas if a pole is in the surrounding, it will go up. Note that since for real valued $h[n]$, the DTFT exhibits an Hermitian symmetry ($H(e^{-j\omega}) = H(e^{j\omega})$) we only show the frequency response from 0 to π (or normalized from 0 to 1).



(a) Amplitude frequency response

Figure 2.4.: DTFT representation of the SMA $h[n] = \frac{1}{3}(\delta[n] + \delta[n-1] + \delta[n-2])$. 2.4a is the amplitude

2.3.1. System Representation

Three different system representations were introduced: the impulse response $h[n]$, the transfer function $H(z)$ or the associated pole and zero diagram, and the frequency response through the DTFT. Depending on the aim of the system, one might want to give $h[n]$ or $H(z)$ as a system representation. For instance moving averages (see Section 2.5) are easy to understand when they are given in the time-domain $h[n]$ since one immediately understands the different importance of past observations for the average. On the other hand frequency selection filters such as lowpass filters can be understood in an easier way when they are given in the frequency domain $H(z)$ or $H(e^{j\omega})$. When evaluating the stability of a system, it is convenient to look at the poles and zeros of the filter on the Z-plane. In this last case, poles tell us at which frequencies the transfer function goes to infinity and zeros where frequencies are damped. Moreover, a sufficient and necessary condition for a system to be BIBO stable is to have all poles lying within the unit circle.

2.4. Physical Meaning of Filtering Financial Time-series

In this thesis, we focus only on financial time-series sampled at a daily frequency[2]. As mentioned before, the aim of filtering is to extract relevant parts of the signal and remove high frequency noise in order to retrieve the underlying trend of a security. In the example (Figure 2.4), since the right part of the amplitude frequency response (note the decibel unit in the y-axis) has lower values than the left part, we call it a *low-pass* filter as low frequencies pass through more easily than higher frequencies (See Section A.1.6 for more details on the different types of filters). All filters we present are low-pass filter since we are interested in filtering out the noise. In our case, the highest possible frequency is the daily frequency and the simplest filter we could imagine is an all-pass filter that would not alter the input. This filter is described as $h[n] = \delta[n]$ so that $y[n] = x[n] * \delta[n] = x[n]$. Its Z-transform is $H(z) = 1$. Thus no frequency component is altered and the output signal is the same as the input signal. However, if we consider the SMA with $N = 3$ example again, the frequency response decreases in the end. The following relation holds between the normalized frequency f_n and the physical period frequency f (physical period T):

$$f = 1/T = f_n T_s \text{ [1/trading day]} \tag{2.21}$$

where T_s is the sampling period (in our case trading days). Concretely, for our SMA example, all normalized frequencies above $f_n = \frac{2}{3}$ are reduced by at least 20 dB (See Figure 2.4a) and thus, physically, $f = \frac{2}{3}$ [1/trading days] which means that we substantially decrease frequencies that are higher than twice every three-days or $\frac{2}{3}$ per day.

[2]We backtested financial time-series of daily close prices securities that are not traded every day but only on trading days. However, for the sake of convenience we refer here to days.

2.5. The SMA as a Finite Impulse Response (FIR) filter

In this section, we describe the SMA from a signal processing point of view and list some of its properties.

As we have already seen before, the simple moving average is one of the simplest FIR filter. It outputs the average of the last N samples of the input, or mathematically

$$y[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[n-k] \tag{2.22}$$

The impulse response is thus

$$h[n] = \frac{1}{N} \sum_{k=0}^{N-1} \delta[n-k] = \begin{cases} \frac{1}{N} & \text{for } n \in [0, N-1] \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

and its Z-transform is

$$H(z) = \frac{1}{N} \sum_{k=0}^{N-1} z^{-k}. \tag{2.24}$$

Since $\sum_{k=0}^{N-1} z^{-k}$ can always be factorized as $\prod_{k=0}^{N-1}(1 - b_k z^{-1})$, it is easy to see that an SMA filter has N zeros at b_k. Moreover, like all FIR filters, the SMA has a trivial pole at $p = 0$. The impulse and frequency responses are depicted on figure 2.5 and we see that it is a low pass filter with the number of lobes being equal to half the number of zeros. When looking at the zeros and poles in the Z-plane, Figure 2.5c, there is one trivial pole in the center and N zeros uniformly distributed, but no zero at $z = 1$.

2.5.1. Parameter of the SMA

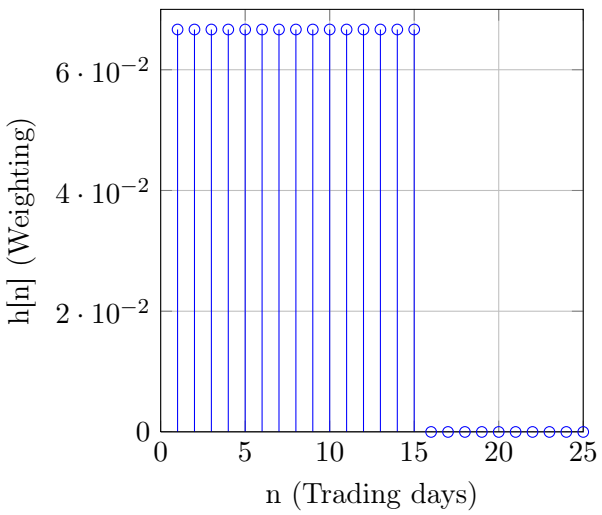
The SMA has only one parameter, N, indicating the length of the impulse response. So, when filtering daily traded time-series we refer to the parameter of the SMA as a length in days. Since, this parameter is easily understandable, in this thesis, we use it as a benchmark and tune all other filters so that we can also express them in terms of days. As explained in Section 2.2.3, we use the concept of energy to do so. Hence, for the SMA, the energy is

$$E_{SMA} = \sum_{k=0}^{N-1} h[n]^2 = \sum_{k=0}^{N-1} \frac{1}{N^2} = \frac{1}{N}. \tag{2.25}$$

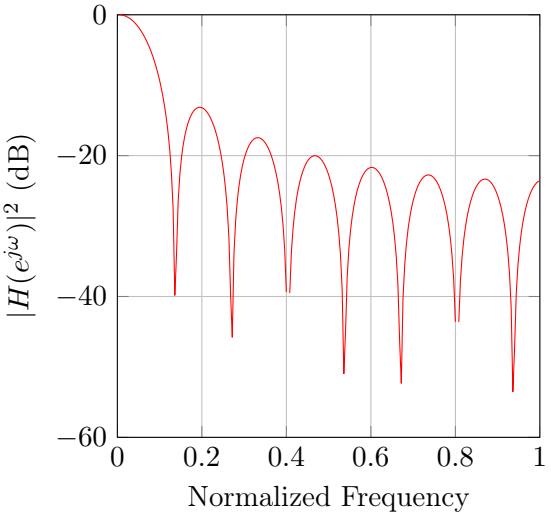
The energy of the SMA is used as a benchmark, so that we can refer to the SMA parameter N for every filters and have a common measure when it comes to compare filters.

2.6. Infinite Impulse Response (IIR) filters

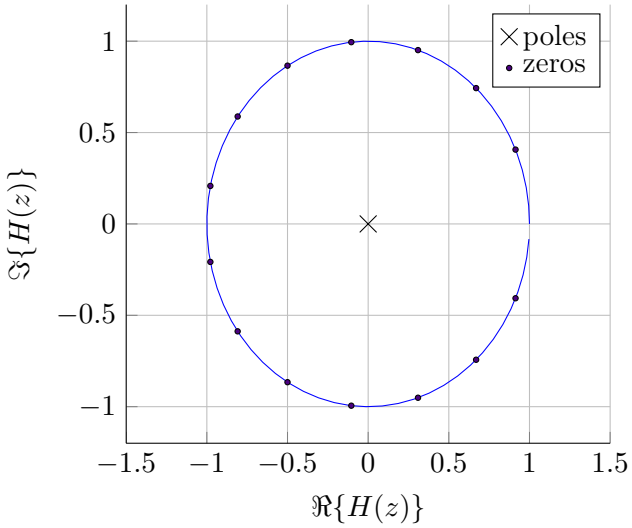
As we outlined before, IIR filters have an impulse response with an infinite support. In the next section, we derive the EWMA, Double EWMA and Triple EWMA formally and introduce an oscillating digital filter, the resonator, that we also investigate in our trading strategies.



(a) Impulse Response

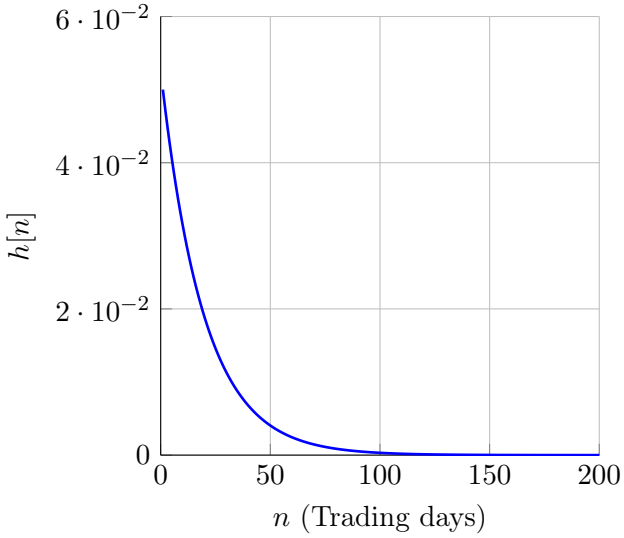


(b) Frequency Response

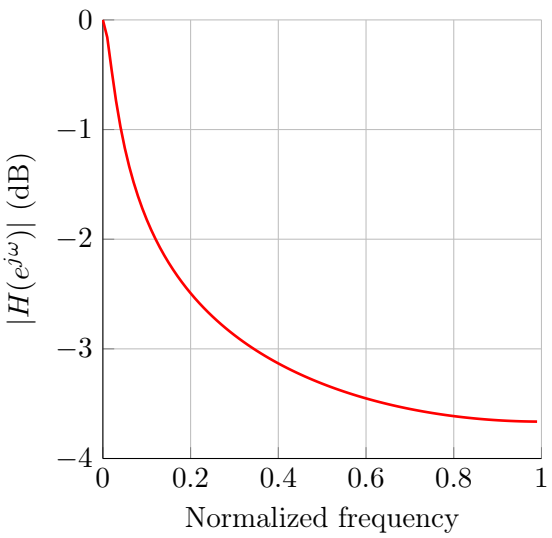


(c) Poles and zeros

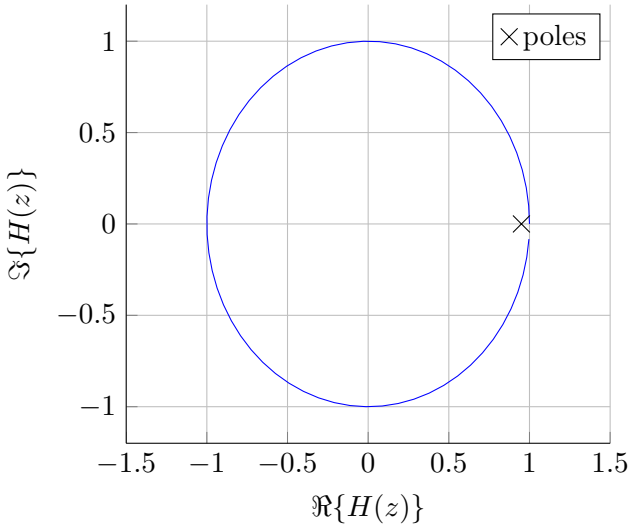
Figure 2.5.: Properties of the simple moving average (SMA) filter ($N = 15$ days)



(a) Impulse Response



(b) Frequency Response



(c) Poles and zeros

Figure 2.6.: Properties of the exponentially weighted moving average (EWMA) filter ($\lambda = 0.95$)

2.6.1. The Exponentially Weighted Moving Average as an IIR filter

The Exponentially weighted moving average (EWMA or EMA) is a moving average with weights that are exponentially decreasing and is described as

$$\begin{aligned} y_n &= (1-\lambda)x_n + \lambda y_{n-1} && (2.26)\\ &= (1-\lambda)\left[x_n + \lambda x_{n-1} + \lambda^2 x_{n-2} + ...\right] && (2.27)\\ &= (1-\lambda)\sum_{k=0}^{\infty} \lambda^k x_{n-k} && (2.28) \end{aligned}$$

This filter is widely used by traders to smooth time-series signals allowing them to identify trends in noisy signals. It is also commonly used to predict variance over time in function of return (see Chapter 4).

In the field of finance, mostly people think of the EWMA only in the time-domain. We provide here another way of deriving it from a signal processing point of view.

The EWMA is the simplest IIR filter one can design. It has only one real pole, λ, and thus only one parameter. In the Z-domain, one can easily write the corresponding transfer function as

$$H(z) = \frac{A}{1-\lambda z^{-1}}. \tag{2.29}$$

where λ is the pole location and A a constant to be determined.

Since, we are interested in moving averages, the filter needs to be normalized so that $\sum_{i=0}^{\infty} h_i = 1$ or $H(1) = 1$ which leads to $A = 1-\lambda$ and hence

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1-\lambda}{1-\lambda z^{-1}}. \tag{2.30}$$

Translating Equation 2.30 in the time domain using the tools explained in the previous sections gives

$$\begin{aligned} y[n] * (\delta[n] - \lambda\delta[n-1]) &= (1-\lambda)x[n] && (2.31)\\ y[n] &= (1-\lambda)x[n] + \lambda y[n-1] && (2.32) \end{aligned}$$

which is exactly the same as Equation 2.26.

From Equation 2.32 we can also get the impulse response in the time domain (from Table A.1, we know $\frac{1}{1-az^{-1}} \Leftrightarrow a^n u[n]$) and thus

$$h[n] = (1-\lambda)\lambda^n \ \forall n \geq 0 \tag{2.33}$$

which is the same as unwrapping Equation 2.26.

When looking at its frequency response (Figure 2.6b), we see that it is much smoother than the one achieved with the SMA. This is mostly due to the introduction of the autoregressive part.

2.6.2. Cumulative Weights

In the case of the EWMA, the cumulative weight function (see Section 2.2.2) is defined as

$$CW(M,\lambda) = \frac{\sum_{i=0}^{M-1}(1-\lambda)\lambda^i}{\underbrace{\sum_{i=0}^{\infty}(1-\lambda)\lambda^i}_{=1}} = (1-\lambda)\frac{1-\lambda^M}{1-\lambda} = 1-\lambda^M. \tag{2.34}$$

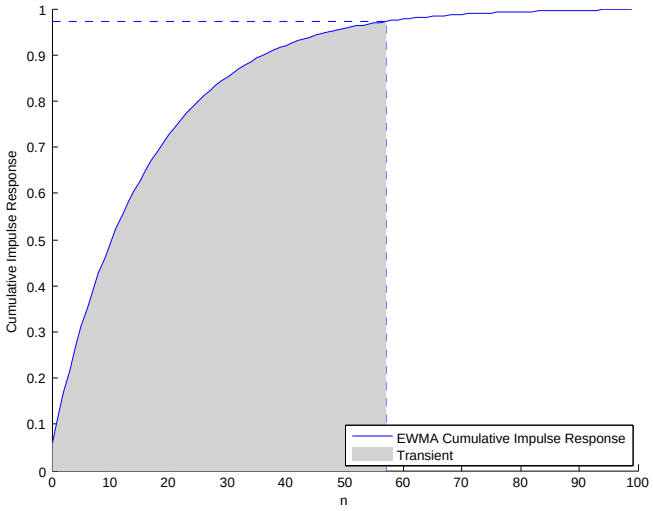


Figure 2.7.: Illustration of the transient for the EWMA with parameters $\Omega = 3\%$, $\lambda = .94$ and thus $M = 57$. These values are the ones recommended by RiskMetrics$^{\text{TM}}$ to compute the daily VaR.

Consequently, $\Omega = \lambda^M$ and thus one can get the minimum required number of days M for a required tolerance Ω by rewriting the previous equation as

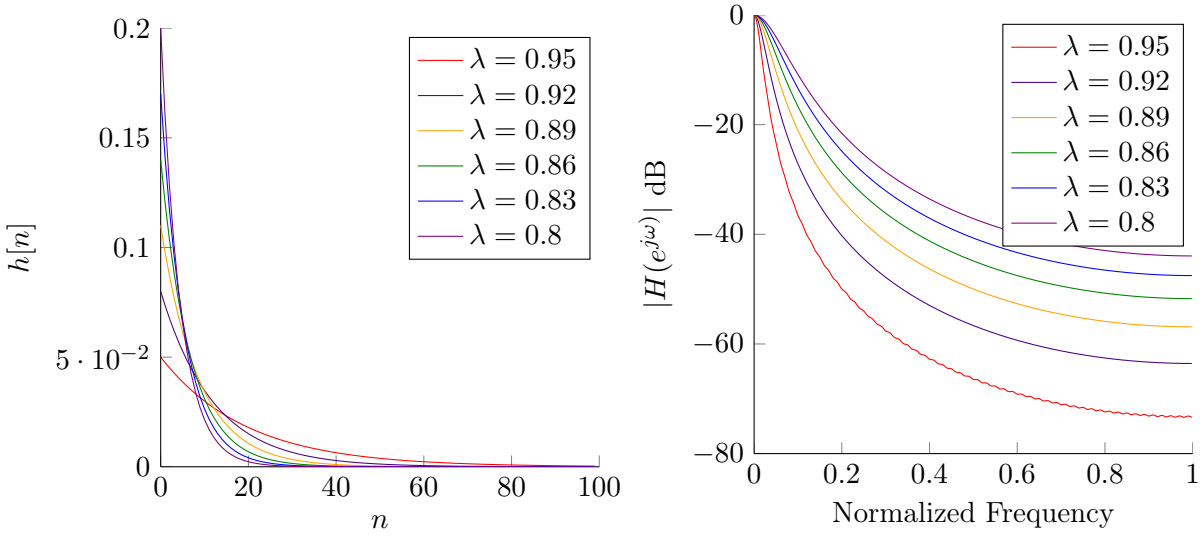
$$M = \frac{\log(\Omega)}{\log(\lambda)}. \tag{2.35}$$

As an example for a $\lambda = .94$ and a tolerance of $\Omega = 3\%$, the EWMA must be fed with at least $M = 57$ observations. In other words, the weights represented by the data-points beyond $n = 57$ only represent 3% of the total weights of the filter. A graphical example is given in Figure 2.7.

2.6.3. Parameter λ

The EWMA has only one parameter (λ) which makes it easy to use. λ defines the position of the only pole of the filter. The closer it is to 1, the smoother is the impulse response as it considers many observations. However, in the frequency domain it is the opposite: the frequency response gets sharper when λ decreases. In this case, it is because higher frequencies are filtered out. Hence, the larger is λ the more high frequencies are filtered out, see Figure 2.8. Note that λ can only be chosen between 0 and 1 since a $\lambda \geq 1$ leads to a filter that is not BIBO stable and $\lambda < 0$ yields a high-pass filter, which in this case is irrelevant.

Practitioners rather describe it in terms of number of days N so that it can easily be compared to the SMA. As for the SMA, a large N will produce a smoother curve as a small N; since at each step more data-points are taken in account. The way N is derived from λ is as a mean of comparison with the SMA. Indeed, if we compare the energy of the EWMA and the energy of the SMA they sum up to the same leaving us with the relation $\lambda = 1 - 2/(N+1)$. For the



(a) Impulse Response

(b) Frequency Response

Figure 2.8.: Visualization of EWMA for different values of λ

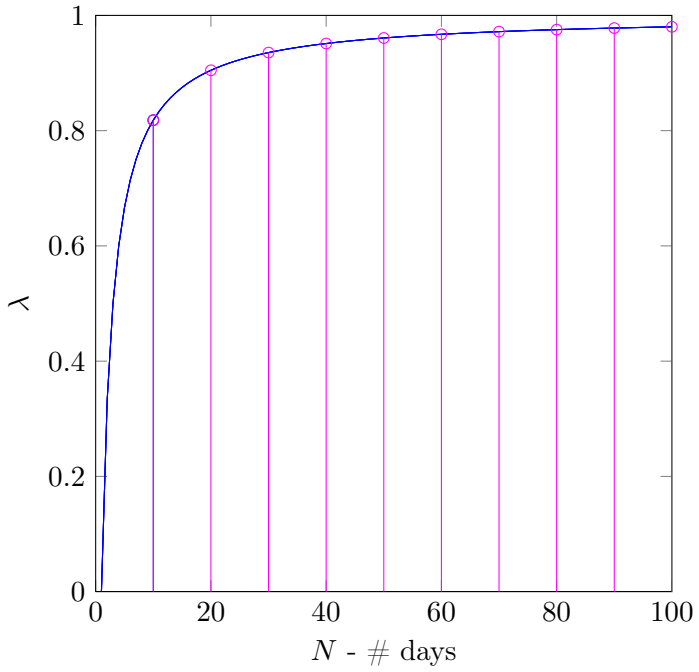


Figure 2.9.: Graphical representation of N in function of λ for the EWMA

EWMA, we have

$$\begin{aligned} E_{EWMA} &= \sum_{i=0}^{\infty} \left((1-\lambda)\lambda^i\right)^2 \\ &= (1-\lambda)^2 \sum_{i=0}^{\infty} \lambda^{2i} \\ &= (1-\lambda)^2 \frac{1}{1-\lambda^2} \\ &= \frac{1-\lambda}{1+\lambda}. \end{aligned} \tag{2.36}$$

Hence, if we set $E_{EWMA} = E_{SMA} = 1/M$, we find

$$\begin{aligned} \lambda &= \frac{N-1}{N+1} \\ &\Leftrightarrow \\ N &= \frac{1+\lambda}{1-\lambda}. \end{aligned} \tag{2.37}$$

The relation is depicted on figure 2.9. Before going further, it is useful to consider the following: $\ln(z) \approx 2\frac{z-1}{z+1}$ [3] which allows us to rewrite Equation 2.37 as

$$N \approx -\frac{2}{\ln(\lambda)}. \tag{2.39}$$

which can be seen as the half-life of EWMA.

Thus, in terms of tolerance, we have

$$\Omega = \lambda^M = \lambda^{-2/\ln\lambda} = e^{-2} \approx 13.53\% \tag{2.40}$$

This means that when using Equation 2.37 which assumes a fair comparison between the EWMA and the SMA in terms of energy, one puts a little bit more than 86% of the weights on the N last values of the time-series and that the rest of the values of the time-series represents less than $1 - .8647 \approx 13.5\%$. It is important here to note this does not mean we neglect each time 13% of the weights but rather that for an SMA of length N, the corresponding EWMA in terms of energy will have a total sum of weights of $1-e^{-2}$ when the SMA finishes. But the remaining will be covered with other observations (not used by the SMA). Nevertheless, that does not prevent from having enough data points so as to avoid the transient effect.

2.6.4. Double Exponentially Weighted Moving Average

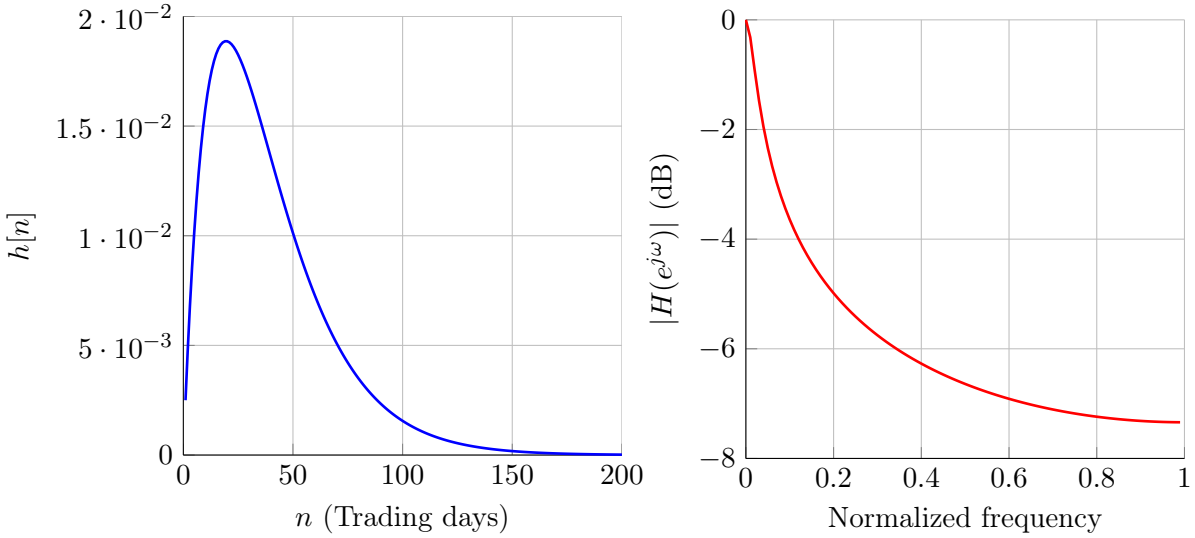
The Double Exponentially Weighted Moving Average (DEWMA) is described as follows

$$\begin{cases} y[n] = (1-\lambda)x[n] + \lambda y[n-1] \\ b[n] = (1-\lambda)y[n] + \lambda b[n-1] \end{cases} \tag{2.41}$$

[3] $\ln z \approx 2\frac{z-1}{z+1}$ is the first order of the following relation:

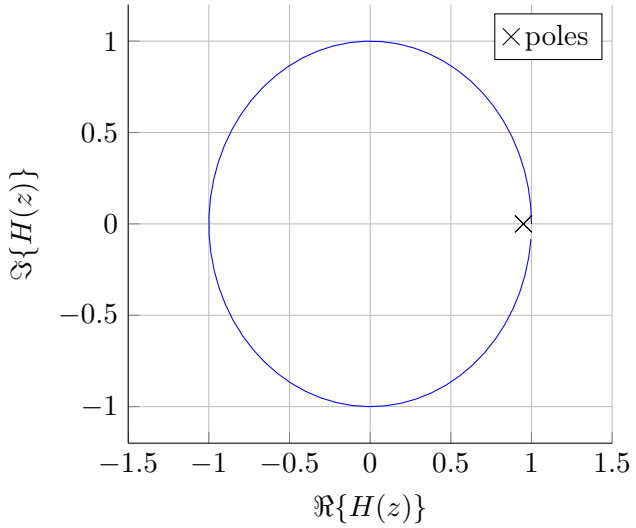
$$\ln z = 2\,\mathrm{arctanh}\left(\frac{z-1}{z+1}\right) = 2\sum_{n=0}^{\infty} \frac{1}{2n+1}\left(\frac{z-1}{z+1}\right)^{2n+1}. \tag{2.38}$$

See [Abramowitz and Stegun, 1964] for more details.



(a) Impulse Response

(b) Frequency Response



(c) Double pole at $r = .95$

Figure 2.10.: Properties of the double exponentially weighted moving average (DEWMA) filter ($\lambda = 0.95$)

which in the Z-domain yields

$$\begin{cases} Y(z) = (1-\lambda)X(z) + \lambda z^{-1}Y(z) \\ B(z) = (1-\lambda)Y(z) + \lambda z^{-1}B(z). \end{cases} \tag{2.42}$$

The above equations simply represents two EWMA filters in series; $x[n]$ being the input, $y[n]$ the output of the first EWMA filter and then inputs into a second EWMA filters with same characteristics (same[4] λ) which gives the output $b[n]$.

If we solve the system for $B(z)$ in function of $X(z)$, we get

$$H_{DEWMA}(z) = \frac{B(z)}{X(z)} = \frac{(1-\lambda)^2}{(1-\lambda z^{-1})^2} = H_{EWMA}(z)^2. \tag{2.43}$$

From this, we see two things. The DEWMA is a filter with a double real pole at $z = \lambda$ and it is the square of the frequency response of the EWMA which means that it is like having two EWMA filters in series.

In the time-domain one can write

$$h[n]_{(DEWMA)} = h[n]_{(EWMA)} * h[n]_{(EWMA)}. \tag{2.44}$$

In terms of frequency, the DEWMA has better smoothing qualities. Indeed in Figure 2.10b, the frequencies decrease twice as fast as for the EWMA (Figure 2.6b). However, this at the cost of an added lag. When looking at the impulse responses of the two filters, one can see that the EWMA responds much quicker than the DEWMA; the EWMA is constantly decreasing whereas the DEWMA has a peak that is lagged. This is a typical tradeoff encountered in signal processing where a better smoothing is achieved at the extent of lag.

2.6.5. Triple Exponentially Weighted Moving Average

The triple EWMA is three EWMA filters in parallel and thus

$$h[n]_{(TEWMA)} = h[n]_{(EWMA)} * h[n]_{(EWMA)} * h[n]_{(EWMA)}. \tag{2.45}$$

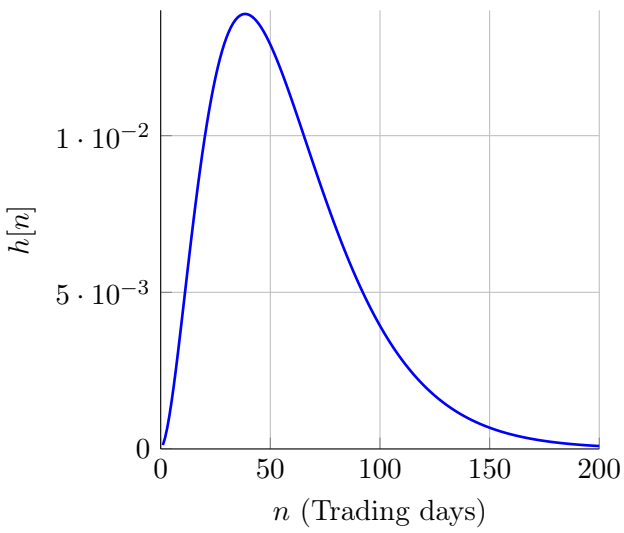
Here, the smoothing qualities are even greater than the double EWMA but the lag is also greater, Figure 2.11.

The reason why we test those three similar moving average, is to see if there are significant improvement in results when the decay in frequency is twice, three times faster than with the EWMA. This idea of putting EWMA in serial can be generalized to the n-EWMA, see Section A.2.4.

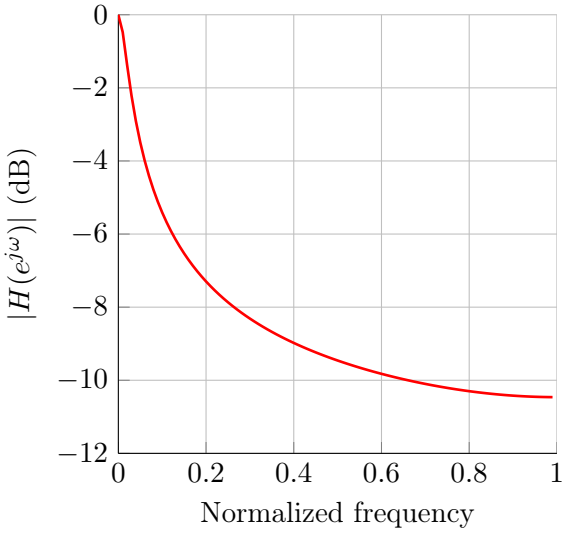
2.6.6. The Resonator - 2 poles IIR filter

The resonator is a generalization of the double EWMA as the two poles of the DEWMA are moved apart. So instead of having $p_1 = p_2$, it has $p_1 = p_2^*$ and it needs two parameters to be described. r, the distance from the center to the poles and θ the angle from the origin to the

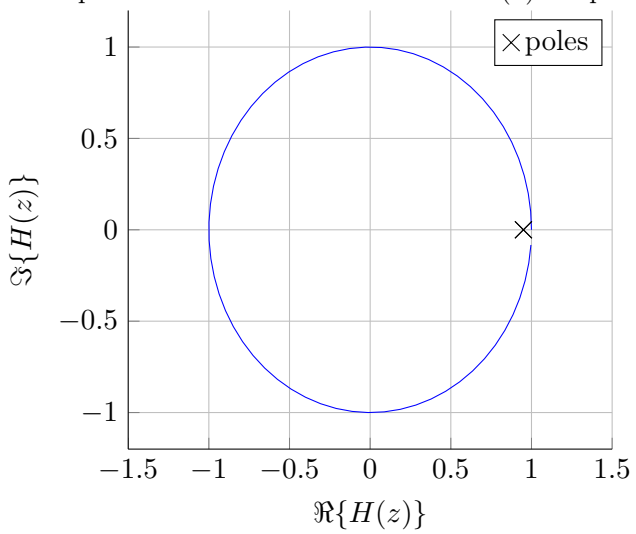
[4]Some authors introduce the DEWMA with two different parameters, λ_1 and λ_2, however we prefer to keep the number of parameters as low as possible. Nevertheless, a generalization with different parameters is straightforward.



(a) Impulse Response

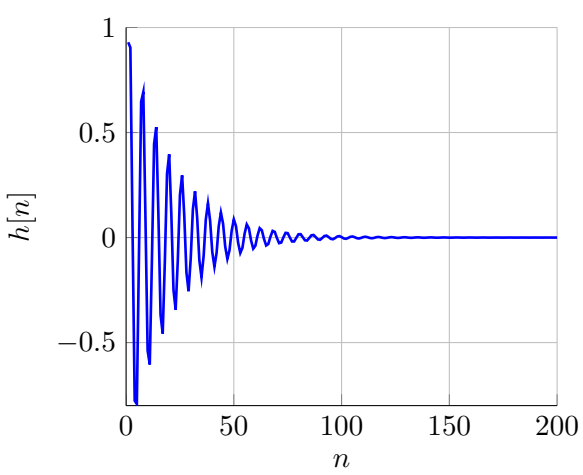


(b) Frequency Response

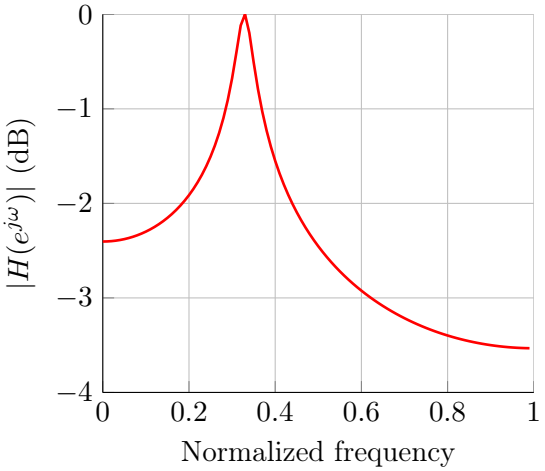


(c) Triple pole at $r = .95$

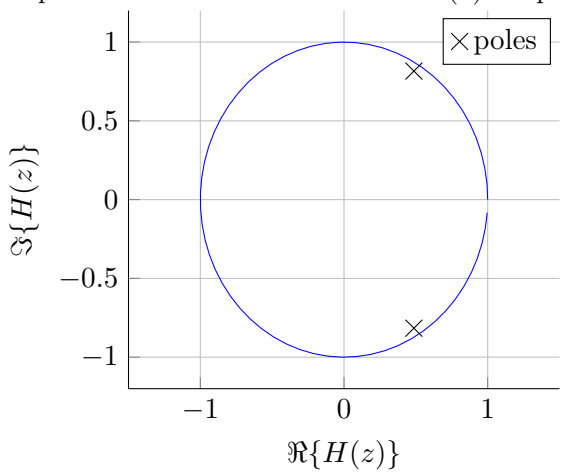
Figure 2.11.: Properties of the triple exponentially weighted moving average (TEWMA) filter ($\lambda = 0.95$)



(a) Impulse Response



(b) Frequency Response



(c) Poles in the Z-domain

Figure 2.12.: Properties of the resonator filter ($r = .9, \omega = \pi/3$)

poles. If $\theta = 0$, the resonator becomes the double EWMA. Since we know the filter has two conjugated poles, it is easy to derive the transfer function

$$\begin{aligned} H(z) &= \frac{A}{(1-pz^{-1})(1-p^*z^{-1})} \\ &= \frac{A}{1-2\Re\{p\}z^{-1}+|p|z^{-2}} \qquad (2.46) \\ &= \frac{A}{(1-re^{j\theta}z^{-1})(1-re^{-j\theta}z^{-1})} \\ &= \frac{A}{1-2r\cos(\theta)z^{-1}+r^2z^{-2}} \qquad (2.47) \end{aligned}$$

To find the value of A, the filter is normalized so that the sum of the coefficients of the impulse response sums to one.

$$H(1) = \frac{A}{1-2r\cos\theta+r^2} = 1 \qquad (2.48)$$

so that $A = 1 - 2r\cos\theta + r^2$ and hence

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1-2r\cos(\theta)+r^2}{1-2r\cos(\theta)z^{-1}+r^2z^{-2}}. \qquad (2.49)$$

From this the DTFT is

$$H(e^{j\omega}) = \frac{A}{(1-re^{j(\omega_0-\omega)})(1-re^{j(-\omega_0-\omega)})}. \qquad (2.50)$$

Finally, using Equation 2.49 and the inverse Z-transform[5] to get the impulse response

$$h[n] = \frac{A}{r\sin(\omega_0)}r^n\sin(\omega_0 n) \quad \forall n \geq 0. \qquad (2.52)$$

The impulse response is represented in Figure 2.12a. We see that oscillations were introduced, so that the condition $h[n] \geq 0$ is no longer fulfilled and thus this filter is not a moving average. Since this filter is only used in the trading part, it is acceptable to use it even if it is not a moving average since only the sign of the difference of two resonator filters will be taken in account. Hence the normalization, is actually also not relevant since it gets cancelled out anyway.

In the frequency domain, the filter no longer has a constantly decreasing frequency response but a peak at θ. This is actually a shifted version of the EWMA frequency response. Thus the resonator is no longer a low-pass filter but a bandpass filter, i.e. only a certain frequency range passes without any problem. The idea behind using this filter is to seek the good frequency range that represents the underlying trend of the examined time-series whereas for all the other filter we tune only for the good cutoff frequency.

[5] From Table A.1, one can use the relation

$$a^n\cos(\omega_0 n)u[n] \Leftrightarrow \frac{a\sin(\omega_0)}{1-2a\cos(\omega_0)z^{-1}+a^2z^{-1}} \qquad (2.51)$$

2.7. Windows

In signal processing, windows are functions that have finite support that are generally used to truncate IIR filters into FIR filters. It does so by multiplying the impulse response of an IIR filter, $h[n]$, with a window of size M, $w[n]$. The new FIR filter is thus

$$h[n]_{FIR} = w[n]h[n]. \tag{2.53}$$

Since the equivalent of a multiplication in the z-domain is a convolution, the resulting frequency response is

$$H_{FIR}(z) = \frac{1}{2\pi} W(z) * H(z). \tag{2.54}$$

2.7.1. Rectangular Window

The simplest window is the rectangular window (also called boxcar) and has only coefficients valued at one:

$$w[n] = \begin{cases} 1 & 0 \leq n \leq M-1 \\ 0 & \text{else} \end{cases}. \tag{2.55}$$

The windowed impulse response is then left with either the same coefficients as the initial impulse response or zero. The problem with this approach is that the windowed impulse response almost inevitably has jumps and those jumps are badly replicated in the frequency domain.

The reason for this is that the frequency response of H_{FIR} is the superposition of M different sine functions (M being the length of the window w_n) and – as a result of the Fourier transform – the longer is M, the less oscillations are introduced in H_{FIR}. Those oscillations that appear at sharp edges are called ripples and the effect is called the Gibbs phenomenon (Figure 2.13). In order to try to diminish the effect of ripples, practitioners have come up with other windows that minimize it. One of the tricks is to gradually decrease the weights on each side of the window so that the resulting impulse response does not include big jumps and hence reduce the Gibbs phenomenon in the frequency domain.

However, when designing a window there is always a trade-off between, what is called, the passband ripple, the transition band roll-off and the stopband attenuation ; this is shown in Figure 2.14. Ideally, in the frequency domain, the window should look like a dirac so that the convolution of the window with filter results in the exact same filter. This is however not possible with a finite length window, so that good window minimize the width of the main lobe and maximize the difference between the height of the main lobe and the adjacent ones. Some windows have become famous in the signal processing field as they minimize some unwanted effect of the Gibbs phenomenon. For instance the Hann window has very low aliasing at the extent of having a wider main lobe. The Hamming window is designed so as to minimize the height of the nearest lobe. The Blackman window places zeros at third and fourth sidelobes. In the Kaiser function, the parameter β determines the tradeoff between the main lobe width and the side lobe level.

2.7.2. Using Windows as Smoothing Moving Averages

The simplest window, the rectangular or boxcar window, is very similar to the simple moving average. Its impulse response h_n is almost the same as the one of the rectangular window.

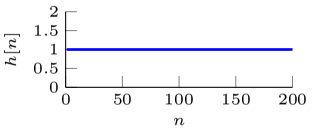
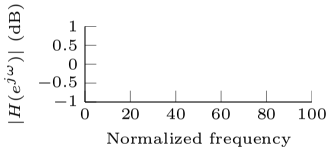
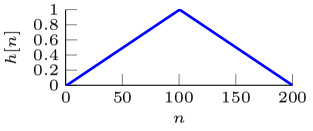
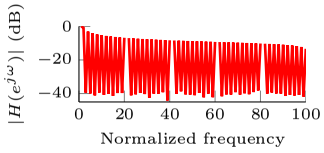
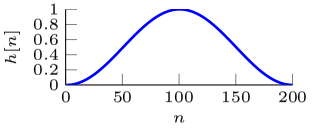
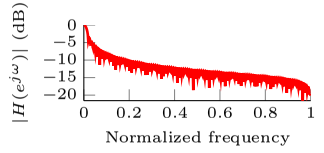
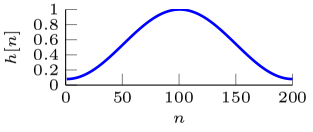
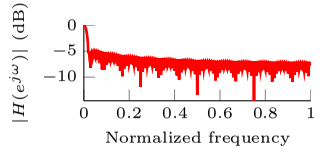
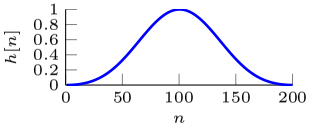
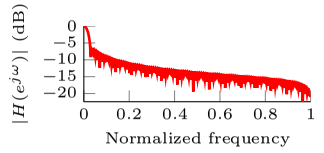
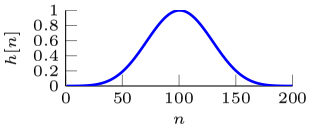
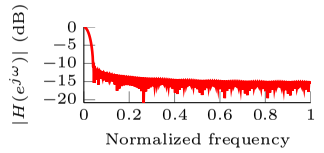
Name	Equation	Parameters	Plot	
Rectangular	$w[n] = \begin{cases} 1 & 0 \leq n \leq M-1 \\ 0 & \text{else} \end{cases}$	M		
Bartlett	$w[n] = \begin{cases} \frac{2n}{M-1} & 0 \leq n \leq \frac{M-1}{2} \\ 2 - \frac{2n}{M-1} & \frac{M-1}{2} \leq n \leq M-1 \\ 0 & \text{else} \end{cases}$	M		
Hann	$w[n] = \begin{cases} \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{M-1}\right)\right) & 0 \leq n \leq M-1 \\ 0 & \text{else} \end{cases}$	M		
Hamming	$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{M-1}\right) & 0 \leq n \leq M-1 \\ 0 & \text{else} \end{cases}$	M		
Blackman	$w[n] = \begin{cases} 0.42 - 0.5\cos\left(\frac{2\pi n}{M-1}\right) + 0.08\cos\left(\frac{4\pi n}{M-1}\right) & 0 \leq n \leq M-1 \\ 0 & \text{else} \end{cases}$	M		
Kaiser	$w[n] = \frac{I_0\left(\pi\beta\sqrt{1-\left(\frac{2n}{M-1}-1\right)^2}\right)}{I_0(\pi\beta)}$	M, β		

Table 2.1.: Properties of some of the most popular windows. Plots are shown for $M = 200$. Note that the rectangular window is also called boxcar and the Bartlett window is a linearly decreasing moving average. For the Kaiser filter the parameter is set to $\beta = 4$.

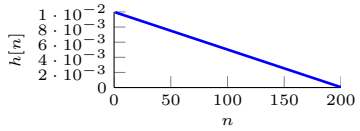
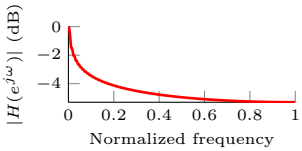
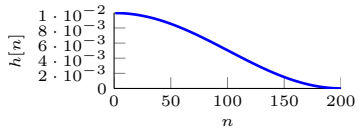
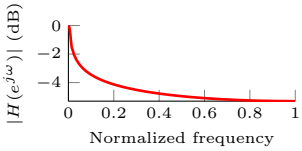
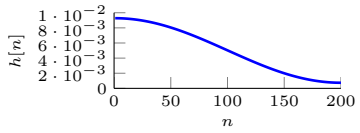
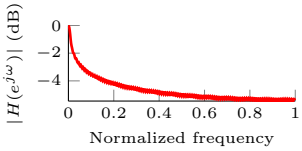
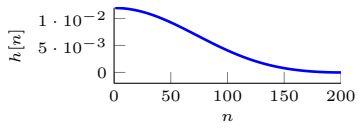
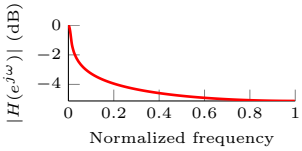
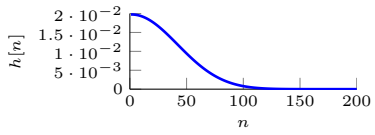
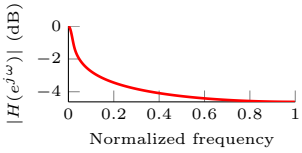
Name	Impulse response	Frequency response
half Triangular		
half Hann		
half Hamming		
half Blackman		
half Kaiser		

Table 2.2.: Table listing the half-windows we test against the SMA and EWMA later in this study

In fact, the impulse response of the SMA $h_n^{(SMA)}$ is simply the rectangular window's impulse response w_n divided by its parameter M (length of the window) so that

$$h[n]_{(SMA)} = w[n]/M. \tag{2.56}$$

One can generalize the previous consideration for all windows as

$$h[n]_{(WMA)} = \frac{1}{\sum_{k=0}^{M-1} w[k]} w[n]. \tag{2.57}$$

As it can be seen from Table 2.1, all windows can also be seen as low-pass filters with one parameters (except Kaiser, that has got two parameters). Note that in this case, windows are not used in the window sense but as filters; i.e. in the time-domain a convolution is applied between the input and the input time-series and the window function $y[n] = w[n] * x[n]$ (and not a multiplication as it is generally thought for window functions).

2.7.3. Half-windows

In this study, we try to beat the commonly used EWMA and SMA in finance when predicting future stock direction or risk.

One idea is to substitute the exponentially decaying impulse response of the EWMA by a filter that has a less abrupt decay. However, it is desirable to have a filter with weights that are constantly decreasing to fulfill the assumption that recent observations are more important than old observations. In addition, we want something that is more sensible to the influence of time than the SMA; the SMA applies the same weights to all observations thus not making any difference between events that happened a long time ago and recent events. Hence, we think that reverted logistic-shaped impulse responses can be good candidates since they are constantly decreasing but have an inflection point; their impulse responses is first concave and after the inflection point convex. With this property they share desirable characteristics of both the EWMA and of the SMA. To avoid over-fitting (see Section 3.3.5), we are interested in investigating only one-parameter filters.

Since the right half of windows introduced earlier all are logistically shaped and have only one parameter, we construct *half windows* as our logistically shaped filters. We define a half-window as a function defined by a window taken only from its highest to its lowest point, thus constructing a reverted logistic shaped impulse response. The ones that are investigated later in this study are presented in Table 2.2.

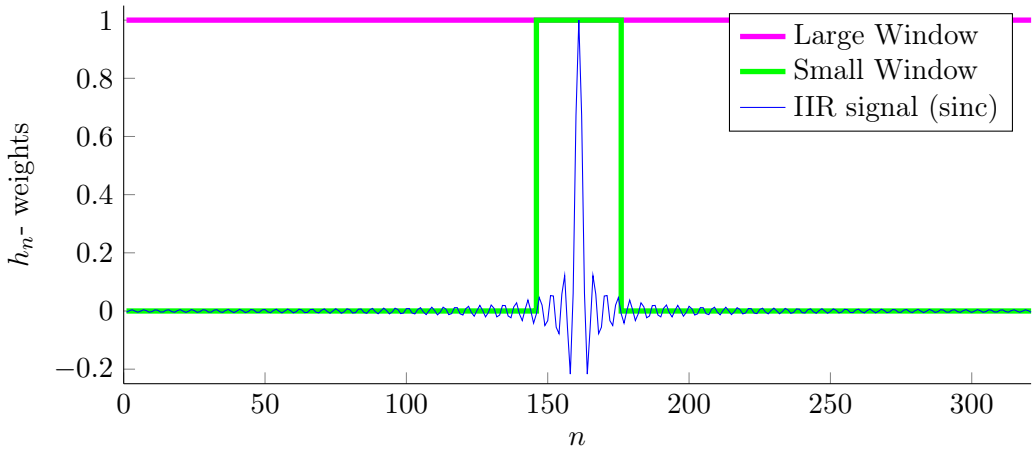
2.7.4. Comparison of half-window filters with the SMA

As we have already pointed before when comparing the EWMA and SMA (see Section 2.6.3), the filter energy is often used to compare two different filters. This ensure that both in the time and frequency domain the amount of energy is the same for the two compared filters.

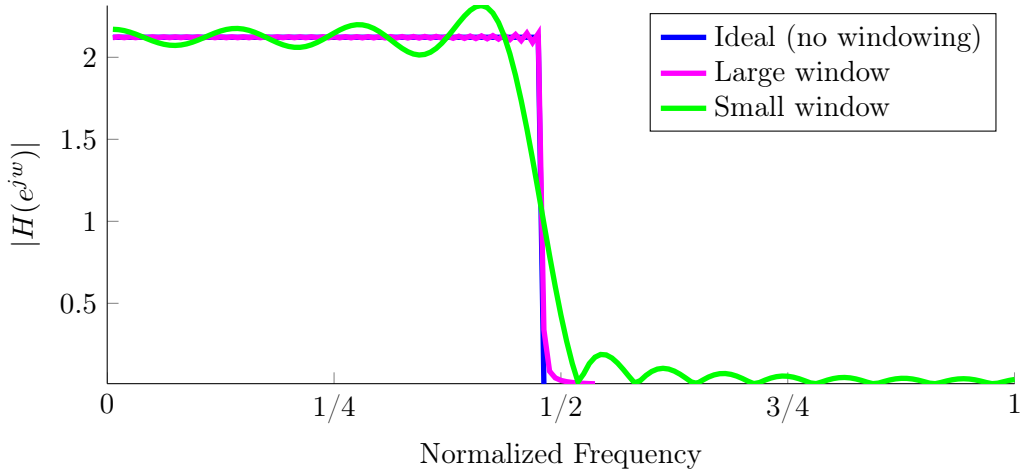
Hence, we will calibrate the half-window filters so that the energy value evaluated at the length of the SMA equals the one of the filter of interest. We demonstrated in Section 2.5.1 that the energy of an SMA filter of length N is $E = 1/N$. Thus we need to calibrate our filters, h, so that

$$\sum_{i=0}^{M-1} h[k]^2 = \frac{1}{N}. \tag{2.58}$$

The calibrated moving averages are depicted in Figure 2.15 for a value of $N = 32$ and an associated energy of $E = 1/N$. We immediately notice that all the windows look very much alike and it is hard to differentiate them. This suggests that results should be very similar. Moreover, note that the SMA stops well before the other filters and that EWMA keeps on forever since (IIR filter). In this study, when talking about half-windows, M represents the length of the window, i.e. number of weights that are non-zero valued and N represents the length in day when the filter has been calibrated with the SMA.



(a) This figure shows an input signal of (theoretical) infinite length and two different windows of different sizes applied to it



(b) This figure shows the difference in ripples depending on the amount of information that is available to reconstruct the signal. The ideal signal (of infinite length) has straight edges, the with a large window small oscillations at the edges and the signal windowed with a small window has large oscillations at its edges.

Figure 2.13.: Illustration of input signal windowed with a small and a large Boxcar window.

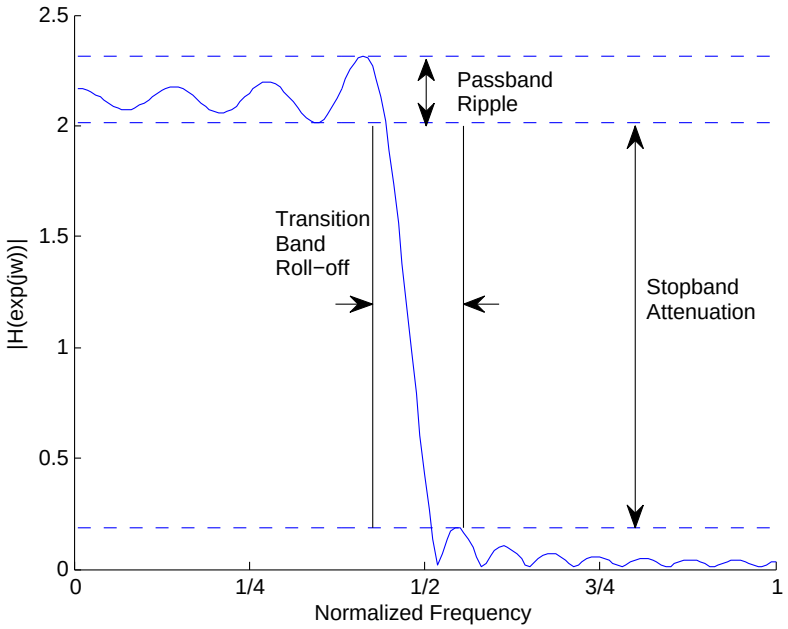


Figure 2.14.: Illustration of the tradeoff between the passband ripple, the transition band roll-off and the stopband attenuation

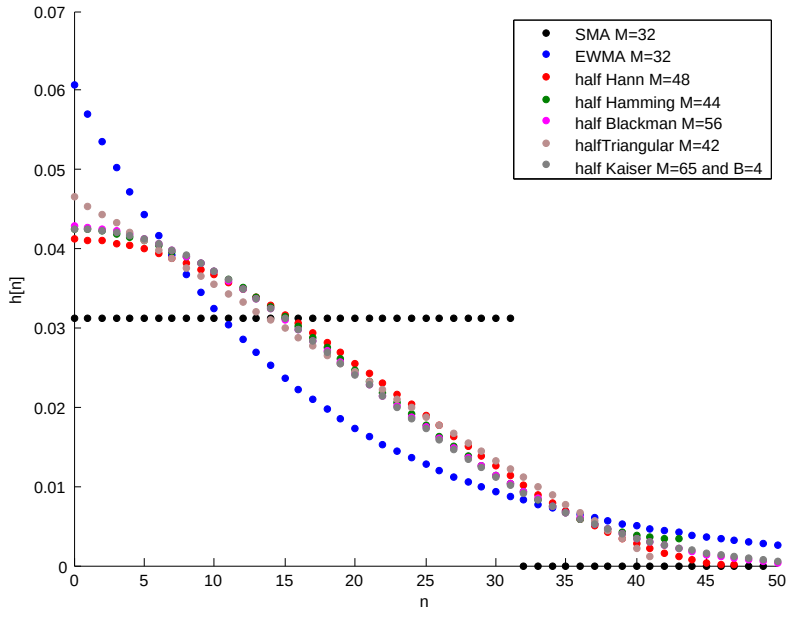


Figure 2.15.: Comparison of the different moving averages being investigated after calibration for $N = 32$ as prescribed by the RiskMetrics methodology

3. Trading Strategies using Moving Averages

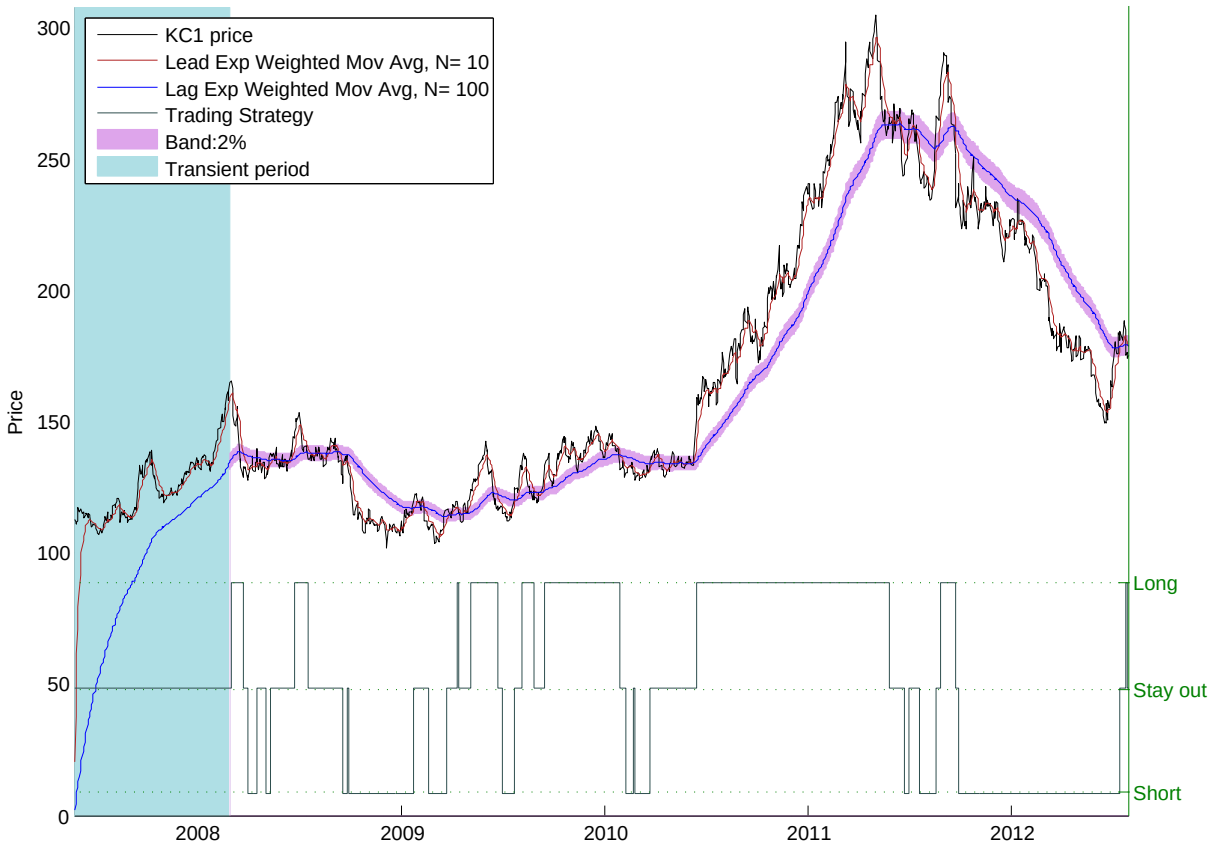


Figure 3.1.: Illustration of trading rule on Coffee (KC1) for a period of 4 years. Note the transient period for which nothing can be computed as Ω has not yet reached a value close enough to zero.

Moving averages and especially the simple and exponentially weighted moving averages (SMA and EWMA) have been used for over 30 years to extract trends from price time-series and create investment strategies, [Thomas et al., 2012, Hurst et al., 2012]. In this part of the study, we investigate the forecasting power of those two moving averages and compare it to other moving averages that – to our knowledge – have not been used in finance yet. Specifically, we want to see if the half windows, the double and triple EWMA, and the resonator introduced in Chapter

2 can outperform the EWMA within a simple trading framework.

Applying a moving average on a time-series outputs a smoothed version of the input and allows to extract and visualize trends of the underlying time-series. From a signal processing point of view and as explained in the previous chapter, a moving average is a low pass filter and thus removes high frequencies from the input signal. The level of smoothing is mostly determined by the length of the impulse response but also the values of the coefficients of the impulse response; e.g. an SMA compared to a decreasing moving average with same length outputs a smoother signal. To have a precise idea of the smoothing qualities of a moving average, one can look at the frequency response of the filter.

In practice, once the signal is filtered, the trends are extrapolated to infer future information about the market such as future price direction. It is generally accepted that if the smoothed signal is lower than the price for a certain time, the market is in a positive trend; the current price is constantly higher than the average of past prices thus suggesting that it will keep on rising. The opposite is true for downtrends. By taking advantage of this effect, one can design automated trading strategies.

In this chapter, we explain the different trading strategies we used as well as the structure and the results of the tests. Technical details about the backtesting framework are given in the appendix.

3.1. Trading Strategies

A trading strategy is a quantized signal that tells the trader (or the trading algorithm) to perform a certain action.

A trading strategy algorithm takes as input the time and all sorts of indicators. Those inputs can be multiple and very sophisticated (e.g. price, moving average, risk indicator, performance of similar securities, weather forecasts, ...). The output is a discrete number that is associated with a trading action (i.e. strong/moderate long position, strong/moderate short position, unwind, stay out of the market etc...). For the sake of simplicity, we will here only use three actions: *long*, *short* and *out of the market* positions.

Let the trading strategy signal be defined by the signal $s : \mathbb{N} \to \{-1, 0, 1\}$, the input being the time and the output the action for the trading algorithm to execute [Yao and Tan, 2000].

We define the following output actions:

$$s[n] = \begin{cases} 1 & \Rightarrow \text{long position} \\ 0 & \Rightarrow \text{stay out of market} \\ -1 & \Rightarrow \text{short position.} \end{cases} \tag{3.1}$$

With this setting, the position is held as long as $s[n] = s[n-1]$. If $s[n] \neq s[n-1]$, the position is closed and a new one is opened depending on the new value of $s[n]$, see Figure 3.1.

Generically, a trading strategy looks like

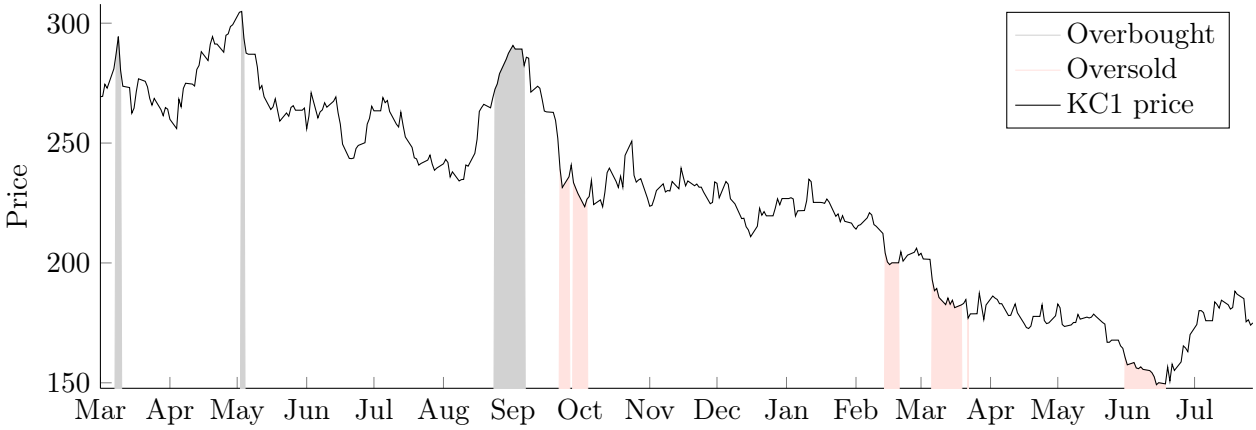
$$s[n+1] = f[n, \text{filtered signals}, \text{indicators}] \tag{3.2}$$

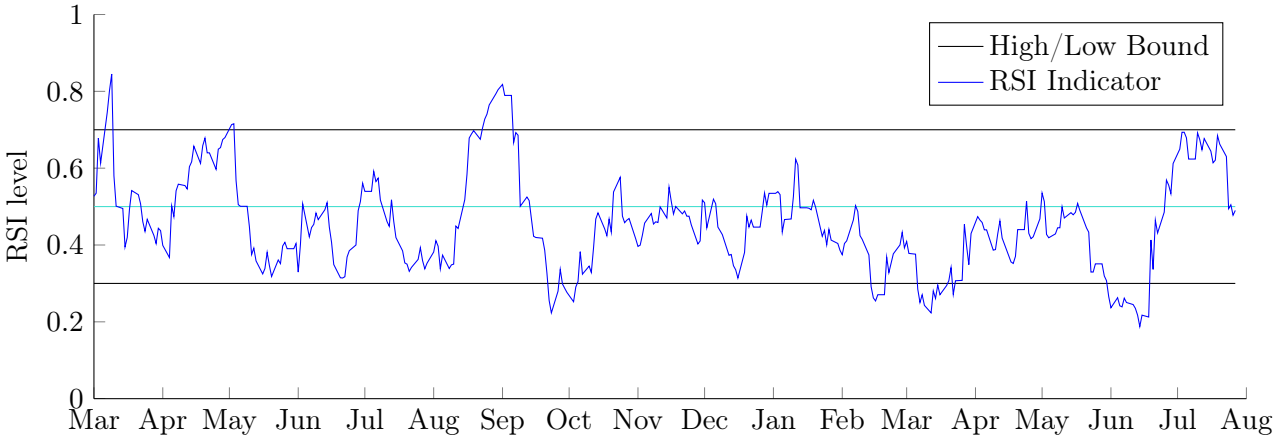
Note that the incrementation of s is shifted by one as the trading strategy is associated with the next return generated by the market. f is a trading rule and maps the different indicators into a trading action. They are detailed in the next section.

Since the main part of this work is to compare different filters against the EWMA, the trading strategy should not be too complex as it could be misleading to draw conclusions regarding the

Figure 3.2.: Coffee (KC1) times series and associated RSI indicator

(a) Price time series for Coffee between 2011 and 2012





(b) RSI indicator with $N = 14$ and upper and lower bound set to 30 and 70%.

different performances of the filters. Consequently, only relatively simple trading strategies with as few parameters as possible were designed to reduce the risk of testing the forecasting quality of the trading strategy rather than the effect of the filters (see also Section 3.3.5). Moreover, it is very difficult to measure to which extent the strategy influences the performance of the filter or to determine if the strategy is independent of the different filtered signals it takes as input.

3.1.1. Relative Strength Index (RSI)

As shown in Equation 3.2, trading rules can take any kind of indicators. To have some more sophisticated trading rules, in addition to filtered signals, we also integrated the relative strength index (RSI) indicator in some of the trading rules.

The RSI is a momentum oscillator indicator [Wilder, 1978] that tells if a security is overbought or oversold. As such it tries to anticipate a market turn. The RSI is defined between 0 and 100% with a low and high threshold. When the RSI is above the high (low) threshold, the security is said to be overbought (oversold). This process is depicted in Figure 3.2 where the area in grey (pink) indicates that the security is overbought (oversold) and thus that it should go down (up).

Computation of the RSI In order to compute the RSI, first a difference vector must be created (difference between two consecutive prices P)

$$\mathcal{D}_n = P[n] - P[n-1]. \tag{3.3}$$

Then upticks and downticks are defined as U and D or

$$U_n = \mathbf{1}\{\mathcal{D}_n\}\mathcal{D}_n \tag{3.4}$$
$$D_n = \mathbf{1}\{\mathcal{D}_n\}\mathcal{D}_n \tag{3.5}$$
$$\tag{3.6}$$

where $\mathbf{1}\{\cdot\}$ is the Heaviside function[1]. The Relative strength (RS) can then be computed as

$$RS[n] = \frac{(h_{(\text{EWMA}_N)} * U)[n]}{(h_{(\text{EWMA}_N)} * D)[n]} \tag{3.7}$$

where $h_{(\text{EWMA}_N)}$ is the computation of the EWMA with parameter N in days on the vector U. In this study all tests were performed with $N = 14$ days and 0.3 and 0.7 are chosen as low and high thresholds as recommended by [Stevens, 2002].

Finally, the Relative Strength Index (RSI) is defined as

$$RSI[n] = 1 - \frac{1}{1 + RS[n]}. \tag{3.8}$$

3.1.2. Trading Rules

Trading rules take different signals as input and output a trading action. In this paper, as trading strategy is not the main focus, we investigate only two inputs: filtered price and RSI (see Section 3.1.1). Let $P[n]$ denote the price at time n and $\hat{P}[n]$ the associated output of the moving average at time n.

Buy and Hold Position A buy and hold position means having a long position throughout the whole sample. In this case, $s[n] = 1\ \forall n$.

Simple Trading Rule on Prices One of the simplest trading rule one can think of using moving averages is a simple trend-following algorithm using the price time-series and a moving average applied to it. A trend-follower on prices means that when the price is above (below) the moving average (that is always lagging) the market is rising (declining) and thus one should keep a long (short) position. Mathematically, this can be described as:

$$\text{Trading strategy} = s[n+1] = sign(P[n] - \hat{P}[n]) = \begin{cases} 1 & \text{long} \\ 0 & \text{stay out of market} \\ -1 & \text{short.} \end{cases} \tag{3.9}$$

This strategy has two parameters: the moving average type (SMA, EWMA, something else) and the length of the moving average. The length of the moving average is generally taken between 50 and 200 days. Nevertheless, the *optimal* parameter value is not constant over different time-series and changes over time for the same time-series as markets exhibit different regimes and change unexpectedly (See Section 3.3.5) and it usually accepted that no *magic* parameter value produce the best performance.

[1] $\mathbf{1}\{x\} = \begin{cases} 1 & \text{if } x \leq 0 \\ 0 & \text{else} \end{cases}$

Crossing of two different moving averages

In this study, we use a slightly more elaborated version of the previous rule as a basis for the trading framework. According to [Brock et al., 1992] and [Fraenkle and Svetlozar, 2009], this rule has been widely used in the last 30 years. Instead of comparing the moving average directly to the value of the price, we compare it to a short moving average, called the lead. The length, N_{lead}, of the short moving average is typically between 1 and 10 days. The other moving average is called the lag and its length N_{lag} is taken between 50 and 200 days. Note that when the length of the lead is $N_{\text{lead}} = 1$, we fall back on the previous trading strategy.

So the price is filtered twice, once with the long moving average and once with the short moving average. The trader interprets the moment when the lead crosses the lag as a signal. To avoid generating trading signals in a sideways market, a tolerance band $b \in (0, 1)$ is sometimes added around the lag so as to avoid trading too often and accumulating trading costs. When the lead is in this region, no trading signal is generated. Hence if s denotes our trading strategy where $+1$ means holding a long position and -1 holding a short position, we have

$$s[n+1] = \begin{cases} 1, & \text{if } \hat{P}_{(\text{lead})}[n] > (1+b)\hat{P}_{(\text{lag})}[n] \\ -1, & \text{if } \hat{P}_{(\text{lead})}[n] < (1-b)\hat{P}_{(\text{lag})}[n] \\ 0, & \text{else.} \end{cases} \tag{3.10}$$

This is depicted in Figure 3.1 where we see the trading strategy being triggered by the crossing of the long and short moving average outside of the tolerance band.

It is important to note that moving averages can only be compared once they have reached an acceptable tolerance level Ω. As an illustration, see the blue region in Figure 3.1 for which nothing can be computed since the lag is in its transient (also called warm-up) period. In this work, we have set the same transient period for all tests so that the longest moving averages investigated has $\Omega \approx 2\%$ when the trading rule takes effect.

Mean Reversion

Until now, only trend-following strategies have been discussed. However, some markets exhibit a mean reverting behavior which is when the time-series tends to converge to its mean [Miller et al., 1994].

The associated rule is basically the inverse of a trend following rule, stating that if the price is above the moving average it will eventually converge back to it:

$$s[n+1] = \begin{cases} 1, & \text{if } \hat{P}_{(\text{lead})}[n] < (1+b)\hat{P}_{(\text{lag})}[n] \\ -1, & \text{if } \hat{P}_{(\text{lead})}[n] > (1-b)\hat{P}_{(\text{lag})}[n] \\ 0, & \text{else.} \end{cases} \tag{3.11}$$

Trading Rule Using the RSI We designed two trading rules using the RSI and both can be applied on top of one of the moving average trading rules denoted by s.

RSI trading rule type I The first one tries to prevent taking any position when the RSI indicates that the underlying is either overbought or oversold. The rationale behind is that moving averages are lagging and thus often indicate the wrong direction when there is an upturn or a downturn.

$$\tilde{s}[n+1] = \begin{cases} 0, & \text{if } RSI[n] \notin [0.3, 0.7] \\ s[n+1] & \text{else.} \end{cases} \tag{3.12}$$

RSI trading rule type II The second one relies solely on the RSI indicator when it is out of the confidence interval $[0.3; 0.7]$ which means that when the signal indicates that the underlying is oversold, the trading rule tells to go long and vice versa, else the moving average rule is followed.

$$\tilde{s}[n+1] = \begin{cases} 1, & \text{if } RSI[n] < .3 \\ -1, & \text{if } RSI[n] > .7 \\ s[n+1] & \text{else.} \end{cases} \tag{3.13}$$

3.2. Measuring Performance

Since we are interested in assessing the performance of the windows and double EWMA against the EWMA and SMA benchmarks, we need to be able to measure their performances in a fair way and then compare them with each other. We present here the different performance indicators and statistical tests we use throughout the experiments. First, we define

3.2.1. Return computation

In this part of the study, returns, r, are always computed as

$$r[n] = \frac{P[n]}{P[n-1]} - 1. \tag{3.14}$$

3.2.2. Computing the profit and loss

We call the profit or loss of a trading strategy at time n, taken return $t[n]$. We define taken returns, t, as returns taken at each timestep when the trading strategy is followed. Let $t[n]$ denote the taken returns, $s[n]$ the trading strategy at time n and c the transaction cost (see 3.3.2). We have

$$t[n] = s[n]r[n] - \big|s[n] - s[n-1]\big|c. \tag{3.15}$$

This means that when the trading strategy tells the algorithm to take a long position ($s[n] = 1$), the taken return equals the associated return (i.e $t[n] = r[n]$). This is what happens for a buy and hold strategy. When the trading strategy tells to take a short position, then the taken return is minus the return of the underlying security and thus $t[n] = -r[n]$. When the trading strategy changes position (e.g. from long to short or short to out of the market), a transaction cost occurs to account for the bid-ask spread and the broker fee. Transaction costs are always positive, hence the absolute value and when the trading strategy goes from a long to a short position – closing and then opening a position – the transaction cost occurs twice. Note that when transaction cost are neglected, taken returns obtained with a trend-following strategy are the negative of those following a mean reversion strategy. This no longer applies when transaction costs are included.

3.2.3. Sharpe Ratio

The Sharpe ratio is a widely used indicator in finance to measure the performance of stocks or funds, [Lo, 2002], and is defined as

$$S = \frac{\mathbb{E}(r[n] - r_f)}{\sigma_{r[n]-r_f}} \tag{3.16}$$

where $r[n]$ are the returns associated with the security or fund of interest and r_f the risk-free interest rate. For our use, we define it as

$$S = \frac{\overline{r}}{s_r} \tag{3.17}$$

where $\overline{r}$ is the mean of the taken daily returns, s_r the standard deviation of daily returns[2] and r_f is set to zero. In this chapter, we computed the Sharpe ratio annualized which is $S_{\text{annualized}} = \sqrt{255}S$. As a rule of thumb when evaluating trading strategies, it is generally accepted than an annualized Sharpe ratio higher than one is considered as good [Chan, 2009a].

3.2.4. Directional quality

The directional quality Q_d, also called success rate, computes the number of positive taken returns over the number of taken returns that are different from zero or

$$Q_d = \frac{\sum_{n=1}^{N} \mathbf{1}\{t[n] > 0\}}{\sum_{n=1}^{N} \mathbf{1}\{t[n] \neq 0\}} \tag{3.18}$$

where $\mathbf{1}\{\cdot\}$ is the Heaviside function , $r[n]$ are the actual returns of the time-series and $t[n]$ the taken returns. This measure is defined between 0 and 1. A Q_d of one means that the strategy is perfect and that it always wins. Note that following the way we defined the taken returns $t[n]$, transaction costs are included in this measure.

An unbiased random strategy, such as predicting the outcome of a fair toss of a coin, would eventually lead to a $Q_d = 0.5$. Thus when elaborating a trading strategy, it is important to make sure that it is significantly higher than this level.

Hypothesis testing and Confidence Interval for the Q_d

The directional quality means that at every timestep there is an estimated probability of Q_d that the prediction is in the right direction. Hence, if we denote X by the number of correct sign predictions ($\sum_{i=1}^{n} \mathbf{1}\{t[n]r[n]\}$) then $X \sim \mathcal{B}(N, p)$ where $\mathcal{B}$ is a binomial, N the length of the sample and p is the directional quality. For an N that is large enough, the de Moivre–Laplace theorem can be applied and the approximation $X \sim \mathcal{N}(Np, Np(1-p))$ is reasonable. The latter can be normalized[3] , so that

$$q = \frac{X}{n} \sim \mathcal{N}\left(p, \frac{1}{n}p(1-p)\right) \tag{3.19}$$

It follows that an approximate confidence interval for the directional quality is

$$I = \left[-\infty; p + \sqrt{\frac{p(1-p)}{n}}\Phi^{-1}(1-\alpha)\right] \tag{3.20}$$

[2] $s_r = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \overline{r})^2}$

[3] We have

$$\begin{aligned} \mathbb{E}\left(\frac{X}{n}\right) &= \quad \tfrac{1}{n}\mathbb{E}(X) = \tfrac{np}{n} \quad &= p \\ Var\left(\frac{X}{n}\right) &= \tfrac{1}{n^2}\,Var(X) = \tfrac{np(1-p)}{n^2} &= \frac{1}{n}p(1-p). \end{aligned}$$

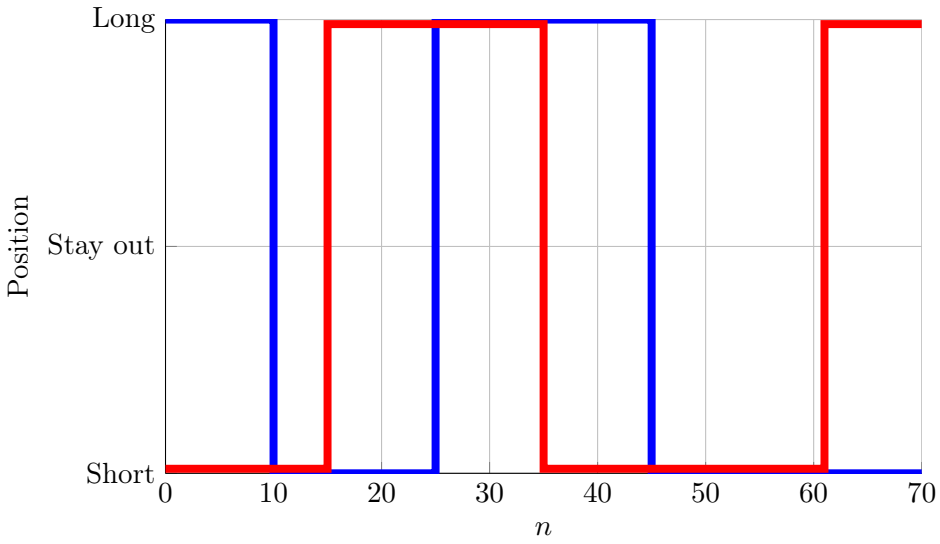


Figure 3.3.: Illustration of a hypothetical trading strategy (blue) and an associated random strategy (red).

where $\Phi(z) = \mathbb{P}(Z \leq z)$, $Z \sim \mathcal{N}(0,1)$ and p is set at the level we significantly want to beat. As we are interested in having $Q_d > .5$, we have ($p = 0.5$):

$$I = \left[-\infty; \frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right)\right]. \tag{3.21}$$

If the estimated Q_d of a trading strategy lies outside of I, one can conclude that the Q_d significantly differs from $Q_d = \frac{1}{2}$ at an α level.

In the tests, time-series were taken on 10 years, which is about 2550 trading days and thus – according to the previous development – all Q_d that are below $\approx 51.7\%$ can not be rejected at the 95% confidence level.

3.2.5. Random Strategies

In order to assess the performance of the strategies in a fair way, they are compared to random strategies. Random strategies are formed so that they are as similar to the observed strategy as possible. To do this, the ts vector is shuffled to produce a random strategy $\tilde{ts}_i$. The ts vector is made of blocks that either indicates to take a long, short or out-of-the-market position. The rule we apply when creating a random strategy, is to shuffle the blocks and randomly re-assign them a position. However, the ratio of short to long positions is to remain constant. An example is shown in Figure 3.3.

Q_d Percentile

For each tested strategy, $5,000$ random strategies and their associated $\hat{Q}_{d_i}$ are computed. If $F_X(x)$ denotes the distribution of $\hat{Q}_d$, one can find the percentile p of the Q_d of the trading strategy so that

$$p = \inf\{F_X(x) : F_X(x) \geq Q_d\}. \tag{3.22}$$

p is a percentile indicating how well the strategy performs in comparison with all the associated random strategies. Typically, a good strategy has a p that is close to one, meaning that the

investigated strategy performs better than most of the associated random strategies. A $p = 0$ would mean that all associated random strategies performed better than the investigated strategy.

3.2.6. Statistical Tests

In order to statistically test if one class of results is significantly better or worse than an other one, we used a variant of a two-sample t-test, the Welch test [WELCH, 1947].

The Welch test is a generalization of the t-test for two samples having possibly different variances. The statistic is defined as

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/N + s_2^2/N}} \tag{3.23}$$

and t follows a t-distribution with a degree of freedom that is approximated by the Welch-Satterhwaite equation

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 \cdot (N_1 - 1)} + \frac{s_2^4}{N_2^2 \cdot (N_2 - 1)}}. \tag{3.24}$$

This statistic allows to test the hypothesis $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 < \mu_2$ or $H_b : \mu_1 > \mu_2$ and we use it to test if the performance of a filter is different from the benchmark filter (the EWMA).

3.3. Backtesting Methodology

In order to be able to assess the different trading strategies and filters presented in the previous sections, tests were conducted on a portfolio of over 100 securities containing different asset classes (equities, currencies and commodities) and spanning 10 years (some are shorter if they had their IPO after 2002 or if they went bankrupt, delisted, merged or acquired between 2002 and 2012). Tests were conducted on daily close prices. More details about the portfolio are given in the appendix in Table A.4.

3.3.1. Chosen Trading Rules

We tested five different trading rules. They are shown in Table 3.1. Three are trend-following and two are mean-reversion strategies and they are combined with the RSI trading rules discussed in Section 3.1.1.

3.3.2. Trading Costs

Every time one enters or leaves a position on the market, a trading cost is lost. This trading cost is usually made of two parts: the bid-ask spread and the broker's fee. To facilitate the simulation and to be able to assess the influence of trading cost on the final outcome, a constant c of either 0 (no transaction cost), 50bps (moderate transaction cost) or 90bps (high transaction) was added to the taken returns whenever there was a change of position [4]. Every time the algorithm enters or exits the market this constant fraction is removed from the the taken return; see [Agrawal

[4] Bps means basis point and is a unit that equals one in ten thousand often used in finance to quantify transaction cost: [bps] $\equiv$ ‱ $\equiv \frac{1}{10{,}000}$. Hence, the absolute transaction cost is the associated closing price multiplied by $tc/10,000$ where tc is the transaction cost in bps.

# Identifier	Type	RSI	Trading Strategy
1	Trend-following	∅	$s_1[n+1] = \begin{cases} 1, & \text{if } \hat{P}_{(\text{lead})}[n] > (1+b)\hat{P}_{(\text{lag})}[n] \\ -1, & \text{if } \hat{P}_{(\text{lead})}[n] < (1-b)\hat{P}_{(\text{lag})}[n] \\ 0, & \text{else.} \end{cases}$
2	Trend-following	Type 1	$s_2[n+1] = \begin{cases} 1, & \text{if } RSI[n] < 0.3 \\ -1, & \text{if } RSI[n] > 0.7 \\ s_1[n+1] & \text{else.} \end{cases}$
3	Mean reverting	Type 1	$s_3[n+1] = \begin{cases} 1, & \text{if } RSI[n] < 0.3 \\ -1, & \text{if } RSI[n] > 0.7 \\ s_2[n+1] & \text{else.} \end{cases}$
4	Mean reverting	∅	$s_4[n+1] = \begin{cases} 1, & \text{if } \hat{P}_{(\text{lead})}[n] < (1+b)\hat{P}_{(\text{lag})}[n] \\ -1, & \text{if } \hat{P}_{(\text{lead})}[n] > (1-b)\hat{P}_{(\text{lag})}[n] \\ 0, & \text{else.} \end{cases}$
5	Trend-following	Type 2	$s_5[n+1] = \begin{cases} 1, & \text{if } RSI[n] < 0.3 \\ -1, & \text{if } RSI[n] > 0.7 \\ s_1[n+1] & \text{else.} \end{cases}$

Table 3.1.: List of the trading rules tested with their associated identifier

et al., 2011] for a similar approach and Equation 3.15 for the mathematical implementation. Moreover, when computing the directional quality, the taken return includes the transaction costs (e.g. if the return of the security is bigger than 0 but smaller than the transaction cost, the associated taken return will be negative). This allows to have a real overview of the performance when looking at the Q_d. Indeed, a performance that has a $Q_d > 50\%$ should be profitable. This is not necessarily the case when transaction costs are not included in taken returns.

3.3.3. Looking Ahead and In-sample optimization

When designing a backtesting framework, one must make sure it replicates the historical level of information for each point in time and that it does not include elements of future history.

Common errors include in-sample optimization (apply optimized parameters for a data-set on the same data-set) or more generally having a looking-ahead bias (using information from the future). Unfortunately, data misuse can easily happen and it is not rare to see data snooping in backtesting frameworks, [Daniel et al., 2008].

In our testing framework, we tried to avoid those biases as much as possible. However, it is almost impossible to totally avoid it, since, by definition, a backtest is done on data that are already known. Moreover, the design of the algorithms to be tested is made already knowing the information of the backtest period. For instance, it is known that trend-following strategies worked in the past and, using this information, most of the trading rules were based on this idea. The same remark applies to the use of the RSI and its parameter $N = 14$ that have been empirically found to perform well in the past. Moreover, prices are sometimes not immediately available and are sometimes slightly modified after a certain time. Those last examples are good illustrations of data snooping biases that are very difficult to avoid and that the reader must keep in mind while interpreting backtest results.

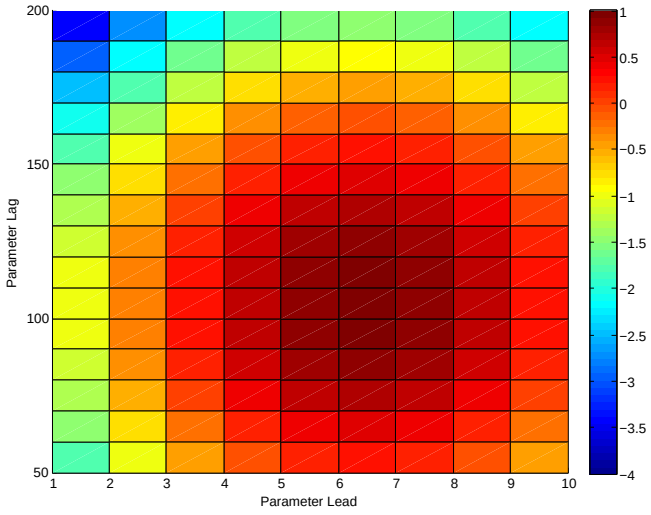


Figure 3.4.: Selection of the best parameter set in the tuning region. In the figure, colors represents the intensity of the Sharpe ratio. Red: good Sharpe ratio, Blue: bad Sharpe ratio. In this case, the algorithm would pick a parameter of $N_1 = 6$ days for the lead and $N_2 = 100$ days for the lag.

3.3.4. Parameter Sweep

Since it is often assumed that securities change behavior over time, it is very rare that a combination of parameters can be kept and utilized as a *magical* parameter set on the whole time-series.

To overcome this problem, we introduced a training and a testing window.

The training window comes always before in time the testing window and they are adjacent. For every new testing window, the model is trained (parameters are swept) on the training window, the best performing set of parameters is kept and then applied on the corresponding testing window.

Mathematically, if the function $f(t_1, t_2, \mu)$ computes the performance indicator for the time-series from time t_1 to time t_2 with parameters μ, the parameter sweep solves the following:

$$\mu^*_{t_1,t_2} = \arg\max_{\mu} f(t_1, t_2, \mu). \tag{3.25}$$

$\mu^*_{t_1,t_2}$ is then used in the testing window which is the interval $[t_2+1; t_2+\tau]$ where τ is the length of the testing window.

To avoid in-sample optimization, the two processes are completely separated and in any case information about the testing window is known while doing the optimization over the training window.

Practically, the different parameters that are swept are the length of the two moving averages (N_1 and N_2[5]). For each combination of parameter an associated Sharpe ratio is computed. The set of parameters associated with the highest Sharpe ratio are used for the corresponding testing window. A graphical illustration of the process of picking the best set of parameter in the training window according to the performance measure is shown in Figure 3.4.

[5]Since the resonator filter has two parameters (the radius, r, and the angle, θ), the swept parameters are r_1, r_2, θ_1 and θ_2.

3.3.5. Parameter Space and Overfitting

The law of parsimony states that the minimum amount of parameters should be chosen in order to describe a system and all sort of redundancy should be avoided. For this study this means that having a very good performance in-sample (training window) does not necessarily mean that the performance will be as good out of sample. Most of the time, a very good performance in-sample means that the probability of having overfit is high. Overfitting occurs when a model has too many parameters and is thus even able to fit the noise or meaningless observations in the training set. Those models are then totally useless out-of-sample and thus, while doing predictions, it is of crucial importance to try not to overfit data in the training window.

Overfitting with price or returns time-series occurs very often and even with few parameters because those time-series are known to have a signal to noise (SnR) ratio that is very low. For the same reason it is hard to find underlying trends by using typical digital filter techniques [DeBrunner and Charpentier, 2002].

Hence whenever it was possible, we resisted the temptation to add additional parameters in the model. For instance, one could investigate different tuning window lengths and apply the chosen parameters on different testing windows of different length. Moreover, some algorithms impose a minimum holding period which probably increases the profitability as transaction costs are minimized but add up another new parameter. All those are additional parameters that need to be optimized and in most cases also increase the risk of overfitting.

At the light of this, we decided to have a window of length 1 day: for every day the optimization is conducted over the tuning window. We chose 30 days as a length for the training window. The choice is a tradeoff between efficiency of the algorithm (the longer the window, the longer it takes to compute the strategy) and accuracy of the results.

It is tempting to design filters with many parameters as the shape then becomes much more modular. For instance, instead of using window filters, we could have built a reversed logistic function on 3 or 4 parameters. However, all the filters we tested (except the resonator that is a two-parameter filter) are one-parameter filters. The reason for this is to decrease the risk of overfitting during the training phase and then having totally inaccurate results in the testing window.

3.4. Simulation Results

3.4.1. Overview

For every security in the portfolio, all possible combinations of trading rules listed in Table 3.1 were tested for different filters (Table 3.2a) , transaction costs (Table 3.2b) and tolerance bands (Table 3.2c). This sums up to about $45,000$ different experiments. For each of them, performance measures were computed, i.e. Q_d, Sharpe ratio, Q_d percentile with respect to random strategies. Note that all windows and the EWMA, for fairness of comparison, were tuned according to their energy so that they all use the same parameter N.

As mentioned in Section 3.3.4, for each security the trading algorithm runs along the time-series taken between 2002 and 2012 and, for each day, applies the best set of parameters that was found by doing a parameter sweep on a training window. If the considered trading day is n, the training window comprises trading days $n-\tau$ to $n-1$ where τ is the length of the training window.

For the short moving average, the algorithm sweeps between $N_1 = 1$ and $N_1 = 10$ days and for the long moving average between $N_2 = 50$ and $N_2 = 200$ days with an increment of 10 days.

Name	Characteristic	Parameters
EWMA	Exponentially weighted (Benchmark)	λ or N
SMA	Constant weights	N
Half Triangular	Linearly decreasing	N
Double EWMA	Two EWMA in series	λ or N
Triple EWMA	Three EWMA in series	λ or N
Half Hann	Half Window	N
Half Hamming	Half Window	N
Half Blackman	Half Window	N
Half Kaiser	Half Window	N and $\beta = 4$
Resonator	Two conjugated poles IIR	r and θ

(a) Filters investigated

Transaction cost	
0 bps	No transaction cost
50 bps	Moderate transaction cost
90 bps	High transaction cost

(b) Transaction costs tested

Bands
0 %
2 %
5 %

(c) Bands tested

Table 3.2.: Parameters tested for each combination of security and trading strategy.

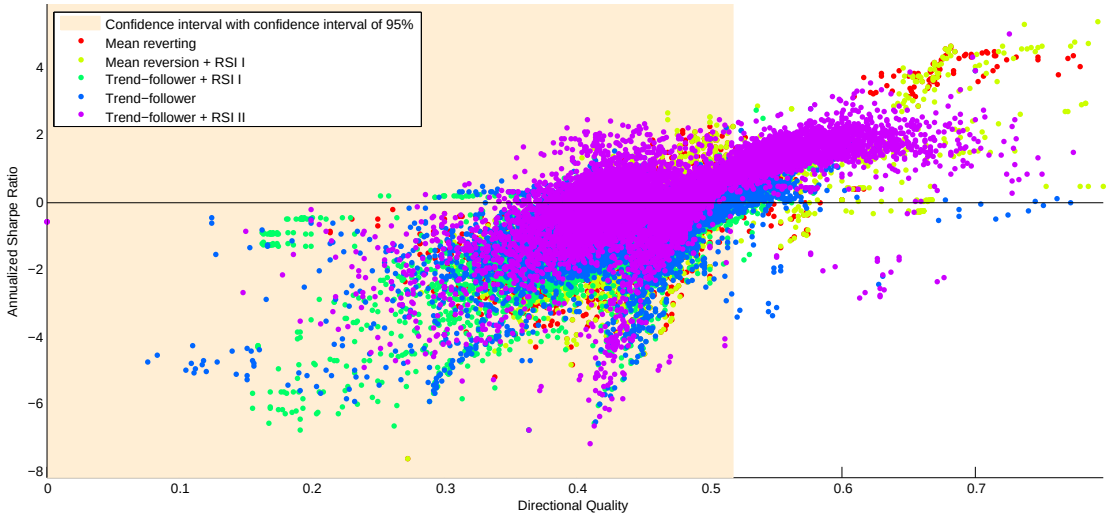
Since the resonator filter has two parameters, the algorithm sweeps over $r = .5$ to $r = .99$ with increment of 0.01 and $\theta = \pi/10, 2\pi/10, ..., \pi$.

3.4.2. Exploring the Sharpe ratios versus directional Qualities

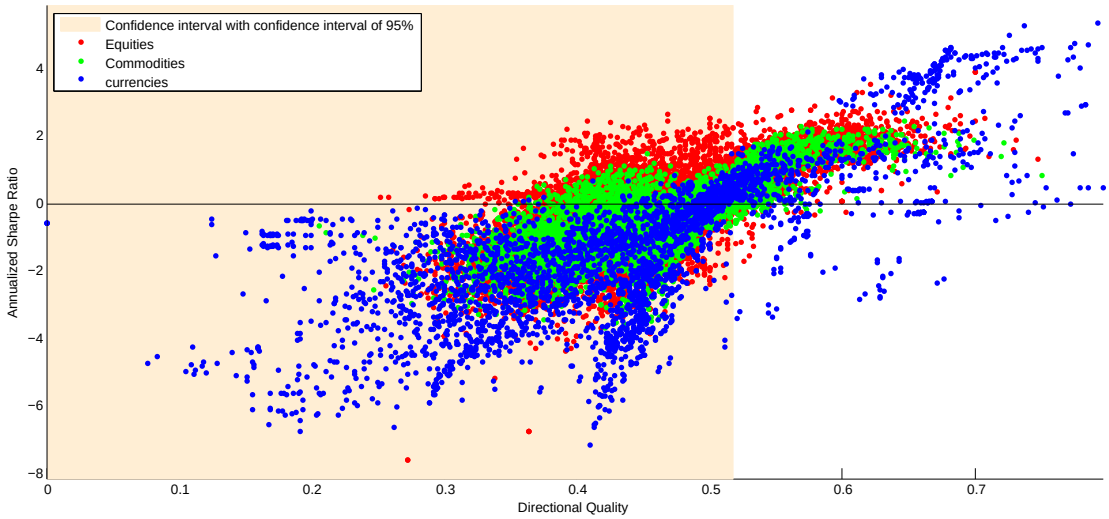
Figure 3.5 shows all the results of the simulation and their associated Sharpe ratios and directional quality on a scatter plot. To be able to graphically visualize the formation of possible clusters, subplots highlight the influence of a parameter on the final results. The light yellow background indicates the region where the hypothesis $H_0 : Q_d = 50\%$ can not be rejected whereas the white background indicates where it can be rejected in favor of $H_a : Q_d > 50\%$ at a 95% confidence level using the reasoning of Section 3.2.4. Finally, the black line indicates the boundary between positive and negative Sharp ratios. Ideally, strategies should have a positive Sharpe ratio and lie outside of the confidence interval.

When looking at the overall effect of the different trading rules (Figure 3.5a), asset classes (Figure 3.5b), or filters (Figure 3.5c), it is difficult to argue in favor of some clustering patterns as it seems that those parameters have very little influence on the results. It is important to note that in Figure 3.5a it seems that in the upper right corner there is a cluster of mean reverting strategies outperforming all other strategies. Those are actually exclusively represented by one time-series (USD-BRL) for which the change of parameter do not influence the final results greatly (recall that for every time-series, there are about 400 experiments). Interestingly, looking at the lower right corner, there seems to be a cluster of poor performing trade-following strategies that are actually mostly represented by USD-BRL and are the counterpart of the well performing strategies.

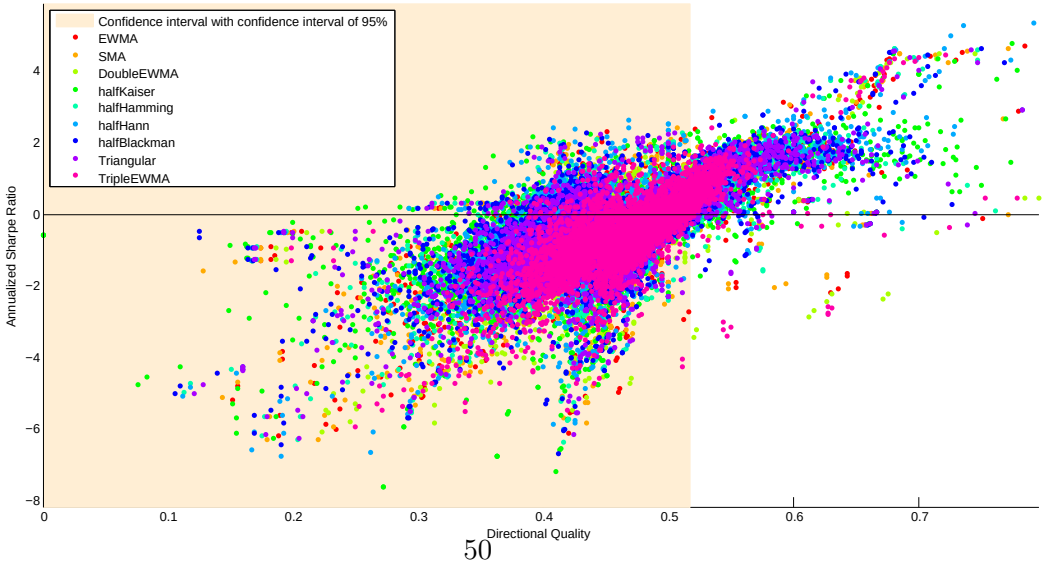
Looking at Figure 3.5e, one can see that the larger the tolerance band, the smaller is the variation in performance. This effect is expected since the introduction of the band should refrain from trading in sideways markets. Hence, since less risks are taken some potential



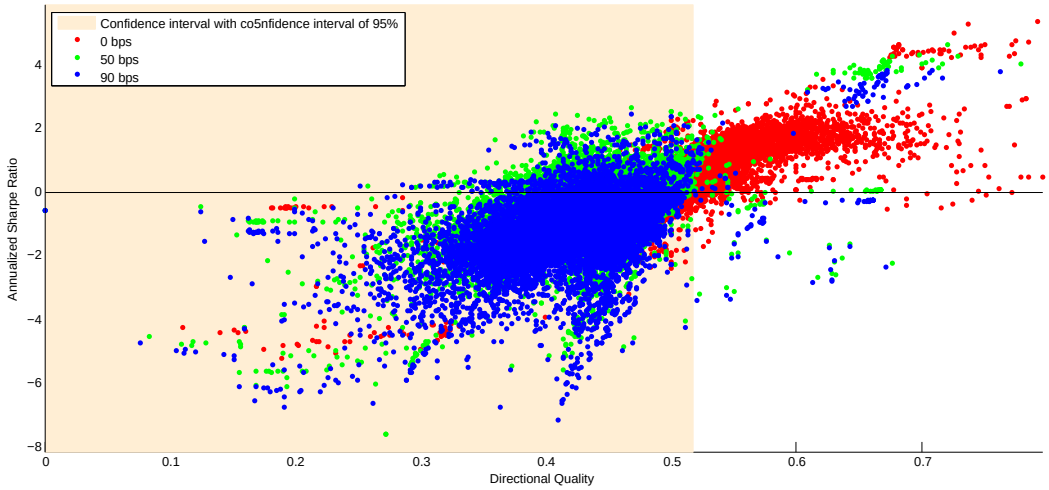
(a) Emphasis on trading rules



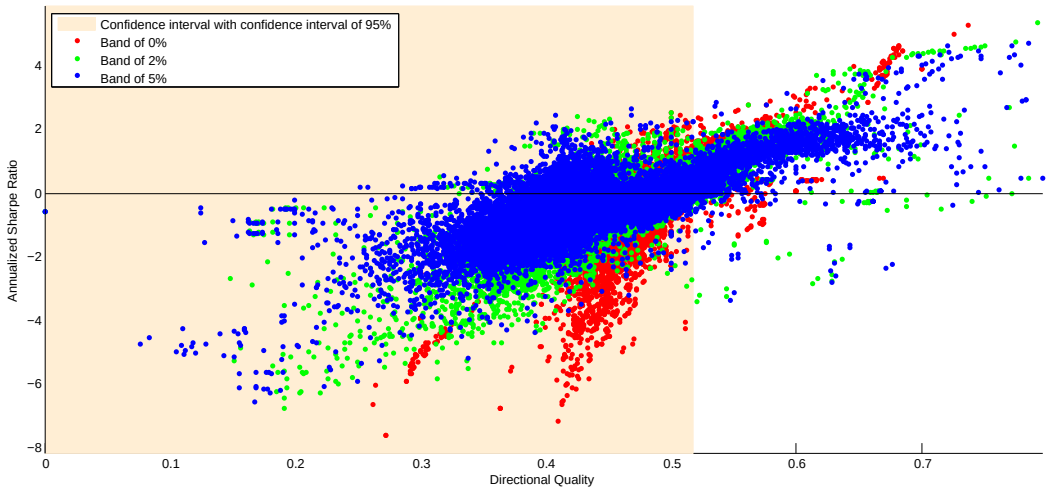
(b) Emphasis on asset classes



(c) Emphasis on filters



(d) Emphasis on transaction costs



(e) Emphasis on tolerance bands

Figure 3.5.: Scatter plot showing the annualized sharpe ratio in function of the directional quality Q_d. Every point represent the outcome of one experiment. Colors emphasis one parameter.

positive returns are lost but on the other hand some potential negative returns are also avoided and thus most of the very good and very bad performances were achieved when no tolerance band was added. The lower right corner of Figure 3.5e shows strategies with a negative Sharpe ratio but a very good Q_d. This is explained when the magnitude of negative taken returns is higher than positive taken returns. However, most of those dots are once again explained only by one security, namely the exchange rate between the US-dollar and the Chinese Yuan (USD-CNY). The time-series exhibits a very long artificial plateau between 2002 and 2004 and another one from 2009 to 2010 and when a band is added in those conditions, the trading algorithm never trades as it recognizes the sideways market but since moving averages are always lagging the effect persists out of the plateau and the algorithm can not cope with it.

Figure 3.5d indicates that strategies without transaction costs perform much better than strategies with transaction costs. This makes sense since for two similar strategies – one with transaction costs and the other without– the one with transaction costs will be penalized in its Sharpe ratio and Q_d every time a change of position occurs.

Finally, it is interesting to note that almost no strategy including transaction costs lies outside of the confidence interval, suggesting that the trading rules we investigate are not significantly better than having a trading strategy based on the outcome of the flip of a coin and thus could not be implemented per se in a trading framework. Nevertheless, strategies without transaction costs can be used as trend indicators or as input to some more sophisticated trading strategies.

Results of the Resonator Filter

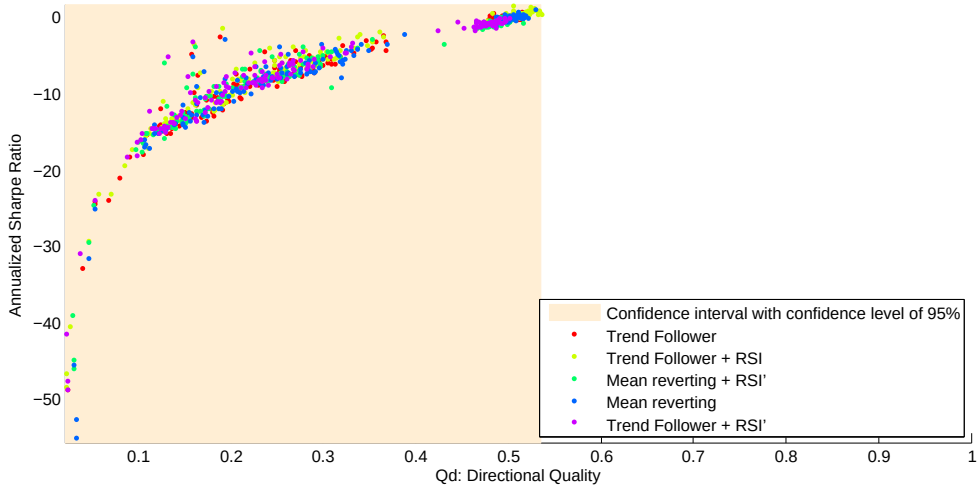
In the previous scatter plots we did not include the resonator filter as the result pattern is completely different from the other filters as shows Figure 3.6. Not a single strategy lies outside of the confidence interval. Strategies without transaction cost form a clear cluster and the difference in transaction cost is clearly outlined (Figure 3.6c). At least two reasons explain the notable difference in performance between the resonator and the other filters. The first one is that the resonator is not a real moving average since the condition $h[k] \geq 0$ is not fulfilled and the inclusion of negative weights can create some unexpected results. Moreover, the resonator has two parameters r and θ and since our algorithm uses two filters, this means four parameters are optimized on a relatively small window. This greatly raises the chance of overfitting which would also partly explain the worse performances.

At the light of the results and for the reasons enumerated above, the resonator is not discussed any further.

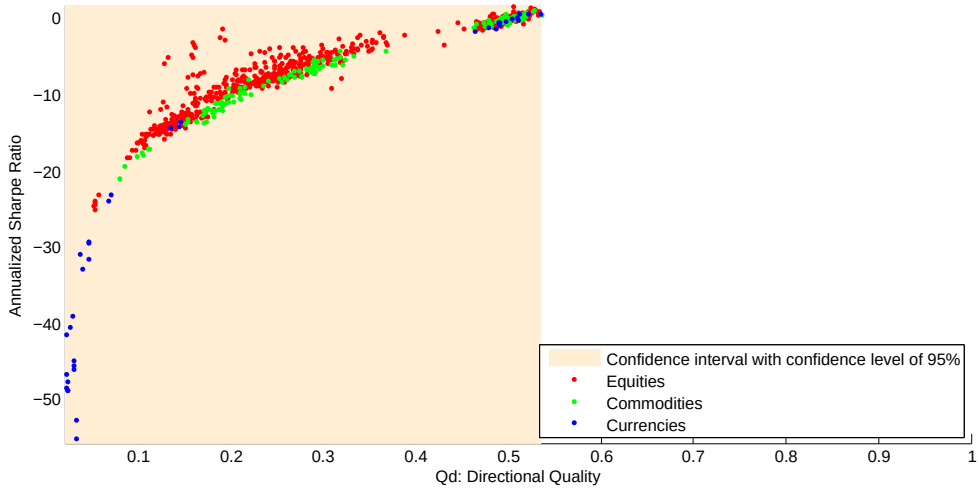
3.4.3. Results obtained with Q_d percentile performance measures

Since we noticed a big influence of the transaction cost on the final results and bands are not significant for most strategies, Figure 3.7 shows a scatter plot of the experiment when there is no transaction costs and without band. The y-axis is the Sharpe ratio and the x-axis is the Q_d percentile, p, as explained in Section 3.2.5. When p is close to one, the strategy is considered as good as it outperforms most of the associated random strategies. To help distinguish good strategies from bad ones, we included an arbitrarily set 95% confidence interval for the percentiles (shown in yellow).

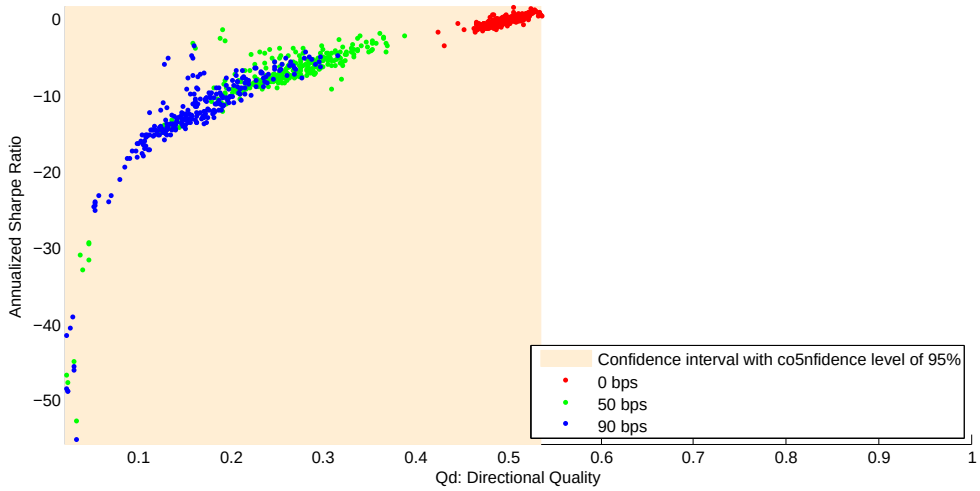
In Figure 3.7a we see that the trading strategy that is trend-following and includes the RSI II on top of it (#10 in Table 3.1) exhibits significantly better performances than its competitors. The percentile scores are really high and most of the results lie outside the 95% confidence interval. Since the only difference with this trading rule and all the other ones, is that it puts a



(a) Emphasis on trading rules

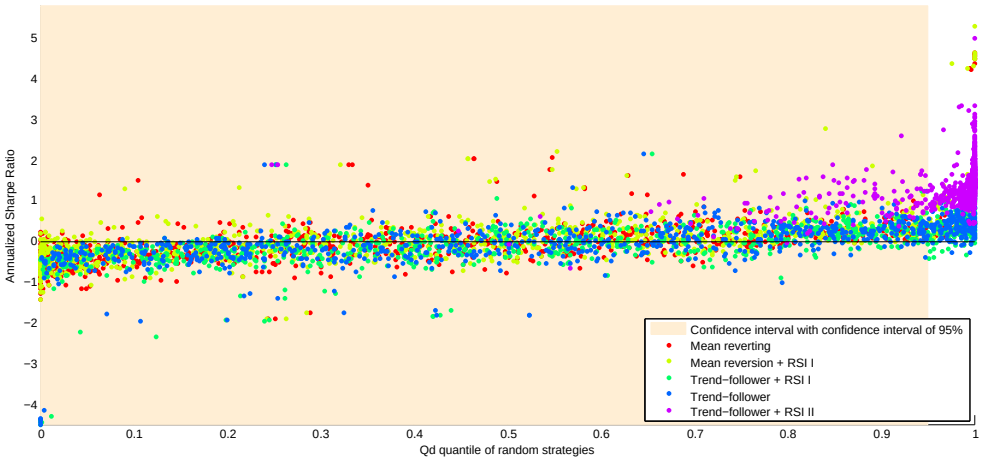


(b) Emphasis on asset classes

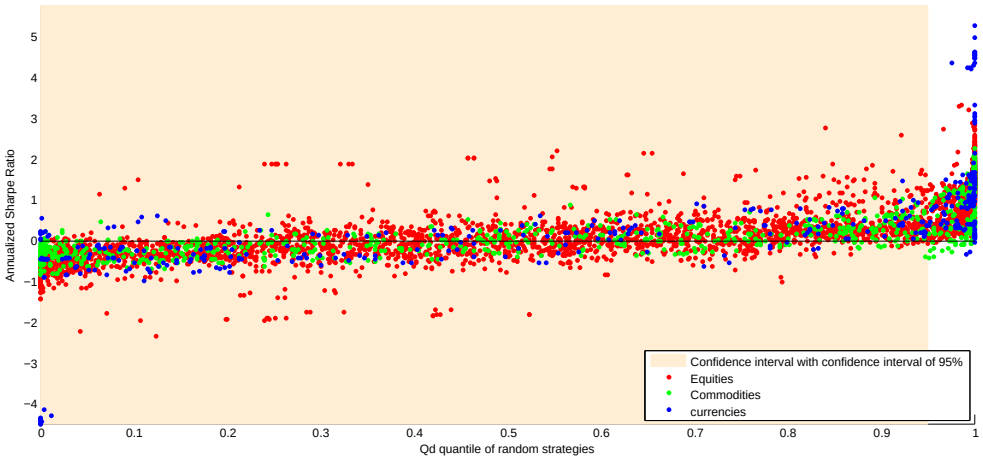


(c) Emphasis on transaction costs

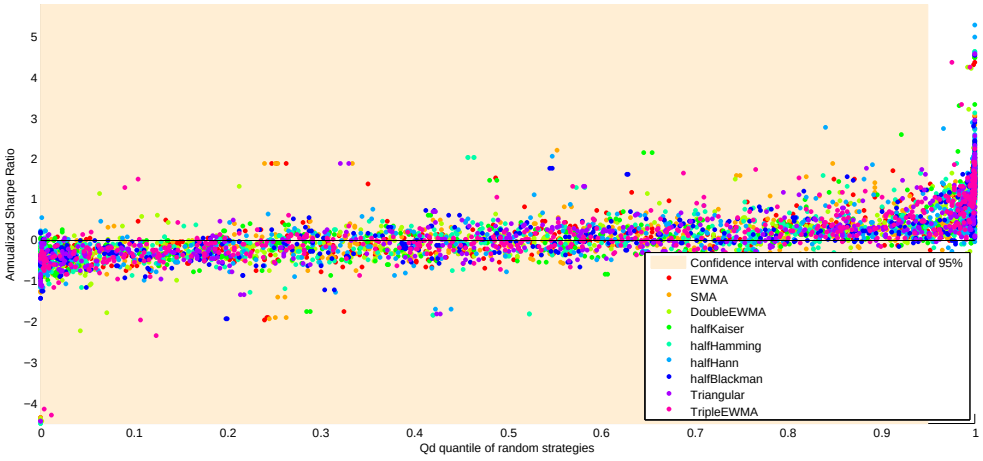
Figure 3.6.: Results for the trading strategies using the resonator.



(a) Emphasis on trading strategies



(b) Emphasis on asset classes



(c) Emphasis on filters

Figure 3.7.: Scatter plot showing the annualized Sharpe ratio in function of the Q_d percentile

Rule	EWMA		Resonator		SMA		DoubleEWMA		halfKaiser		halfHamming		halfHann		halfBlackman		halfTriangular		TripleEWMA	
#	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR		
1	0.5	-0.064	0.5	-0.038	0.5	-0.027	0.5	0.023	0.49	-0.15	0.5	-0.075	0.49	-0.086	0.49	-0.099	0.5	-0.064	0.5	0.074
2	0.5	0.11	0.51	0.6	0.5	0.13	0.51	0.21	0.51	0.15	0.5	0.09	0.53	0.82	0.5	0.1	0.5	0.11	0.51	0.31
3	0.49	-0.2	0.49	-0.59	0.49	-0.24	0.49	-0.36	0.49	-0.18	0.49	-0.19	0.49	-0.27	0.49	-0.15	0.49	-0.2	0.48	-0.5
4	0.5	0.049	0.51	0.085	0.5	0.042	0.5	-0.033	0.51	0.13	0.5	0.058	0.51	0.088	0.51	0.092	0.5	0.056	0.5	-0.083
5	0.56	1.4	0.48	-0.66	0.55	1.4	0.54	1.2	0.59	1.6	0.57	1.5	0.58	1.7	0.58	1.6	0.57	1.5	0.53	1

(a) No transaction cost

Rule	EWMA		Resonator		SMA		DoubleEWMA		halfKaiser		halfHamming		halfHann		halfBlackman		halfTriangular		TripleEWMA	
#	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR		
1	0.45	-0.7	0.26	-8.2	0.45	-0.65	0.47	-0.48	0.42	-0.99	0.44	-0.76	0.44	-0.79	0.43	-0.84	0.44	-0.74	0.47	-0.45
2	0.45	-0.57	0.26	-7.4	0.45	-0.54	0.47	-0.34	0.43	-0.71	0.45	-0.62	0.44	-0.12	0.44	-0.68	0.45	-0.61	0.47	-0.27
3	0.43	-0.98	0.24	-8.2	0.42	-1.1	0.43	-1.1	0.41	-1.1	0.42	-1	0.41	-1.2	0.42	-1	0.42	-1	0.43	-1.3
4	0.43	-0.73	0.27	-7.7	0.43	-0.84	0.44	-0.78	0.41	-0.8	0.43	-0.78	0.43	-0.77	0.42	-0.78	0.43	-0.78	0.44	-0.89
5	0.45	0.25	0.24	-8.2	0.44	0.097	0.45	0.098	0.43	0.25	0.44	0.23	0.43	0.29	0.44	0.26	0.44	0.22	0.45	-0.1

(b) Transaction cost: 50 bps

Rule	EWMA		Resonator		SMA		DoubleEWMA		halfKaiser		halfHamming		halfHann		halfBlackman		halfTriangular		TripleEWMA	
#	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR	Q_d	SR		
1	0.44	-1.2	0.16	-14	0.44	-1.1	0.46	-0.86	0.41	-1.6	0.43	-1.3	0.43	-1.3	0.42	-1.4	0.43	-1.2	0.46	-0.82
2	0.44	-1.1	0.17	-12	0.44	-1	0.46	-0.75	0.42	-1.3	0.44	-1.1	0.43	-0.81	0.43	-1.2	0.44	-1.1	0.46	-0.7
3	0.42	-1.5	0.16	-13	0.41	-1.7	0.42	-1.6	0.4	-1.7	0.41	-1.6	0.4	-1.8	0.41	-1.6	0.41	-1.6	0.42	-1.9
4	0.42	-1.3	0.17	-13	0.42	-1.5	0.43	-1.3	0.4	-1.5	0.42	-1.4	0.41	-1.4	0.41	-1.4	0.42	-1.4	0.43	-1.5
5	0.43	-0.6	0.15	-13	0.42	-0.82	0.44	-0.68	0.41	-0.73	0.42	-0.69	0.41	-0.74	0.41	-0.69	0.42	-0.69	0.44	-0.88

(c) Transaction cost: 90 bps

Table 3.3.: Table showing the averaged Q_d and SR for every filter and strategy for every transaction cost. Green (red) cells are when the average is significantly larger (smaller) than the corresponding one of the EWMA (the benchmark) using a one-sided t Welch test at a 95% confidence level.

lot of emphasis on the RSI indicator. Those results suggest that this indicator is indeed a very good forecasting indicator.

However, from Figure 3.7c and Figure 3.7b no significant clusters can be seen.

3.4.4. Confirming Graphical Observations with Statistical Tests

Since, no clear graphical evidence on the different performance of the filter can be inferred, Table 3.3 shows the mean Sharpe ratios and mean Q_d for all the different tested filters, trading rules and transaction costs. It is essentially a different representation of Figure 3.5c, Figure 3.5a and Figure 3.5d. Green cells indicate that the hypothesis $H_0 : \mu_{EWMA} = \mu_{filter}$ can be rejected in favor of $H_a : \mu_{EWMA} < \mu_{filter}$ with a p-value less than 5%. On the other hand, red cells indicate that the associated H_0 can be rejected in favor of $H_a : \mu_{EWMA} > \mu_{filter}$. Tests are done using the Welch's Test (Section 3.2.6).

In table 3.3a, very few Q_d are above 52% indicating that the hypothesis of randomness as introduced in Section 3.2.4 can not be rejected. Except for the last strategy, the EWMA is mostly always superior than any other filter. However, Strategy #5 is the best performing strategy and it is interesting to see that all half window filters significantly perform better than the EWMA.

When transaction costs are added, the Double and Triple EWMA seem to be the best performing filter and mostly all the other possibilities are worse than the EWMA. Nevertheless, when transaction costs are included, performances in general are quite disappointing as they never reach a $Q_d = 50\%$ and very rarely a positive Sharpe ratio. In those conditions, even if some new filters significantly outperform the EWMA, it does not make sense to comment further more on them as their performance is not good enough. Note also, that most of the windows (be it on the Sharpe ratio or Q_d side) significantly underperform the EWMA, suggesting that

EWMA can not be outperformed with windows when transaction cost of those sizes are added.

3.4.5. Strategy #5 without transaction cost

Strategy #5 is the best performing strategy when transaction costs are not included. In this section, we take a closer look at the distribution of its results. As already suggested by Table 3.3a, the performance of the half windows beating the EWMA can be seen in Figure 3.8b. Moreover, most of the results obtained with half windows lie outside the confidence interval. From Figure 3.8c, one can see a clear evidence that the larger is the band the better is the performance. Hence – according to our results – using a half window (or more generally a one-parameter inverted logistic shaped filter) and a tolerance band gives the best forecasting results and outperforms both the SMA and EWMA. Also, half windows have very similar performances, this can be partly explained graphically looking at Figure 2.15 where, when calibrated, the impulse response of the different half-windows look alike.

3.4.6. Mean Trade Per Return and the Problem of Transaction Cost

As explained earlier, all strategies including transaction costs are not satisfying. However, the choice of the value of the transaction costs (50 and 90bps) is quite conservative and it is not rare to see transaction costs that are lower than those values. Thus, to remove this additional parameter, we now examine our results with their mean return per trades.

Looking at Figure 3.9a, trend-following strategies have mostly Sharpe ratios higher than zero whereas mean-reverting strategies show the inverse pattern. This is understandable since their trading rules are the inverse of each other. However, resulting trading strategies are not reversed because they are always optimized for the best set of parameter in the training window. This plot, strongly argues in favor of trend-following strategies. Note that we could not see those patterns before since the addition of bands and different transaction costs blurred the plot.

Figure 3.9b clearly shows that the investigated trading strategies perform poorly on currencies. The other two asset classes have slightly better results. Here again, very good performing strategies come from small time-series where the number of trade is small, Figure 3.11b.

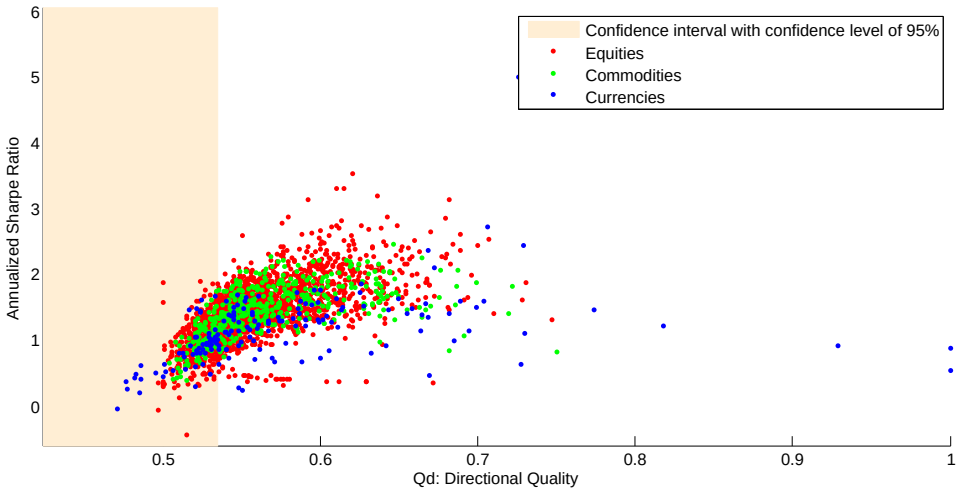
Here again, no cluster with a specific filter can be seen here, Figure 3.9c.

When looking at Figure 3.10, the same remarks as in Figure 3.9 can be drawn. It is interesting to see that in Figure 3.10a, strategies #5 have similar mean return per trade than the other trend following strategies even if their Sharpe ratio is significantly better. Moreover, it can be seen that if transaction set were 5bps or less mostly all trend-following strategies would be profitable.

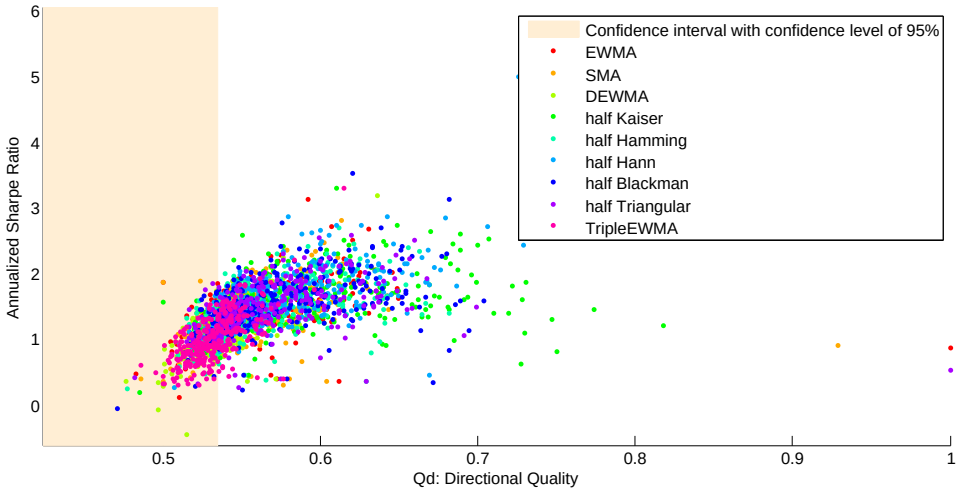
3.4.7. Conclusion

From the results of the previous section, it seems that simple trading rules generally have quite poor performances and do not represent a good investment strategy. Moreover, in general no clearly differentiated cluster could be found between different asset classes, tolerance bands, strategy or filter.

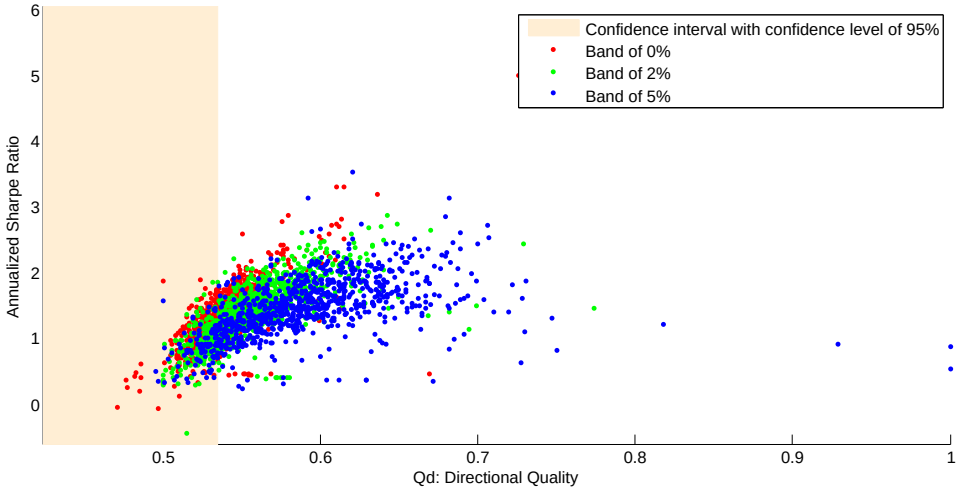
However, as expected, transaction costs play a significant role in the performance results and when looking only at strategies without transaction costs, rule #5 outperformed all the others. Strategy #5 heavily relies on the RSI indicator suggesting that the RSI is a powerful indicator and actually maybe better than the crossing of moving averages we used. For the same rule, all the half windows deliver significantly better results than the EWMA and the SMA and thus



(a) Emphasis on asset classes

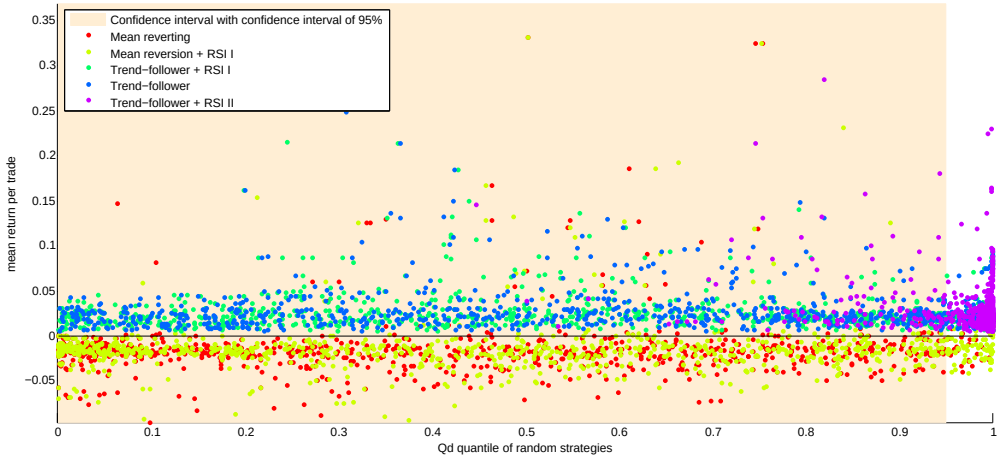


(b) Emphasis on filters

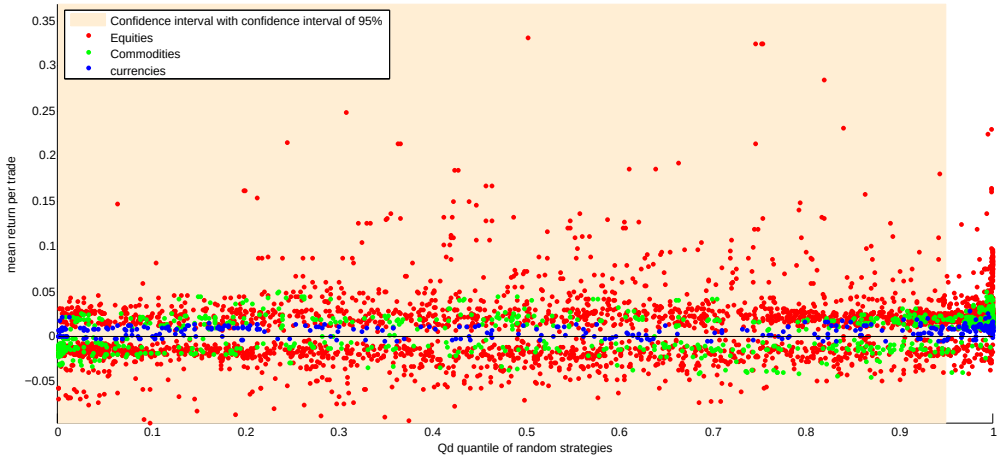


(c) Emphasis on Bands

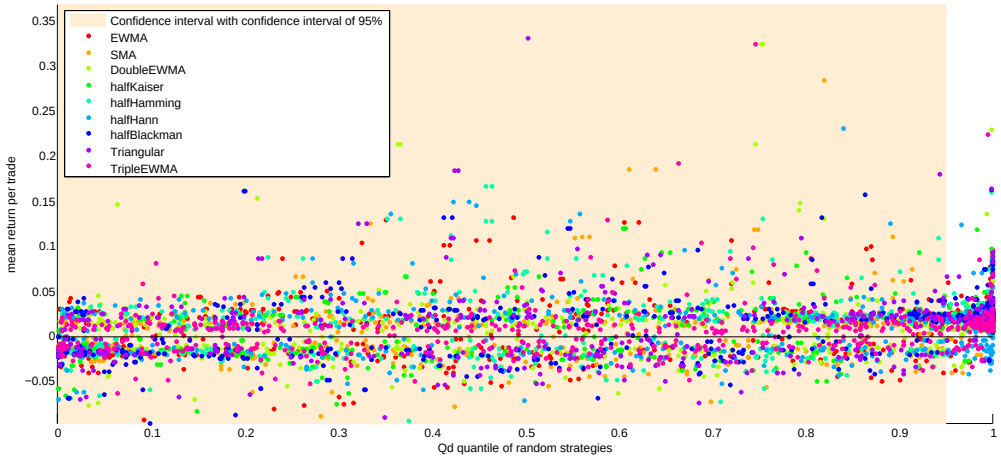
Figure 3.8.: Results for the best performing trading strategy when transaction costs are set to zero.



(a) Emphasis on trading strategies

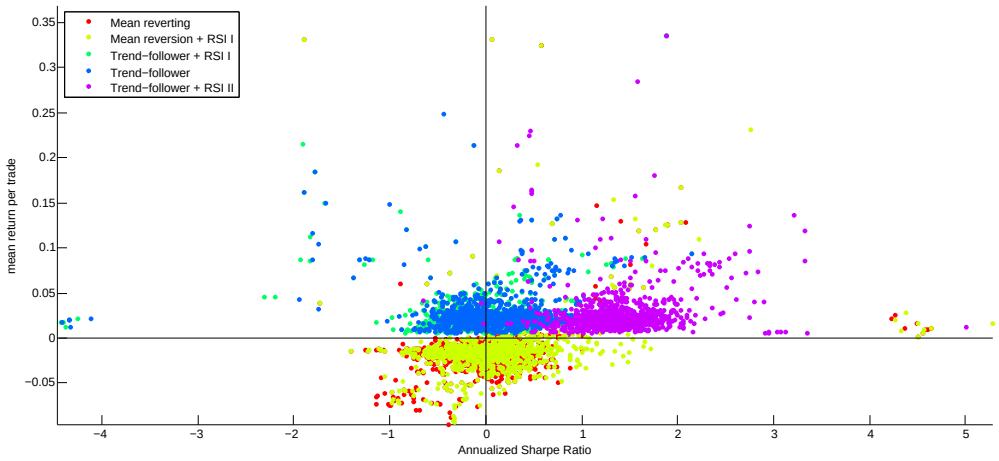


(b) Emphasis on asset classes

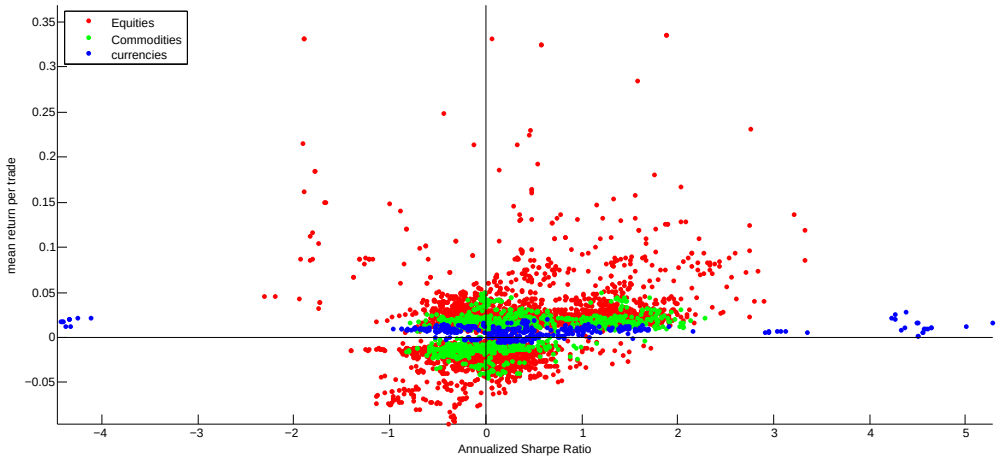


(c) Emphasis on filters

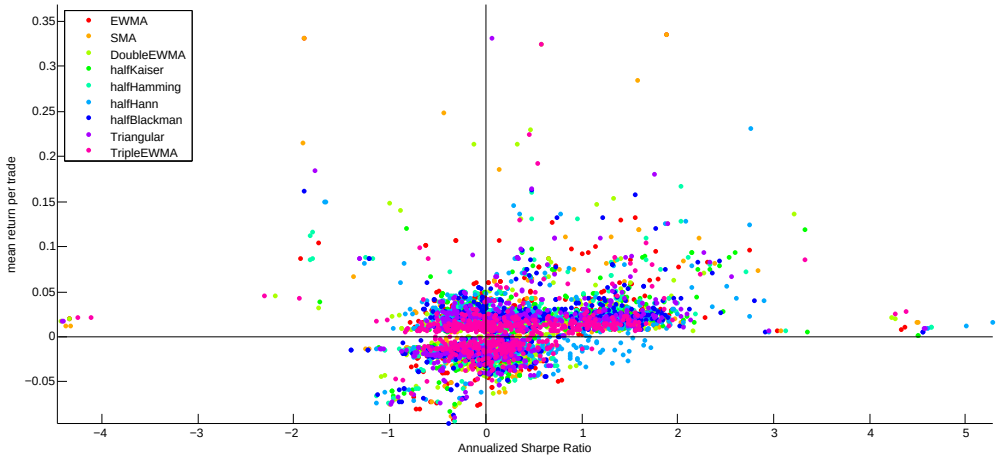
Figure 3.9.: Scatter plot showing the mean return per trade in function of the Q_d percentile



(a) Emphasis on trading strategies

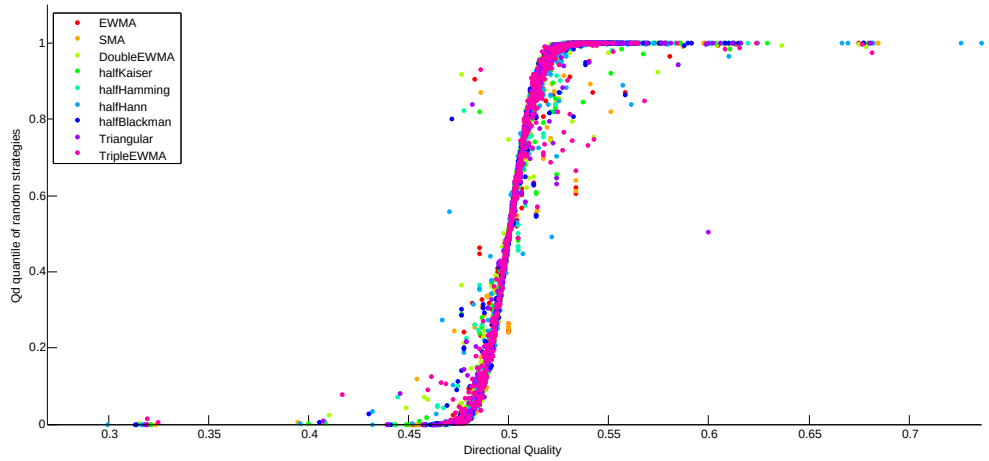


(b) Emphasis on asset classes

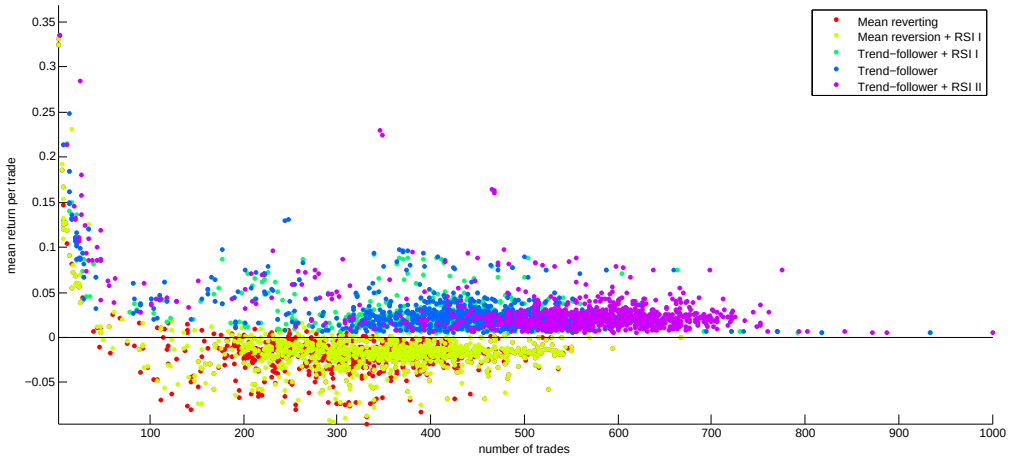


(c) Emphasis on filters

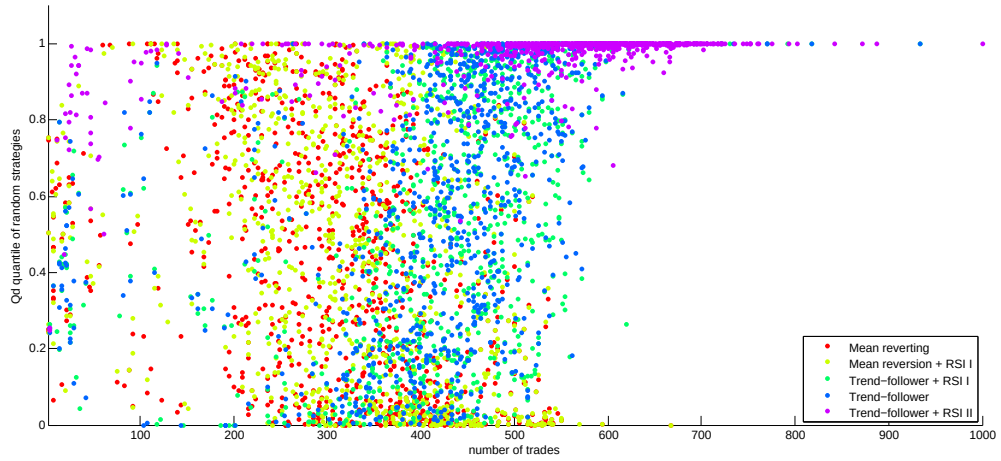
Figure 3.10.: Mean Return per Trade versus Sharpe ratio for the experiments without transaction cost and without band.



(a) Distribution of the Q_d percentiles in function of the realized Q_d



(b) Mean return per trade versus number of trades



(c) Q_d versus number of trades

Figure 3.11.: Scatter plot representing the outcome of the experiments without transaction cost and without band.

could potentially be substituted for the computation of trend indicators or as input to more sophisticated trading strategies.

When looking at three performance dimensions: Sharpe ratio (risk), mean return per trade (profit and loss) and Q_d percentile (relative performance), it can be seen that trend-following strategies exhibit much better performances on the profit and loss side than mean reverting strategies. Even though, strategy #5 was the best with Sharpe ratio and Q_d, its mean return per trade is comparable to other trade following strategies.

Finally, the results obtained with the resonator corroborate the fact that predicting algorithm with many parameters often have the unpleasing effect of overfitting.

A list of the exhaustive best results is given in the Appendix in Table A.5.

4. Substitution of the EWMA with other Moving Averages for the Computation of the Value at Risk

In risk management, it is of great importance to have a coherent and widely recognized measure to assess financial market risk. The Value at Risk (VaR) framework intends to fulfill this need. It was first introduced in 1989 by J.P. Morgan to have a concise view of the risks of the firm [Financial Times]. The aim of the VaR is to have a single figure that gives an immediate and understandable overview of the risk of a certain financial instrument. In the 90s, the VaR quickly established itself as a benchmark in the financial industry, [Jorion, 2001, McNeil, 2005].

In this chapter, we investigate the substitution of the Exponentially Weighted moving Average (EWMA) with other moving averages for the computation of VaR within the RiskMetrics^{TM} framework.

4.1. Value at Risk

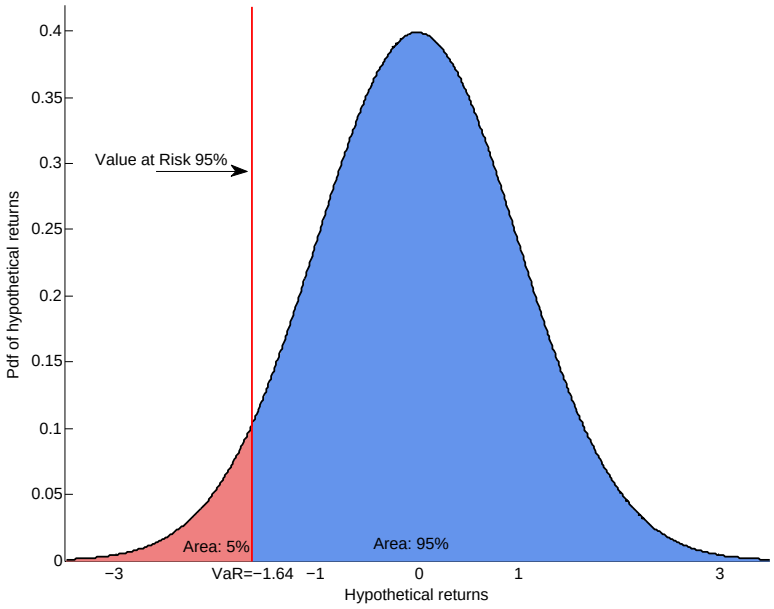


Figure 4.1.: Illustration of the VaR-95% measure on a standard normal distribution.

The Value at Risk (VaR) is defined as the α quantile of a return distribution and gives an indication of the loss threshold at an α-level one should expect when investing in a certain instrument or portfolio over a pre-defined time horizon (Figure 4.1). Typically, $1 - \alpha$ is chosen

to be either $1-\alpha = 95\%$ or $1-\alpha = 99\%$. Hence, in the case of the $VaR_{95\%}$, $1-\alpha = 95\%$ of the returns for the considered period should be above the value and $\alpha = 5\%$ below.

When computing the daily VaR, one needs to forecast the return distribution for the next day. Several approaches exist. Among them it is worthwhile to cite the historical method, the EWMA (see section 4.2), GARCH models [Engle, 1982], Monte-Carlo simulation [Glasserman et al., 2000] or multi-fractals [Boisbourdain, 2008].

Most of those methods require two steps. In a first step, the variance of the returns is forecasted. Then, from this estimate and assumptions about the shape of the distribution, the α-quantile that represents the VaR is inferred.

In this study, we will solely focus on the RiskMetrics^{TM} framework, [J.P. Morgan and Reuters, 1996], as it is widely used in financial institutions throughout the world and uses the EWMA to forecast the variance. We then investigate if substituting the EWMA with reverted logistic-shaped moving averages such as the windows introduced in section 2.7 yields more accurate results.

4.2. Risk Metrics Framework

A few years after the introduction of VaR, in 1994, J.P. Morgan introduced the RiskMetrics^{TM} methodology with the aim to "[..] promote greater transparency for market risk measurement" and "[..] to establish a benchmark for market risk measurement" [J.P. Morgan and Reuters, 1996]. A brief overview of the main ideas of the methodology is explained in the next section. For exhaustive details on the matter, the reader is invited to consult [J.P. Morgan and Reuters, 1996].

4.2.1. Assumptions of the RiskMetrics^{TM} model

This section presents the RiskMetrics methodology for the computation of a one-day VaR forecasting as well as the underlying statistical assumptions. RiskMetrics^{TM} assumes that the dynamic of prices follows a random walk [Brandimarte, 2006] :

$$\ln P_t = \ln P_{t-1} + \mu + \sigma_t \epsilon_t. \tag{4.1}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$. Consequently this model does not take into account the heavy-tailed character of most return time-series. However, the assumption that returns are normally distributed is useful because only the first and second moments are required to describe the whole distribution. Moreover, the methodology also assumes that the squared returns are autocorrelated (see Section 4.2.6) and that $\mu = 0$. With all this, one can rewrite the dynamics of the price described in Equation 4.1 as

$$\begin{aligned} r_t &= \sigma_t \epsilon_t & (4.2)\\ \epsilon_t &\sim \mathcal{N}(0,1) & (4.3) \end{aligned}$$

where r_t are computed continuously, $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$, and ϵ_t follows a Gaussian white noise process.

4.2.2. Variance forecasting

The first step of the methodology is to forecast the variance of the returns. For this, RiskMetrics^{TM} applies the EWMA on the squared returns, meaning that recent observations will be given more

weights and the weights decrease exponentially. Hence the variance predicting equation looks like:

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2. \tag{4.4}$$

If one develops the recursive term, Equation 4.4 can also be written

$$\sigma_{t+1}^2 = (1-\lambda)\sum_{k=0}^{t}\lambda^k r_k^2. \tag{4.5}$$

It is easy to notice that Equations 4.4 and 4.5 are very similar to Equations 2.32 and the impulse response 2.33. Indeed they are the application of an EWMA filter on the input r^2 and output σ^2 as introduced in Section 2.6.1, but lagged, since we do not know the value of the return at time $t+1$ and thus their impulse response is

$$h_{\text{EWMA lagged}}[n] = h_{\text{EWMA}}[n] * \delta[n-1]. \tag{4.6}$$

For the sake of convenience, we will now refer to Equation 4.6 when mentioning the EWMA throughout this chapter.

4.2.3. New Moving averages investigated

As mentioned above, the RiskMetrics$^{\text{TM}}$ methodology is divided into two steps: forecasting the variance and then inferring the quantile from the variance. To forecast the variance, the RiskMetrics$^{\text{TM}}$ framework uses the EWMA but, as explained in the previous chapter, the EWMA is nothing more than a low-pass digital filter. Thus it can be substituted with other filters with similar functions to see if the forecasting power of those other moving averages is better.

We test the substitution of the EWMA with the SMA and half Windows. Figure 4.2 shows the impulse responses of the different moving averages that we tested against the EWMA. We see that the EWMA sharply decays whereas windows take more time to decrease, thus suggesting that past observations will be more relevant to the output. It is important to note that those filters have all been extensively used as windows in digital filtering (see Section 2.7) and are also one-parameter filters [1].

4.2.4. Comparison of filters

The RiskMetrics methodology recommends a value of $\lambda = .94$ for the EWMA parameter to compute the daily var. For details on why this particular value was chosen for the daily VaR, the reader is invited to consult [J.P. Morgan and Reuters, 1996] p.99-100. As we described in Section 2.6.1 it is equivalent in terms of energy to an SMA of length of $N \approx 32$ days. Hence all half-windows were calibrated so that they have the same energy as the SMA with $N = 32$. The resulting impulse responses are shown in Section 2.7.4.

4.2.5. Inferring VaR from the forecasted variance

The definition of VaR states that it is the quantile at a certain level (α) of the return distribution. The RiskMetrics$^{\text{TM}}$ assumes that the VaR can be retrieved by using the mean, μ, the variance,

[1] The Kaiser filter has two parameters, but in this chapter we restrain the second parameter to $\beta = 4$ as in Figure 4.2

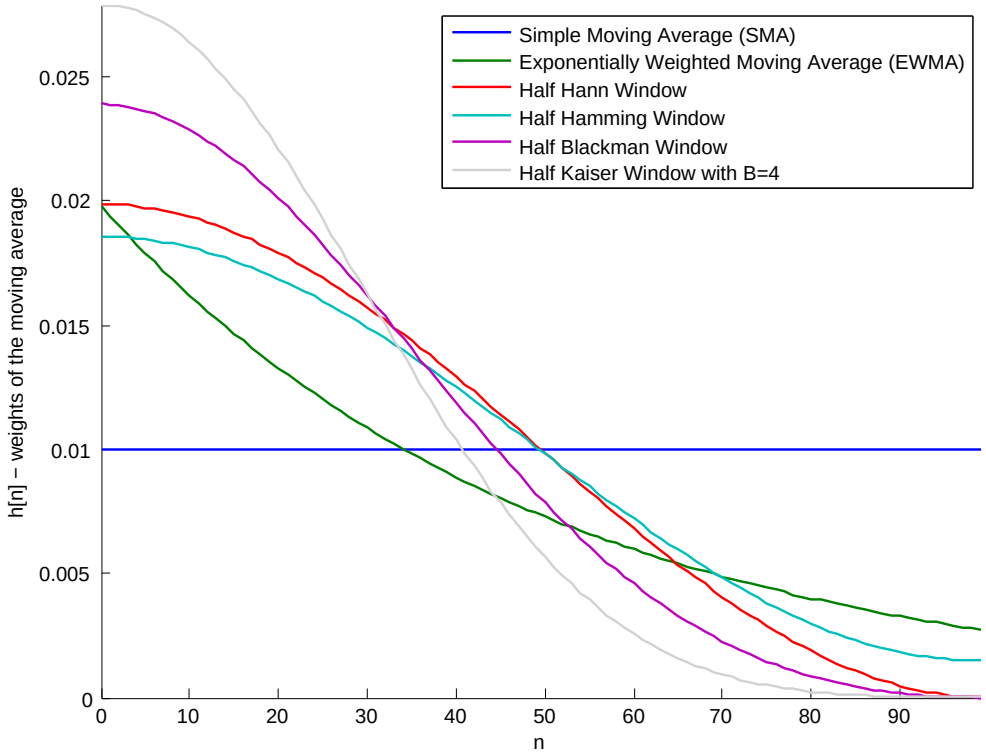


Figure 4.2.: Impulse function $h[n]$ (weights) of the different moving averages investigated. Note: moving averages are shown so that the distinctive shape of their impulse response is shown and are not calibrated to have the same energy.

σ^2 and the α-quantile of the distribution of the residuals ϵ_t that follows a centered normal distribution. The VaR is then given as

$$\underset{1-\alpha}{VaR}(t+1) = \mu + q_\alpha(\epsilon_{t+1})\sigma_{t+1} \tag{4.7}$$

$1-\alpha$	$q_\alpha(Z) = \Phi^{-1}(\alpha)$
68.3%	−1.00
95.0%	−1.64
97.5%	−1.96
99.0%	−2.33
99.5%	−2.58

Table 4.1.: List of some quantiles of $Z \sim \mathcal{N}(0,1)$

Since residuals ϵ_t are normally distributed, $\epsilon_t \sim \mathcal{N}(0,1)$ (from Equation 4.2), the α-quantile is by definition simply the α-quantile of the standard normal distribution or $q_{1-\alpha}(\epsilon_t) = \Phi^{-1}(1-\alpha)$ where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function. As a quick reference, we listed some values of $\Phi^{-1}(1-\alpha)$ in Table 4.1. For μ_{t+1}, which is the expected

value of the returns at time $t+1$, we can use equation 4.2 and equation 4.3 to show that

$$\mu_{t+1} = \mathbb{E}(r_{t+1}) = \mathbb{E}(\sigma_{t+1}\epsilon_{t+1}) = \sigma_{t+1}\mathbb{E}(\epsilon_{t+1}) = 0. \tag{4.8}$$

Hence to compute a daily $VaR_{1-\alpha}$ using the RiskMetrics^{TM} method, we have

$$\underset{1-\alpha}{VaR} = \Phi^{-1}(1-\alpha)\sigma_{t+1} \tag{4.9}$$

where σ_{t+1} is the estimated variance from the EWMA and $\Phi^{-1}(1-\alpha)$ the quantile from the inverse cdf.

The problem with the normality assumption is that it underestimates the heavy-tailed character of a typical return time-series; [Cont, 2001] lists this property as a typical stylized-fact. Thus, in addition to the normality assumption, we also relax assumption 4.2: the residuals, ϵ_t, are still assumed homoskedastic and following a strong white noise process but with an unknown distribution. Thus, for every new time-step where the daily VaR is to be computed an empirical distribution of ϵ_t is computed. To do this, the last 200 observed returns were taken and divided by the associated past forecasted variances to retrieve ϵ_t. By constructing an Kaplan-Meier empirical distribution of the residuals, the quantile of interest $q_\alpha(\epsilon_{t+1})$ can easily be found. To compute the quantile of the residual distribution, the last 200 observed returns r_t were taken and the estimation of the variance and the estimated variance $\hat{\sigma}_t^2$ and using Equation 4.2 obtained the last 200 estimated residuals $\hat{\epsilon}_t$. The expected value of the residuals can be estimated as

$$\hat{\mu}_{t+1} = \hat{\mathbb{E}}(r_{t+1}) = \hat{\mathbb{E}}(\sigma_{t+1}\epsilon_{t+1}) = \sigma_{t+1}\hat{\mathbb{E}}(\epsilon_{t+1}) = \sigma_{t+1}\bar{\epsilon}. \tag{4.10}$$

Practically $\bar{\epsilon}_t$ is computed as the average over the last 200 k of ϵ_{t-k} ($\bar{\epsilon}_t = \frac{1}{N}\sum_{k=0}^{N-1}\epsilon_{t-k}$) and the quantile. Figure 4.3 shows the the evolution of returns for a security where the VaR-95% is computed using the methods explained above, one with historical data and one with the normality assumption.

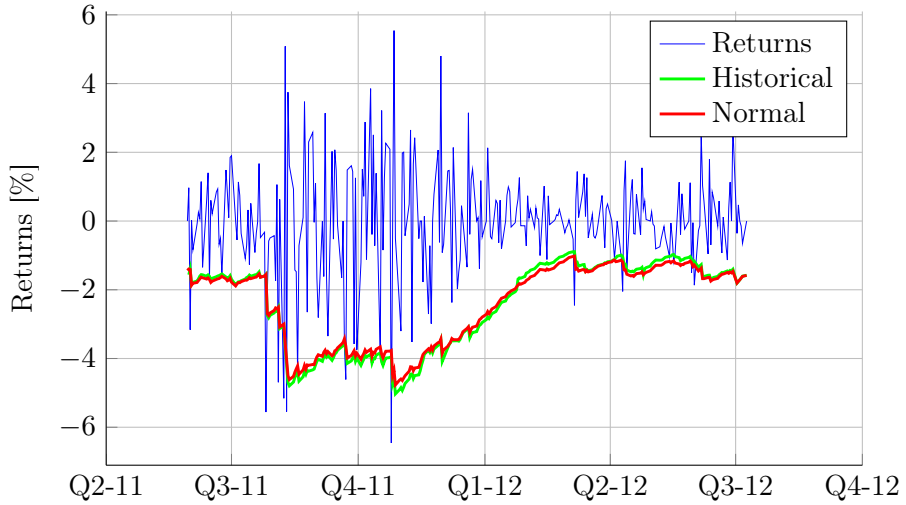
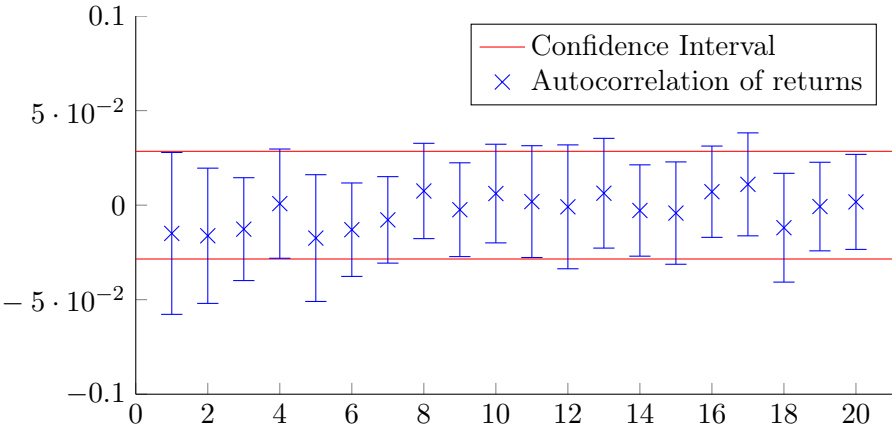
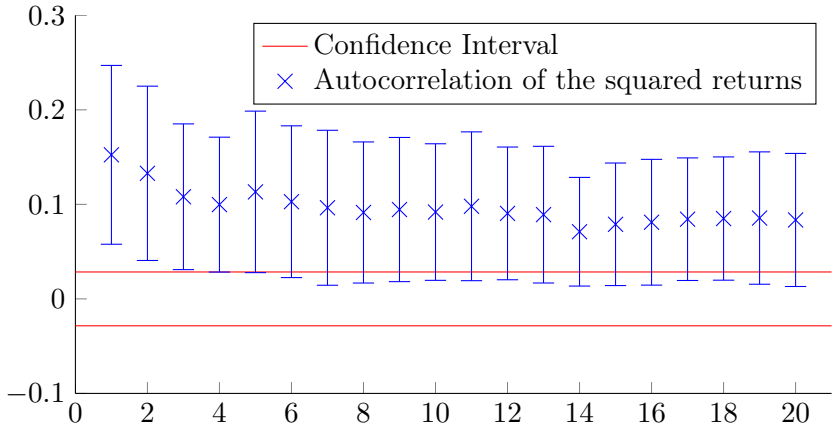


Figure 4.3.: Illustration of the VaR-95% computation using the RiskMetrics methodology (assuming normality and with historical data). As most return time-series it exhibits heteroskedasticity. The time-series is 3M Company between fall 2011 and summer 2012.



(a) Autocorrelation of returns



(b) Autocorrelation of squared returns

Figure 4.4.: Autocorrelation of returns and squared returns of portfolio of 109 securities over a period of 10 years. For each security, coefficients are averaged and the error bars represent the standard deviation of the coefficients. For clarity, $R_x(0) = 1$ is not shown on the graph

4.2.6. Autocorrelation

By using Equation 4.4 to forecast future volatility σ_{t+1}, one implicitly assumes that the output is linearly correlated with the input of the filter or that squared returns at least partly explain the future squared returns which in turn is linearly related to volatility. To show that this actually makes sense, we computed the autocorrelation for all the securities present in our portfolio. The autocorrelation gives information about the similarity between observations separated by a

certain time interval. It is defined as[2]

$$R_x[k] = \underset{x}{autocorr}[k] = \sum_n x[n]x^*[n-k]. \tag{4.13}$$

Hence, $R_x(k)$ tells how much information in the signal can be inferred from an observation $x[n-k]$ to explain $x[n]$. $R_x(0)$ always equal to one as all observations $x(n-0)$ explain entirely $x(n)$. For every security in our portfolio, we computed the first 21 coefficients of the autocorrelation. In Figure 4.4b we show the distribution of the coefficients across the portfolio by displaying their mean and associated standard deviation. We see that squared returns are indeed significantly correlated since most of the coefficient means and their standard deviation lie outside of the confidence interval (Confidence bounds are computed using the approach in [Box and Reinsel, 1994] equations 2.1.13 and 6.2.2, pp. 33 and 188). Hence, it does make sense to try to predict future squared returns since information about the future is actually contained in lagged observations.

On the other hand, in our price dynamics from Section 4.2.1 we also assumed through Equation 4.3 that returns follow a white noise. This is supported by Figure 4.4a which essentially shows the same as Figure 4.4b but for returns rater than squared returns. All coefficient means lie in the interval. Additionally most of the standard deviations also lie in the confidence interval, this strongly suggests that autocorrelation is mostly absent from returns and acts in favor of the white noise assumption. From this statement, one can also imply that it is hard to predict stock market returns using linear methods such as the ones we used in Chapter 3 (at least for the securities included in our portfolio and at a daily frequency). Those two remarks are also expounded as a stylized fact in [Cont, 2001].

4.2.7. Comparison with GARCH models

General Autoregressive Conditional Heteroskedasticity (GARCH) models were introduced by Bollerslev [Bollerslev, 1986] in 1986 after the introduction of ARCH models by Engle in 1982 [Engle, 1982]. They try to overcome the problem of non-constant variance over-time in typical return time-series (see Figure 4.3). The GARCH model is made of two parts: the autoregressive part and the moving average part. The GARCH(p,q) can be defined as in [Farkas, 2011]:

$$\sigma_t^2 = \underbrace{\underbrace{\alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2}_{ARCH} + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2}_{GARCH}. \tag{4.14}$$

It is interesting to note that the EWMA is a special case of GARCH(1,1) with $\alpha_0 = 0, \alpha_1 = (1-\lambda)$ and $\beta_1 = \lambda$, [Farkas, 2011]. Moreover, comparing Equation 4.14 and 2.4, it is easy to see that the GARCH model can be seen as a digital filter with inputs r_t^2 and outputs σ_t^2 and $\alpha_0 = 0$. So,

[2] Autocorrelation is a correlation between two similar signals. Correlation in turn is the convolution of a first signal with a reversed second signal so that

$$\underset{x,y}{corr}(\tau) = (x * y(-\cdot))(\tau) \tag{4.11}$$

and

$$\underset{x}{autocorr}(\tau) = \underset{x,x}{corr}(\tau). \tag{4.12}$$

Name	Properties	Parameters adjusted to EWMA
EWMA	Benchmark, Exponentially decreasing	$\lambda = .94, N = 32$
SMA	Constant weight	$N = 32$
half Triangular	Linearly decreasing	$M = 42$ or $N = 32$
half Hann	Linearly decreasing	$M = 48$ or $N = 32$
half Hamming	Linearly decreasing	$M = 44$ or $N = 32$
half Blackman	Linearly decreasing	$M = 56$ or $N = 32$
half Kaiser	Logistic-shaped	$M = 65$ or $N = 32$, and $B = 4$

Table 4.2.: List of investigated filters, associated properties and the value of the parameter when energy-adjusted with the EWMA with $\lambda = .94$ or $N = 32$.

all filters we investigated are special cases of a GARCH model and, more specifically, FIR filters are special cases of ARCH models. However, the approach is different. Practitioners usually fit GARCH models on past data using a maximum likelihood approach whereas we try to beat the EWMA within a certain framework when the parameter is given. Hence, in this study we do not try to beat GARCH models but rather try to come up with improved VaR estimates than the RiskMetrics^{TM} methodology using the same assumptions. Note also that since half windows are FIR filters, the autoregressive part is missing and thus represent a constrained ARCH model.

4.3. Backtesting Methodology

In order to be able to compare the different filters, we computed the daily VaR for each day for a portfolio of 109 securities over a period spanning 10 years (2002-2012). More information about the content of the portfolio is given in Section A.4.1 of the Appendix. We computed the VaR using six different moving averages, the benchmark being the EWMA. A list of the moving averages and their properties is given in Table 4.2, their impulse responses are depicted in Figure 2.15. In order to avoid unwanted transient effects, the VaR is only computed after the 200th observation. In this case, for the EWMA the tolerance level is very low $\Omega < 5 \cdot 10^{-5}$. For the FIR filters the tolerance Ω equals zero because their length is always smaller than 32.

Once the VaR is estimated, we assign a violation binary variable for each day. The value is one if the return is lower than the VaR (VaR violation) and zero otherwise. Thus repeating the process for every day of a time-series and taking the mean of the violation variables should eventually lead to α if the estimate is perfect.

This can be described as a bernoulli process where $X(t)$ is a binary variable denoting the violation. The distribution looks like:

$$f(X_i|\alpha) = \begin{cases} \alpha^{X_i}(1-\alpha)^{1-X_i} & X_i = 0, 1 \\ 0 & \text{else} \end{cases}. \tag{4.15}$$

and thus $\mathbb{E}(X_i) = \alpha$ where α is the tolerance defined in the VaR.

In order to estimate $\mathbb{E}(X_i)$, we compute the average $\overline{x}_i = \sum_{k=1}^{T} x_{ik}$ for every instrument where x_{ik} is the violation at time step k for instrument i. Moreover, since the time-series are long (mostly about 10 years) the average over the whole sample might be misleading. Some regions might largely overestimate VaR and some other largely underestimate it but the mean over the whole sample might give a good estimate. So in addition to the mean $\overline{x}_i$, we also

generate a vector of means of random resampling (bootstrapping) for a single security

$$\tilde{x}_{il} = \sum_{k \in \Gamma_l} x_{ik} \tag{4.16}$$

where Γ_l is a set containing randomly drawn indices. Finally, in order to assess the dispersion of the estimates, the standard deviation for each instrument is calculated as

$$s_{x_i} = \sqrt{\frac{1}{L-1} \sum_{l=1}^{L} (\tilde{x}_{il} - \overline{x}_i)^2}. \tag{4.17}$$

The reason why the standard deviation is not directly computed from the x_i is that the dispersion of the individual day violation is not of interest to us. We are interested in knowing whether the violations over the whole sample are more or less constant or if there are big jumps. Hence, a security on which the VaR computation worked well should have an estimated mean of $\overline{x}_i = \alpha$ and a standard deviation close to zero. A large standard deviation would mean that the VaR computation is not robust over the whole sample and thus small values of s_{x_i} are favored. In order to be able to see if one of the tested filters improve on the performance of the EWMA, we computed the mean violation achieved with the filter of interest and compared it to the benchmark. To see if the result is significantly different, we used the Welch test again (Section 3.2.6).

4.4. Results

We first display the results for every security on a scatter plot with the average violation on the x-axis and its associated standard deviation on the y-axis for two different α-levels: 1 and 5%. See Figures 4.6 and 4.5.

Results when the α-level is 5% (Figure 4.5) mostly have a mean $\overline{x}$ that is in the interval $[\alpha \pm 1\%]$ and a standard deviation that is below $s_{\overline{x}} = 0.6\%$, suggesting that the spread of violations for the same security is pretty robust over time. Looking at the different clustering, nothing can be inferred from the scatter plots. It seems that different asset classes or different filters produce similar results. However, the results of the VaR computation using the historical assumption seem to be shifted to the right (higher mean) compared to the ones using the normality assumption.

For $\alpha = 1\%$, results are similar to $\alpha = 5\%$. Most of the observations are confined in the interval $\overline{x} \in [\alpha \pm 1\%]$ and $s_{\overline{x}} < 0.6$. Nonetheless, the difference between the computation using the historical assumption and the normality assumption is more pronounced; it seems clear from the graph that the historical assumption yields much poorer results.

For both the $\alpha = 1\%$ and $\alpha = 5\%$ level, it is interesting to note that the standard deviation seem to depend on the length of the sample. The longer the sample, the smaller is the standard deviation (this is supported with Figures 4.5d and 4.6d). One of the reasons why such a behavior is exhibited is mainly because when doing the resampling, the sample is too small to produce good mean estimates and thus the associated standard deviation estimate is not reliable. This also explains the outliers on the other scatter plots that represent exclusively time-series with less than thousand points.

In order to assess the performances for the different filter more precisely, histograms of the violations of the different filters are shown in Figures 4.7, 4.8, 4.9 and 4.10. In general, the distribution is pretty much centered and symmetric around the α-level. However, for $\alpha = 1\%$

Filter	$\overline{x}_{\text{violations}}$	$s_{\text{mean violations}}$	t	p-value
EWMA	0.01	0.0033	0	0.5
SMA	0.011	0.0031	0.48	0.32
Half Kaiser	0.01	0.0032	-0.66	0.25
Half Hamming	0.01	0.0032	-0.12	0.45
Half Hann	0.01	0.0032	0.036	0.49
Half Blackman	0.01	0.0032	0.024	0.49
Half Triangular	0.01	0.0032	0.0092	0.5

(a) Normal, $\alpha = 1\%$. See Figure 4.7

Filter	$\overline{x}_{\text{violations}}$	$s_{\text{mean violations}}$	t	p-value
EWMA	0.048	0.0065	0	0.5
SMA	0.047	0.0065	-0.34	0.37
Half Kaiser	0.047	0.0058	-0.72	0.24
Half Hamming	0.047	0.0064	-0.38	0.35
Half Hann	0.047	0.0063	-0.23	0.41
Half Blackman	0.047	0.0064	-0.34	0.37
Half Triangular	0.047	0.0063	-0.32	0.37

(b) Normal, $\alpha = 5\%$. See Figure 4.9

Filter	$\overline{x}_{\text{violations}}$	$s_{\text{mean violations}}$	t	p-value
EWMA	0.018	0.0034	0	0.5
SMA	0.02	0.0038	3.3	0
Half Kaiser	0.02	0.0039	4.4	0
Half Hamming	0.019	0.0035	2	0.022
Half Hann	0.019	0.0036	2.1	0.017
Half Blackman	0.019	0.0036	1.9	0.031
Half Triangular	0.019	0.0036	1.9	0.032

(c) Historical, $\alpha = 1\%$. See Figure 4.8

Filter	$\overline{x}_{\text{violations}}$	$s_{\text{mean violations}}$	t	p-value
EWMA	0.05	0.0058	0	0.5
SMA	0.052	0.0057	3	0.0017
Half Kaiser	0.054	0.0052	4.9	0.00
Half Hamming	0.051	0.0058	1.7	0.049
Half Hann	0.051	0.0058	1.9	0.032
Half Blackman	0.051	0.0057	1.6	0.058
Half Triangular	0.051	0.0058	1.9	0.033

(d) Historical, $\alpha = 5\%$. See Figure 4.10

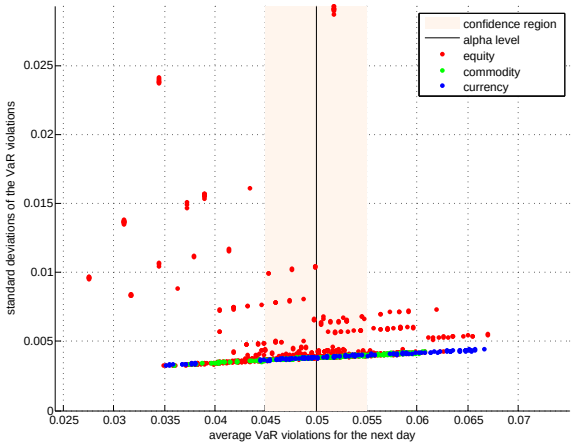
Table 4.3.: Results of the backtests. Mean violation and dispersion $s_{\text{mean violations}}$ are shown along with their associated t statistic and p-value. The t-test is a two-sample test that compares the filter of interest with the EWMA, the hypothesis being H_0 : $\mu_{EWMA} = \mu_{\text{filter}}$

using the historical assumption, the estimation of VaR is totally wrong. It is overestimated by roughly 1% which is unacceptable, since it means that the allowance is twice as big as what is to be forecasted. The shape of the distribution does not seem influenced by the shape of the impulse response except in the case of $\alpha = 1\%$ with the normality assumption where all the windows present skewed distributions.

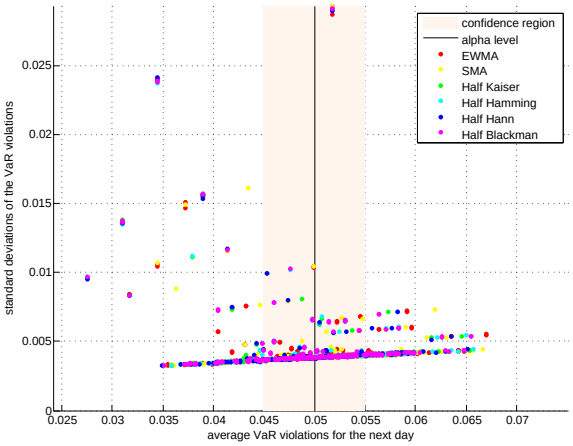
Table 4.3 summarizes the information present in the histograms and scatter plots. One sees that in the case $\alpha = 1\%$ with the normality assumption, the estimate is very close to α and all the filters have very similar results. Furthermore, when conducting t statistics, none of the p-values obtained is low enough to reject the hypothesis $H_0 : \mu_{EWMA} <$ or $> \mu_{newfilter}$ suggesting that all the filters have similar performances. The same remarks can be made for $\alpha = 5\%$ and normality assumption.

However, when using the historical assumption, results are totally different. First, in the $\alpha = 1\%$ case, $\overline{x}_{\text{violations}}$ is much higher than the targeted value. Moreover, p-values are much smaller and all the filters could reject the H_0 hypothesis at a 5% level. This suggests that EWMA has the best performance and that all the other filters have worse results.

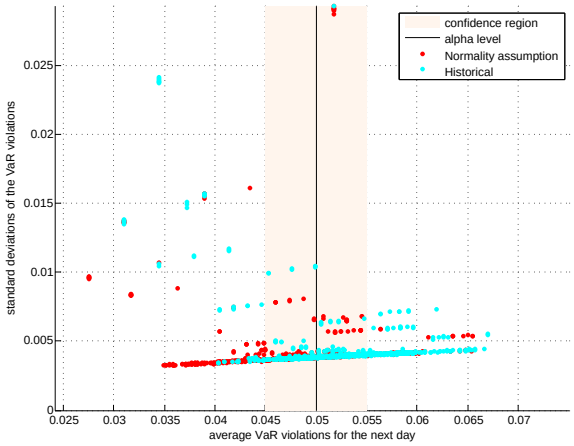
When $\alpha = 5\%$ using the historical assumption, results are also bad. EWMA has a good estimate but all the other filters are significantly higher thus suggesting that EWMA is still the best.



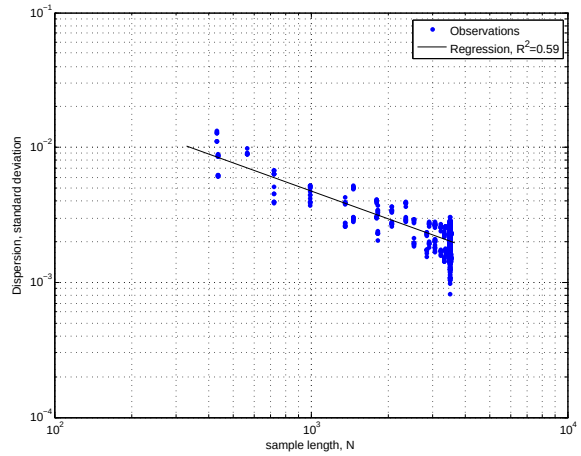
(a) Comparison within asset classes



(b) Comparison for different filters

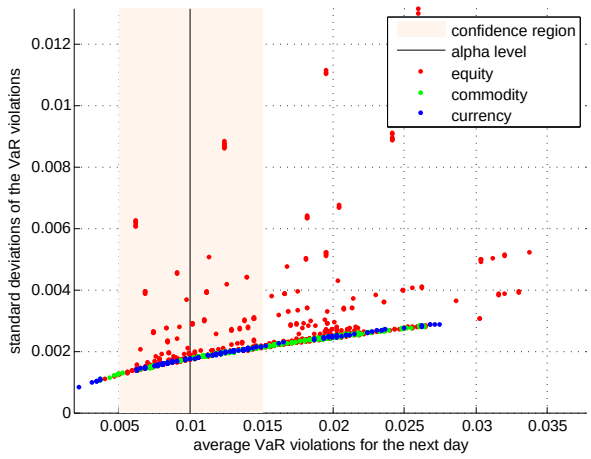


(c) Comparison between the normality and historical assumption

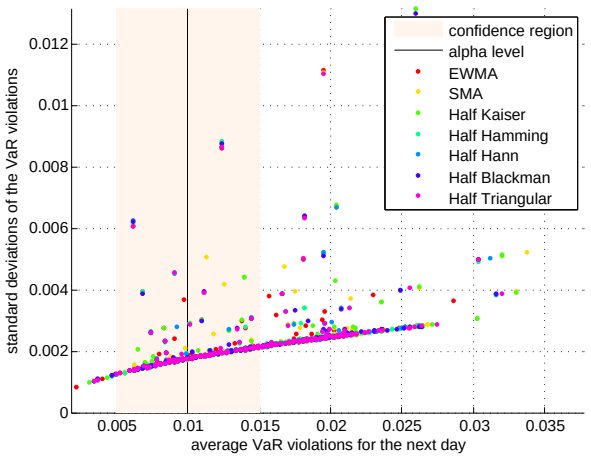


(d) Illustration of the dependency of the sample length on the standard deviation

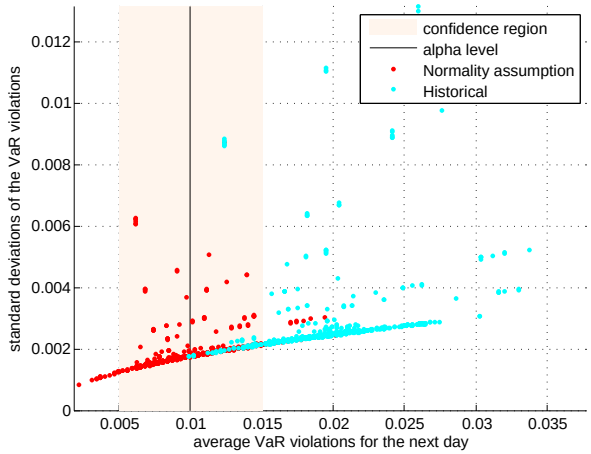
Figure 4.5.: Scatter plot representing the outcome of the experiment when computing $VaR_{1-\alpha}$, $\alpha = 5\%$ with the methods described in the previous sections. The x-axis is the relative number of violations. Ideally this should not exceed the α-level (here 5%). The y-axis is the standard deviation of the violations when doing random sampling of the violation vector (bootstrapping); ideally it should be as low as possible. The confidence region is an arbitrarily set $\pm 0.5\%$ interval around thee α-level to facilitate visualization of seemingly good performances from the rest.



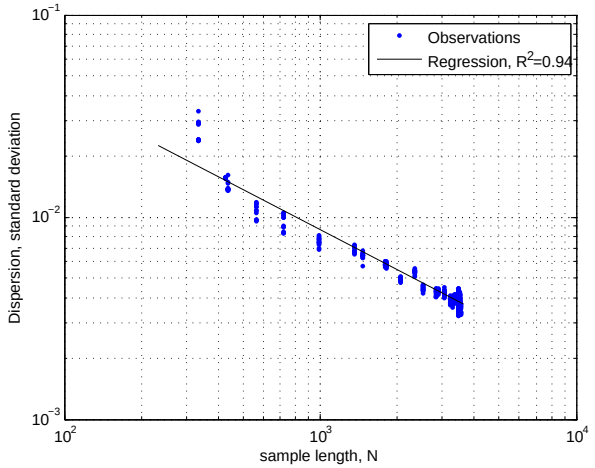
(a) Comparison within asset classes



(b) Comparison within different filters for prediction of the variance σ^2

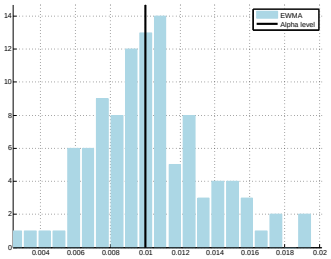


(c) Difference of computation of VaR when using the normality assumption and historical data (200 days)

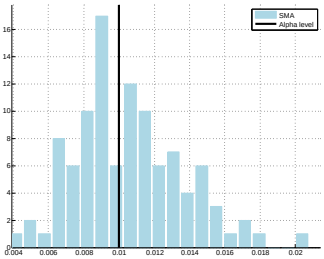


(d) Illustration of the dependency of the sample length on the standard deviation

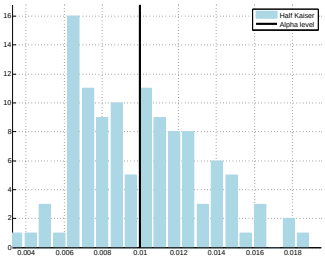
Figure 4.6.: Scatter plot representing the outcome of the experiment when computing $VaR_{1-\alpha}$, $\alpha = 1\%$ with the methods described in the previous sections. The x-axis is the relative number of violations. Ideally this should not exceed the α-level (here 1%). The y-axis is the standard deviation of the violations when doing random sampling of the violation vector (bootstrapping); ideally it should be as low as possible. The confidence region is an arbitrarily set $\pm 0.5\%$ interval around thee α-level to facilitate visualization of seemingly good performances from the rest.



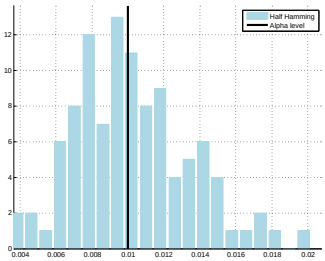
(a) EWMA (benchmark)



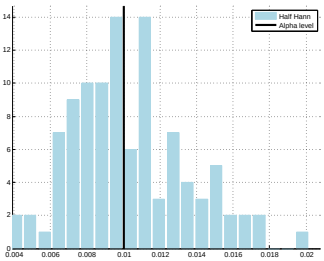
(b) SMA



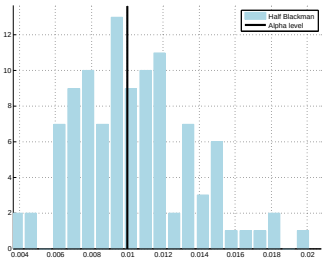
(c) half Kaiser window



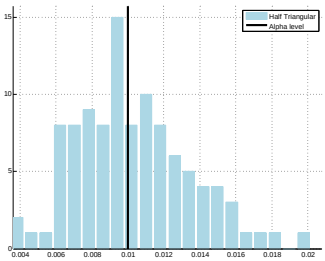
(d) half Hamming window



(e) half Hann window

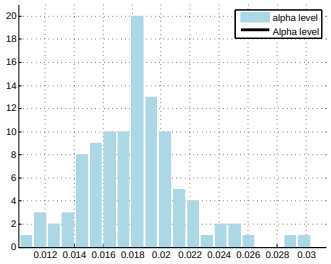


(f) half Blackman window

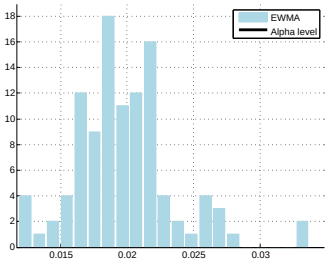


(g) half Triangular window

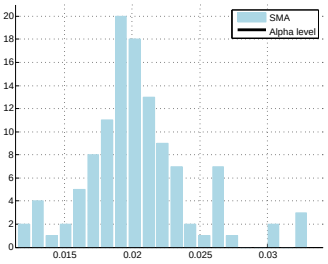
Figure 4.7.: Histogram of mean violation for complete portfolio. $\alpha = 1\%$ and computed using the normality assumption.



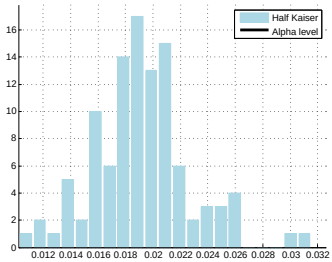
(a) EWMA (benchmark)



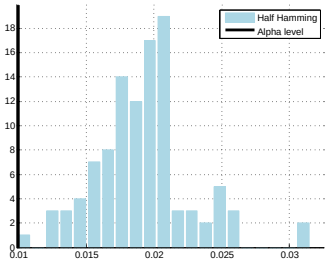
(b) SMA



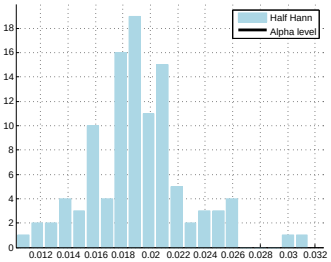
(c) half Kaiser window



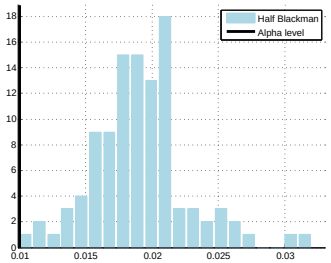
(d) half Hamming window



(e) half Hann window

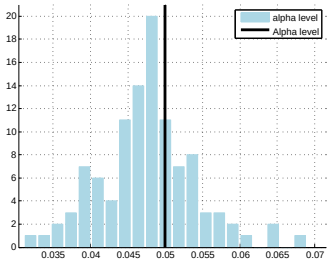


(f) half Blackman window

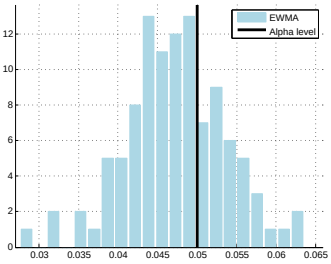


(g) half Triangular window

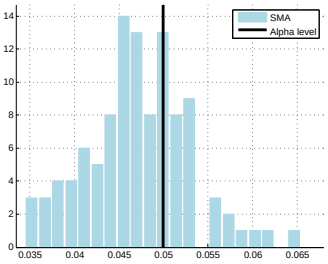
Figure 4.8.: Histogram of mean violation for complete portfolio. $\alpha = 1\%$ and using the historical assumption



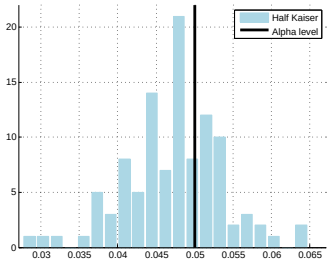
(a) EWMA (benchmark)



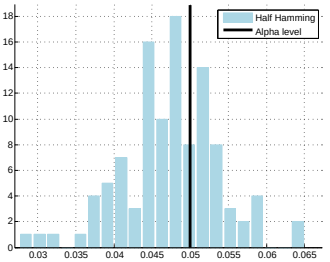
(b) SMA



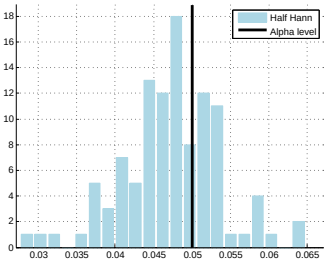
(c) half Kaiser window



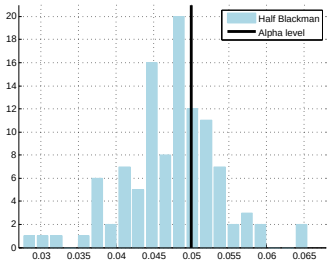
(d) half Hamming window



(e) half Hann window

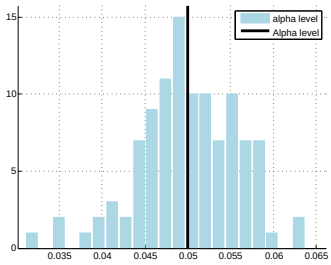


(f) half Blackman window

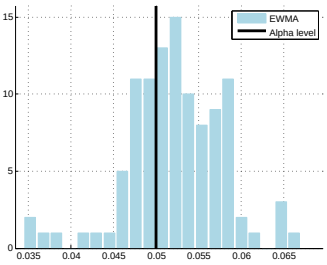


(g) half Triangular window

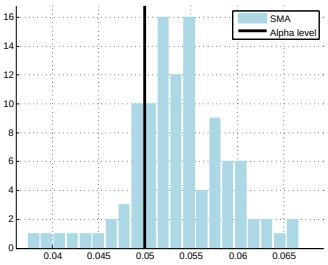
Figure 4.9.: Histogram of mean violation for complete portfolio. $\alpha = 5\%$ and using the normality assumption.



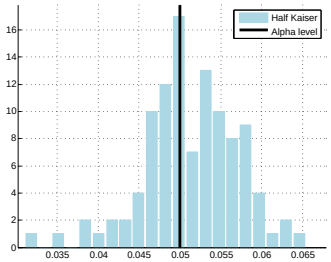
(a) EWMA (benchmark)



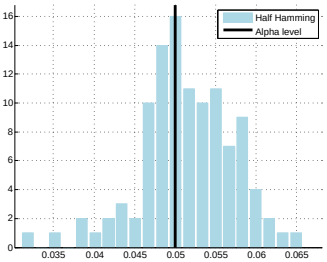
(b) SMA



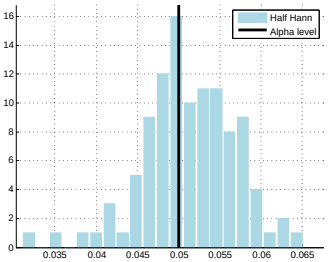
(c) half Kaiser window



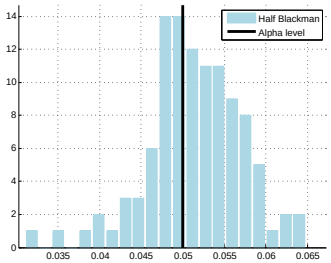
(d) half Hamming window



(e) half Hann window



(f) half Blackman window



(g) half Triangular window

Figure 4.10.: Histogram of mean violation for complete portfolio. $\alpha = 5\%$ and using the historical assumption.

5. Conclusion

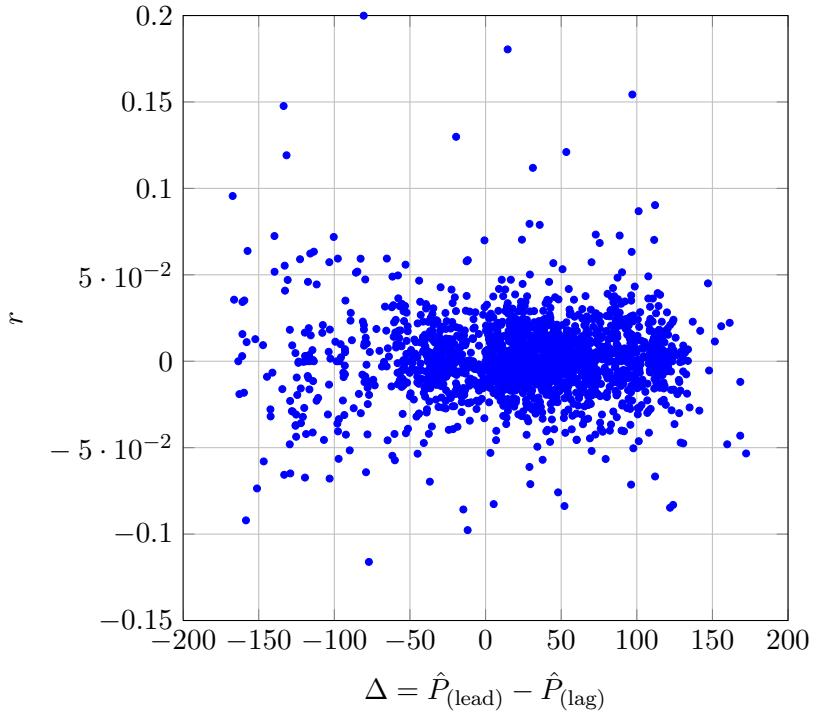


Figure 5.1.: Illustration of the (non-)correlation of the moving average indicator and actual returns for the first two quarters of 2012 of Google

The EWMA is well established in many fields of quantitative finance. In this thesis, we tested it against other digital filters in an automated trading environment and as a variance predictor within the RiskMetrics™ framework.

In the risk part, generally the EWMA remains the filter with the best results both with the normality assumption or using historical data. Nonetheless, results obtained with the other filters were only slightly worse suggesting that the moving average method used within the RiskMetrics™ framework is not very important and the crucial part is to estimate the quantile correctly as the difference in results between the two different methods suggest. Nevertheless, the simplicity of the EWMA makes it the perfect candidate as it can almost be computed by hand.

In the trading part, results for different filters were generally similar (Figure 3.5c) and not very convincing. When transaction costs are neglected, good results could be obtained with trading strategy #5. For this strategy, half windows significantly outperformed the EWMA and the SMA. Hence, where the EWMA is used as a trend indicator it could be substituted by one of the half windows for better performances. We haven't found significant differences in performance within the different half-window moving averages and any of those could be used and still have similar results. When transaction costs were added, the double and triple

EWMA were the best performing moving averages. However, performances were really low. The resonator and its two parameters exhibited very bad performances under every condition which is mostly due to its non-moving-average characteristics and an inclination towards overfitting.

Forecasting stock market returns with digital filters turned out to be challenging and as an illustration consider Figure 5.1 that shows Google stock returns for the first two quarters of 2012. The x-axis represents our indicator, $\Delta = \hat{P}_{(\text{lead})} - \hat{P}_{(\text{lag})}$ and the y-axis the associated returns of the next day. The trend-following rule assumes that whenever $\Delta > 0$, returns should be positive and when $\Delta < 0$, returns should be negative. It is quite clear from the graph, that there is no such relation even though the indicator was optimized in-sample. This, along with the absence of autocorrelation in returns (Section 4.2.6), and the generally recognized very low signal to noise ratio stylized fact, makes it very difficult to design simple linear trading strategies.

Finally, the fact that the EWMA is well established, easy to compute and somewhat elegant makes its substitution with another less known filter challenging.

A. Appendix

A.1. Digital Filter Theory: complement

A.1.1. Short reference for complex numbers

$$
\begin{aligned}
z &= \Re(z) + \Im(z) = a + jb = \sqrt{a^2 + b^2} e^{j \arctan b/a} && \text{(A.1)}\\
z &= |z| e^{\angle z} = r e^{j\omega} = \cos\omega + j \sin\omega && \text{(A.2)}\\
|z| &= \sqrt{a^2 + b^2} = r && \text{(A.3)}\\
|z_1 \cdot z_2| &= |z_1||z_1| && \text{(A.4)}\\
|z_1 + z_2| &= \sqrt{(a_1 + a_2)^2 + (b_1 + b_2)^2} && \text{(A.5)}\\
&= \sqrt{r_1^2 + r_2^2 + 2 r_1 r_2 \cos(\omega_2 - \omega_1)} && \text{(A.6)}\\
&\neq |z_1| + |z_2| && \text{(A.7)}\\
\angle z &= \arctan\frac{b}{a} + k\pi = \omega && \text{(A.8)}\\
\angle z_1 z_2 &= \angle z_1 + \angle z_2 = \omega_1 + \omega_2 && \text{(A.9)}\\
\angle(z_1 + z_2) &= \arctan\left(\frac{r_1 \sin(\omega_1) + r_2 \sin(\omega_2)}{r_1 \cos(\omega_1) + r_2 \cos(\omega_2)}\right) && \text{(A.10)}
\end{aligned}
$$

A.1.2. The inverse Z-transform

The inverse Z-transform does exactly the opposite of the Z-transform and allows to retrieve $x[n]$ from $X(Z)$. It is defined as

$$x[n] = \mathcal{Z}^{-1}\{X(z)\} = \frac{1}{2j\pi} \oint_\Gamma X(z) z^{n-1} dz. \tag{A.11}$$

where Γ is a counterclockwise path encircling the origin and entirely in the region of convergence (ROC).

Practitioners rarely use the analytic inverse Z-transform formula but rather use a table of Z-transforms along with Z-transform properties (See Table A.1).

A.1.3. Direct Computation of the Discrete Time Fourier Transform

The DTFT can also be computed directly from a discrete-time signal:

$$X(e^{j\omega}) = \mathcal{F}\{x[n]\} = \sum_{n \in \mathbb{Z}} x[n] e^{-j\omega n}. \tag{A.12}$$

and the inverse DTFT is

$$x[n] = \mathcal{F}^{-1}\{X(e^{j\omega})\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega. \tag{A.13}$$

	h[n]	**H(z)**	**ROC**
1.	$\delta[n]$	1	all z
2.	$\delta[n-n_0]$	z^{-n_0}	$z \neq 0$
3.	$u[n]$	$\frac{1}{1-z^{-1}}$	$\|z\| \leq 1$
4.	$e^{-\alpha n}u[n]$	$\frac{1.}{1-e^{-\alpha}z^{-1}}$	$\|z\| \geq \|e^{-\alpha}\|$
5.	$-u[-n-1]$	$\frac{1}{1.-z^{-1}}$	$\|z\| = 1$
6.	$nu[n]$	$\frac{z^{-1}}{(1-z^{-1})^2}$	$\|z\| \geq 1$
7.	$-nu[-n-1]$	$\frac{z^{-1}}{(1.-z^{-1})^2.}$	$\|z\| = 1.$
8.	$n^2.u[n]$	$\frac{z^{-1}(1.+z^{-1})}{(1.-z^{-1})^3}$	$\|z\| \geq 1$
9.	$-n^2.u[-n-1]$	$\frac{z^{-1}(1.+z^{-1})}{(1.-z^{-1})^3}$	$\|z\| = 1$
10.	$n^3.u[n]$	$\frac{z^{-1}(1.+4.z^{-1}+z^{-2})}{(1-z^{-1})^4}$	$\|z\| \geq 1$
11.	$-n^3.u[-n-1]$	$\frac{z^{-1}(1.+4.z^{-1}+z^{-2})}{(1-z^{-1})^4}$	$\|z\| = 1$
12.	$a^n u[n]$	$\frac{1}{1-az^{-1}}$	$\|z\| \geq \|a\|$
13.	$-a^n u[-n-1]$	$\frac{1}{1-az^{-1}}$	$\|z\| = \|a\|$
14.	$na^n u[n]$	$\frac{az^{-1}}{(1-az^{-1})^2.}$	$\|z\| \geq \|a\|$
15.	$-na^n u[-n-1]$	$\frac{az^{-1}}{(1-az^{-1})^2.}$	$\|z\| = \|a\|$
16.	$n^2.a^n u[n]$	$\frac{az^{-1}(1.+az^{-1})}{(1-az^{-1})^3}$	$\|z\| \geq \|a\|$
17.	$-n^2.a^n u[-n-1]$	$\frac{az^{-1}(1.+az^{-1})}{(1-az^{-1})^3}$	$\|z\| = \|a\|$
18.	$\cos(\omega_0.n)u[n]$	$\frac{1-z^{-1}\cos(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}$	$\|z\| \geq 1$
19.	$\sin(\omega_0.n)u[n]$	$\frac{z^{-1}\sin(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}$	$\|z\| \geq 1$
20.	$a^n\cos(\omega_0.n)u[n]$	$\frac{1-az^{-1}\cos(\omega_0)}{1-2az^{-1}\cos(\omega_0)+a^2.z^{-2}}$	$\|z\| \geq \|a\|$
21.	$a^n\sin(\omega_0.n)u[n]$	$\frac{az^{-1}\sin(\omega_0)}{1-2az^{-1}\cos(\omega_0)+a^2.z^{-2}}$	$\|z\| \geq \|a\|$

Table A.1.: Table of common z-transform pairs. $\delta[n]$ is the delta Kronecker and $u[n]$ is the Hamming function. *ROC* denotes the region of convergence. [Wikipedia]

A.1.4. The Discrete Fourier Transform (DFT)

Since both the DTFT and the Z-transform are continuous functions, they are not numerically computable. Therefore, we need a function that gives a frequency representation and that is also numerically computable so that it can be modeled. The DFT overcomes this problem, as it is a frequency sampled version of the DTFT. The DFT of $x[n]$, denoted as $X[m]$, and its inverse can be computed as follows

$$X[m] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi mn/N} \tag{A.14}$$

$$x[n] = \frac{1}{N}\sum_{m=0}^{N-1} X[m]e^{-j2\pi mn/N}. \tag{A.15}$$

The DFT and inverse DFT are always defined for finite-length signal. From the previous Equations, it easy to see that the exact sampling relationship between the DFT and DTFT is

$$F[m] = F(e^{j2\pi m/N}). \tag{A.16}$$

Practically, the DFT gives the same information as the DTFT, namely amplitude and phase response in the frequency domain.

Working digitally sometimes makes it hard to keep track of units, as the only *unit* n is the sampling interval. In the case of $X[m]$ we have the following relationship to a physical quantity:

$$fT_s = \frac{m}{N} \tag{A.17}$$

where N is the sample length, T_s the sampling period (so that $x[n] = x(nT_s)$, where $x(\cdot)$ is the continuous signal) and f is the associated frequency with m. Moreover, the quantity ω is called angular velocity and is defined as $\omega = 2\pi f$. In this paper, except where indicated otherwise, we will assume n to be trading days ($T_s = 1$ [day]).

From a computational point of view, Equations A.14 and A.15 are computed using a more efficient algorithm called the Fast Fourier Transform (FFT) which significantly reduces the computation time. Nevertheless, the output of a DFT or FFT is exactly the same.

A.1.5. Parseval's Theorem

The Parseval's theorem states that the energy in the time-domain is the same as the one in the frequency domain. It is also known as Rayleigh's identity.

$$\sum_{n=0}^{N-1} |h[n]|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |H[k]|^2 \tag{A.18}$$

where $H[k]$ is the DFT of $h[n]$.

A.1.6. Main Types of Filters

Digital filters are generally ordered in three categories: low- high- and bandpass filters. A lowpass filter filters out high frequencies, a high pass filters filters out low frequencies and bandpass filters only allow a certain frequency to go through. See Figure Figure A.1 for a graphical example.

Typical examples of those filter in our day-to-day life are very common. For instance, the equalizer of any radio uses a mix of the filters to produce a balanced sound output. Low-pass

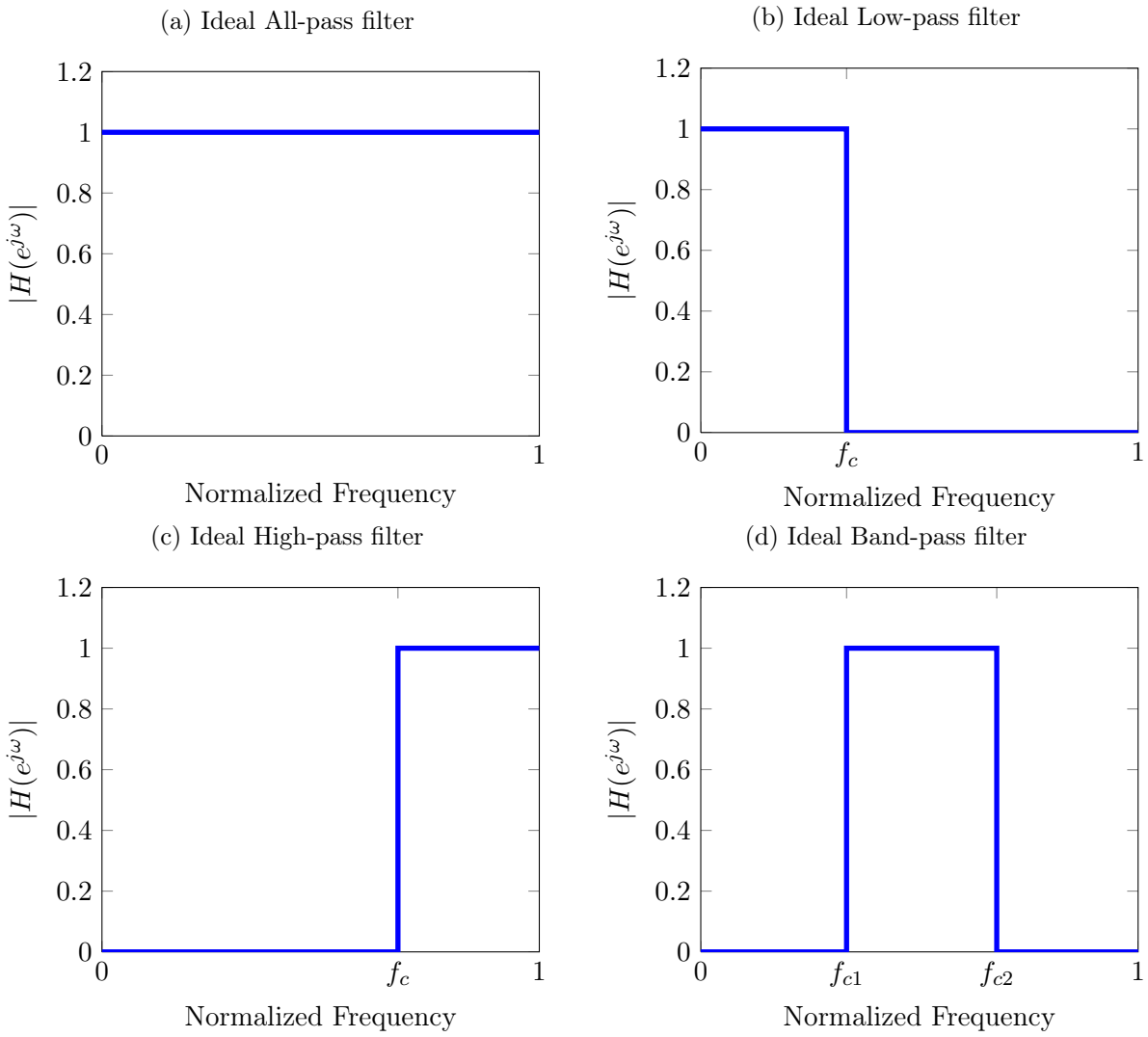
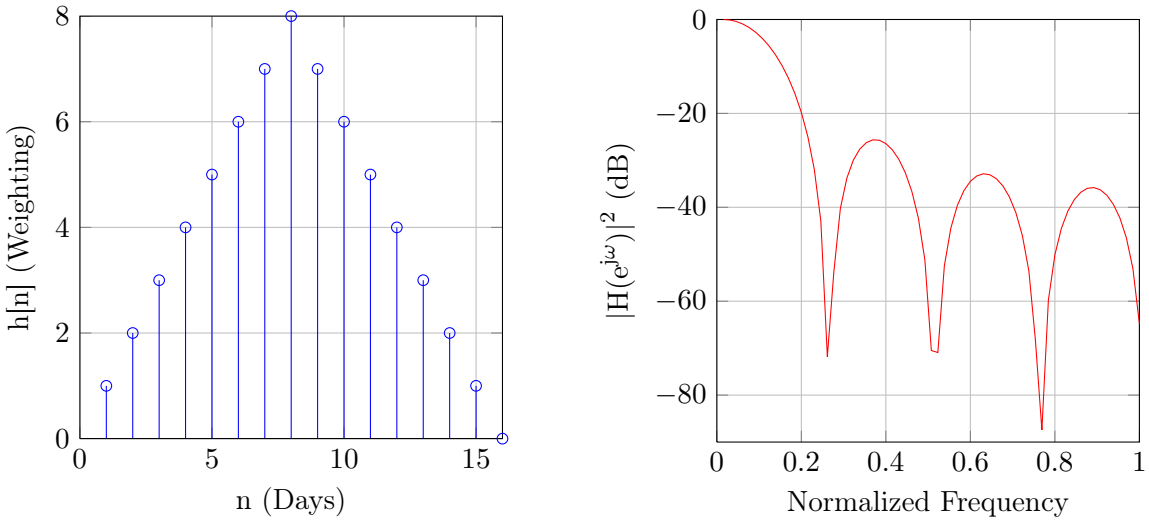


Figure A.1.: Illustration of the main different types of filters in the frequency domain.



(a) Impulse Response

(b) Frequency Response

Figure A.2.: Properties of the triangular moving average (TMA) filter ($N = 16$)

filters are used ECG program to be able to monitor the frequency of the heart without being disturbed by higher frequency noises.

A.2. Additional filters and derivations

A.2.1. Triangular Moving Average (TMA)

The triangular moving average is a double SMA; two SMA filters are put in parallel which gives an impulse response that looks like a triangle. Hence the filter equation is

$$y[n] = \underbrace{h_{SMA}[n] * h_{SMA}[n]}_{h_{TMA}[n]} * x[n] \tag{A.19}$$

where $h_{SMA}[n]$ is the impulse response of the SMA filter and $h_{TMA}[n]$ the impulse response of the triangular moving average.

The derivation of the TMA impulse response from the SMA is given here below:

$$\begin{aligned}
h[n] &= \frac{1}{N}(u[n] - u[n-N]) * \frac{1}{N}(u[n] - u[n-N]) \\
&= \frac{1}{N}\left(u[n] * u[n] - 2(u[n] * u[n]) * \delta[n-N] + (u[n] * u[n]) * \underbrace{\delta[n-N] * \delta[n-N]}_{\delta[n-2N]}\right) \\
&= u[n] * u[n] * (\delta[n] - 2\delta[n-M] + \delta[n-2M]) \\
&= n \cdot u[n] * (\delta[n] - 2\delta[n-N] + \delta[n-2N]) \\
&= n \cdot u[n] - 2n \cdot u[n-N] + n \cdot u[n-2N]
\end{aligned}$$

note that $u[n] * u[n] = n \cdot u[n]$[1].

[1]

$$u[n] * u[n] = \sum_{k\in\mathbb{Z}} u[k]u[n-k] = \sum_{k=0}^{n} u[n-k] = n \cdot u[n]. \tag{A.20}$$

In the z-domain the impulse response is

$$H(z) = H_{SMA}(z)^2 = \frac{1}{N^2}\left(\sum_{k=0}^{N-1} z^- k\right)^2. \tag{A.21}$$

Practically, this means that a TMA with parameter N has N double zeros .

A.2.2. Generalization of the order of the SMA - the n-SMA

We showed in the previous paragraph that the frequency response of the TMA is nothing more than the square of the frequency response of the SMA. Thus, by adding one parameter to the SMA, one can construct the n-SMA, n being the power at which the frequency response of the SMA is raised.

Hence,

$$H_{nSMA}(z) = \frac{1}{N^n}\left(\sum_{k=0}^{N-1} z^- k\right)^n \tag{A.22}$$

and

$$h_{NSMA}[n] = \underbrace{(h_{SMA} * h_{SMA} * ... * h_{SMA})}_{n \text{ times}}[n]. \tag{A.23}$$

This leads to a SMA filter that takes two parameters: N and n. N being the number of poles and N how many times they are repeated.

The advantage of this is that the smoothness of a filter can be improved by putting in series the same filter (basically doubling, tripling, ... the poles and zeros) but to the cost of phase lag [Tillson].

A.2.3. Linear Moving Average

The linear moving average gives weights that decrease linearly. After N points the impulse response is zero. Thus one has to solve the following equation system to find the impulse response

$$h[n] = \begin{cases} b - an, & n \leq N \\ 0, \text{else} \end{cases} \tag{A.24}$$

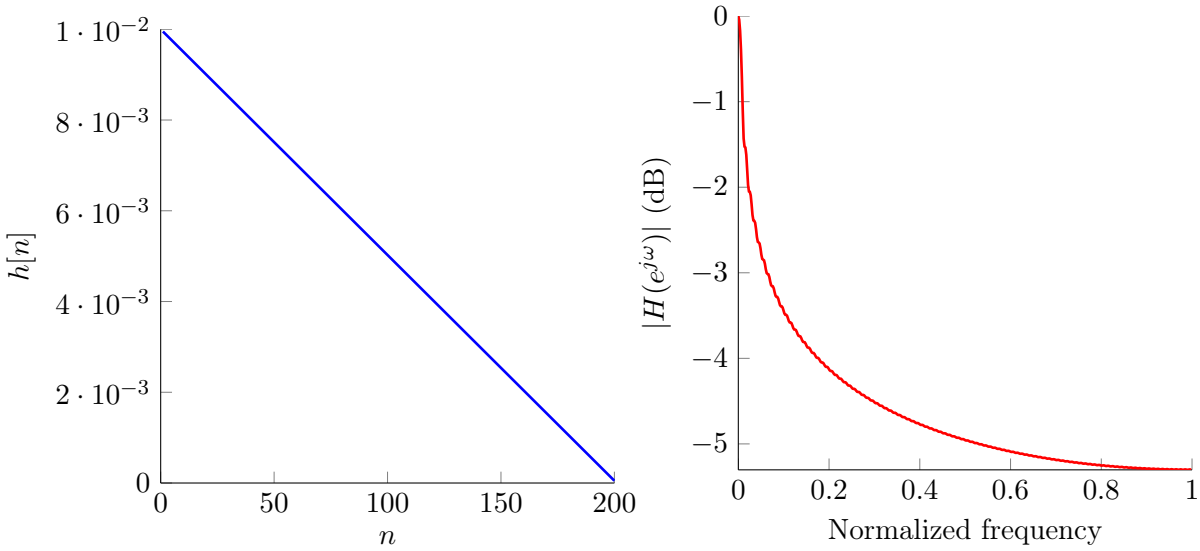
$$\sum_{k=0}^{\infty} h[k] = N\left(b - \frac{N+1}{2}a\right) = 0 \tag{A.25}$$

$$h[N] = b - a(N+1) = 0 \tag{A.26}$$

One finds $b = 2/N$ and $a = 2/(N(N))$ and thus

$$h[n] = \begin{cases} \frac{N-n}{\sum_{k=0}^{N} k} = \frac{2}{N}\left(1 - \frac{n}{N+1}\right), & n \leq N \\ 0, \text{else.} \end{cases} \tag{A.27}$$

This moving average can also be seen as a half triangular filter. This idea of taking only one half of the impulse response will be extended in Section 2.7.3.



(a) Impulse Response

(b) Frequency Response

Figure A.3.: Properties of the linear moving average filter ($N = 200$)

A.2.4. Generalization of the EWMA to the n-EWMA

From the DEWMA and TEWMA, it is easy to generalize it to the n-order exponentially weighted moving average. In this case, the system looks as following

$$\begin{aligned} y_1[m] &= (1-\lambda)x[m] + \lambda y_1[m] \\ y_2[m] &= (1-\lambda)y_1[m] + \lambda y_2[m] \\ &\dots \\ y_n[m] &= (1-\lambda)y_{n-1}[m] + \lambda y_n[m] \end{aligned} \tag{A.28}$$

or simpler

$$y_n[m] = x_m * (y_{n-1}[m] * \dots * y_1[m]) \tag{A.29}$$

$$\Leftrightarrow$$

$$Y_n(z) = X(z) \prod_{k=1}^{n-1} Y_k(z) \tag{A.30}$$

From Equation A.28, it is easy to see that every equation has the same transfer function, i.e. $y_{k+1}[m] = y_k[m] * h[m]$ where $h[m]$ is the EWMA transfer function. Using this result and eq A.30, one can write

$$H_{nEWMA} = H^n_{EWMA}(z) = \frac{Y(z)}{X(z)} = \frac{(1-\lambda)^n}{(1-\lambda z^{-1})^n}. \tag{A.31}$$

Note that the n-EWMA is for the EWMA what the n-SMA is for the SMA (see section A.2.2). Even so – as for the n-SMA – there is a tradeoff between the number of poles and thus the smoothness of the filter and the total lag of the filter.

A.3. Backtesting Framework Structure

In order to effectively backtest the different filters under different conditions, a framework has to be set up that allows to launch experiments and do statistical tests on the performance of the different filters in an automated way and on an entire cluster of data, e.g. all soft commodities. The framework is designed on the model of the project structure (see fig A.4). An overview of the workflow of the framework is depicted on fig A.5. The framework design was inspired by [Chan, 2009b]. The main idea is to run an optimization of parameters over the time-series for two filters. One benchmark (the EWMA) against a newly designed one (windows and two-poles IIR filters). The output will always show the performance of the buy and hold strategy in comparison with a chosen trading strategy applied to both the benchmark and the new filters. For fairness of comparison, both filters are tested under the same conditions.

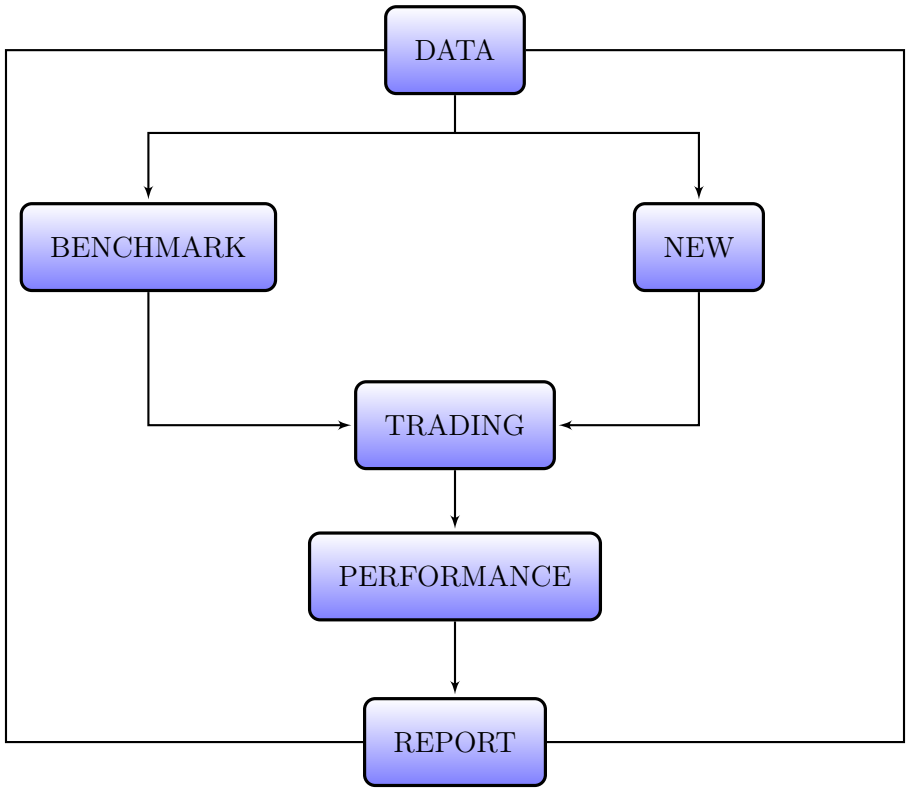


Figure A.4.: Backtesting Framework Structure

A.3.1. Settings

This is the first step of the process where all the parameters have to be set and the scope of the experiment is to be determined. An object is initialized with parameters that designs the experiment to be carried out. A list of the most important attributes is given in table A.2. All along the same experiment the parameters are to remain constant.

Attribute	Meaning
`start_date` and `end_date`	Indicates the start and end time of the fetched time-series
`transaction_cost`	Price of the transaction cost (in bp)
`nTune`	length in days quarter of one window over which parameters are fitted
`transient`	Delay before starting to trade (filter transient)
`trading_strategy`	Defines a trading strategy
`filter`	Defines filter to use

Table A.2.: Listing of the most important setting attributes

id	ticker	name	currency	exchange	sector	asset class
6537	KC1	Coffee	USD	ICE	Softs	Commodity

Table A.3.: Example of a one line output of the database. More information about the database and its structure in section A.4

A.3.2. Tickers Selection

In this step, a list of tickers is selected according to a set of criteria. Those criteria can select only one ticker (a specific security, e.g. coffee), some tickers (e.g. all soft commodities), or a whole family of tickers (e.g. all commodities). A query with the criteria is then generated and sent to the database which returns a set of lines with the desired information including the asset class and sector it belongs to, the exchange it is traded at and the currency it is traded in. An example of one database output line is given in table A.3.

A.3.3. Fetch Data and Formatting

Once the tickers of interest are retrieved, the associated time-series is fetched from the database. Returns and NAV, see Section 3.2.1, are computed.

Formatting also includes the removing of weekends and bank holidays (where returns are zero) and NaNs. Moreover, for commodities rolling dates are removed so that the time-series is smooth and does not include regular gaps. If not removed, those gaps could be captured by the filter and generate abnormally high virtual profits.

A.4. Database Structure and Content

The complete database structure is given in figure A.6.

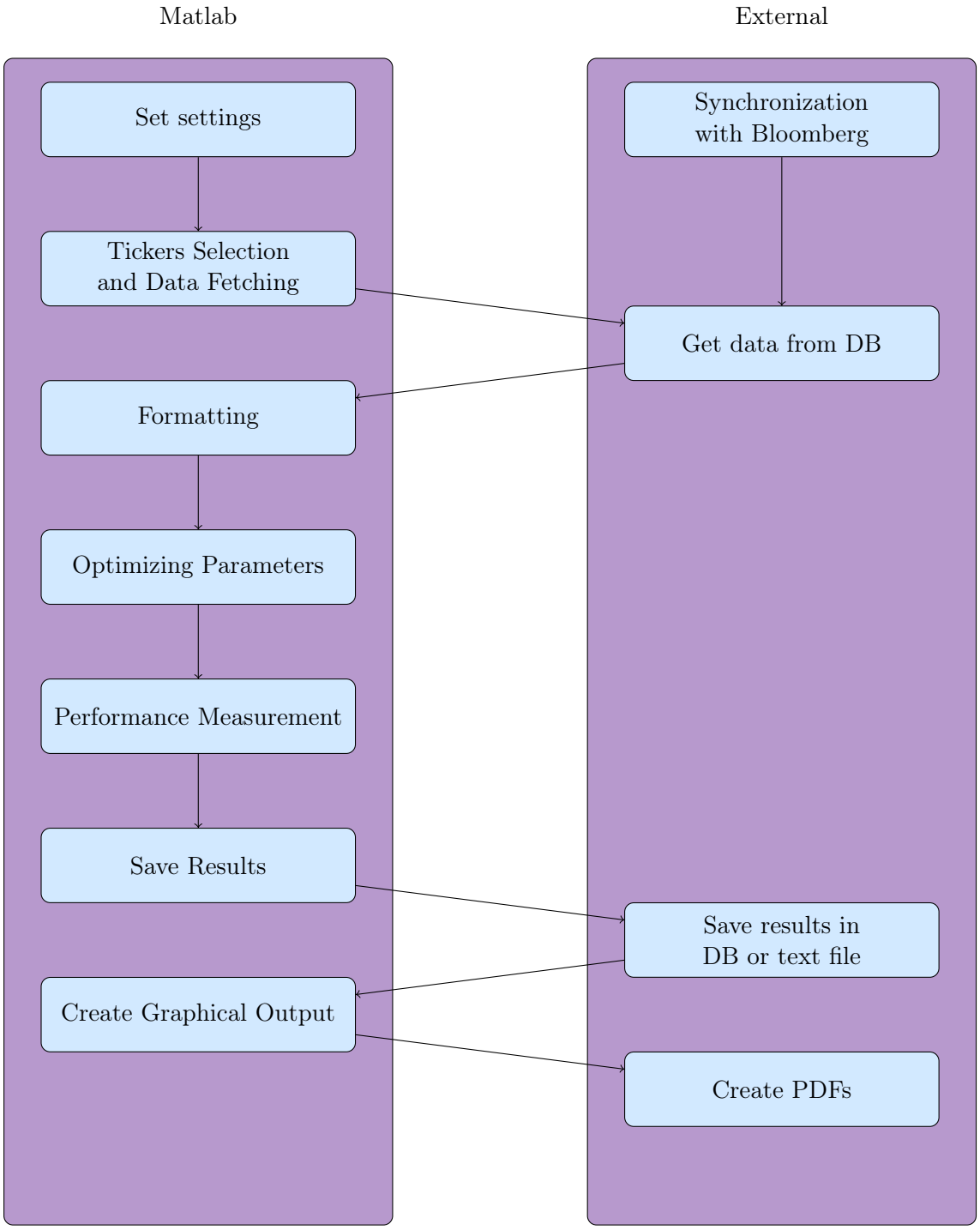


Figure A.5.: Overview of Framework Structure

A.4.1. Portfolio

List of all the securities contained in the portfolio used for the tests in this study

Table A.4.: Portfolio content

Ticker	Company name	Asset Class	Sector
ADTN	ADTRAN, Inc.	Equity	Public Utilities
COLM	Columbia Sportswear Company	Equity	Consumer Non-Durables
DEER	Deer Consumer Products, Inc.	Equity	Consumer Durables
ESYS	Elecsys Corporation	Equity	Capital Goods
GOOG	Google Inc.	Equity	Technology
GRPN	Groupon, Inc.	Equity	Technology
GPOR	Gulfport Energy Corporation	Equity	Energy
HAS	Hasbro, Inc.	Equity	Consumer Non-Durables
NVDQ	Novadaq Technologies Inc	Equity	Health Care
QCOM	QUALCOMM Incorporated	Equity	Technology
RYAAY	Ryanair Holdings plc	Equity	Transportation
GCVRZ	Sanofi	Equity	Consumer Durables
SIGM	Sigma Designs, Inc.	Equity	Technology
SUNH	Sun Healthcare Group, Inc.	Equity	Health Care
YHOO	Yahoo! Inc.	Equity	Technology
MMM	3M Company	Equity	Health Care
AME	AMTEK, Inc.	Equity	Consumer Durables
AT	Atlantic Power Corporation	Equity	Energy
BAC	Bank of America Corporation	Equity	Finance
BRY	Berry Petroleum Company	Equity	Energy
BA	Boeing Company (The)	Equity	Capital Goods
BP	BP p.l.c.	Equity	Energy
CPE	Callon Petroleum Company	Equity	Energy
CP	Canadian Pacific Railway Limited	Equity	Transportation
CAT	Caterpillar, Inc.	Equity	Capital Goods
CX	Cemex S.A.B. de C.V.	Equity	Capital Goods
CVX	Chevron Corporation	Equity	Energy
CEA	China Eastern Airlines Corporation Ltd.	Equity	Transportation
CHA	China Telecom Corp Ltd	Equity	Public Utilities
BLW	Citigroup Inc.	Equity	Finance
CS	Credit Suisse Group	Equity	Finance
DAL	Delta Air Lines Inc. (New)	Equity	Transportation
DOW	Dow Chemical Company (The)	Equity	Basic Industries
DD	E.I. du Pont de Nemours and Company	Equity	Basic Industries
EW	Edwards Lifesciences Corporation	Equity	Health Care
XOM	Exxon Mobil Corporation	Equity	Energy
FDX	FedEx Corporation	Equity	Transportation
F	Ford Motor Company	Equity	Capital Goods
FTE	France Telecom S.A.	Equity	Public Utilities
GNK	Genco Shipping & Trading Limited	Equity	Transportation
GM	General Motors Company	Equity	Capital Goods
GS	Goldman Sachs Group, Inc. (The)	Equity	Finance

Continued on next page

Table A.4 – continued from previous page

Ticker	Company name	Asset Class	Sector
HLS	HealthSouth Corporation	Equity	Health Care
HPQ	Hewlett-Packard Company	Equity	Technology
HMC	Honda Motor Company, Ltd.	Equity	Capital Goods
HBC	HSBC Holdings plc	Equity	Finance
JPM	J P Morgan Chase & Co	Equity	Finance
JW/A	John Wiley & Sons, Inc.	Equity	Consumer Services
JNJ	Johnson & Johnson	Equity	Consumer Durables
K	Kellogg Company	Equity	Consumer Non-Durables
KNX	Knight Transportation, Inc.	Equity	Transportation
KFT	Kraft Foods Inc.	Equity	Consumer Non-Durables
LNKD	LinkedIn Corporation	Equity	Technology
MAR	Marriot International	Equity	Consumer Services
MDT	Medtronic, Inc.	Equity	Health Care
MON	Monsanto Company	Equity	Basic Industries
MS	Morgan Stanley	Equity	Finance
NFG	National Fuel Gas Company	Equity	Public Utilities
NKE	Nike, Inc.	Equity	Consumer Non-Durables
NOK	Nokia Corporation	Equity	Technology
NSC	Norfolk Souther Corporation	Equity	Transportation
TLK	P.T. Telekomunikasi Indonesia, Tbk.	Equity	Public Utilities
PCG	Pacific Gas & Electric Co.	Equity	Public Utilities
RA	RailAmerica, Inc.	Equity	Transportation
ENL	Reed Elsevier NV	Equity	Consumer Services
RIO	Rio Tinto Plc	Equity	Basic Industries
SIX	Six Flags Entertainment Corporation New	Equity	Consumer Services
SNN	Smith & Nephew SNATS, Inc.	Equity	Health Care
SNE	Sony Corp Ord	Equity	Consumer Non-Durables
SWK	Stanley Black & Decker, Inc.	Equity	Capital Goods
SYT	Syngenta AG	Equity	Basic Industries
TEF	Telefonica SA	Equity	Public Utilities
TRI	Thomson Reuters Corp	Equity	Consumer Services
TM	Toyota Motor Corp Ltd Ord	Equity	Capital Goods
TUP	Tupperware Corporation	Equity	Consumer Non-Durables
TSN	Tyson Foods, Inc.	Equity	Consumer Non-Durables
VZ	Verizon Communications Inc.	Equity	Public Utilities
WAG	Walgreen Co.	Equity	Health Care
WPO	Washington Post Company (The)	Equity	Consumer Services
W 1 COMB	Wheat	Commodity	Grains
C 1 COMB	Corn	Commodity	Grains
O 1 COMB	Oats	Commodity	Grains
S 1 COMB	Soybean	Commodity	Grains
CC1	Cocoa	Commodity	Softs
KC1	Coffee	Commodity	Softs
SB1	Sugar	Commodity	Softs
CT1	Cotton	Commodity	Softs

Continued on next page

Table A.4 – continued from previous page

Ticker	Company name	Asset Class	Sector
QS1	Gas Oil	Commodity	Energy
CO1	Brent Crude	Commodity	Energy
CL1 COMB	WTI Crude Oil	Commodity	Energy
NG1 COMB	Natural Gas	Commodity	Energy
LA1	Aluminium	Commodity	Industrial Metals
LL1	Lead	Commodity	Industrial Metals
LP1	Copper	Commodity	Industrial Metals
LN1	Nikel	Commodity	Industrial Metals
GC1 COMB	Gold	Commodity	Precious Metals
SI1 COMB	Silver	Commodity	Precious Metals
PL1 COMB	Platinum	Commodity	Precious Metals
PA1 COMB	Palladium	Commodity	Precious Metals
USDCHF	USD CHF	Currencies	Europe
USDEUR	USD EUR	Currencies	Europe
USDJPY	USD JPY	Currencies	Asia
USDCNY	USD CNY	Currencies	Asia
USDINR	USD INR	Currencies	Asia
USDBRL	USD BRL	Currencies	Americas
USDGBP	USD GBP	Currencies	Europe
USDCAD	USD CAD	Currencies	Americas
USDZAR	USD ZAR	Currencies	Africa
USDDZD	USD DZD	Currencies	Africa

A.4.2. Results

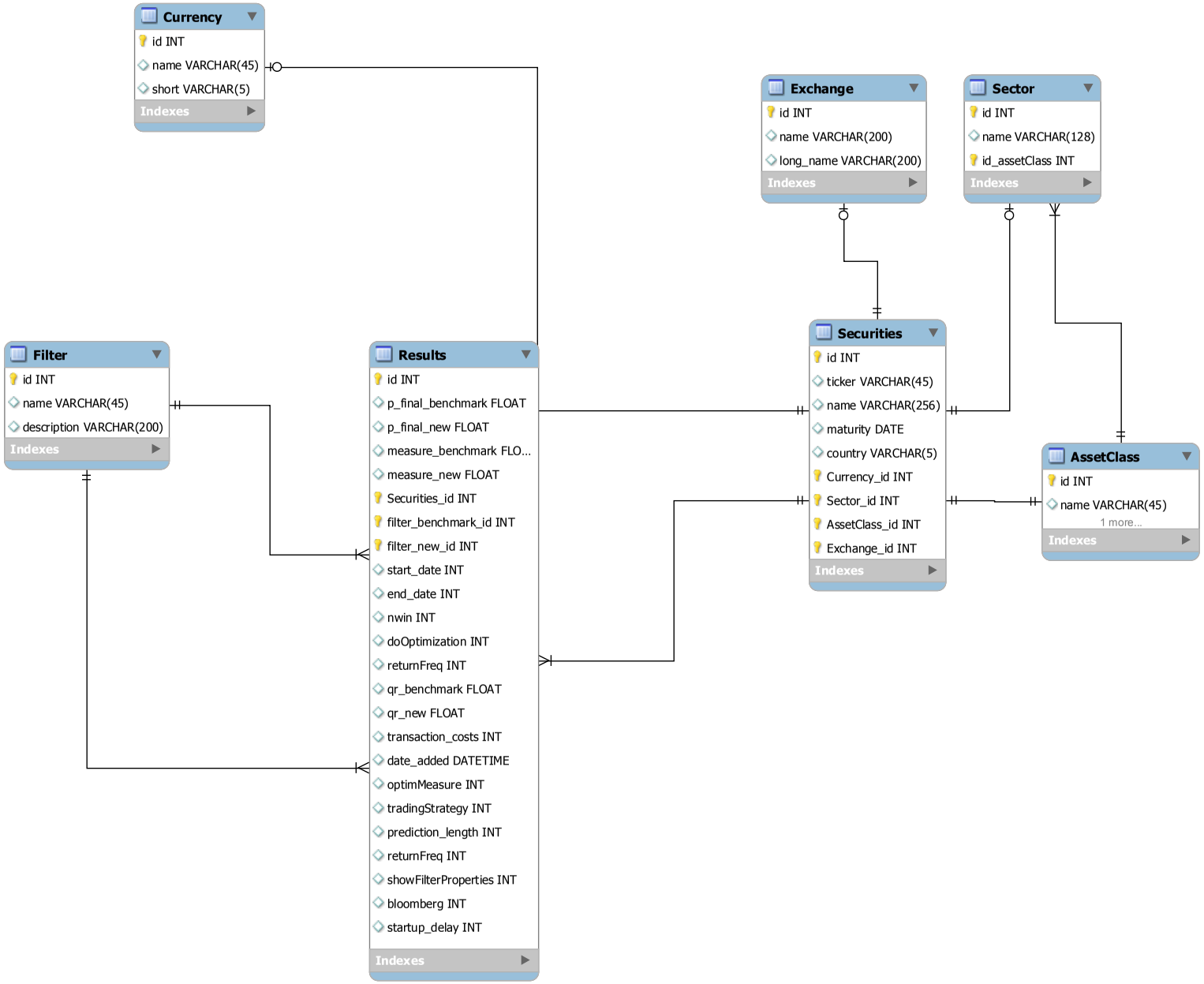


Figure A.6.: Structure of the database

List of the best results obtained with the simulation with transaction costs higher than zero.

Table A.5.: Best performing securities

Ticker	Asset class	Sector	Filter	Band %	Transaction cost, bps	Trading strategy	Q_d	Q_r	Sharpe ratio	Q_d quantile
SUNH	Equity	Health Care	halfHann	5	50	2	0.47	0.21	2.65	0.26
SUNH	Equity	Health Care	DoubleEWMA	0	50	5	0.55	-0.03	2.64	0.85
SUNH	Equity	Health Care	halfHann	0	50	2	0.50	0.20	2.55	0.46
SUNH	Equity	Health Care	halfHann	2	50	2	0.50	0.20	2.55	0.47
SUNH	Equity	Health Care	halfHann	5	90	2	0.47	0.21	2.45	0.27
NVDQ	Equity	Health Care	halfHann	5	50	5	0.41	0.03	2.45	0.00
NVDQ	Equity	Health Care	SMA	0	50	5	0.51	0.22	2.44	0.64
NVDQ	Equity	Health Care	halfHann	2	50	5	0.43	0.07	2.42	0.00
NVDQ	Equity	Health Care	EWMA	0	50	5	0.52	0.22	2.39	0.84
SUNH	Equity	Health Care	halfHann	0	90	2	0.50	0.20	2.36	0.46
SUNH	Equity	Health Care	halfHann	2	90	2	0.50	0.20	2.36	0.46
NVDQ	Equity	Health Care	EWMA	2	50	5	0.45	0.14	2.34	0.02
SUNH	Equity	Health Care	halfKaiser	0	50	5	0.52	-0.02	2.33	0.68
NVDQ	Equity	Health Care	halfBlackman	2	50	5	0.43	0.14	2.31	0.01
NVDQ	Equity	Health Care	halfHamming	0	50	5	0.51	0.21	2.30	0.70
NVDQ	Equity	Health Care	halfTriangular	2	50	5	0.45	0.16	2.26	0.02
ESYS	Equity	Capital Goods	halfKaiser	2	50	5	0.44	0.10	2.25	0.00
SUNH	Equity	Health Care	halfBlackman	5	50	5	0.48	-0.13	2.24	0.42
NVDQ	Equity	Health Care	halfKaiser	5	50	5	0.39	0.11	2.20	0.00
SIGM	Equity	Technology	halfHann	5	50	5	0.45	0.00	2.20	0.00
SUNH	Equity	Health Care	DoubleEWMA	0	90	5	0.54	-0.03	2.17	0.79
GM	Equity	Capital Goods	halfBlackman	5	50	5	0.45	0.06	2.16	0.44
SUNH	Equity	Health Care	halfTriangular	5	50	2	0.48	0.17	2.15	0.22
NVDQ	Equity	Health Care	SMA	2	50	5	0.44	0.15	2.15	0.02
SUNH	Equity	Health Care	halfTriangular	5	50	1	0.48	0.17	2.14	0.28
NVDQ	Equity	Health Care	halfTriangular	0	50	5	0.49	0.19	2.13	0.37
SUNH	Equity	Health Care	SMA	0	50	2	0.51	0.24	2.12	0.62
NVDQ	Equity	Health Care	halfBlackman	5	50	5	0.41	0.13	2.12	0.00
NVDQ	Equity	Health Care	halfHamming	2	50	5	0.44	0.14	2.10	0.02
NVDQ	Equity	Health Care	EWMA	0	90	5	0.51	0.21	2.10	0.68
NVDQ	Equity	Health Care	halfHann	5	90	5	0.41	0.03	2.09	0.00
SUNH	Equity	Health Care	halfKaiser	5	50	1	0.49	0.18	2.09	0.38
SUNH	Equity	Health Care	halfKaiser	5	50	2	0.49	0.18	2.09	0.37
NVDQ	Equity	Health Care	halfHann	0	50	5	0.47	0.08	2.09	0.06
SUNH	Equity	Health Care	halfHann	0	50	5	0.51	-0.05	2.08	0.57
ESYS	Equity	Capital Goods	halfKaiser	0	50	5	0.53	0.22	2.06	0.98

Continued on next page

Table A.5 – continued from previous page

Ticker	Asset class	Sector	Filter	Band %	Transaction cost, bps	Trading strategy	Q_d	Q_r	Sharpe ratio	Q_d quantile
SUNH	Equity	Health Care	halfTriangular	5	90	1	0.48	0.17	2.05	0.29
NVDQ	Equity	Health Care	halfKaiser	2	50	5	0.42	0.14	2.05	0.00
SUNH	Equity	Health Care	SMA	0	90	2	0.51	0.24	2.04	0.63
NVDQ	Equity	Health Care	EWMA	2	90	5	0.45	0.14	2.04	0.02
NVDQ	Equity	Health Care	halfHann	2	90	5	0.43	0.07	2.03	0.00
NVDQ	Equity	Health Care	halfHamming	5	50	5	0.41	0.09	2.03	0.00
SUNH	Equity	Health Care	halfTriangular	5	90	2	0.48	0.17	2.03	0.23
ESYS	Equity	Capital Goods	halfKaiser	5	50	5	0.41	0.06	2.03	0.00
GPOR	Equity	Energy	halfHamming	2	50	5	0.41	0.13	2.01	0.00
ESYS	Equity	Capital Goods	halfBlackman	2	50	5	0.45	0.11	2.01	0.01
GPOR	Equity	Energy	halfHann	0	50	5	0.45	0.15	2.00	0.00
ESYS	Equity	Capital Goods	halfBlackman	5	50	5	0.42	0.06	1.99	0.00
SUNH	Equity	Health Care	halfHann	0	50	1	0.50	0.22	1.99	0.51
GPOR	Equity	Energy	SMA	0	50	5	0.49	0.19	1.98	0.21
GPOR	Equity	Energy	halfBlackman	0	50	5	0.48	0.18	1.98	0.10
NVDQ	Equity	Health Care	DoubleEWMA	0	50	5	0.50	0.18	1.97	0.47
NVDQ	Equity	Health Care	halfTriangular	2	90	5	0.44	0.16	1.96	0.01
NVDQ	Equity	Health Care	halfBlackman	2	90	5	0.43	0.14	1.96	0.01
NVDQ	Equity	Health Care	SMA	5	50	5	0.42	0.12	1.96	0.00
NVDQ	Equity	Health Care	halfHamming	0	90	5	0.50	0.21	1.96	0.47
SUNH	Equity	Health Care	halfKaiser	5	90	2	0.49	0.18	1.95	0.38
SUNH	Equity	Health Care	halfKaiser	5	90	1	0.49	0.18	1.95	0.38
SUNH	Equity	Health Care	halfHamming	0	50	1	0.50	0.22	1.94	0.53
SUNH	Equity	Health Care	halfHamming	0	50	2	0.50	0.22	1.94	0.53
GPOR	Equity	Energy	halfKaiser	0	50	5	0.47	0.18	1.93	0.04
GPOR	Equity	Energy	EWMA	0	50	5	0.48	0.18	1.93	0.20
GPOR	Equity	Energy	halfBlackman	2	50	5	0.41	0.11	1.93	0.00
SUNH	Equity	Health Care	halfHamming	2	50	2	0.49	0.21	1.91	0.38
SUNH	Equity	Health Care	halfHann	0	90	1	0.50	0.22	1.91	0.50
SIGM	Equity	Technology	halfHann	2	50	5	0.46	0.01	1.91	0.01
NVDQ	Equity	Health Care	halfKaiser	5	90	5	0.39	0.11	1.91	0.00
NVDQ	Equity	Health Care	halfBlackman	0	50	5	0.50	0.17	1.91	0.49
NVDQ	Equity	Health Care	halfTriangular	5	50	5	0.40	0.08	1.91	0.00
ESYS	Equity	Capital Goods	halfTriangular	0	50	5	0.53	0.20	1.90	0.99
NVDQ	Equity	Health Care	halfKaiser	0	50	5	0.50	0.17	1.90	0.47
GPOR	Equity	Energy	halfTriangular	0	50	5	0.48	0.16	1.90	0.13
ESYS	Equity	Capital Goods	halfTriangular	5	50	5	0.42	0.05	1.90	0.00

Continued on next page

Table A.5 – continued from previous page

Ticker	Asset class	Sector	Filter	Band %	Transaction cost, bps	Trading strategy	Q_d	Q_r	Sharpe ratio	Q_d quantile
GPOR	Equity	Energy	halfHann	2	50	5	0.39	0.09	1.90	0.00
GPOR	Equity	Energy	halfKaiser	2	50	5	0.40	0.11	1.90	0.00
SUNH	Equity	Health Care	halfTriangular	2	50	1	0.49	0.21	1.89	0.35
SUNH	Equity	Health Care	halfTriangular	2	50	2	0.49	0.21	1.89	0.35
GPOR	Equity	Energy	SMA	2	50	5	0.42	0.12	1.88	0.00
ESYS	Equity	Capital Goods	EWMA	2	50	5	0.47	0.11	1.88	0.03
GPOR	Equity	Energy	halfHamming	0	50	5	0.48	0.16	1.88	0.10
SUNH	Equity	Health Care	halfHann	2	50	5	0.45	0.03	1.87	0.21
SIGM	Equity	Technology	halfKaiser	5	50	5	0.43	-0.01	1.87	0.00
NVDQ	Equity	Health Care	SMA	2	90	5	0.44	0.15	1.87	0.01
GPOR	Equity	Energy	halfTriangular	2	50	5	0.41	0.13	1.86	0.00
ESYS	Equity	Capital Goods	halfHamming	2	50	5	0.46	0.12	1.86	0.02
SUNH	Equity	Health Care	halfHann	5	50	1	0.46	0.17	1.86	0.13
SUNH	Equity	Health Care	halfHamming	0	90	2	0.50	0.22	1.86	0.52
SUNH	Equity	Health Care	halfHamming	0	90	1	0.50	0.22	1.86	0.52
SUNH	Equity	Health Care	halfHamming	2	50	1	0.48	0.21	1.85	0.29
NVDQ	Equity	Health Care	halfBlackman	5	90	5	0.40	0.13	1.85	0.00
SUNH	Equity	Health Care	halfHann	2	50	1	0.50	0.20	1.84	0.47
ESYS	Equity	Capital Goods	halfKaiser	2	90	5	0.44	0.10	1.83	0.00
ESYS	Equity	Capital Goods	halfHamming	5	50	5	0.42	0.05	1.83	0.00
SUNH	Equity	Health Care	halfHamming	2	90	2	0.49	0.21	1.82	0.39
SUNH	Equity	Health Care	halfTriangular	2	90	1	0.49	0.21	1.81	0.35
SUNH	Equity	Health Care	halfTriangular	2	90	2	0.49	0.21	1.81	0.34
NVDQ	Equity	Health Care	halfHamming	2	90	5	0.44	0.14	1.81	0.01
SUNH	Equity	Health Care	halfTriangular	0	50	1	0.50	0.21	1.79	0.39
SUNH	Equity	Health Care	SMA	0	50	1	0.50	0.21	1.79	0.38
SUNH	Equity	Health Care	halfTriangular	0	50	2	0.50	0.21	1.79	0.40

Acknowledgements

I am indebted to Dr Peter Cauwels and Prof Didier Sornette for their help, support and precious comments throughout the master's thesis.

I also would like to thank the swissQuant Group AG and particularly Dr Donnacha Daly for guiding the project, supporting me, and providing me with a favorable working environment. A special thank goes to all the people at swissQuant with whom I had enlightening conversations that directed some of the choices I made during this master's thesis.

Finally, I would like to thank my parents and my brother who have always believed in me and supported me throughout the course of my studies.

Bibliography

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

Mayur Agrawal, Debabrata Mohapatra, and Ilya Pollak. Empirical evidence against CAPM: Relating alphas and returns to betas. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 6, pages 5732–5735. IEEE, May 2011. ISBN 978-1-4577-0538-0. doi: 10.1109/ICASSP.2011.5947662. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5947662.

Beitollah Akbari Moghaddam and Saeed Ebrahimijam Hassan Haleh. Forecasting Trend and Stock Price with Adaptive Extended Kalman Filter Data Fusion. URL http://www.ipedr.com/vol4/24-F00046.pdf.

Vincent Boisbourdain. VaR multi-fractal, un remède en temps de crise ? *Lettre OTC conseil*, 2008. URL http://www.otc-conseil.fr/fre/High/publications/lettre-otc/lettre-n-37-septembre-2007/2175/multi-fractales.pdf.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986. ISSN 03044076. doi: 10.1016/0304-4076(86)90063-1. URL http://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdfhttp://linkinghub.elsevier.com/retrieve/pii/0304407686900631.

G. M. Jenkins Box, G. E. P. and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Upper Saddle River, NJ: Prentice-Hall, 3rd edition edition, 1994.

Paolo Brandimarte. *Numerical methods in finance and economics : a MATLAB-based introduction*. Wiley Interscience, Hoboken, N.J, 2006. ISBN 0471745030.

William Brock, Josef Lakonishok, and Blake LeBaron. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, 47(5):1731, December 1992. ISSN 00221082. doi: 10.2307/2328994. URL http://www.technicalanalysis.org.uk/moving-averages/BrLL92.pdfhttp://www.jstor.org/stable/2328994?origin=crossref.

Mark M. Carhart. On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57, March 1997. ISSN 00221082. doi: 10.2307/2329556. URL http://www.jstor.org/stable/2329556?origin=crossref.

Ernest Chan. *Quantitative trading : how to build your own algorithmic trading business*. John Wiley & Sons, Hoboken, N.J, 2009a. ISBN 0470284889.

Ernest P. Chan. *Quantitative Trading*. John Wiley & Sons, Inc, 2009b. ISBN 9780470284889.

Andrew Clare, James Seaton, Peter N Smith, and Stephen Thomas. The Trend is Our Friend : Risk Parity , Momentum and Trend Following in Global Asset Allocation. pages 1–32, 2012.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, February 2001. ISSN 1469-7688. doi: 10.1080/713665670. URL http://www.informaworld.com/openurl?genre=article&doi=10.1080/713665670&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3.

Gilles Daniel, Didier Sornette, and Peter Wohrmann. Look-Ahead Benchmark Bias in Portfolio Performance Evaluation. page 16, October 2008. URL http://arxiv.org/abs/0810.1922.

V. DeBrunner and Tristan Charpentier. Combining FIR filters and artificial neural networks to model stochastic processes. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, volume 1, pages 333–336. IEEE, 2002. ISBN 0-7803-7576-9. doi: 10.1109/ACSSC.2002.1197201. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1197201.

Robert F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1008, 1982.

Erich Walter Farkas. Lecture Quantitative Finance - Chapter 8 Estimating volatility and correlations, 2011.

Financial Times. The 'monster' of VaR has not gone away - FT.com. URL http://www.ft.com/cms/s/0/10145926-1df6-11e2-8e1d-00144feabdc0.html#axzz2AgGIaLPq.

Jan Fraenkle and T. Rachev Svetlozar. Review: algorithmic trading. *Investment Management and Financial Innovations*, 6(1), 2009. URL http://businessperspectives.org/journals_free/imfi/2009/imfi_en_2009_01_Fraenkle.pdf.

Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Efficient Monte Carlo Methods for Value-at-Risk. *IBM*, 2000.

Brian Hurst, Yao Hua Ooi, and Lasse Pedersen. A Century of Evidence on Trend-Following Investing. 2012.

Philippe Jorion. *Value At Risk*. 2nd edition, 2001.

J.P. Morgan and Reuters. RiskMetrics — Technical Document. 1996.

Jingyi Liu. Forecasting Volatility Using Competitive Time Series models –Evidence from Foreign Exchange Market. 1999.

Andrew W Lo. "The Statistics of Sharpe Ratios": Author's Response. *Financial Analysts Journal*, 58(6):18–18, November 2002. ISSN 0015-198X. doi: 10.2469/faj.v58.n6.2481. URL http://www.cfapubs.org/doi/abs/10.2469/faj.v58.n6.2481.

Alexander McNeil. *Quantitative risk management : concepts, techniques and tools*. Princeton University Press, Princeton, N.J, 2005. ISBN 0691122555.

Merton H. Miller, Jayaram Muthuswamy, and Robert E. Whaley. Mean Reversion of Standard & Poor's 500 Index Basis Changes: Arbitrage-Induced or Statistical Illusion? *The Journal of Finance*, 49(2):479, June 1994. ISSN 00221082. doi: 10.2307/2329160. URL http://www.jstor.org/stable/2329160?origin=crossref.

Binoy B Nair, V P Mohandas, N R Sakthivel, S Nagendran, A Nareash, R Nishanth, and S Ramkumar. Application of Hybrid Adaptive Filters for Stock Market Prediction. pages 443–447, 2010.

Alan V. Oppenheimer and Ronald W. Schafer. *Discrete-Time Signal Processing*. 1999.

Susumu Saito. Verification of Possibility of Forecasting Economical Time Series Using Neural Network and Digital Filter. (Icnc), 2007.

Leigh Stevens. *Essential technical analysis : tools and techniques to spot market trends*. Wiley, New York, NY, 2002. ISBN 047115279X.

Steve H. Thomas, Andrew D. Clare, James Seaton, and Peter N. Smith. Trend Following, Risk Parity and Momentum in Commodity Futures. *SSRN Electronic Journal*, pages 1–34, 2012. ISSN 1556-5068. doi: 10.2139/ssrn.2126813. URL `http://www.ssrn.com/abstract=2126813`.

Tom Tillson. Smoothing Techniques for more accurate Signals. *Stocks and Commodities*, 1(c): 1–5.

B. L. WELCH. THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika*, 34(1-2):28–35, 1947. ISSN 0006-3444. doi: 10.1093/biomet/34.1-2.28. URL `http://biomet.oxfordjournals.org/content/34/1-2/28.full.pdfhttp://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/34.1-2.28`.

J E Wesen, H M De Oliveira, and Bolsa De Valores. Adaptive Filter Design For Stock Market Prediction Using A Correlation-Based Criterion.

Wikipedia. Z-transform - Wikipedia, the free encyclopedia. URL `http://en.wikipedia.org/wiki/Z-transform#Table_of_common_Z-transform_pairs`.

J Wilder. *New concepts in technical trading systems*. Trend Research, Greensboro, N.C, 1978. ISBN 0894590278.

Jingtao Yao and Chew Lim Tan. A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1-4):79–98, September 2000. ISSN 09252312. doi: 10.1016/S0925-2312(00)00300-3. URL `http://linkinghub.elsevier.com/retrieve/pii/S0925231200003003`.