Indicators of Risk Appetite and Applications in Trading

by

Alexandre Max Maraval (CID: 01296924)

Supervisor : Dr Pietro Siorpaes

Department of Mathematics

Imperial College London

London SW7 2AZ

United Kingdom

Thesis submitted as part of the requirements for the award of the MSc in Mathematics and Finance, Imperial College London, 2016-2017

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Alexandre Maraval

Acknowledgements

This thesis has been internally supervised by Dr Pietro Siorpaes of Imperial College London. The research work was done during an internship at Jetstone Asset Management LLP under the external supervision of Igor Pikovsky and Gaurav Bamania.

I would like to express my gratitude to Dr Pietro Siorpaes for his time and support as well as Dr Antoine Jacquier for his advice. I would also like to express gratitude to Igor Pikovsky and Gaurav Bamania for their help and guidance during this project. I would like to thank Ahmad Butt for offering me the opportunity to research for Jetstone Asset Management. I would also like to thank colleagues at Jetsone Asset Management, especially Lora Parvanova, Will Newington-Bridges and Jason Tannen for their help accessing and gathering the data, James Seymour for his help with the software and version control, Andrew Smith for his feedback and Sacha Tungsong for her support.

Abstract

Risk Appetite is a poorly defined concept in financial mathematics. It is often used interchangeably with the concept of Risk Aversion contributing to the fog surrounding its clear meaning. In this thesis we begin by trying to explain what appetite for risk is, why it can be considered as distinct from aversion to risk and how it can be measured. We then follow a set of practitioners' papers and in particular the survey done by Aaron and Illing in [1] gathering a set of various techniques to compute risk appetite indicators. We implement some of these methods and try to improve them when we feel it is relevant, by adding some robustness or by using more recent techniques than the ones used at publication time. It is important to note that the majority of the papers we work with were written prior to the 2008 financial crisis. The general method of validation of these risk appetite'metrics' is to observe if they capture known past financial events like crashes or any other event having large-scale repercussion on its financial environment. In practice, looking at the risk appetite indicator we should see a reaction at these times, manifested by spikes in the indicator. Instead, we will differ from this method and focus on constructing an systematic (or automated) trading routine based on each index. This will give us a realistic way to compare them and assess if they indeed capture those market moves (by avoiding drawdowns mainly). We will still check, however, if they are able capture the financial crash of 2008 because otherwise the indicator is considered useless.

Contents

1 Introduction

Monitoring risk should always be part of every reasonable investment strategy. Furthermore, massive and sudden changes in investors' attitude towards risk is also something one should be aware of. Indeed, it is clear from the (too-many examples of) financial crashes, that investors behave differently during stress periods. It is then potentially worth looking for premises of investors' changes in attitude towards risk. This is essentially what *Risk Appetite* tries to capture. We will try to define it, explain how to measure it and see how it can be used in trading.

We will follow the path set by Aaron and Illing in [1] though other surveys on the topic have been made by the European Central Bank in [18] for example or by Detusche Bundesbank in [15]. In an environment of low appetite for risk for example, the pace of investments is slowed down by a higher cost of capital (high rates for borrowing). Conversely, in a high risk appetite environment, the demand for riskier assets is important and markets can volatile (see the *stylized facts* on clustered volatility in Cont [12]). In our work, we will see that Risk Appetite and Risk Aversion are two concepts that are closely linked together. Some authors do not differentiate them (like Kumar & Persaud in [34]) and some others do (see Gai & Vause [23]) but they always give their intuition on why they should differ or not and on the information they can provide and how it can be extracted. We will then discuss a number of different ways to extract and measure appetite for risk based on the different insights and techniques proposed by the literature. It is interesting to note the work of Misina in [42], [43], [44] as he tried to reconcile previous attempts to compute such a measure of risk appetite and gave a more precise framework for its study. He then comes up with a concept of *implied* risk aversion which is similar to risk appetite (terminology also used by Ait-Sahalia & Lo in [2] and Jackwerth in [28] for example). This idea of *implied* risk aversion is an interesting piece of the puzzle the definition of risk appetite but it is definitely not the only valid intuition. We will therefore also try to implement other techniques.

Apart from our attempt to untangle and synthesize the connections between different risk appellations, our main purpose will be to explain how to create automated trading strategies with information extracted from these risk indicators. Indeed, we will not try to replicate exactly the results obtained by the cited references nor try to show what has already been shown by their analysis. In the end, we will treat all of the indices on an equal footing as we will use them separately to generate trading signals for our strategies. This way, we will be able to compare them by their performances instead of repeating the arguments already made by Aaron and Illing in their survey [1] for example where they show which indices captures which past known markets events. Therefore we will focus on the trading part rather on that 'validation'. We will consider that our index is 'valid' if it *performs* better than a *benchmark* which will represent a passive strategy (*buy and hold* a given portfolio).

This paper is organised as follows. Section 2 starts our work with a digression on Choice Theory in order for us to set the framework for defining risk aversion. Once this is understood, we will move away from it and try to measure risk appetite. It will remain useful to keep in mind the different concepts outlined in section 2 because 3 will be a description of the existing methods and the choices we made trying to improve them. Finally, section 4 will explain our trading framework, how we construct our portfolio, the way we extract our signal and what trade orders are placed in the market.

2 Risk Appetite

We believe that to understand risk appetite and discuss its differences from risk aversion, it is important to define what risk aversion is. We would like to take some time in this section to give the reader an quick review of Choice Theory and how to define risk aversion from it because it is a concept that can be found in classical literature (Arrow [4], Von Neumann & Morgenstern [45], Kreps [32]. We will present a short overview of the concept but a sufficient one to be able to grasp its key properties.

2.1 A Digression on Theory of Choice

In their book [45], Von Neumann & Morgenstern set the basis of Game Theory, they define the concept of expected utility and set axioms to study expected utility functions in the context of modelling economical systems through games (lotteries or gambles). They discuss quite in details the reasons that led them to defining these concepts and since it is not the main purpose of our research, we will refer to their work for the rest of this digression (specifically chapter 3 and appendix A). Nevertheless, we will go through the concepts concerning choice theory in order to set the context for risk-aversion.

2.1.1 The Preference Relation $\succ$ and utility

It is intuitive to think about the outcomes of a game and the probabilities associated to them. It is also intuitive to consider that an actor of the game will *prefer* a certain outcome of the game rather than another. Suppose for now that we have a finite number of different outcomes and let $\mathcal{O}$ be the set of all outcomes. Typically in economical applications, $\mathcal{O}$ would be a subset of $\mathbb{R}^n$ but for a simpler picture take $\mathcal{O} = \{o_1, \ldots, o_n\}$. The reader can refer to Kreps [32] §3 for a rigorous extension to uncountable sets. We assume that our actor (investor) has interest in certain outcomes rather than others. He has what is called a *strict preference* over $\mathcal{O}$. We denote this binary relation by $\succ$. We also denote not $\succ$ (or $\neg \succ$) by the symbol $\not\succ$. The absence of *strict preference* or *weak preference*, is denoted $\succeq$. The *indifference* relation is denoted $\sim$.

Definition 2.1. As defined in [32], a binary relation $\succ$ on $\mathcal{O}$ (i.e. a relation between two elements of $\mathcal{O}$) is a *preference* relation if it is asymmetric ($x \succ y \implies \neg(y \succ x)$) and negatively transitive ($\neg(x \succ y) \cap \neg(y \succ z) \implies \neg(x \succ z)$).
The *weak preference* $\succeq$ is defined as the relation satisfying $x \succeq y \Leftrightarrow x \succ y$ or $x \sim y$ for any x, y in $\mathcal{O}$. Hence the *indifference* relation can be defined as the relation satisfying $x \sim y \Leftrightarrow x \succeq y$ and $y \succeq x$ for any x, y in $\mathcal{O}$.

It can be shown then that a preference relation has nice and intuitive properties as if it was an order relation (see [32] §2 Proposition 2.3-2.4) like irreflexivity ($x \not\succ x$), transitivity ($x \succeq y$ and $y \succeq$

$z \Rightarrow x \succeq z$), completeness (for any $x, y \in \mathcal{O}$, either $x \succ y$ or $y \succ x$ or $x \sim y$) and so on.

We would now like to express this *preference* relation with real numbers so we can measure it. A *utility* function is a way to do exactly that.

Definition 2.2 (Utility)**.** A utility function is a *map* $u : \mathcal{O} \to \mathbb{R}$ such that for any $x, y \in \mathcal{O}$,

$$x \succ y \quad \Leftrightarrow \quad u(x) > u(y),$$

a representation of the *preference* relation with ordered real numbers.

2.1.2 Von Neumann-Morgenstern Expected Utility

Now we consider that the outcomes $\mathcal{O}$ are random events with associated probability measure $\mathcal{P}$ defined as

$$\mathcal{P} = \left\{ p \;\middle|\; p : \mathcal{O} \to [0,1],\ \sum_{o \in \mathcal{O}} p(o) = 1 \right\}.$$

If we then consider the preference relation $\succ$ over the set $\mathcal{P}$ instead of $\mathcal{O}$ we can define what Von Neumann & Morgenstern [45] call the *expected utility*. It is common to assume that it is *complete* and *transitive* as well as *continuous* and *independent*. The last two axioms are often referred to as the *Archimedean* axiom and the *substitution* axiom. These are not the original axioms as presented in [45] but a sensibly simpler version of them from Levin [36]. There is also a discussion on this in Kreps [32] §5 for the interested reader.

Definition 2.3 (Axioms of preference)**.** The preference relation $\succ$ on the elements of $\mathcal{P}$ satisfies the following axioms.

- **(Completeness)** The relation $\succeq$ is a complete ordering of $\mathcal{P}$, meaning that for any $x, y \in \mathcal{P}$ they must be in only one of the following cases

$$x \sim y \quad \text{or} \quad x \prec y \quad \text{or} \quad x \succ y.$$

- **(Transitivity)** For any $x, y, z \in \mathcal{P}$ we have the relationship

$$x \succeq y \quad \text{and} \quad y \succeq z \quad \text{implies} \quad x \succeq z.$$

- **(Continuity)** For any $x, y, z \in \mathcal{P}$,

$$\text{if } x \succ y \succ z \quad \text{then there exists an } \alpha \in (0,1) \text{ with} \quad y \prec \alpha x + (1-\alpha) z.$$

- **(Independence)** For any $x, y, z \in \mathcal{P}$ and $\alpha \in (0,1)$ we have

$$x \succ y \quad \Leftrightarrow \quad \alpha x + (1-\alpha) z \ \succ\ \alpha y + (1-\alpha) z.$$

To generalise that to uncountable spaces, see Kreps [32] §5. The VNM utility function is then defined through the following theorem ([32] §5).

Theorem 2.4 (Von Neumann-Morgenstern (VNM) Utility)**.** *Let $\succ$ be a binary relation on $\mathcal{P}$. This relation satisfies the axioms of Definition 2.3 if and only if there exists a utility function $u : \mathcal{O} \to \mathbb{R}$ such that for any $p, q \in \mathcal{P}$,*

$$p \succ q \quad \Leftrightarrow \quad \sum_{o \in \mathcal{O}} p(o)u(o) > \sum_{o \in \mathcal{O}} q(o)u(o).$$

The VNM utility is the quantity $U(p) = \sum_{o \in \mathcal{O}} p(o)u(o)$ for any $p \in \mathcal{P}$. This definition holds up to an affine transformation as precised in [45] and [32].

2.1.3 Risk Aversion and Arrow-Pratt Coefficient

Assume now that the outcome space $\mathcal{O}$ is a portion of the real line ($\mathbb{R}_+$ for the price of an asset). Assume also that we have an associated set of continuous (cumulative) probability functions $\mathcal{P} = \{p(\cdot) \mid p(o) = \mathbb{P}((-\infty, o])\}$. We write then

$$U(p) = \int u \; dp$$

for the expected utility representation of preferences on probabilities and we assume they exist.

Definition 2.5 (Risk Aversion)**.** Let us consider a game with outcome space $\mathcal{O}$ and its associated set of probabilities $\mathcal{P}$. Its expected value with respect to the probability function $p \in \mathcal{P}$ would be the expected utility $U(p) = \mathbb{E}^p[\mathcal{O}] = \int x dp(x)$. The player is said to be *risk-averse* if he *prefers* earning the guaranteed outcome $\mathbb{E}^p[\mathcal{O}]$ rather than playing at the original game. That is, he would rather have the expected value of the original game with probability one than an uncertain gain even if potentially higher. In terms of his utility function, if it exists it satisfies

$$\int u \; dp \leq u \left(\int x \; dp(x) \right), \text{ for all } p \in \mathcal{P}.$$

We recognise *Jensen's inequality* characterising *concave* functions. In the same fashion, we can define the utility function of a *risk loving* investor (see Figure 1).

Pratt [46] and Arrow [4] have independently characterised risk aversion using the local curvature of the utility function. This is now known as the Arrow-Pratt risk aversion coefficient. The idea behind it is simple, looking at the previous Figure 1. Say that we have two investors A and B with utility function u_A and u_B respectively. Then investor A will be *globally more risk-averse* than investor B if his utility function is 'more concave' than the other one, i.e. if there exists another concave function T such that u_A is a composition of u_b and T ($u_A = T \circ u_B$).

Definition 2.6 (Arrow-Pratt (local) measure of risk aversion)**.** Suppose that $u(\cdot)$ is a twice continuously differentiable utility function. Then the Arrow-Pratt *measure of absolute risk aversion* exists and is defined as

$$\rho_{abs}(x) = -\frac{u''(x)}{u'(x)}.$$

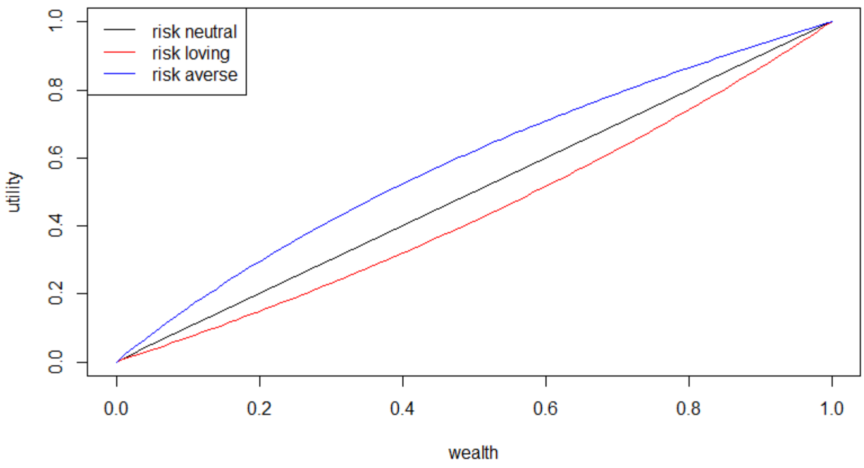

Figure 1: Qualitative view of attitude towards risk represented by utilities.

This is also known as the curvature of the utility function at point x. The Arrow-Pratt *measure of relative risk aversion* exists and is defined as

$$\rho_{rel}(x) = -x\frac{u''(x)}{u'(x)} = xR_{abs}(x).$$

The latter is an adjusted version of the former that takes care of the differences in wealth levels. Indeed, the preferences of two players with very different initial wealth will potentially be different and so will their utility functions be. For example think of one player with initial wealth of 1$ and another player with initial wealth of 1000$. Their preference on a game with expected value of 1$ will probably be very different. This difference of wealth is taken into account in the relative risk aversion coefficient. So if our utility functions are in terms of wealth, the absolute risk aversion should decrease with the wealth but not the relative one.

Example 2.1. *Here are examples of functions we can use to illustrate the difference between both coefficients. For the absolute risk aversion for example we can use*

- *increasing absolute risk aversion*

  $$u(x) = \frac{x^{1-\rho} - 1}{1-\rho} \Rightarrow u'(x) = x^{-\rho} \Rightarrow u''(x) = -\rho x^{-\rho-1}, \quad \rho \in (0,1)$$

  so the Arrow-Pratt coefficient is $\rho_{abs}(x) = \rho x$ *is increasing in* x*,*

- *constant absolute risk aversion*

  $$u(x) = -e^{-x} \Rightarrow u'(x) = e^{-x} \Rightarrow u''(x) = -e^{-x}$$

  so the Arrow-Pratt coefficient is $\rho_{abs}(x) = 1$ *is constant,*

- *decreasing absolute risk aversion*

$$u(x) = \ln(x) \Rightarrow u'(x) = 1/x \Rightarrow u''(x) = -1/x^2$$

so the Arrow-Pratt coefficient is $\rho_{abs}(x) = 1/x$ *is decreasing in* x*.*

And for the relative risk aversion we can use for example

- *increasing relative risk aversion*

$$u(x) = x - x^2 \Rightarrow u'(x) = 1 - 2x \Rightarrow u''(x) = -2$$

so the Arrow-Pratt coefficient is $\rho_{rel}(x) = \frac{2x}{1-2x}$ *is increasing in* x*,*

- *constant relative risk aversion*

$$u(x) = \ln(x) \Rightarrow u'(x) = 1/x \Rightarrow u''(x) = -1/x^2$$

so the Arrow-Pratt coefficient is $\rho_{rel}(x) = 1$ *is constant,*

- *decreasing relative risk aversion*

$$x > 0, \quad u(x) = e^{-1/x} \Rightarrow u'(x) = \frac{1}{x^2}e^{-1/x} \Rightarrow u''(x) = -e^{-1/x}\left(\frac{1}{x^4} + \frac{2}{x^3}\right)$$

so the Arrow-Pratt coefficient is $\rho_{rel}(x) = \frac{1}{x} + 2$ *is decreasing in* x*.*

Remark 2.7. We can keep in mind, and it is a good thing to remember, that the utility function can be seen as the solution to a second order differential equation of the form

$$u''(x) + \rho_{abs}(x)u'(x) = 0$$

in the absolute risk aversion case or

$$xu''(x) + \rho_{rel}(x)u'(x) = 0$$

in the relative risk aversion case. This means that if we can find a general solution, it is twice continuously differentiable and that has the utility properties.

2.2 From aversion for risk to appetite for risk

Risk appetite and risk aversion are both concepts that financial market investors will be familiar with. But the difference between both of them is not well defined. Some authors use both terms interchangeably as the opposite of one another as if they were the two sides of the same coin like Kumar & Persaud in [34]. However others disagree with that and argue that they are in effect sensibly different and they should be distinct concepts while providing ways of measuring one and not the other, like Gai & Vause in [23].

2.2.1 Historical Volatility as a first risk measure

A standard way to estimate the risk of an asset would be to estimate his historical volatility, i.e. the standard deviation of his movements during a period in the past. We will give a definition of this concept along with other useful definitions that will continue to be used for the rest of this project.

Definition 2.8 (Basic concepts and notations)**.** Suppose that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Ω is the set of al possible outcomes, $\mathcal{F}$ is a set of events, $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure and $(\mathcal{F}_t)_{t\geq 0}$ is a *filtration*. We assume that it satisfies the *usual* conditions. Suppose we have *price process* $(S_t)_{t\geq 0}$ on this *filtration*. Then the *returns* of this price are defined as

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

It is also common to work with the log-returns instead for their additive properties. They are defined as

$$r_t = \ln(1 + R_t) = \ln\left(\frac{S_t}{S_{t-1}}\right)$$

and R_t is very close to r_t for very small values of R_t (i.e. $\ln(1 + R_t) \approx R_t$, for $R_t << 1$). Now for the returns time series $(r_t)_{t\geq 0}$, for a past period of n time steps with a n days *historical mean* defined as $\overline{r}_t = \frac{1}{n}\sum_{i=t-n}^{t} r_i$, the *historical volatility* can be computed as

$$\sigma_h = \sqrt{\frac{1}{n-1}\sum_{i=t-n}^{t}(r_i - \overline{r}_t)^2} \tag{2.1}$$

where the *unbiased* version of the empirical variance has been used. This will be the basic framework in which we will always assume to be throughout this work as it is the usual one for studying time series in financial mathematics.

Typically the returns exhibit a *stationary* (or *covariance-stationary*, see Cont [12]) behaviour with mean usually close to zero. If it is verified, then σ_h^2 is said to be the *realised variance*. We can see these empirically in Figure 2 where the price of the S&P 500 is compared to its log-returns, its 1-year historical volatility estimate as well as the VIX index for a volatility benchmark.

To explain the naive concept of risk aversion, let's look at a naive example.

Example 2.2. *Suppose that investors A and B hold currently two assets X and Y. The 1 Year historical volatilities are* $\sigma^X_{hist} = 15\%$ *and* $\sigma^Y_{hist} = 2.5\%$ *respectively.*
Investor A has a 50-50 allocation, which means the proportion of asset X and Y is equal in his portfolio. That is at time t his wealth process is $w^A_t = \frac{1}{2}w^X_t + \frac{1}{2}w^Y_t$*, one half of it comes from the value of X and the other half from the value of Y.*

Now investor B has a different allocation, he sees that asset X has been more volatile in the past year and wants to take profit from that so he decides to have a 75-25 allocation, i.e. $w^B_t = \frac{3}{4}w^X_t + \frac{1}{4}w^Y_t$.

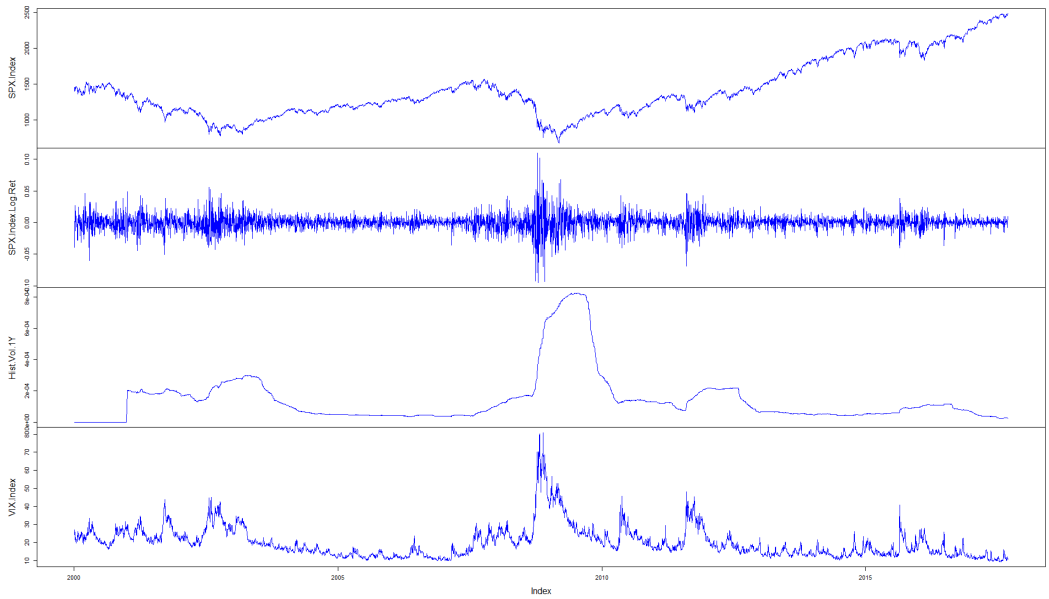

Figure 2: S&P 500 Price, Log-Returns, 1 year historical volatility and VIX from 2000-01-01 to today.

The naive concept of risk appetite is illustrated here saying that investor B has more appetite for risk than investor A because he is willing to allocate more of his wealth in asset X which has a larger historical volatility than asset Y.

In the previous example, risk *appetite* and risk *aversion* are used interchangeably. But is this really the case ? In a complex financial landscape where the actors have potentially different behaviors towards risk, how do we measure the consequences of their anticipations reflected in their actions ?

2.2.2 Measuring Risk Appetite

Throughout our research, we have come to realise that this topic is discussed by the practitioner's literature but not really by the academic one. Misina [44] expresses his view as to the fact that academicians reject the idea of time-varying preferences because it would break the classical asset pricing theory (for example Cochrane [10]). Consequently, some methods are only statistical methods extracting information from aggregates of market indices but not relying on any kind of model.

First we have to separate risk and risk aversion and their consequences (see Misina [44] on *separability*). If there is a change in the riskiness (or a estimate of this riskiness recognised by enough investors, such as *historical volatility* or future volatility like the information given by the

VIX index [51]), investors which are risk averse will ask for more stable assets, driving their price up. And by abandoning (liquidating) their positions in risky assets, they will drive their prices down. The opposite scenario is also true. Then, when there are large moves in prices, we want to be able to capture if a change in the behavior towards risk has led to it. But in order to do so we have to allow something to change in time. It is interesting to read the steps described in Kumar & Persaud [34] and discussed in Misina [42]. They do *not* distinguish risk appetite from risk aversion but they make it clear that it is possible to differentiate price movements caused by changes in risk and price movements caused by changes in risk aversion by allowing the risk aversion to be time-varying. Measuring risk appetite is therefore possible as a *quantitative* measure as precised in [42] (i.e. we will get a unitless value, not a metric but rater a 'score'). Other authors like Jackwerth [28] and Ait-Sahalia & Lo [2] talk about the concept of *implied* risk aversion, which is also theoretically described by Misina [44]. This allows the expected utilities of the actors to be time-varying by considering that their subjective probabilities change over time, while their utilities remain constant (preferences are not time-varying). Finally, approaches like the one by Gai & Vause [23] mark a clear distinction between risk appetite and risk aversion, which is basically constraining the aversion to be constant (or slowly moving) and the appetite to be more reactive and volatile (time-varying).

The methods to extract information from time-varying behaviours towards risk is the object of this work and we will go through the methods of the different authors above. While they differ by their intuition, they all have the same purpose which is to try to capture large moves in investor behaviour predicting a large move in prices.

3 Risk Appetite Indices

Inspired by the survey from Aaron & Illing [1] (a broad and complete survey which is very precious in such an heterogeneous field), we have tried to implement some of the risk indices it describes. We tried to replicate them when we had access to the correct data, modifying others when we found it sensible to do so and eventually going beyond the scope this a survey by using them to build trading strategies (see Section 4). As a consequence, we will use the terms 'indices', 'measure' and 'indicators' interchangeably in general discussions because none of these terms is specific to one risk appetite index. Plus we will not repeat what each index does all the time, we will refer to them by labels and the reader can jump back and forth to the dedicated paragraph to see how it is built. We are forced to do so as we will soon see that these 'risk appetite' measures are based on very different methods. We will of course make the distinction whenever required by the context.

3.1 Market data : investors 'memory' and window length

All our data is provided by Bloomberg L.P. private database which we had access to during our research. All the tickers of the financial time series used for each index can be found in the appendix A.4. Although we tried to make the fewest arbitrary choices when downloading our data, we have faced certain difficulties with time-indexing. We want to address these details here as we had to use more recent data sets than used by the majority of the authors from which our work is based on. Indeed, most of the risk appetite indices have been implemented more than 10 years ago with data that is nowadays considered 'useless' to some investors as it is prior to the financial crisis. We faced two main types of problems during this project and they are the reason why we couldn't reproduce exactly the work done in past research.

- The first type of problem we encountered was when trying to download the same data set as the authors. Either the data they used is not available anymore today or too much of it is missing. It can happen that several time series used by some authors stopped being traded for some time after their research was published or some global indices stopped being provided by Bloomberg or the data has just never been public and the authors used their own database (for example indices computed by some banks and made available on Bloomberg stopped for some reason). So we had to use a smaller data set or substitute the missing parts by what we though was reasonable when we could.

- The second problem faced is a more profound one and has to do with the financial crisis data and the investors 'memory'. Although risk appetite indicators were created for this very purpose, i.e. try to predict large movements using market indicators, the authors could not have thought about an event of this amplitude happening a few years after their respective papers were published. What we mean by investors 'memory' is that all our methods are using so-called *rolling windows*. This is a way to select the data we want to use based on the time step at which we stand. Say we stand at time t and we want to compute for some method the risk appetite estimate at time $t+1$, we will use all the data available from time $t-L+1$ up to t, where L is the length of that window. That is, we consider that the actors of the market *forget* what happened before the beginning of the window as the data will not be included in the computations. This is then a very sensitive parameter because, as we will see in the following parts, the size of this window tells a lot about the kind of information we wish to extract from our data : a short window (approximately a few months long) will be used to capture recent trends and a long window (one year up to ten years) will be used to capture long-term trends. Figure 3 shows exactly this, the choice of the window matters as we can see the 1-year estimate is smoother than the 1-month estimate of historical volatility.

So we have proceeded as closely as possible to what was done in previous research. And when we deviated from the initial methodology, it was either because the data required it as aforementioned, or because we had access to more recent and potentially better methodologies than the authors did at the time, or even just because we thought the methodology could be improved. These modifications will be presented and justified as they become relevant along our work.

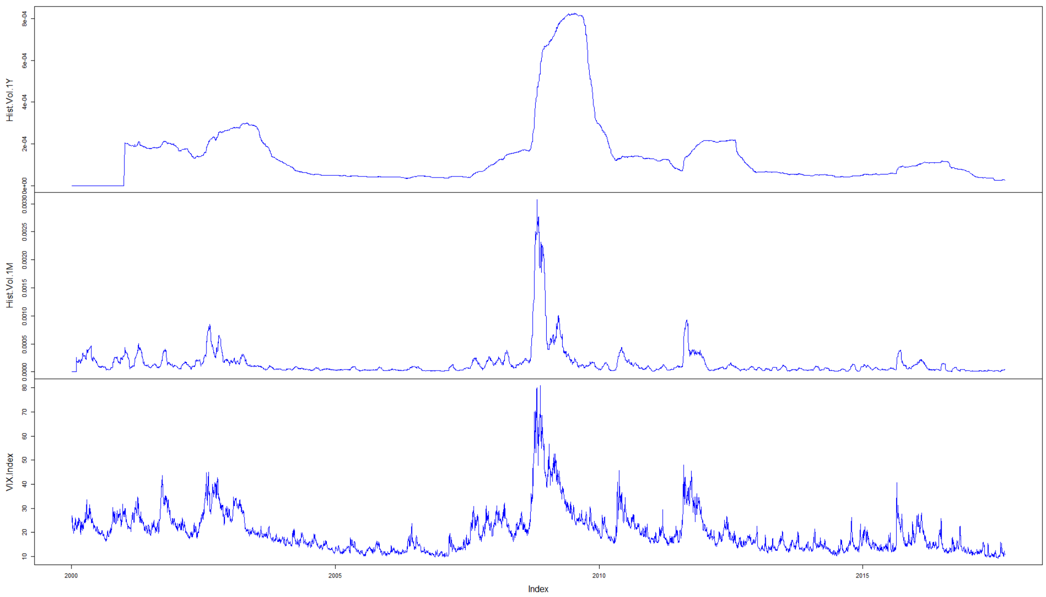

Figure 3: S&P 500 1-year historical volatility, 1-month historical volatility and VIX from 2000-01-01 to today.

3.2 Market Indices

The terminology comes from the European Central Bank (ECB) [18] who described a panel of risk appetite and other variations of such indicators, as Illing and Aaron did in their survey [1]. We have encountered a number of methodologies to built a risk appetite measure that were based on very simple principles and a large degree of intuition. We call these indices *market indices* (or *empirical indices*) as they are not based on any type of model but rather on data. The methodology is described in [1] and we shall refer to them as done in this paper. Incidentally they will carry the name of the bank or fund that has developed the method. They all use a small database composed of well-know market indicators. Since we didn't have access to the documentation the author had, we took their table in BOX1 on page 39 for granted. We describe here the simple methodology for these four indicators. The data tickers used for each one of them is found in appendix A.4.

- The UBS Investor Sentiment Index (UBS) as well as the Merrill Lynch Financial Stress Index (ML) both use a standard normalisation procedure that is known as the *z-transformation* method which consists in subtracting the mean a dividing by the variance every observation of our sample. Doing this for each time series independently gives a way to compare them. We therefore take the average across all series every day and take that number as the risk appetite measure. We also came up with a more robust method we call *robust z-scoring* which takes out outliers before transforming the data into a percentile and then a *z-score.*
- JPMorgan's Liquidity Credit and Volatility Index (LCVI) is based on a *percentile scoring* method. It transforms the observations into quantiles of their own historical distribution and then averages them across assets every day to give a risk appetite measure. Here for example we use an *expanding window* because otherwise the measure is to noisy. It will still be noisy at the beginning though but at least it will become more and more representative of the whole distribution.

3.2.1 Z-scores

We will have to adapt the general definition of a *z-transformation* to our empirical cases on a rolling window basis. But to start, let's see the general way of defining such a *z-score.*

Definition 3.1 (Z-Transformation)**.** Let X be a random variable with distribution f_X, mean μ_X and variance σ_X^2. Suppose a set of observations $S = \{x_1, \dots, x_n\}$ is sampled from f_X. Then the *z-transformation* of the observations is

$$z_i = \frac{x_i - \mu_X}{\sigma_X}, \; \forall i \in \{1, 2, \dots, n\}.$$

Remark 3.2. Note that to perform a *z-transformation* we need to know the mean and the variance of our sample. This is not a problem immediately but it implies that the distribution we hope to compare this score to has a defined mean and variance. Indeed, it is common to have an idea of the distribution that our sample follows. For example let's say that we study a sample of three months of daily log-returns of a price process. If we think the sample distribution should be Gaussian, we will calibrate parameters of the Gaussian distribution μ and σ as seen before in (Equation 2.1), i.e. we will use a so-called *out-of-sample* long-term estimate of those two parameters. They are called the *population* parameters as we think that our sample data - or the so-called *in-sample* observations - is a collection of observations sampled from the same distribution. We then score our sample points using the *population* mean and standard deviation as in the definition above. The *z-scores* then measure a standardised distance at which the sample points are from the supposed density.

Inspired by this idea of scoring, and since these score can be defined without the assumption of normality (still we have to assume that our population distribution has a finite mean and a finite variance) we will use the data available at each *rolling window* to score the next observations.

Definition 3.3. We select one window of observations from a time series X_t, our window is of length L, each point of the window is x_i with $i \in \{t-L+1, t-L+2 \ldots, t\}$. Then the *z-score* of the observation at time $t+1$ (out-of-the-window observation x_{t+1}) is defined as

$$z_{t+1} = \frac{x_{t+1} - \hat{m}_t^L}{\hat{\sigma}_t^L},$$

where the empirical *rolling mean* is defined by

$$\hat{m}_t^L = \frac{1}{L} \sum_{i=t-L+1}^{t} x_i$$

and the empirical *rolling standard deviation* by

$$\hat{\sigma}_t^L = \sqrt{\frac{1}{L-1} \sum_{i=t-L+1}^{t} \left(x_i - \hat{m}_t^L\right)^2}.$$

That is a way of saying we assume the data inside the window $X_t^L = \{x_{t-L+1}, x_{t-L+2}, \ldots, x_t\}$ to be the *population* and the observation just outside of it, x_{t+1}, to be the *sample.* So we use the empirical mean and standard deviation of the window to score the point just after it. Rolling this operation for all t we transform all our data (except for the first L ones as we need them to compute the first index value) into a *z-score* time series.

We can see the result of such a method in Figure 4 for UBS and ML indices. We can see that they are very similar and capture great moves during the financial crisis.

Since this method of scoring can sometimes be noisy (see Figure 4 ML z-score) and uses the empirical mean which can be very sensible to outliers, we also tried to somehow improve the robustness of this method by taking out some extreme values. We will refer to them as the *robust z-scores.*

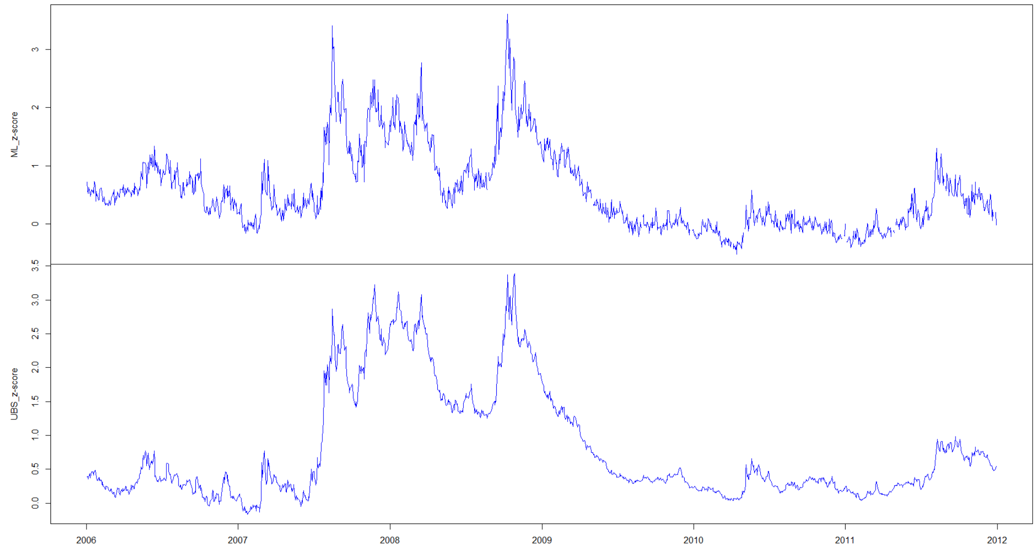

Figure 4: ML z-score index and UBS z-score index from 2006-01-01 to 2011-12-31 with 3 years rolling window.

Definition 3.4 (Quantiles)**.** Let $\alpha \in (0,1)$ and X be a random variable on a probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P})$. Then the *$\alpha$-quantile* of X is defined as

$$q_\alpha := q(\alpha) = \inf\{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) \geq \alpha\}.$$

It is the smallest observation such that the probability of another observation sampled from this random variable X being smaller than this is larger than α. If we agree that $\alpha \in \{0.01, 0.02, \ldots, 1\}$ then all the possible *α-quantiles* are called *percentiles*.

Suppose we have a set of observations $X_N = \{x_1, x_2, \ldots, x_N\}$ sampled from X. We order them in increasing order to get the ordered set $X_N^{<} = \{x_{[1]}, x_{[2]}, \ldots, x_{[N]}\}$. If we want to compute the *5-percentile* q^{05} and *95-percentile* q^{95} of our set we compute

$$\begin{cases} q^{05} := q_{0.05} = \inf\{x_i \in X_N \mid \mathbb{P}(X_N \leq x_i) \geq 0.05\} = \inf\{x_{[i]} \in X_N^{<} \mid i \geq 0.05N, i \in \mathbb{N}\}, \\ q^{95} := q_{0.95} = \inf\{x_i \in X_N \mid \mathbb{P}(X_N \leq x_i) \geq 0.95\} = \inf\{x_{[i]} \in X_N^{<} \mid i \geq 0.95N, i \in \mathbb{N}\}. \end{cases}$$

This means that the 5% of the data with lowest value is below q^{05} and 5% of the data with highest value is above q^{95}.

Coming back at out *robust z-scores*, they are computed as follow. First we compute a rolling upper and lower bound to filter out outliers in our data. We do this by keeping track of the *rolling* 95%-*quantile* q_t^{95} and the *rolling 5%-quantile* q_t^{05} at each time step t. If we are at time t

with window $X_t^L = \{x_{t-L+1}, x_{t-L+2}, \ldots, x_t\}$ then the points of the sample in order are written $X_t^{L,<} = \{x_{[1]}^t, x_{[2]}^t, \ldots, x_{[L]}^t\}$. The rolling quantiles are then defined as

$$q_t^{95} = x_{[\lceil 0.95L \rceil]}^t \quad \text{and} \quad q_t^{05} = x_{[\lceil 0.05L \rceil]}^t,$$

where $\lceil \cdot \rceil$ is the *ceiling* function. We then keep the data points that are within these boundaries in order to eliminate the outliers. In order to do this, instead of deleting extreme values plain and simple, we transform it into a relative position between lower and upper quantiles. That is the point x_t will be transformed into

$$\begin{cases} p_t = 1 \text{ if } x_t > q_t^{95}, \\ p_t = 0 \text{ if } x_t < q_t^{05}, \\ p_t = \frac{x_t - q_t^{05}}{q_t^{95} - q_t^{05}}, \text{ otherwise.} \end{cases} \tag{3.1}$$

This will have the effect of mapping our data into the interval $[0, 1]$ and 'cut' the extreme values. Finally we transform these points into a *rolling z-score* as we explained before. Note that in order to do this we will need a second *rolling window*. Indeed, we are using a first window to map our observations to a relative position. Then, transforming the positions into scores require a second window (typically shorter than the first one), so we will end up having less points than for the usual *z-scoring*. However this time series will be potentially more robust as extreme moves will be disregarded. We can see an illustration in Figure 5. We can already spot the difference between both time series, especially during the financial crisis where we can expect a lot of observations to be ignored by the robust version of the index. The resulting indicator is perhaps more revealing of the shift in appetite for risk during the crisis as we can see that the amplitude of the movements is less heterogeneous. This might be an advantage thereafter as it might trigger other trades.

Finally, for the LCVI index, we transform each time series into what we call a *percentile score*. This is based on the same idea as before: try to measure the distance of x_{t+1} to its past distribution X_t^L. This will be done in this case by transforming x_{t+1} into a quantile of the empirical distribution of the population X_t^L, namely $f_{X_t^L}$. The *percentile score* is then

$$p_{t+1}^c = f_{X_t^L}(x_{t+1}).$$

This will detect if x_{t+1} is behaving as a member of the population X_t^L or an outlier, which will result in an extreme value in our index, see Figure 6 for an illustration with the LCVI data base (see appendix A.4 for the tickers).

3.2.2 Mahalanobis scores

One can then wonder if these methods of scoring are enough. What we have been trying to do until now is to find a good way of comparing each observation x_t of a time series with its past to

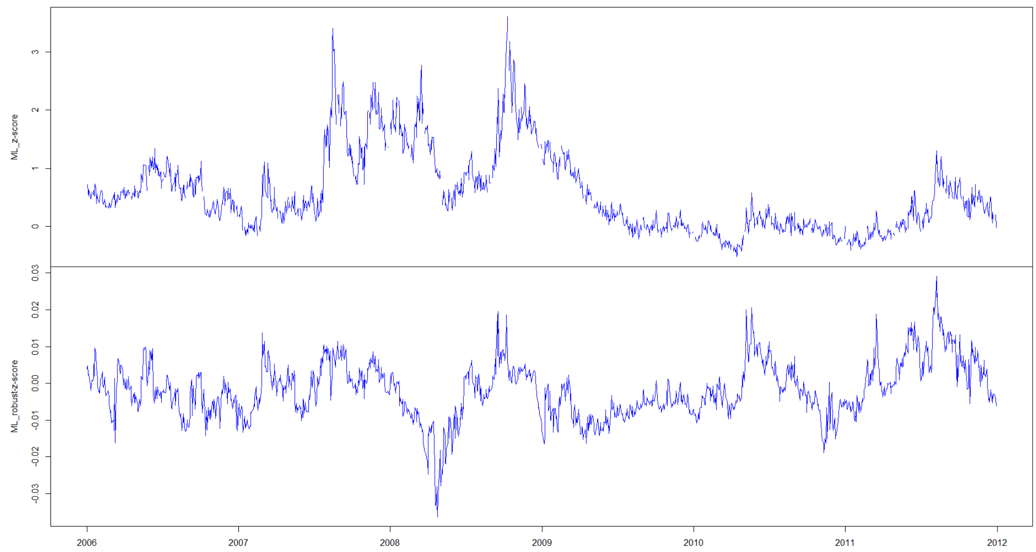

Figure 5: ML z-score index and ML robust z-score index from 2006-01-01 to 2011-12-30 with 3 years rolling window, second rolling window of 6 months and a 95%-quantile as well as a 5%-quantile.

see if this value is an outlier or not by keeping track of an estimate for the mean and the variance and using them to compute distance to the past distribution. There exist a more general way of measuring such a distance in a multivariate way : the Mahalanobis distance [38]. It will give us a potentially better way of scoring as we consider the correlations between our time series instead of scoring each one individually.

Definition 3.5 (Mahalanobis distance)**.** As discussed in Mahalanobis [38] and McLachlan [40] the Mahalanobis distance between two populations, the vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, drawn from the same multivariate distribution with constant and non-singular covariance matrix Σ is defined as

$$\Delta = \sqrt{(\mathbf{x} - \mathbf{y})\Sigma^{-1}(\mathbf{x} - \mathbf{y})^T}.$$

Since Σ is a non-singular covariance matrix, it is positive-definite so Δ is a well defined distance measure.

Remark 3.6. The intuition behind it is that plugging this inverse covariance matrix in the metric will transform the space we are measuring the distance in and potentially give a better measure that the Euclidean one. We can also see it as the Euclidean distance between vectors in a distorted space. Note that if the covariance matrix is the identity matrix then the Mahalanobis distance is nothing else but the Euclidean distance in $\mathbb{R}^n$.

Figure 6: LCVI index 2006-01-01 to 2011-12-30 with 3 years rolling window.

In our case, we want to asses if our observations at time $t+1$ (across all time series) are drawn from the same distribution as the one we suppose the past multivariate one. More precisely, if we have N assets, the population (observations in our window) is

$$\mathbf{X}_t^L = (\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N) = \begin{pmatrix} x_{t-L+1}^1 & x_{t-L+1}^2 & \cdots & x_{t-L+1}^N \\ x_{t-L+2}^1 & x_{t-L+2}^2 & \cdots & x_{t-L+2}^N \\ \vdots & \vdots & & \vdots \\ x_t^1 & x_t^2 & \cdots & x_t^N \end{pmatrix} \in \mathbb{R}^{L \times N},$$

with empirical mean $\hat{\mu}_{\mathbf{X}} \in \mathbb{R}^N$ (the vector or empirical mean of each column of $\mathbf{X}_t^L$) and (unbiased) empirical covariance matrix $\hat{\Sigma}_N \in \mathbb{R}^{N \times N}$ with $\left(\hat{\Sigma}_N\right)_{ij} = \text{cov}(\mathbf{x}_i, \mathbf{x}_j)$. We then have a vector of observations $X_{t+1} = (x_{t+1}^1, x_{t+1}^2, \dots, x_{t+1}^N)$ and we want to measure its distance to the window population. The Mahalanobis distance is then

$$\hat{\mathbf{\Delta}}_{t+1} = \sqrt{(X_{t+1} - \hat{\mu}_{\mathbf{X}})\hat{\Sigma}_N^{-1}(X_{t+1} - \hat{\mu}_{\mathbf{X}})^T}.$$

The results are shown in Figure 7. We can see that the index based on the Mahalanobis distance has a smaller (relative) variance and more visible outliers. This is useful since we aim to detect more ‘important’ shifts in investors appetite for risk. It supports our insight that the Mahalanobis for scoring is a potential method.

It is then a matter of using the techniques outlined above to compute all the indices we need. They are resumed in one graph, Figure 8.

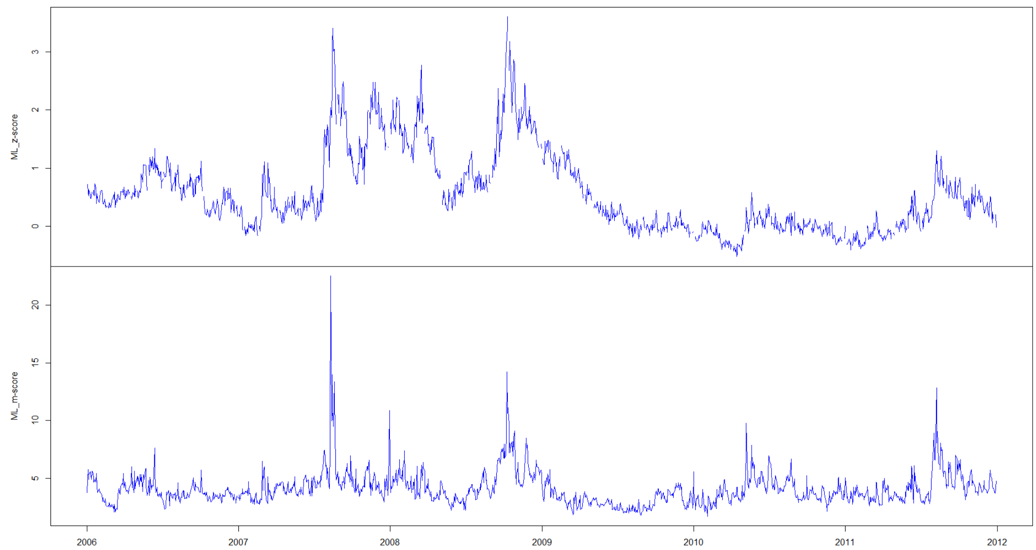

Figure 7: ML z-score index and ML m-score index from 2006-01-01 to 2011-12-30 with 3 years rolling window.

3.3 Risk-On / Risk-Off Index

In HSBC's paper [26] as well as in Kritzman *et al.* [33], the authors use a large collection of market indicators as a proxy for the behaviour of the market. We decided to differentiate this category of risk indicators as they maybe move slightly away from the risk appetite principle towards a more general measure of the inherent risk in the global markets, a more *systemic* risk as said in [33]. The terminology *risk-on/risk-off* is also found as in [26] where the authors describe their method of assessing risk in the markets by looking at a heat-map which is just a large correlation matrix. But the literature on *systemic* risk and *contagion* is very furnished and has been an active field of study since the latest financial crisis (see [27]). However, what is interesting to see is that during known periods of stress the correlations tend to be very high or low between classes of assets. Even though it is a simple idea, it is reasonable to think that observing the correlation of a sufficiently large sample of the market, or at least a selection of driving assets would give an idea of the state of stress experienced by the market. This can be related to liquidity and crowdedness concepts that are also interesting concepts one can consider as potential indicators of unusual behaviour towards risk (see MSCI [5] or Deutsche Bank [14]). Having this in mind, we are going to study the concept of Principal Component Analysis because it is a powerful way to reduce the dimension of our data so that instead of having large (average) correlation matrices as indicator we will have only one number each day.

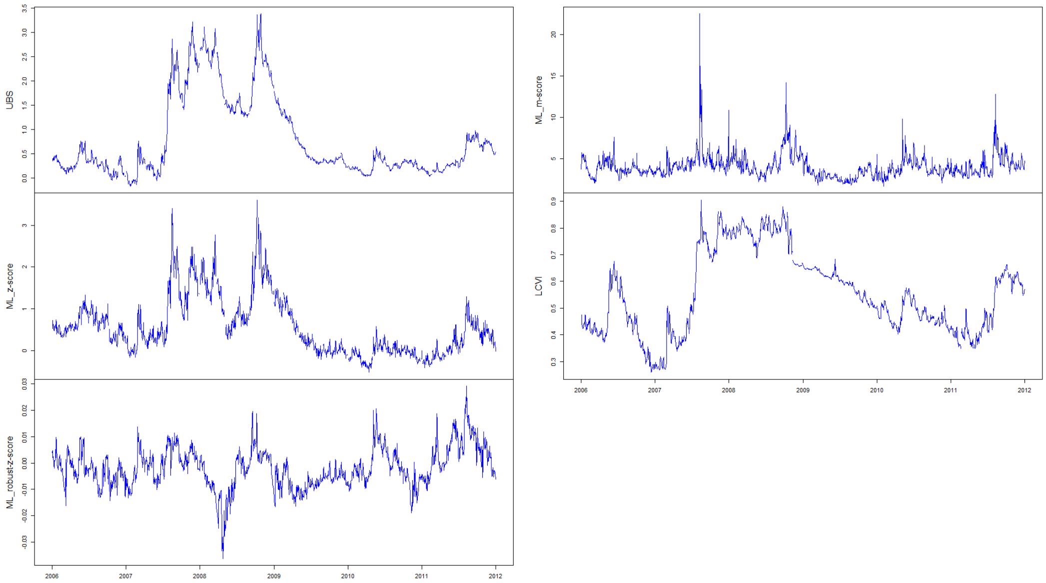

Figure 8: UBS z-scores, ML z-scores, ML robust-z-scores, ML m-scores, LCVI percentile-scores indices from 2006-01-01 to 2011-12-30 with 3 years rolling window.

- The *risk-on/risk-off* index by HSBC (referred to as HSBC from now on) uses a data base described in [26], appendix B, where the authors also describe their methodology. The idea is to do a *rolling PCA* which means do a PCA on a window that will slide at every time step as we have seen before. Then we extract the first principal component and his variance that will be accepted as the risk measure. It will be the the linear combination of the asset that will *explain* the most of the variance of the window.

- The idea of [33] is very similar, we will refer to this index as AR (for Absorption Ratio) from now on. Instead of taking only the fist principal component of the PCA each day, we take more (exact number has to be determined) and we do a ratio of their total variance over the total variance of the data in the window. A more complete study of this ratio is done in the original paper [33] but we will restrain ourselves here at the same data set as for HSBC index.

3.3.1 Principal Component Analysis (PCA)

We are not going to define again PCA here as it is very well done in Bibby *et al.* [6] for example. We consider that we understand the concept and just outline the main steps and results.

- We decompose the covariance matrix of a population using the Spectral Decomposition Theorem,

- this decomposition can be seen as a change of basis and in it the basis vectors are the eigenvectors, each of them having a different eigenvalue and all orthogonal to each other, each eigenvector will be a linear combination of the initial variables,
- each eigenvalue is the variance of its eigenvector and represents the contribution of the total variance 'explained' by the eigenvector,
- so ranking them in decreasing order we can select the eigenvector having the largest eigenvalue: this is the principal component.
- To the second largest eigenvalue correspond another eigenvector orthogonal to the second one: this is the second component, and so on ...

Remark 3.7. Note that PCA is sensitive to the scale of the variables we start from because it looks at the linear combination that has the largest variance. This is leads us to use the correlation matrix instead of the covariance. However, this implies that we give the same importance to each time series - which is what we want here in our case. So prior to a PCA we scale our data by taking the log-returns and weighting them using an exponential weighting as described earlier (the exponential weighting has nothing to do with PCA it is a time re-scaling not a variable re-scaling but we mention it still because it's important to do it after we have selected a window of data and scaled the observations in space).

The HSBC Index then only looks a λ_1 the first eigenvalue and monitors it through time. The AR index however selects more components and then defines the *absorption ratio.*

Definition 3.8 (Absorption Ratio)**.** Following [33] the absorption ratio at time t is defined as

$$r_t = \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{j=1}^{N} \lambda_j}$$

where the λ_j for $j \in \{1, 2, \ldots, N\}$ are the variances of all the N components and λ_i for $i \in \{1, 2, \ldots, K\}$ are the variances of the first K components of the PCA. The number r_t is the *proportion* of the total variance of the total portfolio that is due (or 'explained') by the first K components (the term portfolio here only refers to all the assets we are considering, not to any investment strategy).

However, as we can see in Figure 9, this yields a very smooth indicator for HSBC index, too smooth to be reliable because we took a wide window. But actually the problem doesn't disappear even when we take a shorter window so we need to correct this using another technique. This will be presented in the following part where we talk about exponential weighting.

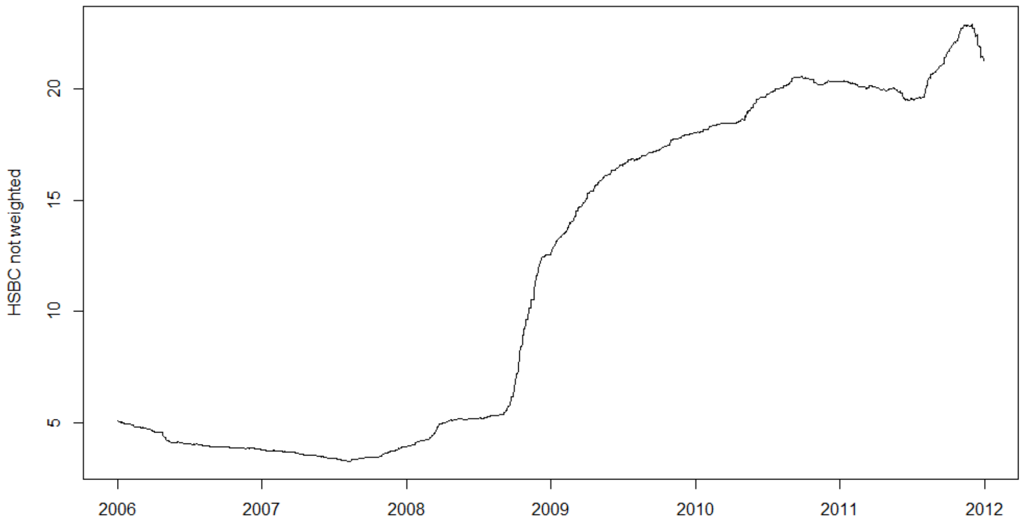

Figure 9: HSBC index from 2006-01-01 to 2011-12-30 with 3 years rolling window.

3.3.2 Data weighting

This idea is fundamentally an empirical intuition rather than a theoretical one but it has an economical meaning as discussed previously when talking about investors' memory. We use *rolling windows* so that at every time step the data we study will be the last L available points and this population will be used as a proxy for our historical estimates. For example, if our data-set is constituted of log-returns and we are standing today 12$^{\text{th}}$ of September 2017 and we look at daily log-returns of the SPX Index with a window of 30 years and draw the historical distribution of those, what Figure 10 shows is a quite smooth density (see the *stylized fact* 'aggregational gaussianity' from Cont [12]). But this is not ideal since it will probably not contain information specific to our direct/local market environment. But if we use a shorter window, say 6 months, what we get is a more skewed distribution with heavier tails than the Gaussian one (see again [12]) which can even be multi-modal distribution depending on the technique we use (kernel estimators, see section 3.5.2), see Figure 11.

It is quite clear from VIX data for example (see Figure 3, lowest time series) that the market environment pre-crash, in-crash and post-crash exhibit clear general different behaviors. Whereas it is not the purpose of this work to try and identify these regimes and regime-switching points, we believe that this is closely related. If we could identify such periods where markets behave uniformly we could then have a better idea of the window length to select, but this goes beyond the scope of our study. So there is an important decision to be made as for the length of that

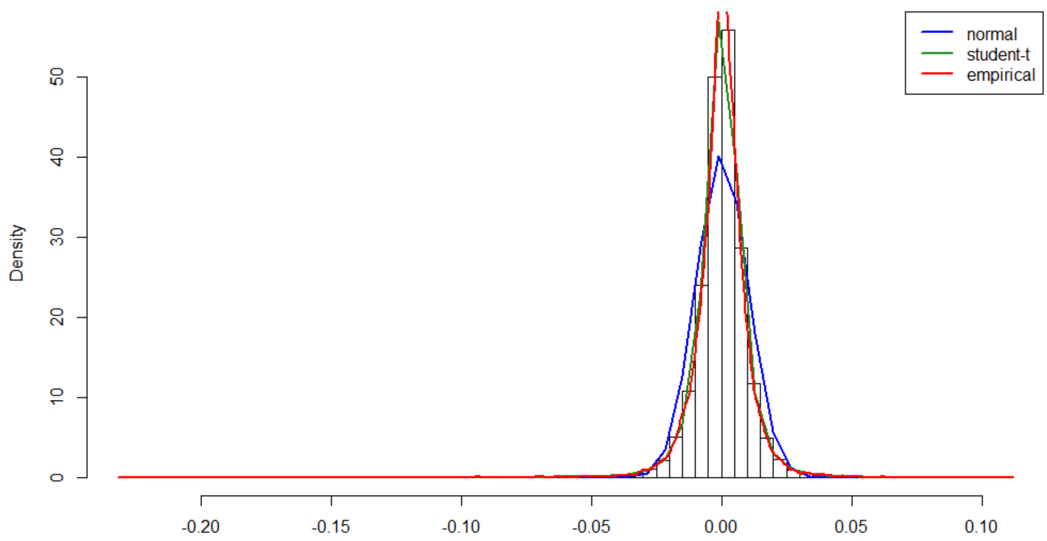

Figure 10: S&P 500 log-returns distribution from 1957-01-01 to 2017-08-29.

window : should it be short and should we forget about recent events as we need a reactive (and potentially noisy) indicator ? Or should we *a contrario* use a longer window because we want to ‘remember’ the past to learn and do better in the future (at the potential cost of a too smooth indicator) ? Following work done in Aste *et al.* [3], we can improve robustness of our data selection by taking maybe a longer window than we previously did but weighting the data points inside it in order to smoothly decrease the importance given to each point as time advances. We use a smooth function, typically an exponential, so that the data enters from the right of the window and is given the most weight and exits smoothly the window by the left side of it as the time passes.

Definition 3.9 (Exponentially Weighted Window)**.** Suppose we are standing at time t (fixed), our sample data takes all points from time t back to time $t-L+1$. The weight w_{t-k} we give to the sample data point x_{t-k} for all $k \in \{0, 1, \dots, L-1\}$ is defined as

$$w_{t-k} = w_0 e^{\alpha(t-k-L)},$$

with $\alpha > 0$ a decaying factor and w_0 a parameter. To find w_0, following [3] we put the following constraint

$$\sum_{k=0}^{L-1} w_0 e^{\alpha(t-k-L)} = 1$$

on the weights. Then we are able to find (see appendix A.1)

$$w_0 = w_0(\alpha) = \frac{1-e^{-\alpha}}{1-e^{-\alpha L}},$$

giving the final result

$$w_{t-k} = \left[\frac{1-e^{-\alpha}}{1-e^{-\alpha L}}\right] e^{\alpha(t-k-L)}, \quad \text{for all } k \in \{0, 1, \dots, L-1\}.$$

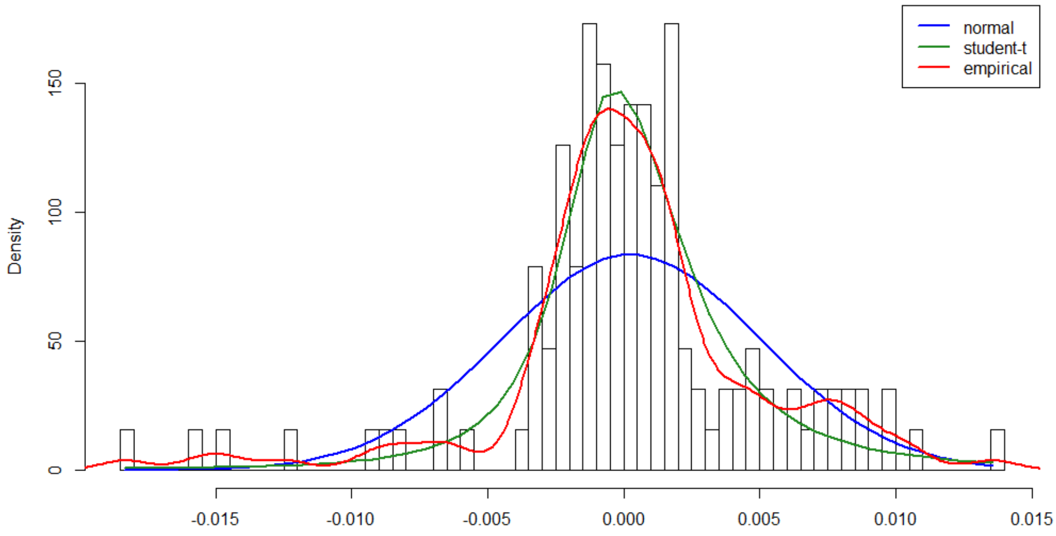

Figure 11: S&P 500 log-returns distribution from 2017-02-28 to 2017-08-29.

Remark 3.10. Note that for each new time step t we have a new set of weights w_{t-k} for all k in $\{1, \dots, L\}$. Since $t > L$ and we work in discrete time, we always have $t - k > 0$ therefore the weighting is always decreasing as k increases which is what we want. Also note that for t, k fixed,

$$\lim_{\alpha \to 0} w_{t-k} = \lim_{\alpha \to 0} w_0 = \frac{1}{L} \qquad \text{and} \qquad \lim_{\alpha \to +\infty} w_{t-k} = \lim_{\alpha \to +\infty} e^{\alpha(t-k)} = 0,$$

which means that for a small α, weights tend to be uniform so equal importance is allocated to every point, recent or old. On the other hand, when α becomes larger, the weights of past events become extremely small.

This method has shifted our question on what size L of window to use towards what decay α to use if we always use the same window length. This is discussed in [3] for the purpose of estimating large covariance matrices. In the Figure 12, we show the results for different values of α on S&P 500 log-returns during the crisis of 2008. For a window length of three years - which is the kind of length we want to have in a mid-to-long-term type of trading strategy - we choose $\alpha = 0.01$ because it gives importance to the financial crisis (so this period will be 'remembered') but it gives slightly more importance to post-crash data which is interesting for direct use when we are standing at the right end of the window. Also Figure 13 supports our choice of a small α. The comparison between the indices with and without weighting is made in Figure 14. Note that they are extremely similar, this tells us that the AR and HSBC methodologies are in fact nearly equivalent in measuring risk appetite. However, as we can see in Figure 14, the HSBC and the AR indices are nearly identical. That could be expected as the AR is rather a small precision over the HSBC index. We will not carry AR index further and keep HSBC weighted exponentially only for

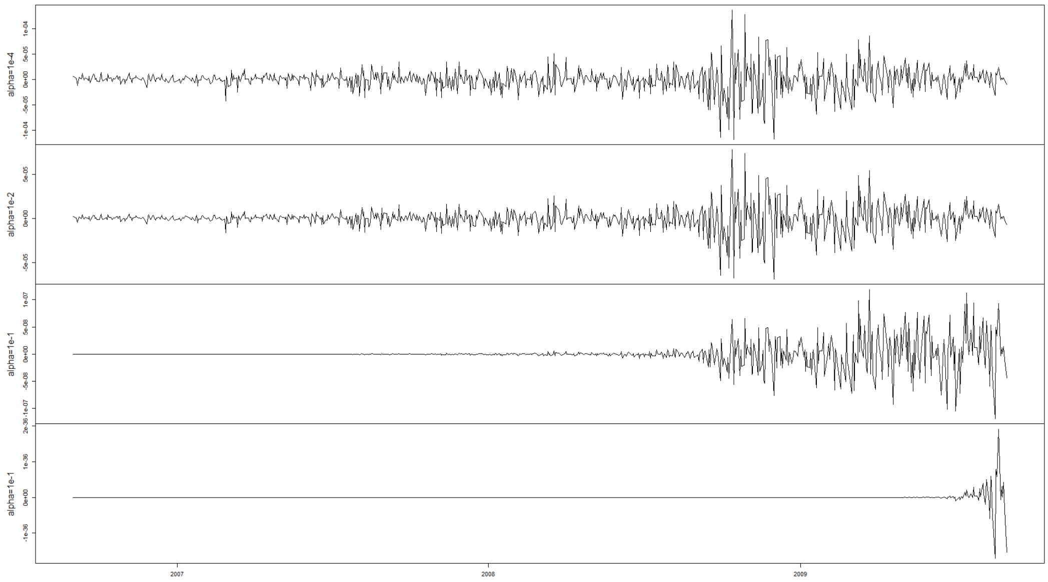

Figure 12: S&P 500 log returns weighted with $\alpha = [0.0001 \quad 0.001 \quad 0.01 \quad 0.1]$ from 2015-08-31 to 2017-08-31.

our tests.

3.4 Global Risk Appetite Indices

We will now see two indices that aim to measure the risk appetite at a global level in the markets, the Global Risk-Appetite Index (GRAI) from Kumar & Persaud [34], used by both JPMorgan and the IMF, a ‘preprocessed’ version of it by Misina [43] (RAI-MI) as well as the risk appetite index from Credit Suisse (CSFB) described in Mielczarski *et al.* [41] and Chen & Poon [9].

3.4.1 Change is risk appetite and rank correlation

In their paper, the authors of [34] do not differentiate risk appetite from risk aversion. However they do differentiate it from the general level of risk in the market. They clearly express a desire a need to move away from using spreads between high and low yield bonds as a proxy for risk appetite, as it was used before. They also consider that measuring changes in risk appetite makes more sense than measuring its level because the shifts in risk appetite are the potential precursors of large market moves (they should at least appear at the same moment). The level itself is not important, what is important is the time at which the risk aversion of investors changes from one state to the other. When there is an increase in risk appetite, investors ask for more risk to bear and the price of the required assets is inflated. But when risk appetite falls, the opposite

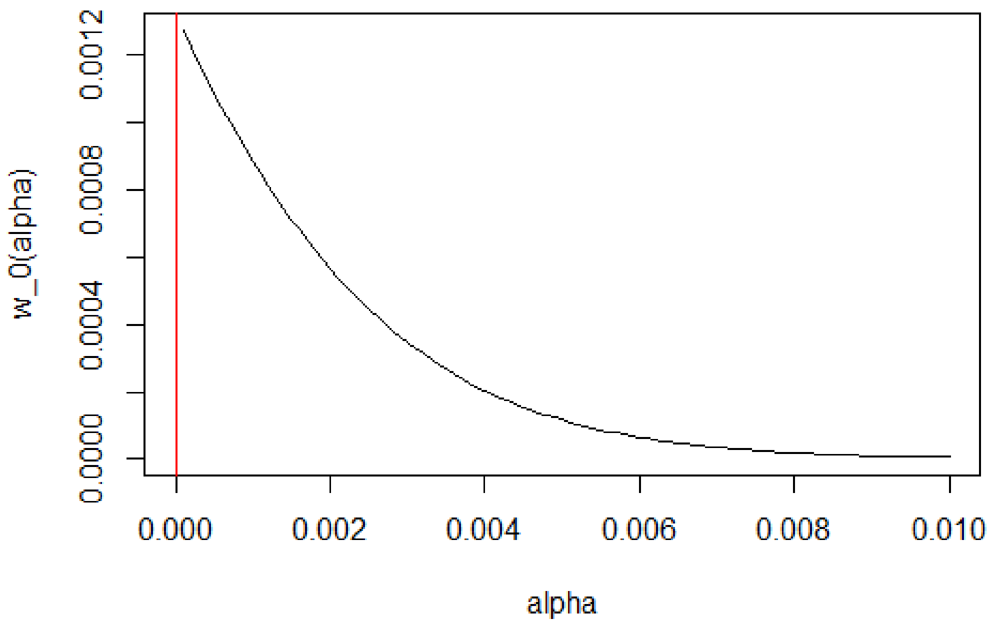

Figure 13: Function $w_0(\alpha)$ for different values of α with a rolling window of 3 years.

scenario happens and prices fall altogether. Note that they mention in their paper the possibility of inferring a risk appetite measure from option implied volatilities. But at the time of the writing of their paper, the derivatives market was far less liquid and broad than today making it difficult or even impossible to estimate a global risk appetite shift measure. We will explore this method in another part.

Anticipating a future need for it, we will define the rank correlation right away. We will use Spearman's rank correlation definition in our case. As he (Spearman) points out in [50], it eliminates the danger of 'accidental error' or outliers that could potentially pollute the estimation of correlation because it disregards the value by only looking at the rank. A value may be very far from the mean of the distribution but its rank will be less distant. Another advantage of this method is that it does not differentiate the types of variables meaning that it enables one to compare populations with different distributions.

Definition 3.11 (Spearman's *rho*)**.** Let X and Y be two random variables. We make no assumption on their distribution. We sample one set of n observations for each random variable, $X_n = \{x_1, x_2, \ldots, x_n\}$ and $Y_n = \{y_1, y_2, \ldots, y_n\}$ respectively. Let $\mathrm{r}(\cdot)$ be the rank operator, therefore $\mathrm{r}(X) = \{\mathrm{r}(x_1), \mathrm{r}(x_2), \ldots, \mathrm{r}(x_n)\}$ and $\mathrm{r}(Y) = \{\mathrm{r}(y_1), \mathrm{r}(y_2), \ldots, \mathrm{r}(y_n)\}$ are permutations of

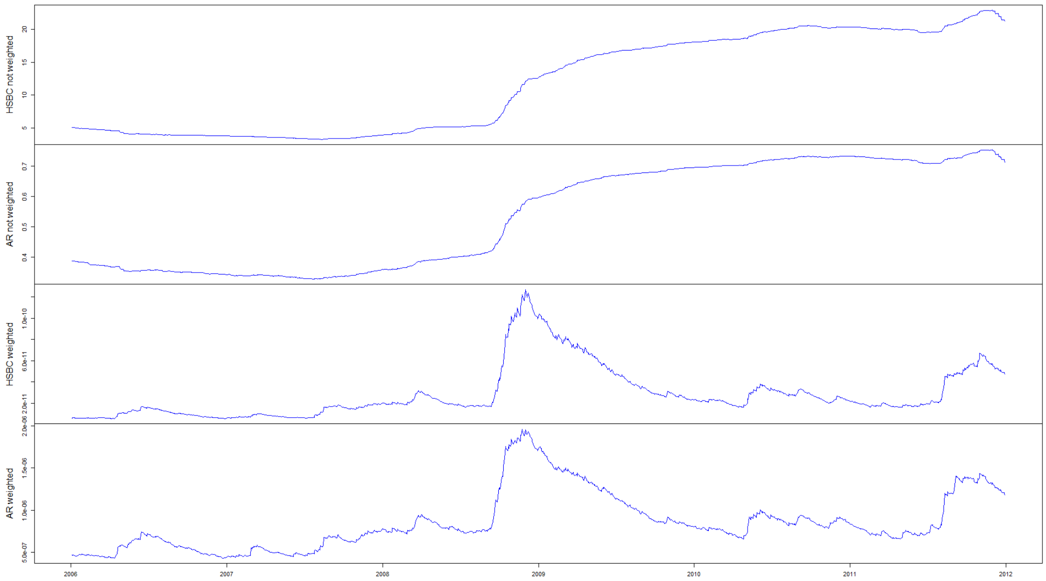

Figure 14: HSBC not weighted, AR not weighted, HSBC weighted, AR weighted from 2006-01-01 to 2011-12-30.

$\{1, 2, \ldots, n\}$. Then Spearman's *rho* is defined as

$$\rho_S(X, Y) = \text{corr}[\text{r}(X), \text{r}(Y)] = \frac{\text{cov}[\text{r}(X), \text{r}(Y)]}{\sqrt{\text{Var}[\text{r}(X)]}\sqrt{\text{Var}[\text{r}(Y)]}}.$$

Remark 3.12. Note that there exist other methods of computing rank correlation between variables, for example the famous Kendall's *tau* τ is mentioned in the appendix A.2.

It is the work done by Misina [42] that gives a real valid mathematical framework to the intuition of Kumar and Persaud [34] by proving two key results and using them as the hypotheses of a statistical test (only mentioned in [34]). The crucial point is what he calls the *separability* of the *observational equivalence* between changes in risk aversion and just changes in risk. This is fundamental since the whole point is to differentiate shocks in risk from shocks in risk aversion by observing something in the data. We need then a theoretical distinction that can be justified and that is what Misina did in his paper ([42] §2.2 Proposition 1 & 2). He breaks this *observational equivalence* and formulates a hypothesis test capable of determining which was the cause of the observation. This argument is also briefly explained in Kumar & Persaud [34] §3 with a heuristic computation supporting the fact that for a given level of risk, a change in risk appetite will have more impact on the change in price than a change in risk. We will then monitor the rank correlation between excess returns and the *past* level of risk (with being careful not to take overlapping periods) for each asset time series.

- For a given level of risk, a shift in risk aversion will have effects on asset returns proportional to their individual level of risk. So if we look at two assets 1 and 2 with levels of risk $\sigma_1 > \sigma_2$, a change in investor risk aversion will impact more the returns of asset A than it will for asset B. We will have the *rank effect* (as said in [42]) $\sigma_1 > \sigma_2 \Rightarrow \Delta r_1 > \Delta r_2$, where Δr_i is the change in the excess returns of asset $i = 1, 2$.

- For a given level of risk aversion, a change in risk will *not* affect the excess returns monotonically.

- So the intuitive thing to test now is whether we have a rank correlation between excess returns r and risk levels σ across all our assets, or not. The test is then

$$[H_0 : \rho_S(r, \sigma) = 0, \qquad H_1 : \rho_S(r, \sigma) \neq 0] \,.$$

Starting from the classical asset pricing relationship (see Cochrane [10])

$$p_t = \mathbb{E}_t[m_{t+1} x_{t+1}], \tag{3.2}$$

where $m_{t+1} = \beta \frac{u'(c_t+1)}{u'(c_t)}$ is the *stochastic* discount factor, c_t the consumption, u the utility function of the representative agent (the investor), β is the *subjective* discount factor and x_{t+1} a random future payoff; and making some assumptions on the utility function structure and the returns distribution (see [42] §3), he is able to prove that in some cases the statistical test is valid and can effectively test for the origin of the changes in excess returns. We refer to his discussion in [42], §3 and §4 but the necessary condition is independence of excess returns. This can be tested for empirically by testing whether the sample correlation matrix is diagonal. Kumar & Persaud do not take that fact into account though. Their index GRAI is computed as follows (see appendix A.4 for the data tickers).

- The excess returns r_t^i for each asset will be the log-difference between the spot rate and the 1-month forward rate 1 month ago. There will essentially be an excess return if the spot has overcome its forward rate.

- We estimate the *inherent* risk by calculating the historical volatility σ_t^i of each asset over a window *ending a day before* the forwards are considered. Doing so, the periods to compute the excess returns and the risk are *not overlapping* so there is no risk of having spurious rank correlation between shifts in risk and shifts in excess returns. Additionally, the window we use for the historical volatility should be long enough to differ from the 'present' volatility (past month in our case) which will have rank correlation with excess returns.

- We are standing at time t, we consider n assets so the excess return of asset i is defined as

$$r_t^i = \ln(S_t^i) - \ln(F_{t-l}^i),$$

where S_t^i is the spot rate of asset i at time t and F_{t-l}^i is the 1-month forward rate 1 month ago of the same asset ($l = 22$ business days, i.e. 1 month). The risk estimation of asset i is computed using a 1-year *non-overlapping window* ($L = 252$ business days, i.e. 1 year from $t-(L+l)$ to $t-(l+1)$) and is given by

$$\hat{\sigma}_t^i = \sqrt{\frac{1}{L-1}\sum_{k=t-(L+l)}^{t-(l+1)}\left(r_k^i - \overline{r}_{t-(l+1)}^i\right)^2}$$

with $\overline{r}_{t-(l+1)}^i = \frac{1}{L}\sum_{k=t-(L+l)}^{t-(l+1)} r_k^i$. Doing this for each one of the n assets, we have n excess returns and n risk estimates at each time t, $r_t = [r_t^1\ r_t^2\ \cdots\ r_t^n]^T$ and $\hat{\sigma}_t = [\hat{\sigma}_t^1\ \hat{\sigma}_t^2\ \cdots\ \hat{\sigma}_t^n]^T$ respectively. The risk appetite index GRAI is therefore the Spearman's rank correlation

$$\rho_t = \rho_S(r_t, \hat{\sigma}_t).$$

The resulting index is shown in Figure 15. We mentionned earlier the necessity of independent returns for the hypothesis test of Misina [42] to be valid. Misina comes back to it in [43] in which he constructs his version of Kumar & Persaud's GRAI, the RAI-MI. As it is not realistic to expect finding independent assets, his idea is to transform their returns before applying the method of rank correlation. This is explained in [43]. He essentially does a change of basis, or as we have seen before a PCA, on the standardised returns time series (*z-scores*). The principal components are then orthogonal vectors in the new space and ready to be used, we just do the above procedure with these 'pre-processed' returns. The result is also shown in Figure 15 for comparison. We use the same data set as for the GRAI index. The justification for such a method can be found in his other paper [44]. In the latter, he develops his definition of *implied* risk aversion which is the Arrow-Pratt risk aversion coefficient we saw before, plus a term he calls the *equivalent variation* and represents the variantion of the risk aversion with respect to the change of investors' *probabilities*. Indeed, in order to keep the utility and risk aversion constant, he allows the probabilities associated with states of the market (outcomes of a lottery if we want) to change over time. However, this index is more noisy than GRAI which is pointed out by the author. He presents a smoothed version of it in his paper [43] but we feel it defies the purpose. So despite our previous efforts to reduce the noise (with market indices in Section 3.2), we keep this one as it is, we do not want to smooth out potential information. He does mention that it gives the same information as Gai & Vause's risk appetite (see [23]), which is our next topic.

3.4.2 Cross-sectional Regression

The CSFB Risk-Appetite from Mielczarski *et al.* [41] is similar to the GRAI index described above; the data can also be found in the appendix A.4. Since [41] is not very precise on their exact method, we will follow the steps detailed in Chen & Poon [9] in which the authors change slightly the data set to tune it to their specific problem which is detecting the Asian crisis of October 1997 and the Russian crisis of September 1998. One essential point is that the authors of the Credit Suisse paper [41] use weighted data, probably by volume or capital. Contrary to this idea, we wish to weight equally every time series as to capture more of the movements from the developing markets. Also, the size of the markets has well evolved and is now very different from the ones at the time this paper [41] was written. In particular, we expect the Asian markets to have more importance relatively to other ones nowadays than they did in 1997. Nevertheless, the underlying idea is the same and is illustrated by charts 3 & 4 of [41]. If we plot for each asset, their excess returns versus their volatility we can clearly see groups between bonds and stocks. The returns of the bonds are low and positive while the returns of the stocks are large and negative while in a market with low risk appetite. The opposite situation is also true. Therefore, the measure of risk appetite will be the slope of the regression line of the regression of the excess returns by the volatilities. The method below follows what is done in [9].

- We are at time t, we consider n assets in our data base and a rolling window of length $L = 252$ business days or 1 year. The excess return of the asset i at time t is defined as

$$r_t^i = \ln\left(\frac{S_t^i}{S_{t-L}^i}\right) - r_t^f,$$

  where S^i is the price of asset i and r_t^f is the risk free rate (we use the 10-year US Treasury as a proxy).

- The risk estimation for asset i at time t is defined by the *non-overlapping* 2-year estimate

$$\hat{\sigma}_t^i = \sqrt{\frac{1}{2L-1}\sum_{k=t-(3L)}^{t-(L+1)}\left(r_k^i - \overline{r}_{t-(L+1)}^i\right)^2},$$

  where $\overline{r}_{t-(L+1)}^i = \frac{1}{2L}\sum_{k=t-3L}^{t-(L+1)} r_k^i$.

- Then we regress the excess returns on the volatilities as

$$r_t^i = \alpha_t + \beta_t \hat{\sigma}_t^i, \text{ for } i = 1, 2, \ldots, n.$$

  We take β_t to be our risk appetite measure at time step t.

Remark 3.13. Note that we need $4L$ (four years) of data before being able to compute the first risk appetite level. Indeed, we need the returns from t back to $t-L$ for the excess returns. Then we also need the volatility over the period $t-(L+1)$ back to $t-3L$. To compute this volatility we need excess returns $r_{t-(L+1)}$ back to r_{t-3L} and the oldest excess return is $r_{t-3L} = \ln\left(\frac{S_{t-3L}}{S_{t-4L}}\right) - r_{t-3L}^f$. So the oldest data point needed is from time $t-4L$. Note also that in this part we have used 1-year rolling windows because it is more coherent with the data. Indeed, we could have used 6-months futures and 3-years windows to stay consistent in our window size across this research. However the and 1-month future prices are more liquid than 6-month ones.

We end up having the following three indices that are presented in Figure 15. We can see reasonable correlation between GRAI and CSFB. RAI-MI does not exhibit any particularities except for its noisiness.

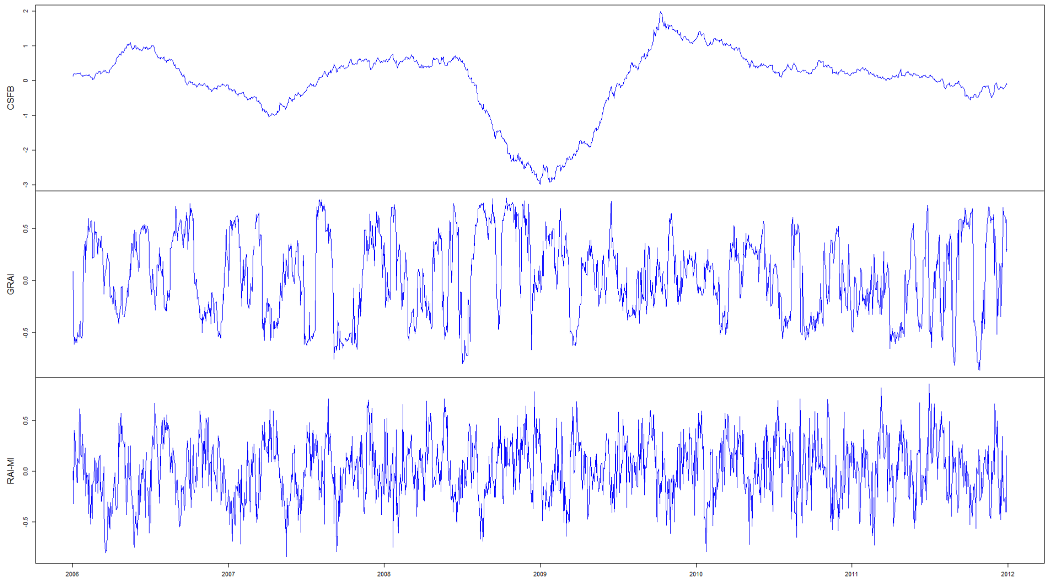

Figure 15: CSFB, GRAI and RAI-MI from 2006-01-01 to 2011-12-30.

3.5 Risk Appetite from option prices

The distinction between risk appetite and risk aversion will be again discussed in this part as we interest ourselves to a more advanced technique. Briere & Chancari [8] and Karampatos, Tarashev, Tsatsaronis [30] talk about risk perception and attitude towards risk while Jackwerth [28] and Ait-Sahalia & Lo [2] talk about implied risk aversion and Gai & Vause [23] talk about risk appetite. Even if these authors use different terms, their ideas are similar. It relies on the need to go further than estimating risk appetite through market indices that rely only on statistical properties of aggregates of market data, which could be limited in the sense that they can't capture investors sentiment but only their behaviour. It might be enough to have a very good indicator relying only on reaction of groups of investors (for example the VIX is a very simple indicator of such type and is consider in our market indices) and this is the incentive behind the market indices. But if we seek to capture investors' sentiment and the importance they give to their sentiment, we have to extract it using a more advanced technique. And the idea is that options prices contain this information. As investors price Call and Put options, they will reveal their preferences: the price given to these options by the market reflects its attitude towards risk because option prices are basically insurances on future states of the market. And the way to price them (as described for example in Breeden & Litzenberger [7] or Cochrane [10] or Follmer & Schied [19]) is by computing their fair value as an expectation of discounted future cash-flows or payoffs, under a certain measure, the *risk-neutral* one. In his paper [28], Jackwerth gives the motives for such a study. Estimating the risk aversion or the *risk-neutral* densities is a complex problem but before and after the 1987 crash, the difference between both densities (*risk-neutral* and *historical*) became larger and the simple utility functions used to compute the Arrow-Pratt coefficient were not sufficient any more.

Let us pause a moment on this concept of *risk-neutral* measure as it is at the very core of modern asset-pricing and therefore at the core of our method to derive the risk appetite measure. This can be traced back to Arrow-Debreu prices of elementary securities in [4] and [13]. The prices are determined by a market, which means they contain information from both buyers and sellers as they agree on what is the fair price for such a security. They also take into account the preference of investors and vary according to the level of wealth. The *risk-neutral* density is linked with the one of an *average investor* (or representative agent) with an unknown utility function, as introduced by Constantinides [11]. In the words of Jackwerth ([28]),

> "The risk-neutral probability is the price that an 'average investor' would pay for receiving one dollar in the state multiplied by the risk-free return. The subjective probability is simply the assessment of the 'average investor' of how likely a state will occur. These two probabilities would be identical if the investor was indifferent to risk. However, the investor might value a dollar more in certain states, namely ones where wealth is low. The risk aversion indicates these preferences of the investor."

Or in the words of Follmer & Schied ([19] Remark 1.5)

> "[The risk-neutral pricing formula] can be seen as a classical valuation formula which does not take into account any risk aversion, in contrast to valuations in therms of expected utility[...]."

In their book [19] §3, they also define this measure in the framework of optimal portfolio construction. If we consider an investor with preferences represented by an expected utility function (remember Definition 2.4), the optimal portfolio maximises its expected utility of the future cash-flows. This refers to section 2.1 on preferences and expected utility and is another way of working out the proof of the Fundamental Asset Pricing Theorem (FTAP) without assuming a complete market but assuming the existence of this optimal portfolio. Both [19] Proposition 3.9 and Corollary 3.12 help us define the *risk-neutral* measure with respect to the utility. And this is very useful for us since we can then link risk aversion and *risk-neutral* density as it is done by most of the papers we are considering ([28],[2],[30],[8]). For example we can find in Jackwerth's paper [28] the following optimisation problem.

Consider a representative investor in a complete market (see Constantinides [11]) with initial wealth $W_0 = 1$ investing all his wealth from time $t = 0$ to time $t = T$. He wants to maximise his expected utility U over all possible terminal values W_T (future wealth). We then want to find

$$\max \int_{\mathbb{R}} q(W_T)U(W_T)dW_T, \tag{3.3}$$

with the constraint that he invests all his initial wealth

$$\frac{1}{(1+r_f)^T}\int_{\mathbb{R}} p(W_T)W_T dW_T = 1,$$

where q is the *subjective* density (or *historical* as we call it), p is the *risk-neutral* density, r_f is the *risk-free rate*. Using the Lagrange multiplier method (see appendix A.3) we find the equilibrium. We know that in this state, the average investor must hold the market portfolio, portfolio that represents the aggregate claim in the market. We invite the reader to refer to Follmer & Schied §3.4 p.141-156 for more precisions as the notations in [8] and [28] can be confusing.

If the market portfolio has return k then the solution reads

$$U'(k) = \frac{\lambda}{(1+r_f)^T}\frac{p(k)}{q(k)}$$

where λ is the Lagrange multiplier. It is then easy to find U'' and we have our (implied) risk aversion coefficient

$$\rho_{abs} := \rho_{abs}(k) = -\frac{U''(k)}{U'(k)} = \frac{q'(k)}{q(k)} - \frac{p'(k)}{p(k)}. \tag{3.4}$$

Note that this was initially found in Leland [37] §4. In practice, the *risk-neutral density* gives greater weight to undesirable outcomes typically negative returns, it is *left-skewed.* It is very useful to compare it with the *historical* distribution of realised performance, which is typically less skewed and more peaked, i.e. has smaller standard deviation and a higher maximum.

We skip for the time being the details of how to estimate those probabilities as we will make different choices from the authors in general. Note that Equation (3.4) involves derivatives of the densities. This will potentially lead to numerical errors. Based on the idea of [30], we estimate the implied risk aversion by what they call a *ratio of downside risk* which is a comparison of the left tails of both densities by a ratio

$$R_{AV} = \left(\frac{\mathcal{A}_p^l}{\mathcal{A}_p^l + \mathcal{A}_p^r}\right) \Big/ \left(\frac{\mathcal{A}_q^l}{\mathcal{A}_q^l + \mathcal{A}_q^r}\right) \tag{3.5}$$

where $\mathcal{A}_p^l$ is the area under the density p (*risk-neutral*) but only for the left tail, $\mathcal{A}_p^r$ is the area under the density p (*risk-neutral*) for the rest of the values k and same for the *historical* distribution q. The way we chose the limits of these areas is the same as described in [8]. We have three points a, b and c to determine, where $a < b < c$ and for example the left-tail area will be the area $\mathcal{A}_p^l = \mathcal{A}_p^{a,b}$ which is the area under the *risk-neutral* density between $k = a$ and $k = b$ and $\mathcal{A}_p^r = \mathcal{A}_p^{b,c}$ will be the area under the *risk-neutral* density between $k = b$ and $k = c$. Same idea for the *historical* density. We take then b to be the *at-the-money* strike (or forward log-moneyness as we will work in that space for the SVI model) and $a = b - \delta\sqrt{\mathrm{Var}(q)}, c = b + \delta\sqrt{\mathrm{Var}(q)}$, where δ adjusts the width (we will take $\delta = 0.35$).

If both densities are different from each other then it means that the *historical* had missed information that the *risk-neutral* captured and the ratio will move away from one. Apart from [23] where the authors make the distinction between appetite for risk and aversion for it, the other papers clearly aim to compute the *implied* risk aversion from the start. For a more rigorous definition of this, one can read Misina [44].

Note that the ratio in Equation (3.5) focuses on the left part of the distribution. This is criticised by Gai & Vause in [23] §3.1. Their idea is that the whole density must be studied and not give such importance to the left tail. They base their derivation of the risk appetite measure on Cochrane's work [10] which is the counterpart of the asset pricing with utility function maximisation. In this framework, in an efficient market, the prices of assets at time t are the conditional expectations of the future payoffs x_{t+1} discounted with the *stochastic* discount factor m_{t+1}, i.e.

$$p_t = \mathbb{E}_t[m_{t+1}x_{t+1}] = \mathbb{E}_t\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}x_{t+1}\right]. \tag{3.6}$$

This is the link between both approaches, Equations (3.22) and (3.2). Gai & Vause come up with something that is close to (3.5) but depends on the whole distribution. They define the risk appetite as the *willingness to bear risk* and in [23]§2 the come up with the definition

$$R_{AP}(k) = \frac{1}{r_f}\mathrm{Var}\left(\frac{p(k)}{q(k)}\right), \quad \forall k \in \mathbb{R}. \tag{3.7}$$

We will eventually do both methods and compare them to see they will be very close.

We are left with the problem of estimating the two probabilities, the *risk-neutral* one as well as the *historical* one. This can become intensive computationally in practice. Ait-Sahalia and Lo [2] use a non-parametric approach based on kernel density estimators (including the Nadaraya-Watson for the *risk-neutral* one). Authors of [30], [8] and [23] estimate the *risk-neutral* density by fitting cubic and quartic polynomials to their implied volatility points but as pointed out specifically in [8] this can lead to negative probabilities in the tails because of the simplicity of the smooth curve (cubic polynomial) they fit and its potentially arbitrageable tail behaviour. They simply discard negative points when there happen to be one. It is also mentioned in [28], deep-in-the money Put prices tend to have an over-priced implied volatility and so discarding some of the data might be necessary sometimes even if it is not ideal given the importance of the tails. We will be using the SVI model which also deals with that problem by respecting the Lee moment formula for wing asymptotic slopes (see [35]). For the *historical* density, we will use a kernel density estimator with Gaussian kernel instead of the various GARCH models used in [30], [8] and [23] which rely on fitting and simulating.

Remark 3.14. As these methods rely on option prices, it limits the scope of markets on which we can actually compute the measure. On the other hand, unlike many of the authors, we have directly access to implied volatility data from data provider Bloomberg so we do not need the market to be particularly liquid. Nevertheless, after discussion with Bloomberg Service Teams, it is unclear how they actually extract the implied volatility data when having too few quoted strikes. Our guess is that they use some sort of local smoother but since we were not sure, we tried to stick with options on underlying known for being highly traded (or liquid options), typically S&P 500. We stayed on the stock market for this risk appetite index, stressing the limitations of this technique. We used 3 month-implied volatility data (i.e. extracted from option prices with a maturity of 3 months) as the corresponding futures on indices are usually 3 month-futures. We will see that the price of the future is involved in the SVI model to pass into the log-forward-moneyness space. Data is described in A.4. Authors of [8] explain also their method for the Foreign Exchange (FX) market, based on fitting a smooth curve on implied volatility data but in the delta space (delta being the sensitivity of the option price to variations in the strike) as it is a standard way to keep track of prices for FX traders. Moreover, at the time of the publications of these papers, derivatives markets were less developed than today. We can therefore expect more participants and more liquidity hence more reliable prices today than before (depending on the degree of confidence we have in today's markets, but this is another topic entirely).

3.5.1 Risk-Neutral Density (RND) extraction and SVI model

Our *risk-neutral* density extraction will be trough a parametric model called the *Stochastic Volatility Inspired* or SVI model. This model was first presented by Merrill Lynch in 1999 and first made public in [21]. It is a popular choice of volatility smile because it can model linear wing asymptotic behaviour consistent with boundaries on the slope set by Roger Lee in [35]. Also it is a parametric model with 5 parameters so it is relatively fast to fit implied volatility data points. In their paper [22], the authors also show that this model is free of butterfly arbitrage (static arbitrage) and also free of calendar arbitrage. That is, the volatility smile is strictly increasing in the maturity. Therefore, the interpolation between the slices (one volatility smile for one maturity is called a slice, as a slice of the implied volatility surface) can be done so as to build an arbitrage-free implied volatility surface. We will not interest ourselves in the latter form of arbitrage however, as we only need one slice per time step. Indeed, our goal is not to build an arbitrage-free surface but merely to extract the *risk-neutral density*. We will then not use the model to its full potential. We chose this model for the simplicity of its formulation and implementation as well as its behaviour for extreme strikes yielding 'safe' risk-neutral densities - and we now know this is of most importance for our problem.

We will use the same notation as in [22]. We consider that $(S_t)_{t\geq 0}$ describes our price process on the natural filtration $(\mathcal{F}_t)_{t\geq 0}$. Hence the forward price can be written as $F_t = \mathbb{E}_0[S_t] = \mathbb{E}[S_t|\mathcal{F}_0]$. Let $k \in \mathbb{R}$ be the log-moneyness, $e^k = \ln\left(\frac{K}{F_t}\right)$ and $t > 0$. The Black-Scholes *implied volatility* is $\sigma_{BS}(k,t)$ and is the volatility you have to plug into the Black-Scholes price of a European Call option with underlying S, strike $K = F_t e^k$ and maturity t to get the market price again, i.e. $C_{BS}\left(k, \sigma^2_{BS}(k,t)t\right) = C_{mkt}$. The *total implied variance* is

$$w(k,t) = \sigma^2_{BS}(k,t)t \tag{3.8}$$

and the *implied variance* is just $v(k,t) = \sigma^2_{BS}(k,t) = w(k,t)/t$. The map $k \mapsto w(k,t)$ is a *slice* and $(k,t) \mapsto w(k,t)$ is the *volatility surface*. Definition 2.3 of [22] states that a slice is free of butterfly arbitrage if its corresponding density is always non-negative. It then gives the function $g : \mathbb{R} \to \mathbb{R}$ that is needed to characterise the no-butterfly-arbitrage condition. For t fixed,

$$g(k) := \left(1 - \frac{kw'(k)}{2w(k)}\right)^2 - \frac{w'(k)^2}{4}\left(\frac{1}{w(k)} + \frac{1}{4}\right) + \frac{w''(k)}{2}. \tag{3.9}$$

Then Lemma 2.2 of [22] gives us all the elements we need, that are the slice is free of butterfly arbitrage if and only if g is non-negative for all k and $\lim_{k\to+\infty} d_+(k) = -\infty$, where for any $k \in \mathbb{R}$,

$$C_{BS}(k, w(k)) = S(\mathcal{N}(d_+(k)) - e^k\mathcal{N}(d_-(k))), \quad \text{and} \quad d_\pm(k) = -\frac{k}{\sqrt{w(k)}} \pm \frac{\sqrt{w(k)}}{2}.$$

Then from Breeden & Litzenberger [7] we know that the density will be the second derivative of the price with respect to the strike K so the density is

$$p(k) = \frac{\partial^2 C_{BS}(w(k))}{\partial k^2} \cdot \frac{\partial^2 k}{\partial^2 K} = \frac{g(k)}{\sqrt{2\pi w(k)}} \exp\left\{-\frac{d_-^2(k)}{2}\right\}. \tag{3.10}$$

In practice we will have to re-scale this estimate of p to make it integrate to one, being very careful with the tails data. As we move away from the zero-moneyness (deep in-the-money or out-of-the-money options) strikes become less quoted by actors in the market, consequently prices and then implied volatilities are less frequent. We have to be aware of the potential need for extrapolation of the model in the wings as to avoid bad re-scaling.

The SVI model is a parametrisation of the total implied variance $w(k)$. We will use three different parametrisations and continue to follow the exact same notation as in [22].

Definition 3.15 (SVI model Raw parametrisation, [22])**.** The *raw* parametrisation is defined using five parameters $\chi_R = \{a, b, \rho, m, \sigma\}$ and reads

$$w(k;\chi_R) := a + b\left(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right),$$

with $a, m \in \mathbb{R}$, $b \geq 0$, $|\rho| \leq 1$, $\sigma > 0$ and $a + b\sigma\sqrt{1-\rho^2} \geq 0$ to ensure that $w(k;\chi_R) \geq 0$ for all $k \in \mathbb{R}$.

The parameters allows one to calibrate this curve to data points using an optimisation routine. However they do not carry any financial meaning. The Jump-Wings parametrisation was then created to make it more readable or an investor.

Definition 3.16 (SVI model Jump-Wings parametrisation, [22])**.** The JW parametrisation is also defined using five parameters $\chi_J = \{v_t, \psi_t, p_t, c_t, \tilde{v}_t\}$ with v_t the ATM variance, ψ_t the ATM skew, p_t and c_t the put and call wings slope respectively and $\tilde{v}_t$ the minimum implied variance ($v_t = w_t/t$). The parameters are defined from χ_R as

$$\begin{aligned}
v_t &= \tfrac{1}{t}\left(a + b(-\rho m + \sqrt{m^2+\sigma^2})\right),\\
\psi_t &= \tfrac{1}{\sqrt{w_t}}\tfrac{b}{2}\left(-\tfrac{m}{\sqrt{m^2+\sigma^2}} + \rho\right),\\
p_t &= \tfrac{1}{\sqrt{w_t}} b(1-\rho),\\
c_t &= \tfrac{1}{\sqrt{w_t}} b(1+\rho),\\
\tilde{v}_t &= \tfrac{1}{t}\left(a + b\sigma\sqrt{1-\rho^2}\right).
\end{aligned}$$

There is also a correspondence in the other way (from χ_J to χ_R) in [22]§3.3.

The interesting part of [22] relates to a parametrisation that is directly free of static arbitrage, the SSVI parametrisation. We will not discuss it here nut suppose we chose to start with the JW parametrisation, we are able to reduce the numbers of parameters we need down to three. As described in [22]§5, if we choose v_t, ψ_t, p_t, then there is a 'clever' way to pick parameters $c_t, \tilde{v}_t$ to avoid any butterfly arbitrage,

$$c_t' = p_t + 2\psi_t, \qquad \text{and} \qquad \tilde{v}_t' = v_t \frac{4p_t c_t'}{(p_t + c_t')^2}. \tag{3.11}$$

In practice, before we consider re-picking these parameters to eliminate for sure static arbitrage, we need to calibrate the model. For a fixed maturity t, the implied volatility points we get from the market (Bloomberg data) are the implied volatilities corresponding to the mid options prices. They are noted $\sigma_i = \sigma_{BS}^{mid}(k_i, t)$ for $i = 1, 2, \dots, n$ if we have n strikes quoted. In our case, we will have 9 points corresponding to 120%, 110%, 105%, 102.5%, 100% (at-the-money), 97.5%, 95%, 90% and 80% moneyness. To calibrate the model we will solve the following problem by numerical optimisation.

In practice, we use a numerical optimiser and we formulate the problem

$$\underset{\chi_J=\{v_t,\psi_t,p_t,c_t,\tilde{v}_t\}}{arg\min} \quad \left[\frac{1}{2n}\sum_{i=1}^{n}\left(w(k_i;\chi_J) - \sigma_i^2 t\right)^2\right]. \tag{3.12}$$

However, we need a starting point for this problem. This leads us to consider what is done in Zeliade's white paper [53]. As stated by the author, even though calibration of the SVI model to options on an index (or any liquid asset really) is remarkable, there is a possibility of existence of multiple local minima when using a least square optimiser. This can lead to the program finding a set of parameters that minimise the objective function but that are clearly not fitting the smile correctly. This sensitivity to the starting point is a flaw we want to deal with in our case because we will be calibrating this model daily on several years of trading days. It would make no sense to suddenly have a straight line or a concave curve in the middle of our set of smiles. To prevent that from happening, we follow the steps described in [53] as the author points out it produces a 'trustworthy and stable' calibration. We refer to [53]§2-§3 for more details but eventually, the model is reduced to a set of two parameters (σ and m) while the other three are determined explicitly without the need for a calibration procedure. The author derives bounds for the parameters in [53]§2 and then operates a change of variable $s = \frac{k-m}{\sigma}$ that transforms the problem into a linear one.

Definition 3.17 (SVI model Quasi-Explicit parametrisation, [53])**.** The total implied variance is

$$w(s; \chi_{QE}) = a + ds + c\sqrt{s^2 + 1},$$

with $d = \rho b\sigma$ and $c = b\sigma$. So $\chi_{QE} = \{\sigma, m;\ a, c, d\}$ is split.

Then, given values of the pair (m_0, σ_0), we solve an optimisation problem on a convex set

$$\begin{aligned} & \underset{(a,c,d)\in D}{arg\min} && \left[\frac{1}{2n}\sum_{i=1}^{n}\left(w(s_i; \{m_0, \sigma_0; a, c, d\}) - \sigma_i^2 t\right)^2\right] \\ & && \textbf{s.t.} \quad D = \begin{cases} 0 \leq c \leq 4\sigma, \\ |d| \leq c \text{ and } |d| \leq 4\sigma - c, \\ 0 \leq a \leq \max_i\{\sigma_i^2 t\}, \end{cases} \end{aligned} \tag{3.13}$$

where D is the set of bounds determined before in [53]§2. D is convex and constraints are linear so there is a unique solution to(3.13). So we can find the optimal set of parameters (a^*, c^*, d^*) and then find (a^*, b^*, ρ^*). Then it's just a matter of solving

$$\underset{m\in\mathbb{R},\sigma>0}{arg\min} \quad \left[\frac{1}{2n}\sum_{i=1}^{n}\left(w(s_i; \{m, \sigma\ a^*, b^*, \rho^*\}) - \sigma_i^2 t\right)^2\right]. \tag{3.14}$$

Once we have solved the optimisation problems (3.13) and (3.14), it is easy to retrieve the density using both equations (3.10) and (3.9). For example, doing this for every day from 2nd of January 2015 until 8 of August 2017, we get the following surfaces presented in Figures 16 and 17. We can already see the skewness of the *risk-neutral* densities. Note well that these are not the usual implied volatility surface and *risk-neutral* density surface. The latter are fitted on one day for different maturities. We here present fitted volatilities and densities on different days for one maturity.

3.5.2 Kernel Density Estimation (KDE) of the Historical Density

Suppose we want to estimate a function f of some random variable X over a domain, say $\mathbb{R}^n$ using only realisations of such transformed random variable. Kernel smoothing is a *non-parametric* and *local* method that can be used to do so. As we will only see the case of kernel density estimation in one dimension, let's only focus on explaining this idea as the applications of kernel smoothing techniques are numerous, see Wand and Jones [29] or more recently [20]§6. So we have

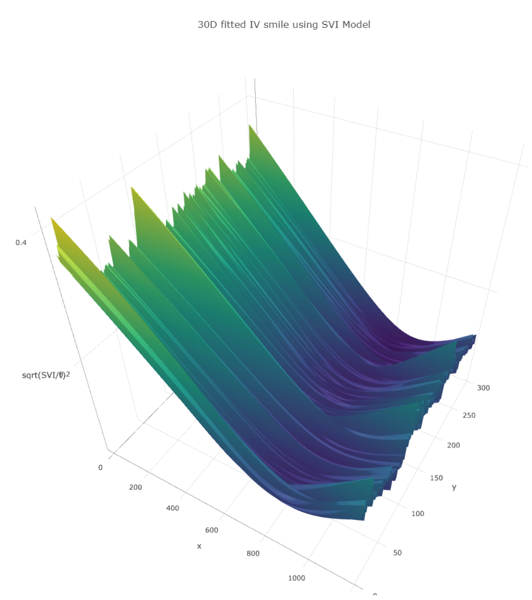

Figure 16: Daily SVI-fitted 3M IV.

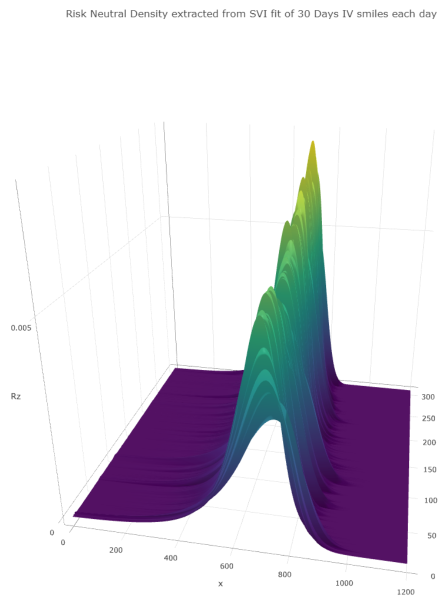

Figure 17: Daily SVI-fitted 3M RND.

a random variable X that has values in $\mathbb{R}$ and an unknown density function f. Using observations $\{f(x_1), f(x_2), \ldots, f(x_n)\}$, we want to compute a local estimate $\hat{f}(x_0)$ at any point $x_0 \in \mathbb{R} \cap [x_1, x_n]$ using the observations in its neighbourhood. And finally, we want to do this in such a way that the function $\hat{f}$ is smooth. The kernels are functions K_h that are used to select the data around the point at which we are estimating $\hat{f}$ with h being the length on the *local window*, also called the *width* of the kernel or the *smoothing parameter*. It is the only feature that will have to be learned. Essentially, $\hat{f}(x)$ will be a weighted average using an appropriate localisation function or kernel.

Definition 3.18 (Univariate kernel density estimator, [20], [29])**.** The kernel density estimator at point x_0 with kernel function K_h is the average

$$\hat{f}(x_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(x_0, x_i) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_0 - x_i}{h}\right), \tag{3.15}$$

where $h > 0$ and the kernel K is a Borel measurable and non-negative function such that $\int K = 1$.

It can be shown (see Devroye & Gyorfi [16] §3 Theorem 1) that this estimator is consistent as it converges in probability, almost surely and exponentially in L^1 as $h \to 0$. The choice of the kernel function dictates the number of data points selected for the estimation and their weight in the average so it should be done carefully. It is intuitive that kernels should be symmetric, unimodal and as smooth as possible. Symmetric kernels are preferred as they will select as much data on both sides of the estimation point x_0 and smooth kernels are chosen because of the local smoothness they provide.

Example 3.1. *Popular choices of kernels include for example*

- *the standard normal kernel is the Gaussian density* $\mathcal{N}(0, h^2)$ *with non-compact support*

$$K_h(x_0, x) = \frac{1}{\sqrt{2\pi}h} \exp\left\{ -\frac{(x-x_0)^2}{2h^2} \right\},$$

- *the Epanechnikov quadratic kernel (used as a reference for historical reasons) with compact support*

$$K_h(x_0, x) = D\left(\frac{|x-x_0|}{h}\right), \quad D(s) = \frac{3}{4}(1-s^2)\mathbf{1}_{\{|s|\le 1\}},$$

- *the tri-cube kernel with compact support too*

$$K_h(x_0, x) = D\left(\frac{|x-x_0|}{h}\right), \quad D(s) = \frac{70}{81}(1-|s|^3)^3\mathbf{1}_{\{|s|\le 1\}}.$$

An example is shown in Figure 18.

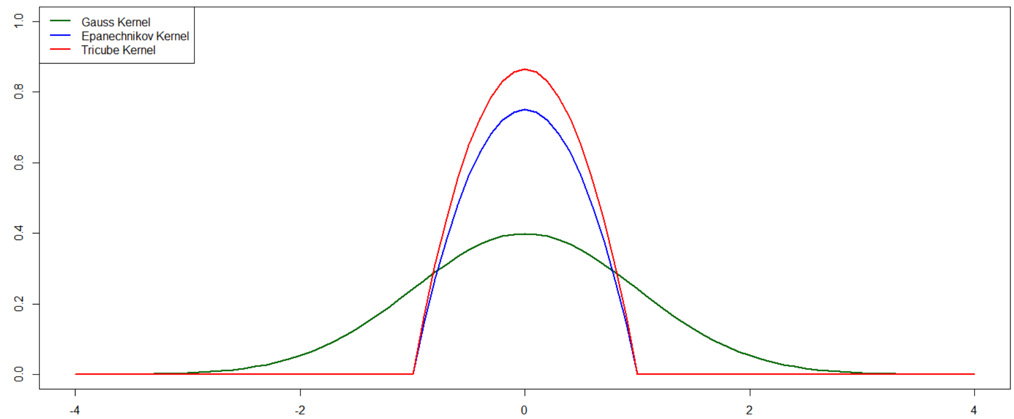

Figure 18: Example of Gauss, Epanechnikov and Tricube kernels with bandwith $h \equiv 1$.

Remark 3.19. Remember that kernels are only used to select data used in the average (3.15). The fact that we chose a kernel to be a density doesn't imply any assumptions about the random variable we are estimating the density of. The bell curves or parabolic or quartic curves are used so that the weights of the points we used in (3.15) decrease with respect to the distance from x. This has for effect, as we move x on the interval $[x_0, x_n]$, to make the neighbors enter and exit the selection window smoothly. The result is a smooth local average of the density function.

We can they ask ourselves if there is a better kernel shape than another for example. Wand and Jones [29] discuss this in §2.7 where they describe canonical kernels (also see Marron & Nolan [39]). The optimal kernel function was first described and proved to be optimal under the constraints that

$$\int K(x)dx = 1, \qquad \int xK(x)dx = 0, \qquad \int x^2K(x)dx < \infty$$

by the mathematician Epanechnikov ([17]). There are also issues with estimating points x_0 that are near the boundaries of the interval $[x_1, x_n]$ because a symmetric kernel will select data asymmetrically in that region. As it will become of first importance for us later - as we need good approximation on the tails where we have less observations - we will need to overcome boundary error problems. A way do deal with this is to do a *local linear regression* as described in [20]§6.1. The idea is to fit an straight line on the boundaries to avoid relying on the kernel that essentially fits a constant (horizontal line) weighted by the kernel function (see [20] Figures 6.3-5). The degree of this polynomial is one for the linear local regression but this can be extended to higher degrees (usually odd degrees work better). This is actually a very powerful idea and an extension of the kernel smoothing as we have seen so far to a tool capable of estimating not only densities but a large class of functions. The *local linear regression* becomes then the *local polynomial regression*.

This lets us with the task of selecting the *smoothing parameter h*. There is an interesting discussion about the error of estimation for univariate kernel density in Wand & Jones [29] §2.3-2.6 and how the errors depend on h. It is important as a small width will produce a noisy density estimate while a large width will produce a density estimate that is too smooth. So there is a so-called *bias-variance trade-off* This parameter control the standard deviation in the normal standard kernel for example, or the support size in the Epanechnikov and tri-cube kernels. This is covered in Wand and Jones [29] as well as in Silverman [49]. In practice, we will use the `R` package `ks` and more precisely the function `kde` which will automatically select eh bandwith using the *plug-in selection* method proposed by Wand & Jones §3.6 which is an automatic *data-driven* method. The kernel density estimator is then an *non-parametric* estimator.

In the end, to visualise it in the case of the S&P 500 from 2015-01-02 to 2017-08-20, we prefer to plot it in 2D to see the full extent of the difference between the both densities (*risk-neutral* in orange and *historical* in blue). We observe an impressive difference between the two for that period in Figure 19. Figure 20 represents the are used in the ratio seen before in Equation (3.5).

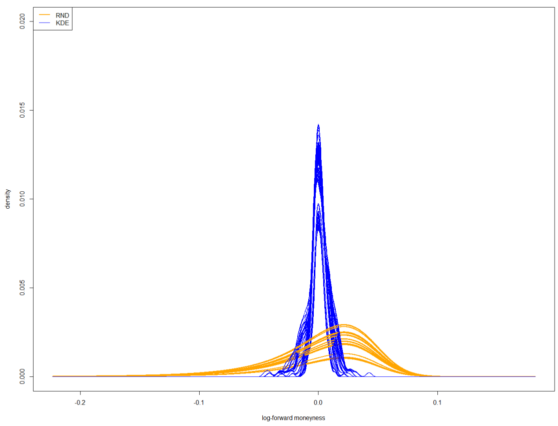

Figure 19: Daily SVI-fitted 3M RND and corresponding returns KDE.

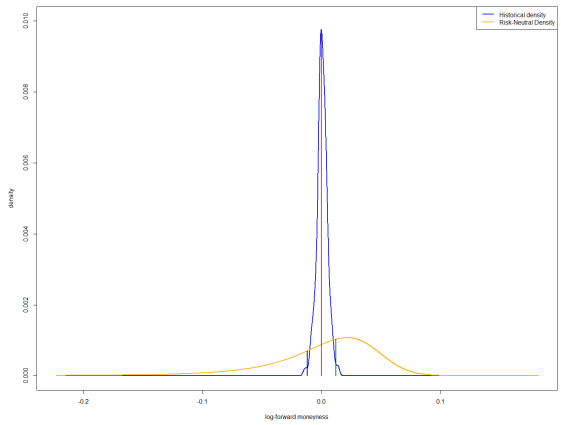

Figure 20: One example of the ratio (3.5) on 2015-01-02.

4 Implementation of trading strategies and back-testing

In this part, we are going to explain how we implemented trading strategies based on our risk appetite indices and present our main results. We use the term risk appetite carefully now as we have seen some of these indices actually measure risk aversion and others do not clearly distinguish both concepts. Our method is based on the time series measuring the risk appetite each day, we will extract a *trading signal*.

As we have seen that risk appetite should be moving around its mean most of the time and then exhibit extreme behaviours around extreme events, events that have a *systemic* effect on the market (in effect market indicator start being more correlated or uncorrelated). So we monitor the level of the γ^{th}-moving-quantile as seen before by computing the γ^{th}-quantile at each time t with the data in the window X_t^L. When the level of our indicator exceeds this *threshold*, we get a signal and then take action based on this. Note that this method is suited to our case since the risk appetite measure (or aversion or change in risk appetite, depending on the index) is a unitless number, hence we compare it to itself and not some general level. In fact, as we can see in practice, each index has very different values but they don't necessarily mean anything. Using a γ-quantile has the same effect as re-scaling each risk appetite measure.

We have three types of trading signals we will test: *long only*, *short only* and *long-short* signals. Each one of them is self-explanatory. The *long-only* signal is only triggered when crossing the upper *threshold* forbidding us of being short the security, i.e. selling a security we do not hold (which is a standard trading behaviour but requires that the short seller posts capital for example so there are costs linked to holding a short position). The *short-only* signal is set of when the lower *threshold* is reached and forbids us to ever be long the trading object. Finally, the *long-short* signal allows us to go long or short depending on our signal. This raises already a few questions: what thresholds do we use ? What time-horizon are we using ? How do we interpret the signals in terms of trading orders ? What orders do we exactly execute when our signal tells us to buy or sell ? What do we even buy or sell ? etc.

4.1 The code structure

Before we start, we would like to say a word about the code. Our program is written in the programming language `R` [54]. We orgnaised our code inside a package called {`rappetite`} documented using the `roxygen2` [56] library for package creation (see also [52]). We will not mention all the packages we used in our program for there are a lot (nearly 45) however we want to emphasize the usefulness of the {`xts`} package [57] when dealing with time series data. Because `R` is famously known for its slow speed `for` loops, we also spent some time optimising the performances of our code using vectorisation of our procedures and calling to packages designed to call procedures in `C++`, for example the {`RcppRoll`} package [55] which is very useful. Another useful discussion is made on the website [58]. Our code is organised into core functions that the user will not be able to see and wrapper or user-friendly functions with description and documentation. In this way, the programme is scalable and can be integrated to other portfolio management programs (in `R` though).

4.2 Risk-Parity portfolio

To back-test we need to specify a portfolio to be traded. Once we have chosen the assets we want to trade, we classify them into two sub-categories, the *risk-on* assets and the *risk-off* assets. We differentiate these two categories using long-term historical volatility estimates (see Equation (2.1)) of each asset and ranking them by this feature. We are then in a position to decide which assets have been more volatile overall (they will go in the *risk-on* class) and which are the ones that have been more stable (they will go in the *risk-off* class). We then compute the *risk-parity* allocation of this portfolio at time $t = 0$, the first day of trading. As described in the book [47] by Qian, the intuition is to weight the assets of your portfolio according to their relative riskiness. $W_{t=0}$ will describe our reference portfolio allocation, we call it our *benchmark* and it will always be (in the case of multiple assets being traded) equal to the *risk-parity* portfolio for our back-testing environment.

Definition 4.1 (Risk-Parity Portfolio)**.** Let us consider n assets, whose returns through time will be noted r_t^i, for any $t > 0$, and $i \in \{1, 2, \ldots, n\}$. Their respective risk will be estimated by their realised variance or historical volatility squared $(\sigma_{h,t}^i)^2$ (see Equation (2.1)). We can drop the subscript t as the window used for this estimate is typically twenty years long so we do not assume the estimate will change a lot during the trading period. Hence we consider this risk estimate as *fixed*. We also drop the subscript h as we know we are talking in this context of historical volatility estimates. We will use the notation $(\hat{\sigma}_t^i)^2$ for the variance estimate of the asset i over the trading period if needed, so the subscript t and the hat notation will differentiate both. Let us then note the long term variance estimate of each n asset by σ_i^2. The total variance of the portfolio will then be

$$\sigma_P^2 = \sigma^T \Sigma \sigma = \sum_{i=1}^{n} \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \rho_{ij}$$

where $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)^T$ and Σ is the correlation matrix between assets with correlation between asset i and j noted ρ_{ij}. To build a *risk-parity* portfolio, we want to find a set of weights $W = (w_1, w_2, \ldots, w_n)^T$ such that every asset has the same participation in the total portfolio variance. The participation of asset i is defined as

$$p_i = \frac{\sigma^T \Sigma \mathrm{e}_i \sigma}{\sigma^T \Sigma \sigma}$$

where $\mathrm{e_i}$ is a $n \times n$ matrix with zeros everywhere expect for a one on the i^{th} element of the diagonal. We therefore want to solve the following problem

$$\text{find a set of weights } W = (w_1, w_2, \ldots, w_n)^T \text{ such that } p_1 = p_2 = \cdots = p_n = \tfrac{1}{n}.$$

In practice, we use a numerical optimiser and we formulate the problem

$$\begin{aligned} \underset{W=(w_1,w_2,\ldots,w_n)}{arg\min} \quad & \left[\sum_{i=1}^{n} \left(p_i - \frac{1}{n} \right)^2 \right] \\ \textbf{s.t.} \quad & \mathbf{0} \leq W \leq \mathbf{1}, \\ & \sum_{i=1}^{n} w_i = 1. \end{aligned} \tag{4.1}$$

So the solution W_{opt} to Equation (4.1) is a convex combination of positive weights, it belongs to the hyper-cube $[0,1]^n$.

Remark 4.2. Note that this way of weighting takes into account the correlations between assets unlike the so-called *volatility-weighted* portfolio allocation which only takes the individual volatilities assets to build their weights.

As we get trading signals s_t we will send orders in the market and this will have for effect to change our portfolio allocation hence moving away from the benchmark trying to capture trends or avoid drawdowns. This is where the two classes of assets become relevant as the idea is to allocate more weight to one subclass depending on the signal we get. And this is reasonable since we want to allocate more of our wealth in riskier assets when market is predominantly not volatile (to capture more risk premia) and on the contrary we want to protect ourselves in turbulent markets, hence allocate more of our wealth in stable assets. We choose a trading parameter δ that will fix the percentage of our wealth we want to reallocate each time we have a non-zero trading signal.

Remember we have 3 trading signals so if we take the most general one (*long-short*) we can be in three different states. Say $\delta = 0.8$, then

$$\begin{cases} \text{in the 'long' state, we allocate 80\% of our wealth in risky assets and 20\% in stable ones,} \\ \text{in the 'neutral' state, we allocate our wealth as dictated by the original weights } W_0, \\ \text{in the 'short' state, we allocate 20\% of our wealth in risky assets and 80\% in stable ones.} \end{cases}$$

The trading signal will determine in which state we are and hence we will reallocate our wealth accordingly. It is perhaps better to illustrate with an example. Note that ***we trade the day after we receive a signal*** as the data we use are the daily closing prices, this is how we deal with ***slippage*** effects.

Example 4.1. *Say we hold one stock with price S_t and one bond with price B_t so the initial price vector is $P_0 = (S_0, B_0)^T$. Suppose the initial portfolio weights (initial risk-parity allocation) are $W_{t=0} = (0.25, 0.75)^T$ and our trading parameter is $\delta = 0.8$. Then in the 'neutral' state, our wealth at time $t = 0$ is $\mathbb{W}_0 = W_0 P_0^T = 0.25S_0 + 0.75B_0$. We wait one time step and the prices move to $P_1 = (S_1, B_1)^T$ and our wealth to $\mathbb{W}_1 = W_0 P_1^T = 0.25S_1 + 0.75B_1$. Imagine now that at time $t = 1$ we receive a 'buy' signal from our risk appetite indicator. We have decided to interpret this signal as a sign to bear more risk. So we want to go from this 'neutral' state in $t = 1$ to the 'long' state at $t = 2$ and to do this we want to allocate more wealth into risky assets, here the stock S. At time $t = 2$ we will then have $W_2 = (0.8, 0.2)^T$. Our wealth will then be $\mathbb{W}_2 = \mathbb{W}_1(W_2 P_2^T) = \mathbb{W}_1 0.8S_2 + \mathbb{W}_1 0.2B_2$ which means that 80% of the wealth at time $t = 1$ will be reinvested in the stock and 20% in the bond.*

Remark 4.3. If we are trading more than two assets, say one risky and two stable, the relative proportion inside the *risk-off* assets must not change when we change state. Say we are in the 'neutral' state at time 0, $W_0 = (0.2, 0.6, 0.2)$ so we have a 20-80 allocation which means 20% in *risk-on* assets and 80% in *risk-off* assets. If we go in the 'long' state at time $t = 1$, the allocation will become a 80-20 allocation and the weights will be $W_1 = (0.8, 0.15, 0.05)$ so that inside the *risk-off* class, the relative importance is kept constant, i.e. asset 2 is still 75% of the *risk-off* class and asset 3 is still 25% of the *risk-off* class.

Note that we also can trade only one asset if we want. In that case, to avoid having to take into account the cost of leverage, we will not allocate all of our wealth all the time. In fact, instead of holding one share of the asset we will hold δ shares of the asset in the 'neutral' state, one share in the 'long' state and $2\delta - 1$ shares in the 'short' state.

4.3 Trading Signal and Strategy

What remains now is to extract the signal s_t from the index. The signal gives us triggering times and direction for orders but to actually trade in our back-testing framework, we need to construct a weight vector W_t that will represent our allocation at each time t. We will use a three-year-long *rolling windows* (with exponential weighting for some of the indices) $L = 3 \times 252$ (expect for GRAI, RAI-MI and CSFB as mentioned before) and $\alpha = 0.01$ the decay factor in exponential weighting; the trading period is from $t = 0$ up to $t = T$. We will use four levels γ for our thresholds : 0.75-rolling-quantile, 0.85-rolling-quantile, 0.95-rolling-quantile and 0.99-rolling-quantile as we expect risk appetite information to be more evident and observable when our index moves consequently with respect to its three-year distribution. The $(1-\gamma)$-thresholds are the 0.25-rolling-quantile, 0.15-rolling-quantile, 0.05-rolling-quantile and 0.01-rolling-quantile respectively and are also monitored for the short signal. But let's be more precise.

Consider that we are standing at time t, our data-window is X_t^L, the current value of the γ-moving quantile is q_t^γ and we already have signal s_t and weights W_{t+1}. Indeed, we consider one day lag between the reception of the signal and the actual trade. This is because, as mentioned earlier, we use daily closing price data. So having already the signal s_t we know how to trade at $t+1$ to adjust our portfolio. Now, want to compute the signal for the next time step s_{t+1} when the index value is x_{t+1}. Based on s_{t+1} we will determine if we trade, hence the set of weights W_{t+2}, and so on (in the stochastic analysis terminology, our trading strategy is called a *previsible* process).

There is also a practical problem we want to tackle. As mentioned earlier, the way we get a trading signal is based on extreme values appearing in the risk appetite indicator time series. However in practice, this can lead to interpreting noise as a signal. Consider for example that we look at the upper threshold only (we want to test a *long-only* signal). If the risk appetite indicator we are looking at shoots up at some time by a sufficiently high value it will trigger a trade. Consequently, the threshold then also moves up with this jump. Depending on the γ level we are considering, this can move the threshold more or less. If we consider $\gamma = 0.75$, then we will need 25% of the data inside the window to move up so the threshold follows, it is a 'slow' threshold if we want. On the other hand, if $\gamma = 0.99$ then one or two consecutive extreme points can potentially make the threshold move because only 1% of the population need to increase to make the 99^{th}-percentile move. What we want to illustrate is that for an 'slow' threshold, the risk appetite indicator actually has the time move before the threshold catches up. This can lead to the indicator oscillating around the threshold level, triggering useless trades. To avoid that effect, we introduce a *friction* parameter ϵ that will be used to this end by acting as a safeguard.

These considerations lead to the following steps. Consider the case of the *long-only* signal first, *short-only* and *long-short* will be easier to explain afterwards.

- We compute the position p_{t+1} of the index relative to the current threshold. We compute it as follows,

$$\begin{cases} \text{if} & x_{t+1} & \in & (q_t^\gamma(1+\epsilon), +\infty) & \text{then} & p_{t+1} = 1, \\ \text{if} & x_{t+1} & \in & (q_t^\gamma(1-3\epsilon), q_t^\gamma(1+\epsilon)) & \text{then} & p_{t+1} = 0 \\ \text{if} & x_{t+1} & \in & (-\infty, q_t^\gamma(1-3\epsilon)) & \text{then} & p_{t+1} = -1. \end{cases} \tag{4.2}$$

Note that the *friction* is set asymmetrically so that it is easier to trigger a trade when moving away from the benchmark than when coming back to it. Our *holding periods* - periods during which we hold a portfolio that is different from the benchmark - are then extended.

- Once we have our relative position p_t, we can compute s_{t+1}. We must be careful as being in the case $p_{t+1} = 1$ does not necessarily mean $s_{t+1} = 1$ because we could already have been in that position before, $p_t = 1$. It is important to consider what was the last trade made since we can't go 'long' if we already are and we cannot go 'short' when in 'neutral' state. Since we have all the signals up to s_t in memory, we know what the last trading direction was. It is not sufficient to keep track only of the last signal because the last trade may have happened before time t. Let's call d the last trade, it can only be equal to 1 or -1 as a signal being equal to 0 is interpreted as 'hold position'. Then the algorithm can be written as

$$\begin{cases} \text{if } p_{t+1} = 1 & \begin{cases} \text{if } d = 1 \text{ then } s_{t+1} = 0, \\ \text{if } d = -1 \text{ then } s_{t+1} = 1, \end{cases} \\ \text{if } p_{t+1} = 0 & \text{then } s_{t+1} = 0 \\ \text{if } p_{t+1} = -1 & \begin{cases} \text{if } d = 1 \text{ then } s_{t+1} = -1, \\ \text{if } d = -1 \text{ then } s_{t+1} = 0. \end{cases} \end{cases} \tag{4.3}$$

- Then we compute what we call our *state* signal S_{t+2} which is just the cumulative sum of the signal vector up to $t+1$, i.e. $S_{t+1} = \sum_{i=0}^{t+1} s_i$. It tells the program in what state should the portfolio be in next. If we have a positive signal at time $t-k$ and then signals are null until time $t+1$ where you have a negative signal, then your $(k+1)$-long holding period is finished. But during that period, your state signal S_t was positive all the time. Weights are then computed as explained in the previous section 4.2 on *risk-parity* as we change states.

Now that we have the *long-only* trading signal, we do the same for the lower thresholds $(1-\gamma)$ to compute the *short-only* signal. And we can compute the *long-short* signal by adding the both together and taking care of the cases where we cross both thresholds in one jump (for example from a low outlier to a high outlier) as it can potentially happen in very noisy and stationary risk appetite indicators.

Last but not least, we have to decide in what *direction* we want to trade, i.e. what i sour *trading strategy*. Essentially, as the author of [42] says, we want to test two fundamental directional trades

State at t	long			neutral			short		
signal at t	1	0	-1	1	0	-1	1	0	-1
momentum	long	long	neutral	long	neutral	short	neutral	short	short
contrarian	neutral	long	long	short	neutral	long	short	short	neutral

Table 1: State transition for each strategy.

which are the *'riding the wave'* and the *'trading against the crowd'* strategies, or as we prefer to call them, *momentum* and *contrarian* strategies. The idea is simple, Table 1 resumes the effect of both strategies, for a given state and signal at time t (first and second row) we have the resulting state in the table for momentum and contrarian strategies respectively. Say we are trading with a momentum strategy and we are currently in the neutral state and we receive a negative signal, we will go 'short'. Of course, it is implicit that depending on the type of our signal (remember we have three types *long-only*, *short-only* and *long-short*) some states will be never be visited because we will never receive the signal to transit to them. We can see that the *contrarian* strategy is to trade in the exact opposite direction as the *momentum* one. Once we have all of that it's just a mater of keeping track of one strategy and the benchmark every day and we can compare both as we will see in the next section.

Remark 4.4. Note that we do *not* take care of the trading cost as we will trade mainly futures on indices (aggregates of assets) which require no funding and if we do want to use our program to backtest with stocks, we have limited our strategy to investing only our current wealth at each time step so no leverage is needed. Also the contracts we trade are very liquid so the transaction costs will be negligible.

We want to mention that the plots of every index during period 2006-01-01 to 2011-12-31 are available in appendix A.5. Also, it is interesting to compare the plots in appendix A.7 where we can see the same indices with their upper and lower thresholds (90-percentile and 10-percentile with friction parameter of 0.05 as well as a window of 15% of the data length, purely for the sake of the plot) and the corresponding signals they have triggered (blue for positive and red for negatives). We can see which ones are the trading the most more graphically.

4.4 Results and Sharpe Ratio

Now that we have an *benchmark* portfolio and a weight vector W_t each trading day, we can compute our daily wealth and more importantly our daily returns for each strategy. The purpose of the *benchmark* is to represent the so-called *buy-and-hold* strategy which is a passive strategy (similar in practice to what Exchange Traded Funds (ETFs) do). We will then compare this passive strategy to our active ones based on dynamic portfolio reallocation dictated by our signals for different parameters as we have seen (γ levels, direction of trading, trading reallocation parameter δ, trading strategies) and see if they perform better. We will now see how to compare them between them and to the *benchmark*.

The Sharpe ratio is a measure proposed by Sharpe in [48] in an attempt to measure the performance of a portfolio given that we only care about two things: expected excess returns $\mathbb{E}[r_x] = \mathbb{E}[r_p - r_b]$ (where r_p are the returns of the portfolio and r_b are the returns of a benchmark portfolio) and the volatility of these expected returns σ. He also assumes a market with complete information and all investors act as optimal portfolio managers which means that the cost of capital is the *risk-free rate* r_f so their expected excess returns are $\mathbb{E}[r_x] = r_f + \beta\sigma$, where β is the risk premium. Therefore, the portfolio is optimal in that sense whenever the risk premium is

$$\beta = \frac{\mathbb{E}[r_x] - r_f}{\sigma}.$$

Note that this assumes that one only needs the expected return and the standard deviation to measure portfolio performance. But in our case it will be considered sufficient. In practice, the average past excess returns $\overline{r}_x$ will be used. Also, estimate of risk, given by the historical volatility of the excess returns $\hat{\sigma}$ will be used. Our *Sharpe ratio* is then

$$\beta = \frac{\overline{r}_x - r_f}{\hat{\sigma}}.$$

So the higher the Sharpe ratio the better. Our trading test will be divided in two parts. First part we will test the indices that do not depend on option data and then we will test those separately. The reason why we do this is mainly because of the time it takes for our algorithm to run the test if we do the same period as for the other indices. The procedure will be to perform an first backtest on period 2008-2011, select the best performing indices and then test them on 2012-2017 period. We choose to proceed this way because it gives a good way of testing the performance in a realistic way and to give an answer to the question "if it has worked in the past will it still work in the future ?".

- We backtest the following indices CSFB, GRAI, RAI-MI, HSBC, LCVI, ML z-score, ML robust-zscore, ML m-score, UBS. The trading period varies is 2008-2011. The assets we trade are ["ESU7 Index", "DNU7 Index"] for the *risk-on* assets and "TYZ7 Index" for the *risk-off* assets (note that we have to roll the futures to be able to trade, these contracts have a fixed maturity and exist only for a limited period of time). The benchmark is presented in Figure 21. We extract the trading signal for each different combination of parameters presented in Table 2, which means that we have 24 possibilities (2 strategies, 3 signals, 4 thresholds).

- We test all these strategies separately and compute their (annualised) Sharpe ratio as well as the Sharpe ratio of the benchmark. We then keep only the cases where we beat the

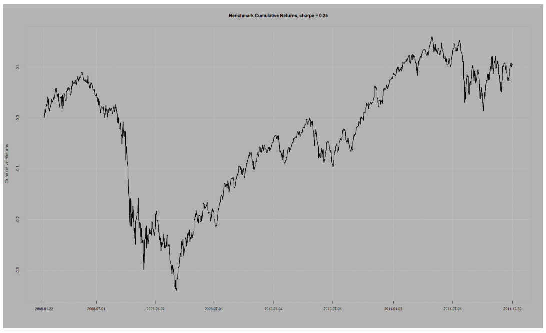

Figure 21: Risk-Parity Weighting Benchmark.

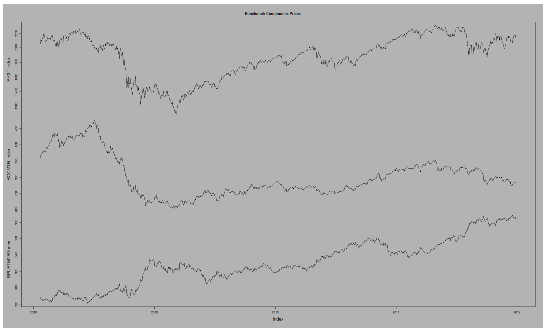

Figure 22: Benchmark Breakdown.

Feature	Values
strategies	momentum, contrarian
signals	long-only, short-only, long-short
thresholds	0.75, 0.85, 0.95, 0.99-quantiles
friction parameter	0.05
reallocation weight	0.9

Table 2: Features for the in-sample backtest 2008-2011.

benchmark, i.e. our Sharpe ratio is better. For example, with the GRAI index, the benchmark achieves an annualised Sharpe ratio of 0.25. The Sharpe ratios of different backtests are summarised in Figure 30. It can seem a bit hard to read but it is actually very simple. Each coloured line represents one backtest and its points are the Sharpe ratios for the corresponding threshold level (the x-axis of the graph). There are then only 6 lines, representing each one arrangement of strategies with signals (2×3). The dashed horizontal line represents the Sharpe ratio level of the benchmark portfolio. We can see here that the contrarian strategies tend to perform better whereas the momentum ones (intuitively) do not as they are the exact opposite strategies. Also note that better trends are captured for lower thresholds as the Sharpe ratios are larger for 0.75 and 0.85-quantile thresholds than for 0.99-quantile thresholds. Remember that when we say 0.99-quantile threshold, this is for a positive signal, but of course our program also computes the 0.01-quantile threshold for negative signals. Every case is tested. We can see the results in Figure 31 where the cumulative returns over the trading period of the benchmark (in black) and the winning arrangements of parameters (in color) are plotted. We can also see a distribution of trades in the middle as well as the drawdown (cumulative negative returns only) chart below. In this example the best performing strategy is the long-short signal, contrarian strategy and 0.75-quantile threshold (so also 0.25-quantile threshold as we are in a long-short signal) with a Sharpe ratio close to 0.9 and cumulative return of +144% (dark blue line on Figure 31) over the trading period,

Parameter arrangement	Annualised Sharpe Ratio (%)
benchmark	0.194
GRAI long-short contrarian 0.75	0.870
ML robust-z-score long-short contrarian 0.85	0.372
ML m-score short-only contrarian 0.75	0.292
LCVI short-only contrarian 0.75	0.360
LCVI long-short contrarian 0.75	0.274

Table 3: Top 5 Sharpe ratios of winning strategies for the out-of-sample backtest.

beating the benchmark with 0.25 Sharpe ratio and +110% cumulative return.

- We repeat this idea for every index mentioned before (CSFB, GRAI, RAI-MI, HSBC, LCVI, ML z-score, ML robust-zscore, ML m-score, UBS) keeping the winning arrangements of parameters every time. This gives us the results presented in Table 3 and in Figures 33 and 32. It turns out that GRAI has the best Sharpe ratio and contrarian strategies perform better in general. This can be expected since the Financial crash is included in the trading period. Nevertheless, all of these five selected strategies remarkably outperform the market represented by our benchmark portfolio during the trading period 2008-2011. Especially, we can see on Figure 33 on the lower part, the drawdown chart outlines the capacity of these indices to avoid crashes.

- As a comparison with the conclusions of Aaron and Illing in [1] who concluded that CSFB was the best index, we present its performances in Figures 34 and 35. CSFB best performing arrangement is long-short signal, contrarian strategy and 0.95-quantile threshold, with a Sharpe ratio slightly over 0.8 and cumulative return of +128% when the benchmark. But this is rather an exception as we can see from Figure 35, the majority of the other Sharpe ratios are largely smaller.

- Now that we have selected GRAI, ML-robust-z, ML-m, LCVI indices, we take them to the next stage of our test and run a backtest on the period 2012-2017 with same set of parameters as in Table 2. We can see some serious moves in these indices around the beginning of 2016 in Figure 23, which is alarming because comparable in amplitude to the ones seen in 2009 respectively. Final results of the backtest are presented in Figures 24. We can see that these indices continue to beat the benchmark and with a significantly higher Sharpe ratio. This is a good element in favour of assessing that our risk appetite indices contain information that is valuable and exploitable by a simple automated trading strategy such as this one. Notable, from Table 4.

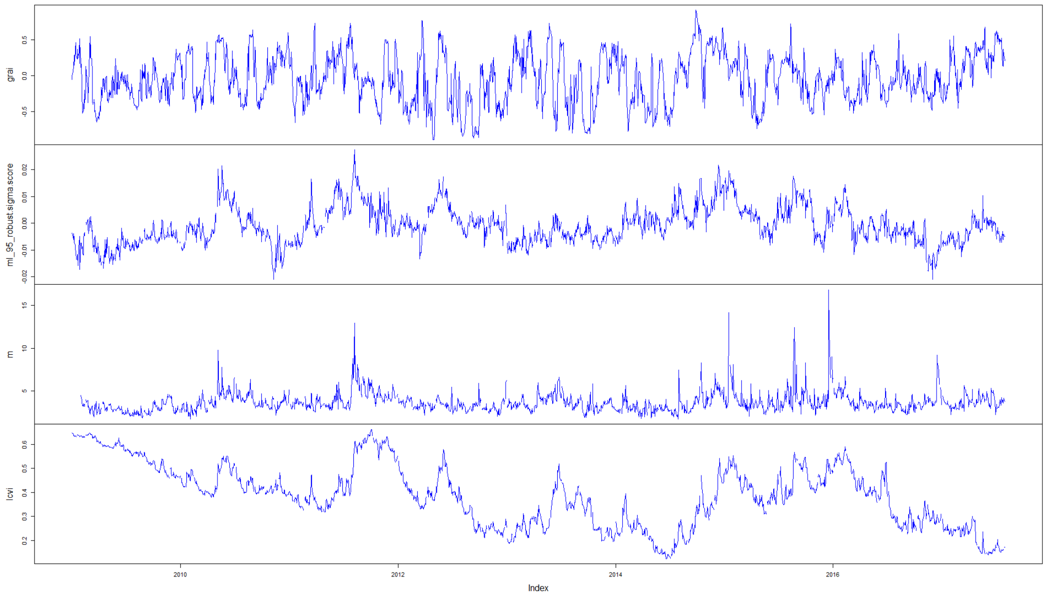

Figure 23: Selected Risk Appetite Indices on 2009-2017.

Parameter arrangement	Annualised Sharpe Ratio (%)
benchmark	1.643
LCVI short-only momentum 0.85	2.081
GRAI long-only contrarian 0.75	1.933
ML m-score long-short contrarian 0.95	1.804
LCVI long-short momentum 0.85	1.951
LCVI long-only contrarian 0.99	1.902

Table 4: Top 5 Sharpe ratios of winning strategies for the out-of-sample backtest 2012-2017.

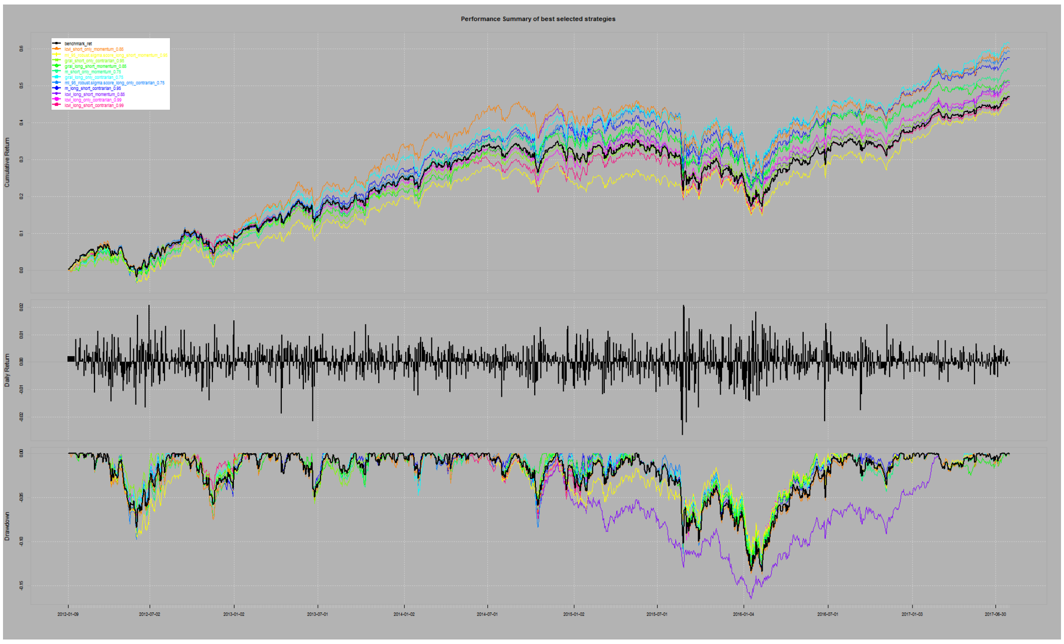

Figure 24: Selected Risk Appetite Indices on 2009-2017.

The tests for the index derived from option prices (KTT Index uses Equation (3.5) and BoE Index uses Equation (3.7)) is done only on the S&P500 future as it is the only time series in our data. It it still useful to observe its performance against the buy-and-hold strategy which is known for being hard to beat. We compared both approaches of Gai & Vause [23] against Karampatos *et al.* [30]. The backtest procedure is the same as before but we do only one step because of data and computational limitations. However we show a very short period. Nonetheless, trading produces very good performance on a short term, see Figures 25 and 26. They reveal that starting the trading in April 2017, all the contrarian strategies have beaten the benchmark, which is an excellent result. The Sharpe ratios are displayed in the Table 5. However, it could also be that the period is luckily helping this kind of strategy so we should also try a longer trading period for our backtest but we are limited by our data set. Therefore, we also choose not to display BoE Index because of the too-little data usable for computing the distributions and their variance from Equation (3.7). KTT Index is somehow more robust. We show also results for a slightly longer trading period (November 2016- July 2017) in 27 and 28. They show more mitigated results.

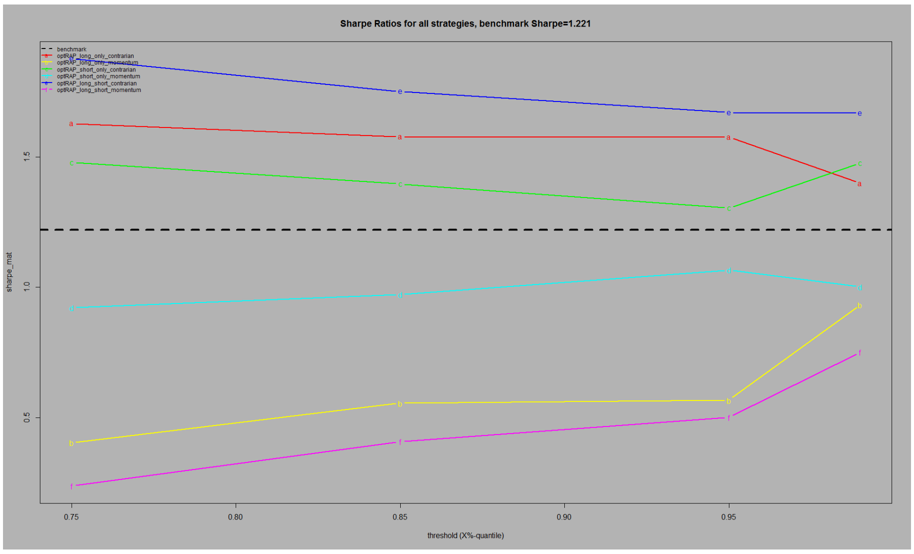

Figure 25: Sharpe Ratios of KTT Index on 2017.

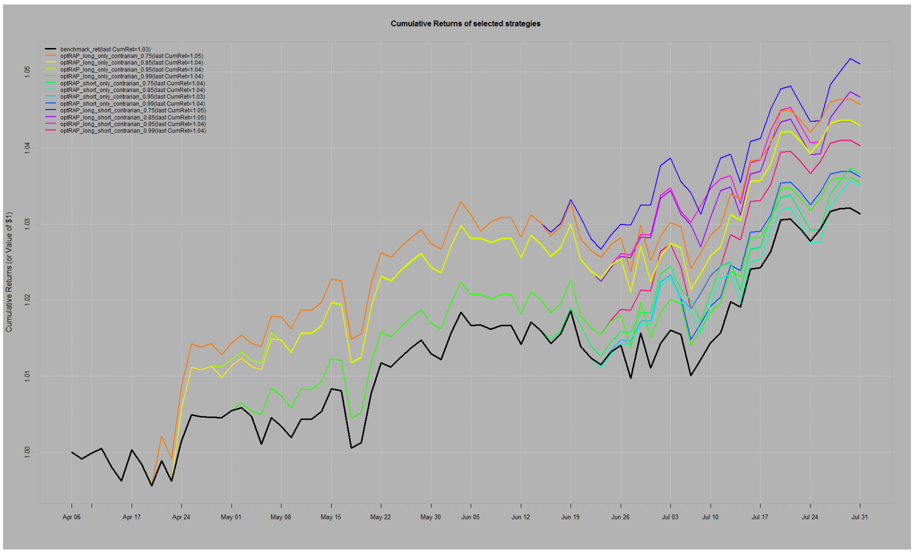

Figure 26: Selected Cumulative Returns of KTT Index on 2017.

Parameter arrangement	Annualised Sharpe Ratio (%)
benchmark	1.221
long-short contrarian 0.95	1.671
long-short contrarian 0.75	1.877
long-short contrarian 0.85	1.751
long-short contrarian 0.99	1.669

Table 5: Top 4 Sharpe ratios of winning strategies for KTT Index backtest on 2017.

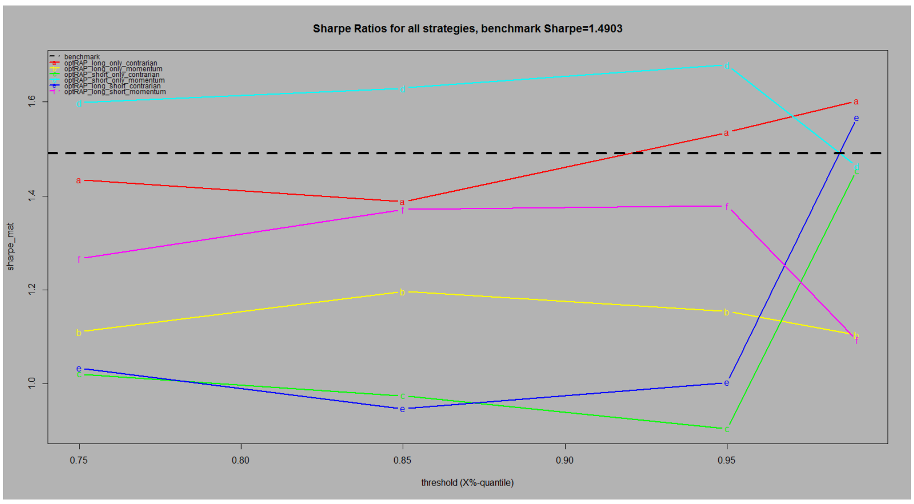

Figure 27: Sharpe Ratios of KTT Index on 2016-2017.

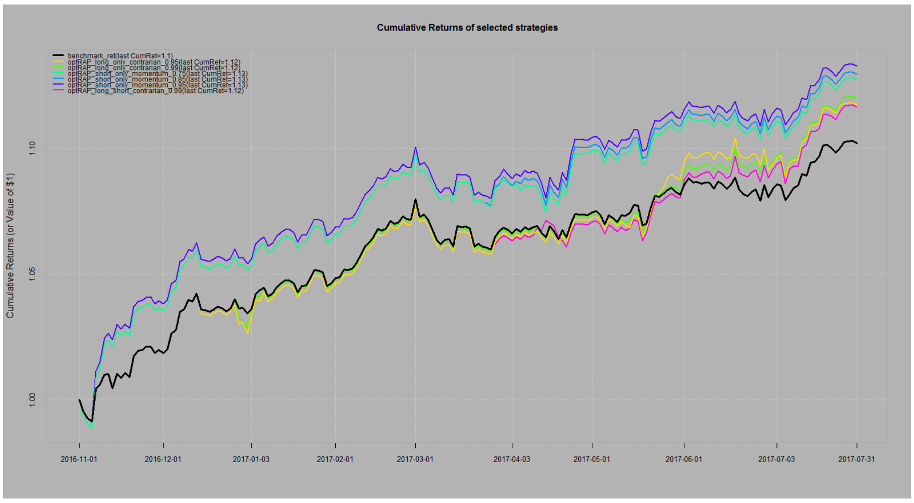

Figure 28: Selected Cumulative Returns of KTT Index on 2016-2017.

5 Conclusion

Clearly, understanding the relation between risk aversion and risk appetite is a deep topic and constructing a rigorous mathematical framework for risk appetite would require a more theoretical study. Nevertheless, we have implemented existing risk appetite indicators and seen how they can be used to construct simple automated active portfolio management strategies. Some of them are performing better than others with respect to the Sharpe Ratio metric. We have shown that they work well when trading classical assets such as the S&P500 or the U.S. 10-year note. But different assets could be used, as long as they are related in some way in the data base of the indices. We do not expect our examples to work on a very illiquid Brazilian bond for example, but why not with Gold or the Swiss Franc. We advise interested investors to use these risk appetite indices for strategies considering diversified portfolios and contrarian strategies, meaning that they have been shown to be most useful for deciding when to go against the market. They also have been shown to be limited to long-term considerations so we recommend not to use them as a signal for short-term investments. This might explain why the literature on these indicators mainly comes from central banks and large investment banks.

We have outlined the advantages and disadvantages of each method and we hope we gave the reader a better understanding of the way they operate. The few indicators we selected exhibit good results for long periods either with or without memory of the financial crisis and we can expect them to continue to work in any state of the market as long as we use their signals for a long-term strategy and a diversified portfolio. Note however that our conclusion on CSFB differs from the one made by Aaron & Illing in [1] so these indices can potentially stop performing as well as they have done in the past if the method they use to extract risk appetite is less adapted to a new market environment. We suspect that the post-crisis environment has led this index to perform poorly compared to GRAI for example.

Possible areas of future study could involve, in our opinion, combining indices to create a composite one. Also adding features like longest runs (i.e. periods of consecutive positive returns will potentially inform us on imminence of *mean-reversion* effect), regime switching insights (see for example Hamilton [24]) and crowding measures (see for example [5], [14], [25]) of the traded assets might give an even better metric that would anticipate potential signs of stress. We want to add that in a financial environment where algorithms make more investment decisions than actual humans, this whole concept of risk appetite and investor behaviour can seem obsolete. However, we think that with the right type of (*high-frequency*) data, this idea could be applied to any type of investor, computer or human and potentially be scaled to shorter investment strategies.

A Appendix

A.1 Exponential Weighting

The justification for the value of w_0 in (Definition 3.9) is the following (these steps are described in the appendix of [3] but as we changed the indices we felt compelled to re-write it to be consistent with our notation)

$$\begin{aligned}\frac{1}{w_0} &= \sum_{k=0}^{L-1} e^{\alpha(t-k-L)} \\ &= e^{\alpha(t-L)} + \cdots + e^{\alpha(t+1-2L)} \\ &= \sum_{i=t+1-2L}^{t-L} (e^{\alpha})^i \\ &= \frac{1-(e^{\alpha})^{(t-L)-(t+1-2L)+1}}{1-e^{\alpha}} \\ &= \frac{1-e^{\alpha L}}{1-e^{\alpha}}.\end{aligned}$$

A.2 Another Rank Correlation Measure

Definition A.1 (Kendall's *tau*). Suppose we have two time series x_k and y_k indexed by a discrete time index $k \in \{1, \ldots, n\}$. Then Kendall's τ measures the rank correlation between x and y. For each observation, $r(x_i)$ is the rank of x_i with respect to his own distribution and similarly $r(y_i)$ is y_i's rank in y, for any $i \in \{1, \ldots, n\}$. Then for any two pairs of elements (x_i, y_i) and (x_j, y_j) with $(i, j) \in \{1, \ldots, n\}^2$ we call them

$$\begin{cases} \textit{concordant} & \text{if } x_i > x_j \text{ and } y_i > y_j \text{ or } x_i < x_j \text{ and } y_i < y_j, \\ \textit{discordant} & \text{if } x_i > x_j \text{ and } y_i < y_j \text{ or } x_i < x_j \text{ and } y_i > y_j, \\ \textit{tied} & \text{if } x_i = x_j \text{ or } y_i = y_j. \end{cases}$$

We then associate with each couple of pairs the numbers

$$u_{ij} = \operatorname{sgn}(r(x_j) - r(x_i)) \text{ and } v_{ij} = \operatorname{sgn}(r(y_j) - r(y_i))$$

which are not defined for $i = j$ It can be proven that $\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}v_{ij} = n_c - n_d$, i.e. the number of concordant pairs minus the number of discordant pairs and that $\sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij}^2 = n(n-1)$. As stated in [31], the original definition of this measure was then

$$\tau = \frac{2}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}v_{ij}.$$

To take care of ties, the most common definition given by [31] is

$$\tau = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} u_{ij}v_{ij}}{\sqrt{\left(\frac{1}{2}n(n-1) - n_t^x\right)\left(\frac{1}{2}n(n-1) - n_t^y\right)}}$$

with n_t^x and n_t^y the number of tied pairs for variable x and y respectively.

A.3 Leland Economic Model Maximisation Solution

This is described in Leland [37] §4 but as the notations are sometimes confusing we make it clear here by explaining the framework. Consider the same portfolio with initial value $W_0 = 1$ and terminal value W_T. Let then $p(W_T)$ denote the price at time 0 of the contingent claim paying $W_0 = 1$ at maturity on the terminal value of the portfolio. Consider also an investor whose terminal wealth is a function of the reference portfolio's terminal wealth, i.e. $Y(W_T)$. The investor will optimise Y over all outcomes of the terminal value of the reference portfolio W_T under a budget constraint, meaning

$$\max_{Y(\cdot)} \int_{\mathbb{R}} q(W_T)U(Y(W_T))dW_T,$$

with the constraint

$$\int_{\mathbb{R}} p(W_T)Y(W_T)dW_T = Y(W_0) = I,$$

where q is the *subjective* density of the investor and p is the *risk-neutral* density or the density of the average investor. Following Lagrange multipliers method we solve the equation

$$\frac{\partial}{\partial Y}\left[\int_{\mathbb{R}} U(Y(w))q(w)dw - \lambda\left(\int_{\mathbb{R}} Y(w)p(w)dw - I\right)\right] = 0.$$

This is an informal derivation but very simply we have

$$\begin{aligned} & \frac{\partial}{\partial Y}\left[\int U(Y)q - \lambda\left(\int Yp - I\right)\right] = 0 \\ \Rightarrow \quad & U'(Y)q - \lambda p = 0 \\ \Leftrightarrow \quad & U'(Y) = \lambda\frac{p}{q}. \end{aligned}$$

Then we can get derive each side in W_T to get

$$\begin{aligned} & \partial_w(U'(Y)) = \partial_w\left(\lambda\frac{p}{q}\right) \\ \Rightarrow \quad & U''(Y) = \frac{\lambda}{Y'}\frac{p'q - pq'}{q^2}. \end{aligned}$$

Therefore we have

$$\rho_{abs} = -\frac{U''}{U'} = \left(\frac{q'}{q} - \frac{p'}{p}\right)\frac{1}{Y'},$$

therefore considering $Y(w) \equiv w$ yields the result.

A.4 Data Tickers by Index

All of our data is from Bloomberg L.P. and this source only.

List of Bloomberg Tickers for CSFB
"SPX index", "AS51 Index", "BEL20 Index", "TXEQ Index", "CAC Index", "DAX Index", "FTSEMIB Index", "NKY Index", "AEX Index", "IBEX Index", "OMX Index", "SMI Index", "UKX Index", "ATX Index", "KFX Index", "HEX Index", "ASE Index", "ISEQ Index", "NZSE Index", "OBX Index", "BVLX Index", "HSI Index", "SENSEX Index", "SHSZ300 Index", "JCI Index", "KRX100 Index", "FBMKLCI Index", "PCOMP Index", "STI Index", "TWSE Index", "SET Index", "IPSA Index", "MEXBOL Index", "MERVAL Index", "IBOV Index", "RTSSTD Index", "MXZA Index", "BESWGA Index", "GTCAD10Y Corp", "GSWISS10 Index", "GTDEM10Y Corp", "GFIN10YR Index", "GSGB10YR Index", "GDGB10YR Index", "GTNLG10Y Corp", "GTATS10Y Corp", "GTBEF10Y Corp", "GTFRF10Y Corp", "GIGB10YR Index", "GTGBP10Y Corp", "GTESP10Y Corp", "GNOR10YR Index", "GTITL10Y Corp", "GTPTE10Y Corp", "GGGB10YR Index", "GSAB10YR Index","GTJPY10Y Corp", "GTTWD10Y Govt", "GTHKD10Y Corp", "MASB10Y Index", "GVSK10YR Index", "GVTL10YR Index", "GTAUD10Y Govt", "GTNZD10Y Corp", "GTCNY10Y Govt", "GTIDR10Y Corp", "GTMYR10YR Corp", "GTINR10Y Corp", "GTCLP10YR Corp", "GTTWD10Y Govt" and "USC0TR03 Index" as the risk-free rate

Table 6: Data used for the CSFB Index.

List of Bloomberg Tickers for GRAI
spot tickers : "EURUSD Curncy", "JPY Curncy", "GBP Curncy", "CHF Curncy", "CAD Curncy", "AUD Curncy", "NZD Curncy", "ZAR Curncy", "ARS Curncy", "CZK Curncy", "HKD Curncy", "MXN Curncy", "NOK Curncy", "PLN Curncy", "SGD Curncy", "SEK Curncy", "TWD Curncy";
forward tickers : "EURUSD1M BGN Curncy", "JPY1M BGN Curncy", "GBP1M BGN Curncy", "CHF1M BGN Curncy", "CAD1M BGN Curncy", "AUD1M BGN Curncy", "NZD1M BGN Curncy", "ZAR1M BGN Curncy", "APN+1M CMPN Curncy", "CZK1M BGN Curncy", "HKD1M BGN Curncy", "MXN1M BGN Curncy", "NOK1M BGN Curncy", "PLN1M BGN Curncy", "SGD1M BGN Curncy", "SEK1M BGN Curncy", "NTN1M BGN Curncy"

Table 7: Data used for the GRAI Index.

List of Bloomberg Tickers for HSBC
”SPX Index”, ”CCMP Index”, ”INDU Index”, ”RTY Index”, ”SX5E Index”,
”UKX Index”, ”CAC Index”, ”DAX Index”, ”USGG10YR Index”,
”GCAN10YR Index”, ”GUKG10 Index”, ”GFRN10 Index”, ”GDBR10 Index”,
”MOODCAAA Index”, ”MOODCBAA Index”, ”TWI USSP Index”, ”TWI EUSP Index”,
”TWI SFSP Index”, ”TWI BPSP Index”, ”TWI JPSP Index”, ”TWI ADSP Index”,
”TWI CDSP Index”, ”TWI NDSP Index”,”XAU Curncy”, ”XAG Curncy”,
”LMCADS03 LME Comdty”, ”VIX Index”, ”CL1 Comdty”, ”CO1 Comdty”,
”NG1 Comdty”, ”HO1 Comdty”, ”W 1 Comdty”, ”S 1 Comdty”, ”CT1 Comdty”

Table 8: Data used for the HSBC Index.

List of Bloomberg Tickers for LCVI
”CSI BARC Index”, ”USSW10 CMPN Curncy”,”USSP10 CMPN Curncy”,
”CEMBTOBS Index”, ”VIX Index”, ”FXVIX Index”

Table 9: Data used for the LCVI Index.

List of Bloomberg Tickers for ML
”FEDL01 Index”, ”BASPTDSP Index”, ”VIX Index”, ”PCUSEQTR Index”,
”CHFAUD Curncy”, ”XAU Curncy”, ”CSI BARC Index”, ”SPUSTTTR Index”

Table 10: Data used for the ML Index.

List of Bloomberg Tickers for UBS
”CSI BARC Index”, ”CEMBTOBS Index”, ”VIX Index”, ”FXVIX Index”,
”XAU Curncy”, ”SPXT Index”, ”SPUSTTTR Index”

Table 11: Data used for the UBS Index.

List of Bloomberg Tickers
”SPX Index”, ”ESU7 Index”

Table 12: Data used for the BoE Index.

A.5 All Risk Appetite Indices Plot

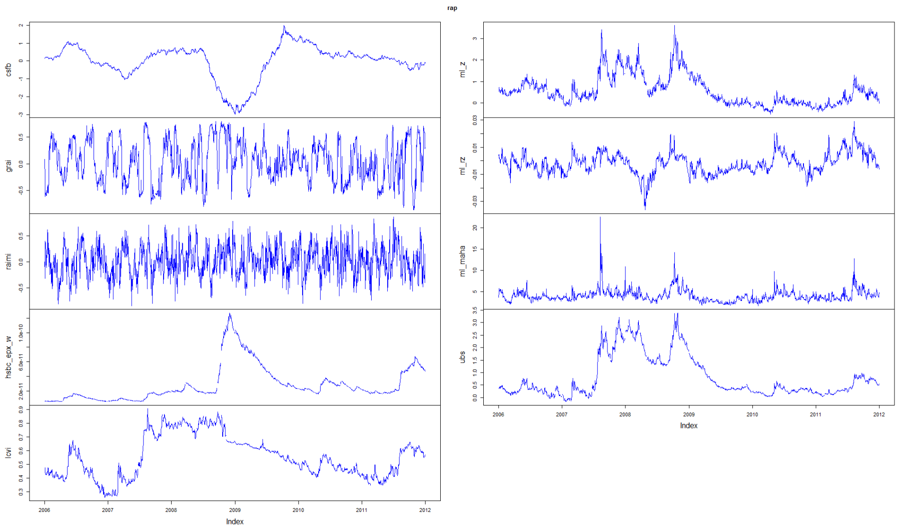

Figure 29: All Risk Appetite Indices.

A.6 Results Graphs

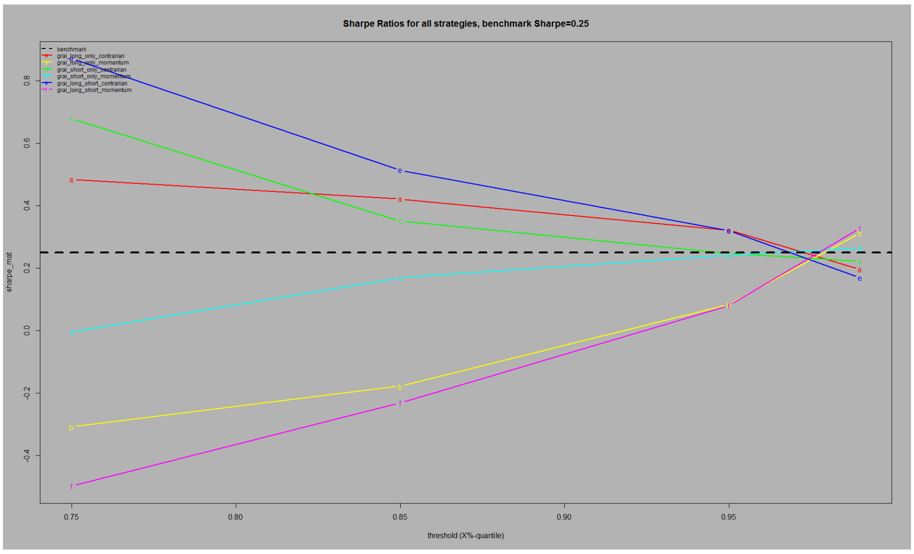

Figure 30: All Sharpe ratios of GRAI variations.

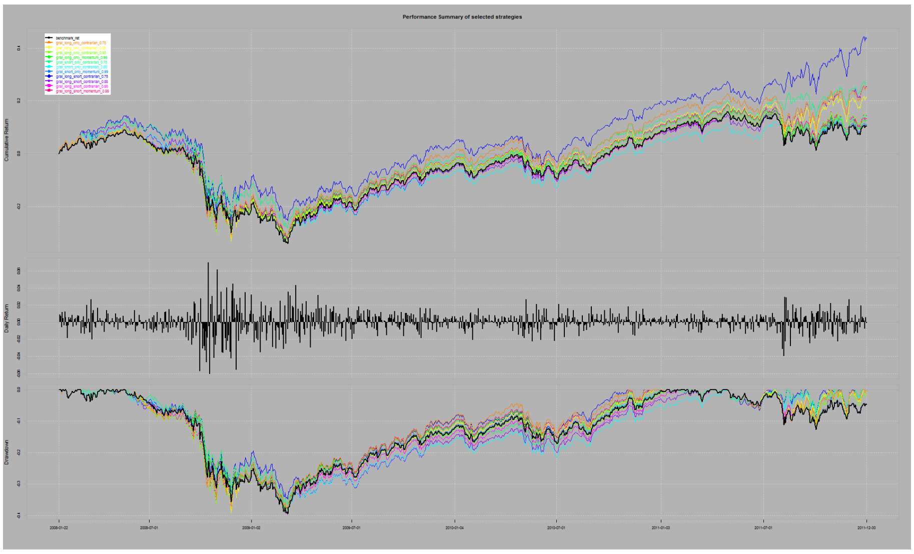

Figure 31: Summary of winning strategies.

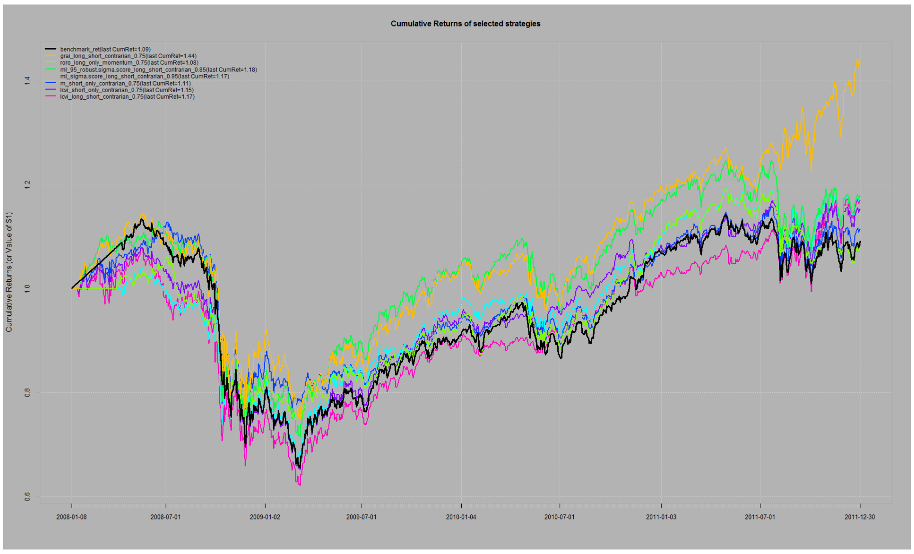

Figure 32: All Winning Parameter arrangements.

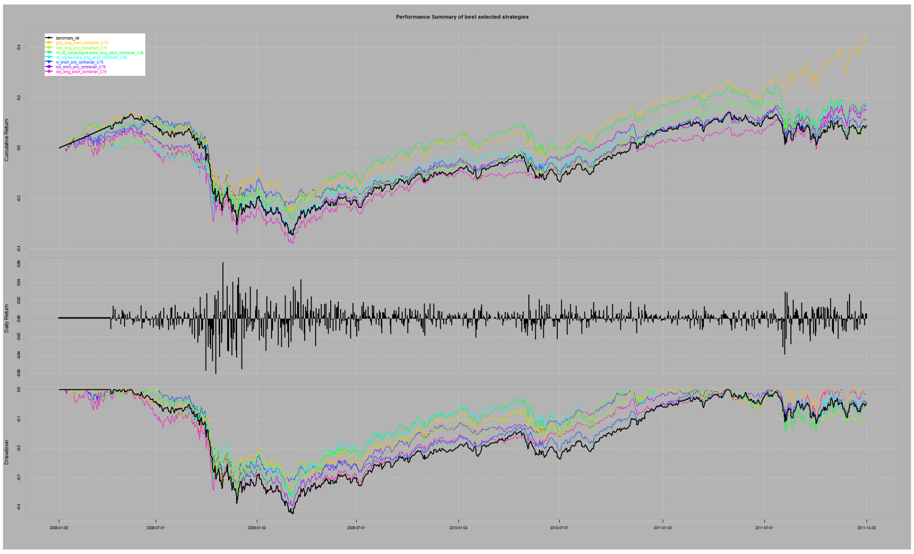

Figure 33: Summary of winning strategies.

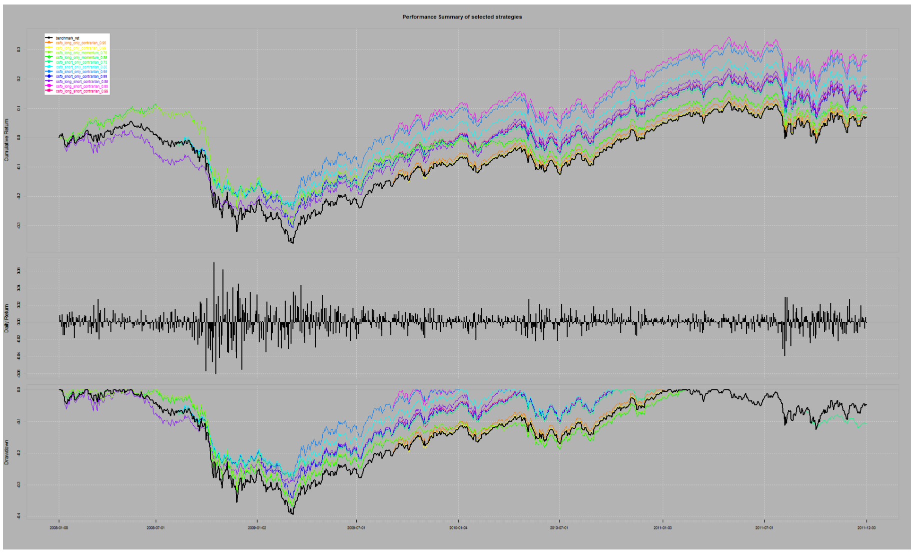

Figure 34: CSFB Winning Parameter arrangements summary.

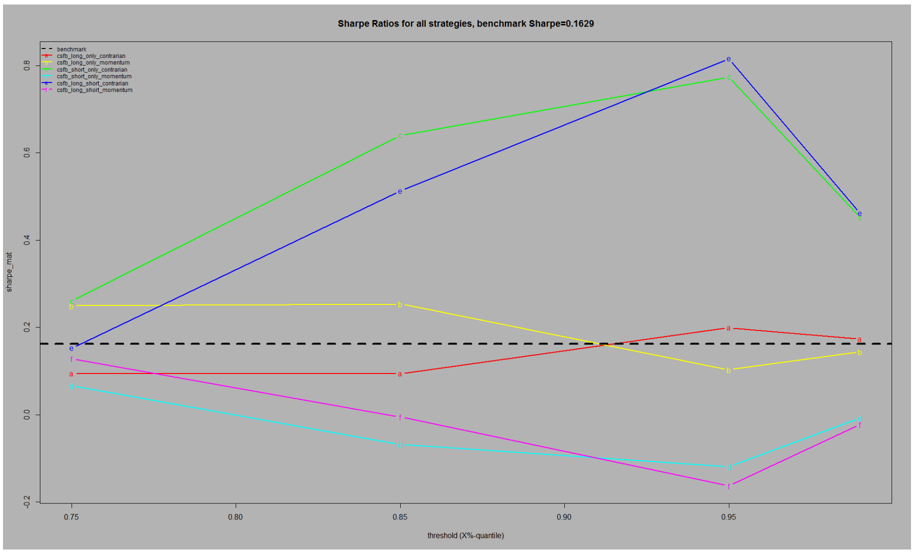

Figure 35: CSFB Sharpe ratios of all arrangements.

A.7 All Trading Signals

Figure 36: CSFB signal.

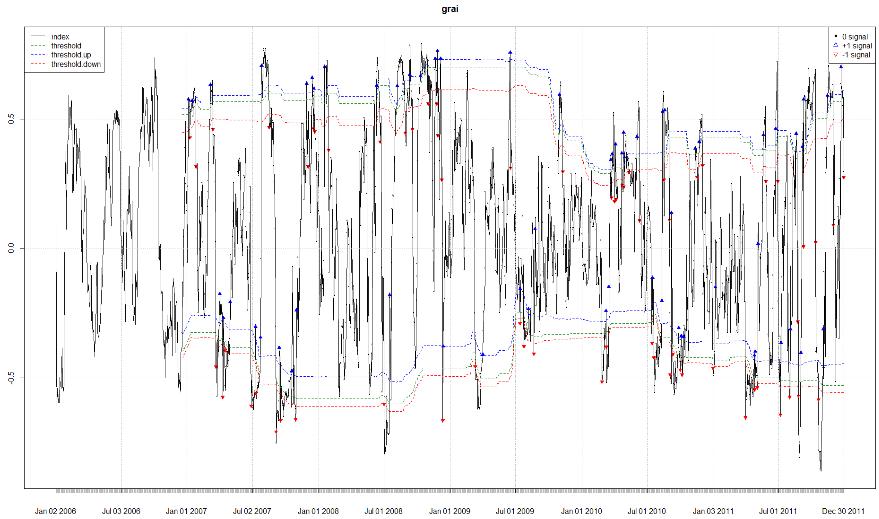

Figure 37: GRAI signal.

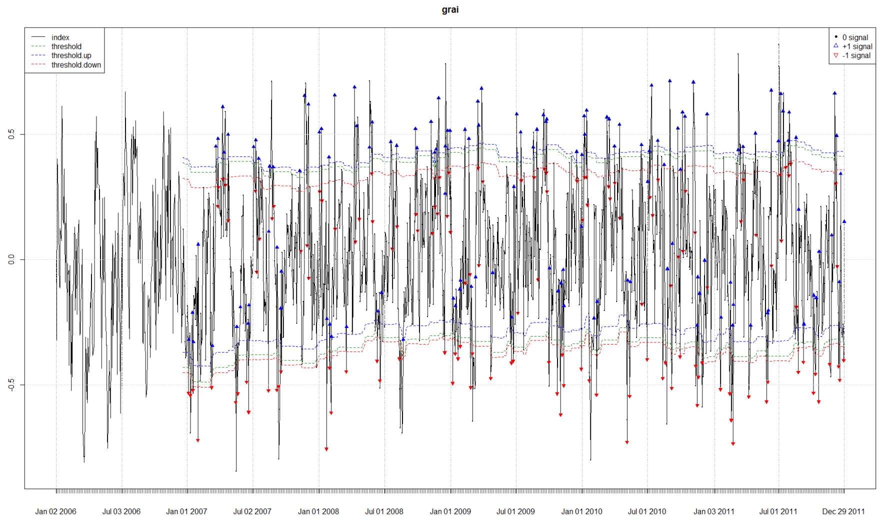

Figure 38: RAI-MI signal.

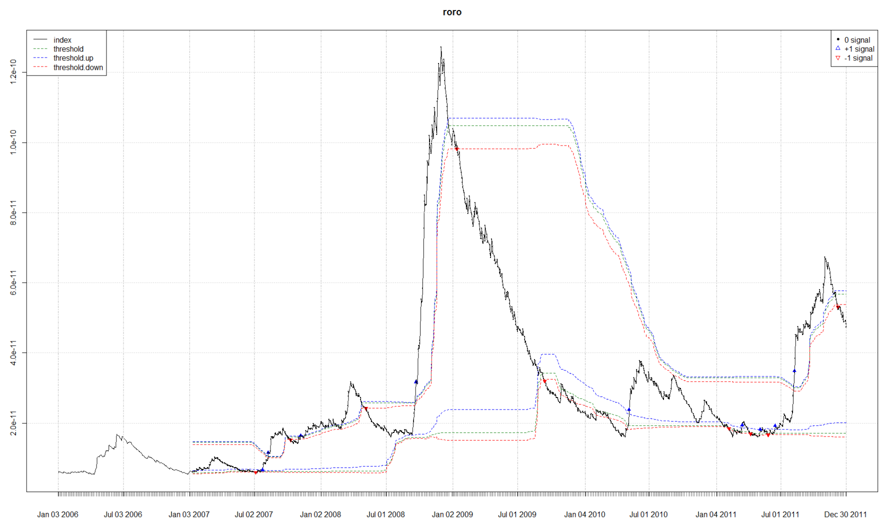

Figure 39: HSBC signal.

Figure 40: LCVI signal.

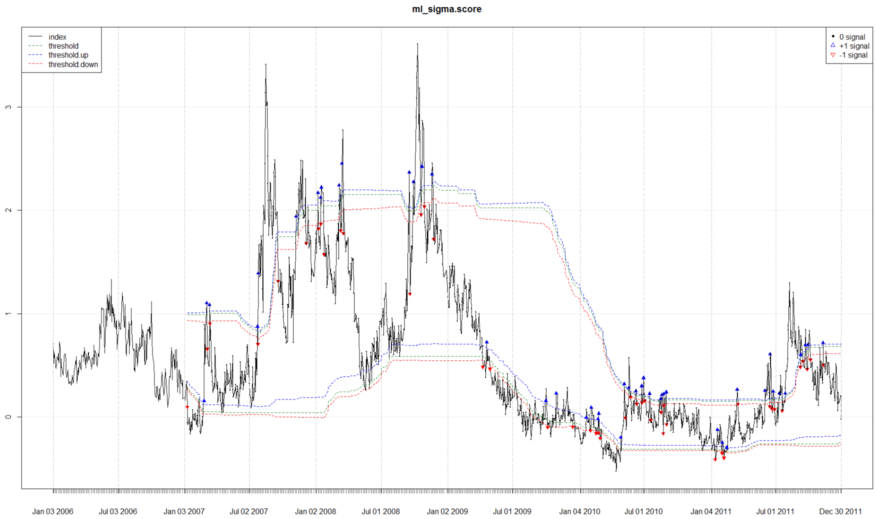

Figure 41: ML z-score signal.

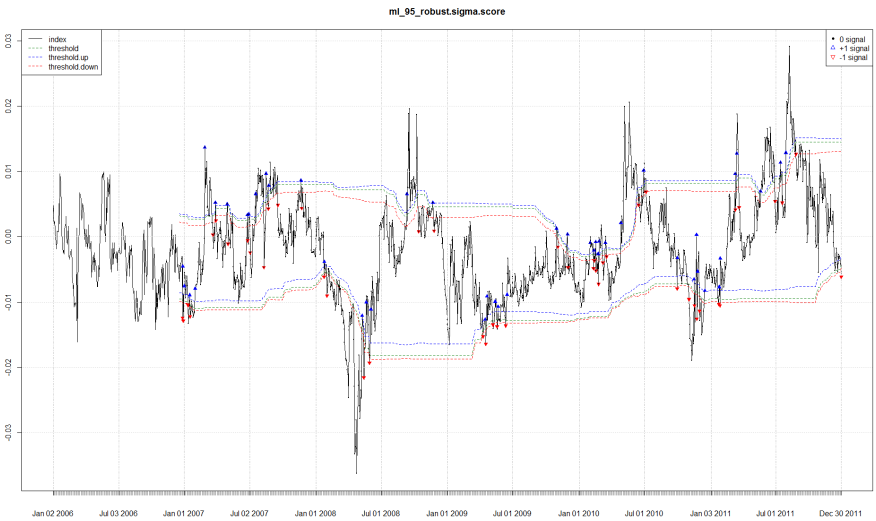

Figure 42: ML robust-z-score signal.

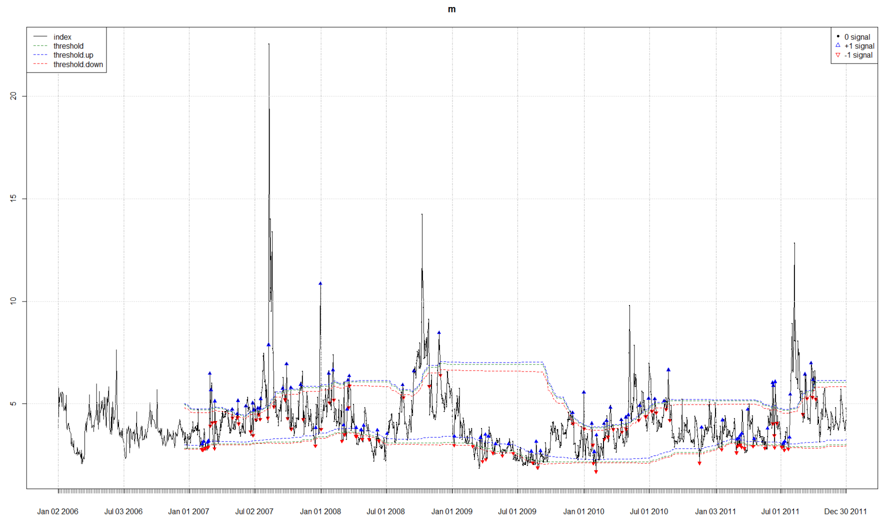

Figure 43: ML m-score signal.

Figure 44: UBS signal.

References

[1] M. Aaron, M. Illing,*A Brief Survey of Risk-Appetite Indexes*, Bank of Canada Financial System Review, 2003.

[2] Y. Ait-Sahalia, A. Lo, *Nonparametric Risk Management and Implied Risk Aversion*, 1997.

[3] T. Aste, T. Di Matteo, F. Possi, *Exponential smoothing weighted correlations*, The European Physical Journal, 2012.

[4] K. J. Arrow, *The role of securities in the optimal allocation of risk bearing*, Review of Economic Studies, vol. 31, p. 91-96, 1964.

[5] M. K. Bayraktar, S. Doole, A. Kassam, S. Rdchenko, *Lost in the crowd ? Identifying and Measuring Crowded Strategies and Trades*, MSCI, June 2015.

[6] J. M. Bibby, J. T. Kent, K. V. Mardia, *Multivariate Analysis*, Academic Press, 1979.

[7] D. Breeden, R. Litzenberger, *Price of State-Contingent Claims Implicit in Option Prices*, The Journal of Business, vol.51, no.4, p.621-651, 1978.

[8] M. Briere, K. Chancari, *Perception des risques sur les marches, construction d'un indice elabore a partir des smiles d'options et test de strategies*, Revue d'economie politique, Vol. 114, p. 527-555. DOI 10.3917/redp.114.0527, April 2004.

[9] S. Chen, S-H. Poon, *Modelling International Stock Market Contagion Using Copula and Risk Appetite*, October 2007.

[10] J. H. Cochrane, *Asset Pricing*, June 2000.

[11] G. M. Constantinides, *Intertemporal Asset Pricing with Heterogeneous Consumers and without Demand Aggregation*, Journal of Business, vol. 55, p. 53-568, 1982.

[12] R. Cont, *Empirical properties of asset returns: stylized facts and statistical issues*, Centre de Mathematiques appliquees, Ecole Polytechnique, 2001

[13] G. Debreu, *Theory of Value*, John Wiley & Sons, 1959.

[14] Deutsche Bank Global Markets Research, *Signal Processing: Strategy crowding*, May 2016.

[15] Deutsche Bundesbank, *Monthly Report*, p. 85-97, October 2005.

[16] L. Devroye, L. Gyorfi, *Nonparametric Density Estimation, the L_1 view*, John Wiley & Sons, 1985.

[17] V. A. Epanechnikov, *Non-Parametric Estimation of a Multivariate Probability Density*, Society for Industrial and Applied Mathematics, 1969.

[18] European Central Bank Financial Stability Review, p.166-171, June 2007.

[19] H. Follmer, A. Schied, *Stochastic Finance, an introduction in discrete time*, Walter de Gruyter, 2002.

[20] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, Springer, Second Edition, August 2008.

[21] J. Gatheral, *A parsimonious arbitrage-free implied volatility parametrization with application to the valuation of volatility derivatives*, Presentation at Global Derivatives, 2004.

[22] J. Gatheral, A. Jacquier, *Arbitrage-free SVI volatility surfaces*, March 2013.

[23] P. Gai, N. Vause, *Measuring investors' risk appetite*, Bank of England Working paper no. 283, 2005.

[24] J. D. Hamilton, *Macroeconomic Regimes and Regime Shifts*, Department of Economics, University of California, 2015.

[25] S.G.Hanson, A. Sunderam, *The Growth and Limits of Arbitrage: Evidence from short Interest*, Harvard Business School, 2013.

[26] HSBC FX Quantitative Strategy, November 2010.

[27] International Monetary Fund, *Global Financial Stability Report*, October 2008.

[28] J. C. Jackwerth, *Recovering Risk Aversion from Option Prices and Realized Returns*, October 1997.

[29] M. C. Jones, M. P. Wand, *Kernel Smoothing*, Chapman & Hall/CRC, 1995.

[30] D. Karampatos, N. Tasharev, K. Tsatsaronis, *Investors' attitude towards risk: what can we learn from options?*, BIS Quarterly Review, June 2003.

[31] M. G. Kendall, *Rank correlation methods*, C. Griffin, 1948.

[32] D .M. Kreps, *Notes on the Theory of Choice*, Westview Press, 1988.

[33] M. Kritzman, Y. Li, S. Page, R. Rigobon, *Principal Component as a Measure of Systemic Risk*, The Journal of Portfolio Management, Vol.37, no.4, Summer 2011.

[34] M. S. Kumar, A. Persaud, *Pure Contagion and Investors' Shifting Risk Appetite: Analytical Issues and Empirical Evidence*, IMF Working Paper, September 2001.

[35] R. Lee, *The moment formula for implied volatility at extreme strikes*, Mathematical Finance, vol. 14, p. 469-480, 2004.

[36] J. Levin, *Choice Under Uncertainty*, Stanford Module on Microeconomics Lecture Notes, October 2006.

[37] H. E. Leland, *Who Should Buy Portfolio Insurance ?*, The Journal of Finance, Vol. 35, no. 2, p. 581-594, Wiley for the American Finance Association, May 1980 .

[38] P. C. Mahalanobis, *On the Generalized Distance in Statistics*, 1936.

[39] J. S. Marron and D. Nolan, *Canonical kernels for density estimation*, Statistics and Probability Letters, vol. 7, p. 195-199, 1989.

[40] G. J. McLachlan, *Mahalanobis Distance*, General Article, Resonance, June 1999.

[41] P. Mielczarski, J.Sweeny, J.Wilmot, *Global Risk Appetite Index*, Credit Suisse First Boston Market focus, Global Strategy Research, February 2004.

[42] M. Misina, *What Does the Risk-Appetite Index Measure ?*, Bank of Canada Working Paper 2003-23, August 2003.

[43] M. Misina, *Benchmark Index of Risk Appetite*, Bank of Canada, 2005.

[44] M. Misina, *Risk Perceptions and Attitudes*, Bank of Canada Working Paper, 2005-17 .

[45] O. Morgenstern, J. Von Neumann, *Theory of Games and Economic Behaviour*, Princeton University Press, 1944.

[46] J. W. Pratt, *Risk aversion in the small and in the large*, Econometrica, vol. 32, p.122-136, 1964

[47] E. E. Qian, *Risk Parity Fundamentals*, CRC Press, 2016.

[48] W. F. Sharpe, *Mutual Fund Performance*, Journal of Business, p. 119-138, January 1966

[49] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman and Hall, 1986.

[50] C. Spearman, *The Proof and Measurement of Association between Two Things*, The American Journal of Psychology, Vol. 15, no. 1, p. 72-101, January 1904 (published by University of Illinois Press).

[51] VIX White Paper, `http://www.cboe.com/micro/vix/vixwhite.pdf`, Copyright CBOE, 2014.

[52] H. Wickham, *R Packages: Organize, Test, Document, and Share Your Code*

[53] Zeliade Systems White Paper, *Quasi-Explicit Calibration of Gatheral's SVI model*, September 2009.

Packages & Links

[54] R description, https://www.r-project.org/about.html

[55] {RcppRoll} package, https://cran.r-project.org/web/packages/RcppRoll/RcppRoll.pdf

[56] {roxygen} package, https://cran.r-project.org/web/packages/roxygen2/vignettes/roxygen2.html

[57] {xts} package, https://cran.r-project.org/web/packages/xts/xts.pdf

[58] https://swcarpentry.github.io/r-novice-inflammation/15-supp-loops-in-depth/, Copyright Software Carpentry Foundation, 2016 - 2017 .