SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Statistics and Quantitative Methods

PhD program Statistics and Mathematical Finance Cycle XXXI

Curriculum in Mathematical Finance

Relevant Properties of the Lambda Value at Risk and Markov Switching Mixture of Multivariate Gaussian Distributions in a Bayesian Framework.

Ruffo Chiara Maria

Registration number: 808680

Tutor: **Emanuela Rosazza Gianini**

Co-tutor**: Andrea Trovato**

Supervisor**: Massimiliano Marzo**

Coordinator: **Giorgio Vittadini**

ACADEMIC YEAR 2017-2018

Contents

Response to Referee

We would like to thank the reviewers for their helpful comments and positive suggestions on the manuscript. In this version of the manuscript we have tried to consider all the points. Here we want to explain step by step the improvements made on the manuscript.

Response to Referee1

1) The Chapter one is published in M. Burzoni, I. Peri, C. Ruffo, " On the properties of the Lambda value at risk: robusteness, elicitability and consistency", Quantitative Finance, 17(11),1735-1843, 2017.

2) In the financial industry the typical effort is to compare quantitative strategies against the market in order to measure their ability to gain money with respect to the buy-and-hold portfolio which is the portfolio full invested into market. However, we also linked our trading strategy with another on.

3) We handle the problem of different closing days in different markets using the SP500 Calendar, which is the most relevant market around the world. In addition, in order to manage weekends data, we use the Bloomberg setup for the download of trading days.

4) The two-State Markov Switching model is almost common in finance for at least a couple of reasons: first of all, this solution is able to highlight clusters in returns in a very intuitive and useful way by describing the pattern of performance in terms of bull and bear market. This way of describing returns is particular useful for conditional portfolio construction issues and conditional tactical asset allocation strategies. Secondly, parameters of Markov Switching models tend to increase quite rapidly when the number of regimes does increase, so

we decided to focus our attention on the 2-state version in order to deal with a number of free parameters under control. Anyway, we run a 3-State Markov Switching model to highlight how the third state is not able to add a strong value to the analysis because its state conditioned normal distribution is really close to the distribution of one of the other two states.

5) All routines are written in Matlab and we provide in the Appendix some of them, but other routines are available upon request. All the routines were implemented by the candidate.

6) In Bayesian Econometrics, it is recommended to work with the so called "natural conjugate priors", so priors once combined with the likelihood generate posterior distributions of the same form of the priors, with posterior parameters easily calculated given prior parameters and sample data. Working with known posterior distributions is crucial for model estimation via simulation. In the Bayesian literature, variance is usually described by the gamma or inverse gamma distribution, while the first order parameter of the normal distribution (μ) is typically described by the normal distribution in conditional terms with respect to the variance.

7)-8) At page 64, we set the parameter of prior gamma distribution in order to put the expected value of the variance of returns equals to 10% in the "bull regime" and to 30% in the "bear regime". These priors in the expected value are typical values of the VIX Index in these king market environment. At page 63-65, we study from a historical point of view the pattern of the VIX Index when the SP500 moves higher or exhibits strong negative corrections. Our analysis confirms how 10% and 30% are good priors for the bull and bear market. Anyway, the impact of these expectations into the estimating process is quite limited and under control because the parameters describing the level of confidence around these expected values are set to be quite large.

9) In Markov Switching Models the variable S_t is a latent variable, so it has to be estimated as the other parameters of the model like μ and σ of the two state conditional normal distributions . The Gibbs sampling estimates the latent variable S_t at the iteration-i via uniform distribution, given the other parameters of model at the iteration $i-1$. At page 53-57 you find the details of the Gibbs Sampling algorithm.

11)-12) Added, Page 66 and Pages 88-91

13) We propose a version of ΛVaR using Markov switching model as process of generating data (MS-ΛVaR). Markov Switching allows to use a dynamic switching of the direction of Λ conditioning on the prevailing regime in each t. We calculate the number of violations of our trading strategy and perform a backtesting exercise both with 1%-MS-ΛVaR and with $1\% - \Lambda VaR$ with increasing Λ.

Response to Referee2

- The aim of the model is to generate a trading strategy able to follow the market when investors are in a "risk-on" mood, condition where the risk-appetite improves sharply and investors' demand for risky assets become supportive. On the other hand, our strategy aims to reduce risk and portfolio drawdowns when the market moves down and investors' risk aversion increases dramatically as the demand for safe assets.

- Figure 2.5 shows the result of the Gibbs Sampling for the posterior distribution of parameters μ and σ. The first chart shows the 2 histograms relative to the draws of μ considering the posterior distributions of the 2 states, the second chart shows the same picture for the variance.

- The preliminary analysis of market returns via Garch model highlights that volatility of returns in not constant but rather it tends to change over time. The Garch model highlights this point assuming the presence of autocorrelation in the patter of the volatility and according to this assumption the memory of shocks in returns on the future volatility tends to decrease gradually overtime. Given this picture, we decided to move ahead in order to describe returns in a more accurate way, not just to manage the time variation of volatility but in the effort to capture the presence of cluster in returns distribution. The state space representation of the pattern of returns and the introduction of the latent variable S_t move in this way. Given the overall evidence of non normality, our solution allows us to recover the normal distribution of returns conditional on any state. In other words, the presence of clusters in returns is not true by construction but it is confirmed by the statistical relevance of the state variable S_t and by the statistical difference of the parameters μ and σ of the normal distributions of the 2 states. The two-State Markov Switching model is almost common in finance for at least a couple of reasons: first of all, this solution is able to highlight clusters in returns in a very intuitive and useful way by describing the pattern of performance in terms of bull and bear market. This way of describing returns is particular useful for conditional portfolio construction issues and conditional tactical asset allocation strategies. Secondly, parameters of Markov Switching models tend to increase quite rapidly when the number of regimes does increase, so we decided to focus our attention on the 2-state version in order to deal with a number of free parameters under control. Anyway, we run a 3-State Markov Switching model to highlight how the third state is not able to add a strong value to the analysis because its state conditioned normal distribution normal distribution is really close to the distribution of one of the other two states.

- The distinction between risky and defensive assets is based on the result of the univariate version of the Markov Switching model, which highlights a different pattern of volatility among the various asset classes. In particular, the estimated volatility for the High Volatility regime highlights how the euro and the yen have shocks in returns which remain under control even in the presence of markets drawdown.

- The different volatility of the two portfolios is the result of the allocation scheme proposed by our model, which tries to manage the relative allocation of the risky portfolio with respect to the defensive one, given the state allocation of the variable S_t. The ability to forecast the regime switching from the normal volatility regime to the high volatility regime allows the model to put money out of risky portfolio in favor of the defensive one, in order to reduce the overall drawdown.

Introduction

Risk management, distribution of returns and trading strategies are the crucial aspects of the complex world of financial markets. Portfolio manager and Risk manager are the beating heart of asset management. The global financial crisis highlighted the limits of the Value at Risk as measure of porfolio risk and led some to doubt about the real benefits of the international and cross assets diversification.

On the one hand, the research of alternative risk measures which overcome lacks of the Value at Risk became questions of primary concern. On the other hand the study of expected mean, volatility and correlations cross assets conditioned on the business cycle was therefore warranted. In this work both this fundamental arguments are bring up.

The first part is focused on important properties for a risk measure. Specifically, a new risk measure, the Lambda Value at Risk, is introduced by [Frittelli et al.2014] as generalization of VaR. This risk measure could be interesting because it is based on the same idea of the VaR, solves some of its problem and keep its good properties. We showed that Lambda Value at Risk is robust and elicitable within particular classes of distributions. In addition, it also satisfies the consistency property without any condition on the mechanism generating data. This first part is taken from the paper "On the properties of the Lambda value at risk: robusteness, elicitability and consistency" published in 2017 on Quantitative Finance, 17(11),1735-1843, with Matteo Burzoni and Ilaria Peri.

The behavior of financial markets may change dramatically when wars, economical or political crises and other events occur. The key issues of quantitative finance regarding to the correct distribution of returns and the correct estimation of correlations in each market status are the central aims of the second part of this work. More specifically, given the heteroskedasticity and the non-normality of returns, we decided to describe returns according to Markov Switching mixture of gaussian distributions. Both univariate and multivariate Markov Switching models are estimate via Gibbs sampling algorithm. Based on the results, a regime-based trade rule is introduced and compared with a buy-and-hold-strategy.
At the end, we propose the use of Markov Switching model in the estimation of the ΛVaR as process of generation data. More specifically, Markov Switching models give the maximum flexibility to the ΛVaR since allow to choice two different Λ conditioned on the prevailing regime in each t, in order to maintain the same numbers of violations respect to the ΛVaR with one function Λ, but reducing the capital aside that could be invested.

Chapter 1

Relevant Properties of Lambda Value at Risk

1.1 Introduction

Risk measurement is a matter of primary concern to the financial services industry. The most widely used risk measure is the value at risk (VaR), which is the negative of the right λ-quantile q_λ^+, for some conventional confidence level λ (e.g. 1%). VaR became popular as a law invariant risk measure for its simple formulation and facility of computation, however, it presents several limits. First, VaR lacks convexity with respect to random variables which, in general, penalize diversification. VaR satisfies, instead, the quasi-convexity property with respect to distributions [Drapeau and Kupper,2012, Frittelli et al.2014]. This condition has a natural interpretation in terms of compound lotteries: the risk of the compound lottery is not higher than the one of the riskiest lottery. Another relevant issue of VaR is the lack of sensitivity to the tail risk as it attributes the same risk to distributions having the same quantile but different tail behavior. Recently, a new risk measure, the Lambda value at risk (ΛVaR), has been proposed by [Frittelli et al.2014]. ΛVaR seems to be interesting for its ability to capture the tail risk by generalizing VaR. Specifically, ΛVaR is defined as follows:

$$\Lambda VaR(F) := -\inf\{x \in \mathbb{R} : F(x) > \Lambda(x)\}$$

where $\Lambda : \mathbb{R} \to [\lambda^m, \lambda^M]$ with $0 < \lambda^m \leq \lambda^M < 1$ is a right continuous and monotone function. When the Λ function is constantly equal to some $\lambda \in (0, 1)$ it coincides with the definition of VaR with confidence level λ. The main idea is that the confidence level can change and it is a function of the asset's losses. In this way, ΛVaR is able to discriminate the risk among P&L distributions with the same quantile but different tail behaviour. In this regard, the sensitivity of ΛVaR is up to the λ^m-quantile of a distribution, since, by definition, $\Lambda VaR(F) \leq VaR_{\lambda^m}(F)$. Nevertheless, the requirement $\lambda^m > 0$ is only technical and λ^m can be chosen arbitrarily close to 0. Properties of ΛVaR such as monotonicity and quasiconvexity are obtained in [Frittelli et al.2014] in full generality (i.e. allowing also for $\lambda^m = 0$).

The purpose of this paper is to study if ΛVaR satisfies other important properties for a risk measure also satisfied by VaR. We first focus on the so-called *robustness* that refers to the insensitivity of a risk estimator to small changes in the data set. We adopt the Hampel's classical notion of qualitative robustness [Hampel et al.,1986, Huber1981], also considered by [Contet al.,2010] for general risk measures (a stronger notion has been later proposed by [Krätschmer et al.2014] for convex risk measures). We show that the historical estimator of ΛVaR is robust within a family of distributions which depends on Λ. In particular, we recover the result of [Contet al.,2010] for VaR, in the case of $\Lambda \equiv \lambda \in (0, 1)$.

A second property we investigate is the *elicitability* for ΛVaR . Several authors underlined the importance of this property in the risk management and backtesting practice [Gneiting2011, Ziegel, 2014, Embrechts and Hofert,2014, Bellini and Bignozzi, 2015]. Specifically, the elicitability allows the comparison of risk measure forecasts and provides a natural methodology to perform the backtesting. As for the case of VaR, also ΛVaR is elicitable in a particular family of distributions which depends on Λ and, for the particular case of $\Lambda \equiv \lambda \in (0,1)$, we recover the results of [Gneiting2011]. Note that the elicitability for Λ decreasing was already observed by [Bellini and Bignozzi, 2015], we extend here to the most interesting case of Λ increasing.

Finally, we study the consistency property, as recently proposed by [Davis 2016][1]. In this study, Davis argues that the decision-theoretic framework of elicitability assumes the strong assumption that the theoretical P&L distribution is known and remain unchanged at any time. He thus suggests, under a more refined framework, the use of the so-called *consistency* property, in order to verify if a risk measure produces accurate estimates.We show that ΛVaR satisfies the consistency property without any assumption on the P&L generating process, as in the case of VaR.

The structure of the paper is as follows. After introducing the basic notions and definitions, in Section 1.2, we start examining the robustness property in Section 1.3. We dedicate the Section 1.4 to the elicitability of ΛVaR. Finally, in Section 1.5, we refine the theoretical framework and we verify the consistency of ΛVaR.

1.2 Notations and Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a non-atomic probability space and $L^0 := L^0(\Omega, \mathcal{F}, \mathbb{P})$ be the space of $\mathcal{F}$-measurable random variables that are $\mathbb{P}$-almost surely finite. We assume that $X \in L^0$ represents a financial position (i.e. a loss when $X < 0$ and a profit when $X > 0$). Any random variable $X \in L^0$ induces a probability measure P_X on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ by $P_X(B) = \mathbb{P}(X^{-1}(B))$ for every Borel set $B \in \mathcal{B}_{\mathbb{R}}$ and $F(x) := P_X(-\infty, x]$ denote its distribution function. Let

[1]During the review process of this paper the consistency property has been renamed by Davis as *calibration* of predictions in a dynamic setting.

$\mathcal{D} := \mathcal{D}(\mathbb{R})$ be the set of distribution functions and $\mathcal{D}_1$ those with finite first moment.

A risk measure is a map $\rho : L \subseteq L^0 \to \overline{\mathbb{R}}$ that assigns to each return $X \in L$ a number representing the minimal amount of capital required by the regulator in order to cover its financial risk. The majority of risk measures used in finance are distribution-based risk measures, that is, they assign the same value to random variables with the same distribution. Such risk measures ρ are called law-invariant, more formally they satisfy:

$$X \sim_d Y \Rightarrow \rho(X) = \rho(Y).$$

In this way, a risk measure ρ can be represented as a map on a set $\mathcal{M} \subseteq \mathcal{D}$ of distributions. With a slight abuse of notation, we still denote this map by ρ and set:

$$\rho(F) := \rho(X)$$

where F is the distribution function of X. Since the seminal paper by [**?**], the theory of risk measures has been based on the study of their minimal properties. Also when risk measures are defined on distributions, monotonicity is generally accepted; formally, for any $F_1, F_2 \in \mathcal{M}$, ρ is monotone if:

$$F_1(x) \geq F_2(x), \ \forall x \in \mathbb{R} \ \text{ implies } \ \rho(F_1) \leq \rho(F_2).$$

Other properties have been discussed by academics. As pointed out in [Frittelli et al.2014], the convexity property, for risk measures defined on distributions, is not compatible with the translation invariance property. Thus, we might require ρ to satisfy quasiconvexity [Drapeau and Kupper,2012, Frittelli et al.2014]:

$$\text{for any } \gamma \in [0,1], \quad \rho(\gamma F_1 + (1-\gamma)F_2) \leq \max(\rho(F_1), \rho(F_2)).$$

It is widely accepted in the financial industry to adopt the risk measure Value at Risk (VaR) at a confidence level $\lambda \in (0,1)$, that is defined as follows [Artzner et al.,1999]:

$$VaR_\lambda(F) := -\inf\{x \in \mathbb{R} : F(x) > \lambda\}. \tag{1.1}$$

VaR is monotone and quasiconvex [Frittelli et al.2014] but, obviously from the definition, it is not tail-sensitive. In order to overcome its limits, the [Basel Committee, 2013] recommends the use of Expected Shortfall (ES), formally given by:

$$ES_\lambda(F) := \frac{1}{\lambda}\int_0^\lambda VaR_s(F)ds. \tag{1.2}$$

ES is able, by definition, to evaluate the tail risk and it satisfies the subadditivity property on random variables [Artzner et al.,1999].

Another tail sensitive risk measure is Lambda Value at Risk (ΛVaR), recently introduced by [Frittelli et al.2014], whose properties are the main topic of this paper. ΛVaR generalizes VaR by considering a function Λ instead of a constant λ in the definition of VaR. The advantages of considering the Λ function are twofold: on the one hand, ΛVaR provides a criterion to change the confidence level when the market condition changes (e.g. putting aside more capital in case of expected greater losses), on the other hand, it allows differentiating the risk of P&L distributions with different tail behaviour. Formally, ΛVaR is defined by:

Definition 1.

$$\Lambda VaR(F) := -\inf\{x \in \mathbb{R} : F(x) > \Lambda(x)\} \tag{1.3}$$

where $\Lambda : \mathbb{R} \to [\lambda^m, \lambda^M]$ *with* $0 < \lambda^m \leq \lambda^M < 1$ *is a right continuous and monotone function.*

Intuitively, if both F and Λ are continuous, ΛVaR is given by the smallest intersection between F and Λ. Unlike ES, ΛVaR lacks subadditivity, positive homogeneity and translation invariance when defined on random variables, nevertheless, ΛVaR is monotone and quasiconvex on the set of distributions [Frittelli et al.2014].

1.3 Robustness

Evaluating the goodness of a risk measure involves determining how its computation can be affected by estimation issues. The problem consists in examining the sensitivity of a risk measure to small changes in the available data set; for this reason, *robustness* seems to be a key property. In this context, the first rigorous study is given by [Contet al.,2010]. The authors pointed out that the notion of robustness should be referred to the "risk estimator", as outcome of a "risk measurement procedure" [Contet al.,2010], and they founded the problem on the Hampel's classical notion of qualitative robustness [Hampel et al.,1986, Huber1981]. Basically, a risk estimator is called robust if small changes in the P&L distribution imply small changes in the law of the estimator. They consider the case of historical estimators $\hat{\rho}^h$, those obtained by applying the risk measure ρ to the empirical distribution $\hat{F}$, and they conclude that historical estimators of VaR lead to more robust procedures than alternative law-invariant coherent risk measures.

Afterwards, [Krätschmer et al.,2012] and [Krätschmer et al.2014] argued that the Hampel's notion does not discriminate among P&L distributions with different tail behaviour and, hence, is not suitable for studying the robustness of risk measures that are sensitive to the tails, such as ES. So they focused on the case of law-invariant coherent risk measures and they showed that robustness is not entirely lost, but only to some degree, if a stronger notion is used.

Substantially, the robustness of a risk estimator is based on the choice of a particular metric and different metrics leads to a more or less strong definition. However, as pointed out by [Embrechts et al.2014], a proper definition of robustness is still a matter of primary concern. The aim of this section is to study the robustness of ΛVaR, where we use the weakest definition of robustness proposed by [Contet al.,2010].

Let us denote with $\mathbf{x} \in \mathcal{X}$ the n-tuple representing a particular data set, where $\mathcal{X} = \cup_{n\geq 1}\mathbb{R}^n$ is the set of all the possible data sets. The estimation of F given a particular data set $\mathbf{x}$ is denoted with $\hat{F}$ and represents the map $\hat{F} : \mathcal{X} \to \mathcal{D}$. We call risk estimator the map $\hat{\rho} : \mathcal{X} \to \mathbb{R}$ that associates to a specific data set $\mathbf{x}$ the following value:

$$\hat{\rho}(\mathbf{x}) := \rho(\hat{F}(\mathbf{x})).$$

In particular, the historical estimator $\hat{\rho}^h$ associated to a risk measure ρ is the estimator obtained by applying ρ to the empirical P&L distribution, F^{emp}, defined by $F^{emp}(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(x\geq x_i)}$ with $n \geq 1$, that is:

$$\hat{\rho}^h(\mathbf{x}) := \rho(F^{emp}(\mathbf{x})).$$

Let us denote with $d(\cdot,\cdot)$ the Lévy metric, such that for any two distributions $F, G \in \mathcal{D}$ we have

$$d(F,G) := \inf\{\varepsilon > 0 \mid F(x-\varepsilon) - \varepsilon \leq G(x) \leq F(x+\varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}.$$

Hereafter, we recall the definition of $\mathcal{C}$-robustness of a risk estimator as proposed by [Contet al.,2010], where $\mathcal{C}$ is a subset of distributions.

Definition 2. *[Contet al.,2010] A risk estimator $\hat{\rho}$ is $\mathcal{C}$-robust at F if for any $\varepsilon > 0$ there exists $\delta > 0$ and $n_0 > 1$ such that, for all $G \in \mathcal{C}$:*

$$d(G,F) \leq \delta \;\Rightarrow\; d(\mathcal{L}_n(\hat{\rho},G), \mathcal{L}_n(\hat{\rho},F)) \leq \varepsilon \quad \forall n \geq n_0$$

where d is the Lévy distance and $\mathcal{L}_n(\hat{\rho},F)$ is the law of the estimator $\rho(\hat{F}(\mathbf{X}))$ with $\mathbf{X} := (X_1,\ldots,X_n)$ a vector of independent random variables with common distribution F.

As a consequence of a generalization of the Hampel's theorem, [Contet al.,2010] obtained the following result:

Corollary 3. *[Contet al.,2010] If a risk measure ρ is continuous in $\mathcal{C}$ respect to the Lévy metric, then the historical estimator, $\hat{\rho}^h$ is $\mathcal{C}$-robust at any $F \in \mathcal{C}$*

.

Hence, they show that the historical estimator of VaR_λ is robust with respect to the following set:

$$\mathcal{C}_\lambda := \left\{F \in \mathcal{D} \mid q_\lambda^-(F) = q_\lambda^+(F)\right\} \tag{1.4}$$

where $q_\lambda^+(F) := \inf\{x \mid F(x) > \lambda\}$ and $q_\lambda^-(F) := \inf\{x \mid F(x) \geq \lambda\}$. Substantially, when the quantile of the true P&L distribution is unique, then the empirical quantile is robust. In addition, they showed that the historical estimator of ES_λ is not robust. More important, they pointed out a conflict between convexity (on random variables) and robustness: any time the convexity property is required on distribution-based risk measures, its historical estimator fails to be robust.

We use the result by [Contet al.,2010] in Corollary 3 to prove under which conditions the historical estimator of ΛVaR is robust.

Assumption 4. *In this section we assume that $\Lambda : \mathbb{R} \mapsto [\lambda^m, \lambda^M]$ is a continuous function.*

First, let us consider the following set:

$$E_F := \{x \in \mathbb{R} \mid F(x) = \Lambda(x) \text{ or } F(x^-) = \Lambda(x)\}$$

which consists of those points where the distribution F (or the left-continuous version of F) intersects Λ. We introduce the following class $\mathcal{C}_\Lambda$ of distributions:

$$\mathcal{C}_\Lambda := \{F \in \mathcal{D} \mid F((x, x+\varepsilon)) > \Lambda((x, x+\varepsilon)) \quad \text{for some } \varepsilon = \varepsilon(x) > 0,\ \forall x \in E_F\} \tag{1.5}$$

where $F((x, x+\varepsilon))$ and $\Lambda((x, x+\varepsilon)$ are the images of the interval $(x, x+\varepsilon)$ through F and Λ respectively. The set $\mathcal{C}_\Lambda$ consists of those distributions that do not coincide with Λ on any interval. In the special case of $\Lambda \equiv \lambda \in (0, 1)$, it simply means that the quantile is uniquely determined, thus, the family $\mathcal{C}_\Lambda$ coincides with the one in (1.4) considered by [Contet al.,2010] for the robustness of VaR_λ. Note also that for Λ decreasing this condition is automatically satisfied and hence $\mathcal{C}_\Lambda = \mathcal{D}$.

In the following proposition we show that the historical estimator of ΛVaR is robust in the class $\mathcal{C}_\Lambda$ of distribution functions.

Proposition 5. *ΛVaR is continuous on $\mathcal{C}_\Lambda$. Hence, $\widehat{\Lambda VaR}^h$ is $\mathcal{C}_\Lambda$-robust.*

Proof. We only need to show continuity of ΛVaR respect to the Lévy metric the rest follows from Corollary 3 by [Contet al.,2010].
Fix $\varepsilon > 0$ and $F \in \mathcal{C}_\Lambda$. Let $\overline{x} := -\Lambda VaR(F)$. For any $n \in \mathbb{N}$, define the sets $A_n := \{x \in (-\infty, \overline{x} - \varepsilon] \mid \Lambda(x) - F(x + 1/(2n)) \geq 1/n\}$. Observe that, for $x \in A_n$, we have

$$\frac{1}{n+1} \leq \frac{1}{n} \leq \Lambda(x) - F\left(x + \frac{1}{2n}\right) \leq \Lambda(x) - F\left(x + \frac{1}{2(n+1)}\right)$$

and hence $A_n \subseteq A_{n+1}$. We first show that

$$(-\infty, \overline{x} - \varepsilon] = \bigcup_{n \in \mathbb{N}} A_n.$$

The inclusion $\supseteq$ is obvious. Fix $x \in (-\infty, \overline{x} - \varepsilon]$ and let $\gamma := \Lambda(x) - F(x)$. By definition of $\overline{x}$ and $\mathcal{C}_\Lambda$ we have that $\Lambda(x) > F(x)$ and hence $\gamma > 0$. From the right-continuity of $\Lambda - F$, and the continuity of Λ, for any $\varepsilon' > 0$ there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, $\Lambda(x + 1/(2n)) - F(x + 1/(2n)) \geq \gamma - \varepsilon'$ and $\Lambda(x) - \Lambda(x + 1/(2n)) \geq -\varepsilon'$. Take now $\varepsilon' = \gamma/4$ to obtain

$$\Lambda(x) - F(x+1/(2n)) = \Lambda(x+1/(2n)) - F(x+1/(2n)) + \Lambda(x) - \Lambda(x+1/(2n)) \geq \gamma - \gamma/4 - \gamma/4 = \gamma/2$$

since $\gamma > 0$, for a sufficiently large n we get $\Lambda(x) - F(x + 1/(2n)) \geq 1/n$ and hence $x \in A_n$ for some $n \in \mathbb{N}$, as claimed.

We now show that there exists $n_0 \in \mathbb{N}$ such that

$$(-\infty, \overline{x} - \varepsilon] = \bigcup_{n=1}^{n_0} A_n.$$

If indeed $A_{n+1} \setminus A_n \neq \emptyset$ for infinitely many $n \in \mathbb{N}$, then there exists a convergent subsequence $\{x_k\}$ with $x_k \in A_{n_k+1} \setminus A_{n_k}$ and $\tilde{x} := \lim_{k\to\infty} x_k$ such that: i) $-\infty < \tilde{x} \leq \overline{x} - \varepsilon$ and ii) $F(\tilde{x}) \geq \Lambda(\tilde{x})$. i) follows from the fact that Λ has a lower bound λ^m while F obviously tends to 0 as x approaches $-\infty$. There exists therefore $M > 0$ and n_M such that $(-\infty, M] \subseteq A_n$ for every $n \geq n_M$; ii) follows from $x_k \notin A_{n_k}$ which implies $\Lambda(x_k) - F(x_k + 1/(2n_k)) < 1/n_k$ and the right-continuity of F which implies

$$F(\tilde{x}) \geq \limsup F(x_k + 1/(2n_k)) \geq \limsup \Lambda(x_k) - 1/n_k = \Lambda(\tilde{x})$$

where the last inequality follows from the continuity of Λ. If $F(\tilde{x}) = \Lambda(\tilde{x})$, by definition of $\mathcal{C}_\Lambda$ we obtain $-\Lambda VaR(F) \leq \tilde{x} \leq \overline{x} - \varepsilon$ which is a contradiction. The same conclusion obviously follows when $F(\tilde{x}) > \Lambda(\tilde{x})$.

We have therefore shown the existence of $n_0 \in \mathbb{N}$ such that $\Lambda(x) - F(x + 1/(2n_0)) \geq 1/n_0$ for every $x \in (-\infty, \overline{x} - \varepsilon]$ Take now $\delta_1 := 1/(2n_0)$ and $G \in \mathcal{C}_\Lambda$ such that $d(F, G) < \delta_1$. We thus have, for any $x \leq \overline{x} - \varepsilon$,

$$\Lambda(x) - G(x) \geq \Lambda(x) - F(x + \delta_1) - \delta_1 \geq \frac{1}{n_0} - \delta_1 = \frac{1}{2n_0} > 0.$$

It follows

$$\Lambda VaR(G) \leq \Lambda VaR(F) + \varepsilon \tag{1.6}$$

which is the upper semi-continuity.

By showing the lower semi-continuity we conclude the proof. From Definition 1, for any $\varepsilon > 0$, there exists $\hat{x} \in [\overline{x}, \overline{x} + \varepsilon]$ such that $\gamma := F(\hat{x}) - \Lambda(\hat{x}) > 0$. Since Λ is continuous, there exists $\delta > 0$ such that for all $\delta' \leq \delta$, $\Lambda(\hat{x}) - \Lambda(\hat{x} + \delta') \geq -\gamma/4$. Take now $\delta_2 \leq \min\{\delta, \gamma/4, \varepsilon\}$ so that $\hat{x} + \delta_2 \in [\overline{x}, \overline{x} + \varepsilon]$. By observing that, for $G \in \mathcal{C}_\Lambda$ with $d(F, G) < \delta_2$ we have

$$G(\hat{x}+\delta_2) - \Lambda(\hat{x}+\delta_2) \geq F(\hat{x}) - \delta_2 - \Lambda(\hat{x}+\delta_2) \geq F(\hat{x}) - \Lambda(\hat{x}) - \gamma/4 - \delta_2 \geq \gamma/2$$

we obtain

$$\Lambda VaR(G) \geq -\hat{x} - \delta_2 \geq \Lambda VaR(F) - \varepsilon. \tag{1.7}$$

By taking $\delta := \min\{\delta_1, \delta_2\}$ and combining (1.6) and (1.7), we have that

$$\forall G \in \mathcal{C}_\Lambda \text{ with } d(F,G) < \delta \Longrightarrow |\Lambda VaR(F) - \Lambda VaR(G)| < \varepsilon$$

as desired. □

The Λ function adds flexibility to ΛVaR, however, when robustness is required, ΛVaR should be constructed as suggested by the set $\mathcal{C}_\Lambda$. The Λ function has to be chosen continuous and, on any interval, it cannot coincide with any distribution F under consideration. We refer to Example 11 to show how this condition can be guaranteed given a set of normal distributions of P&Ls.

1.4 Elicitability

The importance of this property from a financial risk management perspective has been highlighted by [Embrechts and Hofert,2014] as a consequence of the surprising results obtained by [Gneiting2011] and [Ziegel, 2014]. Indeed, [Embrechts and Hofert,2014] pointed out that the elicitability allows the assessment and the comparison of risk measure forecasting estimations and a straightforward backtesting.

The term *elicitable* has been introduced by [Lambert et al.,2008] but the general notion dates back to the pioneering work of [Osband,1985]. In accordance with some parts of the literature, we introduce the notation $T : \mathcal{M} \subseteq \mathcal{D} \to 2^{\mathbb{R}}$ to describe a set-valued statistical functional. Let us denote with $S(x,y)$ the realized forecasting error between the ex-ante prediction $x \in \mathbb{R}$ and the ex-post observation $y \in \mathbb{R}$, where S is a function $S : \mathbb{R} \times \mathbb{R} \to [0, +\infty)$ called "scoring" or "loss". According to [Gneiting2011] a scoring function S is consistent for the functional T if

$$\mathbb{E}_F[S(t,Y)] \leq \mathbb{E}_F[S(x,Y)] \tag{1.8}$$

for all F in $\mathcal{M}$, all $t \in T(F)$ and all $x \in \mathbb{R}$. It is strictly consistent if it is consistent and equality of the expectations implies that $x \in T(F)$.

Definition 6. *[Gneiting2011] A set-valued statistical functional $T : \mathcal{M} \to 2^{\mathbb{R}}$ is elicitable if there exists a scoring function S that is strictly consistent for it.*

[Bellini and Bignozzi, 2015] have recently proposed a slightly different definition of elicitability. They consider only single-valued statistical functionals as a natural requirement in financial applications. In addition, they adopt additional properties for the scoring function. We also consider single-valued statistical functionals but without imposing any restriction on the scoring function.

Definition 7. *A statistical functional $T : \mathcal{M} \to \mathbb{R}$ is elicitable if there exists a scoring function S such that*

$$T(F) = \arg\min_x E_F[S(x, Y)] \quad \forall F \in \mathcal{M}. \tag{1.9}$$

Definition 6 restricted to the case of single-valued statistical functional is equivalent to Definition 7 when the minimum is unique. The statistical functional associated to a risk measure is the map $T : \mathcal{M} \to \overline{\mathbb{R}}$ such that $T(F) = -\rho(F)$ for any distribution F. We adopt this sign convention in accordance with part of the literature. We say that a risk measure is elicitable if the associated statistical functional T is elicitable. In the following we will restrict to $\mathcal{M} \subseteq \mathcal{D}_1$ in order to have a finite expectation of the considered scoring functions.

The statistical functional associated to VaR, $T(F) := q_\lambda^+(F)$, is elicitable on the following set:

$$\mathcal{M}_\lambda := \{F \in \mathcal{D}_1 : F \text{ strictly increasing } \} \subseteq \mathcal{C}_\lambda$$

with $\mathcal{C}_\lambda$ as in (1.4), and with the following scoring function [Gneiting2011]:

$$S(x, y) = \lambda(y - x)^+ + (1 - \lambda)(y - x)^-. \tag{1.10}$$

Let us denote with $T_\Lambda : \mathcal{D} \to \mathbb{R}$ the statistical functional associated to ΛVaR such that:

$$T_\Lambda(F) = -\Lambda VaR(F) \tag{1.11}$$

and consider the set $\mathcal{M}_\Lambda \subseteq \mathcal{D}_1$ defined as follows:

$$\mathcal{M}_\Lambda = \{F \in \mathcal{D}_1 : \exists\, \bar{x} \text{ s.t. } \forall x < \bar{x},\ F(x) < \Lambda(x) \text{ and } \forall x > \bar{x},\ F(x) > \Lambda(x)\}. \tag{1.12}$$

Once again this set coincides with $\mathcal{M}_\lambda$ when $\Lambda \equiv \lambda$. In [Bellini and Bignozzi, 2015] it has been shown that ΛVaR is elicitable under a stronger definition of elicitability and for the special case of Λ continuous and decreasing. In the next theorem we prove that ΛVaR is elicitable using the general Definition 7 and under less restrictive conditions on Λ. Specifically, we show that ΛVaR is elicitable on the particular class of distribution $\mathcal{M}_\Lambda$ in (1.12) depending on Λ.

Theorem 8. *For any monotone and right continuous function $\Lambda : \mathbb{R} \to [\lambda^m, \lambda^M]$, with $0 < \lambda^m \leq \lambda^M < 1$, the statistical functional $T_\Lambda : \mathcal{D} \to \mathbb{R}$ defined in* (1.11) *is elicitable on the set $\mathcal{M}_\Lambda \subseteq \mathcal{D}_1$ defined in* (1.12) *with a loss function given by*

$$S(x, y) = (y - x)^- - \int_y^x \Lambda(t)dt. \tag{1.13}$$

Proof. We need to prove that

$$T(F) = \arg\min_x \int_{\mathbb{R}} S(x, y)dF(y).$$

In order to find a global minimum we first calculate the left and right derivatives of $\int_{\mathbb{R}} S(x, y)dF(y)$. Applying dominated convergence theorem we obtain:

$$
\begin{aligned}
\frac{\partial^-}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) &= \frac{\partial^-}{\partial x}\int_{\mathbb{R}} \Big((y-x)^- - \int_y^x \Lambda(t)dt\Big)dF(y)\\
&= \int_{\mathbb{R}} \Big(\frac{\partial^-}{\partial x}(y-x)^- - \frac{\partial^-}{\partial x}\int_y^x \Lambda(t)dt\Big)dF(y)\\
&= \int_{\mathbb{R}} \Big(\mathbf{1}_{(y<x)} - \Lambda(x^-)\Big)dF(y)\\
&= \lim_{t\uparrow x} F(t) - \Lambda(x^-) = F(x^-) - \Lambda(x^-).
\end{aligned}
$$

Analogously for the right derivative

$$
\begin{aligned}
\frac{\partial^+}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) &= \int_{\mathbb{R}} \Big(\mathbf{1}_{(y\le x)} - \Lambda(x)\Big)dF(y)\\
&= F(x) - \Lambda(x).
\end{aligned}
$$

Observe now that $x^* = \inf\{x \in \mathbb{R} : F(x) > \Lambda(x)\}$, that is the statistical functional associated to ΛVaR, satisfies, for every $F \in \mathcal{M}_\Lambda$,

$$
\begin{aligned}
&\forall x < x^* \quad F(x) < \Lambda(x), \quad F(x^-) \le \Lambda(x^-)\ ;\\
&\forall x > x^* \quad F(x) > \Lambda(x), \quad F(x^-) \ge \Lambda(x^-)\ ;
\end{aligned}
\tag{1.14}
$$

from which we deduce

$$
\begin{aligned}
&\forall x < x^* \quad \frac{\partial^-}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) \le 0, \quad \frac{\partial^+}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) < 0;\\
&\forall x > x^* \quad \frac{\partial^-}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) \ge 0, \quad \frac{\partial^+}{\partial x}\int_{\mathbb{R}} S(x,y)dF(y) > 0.
\end{aligned}
\tag{1.15}
$$

This implies that x^* is a local minimum. By showing that there are no other local minima we obtain that x^* is the unique global minimum. Take first $x < x^*$. Observe that, by applying dominated convergence theorem, $I(x) := \int_{\mathbb{R}} S(x,y)dF(y)$ is a continuous function. Moreover, I is not constant on any interval in $(-\infty, x^*]$ since, from (1.15), we have $\frac{\partial^+}{\partial x} I < 0$. Since I is continuous and, from (1.15), the left and right derivatives are non-positive, we have that any sequence converging to x^- is decreasing. Analogously, any sequence converging to x^+ is increasing. In other words, there exists $\delta > 0$ such that, $I(x_1) > I(x) > I(x_2)$ for all $x - \delta < x_1 < x < x_2 < x + \delta$. Thus

x is not a local minimum. The case $x > x^*$ is analogous. We can conclude that ΛVaR is elicitable on the class of probability measures $\mathcal{M}_\Lambda$ defined in (1.12). $\square$

Remark 9. *It is easy to prove that the scoring function in* (1.13) *can be rewritten as follows:*

$$S(x,y) = \frac{\int_y^x \Lambda(t)dt}{x-y}(y-x)^+ + \left(1 - \frac{\int_y^x \Lambda(t)dt}{x-y}\right)(y-x)^-. \tag{1.16}$$

if $x \neq y$ *and* $S(x,x) = 0$*. It is evident the similarity with the scoring function of* VaR *in* (1.10)*. Moreover, note that If* Λ *is non-increasing obviously* (1.12) *is satisfied by every* $F \in \mathcal{D}_1$ *increasing so that we recover the result of [Bellini and Bignozzi, 2015].*

In general, the elicitability of ΛVaR using the scoring function (1.13) requires that Λ is crossed only once by any possible F at the level $\bar{x} = -\Lambda VaR(F)$ as shown in (1.12).

Remark 10. ΛVaR *with a decreasing function* Λ *is elicitable on the set of all the distributions. In this case,* $\mathcal{M}_\Lambda \equiv \mathcal{D}_1$*, since* F *is non-decreasing and the derivatives of* Λ *are negative. When* Λ *is non-increasing,* ΛVaR *is elicitable on the set of increasing distribution functions.*

If we additionally require continuity of Λ we observe that $\mathcal{M}_\Lambda \subseteq \mathcal{C}_\Lambda$ where $\mathcal{C}_\Lambda$ is defined in (1.5). This implies that the set of distributions where ΛVaR is elicitable guarantees also that ΛVaR is robust. Hereafter, we provide an example of a construction of ΛVaR with non-decreasing Λ that is elicitable and robust given a set of normal distributions of P&Ls.

Example 11. *Denote by* $\Phi(x)$ *the distribution function of a standard normal distribution. Let* $\mathcal{M} := \{\Phi(\frac{x-\mu_i}{\sigma_i})\}_{i \in I}$ *for some collection* I *such that* $\overline{\mu} :=$

$\sup \mu_i < \infty$ *and* $\underline{\sigma} := \inf \sigma_i > 0$. *Set* $\mu > \overline{\mu}$, $0 < \sigma < \underline{\sigma}$ *and define*

$$\Lambda(x) := \begin{cases} \lambda^m & x \leq x^m \\ \Phi\left(\dfrac{x-\mu}{\sigma}\right) & x^m \leq x < x^M \\ \lambda^M & x \geq x^M. \end{cases}$$

If $x^m \leq x^M$ *are such that* $0 < \lambda_m \leq \Phi(\frac{x^m-\mu}{\sigma})$ *and* $\Phi(\frac{x^M-\mu}{\sigma}) \leq \lambda^M$ *then* Λ *is non-decreasing and continuous. Moreover, from Theorem 8,* ΛVaR *is elicitable on* $\mathcal{M}$.

In order to have an elicitable ΛVaR with the scoring function (1.13) we need to build the Λ function under a certain condition that depends on the set of the P&L distributions. In particular, the scoring function (1.13) guarantees the elicitability of ΛVaR with non-decreasing Λ only in the class of probability measures $\mathcal{M}_\Lambda$ in (1.12) as shown by the following counterexample.

Example 12. *Let* $0 < \varepsilon < 0.5\%$. *Let* $\Lambda(x)$ *and* $F(x)$ *as follows*

$$F(x) = \begin{cases} 0 & x < -100 \\ 1.5\% & -100 \leq x < 4 \\ 1 & x \geq 4 \end{cases} \qquad \Lambda(x) = \varepsilon + \begin{cases} 0 & x < -101 \\ (x+101)/100 & -101 \leq x < -99 \\ 2\% & x \geq -99. \end{cases}$$

$F(x)$ is the cumulative distribution function of a random variable Y with distribution: $Y = -100$ with probability $p = 1.5\%$ and $Y = 4$ with probability $1 - p = 98.5\%$.

It is easy to compute that the statistical functional associated to ΛVaR is $T_\Lambda(F) = -100$. If ΛVaR is elicitable T_Λ should be the minimizer of

$$g(x) := \mathbb{E}[S(x,Y)] = S(x,-100)\frac{1.5}{100} + S(x,4)\frac{98.5}{100}.$$

Since S for ΛVaR is defined as in (1.13), *we need compute the primitive for*

Λ that is given by

$$\Psi(t) = \int \Lambda(t) = \varepsilon t + \begin{cases} 0 & t < -101 \\ \dfrac{(t^2/2 + 101t)}{100} & -101 \leq t < -99 \\ \dfrac{2}{100}t & t \geq -99. \end{cases}$$

Hence, $\Psi(-100) = -51 - 100\varepsilon$ and $\Psi(4) = 8/100 + 4\varepsilon$, thus, we have $S(x, -100) = (-100 - x)^- - \Psi(x) - 51 - 100\varepsilon$ and $S(x, 4) = (4 - x)^- - \Psi(x) + 8/100 + 4\varepsilon$ and

$$g(x) = -\Psi(x) + (-100 - x)^- \frac{1.5}{100} + (4 - x)^- \frac{98.5}{100} + c$$

where $c = (-51 - 100\varepsilon) \cdot 1.5\% + (0.08 + 4\varepsilon) \cdot 98.5\%$. Observe now that ΛVaR is not the global minimum, since $g(-100) > g(4)$. Indeed:

$$g(-100) - g(4) = -\Psi(-100) + \Psi(4) - 104 \cdot \frac{1.5}{100} = 51 + \frac{8}{100} - 104 \cdot \frac{1.5}{100} > 0.$$

We have shown that the scoring function in (1.13) guarantees elicitability of ΛVaR only on the set of distributions $\mathcal{M}_\Lambda$ in (1.12). Whether there exists another scoring function that guarantees the elicitability of ΛVaR on a larger class of distributions is an interesting question which might be object of further studies. We conclude this Section by discussing some insights on this problem and the difficulties that might arise for such an extension. In particular we investigate a necessary condition for elicitability, namely, the convex level sets property [Osband,1985].

Definition 13. *If $\mathcal{M} \subseteq \mathcal{D}$ is convex we say that T has convex level sets if, for any $\gamma \in \mathbb{R}$, the level sets*

$$\{T = \gamma\} := \{F \in \mathcal{M} : T(F) = \gamma\}$$

are convex, i.e. for any $\alpha \in [0, 1]$ and $F_1, F_2 \in \mathcal{M}$

$$T(F_1) = T(F_2) = \gamma \Rightarrow T(\alpha F_1 + (1 - \alpha)F_2) = \gamma.$$

Proposition 14. *[Osband,1985] If a statistical functional $T : \mathcal{M} \subseteq \mathcal{D} \to \mathbb{R}$ is elicitable, then T has convex level sets.*

[Gneiting2011] showed that ES does not satisfy this necessary condition, as a consequence, ES is not elicitable.

We have shown in Theorem 8 that ΛVaR is elicitable in $\mathcal{M}_\Lambda$, hence, it also has convex level sets in this class of distributions. The following example shows that, in general, ΛVaR might not satisfy this condition on a larger set of distributions and, thus, neither elicitability.

Example 15. *Fix $0 < \varepsilon < \frac{1}{2}$ and $\lambda^M < 1$. Consider*

$$F_1(x) := \sum_{k=1}^{\infty} \frac{1}{2^k} \mathbf{1}_{[\frac{1}{k+1}, \frac{1}{k})}(x) + \varepsilon \mathbf{1}_{[0,1)} + \mathbf{1}_{[1,\infty)}$$

and

$$F_2(x) := F_1(x) + \sum_{k=1}^{\infty} (-1)^k \frac{1}{10^k} \mathbf{1}_{[\frac{1}{k+1}, \frac{1}{k})}(x).$$

As a function Λ take $\Lambda := \varepsilon \mathbf{1}_{(-\infty,0)} + \frac{1}{2}(F_1 + F_2)\mathbf{1}_{[0,1)} + \lambda^M \mathbf{1}_{[1,\infty)}$. Observe that $\forall k \in \mathbb{N}$ $F_1(\frac{1}{2k}) > \Lambda(\frac{1}{2k})$ and $F_2(\frac{1}{2k+1}) > \Lambda(\frac{1}{2k+1})$. Moreover, $0 = F_1(x) = F_2(x) < \Lambda(x)$ for all $x < 0$. This implies $\Lambda VaR(F_1) = \Lambda Var(F_2) = 0$. Nevertheless, since $\Lambda(x) = \lambda^M < 1$ for $x \geq 1$, we have $\Lambda Var(\frac{1}{2}F_1 + \frac{1}{2}F_2) = -1$, from which the convex level set property fails.

A positive answer for the convex level sets property is given by the choice of a particular class of Λ for which the condition is satisfied on the set of increasing distribution functions.

Lemma 16. *If Λ is non-decreasing and piecewise constant with a finite number of jumps, then ΛVaR has convex level sets on the set of increasing distribution functions.*

Proof. We first observe that, in general, $T_\Lambda(F_i) = \gamma$ for $i = 1, 2$ implies $T_\Lambda(\alpha F_1 + (1-\alpha)F_2) \geq \gamma$ for every $\alpha \in [0,1]$ and $F_1, F_2 \in \mathcal{D}$. To this end, we prove that $\inf\{x : \alpha F_1(x) + (1-\alpha)F_2(x) > \Lambda(x)\} \geq \gamma$, with

$\gamma = T_\Lambda(F_i) := \inf\{x : F_i(x) > \Lambda(x)\}$ for $i = 1, 2$. Note that by definition of $T_\Lambda(F_i)$ for $i = 1, 2$, we have $F_i(x) \leq \Lambda(x)$ for every $x \leq \gamma$. We thus get, for an arbitrary $0 \leq \alpha \leq 1$, $\alpha F_1(x) + (1-\alpha)F_2(x) \leq \Lambda(x)$ for every $x \leq \gamma$ from which $T_\Lambda(\alpha F_1 + (1-\alpha)F_2) \geq \gamma$.

For the converse inequality observe that there exists $\varepsilon > 0$ such that Λ is constant on $[\gamma, \gamma+\varepsilon)$. Since $\gamma = \inf\{x : F_i(x) > \Lambda(x)\}$ and F_i is non-decreasing, for $i = 1, 2$, then $\alpha F_1(x) + (1-\alpha)F_2(x) > \Lambda(x)$ for every $x \in (\gamma, \gamma+\varepsilon)$ from which $T_\Lambda(\alpha F_1 + (1-\alpha)F_2) \leq \gamma$. □

In conclusion, we have observed that extending the class of distributions for which the convex level sets property holds depends heavily on the specific choice of Λ and hence it seems to necessitate a case-by-case study.

1.5 Consistency

In this section we refer to the notion of *consistency* recently studied by [Davis 2016]. Davis recognized the importance of the elicitability property in the backtesting context of risk measures, but he argued that the problem can be better addressed from a different perspective. The motivation of Davis' study relates to the difficulties of predicting the "true" distribution F of portfolio financial returns. Suppose indeed you are given the information up to time $k-1$, at time k only one realization occurs and so there is not enough information to claim if the prediction of F was correct or not. Thus, Davis introduces the notion of consistency of a risk estimator that is based on the daily comparison between the realization of the risk estimator and the realized outcome, but without consideration how the predictions were arrived at. Hence, the fundamental difference with the elicitability property is that the assumption on the model generating the conditional distribution of the portfolio returns can change at any time and one should just check if the prediction is performing well or not [Davis 2016].

In this section we adopt the framework of Davis. Namely, we fix $(\Omega, \mathcal{F}, \{\mathcal{F}_k\}_{k\in\mathbb{N}})$ where $\Omega = \prod_{k=1}^{\infty} \mathbb{R}_{(k)}$ is the canonical space for a real-valued data process $Y = \{Y_k\}_{k\in\mathbb{N}}$; $\mathcal{F}$ is the product sigma-algebra generated by the Borel sigma-algebra in each copy of $\mathbb{R}$ (denoted by $\mathbb{R}_{(k)}$); $\{\mathcal{F}_k\}_{k\in\mathbb{N}}$ is the natural filtration of the process Y and $\mathcal{F}_0$ the trivial sigma-algebra. The class of possible models, for this data process, is represented by a collection $\mathcal{P}$ of probability measures denoted by $\mathcal{P} := \{\mathbb{P}^\alpha, \alpha \in \mathfrak{A}\}$, where $\mathfrak{A}$ is an arbitrary index set. We denote with $\mathbb{E}^\alpha$ the expectation with respect to $\mathbb{P}^\alpha$. For every $\mathbb{P}^\alpha$ it is possible to define, for each $k \geq 1$, the conditional distribution of the random variable Y_k given $\mathcal{F}_{k-1}$, as a map $F_k^\alpha : \mathbb{R} \times \Omega \mapsto [0,1]$ satisfying: for $\mathbb{P}^\alpha$-a.e. ω, $F_k^\alpha(\cdot, \omega)$ is a distribution function, and for every $x \in \mathbb{R}$, $F_k^\alpha(x) = \mathbb{P}^\alpha(Y_k \leq x|\mathcal{F}_{k-1})$ $\mathbb{P}^\alpha$-a.s.

Definition 17. *[Davis 2016] Let $\mathfrak{B}(\mathcal{P})$ be a set of strictly increasing predictable processes $b = \{b_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} b_n = \infty$ $\mathbb{P}^\alpha$-a.s. for every $\alpha \in \mathfrak{A}$, and $l : \mathbb{R}^2 \to \mathbb{R}$ a calibration function, that is a measurable function such that $\mathbb{E}^\alpha[l(T(F_k^\alpha), Y_k)|\mathcal{F}_{k-1}] = 0$ for all $\mathbb{P}^\alpha \in \mathcal{P}$. A risk measure ρ is $(l, b, \mathcal{P})$-consistent if the associated statistical functional T satisfies*

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} l(T(F_k^\alpha), Y_k) = 0 \quad \mathbb{P}^\alpha\text{-a.s. } \forall \mathbb{P}^\alpha \in \mathcal{P}. \tag{1.17}$$

Denote by $\mathfrak{P}$ the set of all probability measures and define:

$$\mathcal{P}^0 = \{\mathbb{P}^\alpha \in \mathfrak{P} : \forall k\ F_k^\alpha(x, \omega) \text{ is continuous in } x \text{ for } \mathbb{P}^\alpha\text{-almost all } \omega \in \Omega\}.$$

[Davis 2016] showed that VaR satisfies this consistency property for a large class of processes $\mathcal{B}(\mathcal{P})$ and for the large class of data models $\mathcal{P}^0$ with the following calibration function:

$$l(x, y) = \lambda - \mathbf{1}_{(y\leq x)}.$$

The statistical functional associated to ΛVaR is given by (1.11), hence we define for every k and $\alpha \in \mathfrak{A}$:

$$T_\Lambda(F_k^\alpha) := \inf\{x \mid F_k^\alpha(x) > \Lambda(x)\}.$$

Notice that $\{T_\Lambda(F_k^\alpha)\}_{k\in\mathbb{N}}$ and $\{\Lambda(T_\Lambda(F_k^\alpha))\}_{k\in\mathbb{N}}$ are predictable process, as shown in the following lemma.

Lemma 18. *For every $k \geq 1$, $T_\Lambda(F_k^\alpha)$ and $\Lambda(T_\Lambda(F_k^\alpha))$ are $\mathcal{F}_{k-1}$-measurable random variables.*

Proof. Fix a probability $\mathbb{P}^\alpha$ with $\alpha \in \mathfrak{A}$. Notice first that for any $y \in \mathbb{R}$, for $\mathbb{P}^\alpha$ a.e. ω, we have

$$\begin{aligned} T_\Lambda(F_k^\alpha) \geq y &\iff F_k^\alpha(x) \leq \Lambda(x) \qquad \forall x \leq y \\ &\iff F_k^\alpha(q) \leq \Lambda(q) \quad \forall q \in \mathbb{Q},\ q \leq y \end{aligned}$$

where the last equivalence follows from the right-continuity of F_k^α and Λ. We therefore have

$$\{\omega \mid T_\Lambda(F_k^\alpha) \geq y\} = \bigcap_{q\in\mathbb{Q}\cap(-\infty,y]} \{\omega \mid F_k^\alpha(q) \leq \Lambda(q)\} \in \mathcal{F}_{k-1}$$

from which $T_\Lambda(F_k^\alpha)$ is an $\mathcal{F}_{k-1}$-measurable random variable.

$\Lambda(T_\Lambda(F_k^\alpha))$ is also $\mathcal{F}_{k-1}$-measurable: since Λ is right-continuous $\Lambda(x) \geq y$ iff $x \geq \Lambda^-(y)$ where $\Lambda^-(y) := \inf\{x \in \mathbb{R} \mid \Lambda(x) \geq y\}$ is the generalized inverse [Embrechts and Hofert, 2013, Proposition 1] and thus

$$\{\omega \mid \Lambda(T_\Lambda(F_k^\alpha)) \geq y\} = \{\omega \mid T_\Lambda(F_k^\alpha) \geq \Lambda^-(y)\} \in \mathcal{F}_{k-1}.$$

□

By following the methodology suggested by [Davis 2016], we are able to show that ΛVaR is consistent for the large class of data models $\mathcal{P}^0$, as shown in the following theorem.

Theorem 19. *For each $\mathbb{P}^\alpha \in \mathcal{P}^0$,*

$$\frac{1}{n}\sum_{k=1}^{n} \Lambda(T_\Lambda(F_k^\alpha)) - \mathbf{1}_{(Y_k \leq T_\Lambda(F_k^\alpha))} \to 0 \quad \mathbb{P}^\alpha\text{-}a.s. \tag{1.18}$$

Thus, ΛVaR is $(l, n, \mathcal{P}^0)$-consistent with

$$l(x,y) = \Lambda(x) - \mathbf{1}_{(y \le x)}. \tag{1.19}$$

Before giving the proof of the theorem we show the following lemma.

Lemma 20. *For each $\mathbb{P}^\alpha \in \mathcal{P}^0$,*

$$\mathbb{E}^\alpha \left[\mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))} \mid \mathcal{F}_{k-1} \right] = \Lambda(T_\Lambda(F_k^\alpha)), \qquad \mathbb{P}^\alpha\text{-}a.s.$$

Proof. Fix $\mathbb{P}^\alpha \in \mathcal{P}^0$. Since there is no confusion, for ease of notation, we omit the dependence on α. Observe that $U_k := F_k(Y_k)$ is uniformly distributed and

$$Y_k \le T_\Lambda(F_k) \Longleftrightarrow U_k \le F_k(T_\Lambda(F_k)) = \Lambda(T_\Lambda(F_k)).$$

Note now that U_k is independent of $\mathcal{F}_{k-1}$ since, from the continuity of F_k, $\mathbb{P}(U_k \le u_k \mid \mathcal{F}_{k-1}) = \mathbb{P}(Y_k \le F_k^-(u_k) \mid \mathcal{F}_{k-1}) = F_k(F_k^-(u_k)) = u_k = \mathbb{P}(U_k \le u_k)$ (where F_k^- denotes the generalized inverse of F_k). Since $\Lambda(T_\Lambda(F_k))$ is $\mathcal{F}_{k-1}$-measurable from Lemma 18, we can compute the desired conditional expectation through the application of the freezing lemma [Williams,1991, Section 9.10]. Namely, define $h(x,y) := \mathbf{1}_{\{y \le x\}}$ and let $\hat{h}(x) := \mathbb{E}[\mathbf{1}_{\{U_k \le x\}}] = x$. Since h is a bounded Borel-measurable function and U_k is independent of $\mathcal{F}_{k-1}$, then

$$\mathbb{E}\left[\mathbf{1}_{(Y_k \le T_\Lambda(F_k))} \mid \mathcal{F}_{k-1} \right] = \hat{h}(\Lambda(T_\Lambda(F_k))) = \Lambda(T_\Lambda(F_k))$$

where equalities are intended in the $\mathbb{P}$-a.s. sense. □

Proof of Theorem 19. Define:

$$Z_k := \Lambda(T_\Lambda(F_k^\alpha)) - \mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))},$$

$$S_n := \sum_{k=1}^n Z_k, \qquad Q_n := \sum_{k=1}^n (Z_k)^2, \qquad \langle S \rangle_n := \sum_{k=1}^n \mathbb{E}^\alpha[(Z_k)^2 \mid \mathcal{F}_{k-1}].$$

From Lemma 20, S_n is a martingale since $\mathbb{E}^\alpha[S_n - S_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}^\alpha[Z_n \mid \mathcal{F}_{n-1}] = 0$. We now compute $(Z_k)^2$, we use the shorthand $W := \Lambda(T_\Lambda(F_k^\alpha))$.

$$\begin{aligned}(Z_k)^2 &= \mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))} + W^2 - 2W\mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))} \\ &= W^2\mathbf{1}_{(Y_k > T_\Lambda(F_k^\alpha))} + (1-W)^2\mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))}.\end{aligned}$$

Note that since $\lambda^m \le W \le \lambda^M$ we obtain $\mathbb{E}^\alpha[(Z_k)^2] \le \max\{(\lambda^M)^2, (1-\lambda^m)^2\} < \infty$ so that S_n is a square integrable martingale. Moreover, observe that,

$$(Z_k)^2 \ge \min\{(\lambda^m)^2, (1-\lambda^M)^2\}. \tag{1.20}$$

Since W is $\mathcal{F}_{k-1}$-measurable, using Lemma 20,

$$\begin{aligned}\mathbb{E}^\alpha[(Z_k)^2 \mid \mathcal{F}_{k-1}] &= W^2\mathbb{E}^\alpha[\mathbf{1}_{(Y_k > T_\Lambda(F_k^\alpha))} \mid \mathcal{F}_{k-1}] + (1-W)^2\mathbb{E}^\alpha[\mathbf{1}_{(Y_k \le T_\Lambda(F_k^\alpha))} \mid \mathcal{F}_{k-1}] \\ &= W^2(1-W) + (1-W)^2 W \\ &= W(1-W).\end{aligned}$$

It follows that

$$\lambda^m(1-\lambda^M) \le \mathbb{E}^\alpha[(Z_k)^2 \mid \mathcal{F}_{k-1}] \le \lambda^M(1-\lambda^m)$$

which firstly implies $\langle S\rangle_n \ge n\lambda^m(1-\lambda^M) \to \infty$, and, secondly, combined with (1.20),

$$\frac{Q_n}{\langle S\rangle_n} \ge \frac{n\min\{(\lambda^m)^2, (1-\lambda^M)^2\}}{n\lambda^M(1-\lambda^m)} = \frac{\min\{(\lambda^m)^2, (1-\lambda^M)^2\}}{\lambda^M(1-\lambda^m)} =: \varepsilon_\alpha > 0.$$

Notice that Z_k is bounded from above by 1 for every $k \in \mathbb{N}$, thus, we have $Q_n \le n$. We can therefore conclude

$$\left|\frac{S_n}{n}\right| \le \left|\frac{S_n}{Q_n}\right| = \frac{\langle S\rangle_n}{Q_n}\left|\frac{S_n}{\langle S\rangle_n}\right| \le \frac{1}{\varepsilon_\alpha}\left|\frac{S_n}{\langle S\rangle_n}\right| \to 0 \quad \mathbb{P}^\alpha\text{-a.s.}$$

where the last term converges to 0 from Proposition 6.3 in [Davis 2016]. □

As a consequence of the theorem, similarly to what observed by [Davis 2016] for VaR, a risk manager could use the following relative frequency measure

$$\frac{1}{n}\sum_{k=1}^{n}\Lambda(T_\Lambda(F_k^\alpha)) - \mathbf{1}_{(Y_k \leq T_\Lambda(F_k^\alpha))} \tag{1.21}$$

as test statistic in a finite-sample hypothesis test [Corbetta and Peri (2017)]. Obviously ΛVaR is also $(l, b', \mathcal{P}^0)$-consistent with $b'_n = nb_n$ and $b = \{b_n\}_n \in \mathfrak{B}(\mathcal{P})$.

Therefore, the consistency of ΛVaR, as the quantile forecasting, can be obtained under essentially no conditions on the mechanism generating the data. This is not the case of the estimates of the statistical functional T_m associated to the conditional mean (such as ES) and defined as follows:

$$T_m(F_k^\alpha) := \int_{\mathbb{R}} xF_k^\alpha(dx).$$

Indeed, [Davis 2016] showed that $T_m(F_k^\alpha)$ satisfies the condition (1.17) with $l(x,y) = x - y$, $Q_n = \sum_{k=1}^n Z_k^2$, where $Z_k := Y_k - T_m(F_k^\alpha)$, and, remarkably, $\mathcal{P}^1 \in \mathfrak{P}$ is the set of probability measures such that:

i) for any k, $Y_k \in L^2(\mathbb{P}^\alpha)$,

ii) $\lim_{n\to\infty}\langle S\rangle_n = \infty$ $\mathbb{P}^\alpha$-a.s., with $\langle S\rangle_n := \sum_{k=1}^n \mathbb{E}[Z_k^2|\mathcal{F}_{k-1}]$,

iii) there exists $\varepsilon_\alpha > 0$ such that $\frac{Q_n}{\langle S\rangle_n} > \varepsilon_\alpha$ for large n, $\mathbb{P}^\alpha$-a.s.

In general, the validity of conditions i), ii), iii) might be difficult to check. In addition, the process Q_n is not predictable, thus, it is not possible to conclude that statistical functionals that depends on the mean (such as ES) satisfy the consistency property as in Definition 17 using this methodology [2]. Hence, in line with the elicitability framework, verifying the accuracy of mean-based estimates is definitely more problematic than the same problem for quantile-based forecasts. For the case of ΛVaR this is possible and all the conditions are satisfied so that the methodology can be successfully applied.

[2] We thank an anonymous referee that pointed out this issue.

1.6 Conclusions

We have shown that ΛVaR, satisfies robustness and elicitability in particular classes of distributions. Robustness requires that the Λ function is continuous and does not coincide with the distribution F on any interval. Elicitability requires a bit more, that is, Λ is crossed only once by any possible F. We have also proposed an example of construction of an elicitable and robust ΛVaR given a set of normal distributions. In addition, we have shown that ΛVaR satisfies the consistency property without any conditions on the mechanism generating data, allowing a straightforward back-testing.

After the recent financial crisis, the [Basel Committee, 2013] has suggested that banks should abandon VaR in favour of the ES as a standard tool for risk management since ES is able to overcome two main shortcomings of VaR: lack of convexity on random variables and insensitivity with respect to tail behaviour. However, ES has also some issues. Specifically, ES is not robust, or only for small degrees when a stronger definition of robustness is required, and it is not elicitable. Recently, [Acerbi and Székely,2014] showed that the elicitability of ES can be reached jointly with VaR [Di Persio and Frigo (2016), see also]for an extended result]FZ14. In addition, verifying the consistency property for ES is more problematic. Moreover, a recent study by [Koch-Medina and Munari,2016] pointed out that not all the aspects of ES are well understood. For instance, for positions with a high probability of losses but also high expected gains in the tails, ES does not necessarily perform better than VaR from a liability holders' perspective. Other risk measures which consider the magnitude of losses beyond ES are the expectiles, recently studied by [Bellini and Di Bernardino, 2015].

In any case, the issue of capturing tail risk remains crucial and cannot be accomplished through VaR. The new risk measure, ΛVaR, may solve this issue since it is able to discriminate the risk among distributions with the same quantile but different tail behaviour and shares with VaR other important properties such as quasi-convexity. On the other hand, ΛVaR lacks subadditivity and the flexibility introduced by the Λ function requires additional criteria for determining its upper and lower bound. However, we think that ΛVaR may be considered as an alternative risk measure valuable for further studies.

Bibliography

[Acerbi and Székely,2014] Acerbi, C. and Székely, B., Backtesting Expected Shortfall. In *Risk*, 2014.

[Artzner et al.,1999] Artzner, P., Delbaen, F., Eber, J.-M. and Heath., D., Coherent measures of risk. In *Mathematical Finance*, 1999, **9**, 203–228.

[Basel Committee, 2013] Basel Committee on Banking Supervision, Fundamental review of the trading book, second consultative document. Bank for International Settlements, 2013.

[Bellini and Bignozzi, 2015] Bellini, F. and Bignozzi, V., Elicitable risk measures. In *Quantitative Finance*, 2015, **15**.

[Bellini and Di Bernardino, 2015] Bellini, F. and Di Bernardino, E., Risk management with expectiles. In *The European Journal of Finance*, 2015, forthcoming.

[Corbetta and Peri,2016] Corbetta, J. and Peri, I., Backtesting Lambda Value at Risk. Working paper, 2016. Available at arxiv.org/abs/1602.07599.

[Contet al.,2010] Cont, R., Deguest, R. and Scandolo, G., Robustness and sensitivity analysis of risk measurement procedures. In *Quantitative Finance*, 2010, **10**, 593–606.

[Davis 2016] Davis, M.H.A., Verification of internal risk measure estimates. In *Statistics & Risk Modeling with Applications in Finance and Insurance*, 2016, **33**, 67–93.

[Drapeau and Kupper,2012] Drapeau, S. and Kupper, M., Risk Preferences and their Robust Representation. In *Mathematics of Operations Research*, 2012, 28–62.

[Embrechts and Hofert, 2013] Embrechts, P. and Hofert, M., A note on generalized inverses. In *Mathematical Methods of Operations Research*, 2013, **77**, 423–432.

[Embrechts and Hofert,2014] Embrechts, P. and Hofert, M., Statistics and Quantitative Risk Management for Banking and Insurance. In *Annual Review of Statistics and Its Application*, 2014, **1**, 493–514.

[Embrechts et al.2014] Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A., An Academic Response to Basel 3.5. In *Risks*, 2014, **2**, 25–48.

[Fissler and Ziegel2016] Fissler, T. and Ziegel, J., Higher order elicitability and Osbands principle. In *The Annals of Statistics* , 2016, **44**, 1680–1707.

[Frittelli et al.2014] Frittelli, M., Maggis, M. and Peri, I., Risk measures on P(R) and Value at Risk with Probability/Loss function. In *Mathematical Finance*, 2014, **24**, 442–463.

[Gneiting2011] Gneiting, T., Making and evaluating point forecasts. In *Journal of the American Statistical Association*, 2011, **106**, 746–762.

[Hampel et al.,1986] Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W., *Robust Statistics: The Approach Based on Influence Functions*, New York, 1986.

[Huber1981] Huber, P., *Robust Statistics*, Wiley, NJ., 1981.

[Koch-Medina and Munari,2016] Koch-Medina, P. and Munari, C., Unexpected shortfalls of Expected Shortfall: Extreme default profiles and regulatory arbitrage. In *Journal of Banking & Finance*, 2016, **62**, 141–151.

[Krätschmer et al.,2012] Krätschmer, V., Schied, A. and Zähle, H., Qualitative and infinitesimal robustness of tail-dependent statistical functionals. In *Journal of Multivariate Analysis*, 2012, **103**, 35–47.

[Krätschmer et al.2014] Krätschmer, V., Schied, A. and Zähle, H., Comparative and qualitative robustness for law invariant risk measures. In *Finance and Stochastics*, 2014, **18**, 271–295.

[Lambert et al.,2008] Lambert, N., Pennock, D. M. and Shoham, Y., Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, Chicago, Illinois, 8-12 July 2008.

[Osband,1985] Osband, K. H., Providing incentives for better cost forecasting. PhD thesis, University of California, Berkeley, 1985.

[Williams,1991] Williams, D., *Probability with Martingales*, Cambridge University Press, 1991.

[Ziegel, 2014] Ziegel, J., Coherency and elicitability. In *Mathematical Finance*, 2014. Available at doi: 10.1111/mafi.12080.

Chapter 2

Markov Switching Mixture of Multivariate Gaussian Distributions in a Bayesian Framework.

2.1 Introduction

Returns of financial markets tend to change their patters over time in terms of average returns, serial correlation and volatility. Typically, assets returns exhibit a stochastic behavior which is not consistent with a Gaussian distribution, particularly due to the volatility of returns changing over time according to the typical phenomena of the heterskedasticity. In addition, the Jarque-Brera test highlights that the distribution of rerturns is not symmetrical around the expected value (skewness) and with fat tails (kurtosis), so extreme values are more likely than in case of normal distribution.

Markov Switching models are a powerful solution in order to manage the patter of returns described above. In this work in particularly, the attention is focused on Markov switching Mixtures of Normal distributions. According to this type of models, returns are assumed to follow a state-space representation where in any state or regime the normal distribution of returns is described by a different configuration of parameters. In this framework, the unconditional likelihood of returns is a weighted average of the likelihood of any state with weights equal to the state probabilities. This type of unconditional distribution is a powerful way to described returns in a non-normal way, dealing with their non-stationary stochastic process in terms of expected value, auto-correlation and volatility. But on the other hand, it is worth highlighting that the conditional distribution of returns on a single state is still normal and so we can apply all the results that in term of portfolio construction, equilibrium and pricing require the hypothesis of normal returns. This way of describing returns is particular useful for conditional portfolio construction issues and conditional tactical asset allocation strategies.

Considering a 20 years database of daily returns for 11 asset classes, we estimate the Markov switching model both in the univariate and multivariate case. In the former, we study the time variation of returns in term of expected value and volatility. In the latter we highlight the fenomeno of correlation clustering, according to the co-movement of different asset classes tend to change quite dramatically in reaction to shocks affecting financial markets over time.

In this work we follow a bayesian approach in order to estimate the model by combining sample data with extra-sample priors about parameters of the model. Following [Kim and Nelson (1999)], we estimate the model thanks to Markov chain Monte Carlo methods (MCMC).

The present study contributes to the sensitive issue of the estimation of correlations matrix with an high number of assets and to the empirical analysis of the performance of a asset allocation strategy based on Markov Switching regime filtering.

After introducing the notion of Markov Switching models and Gibbs Sampling in Section 2.2, we estimate for any index a 2-State univariate Markov Switching Model and a 2- State multivariate Markov switching model, in order to take correlation switching into account. All estimations are due through Gibbs sampling algorithm. At the end a regime-based trading rule is presented and compared with an unconditional buy-and-hold strategy.

2.2 Markov Switching and Gibbs Sampling

2.2.1 Markov Switching: Definition and Properties

Let us consider the time series $\{r_t\}_{t=1}^T$ satisfying

$$\mathbf{r_t} = \mu_{\mathbf{S_t}} + \varepsilon_{\mathbf{t}},\ \varepsilon_{\mathbf{t}} \sim \mathcal{N}(0, \Sigma_{S_t}) \tag{2.1}$$

where $\mathbf{r_t}, \mu_{\mathbf{S_t}}, \varepsilon_{\mathbf{t}} \in \mathbb{R}^{\mathbf{K}}$, $\Sigma_{S_t} \in \mathbb{R}^{K \times K}$ and K is the number of assets. Returns depend on a shock $\varepsilon_{\mathbf{t}}$ and a latent variable, called state or regime S_t, that affects the mean, the volatility and the correlations of the process. The state S_t is a discrete random variable such that $S_t \in \{1, \dots, N\}$. We assume that it follows a first order N-state Markov chain

$$Pr(S_t = j | S_{t-1} = i, S_{t-2} = k, \dots) = Pr(S_t = j | S_{t-1} = i) = p_{ij} \tag{2.2}$$

with

$$0 \leq p_{ij} \leq 1, \quad \sum_{j=1}^{N} p_{ij} = 1. \tag{2.3}$$

Hence, the future state depends on the past only through the present. The probabilities, mentioned in (2.2) and (2.3), are known as the transition probability and give the probability that the process will switch from the state i to the state j. These quantities are usually collected in a $N \times N$ matrix P, called transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \cdots & \cdots & \cdots\cdots & \\ p_{1N} & p_{2N} & \cdots & p_{NN}. \end{bmatrix} \tag{2.4}$$

The transition probabilities p_{ii}, corresponding to the diagonal element of P, are a measure of persistence of each state. If $p_{ii} > p_{jj}$ then the expected sojourn in the state i is longer than in state j. Following [Kim and Nelson (1999)] we define the expected duration

$$\mathbb{E}[D_i] = \frac{1}{1 - p_{ii}}.$$

It is clear that transition probabilities and the latent variable S_t are included in the parameters space

$$\Theta = \{\theta_i, p_{ij} \forall i, j, S_1 \ldots, S_T\} \tag{2.5}$$

$$\theta_i = \{\mu_{S_i}, \sigma^2_{S_i}\} \quad \forall i = 1, \ldots, N. \tag{2.6}$$

As pointed out in 2.5 and 2.6, two issues arise in the estimation of a Markov Switching model: the estimation of the model parameters and the estimation of the latent variable S_t in each t.

In the particular case of mixture of normal distributions, conditional on the state S_t, the stock returns follow a Gaussian distribution with specific parameters μ and Σ

$$\mathbf{r_t}|S_t \sim \mathcal{N}(\mu_{\mathbf{S_t}}, \Sigma_{S_t})$$

$$f(\mathbf{r_t}|S_t = j) = \frac{1}{\sqrt{2\pi}}|\Sigma_{S_j}| \exp(-(\mathbf{r}_t - \mu_{S_j})'\Sigma_{S_j}(\mathbf{r}_t - \mu_{S_j}))$$

All the conditional density functions of $\mathbf{r_t}$ are collected in a vector $\eta_\mathbf{t}$

$$\eta_\mathbf{t} = \begin{bmatrix} f(\mathbf{r_t}|\mathbf{S_t} = \mathbf{1}; \theta_\mathbf{1}) \\ f(\mathbf{r_t}|\mathbf{S_t} = \mathbf{2}; \theta_\mathbf{2}) \\ \cdots \\ f(\mathbf{r_t}|\mathbf{S_t} = \mathbf{N}; \theta_\mathbf{N}) \end{bmatrix}. \tag{2.7}$$

One important property of Markov chains is irreducibility. Roughly speaking the irreducibility allows each state to repeat over time.

Definition 21. *A N-state Markov chain is reducible if there exists a way to label the states such that the transition probabilities can be written in the form*

$$P = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where $B \in \mathbb{R}^{H \times H}$ with $1 \leq H \leq N$. A Markov-chain is irreducible if it is not reducible.

For example, if $N = 2$ the Markov chain is irreducible if $p_{ii} < 1 \quad \forall\, i$. In Finance one possible choice is to identify two states, corresponding to bull and bear market. It is well know that both these situation are not permanent, thus irreducible Markov chains are used.

2.2.2 Ergodic Probabilities

We know from equations 2.3 and 2.4 that every column of the transition matrix P sums to unit

$$P'\mathbf{1} = \mathbf{1}$$

In other words, the unit is an eigenvalue of P because a matrix and its transpose share the same eigenvalues. The normalized eigenvector π, associated with the unit eigenvalue, is called the vector of ergodic probability.

$$P\pi = \pi$$

It can be shown (see [Hamilton 1989]) that

$$\lim_{m\to\infty} P^m = \pi\mathbf{1}' \tag{2.8}$$

and that equation 2.8 implies that the long-run forecast of a Markov chain is independent of the current state and is equal to π. Moreover, the vector of the ergodic probability can be interpreted as the unconditional probability of each different state:

$$\pi_i = Pr(S_{t+m} = i; \theta), \quad m \to \infty.$$

In the particular case of $N = 2$ the ergodic probabilities are

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}$$

$$\pi_2 = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$$

Hamilton Filter

Let assume that the parameters are known. One important issue of Markov Switching models is the nature of the latent variable S_t. Since S_t can not be directly observed, the following procedure, known as Hamilton Filter, allows to make inference on S_t based on the behavior of r_t. This inference takes the form of the following probability:

$$\hat{\xi}_{j,t|t} = Pr(S_t = j|\bar{r}_t; \theta_j)$$

where $\bar{r}_t = \{\mathbf{r}_1, \dots, \mathbf{r}_t\}$ is the set of all information available in t. We collect all these probabilities in a vector $\hat{\xi}_{t|t}$.

$$\hat{\xi}_{\mathbf{t|t}} = \begin{bmatrix} Pr(S_t = 1|\bar{r}_t; \theta_1) \\ Pr(S_t = 2|\bar{r}_t; \theta_2) \\ \dots\dots \\ Pr(S_t = N|\bar{r}_t; \theta_N). \end{bmatrix}$$

It is also interesting to make inference about future states S_{t+1} based on $r_1, \dots, r_t$:

$$\hat{\xi}_{j,t+1|t} = Pr(S_{t+1} = 1|\bar{r}_t; \theta_j)$$

All these probabilities are collected in a vector $\hat{\xi}_{t+1|t}$

$$\hat{\xi}_{\mathbf{t+1|t}} = \begin{bmatrix} Pr(S_{t+1} = 1|\bar{r}_t; \theta_1) \\ Pr(S_{t+1} = 2|\bar{r}_t; \theta_2) \\ \dots\dots \\ Pr(S_{t+1} = N|\bar{r}_t; \theta_N). \end{bmatrix}$$

It can be show (see [Hamilton 1989]) that the Hamilton Filter could be obtain as follow:

$$\hat{\xi}_{t|t-1} = P\hat{\xi}_{t-1|t-1} \quad \text{Prediction} \tag{2.9}$$

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|_t-1} \odot \eta_t}{\mathbf{1}'(\hat{\xi}_{t|_t-1} \odot \eta_t)} \quad \text{Up-Date} \tag{2.10}$$

where η_t is the vector of conditional densities, introduced in equation 2.7 and $\odot$ denotes the element by element product. Given a starting point $\hat{\xi}_{0|0}$, this procedure is runned iteratively for $i = 1, \ldots, T$.

2.3 Markov Chain Monte Carlo: Gibbs Sampling

Markov Switching models can be estimated by different econometric methods. Maximum Likelihood and EM algorithms are exploited by [Hamilton 1989] and [Guidolin et al. (2011)]. In this work we follow a Bayesian approach in order to estimate the model by combining sample data with extra-sample priors about some parameters of the model. Following [Kim and Nelson (1999)], we estimate the model using Markov chain Monte Carlo methods (MCMC). Specifically, Gibbs Sampling algorithm is used.

2.3.1 The Bayesian Framework

In a Bayesian approach the parameters, $\theta = \{\theta_1, \ldots, \theta_h\}$, are treated as random variables having their probability distributions. As in the classical inference, the central aim is to learn about θ but in the Bayesian framework we have two information sources: sample information r_t and researcher knowledge about θ . The Bayesian summarize all the available information about θ through the posterior distribution using the Bayes theorem [1]:

$$\begin{aligned} p(\theta|\mathbf{r}) &= \frac{p(\theta)L(\theta|\mathbf{r})}{f(\mathbf{r})} \\ &\propto p(\theta)L(\theta|\mathbf{r}) \end{aligned} \tag{2.11}$$

[1]Let us consider two random variable A, B such that $P(B) \neq 0$.
The conditional probabilities of A given B is

$$Pr(A|B) = \frac{Pr(B|A)P(A)}{Pr(B)}$$

where we can ignore $f(\mathbf{r})$ because it does not depend on $\theta = \{\theta_1, \ldots, \theta_h\}$. The term $L(\theta|r)$ is the likelihood function, while the term $p(\theta)$ is called prior distribution and summarize the subjective knowledge of θ (extra-data information). Prior reflects any information the researcher has before seeing the data. Different researchers could have different prior distributions or different confidence levels about their subjective knowledge. The posterior distribution summarizes all we know about θ, that is data and extra-data information, and their contribution is proportional to the prior distribution and the likelihood.

The choice of the prior is an important issue in Bayesian framework. Prior distribution could take any functional form but, there is a particular class of prior, called natural-conjugate prior, that are easy to deal with. Specifically, when a natural conjugate prior is combined with the likelihood in equation 2.11 the posterior distribution takes the same functional form of the prior.

Once the joint posterior in equation 2.11 has been calculated, we need to obtain the marginal posterior distribution of a single parameter to make inference on it. This means that the following integral have to be solved

$$p(\theta_i|\mathbf{r}) = \int_{-\infty}^{\infty} p(\theta_1, \ldots, \theta_h|\mathbf{r}) d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_h. \qquad (2.12)$$

Apart from a few simple cases, solve the integral 2.12 analytically is difficult or even impossible. Several solutions have been proposed in the literature to solve the integral above, but the predominant approach in the modern Bayesian framework is posterior simulations.

In the following example, we consider not-independent and independent Normal Gamma as prior distributions and we see the additional complication we have in the case of independent prior.

Example 22. Not independent Normal-Gamma vs Independent Normal-Gamma.

Let us consider the follow simple linear regression model

$$r_t = \mu + \varepsilon_t \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

with $r_t, \varepsilon_t, \mu \in \mathbb{R}$ *and t=1,...,T. The parameter vector of this simple model is clearly* $\theta = \{\mu, \sigma^2\}$. *For technical reasons is easier to work with the precision*

$$h = \frac{1}{\sigma^2}.$$

Let assume a non-independent Normal-Gamma prior distribution for μ *and* h *that is:*

$$\begin{aligned} &p(\mu, h) = p(\mu|h)p(h) \quad \textit{where} \\ &\mu|h \sim \mathcal{N}(\mu_0, h^{-1}\sigma_0^2), \\ &h \sim G(\nu_0, \delta_0) \end{aligned}$$

It is easy to show that Normal-Gamma distribution is a natural conjugate prior, so the jointly posterior is also a Normal-Gamma distribution

$$\mu, h|\mathbf{r} \sim \mathcal{N}\mathbf{G}(\boldsymbol{\mu_1}, \boldsymbol{\sigma_1^2}, \boldsymbol{\nu_1}, \boldsymbol{\delta_1})$$

with

$$\begin{aligned} &\sigma_1^2 = (\sigma_0^{-2} + T)^{-1} \\ &\mu_1 = \sigma_1^2(\sigma_0^{-2}\mu_0 + T\hat{\mu}) \\ &\nu_1 = \nu_0 + T \\ &\delta_1 = \delta_0 + \hat{\varepsilon}'\hat{\varepsilon} + (\hat{\mu} - \mu)'(\Sigma_0 + T)^{-1}(\hat{\mu} - \mu) \end{aligned}$$

where $\hat{\mu}$ *is the OLS estimator of* μ.

If we are interested in the marginal posterior distribution the integral 2.12 has to be calculated and we obtain:

$$\begin{aligned} &\mu|\mathbf{r} \sim t(\mu_1, \delta_1\sigma_1^2, \nu_1) \\ &h|\mathbf{r} \sim \mathbf{G}(\boldsymbol{\nu_1}, \boldsymbol{\delta_1}). \end{aligned}$$

In this simple case we are able to obtain analytically joint and marginal posterior. However, if we also ask that μ and h are independent

$$p(\mu, h) = p(\mu)p(h) \quad \textit{where}$$
$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$
$$h \sim G(\nu_0, \delta_0)$$

the situation is completely different. We are not able to find a well-know distribution for the jointly posterior distribution but, the condition posteriors are simple. It can be show that

$$\mu|\sigma^2, \mathbf{r} \sim \mathcal{N}(\mu_{\mathbf{1}}, \sigma_{\mathbf{1}}^{\mathbf{2}}) \tag{2.13}$$
$$h|\mu, \mathbf{r} \sim \mathbf{G}(\nu_{\mathbf{1}}, \delta_{\mathbf{1}}), \quad \mathbf{h} = \frac{\mathbf{1}}{\sigma^{\mathbf{2}}} \tag{2.14}$$

where

$$\sigma_1^2 = (\sigma_0^{-2} + T\sigma^{-2})^{-1} \tag{2.15}$$
$$\mu_1 = \sigma_1^2(\sigma_0^{-2}\mu_0 + h\sum_{t=1}^{T} r_t) \tag{2.16}$$
$$\nu_1 = \nu_0 + T \tag{2.17}$$
$$\delta_1 = \delta_0 + \sum_{t=1}^{T}(r_t - \mu)^2 \tag{2.18}$$

Observe that despite the similarity of the previous formulas, the latter equations 2.13 - 2.14 are not the marginal posterior distributions $p(\mu|\mathbf{r})$ and $p(h|\mathbf{r})$ but only the conditional posterior distributions $p(\mu|h, \mathbf{r})$ and $p(h|\mu, \mathbf{r})$

The Gibbs-Sampling methodology allows to skip this problem and offers an alternative approach to obtain the posterior marginal distributions without knowing the posterior distribution 2.11 and without solving the integral 2.12

2.4 Markov chain Monte Carlo methods: Gibbs Sampling

Markov chain Monte Carlo (MCMC) methods are powerful tools of posterior simulations and allow to sample from a target distribution by a Markov chain. In many situations, Markov chain Monte Carlo methods are the only possible choice to draw from a target distributions π^*. In the Bayesian framework the target distribution is the posterior distribution of the parameters. Markov chains Monte Carlo were introduced in the 50s by [Metropolis,Rosenbluth,Teller 1953], who proposed an algorithm to simulate multivariate discrete distributions and they found application in physics. They become popular in Bayesian statistics at the early 1990s.

The output of those algorithms consists in a sample of correlated draws, $\{\hat{\theta_1}, \hat{\theta_2}, \ldots, \hat{\theta_k}\}$ from the target distribution π^*. A independent sample can be obtain by splitting the sample in groups of L elements and keeping only the first element of each group. The group size L depends on the autocorrelations of the sample.

One of the most popular MCMC method is Gibbs Sampling algorithm, introduced by [Geman and Geman]. It allows the estimation of the joint and marginal posterior distribution only through the conditional posterior distributions.

Let the parameters space Θ be grouped into H blocks:

$$\theta = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(H)})$$

If the complete set of conditional posterior densities is known:

$$p(\theta_{(1)}|\theta_{(2)},\theta_{(3)},\dots,\theta_{(H)})$$
$$p(\theta_{(2)}|\theta_{(1)},\theta_{(3)},\dots,\theta_{(H)})$$
$$\dots \quad \dots \quad \dots \quad \dots$$
$$p(\theta_{(H)}|\theta_{(1)},\theta_{(2},\dots,\theta_{(H-1)})$$

then Gibbs Sampling allows us to draw a sample $(\theta_{(1)},\theta_{(2)},\dots,\theta_{(H)})$ without the joint and marginal posterior distributions. Specifically we draw $(\theta_{(1)},\theta_{(2)},\dots,\theta_{(H)})$ from the conditional posterior distribution, using the last update of each parameter.
Formally, given an arbitrary starting point

$$\{\theta_{(2)}^{(0)},\theta_{(3)}^{(0)},\dots,\theta_{(H)}^{(0)}\},$$

the Gibbs Sampling iterates S times the following steps,

- Draw $\theta_{(1)}^{(s)}$ from $p(\theta_{(1)}^{(s)}|\theta_{(2)}^{(s-1)},\theta_{(3)}^{(s-1)},\dots,\theta_{(H)}^{(s-1)})$,
- Draw $\theta_{(2)}^{(s)}$ from $p(\theta_{(2)}^{(s)}|\theta_{(1)}^{(s)},\theta_{(3)}^{(s-1)},\dots,\theta_{(H)}^{(s-1)}))$
- $\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$
- Draw $\theta_{(H)}^{(s)}$ from $p(\theta_{(H)}^{(s)}|\theta_{(1)}^{(s)},\theta_{(2)}^{(s)}\dots,\theta_{(H-1)}^{(s)})$

This procedure, iterated for S sufficiently large, will yield a sequence $\theta^{(1)},\theta^{(2)},\dots\theta^{(S)}$. To eliminate the effect of the starting point, it is necessary to eliminate the first S_0 draw, called burn-in-replication. We also keep one value every L to eliminate the dependent path of the output and generated a i.i.d sample. So the final output of the Gibbs Sampling algorithm is

$$\theta^{(s_0)},\theta^{(s_0+L)},\theta^{(s_0+2L)},\dots,\theta^{(S)} \tag{2.19}$$

It can be show that 2.19 converge to the joint and marginal distribution for $j\to\infty$

2.4.1 Gibbs-Sampling in Linear Regression Models

Since conditional on S_t, equation 2.1 is a regression model with a known dummy variable S_t, let explain Gibbs-Sampling procedure for the simple linear regression model.

Univariate Case

Let us consider the simple linear regression model in 2.1 with $K = 1$

$$r_t = \mu + \epsilon_t \quad \varepsilon \sim \mathcal{N}(\mu, I_n\sigma^2)$$

with $r, \varepsilon, \mu \in \mathbb{R}$ and $t = 1, \ldots, T$

The natural choice is to set $H = 2$ blocks with $\theta_1 = \mu$ and $\theta_2 = h = \frac{1}{\sigma^2}$. Let assume independent Normal-Gamma prior for μ and h

$$\mu \sim \mathcal{N}(\mu_0, \Sigma_0)$$
$$h \sim G(\nu_0, \delta_0), \quad h = \frac{1}{\sigma^2}$$

where the prior parameters $\mu_0, \sigma_0, \nu_0, \delta_0$ are known. We have seen that

$$\mu|\sigma^2, \mathbf{r} \sim \mathcal{N}(\mu_{\mathbf{1}}, \mathbf{\Sigma_1}) \tag{2.20}$$

$$h|\mu, \mathbf{r} \sim \mathbf{G}(\nu_{\mathbf{1}}, \delta_{\mathbf{1}}) \tag{2.21}$$

where $\mu_1, \sigma_1, \nu_1, \delta_1$ are expressed in equations 2.16, 2.15, 2.17 ,2.18. Given the conditional posterior distributions, it is very easy to implement Gibbs-Sampling. Let chose an arbitrary starting point, for example, $h^{(0)} = \frac{1}{\sigma^{2(0)}}$. For $s = 1, \ldots, S$ the following steps are repeated:

1 Draw μ^s from equation 2.20 conditional on σ^{s-1}

2 Draw $h^{(s)} = \frac{1}{\sigma^s}$ from equation 2.21 conditional on μ^s

3 Set $s = s + 1$ and go back to 1.

Once this procedure is finished we discard the first S_0 value and keep one value every L to deal with a i.i.d sample. We are now ready to make inference on parameters.

Multivariate Case

Let us consider the simple multivariate model in 2.1 with $K > 1$

$$\mathbf{r_t} = \mu_\mathbf{t} + \varepsilon_\mathbf{t} \quad \varepsilon \sim \mathcal{N}_\mathbf{K}(\mu, \mathbf{\Sigma^2})$$

with $\mathbf{r_t}, \varepsilon_\mathbf{t}, \mu \in \mathbb{R}^K$ and $t = 1, \ldots, T$. The natural extension of the independent Normal-Gamma prior is the independent Normal-Wishart prior:

$$p(\mu, H) = p(\mu)p(H) \quad \text{where}$$

$$\mu \sim \mathcal{N}_K(\mu_\mathbf{0}, \Sigma_0), \quad \Sigma_0 \in \mathbb{R}^{K\times K} \tag{2.22}$$

$$H \sim W(\nu_0, H_0), \quad H = \Sigma^{-1} \tag{2.23}$$

It can be show that the conditional posterior of μ and H are,respectively, Normal and Wishart distribution:

$$\mu | H, \mathbf{r} \sim \mathcal{N}_\mathbf{K}(\mu_\mathbf{1}, \mathbf{\Sigma_1}) \tag{2.24}$$

$$\Sigma_1 = (\Sigma_0^{-1} + TH)^{-1}$$

$$\mu_\mathbf{1} = \Sigma_1(\Sigma_0^{-1}\mu_0 + H\sum_{i=1}^{T}\mathbf{r}_i)$$

$$H|\mu, \mathbf{r} \sim \mathbf{W}(\nu_\mathbf{1}, \mathbf{H_1}) \tag{2.25}$$

$$\nu_1 = \nu_0 + T$$

$$H_1 = [H_0^{-1} + \sum_{i=1}^{T}(\mathbf{r}_i - \mu))^\mathbf{2}]^\mathbf{-1}$$

One the conditional posterior distributions are known, the Gibbs-Sampling procedure can be implemented as in the univariate case.

2.5 Markov Switching Models and Gibbs Sampling

The last step is to consider Gibbs Sampling procedure for Markov Switching models. Let us consider a simple Markov Switching model 2.1-2.3

$$\mathbf{r_t} = \mu_\mathbf{S_t} + \varepsilon_\mathbf{t}, \; \varepsilon_\mathbf{t} \sim \mathcal{N}(0, \Sigma_{S_t}) \quad t = 1, \ldots, T$$

$$P(S_t = j | S_{t-1} = i) = p_{ij} \quad \forall i, j$$

In the Bayesian approach both the parameters of the model, $\mu_{S_1}, \mu_{S_2}, \dots, \mu_{S_N}$, $\sigma_{S_1}, \sigma_{S_2}, \Sigma_{S_N}, p_{ij}$ with $i, j = 1, \dots, N$ and the latent variables $S_1 \dots, S_T$ are viewed as random variable. Following [Kim and Nelson (1999)] it can be show that the joint posterior distribution can be written as follow

$$\begin{aligned} & p(S_1, \dots, S_T, \mu_{S_1}, \dots, \mu_{S_N}, \sigma_{S_1}, \dots, \sigma_{S_N}, p_{ij}) \\ & = p(\mu_{S_1}, \dots, \mu_{S_N}, \sigma_{S_1}, \dots, \sigma_{S_N} | \bar{r}_T, \bar{S}_T) p(p_{ij}, | \bar{S}_T) p(S_1, \dots, S_T | \bar{r}_T) \end{aligned}$$

where $\bar{r}_T = \{r_1, \dots, r_T\}$ and $\bar{S}_T = \{S_1, \dots, S_T\}$. Thus, using an arbitrary starting point for the parameters $\mu^0_{S_1}, \dots, \mu^{(0)}_{S_N}, \sigma^{(0)}_{S_1}, \sigma^{(0)}_{S_2}, p^0_{ij}$, the following Gibbs steps can be repeated until convergence occurs:

1 Generate the block $S_1, \dots, S_T$ from

$$p(S_1, \dots, S_T |, \mu_{S_1}, \dots, \mu_{S_N}, \sigma_{S_1}, \dots, \sigma_{S_N}, p_{ij}, \bar{r}_T)$$

2 Generate p_{ij} from

$$p(p_{ij} | \bar{r}_T)$$

3 Generate $\mu_{S_1}, \dots, \mu_{S_N}, \sigma_{S_1}, \dots, \sigma_{S_N}$ from

$$p(\mu_{S_1}, \dots, \mu_{S_N}, \sigma_{S_1}, \dots, \sigma_{S_N} | \bar{r}_T, \bar{S}_T)$$

Note that is not necessary to implement the previous steps in the specific order.

Step1: Generate $S_1, \dots, S_T$

In this step the conditional distribution of the block $S_1, \dots, S_T$ conditional all other parameters is calculated. Following [Kim and Nelson (1999)] it can be show using the Markov property of S_t, that:

$$p(S_1, \dots, S_T | r_1, \dots, r_T) = p(S_T | \bar{r}_T) \prod_{t=1}^{T-1} p(S_t | S_{t+1}, \bar{r}_t)$$

Thus, first of all we run the Hamilton Filter, (see 2.9) and save the UpDate matrix:

$$\hat{\Xi} = \begin{bmatrix} \hat{\xi}^T_{1|1} \\ \hat{\xi}^T_{2|2} \\ . \\ . \\ . \\ \hat{\xi}^T_{T|T} \end{bmatrix}.$$

where

$$\hat{\xi}_{\mathbf{t|t}} = \begin{bmatrix} Pr(S_t = 1|\bar{r}_t; \theta_1) \\ Pr(S_t = 2|\bar{r}_t; \theta_2) \\ \dots\dots \\ Pr(S_t = N|\bar{r}_t; \theta_N). \end{bmatrix}.$$

We use the last Up-Date $\hat{\xi}_{T|T}$ to generated S_T. Specifically, using the inverse transform method, we draw $u \sim U([0\ 1])$ and set

$$S_T = \begin{cases} 1 \quad \text{if } u \leq Pr(S_T = 1|\bar{r}_T) \\ 2 \quad \text{if } Pr(S_T = 1|\bar{r}_T) < u \leq Pr(S_T = 1|\bar{r}_T) + P(S_T = 2|\bar{r}_T) \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ N \quad \text{if } \sum_{i=1}^{N-1} Pr(S_T = i|\bar{r}_T) < u \leq \sum_{i=1}^{N} Pr(S_T = i|\bar{r}_T) \end{cases}$$

Next for $t = T-1, \dots, 1$ using the Bayes theorem and the Markov property we have

$$p(S_t|\bar{r}_T, S_{t+1}) \propto p(S_{t+1}|S_t)p(S_t|\bar{r}_T) \tag{2.26}$$

where $p(S_{t+1}|S_t)$ are the transition probabilities and $p(S_t|\bar{r}_T)$ the Up-Date. Thus we are able to obtain $P(S_t|\bar{r}_t, S_{t+1})$ from equation 2.26 and we draw S_t using again the inverse transform method.

2.5.1 Step2: Generate Transition Probabilities

Since p_{ij} $i, j = 1, \dots, N$ are probabilities, we would like that for all i, j $0 \leq p_{ij} \leq 1$. Beta distribution ensures this feature and it is also self-conjugate. So, let assume beta distribution prior:

$$p_{ij} \sim \beta(u_{ij}, \bar{u}_{ij})$$

where u_{ij} and $\bar{u}_{ij}$ are known.

Let define the quantities $\bar{p}_{ii}$ and $\bar{p}_{ij}$ with $i \neq j$ as follow:

$$\bar{p}_{ii} = P(S_t \neq i | S_{t-1} = i) \tag{2.27}$$

$$\bar{p}_{ij} = P(S_t = j | S_{t-1} = i, S_t \neq i) \tag{2.28}$$

We have that

$$\begin{aligned} p_{ij} &= P(S_t = j | S_{t-1} = i) \\ &= P(S_t = j | S_{t-1} = i, S_t \neq i) P(S_t \neq i | S_{t-1} = i) \\ &= \bar{p}_{ij}(1 - p_{ii}) \end{aligned} \tag{2.29}$$

Moreover, let define the transition counter η_{ij} and $\bar{\eta}_{ij}$.

$$\eta_{ij} = \#\{\text{transitions from } S_{t-1} = i \text{ to } S_t = j\}$$

and

$$\bar{\eta}_{ij} = \#\{\text{transitions from } S_{t-1} = i \text{ to } S_t \neq j\}.$$

It can be show that $\forall\ i, j = 1, \dots, T$

$$p_{ii} | S_1, \dots, S_T \sim \beta(u_{ii} + n_{ii}, \bar{u}_{ii} + \bar{n}_{ii})$$

and $\forall\ i = 1, \dots, T$ $j = 1, \dots, T-1$

$$\bar{p}_{ij} | S_1, \dots, S_T \sim \beta(u_{ij} + n_{ij}, \bar{u}_{ij} + \bar{n}_{ij}).$$

Once p_{ii} and $\bar{p}_{ij}$ are generated, generation of the p_{ij} $i \neq j$ is straightforward. From equation 2.29 we are able to calculate all the these posterior distribution.

2.5.2 Step3: Generate $\mu_{S_1}, \mu_{S_2}, \ldots, \mu_{S_N}, \sigma_{S_1}, \sigma_{S_2}, \ldots, \sigma_{S_N}$

Conditional on the State S_t, the Markov Switching Models is reduced to N simple linear regression model. So once we split the N subsamples relative to the N states, we can draw the parameters of the model like in the previous section. Specifically , given an arbitrary starting point, for example $h_{S_1}^{(0)}, \ldots, h_{S_N}^{(0)}$, the Gibbs algorithms is performed for each State S_t $t = 1, \ldots, N$. From 2.16, 2.15, 2.17 and 2.18 the update of the parameters for $s = 1, \ldots, S$ are

$$\sigma_{1S_t}^{2(s)} = (\sigma_0^{-2} + T_{S_t} h_{S_t}^{(s-1)})^{-1} \tag{2.30}$$

$$\mu_{1S_t}^{(s)} = \sigma_{1S_t}^{2(s)} (\sigma_0^{-2} \mu_0 + h_{S_t}^{(s-1)} \sum_{t=1}^{T_{S_t}} r_{S_t}) \tag{2.31}$$

$$\nu_{S_t} = \nu_0 + T_{S_t} \tag{2.32}$$

$$\delta_{S_t}^{(s)} = \delta_0 + \sum_{t=1}^{T} (r_{S_t} - \mu_{S_t}^{(s)})^2 \tag{2.33}$$

where r_{S_t} are the returns consistent with the State S_t according to the estimation in Step1 and T_{S_t} is the number of return in the State S_t.

2.6 Markov Switching Results

In this section the results are displayed. Specifically, after a preliminary analysis, we run a univariate Markov Switching model for each index and multivariate Markov Switching model, in order to take the correlations switching into account. Both univariate and multivariate Markov Switching models are estimated through Gibbs Sampling algorithm. A state-dependent trade rule is presented and compared with a buy-and-hold strategy, which is the portfolio full invested into the market.

All routines were written in Matlab and all the algorithms were implemented by the candidate. In the Appendix the key routines are presented.[2]

2.6.1 Data and Preliminary Analysis.

Our data consists of daily returns from Jan 1997 to May 2019 and the data provider is Bloomberg. We analyzed different asset classes across different geographic areas. More specifically, we worked on 4 equity indexes, 1 commodity future and 6 currencies. All prices are in local currency and all currencies are expressed versus USD dollar. A complete list of the data is displayed in the Table 2.1

Since we worked on daily returns quoted on different markets, we have to

Table 2.1: Asset Details

Equity	Ticker	Currency	Currency	Ticker	Contry
SPX 500	SPX Index	US Dollar	Australian Dollar	AUDUSD	Australia
STOXX Europe 600	SXXP	Euro	Japanese yen	JPYUSD	Japan
MSCI Emerging Market Index	MXEF	US Dollar	Brazilian real	BRLUSD	Brazil
Tokyo stock Price IndeX	TPX Index	Japanese yen	Euro	EURUSD	Euro Zone
Commodity	Ticker	Currency	Mexican peso	MXNUSD	Mexico
Crude Oil	CL1	US Dollar	Russian ruble	RUBUSD	Russia

handle the closing days of each market. We decided to align all indexes with the US calendar, which is the most relevant market around the world. In addition, in order to manage weekends data, we use the Bloomberg setup

[2]All the routine are available upon request

to doawnload the trading days.
In the world of the emerging Currencies, given they are high linearly correlated, we selected that currencies with the greatest combination in terms of higher liquidity and lower transaction costs. In addition, we decided to leave out the British pound because the Brexit issue has altered its natural pricing.
Table 2.2 shows the descriptive statistics of returns.
Statistics about skewness and kurtosis highlight the typical non-normal distribution of returns with fat tails and asymmetry around the expected value.

Table 2.2: Descriptive Statistics of Stock Returns

	μ	σ	Skewness	Kurtosis	JB Test
SPX	0.027%	1.21%	-0.05	11.15	Not Accept H_0
SXXP	0.017%	1.23%	-0.04	8.13	Not Accept H_0
MXEF	0.027%	1.23%	-0.5	11.09	Not Accept H_0
TPX	0.017%	1.35%	-0.18	9.44	Not Accept H_0
CL1	0.054%	2.4%	0.15	7.2	Not Accept H_0
AUDUSD	0.0063%	0.81%	-0.15	11.74	Not Accept H_0
JPYUSD	0.0058%	0.68%	0.41	8.9	Not Accept H_0
BRLUSD	-0.017%	1.07%	-0.07	13.55	Not Accept H_0
EURUSD	0.0035%	0.6%	-0.07	4.5	Not Accept H_0
RUBUSD	-0.035%	1.43%	1.8	177	Not Accept H_0
MXNUSD	-0.015%	0.70%	-0.62	13.78	Not Accept H_0

This is confirmed by the Jarque-Brera test that strongly rejects the null hypothesis of Normality. Volatilities are very different across asset classes and moreover in Figure 2.1 the 252-rolling volatility of returns suggests the presence of heteroskedasticity. In order to preliminary investigate the time-variation of the volatility of returns, a GARCH analysis is performed. We also consider the T-GARCH and E-GARCH extensions to take asymmetry and leverage effects into account, where leverage effect indicates the situ-

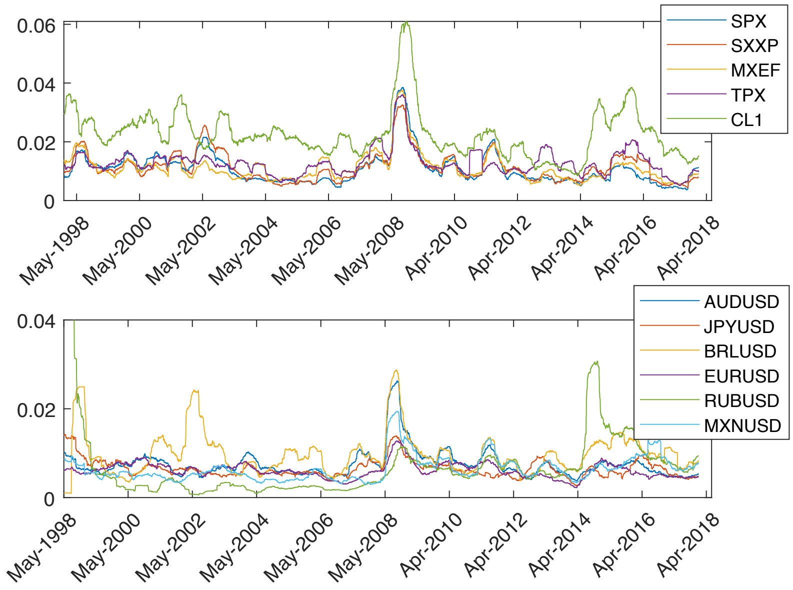

Figure 2.1: Rolling Volatility

ation where low volatility tends to produce further drop of volatility and high volatility tends to increase volatility. In addition to the full sample estimates, all estimates were also made recursively to test the persistence of those effects. The heteroskedasticity is confirmed for all indexes by the GARCH parameters. Moreover, in Figure 2.2 the full sample T-GARCH and E-GARCH parameters are plotted and the presence of asymmetry and leverage effects is pointed out. It is clear the difference between the so-called carry trade currencies (AUD, MXN, RUB, BRL) and the safe currency (EUR, JPY). For the first group we have strong effects of asymmetry and leverage while those effects are absent for Euro and Yen. In particular, it is confirmed the nature of haven currency of the latter with an E-GARCH parameter even positive. Volatility cluster also occurs in equity markets and oil future.

Because the accommodative monetary policy of different central banks around the world over the last ten years has artificially produced a global environ-

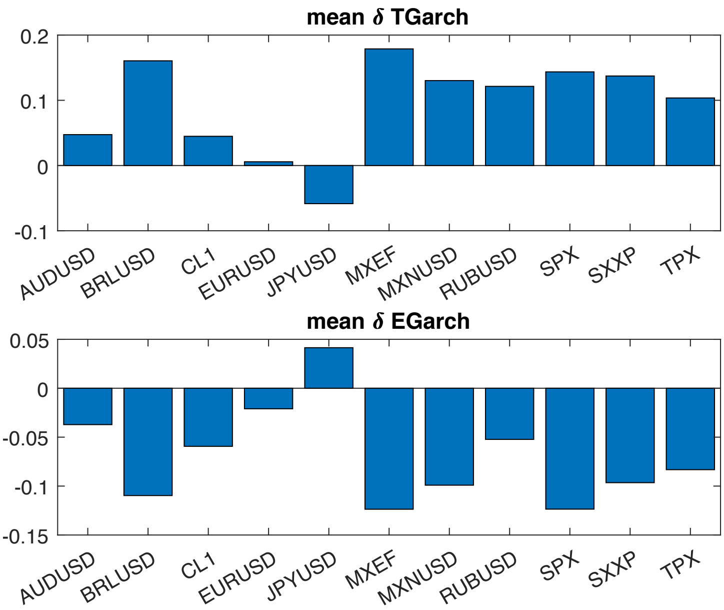

Figure 2.2: T-GARCH and E-GARCH Parameters

ment of ultra-low interest rates, we decide to exclude Bonds from further analysis.

Based on the previous analysis, we move forward describing returns according to a 2-State Markov Switching Mixture of Normal Distribution. The choice of dealing with 2 regimes is almost common in finance for at least a couple of reasons: first of all, this solution is able to highlight clusters in returns in a very intuitive and useful way by describing the pattern of performances in terms of bull and bear market (Low Volatility and High Volatility). Secondly, parameters of Markov Switching models tend to increase quite rapidly when the number of regimes does increase, so we decided to focus our attention on the 2-state version in order to deal with a limited number of free parameters. So the 2-State representation is a powerful solution to capture the non stationarity of returns with the advantage of the limited number of unknown parameters.

Let assume model 2.1,2.2,2.3 with $N = 2$:

$$\mathbf{r_t} = \mu_{\mathbf{S_t}} + \varepsilon_{\mathbf{t}},\ \varepsilon_{\mathbf{t}} \sim \mathcal{N}(0, \Sigma_{S_t}) \quad t = 1, \dots, T$$
$$P(S_t = 1|S_{t-1} = 1) = p_{11} \quad P(S_t = 2|S_{t-1} = 2) = p_{22}$$

All estimates were made in a Bayesian framework, specifically we used the Gibbs-Sampling methodology introduced in the previous chapter.

2.6.2 Univariate Markov Switching

First of all we estimate for each series an univariate Markov Switching model. The first aim is to fix the prior distributions and their parameters. As we explain in the previous chapter, it is useful to consider natural conjugate priors. Let assume a Normal distribution for μ, a Gamma distribution for $h = \frac{1}{\sigma^2}$ and a Beta distribution for the transition probabilities p_{11} and p_{22} as prior distributions:

$$\mu_i \sim \mathcal{N}(\mu_{i0}, \sigma^2_{i0})$$
$$h_i \sim G(\nu_{i0}, \delta_{i0}) \quad h = \frac{1}{\sigma_i^2}$$
$$p_{11} \sim \beta(u_{11}, u_{12})$$
$$p_{22} \sim \beta(u_{22}, u_{21})$$

where $i = 1, 2$. Once the prior distributions of the parameters are fixed, prior means and variances have to be specified. Prior mean represents the expectation of the parameters of interest, while the prior variance is the confidence level of this expectation. Here, prior distributions are designed to be quite uninformative. That is obtained setting parameters in a way that the prior variance is quite large in order to keep the extra-sample information under control when it is combined with the likelihood. Next we provide also a sensitive variance analysis. As regards the mean, since we are working with daily returns, the natural choice is to assume μ_{01} and μ_{02} close to zero, positive the former and negative the latter. From the Gamma distribution we have that [3]

$$\mathbb{E}[h_i] = \frac{v_{i0}}{\delta_{i0}}. \tag{2.34}$$

The parameters of the Gamma distribution are handle using VIX index levels. First off all, Figure 2.3 displays full sample SPX Index versus VIX

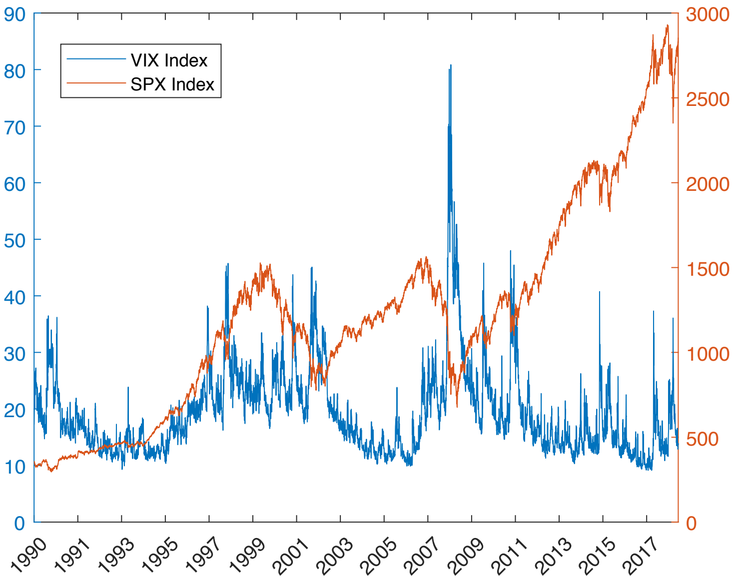

Figure 2.3: Full sample VIX Index vs SPX Index

Index. It is clear that low levels of VIX coincide with SPX rally, while spike of volatility are typical during drawdown. To set those levels we analyze historical performances of VIX Index vs SPX Index in a presampling from Jan 1990 to Dec 2002. Table 2.3 displays that when the SPX Index is in uptrend VIX Index hits its lowest level, around 10%. While negative performances of the SPX Index are accompanied by VIX levels structurally

[3]Let $X \sim G(\nu, \delta)$ we have

$$\mathbb{E}[X] = \frac{\nu}{\delta}$$
$$var(X) = \frac{\nu}{\delta^2}$$

Table 2.3: SPX Index vs VIX Index

YoY	**Perf SPX**	**Perf VIX**	**Max VIX**	**Min VIX**	**Mean VIX**
1990	-6.56%	53.02%	36.47	14.72	23.063
1991	26.31%	-26.80%	36.20	13.95	18.385
1992	4.46%	-34.90%	21.02	11.51	15.452
1991	26.31%	-26.80%	36.20	9.31	18.385
1994	-1.54%	13.21%	23.87	9.94	13.926
1995	34.11%	-5.15%	15.74	10.36	12.389
1996	20.26%	67.09%	21.99	12.00	16.442
1997	31.01%	14.77%	38.20	17.09	22.379
1998	26.67%	1.71%	45.74	16.23	25.603
1999	19.53%	0.90%	32.98	17.42	24.373
2000	-10.14%	8.97%	33.49	16.53	23.315
2001	-13.04%	-11.36%	43.74	18.76	25.750
2002	-23.37%	20.25%	45.08	17.40	27.292

higher. During drawndown VIX Index achieves its maximum exceeding the 30% (see Table 2.4).

So, we suppose a VIX Level of 10% in Normal Volatility and a VIX level of 30% in High Volatility. Thus, prior parameters on h_1 and h_2 are such that

$$\mathbb{E}[h_1] = \frac{v_{01}}{\delta_{01}} = \frac{\sqrt{252}}{0.1}$$
$$\mathbb{E}[h_2] = \frac{v_{02}}{\delta_{02}} = \frac{\sqrt{252}}{0.3}$$

Finally, market experiences lead us to believe that the Normal Volatility is the predominant state. Thus for the transition probabilities, priors parameters are set in order to fix the ergodic probability of the Normal Volatility regime as the predominant one, for example 70% as suggested the empirical evidence.

To check that the Gibbs Sampling procedure converges we compare the

Table 2.4: VIX Index during SPX Index drawndown

YoY	**Max DD SPX**	**Mean VIX during DD**	**Max VIX during DD**
1990	19.92%	26.97	36.47
1991	5.60%	17.30	18.38
1992	6.24%	17.42	20.15
1993	4.99%	13.13	15.66
1994	8.94%	14.82	23.87
1995	2.53%	12.13	14.55
1996	7.64%	16.86	21.55
1997	10.80%	21.54	31.12
1998	19.34%	28.25	44.28
1999	12.08%	24.02	28.75
2000	17.20%	23.26	33.49
2001	29.70%	24.66	43.74
2002	33.75%	26.64	45.08

results of $H = 4$ Gibbs Sampling with different starting point. The starting points are drawn from the prior distribution. With $h = 10.000$ iterations we obtain the same results in each Gibbs Sampling. The first 10% of the draws is discarded in order to delete the effect of the starting point on the data generating process. In addition, in order to generate a iid sample from the posterior, we take a draw any 3 iterations.
The Gibbs Sampling output lead us to obtain point estimations $\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_1, \pi_2$ and the estimation of the latent variable $S_1, \ldots, S_T$. The latter allows us to cluster returns in one of the two regimes in each $t = 1, \ldots, T$.

Table 2.5 and Table 2.6 summarize the point estimates and the confidence level of the parameters. Most indexes exhibit two states clearly separated and Normal Volatility is the predominant state. In general, the volatilities in High Volatility are twice those in Normal Volatility. Except for Yen, the High Volatility regime is characterized by negative returns.

Table 2.5: Univariate Markov Switching Results

	μ_1	μ_2	σ_1	σ_2	π_1	π_2
SPX	0.074%	-0.070%	0.70%	1.82%	67.7%	32.3%
SXXP	0.060%	-0.097%	0.79%	1.89%	69.6%	30.4%
MXEF	0.097%	-0.215%	0.82%	2.04%	76.4%	23.6%
TPY	0.060%	-0.128%	0.97%	2.2%	77.2%	22.8%
CL1	0.091%	-0.12%	1.83%	4.16%	82.5%	17.5%
AUDUSD	0.015%	-0.054%	0.63%	1.60%	87.8%	12.2%
JPYUSD	-0.015%	0.095%	0.52%	1.16%	80.85%	19.15%
BRLUSD	0.02%	-0.1%	0.62%	1.76%	70.6%	29.4%
EURUSD	0.013%	-0.014%	0.48%	0.83%	65.65%	34.35%
RUBUSD	0.06%	-0.25%	0.40%	3.44%	83.2%	16.8%
MNXUSD	0.005%	-0.1%	0.5%	1.27%	80.4%	19.6%

	μ_1	μ_2	$\sigma1$	$\sigma2$
SPX	[0.05%, 0.10%]	[-0.19%, 0.05%]	[0.67%, 0.72%]	[1.72%, 1.92%]
SXXP	[0.03%, 0.08%]	[-0.24%, 0.04%]	[0.75%, 0.82%]	[1.80%, 2.01%]
MXEF	[0.07%, 0.13%]	[-0.33%, -0.09%]	[0.79%, 0.84%]	[1.91%, 2.13%]
TPX	[0.03%, 0.15%]	[-0.30%, 0.03%]	[0.90%, 0.99%]	[1.97%, 2.31%]
CL1	[0.06%, 0.12%]	[-0.19%, -0.05%]	[1.75%, 1.87%]	[3.80%, 4.37%]
AUDUSD	[-0.01%, 0.03%]	[-0.44%, 0.14%]	[0.60%, 0.65%]	[1.44%, 1.71%]
JPYUSD	[-0.03%, 0.002%]	[-0.04%, 0.20%]	[0.48%, 0.52%]	[1.04%, 1.21%]
BRLUSD	[0%, 0.04%]	[-0.19%, -0.01%]	[0.59%, 0.63%]	[1.66%, 1.83%]
EURUSD	[-0.006%, 0.03%]	[-0.05%, 0.02%]	[0.47%, 0.50%]	[0.80%, 0.87%]
RUBUSD	[-0.12%, 0.03%]	[-0.43%, -0.02%]	[0.38%, 0.42%]	[3.24%, 3.64%]
MXNUSD	[-0.01%, 0.11%]	[-0.22%, 0.01%]	[0.50%, 0.55%]	[1.16%, 1.34%]

Table 2.6: Confidence Interval of posterior parameters (95%)

Figure 2.4 display that this regime corresponds to bear market (2000-2001, 2007-2008, 2011). We also observe that both states are highly persistent with $p11 > p22$. This probabilities lead to an expected duration of Normal

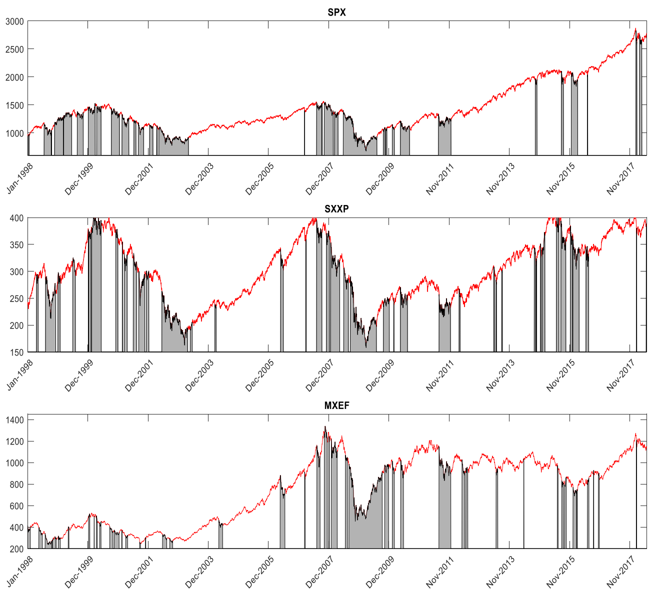

Figure 2.4: Normal Volatility and High Volatility

Volatility greater than that in High Volatility.
We highlight a clear split between the two states for the equity market (Figure 2.5 and Figure 2.6), with greater volatilities for MXEF index e TPX index.

Among currencies we highlight the different split obtained for carry trade and safe currencies. We point out that safe currencies, as Euro and Yen, exhibit low volatilities. Euro do not exhibit a clear threshold between the two states. Specifically, we have returns close to zero and the marginal posterior distribution of μ_2 overlaps the marginal posterior distribution of μ_1 (see Table 2.5 and Figure 2.7). The volatilities are very similar in the two states and moreover, the ergodic probabilities do not identify a predominant state. This is reflected in uniform returns in all the sample period (Figure 2.8). The difference is more glaring for Yen where the dynamics of expected returns of the two gaussian are opposite with respect to all other asset classes. When the risk-off occurs, investors hedge their position taking

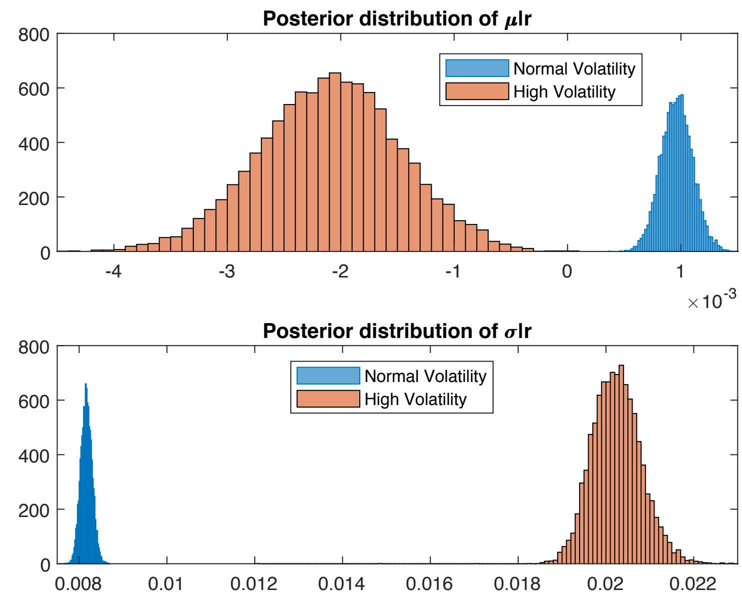

Figure 2.5: Marginal posterior distribution of MXEF

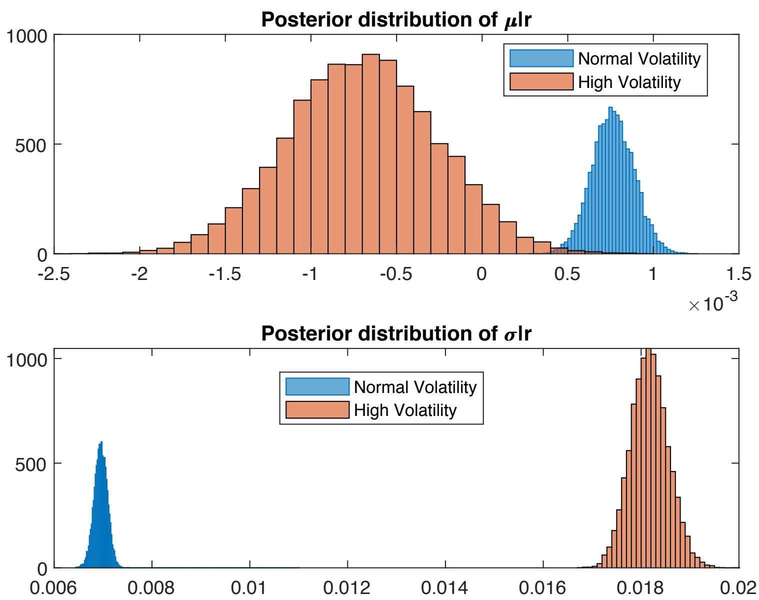

Figure 2.6: Marginal posterior distribution of SPX

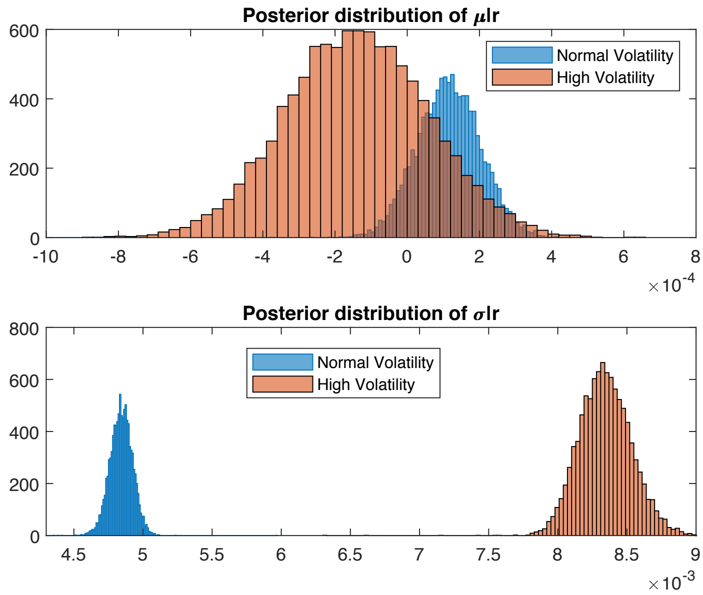

Figure 2.7: Marginal posterior distribution of EURUSD

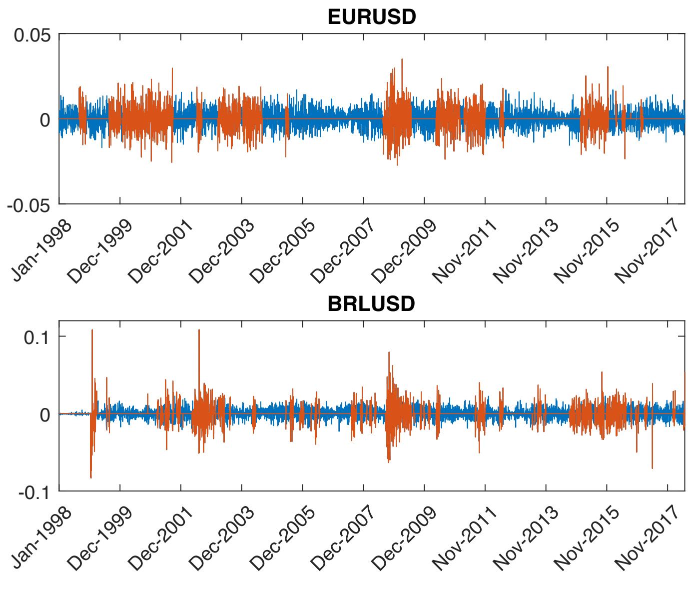

Figure 2.8: In blu Normal Volatility returns and in red High Volatilities returns

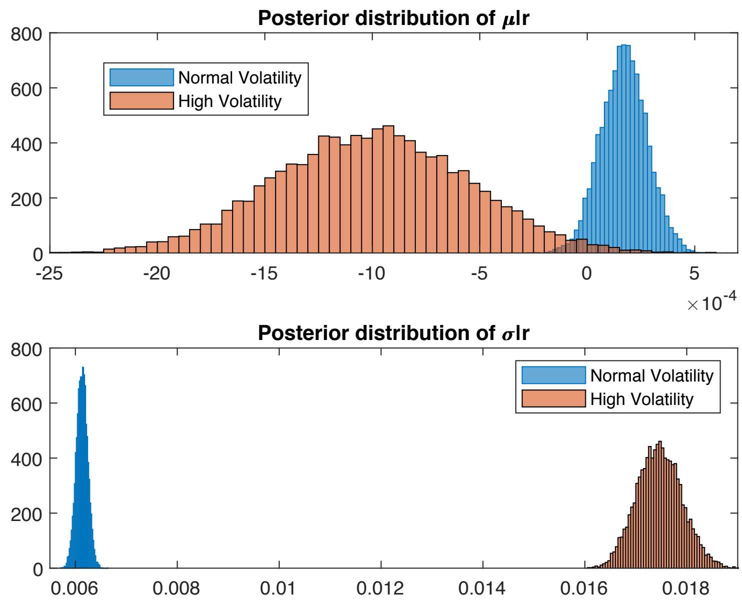

Figure 2.9: Marginal posterior distribution of BRLUSD

long position on Yen.

Figure 2.9 and 2.10 highlight an interesting picture for the carry trade currencies. They perform pretty well against the USD dollar in a Risk-On environment when investors are willing in demanding risky asset, while they suffer a lot when risk-aversion spikes suddenly and demand for safe assets increases quite sharply. That explains why the High Volatility regime is related to the USD dollar appreciation, give its nature of typical safe currency (Figure 2.10).

Skewness and Kurtosis statistics conditioning on any the State are calculated in Table 2.7 to check the Normality of returns in each regime. Skewness statistic exhibits a typical Gaussian profile in both Normal and High Volatility, moreover the Skewnees test is in favor of Normality for all asset indexes in both states. In Normal Volatility also the Kurtosis statistic assumes value compatible with a Normal distribution (around $k = 3$). Note that Normal volatility is the predominant State, Table 2.5 displays that the unconditional

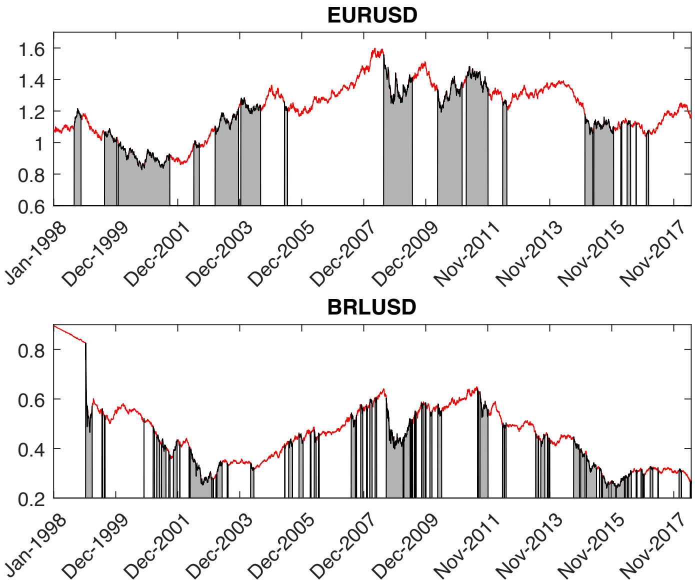

Figure 2.10: Normal Volatility and High Volatility

probabilities of this state, π_1, are between 70% and 80%. This means that the Gaussian is recovered in terms of state conditioned distribution (see Figure 2.11). This is a very important point and justify our choice of working with a mixture of Normal distribution to describe the pattern of returns. This framework is able to capture fat tails and asymmetry and recovers the Normality conditional on any State. One of the most important result of this solution is that we can apply all results in terms of portfolio construction, equilibrium and pricing which requires the assumption of Normality of returns.

Our results are in line with [Bulla et al. 2011] and [Pereiro,Gonzalez (2015)]. The former assumed a Markov Switching model with $N = 2$ states and estimated the model by the method of maximum-likelihood. In particular the results for SPX index are very similar. The latter used SETAR model to test the existence of two or more regimes and carry out the existence two

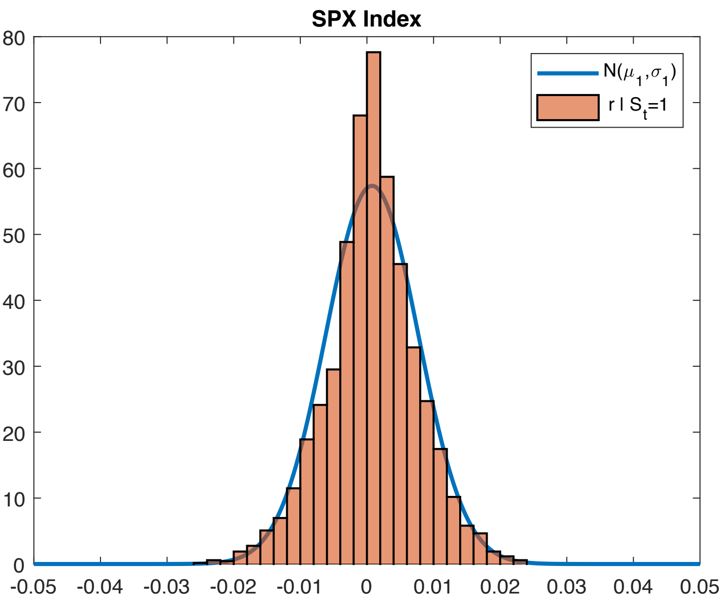

Figure 2.11: Normal Fit SPX Index in Normal volatility

	$Skewness_1$	$Skewness_2$	$Kurtosis_1$	$Kurtosis_2$
SPX Index	-0.001	0.002	3.589	6.225
SXXP Index	-0.001	0.002	3.501	4.601
MXEF Index	-0.001	-0.003	3.330	5.873
TPX Index	0.000	0.000	3.363	5.605
CL1 Comdty	0.000	0.011	3.389	4.136
AUDUSD Curncy	0.000	-0.001	3.359	5.813
JPYUSD Curncy	0.000	0.004	3.391	5.434
BRLUSD Curncy	-0.001	0.001	3.477	6.560
EURUSD Curncy	0.000	0.001	3.325	3.599
RUBUSD Curncy	0.000	0.003	3.8	30
MXNUSD Curncy	0.000	-0.001	3.44	6.3

Table 2.7: Markov Switching Skewness and Kurtosis

different state for SPX Index and TPX Index. They also found the same evidence for China.

Normal Volatility	High Volatility
10%	30%
20%	20%
15%	30%
20%	40%

Table 2.8: Different VIX Level used as Prior

To check the consistency of out estimation, we run Gibbs Sampling algorithm with different prior parameters. Table 2.9 exhibits point estimation of parameters for different variance prior means and large variance. Specifically, VIX levels, used as prior of expected value, are reported in Table 2.8. We observe that the results remain unchanged for different values of VIX. This is due to low confidence level on the prior of the expected value (large prior variance). Specifically when the prior distribution is combined with the likelihood the weight of the prior distribution is quite small. Similar results are obtained changing the expected value of μ and both prior means for μ and h.

The Bayesian framework enables the financial operators to take both their market views and the sample data into account. It is important that the financial operators are aware of the effect of the weight chosen for a subjective view. Tables 2.10 shows how a strong prior on the prior mean requires that a small variance prior is imposed to establishing itself on the sample data. Suppose, for example, that an exceptional event as the US-China trade war, lead a financial operator to a strong view of upside in case of agreement and of a global risk-off in case of growing frictions between parties.

Table 2.10 provides an extreme scenario where $\mu_1 = 3\%$ and $\mu_2 = -3\%$. With this so strong view is necessary a very small variance to allow the posterior to converge versus the prior. In this case the contribute of the data in the posterior distribution is completely neutralized by the prior distribution. This is a important warning for a financial operator because tells him that

	μ_1	μ_2	σ_1	σ_2	P_1	P_2
SPX Index	0.07%	-0.07%	0.74%	1.86%	98.96%	97.78%
SPX Index	0.07%	-0.07%	0.74%	1.86%	98.96%	97.77%
SPX Index	0.07%	-0.07%	0.72%	1.84%	98.88%	97.73%
SPX Index	0.07%	-0.07%	0.72%	1.87%	98.90%	97.63%
MXEF Index	0.10%	-0.21%	0.82%	2.04%	98.70%	95.83%
MXEF Index	0.10%	-0.22%	0.82%	2.04%	98.71%	95.79%
MXEF Index	0.10%	-0.21%	0.82%	2.03%	98.73%	96.00%
BRLUSD Curncy	0.02%	-0.11%	0.61%	1.76%	97.95%	95.12%
BRLUSD Curncy	0.02%	-0.11%	0.61%	1.75%	98.00%	95.32%
BRLUSD Curncy	0.02%	-0.11%	0.62%	1.79%	97.96%	94.94%
MXNUSD Curncy	0.05%	-0.10%	0.49%	1.25%	98.52%	94.26%
MXNUSD Curncy	0.05%	-0.10%	0.49%	1.27%	98.59%	94.25%
MXNUSD Curncy	0.04%	-0.10%	0.50%	1.25%	98.68%	94.85%
MXNUSD Curncy	0.05%	-0.11%	0.51%	1.34%	98.71%	93.94%

Table 2.9: Different Prior mean parameters for σ

his view is completely incompatible with the data.

2.6.3 Three State Univariate Markov Switching.

The choice of dealing with 2 regimes is well justified by the strong evidence that returns exhibit clear cluster in terms of Low versus High Volatility Regime. However we repeat the analysis using 3 regimes. We suppose a Normal Volatility, High volatility and Ultra-High volatility state. As in the previous case we deal with quite uninformative prior. Since also this analysis is provided with daily returns, we assume again μ_{01}, μ_{02} and μ_{03} close to zero, positive the former and negative the last two. We confirm a VIX Level of 10% and 30% for Normal an High Volatility state and we suppose a level of 60% for the ultra High Volatility. SPX Index results are provided in Table 2.11, Table 2.12 and Figure 2.12.

Table 2.11 shows the point estimations of the parameters. Observe that the Normal Volatility parameters remain unchanged respect to the 2-State Markov Switching, while parameters of High Volatility and Ultra High Volatility identify two overlap regimes, Figure 2.12 shows that the two estimated

	μ_1	μ_2	σ_1	σ_2	P_1	P_2	σ_0
SPX Index	0.07%	-0.07%	0.73%	1.86%	98.96%	97.79%	3.00%
SPX Index	0.07%	-0.07%	0.73%	1.86%	98.96%	97.78%	2.79%
SPX Index	0.08%	-0.07%	0.71%	1.84%	98.84%	97.63%	2.57%
SPX Index	0.07%	-0.07%	0.74%	1.86%	98.97%	97.79%	2.36%
SPX Index	0.07%	-0.07%	0.73%	1.86%	98.96%	97.78%	2.15%
SPX Index	0.07%	-0.07%	0.74%	1.86%	98.97%	97.80%	1.93%
SPX Index	0.07%	-0.07%	0.73%	1.86%	98.96%	97.77%	1.72%
SPX Index	0.07%	-0.07%	0.74%	1.86%	98.97%	97.78%	1.51%
SPX Index	0.08%	-0.07%	0.71%	1.84%	98.84%	97.64%	1.29%
SPX Index	0.07%	-0.06%	0.73%	1.86%	98.96%	97.78%	1.08%
SPX Index	0.07%	-0.06%	0.73%	1.86%	98.96%	97.79%	0.86%
SPX Index	0.07%	-0.05%	0.73%	1.86%	98.96%	97.77%	0.65%
SPX Index	0.07%	-0.04%	0.74%	1.86%	98.97%	97.81%	0.44%
SPX Index	0.06%	0.06%	0.74%	1.87%	98.98%	97.80%	0.22%
SPX Index	-2.93%	2.90%	2.75%	2.75%	51.96%	65.88%	0.01%

Table 2.10: Variance sensitive for μ

	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3
SPX Index	0.07%	-0.09%	-0.07%	0.72%	1.66%	2%

Table 2.11: Point estimation results 3-State Markov Switching

Normal are essentially the same. In others words these two regime have the same characteristics and could be summarized in only one State. Moreover the transition probability $p_{22} = 5\%$ for the SPX Index and the number of observations clustered in the High Volatility suggest that this is only a transition state without own features.

Similar observation arise supposing an Ultra Low Volatility State as third state.

	S_1	S_2	S_3	p_{22}
SPX Index	70%	7%	23%	5%

Table 2.12: State frequency

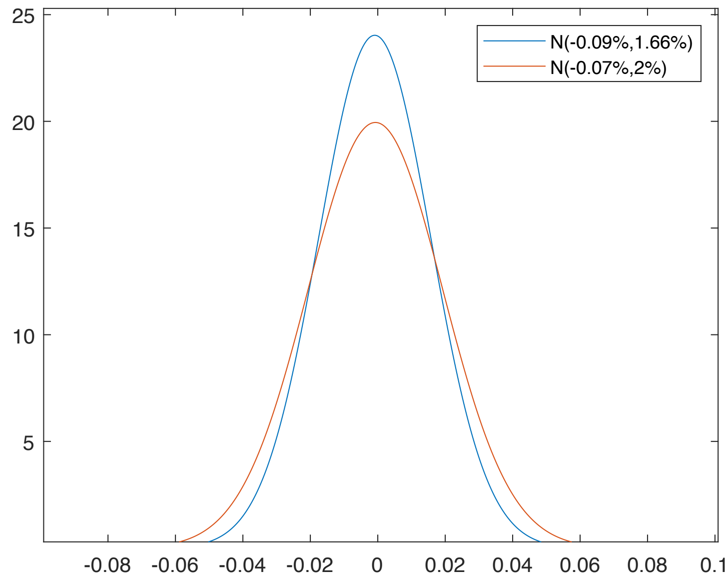

Figure 2.12: pdf of High Volatility vs pdf of Ultra High Volatility

The provided results legitimize the choice of two State corresponding to Low Volatility and High Volatility. This choice not only is most intuitive since the two regime are identified from a financial point of view (Risk-on vs Risk-off) but also allows to take the parameters number under control.

2.6.4 Multivariate Markov Switching

The univariate Markov Switching models do not take into account the correlations between asset classes. As the means and the volatilities, even the correlations are not constant and depend on the current state. The knowledge of the correlations is a fundamental issue in the construction of diversified trading strategy. Market experience lead us to believe that correlations increase in High Volatility regime. As regard the number of state, previous results have confirmed that the 2-State solution is a optimal trade-off between the number of unknown parameters and the catch of non-stationarity features of daily returns.

Thus, we perform a multivariate Markov Switching model which is estimated according to Markov Chain Monte Carlo algorithms, in particular Gibbs Sampling and we set the number of regimes equal to 2. As sad in previous chapter, we suppose an independent Normal-Wishart distribution for μ and Σ and a Beta distribution for the transition probabilities (see equation 2.22, 2.23)

$$\begin{aligned}
&\mu_i = \mathcal{N}(\mu_{i0}, \Sigma_{i0}) \\
&H_i = W(\nu_{i0}, H_{i0}) \quad H = \Sigma^{-1} \\
&p_{11} \sim \beta(u_{11}, u_{12}) \\
&p_{22} \sim \beta(u_{22}, u_{21})
\end{aligned}$$

where $\mu_i \in \mathbb{R}^K$, $H_i \in \mathbb{R}^{K\times K}$ and $p_{ii} \in \mathbb{R}$ with $i = 1, 2$ and K is the number of assets. Given the size of the parameters space, we have decided to move towards the alternative solutions:

1 Run the Gibbs Sampling supposing a proxy for $S_1, \dots, S_T$

2 Impose informative prior

In the first solution $S_1, \dots, S_T$ are fixed equal to the estimated states for the SPX Index in the univariate case. We chose the SPX Index because it could be assumed as general proxy of market sentiment. This solution is computationally simple but inflexible.

In the latter solution, let us consider a pre-sample composed of the first 500 observations $\bar{r}_{ps} = \{r_1, \dots, r_{500}\}$, where $r_t \in \mathbb{R}^{K\times 1}$ is the vector of asset returns in $t = 1, \dots, 500$. Using the states estimated for the SPX Index in the univariate case, S_t, $\bar{r}_t$ is divided in the two following sub-sample

$$r_1 = \{r_t \; s.t \; S_t = 1, \; t = 1, \dots, 500\}$$

$$r_2 = \{r_t \; s.t \; S_t = 2, \; t = 1, \dots, 500\}$$

Thus $\mu_{ps1}, \mu_{ps2}, \Sigma_{ps1}$ ans Σ_{ps2} are calculated from the two sub-sample and the variance priors are set quite large but able to target the model. Here again $H = 4$ Gibbs sampling are performed with different starting points. With $h = 100.000$ iterations we obtain uniform results in all the Gibbs sampling. The first 10% of the draws is discarded

Table 2.13 and Figure 2.13 display the point estimates of mean e volatility

Table 2.13: Multivariate Markov Switching Results

	μ_1	μ_2	σ_1	σ_2
SPX	0.058%	-0.080%	0.77%	1.95%
SXXP	0.065%	-0.147%	0.81%	1.96%
MXEF	0.104%	-0.187%	0.83%	1.91%
TPY	0.076%	-0.168%	1.02%	2.03%
CL1	0.1%	-0.095%	1.83%	3.47%
AUDUSD	0.04%	-0.08%	0.6%	1.24%
JPYUSD	-0.012%	0.03%	0.53%	0.90%
BRLUSD	0.02%	-0.10%	0.72%	1.68%
EURUSD	0.01%	-0.01%	0.53%	0.86%
RUBUSD	0.09%	-0.08%	0.44%	1.34%
MNXUSD	0.01%	-0.08%	0.53%	1.06%

vectors in the two states. The point estimations of ergodic probabilities are

$$\pi_1 = 73.30\%$$

$$\pi_2 = 26.70\%$$

The results are consistent with those in the univariate case and all the previous considerations remain valid.

Let focus on the correlations of the two states. The correlations matrices are given at Table 2.14 and 2.15. The High Volatility state is characterize by an increase of correlations. For example, the correlations in High Volatility between the SPX Index and other assets are, on average, 50%higher those in

Normal Volatility. Thus, in High volatility the diversification could be only apparent, because all assets tend to move in the same direction. We highlight that the Japanese Yen is negative correlated with other Asset Class.

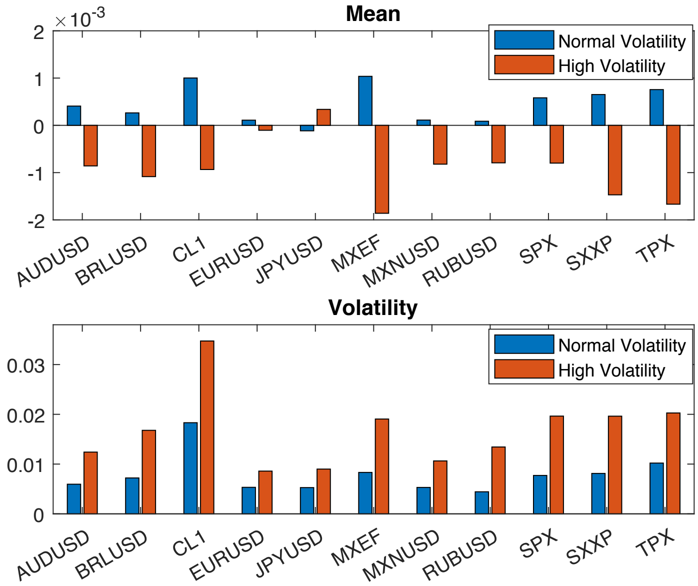

Figure 2.13: Multivariate μ and σ

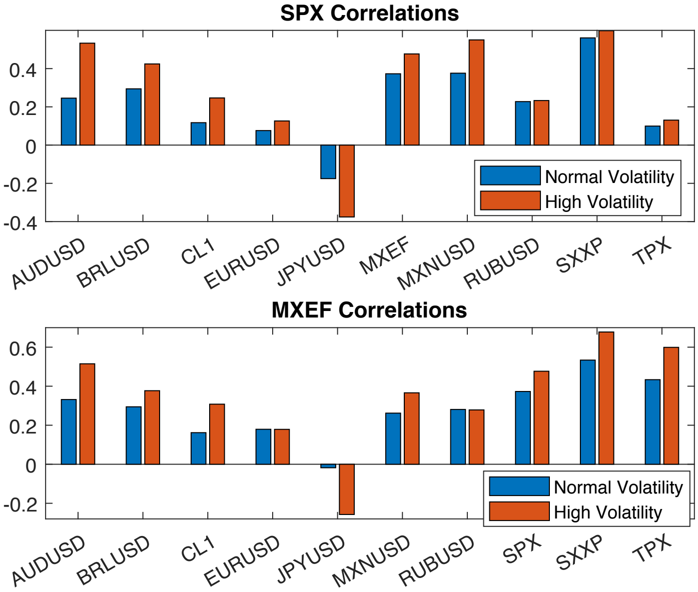

Figure 2.14: Correlations

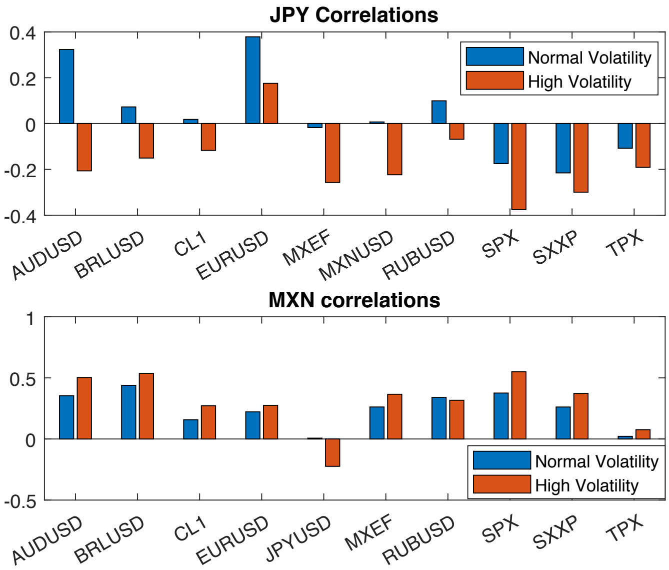

Figure 2.15: Correlations

Table 2.14: Correlations in Normal Volatility

	SPX	SXXP	MXEF	TPX	CL1	AUDUSD	JPYUSD	BRLUSD	EURUSD	RUBUSD	MXNUSD
SPX	100%	56.06%	37.29%	9.94%	11.72%	24.55%	-17.52%	29.42%	7.59%	22.74%	37.60%
SXXP	56.06%	100%	53.38%	23.47%	9.97%	14.82%	-21.54%	17.75%	-8.12%	17.98%	26.18%
MXEF	37.29%	53.38%	100%	43.35%	16.19%	33.17%	-1.78	29.45%	17.93%	28.09%	26.20%
TPX	9.94%	23.47%	43.35%	100%	1.88%	4.47%	-10.77%	2.69%	-0.82%	2.44%	2.20%
CL1	11.72%	9.97%	16.19%	1.88%	100%	24.27%	1.79%	11.60%	15.30%	25.39%	17.71%
AUDUSD	24.55%	14.82%	33.17%	4.47%	24.27%	100%	32.31%	30.92%	57.98%	36.17%	35.39%
JPYUSD	-17.52%	-21.54%	-1.78%	-10.77%	1.79%	32.31%	100%	7.26%	37.87%	9.92%	0.70%
BRLUSD	29.42%	17.75%	29.45%	2.69%	11.60%	30.92%	7.26%	100%	23.58%	27.17%	27.17%
EURUSD	7.59%	-8.12%	17.93%	-0.82%	15.30%	57.98%	37.87%	23.58%	100%	35.70%	22.18%
RUBUSD	22.74%	17.98%	28.09%	2.44%	25.39%	36.17%	9.92%	27.17%	35.70%	100%	34.02%
MXNUSD	37.60%	26.18%	26.20%	2.20%	17.71%	35.39%	0.70%	43.93%	22.18%	34.02%	100%

Table 2.15: Correlations in High Volatility

	SPX	SXXP	MXEF	TPX	CL1	AUDUSD	JPYUSD	BRLUSD	EURUSD	RUBUSD	MXNUSD
SPX	100%	59.80%	47.67%	13.05%	24.66%	53.30%	-37.58%	42.43%	12.62%	23.30%	55.01%
SXXP	59.80%	100%	67.78%	36.25%	29.77%	41.78%	-29.95%	34.45%	2.12%	22.29%	37.31%
MXEF	47.67%	67.78%	100%	59.91%	30.79%	51.49%	-25.73	37.70%	17.89%	27.86%	36.61%
TPX	13.05%	36.25%	59.91%	100%	11.85%	21.52%	-19.11%	8.45%	4.38%	9.67%	7.60%
CL1	24.66%	29.77%	30.79%	11.85%	100%	32.99%	-11.75%	24.4%	20.52%	35.67%	27.21%
AUDUSD	53.3%	41.78%	51.49%	21.52%	32.90%	100%	-20.66%	42.43%	52.42%	30.23%	50.35%
JPYUSD	-37.58%	-29.95%	-25.73%	-19.11%	-11.75%	-20.66%	100%	-15.06%	17.53%	-6.86%	-22.36%
BRLUSD	42.43%	34.45%	37.70%	8.44%	24.4%	42.43%	-15.06%	100%	22.85%	25.89%	53.72%
EURUSD	12.62%	2.12%	17.89%	4.38%	20.52%	52.42%	17.53%	22.85%	100%	21.41%	27.57%
RUBUSD	23.30%	22.29%	27.86%	9.67%	35.67%	30.23%	-6.86%	25.89%	21.41%	100%	31.72%
MXNUSD	55.01%	37.31%	36.61%	7.60%	27.21%	50.35%	-22.36%	53.72%	27.57%	31.72%	100%

2.6.5 Regime-based Strategy

The analysis made in the previous section carries out the existence of two states clearly separated. We want to investigate if the regime affects the performance and the risk profile of each strategy. In this section we compare a regime-based strategy against a simple buy-and-hold which is the portfolio full invested into the market. This is the typical effort in financial industry in order to evaluate the ability of the quantitative strategy to gain money and avoid downtrend with respect to the buy-and-hold portfolio. The aims of the trading strategy is twofold: on one hand, it aims to reduce volatility and drawdown respect to a buy-and-hold strategy when the market moves down and investors' risk aversion increases dramatically as the demand for safe assets. Secondly, it is important that the trading strategy is able to follow the market during the Up-trend when the investors are in a "risk-on" mood, condition where the risk-appetite improves sharply and investors' demand for risky assets become supportive. First of all, the assets were divided in two baskets: risky assets and defensive assets.

Table 2.16: Baskets

Risky Asset	Defensive Asset
SPX 500,SXXP,MXEF,TPX,CL1	JPYUSD
AUDUSD,BRLUSD,RUBUSD,MXNUSD,	EURUSD

The former is composed of equity indexes, oil futures and carry trade currencies. Euro and Yen are in the defensive basket. This split was made based on the volatility structure in High volatility. Recall that Euro and Yen exhibit low volatilities even in this states. For example, the volatility of the Yen is half of that of the SPX index.

A first estimate of the model parameters is made at the end the end of 2011. Then they are kept constant for 1 year and updated every end of December. We present in-sample and out-sample results, the latter since 2011. We use the active parameters in each instant t to run the Hamilton Filter which

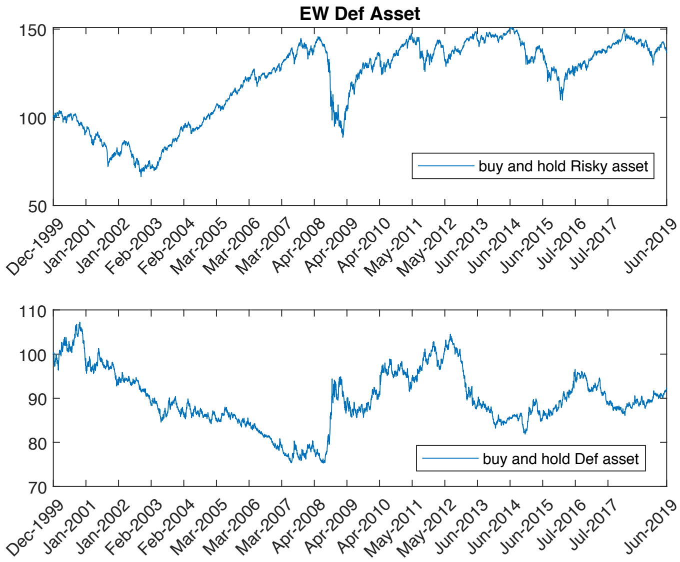

Figure 2.16: Buy-and-hold-strategy

gives the probability to be in Normal and in High Volatility regime (see 2.9).

First of all we studied the features of the two baskets in Normal and High Volatility regime. Figure 2.16 exhibits the equally weighted portfolio of the risky assets and of the defensive assets respectively. It is already clear the difference nature of the two groups. Moreover, we split the EW strategies in two sub-strategies that takes Long positions only in one state: EW Long in Normal Volatility (EWLN) and EW Long in High Volatility (EWLH). With respect to risky assets, the analysis highlights how a portfolio of risky assets strongly outperform the overall equally-weighted portfolio when markets are in normal volatility (Figure 2.17); in this environment it happens that a portfolio of risky assets is able to reduce its volatility by 50% and to improve impressively the active returns with respect to the overall global portfolio. On the other hands, a portfolio of risky assets tends to exhibit absolute negative performance when markets switch into the High Volatility regime

(Table 2.17).

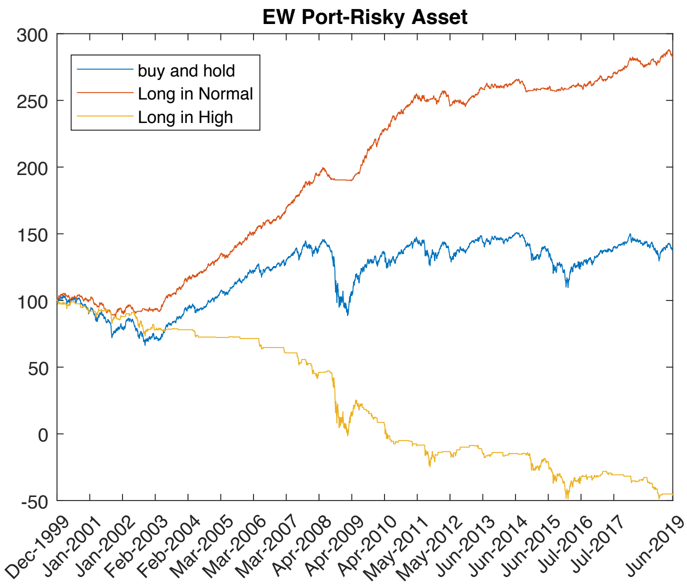

Figure 2.17: Risky Asset

Table 2.17: Risky Assets

	buy-and-hold	buy in Normal	buy in High
α	1.98%	9.75%	-7.77%
TE	11.79%	6.64%	9.71%
SR	0.17	1.47	-0.80

Table 2.18: Defensive Asset

	buy-and-hold	buy in Normal	buy in High
α	-0.42%	-1.99%	1.56%
TE	6.12%	4.12%	4.51%
SR	-0.07	-0.48	0.35

With respect to defensive assets, we have an opposite picture : the portfolio of defensive assets has a negative track record in normal volatility because of a strong demand for risk by the market, but this portfolio is able to provide a solid strong performance in high volatility with an important control of the overall tracking error of the strategy (Figure 2.18, Table 2.18).

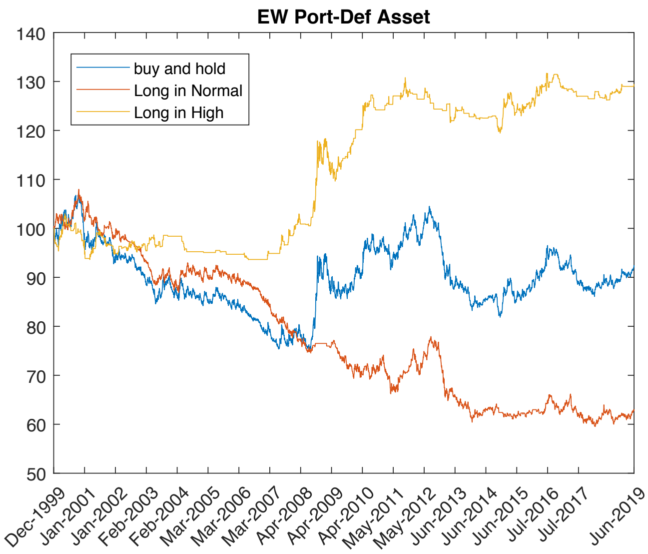

Figure 2.18: Defensive Asset

The main results of this exercise is consistent with the evidence that markets demand of risky assets is particularly strong in the normal volatility regime when the appetite for risk is structurally high, while investors tend to sell risky assets when the volatility of returns is high, fueling risk aversion sentiment and demand for safe assets.

Based on this results, we consider the following portfolio, called Markov Switching Equally weighted portfolio (MSEW):

- Long position in the equally weighted portfolio of risky assets when returns are in normal volatility.
- Long position in the equally weighted portfolio of defensive assets when returns are in high volatility.

More specifically, we consider the system in Normal Volatility if the UpDate of the Hamilton Filter is greater than 50%:

$$P(S_t = 1|\bar{r}_t; \theta_j) > 50\%.$$

The full sample results are displayed in Figure 2.19 where we find the Markov Switching EW (MSEW) portfolio versus the buy-and-hold EW portfolio of all assets (EW). Observe that the MSEW portfolio outperform the EW portfolio. The Sharpe ratio of the MSEW portfolio is twice that of the EW portfolio (Table 2.19). Moreover the MSEW portfolio is able to avoid the drawdown that are in EW portfolio.

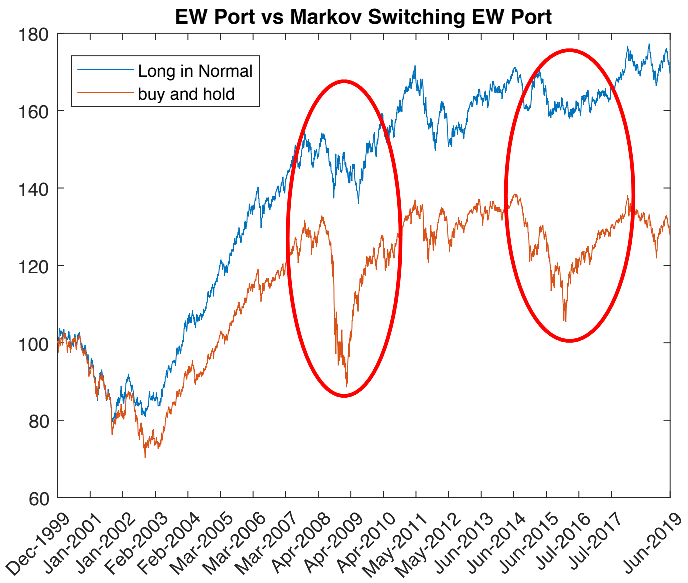

Figure 2.19: Markov Switching EW Portfolio vs EW Portfolio

Table 2.19: Markov Switching EW Portfolio vs EW Portfolio

	Markov Switching EW Portfolio	EW Portfolio
α	3.73%	1.51%
TE	8.52%	9.25%
SR	0.43	0.16

Figure 2.21 and Figure 2.22 display the year on year Information Ratio (IR YoY) and the portfolio drawndown during years of negative performance of the EW portfolio. It is pretty clear that the MSEW portfolio is able

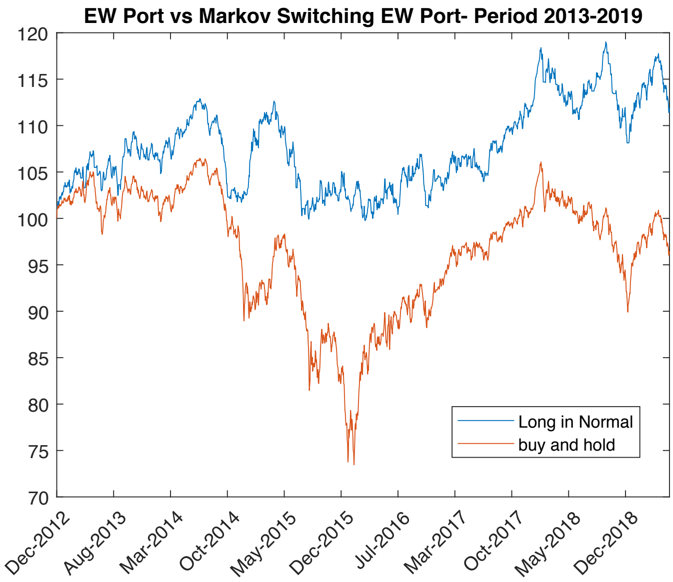

Figure 2.20: Markov Switching EW Portfolio vs EW Portfolio- Period 2013-2019

to reduce dramatically the drawndown and improve by 50% the IR of the strategy.

YoY IR	MSEW	EW
2014	-0.84	-1.67
2015	0.07	-0.81
2018	-0.55	-1.4

Table 2.20: Out of sample MSEW IR vs EW IR

Traditionally, the problem of this kind of smoothing strategies is the inability to follow the market during the up-trend. Figure 2.23 shows that MSEW strategy overcome this trap and thanks to the sensibility of the Hamilton Filter, the Long position on risky asset are taken with a good timing and so the portfolio is able to follow the up-trend of the market.

At the end we consider the transaction cost and compute the net IR.

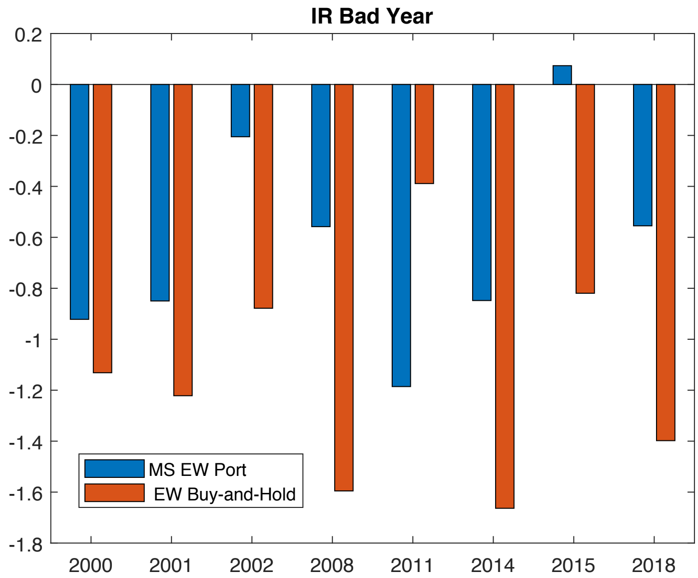

Figure 2.21: Negative IR YoY

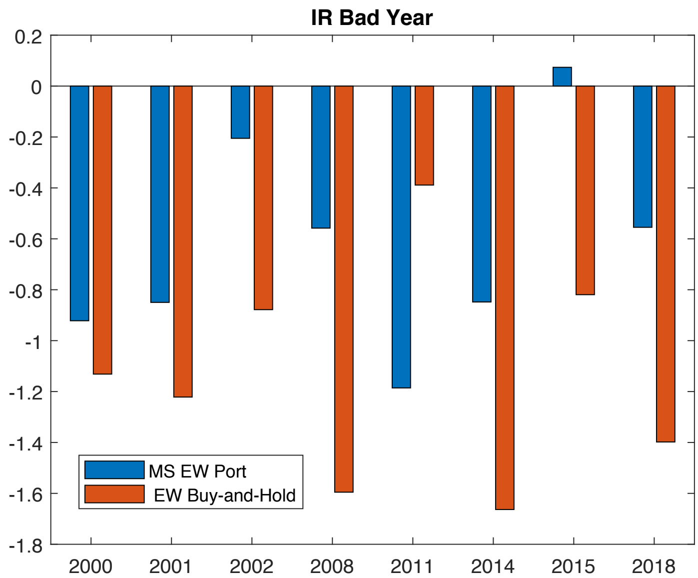

Figure 2.22: Bad years maximum Drawdown

Note that, the turnover of this strategy is quite low by definition, since we are invested in the risky portfolio in the 75% of the time, as pointed out the

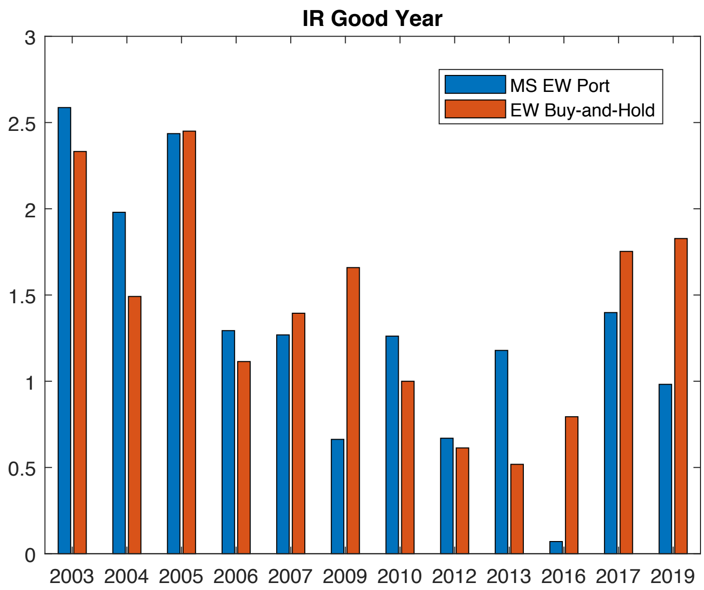

Figure 2.23: Positive IR YoY

estimated ergodic probability in the previous section.

The assets that could more impact on the total cost of the strategies are the carry trade currencies. However, recall that we choose the carry trade currencies with the lower transaction cost and higher liquidity since they are highly linearly correlated. Table 2.21 shows that the transaction costs are under-control and the IR remains interesting.

This trading strategy could be linked with the strategy proposed by [Bulla et al. 2011]. They introduce an univariate state-based trading strategy that take long position on one of the analyzed indexes in Normal Volatility and invest in risk-free asset in high Volatility. As in our case, all these strategy reduce the risk and outperform the corresponding indexes. However, Multivariate Markov Switching model gives a global asset allocation and so a really diversified portfolio in Normal Volatility. Moreover, taking correlations into account, it was possible to identify a investment solution more attractive than the risk-free-strategy.

	Net IR	Gross IR	EW		Net IR	Gross IR	EW
2000	-1.05	-0.92	-1.13	2010	1.15	1.26	1.00
2001	-0.98	-0.85	-1.22	2011	-1.30	-1.19	-0.39
2002	-0.30	-0.21	-0.88	2012	0.58	0.67	0.61
2003	2.44	2.59	2.33	2013	1.01	1.18	0.52
2004	1.90	1.98	1.49	2014	-0.96	-0.85	-1.66
2005	2.30	2.44	2.45	2015	0.01	0.07	-0.82
2006	1.21	1.29	1.11	2016	0.04	0.07	0.79
2007	1.20	1.27	1.39	2017	1.30	1.40	1.75
2008	-0.62	-0.56	-1.60	2018	-0.66	-0.55	-1.40
2009	0.58	0.66	1.66	2019	0.91	0.98	1.83

Table 2.21: Net IR vs Gross IR

2.7 Conclusion

Returns of financial markets tend to change their patters over time in terms of average returns, serial correlation and volatility. Typically, asset returns are used to exhibit a stochastic behavior which is not consistent with a Gaussian distribution, due to time-varying volatility, asymmetry of returns and fat tails.

According to this picture, this study proposed a solution based on a Markov Switching Mixture of Normal distributions in order to manage the non-stationarity of returns. In this solution, returns are supposed to follow a state-space representation where the unconditional likelihood of returns is not normal but rather it is a weighted average of conditional normal distribution of each state with weights equal to the state's probability. The normal distribution of any state is well identified by a particular configuration of parameters μ and σ.

After introducing the general aspects of Markov Switching Models, we implemented a case study based on a 2 State-Markov Switching Mixture of Normal distributions, where the first regime is linked to the Bull Market scenario with positive returns in mean and low volatility across assets, while the second is in connection with the Bear Market case with negative returns in mean and high volatility.

In the multivariate case we highlight how the correlation between asset classes tend to change quite sharply when the structure of returns moves from one regime to another. In particular our study points out that a spike of the general level of volatility of returns generates a strong rise in the correlation between asset classes and that reduce quite dramatically the benefit of the diversification. In this environment the portfolio construction without taking into account this type of switching in the covariance matrix is not able to produce the right allocation into defensive assets in order to reduce the overall portfolio drawndown.

In addition, the solution proposed in this work is able to recover the normality of returns conditional on the single state. This conditional normality is relevant because it allows to deal with any king result based on the assumption of normality and homoscedasticity.

Bibliography

[Albert and Chib (1993)] Albert,J., Chib, J. S., 1993. Bayes inference via gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Buisness & Economic statistics* 11:1-15

[Ang Timmermann (2012)] Ang, A., Timmermann, A., 2012. Regime Changes and Financial Markets. *Annual Reviews* 4:313-337

[Bollerslev (1986)] Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Ecometrics* 31: 307-327

[Bulla et al. 2011] Bulla,J., Mergner,S., Bulla,I., Sesbouè,A.,Chesneau, C., 2011. Markov-switching asset allocation: Do profitable strategies exist?. *Journal of Asset Management* 5:310-321

[Chevallier (2012)] Chevallier, J., 2012. Global imbalances, cross-market linkages, and the financial crisis: A multivariate Markov switching analysis. *Economic Modelling* 29:943-973.

[Chib (2001)] Chib, S., 2001. Markov Chain Monte Carlo Methods: Computation and Inference. Handbook of Econometrics

[Di Persio and Frigo (2016)] lib6] Di Persio, L., Frigo, M., 2016. Gibbs sampling approach to regime switching analysis of financial time series. *Journal of Computational and Applied Mathematics* 300,43-55

[Di Persio and Vettori (2014)] i Persio, L., Vettori, S., 2014. Markov Switching model analysis of implied volatility for market indexes with application to S&P 500 and DAX. *Journal of Mathematics*

[Geman and Geman] Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *Transactions on Pattern Analysis and Machine Intelligence* 12: 609-628.

[Gelman and Rubin (1992)] Gelman, A., Rubin,D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 4:457-511

[Geweke and Whiteman] Geweke, J.,Whiteman, C., 2006. Bayesian Forecasting. Handbook of Economic , vol 1 Elsevier B.V.

[Koop (2003)] Koop, G.,2003. Bayesian Econometric. Wiley

[Koop et al.(2007)] Koop, G., Poirier, D. J., Tobias, J.,L.,2007. Bayesian Econometric Methods. Cambridge University Press

[Metropolis,Rosenbluth,Teller 1953] Metropolis N, Rosenbluth, A. W., Teller, A.H., Telle, E., 1953. Equations of the state calculations by fast computing machine. *Journal of Chemical Machine* 21:1087-1092

[Pereiro,Gonzalez (2015)] Pereiro, L.E., Gonzalez-Rozada M., 2015. Forecasting Price in Regime-Switching Markets. *The Journal of Portfolio Management.* 41: 133-139

[Guidolin et al. (2011)] Guidolin, M., Ria,F., 2011. Regime shifts in mean-variance efficient frontiers: Some international evidence. *Journal of Asset Management* 12,5,322-349.

[Hamilton 1989] Hamilton, J.,D., 1989 A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357-384

[Johannes and Polson] , M., Polson, N., 2010. MCMC Methods for Continuous-Time Financial Econometrics. Handbook of Financial Econometrics: Applications

[Kim and Nelson (1999)] Kim, C., Nelson C., 1999. State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press, Cambridge

[Timmermann (2000)] Timmermann, A., 2000. Moments of Markov switching models. *Journal of Ecometrics* 96:75-111

2.8 Markov Switching ΛVaR

In this section we use Markov Switching models for the estimation of the Lambda Value at Risk. More specifically, Markov Switching models could give the maximum flexibility to the ΛVaR since allow the use of an increasing or decreasing Λ based of the Hamilton Filter UpDate in each t, in order to take the number of overdraft under control and reduce the capital aside which could be invested.

The ΛVaR introduced by [Frittelli et al.2014] and studied by [Burzoni et al. (2017)] has theoretical interesting features, overcomes some lacks of VaR and satisfies important property for a risk measure.

The construction and the backtesting of the ΛVaR are further key issues in the validation of a risk measure. Recall that the ΛVaR is define as follow:

$$\Lambda VaR(F) := -\inf\{x \in \mathbb{R} : F(x) > \Lambda(x)\}$$

where $\Lambda : \mathbb{R} \to [\lambda^m, \lambda^M]$ with $0 < \lambda^m \leq \lambda^M < 1$ is a right continuous and monotone function. When the Λ function is constantly equal to some $\lambda \in (0, 1)$ it coincides with the definition of VaR with confidence level λ. In other words, its confidence level Λ depends on the market returns instead to be a constant λ.

It is clear that in the construction of the ΛVaR an important role is played by the function Λ. One methodological proposal of ΛVaR estimation, called dynamic benchmark approach was introduced by [Hitaj et al. (2018)]. The function Λ is dynamic since is estimated in each t and it is calibrated through the tail distributions of selected benchmarks. This makes ΛVaR sensitive to markets changes. As regards the direction, they recommend to take a decision conditioned on the markets status. Specifically, a decreasing Λ is suggested in Normal Volatility and an increasing Λ in High Volatility. However, a time series dynamic switching between the two direction of Λ is not suggested. In this section we want to exploit at maximum level the flexibility of the ΛVaR and replicating the dynamic benchmark approach of [Hitaj et al. (2018)] we want to add the possibility to use one of the possible direction of Λ in each t. More specifically we have

$$\Lambda VaR_t(F_t) := -\inf\{x \in \mathbb{R} : F_t(x) > \Lambda_t(x)\}$$

where

$$\Lambda \text{ is } \begin{cases} \text{increasing} & \text{in High Volatility} \\ \text{decreasing} & \text{in Normal volatility} \end{cases}$$

where $F_t = F(r|\mathcal{I}_{t-1})$ and $\mathcal{I}_{t-1}$ is all the available information in $t-1$. We consider the system in High Volatility if the UpDate of the Hamilton Filter is greater than 40%. This threshold was chosen to be enough conservative.

$$P(S_t = 2|\bar{r}_t; \theta_j) > 40\%.$$

In the next section we briefly recall the the dynamic benchmark approach estimation and the backtesting procedure introduced by [Hitaj et al. (2018)] and [Corbetta and Peri (2017)].

2.8.1 On the estimation and the backtesting of the ΛVaR

The dynamic benchmarks approach estimation process is based on the following steps:

1 Fix λ_m and Λ_M

2 Decide the direction of Λ

3 Decide the functional form of Λ

4 Estimate the Λ parameters

Note that Λ_M, called Λ confidence level, depend on the risk aversion of the financial operator. As regarding the λ_m [Corbetta and Peri (2017)] suggest to fix it equal to 0.005; The function form used in [Hitaj et al. (2018)] and

[Corbetta and Peri (2017)] is a linear interpolation of n points, (π_i, λ_i) $i = 1, \ldots, n$, in particular we have in the increasing case

$$\Lambda(x) = \begin{cases} \lambda_1 & x < \pi_1 \\ \sum_{i=1}^{n-1} \mathbf{1}_{[\pi_i, \pi_{i+1})} \left((\mathbf{x} - \pi_\mathbf{i}) \frac{\lambda_{\mathbf{i+1}} - \lambda_\mathbf{i}}{\pi_{\mathbf{i+1}} - \pi_\mathbf{i}} + \lambda_\mathbf{i} \right) & \pi_1 \leq x < \pi_n \\ \lambda_n & x \geq \pi_n \end{cases}$$

and in the decreasing case

$$\Lambda(x) = \begin{cases} \lambda_n & x < \pi_1 \\ \sum_{i=1}^{n-1} \mathbf{1}_{[\pi_i, \pi_{i+1})} \left((\mathbf{x} - \pi_\mathbf{i}) \frac{\lambda_{\mathbf{n-i}} - \lambda_{\mathbf{n-i+1}}}{\pi_{\mathbf{i+1}} - \pi_\mathbf{i}} + \lambda_{\mathbf{n-i+1}} \right) & \pi_1 \leq x < \pi_n \\ \lambda_1 & x \geq \lambda_n \end{cases}$$

The estimation of the π_i, $i = 1, \ldots, n$ are made on selected benchmarks while a neutral approach is chosen for the λ_i, $i = 1, \ldots, n$.
In the empirical application [Hitaj et al. (2018)] fix $n = 4$ as good trade off trade off between the fitting accuracy and the number of unknown parameter. They set π_1 as the minimum return of all benchmarks returns in the window analysis and π_2, π_3 and π_4 as the minimum, mean and maximum $\lambda\% - VaR$ of all benchmarks. As regards the y-axix, λ_i are fixed equal to equidistant points from λ_1 and λ_4.
Once the ΛVaR has been calculated a backtesting procedure is necessary to evaluated the quality of the ΛVaR estimation.
The backtesting of a risk measure is based on the comparison of the real returns,r_t, with that estimated by the risk measure in $t-1$ for t. A violation occurs when the risk measure forecast is not able to cover the realized return. Specifically, let r_t and y_t the realized return and the forecast return at time $t = 1, \ldots, T$, a violation I_t occurs if $r_t < y_t$. To perform the backtesting of risk measure, it is necessary to construct the sequence of a random variable $I_1, \ldots, I_T$ such that

$$I_t = \begin{cases} 1 & r_t < y_t \\ 0 & \text{otherwise} \end{cases}$$

Note that each I_t follows a Bernoulli distribution $I_t \sim \mathcal{B}(\lambda_t)$. The flexibility introduced by the function Λ lead to a sequence of violations that are not identically distributed and so the standard procedure can not be directly applied. [Corbetta and Peri (2017)] proposed some backtesting procedures to check if the ΛVaR provides an accurate level of coverage. The authors consider the following null vs alternative hypothesis:

$$H_0 : \lambda_t = \lambda_t^0 \qquad \forall\, t$$
$$H_1 : \lambda_t > \lambda_t^0 \quad \text{for some } t.$$

where λ_t^0 is the ΛVaR confidence level in t. One test proposed by the [Corbetta and Peri (2017)] is based on the total number of violations in a window

$$Z := \sum_{t=1}^{T} I_t. \tag{2.35}$$

Since I_t are not identically distributed we have that Z_1 is a Poisson Binomial. The critical region is given by

$$C = \{z_1 \text{ s.t. } Pr(Z_1 \leq z_1) > 1 - \alpha\}$$

One important issue in implement this backtesting procedure is the calculus of the cdf of the binomial Poisson. The direct calculus of this cdf needs a too high computational time. Different solutions were proposed in the literature: Normal approximation, recursive formula, close-form formula. We follow [Hong (2013)] approach that propose a derivation of the exact cdf of the Poisson Binomial through the discrete Fourier transform of the characteristic function of the distribution. We implement [Hong (2013)] algorithm in Matlab.

2.8.2 Empirical results

In this section the we compute the Markov Switching ΛVaR ($1\%{-}MS\Lambda VaR$) and compare its forecast with the increasing $1\% - \Lambda VaR$. We calculate 1-

day ΛVaR and 1-day MS ΛVaR over a time horizon of 250 days and we use historical simulations.
The selection of the benchmarks is a question of primary concern. The analyzed portfolio is multi-asset and multi-country portfolio. It is necessary that the selected benchmarks are representative of all asset universe. The univariate Markov switching pointed out essentially two groups of asset: risky asset and defensive asset. The former group contains equity market, oil future and carry trade currencies while Euro and Yen are part of the latter group. We decide to select three benchmarks, two related with risky-asset and one with defensive asset. Specifically we select as benchmarks MSCI World Index, US dollar index and JP Morgan Emerging Market currencies Index.
In each t the Hamilton Filter UpDate, calculated through a Markov Switching process, legitimizes the choice of one Λ direction since tells us if we are in Normal or High Volatility. Moreover, since the 2-state Markov Switching model highlights in any state a particular parameters configuration, the parameters of the trading strategy are estimated over the last 250 observation of the regime active in that instant t.
Table 2.22 and Figure 2.24 show the results. Note the the number of violations are consistent with those of the $1\%\Lambda VaR$ and during 2008 and 2011 the number of MS-ΛVaR violations is even smaller. Figure 2.24 shows that the two versions of ΛVaR coincide, by construction for some period, but the MS-ΛVaR is more reactive to switch in bull market and so reduce the capital aside.
At the end a Year on Year backtesting procedure is performed and the MS-ΛVaR has highest accuracy, we accept H_0 in the 98% vs the 95% of the usual ΛVaR.

This first application of the Markov Switching models to the process estimation of the ΛVaR shows how the flexibility of the ΛVaR could be

Year	Violations 1%ΛVaR	Violations 1%$MS\Lambda VaR$
2003	1	1
2004	1	1
2005	1	1
2006	3	1
2007	1	1
2008	3	2
2009	1	0
2010	1	1
2011	1	0
2012	0	1
2013	0	0
2014	0	1
2015	2	2
2016	1	0
2017	0	0
2018	1	2
2019	0	0

Table 2.22: Number of Violations

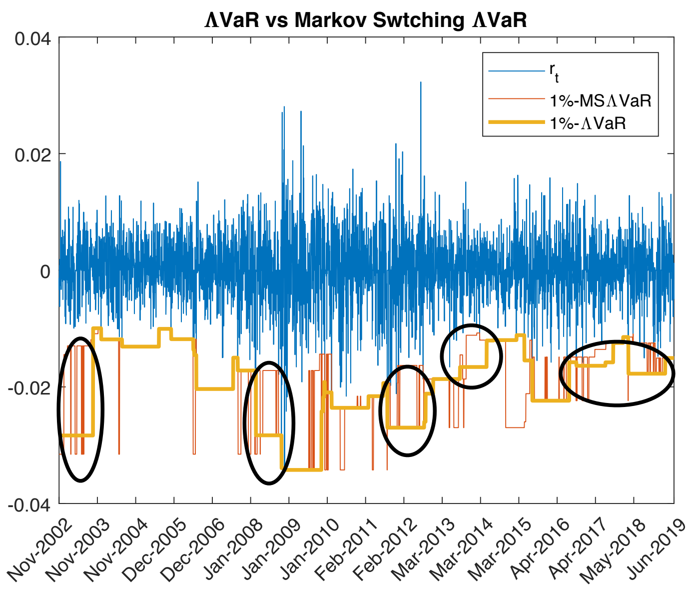

Figure 2.24: ΛVaR vs Markov Switching ΛVaR

exploited fully thought Markov Switching models. In this framework ΛVaR is very reactive to the market changes due to the switching on the Λ direction and the estimation of the empirical cdf of returns conditioned on the state.

Bibliography

[Bracewell] bracewell, R. 2000. The Fourier Transform & its Applications.McGraw-Hill

[Burzoni et al. (2017)] Burzoni, M., Peri, I. Ruffo, C. 2017. On the properties of the Lambda value at risk: robusteness, elicitability and consistency. *Quantitative Finance*, 17(11),1735-1843

[Corbetta and Peri (2017)] Corbetta, J.and Peri, I. 2017., Backtesting Lambda Value at Risk. *European Journal of Finance.*

[Fernández and Williams (2010)] Fernández, M. and Williams, S. 2010. Closed-form expression for the Poisson-binomial probability density function.*IEEE Transctions on Aerospace Electronic System* 46,803-817.

[Frittelli et al.2014] Frittelli, M., Maggis, M. and Peri, I.2014. Risk measures on P(R) and Value at Risk with Probability/Loss function. *Mathematical Finance* **24**, 442-463.

[Hitaj et al. (2018)] Hitaj, A., Mateus, C., Peri,I. 2018. Lambda Value at Risk and Regulatory Capital: A Dynamic Approach to Tail Risk. *Risks* 6,1-18.

[Hong (2013)] Hong, Y. 2013. On computing the distribution function for the Poisson binomial distribution.*Computational Statistics & Data Analysis.*

[Kupiec (1995)] Kupiec, P. 1995. Techniques for Verifying the Accurancy of Risk Measurement Models. *Journal of Derivatives* 3(2):73-84.

2.9 Appendix

In this section we provide the Univariata Markov Switching models estimation and the Hamilton Filter.

```
clear all
close all
clc
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% In questo script stimo i T+6 parametri di un Markov-Switching          %
% univariato attraverso Gibbs Sampling.                                   %
% GIBBS STEPS:                                                            %
% 1- Genero gli stati. Posso generarli uno per volta oppure l'intero ,   %
%     blocco                                                              %
% 2- Genero le probabilit  di transizione dati gli stati                  %
% 3- Genero i paramentri delle normali dei due stati dati i dati e gli   %
%     stati
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
data_start=729757;
%% 1) Carico le serie storiche
cd('Y:\Chiara\TESI\Bloomberg_DB') %%%%%%%%%%%%%%
load('REPORT_DB_Tesi') %%%%%%%%%%%%%%
DateList=REPORT_DB_Tesi.DateList;
[aa, idx2]=ismember(data_start,DateList);
DateList=DateList(idx2:end);
DB_MMName=REPORT_DB_Tesi.DB_MMName;
%% Cerco l'indice che voglio analizzare
idx=find(strcmp('SPX Index',DB_MMName));
Y=REPORT_DB_Tesi.DB_MM(idx2:end,idx);
clear idx
%% Variabili
Nstati=2;
% Regressori
T=length(Y);
X=ones(T,1);
%% 2) Inserisco i parametri delle prior
% a) Prior media --> Normal
mu01=0.0005;
sigma01=0.03; % vola

mu02=-0.001;
sigma02=0.03;
% b) Prior Precision --> Gamma
VixlevelNormal=0.1/sqrt(252);
h1=1/VixlevelNormal^2;
v01=20;
delta01=v01/h1;

VixlevelHigh=0.3/sqrt(252);
h2=1/VixlevelHigh^2;
v02=20;
delta02=v02/h2;

%Prior Transition Probability --> Beta
u012=1;
u011=(0.98 /0.02)*u012;
```

```
u021=1;
u022=(0.95 /0.05)*u021;
%% K numero di iterazioni
NGibbs=4;
K_Gibbs=10000;
kk_Gibbs=1;
mu=zeros(K_Gibbs,Nstati,NGibbs);
sigma=NaN(K_Gibbs,Nstati,NGibbs);
State=NaN(T,K_Gibbs,NGibbs);
P=NaN(K_Gibbs,Nstati,NGibbs);
for i=1:NGibbs
%       i
     %% Estraggo dalle prior gli starting point
     mu1_start=0;
     while mu1_start<=0
         mu1_start=random('Normal',mu01,sigma01);
     end
     mu2_start=0;
     while mu2_start>=0
     mu2_start=random('Normal',mu02,sigma02);
     end
     h1_start=random('Gamma',v01/2,2/delta01);
     sigma1_start=0;
     sigma2_start=0; %%%%%%%%%%%%%%%%%%%%
     while sigma1_start>=sigma2_start
        sigma1_start=1/sqrt(h1_start);
        h2_start=random('Gamma',v02/2,2/delta02);
        sigma2_start=1/sqrt(h2_start);
     end
     p11_start=random('Beta',u011,u012);
     p22_start=random('Beta',u022,u021);

     while kk_Gibbs<=K_Gibbs
           kk_Gibbs
        %% Multimove Gibbs-Sampling --> Simulo S_t con t=1,...,T in blocco
        % Due step: Calcolo l'ultimo UpDate attraverso il filtro di Hamilton e
        % estraggo l'ultimo stato (Teo Trasformata inversa). Poi a catena
        % genero gli altri S_t da T-1....1
        % 1) Filtro di Hamilton
        cd('Y:\Chiara\TESI\MarkovSwitching')
       [UpDate,Prediction]=FiltroHamilton(Y,mu1_start,mu2_start,sigma1_start,...
          sigma2_start,p11_start,p22_start,Nstati);
        % Utilizzo il Teo della trasformata inversa per estrarre l'ultimo
        % stato--> S_T
        u=rand(1);
        if u<=UpDate(end,1)
             State(end,kk_Gibbs,i)=1;
        else
             State(end,kk_Gibbs,i)=2;
        end
        % Stimo da T-1 a 1 tutti gli altri stati
        for t=T-1:-1:1
             if  State(t+1,kk_Gibbs,i)==1
                  probTrans_aux=[p11_start;1-p22_start];
             else
                  probTrans_aux=[1-p11_start; p22_start];
```

```
    end
    num=probTrans_aux(1)*UpDate(t,1);
    den=ones(1,Nstati)*(probTrans_aux.*UpDate(t,:)');
    ProbStato_t=num/den;
   % Utilizzo il Teo della trasformata inversa per estrarre lo
   % stato S_t
    u=rand(1);
    if u<=ProbStato_t
    State(t,kk_Gibbs,i)=1;
    else
    State(t,kk_Gibbs,i)=2;
    end
end
%% Stimo le probabilit  di Transizione
% Calcolo n11 n12 n21 n22
n11=0;
n12=0;
n22=0;
n21=0;
for j=1:T-1
    switch State(j,kk_Gibbs,i)
        case 1
            if State(j+1,kk_Gibbs,i)==1
                n11=n11+1;
            else
                n12=n12+1;
            end
        case 2
            if State(j+1,kk_Gibbs,i)==2
                n22=n22+1;
            else
                n21=n21+1;
            end
    end
 end
% Estraggo una beta
u11=u011+n11;
u12=u012+n12;
u22=u022+n22;
u21=u021+n21;
%estraggo p11
P(kk_Gibbs,1,i)=random('Beta',u11,u12);
%estraggo p22
P(kk_Gibbs,2,i)=random('Beta',u22,u21);
p11_start=P(kk_Gibbs,1,i);
p22_start=P(kk_Gibbs,2,i);

%% Stimo media e Precision dei due stati.
%% Media
% mu_S1
idx1=find(State(:,kk_Gibbs,i)==1);
T1=length(idx1);
mu1=(inv(sigma01^2)+sigma1_start^-2*(X(idx1)'*X(idx1)))^-1...
*(inv(sigma01^2)*mu01+sigma1_start^-2*(X(idx1)'*Y(idx1)));
var1=(inv(sigma01^2)+sigma1_start^-2*(X(idx1)'*X(idx1)))^-1;
sigma1=sqrt(var1);
```

```
%           while mu(kk_Gibbs,1,i)<=0
         mu(kk_Gibbs,1,i)=random('Normal',mu1,sigma1);
%           end
         mu1_start=mu(kk_Gibbs,1,i);
         % mu_S2
         idx2=find(State(:,kk_Gibbs,i)==2);
         T2=length(idx2);
         mu2=(inv(sigma02^2)+sigma2_start^-2*(X(idx2)'*X(idx2)))^-1*...
         (inv(sigma02^2)*mu02+sigma2_start^-2*(X(idx2)'*Y(idx2)));
         var2=(inv(sigma02^2)+sigma2_start^-2*(X(idx2)'*X(idx2)))^-1;
         sigma2=sqrt(var2);
         mu(kk_Gibbs,2,i)=random('Normal',mu2,sigma2);
         mu2_start=mu(kk_Gibbs,2,i);
         %% Sigma
         %sigma_S1  $$$ occhio dividere gli stati
         v1=v01+T1;
         delta1=delta01+(Y(idx1)-X(idx1)*mu(kk_Gibbs,1,i))'*(Y(idx1)-X(idx1)*mu(kk_Gibbs,1,i));
         h=random('Gamma',v1/2,2/delta1);
         sigma1_start=1/sqrt(h);
         sigma(kk_Gibbs,1,i)=sigma1_start;

         % sigma_S2
         v2=v02+T2;
         delta2=delta02+(Y(idx2)-X(idx2)*mu(kk_Gibbs,2,i))'*(Y(idx2)-X(idx2)*mu(kk_Gibbs,2,i));
         h=random('Gamma',v2/2,2/delta2);
         sigma2_start=1/sqrt(h);
         sigma(kk_Gibbs,2,i)=sigma2_start;

         kk_Gibbs=kk_Gibbs+1;


    end
    kk_Gibbs=1;
end
REPORT_UnivariateMS_SPX.mu=mu;
REPORT_UnivariateMS_SPX.sigma=sigma;
REPORT_UnivariateMS_SPX.P=P;
REPORT_UnivariateMS_SPX.State=State;
cd('Y:\Chiara\TESI\MarkovSwitching\UnivariateMS_SCript\REPORT')
save('REPORT_UnivariateMS_SPX','REPORT_UnivariateMS_SPX')
}}

{\small{
function [UpDate,Prediction]=FiltroHamilton(Y,mu1,mu2,sigma1,sigma2,p11,p22,Nstati)
T=length(Y);
UpDate=zeros(T,Nstati);
Prediction=zeros(T,Nstati);
%% Scelgo come Starting point del Filtro di Hamilton le ergodiche
pi1=(1-p22)/(2-p11-p22);
pi2=(1-p11)/(2-p11-p22);
UpDate_start=[pi1; pi2];
P=[p11 1-p22; 1-p11 p22];
pd1 = makedist('Normal',mu1,sigma1);
pd2 = makedist('Normal',mu2,sigma2);
for i=1:T
    Prediction(i,:)=P*UpDate_start;
```

```
    eta=[pdf(pd1,Y(i)); pdf(pd2,Y(i))];
    Update_num=Prediction(i,:)'.*eta;
    Update_den=ones(1,Nstati)*(Prediction(i,:)'.*eta);
    UpDate(i,:)=Update_num./Update_den;
    UpDate_start=UpDate(i,:)';

end
```