Data driven investment strategies using Bayesian inference in regime switching models

Éléonore Blanchard, Pierre-Olivier Goffard

Data driven investment strategies using Bayesian inference in regime switching models

*Eléonore Blanchard[ab] and Pierre-Olivier Goffard[c]

[a] I2M, Université d'Aix Marseille et CNRS, Marseille, France

[b] CPR Asset Management, Paris, France

[c] Institut de Recherche Mathématique Avancée, UMR 7501 Université de Strasbourg et CNRS, Strasbourg, France

*Corresponding author: Eléonore Blanchard; eleonoredblanchard@gmail.com

Abstract

This article presents the benefits of using Bayesian algorithms to fit regime switching models to daily financial returns data in order to design trading strategies. Our study focuses on a Gaussian hidden Markov model. We show how the application of a simple smoothing technique preserves the hidden Markov structure and facilitates regime detection even in instances of highly volatile data. The effectiveness of a trading strategy, based on regime detection, may be hindered by a high rate of false signals, leading to numerous trades and, consequently, an escalation in transaction costs. By reducing variance through data smoothing, we enhance the persistence of regimes over time. We validate our statistical learning procedures using synthetic data prior to their application to real-world financial data.

Keywords: Hidden Markov Models ; Bayesian Inference ; Market Regimes ; Financial Time Series ; Smoothing ; Transaction Costs Mitigation

Contents

1 Introduction **2**

2 Model settings and statistical learning procedure **3**
2.1 Hidden Markov chain and logarithmic returns data 3
2.2 Bayesian inference via Gibbs sampling 4
2.3 How to handle high volatility? 5

3 Investment strategies **8**
3.1 Strategy construction 9
3.2 Binary signal versus continuous signal 9
3.3 Evaluation and metrics 10

4 Simulation experiments on the impact of L and strategies **10**
4.1 Signals study 11
4.2 Strategies study 12

5 Results and investment strategies on real data **15**
5.1 Model selection 15
5.2 Signals study 17
5.3 Strategies study 18

6 Conclusion and perspectives **20**

Appendices **22**

A Gibbs algorithm prior and posterior distributions **22**

B Table of probability distribution **24**

C Algorithmic details **24**

1 Introduction

The price dynamics of stocks and stock indices are known to switch between regimes, depending on the global behaviour of the market. In finance, when dealing with two states switching models, these regimes are usually referred to as bull and bear markets as described, for instance, in the work Gonzalez et al. (2005). A bull market is characterized by rising stock prices, relatively low volatility, and a generally positive outlook among investors. Conversely, a bear market is characterized by falling stock prices, increased market volatility, and a generally negative outlook among investors. Bull and bear markets are part of a normal market cycle and both can present opportunities and challenges for investors. In a bull market, investors may be able to generate significant returns by undertaking higher risk and investing in high-growth stocks. However, in a bear market, investors may need to focus on preserving capital and reducing risk by investing in more defensive assets, such as bonds or gold. Market regimes can change quickly and unexpectedly. The role of a financial advisor is to develop an investment strategy to weather different market conditions over the long term. For a more comprehensive understanding of bull and bear markets, see the analysis of historical turning points after bull and bear markets dating in Gonzalez et al. (2005) and Gonzalez et al. (2006).

Financial assets are usually modeled via stochastic processes. The most commonly used model assumes that stock prices are governed by geometric Brownian motions, we refer the reader to the pioneering works of Merton (1969) and Black and Scholes (1973). This process has two parameters: the expected return μ and volatility σ. This simple model does not account for changes in the regime of the financial markets and is therefore suitable only for short periods. One workaround consists in assuming that the alternation of market regime is governed by a Markov chain. This idea goes back to the work of Hamilton (1989) and has been applied to stock returns by Schaller and Norden (1997). The Markov-modulated geometric Brownian motion has been extensively studied by the financial mathematics community to do option pricing in Jobert and Rogers (2006), optimal trading in Zhang (2001) and portfolio optimization in Yin and Zhou (2004). Our work focuses on statistical calibration. The data at hand consists of daily logarithmic returns which should be normally distributed under the considered model.

One of the main challenges in modeling market regimes arises from their lack of direct observability. To address this issue, statisticians and econometricians commonly use Hidden Markov Models (**HMM**). **HMMs** are probabilistic models that use a sequence of observed data to guess the hidden states which generated the data, based on the assumption that the hidden states follow a Markov process. The parametric inference of such models requires going back and forth between (1) estimating the parameters that characterize each regime and (2) associating each observation to a regime. The learning procedure relies on the optimization of the likelihood function through an expectation maximization algorithm like the famous Baum-Welch algorithm, introduced by Baum et al. (1970). For an overview, we refer the reader to the book of Cappé et al. (2007).

We explore in this paper the benefit of applying a Bayesian approach on smoothed data to detect regime changes and adapt our investment strategies accordingly. The computational burden associated with sampling the posterior distribution is alleviated by the use of Gibbs sampling and conjugate prior distribution, see for instance the works of Richardson and Green (1997) and Rydén (2008). Gibbs sampling is a Markov Chain Monte Carlo technique used for Bayesian inference, where one samples from the posterior distribution by iteratively sampling from the conditional distributions of the variables given the other variables. The incorporation of the uncertainty around the estimated parameters and the Markov chain path provided by the posterior distribution in the investment strategies is one of our contributions. Generating a vast number of potential hidden state sequences contrasts with the frequentist approach that only returns the most probable hidden state sequence.

Our model translates data into a signal used to inform investment decisions. Within a two states model, e.g. bull and bear, we increase exposure when the market is trending upward and progressively shift toward a more defensive position when it is trending downward. This paper explores a trend-following strategy on a benchmark index, adjusting positions according to market regimes, similar to that consid-

ered by Cheng et al. (2018) and Bulla et al. (2011). It aligns with Chen et al. (2020)'s long-short strategy, which involves taking long positions in bull markets and short positions in bear markets (though we exclusively take long positions). Nguyen (2018) implementes a trend-following strategy based on predicted stock prices inferred from an **HMM** rather than relying on the detected signal. Dai et al. (2010) develop a trend-following trading strategy based on market regimes to identify optimal buying and selling times using an optimal stopping problem.

An alternative strategy is factor investing, which assumes that certain risk factors are profitable in bull regimes and protective in bear regimes. Erlwein et al. (2009) and Ammann and Verhofen (2006) recommend using a value factor for protection in bearish phases, and growth or momentum factors for maximizing returns in bullish phases.

Lastly, market regimes can be used for portfolio optimization through stock picking, selecting bullish stocks as discussed in Nguyen and Nguyen (2015) and Kim et al. (2019). Alternatively, Ang and Bekaert (2004) optimize portfolio weights by incorporating signals into the capital asset pricing model (CAPM), while Graflund and Nilsson (2003) does so, based on expectations of future regime probabilities and returns. Finally, the Markov regime-switching framework can be integrated into a factor model for portfolio optimization, as demonstrated by Costa and Kwon (2018) for risk-parity optimization and Costa and Kwon (2019) for mean-variance optimization.

By switching positions, an investor incurs transaction costs. Thus, the level of false alarm, that is to say, non-existent but detected changes of regime, must be controlled. A trade-off must be found between accurately reconstructing the Markov chain path while favoring scenarios with fewer regime changes. Indeed, excessive regime shifts may result in higher transaction costs compared to delaying a regime change.

When the data exhibits high levels of volatility, the signal coming out of the model is noisy. Our second contribution mitigates the impact of data variability by replacing the raw data with a smoothed version of it resulting from a simple moving average procedure. Averaging the data coming from a Gaussian **HMM** outputs data again distributed as a Gaussian **HMM**. The same algorithms can then be used to both fit the raw and the smoothed data. We discuss the pros and cons of replacing the raw data with its moving average when fitting an **HMM**. The novely of this approach is particularly beneficial for practitioners who struggle to obtain a consistent real-time signal on data that is often highly variable, and it has not been utilized in financial literature to our knowledge..

The remainder of the paper is organized as follows. A description of the model, the data, and the estimation procedure is provided in Section 2. Section 3 explains how the outputs of the model are exploited in the implementation of an investment strategy. A simulation study is conducted in Section 4 to assess the soundness of our investment strategies in a controlled environment. Section 5 illustrates our procedure on real-world financial datasets.

2 Model settings and statistical learning procedure

2.1 Hidden Markov chain and logarithmic returns data

The price of a stock or the value of a stock index $(S_t)_{t\geq 0}$ is a stochastic process observed in discrete time

$$s_{0:n} = (s_0, \ldots, s_n).$$

We assume that $(S_t)_{t\geq 0}$ is governed by a Markov-modulated geometric Brownian motion which means that the parameters of the geometric Brownian motion depends on the current state of a Markov chain $(X_t)_{t\geq 0}$ with state space $\mathcal{X}$. Denote by

$$x_{1:n} = (x_1, \ldots, x_n),$$

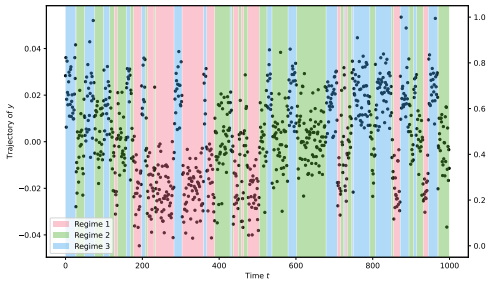
the sequence of market states. Under such model the log-returns defined by

$$y_k = \ln(s_k/s_{k-1}),\ k = 1, 2 \ldots, n,$$

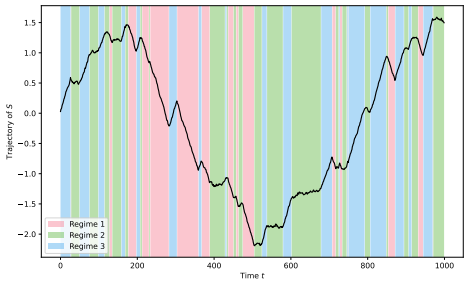
conditionnaly on the state of the Markov chains, are independent and normally distributed. We have

$$y_k | x_k \sim \text{Normal}(\mu_{x_k}, \sigma_{x_k}),\ k = 1, 2 \ldots, n.$$

Denote by $M = (\mu_x)_{x\in\mathcal{X}}$ and $\Sigma = (\sigma_x)_{x\in\mathcal{X}}$ the drift and volatility parameters associated to each state. Further denote by $\nu = (\nu_x)_{x\in\mathcal{X}}$ the initial distribution of the Markov chain and $Q = (q(x,x'))_{(x,x')\in\mathcal{X}^2}$ its transition probability matrix. Figure 1 shows simulated trajectories of $s_{1:n}$, $y_{1:n}$ and $x_{1:n}$ when $\mathcal{X} = \{1,2,3\}$, that is the state space contains three states, with parameters $\nu = \begin{pmatrix}0.025 & 0.025 & 0.95\end{pmatrix}$, $Q = \begin{pmatrix}0.95 & 0.025 & 0.025\\ 0.05 & 0.95 & 0.025\\ 0.025 & 0.025 & 0.95\end{pmatrix}$, $M = \begin{pmatrix}-0.02 & 0 & 0.02\end{pmatrix}$, $\Sigma = \begin{pmatrix}0.01 & 0.01 & 0.01\end{pmatrix}$ and $N = 1000$.



Trajectories of $s_{1:n}$ and $x_{1:n}$ on $\mathcal{X} = \{1,2,3\}$



Trajectories of $y_{1:n}$ and $x_{1:n}$ on $\mathcal{X} = \{1,2,3\}$

Figure 1: Simulated trajectories

On Figure 1, regime 1 corresponds to a downward trend, regime 3 an upward trend and regime 2 a stabilization of the stock value. Our aim is to find the parameters $\theta = (\nu, Q, M, \Sigma)$ that best explains the observed data $y_{1:n}$. The main difficulty is that the state sequence $x_{1:n}$ is unvailable which is why we refer to this model as a Hidden Markov Model (**HMM**). Since the observations are normally distributed, we further refer to the model considered in this work as a Gaussian **HMM** with parameter θ. The next section presents a Gibbs sampler to estimate the model parameters and reconstruct the sequence $x_{1:n}$ of hidden states.

2.2 Bayesian inference via Gibbs sampling

In Bayesian statistics, the model's parameters are random variables having a joint prior distribution $\theta \sim \pi(\theta)$. Inference then relies on the posterior distribution

$$\pi(\theta|y_{1:n}) \propto p(y_{1:n}|\theta)\pi(\theta),$$

where $\propto$ stands for "proportional to", that updates the prior distribution using the likelihood function of the model. In hidden Markov models, the likelihood function depends on the sequence of unknown states $x_{1:n}$. The posterior distribution we aim for becomes

$$\pi(\theta, x_{1:n}|y_{1:n}) \propto p(y_{1:n}|\theta, x_{1:n})p(x_{1:n}|\theta)\pi(\theta), \tag{1}$$

where

$$p(y_{1:n}|\theta, x_{1:n}) = \prod_{n=1}^{N} g_\theta(y_n|x_n), \tag{2}$$

with

$$g_\theta(y_n|x_n) = \frac{1}{\sigma_{x_n}\sqrt{2\pi}} \exp\left[-\frac{(y_n - \mu_{x_n})^2}{2\sigma_{x_n}^2}\right],$$

and

$$p(x_{1:n}|\theta) = \nu(x_1) \prod_{k=2}^{n} q(x_n, x_{n-1}).$$

The posterior distribution in (1) is known up to a constant which prevents us from direct evaluation. A standard workaround in Bayesian statistics is to draw samples from the posterior distribution using Markov Chain Monte Carlo (**MCMC**) algorithms. A Gibbs sampler is a **MCMC** algorithm that samples from a joint distribution $\pi(\theta, x_{1:n}|y_{1:n})$ by sampling sequentially the univariate conditional distributions

$$\pi(\vartheta|\ldots), \text{ for every component } \vartheta \text{ of } \theta, \tag{3}$$

where "..." refers to all the other parameters, the data $y_{1:n}$ and the trajectory of the hidden Markov chain $x_{1:n}$. The trajectory $x_{1:n}$ is not a parameter per se and is recovered by a filtering procedure that we append to the Gibbs iterations. In our model, the prior distribution can be chosen to yield tractable posterior distributions (3). The settings of the prior and posterior distributions are taken from Rydén (2008) originally inspired by Richardson and Green (1997), and are provided in Appendix A.

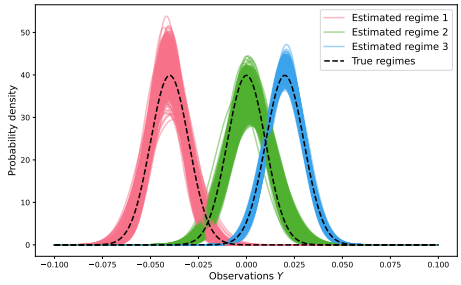
The Gibbs algorithm returns a set of values for the parameters equal to the number of iterations it performs I, minus the burn-in iterations, denoted by $\theta_1, \ldots, \theta_I$, as well as for the sequences of hidden states denoted by $x^1_{1:n}, \ldots, x^I_{1:n}$, each corresponding to a distinct possible scenario. Our objective is to use these scenarios for the investment strategies as described in Section 3.

2.3 How to handle high volatility?

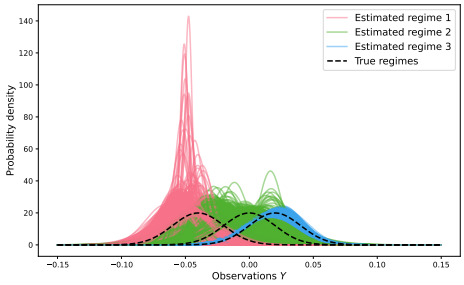
In the context of designing an optimal trading strategy, our focus lies in detecting trend regimes rather than volatility regimes. When σ_x, the standard deviation of a state x, surpasses the gap between the mean parameters associated with this state x and another state x', our algorithm encounters difficulty distinguishing between them. To address this challenge, we introduce a variant of the coefficient of variation commonly used in statistics, that we shall call the state detection score (**SDS**):

$$\mathbf{SDS}_x = \frac{\sigma_x}{\min\limits_{x' \in \mathcal{X} \setminus \{x\}} |\mu_x - \mu_{x'}|},$$

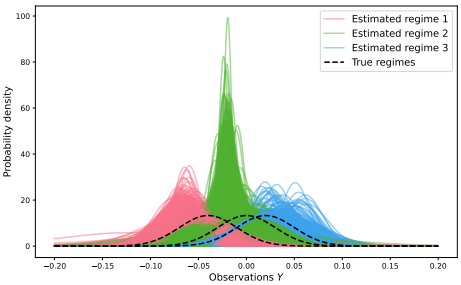
where σ_x represents the standard deviation of the state x and μ_x denote the means associated with state x. This measure allows us to compare the standard deviation to the gap between the mean parameters. Intuitively, a high **SDS** corresponds to a (relatively) small mean difference between the state x and the other states. Conversely, if the mean of a state x differs significally (with regard to its variability σ_x) from those of the other states, we expect a high associated **SDS**.
Figure 2 shows the posterior predictive distribution of the data with different values of **SDS**. The synthetic data $y_{1:N}$ are simulated by an **HMM** on $\mathcal{X} = \{1, 2, 3\}$ with parameters $\nu = \begin{pmatrix} 0.05 & 0.05 & 0.9 \end{pmatrix}$, $Q = \begin{pmatrix} 0.8 & 0.05 & 0.15 \\ 0.05 & 0.8 & 0.15 \\ 0.025 & 0.025 & 0.95 \end{pmatrix}$, $M = \begin{pmatrix} -0.04 & 0 & 0.02 \end{pmatrix}$ and $N = 1000$.



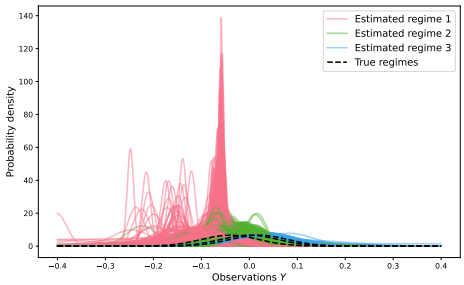
$\Sigma = \begin{pmatrix} 0.01 & 0.01 & 0.01 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 0.25 & 0.5 & 0.5 \end{pmatrix}$



$\Sigma = \begin{pmatrix} 0.02 & 0.02 & 0.02 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 0.5 & 1 & 1 \end{pmatrix}$



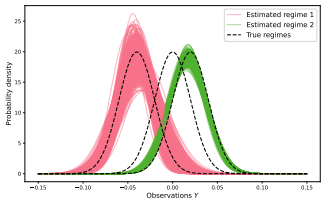
$\Sigma = \begin{pmatrix} 0.03 & 0.03 & 0.03 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 0.75 & 1.5 & 1.5 \end{pmatrix}$



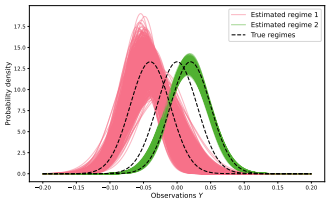
$\Sigma = \begin{pmatrix} 0.06 & 0.06 & 0.06 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 1.5 & 3 & 3 \end{pmatrix}$

Figure 2: Posterior predictive distributions of the three-state model

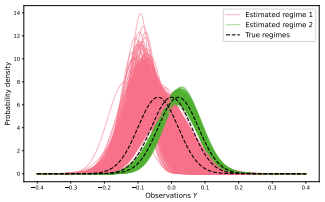
When $\mathbf{SDS} \geq 1$ for any regime, it overlaps with another one, hindering the estimation of the parameters associated with these states. The first workaround involves reducing the number of states. For $\Sigma = \begin{pmatrix} 0.02 & 0.02 & 0.02 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 0.03 & 0.03 & 0.03 \end{pmatrix}$, two **SDS** are greater than or equal to 1, thus transitioning from 3 states to 2 states will naturally merge the two indistinguishable states together (i.e., states 2 and 3). Then $M = \begin{pmatrix} \mu_1 & \frac{\mu_2+\mu_3}{2} \end{pmatrix} = \begin{pmatrix} -0.04 & 0.01 \end{pmatrix}$. We can thus recalculate the values of **SDS** in a two-state model, as depicted in Figure 3.



$\Sigma = \begin{pmatrix} 0.02 & 0.02 & 0.02 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 0.4 & 0.4 \end{pmatrix}$



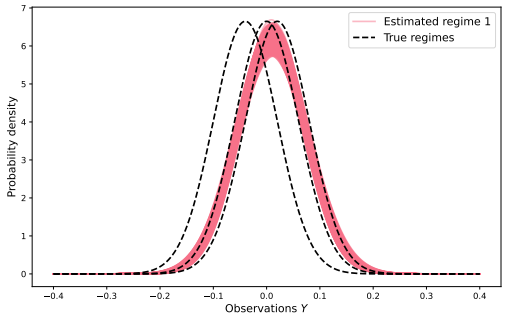
$\Sigma = \begin{pmatrix} 0.03 & 0.03 & 0.03 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 0.6 & 0.6 \end{pmatrix}$



$\Sigma = \begin{pmatrix} 0.06 & 0.06 & 0.06 \end{pmatrix}$
$\mathbf{SDS} = \begin{pmatrix} 1.2 & 1.2 \end{pmatrix}$

Figure 3: Posterior predictive distributions of the two-state model

The **SDS** for $\Sigma = \begin{pmatrix} 0.06 & 0.06 & 0.06 \end{pmatrix}$ are still greater than 1 making the state indistinguishable. We must then settle for a single state, as depicted in Figure 4.



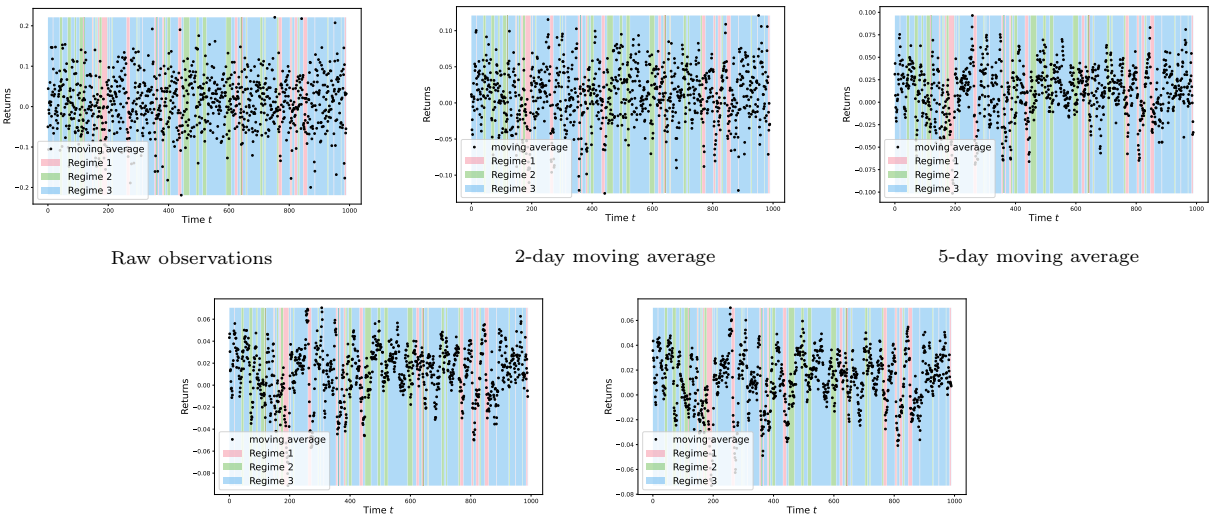
$$\Sigma = \begin{pmatrix} 0.06 & 0.06 & 0.06 \end{pmatrix}$$

Figure 4: Posterior predictive distributions of the one-state model

When we cannot or do not wish to reduce the size of the state space $\mathcal{X}$, we can directly act on the volatility of the data through a simple smoothing procedure to enhance the distinguishability of states. Consider taking a moving average given by

$$z_n^{(L)} = \frac{1}{L} \sum_{l=1}^{L} y_{n-l+1},$$

where $y_{1:n}$ is our initial data. Figure 5 illustrates the influence of this smoothing process on the log returns for various moving average depth $L \in \{2, 5, 8, 10\}$.



Raw observations | 2-day moving average | 5-day moving average

8-day moving average | 10-day moving average

Figure 5: Smoothing with moving averages on some synthetic data $y_{1:n}$ simulated by an **HMM** on $\mathcal{X} = \{1, 2, 3\}$ with parameters $\nu = \begin{pmatrix} 0.05 & 0.05 & 0.9 \end{pmatrix}$, $Q = \begin{pmatrix} 0.8 & 0.05 & 0.15 \\ 0.05 & 0.8 & 0.15 \\ 0.025 & 0.025 & 0.95 \end{pmatrix}$, $M = \begin{pmatrix} -0.04 & 0 & 0.02 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 0.06 & 0.06 & 0.06 \end{pmatrix}$ and $N = 1000$

Figure 5 shows that adequate smoothing make the states nearly observable to the naked eye. This suggests that the decoding algorithm may benefit from smoother returns, increasing states distinguishability.

The interesting feature of the Gaussian **HMM** is that the smoothed observations $z_{1:n}^{(L)}$ yields a Gaussian **HMM** referred to as **MA(L)-HMM** which we describe hereafter. The state space of the Markov chain is $\mathcal{X}^L$, with $x_n^L = x_{(n-L+1):n} \in \mathcal{X}^L$ for $n \geq L$. We denote by $\nu^{(L)}$ and $Q^{(L)}$ the initial distribution and transition matrix. The set up is similar to a higher-order **HMM** discussed in Zhang et al. (2019). The

mean and variance vector of the smoothed observations are given by

$$M^{(L)} = \left\{ \mu_{x^L} = \frac{1}{L}\sum_{i=1}^{L} \mu_{x_i} \right\}_{x^L \in \mathcal{X}^L} \text{ and } \Sigma^{(L)} = \left\{ \sigma_{x^L} = \sqrt{\frac{1}{L}\sum_{i=1}^{L} \sigma_{x_i}^2} \right\}_{x^L \in \mathcal{X}^L}.$$

Despite the increase in the dimension of states, the number of parameters involved in the **MA(L)-HMM** is the same as that of the **HMM**. We illustrate this fact in Example 1 where we consider a two-state space and $L = 2$.

Example 1. Take a Gaussian **HMM** with two state $\mathcal{X} = \{0, 1\}$, the **HMM** associated to the two-day moving average ($L = 2$) has the following state space $\mathcal{X}^{(2)} = \{(0,0), (0,1), (1,0), (1,1)\}$. The transition matrix is given by

$$Q^{(L)} = \begin{pmatrix} q_{(0,0)\to(0,0)} & q_{(0,0)\to(0,1)} & q_{(0,0)\to(1,0)} & q_{(0,0)\to(1,1)} \\ q_{(0,1)\to(0,0)} & q_{(0,1)\to(0,1)} & q_{(0,1)\to(1,0)} & q_{(0,1)\to(1,1)} \\ q_{(1,0)\to(0,0)} & q_{(1,0)\to(0,1)} & q_{(1,0)\to(1,0)} & q_{(1,0)\to(1,1)} \\ q_{(1,1)\to(0,0)} & q_{(1,1)\to(0,1)} & q_{(1,1)\to(1,0)} & q_{(1,1)\to(1,1)} \end{pmatrix},$$

where

$$q_{(i,j)\to(k,l)} = \mathbb{P}(X_n = l, X_{n-1} = k | X_{n-1} = j, X_{n-2} = i).$$

We note that many transitions are impossible, indeed $q_{(i,j)\to(k,l)} = 0, \text{for } j \neq k$. We further note that

$$q_{(i,k)\to(k,l)} = \mathbb{P}(X_n = l, X_{n-1} = k | X_{n-1} = k, X_{n-2} = i) = \mathbb{P}(X_n = l | X_{n-1} = k) = q_{k\to l},$$

where $q_{k\to l}$ is the element in the k^th row and the l^{th} of the transition matrix Q. The matrix $Q^{(L)}$ then becomes

$$Q^{(L)} = \begin{pmatrix} q_{0\to0} & q_{0\to1} & 0 & 0 \\ 0 & 0 & q_{1\to0} & q_{1\to1} \\ q_{0\to0} & q_{0\to1} & 0 & 0 \\ 0 & 0 & q_{1\to0} & q_{1\to1} \end{pmatrix}.$$

The initial probabilities are given by

$$\nu^{(L)} = \begin{pmatrix} \nu_0 q_{0\to0} & \nu_0 q_{0\to1} & \nu_1 q_{1\to0} & \nu_1 q_{1\to1} \end{pmatrix}.$$

Lastly, the vectors of means and standard deviations are given by

$$M^{(L)} = \begin{pmatrix} \mu_0 & \frac{\mu_0+\mu_1}{2} & \frac{\mu_1+\mu_0}{2} & \mu_1 \end{pmatrix} \text{ and } \Sigma^{(L)} = \begin{pmatrix} \frac{\sqrt{2}\sigma_0}{2} & \frac{\sqrt{\sigma_0^2+\sigma_1^2}}{2} & \frac{\sqrt{\sigma_1^2+\sigma_0^2}}{2} & \frac{\sqrt{2}\sigma_1}{2} \end{pmatrix}.$$

Our learning algorithm alternates between (1) estimating parameters based on their posterior conditional distribution using the moving average and (2) decoding hidden states from the smoothed data using the estimated parameters.

In the simulation study discussed in Section 4, we examine the influence of the moving average depth L on the hidden state decoding task, depending on the value of the theoretical **SDS**.

3 Investment strategies

We treat the sequence of hidden states as a signal that we leverage to develop investment strategies. Based on our training sample $y_{1:n}$, we obtain probable sequences of hidden states $x_{1:n}$. Each new data point y_{n+1} must be labeled as x_{n+1}, potentially resulting in a change of position. For simplicity, we consider only two choices: buying shares of a risky asset or holding cash. The next section describes two ways of processing our signal and make an investment call.

3.1 Strategy construction

We consider a portfolio with value V_n and return $R_{p,n}$. A fraction w_n of the portfolio is invested in the risky asset R, and the remainder in a riskfree asset RF. At time n, the portfolio return is expressed as

$$R_{p,n} = w_n \mathrm{R}_n + (1 - w_n)\mathrm{RF}_n,$$

and the portfolio value is

$$V_n = V_{n-1}(1 + R_{p,n}), \text{ for } n > 0,$$

with V_0 the initial investment in the portfolio.

The value w_n corresponds to the portfolio's position in the risky asset and can vary depending on the asset manager's choices. If w_n varies, then we introduce transaction costs f which are proportional to the variation of the value invested in the risky asset $(w_n - w_{n-1})$. The portfolio return becomes:

$$R_{p,n} = w_n \mathrm{R}_n + (1 - w_n)\mathrm{RF}_n - |w_n - w_{n-1}|f, \text{ for } n > 0,$$

and

$$R_{p,0} = w_0 \mathrm{R}_0 + (1 - w_0)\mathrm{RF}_0 - w_0 f.$$

We will vary w_n based on a signal constructed from the sequences of hidden states $x^k_{1:n}$ for $k = 1, \ldots, K$ generated by the Gibbs algorithm. When this signal is favorable w_n will increase, and vice versa.

3.2 Binary signal versus continuous signal

Two ways of processing the signals for the strategies are considered in this work. The first is referred to as the binary signal. The states $x \in \mathcal{X}$ in the scenarios $k = 1, \ldots, K$ are deemed favorable if $\mu^k_x > 0$ and unfavorable otherwise. Scenarios here are understood as values of the parameters and hidden state sequence output by the Gibbs sampler. We denote by K the number of iteration in the Gibbs sampler minus the burn-in period. The binary signals is defined as

$$\mathbf{bs}_n = H\left(\sum_{k=1}^{K} \mathbb{1}_{\mu^k_{x^k_n} > 0} - K/2\right),$$

where

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

denotes the heaviside function. The effective investment strategy involves holding the risky asset ($w_n = 1$) when the binary signal is 1 that is favorable or bullish, and selling the position ($w_n = 0$) when it is 0, that is unfavorable or bearish.

Our second approach consists in defining a continuous signal as

$$\mathbf{cs}_n = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{\mu^k_{x^k_n} > 0}.$$

The continuous signal ranges between 0 and 1, where 0 corresponds to the bearish binary signal (no exposure to the risky asset), 1 represents the bullish binary signal (total exposure to the risky asset), and values in between indicate a gradual exposure.

Our position then follows from

$$w_n = \begin{cases} \mathbf{bs}_n, & \text{if the binary signal is used,} \\ \mathbf{cs}_n, & \text{if the continuous signal is used.} \end{cases}$$

To ensure that the model is as accurate as possible in reflecting real-world conditions, we introduce a one-day delay to the signals. This delay accounts for the fact that the hidden states are estimated based on returns computed on the closing price S_t, but since the market is closed at that time, we have to wait until the next day to process the information using the price S_{t+1}. This introduces a one-day lag between

the detection of the state and the processing date, which is necessary for a more realistic and accurate representation of market conditions. To assess the sensitivity of our proposed strategies to transaction costs, we incorporate various fee levels ranging from 0.5 basis points (bp) to 2 basis points. We then compare the outcomes against those of a strategy with no fees. When using the binary signal, the full charge fees is applied as soon as there is a position change, because $|w_n - w_{n-1}|f$ is either 0 or f. With the continuous signal, on the other hand, only a portion of the fee corresponding to the position change is applied ($0 \leq |w_n - w_{n-1}|f \leq f$).

3.3 Evaluation and metrics

The evaluation of a strategy is conducted through backtesting, involving the retrospective calculation of returns using the signal as if it had been determined at each historical date. Assessing the performance of strategies with historical data provides insights into their effectiveness and potential profitability. The backtest offers valuable information on the strategy's performance across various market regimes, enabling us to evaluate its robustness.

Strategies' performance is assessed by comparing their backtests to corresponding benchmarks', considering relative performance adjusted for volatility. A benchmark is a predefined index or portfolio used to measure the performance of a trading strategy. It serves as a reference point for evaluating the relative success of the strategy against the broader market. In addition, the strategies' performance is compared to that of the riskfree rate to assess their absolute performance, given the level of risk involved. Ultimately, our objective is to determine if we have correctly timed the growth and crash periods of the market and if we have employed suitable strategies to maximize returns while minimizing risk.

To assess the effectiveness of our strategies, we take two key metrics. The first is the widely used Sharpe ratio (Sharpe (1966), Sharpe (1994)), which compares the strategy's performance and volatility to the risk-free rate, allowing us to evaluate the risk-adjusted returns:

$$\mathrm{SR}_n = \frac{R_{p,n} - RF_n}{\sigma_{p,n}}.$$

Here, $R_{p,n}$ represents the expected return of the portfolio, RF_n is the risk-free rate of return, and $\sigma_{p,n}$ is the standard deviation of the portfolio's expected return. Roughly speaking, a Sharpe ratio greater than 1 indicates a good strategy, 2 indicates a very good strategy and 3 indicates an excellent strategy.

The second metric is the information ratio, derived from the Sharpe ratio by substituting the risk-free asset with a benchmark (Sharpe (1966)), assesses strategy effectiveness relative to direct benchmark investment (Sharpe (1966)):

$$\mathrm{IR}_n = \frac{R_{p,n} - R_{b,n}}{\sigma(R_{p,n} - R_{b,n})}.$$

Here, $R_{b,n}$ represents the expected return of the benchmark, and $\sigma(R_{p,n} - R_{b,n})$ represents the standard deviation of the difference between the portfolio's and benchmark's expected return. A positive information ratio indicates a better strategy than the benchmark, while a negative information ratio suggests a less effective strategy than direct investment in the benchmark.

4 Simulation experiments on the impact of L and strategies

We generate synthetic data $y_{1:n}$, where $n = 1,000$, from an **HMM** with parameters given by

$$\nu = \begin{pmatrix} 0.05 & 0.9 & 0.05 \end{pmatrix}, \; Q = \begin{pmatrix} 0.8 & 0.05 & 0.15 \\ 0.05 & 0.8 & 0.15 \\ 0.025 & 0.025 & 0.95 \end{pmatrix}, \text{ and } M = \begin{pmatrix} -0.04 & 0 & 0.02 \end{pmatrix}.$$

Our goal is to evaluate our capability to retrieve the sequence of hidden states across various levels of volatility $\sigma \in \{0.01, 0.02, 0.04, 0.06, 0.08, 0.1\}$. We focus on the trend regimes and thus assign the same volatility σ to all states. We can define an average **SDS** accross the regimes as

$$\mathbf{ASDS} = \frac{\sum_{x \in \mathcal{X}} \mathbf{SDS}_x}{|\mathcal{X}|}.$$

The average values for the different levels of volatilities are presented in Table 1.

σ	0.01	0.02	0.04	0.06	0.08	0.1
ASDS	0.42	0.83	1.67	2.5	3.33	4.17

Table 1: ASDS for the different volatility levels.

The experiment is repeated $1,000$ times for each value of σ. The Gibbs sampler is configured to run $1,500$ iterations, with 500 iterations dedicated to a burn-in phase.

A preliminary study on the accuracy and persistence of the signal generated by the **MA(L)-HMM** is conducted in Section 4.1. Then, in Section 4.2, we implement the strategies outlined in Section 3.1 to examine the performance of the **MA(L)-HMM** and compare strategies on binary and continuous signals.

We compare the **MA(L)-HMM** to the traditional Gaussian **HMM** by considering several lag values $L \in \{2, 5, 8, 10\}$. It is worth noting that the **HMM** corresponds to an **MA(0)-HMM**, and consequently we will use this terminology from now on.

4.1 Signals study

We start by looking into the accuracy, defined as

$$\text{accuracy} = \frac{\sum_{n=1}^{N} \mathbb{1}_{\{x_n = \hat{x}_n\}}}{N},$$

of the estimated hidden state sequence $\hat{x}_{1:N}$ generated by our Gibbs sampler. The accuracy, for each run, is shown as a boxplot on Figure 6 depending on the **ASDS** level.

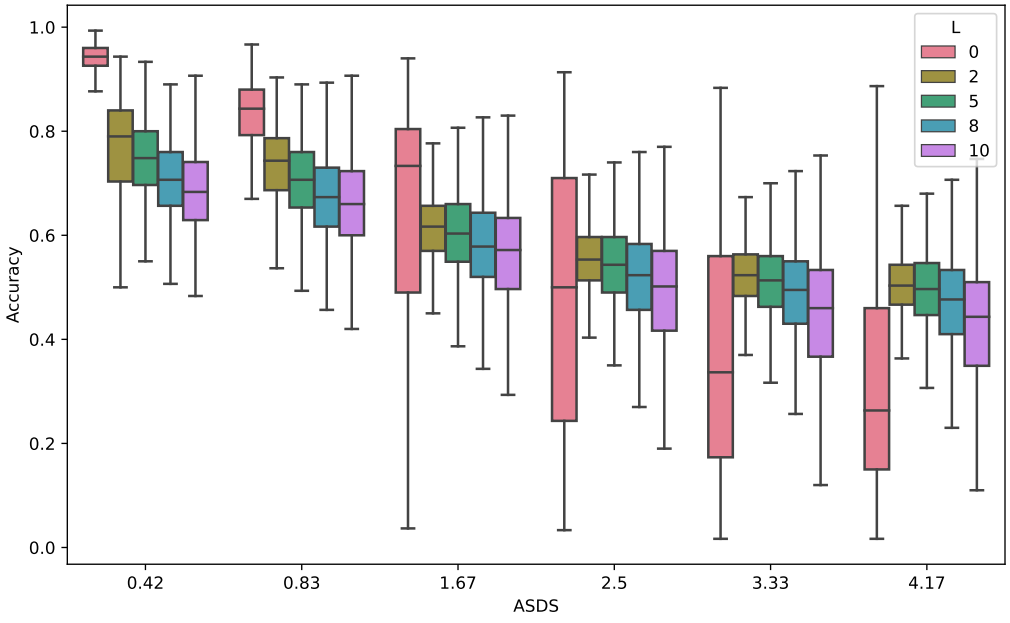


Figure 6: Out-of-sample accuracies

Nearly perfect decoding is achieved when the data volatility is low relative to the differences between the means of the states (**ASDS** ≤ 1). The accuracy sharply declines as volatility increases when using the raw data. Conversely, the use of smoothed data counteract the volatility making accuracy more robust.

As the volatility increases, the gap between the **HMM** and **MA(L)-HMM** results widens.

We further turn our attention to signal stability and persistence through the average state duration given by the inverse of the state switching probability p,

$$\mathrm{p} = \frac{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n \neq x_{n+1}\}}}{N-1},$$

and through the back-and-forth switching rate over one day,

$$\mathrm{baf} = \frac{\sum_{n=1}^{N-2} \mathbb{1}_{\hat{x}_n = \hat{x}_{n+2} \neq \hat{x}_{n+1}}}{N-2}.$$

The latter corresponds to the number of times in the period where a regime lasted only one day, resembling a detection error. This metric is particularly important because these shifts can lead to substantial financial losses, with transaction costs doubling due to entry and exit expenses, potentially outweighing any performance gains. Duration and back-and-forth switching rate are shown on Figure 7 and Figure 8 respectively, depending on the **ASDS** level.

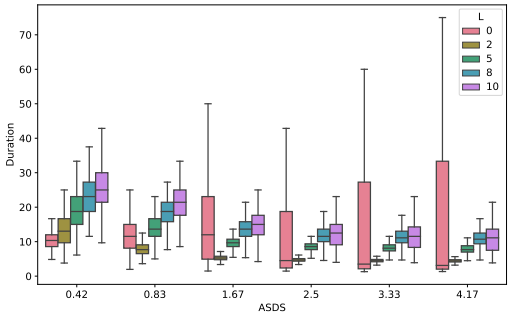


Figure 7: Average state durations

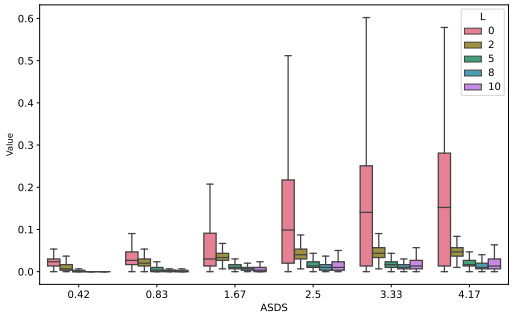


Figure 8: Back-and-forth shifting rates

The average duration of the true hidden state sequence is approximately 11 days. As illustrated in Figure 7, as the volatility of the data increases, the median duration of the sequences of states estimated by the **HMM** decreases, reaching 3 days for **ASDS** = 4.17. The **MA(L)-HMM** tends to produce higher and stable level of duration even for high level volatility. Note that in the context of investment strategies high durations are desirable to reduce trading costs.

This pattern is also discernible in the back-and-forth switching rate in Figure 8. The limited persistence in the regimes of the **HMM** presents a notable challenge for effective strategy implementation, as transaction costs can become substantial. The **MA(L)-HMM** successfully addresses this issue with back-and-forth shifting rates consistently lower than 5% for all values of L and volatilities, and even 2% for $L = \{5, 8, 10\}$.

This simulation study demonstrates the overall effectiveness of smoothing the data to achieve better regime-switching detection when the variance is high. We anticipate that the impact on strategies will follow a similar pattern, and this is precisely what we will now investigate.

4.2 Strategies study

In this section, we aim to examine two aspects: the comparison of the binary signal **bs** and the continuous signal **cs** in strategies, as defined in Section 3.2, and the impact of L in the **MA(L)-HMM** on the strategy outcomes.

First, we explore the influence of transaction fees on strategies employing the binary and continuous signals, by varying the fees from 0, 0.5bp, 1bp and 2bp. The total cost of the strategies in terms of fees are depicted in Figure 9.

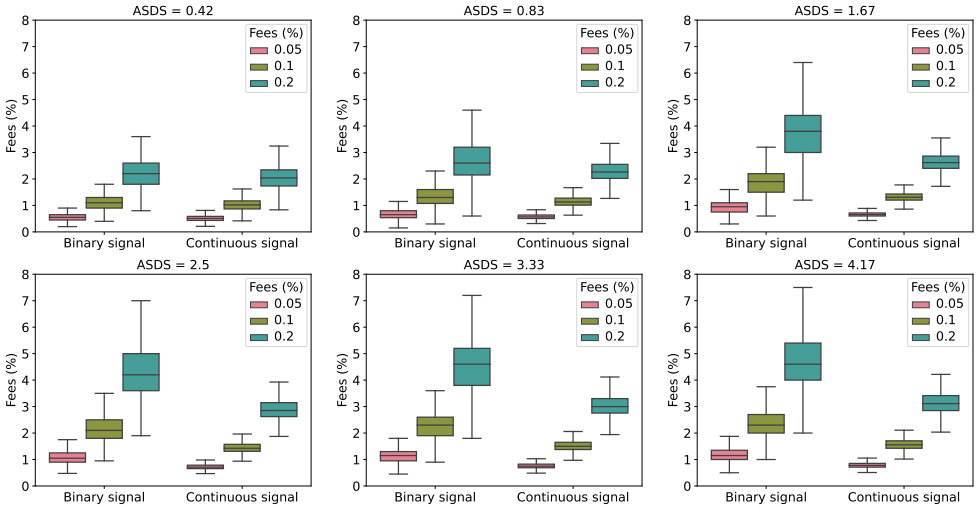
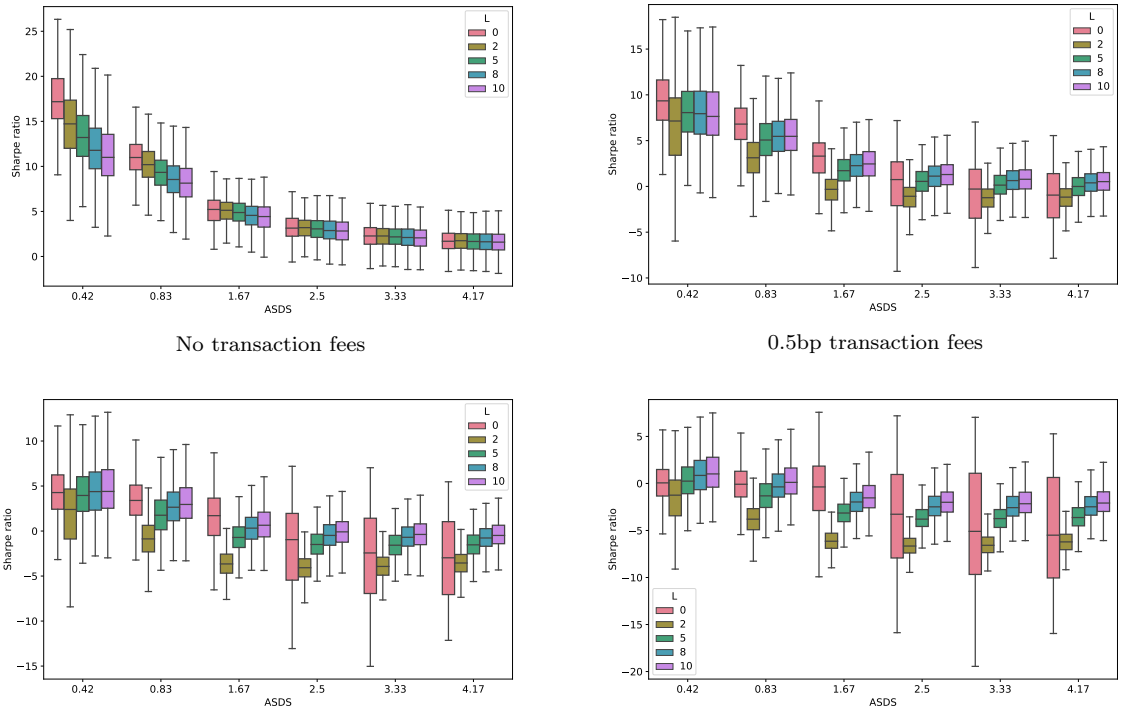


Figure 9: Cumulated transaction fees

In Figure 9, we observe that the continuous signal lower the costs of the strategy compared to the binary signal, particularly as the volatility increases. However, for all strategies, as volatility increases, so do the costs, which is closely related to the distinguishability of states: as the **SDS** increases, states become less distinguishable, leading to more state changes and consequently higher transaction costs.

We will now compare the Sharpe ratios for strategies based on the **MA(L)-HMM** for different values of L, while varying volatility levels and transaction costs. Figure 10 displays the results for the binary signal, and Figure 11 shows the results for the continuous signal.



No transaction fees

0.5bp transaction fees

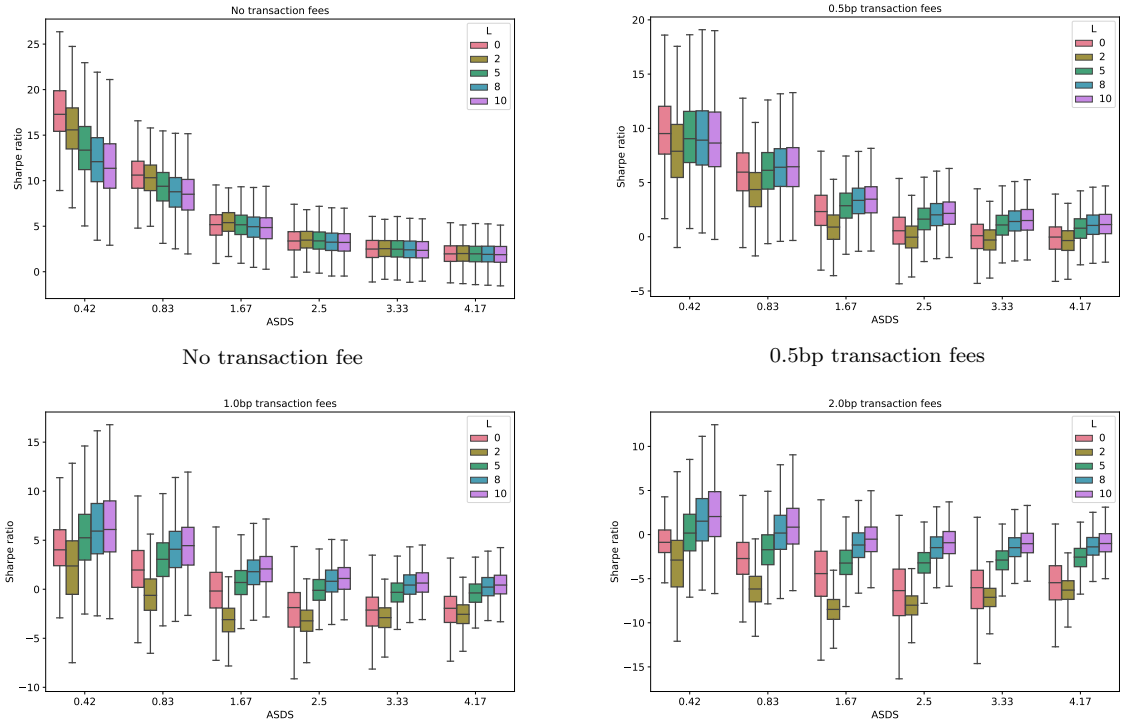
1bp transaction fees

2bp transaction fees

Figure 10: Out-of-sample Sharpe ratio

The Sharpe ratios depicted in Figure 10 show that the **MA(L)-HMM** enables an improvement in

strategy performances as soon as volatility or fees increase. At lower fees, higher volatility is required for the Sharpe ratio of the **MA(L)-HMM** to surpass that of the **HMM**, particularly with lower values of L. However, at higher fees (2bp), the **MA(L)-HMM** consistently maintains a higher Sharpe ratio compared to the **HMM**. In the absence of transaction fees, all models are equivalent in terms of Sharpe ratios when volatility is high. However, the **HMM** outperforms other models when volatility is lower.



No transaction fee | 0.5bp transaction fees

1bp transaction fees | 2bp transaction fees

Figure 11: Out-of-sample Sharpe ratio

The exact same trend appears in Figure 11 for both binary and continuous signals. The only difference observed is in terms of magnitude. When there are no transaction fees, the levels are similar, but as we increase the fees, strategies on the binary signal are more affected compared to those on the continuous signal, resulting in diminished Sharpe ratios.

We also want to examine the performance of our strategies relative to the benchmark. In our simulations, we consider the benchmark to be the sequence of simulated observations upon which we construct the signals and strategies. We look at the information ratio, which is positive when the strategy outperforms the benchmark. The median values for each strategy are reported in Table 2, with bold values corresponding to positive values for better visibility.

	Binary signal					Continuous signal				
	$L=0$	$L=2$	$L=5$	$L=8$	$L=10$	$L=0$	$L=2$	$L=5$	$L=8$	$L=10$
No transaction fee										
ASDS $\simeq 0.42$	**17.18**	**14.72**	**13.21**	**11.78**	**11.00**	**17.29**	**15.58**	**13.36**	**12.09**	**11.36**
ASDS $\simeq 0.83$	**10.99**	**10.19**	**9.33**	**8.55**	**8.14**	**10.61**	**10.33**	**9.39**	**8.78**	**8.51**
ASDS $\simeq 1.67$	**5.22**	**5.13**	**4.88**	**4.56**	**4.43**	**5.17**	**5.39**	**5.16**	**4.94**	**4.85**
ASDS $\simeq 2.50$	**3.15**	**3.20**	**3.06**	**2.88**	**2.83**	**3.38**	**3.47**	**3.39**	**3.25**	**3.22**
ASDS $\simeq 3.33$	**2.28**	**2.27**	**2.17**	**2.11**	**2.07**	**2.49**	**2.53**	**2.48**	**2.41**	**2.35**
ASDS $\simeq 4.17$	**1.69**	**1.75**	**1.66**	**1.63**	**1.59**	**1.96**	**2.00**	**1.97**	**1.91**	**1.88**
0.5bp transaction fees										
ASDS $\simeq 0.42$	**9.34**	**7.14**	**8.06**	**7.94**	**7.64**	**9.52**	**7.88**	**9.04**	**8.91**	**8.64**
ASDS $\simeq 0.83$	**6.81**	**3.13**	**5.07**	**5.46**	**5.47**	**5.96**	**4.35**	**6.13**	**6.41**	**6.46**
ASDS $\simeq 1.67$	**3.31**	-0.32	**1.72**	**2.26**	**2.45**	**2.32**	0.89	**2.86**	**3.35**	**3.46**
ASDS $\simeq 2.50$	**0.75**	-1.08	**0.55**	**1.12**	**1.30**	**0.55**	-0.04	**1.63**	**2.02**	**2.15**
ASDS $\simeq 3.33$	-0.29	-1.23	**0.14**	**0.63**	**0.78**	**0.10**	-0.30	**1.09**	**1.41**	**1.50**
ASDS $\simeq 4.17$	-0.95	-1.17	-0.01	**0.39**	**0.52**	-0.03	-0.35	**0.79**	**1.05**	**1.13**
1bp transaction fees										
ASDS $\simeq 0.42$	**4.27**	**2.40**	**3.95**	**4.38**	**4.40**	**4.02**	**2.38**	**5.25**	**5.93**	**6.09**
ASDS $\simeq 0.83$	**3.40**	-0.87	**1.76**	**2.64**	**2.95**	**1.96**	-0.62	**3.05**	**4.08**	**4.45**
ASDS $\simeq 1.67$	**1.71**	-3.65	-0.71	**0.32**	**0.64**	-0.18	-3.10	**0.69**	**1.79**	**2.07**
ASDS $\simeq 2.50$	-0.95	-4.07	-1.49	-0.49	-0.09	-1.87	-3.22	-0.11	**0.81**	**1.10**
ASDS $\simeq 3.33$	-2.43	-3.93	-1.57	-0.69	-0.38	-2.12	-2.90	-0.31	**0.43**	**0.63**
ASDS $\simeq 4.17$	-2.97	-3.55	-1.53	-0.77	-0.49	-1.94	-2.56	-0.38	**0.22**	**0.43**
2bp transaction fees										
ASDS $\simeq 0.42$	**0.06**	-1.25	**0.25**	**0.85**	**1.01**	-0.88	-2.89	**0.17**	**1.52**	**2.05**
ASDS $\simeq 0.83$	-0.08	-3.80	-1.32	-0.37	**0.11**	-2.71	-6.17	-1.72	**0.17**	**0.84**
ASDS $\simeq 1.67$	-0.37	-6.14	-3.14	-1.97	-1.53	-4.42	-8.48	-3.23	-1.18	-0.52
ASDS $\simeq 2.50$	-3.27	-6.65	-3.79	-2.49	-2.00	-6.35	-8.01	-3.21	-1.48	-0.92
ASDS $\simeq 3.33$	-5.10	-6.58	-3.73	-2.57	-2.16	-6.02	-7.11	-2.89	-1.48	-1.03
ASDS $\simeq 4.17$	-5.50	-6.21	-3.62	-2.48	-2.08	-5.45	-6.29	-2.55	-1.39	-1.01

Table 2: Information ratio

In Table 2, we notice a consistent trend: with no transaction fee, information ratios remain positive and increase further as volatility decreases. However, as fees rise, it becomes progressively challenging to maintain positive information ratios. Increasing the value of L becomes necessary, particularly as volatility levels elevate. Ultimately, at a certain threshold, all ratios turn negative. Combining the continuous signal with an increased value of L in the **MA(L)-HMM** emerges as a promising strategy when dealing with volatile data and transaction fees.

5 Results and investment strategies on real data

Our dataset is the MSCI World Index, denominated in USD and sourced from the Bloomberg platform from January 2, 1990, to Novembre 11, 2023. These prices are transformed into daily logarithmic returns, employing the calculation method outlined in Section 2.1. The model's calibration period is set to one year to capture recent trends and patterns while counteracting the inherent non-stationarity. The learning period is rolled over monthly.

5.1 Model selection

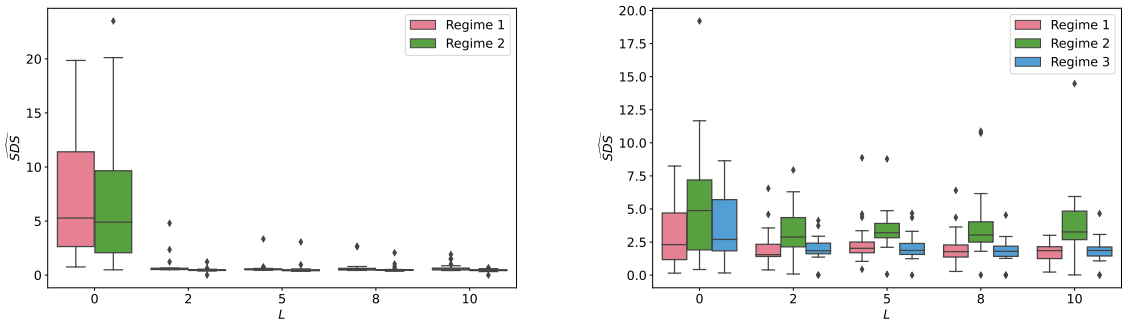
In Section 2.3, we discussed how data volatility exceeding the difference between state means can cause the model to fail in distinguishing between states. In the simulation study of Section 4, we knew the true mean parameters, so we could draw accurate conclusions from the **SDS**. However, this is not the case with real data.

The **SDS** associated to the real data is unkown so we estimate it by

$$\widehat{\mathbf{SDS}}_x = \text{mode}\left\{ \frac{\widehat{\sigma}_x^k}{\min\limits_{x' \in \mathcal{X}\setminus\{x\}} |\widehat{\mu}_x^k - \widehat{\mu}_{x'}^k|},\ k = 1, \ldots, K \right\},$$

with K the number of iteration in the Gibbs sampler, minus the burn-in period. The values of $\widehat{\mu}_x^k$ and $\widehat{\sigma}_x^k$ are the parameters of the log-returns distribution within in each state $x \in \mathcal{X}$ and sweep of the Gibbs sampler $k = 1, \ldots, K$. We consider **MA(L)-HMM** with two- or three-state and for different values of $L = \{0, 2, 5, 8, 10\}$. We assume that this approach provides a reasonable approximation of the **SDS**, and we base our analysis on the findings in Section 2.3.

Figure 12 illustrates the approximated ratio for each year of data obtained with the different models.



$\mathcal{X} = \{1, 2\}$ $\mathcal{X} = \{1, 2, 3\}$

Figure 12: Average yearly $\widehat{SDS}$ for different models

Figure 12 tells us that there are only two distinguishable states in these data, as the $\widehat{\mathbf{SDS}}$ is only lower than 1 for two-state space models. However, even with two states, the **HMM** produces an estimated ratio no lower than 1, so we need to choose a lag $L > 0$. Figure 13 and Figure 14 respectively show the average duration and the back-and-forth switching rate obtained through out-of-sample decoding of the different models.

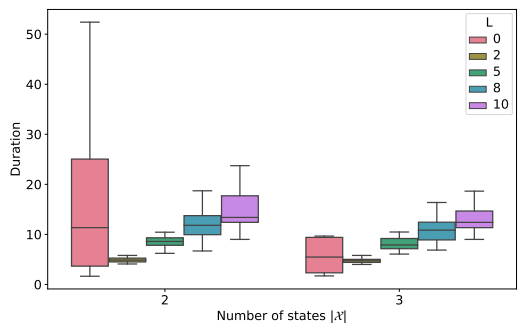


Figure 13: Average state durations

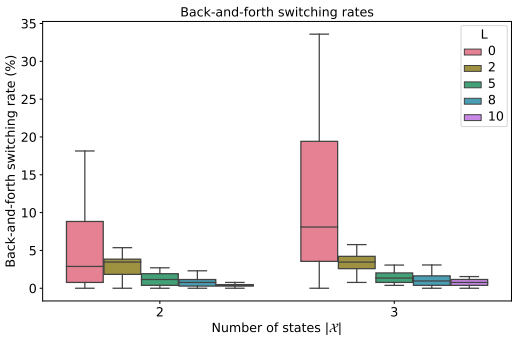


Figure 14: Back-and-forth shifting rates

Figure 13 illustrates that duration gradually increases with L, a trend observed in Section 4, as well as their dispersion across the data series. Figure 14 indicates that lags $L = \{5, 8, 10\}$ for the two-state model significantly reduce the back-and-forth switching rate, with $L = \{8, 10\}$ even approaching zero.

The choice of L is not as straightforward as the choice of number of states. While some series have estimated ratios below 1 with $L = \{2, 5, 8\}$, the dispersion of the ratio for $L = 10$ is much higher. In terms of duration, $L = 10$ generally prevails, but not on all series. Notably, $L = 10$ consistently reduced the back-and-forth switching rate.

In the following sections, we shall compare the obtained signal and its performance in terms of strategy for different models. We compare the **HMM** with the **MA(L)-HMM** for fixed values of L, where $L = \{2, 5, 8, 10\}$, as in the simulation section, along with an **MA(L)-HMM** with a variable L based on the data profile of the learning set.
In the latter case, the selection of L is conducted as follows. We begin by estimating the **ASDS** for **MA(L)-HMM** models with $L \in [5, 10]$, as we believe that from $L \geq 5$, the back and forth switching rate is sufficiently reduced. We choose the value of L that maximizes the **ASDS** and is greater than 1. If there is no such ratio, we reproduce the same procedure for $L \in [0, 4]$. If again no value of L is found, we conclude that the model has only one regime, determining it as bullish if the mean parameter is positive and bearish if negative.

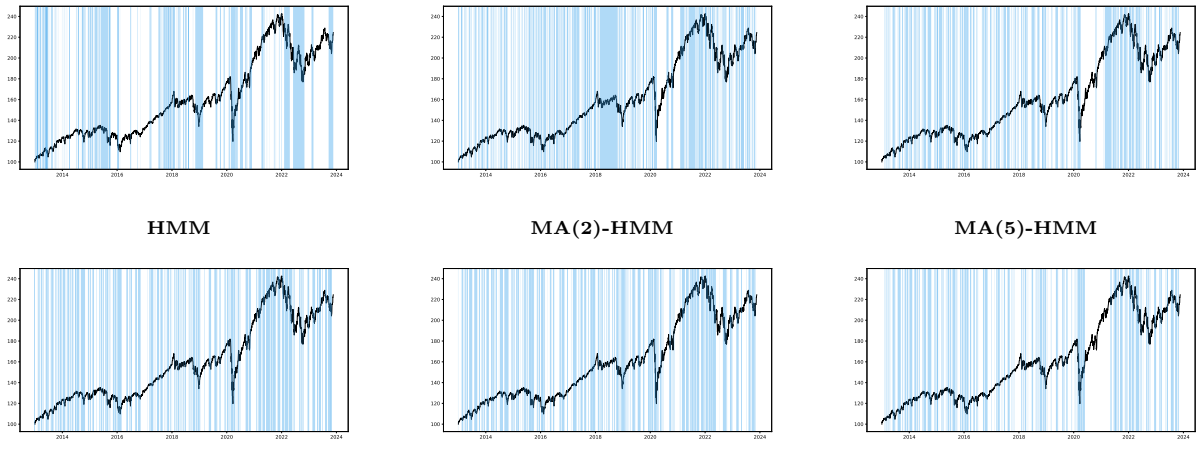
5.2 Signals study

In this section, we compare the signals obtained with the different methods outlined in Section 5.1 over ten years from 2013 to 2023. Table 3 shows the average $\widehat{\mathbf{SDS}}$, duration of states and back-and-forth switching rate for each of the signals, and Figure 15 and Figure 16 respectively display the estimated binary and continuous signal for each model.

	Average $\widehat{\mathbf{SDS}}$		Average state duration (days)		Back-and-forth switching rate
	Bull regime	Bear regime	Bull regime	Bear regime	
HMM	5.27	2.55	19.65	7.47	2.29%
MA(2)-HMM	0.60	0.50	7.94	4.87	1.90%
MA(5)-HMM	0.61	0.48	12.83	8.27	1.06%
MA(8)-HMM	0.64	0.45	15.07	10.29	0.77%
MA(10)-HMM	0.67	0.49	17.18	11.80	0.25%
MA(L)-HMM with varying L	0.54	0.43	13.84	9.91	1.02%

Table 3: Hidden states detection for different models on the MSCI World Index

In Table 3, we first observe that the estimated values of the **SDS** align with what we observed in the historical data in Section 5.1: only the average $\widehat{\mathbf{SDS}}$ of the **HMM** are by far greater than 1. We also observe that the average duration of both states increases with L for $L > 0$. The same inverse trend is observed for the back-and-forth switching rate. Regarding the **HMM**, we notice a long duration in the bull regime and a moderate duration in the bear regime, with a significantly higher back-and-forth change rate compared to all other models. This situation indicates a challenge for the model to accurately distinguish between the two states and to capture changes over an extended period. This is a scenario frequently encountered when the **SDS** exceeds 1. Finally, the model with varying L has the lowest $\widehat{\mathbf{SDS}}$ values without compromising the persistence of states.



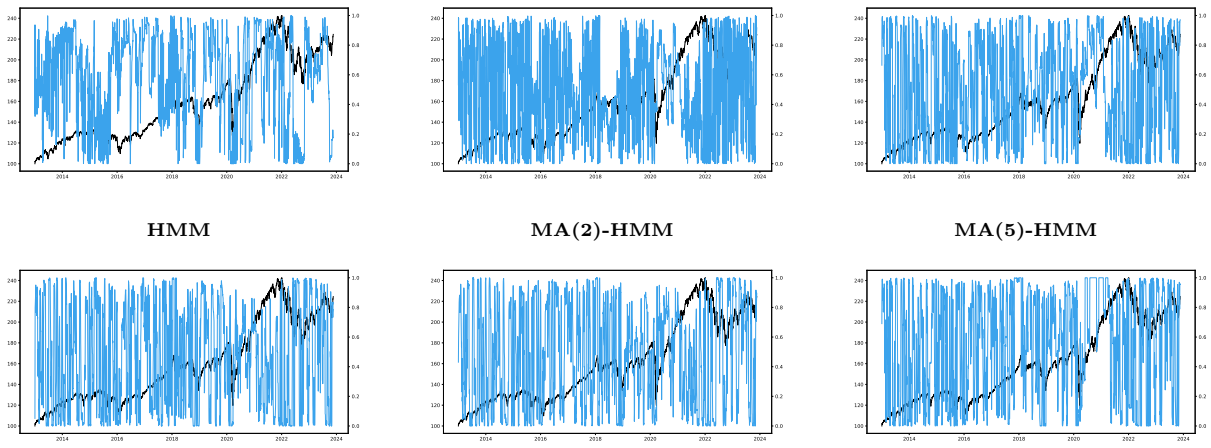
HMM **MA(2)-HMM** **MA(5)-HMM**

MA(8)-HMM **MA(10)-HMM** **MA(L)-HMM** with varying L

Figure 15: Binary signal on the MSCI World Index

In Figure 15, we observe that the **HMM** binary signal appears persistent at first glance, but the detected states are not always consistent, as seen in 2016 or 2022, and it struggles to detect rebounds, particularly

in the beginning of 2019 and mid-2020. It is worth noting that at this scale, the back-and-forth changes are not visible. As the value of L increases, it appears that the decoded states adhere more closely to the data. In the case of **MA(10)-HMM**, distinct bearish phases are clearly depicted in blue and bullish phases in white, with the model seemingly capturing all trends, which is less apparent for the **MA(2)-HMM**, for example. Short-term stays are less frequent with a higher order moving average. The model with a varying L resembles the model with **MA(10)-HMM**, but during strongly defined periods such as in 2021, there is less uncertainty, likely because during such periods, there is only one state and this model allows for switching to a 1-state model when the obtained $\widehat{\mathbf{SDS}}$ are not satisfactory. The model with a varying L is more flexible in adapting to special market periods. Overall, rebounds are the hardest to detect quickly, and this is true for all moving average depths, due to data noise for short moving averages, and lag for long moving averages.



HMM **MA(2)-HMM** **MA(5)-HMM**

MA(8)-HMM **MA(10)-HMM** **MA(L)-HMM** with model selection

Figure 16: Continuous signal on the MSCI World Index

Figure 16 illustrates that as the value of L increases, the continuous signal oscillates less, ensuring better persistence of regimes. The continuous signal of the **MA(L)-HMM** with varying L approaches that of an **MA(L)-HMM** with a high value of L, such as 8 or 10, but it notably remains at 100% during the bullish periods of late 2017 and late 2020, which aligns with the behavior of the one-state model. The selection of this model in this situation is completely justified based on the price history, validating our model selection protocol.

5.3 Strategies study

For the analysis of strategies, we rely on the Sharpe ratio, which represents the performance-volatility tradeoff. We display the annual Sharpe ratio of strategies based on binary signals in Figure 17, and those of strategies based on continuous signals in Figure 18.

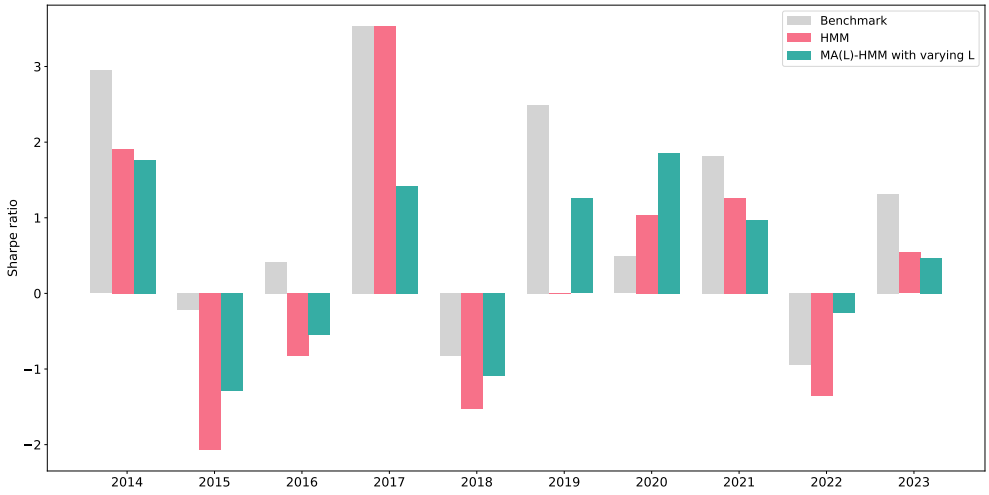


Figure 17: Sharpe ratio of the binary signal strategies on the MSCI World Index.

On Figure 17, we observe that the behavior of the **HMM** is erratic and not well-controlled. During strongly bullish phases, it alternates between being fully invested (2017), not invested at all (2019), and partially invested (2014). This inconsistency highlights a lack of coherence in its approach. Moreover, during bearish market conditions, the model consistently underperforms further, providing no protection against downturns.
The **MA(L)-HMM** consistently reduces losses during bearish periods compared to the **HMM**, notably outperforming the benchmark in 2022. During bullish periods, it significantly outperforms the **HMM** twice and maintains a Sharpe ratio close in other cases, except for 2017 when it mistakenly deactivated unlike the **HMM**.
These findings demonstrate that the **MA(L)-HMM** manages to adopt a more defensive strategy than the **HMM** during bearish periods, while maintaining a similar level of aggressiveness during bullish phases across different years.

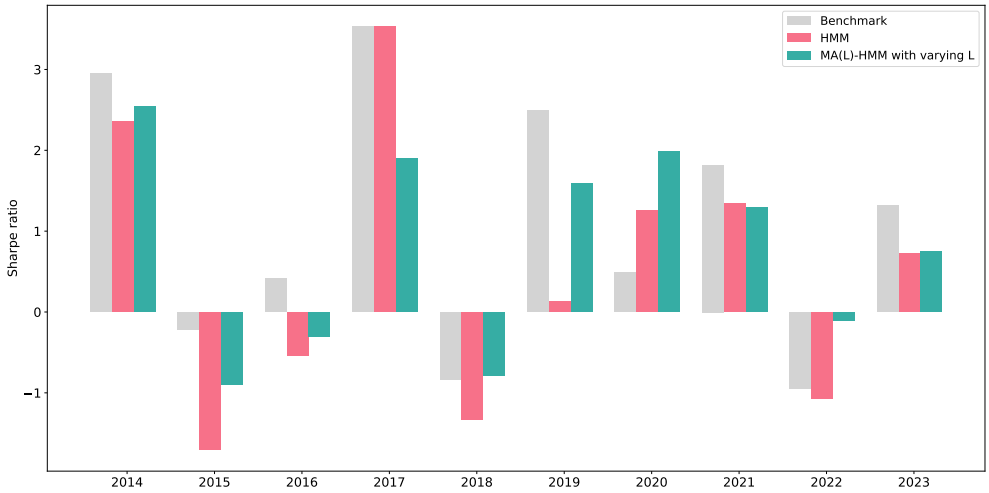


Figure 18: Sharpe ratio of the continuous signal strategies on the MSCI World Index

Figure 18 shows that using the continuous signal leads to better results for both models. The **MA(L)-HMM** manages to outperform the benchmark twice during bearish periods, making it even more pro-

tective. Over the entire period, it outperforms the **HMM** significantly five times, slightly three times, and once it has a Sharpe ratio very close to that of the **HMM**. Overall, except for 2017, it consistently performs at least as well as, and generally better than, the **HMM**.

We also examine the performance of strategies relative to their benchmark, the MSCI World Index. Specifically, we are interested in identifying any positive values among the information ratios, indicating that the benchmark has been outperformed over the period. The annual information ratios of each strategy are displayed in Table 4.

	Binary signal										Continuous signal									
	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
HMM	−0.05	−0.11	−0.09	**0.00**	−0.00	−0.19	**0.00**	−0.12	**0.04**	−0.11	−0.10	−0.07	−0.06	−0.22	**0.03**	−0.18	**0.01**	−0.16	**0.05**	−0.13
MA(L)-HMM with varying L	−0.04	−0.02	−0.05	−0.17	**0.03**	−0.11	**0.03**	−0.07	**0.08**	−0.07	−0.06	−0.03	−0.04	−0.21	**0.05**	−0.12	**0.01**	−0.09	**0.07**	−0.05

Table 4: Information ratio of the strategies on the MSCI World Index

According to Table 4, we observe that we struggle to outperform the benchmark. Apart from 2018, 2020, and 2022, the information ratios are almost always negative, indicating that we have underperformed the benchmark concerning both the volatility and the performance of our strategy. However, in these specific cases, we notice that using the **MA(L)-HMM** improves the information ratio, confirming that it is more protective than the **HMM**. Conversely, it is often the binary signal that displays the highest IR.

6 Conclusion and perspectives

This article explores the use of Bayesian statistics together with moving average smoothing to detect changes in market regimes and design investment strategies. The focus is on tailoring the **HMM** to serve as a foundation for designing a market regime indicator for financial strategies. Decoding and calibration of the **HMM** are done in a Bayesian way which allows us to generate several hidden states sequences and parameter values.

As data volatility increases the accuracy of **HMM** based regime declines. The lack of persistence with increasing volatility is detrimental to the investment strategies performance as the number of false alarm increases. To address these issues, we have adapted the **HMM** to accomodate smoothed data coming from a simple moving average procedure. The resulting model ressembled a higher order **HMM** along the line of models considered in Zhang et al. (2019) for instance. Smoothing the data allowed to act on the volatility to improve decoding accuracy and increase state persistence as a byproduct.

Alongside this work, we have come up with an indicator, referred to as the State Detection Score (**SDS**), to measure the distinguishability of states within our model. This highlights that when the intra-regime volatility estimated by σ_x is higher than the inter-regime gap between the means μ_x , then the states are no longer distinguishable. We use this indicator as a model selection tool to choose how many states to keep.

The outcomes of **MA(L)-HMM** demonstrated significant enhancements over the standard **HMM**. Notably, the **MA(L)-HMM** outperformed the standard **HMM**, showing improved accuracy in both high-volatility offline and online decoding. It enhanced state persistence, making it viable for investment strategies when accounting for transaction fees. Real-world data even demonstrated superior performance of the **MA(L)-HMM** based strategy over the benchmark in some years. Integrating a continuous signal reduces transaction costs and appears to be particularly efficient at mitigating significant losses, as indicated by the Sharpe ratios. The performance gap between continuous and binary decoding remains relatively low, however.

There is room for further improvement. In the future, we aim to explore **HMMs** with diverse stochastic dynamics, integrating factors such as credit spreads and interest rates. This joint modeling approach, encompassing market indices, interest rates, and credit spreads, promises a deeper comprehension and more accurate detection of market regimes. We anticipate that incorporating our detection methodology into investment strategies based on factors, as in the works of Erlwein et al. (2009) and Ammann and Verhofen (2006), could enhance performance, allowing for more efficient capitalization on both bullish and bearish market conditions.

All the results presented in this article are replicable and the code and data are available in the Github directory elchd/Data_driven_investment_strategies_using_Bayesian_inference_in_regime_switching_models.

Acknowledgements

The work of Pierre-O GOFFARD is supported by the French Research Agency through the ANR project DREAMeS (ANR-21-CE46-0002).

References

Liliana Gonzalez, John G. Powell, Jing Shi, and Antony Wilson. Two centuries of bull and bear market cycles. *International Review of Economics ; Finance*, 14(4):469–486, jan 2005. DOI: 10.1016/j.iref.2004.02.003.

Liliana Gonzalez, Philip Hoang, John G. Powell Massey, and Jing Shi. Defining and dating bull and bear markets: Two centuries of evidence. *Multinational Finance Journal*, 10(1/2):81–116, jun 2006. DOI: 10.17578/10-1/2-3.

Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247, aug 1969. DOI: 10.2307/1926560.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, may 1973. DOI: 10.1086/260062.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357, mar 1989. DOI: 10.2307/1912559.

Huntley Schaller and Simon Van Norden. Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177–191, apr 1997. DOI: 10.1080/096031097333745.

A. Jobert and L. C. G. Rogers. Option pricing with markov-modulated dynamics. *SIAM Journal on Control and Optimization*, 44(6):2063–2078, jan 2006. DOI: 10.1137/050623279.

Q. Zhang. Stock trading: An optimal selling rule. *SIAM Journal on Control and Optimization*, 40(1): 64–87, jan 2001. DOI: 10.1137/s0363012999356325.

G. Yin and X.Y. Zhou. Markowitz's mean-variance portfolio selection with regime switching: From discrete-time models to their continuous-time limits. *IEEE Transactions on Automatic Control*, 49(3): 349–360, mar 2004. DOI: 10.1109/tac.2004.824479.

Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, feb 1970. DOI: 10.1214/aoms/1177697196.

Olivier Cappé, Eric Moulines, and Tobias Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer, 2007. ISBN 9780387402642.

Sylvia. Richardson and Peter J. Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 59(4):731–792, nov 1997. DOI: 10.1111/1467-9868.00095.

Tobias Rydén. EM versus markov chain monte carlo for estimation of hidden markov models: a computational perspective. *Bayesian Analysis*, 3(4):659–688, dec 2008. DOI: 10.1214/08-ba326.

Juan Cheng, Chenghu Ma, and Zhibai Wang. The recognition of investor's sentiment and the trading strategy based on hmm. *2018 International Conference on Big Data and Artificial Intelligence (ICBDAI 2018)10.3390/math8071126*, 2018. URL https://www.webofproceedings.org/proceedings_series/ECS/ICBDAI2018/ICBDAI18025.pdf.

Jan Bulla, Sascha Mergner, Ingo Bulla, André Sesboüé, and Christophe Chesneau. Markov-switching asset allocation: Do profitable strategies exist? *Journal of Asset Management*, 12(5):310–321, March 2011. ISSN 1479-179X. DOI: 10.1057/jam.2010.27.

Peng Chen, Dongyun Yi, and Chengli Zhao. Trading strategy for market situation estimation based on hidden markov model. *Mathematics*, 8(7):1126, July 2020. ISSN 2227-7390. DOI: 10.3390/math8071126.

Nguyet Nguyen. Hidden markov model for stock trading. *International Journal of Financial Studies*, 6 (2):36, mar 2018. DOI: 10.3390/ijfs6020036.

M. Dai, Q. Zhang, and Q. J. Zhu. Trend following trading under a regime switching model. *SIAM Journal on Financial Mathematics*, 1(1):780–810, jan 2010. DOI: 10.1137/090770552.

Christina Erlwein, Rogemar Mamon, and Matt Davison. An examination of hmm-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry*, 27(3):204–221, December 2009. ISSN 1524-1904. DOI: 10.1002/asmb.820.

Manuel Ammann and Michael Verhofen. The effect of market regimes on style allocation. *Financial Markets and Portfolio Management*, 20(3):309–337, July 2006. ISSN 2373-8529. DOI: 10.1007/s11408-006-0018-2.

Nguyet Nguyen and Dung Nguyen. Hidden markov model for stock selection. *Risks*, 3(4):455–473, October 2015. ISSN 2227-9091. DOI: 10.3390/risks3040455.

Eun-chong Kim, Han-wook Jeong, and Nak-young Lee. Global asset allocation strategy using a hidden markov model. *Journal of Risk and Financial Management*, 12(4):168, November 2019. ISSN 1911-8074. DOI: 10.3390/jrfm12040168.

Andrew Ang and Geert Bekaert. How do regimes affect asset allocation. *Financial Analysts Journal*, 60 (2):86–99, March 2004. ISSN 1938-3312. DOI: 10.2469/faj.v60.n2.2612.

Andreas Graflund and Birger Nilsson. Dynamic portfolio selection: the relevance of switching regimes and investment horizon. *European Financial Management*, 9(2):179–200, May 2003. ISSN 1468-036X. DOI: 10.1111/1468-036x.00215.

Giorgio Costa and Roy H. Kwon. Risk parity portfolio optimization under a markov regime-switching framework. *Quantitative Finance*, 19(3):453–471, August 2018. ISSN 1469-7696. DOI: 10.1080/14697688.2018.1486036.

Giorgio Costa and Roy Kwon. A regime-switching factor model for mean-variance optimization. *SSRN Electronic Journal*, 2019. ISSN 1556-5068. DOI: 10.2139/ssrn.3609365.

Mengqi Zhang, Xin Jiang, Zehua Fang, Yue Zeng, and Ke Xu. High-order hidden markov model for trend prediction in financial time series. *Physica A: Statistical Mechanics and its Applications*, 517: 1–12, March 2019. ISSN 0378-4371. DOI: 10.1016/j.physa.2018.10.053.

William F. Sharpe. Mutual fund performance. *The Journal of Business*, 39(S1):119, jan 1966. DOI: 10.1086/294846.

William F. Sharpe. The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58, oct 1994. DOI: 10.3905/jpm.1994.409501.

Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966. ISSN 00034851. URL http://www.jstor.org/stable/2238772.

Appendices

A Gibbs algorithm prior and posterior distributions

Suppose that the parameters are mutually independent a priori and distributed as

$$\nu \sim \mathsf{Dir}(1,\ldots,1),\ q(x,\cdot) \sim \mathsf{Dir}(1,\ldots,1),\ \mu_x \sim \mathsf{Normal}(\xi, \kappa^{-1} = R^2),\ \sigma_x \sim \mathsf{Gamma}(\alpha = 2, \beta_x), \tag{4}$$

and $\beta_x \sim \mathsf{Gamma}(g = 0.2, h = 10/R^2)$, where $\xi = (\min y_{1:N} + \max y_{1:N})/2$ and $R = \max y_n - \min y_n$.

The posterior distribution of the initial distribution and the transition matrix of the Markov chain are given by

$$\nu|\ldots \sim \mathsf{Dir}(\{1 + \delta_x(x_1)\}_{x\in\mathcal{X}}), \text{ and } q(x,\cdot)|\ldots \sim \mathsf{Dir}(\{1 + n_{x,y}\}_{x\in\mathcal{X}}), \text{ for } x \in \mathcal{X},$$

where

$$\delta_x(y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{otherwise,} \end{cases}$$

and $n_{x,y} = \#\{2 \leq n \leq N\ ;\ x_{n-1} = x \text{ and } x_n = y\}$ is the number of transitions from state x to state y. Note that $q(x,\cdot)$ corresponds to a row of the transition matrix Q, and that each row is conditionally independent a posteriori.

The posterior distribution of the expected logarithmic returns are given by

$$\mu_x|\ldots \sim \mathsf{Normal}\left(\frac{\sum_{n:x_n=x} y_n + \xi\sigma_x\kappa}{n_x + \sigma_x^2\kappa}, \frac{\sigma_x^2}{n_x + \sigma_x^2\kappa}\right), \text{ for } x \in \mathcal{X}, \tag{5}$$

where $n_x = \#\{1 \leq n \leq N\ ,\ x_n = x\}$.

The posterior distribution of the variance parameters is given by

$$\sigma_x^{-2}|\ldots \sim \mathsf{Gamma}\left(\alpha + \frac{1}{2}N, \beta + \frac{1}{2}\sum_{n:x_n=x}(y_n - \mu_x)^2\right), \text{ for } x \in \mathcal{X}. \tag{6}$$

The conditional distribution is a posteriori independent across the market regimes in $\mathcal{X}$. Lastly, the posterior distribution of the hyperparameters of the variance prior distribution is given by

$$\beta|\ldots \sim \mathsf{Gamma}(g + \alpha, h + \sigma_x^{-2}). \tag{7}$$

Conditionally on the data and the model parameters, the hidden sequence of states is a non-homogeneous Markov chain with probability given by

$$\mathbb{P}(X_1 = x_1|\theta, y_{1:N}) \propto \nu(x_1)g_\theta(y_1|x_1)p(y_{2:N}|\theta, x_1), \tag{8}$$

and

$$\mathbb{P}(X_n = x_n|X_{n-1} = x_{n-1}, \theta, y_{1:N}) \propto q(x_{n-1}, x_n)g_\theta(y_n|x_n)p(y_{(n+1):n}|\theta, x_n). \tag{9}$$

The probabilities $p(y_{1:N}|\theta, x_1)$ and $p(y_{(n+1):N}|\theta, x_n)$ corresponds to the so-called backward variable of the celebrated Baum-Welch algorithm, see for instance Baum and Petrie (1966). The evaluation of these probabilities is done recursively and backward over n starting from the last observations. Since at time N there is no future information available to infer the current state,

$$p(y_N|\theta, x_N = i) = b_i(N) = 1.$$

Subsequently, we utilize this value to evaluate the preceding values with

$$p(y_{n:N}|\theta, x_n = i) = b_i(n) = \sum_{j=1}^{K} b_j(n+1)q(x_n = i, x_{n+1} = j)g_\theta(y_{n+1}|x_{n+1} = j)$$

with $K = 1, ..., d$, the finite state space.

Since at time N there is no future information available to infer the current state,

$$p(y_N|\theta, x_N = i) = b_i(N) = 1.$$

The Gibbs sampler reduces to a for loop that sequentially samples from (4), (5), (6) and (7) followed by a multinomial sampling from (8) and (9) to reconstruct the sequence of hidden states. It is worth noting that a burn-in period is typically necessary for convergence, after which all estimations are retained for further analysis. These estimates can be used by taking their average or through other methods once this burn-in period has passed. The probability distribution involved in the Gibbs sampler is provided in Appendix B. The sampling algorithm is summarized in Appendix C.

B Table of probability distribution

Name	Parameters		p.d.f.
Dirichlet	$\mathsf{Dir}(\alpha)$	$\{\alpha_1, ..., \alpha_K\}_{K\geq 2} > 0$	$\frac{\prod_{i=1}^K x_i^{\alpha_i - 1}}{B(\alpha)},\ \alpha_0 = \sum_{i=1}^K \alpha_i$
Exponential	$\mathsf{Exp}(\delta)$	$\delta > 0$	$\delta e^{-\delta x},\ x > 0$
Gamma	$\mathsf{Gamma}(\alpha, \beta)$	$\alpha, \beta > 0$	$\frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)},\ x > 0$
Normal	$\mathsf{Normal}(\mu, \sigma)$	$\mu \in \mathbb{R},\ \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\ x \in \mathbb{R}$

Table 5: List of distributions.

C Algorithmic details

Algorithm 1 Gibbs sampler for $\pi(\theta, x_{1:N}|y_{1:N})$

Input: The number of states $K \in \mathbb{N}$, the observations $y_{1:N}$, the prior distributions for θ and $x_{1:N}$

Output: The sampled posterior distributions for θ and $x_{1:N}$

1: **initialize** $\theta = \{M, \Sigma, \nu, Q\}$ and β

2: **for** $niter = 1$ to $max\_iterations$ **do**

3: Sample $M = \{\mu_i\}_{i=1}^K$ from $\mu_i|... \sim \mathsf{Normal}\left(\frac{S_i + \kappa\xi\sigma_i^2}{n_i + \kappa\sigma_i^2}, \frac{\sigma_i^2}{n_i + \kappa\sigma_i^2}\right)$

4: Order ascendingly M

5: Sample $\Sigma^2 = \{\sigma_i^2\}_{i=1}^K$ from $\sigma_i^2|... \sim \mathsf{Inv-Gamma}\left(\alpha + \frac{1}{2}n_i, \left(\beta_i + \frac{1}{2}\sum_{n=1}^N (y_n - \mu_i)^2\right)^{-1}\right)$

6: Sample $\beta = \{\beta_i\}_{i=1}^K$ from $\beta_i|... \sim \mathsf{Gamma}\left(g + \alpha, h + \sigma_i^{-2}\right)$

7: Sample $Q = \{q_{ij}\}_{i,j=1}^K$ from $(q_{i1}, ...q_{iK})|... \sim \mathsf{Dir}\left(n_{i1} + 1, ...n_{iK} + 1\right)$

8: Sample $\nu = \{\nu_i\}_{i=1}^K$ from $\nu_i|... \sim \mathsf{Dir}\left(\mathbb{1}_{X_1=1} + 1, ...\mathbb{1}_{X_K=1} + 1\right)$

9: Sample $x_1 \sim \mathsf{Multinomial}\left(\nu_1 G(1, y_1) b_1(1), ...\nu_K G(K, y_1) b_K(1)\right)$

10: Sample $\{x_n\}_{n=2}^N \sim \mathsf{Multinomial}\left(\sum_{i=1}^K q_{i1} G(1, y_n) b_1(n), ... \sum_{i=1}^K q_{iK} G(K, y_n) b_K(n)\right)$

11: **end for**

12: **return** (niter - nburnin) last samples of θ and $x_{1:N}$

$S_i = \sum_{n=1}^N \mathbb{1}_{x_n=i} y_n$, $n_i = \sum_{n=1}^N \mathbb{1}_{x_n=i}$ and $n_{ij} = \sum_{n=1}^N \mathbb{1}_{x_{n-1}=i, x_n=j}$

To maintain consistent correspondence between the sampled parameters and their respective states in a uniform order, we promptly order M upon sampling it, before sampling the remaining parameters given our draw of M. The identification issue in Richardson and Green (1997) is no more in Rydén (2008) due to the underlying Markov chain.