Pairwise Markov Models for Stock Index Forecasting

Ivan Gorynin, Emmanuel Monfrini and Wojciech Pieczynski

SAMOVAR, Telecom SudParis, CNRS, Université Paris-Saclay

Evry, France

Email: ivan.gorynin@telecom-sudparis.eu

Abstract—Common well-known properties of time series of financial asset values include volatility clustering and asymmetric volatility phenomenon. Hidden Markov models (HMMs) have been proposed for modeling these characteristics, however, due to their simplicity, HMMs may lack two important features. We identify these features and propose modeling financial time series by recent Pairwise Markov models (PMMs) with a finite discrete state space. PMMs are extended versions of HMMs and allow a more flexible modeling. A real-world application example demonstrates substantial gains of PMMs compared to the HMMs.

Index Terms—Hidden Markov models, Forecasting, Financial time series, Pairwise Markov models, Technical analysis.

I. Introduction

Stock market prediction remains a significant challenge of modern risk management theory and practice. Universally acknowledged features of financial time series include volatility clustering, autocorrelation in returns and asymmetric volatility phenomenon (AVP). A well-established methodology consists in using a mathematical model to describe available data and to project it into the future. The autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroscedasticity (GARCH) models are popular among practitioners. These models are reviewed in [1]. The GARCH model describes the volatility clustering in the data and some of its variants describe the AVP as well, while the ARIMA model describes autocorrelation in returns. Alternative techniques include artificial neural networks [2], fuzzy logic [3], support vector machines classifiers [4] and their combinations.

In recent years, there was an increasing interest in the regime-switching models, reviewed *e.g.* in [5]. In financial markets, these models allow identifying *bull* and *bear* alternating regimes. A *bull* state is characterized by positive expectation of log-returns and low volatility, while a *bear* state is driven by negative expected log-returns and high volatility. The hidden Markov models (HMMs) provide a suitable framework for modeling regime-switching. An important example of such framework is available in *e.g.* [6]. These models use a hidden sequence of the same length as the sequence of observed log-returns. The HMMs are known to be robust and straightforward to implement. However, the HMMs do not take the following potential features of stock dynamics into account:

- (F1): log-returns may be correlated conditional on the state variables;

- (F2): the future state and current log-return may not be independent conditional on the current state.

The Pairwise Markov models (PMMs) are introduced and studied in [7] as a general statistical concept. Particularly, they are able to include both features (F1) and (F2) in the HMMs for the same processing cost.

The purpose of the paper is to introduce a modeling of financial time series with the PMMs. Specifically, we investigate if the PMMs can allow improving forecasting performance and if both features (F1) and (F2) should be taken into account. Throughout this paper, we assume that the state space is finite discrete in both HMMs and PMMs.

The paper is organized as follows. In Section II we recall the Hidden and Pairwise Markov models. Section III is devoted to modeling stock dynamics with the PMMs and to related estimation methods. Section IV contains experiments on real-world data and Section V is a discussion of the results. Section VI concludes the paper and presents perspectives for further researches.

II. Models

Let $N > 0$, $(Y_1, Y_2, ..., Y_N) = Y_{1..N}$ be a time series and Ω be a finite discrete set. The idea is to describe the probability distribution of $Y_{1..N}$ by using a hidden time series $R_{1..N}$, where for each n in $\{1, .., N\}$, $R_n \in \Omega$. Specifically, one defines the probability distribution $p(r_{1..N}, y_{1..N})$ of the pair $(R_{1..N}, Y_{1..N})$. In this case, we have

$$p(y_{1..N}) = \sum_{r_{1..N} \in \Omega^N} p(r_{1..N}, y_{1..N}).$$

Both HMMs and PMMs are used to define $p(r_{1..N}, y_{1..N})$. In this section we recall the definition and statistical properties of these models.

A. Hidden Markov Models

Definition 1. *A probability distribution $p(r_{1..N}, y_{1..N})$ is a Hidden Markov model of $Y_{1..N}$ if it verifies*

$$p(r_{1..N}, y_{1..N}) = p(r_1) \prod_{n=1}^{N-1} p(r_{n+1} | r_n) \prod_{n=1}^{N} p(y_n | r_n).$$

Any HMM has the following properties:

- (P1): $R_{1..N}$ is a Markov chain;

- (P2): $(Y_1, Y_2, ..., Y_N)$ are independent conditional on $R_{1..N}$;
- (P3): For each n in $\{1, .., N\}$, $p(y_n | r_{1..N}) = p(y_n | r_n)$.

In practice, one specifies the family of distributions to which $p(y_n | r_n)$ belongs to.

B. Pairwise Markov Models

Definition 2. *A probability distribution $p(r_{1..N}, y_{1..N})$ is a Pairwise Markov model of $Y_{1..N}$ if $(R_{1..N}, Y_{1..N})$ is a Markov chain,* i.e.

$$p(r_{1..N}, y_{1..N}) = p(r_1, y_1) \prod_{n=1}^{N-1} p(r_{n+1}, y_{n+1} | r_n, y_n).$$

Figure 1 presents directed dependency graphs of HMM and PMM. Consider the following decomposition of $p(r_{n+1}, y_{n+1} | r_n, y_n)$, for n in $\{1, .., N-1\}$:

$$p(r_{n+1}, y_{n+1} | r_n, y_n) = p(r_{n+1} | r_n, y_n) \, p(y_{n+1} | r_n, r_{n+1}, y_n).$$
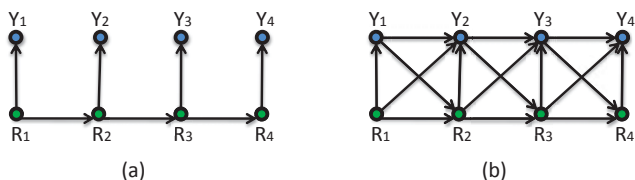

Fig. 1: Directed dependency graphs of HMM (a) and PMM (b).

From the above equation, we see that a PMM is an HMM if, and only if, for each n in $\{1, .., N-1\}$:

$$p(y_{n+1} | r_n, r_{n+1}, y_n) = p(y_{n+1} | r_{n+1}); \qquad (1a)$$

$$p(r_{n+1} | r_n, y_n) = p(r_{n+1} | r_n). \qquad (1b)$$

We also consider two subclasses of PMMs where only one of constraints (1a)-(1b) is relaxed.

Definition 3. *PMM $(R_{1..N}, Y_{1..N})$ is called PMM-F1 (PMM-F2) if it verifies, for each n in $\{1, .., N-1\}$, eq. (1b) (eq. (1a)), respectively.*

Figure 2 presents directed dependency graphs of PMM-F1 and PMM-F2. In practice, one should specify the families of distributions to which $p(y_n | r_n)$, $p(r_{n+1} | r_n, y_n)$ and $p(y_{n+1} | r_n, r_{n+1}, y_n)$ belong to.

III. METHODS

In this section, we introduce a Pairwise Markov modeling of asset log-returns. Specifically, we explain how the PMMs allow modeling features (F1) and (F2) mentioned in Section I. We also outline various types of PMM data processing, such as the state estimation, forecasting and parameter inference.
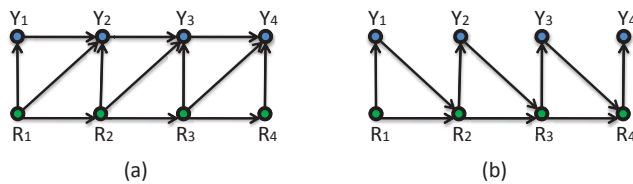

Fig. 2: Directed dependency graphs of PMM-F1 (a) and PMM-F2 (b).

A. Modeling financial time series

Let S_n be the stock price at time n, $n \in \mathbb{N}$. The log-return Y_n at time $n > 0$ is defined by

$$Y_n = \log(S_n) - \log(S_{n-1}). \qquad (2)$$

In the classic Black-Scholes model, the log-returns $Y_{1..N}$ are assumed to be normally distributed and to have the same mean μ and standard deviation σ. In other words, we have, for each $n > 0$,

$$Y_n = \mu + \sigma U_n,$$

where $\{U_n\}_{n>0}$ are zero-mean, unit-variance independent Gaussian random variables, also known as the standard Gaussian white noise. μ and σ are known as the average return (or drift) and the volatility of the stock.

The HMM allows extending the classic Black-Scholes model by making μ and σ dependent on hidden variables. Let $R_{1..N}$ be a Markov chain, then let

$$Y_n = \mu(r_n) + \sigma(r_n) U_n, \qquad (3)$$

with $\{U_n\}_{1 \le n \le N}$ standard Gaussian white noise variables. The parameters of this model include the initial state distribution $p(r_1 = i)$ for each $i \in \Omega$, Markov chain transition matrix $p(r_{n+1} = j | r_n = i)$ for each $i, j \in \Omega$ and the values of the drift and volatility per state $\{\mu(i), \sigma(i)\}_{i \in \Omega}$. For example, if ω_1 is associated with the *bear* market state and ω_2 with the *bull* state, one would expect $\mu(\omega_1) < 0 < \mu(\omega_2)$ and $\sigma(\omega_1) > \sigma(\omega_2)$. The Hidden Markov modeling of $Y_{1..N}$ is given by (P1)-(P3) and

$$\forall n, 1 \le n \le N, p(y_n | r_n) = \mathcal{N}\left(\mu(r_n), \sigma(r_n)^2\right), \qquad (4)$$

where $\mathcal{N}(., .)$ denotes the normal probability distribution with specified mean and variance.

The PMMs provide a more flexible framework than that of HMMs. In order to fulfill the requirement (F1) presented in Section I, we define a first-order autoregressive model of $Y_{1..N}$ conditional on $R_{1..N}$. We set

$$U_{n+1} = \rho(R_n, R_{n+1}) U_n + \sqrt{1 - \rho(R_n, R_{n+1})^2} V_{n+1}, \qquad (5)$$

where $n > 0$, $U_1, \{V_n\}_{n>0}$ are standard Gaussian white noise variables and for each $i, j \in \Omega, |\rho(i, j)| < 1$.

As regards the feature (F2), we make R_{n+1} dependent on Y_n conditional on R_n by using the concept of the logistic

function. Specifically, in the case where Ω has only two elements $\{\omega_1, \omega_2\}$, we set

$$p\left(\mathrm{r}_{n+1} = \omega_1 \,|\mathrm{r}_n, \mathrm{u}_n\right) = \frac{1}{1 + e^{-a(\mathrm{r}_n) - b(\mathrm{r}_n)\mathrm{u}_n}}, \qquad (6)$$

where for each $i \in \Omega$, $a(i) \in \mathbb{R}, b(i) \in \mathbb{R}$.

Finally, we combine (3), (5) and (6) to define a pairwise Markov modeling of $\mathrm{Y}_{1..N}$:

$$p\left(\mathrm{y}_1 \,|\mathrm{r}_1\right) = \mathcal{N}\left(\mu(\mathrm{r}_1), \sigma^2(\mathrm{r}_1)\right); \qquad (7\mathrm{a})$$

$$p\left(\mathrm{r}_{n+1} = \omega_1 \,|\mathrm{r}_n, \mathrm{y}_n\right) = \frac{1}{1 + e^{-a(\mathrm{r}_n) - \frac{b(\mathrm{r}_n)}{\sigma(\mathrm{r}_n)}(\mathrm{y}_n - \mu(\mathrm{r}_n))}}; \quad (7\mathrm{b})$$

$$p\left(\mathrm{y}_{n+1} \,|\mathrm{r}_n, \mathrm{r}_{n+1}, \mathrm{y}_n\right) =$$
$$\mathcal{N}\left(\mu(\mathrm{r}_{n+1}) + \frac{\rho(\mathrm{r}_n, \mathrm{r}_{n+1})\sigma(\mathrm{r}_{n+1})}{\sigma(\mathrm{r}_n)}(\mathrm{y}_n - \mu(\mathrm{r}_n))\right.,$$
$$\left.\sigma(\mathrm{r}_{n+1})^2(1 - \rho(\mathrm{r}_n, \mathrm{r}_{n+1})^2)\right). \qquad (7\mathrm{c})$$

The parameters of this model are

$$\theta = \{\pi(i), \mu(i), \sigma(i), a(i), b(i), \rho(i,j)\}_{i,j \in \{\omega_1, \omega_2\}}, \qquad (8)$$

where $\pi(i) = p\left(\mathrm{r}_n = i\right)$. This model is presented for $\Omega = \{\omega_1, \omega_2\}$, but one can consider a more general definition by using the multinomial logistic function, as explained in [8].

B. State estimation and forecasting

Real-time processing of incoming data $\{\mathrm{Y}_n\}_{n>0}$ in a PMM involves determining $p\left(\mathrm{r}_n \,|\mathrm{y}_{1..n}\right)$, known as the filtering distribution. Algorithm 1, derived in [7], allows computing the filtering distribution with a complexity linear in n.

Algorithm 1. *Filtering in PMMs*
- *Consider, $\forall n > 0, \mathrm{r}_n \in \Omega$, $\alpha_n(\mathrm{r}_n) = p\left(\mathrm{r}_n, \mathrm{y}_{1..n}\right)$;*
- *Initialization: $\forall \mathrm{r}_1 \in \Omega, \alpha_1(\mathrm{r}_1) = p\left(\mathrm{y}_1 \,|\mathrm{r}_1\right) p\left(\mathrm{r}_1\right)$;*
- *Recursion:*
  Given $\{\alpha_n(\mathrm{r}_n)\}_{\mathrm{r}_n \in \Omega}$ and y_{n+1}, compute, $\forall \mathrm{r}_{n+1} \in \Omega$,

  $$\alpha_{n+1}(\mathrm{r}_{n+1}) = \sum_{\mathrm{r}_n} \alpha_n(\mathrm{r}_n)p\left(\mathrm{r}_{n+1}, \mathrm{y}_{n+1} \,|\mathrm{r}_n, \mathrm{y}_n\right).$$

  The filtering distribution is given by

  $$p\left(\mathrm{r}_n \,|\mathrm{y}_{1..n}\right) = \frac{\alpha_n(\mathrm{r}_n)}{\sum\limits_{\mathrm{r}_n \in \Omega} \alpha_n(\mathrm{r}_n)}.$$

Forecasting consists in computing $p\left(\mathrm{y}_{n+1..n+p} \,|\mathrm{y}_{1..n}\right)$ for $p > 0$. An important case of forecasting is the one-step-ahead forecasting, for which $p = 1$. In this case, it is also particularly important to forecast Z_{n+1}, where

$$\mathrm{Z}_{n+1} = \begin{cases} 1 & \text{if } \mathrm{Y}_{n+1} < 0; \\ 2 & \text{otherwise.} \end{cases} \qquad (9)$$

Z_{n+1} represents the direction of the stock price change during the day $n + 1$. The anticipated price change at $n + 1$ given the information available at n is defined by

$$\widehat{z}_{n+1|n} = \begin{cases} 1 & \text{if } p\left(\mathrm{y}_{n+1} < 0 \,|\mathrm{y}_{1..n}\right) > 0.5; \\ 2 & \text{otherwise.} \end{cases} \qquad (10)$$

Algorithm 2. *One-step-ahead forecasting in PMMs*
Let $n > 0$,
- *Compute $p\left(\mathrm{r}_n \,|\mathrm{y}_{1..n}\right)$ by using Algorithm 1;*
- *Compute $p\left(\mathrm{r}_n, \mathrm{r}_{n+1} \,|\mathrm{y}_{1..n}\right)$:*

  $$p\left(\mathrm{r}_n, \mathrm{r}_{n+1} \,|\mathrm{y}_{1..n}\right) = p\left(\mathrm{r}_n \,|\mathrm{y}_{1..n}\right) p\left(\mathrm{r}_{n+1} \,|\mathrm{r}_n, \mathrm{y}_n\right);$$

- *Compute, for each r_n, r_{n+1} in Ω, $\widehat{m}_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1})$ and $\widehat{s}^2_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1})$ by respectively*

  $$\mu(\mathrm{r}_{n+1}) + \frac{\rho(\mathrm{r}_n, \mathrm{r}_{n+1})\sigma(\mathrm{r}_{n+1})}{\sigma(\mathrm{r}_n)}(\mathrm{y}_n - \mu(\mathrm{r}_n))$$
  $$\text{and } (1 - \rho(\mathrm{r}_n, \mathrm{r}_{n+1})^2)\sigma^2(\mathrm{r}_{n+1});$$

- *The predictive distribution $p\left(\mathrm{y}_{n+1} \,|\mathrm{y}_{1..n}\right)$ is a mixture of normal densities $\mathcal{N}\left(\widehat{m}_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1}), \widehat{s}^2_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1})\right)$ with weights $p\left(\mathrm{r}_n, \mathrm{r}_{n+1} \,|\mathrm{y}_{1..n}\right)$. Compute the one-step-ahead forecast*

  $$\widehat{\mathrm{y}}_{n+1|n} \stackrel{\text{def}}{=\!=} \mathbb{E}\left[\mathrm{Y}_{n+1} \,|\mathrm{y}_{1..n}\right]$$

  as the mean of the mixture, that is

  $$\widehat{\mathrm{y}}_{n+1|n} = \sum_{\mathrm{r}_n, \mathrm{r}_{n+1} \in \Omega} p\left(\mathrm{r}_n, \mathrm{r}_{n+1} \,|\mathrm{y}_{1..n}\right) \widehat{m}_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1});$$
  $$(11)$$

- *Let Φ denote the normal cumulative distribution function, compute $p\left(\mathrm{y}_{n+1} < 0 \,|\mathrm{y}_{1..n}\right)$ by*

  $$\sum_{\mathrm{r}_n, \mathrm{r}_{n+1} \in \Omega} p\left(\mathrm{r}_n, \mathrm{r}_{n+1} \,|\mathrm{y}_{1..n}\right) \times \Phi\left(-\frac{\widehat{m}_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1})}{\widehat{s}_{n+1}(\mathrm{r}_n, \mathrm{r}_{n+1})}\right);$$

Contrary to the one-step-ahead forecasting, there is no apparent closed-form expression for $p\left(\mathrm{y}_{n+1..n+p} \,|\mathrm{y}_{1..n}\right)$ in the case of multistep forecasting in PMMs.

C. Parameter estimation

Let $N > 0$, $\mathrm{Y}_{1..N}$ be an observed time series of log-returns.

The goal of a PMM parameter estimation is to infer the parameter vector θ (8) from the observed data $\mathrm{Y}_{1..N}$.

The Expectation-Maximization (EM) and the Iterative Conditional Estimation (ICE) are well-known parameter estimation algorithms and are similar to the maximum likelihood estimation. These algorithms are well suited for both HMMs and PMMs, and the details may be found in [9].

Alternatively, θ can be estimated by using the principle of empirical risk minimization (ERM). Several methods for proving consistency of such estimators are provided in *e.g.* [10]. Let us recall the general idea of the ERM. Assume a training set $(\mathrm{x}_{1..N}, \mathrm{y}_{1..N})$ in $(\mathcal{X} \times \mathcal{Y})^N$, a prediction function $h : \mathcal{X} \to \mathcal{Y}$ and a loss function $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$. The empirical risk associated with the prediction function h is defined as

$$\widehat{R}(h) = \frac{1}{N} \sum_{n=1}^{N} L(h(\mathrm{x}_n), \mathrm{y}_n).$$

Thus, the idea of the ERM is to find a function h for which the risk is minimal.

Regarding the context of forecasting, we have $\mathrm{x}_n = \mathrm{y}_{1..n}$ and $h(\mathrm{x}_n) = \widehat{\mathrm{y}}^\theta_{n+1|n}(\mathrm{y}_{1..n})$, where $\widehat{\mathrm{y}}^\theta_{n+1|n}(\mathrm{y}_{1..n})$ is computed

from θ and $y_{1..n}$ by (11). We consider the following loss functions:

$$L_1(\widehat{y}^{\theta}_{n+1|n}(y_{1..n}), y_{n+1}) = |\widehat{y}^{\theta}_{n+1|n}(y_{1..n}) - y_{n+1}|,$$
$$L_2(\widehat{y}^{\theta}_{n+1|n}(y_{1..n}), y_{n+1}) = (\widehat{y}^{\theta}_{n+1|n}(y_{1..n}) - y_{n+1})^2.$$

The associated risk functions are

$$\widehat{R}_1(\theta) = \frac{1}{N-1} \sum_{n=1}^{N-1} |\widehat{y}^{\theta}_{n+1|n}(y_{1..n}) - y_{n+1}|, \qquad (13a)$$

$$\widehat{R}_2(\theta) = \frac{1}{N-1} \sum_{n=1}^{N-1} (\widehat{y}^{\theta}_{n+1|n}(y_{1..n}) - y_{n+1})^2. \qquad (13b)$$

Let $\lambda > 0$, the following risk function realizes a trade-off between $\widehat{R}_1(\theta)$ and $\widehat{R}_2(\theta)$:

$$\widehat{R}(\theta; \lambda) = \lambda \widehat{R}_1(\theta) + \widehat{R}_2(\theta). \qquad (14)$$

In our study, we estimate θ by minimizing (14) for various values of λ. There is no closed expression known for the corresponding update equations and we solve the optimization problem by the particle swarm optimization (PSO). PSO methods [11] are non-convex global optimization algorithms.

IV. EXPERIMENTS

Let us present our methodology to compare the efficiency of PMM with that of HMM on historical stock quotes. Given a data set $\mathscr{H} = \{y_1, .., y_M\}$ with successive daily log-returns of an asset $\mathscr{E}$, we split $\mathscr{H}$ into two juxtaposed sets as follows: $\mathscr{H}_{\text{training}} = \{y_1, .., y_N\}$ and $\mathscr{H}_{\text{test}} = \{y_{N+1}, .., y_M\}$. The first set is used to estimate the parameter θ by minimizing (14) for a given λ, while the second set serves only to assess the efficiency of each model considered. The models are compared in terms of the outcome produced by the following trading system. At the beginning of each day $n+1$, $N \le n < M$, the system buys asset $\mathscr{E}$ only if the one-day-ahead forecast (10) produced by the model is positive, *i.e.* if $\widehat{z}_{n+1|n} = 2$, and sells the asset at the end of the day. In the case of a negative forecast, the system avoids any trading operations on $\mathscr{E}$. Next, we compute the absolute return of the system on $\mathscr{H}_{\text{test}}$ and compare it with that of the asset. Let us recall that the absolute return of $\mathscr{E}$ relative to date N is defined as

$$\tau(n; N) = \frac{S_n - S_N}{S_N}, \qquad (15)$$

for $n \ge N$. Equivalently, $\tau(n; N)$ can be written as a function of the log-returns:

$$\tau(n; N) = \exp\left(\sum_{t=N+1}^{n} y_t\right) - 1.$$

Thus, the absolute return of the trading system considered can be written as

$$\tau^*(n; N) = \exp\left(\sum_{t=N}^{n-1} y_{t+1} \delta(\widehat{z}_{t+1|t} = 2)\right) - 1. \qquad (16)$$

We apply this methodology to Cliffs Natural Resources Stock prices (NYSE:CLF). Stock quotes are taken from the

Yahoo! database and correspond to the business days from 01/02/1990 to 12/13/1993 for $\mathscr{H}_{\text{training}}$ and from 12/14/1993 to 09/29/1994 for $\mathscr{H}_{\text{test}}$. In this configuration, the size of $\mathscr{H}_{\text{training}}$ is $N = 1000$, the size of $\mathscr{H}_{\text{test}}$ is 200 and the total size of the data set $\mathscr{H}$ is $M = 1200$. In every experiment, the state space consists of only two elements. Figures 3 and 4 display the values of risks $\widehat{R}_1(\theta)$ and $\widehat{R}_2(\theta)$ *cf.* (13) for θ minimizing (14), in function of λ. Absolute returns generated by four models on the test set are given in Table I for various values of λ. Figure 5 displays the returns produced per each model in function of time with $\lambda = 0$.

	$\lambda = 10^{-3}$	$\lambda = 10^{-2}$	$\lambda = 1$	$\lambda = 10^2$	$\lambda = 10^3$
HMM	17%	13%	10%	10%	10%
PMM-F1	16%	14%	11%	9%	9%
PMM-F2	21%	20%	19%	14%	16%
PMM	21%	20%	19%	14%	16%

TABLE I: Absolute returns (16) of HMM, PMM-F1, PMM-F2 and PMM-based trading systems on NYSE:CLF historical prices. The returns are related to the period from 12/14/1993 to 09/29/1994.
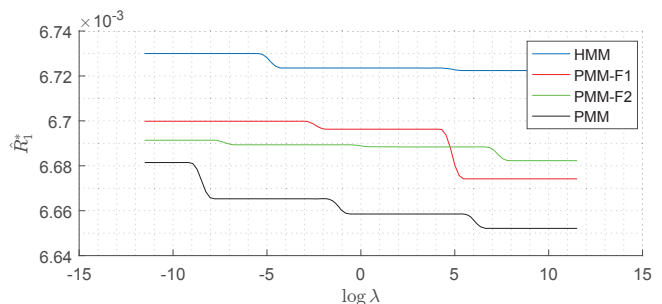

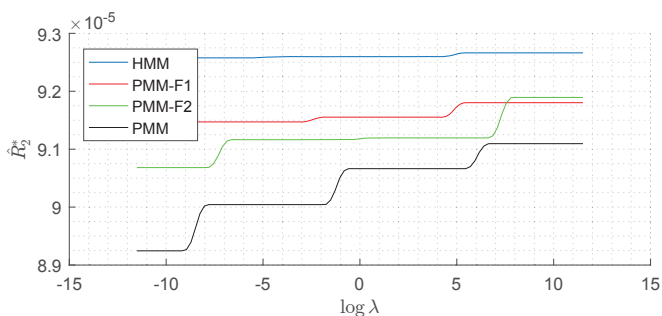Fig. 3: Values $\widehat{R}_1^*(\lambda) = \widehat{R}_1(\theta)$ in function of λ, where θ minimizes (14).

Fig. 4: Values $\widehat{R}_2^*(\lambda) = \widehat{R}_2(\theta)$ in function of λ, where θ minimizes (14).

Let us make several brief observations.

Figures 3 and 4 are consistent with the definition of θ as the minimum of (14). When λ increases, $\widehat{R}_1^*(\lambda) = \widehat{R}_1(\theta)$
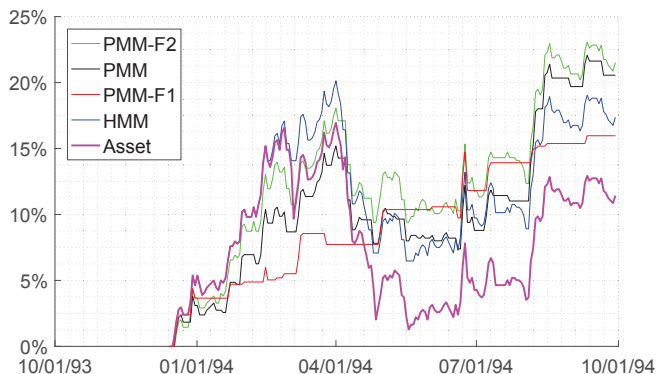
Fig. 5: Absolute returns (16) from 12/14/1993 generated by PMM-based trading systems on NYSE:CLF historical data. PMM models are estimated on the data from 01/02/1990 to 12/13/1993 by minimizing (14) with $\lambda = 0$. Four charts (from top to bottom) relate to the four models. The last chart is the absolute return of the asset (15).

decreases and $\widehat{R}_2^*(\lambda) = \widehat{R}_2(\theta)$ increases, and vice versa, and this holds for the four models.

Progressive inclusion of features (F1) and (F2) in the HMM improves both risk values computed on $\mathcal{H}_{\text{training}}$, as expected, independently of the value of λ.

We can see from Figure 5, that PMM-F1 implies a more risk-adverse trading strategy than that of HMM, and the related generated return increases almost monotonically. However, PMM-F1 may not be well suited for a *bull* market. PMM-F2 and HMM appear to be better suited for *bull* dynamics, while PMM-F2 seems to be less vulnerable than HMM to abrupt drops of asset value.

V. DISCUSSION

We proposed a meaningful parameterization of PMM for modeling financial time series. The results show that both features (F1) and (F2), mentioned in Section I, can be captured by PMMs, which was expected. One can intuitively understand why using the feature (F1) should improve forecasting, while (F2) is more difficult to interpret. Suppose for example that during the *bull* state, the return Y_n appears to be excessively negative compared to the average return of the *bull* market. In this case, the current state may become fairly uncertain in an HMM, *i.e.* $p(r_n = \omega_1 | y_{1..n}) \approx p(r_n = \omega_2 | y_{1..n})$. The PMM incorporates (F2) by using the distribution $p(r_{n+1} | r_n, y_n)$ which allows to decide to which extent Y_n should affect the expectation of R_{n+1}.

Table I indicates that the outcome produced by each model is sensitive to the value of λ. In general, such a parameter should be chosen by a cross-validation procedure accordingly to the application considered.

Our experiments indicate that a more complex structure of PMMs may allow identifying better suited regimes for specific application. We believe that the presented way of use of the

flexibility of PMM will allow overcoming principal constraints of HMMs.

This study has several limitations. Firstly, we assume only two regimes in our models. Next, the Gaussian mixture density and non-Gaussian heavy tailed observation distributions could be considered as well. We consider only closing price per day, while daily opening, low and high prices are also available as well. Finally, our study concerns only one period of stock prices and only one stock was used in the experiment. An upcoming research article will contain more extensive experiments and deal with the outlined points.

VI. CONCLUSION

The paper introduces a Pairwise Markov model for financial time series, obtained by incorporating features (F1) and (F2), mentioned in the Introduction, into the classic Hidden Markov model. The results show that both of these features contribute to improving the performance of the model in some applications. Let us mention the triplet Markov models, which allow, in particular, dealing with the mixture observation distributions and semi-Markovian hidden process simultaneously [12]. Such general models models are also potentially capable of improving the outcome of the classic HMMs.

REFERENCES

[1] D. C. Montgomery, L. A. Johnson, and J. S. Gardiner, *Forecasting and time series analysis*. McGraw-Hill Companies, 1990.

[2] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with applications*, vol. 37, no. 1, pp. 479–489, 2010.

[3] V. Kodogiannis and A. Lolis, "Forecasting financial time series using neural network and fuzzy system-based techniques," *Neural computing & applications*, vol. 11, no. 2, pp. 90–102, 2002.

[4] W. Huang, Y. Nakamori, and S. Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.

[5] R. S. Mamon and R. J. Elliott, *Hidden Markov models in finance: further developments and applications*. Springer, 2014.

[6] M. R. Hassan and B. Nath, "Stock market forecasting using hidden markov model: a new approach," in *5th International Conference on Intelligent Systems Design and Applications, 2005. ISDA'05, Wroclaw, Poland*. IEEE, September 2005, pp. 192–196.

[7] W. Pieczynski, "Pairwise markov chains," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 634–639, 2003.

[8] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 197–200, 1992.

[9] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise markov chains," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2477–2489, 2004.

[10] G. Lugosi and K. Zeger, "Nonparametric estimation via empirical risk minimization," *IEEE Transactions on information theory*, vol. 41, no. 3, pp. 677–687, 1995.

[11] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[12] J. Lapuyade-Lahorgue and W. Pieczynski, "Unsupervised segmentation of hidden semi-markov non-stationary chains," *Signal Processing*, vol. 92, no. 1, pp. 29 – 42, 2012.