

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/306328558>

Hidden Markov models with threshold effects and their applications to oil price forecasting

Article in *Journal of Industrial and Management Optimization* · August 2016

DOI: 10.3934/jimo.2016045

CITATIONS

0

READS

196

5 authors, including:



Wai-Ki Ching

The University of Hong Kong

198 PUBLICATIONS 1,733 CITATIONS

[SEE PROFILE](#)



Robert J. Elliott

University of Adelaide

586 PUBLICATIONS 8,743 CITATIONS

[SEE PROFILE](#)



Tak Siu

Macquarie University

233 PUBLICATIONS 2,153 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Hawkes Processes in Finance [View project](#)



Quantitative Finance [View project](#)

All content following this page was uploaded by [Robert J. Elliott](#) on 01 September 2016.

The user has requested enhancement of the downloaded file.

HIDDEN MARKOV MODELS WITH THRESHOLD EFFECTS AND THEIR APPLICATIONS TO OIL PRICE FORECASTING

DONG-MEI ZHU

School of Economics and Management, Southeast University
Nanjing, China

WAI-KI CHING

Advanced Modeling and Applied Computing Laboratory
Department of Mathematics
The University of Hong Kong, Pokfulam Road, Hong Kong, China

ROBERT J. ELLIOTT

Centre for Applied Financial Studies, University of South Australia
Adelaide 5001, Australia
Haskayne School of Business, University of Calgary
Canada T3A 6A4

TAK-KUEN SIU

Department of Applied Finance and Actuarial Studies
Faculty of Business and Economics
Macquarie University, Sydney, NSW 2109, Australia

LIANMIN ZHANG*

School of Management and Engineering, Nanjing University
Nanjing, China

(Communicated by Hailiang Yang)

ABSTRACT. In this paper, we propose a Hidden Markov Model (HMM) which incorporates the threshold effect of the observation process. Simulated examples are given to show the accuracy of the estimated model parameters. We also give a detailed implementation of the model by using a dataset of crude oil price in the period 1986-2011. The prediction of crude oil spot price is an important and challenging issue for both government policy makers and industrial investors as most of the world's energy comes from the consumption of crude oil. However, many random events and human factors may lead the crude oil price to a strongly fluctuating and highly non-linear behavior. To capture these properties, we modulate the mean and the variance of log-returns of commodity prices by a finite-state Markov chain. The h -day ahead forecasts generated from our model are compared with regular HMM and the Autoregressive Moving Average model (ARMA). The results indicate that our proposed HMM with threshold effect outperforms the other models in terms of predicting ability.

2010 *Mathematics Subject Classification.* Primary: 90C40, 90C15; Secondary: 60J05.

Key words and phrases. Hidden Markov Model, filtering, threshold effect, oil price, forecasting.

* Corresponding author.

1. **Introduction.** Crude oil price has an important impact on the world economy as a significant proportion of energy consumption around the globe comes from crude oil. The fluctuation of the crude oil price is a major concern of both government policy makers and industrial investors, especially in countries of fast-growing economy such as China and India. There are many factors that may affect the crude oil price, for example wars and conflicts, political stability, economic growth, discovery of new energy sources, financial crises and so on. The combined effect of these factors on crude oil prices render the modeling and forecasting of their movements complex and challenging issues. This will have a significant impact on stakeholders in energy markets since investment decisions related to future fluctuations in crude oil price are hard to make.

The prediction of crude oil price becomes extraordinarily important and many different approaches have been proposed for this purpose. In [16], the authors provide probabilistic forecasts of average oil prices by utilizing inherently probabilistic belief network models. In [30, 31], a regression model was proposed to forecast West Texas Intermediate (WTI) price by considering the petroleum inventory level maintained by the Organization for Economic Cooperation and Development (OECD) countries. Later Kaufman et al. [17] take both OPEC's production policies and OECD's stocks into consideration to forecast oil price. The results indicate that OPEC's decisions are important price drivers. However, it is almost impossible to predict the inventory level of the OECD and the OPEC's decisions. During recent decades time series models have been widely used, including the prediction of oil price. Several authors have tried linear time series models [14, 29] to predict oil prices. However, the results indicate that these models cannot capture the nonlinear behavior of the oil price. Recently De Souza e Silva et al. [2] employed nonlinear time series models and Hidden Markov Models (HMMs) to predict the future oil price trends, where the high frequency price movements have been removed in advance by employing wavelet analysis.

In recent years, HMM filtering methods have found diverse applications in electronics, statistics, physics, finance and commodity markets. Elliott et al. [4] provides self-updating estimates for all model parameters of HMM. In [11], an HMM filtering-based method was applied to the asset allocation problem using a mean-variance type utility criterion. [5, 19] study the application of HMMs in option pricing. In [20], the authors illustrate the optimal filtering of log returns of commodity prices in which both the mean and volatility are modulated by a hidden Markov chain with finite state space. In [21, 22], higher-order Markov-switching models were proposed with applications to crediting rating and risk measurement. An interactive hidden Markov model has been proposed in [1] for modeling default data. Erlwein et al. [13] developed investment strategies relying on HMM approaches. Elliott et al. [7] employed an HMM to the mean-variance portfolio selection problem. Elliott et al. [10] discussed ruin theory where the risk process is described by an HMM.

However, the transition probability matrix in most of the mentioned works is time homogeneous, which has its limitation in empirical applications. In this paper, we consider HMM with zero-delay observations and take a threshold effect on the transition matrices into account. The optimal estimation of parameters will be discussed and compared with actual values by considering simulated examples. The model will then be applied to forecast the crude oil price, where both the mean and variance of the log return prices are modulated by a finite-state hidden Markov chain. Here we model the observable feedback by using the threshold principle

[25, 26, 27]; interested readers can also refer to [6, 8, 9]. The optimal estimate of the threshold parameter will be derived using Akaike's information criterion (AIC) [28].

The remainder of this paper is structured as follows. In Section 2, we introduce the HMM with threshold effect. Section 3 presents a reference probability measure under which the observation process is a sequence of identically independent distributed random variables. In Section 4, we give the recursive formulas for the related processes. Section 5 describes the optimal estimation for all parameters. In Section 6, numerical examples based on simulations are given to test the quality of estimated parameters. Then a numerical example is given to demonstrate both the effectiveness and efficiency of our proposed model. Finally concluding remarks are given in Section 7.

2. The Hidden Markov Model with threshold effect. In this section, we present our Hidden Markov Model with threshold effect (THMM). We begin with a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{P} is a real-world probability measure. Let $\{\mathbf{x}_k\}_{k \geq 0}$ be a discrete-time, finite-state, Markov chain on $(\Omega, \mathcal{F}, \mathcal{P})$ with the state space being given by the set of standard unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$. We suppose that \mathbf{x}_0 or its distribution is known. Suppose $\{y_k\}_{k \geq 0}$ is the observation process on $(\Omega, \mathcal{F}, \mathcal{P})$ with the state space \mathfrak{R} .

To specify the probability law of the chain $\{\mathbf{x}_k\}_{k \geq 0}$ under the ‘‘historical measure’’ \mathcal{P} , we consider a family of transition matrices $\{\mathbf{A}_{\mathbf{y}_k}\}_{k \geq 0}$ where

$$\mathbf{A}_{\mathbf{y}_k} := [a_{ji}(\mathbf{y}_k)]_{i,j=1,2,\dots,N}$$

and

$$a_{ij}(\mathbf{y}_k) := \mathcal{P}(\mathbf{x}_{k+1} = \mathbf{e}_i | \mathbf{x}_k = \mathbf{e}_j, \mathbf{y}_k).$$

Here

$$a_{ij}(\mathbf{y}_k) = \begin{cases} a_{ij}^{(1)} & \text{if } \mathbf{y}_k > r \\ a_{ij}^{(2)} & \text{if } \mathbf{y}_k \leq r, \end{cases}$$

and $r \in \mathbb{R}$ is the threshold parameter. Consequently we are incorporating the threshold effect using the threshold principle introduced by Tong [25, 26, 27] to parametric, nonlinear time series analysis. Indeed, the threshold principle plays a fundamental role in nonlinear time series analysis since it can be used to describe a number of important parametric nonlinear time series models.

Let $\{\mathcal{F}_k\}_{k \geq 0}$ and $\{\mathcal{Y}_k\}_{k \geq 0}$ be the \mathcal{P} -completed, natural filtrations generated by the chain $\{\mathbf{x}_k\}_{k \geq 0}$ and the observations process $\{y_k\}_{k \geq 0}$, respectively. Write, for each $k \geq 0$, $\mathcal{G}_k := \mathcal{F}_k \vee \mathcal{Y}_k$. Define a sequence of random variables $\{\mathbf{v}_k\}_{k \geq 1}$ by putting:

$$\mathbf{v}_k := \mathbf{x}_k - \mathbf{A}_{\mathbf{y}_{k-1}} \mathbf{x}_{k-1}. \quad (1)$$

Then it is not difficult to see that $\{\mathbf{v}_k\}_{k \geq 1}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{G}_k\}_{k \geq 0}$ and the measure \mathcal{P} . That is, for each $k \geq 1$, $E[\mathbf{v}_k | \mathcal{G}_{k-1}] = \mathbf{0} \in \mathfrak{R}^N$, where E denotes the expectation under \mathcal{P} .

Let $\{w_k\}_{k \geq 0}$ be a sequence of independent and identically distributed random variables with common distribution, $N(0, 1)$, the standard normal distribution. We further assume that there are vectors

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)' \quad \text{and} \quad \boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)'$$

such that $\mu(\mathbf{x}_k) = \langle \boldsymbol{\mu}, \mathbf{x}_k \rangle$ and $\sigma(\mathbf{x}_k) = \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle$, where $\mu_i \in \mathfrak{R}$ and $\sigma_i > 0$, for each $i = 1, 2, \dots, N$. Here ‘‘ H' ’’ and ‘‘ \mathbf{v}' ’’ denotes the transpose of the matrix H and the

vector \mathbf{v} respectively, and $\langle \cdot, \cdot \rangle$ is the scalar product in \mathfrak{R}^N . We suppose that the observation process \mathbf{y} is governed by the following dynamics:

$$y_k = \mu(\mathbf{x}_k) + \sigma(\mathbf{x}_k)w_k, \quad k = 0, 1, 2, \dots, T. \quad (2)$$

The above relation implies that what we are considering is a zero-delay model. In [18], a zero-delay setting with different observation process is discussed. Our filtering and estimation method to be discussed below also applies to the case with delay.

3. Reference probability approach. In this section, we discuss a reference probability approach for obtaining an optimal filter of the hidden Markov chain $\{\mathbf{x}_k\}_{k \geq 0}$, given information about the observations process $\{y_k\}_{k \geq 0}$. The key idea of the reference probability approach is to start with a probability measure under which the observation process becomes simpler and does not depend on the hidden Markov chain. The idea of the reference probability approach to filtering was introduced by Zakai [32] and was later applied to filtering hidden Markov models by Elliott et al. [4].

We start with a reference probability measure $\bar{\mathcal{P}}$ on (Ω, \mathcal{F}) under which the observation process $\{y_k\}_{k \geq 0}$ is a sequence of i.i.d. standard Gaussian random variables independent of the hidden Markov chain $\{\mathbf{x}_k\}_{k \geq 0}$ and $\{\mathbf{x}_k\}_{k \geq 0}$ has the probability law specified by $\{\mathbf{A}_{\mathbf{y}_k}\}_{k \geq 1}$. In the sequel, we reconstruct the real-world probability measure \mathcal{P} from the reference probability measure $\bar{\mathcal{P}}$ by a density process for a measure change.

Let $\phi(z)$ be the probability density function of a standard normal random variable. For $l = 0, 1, 2, \dots$, we define

$$\bar{\lambda}_l := \frac{\phi(\sigma(\mathbf{x}_l)^{-1}(\mathbf{y}_l - \mu(\mathbf{x}_l)))}{\sigma(\mathbf{x}_l)\phi(\mathbf{y}_l)}. \quad (3)$$

where the functions μ and σ are the same as presented in Section 2.

$$\begin{cases} \bar{\Lambda}_k & := \prod_{l=1}^k \bar{\lambda}_l, \quad k \geq 1, \\ \bar{\Lambda}_0 & := 1. \end{cases}$$

which is a $\{\mathcal{G}_k\}_{k \geq 0}$ -adapted process and will be utilized in changing probability measure.

Then we reconstruct \mathcal{P} by putting: $\frac{d\mathcal{P}}{d\bar{\mathcal{P}}}|_{\mathcal{G}_k} := \bar{\Lambda}_k$. We define a sequence of random variables $\{w_k\}_{k \geq 1}$ by putting:

$$w_k := \frac{\mathbf{y}_k - \langle \mu, \mathbf{x}_k \rangle}{\langle \sigma, \mathbf{x}_k \rangle}.$$

Lemma 3.1. *Under \mathcal{P} , $\{w_k\}$ is a sequence of i.i.d. random variables with common distribution $N(0, 1)$.*

Proof. See Appendix A. □

Our goal is to derive the optimal filter: $\xi_k := E[\mathbf{x}_k | \mathcal{Y}_k]$. This filter is optimal in the mean-square sense. By a version of the Bayes' rule,

$$E[\mathbf{x}_k | \mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k \mathbf{x}_k | \mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k | \mathcal{Y}_k]}.$$

Write, for each $k \geq 1$, $\gamma_k := \bar{E}[\bar{\Lambda}_k \mathbf{x}_k | \mathcal{Y}_k]$. Then it is easy to check that

$$\xi_k := E[\mathbf{x}_k | \mathcal{Y}_k] = \frac{\gamma_k}{\langle \gamma_k, \mathbf{1} \rangle}.$$

Here $\mathbf{1} := (1, 1, \dots, 1)^T \in \mathfrak{R}^N$.

4. Recursive filters and filter-based estimates. In this section, we first derive the recursive filter γ_k and then some recursive relations for some useful processes. We write

$$\Gamma_i(\mathbf{y}_k) = \frac{\phi\left(\frac{\mathbf{y}_k - \mu(\mathbf{e}_i)}{\sigma(\mathbf{e}_i)}\right)}{\sigma(\mathbf{e}_i)\phi(\mathbf{y}_k)},$$

$$\underline{d}^k = (\Gamma_1(\mathbf{y}_k), \Gamma_2(\mathbf{y}_k), \dots, \Gamma_N(\mathbf{y}_k)) \quad \text{and} \quad D^k = \text{diag}(\underline{d}^k)$$

where $\text{diag}(\mathbf{z})$ denotes the diagonal matrix with vector \mathbf{z} as its diagonal.

Proposition 1. *For $k \in N$, we have*

$$\gamma_{k+1} = D^{k+1} A_{\mathbf{y}_k} \gamma_k.$$

Proof. The proof is given in Appendix B. □

In order to estimate the parameters $A_{\mathbf{y}_k} = (a_{ij}(\mathbf{y}_k))$, $\mu = (\mu_i)$ and $\sigma = \sigma_i$, the following processes are required:

$$\mathcal{J}_k^{rs} := \sum_{n=1}^k \langle \mathbf{x}_{n-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_n, \mathbf{e}_s \rangle$$

the number of jumps from state \mathbf{e}_r to \mathbf{e}_s up to time k ;

$$\mathcal{P}_k^r := \sum_{n=1}^k \langle \mathbf{x}_{n-1}, \mathbf{e}_r \rangle$$

the number of times that \mathbf{x} occupies the state \mathbf{e}_r up to time $k-1$;

$$\mathcal{O}_k^r := \sum_{n=1}^k \langle \mathbf{x}_n, \mathbf{e}_r \rangle$$

the number of times that \mathbf{x} occupies the state \mathbf{e}_r up to time k ;

$$\mathcal{T}_k^r(f) := \sum_{n=1}^k \langle \mathbf{x}_n, \mathbf{e}_r \rangle f(\mathbf{y}_n)$$

where f denotes either $f(\mathbf{y}_k) = \mathbf{y}_k$ or $f(\mathbf{y}_k) = \mathbf{y}_k^2$.

We now derive recursive relations for these processes and write

$$\hat{\mathcal{J}}_k^{rs} = E[\mathcal{J}_k^{rs} | \mathcal{Y}_k].$$

From a version of the Bayes' rule, see for instance (Elliott et al., 1995) [4],

$$\hat{\mathcal{J}}_k^{rs} = \frac{\overline{E}[\overline{\Lambda}_k \mathcal{J}_k^{rs} | \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k | \mathcal{Y}_k]} = \frac{\overline{E}[\overline{\Lambda}_k \mathcal{J}_k^{rs} | \mathcal{Y}_k]}{\langle \gamma_k, \mathbf{1} \rangle}. \quad (4)$$

Write $\gamma(C)_k = \overline{E}[\overline{\Lambda}_k C_k | \mathcal{Y}_k]$ and $\langle \gamma(\mathcal{J}^{rs} \mathbf{x})_k, \mathbf{1} \rangle = \overline{E}[\overline{\Lambda}_k \mathcal{J}_k^{rs} | \mathcal{Y}_k]$. The recursive relations for $\gamma(\mathcal{J}^{rs} \mathbf{x})_k$, $\gamma(\mathcal{P}^r \mathbf{x})_k$, $\gamma(\mathcal{O}^r \mathbf{x})_k$ and $\gamma(\mathcal{T}^r(f) \mathbf{x})_k$ are given in the following proposition.

Proposition 2. *With the above notations, we have*

$$\gamma(\mathcal{J}^{rs} \mathbf{x})_k = D^k A_{\mathbf{y}_{k-1}} \gamma(\mathcal{J}^{rs} \mathbf{x})_{k-1} + \langle \gamma_{k-1}, \mathbf{e}_r \rangle \Gamma_s(\mathbf{y}_k) \mathbf{e}_s a_{sr}(\mathbf{y}_{k-1}) \quad (5)$$

$$\gamma(\mathcal{P}^r \mathbf{x})_k = D^k A_{\mathbf{y}_{k-1}} \gamma(\mathcal{P}^r \mathbf{x})_{k-1} + \langle \gamma_{k-1}, \mathbf{e}_r \rangle D^k a_{.r}(\mathbf{y}_{k-1}) \quad (6)$$

$$\gamma(\mathcal{O}^r \mathbf{x})_k = D^k A_{\mathbf{y}_{k-1}} \gamma(\mathcal{O}^r \mathbf{x})_{k-1} + \Gamma_r(\mathbf{y}_k) \mathbf{e}_r a_{r.}(\mathbf{y}_{k-1}) \gamma_{k-1} \quad (7)$$

$$\gamma(\mathcal{T}^r(f)\mathbf{x})_k = D^k A_{\mathbf{y}_{k-1}} \gamma(\mathcal{T}^r(f)\mathbf{x})_{k-1} + \Gamma_r(\mathbf{y}_k) f(\mathbf{y}_k) \mathbf{e}_r a_{r \cdot}(\mathbf{y}_{k-1}) \gamma_{k-1} \quad (8)$$

where $a_{r \cdot}(\mathbf{y}_{k-1})$ and $a_{\cdot r}(\mathbf{y}_{k-1})$ denote the r th row and r th column of the matrix $A_{\mathbf{y}_{k-1}}$ separately.

Proof. The proof is given in Appendix C. \square

5. Parameter re-estimation. In this section, we present the estimation formulas for all the model parameters. Write

$$\theta = \{(a_{ij}(\mathbf{y})), 1 \leq i, j \leq N, \mu_i, 1 \leq i \leq N, \sigma_i, 1 \leq i \leq N\}$$

for the set of the parameters in our model. Furthermore, we note that

$$\sum_{i=1}^N a_{ij}(\mathbf{y}) = 1.$$

Suppose such a set of parameters θ is given. We then wish to determine a new set

$$\hat{\theta} = \{(\hat{a}_{ij}(\mathbf{y})), 1 \leq i, j \leq N, \hat{\mu}_i, 1 \leq i \leq N, \hat{\sigma}_i, 1 \leq i \leq N\}$$

which maximizes the conditional pseudo log-likelihoods, as in [4].

From [4], if $a_{sr}(\mathbf{y}) = 0$ then

$$\hat{a}_{sr}(\mathbf{y}) = 0 \quad \text{and} \quad \frac{\hat{a}_{sr}(\mathbf{y})}{a_{sr}(\mathbf{y})} = 1;$$

otherwise the optimal estimates are given by

$$\hat{a}_{sr}(\mathbf{y}) = \frac{\hat{\mathcal{J}}_k^{rs}}{\hat{\mathcal{P}}_k^r} = \frac{\gamma(\mathcal{J}^{rs})_k}{\gamma(\mathcal{P}^r)_k}. \quad (9)$$

For the parameters $\mu_i, 1 \leq i \leq N$, we consider a new probability measure \mathcal{P}^* defined by

$$\frac{d\mathcal{P}^*}{d\mathcal{P}} \Big|_{\mathcal{G}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*,$$

where

$$\lambda_l^* = \exp \left(\frac{1}{2 \langle \sigma, \mathbf{x}_l \rangle^2} \{ \langle \mu, \mathbf{x}_l \rangle^2 - \langle \hat{\mu}, \mathbf{x}_l \rangle^2 - 2\mathbf{y}_l \langle \mu, \mathbf{x}_l \rangle + 2\mathbf{y}_l \langle \hat{\mu}, \mathbf{x}_l \rangle \} \right).$$

By maximizing

$$E \left[\log \frac{d\mathcal{P}^*}{d\mathcal{P}} \Big| \mathcal{Y}_k \right]$$

we obtain the optimal estimates

$$\hat{\mu}_r = \frac{\hat{\mathcal{T}}_k^r(\mathbf{y})}{\hat{\mathcal{O}}_k^r} = \frac{\gamma(\hat{\mathcal{T}}^r(\mathbf{y}))_k}{\gamma(\hat{\mathcal{O}}^r)_k}.$$

For the parameters $\sigma_i, 1 \leq i \leq N$, consider the probability measure \mathcal{P}^{**} defined by

$$\frac{d\mathcal{P}^{**}}{d\mathcal{P}} \Big|_{\mathcal{G}_k} = \Lambda_k^{**} = \prod_{l=1}^k \lambda_l^{**},$$

where

$$\lambda_l^{**} = \frac{\langle \sigma, \mathbf{x}_l \rangle}{\langle \hat{\sigma}, \mathbf{x}_l \rangle} \exp \left(\frac{1}{2 \langle \sigma, \mathbf{x}_l \rangle^2} (\mathbf{y}_k - \langle \sigma, \mathbf{x}_l \rangle)^2 - \frac{1}{2 \langle \hat{\sigma}, \mathbf{x}_l \rangle^2} (\mathbf{y}_k - \langle \sigma, \mathbf{x}_l \rangle)^2 \right).$$

Again by maximizing

$$E \left[\log \frac{d\mathcal{P}^{**}}{d\mathcal{P}} \middle| \mathcal{Y}_k \right]$$

we can obtain the optimal estimates

$$\hat{\sigma}_i^2 = \frac{\hat{T}_k^i(\mathbf{y}^2) - 2\mu_i \hat{T}_k^i(\mathbf{y}) + \mu_i^2 \hat{\mathcal{O}}_k^i}{\hat{\mathcal{O}}_k^i} = \frac{\gamma(\hat{T}^i(\mathbf{y}^2))_k - 2\mu_i \gamma(\hat{T}^i(\mathbf{y}))_k + \mu_i^2 \gamma(\hat{\mathcal{O}}^i)_k}{\gamma(\hat{\mathcal{O}}^i)_k}.$$

For the threshold parameter r in the transition probability matrix of the hidden sequence, Akaike's Information Criterion (AIC) approach [27, 28] will be used. In the general case, we have

$AIC(m) = -2 \log(\text{maximized likelihood}) + 2(\text{no. of independently adjusted parameters})$, where $m = 1, 2, \dots, M$, the candidate values for the threshold parameter r . The likelihood function is given by

$$f(\mathbf{y}_1, \dots, \mathbf{y}_t) = f(\mathbf{y}_1) f(\mathbf{y}_2 | \mathbf{y}_1) \dots f(\mathbf{y}_k | \mathcal{Y}_{k-1}) \dots f(\mathbf{y}_t | \mathcal{Y}_{t-1}),$$

where $f(\mathbf{y}_k | \mathcal{Y}_{k-1})$ denotes the conditional probability density function \mathbf{y}_k given \mathcal{Y}_{k-1} . By minimizing AIC , the optimal r is determined.

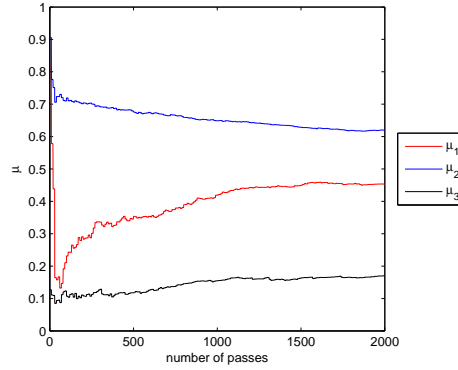
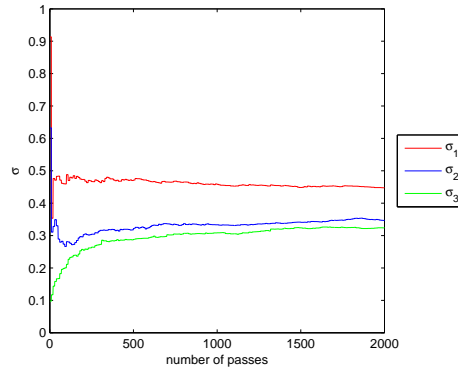
6. Numerical results.

6.1. Simulated examples. In this section, we investigate the quality of the estimated parameters. We start from a "true" model with known parameters and simulate the logarithmic returns from this "true" model. The data will then be used to estimate the parameters using the filtering algorithm. The first thing we need to address is to determine the number of states. Many researchers restrict their attention to a two-state ($N = 2$) or a three-state ($N = 3$). Indeed, in 2005, Taylor [23] pointed out that a two-state model is good enough to distinguish a normal market from one experiencing crisis. In 2009, Erlwein and Mamon, [12] determine the optimal number of regimes by using the Akaike Information Criteria (AIC) and they found that a two-state Markov chain model outperforms other multiple state models. We refer interested readers to [24] for more details about order selection of a Markov chain model. With the development of Internet, the market is changing and becoming more complex, just considering normal market and one with crisis may not be enough. In the following discussions, we set $N = 3$, i.e., we will conduct our research under three state model, which can be seen as considering Expansion, Neutral and Recession markets, respectively.

Under three-state THMM setting, the implementation procedure starts by assigning initial values for $\mu_i, \sigma_i, i = 1, \dots, N$, the transition probability matrices $A^{(1)}, A^{(2)}$ and the threshold parameter r . After a batch of \mathbf{y} , new estimates of $\boldsymbol{\mu}, \boldsymbol{\sigma}, A$ are updated by the above recursive filters, where $A^{(1)}$ is updated by using those data with $\mathbf{y} > r$ and $A^{(2)}$ is updated by using data with $\mathbf{y} \leq r$. After another batch of \mathbf{y} , the threshold parameter r is updated by using the AIC method with new estimates of $\boldsymbol{\mu}, \boldsymbol{\sigma}, A$ and the candidate values for r are given by all the values from -5 to 5 with step size 0.1 . These estimates are then used iteratively to update the model parameters until all the data is used up. It was empirically observed that all the parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}, A^{(1)}$ and $A^{(2)}$, except r , converge to certain values. Furthermore, it appears that this kind of stability does not depend on the initial parameter values. For r , we take the mean of all updated values as the estimate.

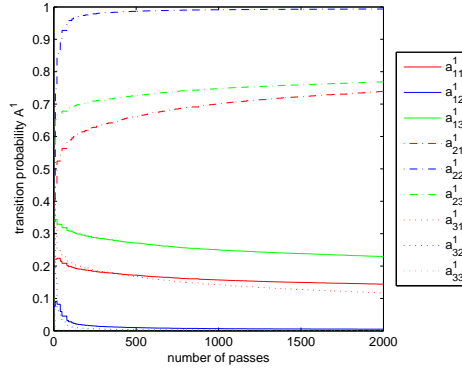
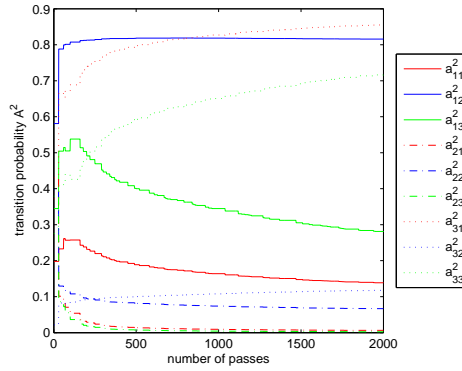
In the following, 50 groups of "true" parameters are given and each group generates 2,000 observable values. To indicate the stable property of the evolution of

μ, σ, A^1 and A^2 , the dynamics of these parameters are shown in Figures 1-4 based on one randomly chosen group of observable values from the simulated data.

FIGURE 1. μ FIGURE 2. σ 

The accuracy of the model parameter estimates under different window sizes is depicted in Table 1, where the differences between “true” parameters and corresponding estimated parameters are measured using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Sum of Squares due to Error (SSE). The number of steps to provide new estimates for parameters μ, σ, A^1 and A^2 is denoted as “It”, which takes values 5, 10 and 15 separately. The corresponding number of steps for updating the parameter r is represented as “It_r”, which is given by 10, 20 and 30 respectively. The data indicates the accuracy of estimation, and it also shows that the estimation tends to be more accuracy by updating more frequently. However, this is not the case for the parameter r . This observation will be taken into consideration in the followings where we will employ our method to predict the oil price by using real data.

6.2. Implementation to a real data set. The estimation method given above was then applied to the log returns series of daily oil prices recorded from 1986 to

FIGURE 3. A^1 FIGURE 4. A^2 

2011. The dataset consisted of 6,524 data values. In the following discussion, we use the data points from 1986 to 2003 as the training data set, and the data points from 2003 to 2011 as the testing data set.

We assume that the drift μ and the volatility σ of the log return are both governed by a hidden Markov chain \mathbf{x} with N states. We observe the logarithmic increments \mathbf{y}_{k+1} of the oil price process S_k , and we suppose it have dynamics

$$\begin{aligned}
 \mathbf{y}_{k+1} &= \log\left(\frac{S_{k+1}}{S_k}\right) \\
 &= \mu(\mathbf{x}_{k+1}) + \sigma(\mathbf{x}_{k+1})w_{k+1} \\
 &= \langle \mu, \mathbf{x}_{k+1} \rangle + \langle \sigma, \mathbf{x}_{k+1} \rangle w_{k+1}.
 \end{aligned} \tag{10}$$

The parameters μ, σ, A and r can be estimated by the same procedure as in the last subsection. In turn they are then used to forecast daily oil prices.

From the representation of \mathbf{x} in Eq. (1), it is easy to see that

$$E(\mathbf{x}_{k+1}|\mathcal{Y}_k) = A_{\mathbf{y}_k} E(\mathbf{x}_k|\mathcal{Y}_k) = A_{\mathbf{y}_k} \xi_k.$$

For $t \in \mathfrak{R}$ consider the following conditional distribution

FIGURE 5. Prediction

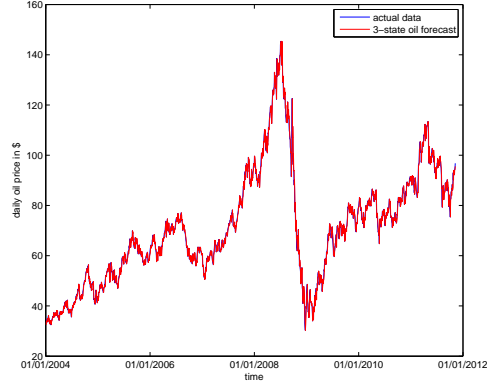
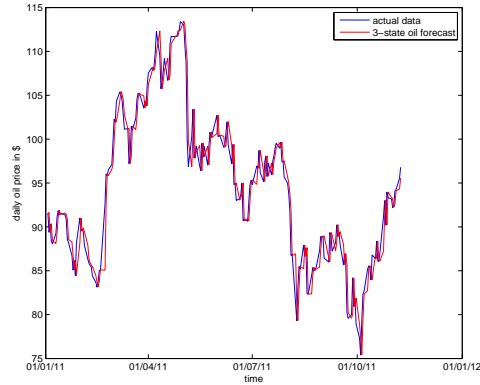


FIGURE 6. Year 2011



$$\begin{aligned}
 \mathcal{P}(\mathbf{y}_{k+1} \leq t | \mathcal{Y}_k) &= \sum_{i=1}^N \mathcal{P}(\sigma_i w_{k+1} \leq t - \mu_i) \mathcal{P}(\mathbf{x}_{k+1} = \mathbf{e}_i | \mathcal{Y}_k) \\
 &= \sum_{i=1}^N \langle A_{\mathbf{y}_k} \xi_k, \mathbf{e}_i \rangle \int_{-\infty}^{t - \mu_i} \phi_i(x) dx
 \end{aligned}$$

where

$$\phi_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{x^2}{2\sigma_i^2}}.$$

Then the density of \mathbf{y}_{k+1} given \mathcal{Y}_k is

$$\sum_{i=1}^N \langle A_{\mathbf{y}_k} \xi_k, \mathbf{e}_i \rangle \phi_i(t - \mu_i).$$

Therefore, the one-step ahead forecast of the asset price based on available information is given by

$$E(S_{k+1} | \mathcal{Y}_k) = S_k E(e^{\mathbf{y}_{k+1}} | \mathcal{Y}_k) = S_k \sum_{i=1}^N \langle A_{\mathbf{y}_k} \xi_k, \mathbf{e}_i \rangle e^{\mu_i + \frac{\sigma_i^2}{2}}. \quad (11)$$

Table 1: Accuracy of estimated parameters

	It=5, It _r =10			It=10, It _r =20			It=15, It _r =30		
	MAE	RMSE	SSE	MAE	RMSE	SSE	MAE	RMSE	SSE
μ_1	0.2105	0.3055	4.6652	0.2373	0.3232	5.2230	0.2583	0.3462	5.9937
μ_2	0.3028	0.3936	7.7456	0.3021	0.3773	7.1170	0.2821	0.3643	6.6375
μ_3	0.2372	0.3330	5.5452	0.2484	0.3258	5.3082	0.2781	0.3745	7.0111
σ_1	0.1641	0.2285	2.6099	0.1804	0.2404	2.8899	0.2321	0.3385	5.7275
σ_2	0.2396	0.3266	5.3347	0.2408	0.3287	5.4033	0.2494	0.3350	5.6124
σ_3	0.2243	0.3114	4.8493	0.2462	0.3250	5.2813	0.2115	0.2788	3.8873
a_{11}^1	0.3038	0.3637	6.6139	0.3768	0.4281	9.1632	0.2863	0.3344	5.5903
a_{12}^1	0.3456	0.4060	8.2436	0.2979	0.3799	7.2170	0.3022	0.3526	6.2180
a_{13}^1	0.2419	0.3059	4.6793	0.3467	0.4280	9.1571	0.3021	0.3642	6.6336
a_{21}^1	0.2338	0.2981	4.4419	0.2488	0.3076	4.7312	0.2556	0.3086	4.7619
a_{22}^1	0.2895	0.3597	6.4692	0.2979	0.3631	6.5927	0.3408	0.4053	8.2130
a_{23}^1	0.2913	0.3469	6.0166	0.3329	0.3895	7.5851	0.3155	0.3786	7.1687
a_{31}^1	0.2140	0.2742	3.7592	0.2687	0.3223	5.1940	0.3125	0.3757	7.0575
a_{32}^1	0.3032	0.3512	6.1654	0.2646	0.3207	5.1414	0.2750	0.3268	5.3387
a_{33}^1	0.2498	0.3233	5.2261	0.2997	0.3510	6.1592	0.3013	0.3629	6.5848
a_{11}^2	0.2499	0.2963	4.3883	0.2597	0.3260	5.3139	0.2878	0.3671	6.7391
a_{12}^2	0.3213	0.3809	7.2547	0.2946	0.3569	6.3691	0.2896	0.3631	6.5922
a_{13}^2	0.2939	0.3453	5.9599	0.2746	0.3301	5.4489	0.3085	0.3638	6.6193
a_{21}^2	0.2613	0.3335	5.5595	0.2789	0.3375	5.6964	0.3076	0.3629	6.5858
a_{22}^2	0.2988	0.3511	6.1621	0.2791	0.3417	5.8363	0.3134	0.3695	6.8273
a_{23}^2	0.3027	0.3638	6.6178	0.3301	0.3938	7.7553	0.3298	0.3947	7.7913
a_{31}^2	0.3093	0.3620	6.5520	0.2088	0.2808	3.9438	0.2558	0.3121	4.8694
a_{32}^2	0.2781	0.3270	5.3472	0.2779	0.3310	5.4777	0.2969	0.3636	6.6100
a_{33}^2	0.3370	0.4104	8.4216	0.3540	0.4139	8.5662	0.3264	0.3925	7.7025
r	1.3506	2.0633	212.8582	0.8303	1.4174	100.4480	0.5571	0.8851	39.1680

The two-step ahead prediction of the price can be obtained by (11) using one-step ahead forecast, and then h -step ($h = 3, \dots, 20$) ahead prediction will be given.

In [20], the predictability performance under the Diebold-Kilian metric of two- and three-state HMMs is compared with that implied by the autoregressive conditional heteroscedasticity (ARCH(1)) and generalized ARCH (GARCH(1,1)) models. Their results indicate that HMMs outperform the ARCH and GARCH models in terms of short-run forecasts. However, they did not mention anything about the Autoregressive Moving Average (ARMA) model, which is supposed to be the most popular model in forecasting application. To show the advantages of our model, ARMA model would be used as a benchmark in the following numerical experiments.

In our numerical experiments, we aim to compare the predictability of the three-state hidden Markov model with threshold parameter (THMM) with the predictability implied by the regular HMM and the benchmark ARMA model. To make the comparison more convincing, we evaluate the forecasting errors for the three different model types using the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Harding and Pagan Test [15] and Diebold-Mariano Test [3]. Harding and Pagan test seeks to identify synchronicity in the turning points of two series, and it is a statistical measure which casts directionally accurate changes. This test is generated by creating two series, X_F for the forecast (or futures) series and X_S for the actual spot price series:

$$\begin{cases} X_{F,t} = 0 & \text{if } F_{t+n|t} - S_t < 0 \\ X_{F,t} = 1 & \text{if } F_{t+n|t} - S_t \geq 0 \\ X_{S,t} = 0 & \text{if } S_{t+n} - S_t < 0 \\ X_{S,t} = 1 & \text{if } S_{t+n} - S_t \geq 0 \end{cases}$$

where F and S are the futures and spot price series, respectively, and n is the forecast horizon. For a given forecast horizon, the concordance statistic is determined

by:

$$C_{S,F} = T^{-1} \left[\sum_{t=1}^T (X_{S,t} X_{F,t}) + \sum_{t=1}^T (1 - X_{S,t})(1 - X_{F,t}) \right] \quad (12)$$

This statistic measures how closely predictions move with actual values in directional terms.

The Diebold-Mariano test is proposed for the comparison of different forecasting models. The basic underlying principle of Diebold-Mariano Test is to base such test on the expectation of the difference between loss functions of the forecasting errors for different models. In the following, let $\{y_t\}$ denote the actual data series, and $\{\hat{y}_{i,t}^h\}$ denote the i^{th} competing h -step forecasting series. The loss differential between two forecasts is given by $d_t = g(e_{it}) - g(e_{jt})$, $i, j = 1, 2, \dots, m$, and the two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t . Suppose that the forecasts are $h(> 1)$ -step-ahead. In order to test the null hypothesis, $H_0 : E(d_t) = 0, \forall t$, the following statistics is utilized by Diebold and Mariano:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}}$$

where $\bar{d} = \sum_{t=1}^T d_t$, and $\hat{f}_d(0)$ is a consistent estimate of

$$f_d(0) = \frac{1}{2\pi} \left(\sum_{k=-\infty}^{\infty} \gamma_d(k) \right)$$

the spectral density of the loss differential at frequency 0, $\gamma_d(k)$ is the autocovariance of the loss differential at lag k . The quantity $\hat{f}_d(0)$ is defined by

$$\hat{f}_d(0) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} I\left(\frac{k}{h-1}\right) \hat{\gamma}_d(k)$$

where

$$\hat{\gamma}_d(k) = \frac{1}{T} \sum_{t=|k|+1}^T (d_t - \bar{d})(d_{t-|k|} - \bar{d})$$

and

$$I\left(\frac{k}{h-1}\right) = \begin{cases} 1 & \text{for } \left| \frac{k}{h-1} \right| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Under the null hypothesis, the test statistics DM is asymptotically $N(0, 1)$ distributed. The null hypothesis of no difference will be rejected if the computed DM statistic falls outside the range of $-Z_{\alpha/2}$ to $Z_{\alpha/2}$.

In our numerical experiment, the number of underlying hidden state is set to be 3 ($N = 3$), and the forecasting errors for the three different model types for the h -step ahead forecasts ($h = 1, 2, \dots, 20$) are reported in Table 2. By analyzing the autocorrelation functions and the partial autocorrelation functions of the price series, the orders of the model ARMA are given by 8 and 12 respectively. Additionally, the results for Harding and Pagan test are reported in Table 3 and the values of DM statistics are presented in Table 4.

Table 2: Accuracy of predictions

Step	MAE			MAPE			RMSE		
	THMM	HMM	ARMA	THMM	HMM	ARMA	THMM	HMM	ARMA
1	0.6571	0.6772	38.5056	70.1171	70.1162	70.6075	1.2772	1.2900	44.6933
2	0.8618	0.8767	38.4588	70.1171	70.1157	70.6071	1.5165	1.5291	44.6933
3	1.0943	1.1075	38.2271	70.1170	70.1152	70.6048	1.7544	1.7672	44.3661
4	1.3337	1.3252	38.5154	70.1191	70.1168	70.6107	2.1103	2.1094	44.4763
5	1.5809	1.5720	38.1649	70.1200	70.1171	70.6049	2.4751	2.4744	44.2075
6	1.7653	1.7512	39.0297	70.1190	70.1157	70.6191	2.6917	2.6858	44.8844
7	1.9091	1.9004	37.9088	70.1182	70.1143	70.6022	2.8747	2.8849	43.8580
8	2.0631	2.0395	36.8175	70.1197	70.1153	70.5854	3.1950	3.1885	42.8849
9	2.3027	2.2690	37.0833	70.1225	70.1176	70.5901	3.4312	3.4062	43.0634
10	2.4098	2.3912	37.0323	70.1221	70.1168	70.5893	3.6031	3.6021	43.0214
11	2.4520	2.4237	38.0120	70.1213	70.1153	70.6041	3.6323	3.6301	43.9126
12	2.7095	2.6502	38.3527	70.1256	70.1192	70.6088	3.9464	3.9046	44.2681
13	2.7674	2.7020	38.2748	70.1234	70.1166	70.6077	3.9701	3.9416	44.1929
14	3.1003	3.0394	38.7028	70.1207	70.1132	70.6143	4.6993	4.7048	44.5676
15	3.3073	3.2407	39.8606	70.1216	70.1138	70.6318	4.8768	4.8855	45.6269
16	3.0475	2.9143	39.8835	70.1256	70.1173	70.6321	4.2991	4.2164	45.6542
17	3.2624	3.2002	38.7757	70.1220	70.1131	70.6147	4.9488	4.9643	44.7068
18	3.8019	3.7099	38.6336	70.1287	70.1195	70.6121	5.3424	5.2526	44.6244
19	3.3387	3.2340	39.8483	70.1258	70.1160	70.6302	4.6491	4.5823	45.7631
20	3.5091	3.4241	39.9252	70.1265	70.1162	70.6313	5.0282	4.9773	45.8393

Table 3: Directional Forecasts
Harding-Pagan Test

Step	THMM	HMM	ARMA
1	0.3475	0.5111	0.4737
2	0.2158	0.2786	0.4569
3	0.1403	0.1494	0.4443
4	0.1479	0.1515	0.4559
5	0.1378	0.1337	0.4382
6	0.1474	0.1489	0.4321
7	0.1312	0.1302	0.4311
8	0.1398	0.1383	0.4265
9	0.1398	0.1393	0.4235
10	0.1530	0.1570	0.4281
11	0.1555	0.1525	0.4326
12	0.1520	0.1459	0.4265
13	0.1575	0.1494	0.4306
14	0.1611	0.1631	0.4250
15	0.1601	0.1575	0.4250
16	0.1702	0.1636	0.4235
17	0.1499	0.1525	0.4159
18	0.1738	0.1651	0.4169
19	0.1651	0.1520	0.4139
20	0.1550	0.1550	0.4144

From Table 2, by comparing the forecasting errors of the THMM with those of HMM and ARMA models, we find that THMM and HMM models dominate ARMA model in terms of the three criteria for all the h -step ahead forecasts ($h = 1, 2, \dots, 20$). Additionally, these criteria also indicate that THMM performs better than the HMM model for short-run forecasts. Nevertheless, the difference of long-run forecasting errors is too small to yield any practical significance, which can be justified by corresponding values of DM statistics from Table 4. On the other hand, from Table 3, we conclude that ARMA model shows the best directional forecasting ability. However, THMM outperforms HMM in directional forecasts for long-run forecasts. By calculating the Diebold-Mariano statistics, the null hypothesis of no difference is rejected at the 5% level for THMM with ARMA model. For THMM with HMM model, the null hypothesis is rejected at the 5% level for short-run

Table 4: The Diebold-Mariano Test

Step	DM Test based on THMM and HMM	DM Test based on THMM and ARMA
1	-7.3944	-39.8539
2	-3.5605	-23.0956
3	-1.9302	-18.0701
4	0.0952	-15.6151
5	0.0506	-13.6254
6	0.3246	-12.6574
7	-0.4322	-11.4710
8	0.2511	-10.4746
9	0.8516	-9.9699
10	0.0270	-9.4191
11	0.0532	-9.0919
12	0.9990	-8.6963
13	0.5196	-8.3548
14	-0.0829	-8.0797
15	-0.1185	-7.9669
16	1.2755	-7.6941
17	-0.1728	-7.2728
18	1.1428	-6.9931
19	0.8583	-6.9341
20	0.5422	-6.7681

forecasts. These results reinforce further the adequacy of the THMM model in capturing the dynamics of the data set.

A comparison of one-step ahead forecasts generated from our model with the actual data is depicted in Figure 5 and the zoom-in view of year 2011 is given in Figure 6. The two figures shows directly how closely the one-step ahead forecasts follow the actual data.

7. Concluding remarks. In this paper, we have added a threshold parameter to a hidden Markov model for oil prices and considered the zero-delay case. The optimal estimators of the model parameters are given and compared with corresponding true values. The prediction results show good forecasting ability of our model. For future research, it will be interesting to extend this model to a higher-order case, where the hidden sequence follows a discrete-time higher-order Markov chain. The embedding technique discussed in [21] can be employed to analyze such a higher-order HMM. The model introduced in this article can also be applied to other research areas, such as optimal portfolio management, option valuation and risk measurement.

Appendix. *Proof of Lemma 1.* Firstly, we note that

$$\mathcal{P}(w_{k+1} \leq t | \mathcal{G}_k) = E[I_{\{w_{k+1} \leq t\}} | \mathcal{G}_k]$$

By a version of the Bayes' rule,

$$\begin{aligned} E[I_{\{w_{k+1} \leq t\}} | \mathcal{G}_k] &= \frac{\overline{E}[\overline{\Lambda}_{k+1} I_{\{w_{k+1} \leq t\}} | \mathcal{G}_k]}{\overline{E}[\overline{\Lambda}_{k+1} | \mathcal{G}_k]} \\ &= \frac{\overline{\Lambda}_k \overline{E}[\overline{\lambda}_{k+1} I_{\{w_{k+1} \leq t\}} | \mathcal{G}_k]}{\overline{\Lambda}_k \overline{E}[\overline{\lambda}_{k+1} | \mathcal{G}_k]} \\ &= \frac{\overline{E}[\overline{\lambda}_{k+1} I_{\{w_{k+1} \leq t\}} | \mathcal{G}_k]}{\overline{E}[\overline{\lambda}_{k+1} | \mathcal{G}_k]}. \end{aligned}$$

Now

$$\begin{aligned}
\bar{E}[\bar{\lambda}_{k+1}|\mathcal{G}_k] &= \bar{E}\left[\frac{\phi(w_{k+1})}{\sigma(\mathbf{x}_{k+1})\phi(\mathbf{y}_{k+1})}\middle|\mathcal{G}_k\right] \\
&= \bar{E}\left[\bar{E}\left[\frac{\phi(w_{k+1})}{\sigma(\mathbf{x}_{k+1})\phi(\mathbf{y}_{k+1})}\middle|\mathcal{G}_k, \mathcal{F}_{k+1}\right]\middle|\mathcal{G}_k\right] \\
&= \bar{E}\left[\int_{-\infty}^{+\infty}\frac{\phi(w_{k+1})}{\sigma(\mathbf{x}_{k+1})\phi(\mathbf{y}_{k+1})}\phi(\mathbf{y}_{k+1})d\mathbf{y}_{k+1}\middle|\mathcal{G}_k\right] \\
&= 1,
\end{aligned}$$

so

$$\begin{aligned}
\mathcal{P}(w_{k+1} \leq t|\mathcal{G}_k) &= \bar{E}[\bar{\lambda}_{k+1}I_{\{w_{k+1} \leq t\}}|\mathcal{G}_k] \\
&= \bar{E}\left[\bar{E}[\bar{\lambda}_{k+1}I_{\{w_{k+1} \leq t\}}|\mathcal{G}_k, \mathcal{F}_{k+1}]\middle|\mathcal{G}_k\right] \\
&= \bar{E}\left[\int_{-\infty}^{+\infty}\frac{\phi(w_{k+1})}{\sigma(\mathbf{x}_{k+1})\phi(\mathbf{y}_{k+1})}\phi(\mathbf{y}_{k+1})I_{\{w_{k+1} \leq t\}}d\mathbf{y}_{k+1}\middle|\mathcal{G}_k\right] \\
&= \int_{-\infty}^t\frac{1}{\sqrt{2\pi}}e^{-\frac{w_{k+1}^2}{2}}dw_{k+1} \\
&= \mathcal{P}(w_{k+1} \leq t).
\end{aligned}$$

The result follows. \square

Proof of Proposition 1. With the notations, we have

$$\begin{aligned}
\gamma_{k+1} &= \bar{E}[\bar{\Lambda}_k \lambda_{k+1} X_{k+1} | \mathcal{Y}_{k+1}] \\
&= \sum_{i=1}^N \bar{E}[\bar{\Lambda}_k \langle X_{k+1}, \mathbf{e}_i \rangle | \mathcal{Y}_{k+1}] \Gamma_i(\mathbf{y}_{k+1}) \mathbf{e}_i \\
&= \bar{E}[\bar{\Lambda}_k \langle A_{\mathbf{y}_k} X_k + \mathbf{v}_{k+1}, \mathbf{e}_i \rangle | \mathcal{Y}_{k+1}] \Gamma_i(\mathbf{y}_{k+1}) \mathbf{e}_i \\
&= \sum_{i=1}^N \sum_{j=1}^N \bar{E}[\bar{\Lambda}_k \langle X_k, \mathbf{e}_j \rangle | \mathcal{Y}_{k+1}] \Gamma_i(\mathbf{y}_{k+1}) \mathbf{e}_i a_{ij}(\mathbf{y}_k) \\
&= \sum_{i=1}^N \sum_{j=1}^N \langle \gamma_k, \mathbf{e}_j \rangle \Gamma_i(\mathbf{y}_{k+1}) \mathbf{e}_i a_{ij}(\mathbf{y}_k) = D^{k+1} A_{\mathbf{y}_k} \gamma_k.
\end{aligned}$$

\square

Proof of Proposition 2. With the given notation,

$$\begin{aligned}
\gamma(\mathcal{J}^{rs} \mathbf{x})_k &= \bar{E}[\bar{\Lambda}_k \mathcal{J}_k^{rs} \mathbf{x}_k | \mathcal{Y}_k] = \bar{E}[\bar{\Lambda}_k (\mathcal{J}_{k-1}^{rs} + \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_k, \mathbf{e}_s \rangle) \mathbf{x}_k | \mathcal{Y}_k] \\
&= \bar{E}[\bar{\Lambda}_k \mathcal{J}_{k-1}^{rs} \mathbf{x}_k | \mathcal{Y}_k] + \bar{E}[\bar{\Lambda}_k \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_k, \mathbf{e}_s \rangle \mathbf{x}_k | \mathcal{Y}_k] \\
&= \sum_{i=1}^N \bar{E}[\bar{\Lambda}_{k-1} \bar{\lambda}_k \mathcal{J}_{k-1}^{rs} \mathbf{x}_k \langle \mathbf{x}_k, \mathbf{e}_i \rangle | \mathcal{Y}_k] \\
&\quad + \bar{E}[\bar{\Lambda}_{k-1} \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_k, \mathbf{e}_s \rangle | \mathcal{Y}_k] \frac{\phi(\frac{\mathbf{y}_k - \mu_s}{\sigma_s})}{\sigma_s \phi(\mathbf{y}_k)} \mathbf{e}_s
\end{aligned}$$

$$\begin{aligned}
\gamma(\mathcal{J}^{rs}\mathbf{x})_k &= \sum_{i=1}^N \bar{E}[\bar{\Lambda}_{k-1} \mathcal{J}_{k-1}^{rs} \langle \mathbf{x}_k, \mathbf{e}_i \rangle | \mathcal{Y}_k] \Gamma_i(\mathbf{y}_k) \mathbf{e}_i \\
&\quad + \bar{E}[\bar{\Lambda}_{k-1} \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_k, \mathbf{e}_s \rangle | \mathcal{Y}_k] \Gamma_s(\mathbf{y}_k) \mathbf{e}_s \\
&= \sum_{i=1}^N \sum_{j=1}^N \bar{E}[\bar{\Lambda}_{k-1} \mathcal{J}_{k-1}^{rs} \langle A_{\mathbf{y}_{k-1}} \mathbf{x}_{k-1} + v_k, \mathbf{e}_i \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_j \rangle | \mathcal{Y}_k] \Gamma_i(\mathbf{y}_k) \mathbf{e}_i \\
&\quad + \bar{E}[\bar{\Lambda}_{k-1} \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle A_{\mathbf{y}_{k-1}} \mathbf{x}_{k-1} + v_k, \mathbf{e}_s \rangle | \mathcal{Y}_k] \Gamma_s(\mathbf{y}_k) \mathbf{e}_s
\end{aligned}$$

then

$$\gamma(\mathcal{J}^{rs}\mathbf{x})_k = \sum_{i=1}^N \sum_{j=1}^N \langle \gamma(\mathcal{J}^{rs}\mathbf{x})_{k-1}, \mathbf{e}_j \rangle \Gamma_i(\mathbf{y}_k) \mathbf{e}_i a_{ij}(\mathbf{y}_{k-1}) + \langle \gamma_{k-1}, \mathbf{e}_r \rangle \Gamma_s(\mathbf{y}_k) \mathbf{e}_s a_{sr}(\mathbf{y}_{k-1})$$

so

$$\gamma(\mathcal{J}^{rs}\mathbf{x})_k = D^k A_{\mathbf{y}_{k-1}} \gamma(\mathcal{J}^{rs}\mathbf{x})_{k-1} + \langle \gamma_{k-1}, \mathbf{e}_r \rangle \Gamma_s(\mathbf{y}_k) \mathbf{e}_s a_{sr}(\mathbf{y}_{k-1}).$$

The other two recursive relations can be obtained similarly. See Elliott et al. (1995)[4]. \square

Acknowledgments. This research work was supported by Research Grants Council of Hong Kong under Grant Number 17301214, HKU CERG Grants and Hung Hing Ying Physical Research Grant, National Natural Science Foundation of China Under Grant number 71501093, the Basic Research Foundation (Natural Science) of Jiangsu Province with Grant number BK20150566.

REFERENCES

- [1] W. K. Ching, T. K. Siu, L. M. Li, T. Li and W. K. Li, [Modeling default data via an interactive hidden Markov model](#), *Computational Economics*, **34** (2009), 1–19.
- [2] E. G. De Souza e Silva, L. Legey and E. A. De Souza e Silva, Forecasting oil price trends using wavelet and hidden Markov models, *Energy Economics*, **32** (2010), 1507–1519.
- [3] F. X. Diebold and R. Mariano, Comparing predictive accuracy, *Journal of Business and Economic Statistics*, **13** (1995), 253–265.
- [4] R. J. Elliott, L. Aggoun and J. Moore, *Hidden Markov Models: Estimation and Control*, Springer, New York, 1995.
- [5] R. J. Elliott, L. L. Chan and T. K. Siu, [Option pricing and Esscher transform under regime switching](#), *Annals of Finance*, **1** (2005), 423–432.
- [6] R. J. Elliott, C. C. Liew and T. K. Siu, [On filtering and estimation of a threshold stochastic volatility model](#), *Applied Mathematics and Computation*, **218** (2011), 61–75.
- [7] R. J. Elliott, T. K. Siu and A. Badescu, [On mean-variance portfolio selection under a hidden Markovian regime-switching model](#), *Economic Modelling*, **27** (2010), 678–686.
- [8] R. J. Elliott, T. K. Siu and A. Badescu, [On pricing and hedging options in regime-switching models with feedback effect](#), *Journal of Economic Dynamics and Control*, **35** (2011), 694–713.
- [9] R. J. Elliott, T. K. Siu and J. W. Lau, [Filtering a double threshold model with regime switching](#), *IEEE Transactions on Automatic Control*, **58** (2013), 3185–3190.
- [10] R. J. Elliott, T. K. Siu and H. L. Yang, [Ruin theory in a hidden Markov-modulated risk model](#), *Stochastic Models*, **27** (2011), 474–489.
- [11] R. J. Elliott and J. van der Hoek, [An application of hidden Markov models to asset allocation problems](#), *Finance and Stochastic*, **1** (1997), 229–238.
- [12] C. Erlwein and R. Mamon, [An online estimation scheme for a Hull and White model with HMM-driven parameters](#), *Statistical Methods and Applications*, **18** (2009), 87–107.
- [13] C. Erlwein, R. Mamon and M. Davison, [An examination of HMM-based investment strategies for asset allocation](#), *Applied Stochastic Models in Business and Industry*, **27** (2011), 204–221.
- [14] G. Frey, M. Manera, A. Markandya and E. Scarpa, Econometric models for oil price forecasting, *A Critical Survey CESinfo Forum*, **10** (2009), 29–44.
- [15] D. Harding and A. R. Pagan, *Synchronisation of Cycles*, mimeo., Australian National University.

- [16] H. G. Huntington, [Oil price forecasting in the 1980s: What went wrong?](#) *The Energy Journal*, **15** (1994), 1–22.
- [17] R. K. Kaufman, S. Dees, P. Karadeloglou and M. Sanchez, [Does OPEC matter? An econometric analysis of oil prices](#), *The Energy Journal*, **25** (2004), 67–90.
- [18] M. W. Korolkiewica and R. J. Elliott, [Smoothed parameter estimation for a Markov model of credit quality](#), *Hidden Markov Models in Finance*, Springer, New York, **104** (2007), 69–90.
- [19] C. C. Liew and T. K. Siu, [A hidden Markov regime-switching model for option valuation](#), *Insurance: Mathematics and Economics*, **47** (2010), 374–384.
- [20] R. S. Mamon, C. Erlwein and R. B. Gopaluni, [Adaptive signal processing of asset price dynamics with predictability analysis](#), *Information Sciences*, **178** (2008), 203–219.
- [21] T. K. Siu, W. K. Ching, E. Fung, M. Ng and X. Li, [A high-order Markov-switching model for risk measurement](#), *Computers and Mathematics with Applications*, **58** (2009), 1–10.
- [22] T. K. Siu, W. K. Ching, E. S. Fung and M. K. Ng, [Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models](#), *Computational Economics*, **26** (2005), 251–284.
- [23] S. J. Taylor, *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, Princeton, 2005.
- [24] H. Tong, [Determination of the order of a Markov chain by Akaike's information criterion](#), *Journal of applied probability*, **12** (1975), 488–497.
- [25] H. Tong, [On a threshold model](#), *Pattern Recognition and signal processing, NATO ASI Series E: Applied Sc. No. 29, ed. C.H.Chen. The Netherlands: Sijthoff & Noordhoff*, (1978), 575–586.
- [26] H. Tong, [A view on non-linear time series model building](#), *Time Series, ed. O. D. Anderson*, Amsterdam: North-Holland, 1980, 41–56.
- [27] H. Tong, *Threshold Models in Non-linear Time Series Analysis*, Lecture Notes in Statistics, No. 21, Springer-Verlag, New York, 1983.
- [28] H. Tong, *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford, 1990.
- [29] W. Xie, L. Yu, L. Xu and S. Wang, [A new method for crude oil price forecasting based on support vector machines](#), *International Conference on Computational Science, (Part IV) 3994* (2006), 444–451.
- [30] M. Ye, J. Zyren and J. Shore, [Forecasting crude oil spot price using OECD petroleum inventory levels](#), *International Advances in Economic Research*, **8** (2002), 324–333.
- [31] M. Ye, J. Zyren and J. Shore, [A monthly crude oil spot price forecasting model using relative inventories](#), *International Journal of Forecasting*, **21** (2005), 491–501.
- [32] M. Zakai, [On the optimal filtering of diffusion processes](#), *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **11** (1969), 230–243.

Received September 2015; revised June 2016.

E-mail address: dongmeizhu86@gmail.com

E-mail address: wching@hku.hk

E-mail address: relliott@ucalgary.ca

E-mail address: ktksiu2005@gmail.com

E-mail address: zhanglm@nju.edu.cn