IMA Journal of Management Mathematics (2007) **18**, 85−97
doi:10.1093/imaman/dpl014
Advance Access publication on November 15, 2006

Interactive hidden Markov models and their applications

W. K. CHING†
The Advanced Modeling and Applied Computing Laboratory and Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong

E. FUNG AND M. NG
Department of Mathematics, The Hong Kong Baptist University, Kowloon Tong, Hong Kong

T. K. SIU
Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

AND

W. K. LI
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

[Received on 21 March 2006; accepted on 20 September 2006]

In this paper, we propose an Interactive hidden Markov model (IHMM). In a traditional HMM, the observable states are affected directly by the hidden states, but not vice versa. In the proposed IHMM, the transitions of hidden states depend on the observable states. We also develop an efficient estimation method for the model parameters. Numerical examples on the sales demand data and economic data are given to demonstrate the applicability of the model.

Keywords: hidden Markov model; categorical time series; steady-state probability distribution; prediction of demand.

1. Introduction

In the world of data and information, data sequences (or time series) occur frequently in many applications. To analyse a data sequence, it is of practical importance to select an appropriate model for the data. Numerical data sequences have been well studied in Brockwell & Davis (1991). Mathematical tools such as Fourier transform and spectral analysis are employed frequently in the analysis of numerical data sequences. For categorical data sequences, there are many situations that one would like to employ Markov models as a mathematical tool, see for instance in Raftery (1985), Li & Kwok (1990), Aggoun & Benkherouf (2003), MacDonald & Zucchini (1999) and Ching *et al.* (2004c). A number of applications such as inventory control, customer classification, marketing, bioinformatics, economics and finance can be found in the literature Ching *et al.* (2002, 2004a,b,c), Ching & Ng (2006), Siu *et al.* (2005a,b), and Waterman (1995). Take, e.g. in the sales demand prediction, products are classified into several states such as, very high sales volume, high sales volume, low sales volume and very

†Email: wching@hkusua.hku.hk

low sales volume. In these applications and many others, one would like to (i) characterize categorical data sequences for the purpose of comparison and classification process or (ii) model categorical data sequences, and, hence to make predictions in the control and planning processes. It has been shown that Markov models can be a promising approach for these purposes.

A well-known class of models is the Hidden Markov model (HMM). HMMs have been widely adopted by practitioners in various fields, for instance, speech recognition (MacDonald & Zucchini, 1999) and bioinformatics (Waterman, 1995). A tutorial paper providing a comprehensive discussion on the model can be found in Rabiner (1989). The monograph by Elliott *et al.* (1994) provides a comprehensive discussion on HMMs. Higher order HMMs have also been proposed in Siu *et al.* (2005a). In a traditional HMM, the observable states are affected directly by the hidden states, but not vice versa. Here, we propose a HMM such that the transitions of hidden states depend on the observable states. Our model can be related to a discrete-time version of the class of the first-order self-exciting threshold auto-regressive (SETAR) models first proposed by Tong (1977, 1978, 1983) for modelling nonlinear time series taking numerical values. In particular, when the observable state can determine the hidden state with probability one, our model can be considered a discrete-state analogy of the class of the first-order SETAR models. The monograph by Tong (1990) provides an excellent and original discussion of the SETAR models and other important nonlinear time series models.

Much of the literature focus on the modelling of continuous-state nonlinear time series. There is a relatively little work on modelling the nonlinear behaviour of categorical time series. In the continuous-state case, the idea of threshold autoregressive model is to provide a piecewise linear approximation to a nonlinear autoregressive time series model by dividing the state space into several regimens via the threshold principle. Here, we also provide a first-order approximation of the nonlinear behaviour of categorical time series by dividing the state-space of the Markov chain process into several regimes, say two regimes.

The rest of the paper is organized as follows, In Section 2, we present the idea of the interactive hidden Markov model (IHMM) through an example. We then propose an estimation method for the model parameters required in our model. In Section 3, we give the general IHMM. In Section 4, numerical examples on the sales demand data and economic data are given to demonstrate the applicability of the model. Finally, concluding remarks are given in Section 5.

2. The IHMM

We present an IHMM for modelling categorical sales volumes, where the transitions of hidden states depend on the current observable state through a numerical example. We extend the results to give a general model in Section 3.

2.1 *An example*

Suppose we are given a categorical data sequence of six possible sales volumes (1, 2, 3, 4, 5, 6) in steady state as follows:

$$1, 2, 1, 2, 1, 2, 2, 4, 2, 5, 6, 2, 1, \ldots.$$

Here,

$$1 = \text{very high},\ 2 = \text{high},\ 3 = \text{moderate high},\ 4 = \text{moderate low},\ 5 = \text{low},\ 6 = \text{very low}.$$

They are the observable states. Suppose further that there are two hidden states: good economic situation (A) and bad economic situation (B). In the good economic situation, the probability distribution of the sales volume is assumed to follow the distribution:

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0\right).$$

While in the bad economic situation, the probability distribution of the sales volume is assumed to follow the distribution:

$$\left(0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

In our model, we assume that the economic situation is unobservable, but the sales volume (observable state) can infer the economic situation.

Here, we aim at modelling the dynamics of the sales demand data sequences by a Markov chain.

In the Markov chain, the states are A, B, 1, 2, 3, 4, 5 and 6. We assume that when the observable state is i, the probabilities that the hidden state is A and B in next time step are given by α_i and $1-\alpha_i$, respectively. The transition probability matrix governing the Markov chain is then given by the following matrix:

$$P_2 = \left(\begin{array}{cc|cccccc}
0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\hline
\alpha_1 & 1-\alpha_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_2 & 1-\alpha_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_3 & 1-\alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_4 & 1-\alpha_4 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_5 & 1-\alpha_5 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_6 & 1-\alpha_6 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

2.2 *Estimation of parameters*

In order to define the IHMM, one has to estimate $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$ from an observed data sequence. We first consider the two-step transition probability matrix:

$$P_2^2 = \left(\begin{array}{cc|cccccc}
\frac{\alpha_1+\alpha_2+\alpha_3+\alpha_4}{4} & 1-\frac{\alpha_1+\alpha_2+\alpha_3+\alpha_4}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{\alpha_3+\alpha_4+\alpha_5+\alpha_6}{4} & 1-\frac{\alpha_3+\alpha_4+\alpha_5+\alpha_6}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & \frac{\alpha_1}{4} & \frac{\alpha_1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_1}{4} & \frac{1}{4}-\frac{\alpha_1}{4} \\
0 & 0 & \frac{\alpha_2}{4} & \frac{\alpha_2}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_2}{4} & \frac{1}{4}-\frac{\alpha_2}{4} \\
0 & 0 & \frac{\alpha_3}{4} & \frac{\alpha_3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_3}{4} & \frac{1}{4}-\frac{\alpha_3}{4} \\
0 & 0 & \frac{\alpha_4}{4} & \frac{\alpha_4}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_4}{4} & \frac{1}{4}-\frac{\alpha_4}{4} \\
0 & 0 & \frac{\alpha_5}{4} & \frac{\alpha_5}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_5}{4} & \frac{1}{4}-\frac{\alpha_5}{4} \\
0 & 0 & \frac{\alpha_6}{4} & \frac{\alpha_6}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_6}{4} & \frac{1}{4}-\frac{\alpha_6}{4}
\end{array}\right).$$

We then extract the one-step transition probability matrix of the observable states from P_2^2 as follows:

$$\tilde{P}_2 = \begin{pmatrix} \frac{\alpha_1}{4} & \frac{\alpha_1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_1}{4} & \frac{1}{4}-\frac{\alpha_1}{4} \\ \frac{\alpha_2}{4} & \frac{\alpha_2}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_2}{4} & \frac{1}{4}-\frac{\alpha_2}{4} \\ \frac{\alpha_3}{4} & \frac{\alpha_3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_3}{4} & \frac{1}{4}-\frac{\alpha_3}{4} \\ \frac{\alpha_4}{4} & \frac{\alpha_4}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_4}{4} & \frac{1}{4}-\frac{\alpha_4}{4} \\ \frac{\alpha_5}{4} & \frac{\alpha_5}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_5}{4} & \frac{1}{4}-\frac{\alpha_5}{4} \\ \frac{\alpha_6}{4} & \frac{\alpha_6}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\frac{\alpha_6}{4} & \frac{1}{4}-\frac{\alpha_6}{4} \end{pmatrix}.$$

The advantage of looking at the matrix $\tilde{P}_2$ is that it gives the information of the one-step transition from one observable state to another observable state. Even though, in this case, we do not have a closed-form solution for the stationary distribution of the process. There are six parameters to be estimated. To estimate the parameter α_i, we first estimate the one-step transition probability matrix from the observed sequence. This can be done by counting the transition frequency of the states in the observed sequence as in Ching *et al.* (2002, 2004a,b,c), Ching & Ng (2006) and MacDonald & Zucchini (1999). Suppose the estimate for this example is given by $\hat{P}_2$. We expect $\tilde{P}_2 \approx \hat{P}_2$ and, hence, α_i can be obtained by solving the following minimization problem:

$$\min_{\alpha_i} \|\tilde{P}_2 - \hat{P}_2\|_F^2$$

subject to:

$$0 \leqslant \alpha_i \leqslant 1.$$

Here, $\|\cdot\|_F$ is the Frobenius norm, i.e.

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{i=1}^{n} A_{ij}^2.$$

We remark that other matrix norms can also be used as the objective function.

3. The general IHMM

In this section, we present our method to a general IHMM for the case of m hidden states and n observable states. First, by following Elliott *et al.* (1994), we shall present the model dynamics for the general IHMM by making use of the canonical representation of the state spaces of the hidden and observable Markov chain processes. Then, we shall demonstrate the least squares method for estimating the general IHMM.

3.1 *Model dynamics*

Fix a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{P}$ is a real-world probability. Let $\mathcal{T}$ denote the time index set $\{0, 1, 2, \ldots\}$ of our model. Suppose $X := \{X_t\}_{t\in\mathcal{T}}$ and $Y := \{Y_t\}_{t\in\mathcal{T}}$ denote a discrete-time hidden Markov chain and a discrete-time observable Markov chain on $(\Omega, \mathcal{F}, \mathcal{P})$ with space

$\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots, y_n\}$, respectively. We suppose that X_0 is known or its distribution is given. Define the following transition probabilities:

$$p_{ij} := \mathcal{P}(Y_t = y_j | X_t = x_i), \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n, \tag{1}$$

and

$$\alpha_{jk} := \mathcal{P}(X_{t+1} = x_k | Y_t = y_j), \quad j = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, m. \tag{2}$$

Then, we define the following transition probability matrices:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix} \tag{3}$$

and

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nm} \end{pmatrix} \tag{4}$$

For each $t \in \mathcal{T}$, let $\mathcal{F}_t^X$ and $\mathcal{F}_t^Y$ denote the information sets generated by the values of $\{X_0, X_1, \ldots, X_t\}$ and $\{Y_0, Y_1, \ldots, Y_t\}$, respectively. Since X and Y are hidden and observable sequence here, $\mathcal{F}_t^X$ and $\mathcal{F}_t^Y$ represent hidden and observable information sets, respectively. Then, we represent the general IHMM using a standard filtering formulation.[1] By the Bayes' rule,

$$\begin{aligned} \mathcal{P}(X_{t+1} = x_i | \mathcal{F}_{t+1}^Y) &= \frac{\mathcal{P}(Y_{t+1} | X_{t+1} = x_i) P(X_{t+1} = x_i | \mathcal{F}_t^Y)}{\sum_{i=1}^m \mathcal{P}(Y_{t+1} | X_{t+1} = x_i) P(X_{t+1} = x_i | \mathcal{F}_t^Y)} \\ &= \frac{\mathcal{P}(Y_{t+1} | X_{t+1} = x_i) P(X_{t+1} = x_i | Y_t)}{\sum_{i=1}^m \mathcal{P}(Y_{t+1} | X_{t+1} = x_i) P(X_{t+1} = x_i | Y_t)}. \end{aligned} \tag{5}$$

Then,

$$\begin{aligned} \mathcal{P}(Y_{t+1} | \mathcal{F}_t^Y) &= \sum_{i=1}^m \mathcal{P}(Y_{t+1} | X_{t+1} = x_i) \mathcal{P}(X_{t+1} = x_i | \mathcal{F}_t^Y) \\ &= \sum_{i=1}^m \mathcal{P}(Y_{t+1} | X_{t+1} = x_i) \mathcal{P}(X_{t+1} = x_i | Y_t). \end{aligned} \tag{6}$$

[1] We would like to thank an anonymous referee for pointing out this.

Equation (6) provides a justification for the equation

$$[\tilde{P}_2]_{ij} = \sum_{k=1}^{m} \alpha_{ik} p_{kj}.$$

It can also be used for parameter estimation.

3.2 *Estimation method*

Now, we shall present the least squares method for the estimation of the general IHMM.

Case I: P is known.

Using the same trick as in the example of Section 2, the one-step transition probability matrix of the observable states is given by:

$$\tilde{P}_2 = \alpha P$$

(P is called the emission matrix), i.e.

$$[\tilde{P}_2]_{ij} = \sum_{k=1}^{m} \alpha_{ik} p_{kj} \quad i, j = 1, 2, \ldots, n.$$

Here, we assume that α_{ij} are unknown and that the probabilities p_{ij} are given. Suppose $[Q]_{ij}$ is the one-step transition probability matrix estimated from the observed sequence. Then, for each fixed i, α_{ij}, $j = 1, 2, \ldots, m$, can be obtained by solving the following constrained least square problem:

$$\min_{\alpha_{ik}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik} p_{kj} - [Q]_{ij} \right)^2 \right\}$$

subject to

$$\sum_{k=1}^{m} \alpha_{ik} = 1 \quad \text{and} \quad \alpha_{ik} \geqslant 0.$$

Case II: P is unknown.

Suppose all the probabilities P_{ij} are also unknown. One can use the bi-level programming technique to solve for all the model parameters.

Initialize $p_{ij}^{(0)}$; $e = 1$; $h = 1$;

Solve $\alpha_{ik}^{(h)}$

$$\min_{\alpha_{ik}^{(h)}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik}^{(h)} p_{kj}^{(h-1)} - [Q]_{ij} \right)^2 \right\}$$

subject to:

$$\sum_{k=1}^{m} \alpha_{ik}^{(h)} = 1 \quad \text{and} \quad \alpha_{ik}^{(h)} \geqslant 0;$$

Solve $p_{ik}^{(h)}$

$$\min_{p_{ik}^{(h)}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik}^{(h)} p_{kj}^{(h)} - [Q]_{ij} \right)^2 \right\}$$

subject to:

$$\sum_{k=1}^{n} p_{ik}^{(h)} = 1 \quad \text{and} \quad p_{ik}^{(h)} \geqslant 0.$$

While e < tolerance,

$h := h + 1;$

Solve $\alpha_{ik}^{(h)}$

$$\min_{\alpha_{ik}^{(h)}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik}^{(h)} p_{kj}^{(h-1)} - [Q]_{ij} \right)^2 \right\}$$

subject to:

$$\sum_{k=1}^{m} \alpha_{ik}^{(h)} = 1 \quad \text{and} \quad \alpha_{ik}^{(h)} \geqslant 0$$

Solve $p_{ik}^{(h)}$

$$\min_{p_{ik}^{(h)}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik}^{(h)} p_{kj}^{(h)} - [Q]_{ij} \right)^2 \right\}$$

subject to:

$$\sum_{k=1}^{n} p_{ik}^{(h)} = 1 \quad \text{and} \quad p_{ik}^{(h)} \geqslant 0$$

$e := (\|\alpha^{(h)} - \alpha^{(h-1)}\|_2^2 + \|P^{(h)} - P^{(h-1)}\|_2^2)/N;$
end.

We remark that $N = \mathrm{O}(mn)$ or $N = 2mn - m - n$ is the total number of the evaluated parameters. Thus, the value of e is the average tolerance for each evaluated parameter in either $\alpha^{(h)}$ or $P^{(h)}$.

4. Practical numerical examples

In this section, we present an application of the IHMM to both the production planning problem (Ching *et al.*, 2002, 2003, 2004a) and the economic data. In the first experiment, based on the prior knowledge on the numerical data in the production planning problem, we can obtain both the product of the soft drink company and the corresponding true unobservable sequence for some sales periods. Therefore, we consider two different scenarios to illustrate the performance of the proposed IHMM, i.e. case I:

TABLE 1 *The average results for both the HMM and the IHMM given P*

	Prediction accuracy (in %)		Computational time (in s)	
T	IHMM	1-HMM	IHMM	1-HMM
538	61.71	55.39	0.2	0.7581

TABLE 2 *The average results for both the HMM and the IHMM*

	Prediction accuracy (in %)		Computational time (in s)		Number of iteration	
T	IHMM	1-HMM	IHMM	1-HMM	IHMM	1-HMM
538	60.19	50.84	5.0208	52.3135	22.0000	326.0100

P is known and case II: P is unknown. In the second experiment, we investigate the use of the IHMM to extract information on the unobservable states of Hong Kong economy during the period 1994–2004, from the observable data on the Hong Kong sovereign ratings data by FitchRatings and the 1-week Hong Kong Inter-Bank Offered Rate (HIBOR) from DataStream.

4.1 *Production planning analysis*

A soft drink company in Hong Kong faces an in-house problem of production planning and inventory control. The company needs to find the interplay between the storage space requirement and the overall growing sales demand. There are product categories due to the sales volume. All products are labelled as very high sales volume (state 1), high sales volume (state 2), moderate high sales volume (state 3), moderate low sale volume (state 4), low sales volume (state 5) or very low sales volume (state 6). Such labelling is useful from both marketing and production planning points of view. Based on some prior knowledge on the data set, some unobservable environmental effects on the sales volume data are known in advance. These external effects can be roughly classified as two hidden states, say 1 (bad economic condition or hot weather) and 2 (good economic condition or cold weather). The binary hidden sequence can be obtained. In our experiment, only the sales volume sequence is used for training the model parameters. The hidden information will be used later for validating our estimated matrix, α and P. The categorical sequence for the demand of the product of the soft drink company and the corresponding unobservable sequence for some sales periods are given in Tables 3 and 4.

To evaluate the performance and effectiveness of the IHMM, a prediction result on the hidden sequence $(H_1, H_2, \ldots, H_T)$, given observable sales demand sequence $(O_1, O_2, \ldots, O_T)$. For the traditional HMM, there is a standard algorithm to estimate the most likely hidden sequence (Rabiner, 1989). While for our IHMM, given the observable state, one estimates the hidden state by using the one having the highest probability (matrix P). The prediction accuracy r is defined as

$$r = \frac{\sum_{t=1}^{T} \delta_t}{T},$$

where T is the length of the data sequence and

$$\delta_t = \begin{cases} 1, & \text{if } \tilde{H}_t = H_t \text{ (correct prediction)}, \\ 0, & \text{otherwise.} \end{cases}$$

TABLE 3 *The observable demand sequence and parameters for the soft drink*

1 4 4 1 4 1 1 1 1 1 1 4 4 4 4 4 4 4 4 4 2 4 4 4 4 4 4 1 2 1 4 4 1 1 1 4 4 2 1 4 4 2 4 4 4 4 2 4 4 4 2 2 4 4 1 1 1 4 1 2
1 4 2 2 2 4 4 4 4 4 4 4 1 4 2 4 4 4 2 2 4 1 4 4 1 2 1 4 4 4 1 4 4 4 4 4 2 4 4 1 2 4 1 4 1 1 4 4 1 1 1 1 4 1 1 2 1 4 1 4
2 4 4 4 4 4 4 4 2 4 4 1 4 1 1 4 2 4 2 4 4 4 3 4 4 2 4 4 1 4 4 4 4 4 4 4 4 4 1 4 4 2 2 2 3 ! 2 2 4 4 3 3 4 3 2 1 2 2 1 4 4
1 4 4 1 1 4 2 4 4 2 4 2 4 4 1 1 4 4 4 4 4 4 1 4 4 4 1 4 3 4 4 1 1 1 2 2 3 2 2 4 2 4 4 4 4 4 4 4 4 4 4 4 2 4 4 2 2 1 4 1 4
1 1 4 2 1 1 1 4 4 4 4 4 4 2 2 4 4 2 4 1 2 2 1 4 1 1 1 1 2 4 1 1 1 2 4 4 4 2 4 4 2 4 4 2 4 1 1 4 1 4 1 2 1 4 1 4 4 1 4 4
2 4 2 1 2 4 4 2 2 4 2 4 1 2 2 2 2 3 3 2 2 2 4 4 4 2 2 4 2 ! 4 4 4 2 2 2 4 4 2 4 ! 2 1 1 4 1 1 1 4 1 4 1 4 4 4 2 2 4 1 2 1 2
4 4 2 2 1 4 1 1 4 4 2 1 1 4 4 1 1 4 4 1 2 2 1 4 2 4 2 2 1 1 2 4 2 4 4 4 1 2 4 1 4 1 1 5 4 3 4 1 4 4 1 2 2 4 1 4 1 1 4 1
4 1 1 4 4 1 4 4 1 2 4 4 1 4 2 4 2 4 4 4 1 4 4 1 4 4 1 1 4 4 2 2 2 2 2 1 4 4 2 2 2 1 2 3 4 1 4 1 2 2 2 2 4 4 4 4 1 4 4 4
1 4 1 4 1 4 4 4 1 1 4 4 4 4 4 4 1 4 4 1 1 4 1 3 3 3 5 4 1 4 4 4 1 4 1 1 4 4 4 4 1 4 4 4 4 4 4 4 4 4 4 4 4 4 1 4 4 1 4

TABLE 4 *The unobservable effect for the soft drink*

1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1
1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2
2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 ! 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2
2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1
1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2
2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 ! 2 2 2 2 2 2 2 2 2 2 ! 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2
2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1
1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1
1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 ! 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2

4.1.1 *Prediction of demand.* We compare the performance of our proposed model and the first-order HMM in two scenarios. In both of the scenarios, we assume that the total number of hidden states is provided and the model parameters of the first-order HMM can be estimated by using the Viterbi algorithm, forward–backward algorithm and the expectation maximization (EM) algorithm in Rabiner (1989). More precisely, the estimation method on the model parameters ensures that the probability $P(O|\lambda_t)$ is higher than or equal to $P(O|\lambda_{t-1})$, where O is the observation sequence and λ_t is the model at time t. The only difference is that the emission probability matrix $P^{(0)}$ is known in advance in the first case, but not in the latter case. While initialization for the emission probability matrix $P^{(0)}$ is required in the second case. For the first-order HMM parameters' estimation, let us denote matrix A as the hidden state transition probability matrix. Here, we choose the initial transition probability matrix A to be a uniform transition probability matrix for both cases. In order to make a fair comparison, the stopping criterion of the first-order HMM is chosen as

$$(\|A^{(t-1)} - A^{(t)}\|_2^2 + \|P^{(t-1)} - P^{(t)}\|_2^2)/M < e,$$

where e is the tolerance and $M = m^2 + mn$ is the total number of the evaluated parameters.

4.1.2 *Case I: P is known.* Based on the proposed estimation method presented in Section 3.2, α can be estimated by one iteration with given P. On the other hand, if the transition probability matrix $A^{(0)}$ are initialized, we iteratively estimate matrix A until it reaches the stopping criterion. Here, we intend to take the most stable or convergent matrix A, and we choose the tolerance e to be equal to $1.00e-20$. The results are reported in Table 1.

Here, IHMM represents the interactive HMM and 1-HMM represents the first-order HMM. From Table 1, we note that the prediction result of our model is over 6% better than that of the first-order HMM while the computational time in the IHMM is less than one-third of the first-order HMM. This shows that if the emission probability matrix is given and the data fulfill our model assumptions, our proposed model outperforms the ordinary first-order HMM on the data.

4.1.3 *Case II: P is unknown.* Based on the proposed estimation method presented in Section 3.2, both α and P can be estimated iteratively. On the other hand, if the initial $A^{(0)}$ is given, the model parameters of the first-order HMM can also be estimated by using the Viterbi algorithm, forward–backward algorithm and EM algorithm in Rabiner (1989). However, the estimation method in both models cannot guarantee the global minimum solution; the obtained local minimum solution is determined by the initial value of $P^{(0)}$ and the value of tolerance. Here, if the value of the tolerance is set to be too large, the algorithm will be terminated before convergence is obtained. Thus, we have chosen the tolerance e to be equal to $1.00e-18$ times the total number of unknown parameters. Further, we repeat the estimation process 100 times with different initial matrix $P^{(0)}$. The average results are reported in Table 2.

From Table 2, we observe that the average prediction result of our model is over 9% better than that of the first-order HMM, while both the computational time and the number of iterations of the IHMM are less than one-tenth of those of the first-order HMM. This shows that our proposed model provides a significant improvement for the first-order HMM based on the data set. On the other hand, the average predication accuracy of the first-order HMM is very close to the random chosen one (50%). The main reason may be due to the structural property of the hidden and observation sequences. Based on both sequences, we found that the transition matrix A can be estimated from $\hat{\alpha}\hat{P}$ and it is equal to

$$\begin{pmatrix} 0.5287 & 0.4713 \\ 0.4691 & 0.5309 \end{pmatrix},$$

which is close to the uniform distribution matrix. Here, $\hat{\alpha}_{ij}$ can be estimated by counting the transition frequency from the observation state i at time t to the hidden state j at time $t+1$ after normalization. Here, $\hat{P}$ can be obtained similarly. We observe that even the matrix A can be well estimated and the true hidden sequence can hardly be obtained owing to the close probability value for either staying at the same hidden state or jumping to another hidden state.

4.2 *Analysis of economic data*

In this subsection, we investigate the use of the IHMM for extracting information about the Hong Kong economy, in particular, the unobservable states of the Hong Kong economy, from the observable economic data, namely, the Hong Kong sovereign ratings data from FitchRatings and the 1-week HIBOR from DataStream. We investigate the impact of the presence of the interactive effect in the HMM on the extraction of the economic states by comparing the economic states implied by the IHMM with those implied by the traditional HMM without interactive effect over the 11-year period. We also discuss the economic implications of the results based on the Hong Kong economy during that period. Table 5 displays the Hong Kong sovereign ratings data and the 1-week HIBOR data with the corresponding dates used in our study. In either the IHMM or the HMM, we need to convert the observable economic data into discrete states. For the Hong Kong sovereign ratings data, there are two states, namely, H = AA− and L = A+. We convert the 1-week HIBOR data (i.e. spot interest rates data) into discrete states

TABLE 5 *Observable economic data*

Date	1-week HIBOR	Hong Kong sovereign ratings
16 May 2004	0.06	AA−
24 April 2003	1.56	AA−
25 June 2001	3.75	AA−
21 September 2000	6.25	A+
14 November 1996	5.25	A+
24 November 1995	5.62	A+
26 October 1995	5.25	A+
09 August 1995	5.38	A+
10 August 1994	4.19	AA−

TABLE 6 *The transformation from two state spaces into an enlarged state space*

Sovereign ratings	Spot interest rates	Enlarged state
H	1	1
H	2	2
H	3	3
L	1	4
L	2	5
L	3	6

according to the following rule:

$$s_t = \begin{cases} 1, & \text{if } r_t > r_{t-1}, \\ 2, & \text{if } r_t = r_{t-1}, \\ 3, & \text{if } r_t < r_{t-1}, \end{cases} \qquad \text{for} \quad t > 1,$$

where $s_t = 1$, $s_t = 2$ and $s_t = 3$ represent an 'up move', 'no move' and 'down move' of the spot interest rate at time t relative to the spot interest rate at time $t-1$. We further assume that the initial state of the spot interest rate at time $t = 1$ is 2 (i.e. $s_1 = 2$). Then, we obtain the following sequences of sovereign ratings data and spot interest rates:

Sovereign ratings: H, L, L, L, L, L, H, H, H.

Spot interest rates: 2, 1, 3, 1, 3, 1, 3, 3, 3.

Here, the length of both data sequences is $T = 9$. We assume that in both the IHMM and the first-order HMM (i.e. the control model), there are three hidden states of economic conditions, namely, 'Recession', 'Neutral' and 'Expansion', which are represented by R, N and E, respectively. Then, we adopt a transformation method to combine the state spaces of the sovereign ratings and the spot interest rates into an enlarged state space. The transformation and the enlarged state space are presented in Table 6.

Note that the observable sequence here is given by the combined states of sovereign ratings and spot interest rates, while the hidden sequence is the one given by the hidden state of economy in Hong Kong.

Based on the states in the enlarged state space, the states of the combined observable sequence over time are given by

$$2, 4, 6, 4, 6, 4, 3, 3, 3.$$

By the estimation method in Section 3.2, we obtain the estimated hidden sequence of Hong Kong economy implied by the IHMM as follows:

$$\mathrm{E},\ \mathrm{N},\ \mathrm{E},\ \mathrm{N},\ \mathrm{E},\ \mathrm{N},\ \mathrm{R},\ \mathrm{R},\ \mathrm{R}.$$

We also obtain the estimated hidden sequence of Hong Kong economy implied by the first-order HMM as follows:

$$\mathrm{E},\ \mathrm{R},\ \mathrm{N},\ \mathrm{R},\ \mathrm{N},\ \mathrm{R},\ \mathrm{N},\ \mathrm{R},\ \mathrm{N}.$$

From the hidden sequence of Hong Kong economy obtained by the IHMM, the economic conditions have gone through fluctuations between Expansion and Neutral from the third quarter of 1994 to the last quarter of 1996. In the third quarter of 2001, the economic condition has turned to Recession. It remains in the state of Recession until the end of the second quarter of 2004. On the other hand, the hidden sequence of Hong Kong economy implied by the first-order HMM reveals that the economic condition was Expansion in the third quarter of 1994 and that the economic condition then fluctuates between Recession and Neutral from the third quarter of 1995 to the second quarter of 2004.

In view of the Hong Kong economy during the period 1994–2004, the economic conditions implied by the IHMM are more consistent with those implied by the traditional HMM without incorporating the interactive effect. At the beginning of the 1990s, Hong Kong has experienced a high economic growth. This growth sustained until the Asian financial crisis in 1998. During the period 1998–2003, Hong Kong experienced economic downturn, which may be attributed to both internal and external factors. Internally, the unemployment rate reached remarkably high level during that period and the internal consumption was weak. A high level of deflation was recorded during that period. Due to the change in the public housing policy in 1997, there was a serious drop in the prices of real estates. Some of the real estates even diminished their values by around 60%. The Hong Kong real estates market remains in downturn during the period 1997–2003. This leads to recession in Hong Kong economy during that period, since the real estates market is one of the major components in the Hong Kong economy. The outbreak of severe acute respiratory syndrome and bird-flu during that period also seriously harmed the Hong Kong economy. Externally, there were Asian financial crises from 1997 to 1999, an unstable postcrisis period in 2000, and the September 11 terrorist attack in United States in 2001. Since Hong Kong is an open economy, its economic condition is greatly affected by the economic conditions in United States and all over the world.

5. Summary

In this paper, we proposed an IHMM. Our model differs from the traditional HMM in the way that the hidden states are also affected directly by the observable states. We also developed an efficient estimation method for solving the model parameters. Numerical examples on the sales demand data and economic data have been given to demonstrate both the efficiency and effectiveness of the proposed model.

Acknowledgements

Research was supported in part by HKU CRCG Grants and Strategic Research Theme Fund on Computational Physics and Numerical Methods, RGC 7046/03P, 7035/04P, 7035/05P and HKBU FRGs.

The authors would like to thank the two anonymous referees for their many helpful and valuable comments and suggestions.

REFERENCES

AGGOUN, L. & BENKHEROUF, L. (2003) Parameter tracking for an inventory model. *Appl. Math. Comput.*, **146**, 393–401.

BROCKWELL, P. & DAVIS, R. (1991) *Time Series: Theory and Methods*. New York: Springer.

CHING, W., FUNG, E. & NG, M. (2002) A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA J. Manage. Math.*, **13**, 187–199.

CHING, W., FUNG, E. & NG, M. (2003) A higher-order Markov model for the Newsboy's problem. *J. Oper. Res. Soc.*, **54**, 291–298.

CHING, W., FUNG, E. & NG, M. (2004a) Building higher-order Markov chain models with EXCEL. *Int. J. Math. Educ. Sci. Eng.*, **35**, 921–932.

CHING, W., FUNG, E. & NG, M. (2004b) Higher-order Markov chain models for categorical data sequences. *Int. J. Nav. Res. Logist.*, **51**, 557–574.

CHING, W. & NG, M. (2006) *Markov Chains: Models, Algorithms and Applications*. International Series on Operations Research and Management Science. New York: Springer.

CHING, W., NG, M. & WONG, K. (2004c) Hidden Markov models and its applications to customer relationship management. *IMA J. Manage. Math.*, **15**, 13–24.

ELLIOTT, R. J., AGGOUN, L. & MOORE, J. B. (1994) *Hidden Markov Models: Estimation and Control*. Berlin: Springer.

LI, W. & KWOK, M. (1990) Some results on the estimation of a higher order Markov chain. *Commun. Stat. Simul. Comput.*, **19**, 363–380.

MACDONALD, I. & ZUCCHINI, W. (1999) *Hidden Markov and Other Models for Discrete-Valued Time Series*. London: Chapman & Hall.

RABINER, L. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, **77**, 257–286.

RAFTERY, A. (1985) A model for high-order Markov chains. *J. R. Stat. Soc. B*, **47**, 528–539.

SIU, T., CHING, W., NG, M. & FUNG, E. (2005a) Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models. *Comput. Econ.*, **26**, 251–284.

SIU, T., CHING, W., NG, M. & FUNG, E. (2005b) On a multivariate Markov chain model for credit risk measurement. *Quant. Financ.*, **5**, 543–556.

TONG, H. (1977) Discussion of a paper by A. J. Lawrence and N. T. Kottegoda. *J. R. Stat. Soc. A*, **140**, 34–35.

TONG, H. (1978) On a threshold model. *Pattern Recognition and Signal Processing* (C. H. Chen ed.). Amsterdam: Sijthoff and Noordhoff, pp. 575–586.

TONG, H. (1983) *Threshold Models in Non-linear Time Series Analysis*. Vol. 21. Lecture Notes in Statistics. Heidelberg: Springer.

TONG, H. (1990) *Nonlinear Time Series: A Dynamical Systems Approach*. Oxford: Oxford University Press.

WATERMAN, M. (1995) *Introduction to Computational Biology*. Cambridge: Chapman & Hall.