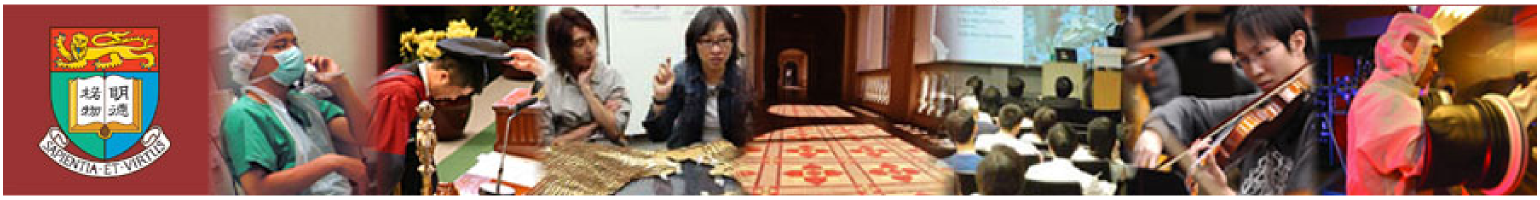

2010 Third International Conference on Business Intelligence and Financial Engineering

A New Estimation Method for Multivariate Markov Chain Model with Application in Demand Predictions

Dong-Mei Zhu*
Advanced Modeling and Applied Computing Laboratory
Department of Mathematics
The University of Hong Kong
Pokfulam Road, Hong Kong
Hong Kong
Email: dongmeizhu86@gmail.com

Wai-Ki Ching
Advanced Modeling and Applied Computing Laboratory
Department of Mathematics
The University of Hong Kong
Pokfulam Road, Hong Kong
Hong Kong
Email: wkc@maths.hku.hk

***Abstract*—In this paper, we propose a new estimation method for the parameters of a multivariate Markov chain model. In the new method, we calculate the correlations of the sequences first and establish multivariate Markov chain models for those positively correlated sequences. The parameters are estimated by minimizing the error of prediction. We apply the method to demand predictions for a soft-drink company in Hong Kong. Numerical experiments are given to show the effectiveness of our proposed method.**

Key Words: Multivariate Markov Chain Model, Demand Prediction.

I. Introduction

Markov chain models have been used successfully in modeling many practical systems such as telecommunication systems and manufacturing systems [1], [2] and many other systems [6]. A Markov chain process is a stochastic process satisfying the "Markov property" [10]. In this paper we consider discrete-time Markov processes having finite discrete states for modeling categorical data sequences. Categorical data sequences (or time series) occur frequently in many real world applications [5]. Suppose a categorical data sequence $\{x_1, x_2, \ldots, x_T\}$ having m states is given. Here $x_i \in \mathcal{M} = \{1, 2, \ldots, m\}$ and T is the length of the sequence. The categorical data sequence can also be represented by a sequence of vectors (the canonical form representation) $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ where $\mathbf{x}_i = \mathbf{e}_k$ ($\mathbf{e}_k$ is the unit vector with the kth entry being equal to one) if it is in state k. This representation facilitates our discussion of the problem in a Markov chain framework. By making use of the transition probability matrix, a categorical data sequence of m states can be modeled by an m-state Markov chain model. We then extend this idea to model multiple categorical data sequences. One would expect categorical data sequences generated by similar sources or same source to be correlated to each other. Therefore by exploring these relationships, one can develop a better model for the categorical data sequences and hence better prediction rules.

Modeling the categorical data sequences is vital for good predictions and optimal planning in a decision process. We note that the conventional first-order Markov chain model for s categorical data sequences of m states has m^s states. The number of parameters (transition probabilities) increases exponentially with respect to the number of categorical sequences. This large number of parameters is a major obstacle and discourages us from using such kind of Markov chain model directly. According to [3], multivariate Markov model capturing both the intra- and inter-transition probabilities among the sequences are developed. And the number of parameters in the model is only ($s^2m^2 + s^2$). Also a parameter estimation method based on linear programming is proposed. Here we still use the multivariate Markov model (see for [3]). And we propose a new estimation method for the parameters of the multivariate Markov model. Due to our purpose of making good predictions, we estimate the parameters by minimizing the error prediction. We also give the results by using the method based on nonlinear programming problems, which is similar to the method introduced in [3]. Then we apply the model and method to solve the sales demand prediction problem in a soft-drink company in Hong Kong. We also compare the results with the methods in [3] and other new method yields better results.

The rest of the paper is organized as follows. In Section 2, we first give a review on the Markov chain model. We then briefly present the multivariate Markov model [3]. In Section 3, we propose estimation methods for the model parameters. In Section 4, numerical examples based on sales demand data are given to demonstrate the effectiveness of our method. Finally, concluding remarks are given to summarize the paper and address further research issues in Section 5.

II. The Multivariate Markov Chain Model

In this section we first give a short review on discrete-time Markov chain. We then briefly present the multivariate

978-0-7695-4116-7/10 $26.00 © 2010 IEEE
DOI 10.1109/BIFE.2010.39

Markov chain model proposed in [3], [6].

A discrete-time stochastic process is a family of random variables $\{\mathcal{X}_t, t \in N\}$ defined on a given probability space and indexed by the parameter $t \in N = \{0, 1, 2, \cdots\}$. A Markov process is a stochastic process whose conditional probability distribution function satisfies the so-called "Markov property". If the state-space of a Markov process is discrete, the Markov process is a discrete-time stochastic process and is also called a Markov chain. In this paper we consider Markov chains having finite number of states $\mathcal{M} = \{1, 2, 3, \cdots, m\}$. In generally a categorical data sequence $x_1, x_2, x_3, \ldots, x_T$ can be logically represented by a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_T$, where T is the length of the sequence, and $\mathbf{x}_i = \mathbf{e}_k$ ($\mathbf{e}_k$ is the unit vector with the kth entry being one) if it is in state k. A first-order discrete-time Markov chain having m discrete states satisfies the following relationship:

$$\begin{aligned}&\text{Prob } (\mathbf{x}_{t+1} = \mathbf{e}_{\mathbf{x}_{t+1}} \mid \mathbf{x}_0 = \mathbf{e}_{x_0}, \mathbf{x}_1 = \mathbf{e}_{x_1}, \ldots, \mathbf{x}_t = \mathbf{e}_{x_t}) \\ &= \text{Prob } (\mathbf{x}_{t+1} = \mathbf{e}_{x_{t+1}} \mid \mathbf{x}_t = \mathbf{e}_{x_t})\end{aligned}$$

where $x_i \in M$. The conditional probabilities

$$\text{Prob}(\mathbf{x}_{n+1} = \mathbf{e}_{x_{n+1}} \mid \mathbf{x}_n = \mathbf{e}_{x_n})$$

are called the single-step transition probabilities of the Markov chain. They give the conditional probability of making a transition from state i to state j when the time parameter increases from n to $n+1$. These probabilities are independent of n and are written as

$$p_{ij} = \text{ Prob } (\mathbf{x}_{n+1} = \mathbf{e}_i \mid \mathbf{x}_n = \mathbf{e}_j), \ \forall \ i, j \in M.$$

The matrix P, formed by placing p_{ij} in row i and column j for all i and j, is called the transition probability matrix. We note that the elements of the matrix P satisfy the following two properties:

$$0 \le p_{ij} \le 1 \quad \forall i, j \in \mathcal{M} \quad \text{and} \quad \sum_{i=1}^{m} p_{ij} = 1, \quad \forall j \in \mathcal{M}.$$

We then apply the multivariate Markov chain model proposed in [3] to represent the behavior of multiple categorical sequences generated by similar sources or same source. Here we assume that there are s categorical sequences and each has m possible states in M. Let $\mathbf{x}_n^{(k)}$ be the state vector of the kth sequence at time n. If the kth sequence is in state j at time n then

$$\mathbf{x}_n^{(k)} = \mathbf{e}_j = (0, \ldots, 0, \underbrace{1}_{j\text{th entry}}, 0 \ldots, 0)^T.$$

In this multivariate Markov chain model, the following relationship is assumed:

$$\mathbf{x}_{n+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} \mathbf{x}_n^{(k)}, \quad \text{for} \quad j = 1, 2, \cdots, s \qquad (1)$$

where $\lambda_{jk} \ge 0, \ 1 \le j, k \le s$ and

$$\sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{for} \quad j = 1, 2, \cdots, s. \qquad (2)$$

The state probability distribution of the kth sequence at the $(n+1)$th step depends on the weighted average of $P^{(jk)}\mathbf{x}_n^{(k)}$. Here $P^{(jk)}$ is a transition probability matrix from the states in the kth sequence to the states in the jth sequence, and $\mathbf{x}_n^{(k)}$ is the state probability distribution of the kth sequences at the nth step. In matrix form, the model can be represented as follows:

$$\begin{aligned}\mathbf{x}_{n+1} &\equiv \left(\begin{array}{cccc} \mathbf{x}_{n+1}^{(1)} & \mathbf{x}_{n+1}^{(2)} & \ldots & \mathbf{x}_{n+1}^{(s)} \end{array} \right)^T \\ &= \begin{pmatrix} \lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \cdots & \lambda_{1s}P^{(1s)} \\ \lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}P^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n^{(1)} \\ \mathbf{x}_n^{(2)} \\ \vdots \\ \mathbf{x}_n^{(s)} \end{pmatrix} \\ &\equiv Q\mathbf{x}_n.\end{aligned}$$

III. Estimations of Model Parameters

In this section we introduce some methods for the estimations of $P^{(jk)}$ and λ_{jk}. For each data sequence, we estimate the transition probability matrix by the following method, see for instance [3]. Given the data sequence, we count the transition frequency from the states in the kth sequence to the states in the jth sequence. Hence one can construct a transition frequency matrix for the data sequences. After making normalization in each column, the estimates of the transition probability matrices can also be obtained. More precisely, we count the transition frequency $f_{i_j i_k}^{(jk)}$ from the state i_k in the sequence $\{x_n^{(k)}\}$ to the state i_j in the sequence $\{x_n^{(j)}\}$ and therefore we construct the transition frequency matrix for the sequences as follows:

$$F^{(jk)} = \begin{pmatrix} f_{11}^{(jk)} & \cdots & \cdots & f_{m1}^{(jk)} \\ f_{12}^{(jk)} & \cdots & \cdots & f_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1m}^{(jk)} & \cdots & \cdots & f_{mm}^{(jk)} \end{pmatrix}.$$

From $F^{(jk)}$, we get the estimates for $P^{(jk)}$ as follows:

$$\hat{P}^{(jk)} = \begin{pmatrix} \hat{p}_{11}^{(jk)} & \cdots & \cdots & \hat{p}_{m1}^{(jk)} \\ \hat{p}_{12}^{(jk)} & \cdots & \cdots & \hat{p}_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{p}_{1m}^{(jk)} & \cdots & \cdots & \hat{p}_{mm}^{(jk)} \end{pmatrix}$$

where

$$\hat{p}_{i_j i_k}^{(jk)} = \begin{cases} \dfrac{f_{i_j i_k}^{(jk)}}{\sum_{i_k=1}^{m} f_{i_j i_k}^{(jk)}} & \text{if } \sum_{i_k=1}^{m} f_{i_j i_k}^{(jk)} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Besides the estimates of $P^{(jk)}$, we need to estimate the parameters λ_{jk}. According to [3], we have seen that the multivariate Markov chain has a stationary vector $\mathbf{x}$. The vector $\mathbf{x}$ can be estimated from the sequences by computing the proportion of the occurrence of each state in each of the sequences, and let us denote it by $\hat{\mathbf{x}} = (\hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)}, \ldots, \hat{\mathbf{x}}^{(s)})^T$. Then one would expect

$$\begin{pmatrix} \lambda_{11}\hat{P}^{(11)} & \lambda_{12}\hat{P}^{(12)} & \cdots & \lambda_{1s}\hat{P}^{(1s)} \\ \lambda_{21}\hat{P}^{(21)} & \lambda_{22}\hat{P}^{(22)} & \cdots & \lambda_{2s}\hat{P}^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}\hat{P}^{(s1)} & \lambda_{s2}\hat{P}^{(s2)} & \cdots & \lambda_{ss}\hat{P}^{(ss)} \end{pmatrix} \hat{\mathbf{X}} \approx \hat{\mathbf{X}}. \quad (3)$$

From (3), it suggests one possible way to estimate the parameters $\lambda = \{\lambda_{jk}\}$ as follows. One may consider solving the following optimization problem (I):

$$\begin{cases} \min_{\lambda} \sum_{i=1}^{m} \left[\sum_{k=1}^{s} \lambda_{jk}\hat{P}^{(jk)}\hat{\mathbf{x}}^{(k)} - \hat{\mathbf{x}}^{(j)} \right]_i^2 \\ \text{subject to} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{and} \quad \lambda_{jk} \geq 0, \quad \forall k. \end{cases} \quad (4)$$

It can actually be formulated as quadratic programming problems. Here we give a new method to estimate the parameters $\lambda = \lambda_{jk}$. As our model can only capture the positive correlations among those sequences, we first find the correlations between the sequences and build multivariate Markov chain models for those positively correlated sequences. Due to our main purpose here is to minimize the error of prediction, we consider solving λ_{jk} by minimizing the error of prediction.

Here we let $\mathbf{X}_t = (\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)} \ldots, \mathbf{x}_t^{(s)})^T$ be the state probability distribution of the system at time t observed from the real dataset. And let

$$\hat{P} = \begin{pmatrix} \lambda_{11}\hat{P}^{(11)} & \lambda_{12}\hat{P}^{(12)} & \cdots & \lambda_{1s}\hat{P}^{(1s)} \\ \lambda_{21}\hat{P}^{(21)} & \lambda_{22}\hat{P}^{(22)} & \cdots & \lambda_{2s}\hat{P}^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}\hat{P}^{(s1)} & \lambda_{s2}\hat{P}^{(s2)} & \cdots & \lambda_{ss}\hat{P}^{(ss)} \end{pmatrix}.$$

We then consider solving λ_{ij} by minimizing the following:

$$r_2 = \sqrt{\sum_{t=1}^{N-1} \frac{1}{N-1} ||\mathbf{X}_{t+1} - \hat{P}\mathbf{X}_t||_2^2}. \quad (5)$$

Then one may consider solving the following optimization problem(II):

$$\begin{cases} \min_{\lambda} \sum_{t=1}^{N-1} \frac{1}{N-1} ||\mathbf{X}_{t+1} - \hat{P}\mathbf{X}_t||_2^2 \\ \text{subject to} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{and} \quad \lambda_{jk} \geq 0, \quad \forall k. \end{cases} \quad (6)$$

IV. Application to Sales Demand Predictions

In this section, we consider a soft-drink company in Hong Kong. The company faces an in-house problem of production planning. An important issue that stands out is the storage space of its central warehouse, which often finds itself in the state of overflow or near capacity. The company is thus in urgent need to study the interplay between the storage space requirement and the overall growing sales demand. The products are categorized into six possible states according to their sales volumes. All products are labeled as either very fast-moving (very high sales volume), fast-moving, standard, slow-moving, very slow-moving (low sales volume) or no sales volume. Such labels are therefore useful from both marketing and production planning points of view.

The company got an important customer and would like to predict sales demand for this customer in order to minimize its inventory build-up and to maximize the demand satisfaction for this customer [3]. The company can analyze the sales pattern of the customer and then develop a marketing strategy to deal with this customer. In Appendix, we present the customer's historical sales demand of five important products of the company for a year. We expect sales demand sequences generated by the same customer to be correlated to each other. Therefore by exploring these relationships, we develop the multivariate Markov model for such demand sequences, hence obtain better prediction rules.

First we compute the correlations of the five sequences. Here we denote a sequence as a vector and the elements in the vector are the states of the sequence. Then we compute the correlations of the five vectors and their correlations are reported in the following 5×5 matrix:

$$\begin{pmatrix} 1.0000 & 0.0652 & 0.0837 & 0.0651 & -0.0284 \\ 0.0652 & 1.0000 & 0.0730 & 0.0359 & 0.0655 \\ 0.0837 & 0.0730 & 1.0000 & 0.1293 & 0.0773 \\ 0.0651 & 0.0359 & 0.1293 & 1.0000 & 0.7854 \\ -0.0284 & 0.0655 & 0.0773 & 0.7854 & 1.0000 \end{pmatrix}.$$

From this matrix, we observe that Sequences A, B, C and D are positively correlated to each other and Sequences B,C,D and E are positively correlated to each other. Therefore two multivariate Markov chains can be built for these two sets of sequences. We first estimate all the transition probability matrices $P^{(ij)}$ by using the method proposed in Section 3.

Now we estimate the parameters λ_{jk} by using the following optimization problem (I):

$$\begin{cases} \min_{\lambda} \sum_{i=1}^{m} \left[\sum_{k=1}^{s} \lambda_{jk}\hat{P}^{(jk)}\hat{\mathbf{x}}^{(k)} - \hat{\mathbf{x}}^{(j)} \right]_i^2 \\ \text{subject to} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{and} \quad \lambda_{jk} \geq 0, \quad \forall k. \end{cases} \quad (7)$$

By solving the above nonlinear programming problems, one can get a multivariate Markov model for sequences

A, B, C, D:

$$\begin{cases} \mathbf{x}_{t+1}^{(1)} &= 0.7500\hat{P}^{(12)}\mathbf{x}_t^{(2)} + 0.2500\hat{P}^{(14)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(2)} &= 1.0000\hat{P}^{(22)}\mathbf{x}_t^{(2)} \\ \mathbf{x}_{t+1}^{(3)} &= 0.2698\hat{P}^{(32)}\mathbf{x}_t^{(2)} + 0.7032\hat{P}^{(34)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(4)} &= 0.5000\hat{P}^{(41)}\mathbf{x}_t^{(1)} + 0.5000\hat{P}^{(42)}\mathbf{x}_t^{(2)} \end{cases}$$

We can also get a multivariate Markov model for sequences B, C, D, E:

$$\begin{cases} \mathbf{x}_{t+1}^{(2)} &= 0.5000\hat{P}^{(22)}\mathbf{x}_t^{(2)} + 0.5000\hat{P}^{(25)}\mathbf{x}_t^{(5)} \\ \mathbf{x}_{t+1}^{(3)} &= 0.1131\hat{P}^{(32)}\mathbf{x}_t^{(2)} + 0.8869\hat{P}^{(34)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(4)} &= 1.0000\hat{P}^{(42)}\mathbf{x}_t^{(2)} \\ \mathbf{x}_{t+1}^{(5)} &= 0.6250\hat{P}^{(52)}\mathbf{x}_t^{(2)} + 0.3750\hat{P}^{(55)}\mathbf{x}_t^{(5)} \end{cases}$$

Now we estimate the parameters by using our new method. Here we have to solve the following nonlinear programming problem (II):

$$\begin{cases} & r_2 = \min_{\lambda_{jk}} \sum_{t=1}^{N-1} \frac{1}{N-1} ||\mathbf{X}_{t+1} - \hat{P}\mathbf{X}_t||_2^2 \\ \text{subject to} & \sum_{k=1}^{s} \lambda_{jk} = 1; \quad \text{and} \quad \lambda_{jk} \geq 0, \forall k. \end{cases}$$

By solving upper nonlinear programming problem, we get the multivariate Markov model for sequences A, B, C, D:

$$\begin{cases} \mathbf{x}_{t+1}^{(1)} &= 0.4282\hat{P}^{(11)}\mathbf{x}_t^{(1)} + 0.5718\hat{P}^{(12)}\mathbf{x}_t^{(2)} \\ \mathbf{x}_{t+1}^{(2)} &= 0.3696\hat{P}^{(21)}\mathbf{x}_t^{(1)} + 0.4481\hat{P}^{(22)}\mathbf{x}_t^{(2)} + 0.1823\hat{P}^{(24)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(3)} &= 0.2945\hat{P}^{(31)}\mathbf{x}_t^{(1)} + 0.1400\hat{P}^{(32)}\mathbf{x}_t^{(2)} + 0.0336\hat{P}^{(33)}\mathbf{x}_t^{(3)} \\ & \quad +0.5319\hat{P}^{(34)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(4)} &= 0.0450\hat{P}^{(41)}\mathbf{x}_t^{(1)} + 0.0002\hat{P}^{(42)}\mathbf{x}_t^{(2)} + 0.0461\hat{P}^{(43)}\mathbf{x}_t^{(3)} \\ & \quad +0.9087\hat{P}^{(44)}\mathbf{x}_t^{(4)} \end{cases}$$

Similarly, the multivariate Markov chain models for sequences B, C, D, E:

$$\begin{cases} \mathbf{x}_{t+1}^{(2)} &= 0.7391\hat{P}^{(22)}\mathbf{x}_t^{(2)} + 0.0095\hat{P}^{(23)}\mathbf{x}_t^{(3)} + 0.2514\hat{P}^{(24)}\mathbf{x}_t^{(4)} \\ \mathbf{x}_{t+1}^{(3)} &= 0.1796\hat{P}^{(32)}\mathbf{x}_t^{(2)} + 0.2375\hat{P}^{(33)}\mathbf{x}_t^{(3)} + 0.0950\hat{P}^{(34)}\mathbf{x}_t^{(4)} \\ & \quad +0.4879\hat{P}^{(35)}\mathbf{x}_t^{(5)} \\ \mathbf{x}_{t+1}^{(4)} &= 0.0444\hat{P}^{(42)}\mathbf{x}_t^{(2)} + 0.5377\hat{P}^{(44)}\mathbf{x}_t^{(4)} + 0.4179\hat{P}^{(45)}\mathbf{x}_t^{(5)} \\ \mathbf{x}_{t+1}^{(5)} &= 0.0321\hat{P}^{(53)}\mathbf{x}_t^{(3)} + 0.6056\hat{P}^{(54)}\mathbf{x}_t^{(4)} + 0.3623\hat{P}^{(55)}\mathbf{x}_t^{(5)} \end{cases}$$

Next we use the multivariate Markov model, to predict the next state $\hat{\mathbf{x}}_t$ at time t which can be taken as the state with the maximum probability, i.e.,

$$\hat{\mathbf{x}}_t = j, \quad \text{if } [\hat{\mathbf{x}}_t]_i \leq [\hat{\mathbf{x}}_t]_j, \forall 1 \leq i \leq m.$$

To evaluate the performance and effectiveness of our multivariate Markov chain model, a prediction result is measured by the prediction accuracy r defined as

$$r_1 = \frac{1}{T} \times \sum_{t=1}^{T} \delta_t \times 100\%,$$

where T is the length of the data sequence and

$$\delta_t = \begin{cases} 1, & \text{if } \hat{\mathbf{x}}_t = \mathbf{x}_t \\ 0, & \text{otherwise.} \end{cases}$$

As we are going to predict the next state, the more states are the same with the given states, the better the method is. Then we can get that the larger r_1 is, the better the method is.

To further differentiate the power of the methods, we can also consider the error as follows:

$$r_2 = \sqrt{\frac{1}{N-1} \times \sum_{t=1}^{N-1} ||\mathbf{X}_{t+1} - \hat{P}\mathbf{X}_t||_2^2}.$$

Our purpose is to minimize the prediction accuracy, so the less r_2 is, the better the method is.

Now we give the values of r_1 and r_2 for these Markov models of four sequences. For the sake of comparison, we give the results of the first-order Markov chain model (Model (0)), the multivariate Markov chain model getting from the nonlinear programming problem (Model (I)) and our proposed model (Model (II)). The results are reported in Tables 1,2,3 and 4.

	Product A	Product B	Product C	Product D
Model (0)	46%	45%	63%	51%
Model (I)	50%	45%	63%	38%
Model (II)	51%	51%	63%	53%

Table 1. Values of r_1 in the Sales Demand Data.

	Product A	Product B	Product C	Product D
Model (0)	0.7789	0.7891	0.7344	0.7765
Model (I)	0.7780	0.7891	0.7323	0.8516
Model (II)	0.7702	0.7852	0.7316	0.7760

Table 2. Values of r_2 in the Sales Demand Data.

	Product B	Product C	Product D	Product E
Model (0)	45%	63%	51%	53%
Model (I)	47%	63%	37%	53%
Model (II)	49%	63%	52%	54%

Table 3. Values of r_1 in the Sales Demand Data.

	Product B	Product C	Product D	Product E
Model (0)	0.7891	0.7344	0.7765	0.7704
Model (I)	0.7912	0.7333	0.8554	0.8008
Model (II)	0.7868	0.7310	0.7731	0.7640

Table 4. Values of r_2 in the Sales Demand Data.

From Table 1, we can find that for products A, B and D, the values of r_1 calculated from our proposed method are the largest, so our method is the best among these three methods. Although the value of r_1 for product C keeps the same, the value itself is large however. Moreover, similar conclusions can be obtained from Table 3.

From Table 2, we can see that by using our method, the values of r_2 for products A,B,C and D are the least. This means that the prediction accuracy of our proposed method is the best. Similar conclusions can be drawn from Table 4.

V. Concluding Remarks

In this paper, we give a new method to estimate the parameters of a multivariate Markov model proposed in [3]. In the proposed method, we do not have to use the estimation of the stationary vector, which may have large error when the period of the given data sequence is not long enough. Additionally, we build Markov models among those positively correlated sequences. As in [3], all the data sequences are assumed to be positively correlated and this is not necessarily true in practice. Then we can avoid the error caused by those negative associations to certain extent. Moreover, according to the numerical results of the data sequences of the soft-drink company in Hong Kong, our proposed estimation method is better in prediction error.

The followings are some future research issues. We can extend the method to high-order case [7] and also other types of multivariate Markov chain models [8], [9]. By using the same thought, we can give new estimations of the parameters λ_{jk}. Then we can compare the new high-order Markov models with old one to see which one is better. Apart from demand prediction, one can also derive optimal inventory policy for the products using the idea in [4].

VI. Appendix

Sales Demand Sequences of the Five Products

Product A: 6 6 6 6 2 6 2 6 2 2 6 2 6 6 2 6 2 4 4 4 5 6 6 1 2 2 6 6 6 2 6 2 6 6 2 6 2 2 6 2 1 2 2 6 6 6 2 1 2 6 2 6 6 2 2 6 2 2 2 6 2 6 2 2 2 2 2 6 2 2 6 6 6 6 1 2 2 6 2 2 2 2 6 2 2 2 2 3 3 2 3 2 6 6 6 6 2 6 2 6 6 2 6 2 6 6 2 6 6 2 2 3 4 3 3 1 3 1 2 1 6 1 6 6 1 6 6 2 6 2 6 2 2 2 6 6 1 6 2 6 1 2 1 6 2 6 2 2 2 2 6 6 1 6 6 2 2 6 2 2 2 3 4 4 4 6 4 6 1 6 6 1 6 6 6 6 1 6 2 2 2 6 6 6 6 2 6 6 2 2 6 2 6 2 2 2 6 2 2 2 6 6 6 6 3 2 2 6 2 2 2 2 2 2 6 2 6 2 2 2 6 2 2 6 6 2 6 6 6 2 2 2 3 3 3 4 1 6 6 1 6 6 1 6 1 6 6 6 6 1 6 6 6 2 1 2 2 2 2 2 2 3 6 6 6 6 6 2 6

Product B: 1 6 6 1 6 1 1 1 1 1 1 6 6 6 1 2 1 6 6 1 1 1 6 6 2 1 6 6 1 1 1 6 1 2 1 6 2 2 2 2 2 6 1 6 6 1 2 1 6 6 6 1 1 1 6 6 1 1 1 1 6 1 1 2 1 6 1 6 1 1 6 2 6 2 6 6 6 3 6 6 1 6 6 2 2 2 3 2 2 6 6 6 1 1 6 2 6 6 2 6 2 6 6 1 3 6 6 1 1 1 2 2 3 2 2 6 2 2 2 1 6 1 6 1 1 6 2 1 1 1 2 2 1 6 1 1 1 1 2 6 1 1 1 1 6 1 6 1 2 1 6 1 6 6 1 6 1 2 2 2 2 3 3 2 2 2 6 6 6 6 2 1 1 6 1 1 1 6 1 6 1 6 1 6 1 1 6 6 2 1 1 6 6 1 1 2 6 2 6 6 6 1 2 6 1 6 1 1 1 1 6 1 6 1 1 6 6 1 6 6 1 6 1 6 6 1 1 6 6 2 2 2 2 2 2 2 2 2 6 6 6 6 1 6 6 6 1 6 6 1 6 6 1 1 6 1 3 3 3 5 1 6 6 6 6 6 6 6 6

Product C: 6 6 6 6 6 6 6 2 6 6 6 6 6 6 6 2 6 6 6 6 2 6 6 6 2 2 6 6 6 6 6 6 6 1 6 2 6 6 6 6 6 6 6 6 2 6 6 1 2 6 1 6 6 1 6 2 6 6 6 6 6 6 6 2 6 6 6 2 6 6 1 6 6 6 6 6 6 6 3 3 6 3 2 1 2 2 1 6 6 1 6 1 6 6 6 6 6 6 1 6 6 6 1 6 6 6 6 6 6 6 6 6 6 6 2 6 6 6 6 6 6 6 6 2 2 6 6 2 6 1 2 6 6 6 2 6 6 2 6 6 2 6 1 6 2 6 2 1 2 6 6 2 2 6 2 6 2 2 6 2 6 6 6 2 2 2 6 6 2 6 6 2 2 6 1 2 1 2 6 6 2 2 6 6 1 2 2 1 6 2 6 2 2 1 1 5 6 3 6 1 6 6 1 2 2 6 1 6 2 6 6 1 6 2 6 2 6 6 6 1 6 1 6 6 2 2 2 1 2 3 6 1 6 1 6 1 6 1 6 6 6 1 1 6 6 6 6 6 1 6 6 6 1 6 1 1 6 6 6 6 6 6 6 6 1 6 6 1 6

Product D: 6 2 2 2 2 3 3 4 4 4 5 4 3 3 6 2 6 6 6 3 4 4 3 3 3 3 3 2 6 6 3 4 4 4 4 3 4 2 6 2 2 6 2 2 6 6 3 4 5 4 4 6 3 6 6 6 2 6 2 6 6 2 2 6 4 4 5 4 3 4 3 4 4 6 2 6 6 2 2 6 2 6 6 2 6 6 2 6 6 2 6 2 6 3 5 5 5 4 4 4 3 6 2 6 6 2 6 2 6 2 2 6 2 6 6 2 6 4 4 4 4 4 4 6 3 6 6 2 6 2 6 2 6 2 6 6 2 2 2 2 2 2 2 2 2 3 3 3 5 5 4 5 3 3 3 6 2 6 6 2 2 6 2 2 2 2 6 2 3 2 2 3 6 3 2 2 3 4 4 4 4 5 5 4 4 6 6 2 6 2 6 2 2 2 2 2 2 2 2 5 5 4 4 5 5 2 6 2 6 6 2 6 2 6 2 2 3 3 4 4 5 4 4 4 3 4 3 6 2 6 2 2 2 2 2 2 2 2 2 2 2 2 3 4 4 4 4 5 4 4 4 3 2 2 2 6 2 2 2 6 2 6 2 6 2 2 2 2 2 2 3 2

Product E: 6 2 2 2 2 3 3 4 4 4 5 4 3 3 6 2 6 6 2 3 4 4 3 4 4 3 3 2 2 6 3 4 4 4 4 3 4 2 3 2 2 6 3 3 6 6 3 4 5 4 5 3 3 2 6 6 2 6 2 6 6 2 2 6 4 4 4 4 4 4 5 4 4 6 2 6 6 2 2 6 2 6 6 2 6 6 2 6 6 2 6 2 6 3 4 4 4 4 4 4 4 6 2 6 6 2 6 2 6 6 6 6 2 6 2 2 6 4 4 4 4 4 4 6 3 3 6 2 2 2 6 2 6 2 2 2 2 2 2 2 2 2 2 2 2 2 3 6 4 5 5 5 5 2 4 6 6 2 6 6 2 2 6 2 2 2 2 6 2 3 2 2 3 6 3 2 2 3 4 4 4 4 5 5 4 3 3 6 2 6 2 2 2 6 3 2 2 2 2 5 5 4 4 4 4 3 6 2 6 6 2 6 2 6 2 2 3 3 4 4 5 4 4 4 4 4 3 6 2 6 2 2 2 6 2 2 2 2 2 2 2 3 4 4 4 4 5 4 4 4 3 2 2 2 6 6 6 2 6 2 6 2 6 2 2 2 2 2 2 2

6=very fast-moving, 5 = fast-moving, 4 = standard, 3 = slow-moving, 2 = very slow-moving and 1= no sales volume.

Acknowledgment

Research supported in part by HKRGC Grant No. 7017/07P, HKUCRGC Grants and HKU Strategy Research Theme fund on Computational Sciences. The authors would like to thank the two anonymous reviewers for their helpful comments.

References

[1] W. Ching (1997) *Markov Modulated Poisson Processes for Multi-location Inventory Problems*, Inter. J. Prod. Econ., **53**, pp. 232–239.

[2] W. Ching (2001) *Iterative Methods for Queuing and Manufacturing Systems*, Springer Monograph in Mathematics, Springer, London.

[3] W. Ching, E. Fung and M. Ng (2002), *A Multivariate Markov Chain Model for Categorical Data Sequences and Its Applications in Demand Prediction*, IMA J. Manag. Math., **13**, pp. 187-199.

[4] W. Ching, E. Fung and M. Ng (2003) *A Higher-order Markov Model for the Newsboy's Problem*, J. Operat. Res. Soc., **54** 291–298.

[5] W. Ching, E. Fung and M. Ng (2004) *Higher-order Markov Chain Models for Categorical Data Sequences*, Naval Research Logistics, 4 **51**, pp. 557–574.

[6] W. Ching and M. Ng (2006) *Markov Chains: Models, Algorithms and Applications*, Springer International Series in Operations Research & Management Series, Springer, New York.

[7] W. Ching, M. Ng and E. Fung (2008) *Higher-order Multivariate Markov Chains and Their Applications*, Lin. Alg. Appl., **428**, pp. 492–507.

[8] W. Ching, S. Zhang and M. Ng (2007) *On Multi-dimensional Markov Chain Models*, Pacific Journal of Optimization, **3** pp. 235–243.

[9] W. Ching, T. Siu, L. Li, H. Jiang T. Li and W. Li (2009) *An Improved Parsimonious Multivariate Markov Chain Model for Credit Risk*, Journal of Credit Risk, **5**, pp. 1–25.

[10] S. Ross (2000) *Introduction to Probability Models*, Academic Press, San Diego, CA.