北京

International Conference
on Industrial Engineering and Systems Management
IESM 2007
May 30 - June 2
BEIJING - CHINA

二00七

A New Multivariate Markov Chain Model with Applications to Sales Demand Forecasting ⋆

Wai-Ki CHING [a], Li-Min LI [c], Tang LI [d], Shu-Qin ZHANG [b]

[a] *Advanced Modeling and Applied Computing Laboratory and Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.*

[b] *Advanced Modeling and Applied Computing Laboratory and Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.*

[c] *Advanced Modeling and Applied Computing Laboratory and Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.*

[d] *Advanced Modeling and Applied Computing Laboratory and Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.*

Abstract

Markov chains are popular tools for modeling a lot of practical systems such as queueing systems, manufacturing systems and categorical data sequences. Multiple categorical sequences occur in many applications such as inventory control, finance and data mining. In many situations, one would like to consider multiple categorical sequences together at the same time. The reason is that the data sequences can be correlated and therefore by exploring their relationships, one can develop better models. In this paper, we propose a new multivariate Markov chain model for modeling multiple categorical data sequences. We then test the proposed model with synthetic data and apply it to practical sales demand data.

Key words: Inventory, Markov Process, Multivariate Markov chains, Categorical Data Sequences, Demand Prediction.

1 Introduction

Categorical data sequences (time series) have many applications in both applied sciences and engineering problems such as inventory control problemss [1,10,11], webpage prediction problems in data mining [12], credit risk problems in finance [9] and many others [2,6,8]. Categorical data sequences can be modeled by using Markov chains, see for instance [10,11,8]. In many occasions, one has to consider multiple Markov chains (categorical sequences) together at the same time, i.e., to study the chains in a holistic manner rather than individually. The reason is that the chains (data sequences) can be "correlated" and therefore the information of other chains can contribute to explain the captured chain (data sequence). Thus by exploring these relationships, one can develop better models. We remark that the conventional Markov chain model for s categorical data sequences of m states has m^s states. The number of parameters

⋆ This paper was not presented at any other revue. Research supported in part by RGC Grants and HKU CRCG Grants. Corresponding author Li-Min LI Email: liminli@hkusua.hku.hk Fax (852) 2559-2225.

Email addresses: `wching@hkusua.hku.hk` (Wai-Ki CHING), `liminli@hkusua.hku.hk` (Li-Min LI), `litang@hkusua.hku.hk` (Tang LI), `sqzhang@hkusua.hku.hk` (Shu-Qin ZHANG).

(transition probabilities) increases exponentially with respect to the number of categorical sequences. Because of this large number of parameters, people seldom use such kind of Markov chain models directly. In view of this, Ching et al. proposed a first-order multivariate Markov chain model in [10] for modeling the sales demands of multiple products in a soft drink company. Their model involves only $O(s^2m^2 + s^2)$ number of parameters where s is the number of sequences and m is the number of possible states. The model can capture both the intra- and inter-transition probabilities among the sequences. They also developed efficient estimation methods for solving the model parameters and applied the model to a problem of sales demand data sequences. A simplified multivariate Markov model based on [10] has also been proposed in [13] where the number of parameters is only $O(sm^2 + s^2)$. However, all the models [10,13] cannot capture negative correlations among the data sequences. We will elaborate the meaning of positive and negative correlations shortly in the next section. Moreover, all the models have been shown to be stationary and steady-state is assumed in the estimation of model parameters. Thus the estimation methods have to be modified when the given data sequences are short. In this paper, we propose a new model which can capture both positive and negative correlations among the data sequences. The model can also be adjusted to handle the case when the data sequences are short easily.

The rest of the paper is organized as follows. In Section 2, we propose the new multivariate Markov model and discuss some important properties of the model. In Section 3, we present the method for the estimation of model parameters. In Section 4, we apply the model and the method to some synthetic data and practical sales demand data. Finally, a summary is given in Section 5 to conclude the paper.

2 The New Multivariate Markov Chain Model

In this section, we first propose our new multivariate Markov chain model and then some of its properties. The properties are important for the estimation of model parameters.

The following multivariate Markov chain model has been proposed in [10]. The model assumes that there are s categorical sequences and each has m possible states in

$$M = \{1, 2, \ldots, m\}.$$

Here we adopt the following notations as in [10]. 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 we write

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

The following relationships among the sequences are assumed:

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

where

$$\lambda_{jk} \geq 0, \quad 1 \leq j, k \leq s \quad \text{and} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{for} \quad j = 1, 2, \ldots, s. \tag{2}$$

Equation (1) simply means that the state probability distribution of the jth chain (sequence) at time $(n+1)$ depends only on the weighted average of $P^{(jj)}\mathbf{x}_n^{(j)}$ and $P^{(jk)}\mathbf{x}_n^{(k)}$. Here $P^{(ij)}$ is the one-step transition probability matrix of the states from the jth sequence to the states of the ith sequence. In matrix form, one may write

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

We note that this model only allows positive correlation among the data sequences as all the λ_{ij} are assumed to be non-negative. This means an increase in a state probability in any one of the sequences at time n can only increase (but never decrease) the state probabilities at time $(n+1)$. To extend the model so as to take care of the negative correlations among the data sequences, our idea here is to use

$$\mathbf{z}_{n+1} = \frac{1}{m-1}(\mathbf{1} - \mathbf{x}_n)$$

to model the case when the state probability vector $\mathbf{x}_n$ is negatively correlated to the state probability vector $\mathbf{z}_{n+1}$. Here $\mathbf{1}$ is the vector of all ones and the factor $(m-1)^{-1}$ is the normalization constant and the number of possible states $m \geq 2$.

We propose the following model for s data sequences $\mathbf{x}_n^{(1)}, \mathbf{x}_n^{(2)}, \ldots, \mathbf{x}_n^{(s)}$. In order to reduce the number of parameters, in the proposed model, we assume that $P^{(ij)} = I$ when $i \neq j$. This idea has been used in [13] and has been shown to be effective. We then assume the following relationship among the data sequences:

$$\begin{pmatrix} \mathbf{x}_{n+1}^{(1)} \\ \mathbf{x}_{n+1}^{(2)} \\ \vdots \\ \mathbf{x}_{n+1}^{(s)} \end{pmatrix} = \underbrace{\Lambda^{+} \begin{pmatrix} \mathbf{x}_{n}^{(1)} \\ \mathbf{x}_{n}^{(2)} \\ \vdots \\ \mathbf{x}_{n}^{(s)} \end{pmatrix}}_{\text{Positive correlated part}} + \frac{1}{m-1} \underbrace{\Lambda^{-} \begin{pmatrix} \mathbf{1} - \mathbf{x}_{n}^{(1)} \\ \mathbf{1} - \mathbf{x}_{n}^{(2)} \\ \vdots \\ \mathbf{1} - \mathbf{x}_{n}^{(s)} \end{pmatrix}}_{\text{Negative correlated part}}$$

where

$$\Lambda^{+} = \begin{pmatrix} \lambda_{1,1}P^{(11)} & \lambda_{1,2}I & \cdots & \lambda_{1,s}I \\ \lambda_{2,1}I & \lambda_{2,2}P^{(22)} & \cdots & \lambda_{2,s}I \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{s,1}I & \cdots & \lambda_{s,s-1}I & \lambda_{s,s}P^{(ss)} \end{pmatrix}$$

and

$$\Lambda^{-} = \begin{pmatrix} \lambda_{1,-1}P^{(11)} & \lambda_{1,-2}I & \cdots & \lambda_{1,-s}I \\ \lambda_{2,-1}I & \lambda_{2,-2}P^{(22)} & \cdots & \lambda_{2,-s}I \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{s,-1}I & \cdots & \lambda_{s,-s+1}I & \lambda_{s,-s}P^{(ss)} \end{pmatrix}.$$

Here $\lambda_{i,j} \geq 0$ for $i = 1, 2, \ldots, s$ and $j = \pm 1, \ldots, \pm s$ and

$$\sum_{j=-s}^{s} \lambda_{i,j} = 1.$$

This is equivalent to

$$
\begin{aligned}
\mathbf{x}^{(n+1)} &= \begin{pmatrix} \mathbf{x}_{n+1}^{(1)} \\ \mathbf{x}_{n+1}^{(2)} \\ \vdots \\ \mathbf{x}_{n+1}^{(s)} \end{pmatrix} \\
&= \begin{pmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,s} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,s} \\ \cdots & \cdots & \cdots & \cdots \\ H_{s,1} & H_{s,2} & \cdots & H_{s,s} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{n}^{(1)} \\ \mathbf{x}_{n}^{(2)} \\ \vdots \\ \mathbf{x}_{n}^{(s)} \end{pmatrix} + \frac{1}{m-1} \begin{pmatrix} J_{1,-1} & J_{1,-2} & \cdots & J_{1,-s} \\ J_{2,-1} & J_{2,-2} & \cdots & J_{2,-s} \\ \cdots & \cdots & \cdots & \cdots \\ J_{s,-1} & \cdots & J_{s,-s+1} & J_{s,-s} \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} \\
&\equiv M_s \mathbf{x}^{(n)} + \mathbf{b}.
\end{aligned}
$$

Here

$$
H_{ij} = \begin{cases} (\lambda_{i,j} - \frac{\lambda_{i,-j}}{m-1})P^{(ii)} & \text{if } i = j \\ (\lambda_{i,j} - \frac{\lambda_{i,-j}}{m-1})I & \text{otherwise} \end{cases}
$$

and

$$
J_{ij} = \begin{cases} \lambda_{i,-j} P^{(ii)} & \text{if } i = j \\ \lambda_{i,-j} I & \text{otherwise.} \end{cases}
$$

We note that

$$
\begin{aligned}
\mathbf{x}^{(n+1)} &= M_s^2 \mathbf{x}^{(n-1)} + (I + M_s)\mathbf{b} \\
&= M_s^3 \mathbf{x}^{(n-2)} + (I + M_s + M_s^2)\mathbf{b} \\
&\;\;\vdots \\
&= M_s^{(n+1)} \mathbf{x}^{(0)} + \sum_{k=0}^{n} M_s^k \mathbf{b}
\end{aligned}
$$

where $M_s^0 = I$.

The model has a stationary distribution if for certain matrix norm $||.||$ (let us say $||M_s||_\infty$) we have $||M_s|| < 1$. Here given an $n \times n$ real matrix M,

$$
||M||_\infty = \max_i \left\{ \sum_{j=1}^{n} |M_{ij}| \right\}.
$$

In this case, we have

$$
\lim_{n\to\infty} \mathbf{x}^{(n)} = \lim_{n\to\infty} \sum_{k=0}^{n} M_s^k \mathbf{b} = (I - M_s)^{-1}\mathbf{b}.
$$

We also note that

$$
||M_s||_\infty \leq \max_{1\leq k\leq s} \left\{ m \left| \lambda_{k,k} - \frac{\lambda_{k,-k}}{(m-1)} \right| + \sum_{k\neq i} \left| \lambda_{k,i} - \frac{\lambda_{k,-i}}{(m-1)} \right| \right\}.
$$

Therefore the smaller $||M_s||_\infty$ is the faster the convergence rate of the process will be. Thus one can control the convergent rate by setting an upper bound $\alpha < 1$ by introducing the extra constraints

$$
m \left| \lambda_{k,k} - \frac{\lambda_{k,-k}}{(m-1)} \right| + \sum_{k\neq i} \left| \lambda_{k,i} - \frac{\lambda_{k,-i}}{(m-1)} \right| \leq \alpha \quad \text{for} \quad i = 1, 2, \ldots, s.
$$

3 Estimation of Model Parameters

In this section, we propose efficient methods for the estimations of $P^{(jj)}$ and λ_{jk}. For each data sequence, one can estimate the transition probability matrix by the following method [10,11]. Given a data sequence, one can count the transition frequencies from one arbitrary state to the other states. Hence we can construct the transition frequency matrix for the data sequence. After making a normalization, the estimates of the transition probability matrices can also be obtained. We note that one has to estimate $O(s \times m^2)$ transition frequency matrices for the multivariate Markov chain model. The vector stationary vector $\mathbf{x}$ can be estimated from proportion of the occurrence of each state in each of the sequences. According to the idea at the end of last section, if we take $||.||$ to be $||.||_1$ we can get the values of λ_{jk} by solving the following optimization problem ([10,11]):

$$\left\{\begin{array}{ll} & \min_{\lambda} \sum_i \left| \left[b_{j,k} - \hat{\mathbf{x}}^{(j)} \right]_i \right| \\ \text{subject to} & \\ & b_{j,k} = \sum_{k=1}^{s} \left((\lambda_{j,k} - \frac{\lambda_{j,-k}}{m-1}) \Delta_{jk} \hat{\mathbf{x}}^{(k)} + \frac{1}{m-1} \lambda_{j,k} \Delta_{jk} \mathbf{1} \right) \\ & \sum_{k=-s}^{s} \lambda_{jk} = 1, \quad \forall j = 1, 2, \ldots, s. \\ & \lambda_{jk} \geq 0, \quad \forall k = \pm 1, \ldots, \pm s, j = 1, 2, \ldots, s. \\ & m \left| \lambda_{k,k} - \frac{\lambda_{k,-k}}{(m-1)} \right| + \sum_{k \neq i} \left| \lambda_{k,i} - \frac{\lambda_{k,-i}}{(m-1)} \right| \leq \alpha \quad \text{for} \quad k = 1, 2, \ldots, s. \end{array}\right. \qquad (3)$$

Here $\mathbf{1}$ is the vector with all the entries being equal to one and

$$\Delta_{jk} = \begin{cases} P^{(jj)} & \text{if } j = k \\ I & \text{if } j \neq k. \end{cases}$$

Problem (3) can be formulated to s linear programming problems as follows, see for instance [3, (p. 221)]. We remark that other vector norms such as $||.||_2$ and $||.||_\infty$ can also be used but they have different characteristics. It is clear that the former will result in a quadratic programming problem while $||.||_\infty$ will still result in a linear programming problem, see for instance [3, (pp. 221-226)]. There is a well-known fact that in approximating data by using a linear function [3, (p. 220)], $||.||_1$ gives the most robust result. While $||.||_\infty$ avoids gross discrepancies with the data as much as possible. If the errors are known to be normally distributed then $||.||_2$ is the best choice. Finally we remark that the complexity of solving a linear programming problem or a quadratic programming problem is $O(n^3 L)$ where n is the number of variables and L is the number of binary bits needed to record all the data of the problem [4].

4 Numerical Examples

In this section, we present both synthetic data sequences and practical sales demand data sequences to demonstrate the effectiveness of the proposed model.

We first estimate all the transition probability matrices $P^{(jj)}$ and the parameters λ_{ij} by using the method proposed in Section 3. To evaluate the performance and effectiveness of the new multivariate Markov chain model, we use BIC (Bayesian Information Criterion) [5,7] as an indicator which is defined as

$$BIC = -2L + q \log n,$$

where

$$L = \sum_{j=1}^{s} \left(\sum_{i_0, k_1, \cdots k_s = 1}^{m} n_{i_0, k_1, \cdots, k_s}^{(j)} \log I \right),$$

$$I = \sum_{l=1}^{m}\sum_{k=1}^{s}(\lambda_{jk} - \frac{1}{m-1}\lambda_{j,-k})p_{i_0,k_l}^{(jk)} + \frac{1}{m-1}\lambda_{j,-k}$$

is the log-likelihood of the model,

$$n_{i_0,k_1,k_2,\cdots,k_s}^{(j)} = \sum x_{n+1}^{(j)}(i_0)x_n^1(k_1)x_n^2(k_2)\cdots x_n^s(k_s).$$

Here q is the number of independent parameters, and n is the length of the sequence. The less the value of BIC, the better the model is. For the sake of comparison, we also give the results for the model proposed by Ching et al. in [10].

4.1 Synthetic Data Sequences

In this subsection, we consider two binary sequences generated by the following model:

$$\begin{pmatrix} \mathbf{x}_{n+1}^{(1)} \\ \mathbf{x}_{n+1}^{(2)} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1}P^{(11)} & \lambda_{1,2}I \\ \lambda_{2,1}I & \lambda_{2,2}P^{(22)} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n^{(1)} \\ \mathbf{x}_n^{(2)} \end{pmatrix} + \begin{pmatrix} \lambda_{1,-1}P^{(11)} & \lambda_{1,-2}I \\ \lambda_{2,-1}I & \lambda_{2,-2}P^{(22)} \end{pmatrix} \begin{pmatrix} \mathbf{1}-\mathbf{x}_n^{(1)} \\ \mathbf{1}-\mathbf{x}_n^{(2)} \end{pmatrix}.$$

Here

$$P^{(11)} = \begin{pmatrix} 0.2 & 0.3 \\ 0.8 & 0.7 \end{pmatrix} \quad \text{and} \quad P^{(22)} = \begin{pmatrix} 0.4 & 0.1 \\ 0.6 & 0.9 \end{pmatrix}$$

and

$$\lambda_{1,1} = 0.4300,\ \lambda_{1,2} = 0.0505,\ \lambda_{1,-1} = 0.4401,\ \lambda_{1,-2} = 0.0794,$$
$$\lambda_{2,1} = 0.0838,\ \lambda_{2,2} = 0.3932,\ \lambda_{2,-1} = 0.1187,\ \lambda_{2,-2} = 0.4043.$$

Beginning with

$$\mathbf{x}_0^{(1)} = (0,1)^T \quad \text{and} \quad \mathbf{x}_0^{(2)} = (1,0)^T,$$

one can generate sequences of

$$\mathbf{x}_n^{(1)} \quad \text{and} \quad \mathbf{x}_n^{(2)}$$

as follows:

A :	21221122212222212122
B :	11121212112121122111

We then apply our to these data sequences and compare it with the model in [10]. The results are then reported in Table 1.

Table 1
The BIC for Synthetic Data Sequences.

Models	BIC
Multivariate Markov Model in [10]	107.93
The New Model ($\alpha = 1.0$)	4.2130e+003
The New Model ($\alpha = 0.9$)	4.2201e+003
The New Model ($\alpha = 0.8$)	4.2258e+003
The New Model ($\alpha = 0.7$)	4.2322e+003
The New Model ($\alpha = 0.6$)	4.2410e+003

4.2 Sales Demands Data Sequences

In this subsection, we present some numerical results based on the sales demand data of a soft-drink company in Hong Kong [10]. The soft-drink company actually facing an in-house problem of production planning and inventory control. A key issue is that the storage space of the central warehouse often finds itself full or near capacity.

Product are categorized into six possible states according to sales volume. 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. The company has a big customer and would like to predict sales demand for this customer so as to minimize the in-house inventory and at same time maximize the demand satisfaction for this customer. More importantly, the company can understand the sales pattern of this customer and then develop a marketing strategy to deal with this customer. 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. We then build the multivariate Markov chain model based on the data in the Appendix of [10].

Table 2
The BIC for Data sequences A,B,C,D,E.

Models	BIC
Multivariate Markov Model in [10]	8.0215e+003
The New Model ($\alpha = 1.0$)	4.1264e+003
The New Model ($\alpha = 0.9$)	4.1762e+003
The New Model ($\alpha = 0.8$)	4.2638e+003
The New Model ($\alpha = 0.7$)	4.3749e+003
The New Model ($\alpha = 0.6$)	4.5112e+003

The results above show the effectiveness of our simplified multivariate Markov model. One can see that the new multivariate Markov model is much better than the multivariate Markov model in [10] in fitting the sales demand data. We remark that when $||.||_\infty$ norm is used instead of $||.||_1$ norm, in the LP, we still get similar results of BIC for both the synthetic data and practical data.

5 Summary

In this paper, we propose a new multi-dimensional Markov chain model for modeling multiple categorical data sequences. We test the proposed model with both synthetic data and practical sales demand data. The new model can capture both positive and negative correlations among the data sequences. The model can also be adjusted to handle the case when the data sequences are short easily.

6 Appendix

Sales Demand Sequences of the Five Products (Taken from [10])

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 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 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 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 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.

References

[1] W. Ching. Markov modulated poisson processes for multi-location inventory problems. *Inter. J. Prod. Econ.*, 53:232–239, 97.

[2] W. Ching and M. Ng. *Markov Chains : Models, Algorithms and Applications*. Springer, 06.

[3] V. Chvátal. *Linear Programming*. Freeman, 83.

[4] S. Fang and S. Puthenpura. *Linear Optimization and Extension : Theory and Algorithms*. Prentice-Hall, 93.

[5] J. Kuha. Aic and bic comparisons of assumptions and performance. *Sociological Methods and Research*, 33:188–229, 04.

[6] I. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London, 97.

[7] A. Raftery. A model of higher-order markov chains. *J. Royal Statist. Soc.*, 47:528–539, 85.

[8] A. Raftery and S. Tavare. Estimation and modelling repeated patterns in high order markov chains with the mixture transition distribution model. *Appl. Statist.*, 43:179–199, 94.

[9] M. Ng T. Siu, W. Ching and E. Fung. On multivariate credibility approach for portfolio credit risk measurement. *Quantitative Finance*, 5:543–556, 05.

[10] E. Fung W. Ching and M. Ng. A multivariate markov chain model for categorical data sequences and its applications in demand prediction. *IMA J. Manag. Math.*, 13:187–199, 02.

[11] E. Fung W. Ching and M. Ng. A higher-order markov model for the newsboy's problem. *J. Operat. Res. Soc.*, 54:291–298, 03.

[12] E. Fung W. Ching and M. Ng. Higher-order markov chain models for categorical data sequences. *Inter. J. Nav. Res. Logist.*, 51:557–574, 04.

[13] S. Zhang W. Ching and M. Ng. On higher-dimensional markov chain models. *to appear in Pacific Journal of Optimization*, 06.