Ann Inst Stat Math (2010) 62:91–107
DOI 10.1007/s10463-009-0244-2

Finding market structure by sales count dynamics —Multivariate structural time series models with hierarchical structure for count data—

Nobuhiko Terui · Masataka Ban · Toshihiko Maki

Received: 27 June 2008 / Revised: 25 December 2008 / Published online: 28 July 2009

Abstract In this paper, we propose a multivariate time series model for sales count data. Based on the fact that setting an independent Poisson distribution to each brand's sales produces the Poisson distribution for their total number, characterized as market sales, and then, conditional on market sales, the brand sales follow a multinomial distribution, we first extend this Poisson–multinomial modeling to a dynamic model in terms of a generalized linear model. We further extend the model to contain nesting hierarchical structures in order to apply it to find the market structure in the field of marketing. As an application using point of sales time series in a store, we compare several possible hypotheses on market structure and choose the most plausible structure by using several model selection criteria, including in-sample fit, out-of-sample forecasting errors, and information criterion.

Keywords Count data · Generalized linear model · Hierarchical market structure · MCMC · Poisson–multinomial distribution · Predictive density · POS time series

1 Introduction

Harvey and Fernandes (1989) proposed a univariate structural time series model for count data—non-negative integer—to follow Poisson distribution, and Ord et al.

Terui acknowledges the financial support of by the Japanese Ministry of Education Scientific Research Grants (C)18530152. The authors thank the associate editor and three anonymous reviewers for their helpful comments.

N. Terui (✉) · T. Maki
Graduate School of Economics and Management, Tohoku University,
Sendai 980-8576, Japan
e-mail: terui@econ.tohoku.ac.jp

M. Ban
Department of Management, Mejiro University, Nakaochiai, Shinjuku-ku 161-8539, Japan

(1993) extended it to a multivariate time series model. They both applied a forecasting method based on the exponentially weighted moving average with maximum likelihood estimation. On the other hand, West et al. (1985) and Cargnoni et al. (1997) developed time series models for variates following multinomial distribution by introducing dynamic linear models in terms of the Bayesian approach.

On the other hand, the stochastic models for count data have an interesting distributional property that the sum of independent Poisson variables produces the Poisson distribution and then, conditional on the sum, the Poisson variables follow a multinomial distribution. In this paper, we propose a multivariate time series model based on this Poisson–multinomial relationship in terms of a dynamic generalized linear model.

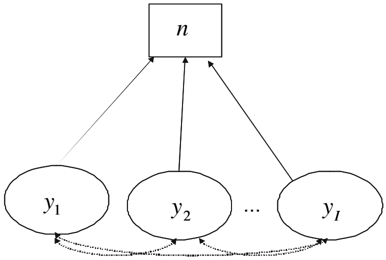
The model is directly built on the natural distributional assumption for discrete outcomes of amount of sales, and it describes the possible market structures with nesting hierarchical submarkets. Setting a Poisson distribution independently to each brand's sales produces the Poisson distribution for their total number, characterized as market sales, and then, conditional on the market sales, we obtain multinomial distribution for brand sales. Then the competitive relations of brand sales inside market are expressed by the correlation of multinomial variables. Extending this Poisson–multinomial modeling to a dynamic model in terms of a generalized linear model, we specify the market structure based on the brand's sales dynamics. We use "market structure" in terms of which we can classify the brands into several submarkets, so that the brands are competitive inside a submarket but not competitive outside it. We also forecast the several levels of sales amounts, that is, brand sales, submarket sales and market sales.

Our modeling contributes to the literature in two ways. The first contribution is through the modeling. West et al. (1985) proposed a dynamic model with multinomial distribution, which forecasts the amount of brand sales relative to competitive brands, i.e., market share among brands, because it models multinomial outcomes, when the number of total counts is given as constant. Cargnoni et al. (1997) deal with several sets of multinomial distributions in the same way; however, they share the properties with West et al. (1985).

On the other hand, our proposed model makes it possible to accommodate not only the market share but also the total amount of brand sales in the market or submarket. That is, it describes not only the competitive relations between brands but also the market and category expansions. Moreover, assuming the existence of different submarkets in the market, we jointly deal with several Poisson–multinomial distributions by means of a nesting hierarchical structure. In this paper, we deal with a three-layer hierarchical structure for the subjects focusing on marketing applications to our dataset; however, our method can be extended to higher-order layers structures. As for the model estimation, compared with Cargnoni et al. (1997), our modeling utilizes an updated MCMC algorithm for computational efficiency to estimate the model and constitute predictive density.

The second contribution is to marketing literature. Finding a particular structure of market is an important step for marketers to consider not only when they try to start up a new business, but also when they face the need to modify their strategy in the existing markets. As shown in Lilien et al. (1992), Kamakura and Wedel (2000), Hanssen et al. (2001) and Lilien and Rangaswamy (2003), numerous models have been developed to define the market structure in the literature, for example,

Fig. 1 Market with no specific structure



Urban et al. (1984), Grover and Dillon (1985), Grover and Srinivasan (1987) and Colombo and Morrison (1989). The standard methods contain the perceptual map by factor analysis of consumer survey data with ratings on the brand evaluations at a specified time point, the segmentation by the clustering and mixture models of survey or panel data, and the use of econometric models to estimate time invariant elasticity across brands.

Most of previous methods typically use static models; however, every market should continue to change over time, and thus it is desirable to incorporate dynamics in the method for finding a market structure. Besides, the market is typically expressed as a tree structure with several orders of hierarchy to define the submarkets. The preceding models do not always reflect these hierarchies. Thus, the model needs to accommodate hierarchical structures together with dynamics. The proposed model utilizes dynamics of sales count time series to classify the brands into homogeneous submarkets in nesting hierarchical structures. In addition, we can conduct sales forecasting at any level (i.e., brand, submarket, and market) by extending the time horizon of model to the future.

2 Models for defining market structure

We assume that there are I brands in the market and denote y_{it} by the amount of sales for the brand i at time t $(t = 1, \ldots, T)$ and y_{it} follows the Poisson distribution independently with a time varying parameter λ_{it} for $i = 1, \ldots, I$. Then we obtain the Poisson distribution for market sales, defined as the aggregate of brand sales, $n_t = \sum_{i=1}^{I} y_{it}$ under the assumption that there is no specific structure between brands in the market, as shown in Fig. 1.

That is, we have marginal distributions,

$$y_{it} \sim \text{Poisson}\,(\lambda_{it}), \quad n_t \sim \text{Poisson}\left(\sum_{i=1}^{I} \lambda_{it}\right). \tag{1}$$

Using the distributional property that when total number n_t is given, the conditional distribution of a respective brand's sales $\tilde{y}_t = \{y_{it}, i = 1, \ldots, I\}$ is derived as the multinomial distribution with parameters $(n_t, \{\pi_{it} = \lambda_{it}/\sum_{it}^{I} \lambda_i, i = 2, \ldots, I\})$,

$$\tilde{y}_t \,|n_t \sim \text{Multinomial}\,(y_{it}\,|n_t, \pi_{it}). \tag{2}$$

The sequential use of (1) and (2) produces the joint distribution for market and brand sales,

$$p\,(n_t, \tilde{y}_t) = p\,(n_t)\,p\,(\tilde{y}_t|n_t). \tag{3}$$

As is common in marketing, we assume that there are submarkets, implying that the brands inside each submarket are correlated with each other in the sense that they are engaged in competition; on the other hand, no substantial correlations are expected between the brands outside their submarkets because their competitive relations should be quite weak.

Extending the framework of the Poisson–multinomial relation in the above manner, we decompose the market structure into L submarkets M_k so that $\tilde{y}_t = \bigcup_{k=1}^{L} \tilde{y}_{it}^{[k]}$, where $\tilde{y}_t^{[k]} = (y_{jt},\ j \in M_k)'$ means the vector of the brands that are grouped in the $M_k, k = 1, \ldots, L$. Then the aggregated sales at submarket level, $m_t^{[k]} = \sum_{i \in M_k} \tilde{y}_{it}^{[i]}$, $k = 1, \ldots, L$, follow independent Poisson distributions.

Conditional on $m_t^{[k]}$, each of $\tilde{y}_t^{[k]}|m_t^{[k]}, k = 1, \ldots, L$, respectively, follow a multinomial distribution, and $\tilde{y}_t^{[k]}$'s are orthogonal each other, i.e.,$(\tilde{y}_t^{[l]}|m_t^{[l]}) \perp (\tilde{y}_t^{[k]}|m_t^{[k]})$ for $l \neq k$. From the definition of submarket, the brands outside submarkets should not be competitive with each other, and this is consistent with the orthogonal property.

Next, let $\tilde{m}_t = (m_t^{[1]}, \ldots, m_t^{[L]})'$ denote the L dimensional vector of submarket sales. Then, $\tilde{m}_t|n_t$ follows a multinomial distribution conditional on the sum of submarket sales, i.e. market sales $n_t = \sum_{k=1}^{L} m_t^{[k]}$. In all, we have the three layer hierarchical market structure model. We note that $\{\tilde{y}_t\}$ and $\{n_t\}$ are independent conditional on $\{\tilde{m}_t\}$.

Figure 2 illustrates the structure of this model. At the bottom layer, brands sales follow independent Poisson distribution as an unconditional distribution. They are aggregated up to submarket level at the middle layer, and further aggregated to market level of top layer. Both kinds of aggregated sales counts unconditionally follow

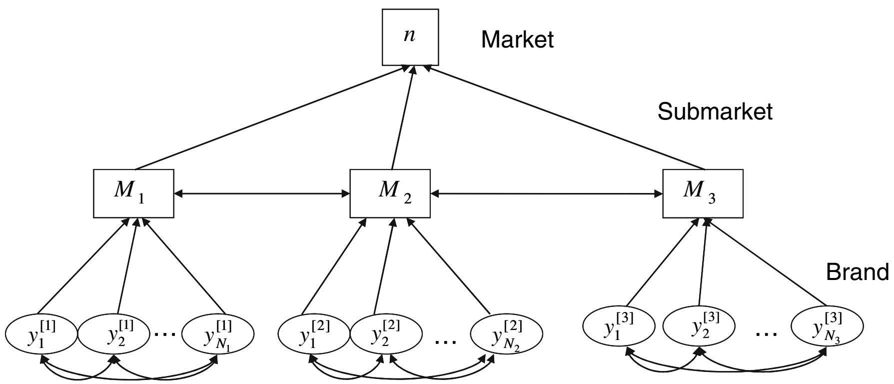


Fig. 2 Three-layer hierarchical market structure

Poisson distributions with corresponding parameters. In the case of bottom layer, conditional on the submarket sales, the brands under each submarket are correlated through multinomial distribution. Then, by aggregating three submarket sales at the market level, conditional on the market sales, the three submarket sales follow a multinomial distribution, which represents a competitive relation between submarkets through correlation between multinomial variables.

In the following, we first formally define the data distribution for variables in the model and specify dynamic structure of generalized linear models. Then we express the induced likelihood function, prior and posterior densities for Bayesian MCMC inference.

Data distribution We consider the three-layer model with L number of submarket $M_k, k = 1, \ldots, L$. Suppose that each submarket M_k is composed of N_k brands, then the total number of brands in the market can be connected by the sum of the number of brands attributed to each submarket $I = \sum_{j=1}^{L} N_j$.

In terms of these notations, the joint density function of I brand sales (bottom layer), L submarket sales (middle layer), and market sales (top layer) are decomposed into

$$\begin{aligned} p(n_t, \tilde{m}_t, \tilde{y}_t) &= p(n_t)\, p(\tilde{m}_t|n_t)\, p(\tilde{y}_t|\tilde{m}_t) \\ &= p(n_t) \prod_{k=1}^{L} \left[p\left(m_t^{[k]}|n_t\right) \left\{ \prod_{j \in M_k} p\left(y_{jt}|m_t^{[k]}\right) \right\} \right] \\ &= p(n_t) \prod_{k=1}^{L} \left[p\left(m_t^{[k]}|n_t\right) p\left(\tilde{y}_t^{[k]}|m_t^{[k]}\right) \right], \end{aligned} \tag{4}$$

where $\tilde{m}_t = (m_t^{[1]}, \ldots, m_t^{[L]})'$ and $\tilde{y}_t^{[k]} = \{y_{jt},\ j \in M_k\}$. We then obtain the marginal and conditional data distributions:

$$\begin{cases} n_t \sim \text{Poisson}\left(\sum_{it}^{I} \lambda_{it}\right) \\ \tilde{m}_t|n_t \sim \text{Multinomial}\left(n_t, \{\pi_{jt}\}\right) \\ \tilde{y}_t^{[k]}|m_t^{[k]} \sim \text{Multinomial}\left(m_t^{[k]}, \left\{\pi_{jt}^{[k]}\right\}\right), \quad k = 1, \ldots, L, \end{cases} \tag{5}$$

where $\pi_{it} = \sum_{j=1}^{N_i} \lambda_{jt}^{[i]} / \sum_{k=1}^{L} \sum_{j=1}^{N_k} \lambda_{jt}^{[k]}, i = 1, \ldots, L-1$, and $\lambda_{jt}^{[k]}$ is the parameter of variable classified to the submarket M_k, and $\pi_{jt}^{[k]} = \lambda_{jt}^{[k]} / \sum_{j=1}^{N_k} \lambda_{jt}^{[k]}$, $j = 1, \ldots, N_k - 1$.

That is, we have a Poisson distribution for market sales $\{n_t\}$ and L multinomial distributions, i.e., a multinomial distribution for submarkets $\tilde{m}_t = (m_t^{[1]}, \ldots, m_t^{[L]})'$ conditional on n_t, and L multinomial distributions for brand sales $\{y_{jt},\ j \in M_k\}$ when the submarket sales $\{m_t^{[k]}\}$ are given.

Next, we set a generalized linear model to have a structural equation and system equation for the parameters in each layer distribution. Below, we define the structure for three kinds of data distribution for 1:$p(n_t)$, 2 : $p(\tilde{m}_t|n_t)$, and 3: $p(\tilde{y}_t|\tilde{m}_t)$.

Structural equation We define the structural equation for the parameters of each layer's data distribution, which is denoted using subscript $s = 1, 2$, and 3, as

$$\tilde{\eta}_t^s = F_t^s \tilde{\theta}_t^s + \tilde{v}_t^s; \quad \tilde{v}_t^s \sim N\left(0, V^s\right), \tag{6}$$

where we assume that V^s's are uncorrelated each other. $\tilde{\eta}_t^s = \{\eta_{it}^s\}$ are linked with the parameters of data distribution;
$s = 1$: $\{n_t\}$

$$\tilde{\eta}_t^1 \equiv \eta_t = \log\left(\sum_{i=1}^{I} \lambda_{it}\right) \quad \text{and} \quad \sum_{i=1}^{I} \lambda_{it} \equiv \bar{\lambda}_t = \exp\left(\eta_t\right), \tag{7}$$

$s = 2$: $\tilde{m}_t = (m_t^{[1]}, \ldots, m_t^{[L]})'$.

The ith element of $\tilde{\eta}_t^2$, $\eta_{it}^2 = f(\pi_{it}^2)$, is connected to the multinomial parameters π_{it} by the relation

$$\pi_{it} = \exp\left(\eta_{it}\right)/(\exp\left(\eta_{1t}\right) + \cdots + \exp\left(\eta_{L-1t}\right) + 1) \quad \text{for } i = 1, \ldots, L-1, \tag{8}$$

$s = 3$: $\tilde{y}_t^{[k]}$ for $k = 1, \ldots, L-1$.

The multinomial parameter inside the submarket M_k is related to the link parameter as

$$\pi_{it}^{[k]} = \exp\left(\eta_{it}^{[k]}\right) \Big/ \left(\exp\left(\eta_{1t}^{[k]}\right) + \cdots + \exp\left(\eta_{N_k-1t}^{[k]}\right) + 1\right) \text{ for } i = 1, \ldots, N_k - 1. \tag{9}$$

The structural equation (6) is formulated by

$$\eta_{it}^s = \mu_{it}^s + X_{it}^{s\prime} \tilde{\beta}_{it}^s + \nu_{it}^s \quad \text{for } i = 1, \ldots, P, \tag{10}$$

where P refers to the number of independent parameters for each layer's distribution. That is, $P = 1$ for $s = 1$, $P = L - 1$ for $s = 2$, and $P = N_j - 1$ for $s = 3$. X_{it}^s means the vector of exogeneous marketing mix variables, and μ_{it}^s represents the trend term with local trend slope κ_{it}^s and $\tilde{\beta}_{it}^s$ is the response parameter vector.

System equation We define the dynamics on these parameters as

$$\begin{cases} \mu_{it}^s = \mu_{it-1}^s + \kappa_{it-1}^s + w_{1it}^s \\ \kappa_{it}^s = \kappa_{it-1}^s + w_{2it}^s \\ \tilde{\beta}_{it}^s = \tilde{\beta}_{it-1}^s + \tilde{w}_{3it}^s. \end{cases} \tag{11}$$

Then, the state vector $\tilde{\theta}_t^s$ is composed of a time varying trend $(\tilde{\mu}_t^{s\prime}, \tilde{\kappa}_t^{s\prime}) = (\{\mu_{it}^s\}, \{\kappa_{it}^s\})$ and market response parameter vectors $\tilde{\beta}_t^{s\prime} = \{\beta_{it}^s\}$, denoted as $\theta_t^s = (\tilde{\mu}_t^{s\prime}, \tilde{\kappa}_t^{s\prime}, \tilde{\beta}_t^{s\prime})'$. F_t^s of (6) is constituted from an explanatory variable matrix of constant and marketing mix variables. We set $\tilde{w}_t^s = (w_{1t}^{s\prime}, w_{2t}^{s\prime}, \tilde{w}_{3t}^{s\prime})'$; then model (11) constitutes the system equation with constant matrix H_t^s

$$\tilde{\theta}_t^s = H_t^s \tilde{\theta}_{t-1}^s + \tilde{w}_t^s; \quad \tilde{w}_t^s \sim N(0, W^s). \tag{12}$$

We note that, from the Bayesian viewpoint, this model is composed of two-stage hierarchical models; that is, (6) represents the structure of the first stage prior on the parameter $\tilde{\eta}_t^s$ with hyper-parameter $\tilde{\theta}_t^s$, and (12) implies the second stage prior for $\tilde{\theta}_t^s$. In all, our model has non-Gaussian likelihood and two-stage hierarchical Gaussian prior distributions on the parameters.

Likelihood function, prior and posterior densities The likelihood function is expressed from data distributions [1] for $t = 1, \ldots, T$ as

$$\begin{aligned} &p(\{n_t\}, \{\tilde{m}_t\}, \{\tilde{y}_t\} | \{\tilde{\eta}_t^1\}, \{\tilde{\eta}_t^2\}, \{\tilde{\eta}_t^3\}) \\ &\quad = \prod_{t=1}^{T} p(n_t, \tilde{m}_t, \tilde{y}_t | \tilde{\eta}_t^1, \tilde{\eta}_t^2, \tilde{\eta}_t^3) \\ &\quad = \prod_{t=1}^{T} p(n_t | \tilde{\eta}_t^1) p(\tilde{m}_t | n_t, \tilde{\eta}_t^2) p(\tilde{y}_t | \tilde{m}_t, \tilde{\eta}_t^3). \end{aligned} \tag{13}$$

Assuming the prior density for covariance matrix of structural and system equations for each layer $p(V^s, W^s) = p(V^s)p(W^s)$ to be independent of other parameters, we can express the prior distribution from (6) and (12) as

$$\begin{aligned} &p(\{\tilde{\eta}_t^s\}, \{\tilde{\theta}_t^s\}, V^s, W^s, \quad s = 1, 2, 3) \\ &\quad = \prod_{s=1}^{3} p(\{\tilde{\eta}_t^s\}, \{\tilde{\theta}_t^s\}, V^s, W^s) \\ &\quad = \prod_{s=1}^{3} \left\{ \prod_{t=1}^{T} p(\tilde{\eta}_t^s | X_t^s, \tilde{\theta}_t^s, V^s) p(\tilde{\theta}_t^s | \tilde{\theta}_{t-1}^s, W^s) p(V^s) p(W^s) \right\}. \end{aligned} \tag{14}$$

Finally, combined with the likelihood (13) and the prior (14), we obtain the posterior density for our model

$$\begin{aligned} &p(\{\tilde{\eta}_t^s\}, \{\tilde{\theta}_t^s\}, V^s, W^s | \{n_t\}, \{\tilde{m}_t\}, \{\tilde{y}_t\}, \{X_t\}) \\ &\quad = \prod_{s=1}^{3} \left\{ \prod_{t=1}^{T} p(z_t^s | \tilde{\eta}_t^s) p(\tilde{\eta}_t^s | X_t^s, \tilde{\theta}_t^s, V^s) p(\tilde{\theta}_t^s | \tilde{\theta}_{t-1}^s, W^s) p(V^s) p(W^s) \right\}, \end{aligned} \tag{15}$$

where

$$p\left(z_t^s|\tilde{\eta}_t^s\right) = \begin{cases} p\left(n_t|\tilde{\eta}_t^s\right) & \text{if } s = 1 \\ p\left(\tilde{m}_t|n_t, \tilde{\eta}_t^s\right) & \text{if } s = 2 \\ p\left(\tilde{y}_t|\tilde{m}_t, \tilde{\eta}_t^s\right) & \text{if } s = 3. \end{cases} \tag{16}$$

3 Estimation and forecasting

In addition to the standard Bayesian inference on state space modeling by DLM (dynamic linear models) by West and Harrison (1989), we employ the MCMC approach to estimate the model, in particular, using Metropolis–Hastings sampling for the conditional posterior density of link functions

$$p(\tilde{\eta}_t^s|X_t^s, \tilde{\theta}_t^s, V^s, \{\tilde{y}_t\}) \propto p(z_t^s|\tilde{\eta}_t^s)p(\tilde{\eta}_t^s|X_t^s, \tilde{\theta}_t^s, V^s). \tag{17}$$

Once the value of $\tilde{\eta}_t^{s(k)}$ is given, the structural equation (6) coupled with the system equation (12) constitute a conventional Gaussian state space model. The multi-move sampler by Carter and Kohn (1994) is employed to sample the state vectors. We assume that the initial values of state vector $\tilde{\theta}_0^s$ follows a multivariate normal distribution $N(\bar{\theta}_0^s, d \times I)$. The mean vector $\bar{\theta}_0^s$ was set as the estimate of coefficient on the static regression, that is, the regression with time invariant coefficient, and we set $d = 0.1$ for empirical application.

As for forecasting, we evaluate the predictive density $p(z_{T+1}^s|\text{data})$ of one-step- ahead forecast, which is defined as

$$\int p(z_{T+1}^s|\, \tilde{\eta}_{T+1}^s, \text{data})\, p(\tilde{\eta}_{T+1}^s\, |\tilde{\theta}_{T+1}^s, V^s, \text{data})$$
$$p(\tilde{\theta}_{T+1}^s|\tilde{\theta}_T^s, W^s, \text{data})P(V^s)P(W^s)\mathrm{d}\tilde{\theta}_T^s\mathrm{d}V^s\mathrm{d}W^s, \tag{18}$$

where "data" means the observed data $(\{\tilde{y}_t\}, \{X_t\})$. The computationally efficient Monte–Carlo integration can be applied to evaluate this predictive density. The details are described in the Appendix. Extending the one-step-ahead prediction, we obtain the joint predictive density $p(z_{T+1}^s, z_{T+2}^s, \ldots, z_{T+h}^s|\text{data})$.

As an application, we can make an inference on future events. For example, if we are interested in the brand level relationship between brand 1 and 2 of submarket M_k in the h step ahead, we can evaluate the posterior probability

$$\Pr\{y_{1,T+h}^{[k]} > y_{2,T+h}^{[k]}\,|\text{data}\}$$
$$= \iiint_{\Omega=\{y_{1,T+h}^{[k]} > y_{2,T+h}^{[k]}\}} p(\tilde{y}_{T+1}^{[k]}, \tilde{y}_{T+2}^{[k]}, \ldots, \tilde{y}_{T+h}^{[k]}|\text{data})\mathrm{d}\tilde{y}_{T+1}^{[k]}$$
$$\times\, \mathrm{d}\tilde{y}_{T+2}^{[k]}, \ldots, \mathrm{d}\tilde{y}_{T+h}^{[k]}. \tag{19}$$

in the process of evaluating joint predictive density as the by-products.

Table 1 Summary of data

Brand	Average weekly sales	Average weekly price (/100g)	Average weekly display	Average weekly feature
A1	33.74	90.04	2.09	0.40
B1	74.11	90.33	2.10	0.40
C1	51.29	89.59	2.09	0.40
A2	61.54	87.29	1.52	0.23
B2	87.00	83.71	2.73	0.23
C2	49.67	80.62	3.39	0.20
A3	27.22	101.22	0.69	0.19
B3	48.13	99.31	0.73	0.19
C3	24.42	99.03	0.72	0.18

4 Analysis of POS data

Data and variables The store level scanner—point of sales (POS)—time series in the curry roux category are applied to our model. The weekly series comprises three makers that produce three brands each, a total of nine brands during 110 weeks. The first 100 weeks are used for estimation and the last 10 weeks are reserved for validation of forecasting. The data contain the amount of brand sales for $\{y_{it}\}$, and "prices", display (in-store promotion)" and "features (advertising in newspaper)" for marketing mix variables $\{X_{it}\}$. The display and feature are binary data taking one when it was on, and zero when it was off. Table 1 displays the summary statistics of these variables.

As for the explanatory variable X_{it}^s in the structural equation (10) for the case of $s = 3$: brand sales, in addition to the second-order stochastic trend term, we use its own price, displays, and features as well as those of all competitive brands inside the corresponding submarket. In case of the $s = 2$: submarket sales, its average price and number of displays and features aggregated over the brands in its own submarket, and those of competitive submarkets, are used. X_{it}^s contains a second-order stochastic trend with no explanatory variables for $s = 1$: market sales. Hence, the matrix F_t^s in the structural equation is composed of known constants constructed by these variables. The matrix H_t^s of system equation (12) is also defined by known constants from the relationship (11).

Each of three makers, A, B, and C, produces three categories of products according to the level of spiciness to accommodate the difference in consumer tastes (product category 1: Not spicy, 2: Medium spicy, 3: Spicy). Thus, we can easily consider that these brands are competitive between product categories as well as between makers, and thus we employ the hypotheses of two kinds of market structure induced by the competitive relations between product categories and between makers. In addition to these, after looking at the data, we can classify the brands by the usage of the product, that is, "ordinary" or "luxury" use. In fact, as shown in Table 1, the brands produced by makers 1 and 2 are not considered too expensive for frequent ordinary use. On the other hand, maker 3 is selling products targeted toward luxury use. In fact, the

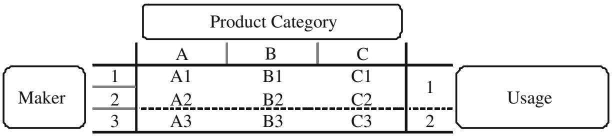


Fig. 3 Market structures—instant curry data

Table 2 Model specification

	ML	DIC	RMSE(sum1)	RMSE(sum2)
Product category	150558	–300810	648.957	480.061
Maker	150575	–300880	858.244	515.756
Usage	150593	–300995	621.130	449.510
Null	150453	–300620	792.709	792.709

RMSE

Product Category		Maker		Usage		Null	
Market	164.356	Market	138.734	Market	171.340	Market	222.646
Category A	65.709	Maker 1	133.650	Usage 1	125.810	A1	47.522
A1	45.745	A1	49.735	A1	38.740	A2	31.213
A2	27.636	B1	62.984	B1	75.440	A3	76.933
A3	9.661	C1	37.204	C1	42.160	B1	105.188
Category B	67.773	Maker 2	141.381	A2	19.210	B2	78.920
B1	71.431	A2	52.435	B2	33.270	B3	83.774
B2	47.722	B2	59.212	C2	19.280	C1	54.631
B3	35.981	C2	31.075			C2	51.565
Category C	35.414	Maker 3	67.457	Usage 2	45.810	C3	40.319
C1	44.846	A3	22.500	A3	17.190		
C2	19.005	B3	39.224	B3	23.710		
C3	13.678	C3	22.653	C3	9.170		
(sum1)	648.957	(sum1)	858.244	(sum1)	621.130	(sum1)	792.709
(sum2)	480.061	(sum2)	515.756	(sum2)	449.510	(sum2)	792.709

brands of maker 3 have higher prices and fewer promotions than those of the other makers. We refer to this criterion as "usage" to define the market. In all, we have three possible market structures: (1) product category, (2) makers, and (3) usage, as shown in Fig. 3.

Model specification The top of Table 2 shows the log of marginal likelihood (ML) as an in-sample fit criterion, and two kinds of predictive measure, i.e. the deviance information criteria (DIC) by Spiegelhalter et al. (2002) and the root mean squared errors (RMSE) of 10-step-ahead forecasts of hold-out samples as out-of-sample criteria.

Based on expected deviance as a measure of predictive accuracy for Bayes modeling, DIC for data y and parameter θ is defined by,

$$\mathrm{DIC} = 2\frac{1}{M}\sum_{l=1}^{M} D(y, \theta^{(m)}) - D(y, \tilde{\theta}), \tag{20}$$

where $D(y, \tilde{\theta}) = -2 \log p(y|\theta)$ and $p(y|\theta)$ means the likelihood of (13), and $\theta^{(m)}$ is the mth draw of MCMC and $\tilde{\theta}$ is posterior mean through M times iterations.

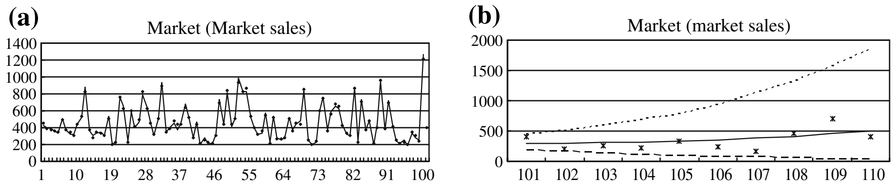


Fig. 4 In-sample performance and forecasting—market **a** in-sample fit, **b** forecasting

As for the RMSE, there are three levels for forecasting error in this case: market, submarket, and brand. These errors including null model of "no" structure are reported in the lower panel of the table. The choice of the best hypothesis of market structure depends on which errors are to be considered; we employ two kinds of measure, "sum1" and "sum2". "sum1" is the sum of all of errors induced by the model in the sense that we have no specific preference on the levels to be predicted. On the other hand, "sum2" is defined to be the sum of market and brand errors by considering that the number of submarket are different between (1), (2) and (3), where the number of errors is the same across the hypotheses.

According to the three criteria, the market structure defined by "usage" is supported. In fact, there is no great difference in the values for ML and DIC; however, RMSE is more evident.

The left panel of Fig. 4 shows the predicted fit of in-sample market sales data, where each observation is denoted by a dot and the estimates are connected by straight lines. We observe that the model fits the market sales quite well over the observational period. The right panel depicts its ten-step-ahead forecasting of market sales, where the mean values of predicted density at each prediction step are connected by a continuous line, and 2.5 and 97.5% quantiles of the density at each step are connected by dashed lines. The hold-out samples are also denoted by the dots in the figure. This shows that the market will gradually expand over the next 10 weeks, and these forecasts are consistent with the movement of hold-out samples. We generate the forecasts keeping the last observation X_{iT}^{s} for the prediction steps.

Figure 5 conveys the same information for submarket and brand sales. For saving spaces, we take up only a few pictures. We observe that the model performs quite well not only for in-sample but also for out-of-sample criteria at these levels.

Figure 6a indicates the trend for the market sale and suggests that the market sale has a little downward trend after 70th week with a cyclical movement. Figure 6b and c depicts time varying coefficient estimates for submarket 1 and one of the brand in the submarket, respectively. The Fig. 6b shows that the feature promotions of both submarkets help to increase of the sales of "usage 1" submarket because the response parameters of feature promotions of both submarket keep taking positive values on the mean parameter of usage 1 submarket. The Fig. 6c suggests that the price cut of B1 promotes the increase of brand A1's sales because the price coefficients of B1 on the A1's mean parameter are negative throughout sample. That is, by considering negative correlation between A1 and B1, we find unexpected relation so that B1 is not hostile to A1 in the sense of pricing strategy. However, other brand's pricing works competitive to A1.

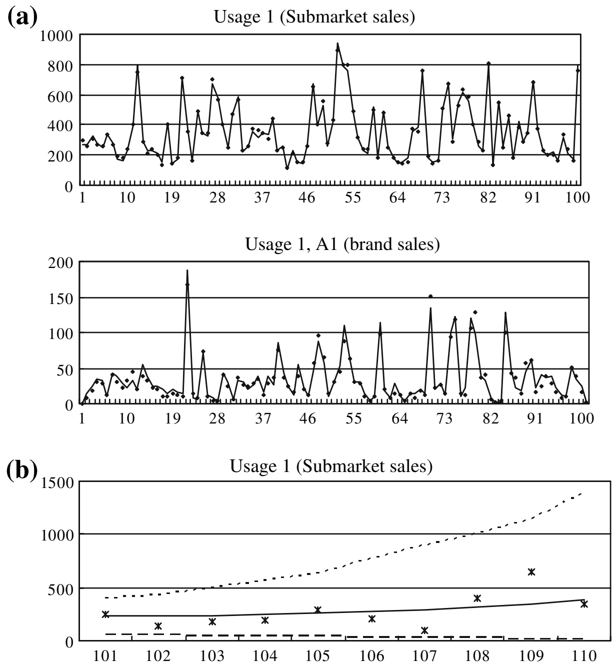


Fig. 5 In-sample performance and forecasting—submarket and brand **a** in-sample fit, **b** forecasting

Forecasting future events Finally, we forecast future events regarding the competitive relation between submarkets and brands. Figure 7a shows the posterior probability of submarket "usage 1" having larger sales than those of "usage 2," $\Pr\{M_{1T+s}(=\sum_{j=1}^{N_1} y_{jT+s}^{[1]}) > M_{2T+s}(=\sum_{j=1}^{N_2} y_{jT+s}^{[2]})|\text{data}\}$ for $s = 1, \ldots, 10$. The figure suggests that the submarket "usage 1" maintains larger sales than "usage 2" with high probability, more than 0.825, over the next 10 weeks; however, its probability is gradually decreasing.

Figure 7b indicates the future competitive relations between brands in "usage 2" submarket, where we evaluate the posterior probability of $\Pr\{y_{jT+s}^{[2]} > y_{kT+s}^{[2]}|\text{data}\}$ for the brand A3, B3 and C3. It shows that there will be still higher probability of the event "B3 > C3"; however, C3 will be getting the share slightly from B3. On the other hand, the relations "A3 > C3" and "A3 > B3" will be relatively stable. The picture also suggests that the probabilities of "B3 > C3" and "A3 > C3" will fall a bit at three and six weeks ahead, although this change is not so significant.

5 Concluding remarks

In this paper, we proposed a multivariate time series model for discrete outcomes to find the market structure between brand sales and to forecast the amount of brand sales as well as their submarket and market sales. The model was directly built on the natural distributional assumption for discrete outcomes of amount of sales. We

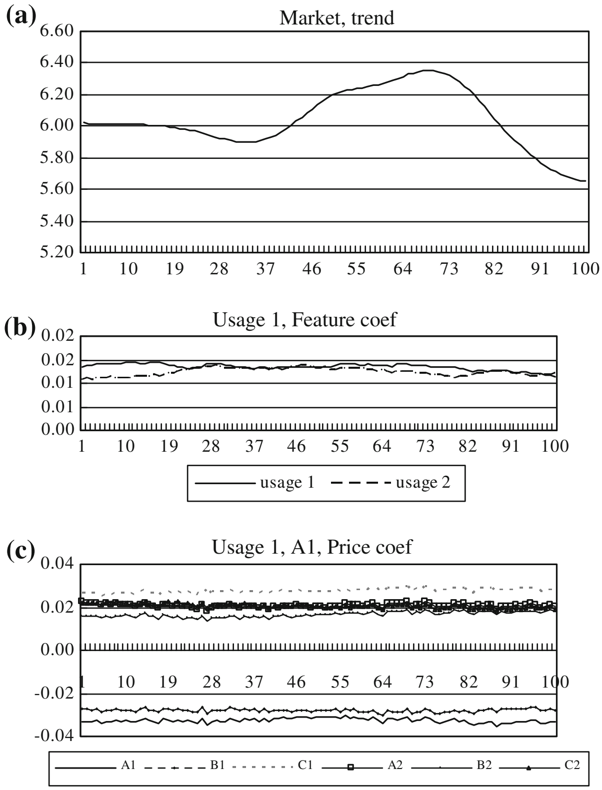


Fig. 6 Trend and time varying market response **a** trend for market sales, **b** time varying coefficient of feature for submarket "Usage 1", **c** time varying coefficient of price on A1

extended Poisson–multinomial modeling to a dynamic model in terms of a dynamic generalized linear model. We employed the Bayesian MCMC approach to estimate the model and constitute predictive density. We showed that a higher-order layers model could be useful to find market structure.

We proposed a three-layer hierarchical structure model with "flat" submarkets, which means that the brands are classified into submarkets so that they are competitive at the same level. However, the model can be easily extended to express more complex hierarchical structures, for example, irregularly bifurcated to describe an "umbrella" structure known in marketing.

There are a couple of problems for future research. One is the extension of the theoretical study. The model could have an over-dispersion of in-sample as well as out-of-sample predictions, which stems from the fact that Poisson distribution has a variance identical with the mean parameter by nature. This means that the variance of predicted sales gets larger whenever the mean level is higher. For this over-dispersion problem, it is shown in, for example, McCullagh and Nelder (1990) that compound Poisson distribution can be employed, whereas several Poisson distributions are mixed by the gamma density so that the induced compound Poisson distribution has different

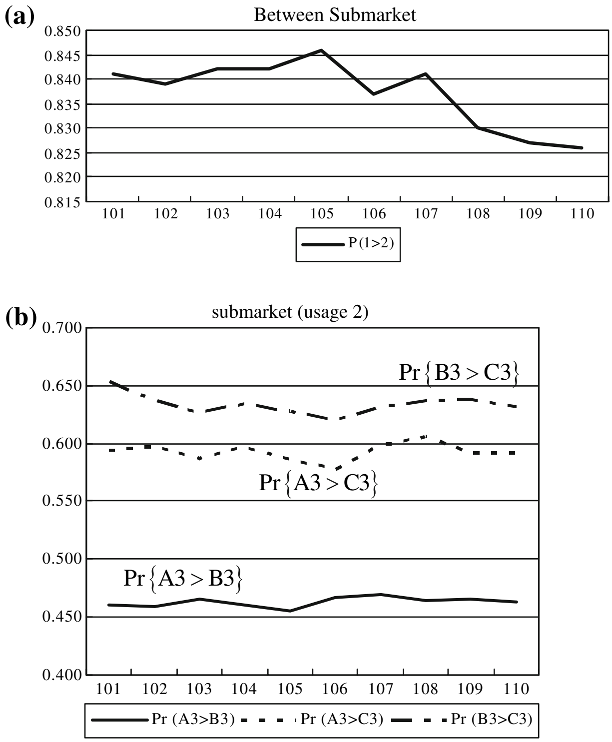


Fig. 7 Forecasting future events **a** submarket sales, **b** brand sales

mean and variance values. It is also known that several independent compound Poisson distributions, when their sum is given, have a similar relationship with compound multinomial distributions characterized as mixed negative binomial distributions. In addition, the Zero-inflated Poisson (ZIP) model by e.g., Lambert (1992) could be also applied to our modeling when the data contain many zeros. This modeling demands more complicated procedures of computation. The expected gains from this extension could not be substantial, compared with the development of a new model, and thus we would like to leave this modification of model for future research.

From the view point of developing marketing models, we would like to explore the problems of dynamic change of market structure, endogeneity between sales and marketing variables, and more elaborated selection of explanatory variables, e.g., the choice of order of stochastic trend for each subject, pricing and promotional variables to extend our model.

Another problem is on the empirical research. As mentioned above, the proposed model can depict more complicated market structure with irregularly bifurcated sub-markets. In particular, it is of much interest to find a market with umbrella structure in the point of brand management for the firms producing several brands in the same category. The existing methods are not easily applicable to that problem, at least, in the way of incorporating sales dynamics. We are not ready to apply this structure to our dataset because it is evident that the dataset has no such relations. The usefulness of our model could be shown further when another dataset with possible umbrella structure gets available.

Appendix A: *MCMC* Algorithm

A.1 *Generalized linear model* Generating $\{\tilde{\eta}_t^s\}$ for link function

When the initial values of $(\tilde{\eta}_t^{s(0)}, \tilde{\theta}_t^{s(0)}, V^{s(0)})$ are given, the conditional posterior density of $\tilde{\eta}_t^s$ is defined by

$$p(\tilde{\eta}_t^s | X_t^s, \tilde{\theta}_t^s, V^s, \{y_{it}\}) \propto p(z_t^s | \tilde{\eta}_t^s) p(\tilde{\eta}_t^s | X_t^s, \tilde{\theta}_t^s, V^s),$$

where $p(z_t^s | \tilde{\eta}_t^s)$ is the data density of n_t with Poisson when $s = 1$, and multinomial distributions of $\tilde{m}_t | n_t$ and $\tilde{y}_t | \tilde{m}_t$ when $s = 2$ and 3 respectively. The last term on the right-hand side is the prior density for $\tilde{\eta}_t$, which is defined as structural equation (6).

Then, we use Metropolis–Hastings with a random walk algorithm,

$$\tilde{\eta}_t^{s(k)} = \tilde{\eta}_t^{s(k-1)} + \omega, \quad \omega \sim N(0, 0.1 I_s),$$

where I_s is an identity matrix with corresponding dimensions to the case of “s”.

Acceptance probability α is defined as

$$\alpha(\tilde{\eta}_t^{s(k)}, \tilde{\eta}_t^{s(k-1)}) = \min\left(\frac{p(\tilde{\eta}_t^{s(k)} | X_t^s, \tilde{\theta}_t^{s(k)}, V^{s(k)}, \{y_{it}\})}{p(\tilde{\eta}_t^{s(k-1)} | X_t^s, \tilde{\theta}_t^{s(k-1)}, V^{s(k-1)}, \{y_{it}\})}, 1 \right).$$

Then, using uniform random number $u \sim U(0, 1)$, we determine the rule of acceptance of random draws as

$$\tilde{\eta}_t^s = \begin{cases} \tilde{\eta}_t^{s(k)} \text{accepted} & \text{if } u \le \alpha\left(\tilde{\eta}_t^{s(k)}, \tilde{\eta}_t^{s(k-1)}\right) \\ \tilde{\eta}_t^{s(k-1)} \text{accepted} & \text{if otherwise.} \end{cases}$$

A.2 *Forecasting sales and constituting predictive density*

Given the mth draw of MCMC, $\{\tilde{\theta}_t^{s(m)}, V^{s(m)}, W^{s(m)}\}$ for $s = 1, 2, 3$,

(i) obtain the forecast of parameter of market sales (total number) $\lambda_{1t+1}^{1(m)} = \exp(\eta_{t+1}^{1(m)})$ from the one-step-ahead forecast $\eta_{1t+1}^{1(m)}$ of structural equation (6) generated using $\theta_{t+1}^{1(m)}$ in the system equation (12),

(ii) get the forecast of market sales $n_{t+1}^{(m)} \sim \text{Poisson}(\lambda_{t+1}^{1(m)})$,

(iii) given $n_{t+1}^{(m)}$ together with the parameter values $\{\pi_{it+1}^{(m)}\}$ derived from $\{\eta_{it+1}^{2(m)}\}$ in (10), generate the submarket sales forecasts $\tilde{m}_{t+1}^{[k](m)}$ by sampling from the multinomial distribution

$$\tilde{m}_{t+1}^{(m)} | n_{t+1}^{(m)} \sim \text{Multinomial}(n_{t+1}^{(m)}, \{\pi_{it+1}^{(m)}\}).$$

(iv) given $\tilde{m}_{t+1}^{[k](m)}$ together with the parameter values $\{\pi_{it+1}^{[k](m)}\}$ of multinomial distribution of submarket M_k derived from $\{\eta_{it+1}^{3(m)}\}$ of (10), generate the respective brand's forecasts by sampling from the multinomial distribution

$$\tilde{y}_{t+1}^{[k](m)} \,|\, \tilde{m}_{t+1}^{[k](m)} \sim \text{Multinomial}\ (\tilde{m}_{t+1}^{[k](m)}, \{\pi_{it+1}^{[k](m)}\}) \quad \text{for } k = 1, \ldots, L.$$

(v) iterate the steps (i)–(iii) M times.

Then, the empirical distribution of $\{\tilde{y}_{t+1}^{[k](m)}, m = b, \ldots, M\}$ approximates the predictive density (18) in case of $z_{T+1} = y_{t+1}^{[k]}$. We set the burn-in parameter $b = 4{,}000$ and the total number of iterations $M = 6{,}000$ for empirical application after checking the convergence. By extending the forecasting steps above up to H step ahead, we obtain the MCMC sample path $\{\tilde{y}_{t+1}^{[k](m)}, \tilde{y}_{t+2}^{[k](m)}, \ldots, \tilde{y}_{t+H}^{[k](m)}\}$ for the joint predictive density. As for the estimate of future events between variables (19), we count the number of times that the event is held in the $(M - b)$ times iterations and its ratio gives the estimate.

References

Cargnoni, C., Muller, P., West, M. (1997). Bayesain forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association, 92*, 640–647.

Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika, 81*, 541–553.

Colombo, R. A., Morrison, D. G. (1989). A brand switching implications for marketing strategies. *Marketing Science, 8*, 89–99.

Grover, R., Dillon, W. R. (1985). A probabilistic model for testing hypothesized hierarchical market structures. *Marketing Science, 4*, 312–335.

Grover, R., Srinivasan, R. V. (1987). A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research, 24*, 139–153.

Hanssen, D. M., Parsons, L. J., Shultz, R. L. (2001). *Market response models: econometric and time series analysis* (2nd ed.). Boston, MA: Kluwer.

Harvey, A. C., Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics, 7*, 407–422.

Kamakura, W., Wedel, M. (2000). *Market segmentation—conceptual and logical foundations* (2nd ed.). Boston: Kluwer.

Lambert, D. (1992). Zero-inflated Poisson regression, with application to defects in manufacturing. *Technometrics, 34*, 1–14.

Lilien, G. L., Rangaswamy, A. (2003). *Marketing engineering*. New Jersey: Prentice Hall.

Lilien, G. L., Kotler, P., Moorthy, K. S. (1992). *Marketing models*. New Jersey: Prentice Hall.

McCullagh, P., Nelder, J. A. (1990). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Ord K., Fernandes C., Harvey A. C. (1993) Time series models for multivariate series of count data. In T. Subba Rao (Ed.), *Developments in time series analysis* (pp. 295–309). London: Chapman & Hall.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series, B 64*, 583–639.

Urban, G. L., Johnson, P. L., Hauser, J. R. (1984). Testing competitive market structures. *Marketing Science, 3*, 83–112.

West, M., Harrison, P. J. (1989). *Bayesian forecasting and dynamic models*. New York: Springer.

West, M., Harrison, P. J., Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association, 80*, 73–97.