

T M & A R G

Discussion Paper No. 113

**Multivariate Time Series Model with
Hierarchical Structure
for Over-dispersed Discrete Outcomes**

Nobuhiko Terui and Masataka Ban

August, 2013

January, 2013 (First version)

TOHOKU MANAGEMENT & ACCOUNTING RESEARCH GROUP

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
KAWAUCHI, AOBA-KU, SENDAI
980-8576 JAPAN

**Multivariate Time Series Model with Hierarchical Structure
for Over-dispersed Discrete Outcomes**

Nobuhiko Terui*

and

Masataka Ban**

August, 2013

January, 2013

*Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan

**College of Economics, Nihon University, Chiyoda-ku, Tokyo 102-8360, Japan

Corresponding Author: Nobuhiko Terui, terui@econ.tohoku.ac.jp

The authors acknowledge useful comments from two referees for revising this manuscript.
Terui also acknowledges the financial support of the Japanese Ministry of Education Scientific Research Grants (A)21243030.

Multivariate Time Series Model with Hierarchical Structure for Over-dispersed Discrete Outcomes

Abstract

In this paper, we propose a multivariate time series model for over-dispersed discrete data to explore the market structure based on sales count dynamics. We first discuss the microstructure to show that over-dispersion is inherent in the modeling of market structure based on sales count data. The model is built on the likelihood function induced by decomposing sales count response variables according to products' competitiveness and conditioning on their sum of variables, and it augments them to higher levels by using Poisson-Multinomial relationship in a hierarchical way, represented as a tree structure for the market definition. State space priors are applied to the structured likelihood to develop dynamic generalized linear models for discrete outcomes.

For over-dispersion problem, Gamma compound Poisson variables for product sales counts and Dirichlet compound multinomial variables for their shares are connected in a hierarchical fashion. Instead of the density function of compound distributions, we propose a data augmentation approach for more efficient posterior computations in terms of the generated augmented variables particularly for generating forecasts and predictive density.

We present the empirical application using weekly product sales time series in a store to compare the proposed models accommodating over-dispersion with alternative no over-dispersed models by several model selection criteria, including in-sample fit, out-of-sample forecasting errors, and information criterion. The empirical results show that the proposed modeling works well for the over-dispersed models based on compound Poisson variables and they provide improved results than models with no consideration of over-dispersion.

Key words:

Compound Poisson, Compound Multinomial, Discrete Outcomes, Dynamic Generalized Linear Model, Hierarchical Market Structure, MCMC, Over-dispersion

1. Introduction

A full Bayesian analysis on the dynamics of discrete responses such as counts has been facilitated by the Markov chain Monte Carlo (MCMC) methods for more than 20 years. The scope is beyond the earlier works with maximum likelihood method by Harvey and Fernandes (1989) and Ord, Fernandes, and Harvey (1993). In particular, West et al. (1985) and Cargnoni et al. (1997) developed time-series models for variables following a multinomial distribution by introducing dynamic linear models using the Bayesian approach. The former proposed a dynamic model with a multinomial distribution and the latter dealt with several sets of multinomial distributions, both of which assumed the total number of variables to be constant.

In contrast, the stochastic models for a discrete response exhibit an interesting distributional property: the reproduction of Poisson variables and the conditional distribution of these variables on their sum follow a multinomial distribution. Terui et al. (2010) used this Poisson–multinomial relationship in a dynamic generalized linear model to propose a multivariate time-series model with a hierarchical structure between variables for specifying market structure on the basis of a product’s sales dynamics. We use the term “market structure” to refer to how we classify products into several groups called “submarkets” or “categories,” so that the products are competitive inside a submarket but not outside it. Their model was a macro model for direct aggregate sales, without consideration of microstructure. On the other hand, the Poisson variable has a limited property of having identical first two moments, and therefore the over-dispersion has been discussed as important issues in the literature particularly in econometrics, as is fully discussed in Winkelmann (2008).

In this paper, we extend the model of Terui et al. (2010) such that the discrete response variables have over-dispersions. We first incorporate the model of an individual consumer’s purchase and then aggregate those up to a product sale, after which we find a microstructure to generate over-dispersions for our application. We prove that over-dispersion is inherent

whenever consumers in the market do not behave independently, as is usually assumed in economics and marketing. Then, we propose to build a multivariate time series model with hierarchical structures by using Gamma compound Poisson variables for discrete responses having over-dispersion. We first develop the statistical modeling of compound distributions to represent a market structure, and then, propose a data-augmentation approach. That is, we retain the Poisson–Multinomial distributional relationship implying the number of total sales and product’s market share, where we augment Poisson and Multinomial parameters, which are generated by Gamma and Dirichlet distributions, respectively. Then, we obtain the joint posterior density by a full Bayesian MCMC procedure.

We discuss the microstructure for the inherent over-dispersion in our problem in Section 2. Section 3 describes the properties of the compound Poisson and compound multinomial distributions used as the building blocks for our model. The structure of the model is explained in Section 4, and data augmentation approach to over-dispersion is proposed in Section 5. Section 6 describes model specification in the dynamic generalized linear model and derives joint posterior density. Section 7 deals with the estimation and forecasting procedures. Section 8 reports the empirical application. Concluding remarks are provided in Section 9.

2. Microstructure for Generating Over-dispersions

Suppose that there are H potential consumers in the market and the number of purchases of product i by consumer h at time t follows a Poisson distribution with parameter λ_{ih} as

$$x_{iht} \sim \text{Poisson}(\lambda_{ih}), \quad (1)$$

on the ground that consumers will make no purchase or small quantity of buying product at a specific time.

Thus, the total number of sales for product i , y_{it} , also follows Poisson distribution

$$y_{it} = \sum_{h=1}^H x_{iht} I(h \in C_t) = \sum_{h^*=1}^{H_t} x_{ih^*_t}, \quad (2)$$

where $I(h \in C_t)$ is an indicator function, taking value “1” when consumer h belongs to a potential consumer set C_t , i.e., when he/she is ready to buy, and “0” otherwise, and $H_t = \sum_{h=1}^H I(h \in C_t)$ with h^* being the index for reordering of consumers in the set C_t . In this circumstance, the over-dispersion phenomenon is derived by evaluating the mean and variance of y_{it} as

$$E(y_{it}) = \sum_{h^*=1}^{H_t} E(x_{ih^*_t}) = \sum_{h^*=1}^{H_t} \lambda_{ih^*_t} \equiv \lambda_{it}^*, \quad (3)$$

$$\begin{aligned} \text{Var}(y_{it}) &= \sum_{h^*=1}^{H_t} \text{Var}(x_{ih^*_t}) + \sum_{h \neq h'} \text{Cov}(x_{ih^*_t}, x_{ih'^*_t}) \\ &= \lambda_{it}^* + \sum_{h \neq h'} \text{Cov}(x_{ih^*_t}, x_{ih'^*_t}) \geq \lambda_{it}^* = E(y_{it}), \end{aligned} \quad (4)$$

because it holds $\text{Cov}(x_{ih^*_t}, x_{ih'^*_t}) > 0$ for any pair of correlated Poisson variables.

That is, equation (4) implies that there is over-dispersion whenever at least one pair of consumers does not behave independently. This is not a strong assumption, as justified by discussions on the existence of a reference group in society and its decision making, going back to Hyman (1942), and its application to social psychology on the basis of the consumer behavior theory, such as Park and Lessig (1977) and Bearden and Etzel (1982). In fact, marketing models have been developed based on a commonality across consumers when they represent heterogeneity in the random effect models. This is well explained in the text book Rossi et al.(2005).

On the other hand, Gamma compound Poisson variables with positive parameters (a, τ) , denoted as $y \sim \text{Compound Poisson}(a, \tau)$, are suitable for over-dispersion as they contain the first two moments

$$\begin{aligned}
E[y] &= a\tau \\
\text{Var}[y] &= a\tau(1+\tau) \quad (> E[y]).
\end{aligned}
\tag{5}$$

In the next section, we assume that the number of product sales, y , by aggregating over an individual consumer's purchase, as shown in equation (2), follows the compound Poisson distribution.

3. Gamma Compound Poisson and Dirichlet Compound Multinomial Distributions

A Gamma compound Poisson variable is defined by the mixture of the Poisson variable having parameter λ in the Gamma distribution with (a, τ) , and its density function is evaluated as a negative binomial distribution.

$$\begin{aligned}
p(y|a, \tau) & \quad (a, \tau > 0) \\
&= \int_0^{\infty} f(y|\lambda) f(\lambda|a, \tau) d\lambda = \frac{\Gamma(y+a)}{y!\Gamma(a)} \left(\frac{\tau}{1+\tau}\right)^y \left(\frac{1}{1+\tau}\right)^a.
\end{aligned}
\tag{6}$$

Hoadley (1969) discussed the reproductive property of Gamma compound Poisson variables and the conditional distribution for a set of these variables when the sum of variables is given. That is, let y_1, y_2, \dots, y_I be mutually independent random variables having a Gamma compound Poisson distribution with the second parameter τ common across subjects, i.e.,

$$y_i \sim \text{Compound Poisson}(a_i, \tau). \tag{7}$$

Then, the sum $n = y_1 + y_2 + \dots + y_I$ follows

$$n \sim \text{Compound Poisson}\left(\sum_{i=1}^I a_i, \tau\right). \tag{8}$$

Furthermore, the conditional distribution of $\tilde{y} = (y_1, y_2, \dots, y_I)'$ when n is given is shown as

$$\tilde{y} | n \sim \text{Dirichlet Compound Multinomial}(\tilde{y} | n, \tilde{a}), \tag{9}$$

where $\tilde{a} = \{a_i, i = 1, \dots, I\}$, and its density is derived by

$$p(\tilde{y} | n, \tilde{a}) = \int_0^\infty f(\tilde{y} | n, \tilde{\pi}) f(\tilde{\pi} | \tilde{a}) d\tilde{\pi} = \frac{n! \Gamma(\sum_i a_i)}{\Gamma(n + \sum_i a_i)} \prod_{i=1}^n \frac{\Gamma(y_i + a_i)}{y_i! \Gamma(a_i)}. \quad (10)$$

4. Models for Defining Market Structure

The products are more or less competitive in their market. These products are grouped according to the degree of competitiveness into several segments, and further categorizing these groups to higher levels to make subgroups leads to a hierarchical structure for the market definition. A tree structure is used to represent the hierarchical nature of competitive relationships among products in a graphical form, as shown in Figure 1. Figure 1(a) indicates the market with no specific structure and each product sale is directly accumulated to market sale. On the other hand, Figure 1(b) shows the situation that some groups of product sales are respectively accumulated to sub-markets first and then sub market sales constitute market sale.

Figure 1 (a), (b) Market Structures

Basic Structure

Let us assume that there are I products in the market and that y_{it} is the number of sales for the product i at time t ($t = 1, \dots, T$), which follows the Gamma compound Poisson distribution independently with a time-varying parameter (a_{it}, τ_t) defined by (7) for $i = 1, \dots, I$ when there is no competitive relationship with each other. Then, we obtain the Gamma compound Poisson distribution with $(\sum a_{it}, \tau_t)$ for market sales, defined as the aggregate of product sales, $n_t = \sum_{i=1}^I y_{it}$, under the assumption of no specific structure among products in the market, as shown in Figure 1(a). That is, we have marginal distributions for

product and market sales by

$$y_{it} \sim \text{Compound Poisson}(a_{it}, \tau_t), \quad n_t \sim \text{Compound Poisson}(a_t^*, \tau_t), \quad (11)$$

where $a_t^* = \sum_{i=1}^I a_{it}$. Furthermore, after n_t is given, the conditional distribution of product

sales $\tilde{y}_t = \{y_{it}, i = 1, \dots, I\}$ follows

$$\tilde{y}_t | n_t \sim \text{Dirichlet Compound Multinomial}(\tilde{y}_t | n_t, \tilde{a}_t), \quad (12)$$

where $\tilde{a}_t = (a_{1t}, \dots, a_{It})'$. The sequential use of equations (11) and (12) produces a joint

distribution for market and product sales,

$$p(n_t, \tilde{y}_t | \tilde{a}_t, \tau_t) = p(n_t | a_t^*, \tau_t) p(\tilde{y}_t | n_t, \tilde{a}_t). \quad (13)$$

We note that the conditioning set $(a_t^*, \tau_t, \tilde{a}_t)$ has equivalent information with (\tilde{a}_t, τ_t) if we take \tilde{a}_t as a full-dimensional vector with a nondegenerated distribution.

In contrast, Terui et al. (2010) proposed a dynamic generalized linear model based on Poisson variables without over-dispersions. This represents a macromodel for aggregate sales directly without considering the microstructure. They used the reproductive property of Poisson variables and conditional multinomial distribution when the sum of variables is given, and proposed a multivariate time-series model with a hierarchical structure based on the discrete outcomes. That is, we have marginal distributions $y_{it} \sim \text{Poisson}(\lambda_{it})$ and

$n_t \sim \text{Poisson}(\lambda_t^*)$ and conditional distribution $\tilde{y}_t | n_t \sim \text{Multinomial}(\tilde{y}_t | n_t, \tilde{\pi}_t)$, where

$\lambda_t^* = \sum_{i=1}^I \lambda_{it}$ and $\tilde{\pi}_t = \{\pi_{jt} = \lambda_{jt} / \sum_{i=1}^I \lambda_{it}, j = 1, \dots, I-1\}$. The likelihood at time t is defined

by

$$p(n_t, \tilde{y}_t | \lambda_t^*, \tilde{\pi}_t) = p(n_t | \lambda_t^*) p(\tilde{y}_t | n_t, \tilde{\pi}_t), \quad (14)$$

where the induced parameters λ_t^* and $\tilde{\pi}_t$, respectively, represent the expected total sales and market shares for each product.

Higher Order Structure

This model is extended to a higher order hierarchical structure, as developed by Terui et al. (2010). Next, we fully explain the model specifications, including the design matrix of state space priors, as this model is applied to actual time-series data in the empirical analysis.

Extending the Poisson–multinomial relationship in the above-described manner, we decompose the market structure into L submarkets M_k such that $\tilde{y}_t = \bigcup_{k=1}^L \tilde{y}_t^{[k]}$, where $\tilde{y}_t^{[k]} = (y_{it}, i \in M_k)'$ represents N_k dimensional vector of the products that are grouped in M_k , $k = 1, \dots, L$. Given aggregated submarket sales $m_t^{[k]} = \sum_{i \in M_k} y_{it}$, the conditional distribution $\tilde{y}_t^{[k]} | m_t^{[k]}$, $k = 1, \dots, L$, follows independent N_k -dimensional multinomial distribution since $\tilde{y}_t^{[k]}$ s are orthogonal to each other, i.e., $(\tilde{y}_t^{[l]} | m_t^{[l]}) \perp (\tilde{y}_t^{[k]} | m_t^{[k]})$ for $l \neq k$ by the definition of a submarket.

Next, let $\tilde{m}_t = (m_t^{[1]}, \dots, m_t^{[L]})'$ denote an 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 brief, we have a three-layer hierarchical market structure model. We note that $\{\tilde{y}_t\}$ and $\{n_t\}$ are independent, conditional on $\{\tilde{m}_t\}$. Then, the joint density function of I product sales (bottom layer), L submarket sales (middle layer), and market sales (top layer) are decomposed into

$$\begin{aligned} p(n_t, \tilde{m}_t, \tilde{y}_t) &= p(n_t) p(\tilde{m}_t | n_t) p(\tilde{y}_t | \tilde{m}_t) \\ &= p(n_t) \prod_{k=1}^L [p(m_t^{[k]} | n_t) p(\tilde{y}_t^{[k]} | m_t^{[k]})]. \end{aligned} \quad (15)$$

Then, we obtain marginal and conditional data distributions: $n_t \sim \text{Poisson}(\lambda_t^*)$,

$\tilde{m}_t | n_t \sim \text{Multinomial}(n_t, \tilde{\pi}_t)$, where $\pi_{jt} = \sum_{i=1}^{N_j} \lambda_{it}^{[j]} / \sum_{k=1}^L \sum_{i=1}^{N_k} \lambda_{it}^{[k]}$ ($j=1, \dots, L-1$),

$\tilde{y}_t^{[k]} | m_t^{[k]} \sim \text{Multinomial}(m_t^{[k]}, \phi_t^{[k]})$ ($k=1, \dots, L$), $\phi_t^{[k]} = (\phi_{it}^{[k]}, i \in M_k)'$,

$\phi_{it}^{[k]} = \lambda_{it}^{[k]} / \sum_{j=1}^{N_k} \lambda_{jt}^{[k]}$ ($i=1, \dots, N_k-1$), and $\lambda_{it}^{[k]}$ is the parameter of the Poisson variable i

classified to the submarket M_k . The structure of this model is illustrated in Figure 1(b).

5. Data Augmentation Approach to Over-dispersion

Data Augmentation

Extending the model for accommodating over-dispersion, instead of direct use of densities (6) and (10), we take a data augmentation approach to keep the original parameters, i.e., implying expected sales and market shares for each product in the modeling, and use the generated sample of augmented variable $z_t = (\lambda_t^*, \tilde{\pi}_t)'$ in the MCMC process to define the likelihood for the parameters (\tilde{a}_t, τ_t) :

$$p(z_t | \tilde{a}_t, \tau_t) = \text{Gamma}(\lambda_t^* | \tilde{a}_t, \tau_t) \times \text{Dirichlet}(\tilde{\pi}_t | \tilde{a}_t). \quad (16)$$

By using the relation of posterior density

$$\begin{aligned} p(\tilde{a}_t, \tau_t | n_t, \tilde{y}_t) &\propto \iint p(\tilde{a}_t, \tau_t, \lambda_t^*, \tilde{\pi}_t | n_t, \tilde{y}_t) d\lambda_t^* d\tilde{\pi}_t \\ &= \left[\int p(n_t | \lambda_t^*) p(\lambda_t^* | \tilde{a}_t, \tau_t) d\lambda_t^* \right] \left[\int p(\tilde{y}_t | n_t, \tilde{\pi}_t) p(\tilde{\pi}_t | \tilde{a}_t) d\tilde{\pi}_t \right], \end{aligned} \quad (17)$$

we evaluate these integrals by augmenting the gamma and Dirichlet parameters in terms of generating the s -th samples $\lambda_t^{*(s)}$ from gamma $p(\lambda_t^{*(s)} | \tilde{a}_t^{(s)}, \tau_t^{(s)})$ and $\tilde{\pi}_t^{(s)}$ from Dirichlet $p(\tilde{\pi}_t^{(s)} | n_t, \tilde{a}_t^{(s)})$ in MCMC iterations. Then, conditional on $z_t^{(s)} = (\lambda_t^{*(s)}, \tilde{\pi}_t^{(s)})'$, we obtain the Poisson–Multinomial likelihood function

$$p(n_t | \lambda_t^{*(s)}) p(\tilde{y}_t | n_t, \tilde{\pi}_t^{(s)}), \quad (18)$$

and this forms a building block to constitute a hierarchical structure.

In case of two-layer models and three submarket ($L=3$), it is extended as follows: for

$$\begin{aligned} z_t &= (\lambda_t^*, \tilde{\pi}_t', \tilde{\phi}_t^{[1]}, \tilde{\phi}_t^{[2]}, \tilde{\phi}_t^{[3]})', \\ p(\tilde{a}_t, \tau_t | n_t, \tilde{m}_t, \tilde{y}_t) &= \iint p(\tilde{a}_t, \tau_t; z_t | n_t, \tilde{m}_t, \tilde{y}_t) dz_t \\ &= \left[\int p(n_t | \lambda_t^*) p(\lambda_t^* | a_t^*, \tau_t) d\lambda_t^* \right] \times \\ &\int p(\tilde{m}_t | n_t, \tilde{\pi}_t) p(\tilde{\pi}_t | \tilde{a}_{mt}) d\tilde{\pi}_t \prod_{k=1}^L \left[\int p(\tilde{y}_t^{[k]} | m_t^{[k]}, \tilde{\phi}_t^{[k]}) p(\tilde{\phi}_t^{[k]} | \tilde{a}_t^{[k]}) d\tilde{\phi}_t^{[k]} \right], \end{aligned} \quad (19)$$

$$\text{where } \tilde{a}_{mt} = \left(\sum_{i \in M_k} a_{it}, k = 1, \dots, L \right)' \text{ and } \tilde{a}_t^{[k]} = (a_{it}, i \in M_k)'.$$

We evaluate these integrals by augmenting the gamma and Dirichlet parameters z_t in terms of generated samples $\lambda_t^{*(s)}$ from Gamma $p(\lambda_t^{*(s)} | a_t^{*(s)}, \tau_t^{(s)})$, $\tilde{\pi}_t^{(s)}$ and $\tilde{\phi}_t^{[k](s)}$ from Dirichlet $p(\tilde{\pi}_t^{(s)} | n_t, \tilde{a}_{mt}^{(s)})$ and $p(\tilde{\phi}_t^{[k](s)} | \tilde{a}_t^{[k](s)})$, respectively. Then, conditional on $z_t^{(s)} = (\lambda_t^{*(s)}, \tilde{\pi}_t^{(s)}, \tilde{\phi}_t^{[1](s)}, \tilde{\phi}_t^{[2](s)}, \tilde{\phi}_t^{[3](s)})'$, we obtain the Poisson–Multinomial likelihood function:

$$p(n_t | \lambda_t^{*(s)}) p(\tilde{m}_t | n_t, \tilde{\pi}_t^{(s)}) \prod_{k=1}^L \left[p(\tilde{y}_t^{[k](s)} | m_t^{[k](s)}, \tilde{\phi}_t^{[k](s)}) \right]. \quad (20)$$

6. Model Specification and Joint Posterior Density

Dynamic Generalized Linear Model

Using the expectation of product sales leads to

$$E[y_{it}] = a_{it} \tau_t. \quad (21)$$

Thus, we interpret that the expected sale is decomposed into an individual mean a_{it} and a common mean τ_t across products. In turn, we model the mean function as $a_{it} \tau_t = f(x_{it}, \varepsilon_{it})$

connecting with marketing mix variables x_{it} and stochastic error ε_{it} . We specify the structure in more detail as

$$\log a_{it} = a_i - \log \tau_t + x_{it}' \beta_{it} + \varepsilon_{it}, \quad i = 1, \dots, I. \quad (22)$$

That is, the individual mean has a product-specific mean a_i and a common time trend $\log \tau_t$, and it is assumed to be mostly explained by marketing mix variables. We set $a_{it}^* = \log a_{it} - a_i$ and denote $\tau_t^* = -\log \tau_t$, and then conditional on a_i , we have the dynamic equation

$$a_{it}^* = \tau_t^* + x_{it}' \beta_{it} + \varepsilon_{it}. \quad (23)$$

This constitutes the structural equation $\tilde{\eta}_t = F_t \tilde{\theta}_t + \tilde{v}_t$, where $\tilde{\eta}_t = (a_{1t}^*, \dots, a_{It}^*)'$ and

$\tilde{\theta}_t = (\beta_{1t}, \beta_{2t}, \dots, \beta_{It}, \tau_t^*)'$, F_t is the matrix defined by the structure on $\tilde{\theta}_t$ in equation (24),

and \tilde{v}_t is the error vector comprising of ε_{it} . As for the dynamic state vector used in the

application, we specify the second-order local common trend for τ_t^* and the first-order local trend model for the response parameter β_{it} , i.e.,

$$\begin{cases} \tau_t^* = \tau_{t-1}^* + \kappa_{t-1} + w_{1t}; \kappa_t = \kappa_{t-1} + w_{2t} \\ \beta_{it} = \beta_{it-1} + w_{3it}, \quad i = 1, \dots, I, \end{cases} \quad (24)$$

and this specifies the system equation $\tilde{\theta}_t = H_t \tilde{\theta}_{t-1} + \tilde{w}_t$.

Coupled with the data distribution (15), we define the dynamic generalized linear model with the state space prior:

$$\begin{cases} \tilde{\eta}_t = F_t \tilde{\theta}_t + \tilde{v}_t, & \tilde{v}_t \sim N(0, V) \\ \tilde{\theta}_t = H_t \tilde{\theta}_{t-1} + \tilde{w}_t, & \tilde{w}_t \sim N(0, W) \end{cases}. \quad (25)$$

Joint Posterior Density

Under the usual assumption that the prior density for the covariance matrix of structural and system equations $p(V, W) = p(V) p(W)$, we can express the prior distribution as

$$\begin{aligned}
& p(\{\tilde{\eta}_t\}, \{\tilde{\theta}_t\}, V, W | \{a_i\}) \\
&= \prod_{t=1}^T p(\tilde{\eta}_t | X_t, \tilde{\theta}_t, V, \{a_i\}) p(\tilde{\theta}_t | \tilde{\theta}_{t-1}, W, \{a_i\}) p(V) p(W)
\end{aligned} \tag{26}$$

Next, we set the prior distribution $p(a_i)$ of the individual mean parameter a_i by

$$a_i | v_i^2 \sim N(\bar{a}_i, v_i^2 / \xi_i^2); \quad (\bar{a}_i \sim N(a_0, b_0); \xi_i^2 \sim IG(n_0 / 2, s_0 / 2)). \tag{27}$$

By arranging the term in (22) as $c_{it} = a_i + \varepsilon_{it}$, where $c_{it} = \log a_{it} + \log \tau_t - x_{it}' \beta_{it}$, for $t = 1, \dots, T$, we constitute the likelihood for the mean a_i conditional on the error variance v_i^2 and derive conditional posterior distribution $p(a_i | \{\tilde{\eta}_t\}, \{\tilde{\theta}_t\}, V, W, \{a_{-i}\}, \xi_i^2)$ in a closed form as

$$a_i | v_i^2 \sim N\left(\frac{\xi_i^2 \bar{a}_i + T \bar{c}_{i\cdot}}{\xi_i^2 + T}, \frac{v_i^2}{\xi_i^2 + T}\right), \tag{28}$$

where $\bar{c}_{i\cdot} = \sum_{t=1}^T c_{it} / T$ and $\{a_{-i}\}$ means the set of $a_k, k = 1, \dots, I$ excluding a_i .

Finally, we obtain the joint posterior density with data augmentation of

$$z_t = (\lambda_t^*, \tilde{\pi}_t, \tilde{\phi}_t^{[1]}, \tilde{\phi}_t^{[2]}, \tilde{\phi}_t^{[3]})' \text{ by equation (29):}$$

$$\begin{aligned}
& p(\{\tilde{\eta}_t\}, \{\tilde{\theta}_t\}, (V, W), \{a_i\}, \{\xi_i^2\}; \{z_t\} | \text{data}) \\
&= \prod_{t=1}^T \left[p(n_t | \lambda_t^*) p(\lambda_t^* | \tilde{\eta}_t, \tau_t) p(\tilde{m}_t | n_t, \tilde{\pi}_t) p(\tilde{\pi}_t | \tilde{\eta}_t, \tau_t) \prod_{k=1}^3 \left\{ p(y_t^{[k]} | m_t, \phi_t^{[k]}) p(\phi_t^{[k]} | \tilde{\eta}_t, \tau_t) \right\} \right. \\
&\quad \times p(\tilde{\eta}_t | X_t, \tilde{\theta}_t, V, \{a_i\}) p(\tilde{\theta}_t | \tilde{\theta}_{t-1}, W) p(V) p(W) \\
&\quad \left. \times \prod_{i=1}^I p(a_i | \{\tilde{\eta}_t\}, \{\tilde{\theta}_t\}, V, W, \{a_{-i}\}, \xi_i^2) p(a_i) p(\xi_i^2) \right].
\end{aligned} \tag{29}$$

In (29), ‘‘data’’ means the observed data $(\{\tilde{y}_t\}, \{\tilde{x}_t\})$.

Our model with over-dispersions is characterized as the dynamic generalized linear models

of the Poisson–Multinomial distribution perturbed by the Gamma–Dirichlet distributions.

7. Estimation and Forecasting

Sampling the link function for MCMC

In addition to the standard Bayesian inference on state space modeling by dynamic linear models (DLMs) by West and Harrison (1997), we use the MCMC approach to estimate the model by using Metropolis–Hastings sampling specifically for the conditional posterior density of link functions based on the data-augmented representation

$$p(\tilde{\eta}_t | F_t, \tilde{\theta}_t, V, \text{data}) \propto \int p(n_t, \tilde{m}_t, \tilde{y}_t | z_t) p(z_t | \tilde{\eta}_t) p(\tilde{\eta}_t | F_t, \tilde{\theta}_t, V) dz_t, \quad (30)$$

where $z_t = (\lambda_t^*, \tilde{\pi}_t', \phi_t^{[1]'}, \phi_t^{[2]'}, \phi_t^{[3]'})'$. Once the values of $\tilde{\eta}_t$ for equation (30) are given, the structural equations coupled with the system equations in their state space priors in equation (25) constitute the conventional Gaussian state space models. The multi-move sampler by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) is used to sample the state vector $\{\tilde{\theta}_t\}$.

We assume that the initial values of the state vector $\tilde{\theta}_0$ follow a multivariate normal distribution $N(\bar{\theta}_0, d \times I)$. The mean vector $\bar{\theta}_0$ was set as the estimate of the coefficient on static regression, i.e., the regression with time-invariant coefficient, and we set $d = 0.1$ for the empirical application.

Predictive Density

Next, one-step-ahead predictive density $p(\tilde{y}_{t+1} | \text{data})$ is evaluated by

$$\int p(\tilde{y}_{T+1} | z_{T+1}, \text{data}) p(z_{T+1} | \tilde{\eta}_{T+1}, \tilde{\theta}_{T+1}) p(\tilde{\eta}_{T+1} | \tilde{\theta}_{T+1}, V, W, \text{data}) \times p(\tilde{\theta}_{T+1} | \tilde{\theta}_T, V, W, \text{data}) p(V, W) d\tilde{\theta}_T dV dW, \quad (31)$$

where $p(\tilde{\eta}_{T+1} | \tilde{\theta}_{T+1}, V, W, \text{data})$ is the conditional predictive density of link parameters when the predicted state vector, and structural and system error covariance matrices are given. The MCMC method is applied to evaluate this predictive density by the augmenting procedure. To evaluate this density, we first extend equation (16) to define the predictive likelihood of $\tilde{\eta}_{T+1}$ conditional on $\tilde{\theta}_{T+1}$ by

$$p(z_{T+1} | \tilde{\eta}_{T+1}, \tilde{\theta}_{T+1}, V, W, \text{data}) = \int \text{Gamma}(\lambda_{T+1}^* | \tilde{a}_{T+1}, \tilde{\theta}_{T+1}(\tau_{T+1})) d\lambda_{T+1}^* \times \iint \text{Dirichlet}(\tilde{\pi}_{T+1} | \tilde{a}_{m,T+1}) \prod_{k=1}^3 \text{Dirichlet}(\tilde{\phi}_{T+1}^{[k]} | \tilde{a}_{T+1}^{[k]}) d\tilde{\pi}_{T+1} d\tilde{\phi}_{T+1}^{[k]} \quad (32)$$

The details of the sampling scheme for MCMC are described in the appendix.

8. Empirical Application

Data and Variables

We use the store level scanner, point of sales (POS), time series in the curry roux category that was applied to our previous model in Terui et al. (2010) for comparison with the model with over-dispersion. The weekly series comprises three manufacturers that produce three products each, for a total of nine products during 110 weeks.

Table 1 Summary of Data

Figure 2 Histogram of Weekly Product Sales Data

Table 1 describes the summary statistics for sales, price display, and feature data. In particular, sales data contains variance to give an evidence of the presence of over-dispersion

in the data. Figure 2 show the histogram of brand sales.

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 product sales for $\{y_{it}\}$, and “prices”, “display (in-store promotion)”, and “features (advertising in newspaper)” for marketing mix variables $\{x_{it}\}$. x_{it} contains not only variables of their own, but also those with other products. The display and features are binary data taking a value of “1” when it was on and “0” when it was off. The logs of price data are used.

Model Comparison

Each of the three makers, A, B, and C, produces three categories of products according to the level of spiciness to accommodate the difference in consumer tastes (1: Not spicy, 2: Medium spicy, 3: Spicy). Following the discussion of Terui (2011), we assume three possible market structures: (1) product category, (2) makers, and (3) usage, i.e., “ordinary” or “luxury” usage, as shown in Figure 3, and compare these models with over-dispersion and without over-dispersion. That is, we have six models to be compared.

Figure 3 Comparative Models

The top of Table 2-1 shows the log of marginal likelihood (LML) as an in-sample fit criterion and two types of predictive measures, 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.

There are three levels for forecasting the RMSE: market, submarket, and product. These errors, including the null model of “no” structure, are reported in the lower panel of the table 2-2. We apply two types of measures: “sum1” and “sum2.” “sum1” is the sum of all errors

induced by the model in which we have no specific preference on the levels to be predicted. “sum2” is defined as the sum of market and product errors by considering that the numbers of submarkets differ between null and other structures.

According to the three criteria, the proposed models based on compound Poisson variables accommodating over-dispersion improve the models with no over-dispersions. In particular, the compound Poisson variable model under the market structure (3) shows the best performance.

Table 2-1 Model Comparison: Overall

Table 2-2 Model Comparison: Decomposition of RMSE

Table 3 Estimates of Product Intercept

Figure 4 In-sample Performance and Forecasting—Market, Submarket, and Product

Figure 5 Estimates of Structural Parameter

The left panels of Figure 4 show the predicted fit of in-sample data for market, submarket of “usage,” and product levels, 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 well over the observational period. The right panel of each element depicts its 10-step-ahead forecasting for these sales, where the mean values of the predicted density at each prediction step are connected by a continuous line, and the 2.5% and 97.5% quantiles of the density at each step are connected by dashed lines. The hold-out samples are denoted by 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} for the prediction steps.

Figure 5(a), (b) show the trajectory of estimated parameters $\{\tau_t\}$ and $\{a_{1t}, \dots, a_{9t}\}$. It appears that τ_t s are fluctuating downward for the first period and then turning upward with local trends around a mean level of 0.9998. $\{a_{1t}, \dots, a_{9t}\}$ move more heterogeneously with large fluctuations, which should be proportional to the observed discrete outcomes.

Figure 5(c) indicates trends for product A2 and B3 sales, which belong to a different category; it shows the opposite trends. Figure 5(d) depicts the time-varying price coefficient estimates in response to price and promotions (“End” display and “Advertising”). We confirm the competitive relationship between submarkets, and more interestingly, we find that products B1 and C2 are not hostile to A1 in the sense of pricing strategy, as they have the same coefficient sign as that of A1.

9. Concluding Remarks

In this study, we proposed a multivariate time series model for over-dispersed discrete data. Therein, we extended the model with the hierarchical structure by Terui et al. (2010) to accommodate the over-dispersion problem inherent in the modeling of market structure based on sales count dynamics. We first discussed the mechanism of the microstructure for generating over-dispersion in a number of discrete sales data.

The Gamma compound Poisson variable for product sales count responses and Dirichlet compound multinomial variables for product share are connected in a hierarchical fashion as a tree structure for depicting a market. The model is based on the likelihood generated by decomposing sales count response variables according to the degree of competitiveness among products and conditioning on their sum, and builds them up to higher levels, represented as a tree structure. State space priors are applied to the likelihood generated by the compound distributions to develop dynamic generalized linear models for discrete responses with a

hierarchical structure.

We first showed that over-dispersion is inherent to the problems where consumers are more or less dependent. Then, we model the compound distributions for accommodating over-dispersions. However, instead of the direct use of the density function of compound distributions, we augment variables to make the numerical integrations for mixing easier, and provide more efficient algorithms compared to the method that makes direct use of compound distributions. The empirical analysis by weekly product sales data in a store showed that our modeling worked well and the models with over-dispersion, which is constructed by compound Poisson variables, performs better than the models without over-dispersion.

There are a few problems for future research. One is the extension of the theoretical study. The zero-inflated Poisson (ZIP) model by, for example, Lambert (1992) could be also applied to our modeling when the data contain many zeros. In particular, this could be important if we further incorporate modeling of individual consumer behavior in the analysis. This could be accommodated by mixture distributions through hierarchical models, as used in this study. However, this additional mixture modeling demands more complicated computation procedures. 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 the model for future research.

Appendix: MCMC Algorithm

A.1 Dynamic Generalized Linear Model

We summarize the prior and conditional posterior distribution used for our proposed model below.

Conditional Posterior Distribution

We run 10,000 MCMC iterations in the model. In all models, we used the last 5,000 iterations to estimate the posterior distribution of model parameters. When the initial values of the parameters are given, the conditional posterior density of the necessary parameters is generated as follows.

(i) $\tilde{\eta}_t \mid -$

$\tilde{\eta}_t$ is defined by equation (23), and we use Metropolis–Hastings with a random walk algorithm,

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

where I is an identity matrix with corresponding dimensions.

Acceptance probability α is defined as

$$\alpha\left(\tilde{\eta}_t^{(s)}, \tilde{\eta}_t^{(s-1)}\right) = \min\left(\frac{p\left(\tilde{\eta}_t^{(s)} \mid F_t, \tilde{\theta}_t^{(s)}, V^{(s)}, \text{data}\right)}{p\left(\tilde{\eta}_t^{(s-1)} \mid F_t, \tilde{\theta}_t^{(s-1)}, V^{(s-1)}, \text{data}\right)}, 1\right).$$

After obtaining the draw of these parameters, we use West and Harrison’s (1997) standard Bayesian inference procedure on state space modeling by DLM.

(ii) $\tilde{\theta}_t \mid -$

The multi-move sampler by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) is used to sample the state vector $\{\tilde{\theta}_t\}$

(iii) $\lambda_t^* \mid -$

Generate $\lambda_t^{*(s)}$ from gamma $p\left(\lambda_t^* \mid a_t^{*(s)}, \tau_t^{(s)}\right)$

(iv) $\tilde{\pi}_t \mid -$

Generate $\tilde{\pi}_t^{(s)}$ from Dirichlet $p\left(\tilde{\pi}_t \mid n_t, \tilde{a}_{mt}^{(s)}\right)$

(v) $\tilde{\phi}_t^{[k]} \mid -$

Generate $\tilde{\phi}_t^{[k](s)}$ from Dirichlet $p\left(\tilde{\phi}_t^{[k]} \mid \tilde{a}_t^{[k](s)}\right)$

(vi) $\bar{a}_i \mid -$

Generate $\bar{a}_i^{(s)}$ from $N\left(\frac{b_0^{-2}a_0 + m\bar{a}_\bullet}{b_0^{-2} + m}, \frac{\xi_i^2}{b_0^{-2} + m}\right)$,

where $\bar{a}_\bullet = \sum_{i=1}^m \bar{a}_i / m$. We set $a_0 = 0, b_0 = 10$ in the empirical analysis.

(vii) $\xi_i^2 | -$

Generate $\xi_i^{2(s)}$ from $IG\left((n_0 + m)/2, \left(s_0 + \sum_{i=1}^m (\bar{a}_i - \bar{a}_\bullet)^2\right)/2\right)$.

We set $n_0 = 2, s_0 = m$ in the empirical analysis.

A.2 Forecasting Discrete Outcomes and Constituting Predictive Density

Given the s -th draw of MCMC $\{\tilde{\theta}_T^{(s)}, V^{(s)}, W^{(s)}\}$,

(i) obtain the forecast $\tilde{\eta}_{T+1}^{(s)}$ from $p(\tilde{\eta}_{T+1} | \tilde{\theta}_T, V, W, \text{data})$ by the algorithm of Gaussian state space model;

(ii) obtain the forecast $\tilde{z}_{T+1}^{(s)}$ from $p(\tilde{z}_{T+1} | \tilde{\eta}_{T+1})$ to get the parameter

$$z_{T+1}^{(s)} = (\lambda_{T+1}^{*(s)}, \tilde{\pi}_{T+1}^{(s)}, \phi_{T+1}^{[1](s)}, \phi_{T+1}^{[2](s)}, \phi_{T+1}^{[3](s)});$$

(iii) generate the random number $n_{T+1}^{(s)} \sim \text{Poisson}(\lambda_{T+1}^{*(s)})$ for the market sales forecast;

(iv) given $n_{T+1}^{(s)}$ together with the parameter values $\{\pi_{iT+1}^{(s)}\}$, generate $\tilde{m}_{T+1}^{(s)}$ by sampling from the multinomial distribution

$$\tilde{m}_{T+1}^{(s)} | n_{T+1}^{(s)} \sim \text{Multinomial}(n_{T+1}^{(s)}, \tilde{\pi}_{T+1}^{(s)}) \text{ for the submarket sales forecasts;}$$

(v) given $\tilde{m}_{T+1}^{[k](s)}$ together with the parameter values $\{\tilde{\phi}_{T+1}^{[k](s)}\}$ of the multinomial distribution of submarket M_k , generate the respective product's forecasts by sampling from the multinomial distribution

$$\tilde{y}_{T+1}^{[k](s)} | \tilde{m}_{T+1}^{[k](s)} \sim \text{Multinomial}(\tilde{m}_{T+1}^{[k](s)}, \tilde{\phi}_{T+1}^{[k](s)}) \text{ for } k=1, \dots, L;$$

(vi) iterate steps (i)–(v) M times.

Then, the empirical distribution of $\{\tilde{y}_{t+1}^{[k](s)}, s = b, \dots, M\}$ approximates the predictive density (31) in $z_{t+1} = y_{t+1}^{[k]}$. We set the burn-in parameter $b = 5,000$ and the total number of iterations $M = 10,000$ for the empirical application after checking the convergence. Seven hours of computation were necessary to implement our empirical analysis. By extending the

above forecasting steps up to H step ahead, we obtained the MCMC sample path

$\left\{ \tilde{y}_{t+1}^{[k](s)}, \tilde{y}_{t+2}^{[k](s)}, \dots, \tilde{y}_{t+H}^{[k](s)} \right\}$ for the joint predictive density.

References

- Bearden, W.O. and Etzel, M.J. (1982), "Reference group influence on product and brand purchase decisions," *Journal of Consumer Research*, vol. 9, pp. 183-194.
- Cargnoni, C., Muller, P. and West, M. (1997), "Bayesian forecasting of multinomial time series Through Conditionally Gaussian Dynamic Models," *Journal of the American Statistical Association*, vol. 92, pp. 640-647.
- Carter, C.K. and Kohn, R. (1994), "On Gibbs sampling for state space models," *Biometrika*, vol. 81, pp. 541-553.
- Fruhwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, vol. 15, pp. 193-202.
- Harvey, A.C. and Fernandes, C. (1989), "Time series models for count or qualitative observations," *Journal of Business and Economic Statistics*, vol. 7, pp. 407-422.
- Hoadley, B. (1969), "The compound multinomial distribution and bayesian analysis of categorical data from finite populations," *Journal of the American Statistical Association*, vol. 64, pp. 216-229.
- Hyman, H.H. (1942), "The psychology of status." *Archives of Psychology*, 269, 5-91. Reprint in H. Hyman & E. Singer (Eds.), *Readings in reference group theory and research* (pp. 147-165). New York: Free Press, London: Collier-Macmillan Limited. (Page citations are to the reprint edition).
- Lambert, D. (1992), "Zero-inflated Poisson regression, with application to defects in manufacturing," *Technometrics*, vol. 34, pp. 1-14.
- Ord, K., Fernandes, C., and Harvey, A.C. (1993), "Time series models for multivariate series of count data," Ed. T. Subba Rao, *Developments in Time Series Analysis*, pp. 295-309.
- Park, C.W. and Lessig, V.P. (1977), "Students and housewives: Differences in susceptibility to reference group influence," *Journal of Consumer Research*, vol. 4, pp. 102-110.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society Series B.*, vol. 64, pp. 583-639.
- Terui, N., Ban, M., and Maki, T. (2010), "Finding market structure by sales count dynamics - multivariate structural time series models with hierarchical structure for count data -," *Annals of the Institute of Statistical Mathematics*, vol. 62, pp. 92-107.
- Rossi, P. E, G. Allenby and R. McCulloch (2005), *Bayesian Statistics in Marketing*, John Wiley & Sons, New Jersey.
- West, M. and Harrison, P.J. (1997), *Bayesian Forecasting and Dynamic Models*, 2nd ed., Springer-Verlag, New York.
- West, M., Harrison, P.J., and Migon, H.S. (1985), "Dynamic generalized linear models and

Bayesian forecasting,” *Journal of the American Statistical Association*, vol. 80, pp. 73-97.

Winkelmann, R. (2008), *Econometric Analysis of Count Data (Fifth edition)*, Springer, Heidelberg.

Table 1 Summary of Data

Brand	weekly Sales		weekly Price (/100g)	weekly Display	weekly Feature
	Average	(Variance)	Average	Average	Average
A1	33.74	(1057.47)	90.04	2.09	0.40
B1	74.11	(4287.59)	90.33	2.10	0.40
C1	51.29	(2034.13)	89.59	2.09	0.40
A2	61.54	(2915.08)	87.29	1.52	0.23
B2	87.00	(4470.92)	83.71	2.73	0.23
C2	49.67	(1551.92)	80.62	3.39	0.20
A3	27.22	(578.82)	101.22	0.69	0.19
B3	48.13	(1947.57)	99.31	0.73	0.19
C3	24.42	(959.64)	99.03	0.72	0.18

Table 2-1 Model Comparison: Overall

	ML	DIC	RMSE(sum1)	RMSE(sum2)
Poisson				
Null	150453	-300620	792.71	792.71
Product Category	150558	-300810	648.96	480.06
Maker	150575	-300880	858.24	515.76
Usage	150593	-300995	621.13	449.51
Compound Poisson				
Null	150797	-301170	653.99	653.99
Product Category	150827	-301202	650.41	467.65
Maker	150829	-301235	770.48	490.00
Usage	150868	-301309	612.19	412.85

ML: the log of marginal likelihood, DIC: Deviation information measure

RMSE: the root mean squared errors of 10 step ahead forecasts

Table 2-2 Model Comparison: Decomposition of RMSE

Poisson							
Null		Product Category		Maker		Usage	
Market	222.65	Market	164.36	Market	138.73	Market	171.34
A1	47.52	Category A	65.71	Maker 1	133.65	Usage 1	125.81
A2	31.21	A1	45.75	A1	49.74	A1	38.74
A3	76.93	A2	27.64	B1	62.98	B1	75.44
B1	105.19	A3	9.66	C1	37.20	C1	42.16
B2	78.92	Category B	67.77	Maker 2	141.38	A2	19.21
B3	83.77	B1	71.43	A2	52.44	B2	33.27
C1	54.63	B2	47.72	B2	59.21	C2	19.28
C2	51.56	B3	35.98	C2	31.08		
C3	40.32	Category C	35.41	Maker 3	67.46	Usage 2	45.81
		C1	44.85	A3	22.50	A3	17.19
		C2	19.01	B3	39.22	B3	23.71
		C3	13.68	C3	22.65	C3	9.17
(sum1)	792.71	(sum1)	648.96	(sum1)	858.24	(sum1)	621.13
(sum2)	792.71	(sum2)	480.06	(sum2)	515.76	(sum2)	449.51

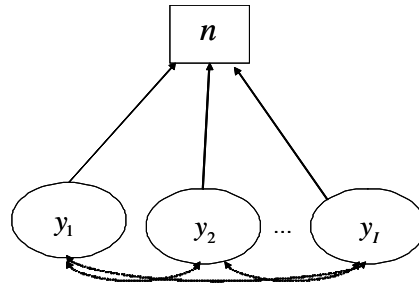
Compound Poisson							
Null		Product Category		Maker		Usage	
Market	305.95	Market	215.32	Market	198.35	Market	203.94
A1	42.34	Category A	49.88	Maker 1	124.41	Usage 1	139.20
A2	93.88	A1	20.04	A1	29.99	A1	26.89
A3	55.41	A2	20.69	B1	63.23	B1	30.30
B1	28.20	A3	25.07	C1	35.25	C1	21.26
B2	52.06	Category B	81.87	Maker 2	75.66	A2	16.61
B3	15.59	B1	50.56	A2	17.35	B2	29.37
C1	22.97	B2	32.68	B2	47.74	C2	20.50
C2	27.33	B3	33.89	C2	13.26		
C3	10.26	Category C	51.01	Maker 3	80.41	Usage 2	60.14
		C1	25.87	A3	14.96	A3	22.01
		C2	19.19	B3	47.40	B3	20.79
		C3	24.34	C3	22.47	C3	21.18
(sum1)	653.99	(sum1)	650.41	(sum1)	770.48	(sum1)	612.19
(sum2)	653.99	(sum2)	467.65	(sum2)	490.00	(sum2)	412.85

Table 3 Estimates of Product Intercept

Category	Brand	Post Mean	Post S.D.
Usage 1	A1	0.17	0.06
	B1	0.81	0.07
	C1	-0.10	0.06
	A2	0.69	0.10
	B2	0.94	0.07
	C2	0.26	0.06
Usage 2	A3	-0.48	0.08
	B3	0.33	0.08
	C3	-1.03	0.08

Figure 1 Market Structure

(a) No Specific Structure



(b) Three-Layer Hierarchical Market Structure

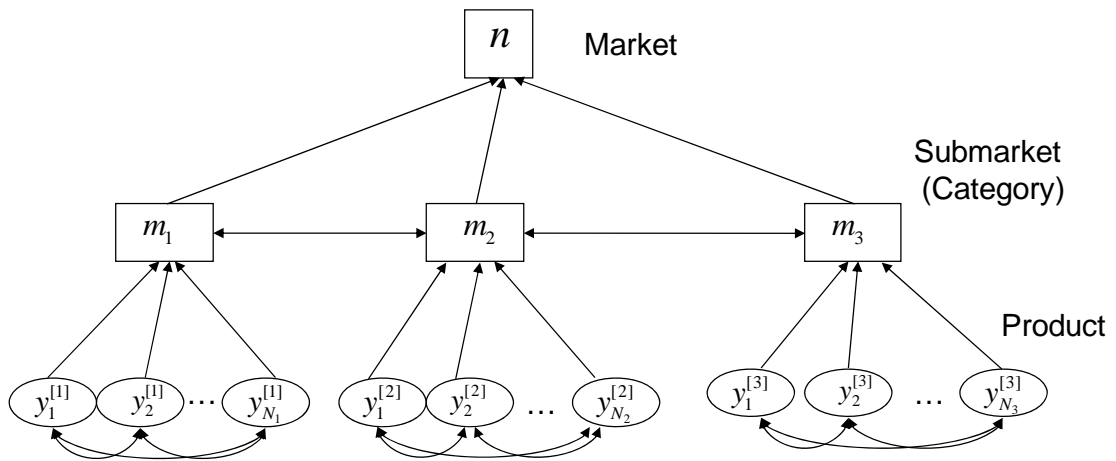
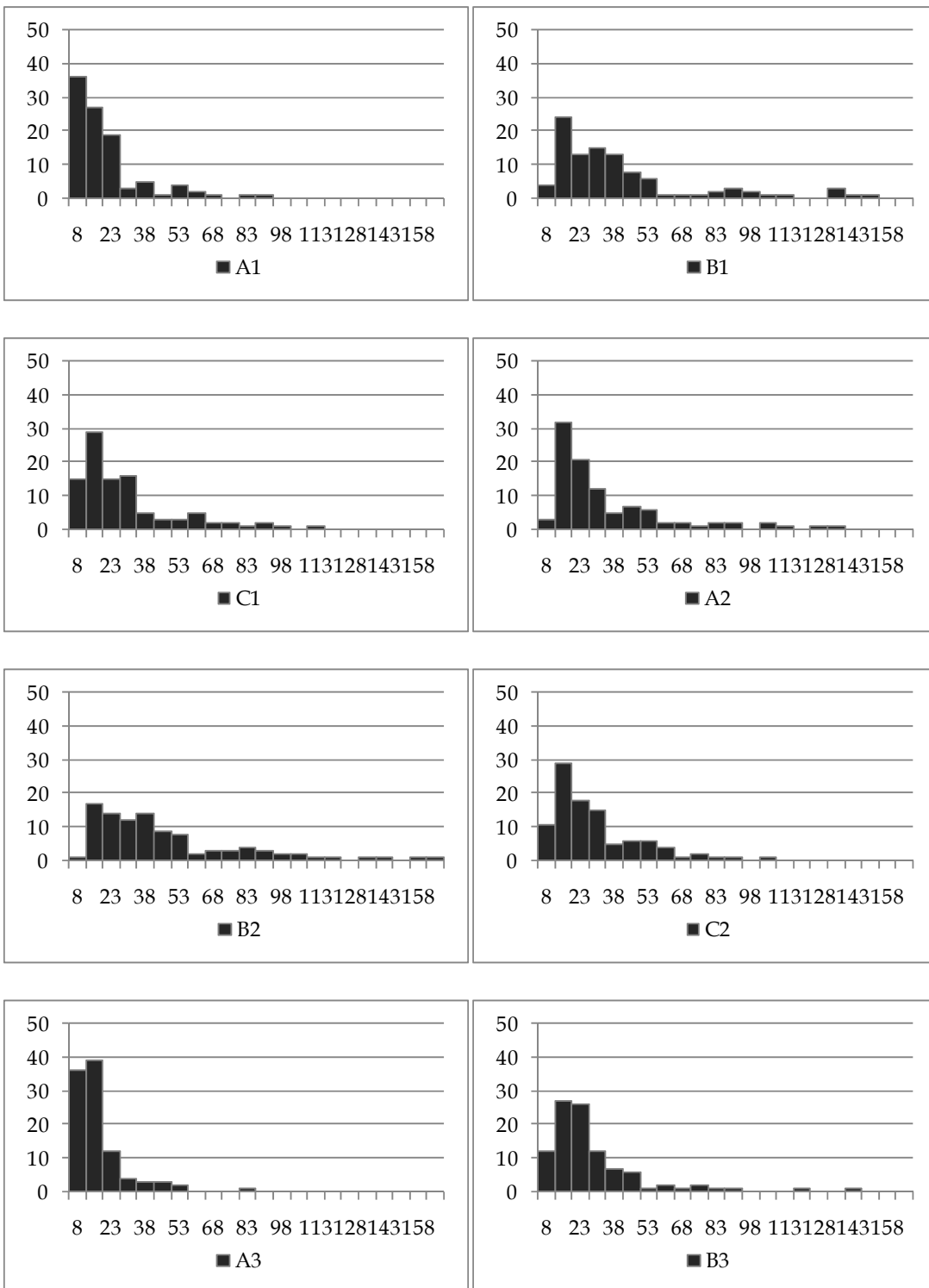


Figure 2 Histogram of Weekly Product Sales Data



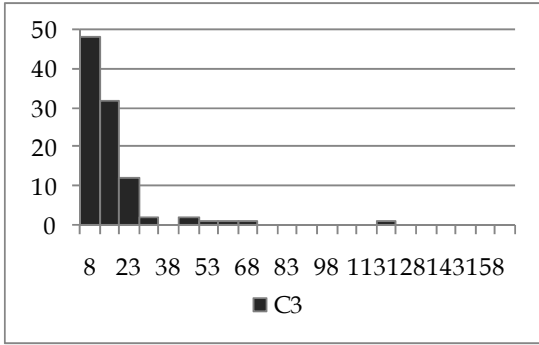


Figure 3 Comparative Models

		Product Category				
		A	B	C		
Maker	1	A1	B1	C1	1	Usage
	2	A2	B2	C2		
	3	A3	B3	C3		

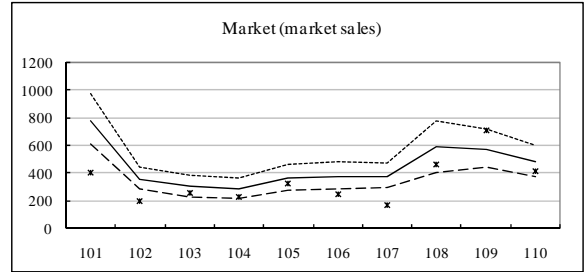
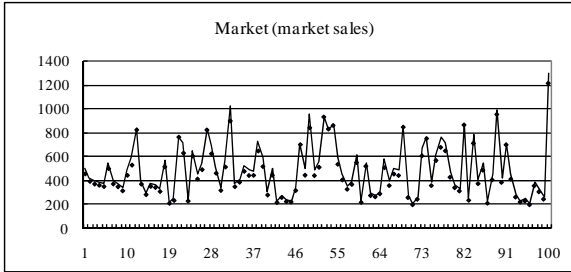
Three makers produce three categories of products and they are classified into two groups by their usages.

Figure 4 In-sample Performance and Forecasting: Market, Submarket, and Product

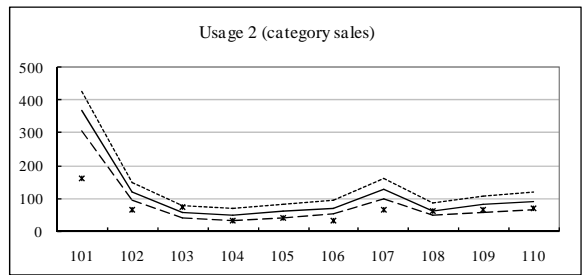
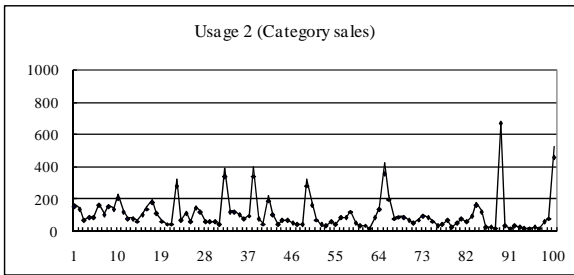
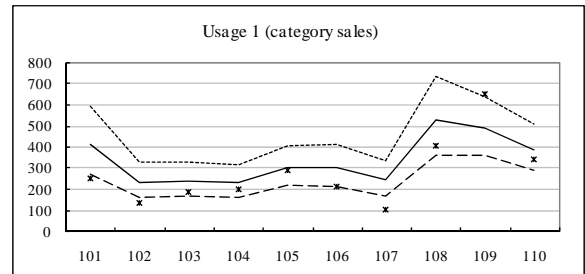
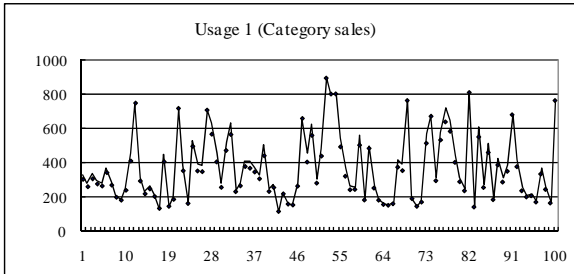
(a) In-sample fit

(b) Forecasting

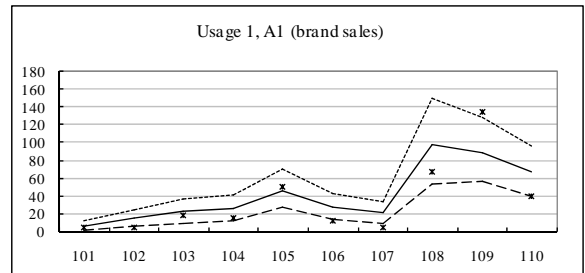
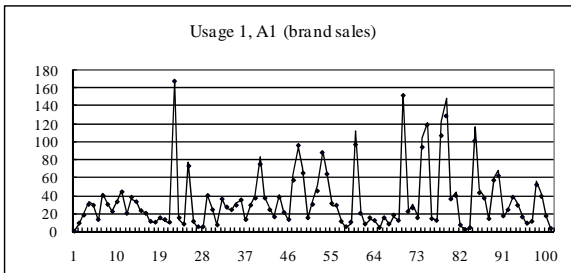
Market Level



Category Level



Product Level



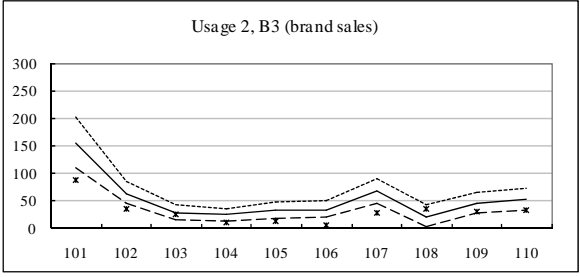
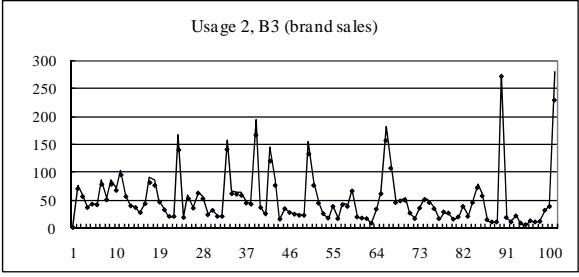
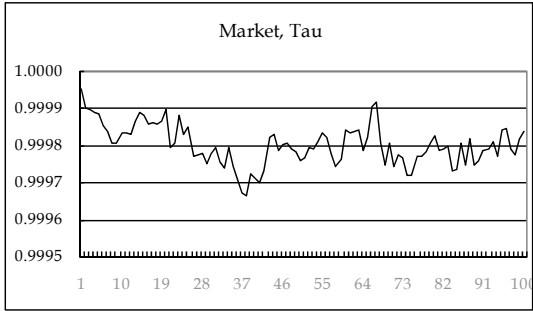
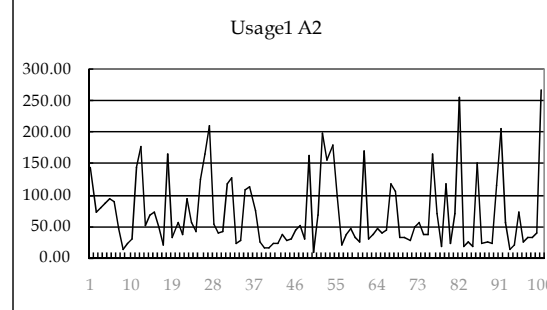
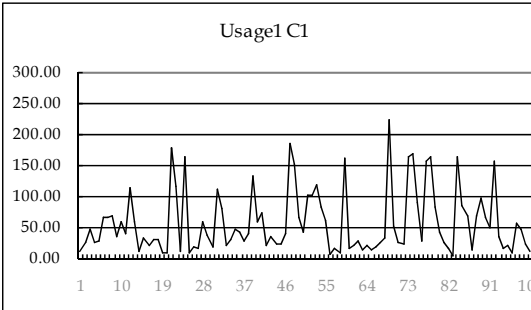
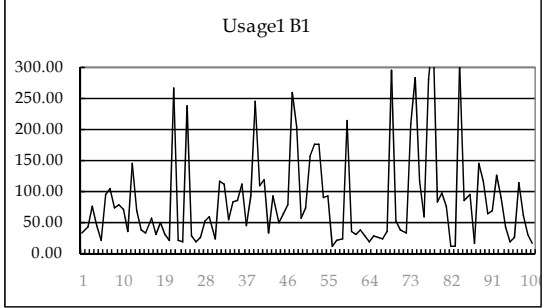
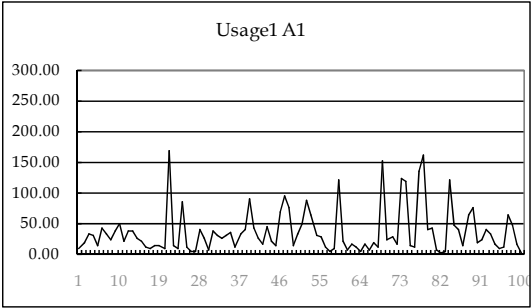


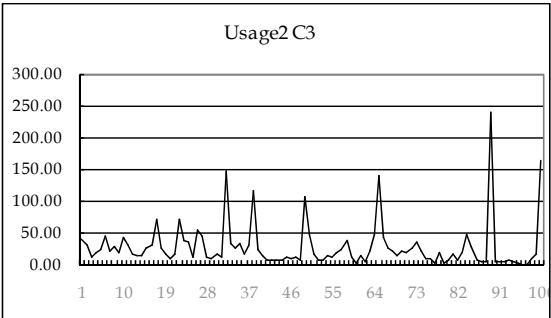
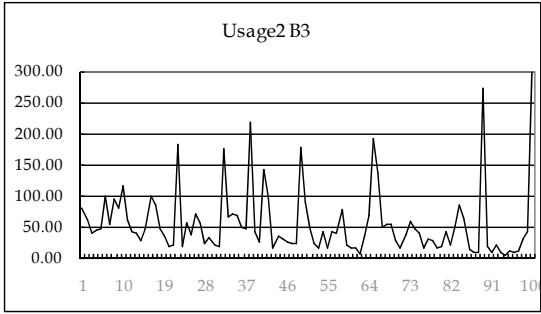
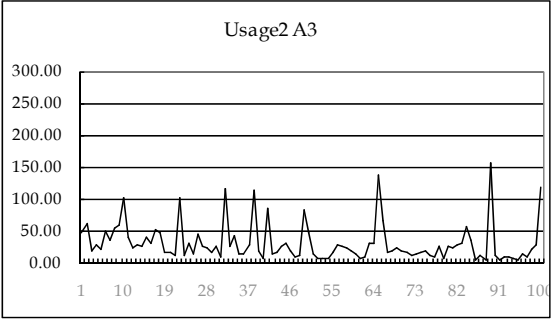
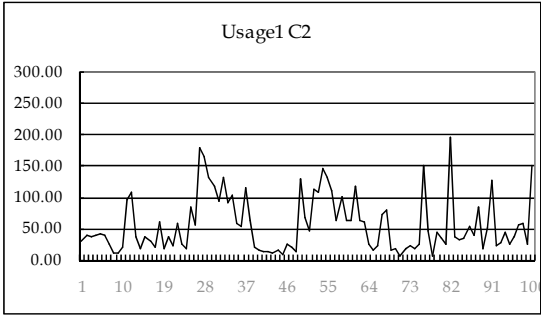
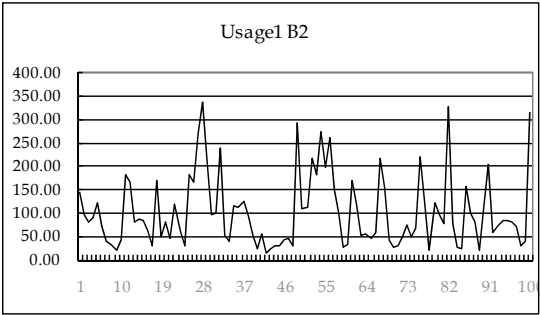
Figure 5 Estimates of Structural Parameter

(a) $\{\tau_t\}$

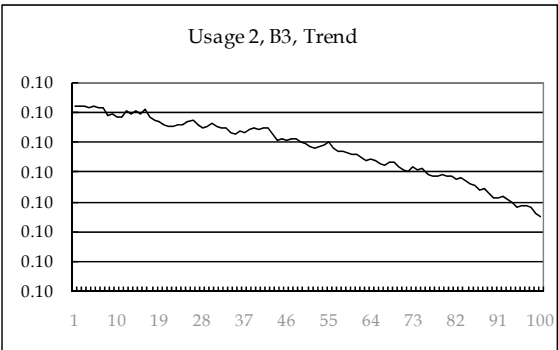
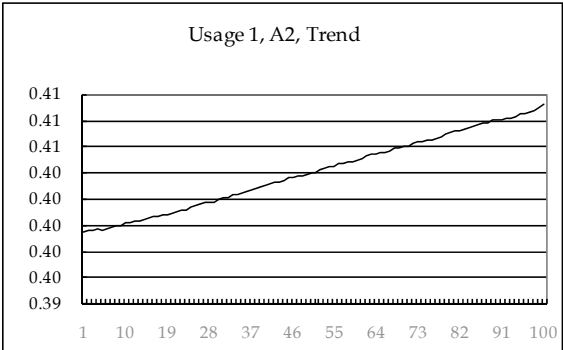


(b) $\{a_{it}\}$





(c) Trend



(d) Response Parameters

