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

Terui, Nobuhiko
Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan
E-mail: Terui@econ.tohoku.ac.jp

Ban, Masataka
Department of Management, Mejiro University, Nakaochiai, Shinjuku-ku, 161-8539, Japan
E-mail: ban@mejiro.ac.jp

Maki, Toshihiko
Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan
E-mail: Terui@econ.tohoku.ac.jp

1. Introduction

In this study, we extend time series model for count data by Terui et al. (2010) so 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 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 usually assumed in economics and marketing. Then, we propose, using Gamma compound Poisson variables for discrete responses having over-dispersion, to build a multivariate time series model with hierarchical structures. We develop the statistical modeling of compound distributions by using hierarchical models rather than the density directly. Our proposed model nests the Terui et al.'s (2010) original model, keeping original parameters in the modeling linked with appropriate covariates at the upper level in the hierarchical models; the lower level model describes the dynamics by setting the state space prior on mixing parameters to generate compound variables. Then we obtain the joint posterior density by the full Bayes MCMC procedure.

2. Microstructure for Generating Over-dispersions

Suppose that there are H numbers of 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

$$x_{iht} \sim \text{Poisson}\left(\lambda_{ih}\right). \tag{1}$$

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

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

where $I\left(h \in C_t\right)$ is an indicator function, taking one when consumer *h* belongs to potential shopper set

C_t, i.e. when she is ready to buy, and zero otherwise, $H_t = \sum_{h=1}^{H} I\left(h \in C_t\right)$ with h^* as the index for reordering of shoppers in the set C_t. In this circumstance, the over-dispersion phenomenon is derived by evaluating the mean and variance of y_{it} as

$$E\left(y_{it}\right) = \sum_{h^*=1}^{H_t} E\left(x_{ih^*t}\right) = \sum_{h^*}^{H_t} \lambda_{ih^*} \equiv \lambda_{it}^* \tag{3}$$

$$\begin{aligned} \mathrm{Var}\left(y_{it}\right) &= \sum_{h^*}^{H_t} \mathrm{Var}\left(x_{ih^*t}\right) + \sum_{h \neq h'} \mathrm{Cov}\left(x_{ih^*t}, x_{ih^{*\prime}t}\right) \\ &= \lambda_{it}^* + \sum_{h \neq h'} \mathrm{Cov}\left(x_{ih^*t}, x_{ih^{*\prime}t}\right) \geq \lambda_{it}^* = E\left(y_{it}\right), \end{aligned} \tag{4}$$

because it holds that $\mathrm{Cov}\left(x_{ih^*t}, x_{ih^{*\prime}t}\right) > 0$ for any pairs of correlated Poisson variables.

That is, equation (4) implies that over-dispersion happens 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 reference group in society and their decision making, going back to Hyman (1942), and its application to social psychology based consumer behavior theory, such as Park and Lessig (1977), and Bearden and Etzel (1982). On the other hand, the Gamma compound Poisson variable with positive parameters $\left(a, \tau\right)$, denoted as $y \sim$ Compound Poisson$\left(a, \tau\right)$, satisfies over-dispersion because it has the first two moments

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

3. Models for Defining Market Structure

We assume that there are I products in the market and denote y_{it} by the number of sales for the product i at time t $(t = 1, \ldots, T)$, which follows the Gamma compound Poisson distribution independently with a time varying parameter $\left(a_{it}, \tau_t\right)$ for $i = 1, \ldots, I$ when there is no competitive relationship to each other. Then we obtain the Gamma compound Poisson distribution with $\left(\sum a_i, \tau\right)$ 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. That is, we have marginal distributions for product and market sales by

$$y_{it} \sim \text{Compound Poisson}\left(a_{it}, \tau_t\right), \quad n_t \sim \text{Compound Poisson}\left(a_t^*, \tau_t\right), \tag{6}$$

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

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

$$\tilde{y}_t \mid n_t \sim \text{Dirichlet Compound Multinomial}\left(\tilde{y}_t \mid n_t, \tilde{a}_t\right), \tag{7}$$

where $\tilde{a}_t = (a_{1t},...,a_{It})'$. The sequential use of (11) and (12) produces the joint distribution for market sale and product sales,

$$p\left(n_t, \tilde{y}_t \mid \tilde{a}_t, \tau_t\right) = p\left(n_t \mid a_t^*, \tau_t\right) p\left(\tilde{y}_t \mid n_t, \tilde{a}_t\right). \tag{8}$$

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

In contrast, Terui et al. (2010) proposed the dynamic generalized linear model based on Poisson variables without over-dispersions, which represents a macro model 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 they proposed a multivariate time series model with hierarchical structure on the discrete outcomes. That is, we have marginal distributions, $y_{it} \sim \text{Poisson}(\lambda_{it})$, $n_t \sim \text{Poisson}\left(\lambda_t^*\right)$ and conditional distribution $\tilde{y}_t \mid n_t \sim \text{Multinomial}\left(y_{it} \mid n_t, \tilde{\pi}_t\right)$, where $\lambda_t^* = \sum_{it}^{I} \lambda_{it}$ and $\tilde{\pi}_t = \left\{\pi_{it} = \lambda_{it} \Big/ \sum_{it}^{I} \lambda_{it}, i = 2,...,I\right\}$. The likelihood at time *t* is defined by

$$p\left(n_t, \tilde{y}_t \mid \lambda_t^*, \tilde{\pi}_t\right) = p\left(n_t \mid \lambda_t^*\right) p\left(\tilde{y}_t \mid n_t, \tilde{\pi}_t\right), \tag{9}$$

where the induced parameters λ_t^* and $\tilde{\pi}_t$, respectively, mean expected total sales and market shares for each product. These parameters, after making appropriate transformations $\tilde{\eta}_t = \left(\tilde{\eta}_t^1\left(\lambda_t^*\right), \tilde{\eta}_t^2\left(\tilde{\pi}_t\right)'\right)'$ so as to be Gaussian, are embedded in the state space model

$$\begin{cases} \tilde{\eta}_t = F_t \tilde{\theta}_t + \tilde{v}_t, & \tilde{v}_t^s \sim N(0,V) \\ \tilde{\theta}_t = H_t \tilde{\theta}_{t-1} + \tilde{w}_t, & \tilde{w}_t \sim N(0,W) \end{cases}, \tag{10}$$

where we set

$$\tilde{\eta}_t^1 = \log\left(\lambda_t^*\right), \tag{11}$$

for the parameters on $\{n_t\}$, and $\eta_{it}^2(\pi_{it})$, *i*th element of $\tilde{\eta}_t^2$, is connected to the multinomial parameters π_{it} by the relationship

$$\pi_{it} = \exp\left(\eta_{it}^{2}\right) \Big/ \left(\exp\left(\eta_{1t}^{2}\right) + \ldots + \exp\left(\eta_{I-1t}^{2}\right) + 1\right) \quad \text{for} \quad i = 1, \ldots, I-1. \qquad (12)$$

Thus the likelihood function constituted by (9) and the state space prior (10) for transformed parameters constitute the dynamic generalized linear models. We note that *i*th element in the first equation of (10) includes covariate X_{it} and stochastic trend μ_{it} terms as $F_t\tilde{\theta}_t = \mu_t + X_t'\tilde{\beta}_t$ and $F_t = (1, X_{it}'), \tilde{\theta}_t = \left(\mu_{it}, \tilde{\beta}_{it}\right)'$, and the settings of the order of stochastic trend and the transition of time varying coefficient define H_t in the system equation of (10).

Instead of using these densities, we take the 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 $\left(\lambda_t^*, \tilde{\pi}_t\right)$ in the MCMC process to define the likelihood for mixing parameters $\left(\tilde{a}_t, \tau_t\right)$:

$$p\left(\lambda_t^*, \tilde{\pi}_t \mid \tilde{a}_t, \tau_t\right) = \text{Gamma}\left(\lambda_t^* \mid \tilde{a}_t, \tau_t\right) \times \text{Dirichlet}\left(\tilde{\pi}_t \mid \tilde{a}_t\right). \qquad (13)$$

Then, we have two kinds of prior on $\left(\lambda_t^*, \tilde{\pi}_t\right)$, i.e., state space prior (10) for dynamics and (13) for over-dispersions. Thus we combine the likelihood function derived from (9) with priors (10) and (13) to derive the posterior density:

$$\begin{aligned} & p\left(\left\{\lambda_t^*, \tilde{\pi}_t\right\}, \left\{\tilde{\theta}_t\right\}, \left\{\tilde{a}_t, \tau_t\right\} \mid \left\{n_t, \tilde{y}_t\right\}\right) \\ & \propto \left[p\left(\left\{n_t\right\} \mid \left\{\lambda_t^*\right\}\right) p\left(\left\{\tilde{y}_t\right\} \mid \left\{n_t\right\}, \left\{\tilde{\pi}_t\right\}\right) p\left(\left\{\lambda_t^*, \tilde{\pi}_t\right\} \mid \left\{\tilde{\theta}_t\right\}\right) p\left(\left\{\tilde{\theta}_t\right\}\right) \right] \\ & \quad \times p\left(\left\{\lambda_t^*, \tilde{\pi}_t\right\} \mid \left\{\tilde{a}_t, \tau_t\right\}\right) p\left(\left\{\tilde{a}_t, \tau_t\right\}\right). \end{aligned} \qquad (14)$$

We note that $p\left(\left\{\tilde{\theta}_t\right\}\right)$ contains prior structure $p\left(\tilde{\theta}_t \mid \tilde{\theta}_{t-1}\right)$ in the state space models (15) and also we set the state space prior on $\left(\tilde{a}_t, \tau_t\right)$ as

$$\begin{cases} \tilde{z}_t = \tilde{f}_t + \tilde{\varepsilon}_t, & \tilde{\varepsilon}_t \sim N\left(0, V_z\right) \\ \tilde{f}_t = H_{zt}\tilde{f}_{t-1} + \tilde{v}_t, & \tilde{v}_t \sim N\left(0, W_z\right), \end{cases} \qquad (15)$$

where $\tilde{z}_t = \left(\log\left(a_{1t}\right), \ldots, \log\left(a_{It}\right), \log\left(\tau_t\right)\right)$ and $\tilde{f}_t$ is the mean function.

However, full description of joint posterior density will be provided after specifying our state space models in higher order structure model, as we apply it in the empirical application. Higher Order Structure is defined in a similar way.

5. Estimation and Forecasting

In addition to the standard Bayesian inference on state space modeling by dynamic linear models (DLM) by

West and Harrison (1997), we use the MCMC approach to estimate the model, using Metropolis–Hastings sampling specifically for the conditional posterior density of link functions for the higher level model and mixing parameters for the lower level model, respectively, by

$$p\left(\tilde{\eta}_t^s \mid X_t^s, \tilde{\theta}_t^s, V^s, \{\tilde{y}_t\}\right) \propto p\left(d_t^s \mid \tilde{\eta}_t^s\right) p\left(\tilde{\eta}_t^s \mid X_t^s, \tilde{\theta}_t^s, V^s\right) \tag{16}$$

$$p\left(\tilde{z}_t \mid \tilde{\eta}_t, \tilde{f}_t, V_z, W_z\right) \propto p\left(\tilde{\eta}_t \mid \tilde{z}_t\right) p\left(\tilde{z}_t \mid \tilde{f}_t, V_z, W_z\right). \tag{17}$$

Now we have two types of dynamic generalized linear models in the upper level for the original model as well as in the lower level for mixing parameters for compound distributions. Through both models, once the values of $\tilde{\eta}_t^{s(k)}$ for (16) and $\tilde{z}_t$ for (17) are given, the structural equations coupled with the system equations in their state space priors constitute conventional Gaussian state space models. The multi-move sampler by Crater and Kohn (1994) and Fruhwirth-Schnatter(1994) is used to sample the state vectors.

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

$$\begin{aligned}\int p(\tilde{y}_{T+1} \mid \tilde{\eta}_{T+1}, \text{data})\, p(\tilde{\eta}_{T+1} \mid \tilde{\theta}_{T+1}, \tilde{z}_{T+1}, V, W, \text{data})\, p(\tilde{\theta}_{T+1} \mid \tilde{\theta}_T, V, W, \text{data}) \\ \times p(\tilde{z}_{T+1} \mid \tilde{z}_T, V_z, W_z, \text{data})\, d\tilde{\theta}_T d\tilde{z}_T dV dW dV_z dW_z,\end{aligned} \tag{18}$$

where "data" means the observed data $\left(\{\tilde{y}_t\}, \{X_t\}\right)$, and $p\left(\tilde{\eta}_{T+1} \mid \tilde{\theta}_{T+1}, \tilde{z}_{T+1}, V, W, \text{data}\right)$ is the conditional predictive density of upper level's link parameter when the predicted state, predicted mixing parameters and state space covariance are given. To evaluate this density, we first define the predictive likelihood of $\tilde{\eta}_{T+1}$ conditional on $\tilde{z}_{t+1}$ by

$$p\left(\tilde{z}_{T+1} \mid \tilde{\eta}_{T+1}\right) = \text{Gamma}\left(\sum_i^I \lambda_{iT+1} \mid \sum_i^I a_{iT+1}, \tau_{T+1}\right) \text{Dirichlet}\left(\{\pi_{iT+1}\} \mid \{a_{iT+1}\}\right). \tag{19}$$

6. Empirical Application

We use the store level scanner - POS: point of sales - time series in the curry roux category that were applied to our previous model in Terui et al. (2010) for the c omparison with the model with over-dispersion. The weekly series comprises three manufacturers that produce three products each, for a total of n ine products during 110 weeks. The first 100 weeks are used for estimation and the last 10 weeks are reserve d 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 $\{X_{it}\}$. The displa y and feature are binary data taking one when it was on, and zero when it was off. The logs of price data are used. Table 1 displays the summary statistics of these variables.

Figure Market Structure - Instant Curry Data

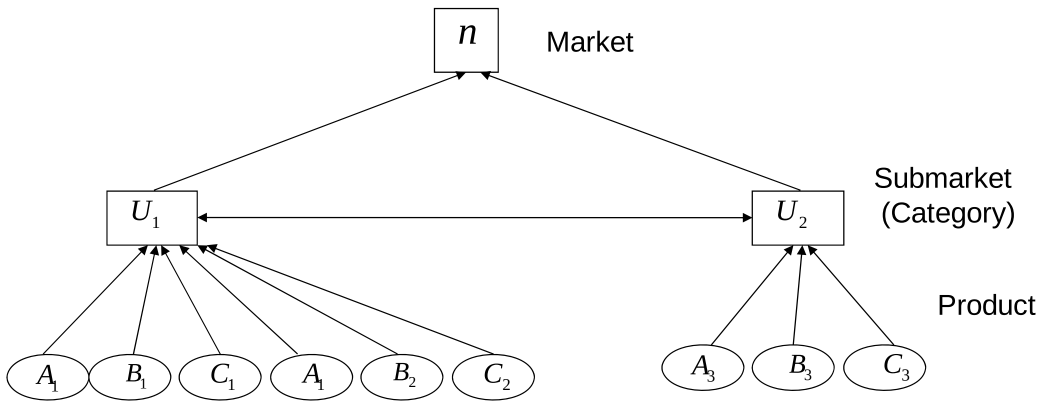

Table Model Specification

	LML	DIC	RMSE(sum1)	RMSE(sum2)
Compound Poisson				
Null	150459	-300686	728.330	728.330
Usage	150706	-301278	613.740	426.120
Piosson				
Null	150453	-300620	792.709	792.709
Usage	150593	-300995	621.130	449.510

REFERENCES (RÉFERENCES)

Bearden,W.O. and Etzel,M.J. (1982). “Reference group influence on product and brand purchase decisions,” *Journal of Consumer Research*, 9, pp.183-194.

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 Associaton, vol.64, pp.216-229.

Hyman, Herbert 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.

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.

West, M. and Harrison, P.J. (1989), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.

West, M., Harrison, P.J. and Migon, H.S. (1985), “Dynamic generalized linear models and Bayesian fore casting,” *Journal of the American Statistical Association*, vol. 80, pp.73-97.