Ann Inst Stat Math (2008) 60:763–780
DOI 10.1007/s10463-008-0194-0

Measuring the baseline sales and the promotion effect for incense products: a Bayesian state-space modeling approach

Tomohiro Ando

Received: 16 October 2007 / Revised: 11 April 2008 / Published online: 2 September 2008

Abstract One of the most important research fields in marketing science is the analysis of time series data. This article develops a new method for modeling multivariate time series. The proposed method enables us to measure simultaneously the effectiveness of marketing activities, the baseline sales, and the effects of controllable/uncontrollable business factors. The critical issue in the model construction process is the method for evaluating the usefulness of the predictive models. This problem is investigated from a statistical point of view, and use of the Bayesian predictive information criterion is considered. The proposed method is applied to sales data regarding incense products. The method successfully extracted useful information that may enable managers to plan their marketing strategies more effectively.

Keywords Bayesian method · General state-space models · Marketing

1 Introduction

A central concern in the planning of any marketing strategy is the creation of a sustainable competitive advantage. To create and possess a competitive advantage, understanding the structure and nature of the market from long-term perspective is an important research area in marketing science. This paper tries to shed light on the following research question: how can we measure the effectiveness of marketing activities, the baseline sales, and the effects of controllable/uncontrollable business factors using available information sources?

T. Ando (✉)
Graduate School of Business Administration, Keio University,
2-1-1 Hiyoshi-Honcho, Kohoku-ku, Yokohama, Kanagawa 223-8523, Japan
e-mail: andoh@hc.cc.keio.ac.jp

One of the major sources of information that measures the past performance of an individual firm is time series data, which can include sales, market shares, and additional marketing-mix variables such as advertising, pricing promotion, display promotion, etc. Various types of models were applied to investigate the relationship between marketing activities and their effects on performance (Bass 1969; Beckwith 1972; Wildt 1974; Hanssens 1980; Blattberg et al. 1981; Leone 1983; Neslin et al. 1985; Gupta 1988; Neslin 2002).

In marketing research fields, baseline sales—the amount of sales when there are no marketing promotions (Abraham and Lodish 1993)—have received considerable attention in recent years (Abraham and Lodish 1993; Tellis et al. 1995; Ando 2006a). Marketing managers widely use baseline sales to assess the profitability and effectiveness of marketing activities by investigating how promotions can impact baseline sales over time.

The main aim of this paper is to develop a method for modeling multivariate time series within the general framework of state-space modeling (Kitagawa 1996; Kitagawa and Gersch 1996). State-space models have been applied to a number of studies to investigate the effectiveness of marketing activities (Kondo and Kitagawa 2000; Kitagawa et al. 2003; Lee et al. 2003; Pauwels et al. 2004; Sato et al. 2004; Van Heerde et al. 2004a,b; Yamaguchi et al. 2004; Ando 2006a). For instance, Xie et al. (1997) and Naik et al. (1998) employed state-space models to estimate the Bass model and the modified Nerlove–Arrow model. Introducing the concept of the 'half-life' of an advertising campaign, Naik (1999) also utilized a state-space model. Neelamegham and Pradeep (1999) and Ando (2006a) applied a general state-space model to predict sales for movies and everyday foods. For the use of time series techniques in a wide range of marketing research, we refer to Dekimpe and Hanssens (2000).

Ando (2006a) developed the method that simultaneously measures the baseline sales and the effectiveness of marketing activities within the framework of Bayesian general state-space modeling. The method is also useful for predicting future sales through the consideration of several factors, such as marketing promotions (temporary price cuts, display promotions, points-of-purchase, advertising catalogs, etc.) and certain uncontrollable business factors (the day of the week, the weather conditions, the season, events, etc.). Such information assists managers not only in planning their marketing strategy but also in planning their strategies for research and development, inventory management, manpower use, and so on.

In contrast to Ando's (2006a) study, where Poisson distribution is employed for predicting the sales, this paper extends this method by allowing various types of distributions. An empirical analysis clearly shows an improvement of Ando's (2006a) method in the sense that the proposed model obtained better model evaluation score, described below.

In the model building process, the Bayesian approach via the Markov chain Monte Carlo (MCMC) method is implemented for estimating model parameters. We do this because the likelihood function depends on integrals of high dimensions. The MCMC method has played a major role in the recent advances in Bayesian analyses of time series models. Fortunately, it is not a computationally intensive task—thanks to increased access to appropriate computational tools.

The critical issue in the model construction process is the method for evaluating the usefulness of the constructed models. Although progress in MCMC simulation methods has made flexible statistical modeling popular, the assessment of the usefulness of the estimated model is still under development. This paper investigates this problem from a statistical point of view and uses the Bayesian predictive information criterion (BPIC; Ando 2007). The advantage of the BPIC is that it is easily calculated from the samples generated by a MCMC simulation. As an alternative criterion for selecting a model, one might consider using the deviance information criterion (DIC; Spiegelhalter et al. 2002). However, Robert and Titterington (2002) and Ando (2007) have pointed to some theoretical problems in the DIC. One of the most crucial issues is over-fitting. To overcome theoretical problems in the DIC, Ando (2007) proposed the use of the Bayesian predictive information criterion.

One of contributions of this article in marketing research is the introduction of a new Bayesian general state-space modeling method. Therefore, various types of probability distributions are available to express the randomness of the sales. The use of the Bayesian predictive information criterion in marketing research is also a new concept. Thanks to this criterion, we can evaluate the goodness-of-fit of the estimated models. Furthermore, to our knowledge, no empirical study has conducted an analysis of the sales of Japanese incense products.

This article is organized into four sections. In Sect. 2, we present the method for modeling multivariate time series within the framework of general state-space modeling. Section 3 applies the proposed method to the daily sales of Japanese incense products. Conclusions are given in Sect. 4.

2 Methodology

2.1 Preliminaries

It is useful to begin with a brief review of the general state-space models (Kitagawa 1987; Kitagawa and Gersch 1996). The general state-space model consists of two stochastic components: an observation equation and a system equation:

$$\begin{cases} \text{Observation equation}: & \boldsymbol{y}_t \sim f(\boldsymbol{y}_t|F_t, \boldsymbol{h}_t, \ldots, \boldsymbol{h}_1), \\ \text{System equation}: & \boldsymbol{h}_t \sim f(\boldsymbol{h}_t|F_{t-1}, \boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_1), \end{cases}$$

where F_t denotes the history of the information sequence up to time t, a sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ is the observable time series while a sequence $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots$, so-called a state vector, is unobserved. Here, $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{pt})'$ is the p-dimensional vector, $\boldsymbol{h}_t = (h_{1t}, \ldots, h_{qt})'$ is the q-dimensional vector, $f(\boldsymbol{y}_t|F_t, \boldsymbol{h}_t, \ldots, \boldsymbol{h}_1)$ and $f(\boldsymbol{h}_t|F_{t-1}, \boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_1)$ are the conditional distribution of $\boldsymbol{y}_t$ given F_t, $\boldsymbol{h}_t, \ldots, \boldsymbol{h}_1$ and of $\boldsymbol{h}_t$ given F_{t-1}, $\boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_1$, respectively. The main focus concerns how to construct these two equations so that the model captures the true structure governing the time series of $\boldsymbol{y}_t$.

2.2 Model description

In this paper, we focus on p-dimensional time series data for daily sales of incense products in stores. Given a mean structure of total sales y_{jt}, say λ_{jt}, we shall decompose it into the baseline sales and other components by incorporating the covariate effects:

$$\lambda_{jt}(h_{jt}, \boldsymbol{\beta}_j, \boldsymbol{x}_{jt}) = h_{jt} + \sum_{a=1}^{b} \beta_{ja} x_{jat} = h_{jt} + \boldsymbol{\beta}_j \boldsymbol{x}_{jt}, \tag{1}$$

where h_{jt} is the baseline sales effect, while $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jb})'$ and $\boldsymbol{x}_{jt} = (x_{j1t}, \ldots, x_{jbt})'$ are the b-dimensional vector of unknown parameters to be estimated and the b-dimensional covariate vector, respectively. The covariate vector may include the information on some marketing-mix variables, price levels, price discount percentages, features, advertising, displays, post-promotion dips, the day of the week, the weather, the season, the regulatory, and so on. The dimension b therefore might depend on data availability.

The purpose of analysis also affects the dimension of covariate vector. Consider, for example, we want to quantify an impact of competitor's marketing action on the total sales y_{jt}. In such a case, incorporating competitor's marketing-mix variables (e.g., price discount rate) into the model (1), the sensitivity of the total sales to competitor's marketing action could be measured by its coefficient β. Although the baseline sales h_{jt} do not contain the effects of competitor's marketing action explicitly, the competitor's marketing actions are implicitly affecting the each of baseline sales through the information on total sales. Because we model the baseline sales (and also a mean structure of total sales, λ_{jt}) jointly, the baseline sales and the mean structures of total sales describe a competitive relation in the competitive market. Therefore, we can learn the competitive market with some knowledge of interactive structure between sales.

We are usually not sure about the distribution of daily sales y_{jt}; we therefore shall consider several density functions. Because the sales data take positive values, truncated distributions are used.

$$\begin{aligned}
&\text{Truncated normal :}\\
&\quad f_N(y_{jt}|\mu_{jt}, \sigma_j^2) = I(y_{jt} > 0) \cdot \frac{1}{2\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(y_{jt}-\mu_{jt})^2}{2\sigma_j^2}\right\},\\
&\text{Truncated Student } t \text{ :}\\
&\quad f_{St}(y_{jt}|\mu_{jt}, \sigma_j^2, \nu_j) = I(y_{jt} > 0) \cdot \frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{2\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu_j}{2}\right)\sqrt{\nu_j\sigma_j^2}} \left\{1 + \frac{(y_{jt}-\mu_{jt})^2}{\sigma_j^2\nu_j}\right\}^{-\frac{\nu_j+1}{2}},\\
&\text{Truncated Cauchy :}\\
&\quad f_C(y_{jt}|\mu_{jt}, \sigma_j^2) = I(y_{jt} > 0) \cdot \frac{1}{2\pi\sigma_j} \left\{1 + \frac{(y_{jt}-\mu_{jt})^2}{\sigma_j^2}\right\}^{-1},
\end{aligned} \tag{2}$$

where $I(y_{jt} > 0)$ is the indicator function, takes value one if $y_{jt} > 0$ and zero otherwise, $\mu_{jt} := \lambda_{jt}(h_{jt}, \boldsymbol{\beta}_j, \boldsymbol{x}_{jt})$ is the mean parameter given in (1), s_j^2 is the variance parameter and ν_j is the degrees of freedom of Student-t distribution. Note that we can also consider other distributions. Hereafter, for the simplicity of presentation, we denote these densities by $f(y_{jt}|\boldsymbol{x}_j, h_{jt}, \boldsymbol{\gamma}_j)$, where $\boldsymbol{\gamma}_j$ is the unknown parameter vector associated with each density function. In contrast to Ando's (2006a) study, where Poisson distribution is employed, this paper allows various types of distributions. Under the data availability, instead of the sales, we can therefore analyze the market (also category) share of each product by using the multinomial logit/probit density for $\boldsymbol{y}_t$.

It is assumed that the state variable h_{jt}, the baseline sales effect for the jth store, follows the rth order trend model:

$$\Delta^r h_{jt} = \varepsilon_{jt},$$

where Δ ($\Delta h_{jt} = h_{jt} - h_{j,t-1}$) is the difference operator (e.g., Kitagawa and Gersch 1996) and $\varepsilon_{jt} \sim N(0, \sigma_{jj})$ is a Gaussian white noise sequence. For $r = 1$, the baseline sales become a well-known random walk model, $h_{jt} = h_{j,t-1} + \varepsilon_{jt}$, For $k = 2$, the model becomes $h_{jt} = 2h_{j,t-1} - h_{j,t-1} + \varepsilon_{jt}$. Another expression of the rth order trend model is

$$h_{jt} = \sum_{s=1}^{r} c_s \times B^s h_{jt} + \varepsilon_{jt},$$

where B ($B^1 h_{jt} = h_{j,t-1}$) is the backshift operator and $c_s = (-1)^{s-1} \times_r C_i$ are binomial coefficients (e.g., Kitagawa and Gersch 1996).

It is natural to assume that the daily sales of each store are mutually dependent on each other. Following Ando (2006a), we therefore introduce the correlation between the noises ε_{jt} and ε_{kt}: $\text{Cov}(\varepsilon_{jt}, \varepsilon_{kt}) = \sigma_{jk}$.

Summarizing the above specifications, we then formulate the following observation and system equations:

$$\begin{aligned} \boldsymbol{y}_t &\sim f(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{h}_t; \boldsymbol{\gamma}), \quad j = 1, \ldots, p, \\ \boldsymbol{h}_t &\sim f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_{t-r}; \Sigma), \quad \Sigma = (\sigma_{ij}), \end{aligned} \tag{3}$$

where $f(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{h}_t; \boldsymbol{\gamma})$ with $\boldsymbol{x}_t = (\boldsymbol{x}'_{1t}, \ldots, \boldsymbol{x}'_{pt})'$ is the p-dimensional density function specified by the components $f(y_{jt}|\boldsymbol{x}_j, h_{jt}; \boldsymbol{\gamma}_j)$ in (2). The system model $f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_{t-r}; \Sigma)$ is the p-dimensional normal density with the mean $\boldsymbol{h}_t = \sum_{s=1}^{r} c_s \times B^s \boldsymbol{h}_t$ and covariance matrix Σ.

The next problem is how to estimate unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \text{vech}(\Sigma))'$ with $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_1, \ldots, \boldsymbol{\gamma}'_p)'$. This problem will be investigated in the following section.

2.3 Bayesian inference via MCMC

As shown in the following equation, the likelihood function depends on the high-dimensional integrals:

$$L(D_n|X_n, \boldsymbol{\theta}) = \prod_{t=1}^{n} f(\boldsymbol{y}_t|F_{t-1}, \boldsymbol{x}_t, \boldsymbol{\theta})$$
$$= \prod_{t=1}^{n} \left[\int \prod_{j=1}^{p} f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j) f(\boldsymbol{h}_t|F_{t-1}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{h}_t \right],$$

where $D_n = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ and $X_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and F_{t-1} denotes the history of the observation sequence up to time $t-1$ (See for e.g., Chib et al. 2002; Kitagawa 1987; Kitagawa and Gersch 1996; Tanizaki and Mariano 1998).

The source of the problem is that we cannot express the density $f(\boldsymbol{h}_t|F_{t-1}, \boldsymbol{\theta})$ in the closed form. It is therefore obvious that the maximum likelihood estimation of the models is very difficult. In contrast, the Bayesian treatment of this inference problem relies solely on the theory of probability. It allows us to estimate the model parameters easily because the inference can be done without evaluating the likelihood function. In particular, the Bayesian approach via the MCMC algorithm is useful for estimating model parameters. Details on the MCMC method can be found in Carlin and Louis (1996), Gilks et al. (1996), Tierney (1994) and in references given therein.

In the Bayesian approach via the MCMC method, both $\boldsymbol{\theta}$ and the state vector $\boldsymbol{h}_t$ are considered to be model parameters. An inference on the parameters is conducted by producing a sample from the posterior distribution

$$\pi(\boldsymbol{\theta}, \boldsymbol{h}|D_n, X_n) \propto \pi(\boldsymbol{\theta}) \times \prod_{t=1}^{n} \prod_{j=1}^{p} f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j) f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_{t-r}; \Sigma).$$

To complete the Bayesian model, we now formulate a prior distribution on the parameters. A prior independence of the parameters is assumed: $\pi(\boldsymbol{\theta}) = \pi(\Sigma)\pi(\boldsymbol{\gamma})$, $\pi(\boldsymbol{\gamma}) = \prod_{j=1}^{p} \pi(\boldsymbol{\gamma}_j)$.

Decomposing the covariance matrix Σ as a product of the variance and the matrix of correlations into $\Sigma = RCR$, where $R = (r_{ij})$ is a diagonal variance matrix and $C = (c_{ij})$ is the correlation matrix (Barnard et al. 2000), we formulate a prior distribution on r_{ii} $(i = 1, \ldots, p)$ and the elements $\{c_{ij}, i < j\}$. Following Ando (2006a), we assume that each of the elements $\{r_{ii}; i = 1, \ldots, p\}$ is independently and identically distributed. We then place a gamma prior with parameters a and b on the diagonal entries of Σ:

$$\pi(\sigma_{ii}) = \frac{b^a}{\Gamma(a)} (\sigma_{ii})^{a-1} \exp\{-b\sigma_{ii}\}, \quad i = 1, \ldots, p,$$

which implies

$$\pi(r_{ii}) = \pi(\sigma_{ii})\frac{\mathrm{d}\sigma_{ii}}{\mathrm{d}r_{ii}} = \frac{2b^a}{\Gamma(a)}(r_{ii})^{2a-1}\exp\{-br_{ii}^2\}.$$

To make the prior uninformative, we shall take $a = 10^{-10}$ and $b = 10^{-10}$. For the prior distribution of $\{c_{ij}, i < j\}$, a uniform prior distribution $U[-1, 1]$ is employed.

When we specify the Student-t density for y_{jt}, the unknown parameter vector $\boldsymbol{\gamma}_j$ include the degree of freedom ν_j as well as the coefficient $\boldsymbol{\beta}_j$ and s_j^2. For the coefficient $\boldsymbol{\beta}_j$, the b-dimensional uninformative normal prior $N(\mathbf{0}, 10^{10} \times I_b)$ is utilized. In addition to σ_{ii}, a gamma prior with parameters $a = b = 10^{-10}$ is used for s_j^2. A uniform prior distribution is used $U[2, 100]$ for $\pi(\nu_j)$. The same prior distributions are employed for other density cases.

The MCMC algorithm is then summarized as follows.

MCMC sampling algorithm:

Step 1. Initialize $\boldsymbol{\theta}$ and $\boldsymbol{h}$.
Step 2. Sample $\boldsymbol{h}_t$ from $\boldsymbol{h}_t|\boldsymbol{\theta}, \boldsymbol{h}_{-h_t}, D_n$, for $t = 1, \ldots, n$.
Step 3. Sample $\boldsymbol{\beta}_j$ from $\boldsymbol{\beta}_j|\boldsymbol{\theta}_{-\beta_j}, \boldsymbol{h}, D_n$, for $j = 1, \ldots, p$.
Step 4. Sample r_{ii} from $r_{ii}|\boldsymbol{\theta}_{-r_{ii}}, \boldsymbol{h}, D_n$, for $j = 1, \ldots, p$
Step 5. Sample c_{ij} from $c_{ij}|\boldsymbol{\theta}_{-c_{ij}}, \boldsymbol{h}, D_n$, for $i, j = 1, \ldots, p$ $(i < j)$
Step 6. Sample s_j^2 from $s_j^2|\boldsymbol{\theta}_{-s_j^2}, \boldsymbol{h}, D_n$, for $j = 1, \ldots, p$
Step 7. Sample ν_j from $\nu_j|\boldsymbol{\theta}_{-\nu_j}, \boldsymbol{h}, D_n$, for $j = 1, \ldots, p$,
Step 8. Repeat Step 2 $\sim$ Step 7 for sufficient iterations.

Here $\boldsymbol{h}_{-h_t}$ denotes the rest of the $\boldsymbol{h}$ vector other than $\boldsymbol{h}_t$. By making a proposal draw from a random walk sampler, the Metropolis–Hastings (MH) algorithm implements steps 2–7. For instance, assume the first-order random walk model for the baseline sales. In step 2, the conditional posterior density function of $\boldsymbol{h}_t$ is

$$\pi(\boldsymbol{h}_t|\boldsymbol{\theta}, \boldsymbol{h}_{-h_t}, D_n, X_n)$$
$$\propto \begin{cases} f(\boldsymbol{h}_{t+1}|\boldsymbol{h}_t, \Sigma) \times \prod_{j=1}^p f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j), \ (t = 1), \\ f(\boldsymbol{h}_{t+1}|\boldsymbol{h}_t, \Sigma) \times f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \Sigma) \times \prod_{j=1}^p f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j), \ (t \neq 1, n), \\ f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \Sigma) \times \prod_{j=1}^p f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j), \ (t = n). \end{cases}$$

At the kth iteration, we make a candidate draw of $\boldsymbol{h}_t^{(k+1)}$ using the Gaussian proposal density function centered at the current value $\boldsymbol{h}_t^{(k)}$ with the variance matrix $0.01 \times I_p$. We then accept a candidate draw with the probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{h}_t^{(k+1)}|\boldsymbol{\theta}, \boldsymbol{h}_{-h_t}, D_n, X_n)}{\pi(\boldsymbol{h}_t^{(k)}|\boldsymbol{\theta}, \boldsymbol{h}_{-h_t}, D_n, X_n)}\right\}.$$

The remaining conditional posterior density functions are

$$\pi(\boldsymbol{\beta}_j|\boldsymbol{\theta}_{-\beta_j}, \boldsymbol{h}, D_n, X_n) \propto \prod_{t=1}^{n}\prod_{j=1}^{p} f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j) \times \pi(\boldsymbol{\beta}_j),$$

$$\pi(r_{ii}|\boldsymbol{\theta}_{-r_{ii}}, \boldsymbol{h}, D_n, X_n) \propto \prod_{t=2}^{n} f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \Sigma) \times \pi(r_{ii}),$$

$$\pi(c_{ij}|\boldsymbol{\theta}_{-c_{ij}}, \boldsymbol{h}, D_n, X_n) \propto \prod_{t=2}^{n} f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \Sigma) \times \pi(c_{ij}),$$

$$\pi(s_j^2|\boldsymbol{\theta}_{-s_j^2}, \boldsymbol{h}, D_n, X_n) \propto \prod_{t=1}^{n}\prod_{j=1}^{p} f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j) \times \pi(s_j^2),$$

$$\pi(\nu_j|\boldsymbol{\theta}_{-\nu_j}, \boldsymbol{h}, D_n, X_n) \propto \prod_{t=1}^{n}\prod_{j=1}^{p} f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \boldsymbol{\gamma}_j) \times \pi(\nu_j).$$

In addition to implementing step 2, the MH algorithm implements the steps . The outcomes from the MH algorithm can be regarded as a sample from the posterior density function after a burn-in period.

The remaining problem is the question of how to evaluate whether the estimated model is good. For example, we have to select the sampling density function among a set of models in (2). In the following section, we assess whether predictions made by the estimated model are close to those made by the true structure.

2.4 Model diagnosis: Bayesian predictive information criterion

In the previous section, we discussed the development of Bayesian models. One of the most crucial issues is the choice of an optimal model that adequately expresses the dynamics of the sales. In this section, we use the Bayesian predictive information criterion (Ando 2007) for evaluating the success of the predictive distribution constructed by the Bayesian methods.

Recently, Ando (2007) proposed the maximization of the posterior mean of the expected log-likelihood

$$\eta = \int \left[\int \log L(Z_n|X_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|D_n)\mathrm{d}\boldsymbol{\theta}\right] g(Z_n)\mathrm{d}Z_n,$$

where $\pi(\boldsymbol{\theta}|D_n)$ is the posterior density function and $Z_n = \{z_1, \ldots, z_n\}$ is the unseen observation generated from a true model. The best model is selected by maximizing this quantity.

Considering a situation in which the prior is assumed to be dominated by the likelihood as increases, and in which the specified parametric models contain the true model, Ando (2007) showed that an estimator of η is given by $\hat{\eta} - \dim\{\boldsymbol{\theta}\}/n$, where

Table 1 Basic statistics

	Store 1	Store 2
μ	61.9908	38.6059
σ	27.5079	19.4827
s	1.5599	1.3337
k	4.5616	3.8594

μ, the mean; σ, standard deviation; s, skewness; k, kurtosis

$\hat{\eta}$ is the posterior mean of the log-likelihood:

$$\hat{\eta} = \frac{1}{n}\int \log L(D_n|X_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|D_n)\mathrm{d}\boldsymbol{\theta}.$$

Multiplying -2, we then obtain a tailor-made version of the Bayesian predictive information criterion, BPIC (Ando 2007):

$$\mathrm{BPIC} = -2\int \log L(D_n|X_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|D_n)\mathrm{d}\boldsymbol{\theta} + 2\mathrm{dim}\{\boldsymbol{\theta}\}. \tag{4}$$

We can see that the BPIC balances the tradeoff between goodness-of-fit and parsimony. The best predictive distribution is selected by minimizing the Bayesian predictive information criterion (BPIC). As an alternative model selection criterion derived in the above framework, Spiegelhalter et al. (2002) proposed the DIC. From a theoretical viewpoint, it has been argued that the model chosen by the DIC is more complex than that chosen by the BPIC (Ando 2007). Therefore, this paper uses the BPIC.

The BPIC is available for the evaluation of various types of Bayesian models. For instance, Ando (2006b) employed this criterion when evaluating the success of several stochastic volatility models. For research on credit ratings, this criterion was applied to a Bayesian ordered probit regression model with a functional predictor (Ando 2006c).

3 Empirical illustration of proposed method

3.1 Data description

In 2006, the size of the market for incense products in Japan was estimated to be about 30 billion yen. Although the market has been shrinking gradually (the 2006 size is only 88% of the size in 1980), the business of producing and selling incense products still provides an opportunity to earn a profit. The data analyzed here consist of the daily sales figures for incense products from January 2006 to March 2007. The data were collected from two department stores (hereafter, Store 1 and Store 2), both located in Tokyo. In both stores, incense manufacturers sell two main products: traditional incense and lifestyle incense. In Japan, traditional incense is used differently from lifestyle incense. Traditional incense is used for religious purposes, for example at Buddhist altars or at the graves of ancestors. In contrast, lifestyle incense is used therapeutically for enjoyment. The positioning of these products is distinct.

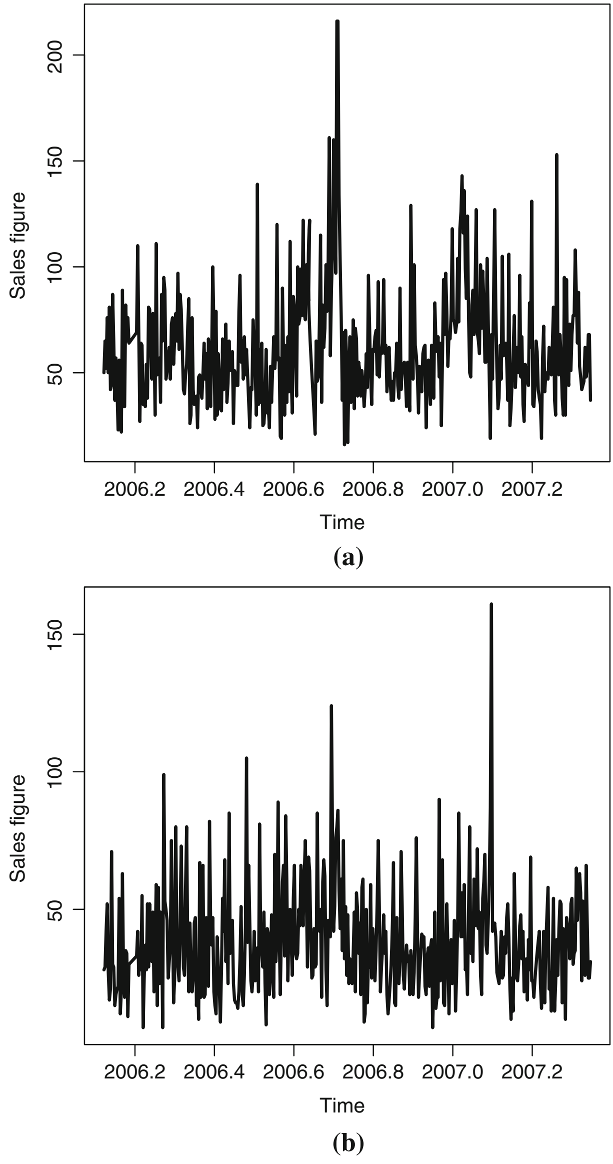

Fig. 1 Time series plots of the daily sales figures for incense products from January 2006 to March 2007. **a** Store 1 and **b** Store 2

Figure 1a, b shows the time series plots of the daily sales at Store 1 and Store 2, respectively. In this analysis, the units are *thousands of yen*. From Fig. 1, it may be seen that the daily sales vary over time. The basic statistics are shown in Table 2. Since the kurtosis of the returns is greater than three, the true distribution the data must be a fat-tailed distribution. Using the Shapiro–Wilk normality test (Patrick 1982) the

Table 2 Summary of the estimation results

	Mean	SDs	95% Conf. interval	INEFs	CD
β_{11}	−1.883	0.944	[-3.740, 0.053]	2.385	−0.584
β_{21}	10.028	0.845	[8.398, 11.810]	2.692	−0.839
β_{12}	2.223	0.893	[0.596, 3.742]	2.452	−0.335
β_{22}	3.127	0.763	[1.739, 4.624]	2.547	−1.332
β_{13}	−0.596	0.831	[-2.243, 1.126]	2.193	0.550
β_{23}	10.099	0.742	[8.573, 11.604]	2.849	1.032
β_{14}	24.396	0.966	[22.592, 26.105]	2.309	−0.697
β_{24}	11.670	0.864	[9.841, 13.421]	2.325	−1.725
s_1^2	25.472	0.080	[25.216, 25.762]	2.604	−1.814
s_2^2	17.061	0.049	[16.964, 17.155]	2.270	−1.745
σ_{11}	25.472	0.063	[25.243, 25.653]	5.857	−0.995
σ_{22}	17.006	0.046	[16.960, 17.155]	5.935	0.056
σ_{12}	0.185	0.010	[0.169, 0.201]	2.783	0.967
ν_1	26.106	0.602	[24.998, 27.042]	25.092	0.653
ν_2	5.001	0.483	[4.049, 6.012]	24.330	0.976

The posterior means, the standard deviations (SDs), the 95% confidence intervals, the inefficiency factors (INEFs) and Geweke's (1992) CD test statistic (CD) are calculated

null hypothesis—that sales were normally distributed—was rejected. The p values for each score were 2.58×10^{-16} and 2.62×10^{-14}, respectively.

In addition to the daily sales data, the following information was tabulated; the weather effect x_{j1t}, the weekly and holiday effect x_{j2t}, the sales promotion effect x_{j3t} and the event effect x_{j4t}. Definitions of each variable are given as follows:

$$x_{j1t} = \begin{cases} 1 \ \text{(Fine)} \\ 0 \ \text{(Cloudy)} \\ \text{-1} \ \text{(Rain)} \end{cases}, \quad j = 1, \ldots, p,$$

$$x_{j2t} = \begin{cases} 1 \ \text{(Sunday, Saturday, National holiday)} \\ 0 \ \text{(Otherwise)} \end{cases}, \quad j = 1, \ldots, p,$$

$$x_{j3t} = \begin{cases} 1 \ \text{(Execution)} \\ 0 \ \text{(Nonexecution)} \end{cases}, \quad j = 1, \ldots, p,$$

$$x_{j4t} = \begin{cases} 1 \ \text{(Holding)} \\ 0 \ \text{(Nonholding)} \end{cases}, \quad j = 1, \ldots, p.$$

As pointed out in Sect. 2.2, information on other variables, price levels, price discount percentages, features, displays, post-promotion dips are important factors. Unfortunately, due to the limitations of the dataset, we considered only these variables. We would like to emphasize that the analysis can be done easily once we could obtain such additional information.

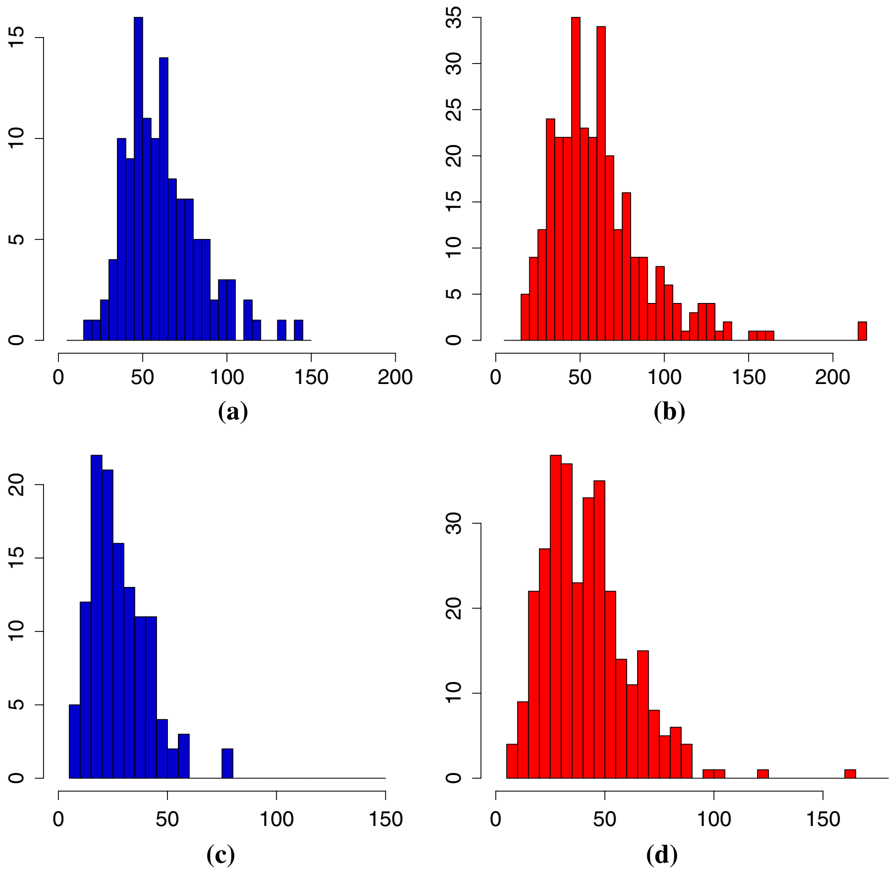

Fig. 2 Histograms show the number of sales for **a** Store 1 when the sales promotions are executed, **b** Store 1 when the sales promotions are not executed, **c** Store 2 when the sales promotions are executed, and **d** Store 2 when the sales promotions are not executed

In both stores, the responsibility for promotion was shared between the manufacturer and the department store. On five of every 7 days, the manufacturers were responsible for promoting their own products. On the remaining 2 days, the department stores were responsible for selling the manufacturers' products. Department stores were obligated to sell the manufactures' products such that the average sales on days of department store promotion were consistent with the average sales on days of manufacturer promotion.

Figure 2 examines the effects of sales promotions. The horizontal axis measures sales. We can see that the shapes of histograms (a) and (b) for Store 1 are similar, while the shapes of histograms (c) and (d) for Store 2 are more disparate. The χ^2 test at a 5% significance level does not reject the null hypothesis (that there is no difference in the distribution of sales regardless of the responsible party) for Store 1. On the other hand, the null hypothesis is rejected for Store 2.

3.2 Estimation results

In this section, we fit the various statistical models given in (2). The largest model evaluation space might be the selections of distributional assumption on $\boldsymbol{y}_t$, the lag of the baseline sales r, and the combination of the covariates $\boldsymbol{x}_{jt}$ in the model. Because one of our aims is to quantify the impacts of each covariate, we consider the selections of distributional assumption on $\boldsymbol{y}_t$, and the lag of the baseline sales $r = \{1, 2, 3\}$.

The total number of MCMC iterations is chosen to be 6,000; of those 6,000 iterations, the first 1,000 iterations are discarded as a burn-in period. To ensure the convergence of the MCMC sampling algorithm, we stored every fifth iteration after the burn-in period. All inferences were therefore derived using the 1,000 generated samples.

It is necessary to check whether the generated posterior sample is taken from the stationary distribution. We assessed the convergence by calculating the convergence diagnostic (CD) test statistics (Geweke 1992). Geweke's (1992) CD test statistic evaluates the equality of the means in the first and last part of the Markov chains. If the samples are drawn from the stationary distribution, the two means calculated from the first and the last part of a Markov chain are equal. It is known that the CD test statistic has an asymptotic standard normal distribution. All of the results that we report in this paper are based on samples that have passed Geweke's (1992) convergence test at a significance level of 5% for all parameters.

Searching the best model, we found that the most adequate model to describe the data is the Student-t model with the lag of the baseline sales $r = 2$, which achieved the minimum value of BPIC, BPIC $= 9,948.833$. We therefore select this model, which is preferred by the BPIC. Table 2 reports the posterior means, the standard errors, the 95% confidence intervals, the inefficiency factor (Kim et al. 1998) and the values of Geweke's CD test statistic. Based on 1,000 draws for each of the parameters, we calculated the posterior means, the standard errors, and the 95% confidence intervals. The 95% confidence intervals are estimated using the 2.5th and 97.5th percentiles of the posterior samples. The inefficiency factor is a useful measure for evaluating the efficiency of the MCMC sampling algorithm. It is defined as $1 + 2\sum_{k=1}^{\infty} \rho(k)$, where $\rho(k)$ is the sample autocorrelation at lag k calculated from the sampled draws. We have used 1,000 lags to estimate the inefficiency factors. As shown in Table 2, the employed sampling procedure achieved a good efficiency.

Figure 3 plots the change in the posterior means of the baseline sales for each store. As shown in Fig. 3, it shows a nonlinear relationship over the sales period. We can also see that the baseline sales for each store are different from each other.

3.3 Discussion

As shown in Fig. 3, the baseline sales for each store stores are time varying. In Japan, it is widely expected that the sales patterns for traditional incense and lifestyle incense will differ over the course of the year. For traditional incense, sales peak during times of religious significance; specifically, they peak during the equinoctial weeks of spring and autumn and during the Bon Festival in August. Sales tend to be highest in March,

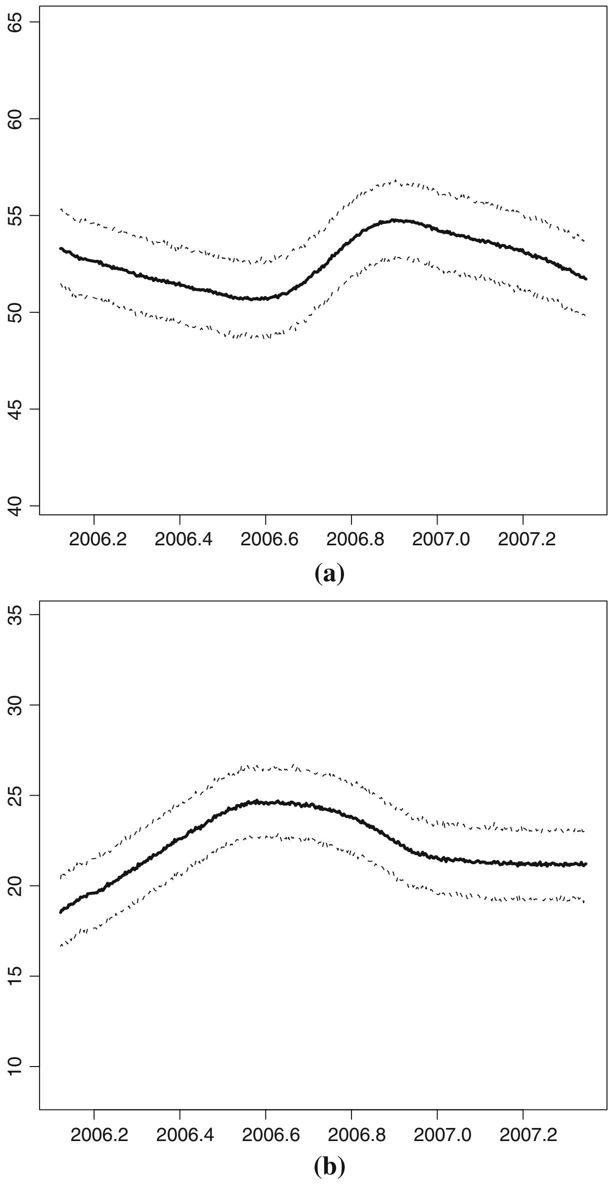

Fig. 3 The fluctuations in the posterior means of baseline sales for each item. The *dashed lines* are the 95% confidence intervals. **a** Store 1 and **b** Store 2

in the months of July through September, and briefly in early December. By contrast, sales for lifestyle incense peak during the rainy season in late May and June.

Figure 3 also indicates that the baseline sales for each store are different from each other. We further investigated the consumer demographics for each of the two stores. Demographic analysis showed that elderly consumers represented the vast majority of sales at Store 1, whereas the majority of consumers at Store 2 were younger, females, or foreigners. The sales data are consistent with the observation that elderly people tend to purchase traditional incense and that younger people tend to purchase lifestyle incense. In Store 1, the sale of traditional incense predominated; in Store 2, lifestyle incense was much more popular.

As shown in Table 2, weather appears to impact demand for lifestyle incense. In Store 2, sales rose during the rainy season. The posterior mean of β_{21} is greater than 0. The estimated coefficients on the weekly effect, β_{12} and β_{22}, indicate that working days have a negative effect on sales. This is to be expected, since working people rarely visit department stores during the workday. There is a significant difference in the coefficients that measure the promotion effects in Store 1 and Store 2. The posterior mean of β_{13} is close to zero, while that of β_{23}'s is far from zero. Moreover, the 95% confidence interval around β_{13} includes 0. This suggests that sales will not increase even if promotion is used for Store 1. On the other hand, the daily sales would increase when promotion is used for Store 2.

There are at least two reasons for this. First, Store 2 is located near many foreign offices. Foreign customers often buy the fancy cassolette and the accompanying incense products as souvenirs. Selling these luxury goods requires substantial acquaintance with and knowledge of the product. The result is therefore expected, since manufacturer employees tend to be more knowledgeable about the product than department store employees. Second, auspicious product display in the department store has an impact on sales. In Store 1, incense products are displayed near the checkout lines. Because most department store employees are stationed near checkout areas, they can easily support customers looking for incense. Unlike in Store 1, the location of incense in Store 2 is far from the checkout area.

Often, department stores will have store-wide promotions. Sales of incense increase in conjunction with store-wide events. Correlation coefficients indicate that the sales of each store are correlated with each other. Since the posterior means of ν_j are around 5 and 26, the sales data have a fat-tailed distribution. Additionally, this conclusion was supported by the BPIC. As described before, the Student t model is superior to the normal model. As a benchmark model, we also fitted Ando's (2006a) model, where Poisson distribution is employed for predicting the sales. BPIC score indicated that the proposed model is also superior to the Ando (2006a) model. The BPIC score of Poisson model with the lag of the baseline sales $r = 2$ was BPIC $= 9,998.964$. The aforementioned results suggest that the proposed method can be used to distill useful information from observed data.

4 Conclusions

This paper considered the problem of simultaneously identifying unobserved baseline sales, the effectiveness of marketing activities, and the effects of controllable/uncontrollable business factors. We developed a method for modeling multivariate time series within the framework of Bayesian general state-space modeling. Since the likelihood function depends on high-dimensional integrals, the Bayesian approach via the MCMC algorithm is proposed. This approach can more easily estimate the model parameters due to recent advancements in computer technology. To determine the most amenable model among a set of candidate models, the use of the Bayesian predictive information criterion is proposed. As shown in the data analysis, interesting and important practical results are obtained. We apply the proposed method to the daily sales of incense products, using information collected from two department stores (both located in Tokyo). In both stores, incense manufacturers sell two main products: traditional incense and lifestyle incense. Generally, elderly people tend to purchase the former and younger people tend to purchase the latter. Traditional incense products are used for religious purposes, while lifestyle incense products are used for pleasure. In this study, the proposed method achieved many results.

First, the daily sales data have a fat-tailed distribution. When we compared the BPIC scores of the normal and the fat-tailed models, the latter was supported. This result was also consistent with the result from the Shapiro-Wilk normality test. Because normality is an essential assumption of the traditional state-space models, the proposed method is a powerful tool for data analysis.

Second, our results suggest that the majority of consumers in the two stores differ. Demographic analysis showed that the majority of consumers at one store were elderly, whereas the majority of consumers in the other store were younger, female, or foreign. Because the positioning of these products is distinct, such information is useful when planning a marketing strategy.

Third, there is a significant difference in the promotion effect between these stores. The results suggest that daily sales will not increase in one store, even if promotions are used. In contrast, promotions are effective at the other store. We believe that the proposed method might allow us to forecast with less uncertainty and greater accuracy as more information becomes available. Unfortunately, a large dataset was not available for this paper.

There is still room for further investigation. First, as demonstrated an empirical study, traditional and lifestyle incense are very different products, and may aimed at different consumer groups for distinctly different purposes. However, due to the limited data, the model used the aggregated series of incense sales instead of modeling two separate series. This aggregation may increase noise in the data. If we have more detailed data, we can incorporate such a strong difference. Second, our sales promotion data are 0/1 dummies, just indicating the execution of a display promotion. More detailed price promotion data will help us to investigate the depth of the price cut matter. If dataset is the combination of price promotion, feature and display; there is an opportunity to extend the proposed model to analyze their separate and synergistic effect on sales. We would like to investigate these problems in a future paper.

Acknowledgments The author would like to acknowledge two anonymous referees for constructive comments that have substantially improved the article. This study was supported in part by a Grant-in-Aid for Young Scientists (B) (No.18700273).

References

Abraham, M. M., Lodish, L. M. (1993). An implemented system for improving promotion productivity using store scanner data. *Marketing Science, 12*, 248–269.

Ando, T. (2006a). Bayesian State space modeling approach for measuring the effectiveness of marketing activities and baseline sales from POS data. In *Proceeding of IEEE International Conference on Data Mining* (pp. 21–32).
Ando, T. (2006b). Bayesian inference for nonlinear and non-Gaussian stochastic volatility model with leverage effect. *Journal of the Japan Statistical Society, 36*, 173–197.
Ando, T. (2006c). Bayesian credit rating analysis based on ordered probit regression model with functional predictor. In *Proceeding of The Third IASTED International Conference on Financial Engineering and Applications* (pp. 69–76).
Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika, 94*, 443–458.
Barnard, J., McCulloch, R., Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica, 10*, 1281–1311.
Bass, F.M. (1969). A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research, 6*, 291–300.
Beckwith, N.E. (1972). Multivariate analysis of sales responses of competapplication brands to advertising. *Journal of Marketing Research, 9*, 168–176.
Blattberg, R.C., Eppen, G., Liebermann, J. (1981). A theoretical and empirical evaluation of price deals in consumer non-durables. *Journal of Marketing, 45*, 116–129.
Carlin, B., Louis, T. (1996). *Bayes and empirical Bayes methods for data analysis*. New York: Chapman and Hall.
Chib, S., Nardarib, F., Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics, 108*, 281–316.
Dekimpe, M.G., Hanssens, M. (2000). Time-series models in marketing: Past, present and future. *International Journal of Research in Marketing, 17*, 183–193.
Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernado, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian statistics*, *Vol. 4* (pp. 169–193). Oxford: Oxford University Press.
Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (Eds.) (1996). *Markov Chain Monte Carlo in practice*. New York: Chapman and Hall.
Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research, 25*, 342–355.
Hanssens, D. M. (1980). Market response, competitive behavior, and time series analysis. *Journal of Marketing Research, 17*, 470–485.
Kim, S., Shephard, N., Chib, S. (1998). Stochastic volatility: likelihood inference comparison with ARCH models. *Review of Economic Studies, 65*, 361–393.
Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics, 5*, 1–25.
Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association, 82*, 1032–1063.
Kitagawa, G., Gersch, W. (1996). *Smoothness priors analysis of time series*. New York: Springer.
Kitagawa, G., Higuchi, T., Kondo, F. N. (2003). Smoothness prior approach to explore mean structure in large-scale time series. *Theoretical Computer Science, 292*, 431–446.
Kondo, F. N., Kitagawa, G. (2000). Time series analysis of daily scanner sales—Extraction of trend, day-of-week effect, and price promotion effect. *Marketing Intelligence & Planning, 18*, 53–66.
Lee, J., Boatwright, P., Kamakura, W. A. (2003). A Bayesian model for prelaunch sales: forecasting of recorded music. *Management Science, 49*, 179–196.
Leone, R. P. (1983). Modeling sales-advertising relationships: an integrated time series—Econometric approach. *Journal of Marketing Research, 20*, 291–295.
Naik, P. A., Murali, K. M., Alan, S. (1998). Planning pulsing media schedules in the presence of dynamic advertising quality. *Marketing Science, 17*, 214–235.
Naik, P. A. (1999). Estimating the half-life of advertisements. *Marketing Letters, 10*, 351–362.
Neelamegham, R., Pradeep, C. (1999). Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science, 18*, 115–136.
Neslin, S., Henderson, C., Quelch, J. (1985). Coupon promotions and acceleration of product purchase. *Marketing Science, 4*, 147–165.
Neslin, S. (2002). *Sales promotion*. Cambridge: Marketing Science Institute.
Patrick, R. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics, 31*, 115–124.

Pauwels, K., Hanssens, D. M., Siddarth, S. (2004). The long-term effects of price promotions on category incidence, brand choice and purchase quantity. *Journal of Marketing Research, 29*, 421–439.

Robert, C. P., Titterington, D. M. (2002). Discussion on "Bayesian measures of model complexity and fit" (by Spiegelhalter, D. J. *et al*). *Journal of the Royal Statistical Society, Series B, 64*, 621–622.

Sato, T., Higuchi, T., Kitagawa, G. (2004). Statistical inference using stochastic switching models for the discrimination of unobserved display promotion from POS data. *Marketing Letters, 15*, 37–60.

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

Tanizaki, H., Mariano, R.S. (1998). Nonlinear and non-Gaussian state-space modeling with Monte Carlo simulations. *Journal of Econometrics, 83*, 263–290.

Tellis, G. J., Fred, F. Z., Zufryden, F. (1995). Tackling the retailer decision maze: which brands to discount, how much, when, and why? *Marketing Science, 12*, 271–299.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics, 22*, 1701–1762.

Van Heerde, H., Leeflang, P., Wittink, D. (2004a). Decomposing the sales promotion bump with store data. *Marketing Science, 23*, 317–334.

Van Heerde, H., Mela, C., Manchandra, P. (2004b). The dynamic effect of innovation on market structure. *Journal of Marketing Research, 41*, 166–183.

Wildt, A.R. (1974). Multifirm analysis of competitive decision variables. *Journal of Marketing Research, 11*, 50–62.

Xie, J., Song, M.S., Wang, Q. (1997). Kalman filter estimation of new product diffusion models. *Journal of Marketing Research, 34*, 378–393.

Yamaguchi, R., Tsuchiya, E., Higuchi, T. (2004). State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors (in Japanese). *Transactions of the Operations Research Society of Japan, 49*, 52–60.