Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 138 (2007) 231–251
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Seasonality and non-linear price effects in scanner-data-based market-response models

Dennis Fok*, Philip Hans Franses, Richard Paap

Econometric Institute, Erasmus University Rotterdam, H11-23, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands

Available online 5 July 2006

Abstract

Scanner data for fast moving consumer goods typically amount to panels of time series where both N and T are large. To reduce the number of parameters and to shrink parameters towards plausible and interpretable values, Hierarchical Bayes models turn out to be useful. Such models contain in the second level a stochastic model to describe the parameters in the first level.

In this paper we propose such a model for weekly scanner data where we explicitly address (i) weekly seasonality when not many years of data are available and (ii) non-linear price effects due to historic reference prices. We discuss representation and inference and we propose a Markov Chain Monte Carlo sampler to obtain posterior results. An illustration to a market-response model for 96 brands for about 8 years of weekly data shows the merits of our approach.
© 2006 Elsevier B.V. All rights reserved.

JEL classification: C11; C22; C23; M31

Keywords: Panels of time series; Weekly seasonality; Threshold models; Non-linearity; Hierarchical Bayes; MCMC

1. Introduction

This paper deals with the econometric aspects of market-response models, when these are calibrated for weekly scanner data, typically for fast moving consumer goods (FMCGs). Market-response models usually seek to correlate sales or market shares with marketing-mix instruments such as price, promotions like feature and display, and

*Corresponding author. Tel.: +31 10 4081333; fax: +31 10 4089162.
E-mail address: dfok@few.eur.nl (D. Fok).

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.05.021

advertising, see Hanssens et al. (2001) and Leeflang et al. (2000). Due to in-store scanner techniques, the data that are typically available to estimate model parameters are weekly data for 4–8 years. The amount of years is due to the fact that the life cycle of products and, sometimes also of brands, does not often extend beyond that time frame. The weekly data are usually provided by a particular retail chain, and they concern the most important brands in several categories across the outlets within that chain. It is common to stack the information on brands across categories, and to consider models for, say, N brands, where these brands thus cover a variety of FMCG categories like margarine, tissues, ketchup and so on, see Pauwels and Srinivasan (2004), Nijs et al. (2001) and Fok et al. (2006). In sum, the relevant market-response models are calibrated for panels of time series, where N ranges from, say, 50 to 300, and where time T covers 52 weeks for 4–8 years. Note that this type of data is not always suited to analyze the market response across all supermarkets in a particular area as some retailers, usually hard discounters like Wall-Mart, do not supply data. There are some recent developments to overcome this problem, such as the use of consumer hand-scan panels, see, for example, Fox et al. (2004) and van Heerde et al. (2005). In this paper we however consider a single retailer.

Given the availability of large N and large T data, one could simply want to consider N different models, or different models per product category. However, in marketing research it is common practice to search for, so-called, empirical generalizations, that is here, common features across the N models. In market-response models such common features could concern the effects of price changes or of promotions. These effects could partly be idiosyncratic, and partly be the same for similar brands in similar categories, for example. This usually means that a useful market-response model has a second layer in which the parameters in the N models are correlated with characteristics of brands and categories which are constant over time, see Fok et al. (2006) among others. In the present paper, we also propose such a two-level Hierarchical Bayes model and we use a Markov Chain Monte Carlo (MCMC) approach to obtain posterior results.

The first advantage of Hierarchical Bayes models for panels of time series is that it often amounts to a plausible reduction of the number of parameters. Hence, there is an increase in the degrees of freedom. This is particularly useful when the first-level parameters are less easy to estimate due to a lack of degrees of freedom. For example, as we will discuss below, the inclusion of 52 weekly dummies to capture seasonality in weekly market-response models amounts to a serious loss of degrees of freedom, particularly when there would be only 4 years of data. Hence, a plausible strategy here is to introduce a second-level model where seasonality is captured by a sinusoid regressor and an error term.

The model that we will propose below further allows for the possibility that past prices have an effect on the current short-run price elasticity. Such an effect is often documented in the marketing literature, see Pauwels et al. (2006) and the references cited therein, and it means that the difference between the current price and the previously observed price has an impact on the current price effect. Such a non-linear effect can occur in many N equations, but perhaps not in all. Hence, a second advantage of a two-level model is that by including all N equations, the parameters in each of these get shrunk towards a common value in the second level, see Blattberg and George (1991), while of course the error term allows for brand-specific variation.

In sum, in this paper we put forward a two-level Hierarchial Bayes model for a panel of weekly time series on sales and marketing-mix instruments, where we use the second level to effectively reduce the number of parameters to capture seasonality and to shrink (potentially difficult to estimate) non-linear effects towards interpretable parameters. In the second level, the latter parameters are correlated with brand-specific and category-specific characteristics. In Section 2, we describe the representation of the model. In Section 3 we propose an MCMC sampler to obtain posterior results. In Section 4, we apply our model to data on 96 brands for close to 8 years of weekly data. We demonstrate that the model yields plausible and reliable estimates. In Section 5 we conclude with some remarks.

2. Representation and interpretation

When modeling weekly sales of FMCGs, a typical model that relates log sales to log prices and promotion variables, amongst other marketing instruments, is

$$\ln S_{it} = \mu_i + \beta_i \ln P_{it} + \mathrm{Promo}_{it}'\psi_i + \varepsilon_{it}, \tag{1}$$

where S_{it} denotes the sales of brand i for $i = 1, 2, \ldots, N$ at time t, for $t = 1, 2, \ldots, T$, and P_{it} denotes the price of brand i at time t, and where $\varepsilon_{it} \sim \mathrm{N}(0, \sigma_i^2)$, see Wittink et al. (1988) and many others. The vector Promo_{it} captures promotion activities for brand i at time t. In recent years it has been recognized that dynamics cannot be ignored when modeling sales, and hence the specification in (1) can be replaced by, for example,

$$\Delta \ln S_{it} = \mu_i + \rho_i \ln S_{i,t-1} + \beta_i \Delta \ln P_{it} + \delta_i \ln P_{i,t-1} + \mathrm{Promo}_{it}'\psi_i + \varepsilon_{it}, \tag{2}$$

where Δ is the first-differencing operator.

Two assumptions in this model may not hold for actual scanner data. They are (i) that expected (unconditional) sales are constant over time, here reflected by μ_i, and (ii) that the short-run price elasticity, here β_i, is constant. In this paper we propose a model for which these assumptions are relaxed. First, we allow sales to show a weekly seasonal pattern. Upon doing so, we need to prevent having to include 52 seasonal dummies to retain degrees of freedom. Second, we allow the price elasticity to depend on the direction and the magnitude of price changes, that is, the current price effect can be different for cases where the past price was lower than when it was higher. The literature surveyed in Pauwels et al. (2006) shows that consumers tend to show an asymmetry in the evaluation of gains and losses. Following this literature the price elasticity must be different for a price increase than for a price decrease. Additionally, it may be that consumers do not notice small price changes, that is, small price changes may not lead to sales changes and sales are only affected if the price change exceeds a certain threshold. In sum, we modify (2) to include weekly seasonality and non-linear price effects. Below, we present these two model extensions in detail.

2.1. Weekly seasonality

A standard approach to capturing seasonality in sales data (that usually do not show seasonal unit roots) is to include seasonal dummies. Denote the number of observations per year by S. Model (2) would then not include μ_i but $\sum_{s=1}^{S} D_{st}\mu_{is}$, where the seasonal dummy variable $D_{st} = 1$ if observation t corresponds to season s, and where $D_{st} = 0$ otherwise. Of course, in case of weekly data, the estimation of the parameters associated with $S = 52$ dummy variables can be cumbersome, in particular when there are not many years of data available. Note however that one may expect that the seasonality can be decomposed into a regular cyclical pattern and an irregular component. The regular component corresponds to seasonal effects that are caused by summer and winter. The irregular component corresponds to other recurring patterns in the data, for example caused by special events. In this paper we propose to take advantage of this decomposition. To keep things simple, we specify the season by a deterministic cycle with a period of 1 year and a stochastic factor which gives the deviation from this perfect cycle. In sum, we propose to model the seasonal component for brand i (μ_{is}) by

$$\mu_{is} = \alpha_{i0} + \alpha_{i1} \cos\left(2\pi \frac{s}{S} - \alpha_{i2}\right) + \eta_{is}, \tag{3}$$

where $\eta_{is} \sim \mathrm{N}(0, \sigma^2_{\eta_i})$. The parameter α_{i0} determines the conditional mean of the series, α_{i1} gives the amplitude of the deterministic part of the cycle. The parameter α_{i2} $(0 \leqslant \alpha_{i2} \leqslant 2\pi)$ determines the phase of the cycle, see Jones and Brelsford (1967) for a similar approach in periodic models. For notational convenience we define $\alpha_i = (\alpha_{i0}, \alpha_{i1}, \alpha_{12})$.

The two parts of the seasonal component have different interpretations. The deterministic part of the seasonal pattern corresponds to a regular cycle, for example a factor determined by summer and winter. The stochastic part deals with recurring spikes or dips in the sales, which may be due to special festivals as Christmas or Easter. Note that instead of including S parameters we now face estimating only four parameters.

2.2. Non-linear price effects

To allow for non-linear price effects due to past prices, we propose to replace $\beta_i \Delta \ln P_{it}$ in (2) by $G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i)$, where G is a certain non-linear function to be discussed below. For the immediate effect of price we want to allow for price thresholds and price gaps, that is, there might be no price effect for some price changes. Furthermore, we want to allow for asymmetric effects. For example, price increases relative to the previous price may have a more prominent effect than price decreases have. Finally, small price changes may have a different effect on sales than large price changes have.

To capture this range of possible non-linear effects, we introduce three regimes, that is, (i) large price decreases, (ii) small price changes, and (iii) large price increases. These regimes are bounded by two thresholds $\tau_{i1} > 0$ and $\tau_{i2} > 0$. If an increase in price is larger than τ_{i1}, that is, if $\Delta \ln P_{it} > \tau_{i1}$, we classify it as a large price increase. And, if the price decrease is larger than τ_{i2}, it is a large price decrease ($-\Delta \ln P_{it} > \tau_{i2}$). In the third case, the price change is classified as being small. Note that the regime of a small price change is not necessarily symmetric, that is, a price increase of 10% may still be classified as small while a price decrease of 5% can be considered as large. Of course, the actual boundaries of the regimes need to be estimated from the data, and they may differ across brands.

As is common in the literature on threshold models, see Granger and Teräsvirta (1993) and Franses and van Dijk (2000), we consider logistic functions to define the three regimes. The logistic function is

$$F(z; \gamma, \tau) = \frac{1}{1 + \exp(-\gamma(z - \tau))}. \tag{4}$$

Assuming that $\gamma, \tau > 0$, the switching function equals 1 for large positive values of the indicator z. Note that, depending on the value of γ, this function allows for a smooth transition from one regime to the other. Using (4), and taking aboard the arguments above, we can now specify the price–effect function as

$$G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i) = \beta_{i0}\Delta \ln P_{it} + (\beta_{i1} - \beta_{i0})F(\Delta \ln P_{it}; \gamma, \tau_{i1})(\Delta \ln P_{it} - \tau_{i1}) + (\beta_{i2} - \beta_{i0})F(-\Delta \ln P_{it}; \gamma, \tau_{i2})(\Delta \ln P_{it} + \tau_{i2}), \tag{5}$$

where $\tau_i = (\tau_{i1}, \tau_{i2})$ and $\beta_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})$. This expression can be interpreted as follows. For large price increases the derivative of the function G with respect to $\ln P_{it}$ equals β_{i1}, for large price decreases it equals β_{i2}, and for small price changes it equals β_{i0}. Fig. 1 graphically depicts the resulting sales–response curve.

Close to the thresholds, the derivative equals a weighted combination of the two derivatives in the adjacent regimes. A good approximation is obtained when the switching function itself is used as the weight. More formally,

$$\begin{aligned}\frac{\partial \ln S_{it}}{\partial \ln P_{it}} &= \frac{\partial G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i)}{\partial \ln P_{it}} \\ &= \beta_{i0} + (\beta_{i1} - \beta_{i0})F(\Delta \ln P_{it}; \gamma, \tau_{i1}) + (\beta_{i2} - \beta_{i0})F(-\Delta \ln P_{it}; \gamma, \tau_{i2}) \\ &\quad + (\beta_{i1} - \beta_{i0})\frac{\partial F(\Delta \ln P_{it}; \gamma, \tau_{i1})}{\partial \ln P_{it}}(\Delta \ln P_{it} - \tau_{i1}) \\ &\quad + (\beta_{i2} - \beta_{i0})\frac{\partial F(-\Delta \ln P_{it}; \gamma, \tau_{i2})}{\partial \ln P_{it}}(\Delta \ln P_{it} + \tau_{i2}) \\ &\approx \beta_{i0} + (\beta_{i1} - \beta_{i0})F(\Delta \ln P_{it}; \gamma, \tau_{i1}) + (\beta_{i2} - \beta_{i0})F(-\Delta \ln P_{it}; \gamma, \tau_{i2}),\end{aligned} \tag{6}$$

where the last line follows from the fact that

$$\frac{\partial F(z; \gamma, \tau)}{\partial z} = \gamma F(z; \gamma, \tau)(1 - F(z; \gamma, \tau)) \approx 0 \tag{7}$$

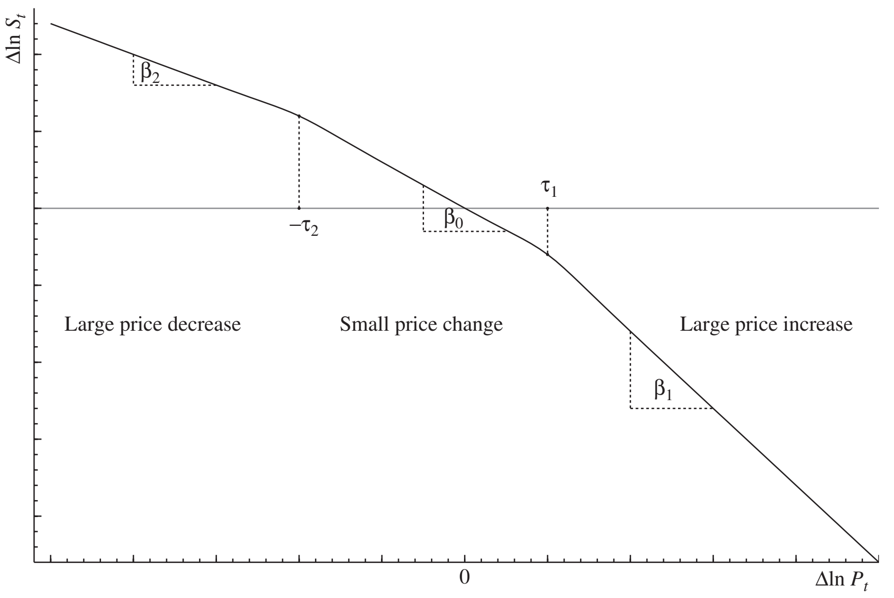

Fig. 1. Example of a non-linear sales response curve.

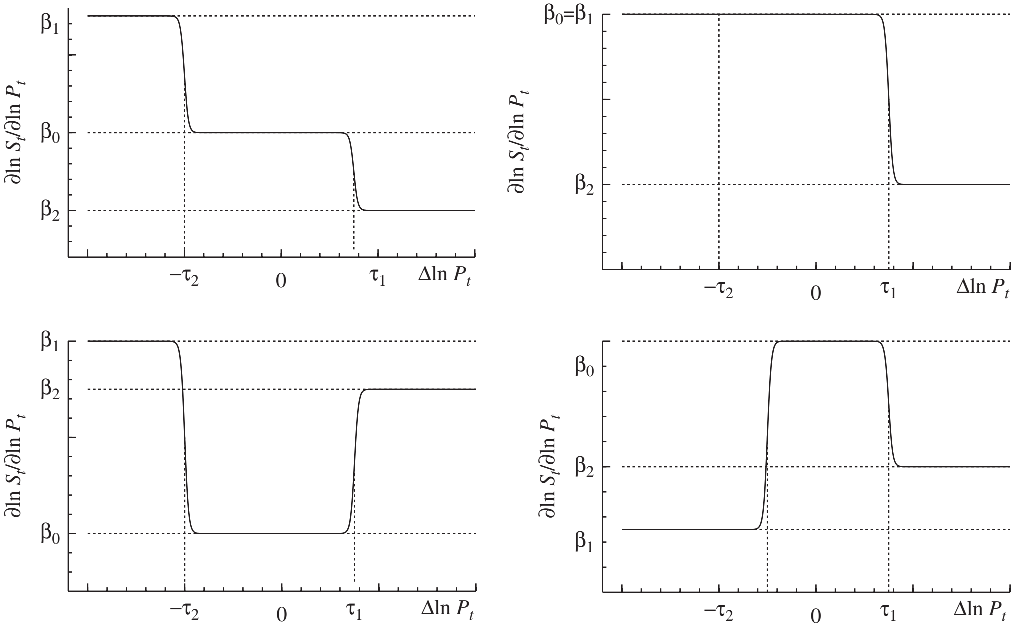

Fig. 2. Four examples of the implied price elasticities.

for γ large, as either $F(z;\gamma,\tau)\approx 0$ or 1. In our application we will indeed fix the value of γ at a relatively large value. As a consequence, given a value for τ, the transition from one regime to the other is immediate.

The expression in (6) and the line in Fig. 1 can be interpreted as the short-run price elasticity. The size of the price elasticity now depends on the size and the direction of the price change, relative to the previous price. The usual definition of a price elasticity only appears for very small price changes. Here we extend this definition by allowing for different regimes. An interpretation is the following. Suppose that a manager is planning a price change of $\Delta \ln P_{it} = -0.1$ (approximately a price cut of 10%), then the derivative $\partial \ln S_{it}/\partial \ln P_{it}$ evaluated at $\ln P_{it} = \ln P_{i,t-1} - 0.1$ gives the (additional) percentage change in sales in case the price would be decreased even further.

Our model with (5) is very flexible, as can be seen from Fig. 2 where we present four different possible patterns for the price elasticity. The graphs show that the model allows for a variety of price elasticity functions. The top-left graph corresponds to a case where the elasticity for a large price increase is larger in magnitude when compared with a similar-sized decrease. The bottom-right graph shows the opposite case. In this case the elasticity of a small price change is also relatively small. Also note that the thresholds defining the regimes can of course take different values.

2.3. A second-level model

In the literature there is much evidence that the price elasticity differs across product categories and even across brands within a product category, see, for example, Nijs et al. (2001) and Fok et al. (2006) among many others. However, in these studies it is assumed that for a brand the elasticity is independent of the price change itself. With our non-linear model, we can see if this assumption holds. Furthermore, we will try to explain possible differences in non-linearities using observable brand and category characteristics. To this end we propose a second-level model, in which we relate the parameters in β_i in (5) to observable characteristics (Z_i), that is,

$$\begin{aligned}\beta_{i0} &= Z_i'\theta_0 + \xi_{i0},\\ \beta_{i1} &= Z_i'\theta_1 + \xi_{i1},\\ \beta_{i2} &= Z_i'\theta_2 + \xi_{i2},\end{aligned} \tag{8}$$

where $\xi_i = (\xi_{i0}, \xi_{i1}, \xi_{i2})' \sim \mathrm{N}(0, \Sigma)$ and Z_i is a k-dimensional vector of explanatory variables and θ_j is a k-dimensional vector of parameters for $j = 0, 1, 2$, see, for example, Hendricks et al. (1979) for a similar approach. We define the matrix $\theta = (\theta_0, \theta_1, \theta_2)$. The covariance matrix Σ is *not* restricted to be diagonal.

We do not impose any restrictions on the sales–response curves. In case one wants to guarantee a monotonic decreasing response curve, one may impose that $\beta_{i0} < 0$, $\beta_{i1} < 0$ and $\beta_{i2} < 0$. We however abstain from this restriction. As will be shown below, in our application the posterior results indicate that in general the sales–response curves are monotonically decreasing.

3. Bayes analysis

The total model is given by

$$\Delta \ln S_{it} = \sum_{s=1}^{S} D_{st}\mu_{is} + \rho_i \ln S_{i,t-1} + G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i) + \delta_i \ln P_{i,t-1} + \mathrm{Promo}_{it}'\psi_i + \varepsilon_{it}, \tag{9}$$

where μ_{is} is given in (3), $G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i)$ is given in (5) together with (8) and $\varepsilon_{it} \sim \mathrm{N}(0, \sigma_i^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The likelihood function belonging to this model is

$$\begin{aligned}\ell(\mathrm{Data}|\zeta) = \prod_{i=1}^{N} \int_{\beta_i} \Bigg(\int_{\mu_{i1}} \cdots \int_{\mu_{iS}} & \prod_{t=1}^{T} \phi(\varepsilon_{it}; 0, \sigma_i^2) \\ & \times \prod_{s=1}^{S} \phi\left(\mu_{is}; \alpha_{i0} + \alpha_{i1}\cos\left(2\pi\frac{s}{S} - \alpha_{i2}\right), \sigma_{\eta_i}^2\right) \mathrm{d}\mu_{i1} \cdots \mathrm{d}\mu_{iS} \Bigg) \\ & \times \phi(\beta_i; \theta' Z_i, \Sigma)\, \mathrm{d}\beta_i,\end{aligned} \tag{10}$$

where $\phi(\cdot; m, \Omega)$ is the density function of a normal distribution with mean m and covariance matrix Ω and

$$\varepsilon_{it} = \Delta \ln S_{it} - \sum_{s=1}^{S} D_{st}\mu_{is} - \rho_i \ln S_{i,t-1} - G(\Delta \ln P_{it}; \beta_i, \gamma, \tau_i) - \delta_i \ln P_{i,t-1} - \mathrm{Promo}_{it}'\psi_i. \tag{11}$$

The model parameters are summarized by $\zeta = (\{\alpha_i, \rho_i, \delta_i, \psi_i, \tau_i, \sigma_i^2, \sigma_{\eta_i}^2\}_{i=1}^N, \Sigma, \theta)$. To estimate these parameters we opt for a Bayesian approach. Posterior results are obtained using MCMC techniques (Tierney, 1994; Smith and Roberts, 1993), in particular, the Gibbs sampling technique of Geman and Geman (1984) with data augmentation (Tanner and Wong, 1987). The latent variables $\mathscr{S} = \{\{\mu_{is}\}_{s=1}^{S}, \beta_i\}_{i=1}^{N}$ are sampled alongside the model parameters, see, for example, Rossi et al. (2005) for a recent overview of the analysis of hierarchical Bayes models in marketing.

3.1. Prior specification and posterior simulation

For the model parameters in the first layer of the model we impose an uninformative prior, that is,

$$p(\rho_i, \delta_i, \psi_i, \sigma_i^2) \propto \sigma_i^{-2} \tag{12}$$

for $i = 1, \ldots, N$. To be able to compute Bayes factors for the absence of seasonal effects in the sales series, see Section 3.2, we assume a normal prior for the α_{i1} parameters

$$\alpha_{i1} \sim \mathrm{N}(0, \sigma_{\alpha_1}^2) \tag{13}$$

for $i = 1, \ldots, N$. For the remaining α parameters we take a flat prior, that is,

$$p(\alpha_{i0}) \propto 1 \quad \text{and} \quad p(\alpha_{i2}) = \frac{1}{2\pi} \times \mathbb{I}[0, 2\pi] \tag{14}$$

for $i = 1, \ldots, N$, where $\mathbb{I}[a, b]$ is an indicator function that equals 1 on the interval $[a, b]$ and 0 otherwise. For smoother convergence of the Gibbs sampler, we define proper but relatively uninformative priors for the variances of the error terms in the two second layers of the model, see Hobert and Casella (1996) for a discussion on this issue. For $\sigma_{\eta_i}^2$ we take an inverted Gamma-2 prior distribution with scale parameter v and degrees of freedom v

$$\sigma_{\eta_i}^2 \sim \mathrm{IG} - 2(v, v) \tag{15}$$

for $i = 1, \ldots, N$ and for Σ we take an inverted Wishart prior distribution with scale parameter V and degrees of freedom λ,

$$\Sigma \sim \mathrm{IW}(V, \lambda). \tag{16}$$

The prior for θ is uninformative and given by

$$p(\theta) \propto 1. \tag{17}$$

Finally, to identify the price regimes we impose a prior on the τ_{ij} parameters. The prior is normal on the region $[0, ub]$, that is,

$$\tau_{ij} \sim \mathrm{N}(\mu_\tau, \sigma_\tau^2) \times \mathbb{I}[0, ub] \tag{18}$$

for $j = 1, 2$ and $i = 1, \ldots, N$. Note that we do not use explanatory variables to explain differences in the τ parameters across brands as our main focus is the non-linear price effect.

As mentioned before, we fix γ at a rather high value. For such a value, the transition function shows an abrupt change from one regime to the other. The reason for fixing the value of γ is that in practice it turns out to be very difficult to conduct inference on this parameter, see Bauwens et al. (1999, Chapter 8) for a discussion in a Bayesian setting. Another motivation follows from the fact that in our model the thresholds are stochastic. In fact, uncertainty in the thresholds leads to a model in which the transition between regimes is not immediate. So, even if we restrict γ to be large our model still allows for a smooth transition between regimes. Note that this feature of the model complicates the practical identification of γ even more.

The joint prior density for ζ denoted by $p(\zeta)$ is given by (12)–(18). The posterior density is equal to $p(\zeta)\ell(\text{Data}|\zeta)$. In Appendix A we derive the full conditional posterior distributions, which are needed to sample from the posterior distribution using MCMC, see also Rossi et al. (2005).

3.2. Testing for weekly seasonality

To test for the presence of weekly seasonality in the sales series we use Bayes factors. We compare the model with weekly seasonality to a model where we restrict the regular seasonal component of series i to be zero, that is, $\alpha_{i1} = 0$. Hence, we analyze the absence of the deterministic seasonal part.

The Bayes factor for $\alpha_{i1} = 0$ is given by

$$\text{BF}_i = \frac{\int p(\zeta)\ell(\text{Data}|\zeta)\,\text{d}\zeta}{\int p_0(\zeta_0)\ell_0(\text{Data}|\zeta_0)\,\text{d}\zeta_0}, \tag{19}$$

where $p_0(\zeta_0)$ and $\ell_0(\text{Data}|\zeta_0)$ denote the prior density and the likelihood function for $\alpha_{i1} = 0$, respectively, and ζ_0 summarizes the parameters in case $\alpha_{i1} = 0$. The prior density $p(\zeta_0)$ follows from (12) and (14)–(18). To compute this Bayes factor we use the Savage–Dickey density ratio of Dickey (1971), see also Verdinelli and Wasserman (1995). The Bayes factor (19) is equal to

$$\text{BF}_i = \frac{p(\alpha_{i1}|\text{Data})|_{\alpha_{i1}=0}}{p(\alpha_{i1})|_{\alpha_{i1}=0}}, \tag{20}$$

that is, the ratio of the height of the marginal posterior density of α_{i1} and the height of the marginal prior density of α_{i1}, both evaluated at $\alpha_{i1} = 0$, see Koop and Potter (1999) for a similar approach. The height of the marginal prior follows directly from (13). The marginal posterior density of α_{i1} is given by

$$\begin{aligned} p(\alpha_{i1}|\text{Data}) &= \int p(\alpha_{i1}|\zeta_0, \text{Data})p(\zeta_0|\text{Data})\,\text{d}\zeta_0 \\ &= \int p(\alpha_{i1}|\zeta_0, \mathscr{S}, \text{Data})p(\zeta_0, \mathscr{S}|\text{Data})\,\text{d}\zeta_0\,\text{d}\mathscr{S}, \end{aligned} \tag{21}$$

where ζ_0 summarizes the model parameters without α_{i1} and $p(\alpha_{i1}|\zeta_0, \mathscr{S}, \text{Data})$ is the full conditional posterior density of α_{i1}. The height of the marginal posterior density of α_{i1} in $\alpha_{i1} = 0$ can easily be computed by averaging the full conditional posterior density of α_{i1} over the MCMC draws of the other parameters, that is,

$$p(\alpha_{i1}|\text{Data})|_{\alpha_{i1}=0} = \frac{1}{M}\sum_{m=1}^{M} p(\alpha_{i1}|\zeta_0^{(m)}, \mathscr{S}^{(m)}, \text{Data})|_{\alpha_{i1}=0}, \tag{22}$$

see Gelfand and Smith (1990).

Bayes factors can be very sensitive to the prior specification. To investigate this sensitivity one may compute Bayes factors for different prior specifications. Geweke (2005, Section 8.4) shows that one can use importance sampling techniques to compute posterior results for a different prior setting. Let $\tilde{p}(\alpha_{i1})$ denote an alternative prior specification and

$\tilde{p}(\zeta|\text{Data})$ the corresponding posterior density. The height of the marginal posterior density in $\alpha_{i1} = 0$ is given by

$$
\begin{aligned}
&\tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \\
&\quad = \int \tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \tilde{p}(\zeta_0|\text{Data})\,\mathrm{d}\zeta_0 \\
&\quad = \int \tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \int \tilde{p}(\zeta_0|\text{Data})\tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})\,\mathrm{d}\alpha_{i1}\,\mathrm{d}\zeta_0 \\
&\quad = \iint \tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \tilde{p}(\zeta_0|\text{Data})\tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})\,\mathrm{d}\alpha_{i1}\,\mathrm{d}\zeta_0 \\
&\quad = \iint \tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \frac{\tilde{p}(\zeta_0|\text{Data})\tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})}{p(\alpha_{i1}, \zeta_0|\text{Data})} p(\alpha_{i1}, \zeta_0|\text{Data})\,\mathrm{d}\alpha_{i1}\,\mathrm{d}\zeta_0 \\
&\quad = \int \tilde{p}(\alpha_{i1}|\zeta_0, \text{Data})|_{\alpha_{i1}=0} \frac{\tilde{p}(\zeta|\text{Data})}{p(\zeta|\text{Data})} p(\zeta|\text{Data})\,\mathrm{d}\zeta, \qquad (23)
\end{aligned}
$$

where we adjust the arguments of Geweke (2005) for the computation of marginal densities. Hence, the height of the marginal density for the prior specification $\tilde{p}(\alpha_{i1})$ can be computed as follows:

$$
\tilde{p}(\alpha_{i1}|\text{Data})|_{\alpha_{i1}=0} = \frac{\sum_{m=1}^{M} w(\alpha_{i1}^{(m)})\tilde{p}(\alpha_{i1}|\zeta_0^{(m)}, \mathscr{S}^{(m)}, \text{Data})|_{\alpha_{i1}=0}}{\sum_{m=1}^{M} w(\alpha_{i1}^{(m)})}, \qquad (24)
$$

where $w(\alpha_{i1}) = \tilde{p}(\alpha_{i1})/p(\alpha_{i1})$, and where $\tilde{p}(\alpha_{i1}|\zeta_0, \mathscr{S}, \text{Data})$ is the full conditional posterior distribution belonging to the prior specification $\tilde{p}(\alpha_{i1})$.

Finally, note that also the irregular part η_{is} contributes to the seasonal pattern. To analyze the influence of this component, we compare its variance $\sigma^2_{\eta_i}$ with the variance of the deseasonalized sales series σ_i^2.

4. An illustration

To illustrate the usefulness of our model, we consider weekly sales volumes for 96 brands of fast moving consumer goods in 24 distinct categories. These data are obtained from the database of the US supermarket chain Dominick's Finer Foods. The data cover the period September 1989 to May 1997 in the Chicago area. The same data are used in Srinivasan et al. (2004). We take the top four brands of each product category. Next to the actual price P_{it} we include in Promo_{it} the typical promotion variables display and feature.

To explain the three price elasticities for each brand, we collect and construct a range of explanatory variables (Z_i). Some of these variables correspond to the characteristics of the product category, while other variables correspond to the characteristics of the brand itself. Table 1 contains a list of the variables with their explanation. The selection of these variables are based on earlier studies on variables influencing price elasticities, see Fok et al. (2006) for a discussion and details on the variables. All variables can be straightforwardly derived from the available data. In case of availability one could want to include detailed information on the product category and brands themselves, but here we cannot.

We now turn to the estimation results. We mainly impose weakly informative priors, that is we set $\sigma^2_{\alpha_1} = 1$, $v = 0.15$, $\nu = 5$, $V = \mathbf{I}_3$, and $\lambda = 6$. For the thresholds we set the

Table 1
Explanatory variables Z for the second-level model (8)

Variable	Description
Category-level characteristics	
Price dispersion	Average distance between the highest and the lowest regular price
Concentration index	Measured through $\sum_i M_i \log M_i$, where M_i denotes the average market share of brand i in a category
Price promotion frequency	Frequency with which at least one brand in the market has a price promotion
Depth of price promotion	Average size of the price promotion
Display frequency	Frequency of at least one product on display in a category
Feature frequency	Frequency of at least one featured product in a category
Hedonic	Dummy variable indicating if the product has a hedonic nature
Brand-level characteristics	
Price index	Average price relative to the average price in the category
Brand size	Average market share of the brand
Relative price promotion frequency	Frequency of price promotion divided by the category price promotion frequency
Relative depth of price promotions	Depth of price promotion divided by the depth of price promotion for the category
Relative feature frequency	Frequency of feature relative to the frequency of at least one feature in the category
Relative display frequency	Frequency of display relative to the frequency of at least one feature in the category
Market leader	Dummy variable for the brand with the highest average market share

prior parameters such that the regimes correspond to small versus large price changes, that is we set $\mu_\tau = 0.1$, $\sigma_\tau^2 = 0.025$ and $ub = 0.4$. Hence, we expect the threshold between small and large price changes to be around 10% and we restrict it not to be larger than 40%. Finally, we set $\gamma = 50$.

These results are based on 40,000 draws of our MCMC sampler, where the first 25,000 draws are discarded and out of the remaining draws we only use each 5th draw to obtain a reasonable random sample from the posterior distribution.

4.1. Weekly seasonality

First we comment on the seasonal component of our model. We use the Bayes factor to compare our model to models where one of the α_{i1} parameter equals 0 for $i = 1, \ldots, 96$, that is a model with no deterministic seasonality for one of the brands. The Bayes factors are computed using the Savage–Dickey density ratio as described in Section 3.2. Of these 96 Bayes factors, 15 are smaller than 1, that is, for 15 brands we prefer the model with a clear sinusoid seasonal pattern. These 15 brands are in the product categories beer, oatmeal, crackers, canned soup, snacks, and frozen dinners. To judge the sensitivity of these results, we calculate the Bayes factor under a number of alternative prior distributions using the importance sampling technique. The alternative priors all have mean 0 but a different prior variance. If we increase the prior standard deviation to 2 or even 5, the Bayes factors indicate the presence of a clear sinusoid pattern in 13 and 10 brands, respectively. This result follows from the well-known fact that under more diffuse prior specifications the Bayes factor will always favor the model with less parameters. For smaller prior standard deviations like 0.1 and 0.5, Bayes factors favor the sinusoid pattern in 25 and 17 brands, respectively. Overall, for moderate values of the prior variance, the results do not seem to be very sensitive on the choice of the prior.

In Fig. 3 we present a histogram of the Bayes factors for all the brands in our data. Note that there are quite some brands for which there is clear evidence that $\alpha_{i1} = 0$, implying that some Bayes factors are very large. Note that this does not automatically imply that all of these brands do not show any seasonal pattern. Indeed, the unexplained stochastic part of (3), that is η_{is}, also contributes to the seasonal pattern. As mentioned before, in contrast to the deterministic part, this part of the seasonality is not smooth. By comparing the variance of η_{is} to the variance of ε_{it} we can evaluate the relative importance of the irregular seasonal pattern. Fig. 4 shows a histogram of $\mathrm{E}[\sigma_i^2|\mathrm{Data}]/\mathrm{E}[\sigma_{\eta_i}^2|\mathrm{Data}]$. Smaller values of this fraction indicate stronger seasonal patterns. All values appear larger than 1 and this indicates that for all brands the variance of the ε_{it} is larger than the variance of η_{is}. For more than half of the brands the fraction is smaller than 5.

These results show that there are brands for which we do not find a deterministic seasonal pattern, but for which the stochastic component of the seasonality is important. In these cases the sales do not show a smooth seasonal pattern but rather a pattern of (seasonally) recurring spikes and dips. These spikes or dips may correspond for example to special holidays. In Fig. 5 we show the seasonal pattern that we find for the first brand in each product category. It is clear that for some brands we find no seasonality at all, that for

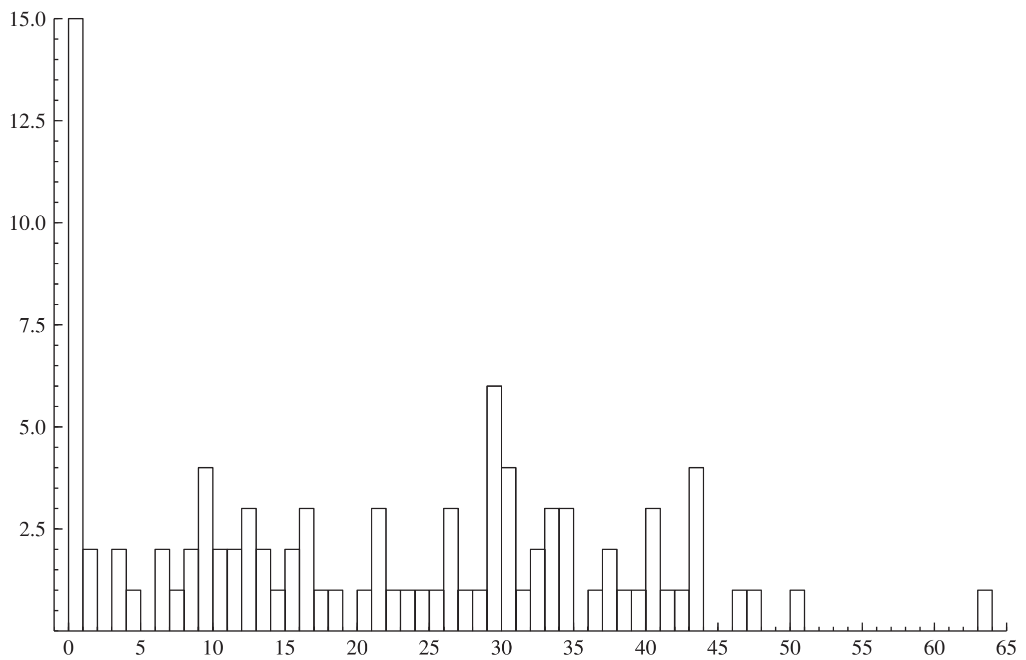

Fig. 3. Histogram of the Bayes factors for $\alpha_{i1} = 0$ for $i = 1, \ldots, 96$.

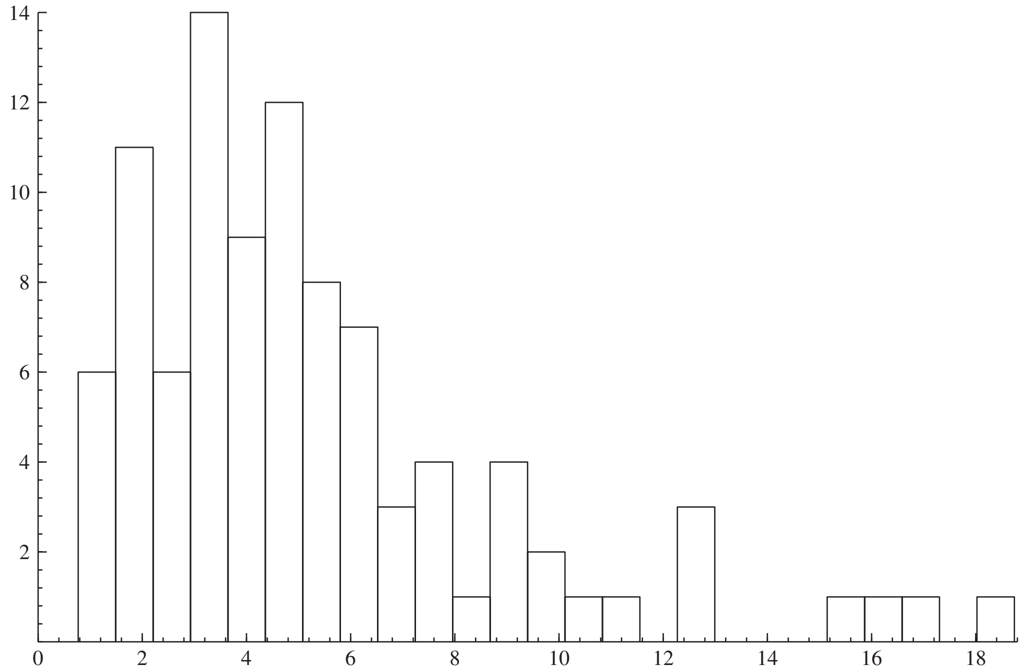

Fig. 4. Histogram of posterior mean of σ_i^2 divided by posterior mean of $\sigma_{\eta_i}^2$.

others we find a relatively smooth cycle, while yet for other brands the seasonal patterns correspond to just a few spikes in sales.

4.2. Non-linear price elasticity

We now continue with the price elasticity. In Table 2 we present the estimation results corresponding to (8). It is important to note that all explanatory variables are standardized, that is, they are transformed to have mean 0 and variance 1. The intercepts in (8) can therefore directly be interpreted as the posterior mean of β_i. When we compare these posterior means we see that, overall, the elasticity for small price changes is the largest, followed by the elasticity of large price decreases. The elasticity of large increases is the smallest. The marginal effect of a price change is smaller for large price changes relative to small changes. These findings are consistent with the marketing literature, where saturation effects are found for large price changes (see for example, van Heerde et al., 2001; Fox et al., 2004). Furthermore, the research on brand loyalty shows that the elasticity of large price increases is limited. It turns out that for large price decreases the saturation effect is stronger than the stockpiling effect that tends to occur for price decreases (Blattberg et al., 1995).

The posterior results show that in general the sales–response curve is monotonically decreasing. The 99% highest posterior density region for the intercepts do not include zero. The posterior probability of a monotonic response can easily be calculated using the posterior draws of β_i, this probability is over 90%.

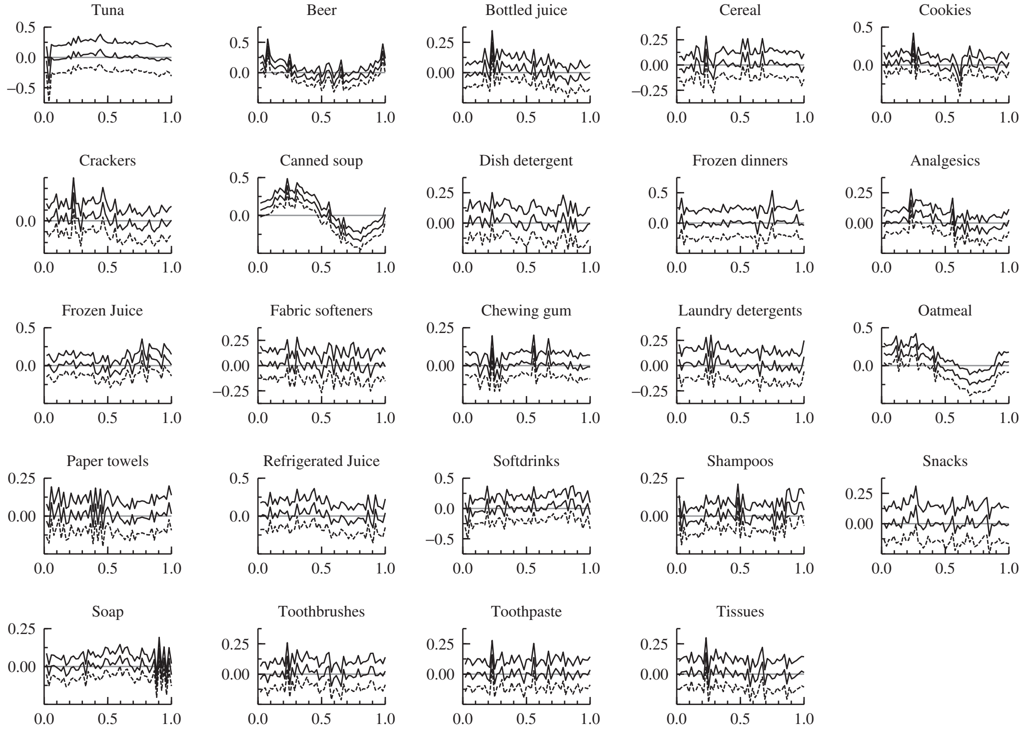

Fig. 5. Some examples of the posterior mean of $\mu_{is} = \alpha_{i1}\cos(2\pi(s/S) - \alpha_{i2}) + \eta_{is}$, with 95% highest posterior density region. The horizontal axis corresponds to 1 year, note that the start of the year may correspond to different months for different categories.

The threshold parameters define the regions which mark large price changes. The posterior means of these threshold parameters vary in the range of 0.009–0.266 for price increases, and in the range 0.02–0.298 for price decreases, the average posterior means are 0.123 and 0.171, respectively. In general, price changes larger than 12% are considered to be large.

In Fig. 6 we show the price effects for the four brands in a representative category (softdrinks), and we take this case to show how estimation results can be interpreted. We show sales–response graphs, similar to Fig. 1, and graphs of the price elasticity, similar to Fig. 2. The domain of each graph corresponds to price changes actually observed in the sample. For three of the four brands we find that the sales–response curve flattens for large price increases. This could imply that for these brands there is a large segment of loyal consumers, who do not switch to another brand even if the price increases to a large extent. Next, for large price increases the sales–response curve is not as steep as for small price changes. The threshold for large price decreases (τ_2) is approximately 0.18, the threshold for large price increases (τ_1) is about 0.10.

Table 2 also shows the relevance of some of the brand and category characteristics for explaining the non-linear price effects. We see that larger brands tend to have smaller price elasticities for large price increases and large price decreases, while the elasticity for small price changes seems to be unrelated to the brand size. The relative price promotion

Table 2
Posterior mean of second-level parameters θ (Eq. (8)), with the posterior standard deviation in parentheses

Z_i	Small price changes θ_0		Large price increase θ_1		Large price decrease θ_2	
Intercept	-2.432^{c}	(0.106)	-1.380^{c}	(0.117)	-2.014^{c}	(0.133)
Brand-level characteristics						
Price index	0.161	(0.116)	−0.103	(0.142)	0.096	(0.140)
Brand size	0.081	(0.174)	0.339^{b}	(0.178)	0.549^{c}	(0.201)
Relative price promotion frequency	0.319^{b}	(0.129)	−0.046	(0.143)	−0.148	(0.172)
Relative depth price promotion	0.024	(0.109)	0.264^{c}	(0.106)	0.150	(0.119)
Relative feature frequency	−0.152	(0.129)	0.108	(0.137)	0.183	(0.156)
Relative display frequency	-0.253^{b}	(0.110)	0.031	(0.124)	0.133	(0.141)
Market leader	0.018	(0.159)	−0.171	(0.168)	−0.038	(0.184)
Category-level characteristics						
Price dispersion	0.252^{b}	(0.113)	−0.145	(0.143)	0.220	(0.149)
Concentration index	−0.022	(0.114)	−0.151	(0.120)	−0.227	(0.151)
Price promotion frequency	−0.238	(0.158)	−0.252	(0.160)	−0.206	(0.185)
Depth price promotion	−0.079	(0.137)	0.150	(0.155)	0.029	(0.174)
Display frequency	−0.124	(0.167)	−0.039	(0.167)	−0.021	(0.191)
Feature frequency	−0.125	(0.149)	0.328^{b}	(0.139)	0.220	(0.158)
Hedonic	−0.043	(0.126)	0.174	(0.139)	0.292^{a}	(0.160)

Covariance of random effects ξ_i

$$\Sigma = \begin{pmatrix} 0.7659 & 0.2307 & 0.0015 \\ 0.2307 & 0.6808 & 0.5496 \\ 0.0015 & 0.5496 & 0.8427 \end{pmatrix}$$

[a,b,c]Zero not contained in 90%, 95% or 99% highest posterior density region, respectively.

frequency and the relative display frequency are only related to the elasticity of small price changes. A high relative price promotion frequency corresponds to a small elasticity. A possible explanation for this is that for brands with many price promotions, consumers learn to anticipate the promotions (Erdem et al., 2003). In this case the stockpiling effect for a price decrease may be small. A high display frequency corresponds to a large elasticity for small price changes. Bolton (1989) already found such a relationship for the price elasticity in general. Note that we do not find a moderating effect of display frequency on the effectiveness of large price changes. Furthermore, in categories with high price dispersion, the elasticity of a small price change is relatively small. Finally, from the second panel we learn that brands, in a category with a high feature activity or brands that have relatively deep price discounts, tend to have a smaller elasticity of large price increases. For large price decreases we find that hedonic categories show smaller elasticities than utilitarian categories.

The covariance matrix of the random component (ξ_i) of the price effects is given in bottom panel of Table 2. Compared to the mean effects and the size of the other parameters, the magnitude of the random component is quite large. Hence, the price effects differ widely across brands, and only a relatively small part of this variation can be explained by our second-level model (8).

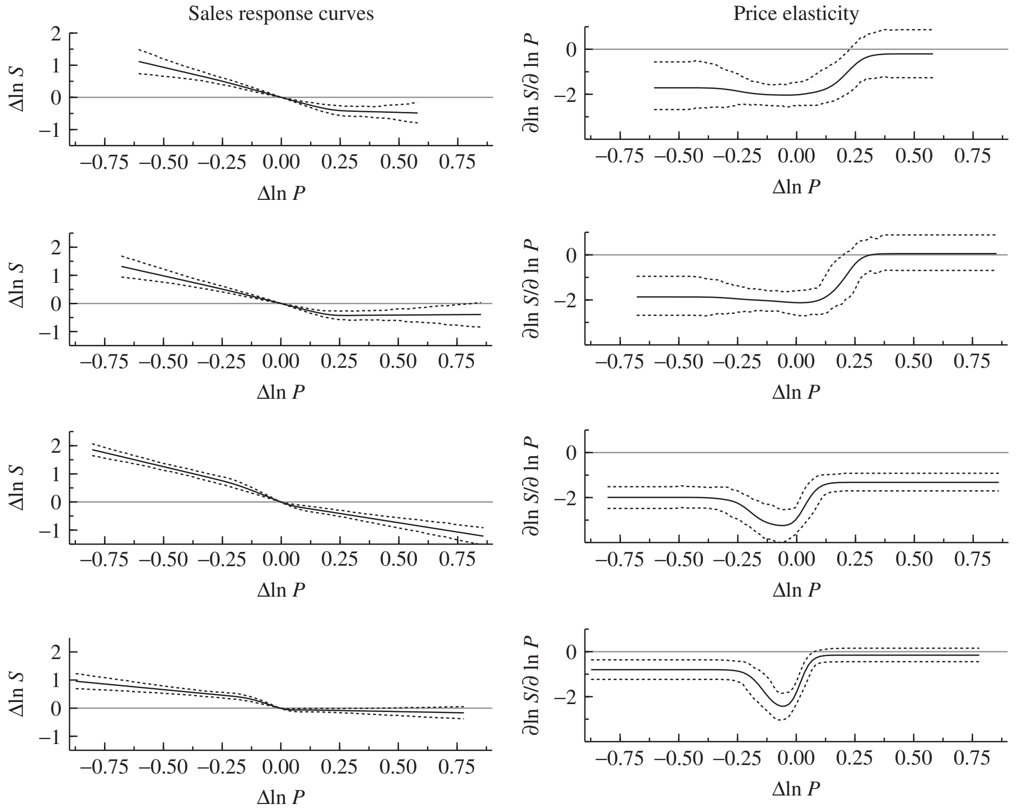

Fig. 6. Posterior mean of price effect for four brands in the softdrinks category (with 95% highest posterior density region).

5. Conclusion

In this paper we have put forward a hierarchical Bayes model for a panel of time series on sales and marketing activities, where we allowed for weekly seasonality and for non-linear asymmetric price effects. As sales data can be available at a weekly basis, the standard dummy variable approach to model seasonality involves too many parameters. Instead, we have proposed a combination of a deterministic cycle and of random effects to capture seasonality. In the empirical section we showed that this specification can capture a wide variety of seasonal patterns ranging from a smooth cycle to a pattern of recurring spikes and dips.

We also introduced flexibility in the sales–price curve, which is usually assumed as linear. Indeed, we allowed price increases to have a different effect than price decreases and we also distinguished between large price changes and small changes. In the empirical section we tried to explain possible differences in these price effects across brands.

We are tempted to conclude that our hierarchical Bayes model, while summarizing thousands of observations with varying features over time, over categories and over elasticities, kept interpretability of the parameters. Further work could address the forecasting power of models like ours as well as model selection issues.

Acknowledgments

We thank Herman van Dijk and two anonymous reviewers for their comments. Over the years, members of the Econometric Institute showed an interest in modeling data that are not necessarily of a macroeconomic nature, and that show seasonality and non-linearity. Also, panel data have been considered. We combine all these research interests into a single model for marketing data in the present paper, which has been prepared for the commemorative special issue of the Journal of Econometrics, celebrating the 50th anniversary of the Econometric Institute.

Appendix A. Full conditional posterior distributions

A.1. Sampling of β_i, ρ_i, δ_i and ψ_i

For notational convenience, we rewrite the model in (9) as

$$Y_{it} = X'_{it}(\beta'_i, \rho_i, \delta_i, \psi'_i)' + \varepsilon_{it} \tag{25}$$

for $t = 1, \ldots, T$, where $Y_{it} = \Delta \ln S_{it} - \sum_{s=1}^{S} D_{st}\mu_{is}$ and

$$X_{it} = \begin{pmatrix} \Delta \ln P_{it} - F(\Delta \ln P_{it}; \gamma, \tau_{i1})(\Delta \ln P_{it} - \tau_{i1}) - F(-\Delta \ln P_{it}; \gamma, \tau_{i2})(\Delta \ln P_{it} + \tau_{i2}) \\ F(\Delta \ln P_{it}; \gamma, \tau_{i1})(\Delta \ln P_{it} - \tau_{i1}) \\ F(-\Delta \ln P_{it}; \gamma, \tau_{i2})(\Delta \ln P_{it} + \tau_{i2}) \\ \ln S_{i,t-1} \\ \ln P_{i,t-1} \\ \text{Promo}_{it} \end{pmatrix}. \tag{26}$$

If we stack the T equations we obtain

$$Y_i = X_i(\beta'_i, \rho_i, \delta_i, \psi'_i)' + \varepsilon_i, \tag{27}$$

where $Y_i = (Y_{i1}, \ldots, Y_{iT})'$, $X_i = (X_{i1}, \ldots, X_{iT})'$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. The second layer of model (8) can be written as

$$-Z'_i\theta = -\beta'_i + \xi'_i. \tag{28}$$

If we collect and standardize Eqs. (25) and (28) we obtain

$$\begin{pmatrix} \sigma_i^{-1} Y_i \\ -\Sigma^{-1/2}\theta' Z'_i \end{pmatrix} = \begin{pmatrix} \sigma_i^{-1} X_i & & & \\ -\Sigma^{-1/2} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_i \\ \rho_i \\ \delta_i \\ \psi_i \end{pmatrix} + \begin{pmatrix} \sigma_i^{-1}\varepsilon_i \\ \Sigma^{-1/2}\xi_i \end{pmatrix}. \tag{29}$$

We define this equation as $Y_i^* = X_i^*(\beta_i, \rho_i, \delta_i, \psi'_i)' + e_i$, where e_i has a multivariate normal distribution with mean zero and an identity covariance matrix. From this final equation it is clear that the full conditional posterior distribution of $(\beta'_i, \rho_i, \delta_i, \psi_i)'$ is normal with mean $(X_i^{*\prime} X_i^*)^{-1}(X_i^{*\prime} Y_i^*)$ and covariance matrix $(X_i^{*\prime} X_i^*)^{-1}$, see, for example, Zellner (1971, Chapter III).

A.2. Sampling of μ_{is}

We use (26) to rewrite model (9) as

$$Y_{it} = \sum_{s=1}^{S} D_{st}\mu_{is} + \varepsilon_{it}, \tag{30}$$

where now $Y_{it} = \Delta \ln S_{i,t} - X'_{it}(\alpha_{i0}, \beta_i, \rho_i, \delta_i, \psi_i)'$ for $t = 1, \ldots, T$. If we stack the T equations we obtain

$$Y_i = D\mu_i + \varepsilon_i, \tag{31}$$

where $Y_i = (Y_{i1}, \ldots, Y_{iT})'$, $D = (D_1, \ldots, D_T)'$ with $D_t = (D_{1t}, \ldots, D_{St})'$, $\mu_i = (\mu_{i1}, \ldots, \mu_{iS})'$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. The stochastic seasonal model (3) can be written as

$$-\alpha_{i0} - \alpha_{i1}\cos\left(2\pi\frac{s}{S} - \alpha_{i2}\right) = -\mu_{is} + \eta_{is} \tag{32}$$

for $s = 1, \ldots, S$. If we collect and standardize the equations we obtain

$$\begin{pmatrix} \sigma_i^{-1} Y_i \\ -\sigma_{\eta_i}^{-1}\left(\alpha_{i0} + \alpha_{i1}\cos\left(2\pi\frac{1}{S} - \alpha_{i2}\right)\right) \\ -\sigma_{\eta_i}^{-1}\left(\alpha_{i0} + \alpha_{i1}\cos\left(2\pi\frac{2}{S} - \alpha_{i2}\right)\right) \\ \vdots \\ -\sigma_{\eta_i}^{-1}\left(\alpha_{i0} + \alpha_{i1}\cos\left(2\pi\frac{S}{S} - \alpha_{i2}\right)\right) \end{pmatrix} = \begin{pmatrix} \sigma_i^{-1} D \\ -\sigma_{\eta_i}^{-1}\mathbf{I}_S \end{pmatrix}\mu_i + \begin{pmatrix} \sigma_i^{-1}\varepsilon_i \\ \sigma_{\eta_i}^{-1}\eta_i \end{pmatrix}, \tag{33}$$

where $\mathbf{I}_S$ denotes an identity matrix of dimension S and where $\eta_i = (\eta_{i1}, \ldots, \eta_{iS})'$. This equation can be written as $Y_i^* = X_i^*\mu_i + e_i$, where e_i has a multivariate normal distribution with mean zero and an identity covariance matrix. Hence, the full conditional distribution of μ_i is normal with mean $(X_i^{*\prime}X_i^*)^{-1}(X_i^{*\prime}Y_i^*)$ and covariance matrix $(X_i^{*\prime}X_i^*)^{-1}$.

A.3. Sampling of τ_i

The full conditional posterior distribution of τ_i for $i = 1, \ldots, N$ does not have a standard form. We draw τ_{i1} and τ_{i2} separately using the Griddy Gibbs sampler of Ritter and Tanner (1992). The full conditional posterior of τ_{ij} is proportional to

$$\phi(\tau_{ij}; \mu_\tau, \sigma_\tau^2) \times \mathbb{I}[0, ub] \prod_{t=1}^{T} \phi(\varepsilon_{it}; 0, \sigma_i^2), \tag{34}$$

where ε_{it} is given in (11). We choose a grid on the region $[0, ub]$. For each value of τ_{ij} on the grid we calculate the relative height of the full conditional posterior density and construct an approximation of the cumulative distribution function (CDF) of the full conditional posterior density function. Finally, we sample a uniform random number and use the inverse CDF technique to generate a draw of τ_{ij} for $j = 1, 2$ and $i = 1, \ldots, N$.

A.4. Sampling of σ_i^2

Conditional on the other parameters, the posterior distribution of σ_i^2 is an inverted Gamma-2 distribution with scale parameter $\sum_{t=1}^{T} \varepsilon_{it}^2$ and degrees of freedom T and hence

$$\frac{\sum_{t=1}^{T} \varepsilon_{it}^2}{\sigma_i^2} \sim \chi^2(T), \tag{35}$$

where ε_{it} is given in (11) for $i = 1, \ldots, N$.

A.5. Sampling of α_i

If we condition on $\{\mu_{is}\}_{s=1}^{S}$, the only relevant part of the model for sampling α_i is

$$\mu_{is} = X'_{is} \begin{pmatrix} \alpha_{i0} \\ \alpha_{i1} \end{pmatrix} + \eta_{is} \tag{36}$$

for $s = 1, \ldots, S$ where $X_{is} = (1, \cos(2\pi(s/S) - \alpha_{i2}))'$ for $i = 1, \ldots, N$. Hence, the full conditional posterior distribution of $(\alpha_{i0}, \alpha_{i1})$ (conditional on α_{i2}) is normal with mean

$$\left(\frac{1}{\sigma_{\eta_i^2}} \sum_{s=1}^{S} X'_{is} X_{is} + \begin{pmatrix} \sigma_{\alpha_0}^{-2} & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1} \left(\frac{1}{\sigma_{\eta_i}^2} \sum_{s=1}^{S} X'_{is} \mu_{is} \right) \tag{37}$$

and covariance matrix

$$\left(\frac{1}{\sigma_{\eta_i^2}} \sum_{s=1}^{S} X'_{is} X_{is} + \begin{pmatrix} \sigma_{\alpha_0}^{-2} & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1} \tag{38}$$

for $i = 1, \ldots, S$. In case one wants to impose a flat prior for α_{i1} one has to replace $\sigma_{\alpha_0}^{-2}$ by 0.

The full conditional distribution of α_{i2} is not of a known form. To sample this parameter we again rely on the Griddy Gibbs sampler. The model restricts $0 < \alpha_{i2} < 2\pi$, this interval serves as natural bounds for our grid. On this grid we evenly distribute 75 points. For each point we calculate the relative height of the full conditional posterior density as

$$\prod_{s=1}^{S} \phi\left(\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos\left(2\pi \frac{s}{S} - \alpha_{i2} \right); 0, \sigma_{\eta_i}^2 \right). \tag{39}$$

Next we again use the inverse CDF technique to transform a uniform random draw into a draw from the full conditional distribution of α_{i2} for $i = 1, \ldots, N$.

A.6. Sampling $\sigma_{\eta_i}^2$

If we condition on $\{\mu_{is}\}_{s=1}^{S}$ and α_i, the only part of the model which is relevant for sampling $\sigma_{\eta_i}^2$ is (36) together with the prior for $\sigma_{\eta_i}^2$ (15). Hence, the full conditional posterior distribution of $\sigma_{\eta_i}^2$ is an inverted Gamma-2 with scale parameter $\sum_{s=1}^{S} (\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos(2\pi(s/S) - \alpha_{i2}))^2 + v$ and $S + v$ degrees of freedom. We can use that

$$\frac{\sum_{s=1}^{S} (\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos(2\pi(s/S) - \alpha_{i2}))^2 + v}{\sigma_{\eta_i}^2} \sim \chi^2(S + v). \tag{40}$$

A.7. Sampling of θ

If we condition on $\{\beta_i\}_{i=1}^N$ and Σ, the only part of the model which is relevant for sampling θ is

$$\beta_i' = Z_i'\theta + \xi_i' \tag{41}$$

for $i = 1, \ldots, N$. Stacking these N equations we obtain

$$\beta = Z\theta + \xi, \tag{42}$$

where $\beta = (\beta_1, \ldots, \beta_N)'$, $Z = (Z_1, \ldots, Z_N)'$ and $\xi = (\xi_1, \ldots, \xi_N)'$ with $\text{vec}(\xi) \sim \text{N}(0, \Sigma \otimes \mathbf{I}_k)$. Hence, the posterior distribution of $\text{vec}(\theta)$ is normal with mean $\text{vec}((Z'Z)^{-1}Z'\beta)$ and variance $\Sigma \otimes (Z'Z)^{-1}$, see, for example, Zellner (1971, Chapter VIII).

A.8. Sampling of Σ

If we condition on $\{\beta_i\}_{i=1}^N$ and θ the only part of the model which is relevant for sampling Σ is given in (42). Hence, the covariance matrix Σ can straightforwardly be sampled from an inverted Wishart distribution with scale parameter $(\beta - Z\theta)'(\beta - Z\theta) + V$ and $N + \lambda$ degrees of freedom.

References

Bauwens, L., Lubrano, M., Richard, J.F., 1999. Bayesian Inference in Dynamic Econometric Models. Advanced Texts in Econometrics. Oxford University Press, Oxford.

Blattberg, R.C., George, E.I., 1991. Shrinkage estimation of price and promotional elasticities-seemingly unrelated equations. Journal of the American Statistical Association 86, 304–315.

Blattberg, R.C., Briesch, R., Fox, E.J., 1995. How promotions work. Marketing Science 14, G122–G132.

Bolton, R.N., 1989. The relationship between market characteristics and promotional price elasticities. Marketing Science 8, 153–169.

Dickey, J., 1971. The weighted likelihood ratio, linear hypothesis on normal location parameters. Annals of Mathematical Statistics 42, 204–223.

Erdem, T., Imai, S., Keane, M.P., 2003. Brand and quantity choice dynamics under price uncertainty. Quantitative Marketing and Economics 1, 5–64.

Fok, D., Horváth, Cs., Paap, R., Franses, P.H., 2006. A hierarchical Bayes error correction model to explain dynamic effects of price changes. Journal of Marketing Research 43.

Fox, E.J., Montgomery, A.L., Lodish, L.M., 2004. Consumer shopping and spending across retail formats. Journal of Business 77, S25–S60.

Franses, P.H., van Dijk, D.J.C., 2000. Non-linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.

Gelfand, A.E., Smith, A.F., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.

Geman, S., Geman, D., 1984. Stochastic relaxations, Gibbs distributions, and the Bayesian restoration of images. IEEE Transaction on Pattern Analysis and Machine Intelligence 6, 721–741.

Geweke, J.F., 2005. Contemporary Bayesian Econometrics and Statistics. Wiley, New Jersey.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford.

Hanssens, D.M., Parsons, L.J., Schultz, R.L., 2001. Market Response Models: Econometric and Time Series Analysis, second ed. Kluwer Academic Publishers, Dordrecht.

Hendricks, W., Koenker, R., Poirier, D.J., 1979. Residential demand for electricity: an econometric approach. Journal of Econometrics 9, 33–57.

Hobert, J.P., Casella, G., 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 91, 1461–1473.

Jones, R.H., Brelsford, W.M., 1967. Time series with periodic structure. Biometrika 54, 403–407.

Koop, G., Potter, S.M., 1999. Bayes factors and nonlinearity: evidence from economic time series. Journal of Econometrics 88, 251–281.

Leeflang, P.S.H., Wittink, D.R., Wedel, M., Naert, P.A., 2000. Building Models for Marketing Decisions. Kluwer Academic Publishers, Dordrecht.

Nijs, V.R., Dekimpe, M.G., Steenkamp, J.B.E.M., Hanssens, D.M., 2001. The category-demand effects of price promotions. Marketing Science 20, 1–22.

Pauwels, K., Srinivasan, S., 2004. Who benefits from store brand entry? Marketing Science 23, 364–390.

Pauwels, K., Srinivasan, S., Franses, P.H., 2006. When do price thresholds matter in retail categories? Marketing Science, forthcomming.

Ritter, C., Tanner, M.A., 1992. Facilitating the Gibbs sampler—the Gibbs stopper and the Griddy–Gibbs sampler. Journal of the American Statistical Association 87, 861–868.

Rossi, P.E., Allenby, G.M., McCulloch, R.E., 2005. Bayesian Statistics and Marketing. Wiley, New Jersey.

Smith, A.F.M., Roberts, G.O., 1993. Bayesian computing via the Gibbs sampler and related Markov Chain Monte Carlo methods. Journal of the Royal Statistical Society B 55, 3–23.

Srinivasan, S., Pauwels, K., Hanssens, D.M., Dekimpe, M.G., 2004. Do promotions benefit manufacturers, retailers, or both? Management Science 50, 617–629.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–550.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.

van Heerde, H.J., Leeflang, P.S., Wittink, D.R., 2001. Semiparametric analysis to estimate the deal effect curve. Journal of Marketing Research 38, 197–215.

van Heerde, H.J., Gijsbrechts, E., Pauwels, K., 2005. Price war: what is it good for? Store incidence and basket size response to the price war in dutch grocery retailing. Working paper, Tilburg University.

Verdinelli, I., Wasserman, L., 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. Journal of the American Statistical Association 90, 614–618.

Wittink, D.R., Addona, M., Hawkes, W., Porter, J., 1988. SCAN*PRO: the estimation, validation and use of promotional effects based on scanner data. Working paper, AC Nielsen, Schaumburg, Illinois, USA.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.