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

ECONOMETRIC INSTITUTE REPORT EI 2005-45

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, multi-level 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 in a limited number of yearly data and (ii) non-linear price effects due to historic reference prices. We discuss representation and inference and we propose an estimation method using Bayesian techniques. An illustration to a market-response model for 96 brands for about 8 years of weekly data shows the merits of our approach.

Key words: Panels of time series; Weekly seasonality; Threshold Models; Non-linearity; Bayes estimation; MCMC

---

*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.

†Corresponding author: Erasmus University Rotterdam, Econometric Institute, H 11-23, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: dfok@few.eur.nl

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 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 four to eight 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.* (2005). 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 to 8 years.

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.* (2005) among others. In the present paper, we also propose such a two-level model and we use a Bayesian approach to estimate the model parameters.

The first advantage of multiple-level 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 harmonic 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.* (2005) 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 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, these 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 a Markov Chain Monte Carlo (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 the potential limitations of our analysis and with a discussion of further research avenues.

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

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

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 $\varepsilon_{it} \sim \text{N}(0, \sigma_i^2)$, 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,

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

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.* (2005) 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 seasonality to show a regular cyclical pattern. In this paper we propose to take advantage of such a possible pattern. To keep things simple, we specify the season by a deterministic cycle with a period of 1 year with a stochastic factor which gives the deviation from this perfect cycle. In sum, we propose to model μ_{is} by

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

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} ($0 \leq \alpha_{i2} \leq 2\pi$) 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 deterministic part of the seasonal pattern corresponds to a regular cycle, while 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 4 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

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

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}. Figure 1 graphically depicts the resulting sales response curve.

– Insert Figure 1 about here –

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,

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

where the last line follows from the fact that

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

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

The expression in (6) and the line in Figure 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 Figure 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.

– Insert Figure 2 about here –

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.* (2005) 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.

3 Bayes analysis

The total model is given by

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

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 $\varepsilon_{it} \sim \text{N}(0, \sigma_i^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The likelihood function belonging to this model is

$$\ell(\text{Data}|\zeta) = \prod_{i=1}^{N} \int_{\beta_i} \Big(\int_{\mu_{i1}} \cdots \int_{\mu_{iS}} \prod_{t=1}^{T} \phi(\varepsilon_{it}; 0, \sigma_i^2) \prod_{s=1}^{S} \phi(\mu_{is}; \alpha_{i0} + \alpha_{i1} \cos(2\pi \frac{s}{S} - \alpha_{i2}), \sigma_{\eta_i}^2) d\mu_{i1} \cdots d\mu_{iS} \Big) \phi(\beta_i; \theta' Z_i, \Sigma) d\beta_i, \quad (10)$$

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

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

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 $\{\{\mu_{is}\}_{s=1}^S, \beta_i\}_{i=1}^N$ are sampled alongside the model parameters.

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

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

for $i = 1, \ldots, N$. To be able to compute Bayes factors for the absence of seasonal effects in the sales series, see Section 3.1, we assume a normal prior for the α_{i1} parameters

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

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

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

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 $\sigma^2_{\eta_i}$ we take an inverted Gamma-2 prior distribution with scale parameter v and degrees of freedom ν

$$\sigma^2_{\eta_i} \sim \text{IG-2}(v, \nu) \tag{15}$$

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

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

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,

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

for $j = 1, 2$ and $i = 1, \ldots, N$.

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) 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 distribution is equal to $p(\zeta)\ell(\text{Data}|\zeta)$. Below, we derive the steps of the MCMC sampler to sample from this posterior distribution.

Sampling of β_i, ρ_i, δ_i and ψ_i

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

$$Y_{it} = X'_{it}(\beta'_i, \rho_i, \delta_i, \psi'_i)' + \varepsilon_{it}, \tag{19}$$

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

$$X_{it} = \begin{pmatrix} \Delta \ln P_{it} - F(\Delta \ln P_{it}; \gamma, \tau_{i1})(\Delta \ln P_{it} - \tau_{i1}) - F(-\Delta \ln P_{it}; \gamma, \tau_{i2}))(\Delta \ln P_{it} + \tau_{i2}) \\ F(\Delta \ln P_{it}; \gamma, \tau_{i1})(\Delta \ln P_{it} - \tau_{i1}) \\ F(-\Delta \ln P_{it}; \gamma, \tau_{i2}))(\Delta \ln P_{it} + \tau_{i2}) \\ \ln S_{i,t-1} \\ \ln P_{i,t-1} \\ \text{Promo}_{it} \end{pmatrix}. \tag{20}$$

If we stack the T equations we obtain

$$Y_i = X_i(\beta_i', \rho_i, \delta_i, \psi_i')' + \varepsilon_i, \tag{21}$$

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 the model (8) we can write as

$$-Z_i'\theta = -\beta_i' + \xi_i'. \tag{22}$$

If we collect and standardize the equations (19) and (22) we obtain

$$\begin{pmatrix} \sigma_i^{-1} Y_i \\ -\Sigma^{-\frac{1}{2}}\theta' Z_i' \end{pmatrix} = \begin{pmatrix} & \sigma_i^{-1} X_i & & \\ -\Sigma^{-\frac{1}{2}} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_i \\ \rho_i \\ \delta_i \\ \psi_i \end{pmatrix} + \begin{pmatrix} \sigma_i^{-1}\varepsilon_i \\ \Sigma^{-\frac{1}{2}}\xi_i \end{pmatrix}. \tag{23}$$

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).

Sampling of μ_{is}

We use (20) to rewrite model (9) as

$$Y_{it} = \sum_{s=1}^{S} D_{st}\mu_{is} + \varepsilon_{it} \tag{24}$$

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

$$Y_i = D\mu_i + \varepsilon_i, \tag{25}$$

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

$$-\alpha_{i0} - \alpha_{i1}\cos(2\pi\frac{s}{S} - \alpha_{i2}) = -\mu_{is} + \eta_{is}. \tag{26}$$

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

$$\begin{pmatrix} \sigma_i^{-1} Y_i \\ -\sigma_{\eta_i}^{-1}(\alpha_{i0} + \alpha_{i1}\cos(2\pi\frac{1}{S} - \alpha_{i2})) \\ -\sigma_{\eta_i}^{-1}(\alpha_{i0} + \alpha_{i1}\cos(2\pi\frac{2}{S} - \alpha_{i2})) \\ \vdots \\ -\sigma_{\eta_i}^{-1}(\alpha_{i0} + \alpha_{i1}\cos(2\pi\frac{S}{S} - \alpha_{i2})) \end{pmatrix} = \begin{pmatrix} \sigma_i^{-1} D \\ -\sigma_{\eta_i}^{-1}\mathbf{I}_S \end{pmatrix} \mu_i + \begin{pmatrix} \sigma_i^{-1}\varepsilon_i \\ \sigma_{\eta_i}^{-1}\eta_i \end{pmatrix}, \tag{27}$$

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 $(X_i^{*\prime}X_i^*)^{-1}$.

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

$$\phi(\tau_{ij}; \mu_\tau, \sigma_\tau^2) \times \mathbb{I}[0, ub] \prod_{t=1}^{T} \phi(\varepsilon_{it}; 0, \sigma_i^2), \tag{28}$$

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 $i = 1, \ldots, N$.

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

$$\frac{\sum_{t=1}^{T} \varepsilon_{it}^2}{\sigma_i^2} \sim \chi^2(T), \tag{29}$$

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

Sampling of $\boldsymbol{\alpha_i}$

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

$$\mu_{is} = X_{is}' \begin{pmatrix} \alpha_{i0} \\ \alpha_{i1} \end{pmatrix} + \eta_{is}. \tag{30}$$

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

$$\left(\frac{1}{\sigma_{\eta_i^2}} \sum_{s=1}^{S} X_{is}' X_{is} + \begin{pmatrix} \sigma_{\alpha_0}^{-2} & 0 \\ 0 & 0 \end{pmatrix}\right)^{-1} \left(\frac{1}{\sigma_{\eta_i}^2} \sum_{s=1}^{S} X_{is}' \mu_{is}\right) \tag{31}$$

and covariance matrix

$$\left(\frac{1}{\sigma_{\eta_i^2}} \sum_{s=1}^{S} X_{is}' X_{is} + \begin{pmatrix} \sigma_{\alpha_0}^{-2} & 0 \\ 0 & 0 \end{pmatrix}\right)^{-1} \tag{32}$$

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 the relative height of the full conditional posterior density as

$$\prod_{s=1}^{S} \phi(\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos(2\pi \frac{s}{S} - \alpha_{i2}); 0, \sigma_{\eta_i}^2). \tag{33}$$

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 $i = 1, \ldots, N$.

Sampling $\boldsymbol{\sigma_{\eta_i}^2}$

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 (30) 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 $\sum_{s=1}^{S}(\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos(2\pi\frac{s}{S} - \alpha_{i2}))^2 + \upsilon$ and $S + \nu$ degrees of freedom. We can use that

$$\frac{\sum_{s=1}^{S}(\mu_{is} - \alpha_{i0} - \alpha_{i1} \cos(2\pi\frac{s}{S} - \alpha_{i2}))^2 + \upsilon}{\sigma_{\eta_i}^2} \sim \chi^2(S + \nu). \tag{34}$$

Sampling of $\boldsymbol{\theta}$

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

$$\beta_i' = Z_i'\theta + \xi_i' \tag{35}$$

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

$$\beta = Z\theta + \xi, \tag{36}$$

where $\beta = (\beta_1, \ldots, \beta_N)'$, $Z = (Z_1, \ldots, Z_N)'$ and $\xi = (\xi_1, \ldots, \xi_N)'$ with $\text{vec}(\xi) \sim N(0, \Sigma \otimes \mathbf{I}_k)$. 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).

Sampling of $\boldsymbol{\Sigma}$

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 (36). 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.

3.1 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

$$\text{BF}_i = \frac{\int p(\zeta)\ell(\text{Data}|\zeta)d\zeta}{\int p_0(\zeta_0)\ell_0(\text{Data}|\zeta_0)d\zeta_0}, \tag{37}$$

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 (37) is equal to

$$\text{BF}_i = \frac{p(\alpha_{i1}|\text{Data})|_{\alpha_{i1}=0}}{p(\alpha_{i1})|_{\alpha_{i1}=0}}, \tag{38}$$

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 height of the marginal posterior density of α_{i1} can easily be computed using the draws from the MCMC sampler, see Gelfand and Smith (1990).

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^2_i.

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. These variables were also used in Fok *et al.* (2005), and details can be found in that paper.

– Insert Table 1 about here –

We now turn to the estimation results. We mainly impose weakly informative priors, that is we set $\sigma^2_{\alpha_1} = 1$, $\upsilon = 0.15$, $\nu = 5$, $V = \mathbf{I}_3$, and $\lambda = 6$. For the thresholds we set the prior parameters such that the regimes correspond to small versus large price changes, that is we set $\mu_\tau = 0.1$, $\sigma^2_\tau = 0.025$ and $ub = 0.4$. That is 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 of the remaining draws we only use each 5th draw to obtain a reasonable random sample from the posterior distribution.

Weekly seasonality

First we comment on the seasonal component of our model. We use the Bayes Factor to compare our model to a model where $\alpha_{i1} = 0,\ i = 1, \ldots, 96$, that is a model with no seasonality. The Bayes factors are computed using the Savage-Dickey Density Ratio as described in Section 3.1. Of these 96 Bayes Factors, 15 are smaller than 1, that is, for 15 brands we prefer the model with a clear sigmoidal seasonal pattern. These 15 brands are in the product categories beer, oatmeal, crackers, canned soup, snacks, and frozen dinners. Note that this does not automatically imply that the remaining 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, in contrast to the deterministic part, this part of the description for 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. Figure 3 shows a histogram of $\mathrm{E}[\sigma_i^2|\text{Data}]/\mathrm{E}[\sigma_{\eta_i}^2|\text{Data}]$. 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.

– Insert Figure 3 about here –

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 Figure 4 we show some examples of the seasonal patterns that we find. It is clear that for some brands we find no seasonality at all, that for others we find a relatively smooth cycle, while yet for other brands the seasonal patterns correspond to just a few spikes in sales.

– Insert Figure 4 about here –

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.

In Figure 5 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 Figure 1, and graphs of the price elasticity, similar to Figure 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 .18, the threshold for large price increases (τ_1) is about .10.

– Insert Figure 5 about here –

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 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, while a large relative display frequency corresponds to a large elasticity. 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).

5 Conclusion

In this paper we have put forward a multi-level 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 multi-level 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.

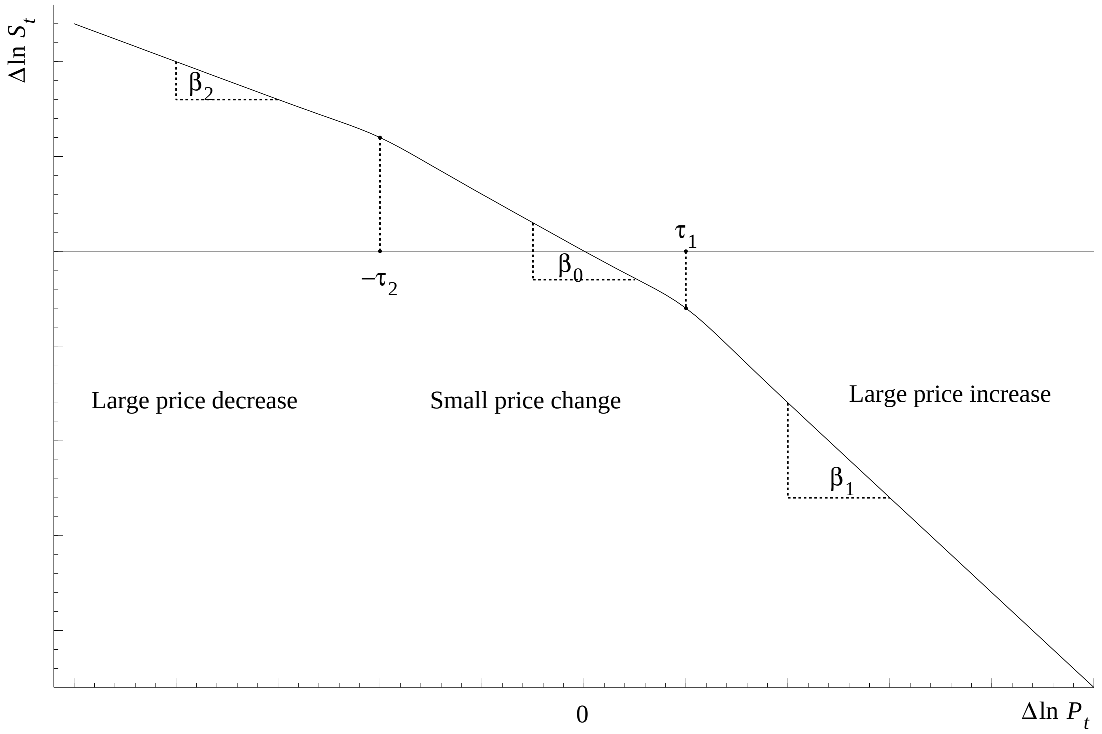

Figure 1: Example of a non-linear sales response curve

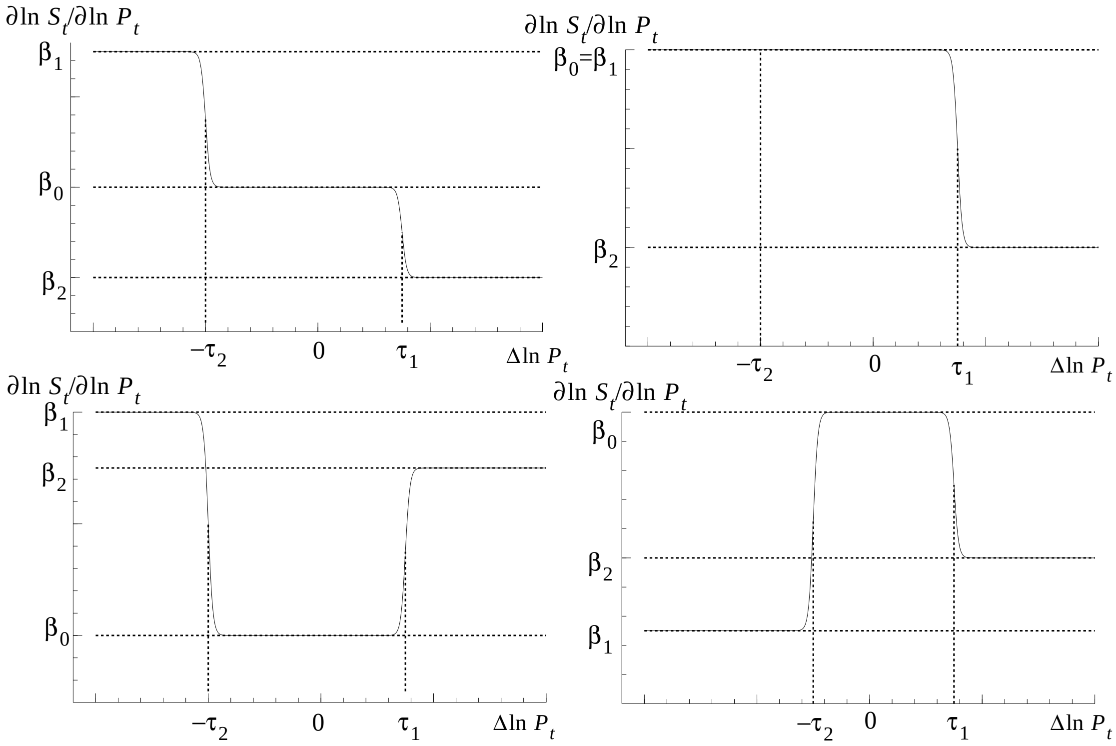

Figure 2: Four examples of the implied price elasticities.

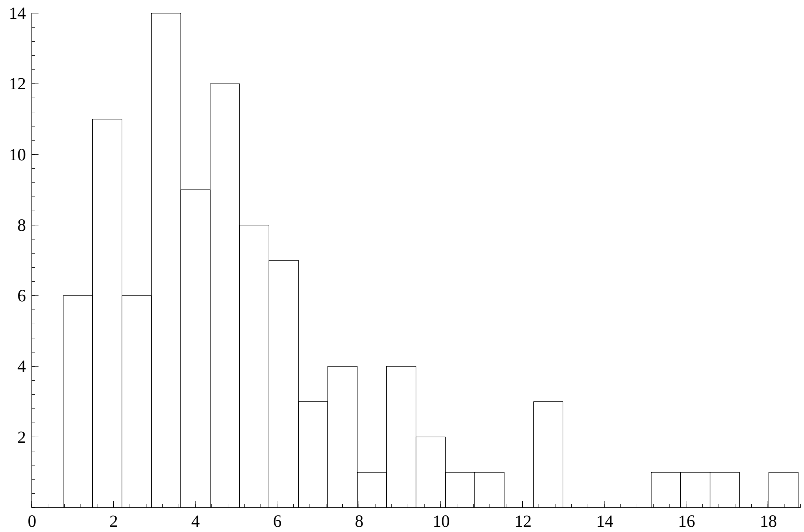

Figure 3: Histogram of posterior mean of σ_i^2 divided by posterior mean of $\sigma_{\eta i}^2$.

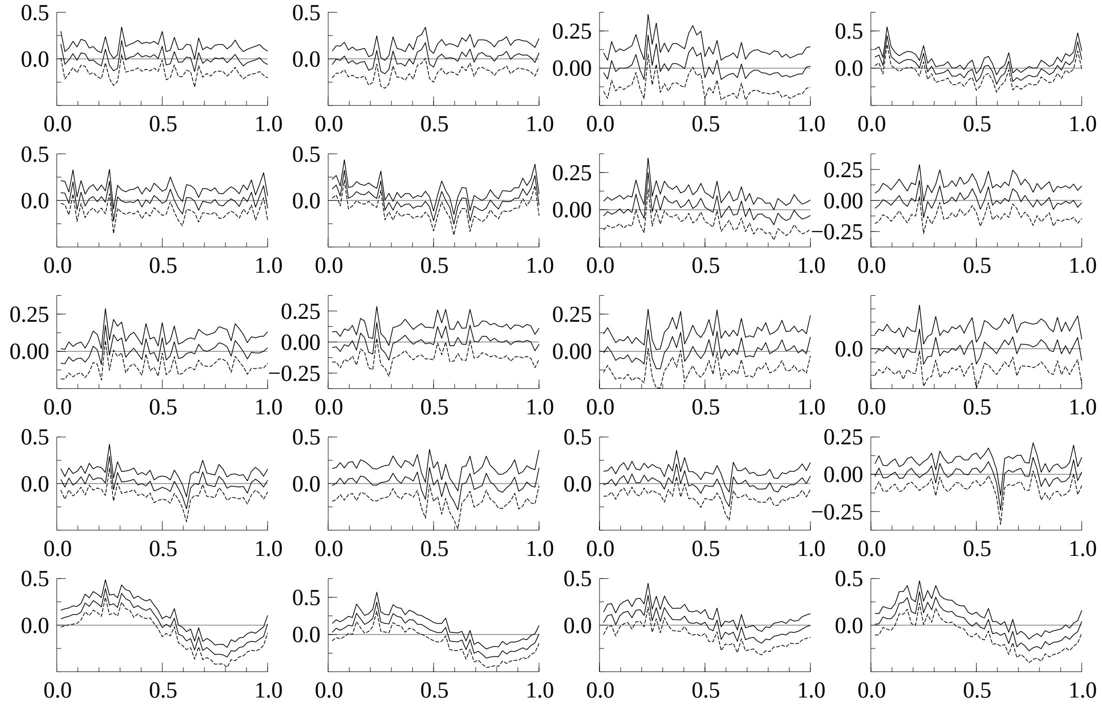

Figure 4: Some examples of the posterior mean of $\mu_{is} = \alpha_{i1}\cos(2\pi\frac{s}{S} - \alpha_{i2}) + \eta_{is}$, with 95% highest posterior density region. The horizontal axis corresponds to one year.

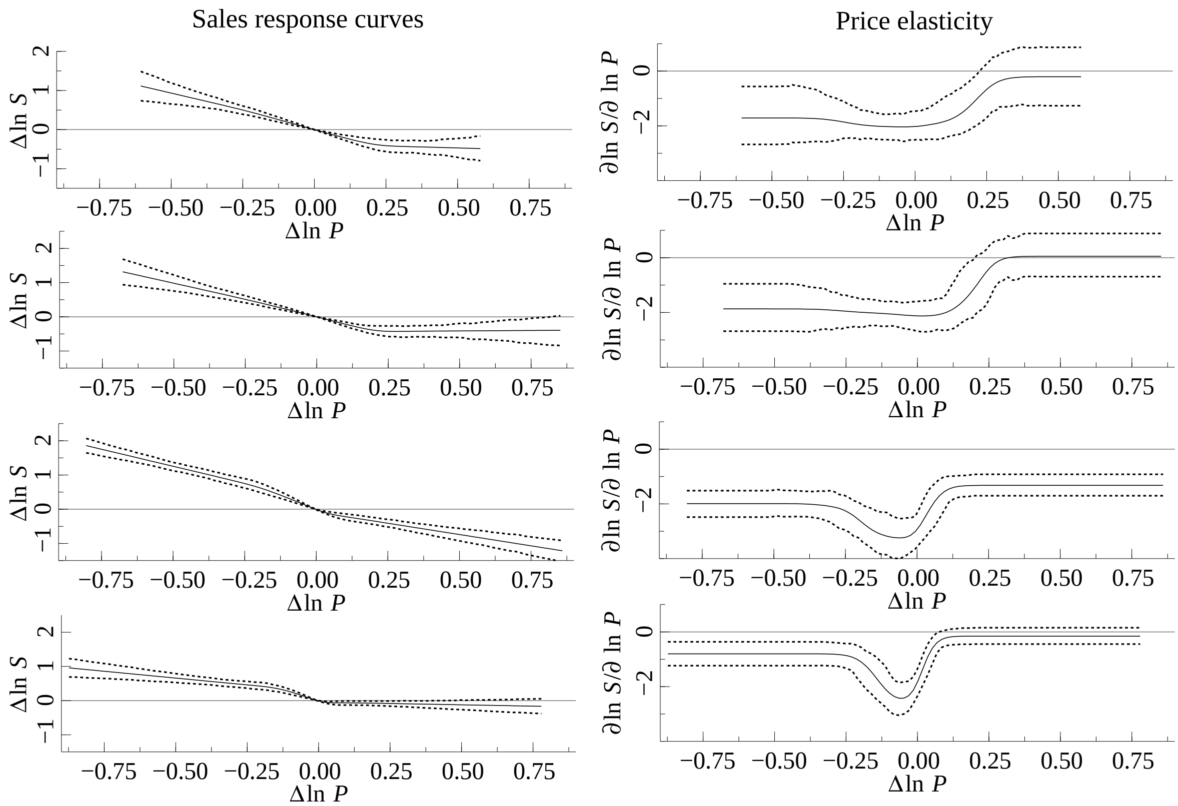

Figure 5: Posterior mean of price effect for four brands in the softdrinks category (with 95% highest posterior density region).

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

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***	(0.106)	-1.380***	(0.117)	-2.014***	(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**	(0.178)	0.549***	(0.201)
Rel. price prom. freq.	0.319**	(0.129)	-0.046	(0.143)	-0.148	(0.172)
Rel. depth price prom.	0.024	(0.109)	0.264***	(0.106)	0.150	(0.119)
Rel. feature freq.	-0.152	(0.129)	0.108	(0.137)	0.183	(0.156)
Rel. display freq.	-0.253**	(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**	(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 prom. freq	-0.238	(0.158)	-0.252	(0.160)	-0.206	(0.185)
Depth price prom.	-0.079	(0.137)	0.150	(0.155)	0.029	(0.174)
Display freq.	-0.124	(0.167)	-0.039	(0.167)	-0.021	(0.191)
Feature freq.	-0.125	(0.149)	0.328**	(0.139)	0.220	(0.158)
Hedonic	-0.043	(0.126)	0.174	(0.139)	0.292*	(0.160)

Covariance of random effects ξ

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

*,**,*** Zero not contained in 90%, 95% or 99% highest posterior density region, respectively.

References

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

Blattberg, R. C. and E. I. George (1991), Shrinkage Estimation of Price and Promotional Elasticities - Seemingly Unrelated Equations, *Journal of the American Statistical Association*, **86**, 304–315.

Dickey, J. (1971), The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters, *Annals of Mathematical Statistics*, **42**, 204–223.

Fok, D., C. Horváth, R. Paap, and P. H. Franses (2005), A Hierarchical Bayes Error Correction Model to Explain Dynamic Effects of Price Changes, *Journal of Marketing Research*, **conditionally accepted**.

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

Gelfand, A. E. and A. F. Smith (1990), Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85**, 398–409.

Geman, S. and D. Geman (1984), Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

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

Hanssens, D. M., L. J. Parsons, and R. L. Schultz (2001), *Market Response Models: Econometric and Time Series Analysis*, 2nd edn., Kluwer Academic Publishers.

Hendricks, W., R. Koenker, and D. J. Poirier (1979), Residential Demand for Electricity: An Econometric Approach, *Journal of Econometrics*, **9**, 33–57.

Hobert, J. P. and G. Casella (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. and W. M. Brelsford (1967), Time Series with Periodic Structure, *Biometrika*, **54**, 403–407.

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

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

Nijs, V. R., M. G. Dekimpe, J.-B. E. M. Steenkamp, and D. M. Hanssens (2001), The Category-Demand Effects of Price Promotions, *Marketing Science*, **20**, 1–22.

Pauwels, K. and S. Srinivasan (2004), Who Benefits from Store Brand Entry?, *Marketing Science*, **23**, 364–390.

Pauwels, K., S. Srinivasan, and P. H. Franses (2005), When do price thresholds matter in retail categories?, Invited to resubmit to Marketing Science.

Ritter, C. and M. A. Tanner (1992), Facilitating the Gibbs Sampler - The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association*, **87**, 861–868.

Smith, A. F. M. and G. O. Roberts (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., K. Pauwels, D. M. Hanssens, and M. G. Dekimpe (2004), Do Promotions Benefit Manufacturers, Retailers, or Both?, *Management Science*, **50**, 617–629.

Tanner, M. and W. Wong (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.

Verdinelli, I. and L. Wasserman (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., M. Addona, W. Hawkes, and J. Porter (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.