JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 717–744 (2000)

A DYNAMIC MULTINOMIAL PROBIT MODEL FOR BRAND CHOICE WITH DIFFERENT LONG-RUN AND SHORT-RUN EFFECTS OF MARKETING-MIX VARIABLES

RICHARD PAAP[a]* AND PHILIP HANS FRANSES[b]

[a]*Rotterdam Institute for Business Economic Studies, Erasmus University Rotterdam, PO Box 1738, NL-3000, DR Rotterdam, The Netherlands*
[b]*Econometric Institute and Department of Marketing and Organization, Erasmus University Rotterdam, The Netherlands*

SUMMARY

In this paper we propose a dynamic multinomial probit model in order to estimate the long-run and short-run effects of marketing mix variables on brand choice. The latent variables, which contain the unobserved perceived utilities, follow a first-order vector error correction autoregressive process of order 1 with current and lagged explanatory variables. The unrestricted autoregressive parameter matrix concerns the intertemporal correlation in perceived utilities of households over purchase occasions and indicates the persistence in brand choice. As explanatory variables we consider relative prices and promotional activities like feature and display. An important and novel feature of our model is that it allows for different long-run and short-run effects of promotional activities, thereby extending the models that are currently available in the literature. Additionally, to account for different base preferences for brands across households, we allow for consumer heterogeneity. Our application concerns a panel of households choosing among several brands of a FMCG. Our estimated model turns out to be an improvement over a static model and over a model with only short-run effects, in terms of in-sample fit and out-of-sample forecasts.

1. INTRODUCTION

A typical data set for marketing research purposes contains information on the actual purchases of a panel of households observed during several weeks. These purchases usually concern various brands in a variety of product categories. It is often known how much each household pays for a particular brand, what the prices are of other brands available at that purchase occasion, and whether certain brands are on promotion (for example, featured or displayed). For a market researcher it is commonly of interest to examine if marketing-mix variables have an impact on brand choice. An example issue concerns the effects of promotion on brand-switching behaviour. For managerial decision making it is additionally important to distinguish long-run effects from short-run effects. Indeed, it would be helpful to the marketing manager if one would have a model that can be used for evaluating various scenarios, the outcomes of which may lead to designing more effective marketing strategies. In this paper, we propose such an econometric model for the above described generic data set, which can be usefully considered for these purposes.

The dynamic impact of marketing-mix variables on marketing performance measures, and the long-run impact in particular, has received considerable attention in the marketing literature

*Correspondence to: R. Paap, Erasmus University Rotterdam, RIBES, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: paap@few.eur.nl

Received 30 September 1999
Revised 5 September 2000

recently. Blattberg and Neslin (1989) and Dekimpe and Hanssens (1995) study the long-run effect of marketing effects on sales, Mela *et al.* (1998) and Erdem (1996) consider this effect on market structure, and Jedidi *et al.* (1999) focus on profitability, among others. There are also a number of studies on the dynamic effects of the marketing mix on brand choice; see Mela *et al.* (1997), Papatla and Krishnamurthi (1996) and Erdem and Keane (1996), among others.

If the marketing measure concerns sales or market shares, one usually relies on the (multivariate) time series model to capture the dynamic structure; see, for example, Dekimpe and Hanssens (1995) and Bronnenberg *et al.* (2000). If the performance measure concerns brand choice, it becomes less evident how one can incorporate dynamics. One may consider the inclusion of a variable which somehow measures loyalty, based on past purchase behaviour; see Guadagni and Little (1983). It is also possible to allow the parameters to be time-varying; see Papatla and Krishnamurthi (1996). Finally, one can explicitly incorporate a dynamic structure into a multinomial brand choice model; see, for example, Erdem and Keane (1996). A common property of the modelling approaches in the aforementioned studies is that no explicit distinction between long-term and short-term effects has been made, at least not within the context of a single model. It is the purpose of the present paper to propose such a model where this distinction can be made.

The model we put forward is a dynamic multinomial probit model which incorporates dynamics and unobserved household heterogeneity, and where we do not impose parameter restrictions on dynamic parameters. Most importantly, in order to allow for different long-run and short-run effects of marketing-mix variables, we describe the vector of unobserved utilities (concerning the brands) as a vector error-correction mechanism (VECM). Indeed, it is quite possible that promotions have long-run effects which may differ in size (and perhaps even in sign) from short-run effects. Our model extends the models that are currently available in the literature, which are, for example, a static model and a model with identical long-run and short-run effects, as these are both nested within the VECM model. We compare these three different models on their in-sample and out-of-sample forecasting performance.

The outline of our paper is as follows. In Section 2, we propose our model, and compare it with several of its restricted variants. In Section 3, we elaborate on the interpretation of the model in terms of long-run and short-run effects of marketing-mix variables on brand choice. In Section 4, we discuss parameter estimation and a method to generate out-of-sample forecasts. These forecasts can be used to evaluate our model with its nested variants. In Section 5, we apply our model to an illustrative data set, and we find that our model yields useful insights. Finally, we conclude our paper in Section 6 with some comments.

2. A DYNAMIC MULTINOMIAL PROBIT MODEL

In this section we put forward a dynamic multinomial probit model which is useful for the purposes outlined in the introduction. We begin with a basic model and subsequently introduce additional features. In Section 2.1 we discuss several aspects of the model including the novel dynamic specification, while in Section 2.2 we discuss parameter identification.

2.1 The Model

Assume that a household i perceives utility U_{ijt} if it buys brand j on purchase occasion t, and that this utility obeys

$$U_{ijt} = X'_{ijt}(\beta + \beta_i) + \varepsilon_{ijt} \tag{1}$$

where X_{ijt} is a k-dimensional vector of explanatory variables, β is a k-dimensional parameter vector, and ε_{ijt} is an unobserved random variable, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $t = 1, \ldots, T_i$. The vector X_{ijt} can contain a brand-specific intercept, the price of product j, and, for example, 0/1 dummy variables to indicate whether product j is on feature or display at purchase occasion t. The unknown parameter vector β describes the effect of explanatory variables X_{ijt} on the utilities and β_i the household-specific effect.

Furthermore, assume that household i chooses brand j on purchase occasion t if the perceived utility of brand j exceeds the perceived utilities of all the other brands in the same product category, that is, if

$$U_{ijt} > U_{imt} \quad \text{for} \quad m = 1, \ldots, j-1, j+1, \ldots, J \tag{2}$$

As the random terms ε_{ijt} are not observed, we do not observe the perceived utilities U_{ijt}. In fact, we observe only the actual purchase. Define the variable d_{it} with $d_{it} = j$ if household i buys brand j at purchase occasion t. Given equation (2), the probability $\Pr[d_{it} = j]$ is equal to

$$\Pr[d_{it} = j] = \Pr[U_{ijt} > U_{i1t}, \ldots, U_{ijt} > U_{i,j-1,t}, U_{ijt} > U_{i,j+1,t}, \ldots, U_{ijt} > U_{iJt}] \tag{3}$$

This probability depends on the assumptions about the distribution of ε_{ijt}. If we take for ε_{ijt} an independent, uncorrelated type-1 extreme value distribution, we obtain the conditional logit model of McFadden (1973). This conditional logit model assumes independence of irrelevant alternatives (IIA); see, for example, McFadden (1984) and Cramer (1991) for a discussion. To avoid the IIA property, one can, for example, assume that the J-dimensional disturbance vector $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt})'$ is normally distributed, that is,

$$\varepsilon_{it} \sim \text{NID}(\mathbf{0}, \Sigma) \quad \text{for all } i \tag{4}$$

where Σ is a $(J \times J)$ covariance matrix. This leads to a multinomial probit (MNP) model; see Hausman and Wise (1978) and Daganzo (1979), among others.

As brand choice is fully determined by utility differences (see equation (2)), it is conventional to measure utility *relative* to the alternative J to identify the model parameters; see, for example, Bunch (1991). Thus, we can define

$$\begin{aligned} U_{ijt} - U_{iJt} &= X'_{ijt}(\beta + \beta_i) - X'_{iJt}(\beta + \beta_i) + \varepsilon_{ijt} - \varepsilon_{iJt} \\ \tilde{U}_{ijt} &= \tilde{X}'_{ijt}(\beta + \beta_i) + \tilde{\varepsilon}_{ijt} \qquad \text{for } j = 1, \ldots, J-1 \end{aligned} \tag{5}$$

where $\tilde{U}_{ijt}$ denotes the perceived *relative* utility of brand j, that is, *relative* to brand J. Hence, household i chooses brand j if $\tilde{U}_{ijt}$ is the maximum of the relative utilities unless all relative utilities are smaller than zero, which corresponds to choosing brand J. Of course, the relative utility of brand J is zero.

If we define $\tilde{U}_{it} = (\tilde{U}_{i1t}, \ldots, \tilde{U}_{i,J-1,t})'$ and $\tilde{X}_{it} = (\tilde{X}_{i1t}, \ldots, \tilde{X}_{i,J-1,t})'$ we can write the MNP model in equation (5) as

$$\tilde{U}_{it} = \tilde{X}_{it}\beta + \tilde{\varepsilon}_{it} \tag{6}$$

$i = 1, \ldots, I$ and $t = 1, \ldots, T_i$. For the vector of stacked $\tilde{\varepsilon}_{ijt}$, denoted by $\tilde{\varepsilon}_{it}$, it now holds that

$$\tilde{\varepsilon}_{it} \sim \text{NID}(\mathbf{0}, \tilde{\Sigma}), \tag{7}$$

with $\tilde{\Sigma} = M\Sigma M'$ where $M = (\mathbf{I}_{J-1} | -\mathbf{i})$ with $\mathbf{I}_{J-1}$ a $(J-1)$-dimensional identity matrix and $\mathbf{i}$ a $(J-1)$-dimensional vector of ones.

Dynamic specification

There are various ways to extend equation (6) in order to allow for dynamics. A rather general specification (see also Hendry *et al.*, 1984), and which is also useful for our purposes, is to include lagged utilities and lagged explanatory variables. Assuming first-order dynamics, the model would then become

$$\tilde{U}_{it} = \Pi \tilde{U}_{i,t-1} + \tilde{X}_{it}(\gamma + \gamma_i) + \tilde{X}_{i,t-1}(\delta + \delta_i) + \eta_{it} \tag{8}$$

where Π is a $((J-1) \times (J-1))$ matrix and where $\eta_{it} \sim \text{NID}(\mathbf{0}, \tilde{\Sigma})$.To ensure stationarity of the relative utilities, we impose that the eigenvalues of Π are within the unit circle. The possibility of eigenvalues of 1 allows for cointegration analysis, but we leave this topic for further research.

In order to analyse possibly different long-run and short-run effects, it is well known to be most convenient to write equation (8) in a so-called vector error-correction format (VECM), that is, to write it as

$$\Delta\tilde{U}_{it} = \Delta\tilde{X}_{it}(\alpha + \alpha_i) + (\Pi - \mathbf{I}_{J-1})\big(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta + \beta_i)\big) + \eta_{it} \tag{9}$$

where Δ is the usual first difference operator. Note that equation (9) does not impose any restrictions on the parameters in equation (8) in the sense that the number of parameters in both models are the same. In the VECM format, the parameters $\alpha + \alpha_i$ measure the short-run effects of $\tilde{X}_{it}$ on $\tilde{U}_{it}$, and the $\beta + \beta_i$ parameters concern the long-run effects. The parameters in $(\Pi - \mathbf{I}_{J-1})$ measure the speed at which deviations from the long-run relation between $\tilde{U}_{it}$ and $\tilde{X}_{it}$ become adjusted. A vector error-correction model is frequently applied in modern time-series econometrics and in this paper we introduce it for modelling dynamic brand choice.

In the marketing literature one can find the application of two restricted versions of equation (9). The first is the model considered in Erdem (1996) and Erdem and Keane (1996), where it assumed that $\delta + \delta_i = 0$ in equation (8) and hence only lagged utilities are included. It is not difficult to derive that this restriction imposes the nonlinear restriction $\alpha + \alpha_i = -(\Pi - \mathbf{I}_{J-1})(\beta + \beta_i)$ on the parameters in equation (9). Hence this model imposes a particular link between the long-run and short-run effects of $\tilde{X}_{it}$.

Perhaps a more plausible restriction on the parameters in equation (9) would be to assume that the short-run and long-run effects are the same. In that case the model becomes

$$\Delta\tilde{U}_{it} = \Delta\tilde{X}_{it}(\beta + \beta_i) + (\Pi - \mathbf{I}_{J-1})\big(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta + \beta_i)\big) + \eta_{it} \tag{10}$$

which effectively amounts to equation (6) with

$$\tilde{\varepsilon}_{it} = \Pi\tilde{\varepsilon}_{i,t-1} + \eta_{it} \tag{11}$$

As the first-order autoregressive structure in equation (11) operates on both $\tilde{U}_{it}$ and $\tilde{X}_{it}$, the model in equation (10) is usually called a common factor dynamic model; see Hendry *et al.* (1984). McCulloch and Rossi (1994), Geweke *et al.* (1997) and Allenby and Lenk (1994) consider such a dynamic structure, where they additionally assume that Π is a diagonal matrix. Needless to say, when the short-run and long-run effects are not the same, this common factor model is misspecified.

In the present paper we will continue to focus on the VECM-MNP model in equation (9) and additionally, we do not restrict Π to be diagonal. In our empirical work below, we will compare our model with the static model in equation (6) and the common factor model in equation (10), where again we do not restrict Π.

Household heterogeneity
The household-specific parameters α_i and β_i allow the effects of the covariates $\tilde{X}_{ijt}$ on the perceived utilities to be different for each household i. Hence, we allow that some households may be more sensitive to promotional activities than others and that some households have more base preference for a specific brand than others. Neglecting such household heterogeneity can be shown to lead to overestimation of the persistence in brand choice, which in our case would be reflected by higher-valued diagonal elements of Π; see also Keane (1997) for a discussion.

In practice it can happen that a household never buys a certain brand. This implies that it is then not possible to estimate the brand-specific intercept for this particular brand; see Rossi and Allenby (1993) for a discussion. Therefore, one often assumes that β_i, and in our case also α_i, are draws from a population distribution. See also Chintagunta *et al.* (1991), Gönül and Srinivasan (1993) and Jain *et al.* (1994), who follow a similar procedure. One reasonable possibility is to assume that

$$\beta_i \sim N(\mathbf{0}, \Sigma_\beta) \quad \text{and} \quad \alpha_i \sim N(\mathbf{0}, \Sigma_\alpha) \tag{12}$$

where Σ_β and Σ_α are $(k \times k)$ covariance matrices. See also McCulloch and Rossi (1994) and Allenby and Rossi (1999) for a similar approach.

2.2 Identification Issues

Before we continue to analyse the properties of the MNP models in equations (6) and (10) and our VECM-MNP model in equation (9) in detail, we have to discuss a few issues concerning parameter identification. In fact, not all parameters of the unrestricted MNP models are identified.

First, we consider the identification issue arising from the fact that we can identify only utilities differences. It follows from the definition of $\tilde{\Sigma}$ below equation (7) that at most $\frac{1}{2}J(J-1)$ parameters of the covariance matrix Σ are identified, where Σ may contain $\frac{1}{2}J(J+1)$ different parameters. Furthermore, the base brand normalization also implies that only $(J-1)$ brand-specific intercepts in β are identified. As a result we can specify only unobserved household heterogeneity via Σ_β on $(J-1)$ brand-specific intercepts and on the effect of any other covariates.

The second identification problem is due to the presence of time-invariant covariates. It follows from equation (8) that time-invariant covariates in $\tilde{X}_{i,t}$, like brand-specific intercepts, imply perfect collinearity. To solve this identification problem we put the α parameters in equation (9), which correspond to the time-invariant covariates, equal to zero. Additionally, we allow only for unobserved household heterogeneity for the non-zero elements in α, which implies that the dimension of Σ_α equals the number of non-zero elements in α.

Finally, we need to impose yet another identification restriction. If we multiply each relative utility $\tilde{U}_{ijt}$ with the same constant s, households would still make the same decision. For the VECM model, the parameters β, α, Σ_β, Σ_α, Π, $\tilde{\Sigma}$ then change to $s\beta$, $s\alpha$, $s^2\Sigma_\beta$, $s^2\Sigma_\alpha$, Π, $s^2\tilde{\Sigma}$. It is easy to see that the brand choice probabilities then also do not change. To identify the model parameters, we can restrict one of the elements of β (or α). This implies that we fix the sign of this single parameter. However, a more convenient approach is to restrict one of the elements of $\tilde{\Sigma}$. The conventional normalization is to impose that $\tilde{\Sigma}_{J-1,J-1}=1$. In this paper, however, we follow a Bayesian approach and handle this scaling problem by imposing a proper prior on $\tilde{\Sigma}$ followed by a renormalization as in McCulloch and Rossi (1994); see also Section 4.1.

3. MODEL INTERPRETATION

In general, a MNP model assumes that a household bases its brand choice on the perceived utilities for each brand. In the static MNP model (6), these utilities depend only on the current covariates $\tilde{X}_{it}$ which may, for instance, contain information on price and promotions. This model implies that households instantaneously react to a promotion denoted by a change in $\tilde{X}_{it}$. If this promotion is held, for example, only on the current purchase occasion t, it is assumed to have no effect on future brand choice decisions of households. Hence, only permanent changes in the covariates affect future brand choice decisions.

To analyse the long-run and short-run properties for the dynamic MNP models, we solve equation (9) for the current relative utilities and obtain

$$\tilde{U}_{it}=\Pi^{t}\tilde{U}_{0t}+\sum_{h=0}^{t}\Pi^{h}\big(\Delta\tilde{X}_{i,t-h}(\alpha+\alpha_i)+(\mathbf{I}_{J-1}-\Pi)\tilde{X}_{i,t-1-h}(\beta+\beta_i)+\eta_{i,t-h}\big) \tag{13}$$

The current relative utilities are seen to be a function of current and past covariates and of unobserved error terms. As the eigenvalues of Π are within the unit circle, we have that $\Pi^{t}\approx\mathbf{0}$ for large t, and hence the influence of past perceived relative utilities on the current brand choice dies out as time proceeds. The speed of this process depends on the eigenvalues of Π.

As the expectation of $\eta_{i,t-h}\ \forall h$ is zero, it follows from equation (13) that the expected value of $\tilde{U}_{it}$ given current and past covariates equals

$$\mathrm{E}\big[\tilde{U}_{it}|\tilde{X}_{i,t-h},h=1,\ldots,t,\beta_i,\alpha_i\big]=\sum_{h=0}^{t}\Pi^{h}\big(\Delta\tilde{X}_{i,t-h}(\alpha+\alpha_i)+(\mathbf{I}_{J-1}-\Pi)\tilde{X}_{i,t-1-h}(\beta+\beta_i)\big) \tag{14}$$

For fixed values of the explanatory variables over time, that is, $\tilde{X}_{i,t}=\tilde{X}_i$ for all t, this expectation simplifies to

$$\mathrm{E}\big[\tilde{U}_{it}|\beta_i,\tilde{X}_i\big]=\tilde{X}_i(\beta+\beta_i) \tag{15}$$

The unconditional variance of $\tilde{U}_{it}$ is given by

$$\mathrm{V}\big[\tilde{U}_{it}|\beta_i\big]=\sum_{h=0}^{\infty}\Pi^{h}\tilde{\Sigma}(\Pi^{h})'=V \tag{16}$$

See also Lütkepohl (1993). This implies that the process $\tilde{U}_{it}$ for fixed $\tilde{X}_{it}$ converges in the long run to

$$\tilde{U}^{*}_{it}=\tilde{X}_i(\beta+\beta_i)+\zeta_{it} \tag{17}$$

with $\zeta_{it}\sim\mathrm{NID}(\mathbf{0},V)$.

The relative utilities of each household i are draws from the distribution of $\tilde{U}^*_{it}$. If this long-run distribution is reached, the mean and variance of $\tilde{U}_{it}$ stay the same on every future purchase occasion as long as $\tilde{X}_{it}$ does not change. It is even true that the distribution of $\tilde{U}_{it}$ does not change, which implies that the probabilities $\Pr[U^*_{ijt} > U^*_{imt}$ for all $m \neq j]$ stay the same for large t as long as $\tilde{X}_{it}$ does not change. Note that this does not imply that individual households buy the same brand at each future purchase occasion or that the brand choice probabilities of individual households are the same over time. This is established because the brand choice probabilities of an individual household depend on the perceived utilities at the previous purchase occasion. Hence, only the distribution of the perceived relative utilities over the households remains the same.

As equation (17) describes the long-run values of the relative utilities for an arbitrary household i given $\tilde{X}_{it}$, we may interpret the probability

$$\int_{\beta_i} \Pr\left[U^*_{ijt} > U^*_{imt} \text{ for all } m \neq j\right] f_N(\beta_i|\mathbf{0}, \Sigma_\beta)\, \mathrm{d}\beta_i \tag{18}$$

with $f_N(\cdot|\mu,\Sigma)$ the p.d.f. of a multivariate normal distribution with mean μ and covariance matrix Σ, as the long-run equilibrium market share of brand j given $\tilde{X}_{it}$. Note that we have to integrate with respect to β_i to remove household heterogeneity. To compute this long-run market share, which can be useful for managerial purposes, we can use a crude frequency simulator or the GHK simulator; see Börsch-Supan and Hajivassiliou (1993) and Geweke *et al.* (1997).

We have to be careful to relate choice probabilities directly to market shares. The model describes brand choice conditional on the purchase of a product. To construct market share we may also need to consider the timing of purchase and the number of items purchased at the same time. Gupta (1988) considers a joint model for brand choice, interpurchase times and amount of items purchased, but his model does not allow for cross-equation correlations and it is static. We leave the construction of a joint dynamic model as an important topic for further research and in the present paper we translate choice probabilities directly to market shares.

Equation (18) gives the long-run market shares as a function of the explanatory variables $\tilde{X}_{it}$. As the expression that determines the long-run properties of the dynamic models (17) does not involve α, it is the same for the common factor dynamic MNP model and the VECM-MNP model. Hence, the common factor MNP model and the VECM-MNP model have the same long-run behaviour for fixed $\tilde{X}_{it}$. The difference between the two models lies in the short-run dynamics. From equation (13) it follows that a change in $\tilde{X}_{it}$ due to promotional activities (price reduction, features) leads to different market shares in the immediate short run as it changes the conditional mean of future $\tilde{U}_{it}$ values for the VECM-MNP specification. If the change in $\tilde{X}_{it}$ is only temporary, the system will converge to the same market share equilibrium after a while, while if the change is permanent, for example, a permanent price reduction, the system will converge to a new market share equilibrium. Due to the specific structure of the model, the speed and decay patterns of the relative utilities to the equilibrium do not depend on which of the k covariates in $\tilde{X}_{it}$ changes, as can be seen from equation (13). This will probably also be the case for the choice probabilities, which are non-linear functions of the relative utilities. This phenomenon is inherent to the model and it also applies to the other autoregressive models discussed in the previous subsection as these are all nested within our VECM-MNP model. Finally, note that an equilibrium, where market shares become constant, may never be attained as $\tilde{X}_{it}$ may change frequently over time, for example due to repeated promotional activities.

If the short-run and long-run effects of marketing variables are the same, equation (13) simplifies to

$$\tilde{U}_{it} = \Pi^t \tilde{U}_{0t} + \tilde{X}_{it}(\beta + \beta_i) + \sum_{h=0}^{t} \Pi^h \eta_{i,t-h} \tag{19}$$

and hence current relative utilities do not depend on the values of lagged covariates. This means, as for the static model (6), that also for this model a temporary promotional activity (reflected by a temporary change in $\tilde{X}_{it}$) has no impact on future utility levels. Of course, permanent changes in $\tilde{X}_{it}$ do have a long-run impact.

4. ESTIMATION AND INFERENCE

In this section we discuss estimation and forecasting with the three MNP models discussed so far. As our vector error-correction MNP model in equation (9) nests the static and the common factor dynamic model, we will focus in this section on this most general model. Results for the other models can be obtained in a similar way. In Section 4.1 we consider parameter estimation and Section 4.2 deals with forecasting.

4.1 Parameter Estimation

Let $d_i = (d_{it}, \ldots, d_{iT_i})'$ denote the vector of brand choices of household i on the T_i purchase occasions. Let j_{it} denote the choice of household i on purchase occasion t, $t = 1, \ldots, T_i$. The probability that household i chooses for d_i is now given by

$$\Pr\left[d_i|\theta_i\right] = \Pr\left[U_{i,j_{it},t} > U_{imt} \text{ for all } m \neq j_{it}, t = 1, \ldots, T_i\right] \tag{20}$$

To compute this probability, we use that $\tilde{U}_{it}$ given $\tilde{U}_{i,t-1}$ is normally distributed (see equation (9)). Hence, the joint distribution of $(\tilde{U}_{it},\ j = 1, \ldots, J-1, t = 1, \ldots, T_i)$ is a $(T_i \times (J-1))$ multivariate normal distribution, with mean and covariance matrix that are functions of the unknown parameters $\theta_i = (\beta, \alpha, \tilde{\Sigma}, \beta_i, \alpha_i, \Sigma_\beta, \Sigma_\alpha, \Pi)$ in the model.[1] Therefore, the computation of equation (20) requires the evaluation of a $(T_i \times (J-1))$-dimensional integral. The likelihood is given by the product of the probabilities in equation (20) for the households $i = 1, \ldots, I$ as

$$\ell(\theta|\text{data}) = \prod_{i=1}^{I} \int_{\alpha_i} \int_{\beta_i} \Pr[d_i|\theta_i] f_N(\beta_i|\mathbf{0}, \Sigma_\beta) f_N(\alpha_i|\mathbf{0}, \Sigma_\alpha)\, \mathrm{d}\beta_i\, \mathrm{d}\alpha_i \tag{21}$$

where $\theta = (\beta, \alpha, \tilde{\Sigma}, \Sigma_\beta, \Sigma_\alpha, \Pi)$ and $f_N(\cdot)$ is defined below equation (18). Again, we integrate over β_i and α_i to take care of household heterogeneity.

The evaluation of the likelihood function (21) requires I times the computation of probability $\Pr[d_i|\theta_i]$ and hence we have to compute I times a $(T_i \times (J-1))$-dimensional integral. Furthermore, we also have to integrate over β_i and α_i. It is easy to understand that analytical

[1] We model the first observation by $\tilde{U}_{i1} = \tilde{X}_{i1}(\beta + \beta_i) + \eta_{i1}$ with $\eta_{i1} \sim \mathrm{N}(\mathbf{0}, \tilde{\Sigma})$. We do not use V in equation (16) instead of $\tilde{\Sigma}$ as it leads to full conditional posterior distributions for $\tilde{\Sigma}$ and Π of an unknown type and hence to computational difficulties in obtaining posterior distributions using the Gibbs sampler.

and even numerical integration methods are intractable. Therefore, we have to use simulation techniques to evaluate the likelihood.

Geweke *et al.* (1997) compare different methods to obtain parameter estimates for a multinomial multiperiod probit model, including simulated maximum likelihood, the method of simulated moments and a Bayesian approach with Gibbs sampling and data augmentation as in McCulloch and Rossi (1994). Here, we opt for the latter approach, as it produces slightly better estimates for a probit model with dynamics. Like Geweke *et al.* (1997), we use posterior means and variances of the Bayesian analysis as estimates for the model parameters.

A standard Bayesian analysis of the MNP model using Gibbs sampling is complicated due to the identifying restriction on the $\tilde{\Sigma}$ matrix. To solve this identification problem, we analyse the model without the restriction on $\tilde{\Sigma}$ and use a proper but weakly informative prior on $\tilde{\Sigma}$ as an identification tool. In this paper we take for $\tilde{\Sigma}$ an inverted Wishart prior or

$$p(\tilde{\Sigma}^{-1}) \sim \mathrm{W}(S, \lambda) \tag{22}$$

This induces proper posterior distributions for the remaining model parameters, for which we take the uninformative priors

$$\begin{aligned} p(\beta, \alpha) &\propto 1 \\ p(\Sigma_\beta) &\propto |\Sigma_\beta|^{-\frac{1}{2}J} \\ p(\Sigma_\alpha) &\propto |\Sigma_\alpha|^{-\frac{1}{2}J} \end{aligned} \tag{23}$$

These priors results from a Wishart density on Σ_β^{-1} and Σ_α^{-1} with the degrees of freedom approaching zero; see Geisser (1965). In the application below we discuss the sensitivity of our results with respect to the prior specification for the Σ_β and Σ_α parameter matrices. The posteriors of the parameters of interest are now given by β/v, α/v, $\tilde{\Sigma}/v^2$, Σ_β/v^2, Σ_α/v^2 and Π with $v = \sqrt{\tilde{\Sigma}_{J-1,J-1}}$; see McCulloch and Rossi (1994) for a justification of this approach.

To obtain posterior means we use the Gibbs sampling technique of Geman and Geman (1984) with data augmentation; see Tanner and Wong (1987) and Albert and Chib (1993). The idea of Gibbs sampling is to sample iteratively from the full conditional posterior distributions of the model parameters θ. This creates a Markov chain which converges under mild conditions such that the draws can be used as draws from the joint distribution; see, for example, Tierney (1994). The unobserved utilities $\tilde{U}_{ijt}$ and the β_i and α_i parameters are sampled alongside the other model parameters. The posterior means and standard deviations of the parameters of interest can be obtained by computing the sample mean and variance of the normalized sampled parameters. In the Appendix we give a brief outline of the derivation of the full conditional posterior distributions. It amounts to quite a general approach for deriving full conditional posterior distributions. A more detailed derivation for models of type (10) can be found in Geweke *et al.* (1997) and McCulloch and Rossi (1994) who show how to deal with household heterogeneity.

4.2 Forecasting

To enable a comparison of the three MNP models it seems useful to perform an out-of-sample forecasting exercise. Given our dynamic model specification, it is interesting to forecast the brand choices of households on the next purchase occasion. From a market research point of view, it can also be interesting to forecast brand choice patterns for new customers.

We consider forecasting the next brand choice of households which are within the estimation sample. As we model the discrete variable brand choice with continuous unobserved utility variables, forecasting with a MNP model results in predictive probabilities that a household chooses brand j for $j = 1, \ldots, J$. We know that the probability that household i purchases brand j on purchase occasion $T_i + 1$ equals

$$\Pr\left[d_{i,T_i+1} = j|\theta_i, U_{iT_i}\right] = \Pr\left[U_{ij,T_i+1} > U_{im,T_i+1} \text{ for all } m \neq j\right] \tag{24}$$

From equation (9) it is easy to see that this probability depends on the parameter θ_i and the perceived utilities at purchase occasion T_i. The predictive probability that d_{i,T_i+1} is j equals

$$\Pr\left[d_{i,T_i+1} = j\right] = \int_{\tilde{U}_{iT_i}} \int_{\theta_i} \Pr\left[d_{i,T_i+1} = j|\theta_i, \tilde{U}_{iT_i}\right] p(\tilde{U}_{iT_i}|\theta_i) p(\theta_i|\text{data})\, \mathrm{d}\theta_i\, \mathrm{d}\tilde{U}_{iT_i} \tag{25}$$

where $p(\tilde{U}_{iT_i}|\theta_i)$ and $p(\theta_i|\text{data})$ denote the full conditional posterior density of $\tilde{U}_{iT_i}$ and the posterior density of θ_i, respectively. To compute this integral, we extend the Gibbs sampler with an extra step. Given a draw for θ_i and $\tilde{U}_i$, we stimulate $\tilde{U}_{ij,T_i+1}$ $j = 1, \ldots, J-1$ using equation (9). The relative number of cases in the Gibbs run for which the sampled relative utilities $\tilde{U}_{i,T_i+1}$ indicate that brand j, $j = 1, \ldots, J$, is chosen, provides us the predictive density for d_{i,T_i+1}.

5. AN ILLUSTRATION

In this section we examine the empirical usefulness of our VECM-MNP model in equation (9) for analysing the impact of marketing-mix variables on long-run and short-run brand choice (or market shares). In Section 5.1 we discuss the data. In Sections 5.2 and 5.3 we present estimation and forecasting results. Finally, the last two subsections deal with the long-run and short-run effects of promotional activities.

5.1 The Data

We consider an optical scanner panel data set on purchases of saltine crackers in the Rome (Georgia) market, collected by Information Resources Incorporated. The data set contains information on all purchases of crackers (3292) of 136 households over a period of two years, including brand choice, actual price of the purchased brand and shelf price of other brands, and whether there was a display and/or newspaper feature of the considered brands at the time of purchase. A subset of this data set was analysed in Jain *et al.* (1994) with a random-coefficients logit brand-choice model.

Table I shows some data characteristics. There are three major national brands in our database, that is, Sunshine, Keebler and Nabisco with market shares of 7%, 7% and 54%, respectively. The local brands are collected under 'Private label', which has a market share of 32%; see the first row of Table I. 'Display' refers to the fraction of purchase occasions that a brand was on display and 'feature' refers to the fraction that a brand was featured. The market leader, Nabisco, is relatively often on display (34%) and featured (9%). The 'average price' denotes the mean of the price of a brand over the 3292 purchases. The Keebler crackers seem to be the most expensive crackers in our data set. The final panel in Table I provides information of the number of brand switches in the sample. For example, in 39% of the cases households buying Sunshine on the current purchase occasion buy the same brand on the next purchase

Table I. Some data characteristics of the saltine crackers

		Sunshine	Keebler	Nabisco	Private label
Market share		0.07	0.07	0.54	0.32
Display[a]		0.13	0.11	0.34	0.10
Feature[b]		0.04	0.04	0.09	0.05
Average price		0.96	1.13	1.08	0.68
Estimated	S	0.39	0.07	0.35	0.19
Brand	K	0.09	0.50	0.30	0.11
Switching	N	0.04	0.04	0.84	0.08
Probabilities[c]	P	0.04	0.03	0.12	0.81

[a] Fraction of purchase occasions on which each brand is on display.
[b] Fraction of purchase occasions on which each brand is featured.
[c] Brand switching transition probability matrix. The (i,j)-th element of this matrix shows the relative frequency that a household buys brand i at the current purchase occasion and brand j on the next purchase occasion.

occasion, while in 35% of the cases they switch to Nabisco. If a household buys Nabisco or Private label, it is more likely to by the same brand on the next purchase than if it buys Sunshine or Keebler. Note that we cannot directly interpret these results as brand loyalty as they depend on the marketing-mix variables. Nabisco, for instance, is featured and displayed more than the three other brands.

5.2 Estimation Results

To analyse this saltine cracker data set we estimate the three versions of the MNP model, that is, the static model in (6), the common factor model in (10) and our vector error-correction model in (9). In all three models we allow for household heterogeneity and we set Private label as the base brand. We use the last purchase of each household for forecasting and model evaluation, that is, this evaluation sample thus contains 136 observations. As explanatory variables, we use the price of the brand on purchase occasion t, a 0/1 display dummy indicating whether the brand was on display at t and a 0/1 feature dummy. Tables II and III give the posterior means and standard deviations of the parameters in these three models. These posterior results are based on the prior specification given in equations (22) and (23) with $S = 10 \times \mathbf{I}_{J-1}$ and $\lambda = 10$ to centre $\tilde{\Sigma}$ around 1. The results are based on 10,000 draws of the Gibbs sampler after burn in.

Table II shows the estimated posterior mean and standard deviations of the α, β, $\tilde{\Sigma}$ and Π parameters. The display, feature and price parameters have the expected sign, indicating that a feature, a display and a price reduction for a brand lead to higher utility for that brand and hence higher probability that the brand is chosen. The posterior mean of the intercept parameters are all positive, indicating that households prefer national brands above private labels. The intercept parameters are smaller for the dynamic models.

If we compare the posterior means of the β parameters across the three models, we observe that they are of similar size (and sign) for the static model and the common factor model but larger (in absolute sense) for the vector error-correction model specification. Also, the posterior uncertainty on the β parameters is higher for the VECM-MNP model. If we compare the display parameters with the posterior standard deviations, we observe that only our VECM-MNP

Table II. Posterior means of the model parameters (with posterior standard deviations in parentheses)

Parameter	Static MNP model[a]	Common factor MNP model[a]	VECM MNP model[a]
β parameter			
Display	0.05 (0.07)	0.04 (0.07)	0.35 (0.16)
Feature	0.27 (0.12)	0.37 (0.10)	0.45 (0.24)
Price	−1.81 (0.36)	−1.79 (0.36)	−1.96 (0.53)
Intercepts			
Sunshine	0.30 (0,26)	0.07 (0.25)	0.01 (0.21)
Keebler	0.84 (0.27)	0.69 (0.20)	0.51 (0.24)
Nabisco	1.93 (0.27)	1.75 (0.25)	1.79 (0.31)
α parameter			
Display			0.08 (0.08)
Feature			0.31 (0.09)
Price			−2.38 (0.33)
$\tilde{\Sigma}$	$\begin{pmatrix} 0.72 & 0.27 & 0.47 \\ (0.21) & (0.21) & (0.17) \\ & 0.65 & 0.64 \\ & (0.14) & (0.09) \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} 0.49 & 0.05 & 0.37 \\ (0.17) & (0.06) & (0.08) \\ & 0.22 & 0.38 \\ & (0.05) & (0.05) \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} 0.77 & -0.11 & 0.52 \\ (0.25) & (0.11) & (0.12) \\ & 0.37 & 0.33 \\ & (0.18) & (0.09) \\ & & 1 \end{pmatrix}$
Π		$\begin{pmatrix} 0.32 & 0.63 & -0.14 \\ (0.15) & (0.09) & (0.14) \\ 0.21 & 0.57 & -0.16 \\ (0.37) & (0.18) & (0.33) \\ 0.16 & -0.23 & 0.29 \\ (0.16) & (0.07) & (0.14) \end{pmatrix}$	$\begin{pmatrix} 0.50 & 0.41 & -0.14 \\ (0.11) & (0.08) & (0.18) \\ 0.28 & 0.58 & -0.19 \\ (0.14) & (0.09) & (0.11) \\ 0.01 & -0.14 & 0.46 \\ (0.05) & (0.08) & (0.07) \end{pmatrix}$
$\ln(\ell(\theta \mid \text{data}))$[b]	−3239	−2972	−2888

[a] The static MNP model is given in equation (6), the common factor MNP model in equation (10) and the VECM-MNP model in equation (9).
[b] The value of the log-likelihood function evaluated in the posterior means of the model parameters.

model suggests that this variable has a significant, in this case long-run, effect on brand choice. Both feature and price seem to have a significant effect on brand choice in all three models, while the results of the VECM-MNP model indicate that long-run and short-run effects are about the same, although the long-run effect of feature is only marginally significant. The posterior mean of the α parameters is smaller for feature and display and larger in absolute value for the price parameter. The posterior uncertainty in the α parameters is smaller than for their corresponding β parameters. Again, display does not have a significant short-run effect on brand choice. Feature and price however have a significant short-run effect.

The final panel of Table II shows the posterior means of the parameters in the Π matrix, which contains the correlation between the perceived relative utilities over time for both dynamic models. The relatively large diagonal elements of the Π matrix indicate that there is some persistence in brand choice, although these elements are not close to unity. If we consider

Table III. Posterior means of household heterogeneity parameters (with posterior standard deviations in parentheses).

Model	Σ_β[b]	Σ_α
Static MNP[a]	$\begin{pmatrix} 0.21 & -0.08 & -0.39 & -0.20 & -0.25 & -0.40 \\ (0.08) & (0.08) & (0.29) & (0.14) & (0.19) & (0.26) \\ & 0.32 & 0.02 & -0.02 & -0.03 & 0.14 \\ & (0.15) & (0.43) & (0.22) & (0.31) & (0.43) \\ & & 9.54 & -2.35 & -2.79 & -1.07 \\ & & (2.44) & (0.96) & (1.23) & (1.20) \\ & & & 3.07 & 3.54 & 3.60 \\ & & & (0.82) & (0.90) & (0.92) \\ & & & & 5.08 & 4.91 \\ & & & & (1.24) & (1.14) \\ & & & & & 6.46 \\ & & & & & (1.41) \end{pmatrix}$	
Common factor MNP[a]	$\begin{pmatrix} 0.23 & -0.12 & -0.27 & -0.04 & -0.07 & -0.23 \\ (0.08) & (0.08) & (0.34) & (0.14) & (0.17) & (0.28) \\ & 0.41 & 0.31 & -0.06 & -0.15 & 0.25 \\ & (0.19) & (0.50) & (0.22) & (0.26) & (0.36) \\ & & 10.82 & -3.30 & -4.22 & -1.81 \\ & & (3.09) & (1.32) & (1.51) & (1.31) \\ & & & 2.64 & 2.79 & 2.85 \\ & & & (0.87) & (0.88) & (0.85) \\ & & & & 3.92 & 3.65 \\ & & & & (1.16) & (1.11) \\ & & & & & 6.25 \\ & & & & & (1.56) \end{pmatrix}$	
VECM-MNP[a]	$\begin{pmatrix} 0.62 & -0.43 & 0.51 & -0.19 & -0.09 & 0.25 \\ (0.31) & (0.36) & (0.73) & (0.31) & (0.31) & (0.44) \\ & 1.52 & 0.70 & -0.39 & -0.20 & 0.64 \\ & (0.88) & (1.07) & (0.39) & (0.45) & (0.64) \\ & & 8.50 & -1.61 & -1.51 & 0.57 \\ & & (4.45) & (1.16) & (1.36) & (1.57) \\ & & & 2.34 & 1.82 & 2.38 \\ & & & (0.73) & (0.65) & (0.78) \\ & & & & 2.41 & 2.14 \\ & & & & (0.93) & (0.88) \\ & & & & & 5.26 \\ & & & & & (1.23) \end{pmatrix}$	$\begin{pmatrix} 0.24 & -0.05 & -0.01 \\ (0.08) & (0.08) & (0.31) \\ & 0.35 & 0.22 \\ & (0.15) & (0.34) \\ & & 8.91 \\ & & (2.65) \end{pmatrix}$

[a] The static MNP model is given in equation (6), the common factor MNP model in equation (10) and the VECM-MNP model in equation (9).

[b] The order of the variables in Σ_β and Σ_α is display, feature, price, and the three intercepts Sunshine, Keebler and Nabisco.

the posterior standard deviations of the off-diagonal elements, we see that the posterior means of the off-diagonal elements are usually not more than two standard deviations away from zero. This is not the case for the element in the second column of the first row for both specifications, which is also relatively high. All eigenvalues for the VECM-MNP model are positive. The posterior mean and mode of the largest eigenvalue of Π are 0.94 and 0.97, respectively. The 90% highest posterior density region is (0.89, 0.99). For the common factor MNP model all eigenvalues are also positive and the posterior mean and mode of the largest eigenvalue are 0.83 and 0.90, respectively. Now the 90% highest posterior density region is (0.68, 0.99). This suggests that there is slightly less persistence in brand choice for the common factor model.

As we have indicated in Section 2.1, neglecting unobserved household heterogeneity usually leads to an overestimate of the persistence in brand choice; see Keane (1997) for an example. Unobserved household heterogeneity captures the fact that individual households buy the same brand on consecutive purchase occasions due to their large base preference for that brand. If, however, we neglect unobserved household heterogeneity in our model, this repeated buying behaviour is now wrongfully completely captured by the dynamic structure. This overestimation of the persistence in brand choice will then inappropriately suggest that promotional activities like price reductions have more dynamic impact than they in fact have. Indeed, our unreported estimation results support this effect. The inclusion of unobserved heterogeneity in a model is therefore very important.

Table III shows the posterior means and standard deviations of the covariance matrices Σ_α and Σ_β modelling the unobserved heterogeneity. As in Allenby and Rossi (1999), we find substantial heterogeneity across the households. The final three diagonal elements of the Σ_β matrices concern the heterogeneity in the intercept parameters. We observe that the posterior means of the static model are larger than for the dynamic models. This suggests that neglecting dynamics may lead to an overestimate of the unobserved heterogeneity; see also Keane (1997) for a similar result. The first three diagonal elements of Σ_β and Σ_α concern the unobserved heterogeneity in the display, feature and price parameters. The pattern in the ubobserved heterogeneity is more or less the same as in the posterior means of β. There is less unobserved heterogeneity in the display and feature parameters for the static and common factor MNP model than for the VECM-MNP model. For the price parameters it is the other way around. The posterior standard deviations are relatively high especially for the VECM specification. If we compare the diagonal elements of the posterior means of Σ_β and Σ_α matrices for the VECM-MNP model, we notice that there is less heterogeneity in the display and feature parameters and more heterogeneity in the price parameters in the short run than in the long run. The differences are, however, relatively small.

To compare the three MNP models, we may consider Bayes factors. As we have imposed improper priors for most of the parameters, Bayes factors are not properly defined. Instead, we will consider predictive likelihoods to compare the three models in the next subsection. To give an idea of the fit of the model, we display the value of the log-likelihood function evaluated in the posterior means in the final row of Table II. As Geweke *et al.* (1997) show, the posterior means provide good estimates for the parameters of a MNP model. The value of the log-likelihood function of the static model is substantially smaller than for both dynamic models. This result is also supported by the value of the posterior density of Π in $\Pi = \mathbf{0}$, which is about equal to zero. Furthermore, the log-likelihood function of the common factor MNP model is smaller than for the VECM-MNP model that allows for different long-run and short-run effects.

This suggests that VECM-MNP model provides a better description of the data than the common factor model.

Finally, we analyse the sensitivity of our results with respect to the prior specification for the household heterogeneity parameters. We impose an inverted Wishart prior with $S = \delta \times \mathbf{I}$ and $\lambda = \delta$ for Σ_β and Σ_α parameters with different values of δ. The remaining prior specification stays the same. We consider again the posterior means and standard deviations of our VECM-MNP model. The posterior means and standard deviations of all parameters turn out to be roughly the same for relative large values of δ. For small values of δ the posterior means of the diagonal elements of Σ_β and Σ_α become somewhat smaller. The differences are, however, never larger than two times the posterior standard deviations.

5.3 Forecast Comparison

To compare the three models on out-of-sample data, we forecast the brand choice of each of the 136 households on the next purchase occasion using the method described in Section 4.2. Forecasts are made conditional on the known future values of the explanatory variables and compared with the actual brand choices in the hold-out sample. We translate the predictive probabilities to discrete choices by assuming that a household chooses brand j if the predictive probability for brand j is the largest. Forecasts are made using posterior knowledge of β_i and α_i obtained from the observed data, that is, we use $p(\beta_i, \alpha_i | \text{data})$ to integrate over the household heterogeneity.

Table IV gives the results of the forecasting exercise. The first row shows the percentage of correct forecasts. We see that the static MNP produces for 87.5% of the 136 households a correct forecast of brand choice. Both dynamic models perform slightly worse and produce in about 85.3% (common factor model) and 86.8% (VECM model) of the cases a correct forecast. The percentage of correct hits can be seen as the risk of a 0/1 loss function. It does not take into account the performance of the model for mis-hit situations.

An overall measure of the forecast performance of the three models can be based on the value of the predictive likelihood function. The predictive likelihood is defined as the predictive density function evaluated in the out-of-sample observations. In fact, they are closely related to non-predictive Bayes factors and they are still interpretable with improper prior specifications. They are therefore suitable for model comparison; see Geweke (1999) for a discussion and Geweke (1996) for an application.

Table IV. Forecast performance of three multinomial probit models

	Static MNP model[a]	Common factor MNP model[a]	VECM MNP model[a]
% correct hits[b]	87.50	85.29	86.76
Log pred. likelihood[c]	−52.04	−52.50	−49.80

[a] The static MNP model is given in equation (6),the common factor MNP model in equation (10) and the VECM-MNP model in equation (9).
[b] The relative number of times that the predictive probability that the true brand choice was chosen is the maximum over all choices.
[c] The log of the predictive likelihood function, that is, the predictive density function evaluated in the out-of-sample observations.

In our case, the predictive likelihood function is simply the product of the predictive probabilities for the choices of the households, that is,

$$\prod_{i=1}^{I} \Pr\left[d_{i,T_i+1} = j_{i,T_i+1}\right] \tag{26}$$

where $\Pr[d_{i,T_i+1} = j]$ are the predictive probabilities defined in equation (25) and j_{i,T_i+1} denotes the actual out-of-sample choice of the households. This predictive likelihood can easily be computed using the output of the Gibbs sampler. The final row of Table IV shows the log of the predictive likelihood functions for the three models. The values are roughly the same for the static model and the common factor model, where the static model is slightly better. The value of the log predictive likelihood of the VECM-MNP model is, however, substantially smaller than for the other two models. Hence, on the basis of predictive likelihoods the dynamic MNP model with different long-run and short-run effects is preferred.

5.4 Long-run Market Shares and Prices

The dynamic MNP models allow an analysis of long-run market shares as defined in Section 3. Here, we will consider our vector error-correction MNP model (9), as it is favoured on the basis of predictive likelihoods and the in-sample value of the log-likelihood function. In Figure 1 we present the long-run market shares for the four brands for a range of price values. In each graph we give the long-run market share for each brand as defined in equation (18) as a function of the price of one brand, where we set the prices of the other brands at their average price given in Table I. We assume that there are no features or displays, and hence the feature and display dummies are set equal to 0. The market shares are computed as discussed below equation (18). As the market shares are functions of the model parameters, we average them with respect to the posterior distribution of the model parameters. To give an indication of posterior uncertainty, we display in Figure 1 the posterior mean together with posterior interquartile ranges. In all cases the posterior mean lies roughly in the middle of the posterior interquartile range.

The left panel of Figure 1 shows the market shares of Sunshine and Keebler, while the right panel shows the market shares of Nabisco and Private label. Note that both panels have a different scale on the vertical axis. There is a variety of conclusions that can be drawn from the graphs, all of which can be relevant for managerial purposes. Here, we will mention only a few. A first is that when Sunshine, Keebler and Private label increase their prices permanently, it leads in all cases to a long-run market leadership of Nabisco. However, when Nabisco increases its price, only Private label will dominate the market in the long run. A second conclusion is that Nabisco keeps its market leadership for a wide range of prices. Note that for very high prices of Nabisco its market share becomes eventually 0 as the price coefficient is negative. A third conclusion is that price changes in Sunshine and Keebler have about the same effect on the long-run market shares of Nabisco. This is also true for Private label. Finally, Sunshine needs a price lower than other brands, to gain a long-run market share of 30%.

5.5 Dynamic Effects of Promotional Activities

Now we turn to analysing the short-run effects of promotional activities on market shares, which in turn can be used in a framework for making decisions concerning the marketing mix;

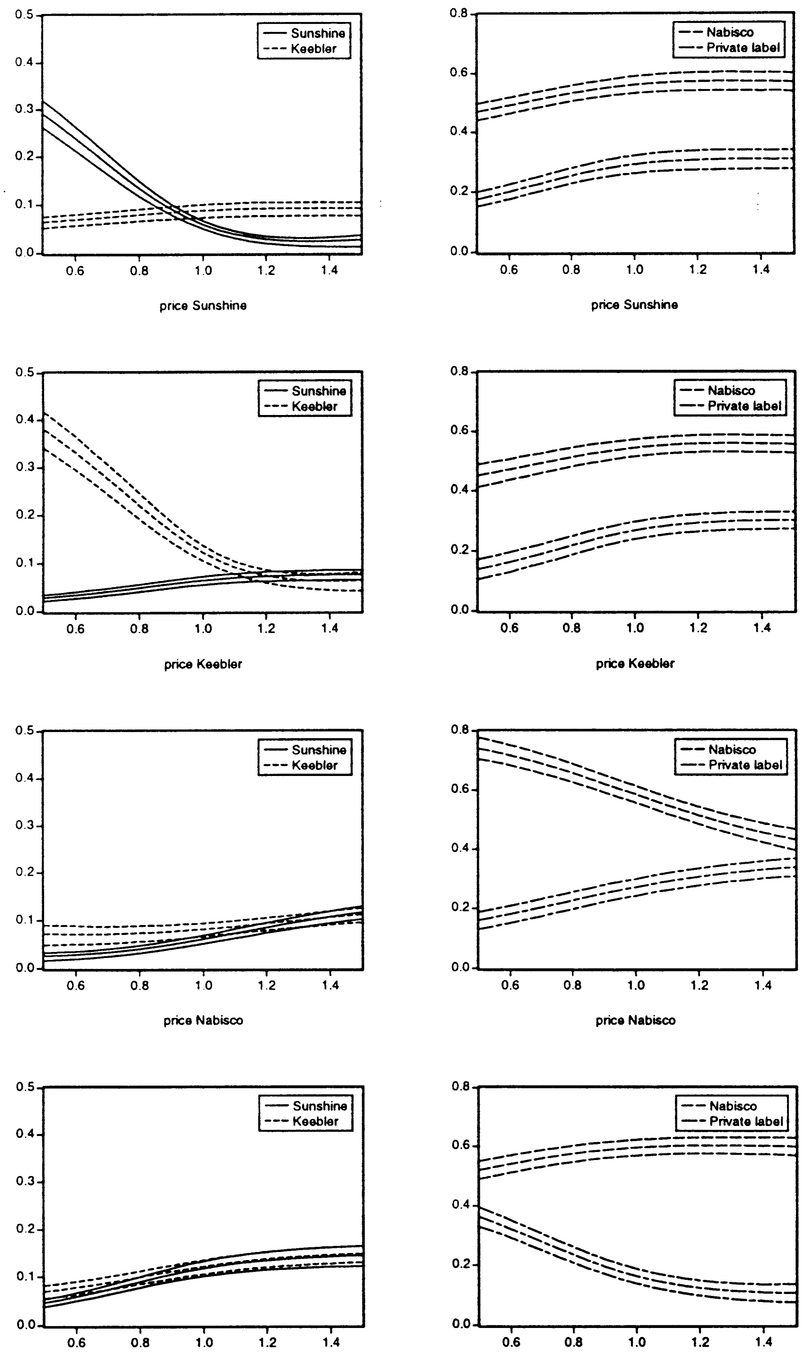

Figure 1. Posterior means and posterior interquartile ranges of the long-run market shares for the four brands as a function of price

see also Keane (1997). Note again that the common factor MNP model (10) assumes that the short-run and long-run effects are the same and is therefore not suitable for this analysis. As the number of possible promotional activities is large, we will consider in this section only two exercises. In the first exercise we investigate the short-run effects of a feature on the market shares of the four brands. In a second exercise we analyse the short-run effects of a price reduction of 50% (2 for the price of 1) on the market shares of the four brands. In both exercises we focus on the brand Sunshine.

Figures 2 and 4 show the gain or loss in market share for the current and next four purchase occasions, caused by a feature during the current purchase occasion. These graphs are constructed by simulation $\tilde{U}^*_{i0}$ from equation (17), where $\tilde{X}_{i0}$ is such that there are no features or displays for any brands, and the prices of the four brands are equal to their sample mean values given in Table I. Next, we simulate recursively the values of $\tilde{U}_{it}$, $t=1,\ldots,5$, using equation (9), where we keep the values of the explanatory variables $\tilde{X}_{it}$ the same as $\tilde{X}_{i0}$ except at $t=1$, where we allow for a feature in one or two of the brands. This is repeated several times and the relative frequency that the sampled U_{ijt} is larger than U_{imt}, $m \neq j$ determines the market share of brand j on purchase occasion t, $t=0,\ldots,5$. Note that at $t=0$ we in fact compute the long-run market share of brand j. Hence, the graphs show the gain or loss in choice probability for brand j on purchase occasion t, that is, $(\Pr[d_{it}=j]-\Pr[d_{i0}=j])$ given that there was a feature on the first purchase occasion ($t=1$).

As the brand choice probabilities are functions of the model parameters, we average $(\Pr[d_{it}=j]-\Pr[d_{i0}=j])$ with respect to the posterior distribution of the model parameters. To give an indication of posterior uncertainty, we display in the graphs the posterior mean (bold lines) together with posterior interquartile ranges. Note that the distribution is often skewed such that posterior mean may lay outside the interquartile range.

The graphs in Figure 2, which shows the effects of a single feature for one of the brands, suggest several conclusions. The first conclusion is that there are indeed short-run effects of features on future brand choice decisions but the effects are not very persistent. The impact of feature promotions dies out already after four periods although the eigenvalues of Π are relatively large. A second conclusion is that the gain in market share of a single feature is smaller for Nabisco than for the other three brands. Features of Sunshine and Keebler have the largest negative impact on the market share of the Nabisco and Private label, while single features of Nabisco and Private label have the largest negative impact on Private label and Nabisco, respectively. When Keebler has a feature, Nabisco witnesses a positive impact after a few purchase occasions, due to the brand switching behavior modelled by the Π matrix.

From a managerial point of view it is also important to analyse the revenue effects of the feature. The total revenue of the feature is the sum of the gain in choice probability times the price of the brand over five purchase occasions, that is,

$$\sum_{i=1}^{5}(\Pr[d_{it}=j]-\Pr[d_{i0}=j])p_{jt} \tag{27}$$

where p_{jt} denotes the price of brand j at time t and $(\Pr[d_{it}=j]-\Pr[d_{it}=0])$ the gain or loss in choice probability at time t. Figure 3 displays the posterior density of equation (27) for the four brands in case of a single feature in the brand under consideration. The distributions are skewed and much of the posterior mass is around zero. The latter is caused by a return to the long-run

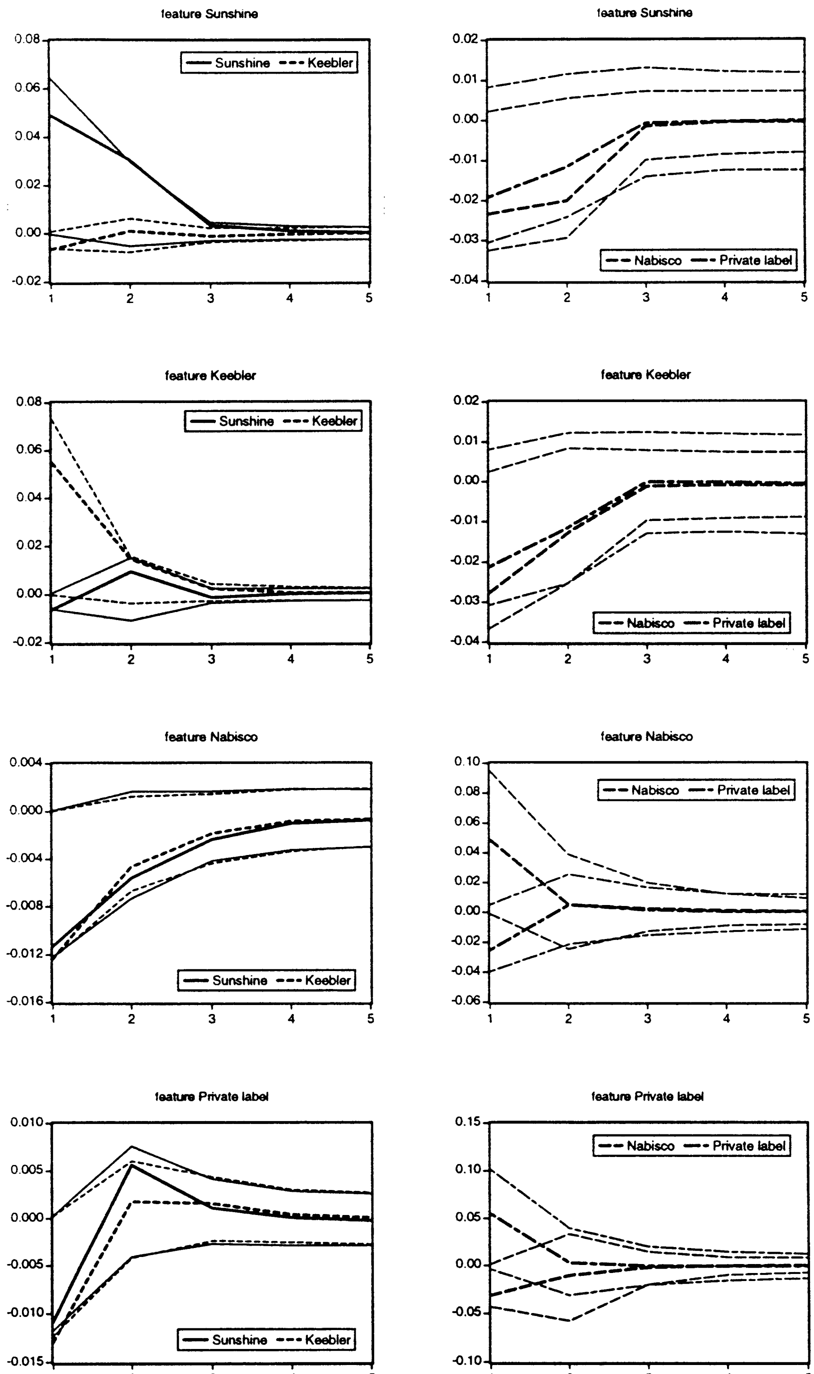

Figure 2. The influence of a feature at time 1 for one brand on future market shares, posterior means (bold) and interquartile ranges

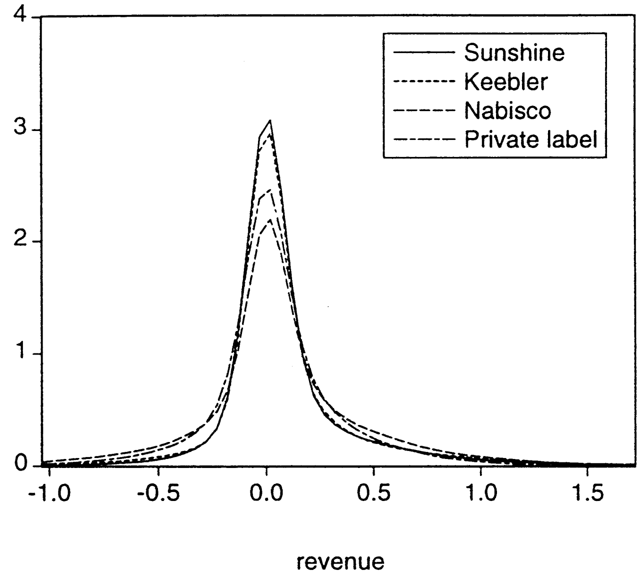

Figure 3. Posterior density of the total change in revenue over five purchase occasions due to a single feature on the first purchase occasion

market share after four periods and the fact that some households are not affected by features which is captured by the household heterogeneity parameters. The posterior mean of the total revenue for Sunshine and Keebler over the five periods is 0.08. The total revenue for Nabisco and Private label is smaller, 0.06 and 0.04 respectively. Additionally, we notice that the posterior uncertainty of the total revenue of a single feature is smaller for the brands with a small market share.

Figure 4 shows the effects of two features at the same time for Sunshine and for one of the other brands. We can learn, for example, that a single feature of Sunshine has more effect than a feature that coincides with a feature of one of the other brands. If the feature of Sunshine coincides with a feature of Keebler, both brands will gain roughly 4% market share in the first period. Private label will experience a loss in market share in period 2 if they have a feature at the same time as when Sunshine has one. This is due to the brand switching behaviour of the households modelled by Π. To a lesser extent this can also be concluded for Nabisco.

Likewise, we may analyse the effects of a price reduction of 50% during the current purchase occasion. We assume no features or displays and set the prices of the other brands at their sample average. Figure 5 shows the effects of a single price reduction in one of the four brands. Again the impact of the price reduction dies out very quickly. We can observe that temporary price reductions in Sunshine and Keebler have more impact on their market share than price reductions in Nabisco and Private label. Furthermore, Nabisco has the largest drop in market shares if one of the brands decides to reduce its price.

Again, we consider the total revenue as defined in equation (27) due to the price reduction. Figure 6 shows the posterior distribution of the total revenue over the five periods. Again, the posterior distributions are skewed and much of the posterior mass is around zero. The posterior

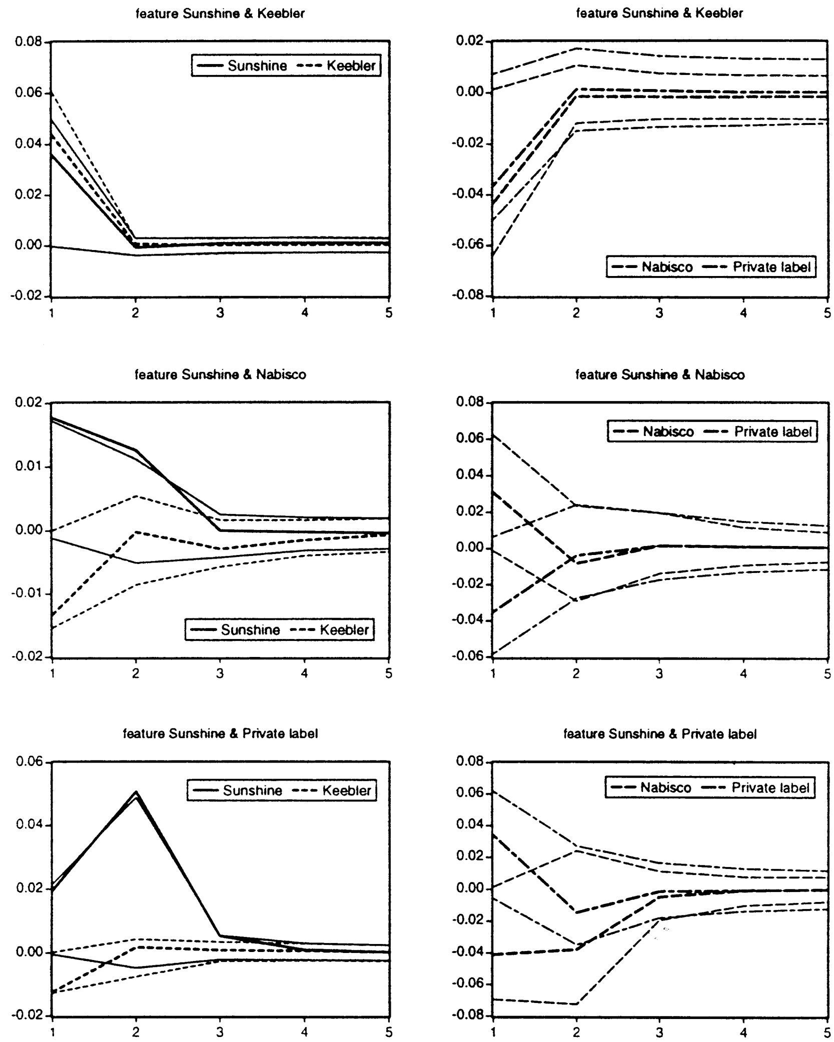

Figure 4. The influence of features at time 1 for two brands on future market shares, posterior means (bold) and interquartile ranges

means of the total revenue are now 0.17 and 0.18 for Sunshine and Keebler, respectively. For Nabisco and Private label the posterior mean of the total revenue is smaller, 0.05 and 0.06 respectively. The brands with the smallest market share (Sunshine and Keebler) seem to have more revenue from a price reduction than Nabisco. Note that we have to be careful in interpreting the results of such a large price reduction. It is maybe likely that households will buy more than one unit of saltine crackers in case of a huge price reduction and may also start to consume more saltine crackers per week. In that case the choice probabilities cannot be directly translated to market shares as discussed earlier.

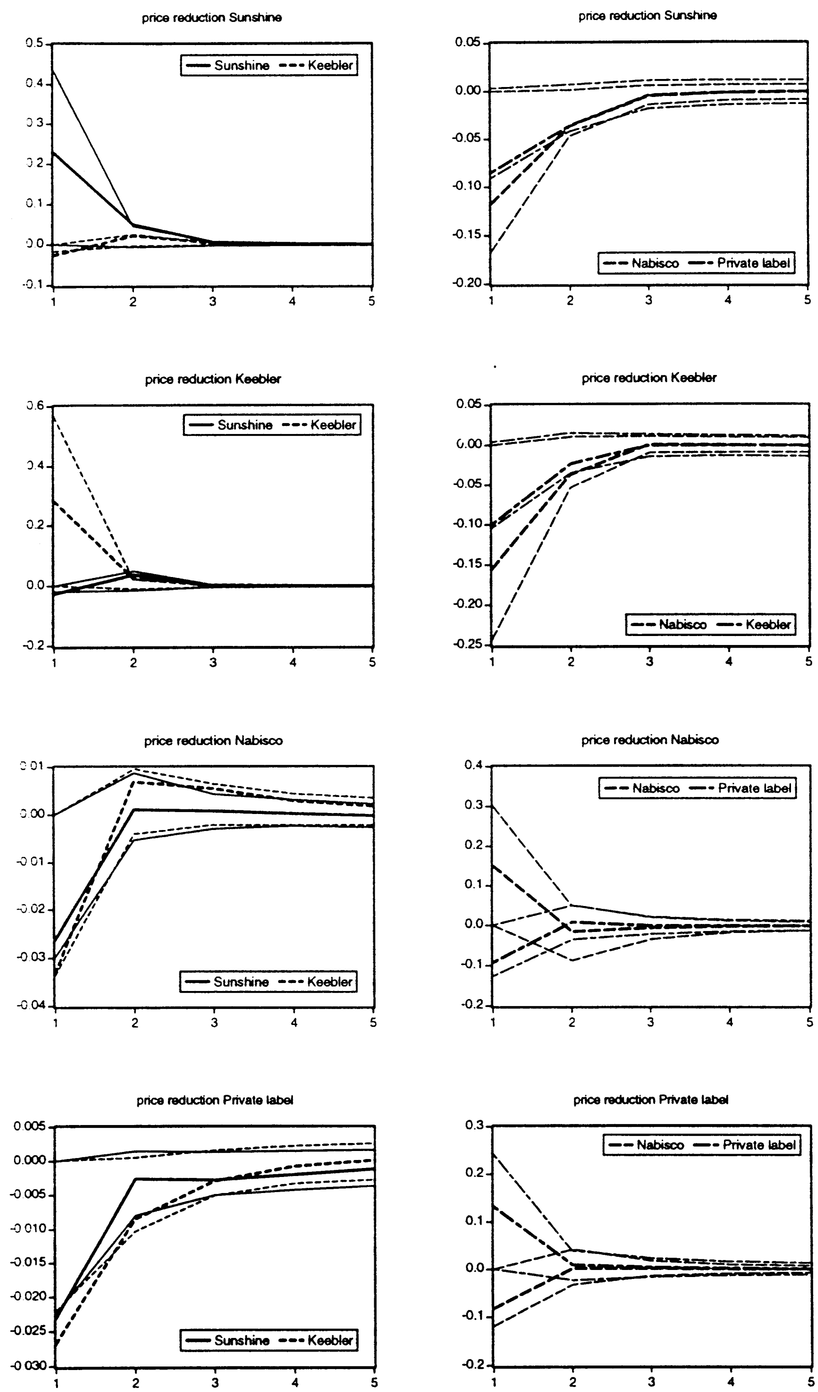

Figure 5. The influence of a price reduction of 50% in one brand at time 1 on future market shares, posterior means (bold) and interquartile ranges

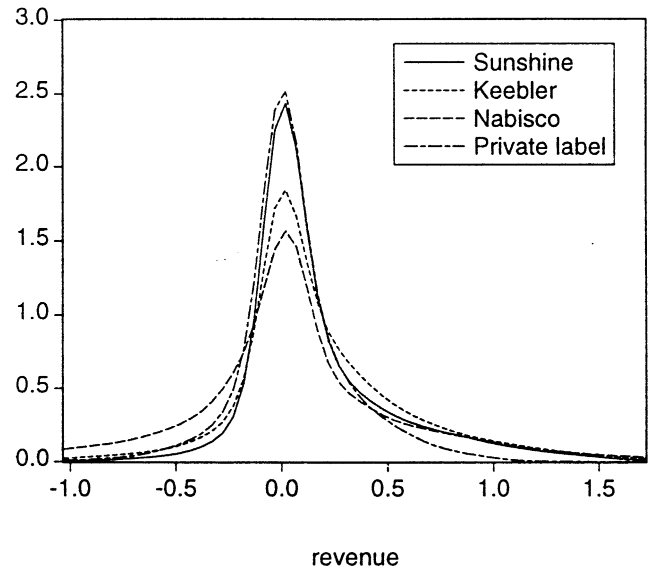

Figure 6. Posterior density of the total change in revenue over five purchase occasions due to a 50% price reduction in one brand at the first purchase occasion

Finally, we show the effects of a simultaneous price reduction in Sunshine and one of the other brands in Figure 7. It is obvious that a single price reduction is better for Sunshine than one that coincides with a price reduction in one of the other brands. A price reduction that coincides with Private label leads to better market shares than a price reduction that coincides with Nabisco or Keebler. Again, the gain in market share is at the cost of the market share of Nabisco, unless Nabisco also introduces a price reduction.

6. CONCLUSION

In this paper we proposed a multinomial probit model with VECM dynamics in unobserved utilities. We discussed the interpretation of the model in terms of the long-run and short-run impact of marketing efforts on dynamic brand choice. We dealt with estimation and inference issues, and we applied our model to a prototypical data set. This application of our model indicated that it can be useful in giving guidance as to which marketing mix strategies may lead to larger probabilities that a brand is favoured above other brands.

There are at least three interesting areas for further research. A first concerns brand loyalty. It seems of interest to examine if marketing efforts have a persistent positive effect on brand loyalty. In terms of our model, this would mean that the vector autoregressive parameters become functions of explanatory variables, and that they occasionally experience levels shifts. A second, and related, area concerns the impact of promotional intensity on long-run and short-run brand choice. It may be that households behave differently at times of high intensity. Finally, it is important to extend the dynamic brand choice model with a dynamic model for interpurchase times and the number of purchases on each purchase occasion. In that case, we

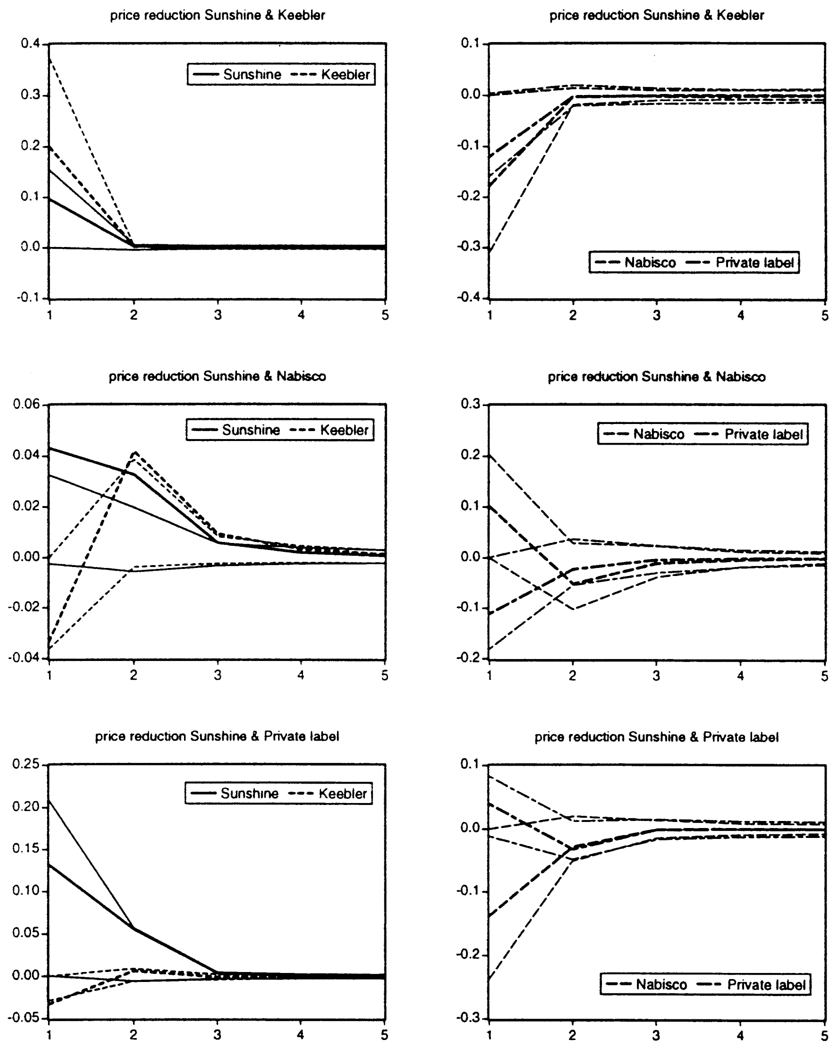

Figure 7. The influence of price reductions of 50% in two brands at time 1 on future market shares, posterior means (bold) and interquartile ranges

can get a better understanding of the effects of (large) price reductions on market shares instead of on choice probabilities.

APPENDIX: DERIVATION OF FULL CONDITIONAL POSTERIOR DISTRIBUTIONS

In this appendix we give a short outline of the derivation of the full conditional posterior distributions needed in the Gibbs sampler; see Geweke *et al.* (1997) and McCulloch and Rossi (1994) for a more detailed overview for a similar model. We only discuss the VECM-MNP model in equation (9) as the static MNP model in equation (6) and the common factor dynamic MNP model in equation (10) are nested within this model. The full conditional posterior distributions for the last model can be derived in a similar way.

To determine the sampling distributions, we rewrite the MNP model so that it represents a standard univariate or multivariate regression model with the parameter to be sampled as regression parameters or variance (covariance matrix) parameters of the error term. For a standard regression model we know that the full conditional posterior distribution of the regression parameter is (matrix) normal with mean and variance resulting from the ordinary least squares (OLS) estimators. The full conditional posterior distribution of the variance (covariance matrix) of the error term is an inverted χ^2 (or inverted Wishart) distribution.

To obtain the full conditional posterior distribution of β we rewrite equation (9) as

$$\tilde{\Sigma}^{-\frac{1}{2}}\big(\Delta\tilde{U}_{it} - \Delta\tilde{X}_{it}(\alpha+\alpha_i) - (\mathbf{I}_{J-1} - \Pi)(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}\beta_i)\big) = \tilde{\Sigma}^{-\frac{1}{2}}(\mathbf{I}_{J-1} - \Pi)\tilde{X}_{i,t-1}\beta + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{it}$$

$i = 1,\ldots,I$, $t = 1,\ldots,T_i$. This represents $(J-1)$ regression equations with regression coefficient β and uncorrelated normally distributed error terms with unit variance. Hence, the distribution of β given α, $\tilde{\Sigma}$, β_i, α_i, Σ_β, Σ_α, Π and $\tilde{U}$ is normal. The mean and variance result from the OLS estimators of β.

Likewise, we can rewrite equation (9) to sample α

$$\tilde{\Sigma}^{-\frac{1}{2}}\big(\Delta\tilde{U}_{it} - \Delta\tilde{X}_{it}\alpha_i - (\Pi - \mathbf{I}_{J-1})(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta+\beta_i))\big) = \tilde{\Sigma}^{-\frac{1}{2}}\Delta\tilde{X}_{it}\alpha + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{it}$$

$i = 1,\ldots,I$, $t = 1,\ldots,T_i$ such that it represents $(J-1)$ regression equations with regression coefficient α. Hence, the distribution of α given β, $\tilde{\Sigma}$, β_i, α_i, Σ_β, Σ_α, Π and $\tilde{U}$ is normal, where the mean and variance follow from the OLS estimators of α.

For $\tilde{\Sigma}$ we note that

$$\Delta\tilde{U}_{it} = \Delta\tilde{X}_{it}(\alpha+\alpha_i) + (\Pi - \mathbf{I}_{J-1})\big(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta+\beta_i)\big) + \eta_{it}$$

for $i = 1,\ldots,I$ and $t = 1,\ldots,T_i$ is a multivariate regression model. The distribution of $\tilde{\Sigma}$ given β, α, Σ_β, Σ_α, β_i, α_i, Π and $\tilde{U}$ is simply an inverted Wishart distribution. The prior information is included by increasing the number of degrees of freedom by λ and the location parameter by S^{-1}.

To sample β_i we rewrite equation (9) as

$$\begin{aligned}\tilde{\Sigma}^{-\frac{1}{2}}\big(\Delta\tilde{U}_{it} - \Delta\tilde{X}_{it}(\alpha+\alpha_i) - (\mathbf{I}_{J-1} - \Pi)(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}\beta_i)\big) &= \tilde{\Sigma}^{-\frac{1}{2}}(\mathbf{I}_{J-1} - \Pi)\tilde{X}_{it}\beta + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{it}\\ \mathbf{0} &= \Sigma_\beta^{-\frac{1}{2}}\beta_i + \Sigma_\beta^{-\frac{1}{2}}w_i\end{aligned}$$

for $i = 1,\ldots,I$. The last line follows from $w_i = (\beta_i - \mathbf{0}) \sim \mathrm{N}(\mathbf{0},\Sigma_\beta)$. This denotes J regression equations with regression parameter β_i and hence the distribution of β_i, $i = 1,\ldots,I$, given β, α, $\tilde{\Sigma}$, Σ_β, Σ_α, α_i, Π, and $\tilde{U}$ is normal. The mean and variance follow from the OLS estimators.

Likewise, to sample α_i we write equation (9) as I regression models with regression coefficient α_i, $i = 1,\ldots,I$

$$\begin{aligned}\tilde{\Sigma}^{-\frac{1}{2}}\big(\Delta\tilde{U}_{it} - \Delta\tilde{X}_{it}\alpha - (\Pi - \mathbf{I}_{J-1})(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta+\beta_i))\big) &= \tilde{\Sigma}^{-\frac{1}{2}}\Delta\tilde{X}_{it}\alpha_i + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{it}\\ \mathbf{0} &= \Sigma_\alpha^{-\frac{1}{2}}\alpha_i + \Sigma_\alpha^{-\frac{1}{2}}w_i\end{aligned}$$

where the last line follows from $w_i = (\alpha_i - \mathbf{0}) \sim \mathrm{N}(\mathbf{0}, \Sigma_\alpha)$. The distribution of α_i, $i = 1, \ldots, I$, given β, α, $\tilde{\Sigma}$, β_i, Σ_β, Σ_α, Π and $\tilde{U}$ is normal, where the mean and variance follow from the OLS estimators.

For Σ_β and Σ_α it holds that

$$p(\Sigma_\beta | \beta, \alpha, \Sigma_\alpha, \tilde{\Sigma}, \beta_i, \alpha_i, \Pi, \tilde{U}) \propto \exp\left(-\frac{1}{2}\beta_i \Sigma_\beta^{-1} \beta_i'\right)$$
$$p(\Sigma_\alpha | \beta, \alpha, \Sigma_\beta, \tilde{\Sigma}, \beta_i, \alpha_i, \Pi, \tilde{U}) \propto \exp\left(-\frac{1}{2}\alpha_i \Sigma_\alpha^{-1} \alpha_i'\right)$$

and hence Σ_β and Σ_α can be sampled from an inverted Wishart distribution.

To obtain the full conditional posterior distribution Π we rewrite equation (9) as

$$\tilde{U}_{it} - \Delta\tilde{X}_{it}(\alpha + \alpha_i) - \tilde{X}_{i,t-1}(\beta + \beta_i) = \Pi\big(\tilde{U}_{i,t-1} - \tilde{X}_{i,t-1}(\beta + \beta_i)\big) + \eta_{it}$$

$i = 1, \ldots, I$, $t = 1, \ldots, T_i$, such that it is a multivariate regression model with regression parameter Π. The distribution of Π given β, α, $\tilde{\Sigma}$, α_i, β_i, Σ_β, Σ_α, and $\tilde{U}$ is matrix normal. The mean and variance results from the OLS estimators for Π with known $\tilde{\Sigma}$. As we want the eigenvalues of Π to be within the unit circle we reject the draw if this requirement is not met and redraw again.

Finally, to sample $\tilde{U}_{it}$ we consider

$$-\tilde{\Sigma}^{-\frac{1}{2}}\big(\Pi\tilde{U}_{i,t-1} + \Delta\tilde{X}_{it}(\alpha + \alpha_i) - (\Pi - \mathbf{I}_{J-1})\tilde{X}_{i,t-1}(\beta + \beta_i)\big) = -\tilde{\Sigma}^{-\frac{1}{2}}\tilde{U}_{it} + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{it}$$
$$\tilde{\Sigma}^{-\frac{1}{2}}\big(\tilde{U}_{i,t+1} - \Delta\tilde{X}_{i,t+1}(\alpha + \alpha_i) + (\Pi - \mathbf{I}_{J-1})\tilde{X}_{it}(\beta + \beta_i)\big) = \tilde{\Sigma}^{-\frac{1}{2}}\Pi\tilde{U}_{it} + \tilde{\Sigma}^{-\frac{1}{2}}\eta_{i,t+1}$$

which can be interpreted as a regression model with $\tilde{U}_{it}$ as regression parameter. Hence, $\tilde{U}_{it}$ is normally distributed. The mean and variance follow from the OLS estimators. Note that $\tilde{U}_{it}$ has to be drawn from a truncated normal distribution as $U_{i,j,t}$ has to be larger than $U_{i,m,t}$ for all $m \neq j$ if $d_{it} = j$ and smaller than $U_{i,m,t}$ if $d_{it} = m$; see Geweke (1991) for details.

ACKNOWLEDGEMENTS

We thank the participants of the conference on Inference and Decision Making in June 1999 in Rotterdam, two anonymous referees, and in particular the discussant and co-editor John Geweke, for their helpful comments. The first author thanks the Netherlands Organizations for Scientific Research (NWO) for its financial support.

REFERENCES

Albert, J. H. and S. Chib (1993), 'Bayesian analysis of binary and polychotomous response data', *Journal of the American Statistical Association* **88**, 669–679.

Allenby, G. M. and P. J. Lenk (1994), 'Modeling household purchase behavior with logistic normal regression', *Journal of the American Statistical Association* **89**, 1218–1231.

Allenby, G. M. and P. E. Rossi (1999), 'Marketing models of consumer heterogeneity', *Journal of Econometrics*, **89**, 57–78.

Blattberg, R. C. and S. A. Neslin (1989), 'Sales promotion: the long and short of it', *Marketing Letters*, **1**, 81–97.

Börsch-Supan, A. and V. A. Hajivassiliou (1993), 'Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models', *Journal of Econometrics*, **58**, 347–368.

Bronnenberg, B. J., V. Mahajan and W. R. Vanhonacker (2000), 'The emergence of new repeat-purchase categories: the interplay of market share and retailer distribution', *Journal of Marketing Research*, **37**, 16–31.

Bunch, D.S. (1991), 'Estimatibility in the multinomial probit model', *Transportation Research B*, **25B**, 1–12.

Chintagunta, P. K., D. C. Jain and N. J. Vilcassim (1991), 'Investigating heterogeneity in brand preferences in logit models for panel data', *Journal of Marketing Research*, **28**, 417–428.

Cramer, J. (1991), *The Logit Model: An Introduction for Economists*, Edward Arnold, New York.

Daganzo, C. (1979), *Multinomial Probit: The Theory and its Application to Demand Forecasting*, Academic Press, New York.

Dekimpe, M. G. and D. M. Hanssens (1995), 'The persistence of marketing effects on sales', *Marketing Science*, **14**, 1–21.

Erdem, T. (1996), 'A dynamic analysis of market structure based on panel data', *Marketing Science*, **15**, 359–378.

Erdem, T. and M. P. Keane (1996), 'Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer good markets', *Marketing Science*, **15**, 1–20.

Geisser, S. (1965), 'A Bayes approach for combining correlated estimates', *Journal of the American Statistical Association*, **60**, 602–607.

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.

Geweke, J. F. (1991), 'Efficient simulation from the multivariate normal and Student-t-distributions subject to linear constraints', in *Computer Science and Statistics: Proceedings of the 23rd Symposium on Interface*, American Statistical Association, Alexandria, VA.

Geweke, J. F. (1996), 'Bayesian reduced rank regression in econometrics', *Journal of Econometrics*, **75**, 121–146.

Geweke, J. F. (1999), 'Using simulation methods for Bayesian econometric models: inference development, and communication', *Econometric Reviews*, **18**, 1–73.

Geweke, J. F., M. P. Keane and D. E. Runkle (1997), 'Statistical inference in the multinomial multiperiod probit model', *Journal of Econometrics*, **80**, 125–165.

Gönül, F. and K. Srinivasan (1993), 'Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues', *Marketing Science*, **12**, 213–229.

Guadagni, P. E. and J. D. C. Little (1983), 'A logit model of brand choice calibrated on scanner data', *Marketing Science*, **2**, 203–238.

Gupta, S. (1988), 'Impact of sales promotions on when, what, and how much to buy', *Journal of Marketing Research*, **25**, 342–355.

Hausman, J. and D. Wise (1978), 'A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogenous preferences', *Econometrica*, **45**, 319–339.

Hendry, D. F., A. R. Pagan and J. D. Sargan (1984), 'Dynamic specification', in Z. Griliches and M. D. Intriligator (Eds), *Handbook of Econometrics*, Vol. 2, Chap. 18, North-Holland, Amsterdam, pp. 1023–1100.

Jain, D. C., N. J. Vilcassim and P. K. Chintagunta (1994), 'A random-coefficients logit brand-choice model applied to panel data', *Journal of Business & Economic Statistics*, **12**, 317–328.

Jedidi, K., C. F. Mela and S. Gupta (1999), 'Managing advertising and promotion for long-run profitability', *Marketing Science*, **19**, 1–22.

Keane, M. P. (1997), 'Modeling heterogeneity and state dependence in consumer choice behavior', *Journal of Business & Economic Statistics*, **15**, 310–327.

Lükepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, (2nd edn), Springer-Verlag, Berlin.

McCulloch, R. and P. E. Rossi (1994), 'An exact likelihood analysis of the multinomial probit model', *Journal of Econometrics*, **64**, 207–240.

McFadden, D. (1973), 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (Ed.), *Frontiers in Econometrics*, Chap. 4, Academic Press, New York; 105–142.

McFadden, D. (1984), 'Econometric analysis of qualitative response models', in Z. Griliches and M. Intriligator (Eds), *Handbook of Econometrics*, Vol. 2, Chap. 18, North-Holland, Amsterdam, pp. 1395–1457.

Mela, C. F. S. Gupta and K. Jedidi (1998), 'Assessing long-term promotional influences on market structure', *International Journal of Research in Marketing*, **15**, 89–107.

Mela, C. F., S. Gupta and D. Lehmann (1997), 'The long-term impact of promotions and advertising on consumer brand choice', *Journal of Marketing Research*, **34**, 248–261.

Papatla, P. and L. Krishnamurthi (1996), 'Measuring the dynamic effects of promotions on brand choice', *Journal of Marketing Research*, **33**, 20–35.

Rossi, P. E. and G. M. Allenby (1993), 'A Bayesian approach to estimating household parameters', *Journal of Marketing Research*, **30**, 171–182.

Tanner, M. A. and W. H. 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.

