Sales Models For Many Items Using Attribute Data

Erjen van Nierop, Dennis Fok, Philip Hans Franses

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2002-65-MKT
Publication	July 2002
Number of pages	28
Email address corresponding author	franses@few.eur.nl
URL	http://www.eur.nl/WebDOC/doc/erim/erimrs20020729172534.pdf
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam P.O. Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31 10 408 1182 Fax: +31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS **R**ESEARCH **I**NSTITUTE OF **M**ANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	Sales models are mainly used to analyze markets with a fairly small number of items, obtained after aggregating to the brand level. In practice one may require analyses at a more disaggregate level. For example, brand managers may be interested in a comparison across product attributes. For such an analysis the number of relevant items in the product category make commonly used sales models difficult to use as they would contain too many parameters. In this paper we propose a new model, which allows for the analysis of a market with many items while using only a moderate number of easily interpretable parameters. This is achieved by writing the sales model as a Hierarchical Bayes model. In this way we relate the marketing-mix effectiveness to item characteristics such as brand, package size, package type and shelf position. In this specification we do not have to impose restrictions on the competitive structure, as all items are allowed to have different own and cross elasticities. The parameters in the model are estimated using Markov Chain Monte Carlo techniques. As a by-product this model allows to make predictions of sales levels and marketing-mix effectiveness of new to introduce items or of attribute changes. For example, one can assess the impact of changing the packaging from plastic to glass, on sales and price elasticity. Besides entering and changing products, our model also allows for items to leave the market. We consider the representation, specification and estimation of the model. We apply the model to a ketchup scanner data set with 23 items at the chain level. Our results indicate that the model fits the sales of most items very well.	
Library of Congress Classification (LCC)	5001-6182	Business
	5410-5417.5	Marketing
	HF 5438.4	Sales Management
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 31 C 44	Marketing Statistical Decision Theory
	M 31	Marketing
European Business Schools Library Group (EBSLG)	85 A	Business General
	280 G 255 A	Managing the marketing function Decision theory (general)
	290 S	Selling
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.40 85.03	Marketing Methoden en technieken, operations research
	85.40	Marekting
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Marketing / Besliskunde	
	Verkoop, Merken, Markov-ketens, Monte Carlo-technieken	
Free keywords	Sales Models, Attribute Data, SKU Level Analysis, Hierarchical Bayes, Markov Chain Monte Carlo	

Sales models for many items using attribute data

Erjen van Nierop

Tinbergen Institute, Erasmus University Rotterdam

Dennis Fok

ERIM, Erasmus University Rotterdam

Philip Hans Franses*

Econometric Institute and Department of Marketing and Organization
Erasmus University Rotterdam

July 4, 2002

*Corresponding author: Econometric Institute, office H11-34, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, email: franses@few.eur.nl

Sales models for many items using attribute data

Abstract

Sales models are mainly used to analyze markets with a fairly small number of items, obtained after aggregating to the brand level. In practice one may require analyses at a more disaggregate level. For example, brand managers may be interested in a comparison across product attributes. For such an analysis the number of relevant items in the product category make commonly used sales models difficult to use as they would contain too many parameters.

In this paper we propose a new model, which allows for the analysis of a market with many items while using only a moderate number of easily interpretable parameters. This is achieved by writing the sales model as a Hierarchical Bayes model. In this way we relate the marketing-mix effectiveness to item characteristics such as brand, package size, package type and shelf position. In this specification we do not have to impose restrictions on the competitive structure, as all items are allowed to have different own and cross elasticities. The parameters in the model are estimated using Markov Chain Monte Carlo techniques.

As a by-product this model allows to make predictions of sales levels and marketing-mix effectiveness of new to introduce items or of attribute changes. For example, one can assess the impact of changing the packaging from plastic to glass, on sales and price elasticity. Besides entering and changing products, our model also allows for items to leave the market.

We consider the representation, specification and estimation of the model. We apply the model to a ketchup scanner data set with 23 items at the chain level. Our results indicate that the model fits the sales of most items very well.

Keywords: sales models, attribute data, SKU level analysis, Hierarchical Bayes, Markov Chain Monte Carlo.

1 Introduction

Sales models are useful tools for analyzing competitive structures, see Wittink *et al.* (1988) among many others. These models are often used to analyze markets with fairly small numbers of items, for example, three national brands and one private label. In practice, the number of items in a product category can easily be very large, and hence the current sales models already concern somehow aggregate data. Suppose one does not want or can aggregate over items, and is aiming at using the same type of sales models for all items, it is easy to imagine that one may run into problems of implementation. A full-blown multivariable sales equation would yield a very large number of parameters when a complete model is analyzed. To reduce the number of parameters, one can assume restricted versions of the general model. For example, one can assume that the covariance matrix is restricted to a diagonal matrix, or that the sales of each brand depend only on its own explanatory variables measuring marketing-mix effects. To simplify matters even more, one can also assume that certain parameters are equal across brands, which entails that one pools parameters. Finally, the dynamics in the equations may be restricted, by removing item-specific effects or some lags in the sales equations.

It is important to understand, though, that if one restricts sales model in this sense, the number of parameters can still be large when analyzing many items. In this paper we address this issue by proposing a new approach, which allows for modeling many items, while keeping the number of parameters at a reasonable level. To this end, we propose a Hierarchical Bayes [HB] method which uses product attribute data, such as package size and type (like glass/plastic/tube), in the parameter equation. This enables us to reduce the number of parameters considerably, while still having item-specific elasticities. Our method also allows the analysis of the consequences of the introduction of a new item. One can vary across different attributes and see which have most potential in terms of sales.

We assume attribute data have an impact on the parameters in the sales equation using the following arguments. Suppose two items are completely the same on their attributes. In this case, there is no reason to assume that their effects of marketing-mix efforts would differ. Now, suppose that two items are not the same, but very similar, that is, they differ only on one attribute. In this case, such a difference may result in different marketing effectiveness parameters for these two items. The magnitude of the difference can then be assigned to the differences across attributes.

Our method improves on an often used approach in marketing research. This approach is that one first regresses sales on marketing instruments, and then regresses the estimated parameter values on product attributes. It is important to understand that this amounts to assuming that the parameter uncertainty in the first step is not taken into account in the second step, and this entails too much confidence in statistical inference. Our HB approach does not require this this assumption, as we take the two steps simultaneously. A by-product is that this also makes calculating confidence intervals for the attribute effect parameters more straightforward, as we will show below.

To our knowledge, there are two related studies in the literature. Fader and Hardie (1996) take a similar approach, but they consider individual level data. Furthermore, they assume the attributes only have impact on the intercepts in the sales model, instead of on all variables. Our approach is different in two respects. First, we use data which are aggregated at the market level. This type of data is more widely available and new data can quite easily be obtained, in contrast to panel scanner data. Furthermore, we allow the attributes to have impact on all marketing instrument parameters.

The second study is Foekens *et al.* (1997), who use market share models to better understand the competition between brands. In their model, consumers evaluate some attributes in a hierarchical way to make their brand choice. This evaluation is however assumed to be in a fixed order, whereas in our model there is no such sequence. Furthermore, in their approach, only few attributes can be evaluated, as it is assumed for a consumer, that it is not possible to evaluate multiple attributes in his/her own preferred order. In contrast, our model allows for a large number of attributes.

The outline of our paper is as follows. In Section 2, we discuss the specification of our sales model. In Section 3, we apply this model to the ketchup category. In this section we consider the impact of marketing instruments, forecasting of sales and the introduction of new items. In Section 4, we conclude our paper with a discussion of the main results.

2 A new model for sales

In this section we introduce our sales model. In Section 2.1, we discuss a simple variant of the model in which we assume that there are no cross effects of marketing instruments. In Section 2.2, we will show how these cross effects can be incorporated. Next, in Section 2.3 we discuss examining the effects of brand repositioning.

2.1 Basic model

First, we discuss notation. The number of items (SKU) in a market will be denoted by I, the number of marketing mix instruments by K, and the number of available attributes by L. Let $\ln s_{i,t}$ be the natural log of sales of SKU i at time $t = 1, ..., T$, and assume it can be respresented by

$$\ln s_{i,t} = \mu_i + x'_{i,t}\beta_i + \varepsilon_{i,t}, \tag{1}$$

where $x_{i,t}$ denotes a $K \times 1$ dimensional vector of explanatory variables for SKU i at time t and where $\beta_i = (\beta_{1,i}, \ldots, \beta_{K,i})'$ with $\beta_{k,i}$ the coefficient measuring the effect of the k-th explanatory variable for SKU i. The vector of explanatory variables will in general contain (log-transformed) marketing instruments and lagged log sales. The parameter μ_i is a SKU-specific intercept. All parameters in (1) are allowed to differ across items. Finally, let the error term $(\varepsilon_{1,t}, ...\varepsilon_{I,t})'$ be normally distributed with zero mean and Σ as covariance matrix.

In a typical application, (1) constitutes the entire model. However, for a market with many different items such a specification may contain too many parameters. The number of parameters may be too large from a substantive or from a statistical point of view. That is, there may be too many parameters to allow for a straightforward interpretation of the model and there may be too many parameters relative to the available data, creating a degrees of freedom problem. These considerations are even more relevant when extended versions of the sales model are considered, for example when considering cross-item effects of marketing efforts, as we will do in the next subsection.

To impose some structure on the parameters, one can use attribute data. In this case, an item is considered to be represented by a specific point in an attribute space. A (linear) mapping then relates the model parameters to this attribute space. This approach is also used when attribute data are not directly available, and in that case a latent attribute space is introduced. The locations in attribute space should then be inferred from the data, see, for example Chintagunta (1999). However, if available, the use of actual attribute data is to be preferred.

An often considered approach in empirical marketing research is to estimate the mapping from an (observed) attribute space to the parameter space from data using a two-step method. First, the competitive model in (1) is estimated using item-specific parameters β_i. In a second step the estimated item-specific parameters $\hat{\beta}_i$ are regressed on the attributes. In this approach, the estimated parameters from the first step are treated as fixed in the next round. There are several statistical and practical problems with this approach. First of all, this approach will break down if there are not enough data to estimate all the item-specific effects. Moreover, the second step ignores the uncertainty about the parameters. These parameters are treated as fixed, while in fact they are estimates. Especially when there are limited data, the uncertainty in these parameters may be large, and hence one might be too confident about the attribute effects in the second step.

Both these problems are solved by combining the two steps in a single approach. This approach accounts for the uncertainty in the SKU-level parameters by directly estimating the attribute space mapping. Furthermore, by explicitly recognizing that items that are close in attribute space will have similar parameters, we can make use of the data more efficiently.

Denote by Z_i an $L \times 1$ vector of the observed attributes of item i. We introduce the following linear relation between the item-specific parameters and the attribute space, that is,

$$\begin{pmatrix} \mu_i \\ \beta_i \end{pmatrix} = \alpha_0 + \alpha_1 Z_i + \eta_i, \tag{2}$$

with α_0 a $(K+1) \times 1$ vector and α_1 a $(K+1) \times L$ matrix of parameters. The coefficients $\alpha_{1,k,l}$ now represent the effect of attribute l on the effectiveness of marketing instrument k. If $k = 0$ the coefficient gives the effect on the intercept. Of course, there may be attributes that we do not observe or intangible attributes such as brand equity that also influence the marketing-instrument effectiveness. To account for that, we represent the joint effect of such attributes by a normally distributed disturbance term, that is, $\eta_i = (\eta_{i,0}, \eta_{i,1}, \ldots, \eta_{i,K})' \sim N(0, \Sigma_\eta)$. The size of uncertainty may differ across instruments, and we therefore allow the variance of $\eta_{i,k}$ to depend on k. Furthermore, we may expect that unobserved attributes affect multiple marketing instruments. For example, if an item has a high feature effectiveness it may also be effective with display. Such relations will lead to positive correlations between $\eta_{i,k}$ and $\eta_{i,h}$. To capture these correlations we allow Σ_η to be non-diagonal.

In sum, the combination of (1) and (2) gives our attribute-based sales model, which is particularly useful in case I is large. Note that, technically speaking, μ_i and β_i are not the key parameters in this model. Indeed, by substituting (2) in (1) we would have a model that does not contain μ_i and β_i. In fact, the only parameters are $\alpha_0, \alpha_1, \Sigma_\eta$ and Σ_ε. Therefore, our model contains less parameters than a model in which β_i is estimated directly. Instead of having to estimate $(K+1)I$ β-parameters, we now only have to estimate $(L+1)(K+1)$ parameters. For large I and with $L+1 < I$, this will of course give a considerable reduction in the number of parameters. Finally, the joint estimation of (1) and (2) gives more precise estimates of the attribute mapping than a two-step approach as it combines all the available information and accounts for uncertainty in the marketing instrument effectiveness.

Although only α_0 and α_1 can be seen as the "true" parameters of the model, we can still get estimates of the marketing instrument effectiveness from the model and the data. Without observing actual sales, (2) gives the prior beliefs about the effects of marketing efforts. Conditional on observed sales, we can update these beliefs to reflect the data. In the Appendix, we discuss an MCMC algorithm that can be used to estimate the model parameters and which, as a by-product, gives draws from the distribution of μ and β conditional on the data.

2.2 Including cross effects

In many practical cases, model (1) could be considered to be too restrictive. For example, in this model it is assumed that the marketing efforts of one SKU do not affect the sales level of competing SKUs. Although a very restrictive assumption, it is often imposed in practical applications, since adding all cross effects to (1) would entail many parameters. A model with all cross-effects in general has $I + KI^2$ parameters, instead of $(K+1)I$ in (1). Consider a market with 25 items and only 3 marketing instruments. A model with all cross effects has $25 + 3 \times 25^2 = 1900$ parameters, while the corresponding model without cross effects only has $4 \times 25 = 100$ parameters. A reasonably sized data base will suffice for the estimation of 100 parameters, but the amount of parameters involved in the most flexible cross-effect model will be more than most available datasets can handle. Indeed, such a data set would have to contain data on a very large number of observations.

It should be evident that the use of attribute data to summarize the competitive structure is especially relevant in case one wants to model cross-item effects of marketing instruments. Including all possible cross effects, the general sales model in (1) becomes

$$\ln s_{i,t} = \mu_i + x'_{i,t}\beta_i + \sum_{\substack{j=1 \\ j \neq i}}^{I} x'_{j,t}\beta_{j,i} + \varepsilon_{i,t}. \tag{3}$$

Denoting by $\beta_{j,i}$ a $K \times 1$ vector representing the effect of item j's marketing instruments on the sales of item i, we now specify the attribute-to-parameter mapping for the cross effects as

$$\beta_{j,i} \quad = \quad \delta + \kappa Z_i + \lambda Z_j + \gamma |Z_i - Z_j| + \xi_{j,i}, \tag{4}$$

where δ is a $K \times 1$ vector, and where κ, λ and γ are $K \times L$ matrices. We will denote $\theta = (\delta, \kappa, \lambda, \gamma)$, which is a $K \times (1 + 3L)$ matrix . The number of parameters to estimate therefore equals $K(1 + 3L)$. The effects of unobserved attributes on the cross effects are captured by $\xi_{j,i}$, for which we again assume that these effects have a normal distribution, now with covariance matrix Σ_ξ. The parameter estimation of the model including (3) and (4) follows along the lines as outlined in the appendix.

The explained part of the cross effects is assumed to depend on the attribute values of both items, that is Z_i and Z_j. Note that the cross effects are not necessarily symmetric, that is in general $E(\beta_{j,i}) \neq E(\beta_{i,j})$. The effect of attributes of SKU i on j only equals the reverse effect if $\kappa = \lambda$. We expect $\gamma < 0$, as we believe that more distant items in the attribute space will have smaller cross elasticities.

To illustrate the interpretation of the parameters in (4), consider a binary attribute denoting the brand of ketchup. This attribute equals 1 if the particular ketchup item is "Heinz" and 0 otherwise. The cross effect of item j on item i now depends on the brand of both items, that is

$j\backslash i$	Brand = 1	Brand = 0
Brand = 1	$\delta+\kappa+\lambda$	$\delta+\lambda+\gamma$
Brand = 0	$\delta+\kappa+\gamma$	δ

Hence, δ measures the "baseline" cross effect, κ and δ the effect of a "Heinz" item and γ measures the change in the cross effect for items that differ on this attribute. Note that the four parameters exactly match the four different cases for the brand attribute. In case of binary attributes, the specification in (4) is the most flexible. Higher order functions of the attributes can be added to (4), in order to model the dependence of the marketing effectiveness on attributes that are measured on a continuous scale.

2.3 Brand repositioning

The store layout may be part of the attribute space. For example, consider the positioning of items on the shelves in the store. Items on the top or bottom shelf may be less attractive to consumers compared to items at the eye level. Furthermore, items that are (physically) close to each other may compete more intensely and therefore have stronger cross elasticities.

Including the store layout in the attribute space leads to some difficulties. Although the store layout is likely to be constant during large periods of time, it does change occasionally. If we assume that the attribute location of an item determines the marketing instrument elasticities, a change in store layout will lead to a change in some of the parameters concerning competition. Note that given the parameters in (4) we can easily predict the effects of such a change.

We do not have to change the model to deal with changes in the attributes, as we will show now. With a change in the attributes at time T_1, $1 < T_1 < T$, the model becomes

$$
\begin{aligned}
\ln s_{i,t} &= \begin{cases} \mu_i^1 + \ln x'_{i,t}\beta_i^1 + \sum_{j=1,j\neq i}^{I} \ln x'_{j,t}\beta_{j,i}^1 + \varepsilon_{i,t} & \text{if } 1 \leq t < T_1 \\ \mu_i^2 + \ln x'_{i,t}\beta_i^2 + \sum_{j=1,j\neq i}^{I} \ln x'_{j,t}\beta_{j,i}^2 + \varepsilon_{i,t} & \text{if } T_1 \leq t \leq T, \end{cases} \\
\beta_{j,i}^1 &= \delta + \kappa Z_i^1 + \lambda Z_j^1 + \gamma |Z_i^1 - Z_j^1| + \xi_{j,i}^1 \\
\beta_{j,i}^2 &= \delta + \kappa Z_i^2 + \lambda Z_j^2 + \gamma |Z_i^2 - Z_j^2| + \xi_{j,i}^2 \\
\begin{pmatrix} \mu_i^1 \\ \beta_i^1 \end{pmatrix} &= \alpha_0 + \alpha_1 Z_i^1 + \eta_i^1 \\
\begin{pmatrix} \mu_i^2 \\ \beta_i^2 \end{pmatrix} &= \alpha_0 + \alpha_1 Z_i^2 + \eta_i^2
\end{aligned} \tag{5}
$$

where Z^1 denotes the attributes before T_1 and Z^2 the attributes after the change, $\xi^r_{j,i}$ and η^r_i are independent and have normal distributions, $i, j = 1, \ldots, I$ and $r = 1, 2$, where the distribution is independent of r .

Clearly, to allow for changes in the parameters due to the attributes in the classical setting, one would have to double the number of parameters, while in our specification the number of parameters to estimate does not change at all. This is an important advantage of our model.

3 An illustration

To illustrate our method, we present a detailed analysis of an interesting dataset. In Section 3.1 we briefly describe the data. Sections 3.2 and 3.3 concern the estimation results and the forecasting performance, respectively. In Section 3.4, we illustrate how our model can be used to predict the performance of new items and of attribute changes.

3.1 Data description

We analyze a dataset with 23 items from a ketchup market. We have sales data for 89 weeks and percentages of stores engaging in display, featuring, promotions and distribution. For each item, we have information on a number of attributes. These are subcategory (curry or tomato), brand, number of facings, shelf position and package size. Table 1 gives some descriptive statistics for our dataset. About halfway the sample period, a change in the category occurs. Four new items enter the market and five items exit the market. The store layout also changes, that is, many items change on the attributes shelf and facings. As described in Section 2.3, our model can easily handle such changes.

3.2 Estimation results

As explanatory variables, we use feature, promotion, distribution and 1-period lagged sales. The variable display, despite the fact that it is 'display only', showed too much correlation with the promotion variable, and is therefore omitted from the analysis. Because of the lagged sales variable, we use the first observation as a starting value and hence end up with 88 observations. For the estimation of the parameters of our HB model, we generate

4000 iterations of the Gibbs sampler for burn in and 8000 iterations for analysis, where we retain every eighth draw. The (unreported) iteration plots are inspected to see whether the sampler converges to stationary draws from the posterior distributions of the model parameters.

Table 2 shows the parameter estimates for α, that is the parameters linking the attributes to the effectiveness of own marketing instruments. From these estimates, it can be seen that the number of facings has a strong positive influence on the constant in the sales equation. This indicates that items that have many facings, have a higher expected sales level when there is no promotion, feature activity, or otherwise. There could be some feedback effects here as well. An item with a high sales level will be granted more shelf space in the store, and thus result in more facings for the item. This feedback effect is not analyzed in our model, since this would require including facings as an endogenous variable. Furthermore, in our dataset the number of facings changes only once, which makes it very hard to analyze the reverse causality. Facings also has a positive influence on the effectiveness of featuring. Items with the Heinz brand name are also more successful with featuring.

Values for $\theta = (\delta, \kappa, \lambda, \gamma)$ are presented in Table 3. Most significant cross effects occur for the instrument feature. To illustrate the interpretation of these parameter estimates, consider the "subcategory" coefficients for feature (own: 0.32, competitor: -2.35 and difference: -0.81). The interpretation of these three coefficients is as follows. Consider two items both having subcategory=1, that is, they are tomato ketchup items. For these two items there will be a strong negative cross effect, as the contribution of the subcategory attribute is $0.32 - 2.35 = -2.33$. For two items with subcategory=0, that is, two curry items, there is a much smaller cross effect, since the contribution of this attribute is only 0.32. Finally, the cross effect between different subcategories is -2.35 - 0.81 = -3.16 for the effect of subcategory=1 on subcategory=0 or 0.32 - 0.81 = -0.49 for subcategory=0 on subcategory=1.

For β_i and β_{ji}, we also report values averaged over items in Table 4, as the dimensions are perhaps too large for interpretation. Notice that the change in the store and entry and exit of the items has most effect on the feature instrument. To give some more insight into the values of β_i, the graphs in Figure 1 display the posterior means per item. The top line is the posterior mean plus two times the standard deviation and the bottom line displays the posterior mean minus two times the standard deviation.

3.3 Forecasting results

In each run of the Gibbs sampler, we simulate a sales forecast for the last 10 periods in the dataset. We use the posterior mode of these forecasts as the out-of-sample prediction. The Mean Average Percentage Error for both in sample fit and out of sample forecast are reported in Table 5. When we compare our HB model with a model which concerns a regression per item, we see that our model performs better, both in sample and out of sample.

3.4 Introducing a new item

As a by-product, our model allows to make predictions of sales levels and marketing-mix effectiveness of new to introduce items and of attribute changes. To illustrate the former, we estimate the model for only 22 of the 23 items. In each iteration of the Gibbs sampler, we use the current values for α and θ to generate values for $\mu_{\text{intro}}, \beta_{\text{intro}}$ and $\beta_{-\text{intro,intro}}$. The sampling is performed using equations (2) and (4) in the following way:

$$\begin{aligned} (\mu_{\text{intro}}, \beta'_{\text{intro}})' &\sim N((\alpha_0, \alpha_1)\left(\begin{smallmatrix}1\\ Z_i\end{smallmatrix}\right), \Sigma_\eta) \\ \beta_{j,\text{intro}} &\sim N((\delta, \kappa, \lambda, \gamma)W_{j,i}, \Sigma_\xi), j \neq \text{intro} \end{aligned} \tag{6}$$

Given the values for $\mu_{\text{intro}}, \beta_{\text{intro}}$ and $\beta_{-\text{intro,intro}}$, we can now generate values for y_{intro}, using (3). The variance of $\varepsilon_{\text{intro}}$ is taken as the average of the other 22 item-specific error variances. In each (stored) iteration of the Gibbs sampler, we simulate sales for the in sample period, as well as for the out-of-sample period. We perform this introduction exercise for each of the 23 items in the market. The MAPE results can be found in Table 6. The first column of forecasts, that is, the "in sample" forecasts, can also be seen as a cross-validation check, while the second column, that is, the "out of sample" forecasts, concerns genuine item introduction. Obviously, the forecast errors are much larger now, but for some items our model gives quite accurate forecasts of these sales levels.

4 Conclusion

In this paper we have shown that it is possible to analyze a parsimonious sales model that describes many items. It appeared even possible to model cross-effects between the items, without having large numbers of parameters. Moreover, with our model, one can gain insight in the impact that attributes can have on the effectiveness of marketing instruments. Also, our model could incorporate store layout changes, without the need to introduce even more parameters. We have shown that the model performs considerably better than an item-specific regression. Finally, our model was able to predict the introduction of a new item reasonably well.

The limitations of our study immediately suggest further avenues for research. Our dataset did not include pricing information and further applications would benefit from a more extensive data set. Furthermore, it would be insightful to compare our "new-item" exercise with alternative methods, for example by using the 2-step-estimation procedure. This would require the development of new evaluation criteria. Finally, it would be interesting to examine why sales of some items can be better predicted than those of others, and whether these findings could be generalized across product categories.

Appendix: Parameter estimation

This appendix describes the algorithm for sampling from a Markov Chain that has the posterior distribution of the model parameters as its stationary distribution. In this appendix we will denote $y_{i,t} = \ln s_{i,t}$, $W_{ji} = \begin{pmatrix} 1 \\ Z_i \\ Z_j \\ |Z_i - Z_j| \end{pmatrix}$ and use $W_{-i,i}$ for the matrix obtained by horizontally concatenating $W_{j,i}$ for $j \in \{1, \dots, I\} \backslash i$. Finally, the matrix X_{-i} gives the matrix consisting of horizontally stacked X_j for $j \in \{1, \dots, I\} \backslash i$.

Sampling of $\alpha_0,\ \alpha_1 | \mu, \beta, \Sigma_\eta$

Rewrite (2) as

$$\begin{pmatrix} \mu_1 & \mu_2 & \dots & \mu_I \\ \beta_1 & \beta_2 & \dots & \beta_I \end{pmatrix} = (\alpha_0, \alpha_1) \begin{pmatrix} 1 & 1 & \dots & 1 \\ Z_1 & Z_2 & \dots & Z_I \end{pmatrix} + \begin{pmatrix} \eta_1 & \eta_2 & \dots & \eta_I \end{pmatrix}. \tag{.1}$$

Writing this model as $\omega = \alpha Z + \eta$, we obtain

$$\text{vec}(\omega) = (Z' \otimes \mathbf{I}_I)\text{vec}(\alpha) + \text{vec}(\eta), \tag{.2}$$

where $\text{vec}(\eta) \sim N(0, \mathbf{I}_I \otimes \Sigma_\eta)$. The posterior distribution of α_0, α_1 conditional on μ, β and Σ_η is therefore normal, where the mean and the variance follow from the OLS estimator of $\text{vec}(\alpha)$ in (.2), that is

$$\begin{aligned} \text{E}(\text{vec}(\alpha)|\mu, \beta, \Sigma_\eta) &= [(Z \otimes \mathbf{I}_I)(\mathbf{I}_I \otimes \Sigma_\eta^{-1})(Z' \otimes \mathbf{I}_I)]^{-1}(Z \otimes \mathbf{I}_I)(\mathbf{I}_I \otimes \Sigma_\eta^{-1})\text{vec}(\omega) \\ &= \text{vec}(\omega Z'(ZZ')^{-1}) \\ \text{Var}(\text{vec}(\alpha)|\mu, \beta, \Sigma_\eta) &= [(Z \otimes \mathbf{I}_I)(\mathbf{I}_I \otimes \Sigma_\eta^{-1})(Z' \otimes \mathbf{I}_I)]^{-1} \\ &= [(ZZ')^{-1} \otimes \Sigma_\eta]. \end{aligned} \tag{.3}$$

Sampling of $(\delta, \kappa, \lambda, \gamma)|\beta_{-i,i}\ (i = 1, \dots, I), \Sigma_\xi$

The model specifies

$$(\beta_{-1,1}, \dots, \beta_{-I,I}) = (\delta, \kappa, \lambda, \gamma)(W_{-1,1}, \dots, W_{-I,I}) + (\xi_{-1,1}, \dots, \xi_{-I,I}). \tag{.4}$$

Denoting this as $\beta_{\text{cross}} = (\delta, \kappa, \lambda, \gamma)W + \xi$, with $\text{vec}(\xi) \sim N(0, \mathbf{I}_{I(I-1)} \otimes \Sigma_\xi)$, we obtain analogous to (.3) that the conditional posterior distribution of $\text{vec}(\delta, \kappa, \lambda, \gamma)$ is normal with

$$\begin{aligned} \text{E}(\text{vec}(\delta, \kappa, \lambda, \gamma)|\beta_{\text{cross}}, \Sigma_\xi) &= \text{vec}(\beta_{\text{cross}} W'(WW')^{-1}) \\ \text{Var}(\text{vec}(\delta, \kappa, \lambda, \gamma)|\beta_{\text{cross}}, \Sigma_\xi) &= [(WW')^{-1} \otimes \Sigma_\xi]. \end{aligned} \tag{.5}$$

Sampling of $\mu_i, \beta_i, \beta_{-i,i} | y_i, \alpha_0, \alpha_1, \delta, \kappa, \lambda, \gamma, \Sigma_\varepsilon, \Sigma_\eta$

Denoting $y_i = (y_{i1}, y_{i2}, \dots, y_{i,T})'$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$ we have

$$\begin{aligned} y_i &= (1, X_i, X_{-i}) \begin{pmatrix} \mu_i \\ \beta_i \\ \mathrm{vec}(\beta_{-i,i}) \end{pmatrix} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2_{\varepsilon_i} \mathbf{I}_T) \\ \begin{pmatrix} \mu_i \\ \beta_i \end{pmatrix} &= (\alpha_0, \alpha_1) \begin{pmatrix} 1 \\ Z_i \end{pmatrix} + \eta_i, \quad \eta_i \sim N(0, \Sigma_\eta) \\ \beta_{-i,i} &= (\delta, \kappa, \lambda, \gamma) W_{-i,i} + \xi_{-i,i}, \quad \mathrm{vec}(\xi_{-i,i}) \sim N(0, \mathbf{I}_{I-1} \otimes \Sigma_\xi) \end{aligned} \tag{.6}$$

Normalizing equations and collecting them, we get

$$\begin{pmatrix} \frac{1}{\sigma_{\varepsilon_i}} y_i \\ \Sigma_\eta^{-1/2} (\alpha_0, \alpha_1) \left(\begin{smallmatrix} 1 \\ Z_i \end{smallmatrix} \right) \\ \mathrm{vec}(\Sigma_\xi^{-1/2} (\delta, \kappa, \lambda, \gamma) W_{-i,i}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_{\varepsilon_i}} (1, X_i) & \frac{1}{\sigma_{\varepsilon_i}} X_{-i} \\ \Sigma_\eta^{-1/2} & 0 \\ 0 & \mathbf{I}_{I-1} \otimes \Sigma_\xi^{-1/2} \end{pmatrix} \begin{pmatrix} \mu_i \\ \beta_i \\ \mathrm{vec}(\beta_{-i,i}) \end{pmatrix} + \zeta, \tag{.7}$$

where $\zeta \sim N(0, \mathbf{I}_{1+KI})$. Again the conditional posterior distribution of $(\mu_i, \beta_i', \mathrm{vec}(\beta_{-i,i})')'$ is normal with mean $(\tilde{X}_i' \tilde{X}_i)^{-1} \tilde{X}_i' \tilde{y}_i$ and variance $(\tilde{X}_i' \tilde{X}_i)^{-1}$, where

$$\begin{aligned} \tilde{X}_i' \tilde{X}_i &= \begin{pmatrix} \frac{1}{\sigma^2_{\varepsilon_i}} \left(\begin{smallmatrix} T & X \\ X' & X'X \end{smallmatrix} \right) + \Sigma_\eta^{-1} & \frac{1}{\sigma^2_{\varepsilon_i}} \left(\begin{smallmatrix} X_{-i} \\ X'X_{-i} \end{smallmatrix} \right) \\ \frac{1}{\sigma^2_{\varepsilon_i}} \left(\begin{smallmatrix} X'_{-i} & X'_{-i} X \end{smallmatrix} \right) & \frac{1}{\sigma^2_{\varepsilon_i}} (X'_{-i} X_{-i}) + (\mathbf{I}_{I-1} \otimes \Sigma_\xi^{-1}) \end{pmatrix} \\ \tilde{X}_i' \tilde{y}_i &= \begin{pmatrix} \frac{1}{\sigma^2_{\varepsilon_i}} \left(\begin{smallmatrix} 1 \\ X' \end{smallmatrix} \right) y_i + \Sigma_\eta^{-1} (\alpha_0, \alpha_1) \left(\begin{smallmatrix} 1 \\ Z_i \end{smallmatrix} \right) \\ \frac{1}{\sigma^2_{\varepsilon_i}} X'_{-i} y_i + \mathrm{vec}(\Sigma_\xi^{-1} (\delta, \kappa, \lambda, \gamma) W_{-i,i}) \end{pmatrix}. \end{aligned} \tag{.8}$$

Sampling of $\Sigma_\varepsilon | y_i, \mu_i, \beta_i, \beta_{-i,i}$ $(i = 1, \dots, I)$

Given $y_i, \mu_i, \beta_i, \beta_{-i,i}$ $(i = 1, \dots, I)$ the disturbances are obtained using

$$\varepsilon_i = y_i - (1, X_i, X_{-i}) \begin{pmatrix} \mu_i \\ \beta_i \\ \mathrm{vec}(\beta_{-i,i}) \end{pmatrix}.$$

As we assume the disturbances to be independent, the posterior distribution of $\sigma^2_{\eta_i}$ is inverted $\chi^2(T)$ that is, a draw from the posterior conditional distribution can be obtained by drawing R from $\chi^2(T)$ and calculating $\varepsilon_i' \varepsilon_i / R$.

Sampling of $\Sigma_\eta | \alpha_0, \alpha_1, \mu, \beta$

For the case of a possibly non-diagonal covariance matrix for η, previous research has shown that the performance of the Markov Chain can be significantly improved by incorporating some prior information on the covariance matrix. In this paper, we use an informative but diffuse inverted Wishart prior, that is $\Sigma_\eta \sim IW(S_\eta, \nu_\eta)$. The conditional posterior distribution is $\Sigma_\eta | \alpha_0, \alpha_1, \mu, \beta \sim IW(\eta'\eta + S_\eta, I + \nu_\eta)$, where $\eta = \omega - \alpha Z$.

Sampling of $\Sigma_\xi | \delta, \kappa, \lambda, \gamma, \beta_{-i,i}$ $(i = 1, \ldots, I)$

For Σ_ξ we also use the inverted Wishart prior, that is $\Sigma_\xi \sim IW(S_\xi, \nu_\xi)$. The conditional posterior is $\Sigma_\xi | \delta, \kappa, \lambda, \gamma, \beta_{-i,i}$ $(i = 1, \ldots, I) \sim IW(\xi'\xi + S_\xi, I(I-1) + \nu_\xi)$, where $\xi = (\xi_{-1,1}, \ldots, \xi_{-I,I})$ with $\xi_{-i,i} = \beta_{-i,i} - (\delta, \kappa, \lambda, \gamma) W_{-i,i}$

Table 1: Descriptive statistics for ketchup dataset

Attribute	**Values**
Subcategory	curry (7 items) or tomato (16 items)
Brand	Heinz (11 items) or other (12 items) [a]
Number of facings	1 to 5
Shelf position	1 to 7
Package size	200 to 875 ml
Number of items	23
Number of weeks	89

[a] We have data for more brands, but for illustration purposes, we focus on Heinz versus other brands. Furthermore, other brands have few items each, which could make estimation of brand effects more difficult.

Table 2: Parameter estimates for α

attribute	**const**	**feature**	**promotion**	**distribution**	**lagged sales**
const	$-2.92*$	-0.16	$2.21*$	$3.47*$	0.19
subcategory	1.56	1.08	-3.13	5.27	0.46
facings	$5.53*$	$12.00*$	$-2.35*$	-1.95	-0.40
shelf	2.73	-2.98	$-1.27*$	-1.13	$0.31*$
package size	2.90	-3.05	$0.88*$	$-2.28*$	$-0.38*$
brand	-0.71	$3.09*$	$-0.77*$	0.07	0.01

Table 3: Parameter estimates for $\theta = (\delta, \kappa, \lambda, \gamma)$

	attribute	**feature**	**promotion**	**distribution**	**lagged sales**
	const	-0.16	-0.07	0.36*	0.00
own	**subcategory**	0.32	0.04	-0.83	0.10
	facings	0.15	-0.07	0.02	-0.02
	shelf	0.09	-0.01	-0.15	-0.01
	package size	0.24	-0.02	0.04	-0.01
	brand	0.17*	0.02	0.02	0.01
competitor	**subcategory**	-2.35*	0.14	-0.54	-0.02
	facings	0.85*	-0.11	0.38	-0.07
	shelf	0.23	0.11	-0.21	0.04
	package size	-0.30	-0.07	-0.15	0.01
	brand	0.17*	0.03*	-0.11	0.00
difference	**subcategory**	-0.81	0.23	-0.49	0.09
	facings	-0.17	0.14*	-0.05	-0.03
	shelf	0.33	0.03	-0.06	0.00
	package size	-0.01	0.06	0.04	-0.01
	brand	-0.21*	0.01	-0.01	-0.01

Table 4: Average parameter estimates for β

	constant	**feature**	**promotion**	**distribution**	**lagged sales**
β_i^1	0.704	1.2	1.09	1.75	0.0737
β_i^2	0.515	1.46	1.06	1.9	0.0482
$\beta_{j,i}^1$		0.105	-0.0151	0.0599	-0.00361
$\beta_{j,i}^2$		0.0782	-0.0259	0.0901	-0.00563

Table 5: MAPE (Mean Average Percentage Error) for two models for ketchup items

			HB		Regression per item	
subcat.	**item**	**description**	**i.s.**	**o.o.s.**[a]	**i.s.**[a]	**o.o.s.**[a]
Curry	1	Others 750	6%	9%	9%	11%
	2	Calve 500	5%	11%	12%	10%
	3	Gouda's Gloria 750	9%	8%	11%	14%
	4	Heinz 450	5%	8%	9%	6%
	5	Heinz 750	6%	10%	7%	13%
	6	Hela 800	5%	6%	7%	7%
	7	Remia 500	7%	19%	10%	22%
Tomato	8	Others 300	23%	[b]	24%	[b]
	9	Calve 500	6%	7%	14%	13%
	10	Private label 300	9%	17%	11%	10%
	11	Private label 450	6%	9%	10%	11%
	12	Gouda's Gloria 750	5%	16%	9%	13%
	13	Heinz hot 450	9%	[b]	11%	[b]
	14	Heinz hot 500	5%	8%	6%	10%
	15	Heinz 300 glass	7%	10%	9%	6%
	16	Heinz 450 glass	7%	[b]	11%	[b]
	17	Heinz 500 glass	6%	17%	8%	201%
	18	Heinz 400 plastic	9%	12%	14%	16%
	19	Heinz 500 plastic	16%	[b]	17%	[b]
	20	Heinz 875 plastic	7%	7%	8%	7%
	21	Heinz 200 tube	11%	13%	10%	13%
	22	Remia 440	8%	[b]	12%	[b]
	23	Remia 500	7%	20%	13%	19%
		average	8%	12%	11%	22%

[a] i.s.: in sample, o.o.s.: out of sample.

[b] For this item no out of sample forecast can be calculated, since it exits the market in period 44.

Table 6: MAPE per ketchup item introduced

subcat.	**item**	**description**	i.s.[a]	o.o.s.[a]
Curry	1	Others 750	34%	43%
	2	Calve 500	20%	17%
	3	Gouda's Gloria 750	45%	66%
	4	Heinz 450	33%	33%
	5	Heinz 750	59%	51%
	6	Hela 800	35%	28%
	7	Remia 500	19%	18%
Tomato	8	Others 300	77%	[b]
	9	Calve 500	22%	24%
	10	Private label 300	45%	81%
	11	Private label 450	66%	65%
	12	Gouda's Gloria 750	13%	13%
	13	Heinz hot 450	64%	[b]
	14	Heinz hot 500	17%	11%
	15	Heinz 300 glass	41%	48%
	16	Heinz 450 glass	55%	[b]
	17	Heinz 500 glass	46%	38%
	18	Heinz 400 plastic	54%	63%
	19	Heinz 500 plastic	77%	[b]
	20	Heinz 875 plastic	44%	52%
	21	Heinz 200 tube	32%	46%
	22	Remia 440	54%	[b]
	23	Remia 500	19%	27%
		average	42%	40%

[a] i.s.: in sample, o.o.s.: out of sample.

[b] For this item no out of sample forecast can be calculated, since it exits the market in period 44.

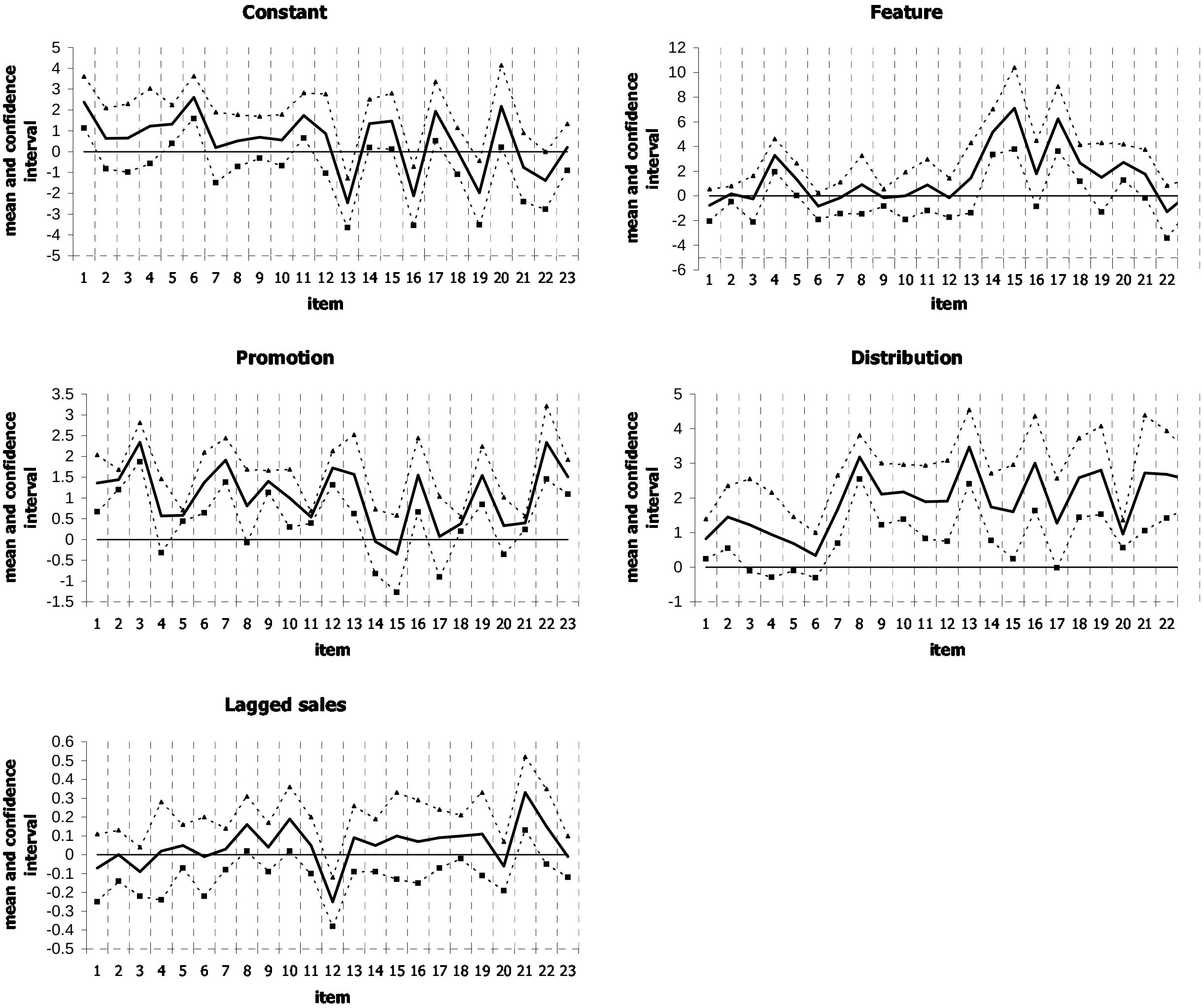

Figure 1: Posterior means and confidence bounds per β_i, for each of the five explanatory variables.

References

Chintagunta, P. K. (1999), Measuring the Effects of New Brand Introduction on Inter-Brand Strategic Interaction, *European Journal of Operational Research*, **118**, 315–331.

Fader, P. S. and B. G. Hardie (1996), Modeling Consumer Choice Among SKUs, *Journal of Marketing Research*, **33**, 442–452.

Foekens, E. W., P. S. H. Leeflang, and D. R. Wittink (1997), Hierarchical Versus Other Market Share Models for Markets with Many Items, *International Journal of Research in Marketing*, **14**, 359–378.

Wittink, D. R., M. J. Addona, W. J. Hawkes, and J. C. Porter (1988), SCAN-PRO: The Estimation, Validation and Use of Promotional Effects Based on Scanner Data, Tech. rep., Cornell University Johnson Graduate School of Management.

Publications in the Report Series Research* in Management

ERIM Research Program: “Marketing”

2002

Suboptimality of Sales Promotions and Improvement through Channel Coordination
Berend Wierenga & Han Soethoudt
ERS-2002-10-MKT

The Role of Schema Salience in Ad Processing and Evaluation
Joost Loef, Gerrit Antonides & W. Fred van Raaij
ERS-2002-15-MKT

The Shape of Utility Functions and Organizational Behavior
Joost M.E. Pennings & Ale Smidts
ERS-2002-18-MKT

Competitive Reactions and the Cross-Sales Effects of Advertising and Promotion
Jan-Benedict E.M. Steenkamp, Vincent R. Nijs, Dominique M. Hanssens & Marnik G. Dekimpe
ERS-2002-20-MKT

Do promotions benefit manufacturers, retailers or both?
Shuba Srinivasan, Koen Pauwels, Dominique M. Hanssens & Marnik G. Dekimpe
ERS-2002-21-MKT

How cannibalistic is the internet channel?
Barbara Deleersnyder, Inge Geyskens, Katrijn Gielens & Marnik G. Dekimpe
ERS-2002-22-MKT

Evaluating Direct Marketing Campaigns; Recent Findings and Future Research Topics
Jedid-Jah Jonker, Philip Hans Franses & Nanda Piersma
ERS-2002-26-MKT

The Joint Effect of Relationship Perceptions, Loyalty Program and Direct Mailings on Customer Share Development
Peter C. Verhoef
ERS-2002-27-MKT

Estimated parameters do not get the “wrong sign” due to collinearity across included variables
Philip Hans Franses & Christiaan Hey
ERS-2002-31-MKT

Dynamic Effects of Trust and Cognitive Social Structures on Information Transfer Relationships
David Dekker, David Krackhardt & Philip Hans Franses
ERS-2002-33-MKT

Means-end relations: hierarchies or networks? An inquiry into the (a)symmetry of means-end relations.
Johan van Rekom & Berend Wierenga
ERS-2002-36-MKT

---

* A complete overview of the ERIM Report Series Research in Management:
http://www.ers.erim.eur.nl

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship

Cognitive and Affective Consequences of Two Types of Incongruent Advertising
Joost Loef & Peeter W.J. Verlegh
ERS-2002-42-MKT

The Effects of Self-Reinforcing Mechanisms on Firm Performance
Erik den Hartigh, Fred Langerak & Harry R. Commandeur
ERS-2002-46-MKT

Modeling Generational Transitions from Aggregate Data
Philip Hans Franses & Stefan Stremersch
ERS-2002-49-MKT

Sales Models For Many Items Using Attribute Data
Erjen Nierop, Dennis Fok, Philip Hans Franses
ERS-2002-65-MKT

The Econometrics Of The Bass Diffusion Model
H. Peter Boswijk, Philip Hans Franses
ERS-2002-66-MKT

How the Impact of Integration of Marketing and R&D Differs Depending on a Firm's Resources and its Strategic Scope
Mark A.A.M. Leenders, Berend Wierenga
ERS-2002-68-MKT