Harald Hruschka *

Relevance of Functional Flexibility for Heterogeneous Sales Response Models

A Comparison of Parametric and Seminonparametric Models

Nr. 394

January 2004

*Dr. Harald Hruschka, Chaired Professor of Marketing, Faculty of Economics, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, Germany

Abstract

So far studies estimating sales response functions on the basis of store-specific data either consider heterogeneity or functional flexibility. That is why in this contribution a model is developed possessing both these features. It is a multilayer perceptron with store-specific coefficients which is specified in a hierarchical Bayesian framework. An appropriate Markov Chain Monte Carlo estimation technique is introduced capable to satisfy theoretical constraints (e.g. sign constraints on elasticities). The empirical study refers to a data base consisting of weekly observations of sales and prices for nine leading brands of a packaged consumer good category. The data were acquired in 81 stores over a time span of at least 61 weeks. The multilayer perceptron is compared to a strict parametric multiplicative model and approaches the maximum value of posterior model probability. This indicates the benefits of using a flexible model even if heterogeneity is dealt with. Estimated sales curves and elasticities demonstrate that both models differ in their implications about price response.[1]

Keywords: Sales Response; Hierarchical Bayes; Multilayer Perceptron; Neural Networks; Marketing

1 Introduction

So far studies estimating sales response functions on the basis of store-specific data can be divided into two groups. Studies belonging to the first group allow heterogeneity across stores, i.e. store-specific coefficients, but assume strict parametric forms, mostly linear, multiplicative or exponential (e.g. Blattberg and George, 1991; Montgomery, 1997; Boatwright et al., 1999). They specify sales response functions as hierarchical Bayesian models so that store-specific coefficients depend on data from all stores (Carlin and Louis, 1996). Markov

[1]This research is supported by a grant of the Deutsche Forschungsgemeinschaft (DFG).

Chain Monte Carlo (MCMC) simulation techniques serve to estimate parameters.

These studies demonstrate the importance of the heterogeneity across stores. Therefore models based on linearely aggregated or pooled data should be avoided as they are known to be biased if effects are heterogeneous even for the case of the linear function (Krishnamurhti et al., 1990). Their main potential weakness lies in the assumed strict parametric form. Though they deal with heterogeneity, if the assumed parametric form differs from the true function a source of bias remains (Hastie and Tibshirani, 1995).

The second group of studies estimating sales response functions allow flexibility of functional form by means of semiparametric or seminonparametric methods, but assume homogeneity across stores. The sales response model of Kalyanam and Shively (1998) uses cubic stochastic splines. The semiparametric models of van Heerde (1999) and van Heerde et al. (2001) include multivariate kernels. Hruschka (2000) and van Heerde et al. (2001) analyze sales response by means of a generalized additive model, whose nonparametric part equals a sum of cubic smoothing splines and univariate kernels, respectively. Lang et al. (2003) introduce a Bayesian semiparametric additive model which consists of P-splines with smoothness priors.

Multilayer perceptrons (MLPs) constitute another flexible modeling approach. They can be understood as series estimators approximating an unknown multivariate function by a linear combination of sigmoid (mostly logistic) functions (Pagan and Ullah, 1999). As series estimators are nonparametric in their orientation, though their modus operandi is parametric, MLPs can be called seminonparametric following the terminology of Gallant (1987). The literature contains some examples of determining market response functions by MLPs. Hruschka (1993) estimates an aggregate sales response model by a MLP. A related stream of research deals with estimation of aggregate market share models (van Wezel and Baets, 1995; Wierenga and Kluytmans 1996;

Natter and Hruschka, 1998; Hruschka, 2001).

With the exception of Hruschka (2000) and Hruschka (2001) who analyzes data of one store, all studies belonging to the second group linearly aggregate data (by sums or averages across stores) or pool data across stores. Though they tackle with bias due to functional form, bias due to heterogeneity across stores still exists. That is why in this paper a model is developed which allows both heterogeneity and functional flexibility. To this end sales response functions are specified as MLPs with store-specific coefficients in a hierarchical Bayesian framework. Moreover, an appropriate MCMC estimation technique is introduced.

We try to answer the question wether for a typical marketing data set functional flexibility is advantageous if heterogeneity is considered also. If a heterogeneous strict parametric model performs at least quite as good as the heterogeneous flexible model or gives roughly the same results about marketing effects, most researchers would stick to the less complex parametric model. Therefore we compare the MLP to a strict parametric model, where both models have store-specific coefficients. Relative performance of the MLP is measured by its posterior probability. If the MLP is better according to this criterion, we investigate if it implies a pattern of price response different from the parametric model.

2 Specification of Models

Originally several parametric sales response models were considered (i.e. linear, multiplicative, exponential, semi-log taking logs of prices, logistic and asymmetric logistic) with price of the respective brand and average price across competitors as predictors (a description of these functional forms can be found in Hanssens et al., 2001). A heterogeneous version of each of these models was estimated by an appropriate MCMC method. Posterior model probabilities for the multiplicative models computed from log marginal model densities approach the best value of one (see section 3). Because of its superiority this paper uses the multiplicative model as standard of comparison for the MLP.

The heterogeneous multiplicative model (abbreviated as HMM) is expressed by:

$$Q_{mit} = \exp(\alpha_{mi0})\, p_{mit}^{\alpha_{mi1}} \prod_{j=1}^{|C_m|} p_{C_{mj}it}^{\alpha_{mij+1}} \qquad (1)$$

Q_{mit} and p_{mit} are sales and price of brand m in store i and week t. C_m denotes the index set of competing brands (in the following briefly called competitors) which affect sales of brand m (this set may be empty). These are the brands for which the 95% credible interval of the mean price coefficient for brand m in the HMM did not include zero (see section 3). C_{mj} is the j-th element of this index set. $|C_m|$ symbolizes the number of brands contained.

Model specifications originally included coefficients for point of sales displays as well as prices and sales both lagged by one week as possible dynamic effects. Estimation of corresponding heterogeneous multiplicative models demonstrated that these variables may be eliminated as 95% credible intervals of their mean coefficients included zero. That is why these variables are ignored here.

Estimation is based on the log-transformation of the HMM:

$$\log(Q_{mit}) = \alpha_{mi0} + \alpha_{mi1}\, \log(p_{mit}) + \sum_{j=1}^{|C_m|} \alpha_{mij+1}\, \log(p_{C_{mj}it}) \qquad (2)$$

MLPs approximate any continuous multivariate function and its derivatives to the desired level of precision given a sufficient number of hidden units with S-shaped activation functions (Cybenko, 1989; Hornik et al., 1989; Ripley, 1993). In order to achieve a certain approximation rate for problems of higher dimension, MLPs approximate unknown functions and their derivatives with errors decreasing at rates as fast as $u^{-1/2}$, with u as number of hidden units.

Rates for standard kernel, spline and Taylor series approximants are higher and increase with the dimension of the input space (Barron, 1993; Hornik et al. 1994).

The heterogeneous multilayer perceptron (abbreviated as HMLP) is written as follows:

$$\log(Q_{mit}) = \alpha^1_{mi0} + \alpha^1_{mi1} \log(p_{mit}) + \sum_{j=1}^{|C_m|} \alpha^1_{mij+1} \log(p_{C_{mj}it}) + \sum_{u=1}^{U_m} \alpha^2_{miu} h_{miut} \quad (3)$$

As can be seen from expression 3 compared to the HMM the HMLP has as additional "predictors" the values of hidden units h_{miut}, which are computed by binomial logistic functions:

$$h_{miut} = 1/[1 + \exp(-(\alpha^3_{mi0u} + \alpha^3_{mi1u} \log(p_{mit}) + \sum_{j=1}^{|C_m|} \alpha^3_{mij+1u} \log(p_{C_{mj}it})))] \quad (4)$$

Note that all coefficients of the HMLP vary across stores.

Store-level elasticities ϵ_{mmi} and cross-elasticities $\epsilon_{mC_{mj}i}$ are defined by:

$$\epsilon_{mmi} = \frac{\partial \log(Q_{mit})}{\partial \log(p_{mit})}, \quad \epsilon_{mC_{mj}i} = \frac{\partial \log(Q_{mit})}{\partial \log(p_{C_{mj}it})} \quad (5)$$

For the HMM store-level elasticities and cross-elasticities are simply equal to the corresponding price coefficients:

$$\epsilon_{mmi} = \alpha_{mi1}, \quad \epsilon_{mC_{mj}i} = \alpha_{mij+1} \quad (6)$$

Store-level elasticities and cross-elasticities for the HMLP are:

$$\epsilon_{mmi} = \alpha^1_{mi1} + \sum_{u=1}^{U_m} \alpha^2_{miu}(1 - h_{miut})\, h_{miut}\alpha^3_{mi1u} \quad (7)$$

$$\epsilon_{mC_{mj}i} = \alpha^1_{mij+1} + \sum_{u=1}^{U_m} \alpha^2_{miu}(1 - h_{miut})\, h_{miut}\alpha^3_{mij+1u} \quad (8)$$

Economic theory postulates that holding other effects constant sales decrease if the price of a brand rises and increase if the price of any competitor rises (e.g. Rao, 1993). Therefore store-level elasticities should not be positive and store-level cross-elasticities should not be negative. For the HMM these restrictions are fullfilled by these obvious restrictions on parameters:

$$\begin{aligned} \alpha_{mi1} &\leq 0 \\ \alpha_{mij+1} &\geq 0 \quad \text{for} \quad j = 1, \cdots, |C_m| \end{aligned} \tag{9}$$

The desired properties of elasticities and cross-elasticities can also be achieved by constraining coefficients of the HMLP utilizing the fact that the first derivative of the logistic function $(1 - h_{miut})\, h_{miut}$ which is part of elasticity formulas 7 and 8 is always positive. These restrictions are:

$$\begin{aligned} \alpha^1_{mi1} \leq 0, &\qquad \alpha^3_{mi1u} \leq 0 \\ \alpha^1_{mij+1} \geq 0, &\qquad \alpha^3_{mij+1u} \geq 0 \\ \alpha^2_{miu} \geq 0 & \end{aligned} \tag{10}$$

These restrictions mean that coefficients for the price of a brand must not be positive, those for prices of competitors and for values of hidden units must not be negative.

If interest focuses on price effects across all stores, one should look at price elasticities w.r.t. total sales. Total sales Q_{mt} of brand m in week t are equal to the sum of store sales of this brand in week t across all I stores:

$$Q_{mt} = \sum_{i=1}^{I} Q_{mit} \tag{11}$$

Price elasticity w.r.t. total sales ϵ_{mm} is the average of store-level elasticities ϵ_{mmi} weighted by the ratio of estimated store-level sales and estimated total sales:

$$\epsilon_{mm} = \sum_i \epsilon_{mmi} \frac{\hat{Q}_{mit}}{\hat{Q}_{mt}} = \sum_i \epsilon_{mmi} \frac{\hat{Q}_{mit}}{\sum_{i=1}^{I} \hat{Q}_{mit}} \tag{12}$$

This expression shows that even for the HMM total sales elasticities are no longer constant.

Coefficients of all models considered here vary across stores. Let coefficients of brand m and store i be collected in a column vector β_{mi}. For the HMM this means:

$$\beta_{mi} \quad = \quad (\alpha_{mi0}, \cdots, \alpha_{mi|C_{mj}|+1})' \tag{13}$$

For the HMLP β_{mi} is defined as:

$$\beta_{mi} = (\alpha^1_{mi0}, \cdots, \alpha^1_{mi|C_{mj}|+1}, \alpha^2_{mi01}, \cdots, \alpha^2_{mi|C_{mj}|+1U_m}, \alpha^3_{mi01}, \cdots, \alpha^3_{mi|C_{mj}|+1U_m})' \tag{14}$$

Store-specific coefficient vectors are assumed to be multivariate normally distributed with mean vector $\bar{\beta}_m$ and covariance matrix Σ_m. This constitutes a hierarchical model as the so-called hyperparameters $\bar{\beta}_m$ and Σ_m affect the dependent variable through the store-specific coefficient vectors only. It is characteristic of such models that each store-specific coefficient depends on data from all stores (Carlin and Louis, 1996). Therefore coefficients are expected to be less noisy and unstable than those based on estimating single store-specific regression models (Hanssens et al., 2001).

3 Estimation and Model Evaluation

There are a few papers estimating MLPs by Bayesian approaches. One of these approaches, the evidence framework of MacKay (1992a,b), is essentially a Laplace approximation to the posterior distribution. Baesens et al. (2002) apply the evidence framework in a direct marketing context. Alternatively, MLPs are estimated by MCMC techniques (Neal, 1996; Müller and Insua, 1998; Lee, 2000; Lampinen and Vehtari, 2001). While Neal (1996) discusses hierarchical models briefly, none of these contributions develop or apply a method to estimate MLPs with heterogeneous coefficients.

The MCMC simulation technique used here estimates heterogenous store-level sales models. It is based on a method developed by Train (2001, 2003), more details on which may be found in the appendix and the references given in this section. After convergence to stationarity it generates random samples of parameters from the joint posterior density. Statistics of sampled values (e.g. means, percentiles) converge to their population values. The MCMC technique consists of several iterations each having four substeps:

1. a Metropolis-Hastings algorithm which samples store specific coefficients β_{mi} (Chib and Greenberg, 1995 and 1996; Gelman et al., 1995), but keeps only samples which satisfy the restrictions given in expression 10 (Gelfand et al., 1992).

2. a Gibbs sampler which draws the error precision (i.e. the inverse of the error variance) from the conditional Gamma distribution whose parameters depend on total sum of squared errors and number of observations.

3. a Gibbs sampler which draws from the conditional multivariate normal distribution of mean coefficients $\bar{\beta}_m$ given store-specific coefficients β_{mi} and covariance matrix Σ_m.

4. a Gibbs sampler which draws from the conditional inverted Wishart distribution of the covariance matrix Σ_m given store-specific and mean coefficients.

Model performance is evaluated by posterior model probabilities in accordance with the dominant approach in Bayesian statistics. Posterior model probabilities penalize models for complexity, i.e. all else being equal the more complex model receives a lower value. In accordance with a proposal made by Raftery (1996) the HMLP is judged to perform better if its posterior model probability is greater than 0.75.

Assuming equal a priori model probabilities the posterior model probability of the HMLP model is computed from marginal model densities $p(y|M_0)$ of the

HMM and $p(y|M_1)$ of the HMLP. Marginal model densities are determined by the harmonic mean estimator of Gelfand and Dey (1994). The posterior probability of the HMLP model is given by:

$$\frac{p(y|M_1)}{p(y|M_0) + p(y|M_1)} \tag{15}$$

Price elasticities are sampled by applying the store-specific coefficients obtained by the MCMC technique to expressions 2(3), 6(7), 11 and 12. This procedure is based on the fact that the posterior of any function of model parameters can be computed by simply plugging sampled parameters into the relevant expressions (Geweke, 1989).

95 % credible intervals for individual coefficients and elasticities are estimated by 2.5 and 97.5 percentiles of their sampled values. The probability that a coefficient (an elasticity) lies in the 95 % credible interval given the observed data is at least 95 % (Carlin and Louis, 1996).

4 Empirical Study

The store-level data analyzed refer to the nine leading brands of a certain category of packaged consumer goods. Data were acquired in 81 stores. Between 61 and 88 weeks per store lead to a total of 62878 observations. Table 1 contains descriptive statistics (means and standard deviations) of sales and prices for these nine brands.

==

put table 1 about here

==

Trace plots and autocorrelations with a maximum lag of 50 iterations for mean coefficients serve to assess wether the MCMC technique does not converge (Kass et al., 1998). 100,000 iterations are used for each model, the last

10,000 of which provide the sampled coefficients used to compute estimates. Trace plots show that for all HMMs and most HMLPs mean coefficients become stable for less than 50,000 iterations. Autocorrelations also do not indicate nonconvergence.

Table 2 contains log marginal densities for both the HMM and the best HMLP (i.e. those with highest log marginal density among HMLPs with between one and four hidden units). None of these MLPs has more than three hidden units, for three brands even one hidden unit is sufficient. Posterior model probabilities given in the last column of table 2 show that HMLPs clearly outperform their multiplicative counterpart. The HMLPs approach the best value of 1.00 for each of the nine brands. These results indicate the superiority of a more flexible compared to a strict parametric model even if heterogeneity is dealt with in both models.

=======================================

put table 2 about here

=======================================

Results are interpreted by means of values of hidden units and sales response curves in the following. Three competitive scenarios are considered called low, medium and high prices of competitors. In these scenarios the price of each competitor is set to its arithmetic mean minus 1.3 times its standard deviation, to its arithmetic mean and to its arithmetic mean plus 1.3 times its standard deviation, respectively.

Interpretation of hidden units is demonstrated for brands 2 and 7. Table 3 gives the output values of hidden units of the best MLP model for each competive scenario and different values of the price of the respective brand on the basis of mean coefficients $\bar{\beta}_m$. Hidden unit 1 can be seen as indicator of a very favorable price of brand 1 compared to the prices of competitors. It attains the maximum value 1.00 at very low prices of competitors and a somewhat lower own price. Hidden unit 2 indicates a favorable price of brand 1. It attains the maximum value at medium prices of competitors and low own price. Proceeding the same way for brand 7, hidden units 1, 2 and 3 may be interpreted as indicators of extremely favorable, very faborable and favorable prices of this brand.

We arrive at quite similar interpretations of hidden units for the other brands. Results for brand 9 are somewhat different as prices of competitors do not have a significant effect on its sales. Hidden units attain the maximum value if the price of brand 9 is low or medium. Values of hidden units decrease if the price becomes very high.

==

put table 3 about here

==

Estimated sales are plotted as function of the price of a brand using the mean coefficients $\bar{\beta}_m$ (figures 1 and 2) and the competitive scenarios introduced above to investigate wether implied price effects differ between the HMM and the best HMLP. Curves for brand 5 are not shown as they are almost equal for the two models.

For brand 9 we only have one curve per model as prices of other brands do not have a significant effect on sales. The HMLP implies higher sales except at very own high prices at which sales decreases are stronger.

==

put figures 1 and 2 about here

==

For the seven remaining brands curves show large (for some brands very large) differences between the two models at high prices of competitors. As a rule, the HMLP expects higher sales. The HMLP expects stronger responses to price changes for brands 6 and 7.

For medium prices of competitors the HMLP implies higher sales for brands 2, 3, 4, 7 and 8 if own price is not too high. For brands 2 and 5 price response in this range is stronger compared to the HMM. At low prices of competitors the HMLP implies smaller sales for brands 2, 3, 4 and 8.

Price elasticities discussed in the following refer to total sales (i.e. sales summed across all 81 stores). They are averaged across all observed prices of competitors. Curves in figures 3 and 4 give the 95 % credible interval of elasticities. Curves are not shown for brand 5 as they are indistinguishable for both models.

==

put figures 3 and 4 about here

==

According to the HMM elasticities decrease very slightly for increasing prices. Except for brand 9 elasticities implied by the HMLP differ from those for the HMM over most of the observed price range. According to the HMLP elasticities tend to follow a bell-shaped curve. This is obvious for brands 3, 7 and 8. The bell shape is incomplete for brands 2, 4 and 6 because observations do not include lower prices. The incomplete bell shape for brand 9 may be explained by the lack of observations at higher prices.

5 Conclusions

This paper focuses on the comparison of two sales response models, a strict parametric and a flexible model (a multilayer perceptron). As effects of marketing instruments are as a rule heterogeneous across stores, both models are specified in a hierarchical Bayesian framework which allows estimation of store-specific coefficients by Markov Chain Monte Carlo techniques. The data base consists of weekly observations of sales and prices for nine brands of a packaged consumer good acquired in 81 stores.

From a statistical point of view the heterogeneous multilayer perceptron turns out to be superior. For all brands its posterior probability approaches the maximum possible value. Moreover, price effects implied by the multilayer perceptron differ for eight out of nine brands, especially at high prices of competitors.

We obtain similar results for elasticities aggregated across stores. 95 % credible intervals of elasticities differ over most of the price range. Elasticities estimated on the basis of the multilayer perceptron follow a (sometimes incomplete) bell-shape for increasing prices for eight brands. Quite contrary, the multiplicative model implies elasticities which decrease very slightly if prices rise.

These results give evidence that considering both functional flexibility and heterogeneity can be beneficial. Hopefully they will motivate researchers to consider more flexible heterogeneous models when analyzing marketing data.

This contribution deals with price response. Given the potential advantages of more flexible models which simultaneously are able to fulfill well founded theoretical constraints, future sales response modelling efforts including other marketing instruments (e.g. sales promotion, sales force, advertising) should be of interest.

Appendix: Markov Chain Monte Carlo Technique

The likelihood value L_{mi} of brand m and store i depends on the number of weekly observations T_i, the sum of squared errors SSE_{mi} and the error precision h_m:

$$L_{mi} = \frac{h_m^{T_i/2}}{(2\pi)^{T_i/2}} \exp\left[-\frac{h_m}{2} SSE_{mi}\right] \tag{16}$$

The sum of squared errors of brand m and store i is defined by:

$$SSE_{mi} = \sum_{t=1}^{T_i} [log(Q_{mit}) - \log(\hat{Q}_{mit})]^2 \tag{17}$$

$\log(Q_{mit})$ and $\log(\hat{Q}_{mit})$ are observed and estimated log sales for brand m, store i and period t, respectively.

Both total sum of squared errors SSE_m and total number of observations are obtained by summing over stores:

$$SSE_m = \sum_{i=1}^{I} SSE_{im}, \quad T = \sum_{i=1}^{I} T_i \tag{18}$$

The MCMC procedure is based on the following priors. The prior of the vector of mean coefficients $\bar{\beta}_m$ is K-variate normal with parameters $\bar{\beta}_m^0$ and Σ_m^0. The prior of $K \times K$ covariance matrix Σ_m is inverse Wishart with K degrees of freedom and the $K \times K$ identity matrix E as scale matrix .

The $K \times K$ diagonal matrix Λ_m denotes the lower Cholesky factor of covariance matrix Σ_m, τ the so-called tuning constant. Indicator function $I_{\beta_{mi}}$ equals one if all restrictions on coefficients given in expression 10 are satisfied, otherwise it equals zero. s_m^2 is a crude initial guess of total error variance.

The steps of the MCMC procedure may be described as follows:

1. Sampling of store-specific coefficients for each store $i = 1, I$

   (a) Generate a trial vector of coefficient values

   $\beta_{mi}^t = \beta_{mi} + \tau \Lambda_m v$

   where

   v is a K-dimensional vector of independent draws from the standard normal distribution

(b) Update coefficients and likelihood

if $(U < R\, I_{\beta_{mi}})$ $\quad \beta_{mi} = \beta_{mi}^t,\ L_{mi} = L_{mi}^t$

where

U is a draw from the uniform distribution $U(0,1)$

$R = (L_{mi}^t\, \phi(\beta_{mi}^t | \bar{\beta}_m, \Sigma_m))/(L_{mi}\, \phi(\beta_{mi} | \bar{\beta}_m, \Sigma_m))$

L_{mi}^t Likelihood value for trial coefficient vector

$\phi(\beta_{mi} | \bar{\beta}_m, \Sigma_m)$ k-variate normal density

2. Sampling of error precision h_m

Draws from the Gamma distribution with $T + 10$ degrees of freedom and mean $(T + 10)/(SSE_m + 10\, s_m^2)$

3. Sampling of mean coefficients $\bar{\beta}_m$

   $\bar{\beta}_m = 1/I \sum_i \beta_{mi} + \Lambda_m v$

   where

   v is a K-dimensional vector of independent draws from the standard normal distribution

4. Sampling of covariance matrix Σ_m

   Draws from the inverted Wishart distribution with K+I degrees of freedom and scale matrix $(KE + I\, V_m)/(K + I)$ and computes its lower Cholesky factor Λ_m

   where

   $V_m = (1/I) \sum_i (\beta_{mi} - \bar{\beta}_m)(\beta_{mi} - \bar{\beta}_m)'$

Ripley (1987) and Train (2003) give details on sampling from probability distributions.

References

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J. and G. Dedene (2002): Bayesian Neural Network Learning for Repeat Purchase Modelling in Direkt Marketing, *European Journal of Operational Research* 134, 191–211.

Barron, A.R. (1993): Universal Approximation Bounds for Superpositions of a Sigmoidal Function, *IEEE Transactions on Information Theory* 39, 930–945.

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

Boatwright, P., McCulloch, R. and P. Rossi (1999): Account-Level Modeling for Trade Promotion: An Application of a Constrained Parameter Hierarchical Model, *Journal of the American Statistiscal Association* 94, 1063–1073.

Carlin, B.P. and Th.A. Louis (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall, London.

Chib, S. and E. Greenberg (1995): Understanding the Metropolis-Hastings Algorithm, *The American Statistician* 49, 327–335.

Chib, S. and E. Greenberg (1996): Markov Chain Monte Carlo Simulation Methods in Econometrics, *Econometric Theory* 12, 409–431.

Cybenko, G. (1989): Continuous Value Neural Networks with Two Hidden Layers are Sufficient, *Mathematics of Control, Signal and Systems* 2, 303–314.

Gallant, A.R. (1987), "Identification and Consistency in Seminonparametric Regression," in T.F. Bewley (ed.), *Advances in Econometrics* Vol I., Cambridge University Press, Cambridge UK, 145–170.

Gelfand, A.E. and D.K. Dey (1994): Bayesian Model Choice: Asymptotics and Exact Calculations, *Journal of the Royal Statistical Society Series B*, 56, 101–514.

Gelfand, A.E., Smith, A.F.M. and T.-M. Lee (1992): Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, *Journal of the American Statistical Association* 87, 523–531.

Gelman, A., Carlin, J.B., Stern, H.S. and D.B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, London.

Geweke, J. (1989): Bayesian Inference in Econometric Models using Monte Carlo Integration, *Econometrica* 57, 1317–1339.

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

Hastie, T.J. and R.J. Tibshirani (1995), *Generalized Additive Models*, Chapman and Hall, London.

van Heerde, H. (1999), *Models for Sales Promotion Effects Based on Store-Level Scanner Data*, Labyrint Publication, Capelle a/d Ijssel, Netherlands.

van Heerde, H., Leeflang, P.S.H. and D.R. Wittink (2001): Semiparametric Analysis to Estimate the Deal Effect Curve, *Journal of Marketing Research* 38, 197–216.

Hornik, K., Stinchcombe, M. and H. White (1989): Multilayer Feedforward Networks are Universal Approximators, *Neural Networks* 3, 359–366.

Hornik, K., Stinchcombe, M., White, H. and P. Auer (1994): Degrees of Approximation Results for Feedforward Networks Approximating Unknown Mapping and Their Derivatives, *Neural Computation* 6, 1262–1275.

Hruschka, H. (1993): Determining Market Response Functions by Neural Network Modeling. A Comparison to Econometric Techniques, *European Journal of Operational Research* 66, 27–35.

Hruschka, H. (2000), "Specification, Estimation and Empirical Corroboration of Gutenberg's Kinked Demand Curve", in Albach, H., Brockhoff, K., Eymann, E., Jungen, P., Steven, M. and A. Luhmer (eds.), *Theory of the Firm. Erich Gutenberg's Foundations and Further Developments*, Springer, Berlin, 153–168.

Hruschka, H. (2001): An Artificial Neural Net Attraction Model (ANNAM) to Analyze Market Share Effects of Marketing Instruments, *Schmalenbach Business Review-zfbf* 53, 27–40.

Kalyanam, K. and Th.S. Shively (1998): Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach, *Journal of Marketing Research* 35, 16–29.

Kass, R.E., Carlin, B.P., Gelman, A. and R.M. Neal (1998): Markov Chain Monte Carlo in Practice: A Roundtable Discussion, *The American Statistician* 52(2), 93–100.

Krishnamurthi, L., Raj, S.P. and R. Selvam (1990): Statistical and Managerial Issues in Cross-Sectional Aggregation. Working Paper, Northwestern University.

Lampinen, J. and A. Vehtari (2001): Bayesian Approach for Neural Networks - Review and Case Studies, *Neural Networks* 14, 257–274.

Lang, St., Adebayo, S.B., Fahrmeir, L. and W.J. Steiner (2003), Bayesian Geoaditive Seemingly Unrelated Regression, *Computational Statistics* 18, 263–292.

Lee, H.K.H. (2000). A Noninformative Prior for Neural Networks. Working Paper, ISDS, Duke University, Durham NC.

MacKay, D.J.C. (1992a): Bayesian Interpolation, *Neural Computation* 4, 415–447.

MacKay, D.J.C. (1992b): A Practical Bayesian Framework for Backpropagation Networks, *Neural Computation* 4, 448–472.

Montgomery, A.L. (1997): Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data, *Marketing Science*, 16, 315–337.

Müller, P. and D.R. Insua (1998): Issues in Bayesian Analysis of Neural Network Models, *Neural Computation* 10, 571–592.

Natter, M. and H. Hruschka (1998), "Using Artificial Neural Nets to Specify and Estimate Aggregate Reference Price Models", in Aurifeille, J.-M. and Ch. Deissenberg (eds.), *Bio-Mimetic Approaches in Management Science*, Kluwer Academic Publishers, Dordrecht, Netherlands, 101–118.

Neal, R.M. (1996), *Bayesian Learning for Neural Networks*, Springer, New York.

Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge UK.

Raftery, A.E. (1996), "Hypothesis Testing and Model Selection," in Gilks, R., Richardson, S. and D.J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, Boca Raton, 163–187.

Rao, V. (1993), "Pricing Models in Marketing", in Eliashberg, J. and G.L. Lilien (eds.), *Marketing*, North-Holland, Amsterdam, 517–552.

Ripley, B.D. (1987), *Stochastic Simulation*, Wiley, New York.

Ripley, B.D. (1993), "Statistical Aspects of Neural Networks," in Barndorff-Nielsen, O.E., Jensen, J.L. and W.S. Kendall (eds.), *Networks and Chaos – Statistical and Probabilistic Aspects*, Chapman and Hall, London, 40–123.

Train, K.E. (2001), A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit. Working Paper, Department of Economics, University of California, Berkeley,

Train, K.E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge UK.

van Wezel, M.C. and W.R.J Baets (1995): Predicting Market Responses with a Neural Network, *Marketing Intelligence & Planning*, 13(7), 23–30.

Wierenga, B. and J. Kluytmans (1996): Prediction with Neural Nets in Marketing Time Series Data. Working Paper, School of Management, Erasmus Universiteit Rotterdam.

Table 1: *Descriptive Statistics*

	Sales		Prices	
Brands	Mean	Standard Deviation	Mean	Standard Deviation
1	55.12	100.76	285.41	33.01
2	195.85	251.50	293.96	57.21
3	264.78	392.411	218.18	38.45
4	290.55	497.38	220.86	42.07
5	99.62	264.19	229.61	35.66
6	48.67	95.08	214.26	31.16
7	56.61	271.32	214.82	41.60
8	307.85	517.74	174.39	41.51
9	53.05	54.79	149.08	8.85

Table 2: *Model Evaluation Results*

Brand	HMM Log of Marginal Density	Best HMLP Number of Hidden Units	Log of Marginal Density	Posterior Probability
1	-23611.33	3	-23348.29	1.00
2	-21763.64	2	-20975.99	1.00
3	-25380.88	2	-25126.25	1.00
4	-24286.86	1	-23774.01	1.00
5	-23617.84	1	-23585.38	1.00
6	-21932.54	1	-21294.56	1.00
7	-27328.39	3	-26833.91	1.00
8	-26790.00	2	-26762.76	1.00
9	-23419.23	3	-23302.14	1.00

Table 3: *Values of Hidden Units of HMLPs*

1	2	3
Brand 2		
Low Prices of Competitors		
224: 0.00	224: 0.00	-
370: 0.00	370: 0.00	-
Medium Prices of Competitors		
224: 0.84	224: 0.99	-
272: 0.10	332: 0.10	-
High Prices of Competitors		
224: 1.00	224: 1.00 -	
370: 0.31	370: 1.00 -	
Brand 7		
Low Prices of Competitors		
165: 0.00	165: 0.00	165: 0.00
269: 0.00	269: 0.00	269: 0.00
Medium Prices of Competitors		
165: 0.00	165: 1.00	165: 1.00
269: 0.00	261: 0.10	239: 0.10
High Prices of Competitors		
165: 1.00	165: 1.00	165: 1.00
227: 0.10	269: 0.10	269: 1.00

price of brand: value of hidden unit

Figure 1: *Sales Response Curves (Part A)*

Top, middle and bottom curves for high, medium and low prices of competitors, respectively.
HMLP (solid curves), HMM (dotted curves).

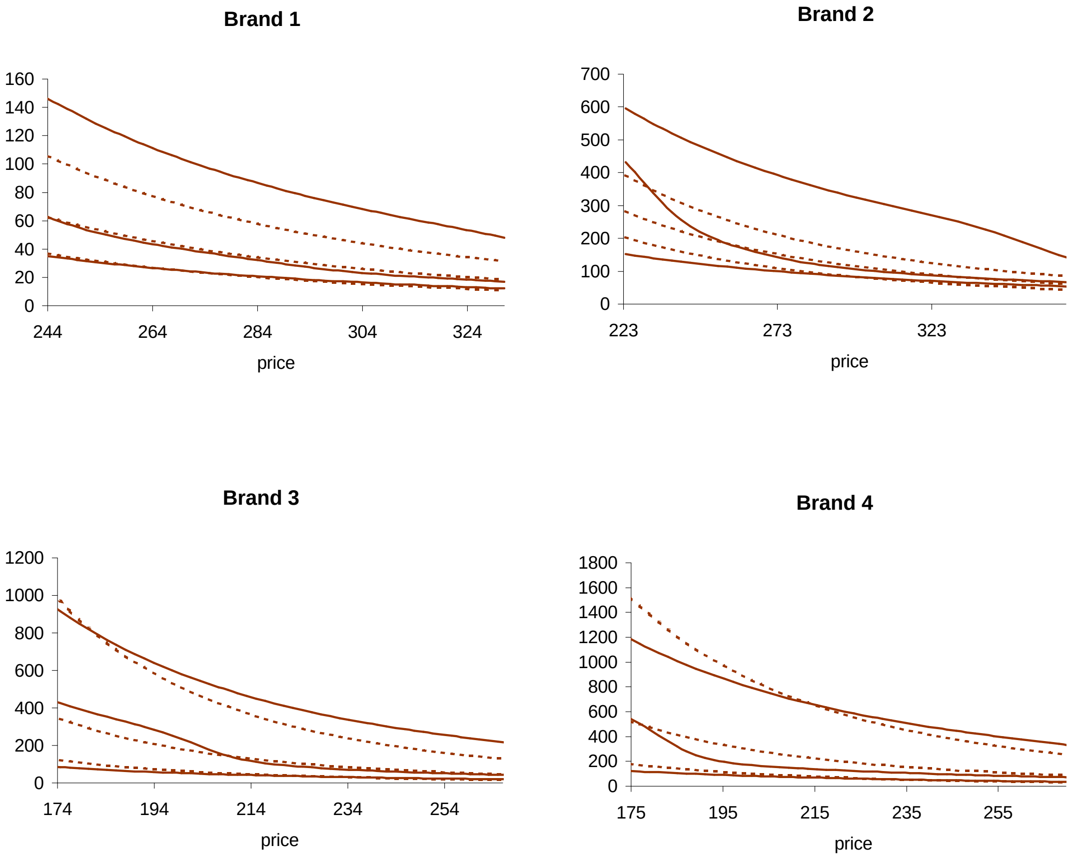

Figure 2: *Sales Response Curves (Part B)*

Top, middle and bottom curves for high, medium and low prices of competitors, respectively.
HMLP (solid curves), HMM (dotted curves).

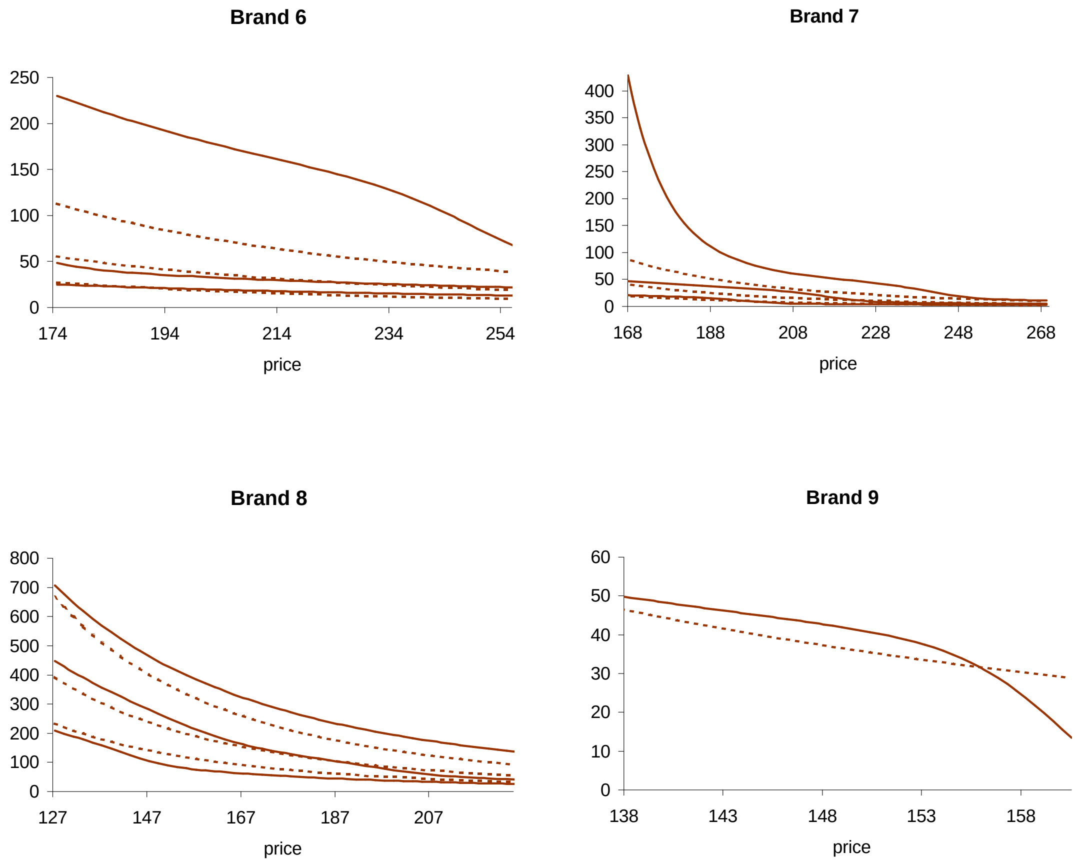

Figure 3: *Elasticity Curves (Part A)*

95 % credible intervals
HMLP (solid curves), HMM (dotted curves).

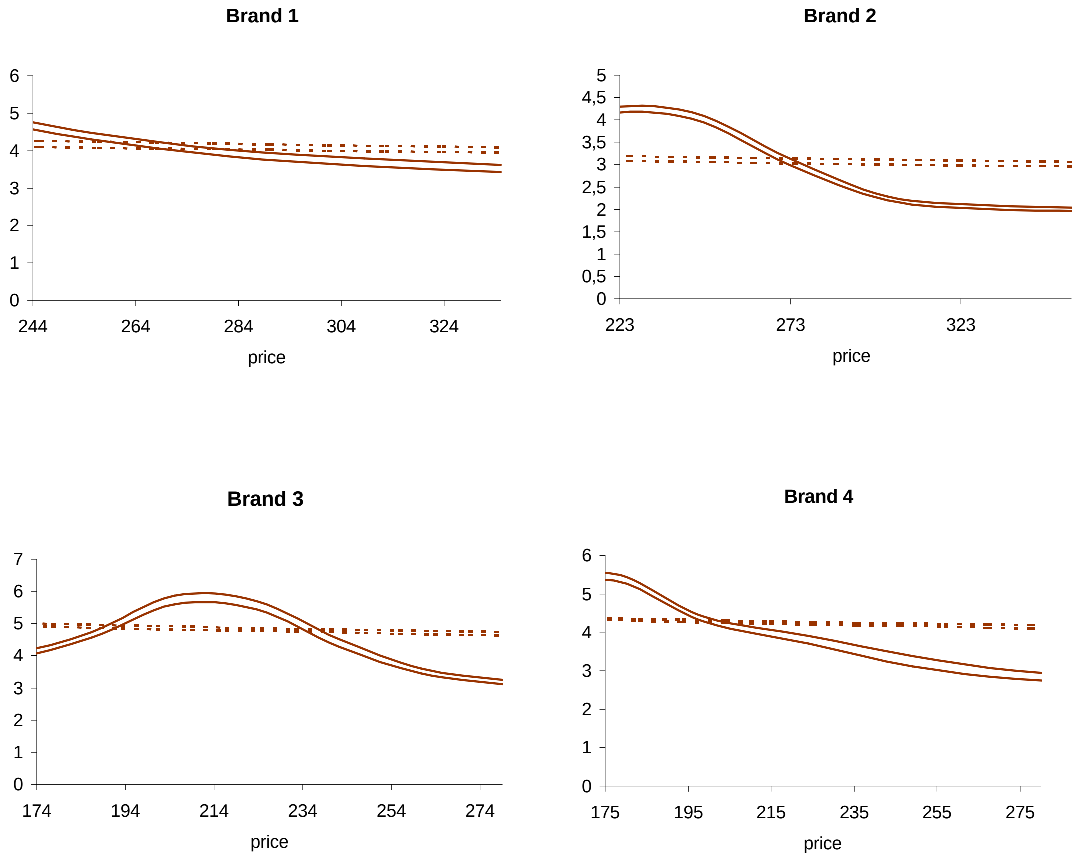

Figure 4: *Elasticity Curves (Part B)*

95 % credible intervals
HMLP (solid curves), HMM (dotted curves).