

Analyzing market baskets by restricted Boltzmann machines

Harald Hruschka

Published online: 9 August 2012
© Springer-Verlag 2012

Abstract We introduce a model which differs from the well-known multivariate logit model (MVL) used to analyze the cross-category dependence in market baskets by the addition of binary hidden variables. This model is called restricted Boltzmann machine (RBM) and new to the marketing literature. Extant applications of the MVL model for higher numbers of categories typically follow a two-step approach as simultaneous maximum likelihood estimation is computationally infeasible. In contrast to the MVL, the RBM can be simultaneously estimated by maximum likelihood even for a higher number of categories as long as the number of hidden variables is moderate. We measure the cross-category dependence by pairwise marginal cross effects which are obtained using estimated coefficients and sampling of baskets. In the empirical study, we analyze market baskets consisting of the 60 most frequently purchased categories of the assortment of a supermarket. For a validation data set, the RBM performs better than the MVL model estimated by maximum pseudo-likelihood. For our data, about 75 % of the baskets are reproduced by the model without cross-category dependence, but 25 % of the baskets cannot be adequately modeled if cross effects are ignored. Moreover, it turns out that both the number of significant cross effects and their relationships can be grasped rather easily.

Keywords Market basket analysis · Retailing · Marketing · Data mining

1 Introduction

The market basket of a shopper, which consists of those product categories which she purchases during a shopping trip at a store (Russell et al. 1997, 1999), is the observable

H. Hruschka (✉)
University of Regensburg, Universitätsstrasse 31, 93053 Regensburg, Germany
e-mail: harald.hruschka@wiwi.uni-regensburg.de

result of multicategory choices. Multicategory choices are often interdependent, i.e., the choice of one category has an effect on the choice of another category. Two categories are considered as purchase complements (purchase substitutes) if their cross effects are positive (negative), i.e., if these categories are purchased jointly more (less) frequently than expected under stochastic independence (Betancourt and Gautschi 1990; Hruschka et al. 1999). Because of these interdependences market basket analysis requires models which measure such cross effects between product categories, whereas using a collection of one-category models would not to be appropriate.

In this paper, we only look at purchase incidences and do not consider other response variables like purchase quantities,¹ or several of these response variables.² Therefore, we represent the market basket as binary vector whose dimension equals the number of considered categories. Measurement of cross effects between product categories constitutes the main objective of our work.

The models most prominent in the marketing literature which are capable of measuring cross-category effects for purchase incidence data are the multivariate logit (MVL) and the multivariate probit (MVP) models. These models may also include explanatory variables such as marketing instruments or consumer demographics. To our knowledge, Manchanda et al. (1999) published the first application of the MVP model (four categories), followed by Chib et al. (2002) and Duvvuri et al. (2007) who analyzed 12 and 6 categories, respectively. In a recent paper, Duvvuri and Gruca (2010) analyzed four subcategories of the instant coffee category. They specify the overall price sensitivity which is indicated by higher-order factor scores as function of household-level covariates. Russell and Petersen (2000) applied the MVL model to four categories. Boztuğ and Hildebrandt (2008) studied cross-category effects by means of the MVL model for six and five categories.

The overwhelming majority of the previous studies on both multicategory purchase incidence and other multicategory response variables consider rather small numbers of categories. For the MVL model, this limitation can be explained by the computational intractability of maximum likelihood estimation, as the number of possible market baskets grows exponentially with the number of categories (see Sect. 3 for more details). Extant publications dealing with higher numbers of categories refrain from simultaneously estimating all parameters of a comprehensive model and follow a two-step approach (Hruschka et al. 1999; Boztuğ and Reutterer 2008). In contrast to simultaneous approaches, the results of the first step are not changed in view of estimation results of the second step. Because results of the first step restrict the models estimated in the second step, two-step approaches usually do not attain the statistical

¹ Brijs et al. (2004) analyzed four product categories by means of a multivariate Poisson mixture model. Wang et al. (2007) introduced a multivariate Poisson-lognormal model and study cross-category effects between five product categories. Niraj et al. (2008) dichotomized the purchase quantity by distinguishing purchases of 1 U and purchases of at least 2 U. On the basis of this dichotomization, these authors specify a two-stage bivariate logit model of purchase incidence and quantity for two product categories.

² Song and Chintagunta (2007) developed an integrated model based on an utility-maximizing framework for multicategory purchase incidence, brand choice, and purchase quantity. In their empirical study, these authors analyzed four product categories.

performance of simultaneous approaches. This weakness of two-step approaches is well known from, e.g., conjoint analysis models (Wedel and Kamakura 1998).

Hruschka et al. (1999) estimated one binary logit model for each of 73 categories investigated and determine which interactions with the remaining 72 categories are significant in the first step. In the second step, these authors estimated a model for all 73 categories with interaction terms selected in the first step setting the remaining interaction terms to zero. Boztuğ and Reutterer (2008) started from basket data on 65 categories. In the first step, they used a vector quantization method to determine clusters of market baskets. In the second step they estimated one MVL model for each cluster considering the four or five categories with highest purchase probabilities conditional on the respective cluster.

The model we introduce is known as restricted Boltzmann machine (RBM). The RBM is related to the MVL model from which it differs by the inclusion of binary hidden variables. So far, the RBM has been mostly applied to solve pattern recognition problems, e.g., recognition of handwritten digits or classification of documents (Hinton and Salakhutdinov 2006). Compared to previous work using MVL models the RBM allows us to analyze a higher number of product categories by simultaneous estimation for all categories investigated.

Due to data limitations, the models estimated by us do not include explanatory variables. This procedure can also be justified by the results of two previous studies. The best MVL model with price effects in and Russell and Petersen (2000) performs only slightly better than the MVL model without explanatory variables in terms of log likelihood for the validation data (the former model has six parameters more and a log likelihood values which is higher by 2.92). Similarly, price effects are not statistically significant in the MVL model of Boztuğ and Hildebrandt (2008).

We begin by juxtaposing the specifications of the well-known MVL model and the RBM. Then, we explain maximum likelihood (ML) estimation of RBMs and the independence model and maximum pseudo-likelihood (MPL) estimation of the the MVL model. We also present the criteria we use to evaluate different RBMs, the independence model and the MVL model. We derive an approach to measure asymmetric marginal cross-category effects based on estimation results of a RBM. In the empirical part of our paper, we present descriptive statistics of the data and select one of the estimated models. Based on this model, we interpret the hidden variables and discuss strong and significant cross effects between pairs of categories. In the final chapter, we give an overview of our results, discuss managerial implications, and mention several related future research opportunities.

2 Model

In this section, we present both the MVL and the RBM models to clarify their relationship. Indices of market baskets are $i = 1, \dots, I$, indices of categories $j = 1, \dots, J$. Basket i is characterized by a binary column vector of purchase incidences $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})'$. If market basket i contains category j , then $y_{ij} = 1$, otherwise $y_{ij} = 0$.

Russell and Petersen (2000) applied the MVL model (which is also known as autologistic model) to market basket analysis drawing upon earlier work of Cox (1972) and Besag (1974). According to the MVL model, the probability of market basket i is:

$$p(\mathbf{y}_i) = \frac{\exp(\mathbf{a}'\mathbf{y}_i + \mathbf{y}_i'\mathbf{V}\mathbf{y}_i)}{Z_{\text{MVL}}} \quad (1)$$

with

$$Z_{\text{MVL}} = \sum_{\mathbf{y} \in \{0,1\}^J} \exp(\mathbf{a}'\mathbf{y} + \mathbf{y}'\mathbf{V}\mathbf{y})$$

Column vector \mathbf{a} contains constants for the categories. The symmetric $J \times J$ matrix \mathbf{V} contains zero diagonal elements ($V_{jj} = 0$) and symmetric pairwise cross-category coefficients ($V_{jl} = V_{lj}$). Paths in the upper part of Fig. 1 symbolize constants and pairwise cross-category coefficients of the MVL model. A positive (negative) cross-category coefficient V_{jl} indicates that categories j and l are purchased together more (less) frequently compared to conditional independence. On the other hand, a cross category coefficient equal to zero implies conditional independence.

The normalization constant Z_{MVL} is obtained by summing over the 2^J possible market baskets. Dividing by this normalization constant ensures that expression (1) provides a valid probability. That is why the MVL model can also be interpreted as multinomial logit model which considers choice of one of 2^J alternatives (Russell and Petersen 2000).

Conditional probabilities of purchase incidences y_{ij} for $j = 1, \dots, J$ given purchase incidences of the other $J - 1$ categories $\mathbf{y}_{i,-j}$ are derived from expression (1). They have binary logit functional forms:

$$p(y_{ij}|\mathbf{y}_{i,-j}) = 1 / \left(1 + \exp \left(- \left(a_j + \sum_{l \neq j} V_{jl} y_{il} \right) \right) \right) \quad \text{for all } i, j \quad (2)$$

The RBM introduced by Smolensky (1986) encompasses a column vector $\mathbf{h}_i = (h_{i1}, \dots, h_{iK})'$ of K binary hidden variables. Each hidden variable is connected to each observed binary variable (here: to each purchase incidence). The RBM is called restricted, because it does not have connections among hidden variables as well as among observed variables (see lower part of Fig. 1).

According to the RBM, the joint probability of market basket \mathbf{y}_i and hidden variables \mathbf{h}_i is:

$$p(\mathbf{y}_i, \mathbf{h}_i) = \frac{\exp(\mathbf{b}'\mathbf{y}_i + \mathbf{h}_i'\mathbf{W}\mathbf{y}_i)}{Z_{\text{RBM}}} \quad (3)$$

with

$$Z_{\text{RBM}} = \sum_{\mathbf{y} \in \{0,1\}^J} \sum_{\mathbf{h} \in \{0,1\}^K} \exp(\mathbf{b}'\mathbf{y} + \mathbf{h}'\mathbf{W}\mathbf{y})$$

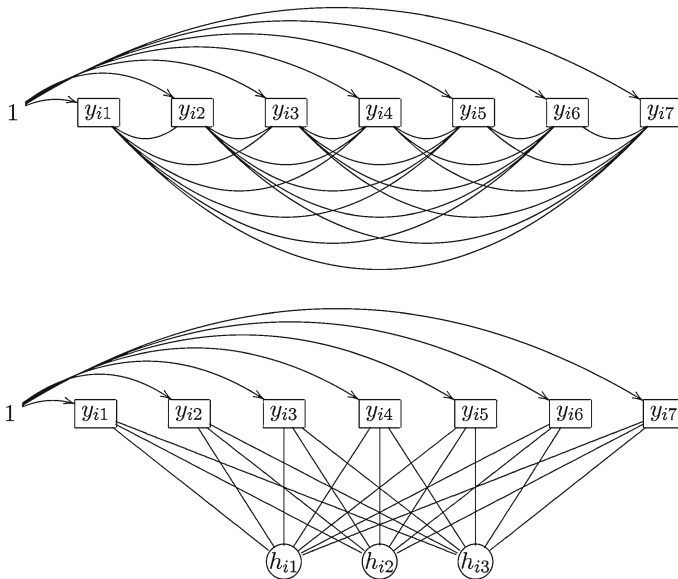


Fig. 1 Path diagrams of MVL and RBM models (*top* MVL, *bottom* RBM)

Column vector \mathbf{b} contains constants for the categories. $K \times J$ Matrix \mathbf{W} contains coefficients which link each hidden variable to each category. Contrary to the MVL model cross effects between categories do not depend on pairwise coefficients, but on coefficients linking purchase incidence and hidden variables (Sect. 4 deals with measurement of cross effects for the RBM). The normalization constant Z_{RBM} is obtained by summing over the 2^J possible market baskets $\mathbf{y} \in \{0, 1\}^J$ and the 2^K possible configurations of hidden variables $\mathbf{h} \in \{0, 1\}^K$. Dividing by the normalization constant Z_{RBM} ensures that expression (3) gives a valid probability.

From expression (3), we obtain the probability $p(\mathbf{y}_i)$ of market basket \mathbf{y}_i by summing over the 2^K possible configurations of hidden variables $\mathbf{h} \in \{0, 1\}^K$:

$$p(\mathbf{y}_i) = \frac{p^*(\mathbf{y}_i)}{Z_{\text{RBM}}} \quad \text{with} \quad p^*(\mathbf{y}_i) = \sum_{\mathbf{h} \in \{0,1\}^K} \exp(\mathbf{b}'\mathbf{y}_i + \mathbf{h}'\mathbf{W}\mathbf{y}_i) \quad (4)$$

In the following, we call $p^*(\mathbf{y}_i)$ unnormalized probability.

Conditional probabilities of purchase incidences y_{ij} given hidden variables \mathbf{h}_i for $j = 1, \dots, J$ and hidden variables h_{ik} given purchase incidences \mathbf{y}_i for $k = 1, \dots, K$ are derived from expression (3). Quite similar to the MVL we get binary logit functional forms:

$$p(y_{ij}|\mathbf{h}_i) = 1 / \left(1 + \exp \left(- \left(b_j + \sum_k w_{kj} h_{ik} \right) \right) \right) \quad \text{for all } i, j \quad (5)$$

$$p(h_{ik}|\mathbf{y}_i) = 1 / \left(1 + \exp \left(- \sum_j w_{kj} y_{ij} \right) \right) \quad \text{for all } i, k \quad (6)$$

3 Estimation and model comparison

Computation of the normalization constant of the MVL model makes ML estimation for higher numbers of categories impractical, because it requires in each iteration to sum over all 2^J possible market baskets, e.g., over more than 10^{18} market baskets in the case of 60 categories (see expression (1)).

Looking at expression (3) one could get the impression that one faces the same problem with the RBM, but by factorizing over categories, we can rewrite its normalization constant with $\mathbf{W}_{\cdot,j}$ denoting the j th column of \mathbf{W} as follows (Larochelle et al. 2010):

$$Z_{\text{RBM}} = \sum_{\mathbf{h} \in \{0,1\}^K} \prod_{j=1}^J \left(1 + \exp(\mathbf{W}'_{\cdot,j} \mathbf{h} + b_j)\right) \quad (7)$$

Expression (7) makes clear that estimation is significantly less-involved compared to the MVL model, if the number of hidden variables is much lower than the number of categories. For example, one has to sum and multiply over only $2^5 \times 60 = 1,920$ terms in the case of five hidden variables and 60 categories.

Therefore a moderate number of hidden variables allows us to estimate RBMs by ML. For ML estimation unnormalized probabilities defined in (4) are rewritten with $\mathbf{W}_{k,\cdot}$ denoting the k th row of \mathbf{W} (Larochelle et al. 2010):

$$p^*(\mathbf{y}_i) = \exp(\mathbf{b}' \mathbf{y}_i) \prod_{k=1}^K \left(1 + \exp(\mathbf{W}_{k,\cdot} \mathbf{y}_i)\right) \quad (8)$$

To our knowledge, our paper is the first to develop and apply ML estimation of RBMs. Of course, estimation of RBMs with a high number of hidden variables is only practical by approximate methods. The most widespread of these approximate methods is contrastive-divergence estimation (Hinton 2002) which in each iteration includes blocked Gibbs updates of expressions (5) and (6).

We estimate coefficients of all considered RBMs by maximizing the log likelihood (LL) across baskets (more details on ML estimation of RBMs can be found in Appendix 1):

$$\text{LL} = \sum_i \log p(\mathbf{y}_i) = \sum_i \log p^*(\mathbf{y}_i) - I \log Z_{\text{RBM}} \quad (9)$$

According to the independence model, the relative frequency of any basket configuration equals the product of relative marginal frequencies of purchases or non-purchases of the categories. The independence model can be obtained as the special case of the RBM by setting $\mathbf{W} = 0$. ML estimates of the coefficients of the independence model are:

$$b_j = \log \left(\sum_i y_{ij} \right) - \log \left(I - \sum_i y_{ij} \right) \quad \text{for } j = 1, \dots, J \quad (10)$$

The normalization constant Z_0 of the independence model equals:

$$Z_0 = 2^J \prod_j (1 + \exp(b_j)) \quad (11)$$

As ML estimation of the MVL model is not possible for a higher number of categories, we estimate its coefficients by the maximum pseudo-likelihood method (MPL) introduced by [Besag \(1974\)](#). The pseudo-likelihood of the MVL model equals the product of conditional probability functions across all J categories which are derived from expression (2). Its log, the log pseudo-likelihood (LPL), can be written as:

$$\begin{aligned} \text{LPL} &= \sum_i \text{LPL}_i \\ \text{with } \text{LPL}_i &= \sum_j [Y_{ij} \log(p(y_{ij}|y_{i,-j})) + (1 - Y_{ij})(1 - \log(p(y_{ij}|y_{i,-j})))] \end{aligned} \quad (12)$$

LPL_i denotes the log pseudo-likelihood of basket i . [Comets \(1992\)](#) showed that MPL estimates are consistent and asymptotically normal. Appendix 2 provides further information on MPL estimation of the MVL model.

The models which we compare consist of RBMs which differ with regard to the number of hidden variables, the independence model and the MVL model. Comparisons of RBMs and the independence model are based on log likelihood values. We measure performance of these models by log likelihood values both for estimation data and for validation data. Log likelihood values for validation data are computed using coefficients determined for the estimation data. In addition, we also compute values of the Bayesian information criterion (BIC) for the models investigated (LL is the loglikelihood value for the estimation data, n the number of parameters of a model):

$$\text{BIC} = -2\text{LL} + n \log I \quad (13)$$

To compare the RBM and the MVL model, we cannot use likelihood for the first, and pseudo-likelihood values for the latter. That is why we follow a predictive model selection approach and use the absolute deviation between observed and model generated purchase incidences as evaluation criterion ([Carlin and Louis 2000](#); [Laud and Ibrahim 1995](#)). We generate 500 artificial datasets each with 10,000 market baskets by Gibbs sampling over the appropriate conditional probability functions (i.e., for the MVL model over expression (2) using MPL coefficients and for the RBM over expressions (5) and (6) using ML coefficients). We define the absolute deviation (AD) with \hat{y}_{ijs} denoting the s th sampled purchase incidence value of category j in basket i as follows:

$$\text{AD} = \frac{1}{500} \sum_{s=1}^{500} \sum_i \sum_j |y_{ij} - \hat{y}_{ijs}| \quad (14)$$

Gibbs sampling produces samples from the joint distribution (expressions (1) and (3) for the MVL and the RBM, respectively) by sampling from the conditional distribution of each variable conditional on current values of the other variables. In their seminal paper, [Geman and Geman \(1984\)](#) proved that samples generated this way are a Markov chain with a unique stationary distribution which equals the joint distribution (also see [Casella and George 1992](#) for another explanation of Gibbs sampling). Blockwise sampling is possible for RBMs, because according to their respective conditional distributions, purchases only depend on hidden variables (expression (5)) and hidden variables only depend on purchases (expression (6)).

We assess the statistical significance of coefficients of the RBM by parametric bootstrap (see, e.g., [Cameron and Trivedi 2005](#)). For each of 500 artificial data sets generated on the basis of the RBM by blockwise Gibbs sampling we compute ML estimates. We calculate t -statistics of coefficients by dividing means by standard deviations across these 500 estimates. The parametric bootstrap allows to assess significance of quantities which are functions of coefficients quite easily using these 500 estimates of coefficient vectors.

4 Marginal cross effects

To derive marginal cross effects, we apply mean field theory ([Hertz et al. 1991](#); [Salakhutdinov and Hinton 2010](#)) and replace the fluctuating values of purchases of categories and hidden variables by shares, i.e., average values. Then the share $\langle y_j \rangle$ of category j and the share $\langle h_k \rangle$ of hidden variable k can be calculated as:

$$\langle y_j \rangle = 1 / \left(1 + \exp \left(- \left(b_j + \sum_k W_{kj} \langle h_k \rangle \right) \right) \right) \quad (15)$$

$$\langle h_k \rangle = 1 / \left(1 + \exp \left(- \sum_j W_{kj} \langle y_j \rangle \right) \right) \quad (16)$$

We define the marginal cross effect of category l on category j by the first derivative of the purchase share of category j given in expression 15 with respect to the purchase share of another category l :

$$\begin{aligned} \frac{\partial \langle y_j \rangle}{\partial \langle y_l \rangle} &= \langle y_j \rangle (1 - \langle y_j \rangle) \sum_k W_{kj} \frac{\partial \langle h_k \rangle}{\partial \langle y_l \rangle} \\ &= \langle y_j \rangle (1 - \langle y_j \rangle) \sum_k W_{kj} W_{kl} \langle h_k \rangle (1 - \langle h_k \rangle) \end{aligned} \quad (17)$$

A positive (negative) cross effect indicates that a higher share of category l is associated with a higher (lower) share of category j . Marginal cross effects are positive (negative) if $W_{kj} W_{kl} > 0$ ($W_{kj} W_{kl} < 0$) for each k . In other words, the cross effects between two categories are positive (negative) if their coefficients have the same (different) signs for each hidden variable. Note that marginal cross effects are asymmetric

(i.e., $\partial \langle y_j \rangle / \partial \langle y_l \rangle \neq \partial \langle y_l \rangle / \partial \langle y_j \rangle$) except for special constellations ($\langle y_j \rangle = \langle y_l \rangle$ or $\langle y_j \rangle = 1 - \langle y_l \rangle$). Marginal cross effects for two independent categories l and j are zero which follows from the fact that their coefficients W_{kj} , W_{kl} equal zero for each k .

Asymmetric marginal cross effects are not peculiar to the RBM. Although coefficients of the MVL are symmetric (i.e., $V_{jl} = V_{lj}$), we obtain asymmetric marginal cross effects for this model as well, because $\langle y_j \rangle (1 - \langle y_j \rangle) V_{jl}$ usually differs from $\langle y_l \rangle (1 - \langle y_l \rangle) V_{jl}$.

For the RBM, estimates of $\langle y_j \rangle$ and $\langle h_k \rangle$ cannot be determined analytically. We therefore set $\langle y_j \rangle$ and $\langle h_k \rangle$ in expression (17) equal to averages of categories and averages of hidden variables in 10,000 pseudo-baskets. We generate these pseudo-baskets by blockwise Gibbs sampling of expressions (5) and (6) using coefficients estimated by ML. We repeat this procedure for each of the 500 coefficient vectors obtained by parametric bootstrap which was described in Sect. 3. Dividing means by standard deviations over 500 estimated marginal cross effects gives t values of marginal cross effects.

We interpret each hidden variable by its larger (in absolute terms) partial marginal cross effects. We define a partial marginal cross effect between categories j and l as contribution of hidden variable k to the overall marginal cross effect between these two categories:

$$\langle y_j \rangle (1 - \langle y_j \rangle) W_{kj} W_{kl} \langle h_k \rangle (1 - \langle h_k \rangle) \quad \text{for } k = 1, \dots, K \quad (18)$$

5 Empirical study

5.1 Data

Twenty-thousand market baskets collected at a medium-sized supermarket during four consecutive Saturdays are randomly divided into two data sets of equal size. One set (estimation data) is required for estimation, the second set (validation data) is used to determine the predictive accuracy of models. Out of a total of 209 categories, we analyze the 60 categories with the highest univariate marginal purchase frequencies (see Table 1). This number of categories is five times as big as the maximum number of categories in the previous studies in which MVP or MVL models are simultaneously estimated (see Sect. 1). Table 2 contains the relative bivariate marginal frequencies for the 60 most frequent pairs of product categories.

5.2 Estimation results

All the RBMs considered perform better than the independence model in terms of log likelihood values and values of the Bayesian information criterion. This result clearly shows that cross effects between categories should not be ignored.

Log likelihood values of RBMs increase with the number of hidden variables. Values of the Bayesian information criterion improve with the number of hidden variables also. RBMs turn out to be robust, as log likelihood values almost stay the same for the validation data (Table 3).

Table 1 Relative univariate marginal frequencies of analyzed categories

Fruits	0.311	Bread	0.308	Vegetables	0.257
Yogurt	0.216	Periodicals	0.212	Milk	0.193
Chocolate	0.152	Soft drinks	0.148	Cigarettes	0.141
Beer	0.141	Juices	0.131	Cheese	0.125
Butter	0.125	UHT milk	0.110	Fat and oil	0.108
Pickles	0.104	Soups and sauces	0.103	Water	0.102
Sliced cheese	0.101	Baking ingredients	0.095	Spice	0.093
Seasonal	0.093	Candy	0.092	Snacks	0.079
Rolls	0.079	Coffee	0.078	Foil	0.076
Chocolate candy	0.073	Pasta	0.072	Rusk	0.072
Hygiene	0.070	Eggs	0.067	Hair care	0.066
UHT milk	0.060	Gifts	0.060	Body care	0.059
Delicatessen 1	0.059	Canned vegetables	0.056	Toilet articles	0.052
Cat food	0.051	Dental care	0.051	Sugar	0.050
Canned fish	0.050	Dried fruits	0.047	Sparkling wine	0.047
Wholemeal bread	0.046	Flour	0.042	Rice and legume	0.041
Tea	0.041	Delicatessen 2	0.041	Spread	0.039
Salt and vinegar	0.038	Frozen poultry	0.035	Red wine	0.033
Canned fruits	0.033	Office supplies	0.032	Liquor	0.031
Greeting cards	0.030	Cleanser	0.028	White wine	0.028

Whereas the RBM with four hidden variables implies 14 additional significant cross effects compared to the RM with three hidden variables (see Table 4), the significant cross effects of the RBM with five hidden variables consist of the same category pairs like the RBM with four hidden variables. Moreover, bootstrapped t values of differences of cross effects between these models are not significant. As we are mainly interested in cross effects, we present only estimation results of the less-complex RBM with four hidden variables in the following, though log likelihood values and values of the Bayesian information criterion of the RBMs with five hidden variables are better.

By computing absolute deviations as explained in Sect. 3, we obtain rounded values of 117, 796 and 97, 818 for the MVL model and the RBM model with four hidden variables, respectively. We think that the lower number of coefficients is one reason for the better performance of the RBM. For 60 categories one has to estimate $1,830 = 60 + 60 \times 59/2$ coefficients for the MVL model, but only $300 = 60 + 4 \times 60$ coefficients for the RBM model with four hidden variables.

Marginal cross effects implied by the RBM with four hidden variables whose absolute t values are >1.65 are given in Table 4 and also illustrated by the graph of Fig. 2.

For positive (negative) W_{kj} , the conditional probability of hidden variable k being equal to one (zero) increases if category j is purchased. The conditional probability of the first hidden variable being equal to one (zero) increases, for example, if beer

Table 2 Relative bivariate marginal frequencies

Fruits, vegetables	0.146	Fruits, bread	0.141	Bread, vegetables	0.122
Fruits, yogurt	0.120	Fruits, milk	0.106	Bread, yogurt	0.105
Bread, milk	0.098	Vegetables, yogurt	0.098	Vegetables, milk	0.097
Yogurt, milk	0.091	Bread, periodicals	0.087	Fruits, periodicals	0.082
Vegetables, periodicals	0.072	Fruits, butter	0.071	Fruits, chocolate	0.071
Fruits, cheese	0.069	Bread, sliced cheese	0.068	Bread, butter	0.066
Vegetables, butter	0.064	Bread, cheese	0.063	Vegetables, cheese	0.062
Fruits, fat	0.060	Yogurt, cheese	0.059	Yogurt, periodicals	0.059
Bread, chocolate	0.059	Fruits, UHT milk	0.059	Fruits, juices	0.059
Fruits, sliced cheese	0.057	Milk, butter	0.057	Fruits, pickles	0.057
Bread, juices	0.057	Yogurt, butter	0.056	Bread, UHT milk	0.056
Fruits, soups	0.055	Bread, cigarettes	0.054	Vegetables, fat	0.054
Periodicals, milk	0.053	Vegetables, soups	0.052	Vegetables, chocolate	0.052
Vegetables, pickles	0.052	Bread, fat	0.052	Vegetables, sliced cheese	0.052
Bread, pickles	0.051	Bread, soups	0.051	Vegetables, UHT milk	0.051
Fruits, bak. ingred.	0.050	Milk, cheese	0.050	Yogurt, chocolate	0.049
Vegetables, juices	0.049	Fruits, spices	0.049	Bread, soft drinks	0.048
Fruits, cigarettes	0.048	Yogurt, Fat	0.048	Yogurt, juices	0.048
Fruits, seasonal	0.047	Yogurt, UHT milk	0.046	Yogurt, sliced cheese	0.045
Vegetables, spices	0.045	Bread, spices	0.044	Yogurt, pickles	0.043

Contains the 60 most frequent pairs of categories. 46.09 % of the baskets with more than one category do not contain any of these 60 pairs

Table 3 Comparison of alternative models

Number of hidden variables	Number of parameters	Log likelihood (LL)		Bayesian Information Criterion (BIC)
		Estimation data	Validation data	
Restricted Boltzmann machines (RBMs)				
1	120	−171,677	−171,695	344,460
2	180	−170,815	−170,825	343,288
3	240	−170,071	−170,090	342,353
4	300	−169,704	−169,710	342,171
5	360	−169,108	−169,116	341,532
Independence model	60	−172,734	−172,842	346,020

LL and BIC values are rounded to the nearest integer value; RBMs differ with respect to the number of hidden variables

(bread) is purchased. To give another example, the conditional probability of the fourth hidden variable being equal to one (zero) increases if cigarettes (vegetables) are purchased. Table 5 provides coefficients W_{kj} linking hidden variables and categories whose absolute values are greater than 1.10. A value greater than 1.10 (less than −1.10)

Table 4 Significant marginal cross effects for the RBM with four hidden variables

Fruits, vegetables	0.357	Vegetables, fruits	0.355	Fruits, cheese	0.196
Vegetables, cheese	0.195	Fruits, milk	0.191	Vegetables, milk	0.190
Fruits, water	0.188	Vegetables, water	0.187	Fruits, butter	0.180
Vegetables, butter	0.179	Milk, vegetables	0.162	Milk, fruits	0.161
Fruits, pasta	0.159	Vegetables pasta	0.158	Fruits, seasonal	0.157
Vegetables seasonal	0.156	Fruits beer	−0.154	Vegetables, beer	−0.152
<i>Fruits, soups</i>	0.152	<i>Vegetables, soups</i>	0.152	Fruits, yogurt	0.152
Vegetables, yogurt	0.151	Fruits, fat	0.149	Vegetables, fat	0.149
Fruits, eggs	0.146	Vegetables, eggs	0.146	<i>Yogurt, vegetables</i>	0.138
<i>Yogurt, fruits</i>	0.137	<i>Fruits, spices</i>	0.133	<i>Vegetables, spices</i>	0.133
<i>Fruits, pickles</i>	0.130	<i>Fruits, tea</i>	0.130	<i>Vegetables, pickles</i>	0.130
<i>Vegetables, tea</i>	0.129	Fruits, sliced cheese	−0.124	Vegetables, sliced cheese	−0.123
<i>Fruits, UHT milk</i>	0.117	<i>Vegetables, UHT milk</i>	0.116	<i>Cheese, vegetables</i>	0.113
<i>Cheese, fruits</i>	0.112				

Absolute t values >1.65; first: independent category, second: dependent category; italic: additional effects compared to the RBM with three hidden units

means that the odds of hidden variable k defined as $p(h_{ik}|y_i)/(1 - p(h_{ik}|y_i))$ are multiplied by a factor greater than 3 (less than 1/3) if category j is purchased (these factors equal $\exp(W_{kj})$).

We interpret the four hidden variables by means of partial marginal cross effects as defined in expression 18 (see Table 6). All four hidden variables contribute to the positive interdependence of fruits and vegetables. Hidden variables 1 and 2 indicate that fruits and vegetables exert negative effects on sliced cheese.

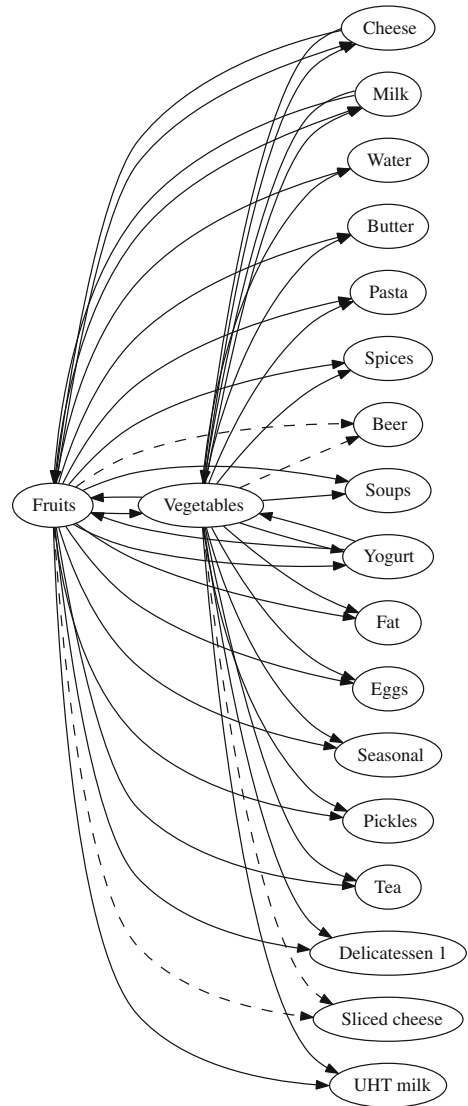
We now deal with partial cross effects which are specific to a hidden variable. Hidden variable 1 contributes on one hand to negative effects of periodicals, bread, milk, and yogurt on beer, on the other hand to negative effects of beer on fruits and vegetables. Hidden variable 3 reflects positive effects of both fruits and vegetables on soft drinks and tea. For hidden variable 4, both fruits and vegetables exert positive effects on fat, soups, and spices, negative effects on chocolate and cigarettes. Moreover, hidden variable 4 contributes to negative effects of both periodicals and cigarettes on vegetables and fruits.

Note that none of the categories having the 26 lowest univariate marginal frequencies is part of any significant cross effect (in Table 1 these are the categories beginning with the category gifts). Fruits and vegetables form the center of the graph of Fig. 2 as they are part of each of the significant cross effects. Fruits and vegetables are interdependent and exert a positive effect on milk, yogurt, and cheese. Milk, yogurt and cheese, on the other hand, have a positive effect on vegetables and fruits.

The majority of the cross effects of fruits and vegetables run one-way. These cross effects are positive with respect to water, butter, pasta, spices, soups, fat, eggs, seasonal, pickles, tea, delicatessen 1 and UHT milk.

By also looking at Table 1, it is possible to examine whether more frequently purchased categories tend to exert significant cross effects or tend to be subject to

Fig. 2 Graph of marginal cross effects (*solid*: positive, *dashed*: negative)



significant cross effects. Equating categories with high purchase frequencies with categories being part of high marginal cross effects turns out to be misleading (e.g., for six of the nine most frequently purchased categories, namely bread, periodicals, chocolate, soft drinks, cigarettes). This result can be explained by a relative high share of purchases whose baskets do not contain another category for these categories. The lowest share of such single category baskets amounts to 0.073 for these six categories. The corresponding highest share for the remaining three of the nine most frequently purchased categories (namely fruits, yogurt, milk) equals 0.003.

Table 5 Selected W_{kj} coefficients

Category	Hidden variable			
	1	2	3	4
Bread	-1.21			
Vegetables	-1.78	-1.98	-2.24	-2.75
Yogurt	-1.20	-1.11		
Periodicals	-1.59		1.19	1.24
Milk	-1.43		-1.27	-1.15
Soft drinks			1.16	
Cigarettes				1.84
Beer	2.57		1.53	
Cheese	-1.17	-1.21	-1.35	-1.22
Butter		-1.16	-1.34	
Fat and oil	-1.21			
Soups and sauces		-1.03		
Water	-1.17	-1.16	-1.18	-1.25
Sliced cheese	1.27	2.37	1.34	
Seasonal	-1.11		-1.19	
Pasta			-1.12	-1.20
Eggs			-1.11	
Canned vegetables			-1.25	
Cat food			-1.19	
Dry fruits			-1.12	

Only coefficients with absolute values >1.10 ; all have absolute t values $\gg 2.00$

Moreover, comparing Table 4 and Fig. 2 to Table 2 makes it clear that the results of our model differ from those of a simple pairwise cross-tabulation of purchase incidences in several respects:

- High pairwise frequencies of bread and periodicals with other categories are due to the high relative marginal purchase frequencies of bread and periodicals and not related to cross effects with other categories, i.e., the corresponding marginal cross effects are not significant (bread with fruits, vegetable, yogurt, milk, periodicals, sliced cheese, butter and cheese; periodicals with bread, fruits and vegetables).
- Of course, it is not possible to discover negative marginal cross effects by means of pairwise purchase frequencies (for our data this holds for the cross effects of fruits and vegetables on both beer and sliced cheese).
- High pairwise purchase frequencies between yogurt and milk, cheese and butter are traced back to the high cross effect of fruits and vegetables on these categories.

6 Conclusion

Extending the well-known MVL model by adding binary hidden variables enables us to analyze a large number of categories. For a moderate number of hidden variables, we can estimate the resulting model by ML and do not have to resort to two-step

Table 6 Partial marginal cross effects for each hidden variable

Hidden variable 1		Hidden variable 2	
Fruits, beer	−0.076	Fruits, sliced cheese	−0.072
Vegetables, beer	−0.074	Vegetables, sliced cheese	−0.071
<i>Periodicals, beer</i>	−0.056	Fruits, vegetables	0.061
<i>Bread, beer</i>	−0.056	Vegetables, fruits	0.061
Fruits, vegetables	0.054		
Vegetables, fruits	0.053		
<i>Beer, fruits</i>	−0.050		
<i>Milk, beer</i>	−0.050		
<i>Beer, vegetables</i>	−0.049		
Fruits, periodicals	0.048		
Vegetables, periodicals	0.046		
<i>Yogurt, beer</i>	−0.045		
Hidden variable 3		Hidden variable 4	
Fruits, vegetables	0.096	Fruits, vegetables	0.147
Vegetables, fruits	0.095	Vegetables, fruits	0.146
Vegetables, beer	−0.066	<i>Vegetables, cigarettes</i>	−0.098
Fruits, beer	−0.065	<i>Fruits, cigarettes</i>	−0.098
Vegetables, cheese	0.058	Vegetables, water	0.067
Vegetables, sliced cheese	−0.058	Fruits, water	0.067
Fruits, cheese	0.058	Vegetables, periodicals	−0.066
Fruits, sliced cheese	−0.058	Fruits, periodicals	−0.066
Vegetables, butter	0.057	Vegetables, cheese	0.066
Fruits, butter	0.057	Fruits, cheese	0.065
Vegetables, milk	0.055	Vegetables, pasta	0.064
Fruits, milk	0.055	Fruits, pasta	0.064
Vegetables, seasonal	0.051	<i>Cigarettes, vegetables</i>	−0.064
Fruits, seasonal	0.051	<i>Cigarettes, fruits</i>	−0.063
Vegetables, water	0.051	Vegetables, milk	0.062
Fruits, water	0.050	Fruits, milk	0.062
Vegetables, periodicals	−0.050	<i>Vegetables, chocolate</i>	−0.058
Fruits, periodicals	−0.050	<i>Fruits, chocolate</i>	−0.057
<i>Vegetables, soft drinks</i>	−0.050	<i>Periodicals, vegetables</i>	−0.057
<i>Fruits, soft drinks</i>	−0.049	<i>Periodicals, fruits</i>	−0.056
Vegetables, pasta	0.048	<i>Vegetables, soups</i>	0.055
Fruits, pasta	0.048	<i>Fruits, soups</i>	0.055
Vegetables, eggs	0.048	<i>Vegetables, spices</i>	0.054
Fruits, eggs	0.047	Vegetables, butter	0.054
Milk, vegetables	0.048	<i>Fruits, spices</i>	0.054
Milk, fruits	0.046	Fruits, butter	0.054

Table 6 continued

Hidden variable 3		Hidden variable 4	
<i>Vegetables, tea</i>	0.046	Milk, vegetables	0.053
		<i>Fruits, tea</i>	0.045
		Milk, fruits	0.052
		<i>Vegetables, fat</i>	0.052
		<i>Fruits, fat</i>	0.052
		Vegetables, eggs	0.050
		Fruits, eggs	0.049
		Vegetables, seasonal	0.049
		Fruits, seasonal	0.049

Contains partial marginal cross effects ≥ 0.045 with absolute t-values ≥ 1.65 ; italic: partial marginal cross effects specific to a hidden variable

estimation methods in contrast to the previous studies which dealt with higher number of categories. To compare the RBM to the MVL, we estimate the latter by maximum pseudo-likelihood. For a validation data set, the RBM performs better than the MVL model estimated this way.

For our data, about 75 % of the baskets are reproduced by the independence model [(75 % is the relative frequency of hidden variables $h_k = 0$ for all k as estimated by blockwise Gibbs sampling over expressions (5) and (6)]. Still, 25 % of the baskets obviously cannot be adequately modeled if cross effects among categories are ignored. Log likelihood values of the RBM and the independence model both for estimation and validation data as well als BIC values of these models also give evidence to this fact.

On the basis of our model, we infer asymmetric marginal cross effects among variables. The number of significant cross effects we obtain appears to be manageable, i.e., 38 cross effects as opposed to the 1,770 possible pairwise effects for 60 categories. We found only four negative cross effects between categories. Our results show that assuming that categories with high marginal cross effects are simply categories with high marginal purchase frequencies or categories with high pairwise purchase frequencies may be misleading.

A study of [Russel and Kamakura \(1997\)](#) shows that due to positively correlated category preferences promotions in one category lead to higher sales not only in that category but also in other positively related categories. [Russell and Petersen \(2000\)](#) emphasize strong positive cross-category effects, because promotional activity in one category simultaneously causes higher sales in other related categories.

Therefore, knowledge on categories which are part of significant cross effects is of benefit to a manager who wants to increase the number of purchases in other categories. To this end, she can make the independent category more attractive to customers by, e.g., sales promotion, advertising or assortment decisions. We give a few examples for the supermarket whose data we analyzed.

Marketing measures in categories which have no effect on purchases of other categories are not appropriate to increase purchases in other categories. For our data, many most-frequent categories share this property, e.g., bread, periodicals, chocolate, and cigarettes.

If purchases across the assortment should be increased, marketing measures for fruits or vegetables are obvious because of the centrality of these two categories in terms of cross effects. Marketing measures for fruits and vegetables are also advisable if purchases should be increased in categories which are positively influenced (milk, yogurt, cheese, water, butter, pasta, spices, soups, fat, eggs, seasonal, tea, pickles, delicatessen 1, UHT milk). Marketing measures for milk, yogurt and cheese, on the other hand, are suitable to increase purchases of fruits and vegetables.

Assuming, e.g., the retailer wants to increase cheese purchases, the manager should undertake marketing measures for fruits or vegetables. Similar measures in categories like, e.g., bread, rolls, red wine, or beer, are not appropriate to increase cheese purchases. Though these products are often consumed together with cheese, they have no effect on cheese purchases.

Marketing measures in other categories do not make sense if the retailer wants to increase purchases of beer or sliced cheese. Management should even avoid such measure for the vegetables or fruits categories, because they have negative effects on beer or sliced cheese purchases.

Applications of the RBM in other application areas, e.g., non-food retailing or financial services, may be fruitful. Adding explanatory variables seems to be an obvious next research task. The development of similar models for market basket analysis based on other response variables (e.g., purchase rates or purchase quantities), though somewhat challenging, constitute interesting research problems.

Acknowledgements I want to thank two anonymous reviewers for their suggestions which helped me to improve this article. Thanks also go to Thomas Reutterer for his comments on an early version.

7 Appendix 1: Maximum likelihood estimation of restricted Boltzmann machines

We apply the BFGS algorithm (see, e.g., [Greene 2003](#)) in 50 random restarts and selected the estimated coefficients which lead to the maximum log likelihood value across these restarts for the estimation data.

Initial values b_j , $j = 1, \dots, J$ are set equal to the ML estimates for the independence model according to expression (10). Coefficients W_{kj} ($k = 1, \dots, K$ and $j = 1, \dots, J$) are initialized to random numbers from the normal distribution with mean zero and standard deviation equal to 0.5.

The first derivative of the log likelihood (9) of a RBM with regard to any coefficient θ is:

$$\frac{\partial LL}{\partial \theta} = \sum_i \frac{\partial \log p^*(y_i)}{\partial \theta} - I \frac{\partial \log Z_{\text{RBM}}}{\partial \theta} \quad (19)$$

In the following, we derive the components of first derivatives of the coefficients. From expression (8) we obtain the unnormalized log probability of a basket:

$$\log p^*(y_i) = \mathbf{b}' y_i + \sum_{k=1}^K \log (1 + \exp(\mathbf{W}_{k,\cdot} y_i)) \quad (20)$$

The various derivatives of the unnormalized log probability are therefore:

$$\frac{\partial \log p^*(\mathbf{y}_i)}{\partial b_j} = y_{ij} \quad (21)$$

$$\frac{\partial \log p^*(\mathbf{y}_i)}{\partial W_{kj}} = 1 / (1 + \exp(-(\mathbf{W}_{k,\cdot} \mathbf{y}_i + c_k))) y_{ij} \quad (22)$$

We investigate $\partial Z_{\text{RBM}} / \partial \theta$ because of

$$\frac{\partial \log Z_{\text{RBM}}}{\partial \theta} = \frac{1}{Z_{\text{RBM}}} \frac{\partial Z_{\text{RBM}}}{\partial \theta} \quad (23)$$

We write the normalization constant as sum over 2^K terms $Z_{\text{RBM}}^{(n)}$:

$$Z_{\text{RBM}} = \sum_{n=1}^{2^K} \prod_{j=1}^J (1 + \exp(\mathbf{W}'_{\cdot,j} \mathbf{h}^{(n)} + b_j)) = \sum_{n=1}^{2^K} Z_{\text{RBM}}^{(n)} \quad (24)$$

Because of $\partial Z_{\text{RBM}} / \partial \theta = \sum_{n=1}^{2^K} \partial Z_{\text{RBM}}^{(n)} / \partial \theta$ we give the various derivatives $\partial Z_{\text{RBM}}^{(n)} / \partial \theta$ with $h_k^{(n)}$ denoting the value that hidden variable k assumes in configuration n :

$$\frac{\partial Z_{\text{RBM}}^{(n)}}{\partial W_{kj}} = 1 / (1 + \exp(-(\mathbf{W}'_{\cdot,j} \mathbf{h}^{(n)} + b_j))) Z_{\text{RBM}}^{(n)} h_k^{(n)} \quad (25)$$

$$\frac{\partial Z_{\text{RBM}}^{(n)}}{\partial b_j} = 1 / (1 + \exp(-(\mathbf{W}'_{\cdot,j} \mathbf{h}^{(n)} + b_j))) Z_{\text{RBM}}^{(n)} \quad (26)$$

8 Appendix 2: Maximum pseudo-likelihood estimation of the multivariate logit model

We estimate the MVL model by means of the BFGS algorithm (see, e.g., [Greene 2003](#)) using first derivatives of the log pseudo-likelihood. After rewriting and simplifying we obtain the following expressions for first derivatives of the log pseudo-likelihood of any basket i with respect to constants a_j and cross-category coefficients V_{jl} from expressions (12) and (2):

$$\frac{\partial \text{LPL}_i}{\partial a_j} = y_{ij} - p(y_{ij} | \mathbf{y}_{i,-j}) \quad \text{for } j = 1, \dots, J \quad (27)$$

$$\frac{\partial \text{LPL}_i}{\partial V_{jl}} = 2y_{ij}y_{il} - y_{il}p(y_{ij} | \mathbf{y}_{i,-j}) - y_{ij}p(y_{il} | \mathbf{y}_{i,-l}) \quad \text{for } l > j \quad (28)$$

References

- Besag J (1974) Spatial interaction and the statistical analysis of lattice Systems. *J R Stat Soc Ser B* 36:192–236
- Betancourt R, Gautschi D (1990) Demand complementarities, household production, and retail assortments. *Mark Sci* 9:146–161
- Boztug Y, Hildebrandt L (2008) Modeling joint purchases with a multivariate MNL approach. *Schmalenbach Bus Rev* 60:400–422
- Boztug Y, Reutterer T (2008) A combined approach for segment-specific market basket analysis. *European J Operational Res* 187:294–312
- Brijs T, Karlis D, Swinnen G, Vanhoof K, Wets G, Manchanda P (2004) A multivariate Poisson mixture model for marketing applications. *Statist Neerl* 58:322–348
- Cameron AC, Trivedi PK (2005) *Microeconometrics*. Cambridge University Press, New York
- Carlin BP, Louis TA (2000) *Bayes and empirical Bayes methods for data analysis*. Cambridge University Press, Cambridge
- Casella G, George EI (1992) Explaining the Gibbs sampler. *The Am Stat* 46:167–174
- Chib S, Seetharaman PB, Strijnev A (2002) Analysis of multi-category purchase incidence decisions using IRI market basket data. In: Franses PH, Montgomery AL (eds) *Econometric models in marketing*. JAI, Amsterdam, p 92
- Comets F (1992) On consistency of a class of estimators for exponential families of Markov random fields on the lattice. *Ann Stat* 20:455–468
- Cox DR (1972) The analysis of multivariate binary data. *J R Stat Soc Ser C* 21:113–120
- Duvvuri SD, Gruca TS (2010) A Bayesian multi-level factor analytic model of consumer price sensitivities across categories. *Psychometrika* 75:558–578
- Duvvuri SD, Ansari A, Gupta S (2007) Consumers' price sensitivities across complementary categories. *Manag Sci* 53:1933–1945
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6:721–741
- Greene WH (2003) *Econometric analysis*, 6th edn. Prentice Hall, Upper Saddle River
- Hertz J, Krogh A, Palmer RG (1991) *Introduction to the theory of neural computation*. Addison-Wesley, Redwood City
- Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comp* 14:1771–1800
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313:504–507
- Hruschka H, Lukanowicz M, Buchta C (1999) Cross-category sales promotion effects. *J. Retail Consumer Serv* 6:99–105
- Larochelle H, Bengio Y, Turian J (2010) Tractable multivariate binary density estimation and the restricted Boltzmann forest. *Neural Comp* 22:2285–2307
- Laud PW, Ibrahim JG (1995) Predictive model selection. *J R Stat Soc Ser B* 57:247–262
- Manchanda P, Ansari A, Gupta S (1999) The "Shopping Basket": a model for multi-category purchase incidence decisions. *Mark Sci* 18:95–114
- Niraj R, Padmanabhan V, Seetharaman PB (2008) A cross-category model of households' incidence and quantity decisions. *Marketing Sci* 27:225–235
- Russel GJ, Kamakura WA (1997) Modeling multiple category brand preference with household basket data. *J Retail* 73:439–461
- Russell GJ, Petersen A (2000) Analysis of cross category dependence in market basket selection. *J Retail* 76:369–392
- Russell GJ, Bell D, Bodapati A, Brown CL, Chiang J, Gaeth G, Gupta S, Manchanda P (1997) Perspectives on multiple category choice. *Marketing Lett* 8:297–305
- Russell GJ, Ratneshwar S, Shocker AD, Bell D, Bodapati A, Degeratu A, Hildebrandt L, Kim N, Ramaswami S, Shankar VH (1999) Multiple category decision-making: review and synthesis. *Marketing Lett* 10:319–332
- Salakhutdinov RR, Hinton GE (2010) An efficient learning procedure for deep Boltzmann machines. Technical Report, Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge

- Smolensky P (1986) Information processing in dynamical systems: foundations of harmony theory. In: Rumelhart DE, McClelland JL (eds) *Parallel distributed processing: explorations in the microstructure of cognition*, vol 1., foundations MIT Press, Cambridge, pp 194–281
- Song I, Chintagunta PK (2007) A discrete-continuous model for multicategory purchase behavior of households. *J. Marketing Res* 44:595–612
- Wang H, Kalwani MU, Akçura T (2007) A Bayesian multivariate Poisson regression model of cross-category store brand purchasing behavior. *J Retail Consumer Serv* 14:369–382
- Wedel M, Kamakura WA (1998) *Market segmentation*. Kluwer, Boston