



# Analyzing joint brand purchases by conditional restricted Boltzmann machines

Harald Hruschka<sup>1</sup> 

Received: 14 July 2020 / Accepted: 2 June 2021  
© The Author(s) 2021

## Abstract

We introduce the conditional restricted Boltzmann machine as method to analyze brand-level market basket data of individual households. The conditional restricted Boltzmann machine includes marketing variables and household attributes as independent variables. To our knowledge this is the first study comparing the conditional restricted Boltzmann machine to homogeneous and heterogeneous multivariate logit models for brand-level market basket data across several product categories. We explain how to estimate the conditional restricted Boltzmann machine starting from a restricted Boltzmann machine without independent variables. The conditional restricted Boltzmann machine turns out to excel all the other investigated models in terms of log pseudo-likelihood for holdout data. We interpret the selected conditional restricted Boltzmann machine based on coefficients linking purchases to hidden variables, interdependences between brand pairs as well as own and cross effects of marketing variables. The conditional restricted Boltzmann machine indicates pairwise relationships between brands that are more varied than those of the multivariate logit model are. Based on the pairwise interdependences inferred from the restricted Boltzmann machine we determine the competitive structure of brands by means of cluster analysis. Using counterfactual simulations, we investigate what three different models (independent logit, heterogeneous multivariate logit, conditional restricted Boltzmann machine) imply with respect to the retailer's revenue if each brand is put on display. Finally, we mention possibilities for further research, such as applying the conditional restricted Boltzmann machine to other areas in marketing or retailing.

**Keywords** Marketing · Retailing · Market basket analysis · machine Learning · Restricted Boltzmann machine · Multivariate logit model

---

I thank one anonymous reviewer for suggestions and hints that really helped to improve the paper.

---

✉ Harald Hruschka  
harald.hruschka@ur.de

<sup>1</sup> University of Regensburg, Universitätsstrasse 31, 93040 Regensburg, Germany

JEL codes M31 · L81 · D12 · C45 · C89

## 1 Introduction

Managers and scientists alike use a wide range of methods to analyze market basket data. One group of methods is based on association rules (Agrawal and Srikant 1994; Hahsler et al. 2006). Another group consists of econometric models like the multivariate logit (Russell and Petersen 2000; Boztuğ and Reutterer 2008) or the multivariate probit model (Manchanda et al. 1999; Hruschka 2017). These econometric models typically include independent variables, e.g., marketing variables or household attributes, which may affect purchases. In contrast to multinomial choice models both the multivariate logit (MVL) and the multivariate probit (MVP) model allow pick-any choices, i.e., they consider that households may purchase multiple products at the same occasion.

Table 1 also provides an overview of several studies applying machine learning methods to market basket analysis.<sup>1</sup> The restricted Boltzmann machine (RBM) is a machine learning method frequently used to solve pattern recognition problems, e.g., recognition of handwritten digits or classification of documents (Hinton and Salakhutdinov 2006). The RBM excludes independent variables. To

**Table 1** Comparison to related studies

Paper	Models	Alternatives	Independent variables	Latent Heterogeneity
Aurier and Mejia (2014)	MVL	1 category, 4 brands	mv	Yes
Dubé (2004)	SM	1 category, 26 brands	ca, mv	Yes
Hruschka (2014)	MVL, RBM	60 categories	no	No
Hruschka (2021)	BF, LDA, CTM, RBM	169 categories	No	Yes
Jacobs et al. (2016)	MDM, LDA	394 brands in different categories	ca	Yes
Kim et al. (2002)	SM	1 category, 1 brand, 5 flavors	mv	Yes
Kwak et al. (2015)	MNL, MVL	1 category, 2 brands, 6 flavors	mv	Yes
Ruiz et al. (2020)	EFE, HPF, SHOPPER	374 categories	mv	Yes
Xia et al. (2019)	MVP, CRBM	100 categories	ca, mv	Yes
This paper	MVL, CRBM	10 categories, 42 brands	ca, mv	Yes

Abbreviations: *BF* binary factor analysis, *CRBM* conditional restricted Boltzmann machine, *CTM* correlated topic model, *EFE* exponential family embeddings, *HPF* hierarchical Poisson factorization, *LDA* latent Dirichlet allocation, *MDM* mixture of Dirichlet multinomials, *MNL* multinomial logit, *MVL* multivariate logit, *MVP* multivariate probit, *SM* structural model, *ca* customer attributes, *mv* marketing variables

<sup>1</sup> We thank one anonymous reviewer for drawing our attention to several references contained in this table.

our knowledge the first application to market basket analysis is Hruschka (2014) using data on 60 product categories and selecting a RBM with four hidden variables. This RBM outperforms a MVL model with category constants and pairwise interactions between categories on a holdout data set. In a later publication of the same author a RBM with three hidden variables turns out to be superior to topic models and binary factor analysis on another market basket data set of 169 grocery categories (Hruschka 2021).

This performance record of the RBM suggests the investigation of the conditional Boltzmann machine (CRBM) that extends the RBM by adding a set of independent variables. Because of the inclusion of independent variables, the CRBM becomes similar to the MVL or MVP models used in market basket analysis. Very recently, Xia et al. (2019) apply the CRBM to market baskets. These authors refer to the study of Hruschka (2014) already mentioned, analyze data at the category level and compare the CRBM to the MVP model. Contrary to Xia et al. (2019) and in line with Hruschka (2014) we compare the CRBM to the MVL model. We take a more detailed perspective by looking at purchases at the brand level instead of the more aggregate category level. In this regard, our papers differs from Xia et al. (2019) and from the majority of studies analyzing market basket data.

Jacobs et al. (2016) apply latent Dirichlet allocation (LDA) which they extend to include customer attributes as independent variables. This way these authors reduce 394 categories of chemists' products offered by a Dutch online retailer to 13 topics. They compare LDA to a mixture of Dirichlet multinomials (MDM) without independent variables. LDA and MDM attain comparable predictive performance.

Ruiz et al. (2020) develop a probabilistic model which they call SHOPPER whose core is a sequential choice model for each item contained in a basket conditional on all previous chosen items. This sequential choice model has a multinomial logit form. Besides a constant term and a latent factor, any item not chosen so far SHOPPER encompasses factors to reproduce both interactions and attributes of the previously chosen items. The model also comprises several factored latent variables. Each of these factored latent variables consists of two components. One component is related to items, the other component is related to shopper preferences, seasonal effects, and price sensitivities, respectively. As market basket data usually lack the information on choice sequences, Ruiz et al. (2020) extend the basic model to consider all possible market baskets. In their empirical application Ruiz et al. (2020) use 100 latent variables related to item attributes, item interactions and shopper preferences. In addition, these authors choose ten latent variables for seasonal effects and price sensitivities.

Using market baskets from a large grocery store for 374 categories Ruiz et al. (2020) compare their model to hierarchical Poisson factorization (HPF) focusing on user preferences and to exponential family embeddings (EFE) focusing on item-to-item interactions. SHOPPER outperforms these related, but less complex models in terms of predicting item probability conditioned on the other observed items of a basket. The complexity of SHOPPER becomes apparent by the huge number of parameters for the latent variables (more than 400,000). For HPF and EFE the number of such parameters amounts to 358,000 and 74,800, respectively.

The few publications dealing with multiple purchases of brands or varieties of brands as a rule investigate one product category only (please also see Table 1). Dubé (2004) mentions that multiple purchases of brands belonging to the same category may be explained by variety seeking, uncertainty on their future tastes, or taste differences of household members. This author develops a structural model that generalizes the typical multinomial choice model by a latent variable, the integer number of consumption occasions. Each of these consumption occasions has its own set of corresponding preferences. Dubé (2004) analyzes 26 carbonated soft drink brands by this structural model. Aurier and Mejia (2014) provide evidence for multiple purchases of four chocolate brands and apply a MVL model at the brand level.

Kim et al. (2002) investigate purchases of five flavors of one yogurt brand. These authors develop a structural model based on an additive utility specification that generalizes the usual linear form. For this specification, utility maximization may lead to interior solution with more than one flavor chosen. Kwak et al. (2015) look at purchases of the yogurt category and combine a multinomial logit model for two brands with a MVL model. The later considers pick-any choices of six flavors of the chosen brand. Note that the total model of Kwak et al. (2015) excludes joint purchases of brands and works satisfactorily as long as their frequency is very low or zero.

We contribute over these publications dealing with brand purchases by considering several categories and allowing for pick-any choices at the brand level. A more restricted approach would account for pick-any choices at the category level only and assume that a household selects a single brand from each category by using multinomial choice models. This way one would ignore that especially in food categories a considerable percentage of households purchase multiple brands.

We measure pairwise interdependence by the purchase probability change of brand  $j' \neq j$  associated to a marginal increase of the probability of brand  $j$ . To compare these two models, we compute probabilities based on both an estimated MVL model and an estimated CRBM. We define two brands to be purchase complements if the probability change is positive and to be purchase substitutes if the probability change is negative. This definition is equivalent to the one put forward by Betancourt and Gautschi (1990), who consider two products as purchase complements (purchase substitutes) if they are purchased jointly more (less) frequently than expected under stochastic independence. Pairwise interdependences may be asymmetric, i.e., the probability change of brand  $j' \neq j$  due to a marginal increase of the probability of brand  $j$  may differ from the probability change of brand  $j$  due to a marginal increase of the probability of brand  $j'$ .

Both the CRBM and the MVL model allow that two brands of the same category may be purchase substitutes or purchase complements. In this regard, these models are more flexible than multinomial choice models that are restricted to substitutive relations between two brands. One brand may also be interdependent with brands of another category and these interdependences may vary across brands of the other category. Econometric models on the category level restrict two categories to be either purchase complements or purchase substitutes.

We present homogeneous MVL models, their finite mixture extension and the CRBM in the next section. Then we explain the basic estimation approach for these

models. The empirical part of our paper characterizes the analyzed data set, compares performance of several models on holdout data and interprets the best performing MVL model and CRBM. In the concluding section, we discuss managerial implications. A cluster analysis of pairwise interdependences shows the competitive structure of brands. We determine the effects of brand-specific marketing decisions on the retailer’s revenue for three selected models by means of counterfactual simulations. Finally, we also hint at possibilities for further research.

## 2 Investigated models

In this section we present homogeneous MVL models, their finite mixture extensions and the CRBM.  $J$  column vector  $y$  denotes a market basket and consists of binary purchases. If product  $j$  is purchased the respective element  $y_j$  equals one. If product  $j$  is not purchased  $y_j$  equals zero. Vector  $x$  consists of independent variables. We omit basket and household indices for the sake of simplicity.

### 2.1 Homogeneous multivariate logit models

In a homogeneous MVL model each coefficient is constant across households. The probability  $P(y|x)$  of one market basket  $y$  conditional on independent variables  $x$  is proportional to:

$$\exp(y^T \alpha + x^T \beta y + 1/2 y^T V y) \tag{1}$$

Coefficients contained in  $(J, J)$  matrix  $V$  measure pairwise interactions between products. As a pairwise interaction of a product with itself does not make sense, all diagonal elements of  $V$  are zero. Off-diagonal elements are symmetric, i.e.,  $V_{j_1, j_2} = V_{j_2, j_1}$ . Column vector  $\alpha$  consists of  $J$  constants. The  $(L, J)$  matrix  $\beta$  holds the effect of  $L$  independent variables on product purchases. Russell and Petersen (2000) apply the homogeneous MVL model to market basket data at the product category level building upon earlier publications in statistics (Cox 1972; Besag 1974).

For the MVL model we can write the purchase probability of product  $j$  conditional on purchases of the other products collected in vector  $y_{-j}$  as follows:

$$P(y_j = 1 | y_{-j}) = \sigma \left( \alpha_j + \beta_j x + \sum_{l \neq j} V_{j,l} y_l \right) \tag{2}$$

$\sigma(Z)$  denotes the binomial logistic function  $1/(1 + \exp(-Z))$ .

By setting all coefficients in  $V$  equal to zero we obtain the independent logit model that excludes interactions between products.

Following the suggestion of an anonymous reviewer, we investigate in addition another version of the MVL model by replacing the pairwise interactions between products in expression (1) by interactions between products and product categories. As the number of categories is lower than the number of products, the resulting

version is more parsimonious. Now the probability  $P(y|x)$  of one market basket  $y$  conditional on independent variables  $x$  is proportional to:

$$\exp(y^T \alpha + x^T \beta y + y^T V^{cat} y^{cat}) \tag{3}$$

The binary column vector  $y^{cat}$  consists of  $C$  elements. Its  $c$ -th element  $y_c^{cat}$  equals one, if the market basket contains at least one product which belongs to category  $c$ . The  $(J, C)$  coefficient matrix  $V^{cat}$  measures the interactions between products and categories. This version of the MVL model allows only interactions between any product and other categories. Consequently, an element  $V_{j,c}^{cat}$  is set to zero if product  $j$  belongs to category  $c$ .

We can write the purchase probability of product  $j$  conditional on purchases of other categories collected in vector  $y_{-ind_j}^c$  as follows:

$$P(y_j = 1 | y_{-ind_j}^c) = \sigma \left( \alpha_j + \beta_{.j} x + \sum_{c \neq ind_j} V_{j,c}^{cat} y_c^{cat} \right) \tag{4}$$

$ind_j$  denotes the index of the category to which product  $j$  belongs. The sum in expression (4) runs over all product categories to which product  $j$  does not belong, whereas the analogous sum in expression (2) runs across all products different from product  $j$ .

### 2.2 Finite mixture multivariate logit models

We also investigate finite mixtures of the MVL models (FM-MVL) in addition to their basic homogeneous versions presented in section 2.1. Because of the computational complexity caused by the high number of alternatives we do not consider MVL models with continuous heterogeneity in contrast to the models in Gentzkow (2007) or Richards et al. (2018), which encompass no more than four alternatives.

Coefficients of the FM-MVL models differ between household segments. The purchase probability of product  $j$  conditional on purchases of the other products collected in vector  $y_{-j}$  is a linear combination of segment-specific conditional probability functions for pairwise interactions and interactions between products and categories, respectively:

$$P(y_j = 1 | y_{-j}) = \sum_{s=1}^S \pi_s \sigma \left( \alpha_{sj} + \beta_{s,j} x + \sum_{l \neq j} V_{s,j,l} y_l \right) \tag{5}$$

$$P(y_j = 1 | y_{-j}) = \sum_{s=1}^S \pi_s \sigma \left( \alpha_{sj} + \beta_{s,j} x + \sum_{c \neq ind_j} V_{s,j,c}^{cat} y_c^{cat} \right) \tag{6}$$

$S$  denotes the number of segments,  $\pi_s$  the relative size of segment  $s$ .

### 2.3 Conditional restricted Boltzmann machine

For the CRBM the probability  $P(y|x)$  of one market basket  $y$  conditional on independent variables  $x$  is proportional to (Mnih et al. 2011):

$$\exp(y^T \alpha + x^T \beta y + \gamma^T h + y^T W h) \tag{7}$$

$K$  binary hidden variables contained in a vector  $h$  affect purchases by means of coefficients in a  $(J, K)$  matrix  $W$  linking each hidden variable to purchases of each product. The CRBM differs from the MVL model by the way it deals with product interdependences. The MVL models investigated consider pairwise coefficients either between products or between categories and products, whereas the CRBM looks at relations between products and hidden variables. Column vectors  $\alpha$  and  $\gamma$  hold  $J$  constants for products and  $K$  constants for hidden variables, respectively.  $W_{.k}$  denotes column  $k$  of matrix  $W$ . The CRBM is homogeneous, i.e., each coefficient is constant across households.

Li et al. (2015) use specification (7) to analyze image and video data with independent variables which directly affect binary dependent variables just like in the MVL model. In a more general specification independent variables may also directly affect hidden variable. Such a specification does not lead to a better performance for our data in spite of its greater complexity. Therefore we only discuss specification (7) in the following. Note that it allows indirect effects of independent variables on hidden variables due to their direct effect on purchases, which on their part influence hidden variables. This mechanism can be seen from the expressions for the conditional probabilities of purchases given hidden variables and for hidden variables given purchases (Li et al. 2015):

$$P(y_j = 1|h) = \sigma \left( \alpha_j + \beta_{j,x} x + \sum_{k=1}^K W_{jk} h_k \right) \tag{8}$$

$$P(h_k = 1|y) = \sigma \left( \gamma_k + \sum_{j=1}^J W_{jk} y_j \right) \tag{9}$$

If we do not allow effects of independent variables by setting all elements of  $\beta$  equal to zero, we obtain the RBM which is defined as joint Boltzmann distribution of hidden and observed variables (purchases). The RBM was introduced by Smolensky (1986) and consists of one layer of observed variables and one layer of binary hidden variables. The RBM is called restricted because variables of the same layer are not connected.

Let us mention several characteristics that the CRBM shares with the RBM. The probability of a market basket is proportional to the product of the probabilities that it would be generated by each of the hidden variables acting alone. If a hidden variable is zero, its separable probability distribution for each product is determined by its constants in  $\alpha$  only. But if a hidden variable is one, this distribution also depends on the coefficients linking the hidden variable to each product (Hinton 2002).

In a RBM distributions each specific to a hidden variable are multiplied first. The product of these distributions is normalized in the next step. This way sharp distributions may be detected. Mixture models on the other hand determine convex combinations of distributions that are normalized beforehand. For high dimensional data the mixture model approach may lead to problems, as the final distribution cannot be sharper than the distributions of the individual hidden variables each of which is adapted to all observed variables (Hinton 2002).

The hidden variables of a RBM produce  $K$  different partitions of the input space which define  $2^K$  possible regions (Bengio 2009). Therefore the RBM is capable to create exponentially many inference regions based on only a polynomial number of parameters. This property differentiates RBMs from mixture models (Montúfar 2016).

Le Roux and Bengio (2007) prove that the RBM can approximate any discrete distribution given a sufficient number of hidden variables. Therefore in contrast to the MVL model specified in expression (1) the RBM is not restricted to pairwise interactions, but is capable to reproduce higher order interactions as well.

### 3 Estimation

Maximum likelihood estimation of the MVL model requires computation of the so-called normalization constant in every iteration that is obtained by summing over all possible market baskets. We exclude the null basket which contains no purchases (for which all purchase indicators  $y_j$  equal zero) in accordance with previous related publications (Russell and Petersen 2000; Dubé 2004; Boztuğ and Reutterer 2008; Kwak et al. 2015). This way we model purchases conditional on the purchase of at least one product. Therefore, the number of possible market baskets is  $2^J - 1$ .

Only when we divide expression (1) by the normalization constant a proper probability results. For 42 products we would have to deal with more than  $4.39 \times 10^{12}$  possible market baskets. Because of the impracticality of this approach, we resort to maximum pseudo-likelihood (MPL) estimation. In a simulation study Bel et al. (2018) compare MPL to maximum likelihood estimation for a maximum number of 12 alternatives. These authors conclude that MPL estimation leads to negligible efficiency losses only.

Recently Kosyakova et al. (2020) have proposed an alternative estimation method for MVL models. This Markov chain Monte Carlo method avoids evaluation of the normalization constant.<sup>2</sup> Kosyakova et al. (2020) analyze menu-based choice experiments by the MVL model. In their empirical study, respondents are exposed to twelve menus. Each menu consists of the same thirteen products, but with different prices. Respondents are free to choose either none of or any combination of these products.

<sup>2</sup> We thank one anonymous reviewer for making us aware of this work.

We decide to use MPL estimation, as it is feasible for a higher number of alternatives (products). Moreover, we can compute pseudo-likelihoods for RBMs and CRBMs, which makes comparison to MVL models possible.

The pseudo-probability  $\tilde{P}_j$  for product  $j$  is defined as probability of  $y_j$  conditional on the observed basket  $y_{-j}$ , i.e., basket  $y$  without product  $j$ :

$$\tilde{P}_j \equiv \frac{P(y|y_{-j})}{P(y|y_{-j}) + P(\tilde{y}_{(-j)}|y_{-j})} \tag{10}$$

MPL estimation is feasible, because the normalization constant drops out in expression (10). The basket  $\tilde{y}_{(-j)}$  corresponds to the observed basket  $y$  except for product  $j$ , whose purchase indicator is flipped, i.e.,  $\tilde{y}_{(j)} = 1 - y_j$ .

For both the homogeneous and the finite mixture MVL models with pairwise interactions the pseudo-probability  $P_j$  for product  $j$  in basket  $y$  is given by:

$$\tilde{P}_j = P(y_j = 1|y_{-j})^{Y_j} + (1 - P(y_j = 1|y_{-j}))^{1-Y_j} \tag{11}$$

For the MVL model with interactions between product and categories we obtain instead:

$$\tilde{P}_j = P(y_j = 1|y_{-ind_j}^c)^{Y_j} + (1 - P(y_j = 1|y_{-ind_j}^c))^{1-Y_j} \tag{12}$$

For the CRBM we can write the pseudo-probability  $\tilde{P}_j$  by means of the so-called free energy  $F(y, x)$  (Mnih et al. 2011):

$$\tilde{P}_j = \frac{\exp(-F(y, x))}{\exp(-F(y, x)) + \exp(-F(\tilde{y}_{(-j)}, x))} \tag{13}$$

with  $F(y, x) = - \left[ y^T \alpha + x^T \beta y + \sum_{k=1}^K \log(1 + \exp(\gamma_k + y^T W_k)) \right]$

For all investigated models the log pseudo-likelihood  $LPL$  of basket  $y$  is obtained by summing the logs of pseudo-probabilities across all products

$$LPL = \sum_{j=1}^J \log(\tilde{P}_j) \tag{14}$$

Estimation of the homogeneous MVL model is straightforward as its pseudo-likelihood function has only one local maximum. By contrast, for the FM-MVL model as well as for the CRBM this function may have multiple local optima. That is why we randomly start estimation ten times both for a FM-MVL model or a CRBM (please see Appendices A, B and C for more details).

**Table 2** Descriptive statistics (1)

Brand					Mean shares		
	Market	Relative	Mean	Standard	Price		
	Share	Frequency	Price	deviation	Feature	Display	Reduction
Softdrinks: Low calorie (Share 57.31 %, Multiple brands purchases 13.89 % )							
Coca Cola	48.05	0.152	3.695	0.456	0.064	0.022	0.149
Pepsi	36.30	0.111	5.166	0.547	0.405	0.452	0.609
Dr Pepper	15.03	0.069	5.434	0.607	0.291	0.302	0.482
Other brands	0.62	0.005	5.870	0.557	0.058	0.242	0.355
Softdrinks: Regular (Share 41.24 %, Multiple Brands Purchases 16.47 %)							
Coca Cola	39.38	0.010	4.925	0.313	0.042	0.039	0.077
Pepsi	37.00	0.088	5.211	0.528	0.397	0.464	0.606
Dr Pepper	20.53	0.071	5.433	0.598	0.290	0.329	0.482
Other brands	3.09	0.017	6.041	0.536	0.074	0.235	0.356
Coffee: Ground (Share 67.61 %, Multiple Brands Purchases 2.04 %)							
Smucker	46.90	0.029	9.966	1.368	0.054	0.274	0.238
Zanetti	15.21	0.009	5.780	0.630	0.067	0.114	0.309
Kraft	13.38	0.007	5.333	1.426	0.076	0.128	0.366
Other brands	24.51	0.022	5.237	0.993	0.070	0.073	0.370
Coffee: Whole Beans (Share 15.02 %, Multiple Brands Purchases 1.56 %)							
Regency	46.26	0.007	9.328	1.062	0.000	0.024	0.061
Tata Tea	25.02	0.003	9.135	0.977	0.000	0.212	0.339
Other brands	28.72	0.004	7.967	0.657	0.0000	0.0000	0.104
Milk: Regular (Share 78.85 %, Multiple Brands Purchases 2.53%)							
Private label	40.59	0.229	0.933	0.089	0.011	0.000	0.071
Hood	32.40	0.198	0.390	0.060	0.034	0.000	0.139
Dean	24.79	0.141	0.558	0.190	0.023	0.000	0.081
Other brands	2.22	0.008	0.451	0.056	0.032	0.015	0.127
Milk: Flavored (Share 18.19 %, Multiple Brands Purchases 1.25 %)							
Hood	53.37	0.117	1.769	0.300	0.032	0.045	0.098
Dean	44.72	0.091	0.581	0.169	0.036	0.118	0.367
Other brands	1.91	0.004	0.507	0.083	0.062	0.116	0.325

## 4 Empirical study

### 4.1 Data

We analyze joint purchases, marketing variables and household attributes which originate from the academic household panel data set provided by SymphonyIRI Group (Bronnenberg et al. 2008). Specifically we consider purchases of brands in ten (sub-) categories made by households in two stores located in Eau Claire, Wisconsin, spanning the whole 2011 calendar year. We also compute yearly revenue

**Table 3** Descriptive statistics (2)

Brand	Mean shares						
	Market Share	Relative Frequency	Mean Price	Standard deviation Price	Feature	Display	Price Reduction
Hotdog (Share 100.00%, Multiple brands purchases 2.16%)							
Kraft	44.14	0.040	3.476	0.390	0.023	0.063	0.211
Buddig	12.45	0.007	2.999	0.424	0.066	0.112	0.328
Cher-Make	11.83	0.008	4.053	0.420	0.012	0.064	0.163
SaraLee	9.65	0.009	3.519	0.296	0.105	0.227	0.454
Other brands	21.93	0.028	3.436	0.451	0.064	0.047	0.293
Spaghetti sauce (share 100.00 %, Multiple brands purchases 3.18 %)							
Campbell	29.76	0.030	1.798	1.789	0.000	0.012	0.142
Unilever	24.00	0.023	1.493	0.162	0.049	0.186	0.308
Conagra	15.86	0.022	1.508	0.138	0.043	0.262	0.330
Heinz	8.73	0.007	0.782	0.010	0.096	0.214	0.370
Lidestri	7.81	0.010	2.246	0.219	0.010	0.023	0.204
Private label	5.65	0.008	1.209	0.259	0.066	0.106	0.312
Newmans	4.09	0.008	1.157	0.166	0.012	0.071	0.275
Other Brands	4.10	0.003	2.012	0.089	0.0000	0.0000	0.165
Soup: Condensed Wet (Share 41.12 %, Multiple brands purchases 5.51 %)							
Campbell	61.56	0.082	2.384	7.061	0.010	0.030	0.169
Private label	37.11	0.058	2.166	2.210	0.046	0.0554	0.18
Other brands	1.33	0.0013	2.074	3.9745	0.086	0.222	0.427
Soup: RTS Wet (Share 30.65 %, Multiple brands purchases 4.65 %)							
General mills	38.43	0.022	2.522	0.380	0.009	0.067	0.159
Campbell	36.39	0.023	2.580	3.871	0.023	0.052	0.293
Conagra	21.78	0.013	2.173	1.813	0.042	0.079	0.224
Other brands	3.40	0.002	2.552	3.550	0.074	0.080	0.286

shares of sub-categories as well as yearly revenue market shares of brands in these two stores.

The product categories considered are soft drinks, coffee, milk, hot dog, spaghetti sauce, and soup. In the panel data set the categories soft drinks, coffee, milk, and soup are further divided into sub-categories. We investigate only sub-categories whose yearly revenue share with respect to their category amounts to at least 15 % (see Tables 2 and 3). This criterion leads to the sub-categories low calorie soft drinks, regular soft drinks, ground coffee, whole beans coffee, regular milk, flavored milk, condensed wet soup, and RTS wet soup. In the following, we no longer distinguish between categories and sub-categories, but simply talk about categories. These ten categories are low calorie soft drinks, regular soft drinks, ground coffee, whole beans coffee, regular milk, flavored milk, hot dog, spaghetti sauce, condensed wet soup, and RTS wet soup.

We remove households who in 2011 did not make a purchase in any of the categories mentioned. We explicitly consider brands in a category with higher revenue market shares. We aggregate the remaining brands of a category calling them "Other Brands". Including "Other Brands" we arrive at a total of 42 brands. Tables 2 and 3 show revenue market shares of these brands. These tables also give the percentage of purchases of multiple brands for each category. Multiple brand purchases are very frequent for the two soft drink categories and are quite remarkable for two soup categories.

Each purchase of a household is represented by a binary purchase vector with 42 elements. An element equal to one indicates a purchase of the respective brand. Basket size, i.e., the number of brands purchased, has an average of 1.88 with a standard deviation of 1.12. The data set we analyze comprises 33,622 purchases made by 1,805 households. Therefore we have on average 18.63 observations per household. 21.22 %, 49.7 % and 11.97 % of households consist of one, two and three persons, respectively. 29.00 %, 34.24 % and 36.76 % of these households have a low, medium and high income, respectively.

Tables 2 and 3 also contain relative purchase frequencies of brands and averages of marketing variables across purchases. Marketing variables comprise price, feature, display, and price reduction. Marketing variables are computed as averages of the different UPCs of a brand weighted by annual dollar sales. Price in dollars refers to the standard package of the category. The other marketing variables are given as shares.

## 4.2 Estimation results and runtimes

Brand specific marketing variables are defined as difference to the average of the other brands of the category. Average marketing variables equal arithmetic means across all brands of the category. Household attributes consist of household size (number of persons) and two binary dummy variables for medium and high income, respectively. Models with independent variables may include the following direct effects:

- brand-specific marketing variables on purchases of the respective brand;
- average marketing variables on purchases of each brand of the respective category;
- household attributes on purchases of each of the 42 brands.

Accordingly, every element of matrix  $\beta$  that contains the effect of a brand-specific marketing variable on other brands or the effect of an average marketing variable on brands of other categories is set to zero and not estimated.

Let  $H$ ,  $M$ ,  $J$ ,  $C$ ,  $S$ , and  $K$  denote the number of household attributes, the number of marketing variables, the number of brands, the number of product categories, the number of segments, and the number of hidden variables, respectively. The number of independent variable coefficients  $L$  equals  $J(H + M)$ . If

**Table 4** Number of model parameters

Model	Interactions	Independent variables	Number of parameters
FM-MVL	No	No	$SJ + S - 1$
	No	Yes	$S(J + L) + S - 1$
	Brands and categories	yes	$S(J + JC + L) + S - 1$
	Brand pairs	No	$S(J + J(J - 1)/2) + S - 1$
	Brand pairs	Yes	$S(J + J(J - 1)/2 + L) + S - 1$
RBM	Brands and hidden variables	No	$J + K + JK$
CRBM	Brands and hidden variables	Yes	$J + K + JK + L$

$S$  number of segments,  $J$  number of brands,  $C$  number of categories,  
 $L$  number of independent variables,  $K$  number of hidden variables

average marketing variables are included as well  $L$  equals  $J(H + 2M)$ . Table 4 shows expressions for the number of parameters for different models. RBM (CRBM) have a lower number of parameters than the homogeneous MVL model ( $S=1$ ) with interactions among brand pairs (and independent variables) if  $K < J(J - 1)/(2(J + 1))$ . For 42 brands a lower number of parameters results for the RBM (CRBM) if the number of hidden variables is not greater than 20.

We randomly form two groups with about 2/3 of the households in the first group. We use data (estimation data) of the first group to estimate models. Data of the second group (holdout data) serve to evaluate models whose coefficients are estimated on data from the first group. We measure performance of models by their LPL value for the holdout data. Table 5 summarizes the evaluation by showing only the best performing model in each row. In terms of holdout LPL values

- the homogeneous (= one segment) MVL model without both interactions and independent variables is better than FM-MVL models which exclude both interactions and independent variables having at least two segments.
- the FM-MVL model for two segments with interactions between brand pairs and no independent variables is better than the FM-MVL models with interactions and no independent variables but a different number of segments.
- the homogeneous FM-MVL model without interactions which includes independent variables is better than FM-MVL models without interactions and including independent variables having at least two segments
- the homogeneous FM-MVL model with interactions between categories and brands that includes independent variables is better than FM-MVL models without interactions and including independent variables having at least two segments.
- the FM-MVL model with interactions between brand pairs, the independent variables household attributes, price reductions, features, and displays and three segments is better than FM-MVL models with interactions between brand pairs with the same independent variables, but a different number of

**Table 5** Model evaluation

Model	Interactions	Independent variables	# of Segments or hidden variables	# of Parameters	Log Pseudo-Likelihood	
					Estimation data	Holdout data
FM-MVL	No	No	1	42	-62,816	-30,812
	No	A, R, F, D	1	378	-61,482	-30,273
	Categories and brands	A, R, F, D	1	798	-61,431	-30,256
	Brand pairs	no	2	1,807	-54,552	-28,236
	Brand pairs	A, R, F, D	3	3,845	-50,367	-27,395
RBM	Hidden variables and brands	No	12	558	-54,752	-27,314
CRBM	Hidden variables and brands	A, R, F, D	12	936	-53,897	-26,983

A household attributes, R price reductions, F features, D displays.

Log pseudo-likelihood values are rounded to the nearest integer.

Each row shows the corresponding model with the highest holdout log pseudo-likelihood.

E.g., first row: no FM-MVL without interactions and without independent variables having a higher number of segments performs better; last row: no CRBM with other independent variables or a different number of hidden variables performs better

- segments as well as FM-MVL models with different independent variables and any number of segments.
- the RBM with 12 hidden variables is better than any RBM with a lower or higher number of hidden variables.
  - the CRBM with 12 hidden variables and household attributes, price reductions, features, displays as independent variables is better than any CRBM with 12 hidden variables and other independent variables.

Table 5 also shows that the homogeneous MVL model with independent variables outperforms the analogous independent logit model that excludes interactions.

All models with independent variables whose results Table 5 shows include average marketing variables. The holdout performance of models that ignore average marketing variables is without exception inferior to models that include both brand-specific and average marketing variables. We emphasize that models with price as one of the independent variables perform worse compared to models with independent variables household attributes, price reductions, features, and displays. This result indicates that the price reduction variable reproduces households' responses to price changes better than the price variable.

Table 5 contains three MVL models with independent variables. We see that pairwise interactions between brands lead to a large improvement after controlling for marketing mix and demographics. The model with interactions between categories and brands, on the other hand, has a much lower number of parameters. Nevertheless, the complexity of this model seems to be too low which can be seen by the rather modest increase of its holdout LPL compared to the MVL model without interactions. Obviously, too much information is lost by not considering interactions between pairs of brands.

Starting with the seminal paper of Guadagni and Little (1983) publications in the marketing literature investigating brand choice models often include brand loyalties as independent variables. In a similar manner, we find category loyalties as independent variables in market basket models at the category level (see, e.g., Russell and Petersen (2000)). We add brand loyalties computed as first order exponentially smoothed brand purchases to the two MVL models with independent variables. The best performing models extended this way result for a smoothing constant of 0.1, meaning that recent brand purchases get a low weight. These models have the same number of segments as the related models without brand loyalties. We do not deal with these extended models in more detail as they do not lead to higher holdout LPL values.

Among all FM-MVL models, the three-segment model with interactions and the independent variables household attributes, price reductions, features, and displays performs best. Though the LPL of this model for the estimation data is higher than the values obtained for RBM and CRBM shown in Table 5, the latter two attain higher holdout LPL values. Obviously, the FM-MVL model has a very high number of parameters and is overly complex. Its high model complexity causes an overfitting problem, which becomes apparent by a deterioration of performance for the holdout data. Both the RBM and the CRBM benefit from their higher flexibility

**Table 6** Estimation run times of selected models

Model	Run time (seconds)
Homogeneous independent logit model	199
Homogeneous MVL model with pairwise interactions	351
Three segment MVL model with pairwise interactions	3301
RBM with 12 hidden variables	571
CRBM with 12 hidden variables	880

with respect to interactions in contrast to the FM-MVL model which is restricted to pairwise interactions (see expression (1)).

The good performance of the RBM is remarkable, because in contrast to the best FM-MVL model it does not include any independent variable. This result shows that flexibility with respect to interactions is more important than the inclusion of independent variables in the FM-MVL model. However, after dealing with interactions in a flexible way, adding independent variables to the RBM to obtain the CRBM leads to a further improvement of holdout performance. The CRBM with 12 hidden variables and household attributes, price reductions, features, and displays as independent variables turns out to be the overall best model with the highest holdout log pseudo-likelihood.

Table 6 shows run times of five selected models for the estimation data, the homogeneous independent logit model, the homogeneous MVL model with pairwise interactions, its finite mixture extension with three segments, the RBM and the CRBM both with 12 hidden variables. The homogeneous independent logit model is identical to the homogeneous MVL model without interactions. With the exception of the RBM these five models include the same set of independent variables (socio-demographics and marketing variables). The CRBM turns out to be superior to the three-segment MVL model with interactions with respect to estimation run time. In fact, estimation of the three-segment MVL model is more than three times slower. Please note that the run time of CRBM estimation includes its first step, i.e., estimating the RBM with the same number of hidden variables by contrastive divergence (see Appendix C).

### 4.3 Interpretation of the conditional restricted Boltzmann machine

In the following, we take three different routes to interpret the selected CRBM with 12 hidden variables. Firstly, we look at  $W_{jk}$  coefficients which link purchases of any brand  $j$  to hidden variable  $k$  (see expression (8)). Secondly, we regard interdependences between brands by means of pairwise marginal probability changes. Thirdly, we investigate the effects of brand-specific marketing variables (price reduction, feature, display) on purchases of the same brand (own effects) and on purchases of other brands (cross effects).

According to expression (9) the conditional probability of a hidden variable  $k$  depends on the  $W_{jk}$  coefficients. Barplots in Figs. 1 and 2 show the  $W_{jk}$  coefficients

for each brand and each hidden variable. Due to high positive coefficients, the conditional probability of hidden variable 1 increases if, e.g., the regular milk brands Hood or Dean, the low calorie brands Coca Cola or Dr Pepper, or the regular soft drink Dr Pepper are purchased. On the other hand, due to negative coefficients this conditional probability gets low if, e.g., private brands of regular milk, spaghetti sauce or soup, or any of the Other soft drink brands (both regular and low calorie) are purchased.

The conditional probability of hidden variable 2 increases with purchases of, e.g., the regular milk brand Dean, the RTS soup brand Conagra, the ground coffee brand Smucker, the low calorie soft drink Dr Pepsi, or the spaghetti sauce brand Conagra. It decreases with purchases of, e.g., the low calorie brand Dr Pepper, the flavored milk brand Dean, the low calorie brand Coca Cola, or the regular soft drink Dr Pepper. In the same manner, one can interpret the relationships between brands' purchases and conditional probabilities of the remaining ten hidden variables.

The  $W_{jk}$  coefficients also hold information about differences between two hidden variables. Let us give two examples here. Whereas the probability of hidden variable 1 increases if the low calorie brand Coca Cola or the low calorie and regular Dr Pepper brands are purchased, such purchases on the other hand decrease the probability of hidden variable 2. As another example, purchases of the soup brand Campbell or the private regular milk brand increase the probability of hidden variable 3, but decrease the probability of hidden variable 6.

As second route to interpretation we measure pairwise interdependence by the probability change of brand  $j' \neq j$  associated to a marginal increase of the probability of brand  $j$ . We also compare probability changes obtained for the CRBM to those for the three-segment FM-MVL model.

For the CRBM we determine the purchase probabilities of brands by fixed point iterations (Tramel et al. 2016) over expressions (8) and (9) with estimated coefficients of the selected CRBM. For the FM-MVL model we generate simulated purchases by iterated Gibbs-sampling from the conditional distribution given by expression (5). We estimate purchase probabilities of each brand by averaging across simulated purchases. Iterations stop if changes of purchase probabilities become very small (see Besag (2004) for Gibbs sampling from the conditional distributions).

We set marketing variables to their arithmetic means and use medium income and two persons as values of the household attributes. In the first round we compute probabilities  $p_0(j)$  for all 42 brands. In the second round we set the purchase probability of brand  $j$  to  $p_0(j) + \delta$  with  $\delta = 0.005$  and hold this purchase probability constant during iterations. The second round then produces new probabilities  $p_1(l)$  for each of the other brands  $l \neq j$ . The probability change is computed as  $pc(j, l) = (p_1(l) - p_0(l))/\delta$ .

As we mentioned in the introduction, positive (negative) probability changes indicate that the two respective brands are purchase complements (purchase substitutes). Figure 3 contains heatmaps of the brand interdependences obtained for three-segment FM-MVL mode and CRBM. One immediately sees that as a rule the interdependences for the CRBM are more varied than those for the FM-MVL model are. According to the latter model interdependences focus on soft drink brands and two regular milk brands (Private Label and Hood). Let us mention a few examples that the CRBM additionally

**Fig. 1**  $W_{jk}$  coefficients for hidden variables 1–6. Abbreviations: fm avored milk, gc ground coffee, hd hot dog, ls low calorie soft drink, PL private label,rm regular milk, rs regular soft drink, s soup, ss spaghetti sauce, wc whole beans coffee

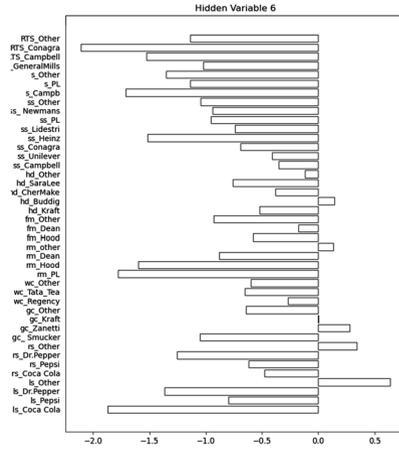
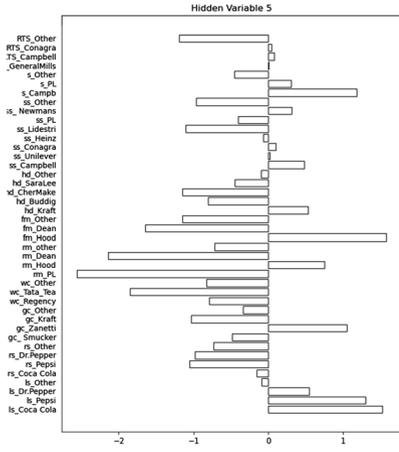
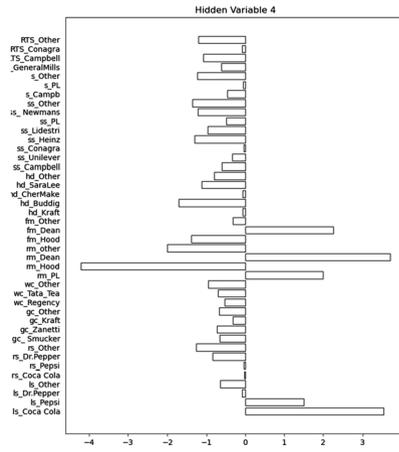
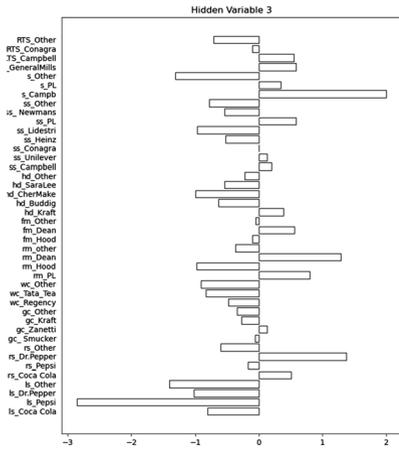
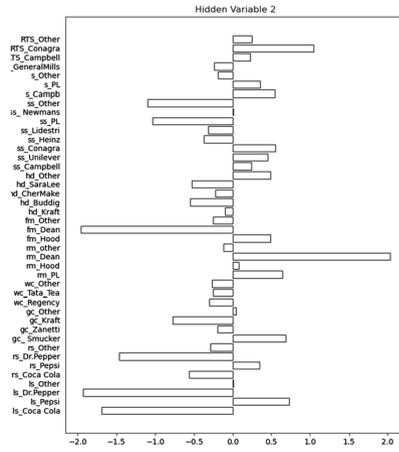
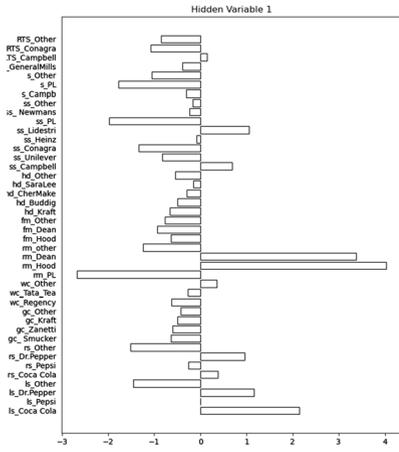
indicates. Regular Cola and Pepsi are substitutes for several soup and RTS soup brands. Two soup brands (Campbell and Private Label) are complements with several spaghetti and RTS soup brands brands and substitutive with several soft drink brands

Taking the third route to interpretation we investigate own effects of the marketing variables feature, display and price reduction on the same brand as well as their cross effects on each of the other brands based on the CRBM. We measure both own effects and cross effects as marginal purchase probability changes. Because of clearly inferior holdout performances we refrain from presenting own and cross effects of FM-MVL models.

Similar to the procedure for pairwise interdependences explained above we determine purchase probabilities of brands in two rounds of fixed point iterations over expressions (8) and (9). However, in the second round we now set the respective marketing variable of brand  $j$  to its arithmetic mean plus  $\delta$  and obtain new probabilities  $p1(j')$  for each of the 42 brands. Marginal probability change is computed as  $(p1(j') - p0(j'))/\delta$ . This expression provides an own effect, if  $j' = j$ , otherwise a cross effect.

Figure 4 shows own effects of marketing variables. All significant own effects of the three marketing variables are positive, i.e., the increase of the marketing variable is accompanied by a probability increase of purchases of the same brand. For all marketing variables own effects are high for the private label soup brand, the hot dog brand Kraft and the ground coffee brand Smucker. In addition, high own effects of display and price reduction turn out for the soup brand Campbell.

Cross effects are much smaller than own effects. This result is similar to Russell and Petersen (2000) as well as Song and Chintagunta (2007) who obtain only small cross effects for product categories and brands, respectively. Table 7 gives cross effects with a minimum absolute value of 0.005 only. We notice that cross effects may assume both positive and negative values. The former indicate a probability increase, the later a probability decrease due to an increase of the respective marketing variable. Taking into account only cross effects of at least 0.005 in absolute size, we obtain the highest (positive) cross effect for features of the private soup brand on purchases of the soup brand Campbell. Although this effect is low, it is nonetheless remarkable, as it means that Campbell benefits from advertising of another brand belonging to the same category. According to a more conventional assumption feature advertising is expected to hurt other brands of the same category. We obtain the lowest (negative) cross effect for displays of the private soup brand on purchases of the regular soft drink brand Dr Pepper.



**Fig. 2**  $W_{jk}$  coefficients for hidden variables 7–12. Abbreviations: fm avored milk, gc ground coffee, hd hot dog, ls low calorie soft drink, PL private label, rm regular milk, rs regular soft drink, s soup, ss spaghetti sauce, wc whole beans coffee

## 5 Managerial implications and conclusion

Russell and Petersen (2000) demonstrate that homogeneous category-level MVL models with interactions lead to better forecasts of the composition of market baskets than independent binary choice models. Our study confirms that this superiority of MVL models persists on the brand level. However, the CRBM shows a better out-of-sample performance than both heterogeneous and homogeneous MVL models which are often used to analyze market basket data. In other words, the CRBM turns out to beat MVL models in accomplishing this forecasting task.

The higher flexibility of the CRBM on the brand level compared to multi-category and multinomial brand-level choice models seems to pay off. In contrast to the later models, the CRBM does not prevent the following relationships between brands. Using the CRBM we find examples of two brands of the same category that are purchase complements (e.g., regular Pepsi with several other soft drink brands). Interdependences of one brand with brands of another category vary. We also come across the extreme case, that a brand is both a purchase substitute with one brand and a purchase complement with another brand of a different category (e.g., regular Coca Cola is a purchase complement with flavoured milk brand Hood, but a purchase substitute with flavoured milk brand Dean.) As presented in the previous section the CRBM indicates a greater number of higher pairwise relationships between brands measured by marginal probability changes.

To obtain a comprehensive insight into the competitive structure we run a K-medoids cluster analysis on dissimilarities. We compute dissimilarities between any two brands  $j$  and  $l$  starting with probability changes obtained for the CRBM as follows:

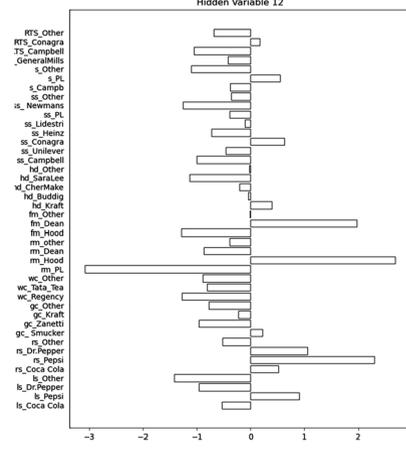
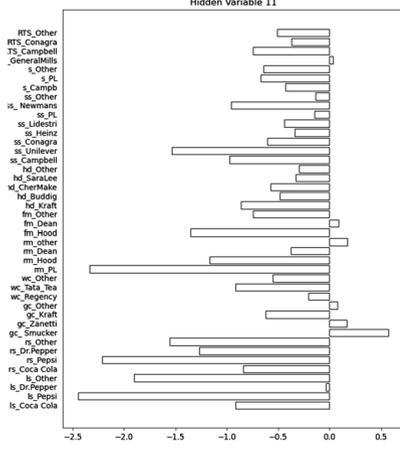
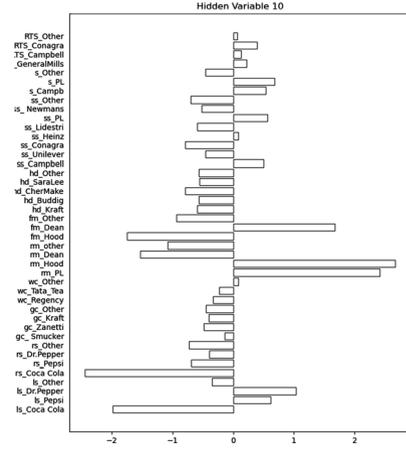
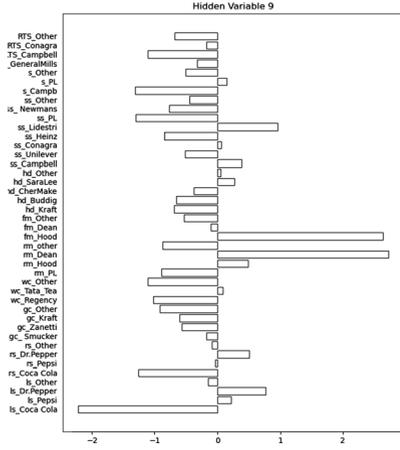
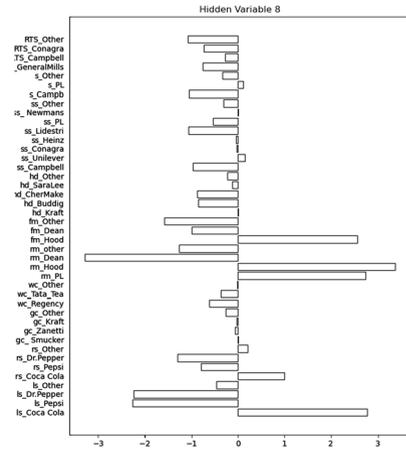
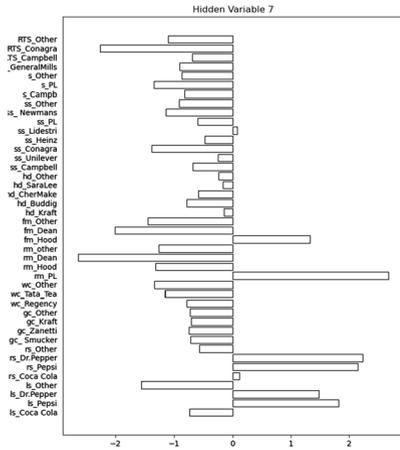
$$dis(j, l) = 1 - (spc(j, l) - spc_{min}) / (spc_{max} - spc_{min}) \quad (15)$$

$$\text{with } spc(j, l) = spc(l, j) = -(pc(l, j) + pc(j, l)) / 2 \quad (16)$$

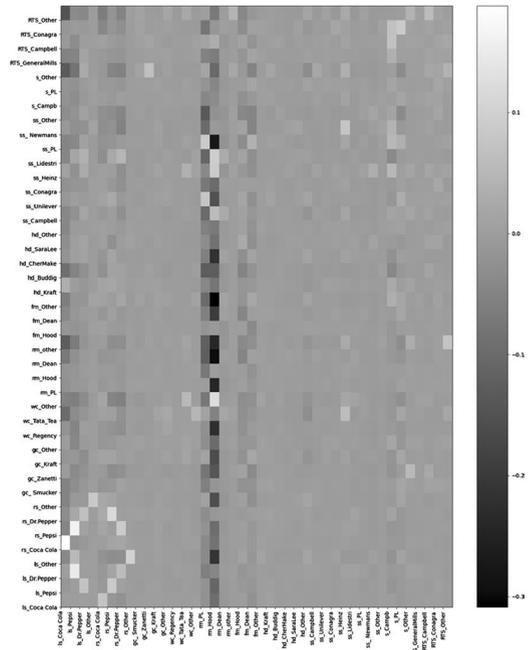
Note that expression (16) reverses the sign of probability changes. Therefore substitutive relations obtain positive signs, complementary relations negative signs. Consequently, two brands with strong substitutive relationships get a high similarity value in expression (16) and a low dissimilarity value in expression (15).  $spc_{max}$  and  $spc_{min}$  denote the maximum and minimum value of pairwise similarities, respectively. Dissimilarities assume values between zero and one. A value of zero (one) for  $dis(j, l)$  means that  $j$  and  $l$  form the most (the least) competitive pair of brands.

Table 8 shows the solution for five clusters. Brands belonging to the same cluster have lower dissimilarities indicating that competition among them is higher than competition with brands belonging to another cluster. Other soups and regular milk Dean each form a singleton, i.e., competition with any of the other brands is weak no matter to which category the latter belongs.

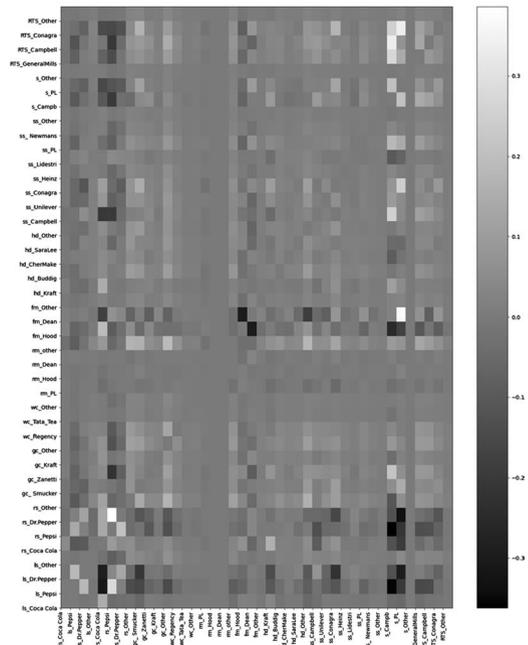
Analyzing joint brand purchases by conditional restricted...



**Fig. 3** Heatmaps of Brand Interdependences according to the FM-MVL Model and the CRBM (bright cells indicate a complementary, dark cells a substitutive interdependence)



FM-MVL Model



CRBM



**Table 8** Competitive clusters (Five cluster K-medoids solution)

Cluster #	Brands
1	s Other
2	ls Coca Cola, ls Dr Pepper, rs RegPepsi, wc Tata Tea, rm PL, rm Hood, fm Hood, hd Cher-Make, hd Sara Lee, ss Lidestri, ss Other, s Campbell
3	ls Pepsi, rs Coca Cola, rs Other, gc Smucker, gc Kraft, gc Other, wc Regency, wc Other, rm Other, fm Dean, fm Other, hd Kraft, hd Buddig, hd Other, ss Unilever, ss Heinz, ss PL, RTS General Mills, RTS Campbell, RTS Other
4	ls Other, rs Dr Pepper, gc Zanetti, ss Campbell, ss Conagra, ss Newmans, s PL, RTS Conagra
5	rm Dean

Abbreviations: fm flavored milk, gc ground coffee, hd hot dog, ls low calorie soft drink, PL private label, rm regular milk, rs regular soft drink, s soup, ss spaghetti sauce, wc whole beans coffee

**Table 9** Revenues due to display of each of 42 brands

Model	Revenue		
	Minimum	Mean	Maximum
Independent logit	68,957	69,580	77,588
Finite mixture multivariate logit	64,588	73,336	81,480
Conditional restricted Boltzmann machine	92,726	92,694	93,282

Motivated by a suggestion of one anonymous reviewer we perform counterfactual simulations for three models, all including independent variables. These three models are the independent logit model (i.e., the MVL model without interactions), the FM-MVL model with three segments and the CRBM with 12 hidden variables. We investigate what putting each brand on display means for the retailer's revenue if no other brand is displayed. We set all feature and price reduction variables to zero. We generate 20,000 baskets by drawing from the conditional distribution of purchases. For the CRBM we also draw from the conditional distribution of hidden variables. To obtain the revenue we multiply the total number of purchases of each brand by its average price and sum across brands.

Table 9 gives the mean of the revenues together with their minimum and maximum values across 42 brands for each of these three models. The mean revenue implied by the MVL model is higher than the corresponding value for the independent logit model. The overall highest sales revenues result for the CRBM. This ranking of sales revenues matches the extent to which the models consider interdependences. The independent logit model rules out interdependences completely. On the other hand, the CRBM is more flexible than the FM-MVL model by also considering higher order dependences.

Results obtained by means of the CRBM can be used to support cross-selling decisions such as cross-selling bundling, cross category promotional programs, cross category loyalty programs and cross-category positioning. Of course, independent brands are not appropriate for cross-selling. Our results show that cross-selling

may encompass brands of the same or related categories if they are purchase complements (e.g., low calorie Pepsi with several other soft drink brands, but not with brands of other categories). Brands may also be purchase complements with brands of other categories (e.g., various soup brands are purchase complements not only with other soup brands, but also with brands of the ground coffee category).

The selected CRBM includes the marketing variables feature, display, and price reductions. Clearly, own effects on the same brand dominate cross effects, which are much lower. As a rule, own effects vary across brands of the same category and are more pronounced for one brand of one category.

The excellent holdout performance of the CRBM suggests continuing research in two respects. One possibility consists in applying the CRBM to other areas in marketing and retailing. Besides non-food retailing and e-commerce media such areas comprise media consumption (Gentzkow 2007; Yang et al. 2010), subscriptions of different (media) services (Schweidel et al. 2011), and menu choice problems (Kosyakova et al. 2020). Developing extensions of the CRBM that are appropriate for different dependent variables such as purchase amounts or purchase quantities constitutes another option.

### A estimation of homogeneous multivariate logit models

We estimate homogeneous multivariate logit models by maximizing the LPL based on average gradients across market baskets. To this end we use the BFGS algorithm contained in the Optimize module of the Python package SciPy (Virtanen 2020). We set initial values of all coefficients to zero.

We show the gradients with respect to the log pseudo-probability  $\log(\tilde{P}_j)$  for one basket omitting the basket index to keep notation simple. Gradients for the different types of coefficients of the MVL model with pairwise interactions are:

$$\begin{aligned} \Delta \log(\tilde{P}_j)_{\alpha_j} &= y_j - P(y_j | y_{-j}) \\ \Delta \log(\tilde{P}_j)_{\beta_{jp}} &= (y_j - P(y_j | y_{-j})) x_{jp} \\ \Delta \log(\tilde{P}_j)_{\gamma_{jl}} &= (y_j - P(y_j | y_{-j})) y_l + (y_l - P(y_l | y_{-l})) y_j \quad \text{for } j \neq l \end{aligned}$$

Please note that gradients for coefficients  $\alpha$  and  $\beta$  look like those for maximum likelihood estimation of the binomial logit model (see, e.g., Greene (2003)). To estimate the independent logit model we only have to set gradients of interaction coefficients constantly to zero.

For the MVL model with interactions between products and categories we obtain:

$$\begin{aligned} \Delta \log(\tilde{P}_j)_{\alpha_j} &= y_j - P(y_j | y_{-ind_j}^c) \\ \Delta \log(\tilde{P}_j)_{\beta_{jp}} &= (y_j - P(y_j | y_{-ind_j}^c)) x_{jp} \\ \Delta \log(\tilde{P}_j)_{\gamma_{j,c}^{cat}} &= (y_j - P(y_j | y_{-ind_j}^c)) y_c^{cat} \quad \text{for } c \neq ind_j \end{aligned}$$

## B estimation of finite mixture multivariate logit models

Our estimation approach is akin to maximizing the classification likelihood (McLachlan and Basford 1988; Ngatchou-Wandji and Bulla 2013). We replace the intractable likelihood by the pseudo-likelihood. We describe estimation by the following pseudo-code:

- Randomly assign each of  $M$  households to one of  $S$  segments to produce binary segment memberships ( $u_{sm} = 1$  if household  $m$  is assigned to segment  $s$ )
- **Repeat**
  - compute relative segment sizes  $\pi_s = \sum_{m=1}^M u_{sm}/M$
  - estimate a MVL model for each segment  $s$  by the method described in Appendix A using all data of households with  $u_{sm} = 1$
  - compute the pseudo-probability  $\tilde{P}_{sm}$  for each household and each segment-specific MVL model
  - assign each household to the segment  $s$  for which  $\pi_s \tilde{P}_{sm}$  is maximal
- **Until** segment assignments do not change

The segment-specific pseudo-probability  $\tilde{P}_{sm}$  is computed for all baskets of household  $m$  in following manner:

$$\tilde{P}_{sm} = \prod_{i=1}^{I_m} \prod_{j=1}^J \tilde{P}_{smij}^{Y_{mij}}$$

$I_m$  denotes the number of baskets of household  $m$ ,  $Y_{mij}$  a binary purchase indicator (set to one if basket  $i$  of household  $m$  contains product  $j$ ),  $\tilde{P}_{smij}$  the pseudo-probability of product  $j$  in basket  $i$  of household  $m$  according for the MVL model for segment  $s$ . The segment-specific MVL model takes one of three alternative forms, the independence logit model, the MVL model with pairwise interactions, and the MVL model with interactions between products and categories.

## C estimation of the conditional restricted Boltzmann machine

Estimation of a CRBM consists of two steps. In the first step we determine coefficients of a RBM that has the same number of hidden variables as the CRBM by contrastive divergence (CD). The objective of CD is related to the Kullback-Leibler divergence between the data distribution and the model distribution (more details on the CD algorithm can be found in Hinton (2002) and Murphy (2012)).

We use the CD algorithm implemented in the Python library NeuPy (Shevchuk 2019). In each iteration the algorithm performs Gibbs sampling of all hidden variables conditional on observed brand purchases followed by Gibbs sampling of all

brand purchases conditional on all hidden variables. Coefficients are updated for mini-batches of 100 observations with a learning rate set to 0.1. The estimation process runs for 500 epochs, i.e., 500 complete passes over all observations. We start CD estimation ten times with different initial random coefficient values and finally choose the solution with the highest LPL.

The second step deals with estimating the CRBM proper. The BFGS algorithm contained in the Optimize module of the Python package SciPy (Virtanen 2020) serves to maximize the LPL based on average gradients across market baskets. Initial values of coefficients  $\gamma_k$  and  $W_{jk}$  are taken from best CD solution. Initial values of the  $\beta$  coefficients of the CRBM are set to zero.

We now show how gradients for one basket are computed omitting the basket index to keep notation simple. The gradient of the log pseudo-probability of product  $j$  with respect to any parameter  $\theta$  of a CRBM is given by (Marlin et al. 2010):

$$\Delta \log(\tilde{P}_j)_\theta = \sum_{j=1}^J P(\tilde{y}_{(j)}|y_{-j}) \left( \Delta F_\theta(y, x) - \Delta F_\theta(\tilde{y}_{(j)}, x) \right)$$

with  $P(\tilde{y}_{(j)}|y_{-j}) = \exp(-F_\theta(\tilde{y}_{(j)}, x)) / (\exp(-F_\theta(y, x)) + \exp(-F_\theta(\tilde{y}_{(j)}, x)))$

$F_\theta(y, x) = -[y^T \alpha + x^T \beta y + \sum_{k=1}^K \log(1 + \exp(\gamma_k + y^T W_{.k}))]$  is the free energy function of the CRBM,  $\Delta F_\theta(y, x)$  symbolizes its gradient. For each type of coefficient we obtain the following expressions for the gradients:

$$\begin{aligned} \Delta \log(\tilde{P}_j)_{\alpha_j} &= P(\tilde{y}_{(j)}|y_{-j}) (2y_j - 1) \\ \Delta \log(\tilde{P}_j)_{\beta_{jp}} &= P(\tilde{y}_{(j)}|y_{-j}) (2y_j - 1) x_{jp} \\ \Delta \log(\tilde{P}_j)_{\gamma_k} &= \sum_{j=1}^J P(\tilde{y}_{(j)}|y_{-j}) \left( \sigma(\gamma_k + y^T W_{.k}) - \sigma(\gamma_k + \tilde{y}_{(j)}^T W_{.k}) \right) \\ \Delta \log(\tilde{P}_j)_{W_{jk}} &= \sum_{j=1}^J P(\tilde{y}_{(j)}|y_{-j}) \left( \sigma(\gamma_k + y^T W_{.k}) y_j - \sigma(\gamma_k + \tilde{y}_{(j)}^T W_{.k}) (1 - y_j) \right) \end{aligned}$$

$\sigma(Z)$  denotes the binary logistic function  $1/(1 + \exp(-Z))$ .

In our study the BFGS algorithm with gradients for coefficients  $\alpha$ ,  $\gamma$  and  $W$  did not improve the LPL of RBMs over its value for the best solution already provided by the CD algorithm.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agrawal R, Srikant R (1994) Fast algorithms for mining association rules in very large databases. In: Proceedings of the 20th international conference on VLDB, Santiago, Chile
- Aurier P, Mejia V (2014) Multivariate logit and probit models for simultaneous purchases: Presentation, uses, appeal and limitations. *Rech Appl Mark* 29:79–98
- Bel K, Fok D, Paap R (2018) Parameter estimation in multivariate logit models with many binary choices. *Econ Rev* 37:534–550
- Bengio Y (2009) Learning deep architectures for AI. *Fdn Trends Mach Learn* 2:1–27
- Besag J (1974) Spatial interaction and the statistical analysis of lattice Systems. *J R Stat Soc B* 36:192–236
- Besag J (2004) An introduction to Markov chain Monte Carlo methods. In: Khudanpur SP, Ostendorf M, Rosenfeld R, Johnson ME (eds) *Mathematical foundations of speech and language processing*. Springer, Berlin, pp 247–270
- Betancourt R, Gautschi D (1990) Demand complementarities, household production, and retail assortments. *Mark Sci* 9:146–161
- Boztuğ Y, Reutterer T (2008) A combined approach for segment-specific market basket analysis. *Eur J Oper Res* 187:294–312
- Bronnenberg BJ, Kruger MW, Mela CF (2008) Database paper: the IRI marketing data set. *Mark Sci* 27:745–748
- Cox DR (1972) The analysis of multivariate binary data. *J R Stat Soc C* 21:113–120
- Dubé JP (2004) Multiple discreteness and product differentiation: demand for carbonated soft drinks. *Mark Sci* 23:66–81
- Gentzkow M (2007) Valuing new goods in a model with complementarity: online newspapers. *Am Econ Rev* 97:713–744
- Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. *Mark Sci* 2:203–238
- Greene WH (2003) *Econometric analysis*, 5th edn. Prentice Hall, Upper Saddle River, NJ
- Hahsler M, Hornik K, Reutterer T (2006) Implications of probabilistic data modeling for mining association rules. In: Spiliopoulou M, Kruse R, Borgelt C, Nürnberger A, Gaul W (eds) *From data and information analysis to knowledge engineering*. Springer, Berlin, pp 598–605
- Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comp* 14:1771–1800
- Hinton GE, Salakhutdinov R (2006) Reducing the dimensionality of data with neural networks. *Sci* 313:504–507
- Hruschka H (2014) Analyzing market baskets by restricted Boltzmann machines. *OR Spect* 36:209–228
- Hruschka H (2017) Multi-category purchase incidences with marketing cross effects. *Rev Manag Sci* 11:443–469
- Hruschka H (2021) Comparing unsupervised probabilistic machine learning methods for market basket analysis. *Rev Manag Sci* 15:497–527
- Jacobs B, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. *Mark Sci* 35:389–404
- Kim J, Allenby GM, Rossi PE (2002) Modeling consumer demand for variety. *Mark Sci* 21:229–250
- Kosyakova T, Otter T, Misra S, Neuerburg C (2020) Exact MCMC for choices from menus - measuring substitution and complementarity among menu items. *Mark Sci* 39:427–447
- Kwak K, Duvvuri SD, Russell GJ (2015) An analysis of assortment choice in grocery retailing. *J Ret* 91:19–33
- Le Roux N, Bengio Y (2007) Representational power of restricted Boltzmann machines and deep belief networks. Technical Report 1294, Département d'informatique et recherche opérationnelle, Université de Montréal

- Li X, Zhao F, Guo Y (2015) Conditional restricted Boltzmann machines for multi-label learning with incomplete labels. In: Proceedings of the 18th AISTATS Conference, San Diego, CA
- Manchanda P, Ansari A, Gupta S (1999) The shopping basket: a model for multi-category purchase incidence decisions. *Mark Sci* 18:95–114
- Marlin B, Swersky K, Chen B, Freitas N (2010) Inductive principles for restricted Boltzmann machine learning. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR workshop and conference proceedings 9:509–516
- McLachlan G, Basford K (1988) Mixture models: inference and applications to clustering. Marcel Dekker, New York
- Mnih V, LaRochelle H, Hinton GE (2011) Conditional restricted Boltzmann machines for structured output prediction. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, Barcelona, Spain
- Montúfar G (2016) Restricted Boltzmann machines: introduction and review. In: Ay N, Gibilisco P, Matúš F (eds) *Inf Geom Appl*. Springer, Cham, Switzerland, pp 75–115
- Murphy KP (2012) Machine learning, a probabilistic perspective. MIT Press, Cambridge
- Ngatchou-Wandji J, Bulla J (2013) On choosing a mixture model for clustering. *J Data Sci* 11:157–179
- Richards TJ, Hamilton SF, Yonezawa K (2018) Retail market power in a shopping basket model of super-market competition. *J Ret* 94:328–342
- Ruiz FJR, Athey S, Blei DM (2020) Shopper: a probabilistic model of consumer choice with substitutes and complements. *Ann Appl Stat* 14:1–27
- Russell GJ, Petersen A (2000) Analysis of cross category dependence in market basket selection. *J Ret* 76:369–392
- Schweidel DA, Bradlow ET, Fader PS (2011) Portfolio dynamics for customers of a multiservice provider. *Manag Sci* 57:471–486
- Shevchuk Y (2019) NeuPy. Neural Networks in Python. <http://neupy.com>. Accessed March 8, 2021
- 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 Mark Res* 44:595–612
- Tramel EW, Drémeau A, Krzakala F (2016) Approximate message passing with restricted Boltzmann machine priors. *Theory Exp J Stat Mechan*. <https://doi.org/10.1088/1742-5468/2016/07/073401>
- Virtanen P et al (2020) SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Meth* 17:261–272
- Xia F, Chatterjee R, May JH (2019) Using conditional restricted Boltzmann machines to model complex consumer shopping patterns. *Mark Sci* 38:711–727
- Yang S, Zhao Y, Erdem T, Zhao Y (2010) Modeling the intrahousehold behavioral interaction. *J Mark Res* 47:470–484

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.