ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Bachelor Thesis Econometrie & Operationele Research

Predicting cross-effects in market basket analysis with Restricted Boltzmann Machines

Name student: Wilko Verheij
Student ID number: 450031

Supervisor: Luuk van Maasakkers
Second assessor: dr. Kathrin Gruber

Date final version: July 4, 2021

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

1 Abstract

Market baskets are the observable outcomes of multicategory choices of people. These choices tend to be interdependent with each other, where we make a distinction in two types of interdependence, substitutes and complements. This paper tries to accurately estimate the cross effects of product categories in these market baskets. It does so by means of the restricted Boltzmann machine (RBM). A model that is similar to the multivariate logit model (MVL), but with the addition of hidden variables. Estimation of these models is done by maximum likelihood and maximum pseudo-likelihood, respectively. A third model is also investigated, namely the deep belief net (DBN). This model is estimated by the greedy layerwise algorithm. The models are compared to each other by means of absolute deviations of 300 Gibbs samples. The cross effects are then estimated and interpreted for the RBM, although the MVL model performs the best of the three models. It is however expected that the DBN will outperform the other models when the number of hidden variables becomes larger. The same can be said of the RBM with regard to the MVL model and therefore we investigate the cross effects of this model, since these cross effects are easier to interpret than those of the DBN.

Contents

2 Introduction

Over the years the number of online retailers grow larger and larger. With this growth comes also an ever growing amount of data for the retailers to analyse. One of the more interesting things to analyse for the owners is the purchase behavior of customers. The market baskets of the shoppers consist of all the product categories they purchase in their visit. It is the observable outcome of the multicategory choices they face every time they go online to do some shopping (G. Russell et al., 1997). These multicategory choices tend to be interdepent, e.g. the purchase of one category might influence the purchase of another category. Two categories are called substitutes if their cross effects are negative and they are complements if their cross effects are positive (Betancourt and Gautschi, 1990). This means that complements are found to be bought more frequent together than substitutes. Two products are called perfect substitutes (complements) if their cross effects is smaller (larger) than one (Gaudet and Salant, 1991). Because of the interdependence of the categories one-category models will not suffice. Therefore, appropriate models have to be formed to analyse these cross effects.

In this paper I only have a look at the purchase incidences of categories. I do not consider other forms of dependent variables, such as purchase quantities and brand choices. The main focus of this paper is to determine the cross effects of the categories, but model comparison is also one of the goals. The research question of this paper is as follows: "*What model performs best for market basket analysis and how can we estimate the cross effects of this model*?"

Two of the most frequently used models for calculating the cross effects is the Multivariate Logit (MVL) and the Multivariate Probit (MVP) model. These models can have the addition of other explanatory variables such as customer characteristics or product pricing. Manchanda et al. (1999) was the first to implement this model in the market basket analysis. They analysed the effects of four categories. Seetharaman et al. (2005) and Duvvuri, Ansari, et al. (2007) quickly followed with the analysis of twelve and six categories, respectively. Duvvuri and Gruca (2010) did this with four categories but also specified the price sensitivity. G. J. Russell and Petersen (2000) did the analysis with a MVL model consisting of four categories and Boztug and Hildebrandt (2008) their model consisted of five and six categories.

Maximum likelihood estimation of these models takes a lot of computational power. This is due to the fact that the number of possible baskets grows exponentially with the rise in the number of categories. Therefore a lot of the papers using these models only consider a small number of categories. If papers do consider more categories, however, they make use of a two-step approach for estimating the coefficients (Hruschka et al., 1999 & Boztuğ and Reutterer, 2008). Because of this two-step approach, however, results of the second step are restricted by the results of the first step. This results in limitations in the performance of these models (Wedel and Wagner A. Kamakura, 2000).

Restricted Boltzmann Machines (RBMs) are another way of estimating these cross effects. This model resembles the MVL model but adds a number of hidden variables. This model is often used in other applications such as analyzing the patterns in handwritten digits or document classification (G. E. Hinton, 2006). This model was introduced to the field of market basket analysis by Hruschka (2014). He already showed that this model is capable of analysing much higher numbers of categories. This paper is also the guideline for my thesis.

The last model I use for the analysis is the Deep Belief Network (DBN). This model is very successful in the machine learning fields like language processing and speech recognition (Najafabadi et al., 2015). The model is related to the RBM model in the sense that the DBN consists of multiple layers of RBMs. To my knowledge Hruschka (2021) is the first to use this model for market basket analysis. He concludes that this model outperforms the RBM when the number of hidden variables in the layers are large. He also states that these two models perform equally good if these numbers are kept low. Because maximum likelihood estimation of the RBM is only feasible for lower numbers of hidden variables, I choose to keep the numbers of hidden variables in the layers of the DBN low as well, although performance will be beter for larger numbers.

I will not use any other explanatory variables for the analysis, since the data does not contain any other information but the orders themselves. Furthermore it is stated that the addition of price effects only slightly improves the performance of the model, if any (G. J. Russell and Petersen, 2000a & Boztug and Hildebrandt, 2008).

In the following I further elaborate on the existing literature. I then explain the data I am using for the empirical part of my paper. In this section I show the descriptive statistics, explain what categories I am using and also show the relation of the products to each other. In the methodology I explain the models that I use more extensively, beginning with the MVL model, followed by the RBM and thereafter I explain the DBN. I also describe how I compare the models and how I estimate the cross effects. Then I discuss the results of the empirical study. Finally, I give an overview of the results, give some implications of this paper and give some further research options.

3 Literature

Manchanda et al. (1999) was the first to implement the multivariate probit model to the field of market basket analysis. They stated that the co-occurrence of two categories in a basket may not happen because these categories are complements but simply because they were part of the same purchase cycle or because of other unobserved factors. They implemented their model in a hierarchical Bayes framework consisting of three levels. Their first level captured the choices of items for the shopping basket during a shopping trip. The second level captures the differences across households and the third level specified the priors for the unknown parameters. They concluded that pricing and promotional changes in one category affected the purchase incidences in related product categories. They also found that their cross effects were symmetrical.

Hruschka et al. (1999) introduced the MVL model with a larger number of categories. They suggested a two step approach for the estimation of the parameters, since simultaneous estimation requires too much computational power. They found similar results for their data as Manchanda et al. (1999).

Hruschka (2014) was the first to introduce the restricted Boltzmann machine to this field and compared the performance to that of the multivariate logit model. He concludes that the RBM outperforms the MVL model in terms of absolute deviations, calculated by means of Gibbs sampling. He also states that cross effects do not have to be symmetric, in contrast to Manchanda et al. (1999) and Hruschka et al. (1999). This paper is also the guideline for this thesis.

Hruschka (2021) goes even further than the RBM in the sense that he decides to stack multiple RBMs on top of each other. This system is called the deep belief network (DBN). He concludes that this network will outperform the other models when the number of hidden variables in the layers are relatively large.

This thesis follows the work of Hruschka (2014) and compares the model to the DBN introduced in Hruschka (2021). It differs from the latter in the sense that the RBM is estimated in a different manner and the number of hidden variables is kept low.

4 Data

The data I am using for this paper is from Instacart, 2017. It contains anonymized data of over three million orders from more than 200,000 Instacart users. Instacart is a company that picks up and delivers groceries for its users. They make use of personal shoppers who do the shopping for the customers at participating retailers. The firm offers its services via an app or a website. The marketplace offers more than 300 retailers, with popular brands like ALDI, Target and Walmart. The data-set provides between 4 and 100 orders per user, together with the sequence of products purchased in each order. They also provide the week and hour of the day the order was placed, as well as a relative measure of time between the orders. For this research I only focus on what products were bought in each order.

First I randomly select 30,000 different orders from the data-set excluding the final orders of users, because keeping all of the baskets will require too much computational power. Here I assume that there will be no difference in order behaviour of customers between the last order and the other orders. I make this assumption because in the data-set the last order and the other orders are split in different files and therefore only selecting one file will be easier implemented. The 30,000 orders selected from the data-set are then split into an estimation and validation set of equal sizes. The almost 50,000 different products are divided into 134 aisles, which can be considered as the categories in this research. I then select the 60 aisles with the highest univariate marginal purchase frequency, just as in Hruschka (2014). These values can be found in table 1. This number of categories is five times as big as the maximum number of categories of other studies in which MVP or MVL models are simultaneously estimated (Hruschka, 2014). Table 2 contains the relative bivariate marginal frequencies of the 60 most frequent pairs of product categories. This number is acquired by first selecting the 60 most frequent categories and then dividing the number of times a pair of these categories is found in a basket by the number of baskets containing more than one of these categories.

Fresh fruits	0.557	Fresh vegetables	0.445	Packages vegetables and fruit	0.372
Yogurt	0.258	Milk	0.245	Packaged cheese	0.228
Seltzer and sparkling water	0.193	Chips and pretzels	0.167	Soy and Lactose-free	0.167
Bread	0.165	Eggs	0.138	Refrigerated produce	0.134
Frozen produce	0.125	Crackers	0.114	Ice cream	0.108
Lunch meat	0.101	Fresh dips and tapenades	0.100	Fresh herbs	0.094
Cereal	0.093	Cream	0.090	Juice and nectars	0.089
Soft drinks	0.086	Other creams and cheeses	0.085	Hot dogs, bacon and sausage	0.084
Energy and granola bars	0.084	Soup broth and bouillon	0.082	Nuts, seeds and dried fruit	0.078
Baking ingredients	0.077	Spreads	0.076	Canned and jarred vegetables	0.074
Butter	0.073	Dry pasta	0.072	Oils and vinegars	0.071
Frozen meals	0.071	Canned meals and beans	0.071	Candy and chocolate	0.070
Breakfast bakery	0.069	Packaged produce	0.065	Paper goods	0.061
Pasta sauce	0.060	Cookies and cakes	0.060	Condiments	0.059
Frozen breakfast	0.058	Coffee	0.056	Tortillas and flat bread	0.056
Tea	0.055	Spices and seasonings	0.054	Instant foods	0.051
Frozen appetizers and sides	0.050	Fruit and vegetable snacks	0.047	Baby food and formula	0.046
Popcorn and jerky	0.045	Asian foods	0.044	Hot cereal and pancake mixes	0.043
Frozen pizza	0.042	Grains, rice and dried goods	0.041	Poultry counter	0.038
Packaged poultry	0.036	Tofu and meat alternatives	0.034	Pickled goods and olives	0.033

Table 1: Relative univariate marginal frequencies of analyzed categories

(Fresh fruits) , (Fresh vegetables)	0.350	(Fresh fruits) , (Packaged vegetables and fruit)	0.301
(Fresh vegetables) , (Packaged vegetables and fruit)	0.264	(Fresh fruits) , (Yogurt)	0.203
(Fresh fruits) , (Milk)	0.181	(Fresh fruits) , (Packaged cheese)	0.169
(Fresh vegetables) , (Yogurt)	0.156	(Fresh vegetables) , (Packaged cheese)	0.148
(Fresh vegetables) , (Milk)	0.141	(Packaged vegetables and fruit) , (Yogurt)	0.135
(Fresh fruits) , (Soy and lactose-free)	0.127	(Fresh fruits) , (Bread)	0.126
(Fresh fruits) , (Seltzer and sparkling water)	0.123	(Packaged vegetables and fruit) , (Packaged cheese)	0.123
(Packaged vegetables and fruit) , (Milk)	0.120	(Fresh fruits) , (Chips and pretzels)	0.115
(Fresh fruits) , (Eggs)	0.107	(Fresh vegetables) , (Soy and lactose-free)	0.105
(Yogurt) , (Milk)	0.105	(Fresh vegetables) , (Bread)	0.102
(Fresh fruits) , (Frozen produce)	0.100	(Fresh fruits) , (Refrigerated produce)	0.095
(Fresh vegetables) , (Eggs)	0.094	(Yogurt) , (Packaged cheese)	0.091
(Fresh vegetables) , (Seltzer and sparkling water)	0.091	(Packaged vegetables and fruit) , (Soy and lactose-free)	0.089
(Fresh vegetables) , (Chips and pretzels)	0.089	(Fresh vegetables) , (Frozen produce)	0.089
(Milk) , (Packaged cheese)	0.088	(Fresh vegetables) , (Fresh herbs)	0.087
(Packaged vegetables and fruit) , (Bread)	0.085	(Packaged vegetables and fruit) , (Chips and pretzels)	0.082
(Fresh fruits) , (Crackers)	0.082	(Fresh fruits) , (Lunch meat)	0.081
(Packaged vegetables and fruit) , (Seltzer and sparkling water)	0.079	(Packaged vegetables and fruit) , (Eggs)	0.077
(Fresh fruits) , (Fresh herbs)	0.077	(Packaged vegetables and fruit) , (Frozen produce)	0.076
(Fresh fruits) , (Fresh dips and tapenades)	0.075	(Fresh vegetables) , (Refrigerated produce)	0.073
(Yogurt) , (Bread)	0.071	(Fresh fruits) , (Ice cream)	0.070
(Fresh fruits) , (Cereal)	0.068	(Milk) , (Bread)	0.067
(Fresh vegetables) , (Lunch meat)	0.065	(Packaged cheese) , (Bread)	0.065
(Fresh vegetables) , (Fresh dips and tapenades)	0.065	(Fresh fruits) , (Hot dogs, bacon and sausage)	0.064
(Fresh fruits) , (Energy and granola bars)	0.063	(Fresh fruits) , (Cream)	0.063
(Fresh vegetables) , (Crackers)	0.062	(Yogurt) , (Chips and pretzels)	0.062
(Packaged vegetables and fruit) , (Refrigerated produce)	0.062	(Packaged cheese) , (Chips and pretzels)	0.062
(Fresh fruits) , (Other creams and cheeses)	0.062	(Fresh vegetables) , (Soup broth and bouillon)	0.061
(Yogurt) , (Seltzer and sparkling water)	0.061	(Fresh fruits) , (Juice and nectars)	0.061
(Fresh fruits) , (Soup broth and bouillon)	0.061	(Fresh vegetables) , (Canned and jarred vegetables)	0.061

Contains the 60 most frequent pairs of categories. 20.18 % of the baskets with more than one category do not contain any of these 60 pairs.

Table 2: Relative bivariate marginal frequencies

5 Methodology

First I will introduce some general notation for the variables and their indices. The indices of market baskets are denoted by $i = 1, ..., I$ and indices of categories by $j = 1, ..., J$. Basket i is characterized by a binary column vector of purchase incidences $\mathbf{y}_i = (y_{i1}, ..., y_{iJ})'$. If market basket i contains category j, then $y_{ij} = 1$, otherwise $y_{ij} = 0$.

5.1 Multivariate Logistic Model

G. J. Russell and Petersen (2000) were the first to introduce the multivariate logit (MVL) model to the field of market basket analysis. They formulated the probability of market basket i as:

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

with

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

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 ($V_{jl} = V_{lj}$). A positive cross-category coefficient indicates that categories are complements of each other and a negative coefficient means they are substitutes.

The normalization constant Z_{MVL} is a sum over all the possible market baskets. This constant as denominator will turn equation (1) into a valid probability. This means that the value will always be larger than 0 and smaller than 1 and that the sum over all the probabilities will add up to 1.

From equation 1 we can derive the conditional probabilities of purchase incidences y_{ij} for all categories given purchase incidences of the other $J - 1$ categories:

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

As stated earlier, estimation of the MVL model is impossible for a higher number of categories. To work around this problem I make use of the maximum pseudo-likelihood (MPL) method (Besag, 1974). For the pseudo-likelihood of the MVL I make use of the conditional probabilities given in equation 2 instead of the unconditional probabilities which is customary for the ML estimation. Taking the log of this results in the log pseudo-likelihood (LPL), this results in the following equation:

$$LPL = \sum_i LPL_i,$$

with

$$LPL_i = \sum_j [Y_{ij} * log(p(y_{ij}|\mathbf{y}_{i,-j})) + (1 - Y_{ij}) * log(1 - p(y_{ij}|\mathbf{y}_{i,-j}))]. \tag{3}$$

Here LPL_i denotes the log pseudo-likelihood of basket i. I estimate this model by means of the BFGS algorithm using the following first derivatives with respect to constants a_j and cross-category coefficients V_{jl}:

$$\frac{\partial LPL_i}{\partial a_j} = y_{ij} - p(y_{ij}|\mathbf{y}_{i,-j}) \quad \text{for } j = 1, ..., J,$$
$$\frac{\partial LPL_i}{\partial V_{jl}} = 2y_{ij}y_{il} - y_{il}p(y_{ij}|\mathbf{y}_{i,-j}) - y_{ij}p(y_{il}|\mathbf{y}_{i,-l}) \quad \text{for } l > j.$$

5.2 Restricted Boltzmann Machine

The restricted Boltzmann machine (RBM) introduced by Smolensky (1986) makes use of K binary hidden variables. These hidden variables are put in a column vector $\mathbf{h}_i = (h_{i1}, ..., h_{iK})'$ and are all connected to the observed binary variables. The RBM does not have any connections from one hidden variable to another hidden variable and the same applies to the observed variables, hence it is called restricted. The joint probability function of the market baskets and the hidden variables is:

$$p(\mathbf{y}_i, \mathbf{h}_i) = \frac{exp(\mathbf{b}'\mathbf{y}_i + \mathbf{h}_i'\mathbf{W}\mathbf{y}_i)}{Z_{RBM}}, \tag{4}$$

with

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

Column vector $\mathbf{b}$ contains constants for the categories. The $K \times J$ matrix $\mathbf{W}$ contains coefficients which link each hidden variable to each category. The normalization constant Z_{RBM} is a sum over all the possible market baskets and all the possible configurations of hidden variables. This constant as denominator will turn equation 4 into a valid probability just like we saw with the MVL model. Since I am interested in the marginal probability function of the market baskets I have to take the sum of equation 4 over all the possible configurations of hidden variables:

$$p(\mathbf{y}_i) = \frac{p^*(\mathbf{y}_i)}{Z_{RBM}} \quad \text{with} \quad p^*(\mathbf{y}_i) = \sum_{\mathbf{h}\in\{0,1\}^K} exp(\mathbf{b}'\mathbf{y}_i + \mathbf{h}'\mathbf{W}\mathbf{y}_i). \tag{5}$$

I will call $p^*(\mathbf{y}_i)$ the unnormalized probability.

From equation 4 we can derive the conditional probabilities of purchase incidences y_{ij} for all categories given hidden variables h_{ik} and vice versa:

$$p(y_{ij}|\mathbf{h}_i) = \frac{1}{1 + exp(-(b_j + \sum_k W_{kj}h_{ik}))} \quad \text{for all } i, j, \tag{6}$$
$$p(h_{ik}|\mathbf{y}_i) = \frac{1}{1 + exp(-\sum_j W_{kj}y_{ij})} \quad \text{for all } i, k. \tag{7}$$

These conditional probabilities are only used for Gibbs sampling (see section 5.4). They are not used for the pseudo-likelihood of the RBM, because we can estimate the model with conventional ML estimation, which is impossible for the MVL model. The difference with the

MVL model becomes more clear when I rewrite the normalization constant as:

$$Z_{RBM} = \sum_{\mathbf{h}\in\{0,1\}^K} \prod_{j=1}^{J} (1 + exp(\mathbf{W}'_{\cdot,j}\mathbf{h} + b_j)). \tag{8}$$

Here $\mathbf{W}'_{\cdot,j}$ denotes the jth column of $\mathbf{W}$. Equation 8 clearly shows that estimation is significantly less-involved than the MVL model, if the number of hidden variables is much lower than the number of categories. Therefore, I ought to estimate the coefficients with ML estimation. In order to do this I have to rewrite equation 5 with $\mathbf{W}_{k,\cdot}$ denoting the kth row of W:

$$p^*(\mathbf{y}_i) = exp(\mathbf{b}'\mathbf{y}_i) \prod_{k=1}^{K} (1 + exp(\mathbf{W}_{k,\cdot}\mathbf{y}_i)). \tag{9}$$

A special case of the RBM is one where the coefficient matrix $\mathbf{W}$ equals to 0. We call this the independence model and ML estimates are:

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

The normalization constant Z_0 of the indepence model equals:

$$Z_0 = 2^K \prod_j (1 + exp(b_j)). \tag{11}$$

The coefficients of the RBM can be estimated by maximizing the log likelihood (LL) across baskets:

$$LL = \sum_i log(p(\mathbf{y}_i)) = \sum_i log(p^*(\mathbf{y}_i)) - I * log(Z_{RBM}). \tag{12}$$

This maximization is done using the BFGS algorithm in three random restarts. I then select the coefficients that achieve the greatest log likelihood. Three random restarts are necessary because log likelihood seem to slightly differ with other initial values. This randomness only applies to the $\mathbf{W}$ coefficient matrix, since the constant terms of $\mathbf{b}$ can be initialized by their values for the independence model. $\mathbf{W}$ values are initialized by a random number from the normal distribution with mean zero and standard deviation equal to 0.5, just like Hruschka (2014) suggests. The first derivatives of the log likelihood can be calculated by:

$$\frac{\partial LL}{\partial b_j} = \sum_i y_{ij} - \frac{I}{Z_{RBM}} \sum_{n=1}^{2^K} \frac{Z_{RBM}^{(n)}}{1 + exp(-(\mathbf{W}'_{\cdot,j}\mathbf{h}^{(n)} + b_j))}, \tag{13}$$

$$\frac{\partial LL}{\partial W_{kj}} = \sum_i \frac{y_{ij}}{1 + exp(-(\mathbf{W}_{k,\cdot}\mathbf{y}_i))} - \frac{I}{Z_{RBM}} \sum_{n=1}^{2^K} \frac{Z_{RBM}^{(n)} h_k^{(n)}}{1 + exp(-(\mathbf{W}'_{\cdot,j}\mathbf{h}^{(n)} + b_j))}. \tag{14}$$

Here $Z_{RBM}^{(n)}$ and $\mathbf{h}^{(n)}$ denote the value that Z_{RBM} and the hidden variable k assume in the n^{th} configuration, respectively. Note the resemblance of the divisions to the conditional probabilities given in equation 6 and 7.

5.3 Deep Belief Network

The deep belief network (DBN) can be considered as an extended version of the RBM. It consists as it were of a network of stacked RBMs, where hidden variables of one layer are connected to hidden variables of the other layer (Hruschka, 2021). For the sake of simplicity I only consider the DBN with two layers. This does not, in fact, disregard the validity of this model. It has been proven by Salakhutdinov and G. Hinton (2012) that networks with more layers do not guarantee better results.

The first layer of the network is based on the observed baskets and is therefore the same as the RBM model discussed in section 5.2. The second layer is based on the hidden variables of the first hidden layer. The conditional probabilities are given by (Hruschka, 2021):

$$p(y_{ij}|\mathbf{h}_{1i}) = \frac{1}{1+exp(-(b_{1j}+\sum_k W_{1kj}h_{1ik}))} \quad \text{for all } i,j, \tag{15}$$

$$p(h_{1ik}|\mathbf{y}_i) = \frac{1}{1+exp(-(d_{1k}+\sum_j W_{1kj}y_{ij}))} \quad \text{for all } i,k. \tag{16}$$

$$p(h_{1ik}|\mathbf{h}_{2i}) = \frac{1}{1+exp(-(b_{2k}+\sum_l W_{2lk}h_{2il}))} \quad \text{for all } i,k, \tag{17}$$

$$p(h_{2il}|\mathbf{h}_{1i}) = \frac{1}{1+exp(-(d_{2l}+\sum_k W_{2lk}h_{1ik}))} \quad \text{for all } i,l. \tag{18}$$

Here b_{1j}, b_{2k}, d_{1k} and d_{2l} denote constant terms for the categories and hidden variables. W_{1kj} and W_{2lk} are coefficients that link each hidden variable of the first layer to each category and each hidden variable of the second layer, respectively. K still denotes the number of hidden variables in the first layer, whereas L stands for the number of hidden variables in the second layer. h_{1ik} and h_{2il} denote the k^{th} first and second layer hidden variable of basket n, respectively. $\mathbf{h}_{1i}$ and $\mathbf{h}_{2i}$ are vectors of first and second layer hidden variables.

A third layer can be added to link the second hidden layer back to the observed baskets. Its probability can be written as:

$$p(y_{ij}|\mathbf{h}_{2i}) = \frac{1}{1+exp(-(d_{3j}+\sum_l W_{3jl}h_{2il}))} \quad \text{for all } i,j, \tag{19}$$

Variable denotation is similar to that of the other layers.

Estimation of the coefficients of a DBN can be done by means of the greedy layerwise algorithm of Geoffrey E. Hinton et al. (2006). In this algorithm one first estimates the parameters of the first hidden layer with the Contrastive Divergence algorithm of Geoffrey E. Hinton (2002). Then the hidden variables of this layer are drawn by Gibbs sampling and used for estimation of the second layer coefficients. All hidden variables of the two layers are set to mean field expected values given by (Geoffrey E. Hinton et al., 2006):

$$\mu_{1ik} = \frac{1}{1+exp(-(d_{1k}+\sum_j W_{1kj}y_{ij}))} \tag{20}$$

$$\mu_{2il} = \frac{1}{1+exp(-(d_{2l}+\sum_k W_{2lk}\mu_{1ik}))} \tag{21}$$

Then, all the conditional probabilities are nested together to make sure that all the coefficients can be estimated simultaneously. Resulting in a nonlinear equation of the coefficients. These coefficients are estimated by nonlinear least squares, where the earlier results of the greedy layerwise algorithm are used as initial values. For the estimation of this model I make use of the deepnet package of R (Rong, 2014).

5.4 Model Comparison

For the comparison of the models I use different techniques. First, I compare the independence model and the different RBMs based on the log likelihood values. I say different RBMs because we would also like to analyze the effect of different numbers of hidden variables. I compare the log likelihood of both the estimation data and the validation data. In addition to that I also compute the Bayesian information criterion (BIC) for these models, this is defined as:

$$BIC = -2LL + n * log(I), \tag{22}$$

where LL denotes the log likelihood, n the number of parameters and I the number of market baskets. I do the same for the MVL model, based on pseudo likelihood. However, one cannot use these methods to compare the RBM with the MVL model, because the likelihoods of these models are not interpreted the same. After all, I concluded that I could not calculate the conventional log likelihood of the MVL model. The same goes for the DBN, where no likelihood is calculated whatsoever. To compare these models one first has to generate 300 artificial datasets by Gibbs sampling over the corresponding conditional probability functions. The absolute deviation (AD) can then be defined as:

$$AD = \frac{1}{300} \sum_{s=1}^{300} \sum_{i} \sum_{j} |y_{ij} - \hat{y}_{ijs}|, \tag{23}$$

where $\hat{y}_{ijs}$ denotes the s^{th} sampled purchase incidence value of category j in baskter i.

For Gibbs sampling (first described by S. Geman and D. Geman (1984)) of the MVL model one calculates the conditional probability of this model (equation 2) for all the observations. This can be done per category where the probabilities of one category depend on all the previous categories of the same Gibbs sample and on all the following categories of the original sample. Baskets of the RBM can be generated by blockwise Gibbs sampling, because purchases only depend on hidden variables and vice versa (Hruschka, 2014)). For this model probabilities of hidden variables are calculated using equation 7 for the original data. Then, the values of the hidden variables are used to calculate the probabilities of the new basket using equation 6. For both of the models one has to turn the probabilities into binary variables directly after calculating them by means of a generated Bernoulli trial.

The 300 Gibbs samples are also used to test the statistical significance of the estimates by parametric bootstrappig (Cameron and Trivedi, 2005). Estimation of the variables for all of these samples result in different estimates. Dividing means by standard deviations of these estimates results in t-values. Values larger than 1.65 are then considered to be significant on a 10 percent significance level.

5.5 Calculation of Cross Effects

Calculation of the cross effects of the categories can be done by first deriving the share $\langle y_j \rangle$ of category j and the share $\langle h_k \rangle$ of hidden variable k:

$$\langle y_j \rangle = \frac{1}{1 + exp(-(b_j + \sum_k W_{kj} \langle h_k \rangle} \tag{24}$$

$$\langle h_k \rangle = \frac{1}{1 + exp(-\sum_j W_{kj} \langle y_j \rangle}. \tag{25}$$

Since the values of $\langle y_j \rangle$ and $\langle h_k \rangle$ cannot be determined analytically for the RBM, Hruschka (2014) suggests to set them equal to averages of categories and hidden variables in 15,000 pseudo-baskets, respectively. These averages are acquired by blockwise Gibbs sampling explained in section 5.4. The statistical significance can again be determined by parametric bootstrapping. Partial marginal cross effects between two categories are then defined as the contribution of hidden variable k to the overall marginal cross effects:

$$\langle y_j \rangle (1 - \langle y_j \rangle) W_{kj} W_{kl} \langle h_k \rangle (1 - \langle h_k \rangle) \tag{26}$$

6 Results

Table 3 shows that all the RBMs perform better than the independence model in terms of both the log likelihood and the BIC. This suggests that cross effects of products should not be ignored, which is in line with Hruschka (2014).

Also in line with Hruschka (2014) is the increase in performance for larger numbers of hidden variables. Table 3 shows an increase in log likelihood as well as a decrease in BIC. RBMs also seem to be robust as the log likelihood of the validation stays approximately the same.

Number of hidden variables	Number of parameters	Log (pseudo-)likelihood (L(P)L)		Bayesian Information Criterion (BIC)	Absolute Deviation	Significant Cross-effects
		Estimation data	Validation data			
Restricted Boltzmann machines (RBMs)						
Independence model	60	-271,334	-271,119	543,244	-	-
1	120	-268,521	-268,506	538,196	149,464	2,640
2	180	-266,568	-266,658	534,867	146,580	1,791
3	240	-265,068	-265,322	532,445	144,386	1,530
4	300	-263,820	-264,103	530,524	142,508	1,744
5	360	-262,759	-263,099	528,980	141,025	1,321
Multivariate Logistic model (MVL model)						
-	1830	-245,858	-248,892	509,313	139,738	-

L(P)L, BIC and AD values are rounded to the nearest integer value; RBMs differ with respect to the number of hidden variables; BIC value of the MVL model is based on the log pseudo-likelihood; Significant Cross-effects with an absolute t-value larger than 1.65.

Table 3: Comparison of alternative models

Table 3 also tells us that the number of significant cross effects decreases with the number of hidden variables (with an exception for three hidden variables). This is most likely because of an increase in variance due to an increase in the number of parameters to be estimated. Out of all the RBMs considered it seems best to further analyse the model with five hidden variables. After all, both the log likelihood and the BIC reach their greatest values for this model. However, I choose to further analyse the RBM model with four hidden variables, because I like to further explain the differences of this paper with that of Hruschka (2014) and it seems best to keep as much factors the same.

First layer \Second layer	1 Hidden variable	2 Hidden variables	3 Hidden variables	4 Hidden variables	5 Hidden variables
1 Hidden variable	156,115	169,205	173,847	168,313	166,029
2 Hidden variables	152,085	148,588	149,048	150,548	149,749
3 Hidden variables	149,790	151,944	147,801	147,543	147,440
4 Hidden variables	150,000	150,602	147,758	145,507	147,324
5 Hidden variables	155,769	150,036	147,360	146,814	144,319

Values are rounded to the nearest integer; rows indicate the number of hidden variables in the first layer and columns in the second layer

Table 4: Absolute deviations of the deep belief network

The absolute deviation of the MVL model appears to be lower than that of the RBMs, in contrast to Hruschka (2014). However, table 3 clearly shows a downward trend in the absolute deviations for the RBMs. It is in the line of expectation that the RBM outperforms the MVL model for larger number of hidden variables. Even though Hruschka (2014) states that the RBM outperforms the MVL model because of the lower number of parameters to be estimated, one must be careful not to choose too few hidden variables, since this might result in the model not getting the full picture. The DBN does not perform any better in terms of absolute deviations (table 4). This is because of the small number of hidden variables in the layers. Hruschka (2021) already stated that RBMs and DBNs might perform just as well when the number of hidden variables is kept small. The absolute deviations do make a big drop when going beyond one hidden variable in the first layer. This can be explained by the way the coefficients are estimated. The deepnet package of Rong (2014) appears to give an error when estimating the coefficients if the number of hidden variables in the first layer equals one. To overcome this error I tried to skip the initialization of the coefficients of the first two layers. This results in a random initialization of these coefficients. This indicates that the greedy layerwise algorithm does indeed outperform nonlinear least squares with random initial values. Although the DBN is outperformed by the other models, there does seem to be a pattern for five hidden variables in either of the layers. Table 4 clearly shows that if one of the layers contains five hidden variables, the absolute deviation decreases with number of hidden variables in the other layer. This could indicate that the benefits of the DBN starts to work for five hidden variables and onwards.

(Fresh herbs) , (Fresh vegetables)	2.567	(Dry pasta) , (Fresh vegetables)	1.891
(Canned and jarred vegetables) , (Fresh vegetables)	1.831	(Tortillas and flat bread) , (Fresh vegetables)	1.796
(Pasta sauce) , (Fresh vegetables)	1.738	(Fresh herbs) , (Packaged vegetables and fruit)	1.121
(Fresh vegetables) , (Fresh herbs)	0.821	(Dry pasta) , (Packaged cheese)	0.608
(Pasta sauce) , (Packaged cheese)	0.596	(Soup broth and bouillon) , (Fresh fruits)	-0.586
(Canned and jarred vegetables) , (Packaged cheese)	0.586	(Fresh herbs) , (Eggs)	0.562
(Spices and seasonings) , (Fresh herbs)	0.554	(Soup broth and bouillon) , (Fresh herbs)	0.542
(Dry pasta) , (Eggs)	0.538	(Tortillas and flat bread) , (Packaged vegetables and fruits)	0.531
(Pasta sauce) , (Eggs)	0.511	(Fresh herbs) , (Seltzer and sparkling water)	-0.491
(Canned meals and beans) , (Packaged cheese)	0.489	(Tortillas and flat bread) , (Eggs)	0.488
(Dry pasta) , (Packaged produce)	-0.486	(Pasta sauce) , (Packaged produce)	-0.482
(Fresh herbs) , (Soup broth and bouillon)	0.481	(Fresh vegetables) , (Canned and jarred vegetables)	0.461
(Fresh vegetables) , (Dry pasta)	0.457	(Dry pasta) , (Packaged vegetables and fruit)	0.457
(Grains, rice and dried goods) , (Packaged cheese)	0.453	(Fresh herbs) , (Frozen produce)	0.447
(Canned and jarred vegetables) , (Eggs)	0.443	(Canned and jarred vegetables) , (Seltzer and sparkling water)	-0.431
(Pickled goods and olives) , (Fresh herbs)	0.425	(Pasta sauce) , (Bread)	0.425
(Dry pasta) , (Bread)	0.422	(Tortillas and flat bread) , (Packaged cheese)	0.422
(Dry pasta) , (Hot dogs, bacon and sausage)	0.418	(Spices and seasonings) , (Fresh vegetables)	0.415
(Fresh herbs) , (Packaged cheese)	0.414	(Pasta sauce) , (Packaged vegetables and fruits)	0.410
(Pasta sauce) , (Hot dogs, bacon and sauasge)	0.407	(Fresh herbs) , (Bread)	0.397

Absolute t values >1.65; first: independent category; second: dependent category

Table 5: The 40 largest, significant marginal cross effects for the RBM with four hidden variables

The 40 largest, significant marginal cross effects of the RBM with four hidden variables can be found in table 5. Six perfect complements appear to be present in the data, as opposed to no perfect substitutes. There are some negative cross-effects, however. Fresh herbs have the most effect on other categories with a total of eight outgoing significant cross effects. Dry pasta and pasta sauce also seem to be having a lot of influence on the other categories. They both influence the purchases of seven other categories. Looking the other way around, table 5 shows that packaged cheese and fresh vegetables are influenced the most by other products with a total of seven and six, respectively. Remarkable is the over-representation of fresh vegetables in the perfect complements. Five of the six of these cross effects influence the purchase of fresh vegetables.

It can be seen from equation 7 that negative (positive) W_{kj} coefficients results in lower (higher) probabilities for the hidden variable to be equal to one when category j is purchased. Table 6 shows these coefficients and what strikes immediately is that all these coefficients are negative. Positive effects of categories on hidden variable appear to be much smaller than the negative effects. In fact, not only their value appears to be smaller, but also the number of positive effects is outnumbered by the negative effects. The threshold of 1.10 in table 6 indicates that all the given categories result in chances of hidden variable k being equal to zero are at least thrice as big as the chances of it being equal to one (Hruschka, 2014). For example, purchasing condiments will result in lower (higher) chances of hidden variable three being equal to one (zero). Looking at the positive coefficients, however, one can conclude that the purchase of packaged produce will only result in twice as greater chances of the first hidden variable being equal to one

Category	Hidden variable			
	1	2	3	4
Fresh vegetables	-1.51	-1.43	-1.44	-1.38
Lunch meat	-1.25	-1.31	-1.30	-1.34
Fresh herbs	-5.50	-5.22	-5.65	-5.24
Soup broth and bouillon	-3.27	-3.00	-3.61	-3.18
Canned and jarred vegetables	-5.96	-5.70	-7.10	-5.96
Dry pasta	-6.25	-5.88	-6.53	-6.14
Canned meals and beans	-5.05	-4.46	-5.40	-4.09
Pasta sauce	-5.49	-5.59	-6.81	-5.87
Condiments			-1.12	
Frozen breakfast	-3.19	-3.19	-3.94	-3.36
Tortillas and flat bread	-4.58	-4.54	-5.02	-4.28
Spices and seasonings	-2.77	-2.37	-3.18	-2.57
Instant foods	-1.58	-1.69	-2.12	-1.70
Frozen appetizers and sides	-2.62	-2.76	-2.78	-2.57
Asian foods	-2.39	-2.36	-3.48	-2.73
Frozen pizza	-1.87	-1.98	-1.49	-1.89
Grains, rice and dried goods	-4.35	-4.15	-5.37	-4.63
Poultry counter	-2.16	-2.07	-1.67	-2.03
Packaged poultry	-1.31	-1.42	-1.49	-1.24
Tofu and meat alternatives	-2.21	-2.56	-2.34	-2.33
Pickled goods and olives	-2.58	-2.3	-3.06	-2.96

Only coefficients with absolute values >1.10

Table 6: Selected $W_{k,j}$ coefficients

Table 7 shows the significant partial marginal cross effects over the four hidden variables given by equation 26. The absence of the 40 largest overall marginal cross effects is remarkable, but can be explained by means of the standard deviation. Summing up the partial marginal cross effects results in a smaller increase in standard deviation than it does in the mean. There are, however, multiple cross effects that have significant cross effects over all the hidden variables. Lunch meat and packaged cheese, for instance, have a significant positive interdependence. Frozen meals and packaged produce on the other hand have a negative interdependence over all the hidden variables. Ice cream and packaged produce have a negative interdependence over the first, second and fourth hidden variable. All other significant effects are positive over all the hidden variables.

Hidden variable 1		Hidden variable 2	
(Lunch meat) , (Packaged cheese)	0.031	(Lunch meat) , (Packaged cheese)	0.032
(Hot dogs, bacon and sausage) , (Packaged cheese)	0.024	*(Fruit and vegetable snacks) , (Energy and granola bars)*	0.024
(Other creams and cheeses) , (Packaged cheese)	0.022	(Hot dogs, bacon and sausage) , (Packaged cheese)	0.023
(Bread) , (Packaged cheese)	0.021	(Bread) , (Packaged cheese)	0.021
(Hot dogs, bacon and sausage) , (Lunch meat)	0.018	(Other creams and cheeses) , (Packaged cheese)	0.021
(Hot cereal and pancake mixes) , (Packaged cheese)	0.018	(Hot cereal and pancake mixes) , (Packaged cheese)	0.018
(Frozen meals) , (Packaged produce)	-0.017	(Frozen meals) , (Packaged produce)	-0.017
(Packaged cheese) , (Eggs)	0.016	(Packaged cheese) , (Bread)	0.016
(Other creams and cheeses) , (Lunch meat)	0.016	(Packaged cheese) , (Lunch meat)	0.016
(Lunch meat) , (Hot dogs, bacon and sausage)	0.015	(Packaged produce) , (Ice cream)	-0.016
(Packaged cheese) , (Lunch meat)	0.015	(Other creams and cheeses) , (Lunch meat)	0.015
(Packaged produce) , (Ice cream)	-0.015	(Packaged produce) , (Frozen meals)	-0.015
(Lunch meat) , (Other creams and cheeses)	0.014	(Lunch meat) , (Other creams and cheeses)	0.013
(Packaged produce) , (Frozen meals)	-0.014	(Hot cereal and pancake mixes) , (Lunch meat)	0.013
(Hot cereal and pancake mixes) , (Bread)	0.013	(Ice cream) , (Packaged produce)	-0.012
(Hot cereal and pancake mixes) , (Lunch meat)	0.013	(Packaged cheese) , (Hot dogs, bacon and sausage)	0.010
(Fruit and vegetable snacks) , (Popcorn and jerky)	0.012	(Packaged cheese) , (Other creams and cheeses)	0.009
(Popcorn and jerky) , (Fruit and vegetable snacks)	0.012	(Lunch meat) , (Soy and lactose-free)	0.008
(Ice cream) , (Packaged produce)	-0.011	(Lunch meat) , (Hot cereal and pancake mixes)	0.006
(Packaged cheese) , (Hot dogs, bacon and sausage)	0.010	(Soy and lactose-free) , (Lunch meat)	0.005
(Packaged cheese) , (Other creams and cheeses)	0.009	(Packaged cheese) , (Hot cereal and pancake mixes)	0.004
(Lunch meat) , (Soy and lactose-free)	0.008		
(Lunch meat) , (Hot cereal and pancake mixes)	0.006		
(Soy and lactose-free) , (Lunch meat)	0.005		
(Bread) , (Hot cereal and pancake mixes)	0.004		
(Packaged cheese) , (Hot cereal and pancake mixes)	0.004		
Hidden variable 3		Hidden variable 4	
(Lunch meat) , (Packaged cheese)	0.032	(Lunch meat) , (Packaged cheese)	0.032
(Hot dogs, bacon and sausage) , (Packaged cheese)	0.024	(Bread) , (Packaged cheese)	0.020
(Other creams and cheeses) , (Packaged cheese)	0.022	(Hot cereal and pancake mixes) , (Packaged cheese)	0.018
(Hot cereal and pancake mixes) , (Packaged cheese)	0.018	(Energy and granola bars) , (Fresh dips and tapenades)	0.017
(Energy and granola bars) , (Fresh dips and tapenades)	0.017	(Frozen meals) , (Packaged produce)	-0.017
(Frozen meals) , (Packaged produce)	-0.017	(Packaged cheese) , (Bread)	0.016
(Packaged cheese) , (Lunch meat)	0.016	(Packaged cheese) , (Lunch meat)	0.016
(Other creams and cheeses) , (Lunch meat)	0.016	(Packaged produce) , (Ice cream)	-0.016
(Fresh dips and tapenades) , (Energy and granola bars)	0.015	(Fresh dips and tapenades) , (Energy and granola bars)	0.015
(Lunch meat) , (Other creams and cheeses)	0.014	(Other creams and cheeses) , (Lunch meat)	0.015
(Packaged produce) , (Frozen meals)	-0.014	*(Energy and granola bars) , (Fruit and vegetable snack)*	0.014
(Hot cereal and pancake mixes) , (Lunch meat)	0.013	(Packaged produce) , (Frozen meals)	-0.014
(Packaged cheese) , (Other creams and cheeses)	0.010	(Lunch meat) , (Other creams and cheeses)	0.013
(Lunch meat) , (Soy and lactose-free)	0.009	*(Hot cereal and pancake mixes) , (Lunch meat)*	0.013
(Lunch meat) , (Hot cereal and pancake mixes)	0.006	(Ice cream) , (Packaged produce)	-0.012
(Soy and lactose-free) , (Lunch meat)	0.005	(Lunch meat) , (Soy and lactose-free)	0.008
(Packaged cheese) , (Hot cereal and pancake mixes)	0.004	(Lunch meat) , (Hot cereal and pancake mixes)	0.006
		(Soy and lactose-free) , (Lunch meat)	0.005
		(Packaged cheese) , (Hot cereal and pancake mixes)	0.004

Contains partial marginal cross effects >= 0.002 with absolute t-values >= 1.65; italic: partial marginal cross effects specific to a hidden variable

Table 7: Partial marginal cross effects for each hidden variable

Only five categories out of the twenty-two with the lowest univariate purchase frequencies have large, significant cross effects with other categories. All of these categories only have an influence on other category purchases and are not in their turn influenced by these other categories. There is a positive interdependence between fresh vegetables and three other categories, namely fresh herbs, canned and jarred vegetables & dry pasta. Fresh herbs also has a positive interdependence with soup broth and bouillon. Out of the 40 largest cross effects, 32 only effect categories in a one way manner. The purchase of fresh vegetables is positively influenced by pasta sauce, tortillas and flat bread & spices and seasonings. Spices and seasonings also has a positive effect on the purchase of fresh herbs, just like pickled goods and olives. Fresh herbs in their turn have a positive effect on packaged vegetables and fruit, packaged cheese, bread, eggs and frozen produce and a negative effect on seltzer and sparkling water. Soup broth and bouillon has the largest negative cross effect on fresh fruits. Canned and jarred vegetables has two positive effects and one negative effect on packaged chees, eggs & seltzer and sparkling water, respectively. Dry pasta has a negative effect on packaged produce, but a positive effect on packaged vegetables and fruits, packaged cheese, bread, eggs & hot dogs, bacon and sausage. The remaining eleven cross effects do not have a direct connection with the interdependent categories. Four of the nine categories with the largest purchase frequencies do not appear in any of the forty largest cross effects. These categories are yogurt, milk, chips and pretzels & soy and lactose-free. This is probably because a large number of baskets do not contain any other categories than one of these four, as this was also the case in Hruschka (2014).

The difference in results between modelling the cross effects as in this paper and by pairwise purchase incidences becomes clearer when comparing table 2 and table 5. Two out of three results of Hruschka (2014) also apply to this dataset:

1. Yogurt and milk do not appear in any of the 40 largest cross effects despite their relatively high pairwise purchase frequencies with other categories (yogurt with fresh fruits, fresh vegetables, packaged vegetables and fruit, milk, packaged cheese, bread, chips and pretzels & seltzer and sparkling water; milk with fresh fruits, fresh vegetables, packaged vegetables and fruit, yogurt, packaged cheese and bread). These high pairwise frequencies can be explained by a relatively high univariate purchase frequency, i.e. these are products that people always tend to buy.

2. Pairwise purchase frequencies are unable to detect negative cross effects. In fact, the pairwise purchase frequency of fresh fruits and soup broth and bouillon is among the top 60 in table 2, but the RBM gives a negative cross effect between these two categories.

3. Hruschka (2014) states that high purchase frequencies between multiple products can be traced back to high cross effects with other products. However, this is not evident by the results in this paper.

7 Conclusion

Contracting the MVL model into a RBM enables us to analyse more categories at once. It is even possible to estimate the RBM by means of maximum likelihood as long as the number of hidden variables remain relatively low. However, it is still possible to estimate the RBM for larger numbers of hidden variables. In this case one has to resort the contrastive divergence algorithm which approximates the marginal log likelihood, just as for the DBN. The MVL model is estimated by maximizing the log pseudo-likelihood in this paper, instead of the two-step approach implemented by earlier papers. In contrast to Hruschka (2014), the MVL model outperforms the RBM in terms of absolute deviation. However, it is expected that the RBM will perform better if more hidden variables are added. The DBN performs worse than the other models. This can again be explained by the fact that too few hidden variables are present in the hidden layers. I expect that the DBN will become the best model of the three if more variables are added to the layers. This is yet to be investigated, however.

Based on the log likelihood values of both the estimation data as the validation data, as well as the Bayesian Information Criterion, we can conclude that cross effects have a clear effect on purchase behavior of people.

We can safely state that cross effects do not appear to be symmetrical, as no pair can be found of which their cross-effects appear to be equal to each other. The number of significant cross effects is considerably larger than the number found by Hruschka (2014). It is, in fact, only less than the 1,770 possible effects in the MVL model in 3 out of 5 RBMs. In total 403 of the 1,321 significant cross effects is negative. Five of these can be found among the 40 largest cross effects. Furthermore, large bivariate purchase frequencies do not have to lead to high cross effects between these products, since some products are just bought in the majority of the shopping trips.

Knowledge over the existing cross effects in a store can be beneficial for the manager. G. J. Russell and Wagner A Kamakura (1997) and G. J. Russell and Petersen (2000) already showed us that promotions in one category will influence the sales of another category. This means that a manager can improve the sales of one category by making another category more attractive with marketing measures such as sales and commercials, but also product placing may benefit from these cross effects. One may think that placing categories with positive cross effects closer to each other might improve the sales of either.

These marketing measure will not have any effect on independent categories even though their pairwise purchase frequency might be relatively high. This applies to fresh fruits and soup broth and bouillon in this data set for instance. Where their purchase frequencies belonged to 60 largest of the data, but their cross effects were found to be negative.

Marketing measures in any of the interdependent categories like fresh vegetables, fresh herbs, canned and jarred vegetables, dry pasta and soup broth and bouillon will always have a positive effect on any of the other categories. It is therefore advised to apply marketing measurements in any of these categories. Furthermore, fresh herbs also has the most influence on other categories outside of the interdepent categories.

Some marketing measure seem to be straightforward, but can actually be improved by looking at the cross effects. For instance, one might think that marketing measures on bread will improve the sales of packaged cheese the most. However measures on a less straightforward category like the canned meals and beans will increase these sales even more. If the manager decides to increase the sales of bread, one will achieve a greater efficiency by applying marketing measures to dry pastas or pasta sauces, compared to the packaged cheese or the hot dogs.

Some marketing measures even have to be avoided when the manager is looking to increase the sales of certain products. In this data set soup broth and bouillon exerts negative effects on the sales of fresh fruits for instance. So when looking to increase these sales one must avoid marketing measures in this category.

This paper contributes to the paper of Hruschka (2014) in the sense that it adds and compares an extra model to the earlier models. It also provides some new perspectives on the results of Hruschka, which leads to other conclusions. Some of the further research ideas of Hruschka still stand though, like the addition of explanatory variables as well as the research of models for analysing product quantities in stead of instances. I would also like to add the comparison with other neural networks to the list of further research ideas.

References

[1] Julian Besag. "Spatial Interaction and the Statistical Analysis of Lattice Systems". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 36.2 (1974), pp. 192–236. ISSN: 0035-9246.

[2] Roger Betancourt and David Gautschi. "Demand Complementarities, Household Production, and Retail Assortments". In: *Marketing Science* 9.2 (1990), pp. 146–161. ISSN: 0732-2399.

[3] Yasemin Boztug and Lutz Hildebrandt. "Modeling Joint Purchases with a Multivariate MNL Approach". In: *Schmalenbach Business Review (sbr)* 60 (Oct. 2008), pp. 400–422. DOI: 10.1007/BF03396777.

[4] Yasemin Boztuğ and Thomas Reutterer. "A Combined Approach for Segment-Specific Market Basket Analysis". en. In: *European Journal of Operational Research* 187.1 (May 2008), pp. 294–312. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2007.03.001.

[5] A Colin Cameron and Pravin K Trivedi. "Microeconometrics : Methods and Applications". en. In: (2005), p. 1062.

[6] Sri Devi Duvvuri, Asim Ansari, and Sunil Gupta. "Consumers' Price Sensitivities Across Complementary Categories". In: *Management Science* 53.12 (Dec. 2007), pp. 1933–1945. ISSN: 0025-1909. DOI: 10.1287/mnsc.1070.0744.

[7] Sri Devi Duvvuri and Thomas S. Gruca. "A Bayesian Multi-Level Factor Analytic Model of Consumer Price Sensitivities Across Categories". en. In: *Psychometrika* 75.3 (Sept. 2010), pp. 558–578. ISSN: 1860-0980. DOI: 10.1007/s11336-010-9167-3.

[8] Gérard Gaudet and Stephen W. Salant. "Increasing the Profits of a Subset of Firms in Oligopoly Models with Strategic Substitutes". In: *The American Economic Review* 81.3 (1991), pp. 658–665. ISSN: 0002-8282.

[9] Stuart Geman and Donald Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (Nov. 1984), pp. 721–741. ISSN: 1939-3539. DOI: 10.1109/TPAMI.1984.4767596.

[10] G. E. Hinton. "Reducing the Dimensionality of Data with Neural Networks". en. In: *Science* 313.5786 (July 2006), pp. 504–507. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1127647.

[11] Geoffrey E. Hinton. "Training Products of Experts by Minimizing Contrastive Divergence". en. In: *Neural Computation* 14.8 (Aug. 2002), pp. 1771–1800. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/089976602760128018.

[12] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. “A Fast Learning Algorithm for Deep Belief Nets”. en. In: *Neural Computation* 18.7 (July 2006), pp. 1527–1554. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.2006.18.7.1527.

[13] Harald Hruschka. “Analyzing market baskets by restricted Boltzmann machines”. en. In: *OR Spectrum* 36.1 (Jan. 2014), pp. 209–228. ISSN: 0171-6468, 1436-6304. DOI: 10.1007/s00291-012-0303-6. URL: http://link.springer.com/10.1007/s00291-012-0303-6 (visited on 05/14/2021).

[14] Harald Hruschka. “Comparing Unsupervised Probabilistic Machine Learning Methods for Market Basket Analysis”. en. In: *Review of Managerial Science* 15.2 (Feb. 2021), pp. 497–527. ISSN: 1863-6691. DOI: 10.1007/s11846-019-00349-0.

[15] Harald Hruschka, Martin Lukanowicz, and Christian Buchta. “Cross-Category Sales Promotion Effects”. en. In: *Journal of Retailing and Consumer Services* 6.2 (Apr. 1999), pp. 99–105. ISSN: 0969-6989. DOI: 10.1016/S0969-6989(98)00026-5.

[16] Instacart. *"The Instacart Online Grocery Shopping Dataset 2017"*. Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on May 3, 2021. 2017.

[17] Puneet Manchanda, Asim Ansari, and Sunil Gupta. “The “Shopping Basket”: A Model for Multicategory Purchase Incidence Decisions”. In: *Marketing Science* 18.2 (May 1999), pp. 95–114. ISSN: 0732-2399. DOI: 10.1287/mksc.18.2.95.

[18] Maryam M. Najafabadi, Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. “Deep Learning Applications and Challenges in Big Data Analytics”. In: *Journal of Big Data* 2.1 (Feb. 2015), p. 1. ISSN: 2196-1115. DOI: 10.1186/s40537-014-0007-7.

[19] Xiao Rong. *Deepnet: Deep Learning Toolkit in R.* Mar. 2014.

[20] Gary Russell, David Bell, Anand Bodapati, Christina Brown, Joengwen Chiang, Gary Gaeth, Sunil Gupta, and Puneet Manchanda. “Perspectives on Multiple Category Choice”. en. In: *Marketing Letters* 8.3 (July 1997), pp. 297–305. ISSN: 1573-059X. DOI: 10.1023/A:1007960529932.

[21] Gary J Russell and Wagner A Kamakura. “Modeling Multiple Category Brand Preference with Household Basket Data”. en. In: *Journal of Retailing* 73.4 (Dec. 1997), pp. 439–461. ISSN: 0022-4359. DOI: 10.1016/S0022-4359(97)90029-4.

[22] Gary J Russell and Ann Petersen. “Analysis of Cross Category Dependence in Market Basket Selection”. en. In: *Journal of Retailing* 76.3 (July 2000), pp. 367–392. ISSN: 0022-4359. DOI: 10.1016/S0022-4359(00)00030-0.

[23] Gary J Russell and Ann Petersen. “Analysis of Cross Category Dependence in Market Basket Selection”. en. In: *Journal of Retailing* 76.3 (July 2000), pp. 367–392. ISSN: 0022-4359. DOI: 10.1016/S0022-4359(00)00030-0.

[24] Ruslan Salakhutdinov and Geoffrey Hinton. “An Efficient Learning Procedure for Deep Boltzmann Machines”. en. In: *Neural Computation* 24.8 (Aug. 2012), pp. 1967–2006. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/NECO_a_00311.

[25] P. B. Seetharaman, Siddhartha Chib, Andrew Ainslie, Peter Boatwright, Tat Chan, Sachin Gupta, Nitin Mehta, Vithala Rao, and Andrei Strijnev. “Models of Multi-Category Choice Behavior”. en. In: *Marketing Letters* 16.3-4 (Dec. 2005), pp. 239–254. ISSN: 0923-0645, 1573-059X. DOI: 10.1007/s11002-005-5888-y.

[26] Paul Smolensky. “Information Processing in Dynamical Systems: Foundations of Harmony Theory”. In: *Parallel Distributed Process* 1 (Jan. 1986).

[27] Michel Wedel and Wagner A. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations.* en. Second. International Series in Quantitative Marketing. Springer US, 2000. ISBN: 978-0-7923-8635-3. DOI: 10.1007/978-1-4615-4651-1.