Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis

International Bachelor Econometrics and Operations Research

The Effect of Attribute Interactions and Demographic Covariates on Modeling Consumer Heterogeneity

Author: Yassine Ben Haddou

Student Number: 523321yh

Supervisor: Prof. dr. D. Fok

Second Assessor: Prof. dr. R. Paap

Abstract

With the rise of technology, marketeers now have more access to data than ever allowing for them to explore novel approaches to renown marketing principles. It is also evident that all customers differ in their needs and desires. Therefore, the aim of this paper is to further investigate a customer heterogeneity modeling technique introduced in Evgeniou, Pontil, and Toubia (2007) called the RR-Het model. This model's framework is rooted in a Ridge Regression with heterogeneous parameters. The basic RR-Het model is further enhanced by adding attribute interactions and demographic covariates. To test these methods, a dataset provided by Lenk, DeSarbo, Green, and Young (1996) is used. However, it is concluded that adding these elements to the model fail to significantly improve it. Attribute interactions decrease estimation accuracy and demographic covariates have little to no effect to the model. Thus, the findings of this research indicate that the basic RR-Het model provides more accurate predictions than more intricate models as available information increases.

May 2022

Contents

1 Introduction

Palmatier and Sridhar (2020) describes customer heterogeneity as the notion that there is variation in the needs and desires of customers; therefore, it is imperative that firms consider this heterogeneity when devising their marketing plan. Failure to do so results in a marketing strategy that is too broad gives the customer a sense of estrangement with the firm and ultimately weakens said firm's sustainable competitive advantage. There are various aspects that need to be addressed in order to tackle the customer heterogeneity problem namely, segmenting, targeting, and positioning. In the most rudimentary form, segmenting is when marketers divide the customers into groups according to certain characteristics, target the most profitable groups, and then position themselves in such a way that it makes informed decisions with these groups in mind.

This paper will focus on the segmenting aspect by investigating a novel method proposed Evgeniou et al. (2007) that will model consumer heterogeneity by means of a conjoint estimation rooted in convex optimization and machine learning based on a ridge regression called RR-Het. The modeling of customer heterogeneity is done by estimating the underlying part-worths that together make up the total worth a consumer gives to a product. In layman's terms, the model will predict to what degree certain characteristics of the product play a role in a consumer choosing this product; for instance, color and price. The dataset for this research is provided by Lenk et al. (1996). A survey was done on first-year MBA students studying at the University of Michigan where they were asked to rate their likelihood (on an 11-point scale, 0 to 10) of purchasing different personal computers. The computers differed in technical attributes such as hard disk size, CPU speed, and more. Due to how closely the attributes are related to each other, it is appropriate to assume that the preference a consumer has for one attribute is linked to how much said consumer prefers another attribute. For example, a consumer could prefer having a big hard disk and high CPU speed. This consumer could greatly prefer having both attributes at the same time rather than either having a big hard disk or high CPU speed individually. Therefore, it is also of interest to investigate the effect of adding interactions to the model on estimation accuracy. This more intricate model will be referred to as iRR-Het.

Similar to how the computers are different so are the students that responded to the survey. Dailey et al. (2006) highlighted the importance of consumer needs on marketing strategy, and more specifically market segmentation. This research is performed on MBA students as well. Accordingly, consumer demographic is added to the RR-Het framework by including covariates within the regression yielding the third model, RR-Het-Cov. Lastly, a fourth model is analyzed called iRR-Het-Cov. This model combines both interactions and covariates. Hess (1994) finds success when implementing interactions and covariates simultaneously. However, the former is performed on a cox-Model compared to the RR-Het framework that is based on ridge regression. Thus, this approach is novel and the literature on the results is sparse. Hence, it remains of interest to further investigate the combination of the elements in this paper. To this end the following research question is investigated:

What is the impact of interactions between attributes and demographic covariates within the RR-Het framework when modeling consumer heterogeneity?

Green and Rao (1971) first proposed the conjoint estimation method and it has become one of the predominant techniques used by marketeers and one that is commonly discussed within the marketing literature. In order to further refine this method, Evgeniou et al. (2007) proposes a method called RR-Het that uses parameters that are set endogenously contrary to exogenously as is often used in convex optimization. This means that the parameters are determined from within the model compared to for instance in hierarchical Bayes, the parameters from the second stage priors are exogenous. This novel method attempts to improve the fit of the model while minimizing the shrinkage by the use of estimated part worths for each respondent. Shrinkage in this context means the amount each individual part-worth is shrunk towards a certain number. This number is usually zero but can also be any other value such as the part-worth mean.

RR-Het is designed by means of a ridge regression where the loss function is tweaked in order to fit the customer heterogeneity setting. The iRR-Het, RR-Het-Cov, and iRR-Het-Cov are created by further adjusting this model by adding extra parameters and feeding more information to the regression. A constraint that all parameters are required to be chosen such that the loss function remains convex allowing for freedom in the choice of the convex optimization method. This paper opts for solving first-order conditions. This approach yields a closed-form solution that provides favorable analytical qualities.

It is concluded that the RR-Het model performs the best. It is the least computationally expensive yet is still able to most accurately predict consumer preferences. Attribute interactions decrease model accuracy and the effect of covariates is negligible when there is sufficient information present.

The remainder of this paper is structured as follows: Section 2 defines the research question in further detail, and describes the data and its corresponding summary statistics. It also contains a detailed description of the methods used in this research, their implementation, and the manner in which they are tested and evaluated. Section 3 considers the results of the research and their implications. Lastly, Section 4 summarises the relevant findings, and presents research limitations and suggestions for further research.

2 Methodology

In this section, the necessary notation is defined allowing for the derivation of the RR-Het model and its implementation to be outlined. Subsequently, the sub-models, iRR-Het, RR-Het-Cov, and iRR-Het-Cov are described in detail. Then, the evaluation of the models is explained. Lastly, this section will be concluded with an explanation of the data used to illustrate the models.

2.1 Notation

Similar notation as described in Evgeniou et al. (2007) is employed in this paper. I represents the set of consumers such that $i \in \{1, 2, \ldots, I\}$. J represents the set of profiles such that $j \in \{1, 2, \ldots, J\}$. There are p partworths for each of the j profiles. X_i is a $j \times p$ design matrix where each row denoted by x_{ij} stores the attribute information. Y_i is a $J \times 1$ vector with all the ratings given by consumer i. z_i is a r-dimensional vector with demographic covariate information where r is the amount of demographic information. Table 4 in the Appendix summarizes the aforementioned notation.

2.2 RR-Het

$$\gamma^* = \arg\min_{\gamma} \text{cross-validation}(\gamma),$$

$$\left(\left\{\mathbf{w}_i^*\right\}, \mathbf{w}_0^*, D^*\right) = \underset{\{\mathbf{w}_i\}, \mathbf{w}_0, \mathrm{D}}{\arg\min} \frac{1}{\gamma^*} \sum_{i=1}^{I} \sum_{j=1}^{I} \left(y_{ij} - \mathbf{x}_{ij}\mathbf{w}_i\right)^2 + \sum_{i=1}^{l} \left(\mathbf{w}_i - \mathbf{w}_0\right)^\top D^{-1} \left(\mathbf{w}_i - \mathbf{w}_0\right),$$

$$\text{subject to } D \text{ is a positive semi definite matrix scaled to have trace 1.} \quad (1)$$

Equation 1 is the main convex optimization problem. It can be observed that this equation has a similar structure to the regular Ridge Regression (RR) formulation. In essence, RR-Het is a RR where certain parameters are adjusted and variables are inserted in order for the formulation to fit the consumer heterogeneity setting. By making these changes, the model shrinks individual part worths towards the population mean and is able to pool information of all the consumers. Intuitively equation 1 does the following. The first part $\left(y_{ij} - \mathbf{x}_{ij}\mathbf{w}_i\right)$ is called the fit. In this part, the model attempts to find a linear relationship between the realized ratings and the respondents underlying preference for different part worths. The second part $\left(\mathbf{w}_i - \mathbf{w}_0\right)^\top D^{-1}\left(\mathbf{w}_i - \mathbf{w}_0\right)$ is the shrinkage part of the model. In the standard ridge regression, the coefficients are shrunk toward zero. In the RR-Het, however, it is shrunk towards the part-worth average with as goal to minimize the error when introducing the model to new data. Lastly, the γ parameter controls the trade-off between the fit and shrinkage. The rest of the sub-section is dedicated to explaining how the variables in equation 1 are optimized.

The estimation of the function is divided into two steps. Firstly, γ needs to be estimated then the following parameters need to be estimated $\left(\{\mathbf{w}_i\}, \mathbf{w}_0, D\right)$. However, the estimation of these parameter requires the entire function to be estimated repetitively until an optimal γ is obtained. This process will be explained in more detail later in the subsection. For now, it is assumed that γ is known.

Another iterative procedure is introduced to find the global optimal solution for the parameters $\left(\{\mathbf{w}_i\}, \mathbf{w}_0, D\right)$. Since the function is jointly convex, various methods can be implemented to find this solution. This paper opts for solving first-order conditions allowing for a well-defined closed solution. The two steps are as follows with the following respective closed-form solution.

1. Solve first order conditions for $\{\mathbf{w}_i\}, \mathbf{w}_0$ given D
2. Given $\{\mathbf{w}_i\}, \mathbf{w}_0$, solve first order conditions for D^*

$$\mathbf{w}_i = \left(X_i^\top X_i + \gamma D^{-1}\right)^{-1} X_i^\top \mathbf{Y}_i + \left(X_i^\top X_i + \gamma D^{-1}\right)^{-1} \gamma D^{-1}\mathbf{w}_0 \tag{2}$$

$$\mathbf{w}_0 = (1/I)\sum_i \mathbf{w}_i \tag{3}$$

$$D = \frac{1}{2\rho}\left(\sum_{i=1}^{1}\left(\mathbf{w}_i - \mathbf{w}_0\right)\left(\mathbf{w}_i - \mathbf{w}_0\right)^\top\right)^{1/2} \tag{4}$$

This two-step procedure will be repeated until there is a convergence of the parameters. Step 1 is done by solving equation 2and equation 3simultaneously. It should be noted that w_0 is simply the sample mean of the part worths. Also for the first iteration, D can be set to a random positive definite matrix to get the process started. In this experiment, D is set to the identity matrix scaled such that the trace is 1. It is possible that the obtained D is not invertible, to circumvent this problem, a pseudo-inverse is calculated. This process is similar to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm where a pseudo inverse is calculated for the hessian as seen in Liu and Nocedal (1989). A possible reason for D to be not invertible is if a row of D is equal to zero. This can happen if for a certain part-worth, all consumers rate it the same meaning heterogeneity is no longer present. This phenomenon goes against the notion that all consumers differ. Therefore, it is also highly unlikely to encounter this situation in practice. Then in Step 2, equation 4 is solved where ρ is chosen such that D has a trace equal to 1. With a trace of one, the problem remains convex.

As stated previously, the aforementioned process is based on the assumption that γ is known. The process to estimate γ will now be outlined. This re-sampling process is done independently of the other parameters because otherwise equation 1 can simply be minimized by setting γ to $-\infty$. Finding an accurate value for γ is highly pertinent as this parameter controls the trade-off between fit and shrinkage. Choose a value that is too high for γ, and there will be excess shrinkage; on the other hand, choose a value too small, there is a chance of over-fitting. This process of finding the optimal γ is called *cross-validation*. *Cross-validation* is an endogenous procedure where apart from the calibration data no other out-of-sample data is necessary. This form of *cross-validation* is called *Leave-one-out cross-validation* (LOOCV). LOOCV is a special case of k-fold cross-validation, where in this setting k is equal to J. When J is sufficiently large, LOOCV becomes computationally expensive however with regard to the current data, LOOCV remains appealing. As outlined in James, Witten, Hastie, and Tibshirani (2013) LOOCV yields favorable properties, contrary to k-fold cross-validation, such as minimal bias, decreased chance of overfitting, and non-stochastic results. Algorithm 1 shows how LOOCV will be implemented in this model. In order to find γ^*, the Algorithm 1is run on a set of potential γ values, and the value that yields the lowest cross-validation is chosen.

Algorithm 1 LOOCV Procedure

Set: $Cross - Validation(\gamma) = 0$.
for $k = 1$ `to` J **do**
  Define $Z^{(-k)} = \bigcup_{i=1}^{I} \left\{ \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i(k-1)}, \mathbf{x}_{i(k+1)}, \ldots, \mathbf{x}_{iJ} \right\}$
  Estimate $\left(\left\{ \mathbf{w}_i^{-k} \right\}, \mathbf{w}_0^{-k}, D^{-k} \right)$ using only $Z^{(-k)}$
  Using $\left\{ \mathbf{w}_i^{-k} \right\}$, compute the left out rating, and set CV(k) to the sum of squared difference of estimated and realised rating.
  set $Cross - Validation(\gamma) = Cross - Validation(\gamma) + CV(k)$
end for

By putting both these steps together, equation 1 is optimized and customer heterogeneity is modeled.

2.3 iRR-Het

When modeling consumer heterogeneity with interactions, equation 1 can also be used. Adding interactions to the model allows for a more holistic overview of how certain part worths interact with each other and the part worths individual relevance to the model. When including all possible pairwise combinations of part worths, the model goes from p to $p + p(p-1)/2$ parameters.

Having $p(p-1)/2$ extra parameters can lead to running time complications. In the situation where there are more parameters than there is data, a rule of thumb is more than $I \times J$ parameters, a dual optimization in combination with kernel theory needs to be implemented as described in Wahba (1990). Since $p = 13$ because the intercept is disregarded, there are 92 parameters, which is less than $190 \times 20 = 3800$. Therefore, the estimation can be performed as described in the previous section where the additional parameters are added to $x_{i,j}$.

In this research, instead of 0 and 1, -1 and 1 are used for dummy variables, respectively. Due to these variables, the interpretation of the interactions can be tricky. Table 1 shows how the interactions should be interpreted. ρ is only positive when both attributes are low or both are high. Thus, it is the additional value a respondent gains from having either both attributes or having neither. In the other cases, when one attribute is low and the other high, ρ is how much value is lost when the attributes are not at the same level. Note that when ρ is negative, the opposite holds.

Table 1: Example portraying how to interpret the interaction within the model

		Attribute 1 (α)	
		Low	High
Attribute 2 (β)	Low	$-\alpha-\beta+\rho$	$\alpha-\beta-\rho$
	High	$-\alpha+\beta-\rho$	$\alpha+\beta+\rho$

2.4 RR-Het-Cov

Contrary to the iRR-Het model, a different function needs to be used in order to incorporate the demographic covariates when estimating the parameters. The variables Θ and z_i as described in Table 4 are added to the function and w_0 is left out. Instead of demeaning the part worths for each respondent with its population mean, $\Theta\mathbf{z}_i$ is subtracted from the part worths. $\mathbf{z}_i$ holds personal demographic information about each respondent which is believed to be able to shrink the model. Therefore, equation 1 then becomes the following

$$\min_{\{\mathbf{w}_i\},\Theta,D} \frac{1}{\gamma} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(y_{ij} - \mathbf{x}_{ij}\mathbf{w}_i\right)^2 + \sum_{i=1}^{I} \left(\mathbf{w}_i - \Theta\mathbf{z}_i\right)^\top D^{-1} \left(\mathbf{w}_i - \Theta\mathbf{z}_i\right)$$

$$\text{subject to } D \text{ is a positive semidefinite matrix scaled to have trace 1} \quad (5)$$

The two-step procedure also becomes a three-step procedure that is as follows and the closed-form solutions become the following.

1. Solve first order conditions for $\{\mathbf{w}_i\}$ given D and Θ
2. Given $\{\mathbf{w}_i\}$ and Θ solve first order conditions for D
3. Given D and $\{\mathbf{w}_i\}$, solve first order conditions for Θ

$$\mathbf{w}_i = \left(X_i^\top X_i + \gamma D^{-1}\right)^{-1} X_i^\top \mathbf{Y}_i + \left(X_i^\top X_i + \gamma D^{-1}\right)^{-1} \gamma D^{-1}\Theta\mathbf{z}_i \quad (6)$$

$$D = \frac{1}{2\rho} \left(\sum_{i=1}^{1} \left(\mathbf{w}_i - \Theta\mathbf{z}_i\right) \left(\mathbf{w}_i - \Theta\mathbf{z}_i\right)^\top \right)^{1/2} \quad (7)$$

$$\text{vec}(\boldsymbol{\Theta}) = \left(ZZ^\top \otimes D^{-1}\right)^{-1} \left(Z \otimes D^{-1}\right) \text{vec}(\mathbf{W}) \quad (8)$$

Equations 2 and 4 become equations 6 and 7 respectively (see Evgeniou et al., 2007). The same restrictions apply to D as described in the previous sub-section, it needs to remain invertible else a pseudo inverse is required. Similarly, as before, D is initially set to the identity matrix with the trace scaled such that it is equal to 1. In this setting, Θ is also set to a random matrix, not necessarily positive definite. Note that in equation8, vec is used to denote the process where a matrix is turned into a vector by stacking each column on top of one another, and $\otimes$ means the Kronecker product operation. The same cross-validation procedure is performed but with the aforementioned equation to find the optimal γ.

2.5 iRR-Het-Cov

In this model, both covariates and interactions are added to the model. These adjustments to the RR-Het model are done by combining the methods described in Sub-section 2.3 and 2.4. Firstly, the $x_{i,j}$ data vector is increased such that it contains the part worths interaction pairs. Then, equation 5, is applied for estimation.

2.6 Evaluation

In this data, there are 20 different computers. From these 20 computers, the rating of only n different computers per customer is known and given as information to the models and the remaining computer ratings are used when testing the prediction accuracy. The computers that are known are randomized for each consumer. n is 8 and 16 leading to both a situation with high sparsity and low sparsity. By setting up the experiment in this manner, a situation is created that is similar to experiments done within the Recommender Systems literature (see Adomavicius & Tuzhilin, 2005). In order to express the model prediction quality, the RMSE is used. The RMSE is calculated for each consumer (190 times) when predicting the rating of the hold-out PCs and then the average of all the RMSEs is presented. Predictions are made by calculating the fit part of the respective model.

2.7 Data

As previously stated, the data is provided by Lenk et al. (1996). A total of 190 students completed the survey. There are 20 different PCs each with slightly different components such that no PC is identical to another. The description of the part worths and demographic covariates can be found in Figures 1 and 2 which can also be found in Table 2 in Lenk et al. (1996). Figure 6 is a correlation heat map of the attributes made from X_{ij} which can be found in the appendix. There is little to no correlation between the attributes, thus, the data set is suitable for this research.

Table 2 MBA Computer Conjoint Analysis

A. Telephone Service Hot Line
- −1 = No
- 1 = Yes

B. Amount of RAM
- −1 = 8 MB
- 1 = 16 MB

C. Screen Size
- −1 = 14 inch
- 1 = 17 inch

D. CPU Speed
- −1 = 50 MHz
- 1 = 100 MHz

E. Hard Disk Size
- −1 = 340 MB
- 1 = 730 MB

F. CD ROM/Multimedia
- −1 = No
- 1 = Yes

G. Cache
- −1 = 128 KB
- −1 = 256 KB

H. Color of Unit
- −1 = Beige
- 1 = Black

I. Availability
- −1 = Mail order only
- 1 = Computer store only

J. Warranty
- −1 = 1 year
- 1 = 3 year

K. Bundled Productivity Software
- −1 = No
- 1 = Yes

L. Money Back Guarantee
- −1 = None
- 1 = Up to 30 days

M. Price
- −1 = $2000
- 1 = $3500

Figure 1: The 13 factors that describe each PC

Subject Level Covariates

Variable	Description	Mean	STD
FEMALE	0 if male and 1 if female	0.27	0.45
YEARS	Years of full-time work experience	4.4	2.4
OWN	1 if own or lease a microcomputer and 0 otherwise	0.88	0.33
TECH	1 if engineer or computer professional 0 otherwise	0.27	0.45
APPLY	Number software applications	4.3	1.6
EXPERT	Sum of two self-evaluations. Each evaluation in on a five-point scale with 1 = Strongly Disagree, 3 = Neutral, and 5 = Strongly Agree. The first evaluation is, "When it comes to *purchasing* a microcomputer, I consider myself pretty knowledgeable about the microcomputer market." The second is, "when it comes to *using* a microcomputer, I consider myself pretty knowledgeable about microcomputers."	7.6	1.9

Figure 2: The six individual-level covariates have been collected for each respondent.

3 Results

In this section, the results of running the different models are presented. The experiment setup is further outlined including choices made when running the models. The estimation accuracy of each model is also described in relation to one another with extra focus on the implication of the selected γ parameter.

The experimental procedure is carried out using MATLAB R2021a with an Intel(R) Core(TM) $i5-9500$ CPU at 3.00 GHz using 16.0 GB RAM, with Windows 10 Enterprise. For all the models and levels of sparsity, γ is chosen from the set $\{0.1, 0.2, \cdots, 1.0\}$. This set is chosen since the LOOCV Procedure is computationally expensive. Running the models with a bigger set is only feasible for the RR-Het and RR-Het-Cov model. The iRR-Het and iRR-Het-Cov models have 92 parameters which are computationally infeasible to be run with a set bigger than the aforementioned set. Thus, for there to be fairness, all models will only be able to choose the optimal γ from the same limited set.

Table 2: The model result with 8 PCs kept for training and 12 for testing

Model	RR-Het	iRR-Het	RR-Het-Cov	iRR-Het-Cov
γ^*	0.7	1.0	0.3	0.3
In-sample RMSE	0.91	0.63	0.55	**0.46**
Out-of-sample RMSE	**1.92**	2.46	1.98	6.24

Table 2 shows the results of the model when $n = 8$. This table presents the results when there is a high level of sparsity. The model tries to predict more ratings than it is given in training. When testing the model on the training data, iRR-Het-Cov performs the best with a RMSE of 0.43. This result is not surprising since this model is provided with most information when training. The RR-Het model performs the best when testing on out-of-sample data with a RMSE of 1.92, followed by RR-Het-Cov with a RMSE of 1.98. The models where interaction among the attributes are introduced perform less well. This phenomenon can be explained by over-fitting. The γ parameter for iRR-Het-Cov is 0.3 with the highest RMSE 6.24. This γ implies that the fit is scaled and given 3.33 times more important than the shrinkage parameter. However, even with this γ, the model still fails to adequately predict the correct ratings. It is also of interest to note that γ is 1.0 for the iRR-Het model, the upper limit of the available γ. After testing this model, on a larger set of γ it is found that the bigger γ is always preferred. This means that in the setting where there is high data sparsity, it becomes increasingly important to control the variables rather than improve the fit.

Table 3: Model result with 16 PCs kept for training and 4 for testing

Model	RR-Het	iRR-Het	RR-Het-Cov	iRR-Het-Cov
γ^*	0.8	0.1	0.8	0.1
In-sample RMSE	1.02	**0.10**	1.00	**0.10**
Out-of-sample RMSE	**1.69**	1.91	**1.69**	1.98

Table 3 shows the results of the model when n is 16. Both the RR-Het and RR-Het-Cov perform the best in this setting with both models obtaining an out-of-sample RMSE of 1.69. The in-sample RMSE is around 1.00 and the optimal γ is 0.8 for both. Similarly, the obtained results for iRR-Het and iRR-Het-Cov are almost identical apart from a slight deviation in the out-of-sample RMSE.

This observation shows that when there is sufficient training data, the influence of covariates decreases and at times iseven negligible. The γ for iRR-Het and iRR-Het-Cov is 0.1. This is the lower limit of the range of the available γ. Testing the models with even smaller values yields that 0.1 is indeed the optimal value for γ. The demographic covariates only enter the model in the shrinkage control and not in the model fit part. Thus, the RR-Het and iRR-Het model are the only two distinct models, and it is obvious that RR-Het outperforms the iRR-Het.

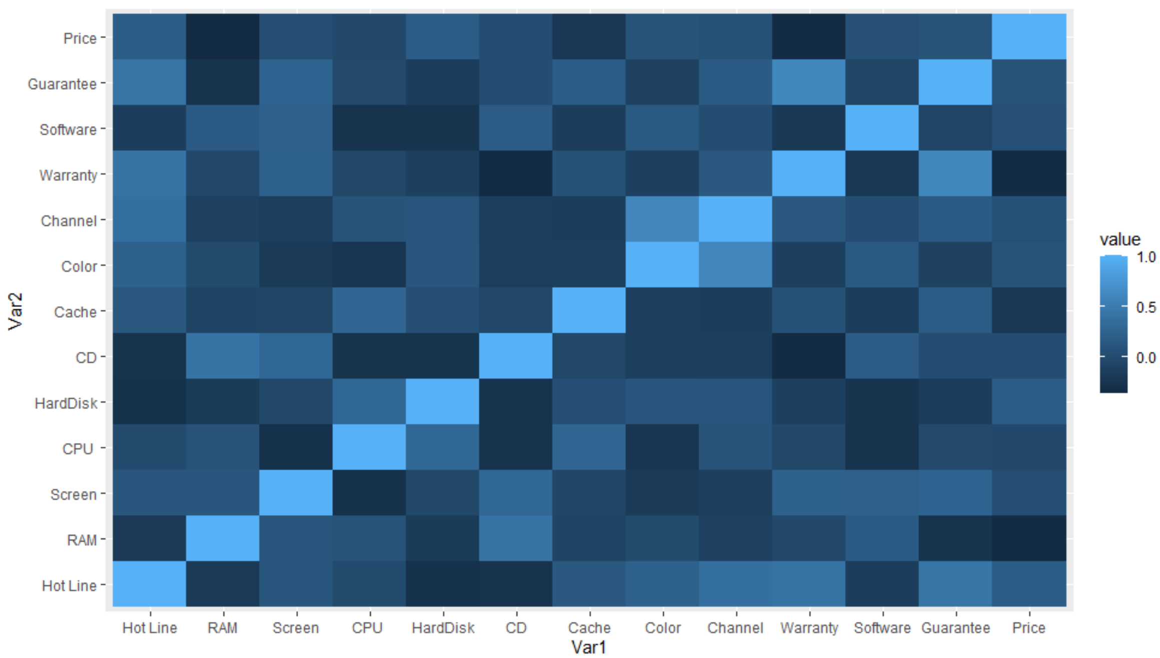

Figure 3: Correlation Matrix of the attribute estimate

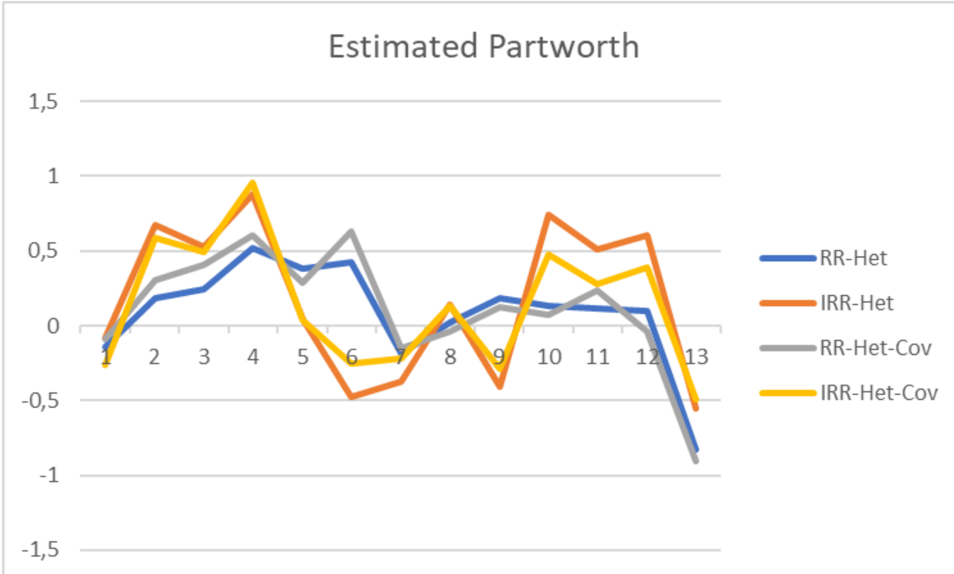

Figure 4: The estimate of the part-worth attributes for the four models where they enter the model in the same order as they are presented in Figure 1

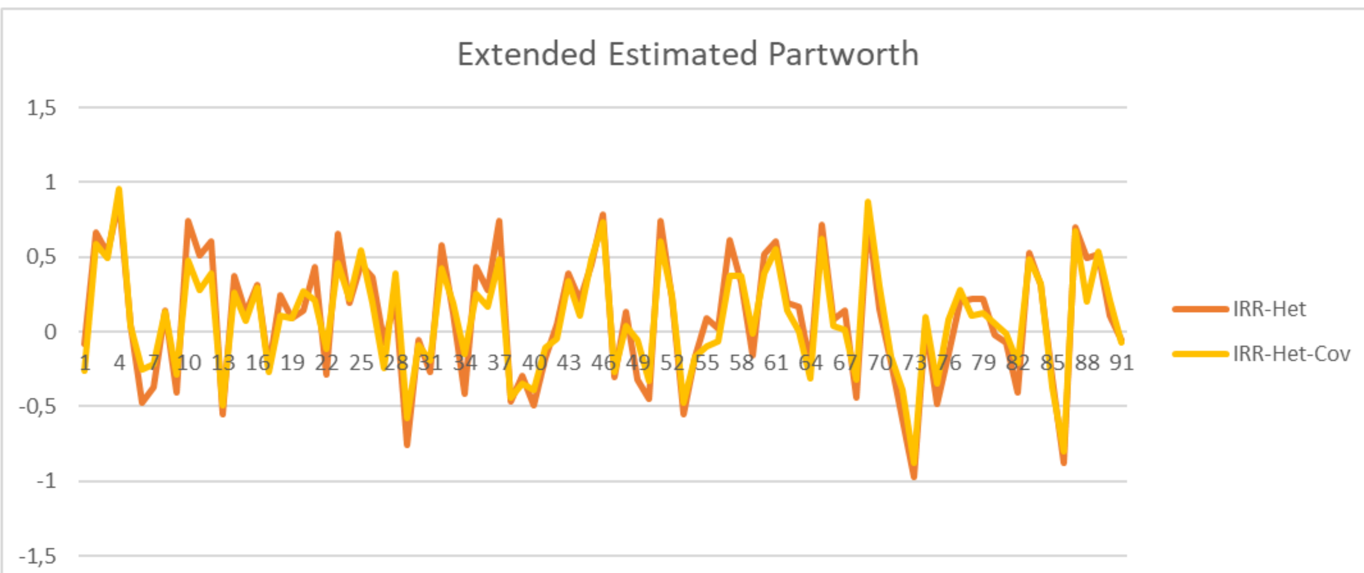

Figure 5: The estimate of the part-worth attributes including all attribute interactions

In order to further inspect the differences between the models, a single consumer is taken and the obtained attribute estimate is investigated. In this example, the first survey respondent is used. Also, it is important to note that due to how for this experiment random 16 PCs are assigned as training data for each consumer, the values obtained will be different each time the code is run. The observed numbers should not be taken as a constant but rather the overall shape of the graph should be looked at. Figure 4 shows the estimated w_i for the first respondent ($i = 1$). Two distinct groups are present in this figure. RR-Het and RR-Het-Cov move together and the iRR-Het and iRR-Het-Cov move together. This observation further demonstrates that including demographic covariates does not significantly influence the estimation result. Moreover, the estimates for iRR-Het and iRR-Het-Cov are more subtle and closer to zero compared to RR-Het and RR-Het-Cov. This result is expected since the part-worth effect can also enter through the interactions in these models. Figure 5 also shows w_1 for the models with interactions. The values in this graph are less subtle and range approximately from -0.5 to 1. This means that the interactions do carry relevant information and attributes affect each other which make sense. For example, CPU Speed goes hand in hand with higher RAM. Figure 6 shows a correlation heat-map of the attribute estimate. It can be seen that the attributes are indeed correlated. For instance, color and channel are highly correlated. This observation is relevant as it gives reason for one to add interactions to the model. However, as described earlier, interactions do not improve the predictive qualities of the model.

Overall, the results are evident. The best model remains the simple RR-Het model. It requires the least information; allowing it to run the fastest compared to the other models. Secondly, the RR-Het shows strong predictive qualities in both the high and low sparsity settings. The iRR-Het model has weaker predictive strength and is computationally expensive to run. However, it is useful to run when attempting to model the underlying effect attributes have on one another. In the dataset used in this paper, these parameters strongly affect one another. This information is highly pertinent and can be used by marketers for instance, but this model should not be used when predicting model rating. The effect of adding covariates to the model is minimal. Demographic covariates do not sufficiently improve the model and does not justify the increase in the necessary computational power. The iRR-Het-Cov model both the issues seen in the iRR-Het and RR-Het-Cov. Combining interactions and demographic covariates do not yield any extra benefits and should be avoided in this setting.

4 Conclusion

This paper investigates the following research question: *What is the impact of interactions between attributes and demographic covariates within the RR-Het framework when modeling consumer het-*

erogeneity?

In order to investigate this question, four distinct models are created and compared on their predictive qualities. The RR-Het model is the simplest model. This model becomes the iRR-Het when introducing attribute interactions and becomes RR-Het-Cov when introducing demographic covariates. Lastly, iRR-Het-Cov is created with as aim to explore whether combining both interaction and covariates will lead to an improved model that combines the theoretical strengths of interactions and covariates.

It is observed that attribute interactions lead to a big increase in the running time of the model with the model also losing predictive accuracy. A similar phenomenon is observed for the RR-Het-Cov model. This model takes longer to run than the RR-Het model yet yields that same predictive accuracy. The iRR-Het-Cov model performs the worst. The theoretical properties fail to uphold when testing on a real dataset. Thus, the conclusion is clear. The basic RR-Het model has the best predictive accuracy and should be used rather than the models with more information.

This paper does have certain limitations. Firstly, a relatively small dataset is used. Each PC has 13 different binary attributes. Thus further research should consider the effect of interactions when there are significantly more attributes. In this case, another optimization method should be used such as a kernel method as introduced Evgeniou et al. (2007)to allow for convergence of parameters. A dataset with more attributes can also be simulated if field data is not available. Secondly, it is assumed that the heterogeneity is uni-modal for all respondents. Thus, future research could consider the effect of clustering the consumers and finding the heterogeneity of these clusters instead of individual consumers. Lastly, the same penalty is applied to the main effect and the interactions. Future research could look at improving the interaction models by defining the main effect in such a way that there is a different penalty for this main effect and the interactions.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, *17*(6), 734–749.

Dailey, L., Anderson, M., Ingenito, C., Duffy, D., Krimm, P., & Thomson, S. (2006). Understanding mba consumer needs and the development of marketing strategy. *Journal of Marketing for Higher Education*, *16*(1), 143–158.

Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, *26*(6), 805–818.

Green, P. E., & Rao, V. R. (1971). Conjoint measurement-for quantifying judgmental data. *Journal of Marketing research*, *8*(3), 355–363.

Hess, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in medicine*, *13*(10), 1045–1062.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, *15*(2), 173–191.

Liu, D. C., & Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, *45*(1), 503–528.

Palmatier, R. W., & Sridhar, S. (2020). *Marketing strategy: Based on first principles and data analytics.* Bloomsbury Publishing.

Wahba, G. (1990). *Spline models for observational data.* SIAM.

A Notation

Table 4: Notation.

Symbol	**Explanation**
I	Set of consumers
J	Set of profiles
p	partworths
D	Scaled weighing matrix
$x_{i,j}$	vector with p columns that represents the rating of consumer i on profile j
X_i	$J \times p$ design matrix of consumer i
w_i	$p \times 1$ column vector of the partworths for respondent i;
Y_i	$J \times 1$ vector with the profile rating of consumer i
z_i	r-dimensional vector of covariates for respondent i
Θ	$p \times r$ matrix of regression coefficients

B Correlation Heat Map

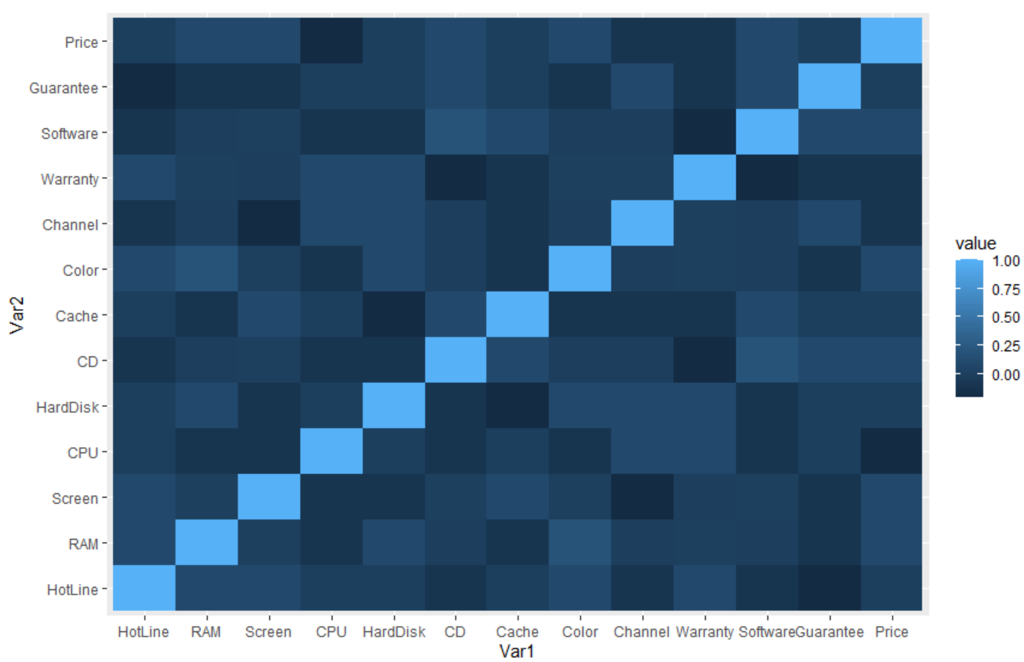

Figure 6: Correlation Matrix of the attributes calculated from X_i

C Code

Main code for the RR-Het and RR-Het-Cov models was provided by the authors of the paper. See footnote 5 in Evgeniou et al. (2007). Code for preparing/sorting the data, and testing the models is written by me