A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 285, 2013

Artur Zaborski[*], *Marcin Pełka*[**]

GEOMETRICAL PRESENTATION OF PREFERENCES BY USING PROFIT ANALYSIS AND R PROGRAM

Abstract. PROFIT is a kind of external vector analysis of preference mapping. It is a combination of multidimensional scaling and multiple regression analysis. PROFIT takes as input both a configuration of stimulus points and a set of preference rankings of the different properties of the stimuli. For stimulus space obtained by multidimensional scaling multiple regression is performed using the coordinates as independent variables and attribute as the dependent variable. The program locates each property as a vector through the configuration of points, so that it indicates the direction over the space in which the property is increasing.

The article presents PROFIT analysis and the R code to carry out the method. The function is illustrated with an example of application in the analysis of consumer preferences.

Key words: multidimensional scaling, property fitting, preference maps.

I. PREFERENCE MAPS

Preference analysis is facilitated by various mapping techniques that provide a visual representation of customer perceptions and preferences. There are two kinds of models for mapping preferences: ideal point model and vector model (see Coxon (1982), p. 223; Kuhfeld (2005)).

The ideal point model attempts to produce a configuration **Y** of points in the space with each point $\mathbf{y}_k$ $(k = 1, \dots, m)$ representing one of *m* judges, together with another configuration **X** of points $\mathbf{x}_i$ $(i = 1, \dots, n)$ in the same space, these points representing choice objects. Individuals are represented as „ideal" points in the multidimensional space, so that the distances from each ideal point to the object points correspond to the preference scores. The ideal point model is used to find a point in a stimulus space which is most like an attribute. If the attribute is a subject's preference for the stimuli, then this point is interpreted as a subject's ideal stimulus. It is the hypothetical stimulus which, if it existed, the subject would prefer most.

In the ideal point model, preferences have the following form (Davison (1984), p. 163):

[*] Ph.D., Chair of Econometrics and Computer Science, Wroclaw Univeristy of Economics.
[**] Ph.D., Chair of Econometrics and Computer Science, Wroclaw Univeristy of Economics.

$$\delta_{ki} = \sum_{a=1}^{r} \left(y_{ka} - x_{ia}\right)^2 + e_k , \tag{1}$$

where:

δ_{ki} – the strength of person *k*' preference $(k = 1, 2, \dots, m)$ for stimuli *i* $(i = 1, 2, \dots, n)$,

y_{ka} – the level along dimension *a* that the subject *k* considers ideal (ideal point coordinate),

x_{ia} – location of stimulus *i* along attribute *a* $(a = 1, 2, \dots, \mathrm{r})$,

e_k – the additive constant unique to subject *k*.

The model assumes that subjects share the same set of reference dimensions but they differ in terms of where their ideal points are located in the space. Each subject has one most preferred point in the space (ideal point) which serves as a reference point to preference objects' scores by comparing their distances from ideal point.

The vector model assumes that the subjects collapse the multidimensional stimulus space into one dimension representing the order of preference. We can use this model when a subject's liking for a stimulus is presumed to increase or decrease linearly along each dimension. The model assumes that the preferences would have the following form (Davison (1984), p. 162):

$$\delta_{ki} = \sum_{a=1}^{r} w_{ka} x_{ia} + e_k , \tag{2}$$

where w_{ka} – linear regression weight,

The vector model represents each subject's preference as a vector directed towards his region of maximum preference. It is a special case of the ideal point model whose ideal points are all infinitely far away from the points representing the choice objects. The projections of the stimulus points onto the vector reproduce the subject's preference values, and a preference ranking is interpreted as the order of the projections of the stimuli points on this line. Moreover, the angle which the vector makes with each dimension can be thought of as representing the salience of that dimension in the preference judgment. Individual differences in preference are expressed by the differing directions which the vectors have in the common space.

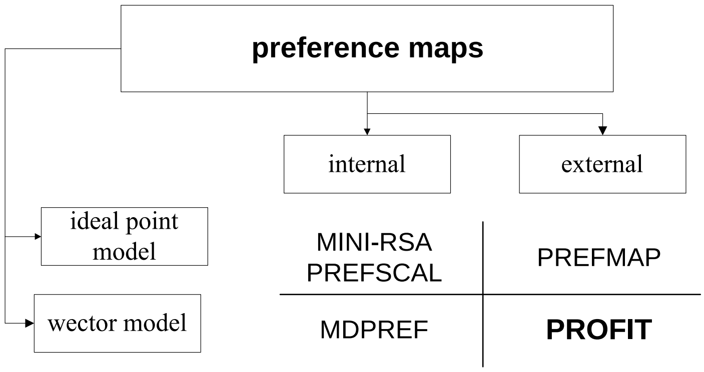

Figure 1. Preference mapping programs

Source: own elaboration.

Ideal points and vectors can be determined in external or internal way (see Figure 1). In external analysis we assume that a similarity configuration of choice object is given. If we have preference data on these objects than external models puts the ideal points or vectors in the space so that it corresponds as much as possible to the preference data. In internal models, both the object configuration and the ideal points or vectors are derived only from preference matrix. We can conceive preference matrix as a submatrix of dissimilarity matrix in which the dissimilarity between objects and between respondents (or attributes) are treated as missing values (see Borg, Groenen (2005), p.335-336).

II. THE CHARACTERISTIC OF PROFIT ANALYSIS

PROFIT (*PROperty FITting*) is a kind of external vector analysis of preference mapping. It is a method of testing hypotheses about the attributes that influence people's judgement of the similarities among a set of items. PROFIT, which is a combination of multidimensional scaling and multiple regression analysis, consist of two phases. In first phase, for *n* objects $O = \{O_1, O_2, \ldots, O_n\}$ and dissimilarities δ_{ij} between all pairs of objects, multidimensional scaling is to find a configuration of *n* points $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^T$ in *r*-dimensional space (*r* is usually 2 or 3) that the distances d_{ij} match, as well as possible, to the dissimilarities δ_{ij}. It means, that $d_{ij} \approx \hat{d}_{ij} = f(\delta_{ij})$, where:

d_{ij} – distance between $\mathbf{x}_i$ and $\mathbf{x}_j$,

$\hat{d}_{ij}$ – monotonic regression of d_{ij} on δ_{ij}.

In the second phase PROFIT takes as input both a configuration of stimulus points $\mathbf{X}$ and a set of attribute preferences data $\mathbf{p}_k = (p_{k1}, \ldots, p_{kn})^T$, where $k = 1, 2, \ldots, m$ is a number of attribute. Then it performs a multiple regression using the coordinates of **X** as independent variables and the attribute as the dependent variable. The program performs a separate regression for each one attribute. The regression coefficients:

$$\mathbf{t}_k = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{p}_k \tag{3}$$

are the coordinates of attribute vector.

For nonlinear regression (see Green, Rao (1972), p. 211) to find the direction of vector $\mathbf{t}_k$ the symmetric matrix $\mathbf{X}^T\mathbf{A}_k\mathbf{X}$ is constructed, The elements of $\mathbf{A}_k$ are defined:

$$\mathbf{A}_k = \begin{cases} -w_{kij} & \text{dla } i \neq j \\ \sum_{j \neq i} w_{kij} & \text{dla } i = j \end{cases}, \tag{4}$$

where $w_{kij} = \dfrac{1}{\left(p_{ki} - p_{kj}\right)^2 + a}$ is a monotonically function of absolute difference between the original ratings of O_i and O_j (*a* is a constant). The vector corresponding to the smallest nonzero characteristic root of $\mathbf{X}^T\mathbf{A}_k\mathbf{X}$ is the attribute vector.

The stimulus ratings on the fitted vector are obtained by projections:

$$\mathbf{h}_k = \mathbf{X}\mathbf{t}_k = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{p}_k\,. \tag{5}$$

III. EMPIRICAL EXAMPLE BY USING R PROGRAM

Consumers were asked to express their preferences for a group of cars on account of eight attributes (suspension, drive, comfort, car body, air conditioning, visibility, electronics, reliability and cost repair) on a 0 to 10 scale, where 0 means no preference and 10 means high preference. Based on the average preference estimations multidimensional scaling analysis was carry out using the following R code:

```
library(MASS)
library(clusterSim)
options(OutDec=",")
library(smacof)
x<-read.csv2("samochody_pref1.csv", header=T, row.names=1)
odl <- dist.GDM(x, method="GDM1")
skalowanie <- smacofSym(delta=odl, ndim=2, metric=TRUE)
x1<-skalowanie$conf
rownames(x1)<-rownames(x)
```

In the second phase a multiple regression for every attribute was made. The coordinates of cars in 2-dimensional space obtained by multidimensional scaling analysis were the independent variables and the attributes were the dependent variable. The following R code was used:

```
y<-x
ile_modeli<-ncol(y)
wyniki<-array(0, c(ile_modeli, 2))
wyniki1<-array(0, c(ile_modeli, 4))
colnames(wyniki)<-c("x1", "x2")
for (i in 1:ile_modeli)
{
   model<-lm(y[,i]~x1[,1]+x1[,2])
   wyn<-summary.lm(model)
   wyniki[i,1]<-model$coefficients[2]
   wyniki[i,2]<-model$coefficients[3]
   wyniki1[i,3]<-summary.lm(model)$r.squared
   wyniki1[i,4]<-summary.lm(model)$adj.r.squared
}
p1<-sqrt(wyniki[,1]^2+wyniki[,2]^2)
nowe_p<-array(0, c(ile_modeli, 2))
for(i in 1:ile_modeli)
{
nowe_p[i,1]<-wyniki[i,1]/p1[i]
nowe_p[i,2]<-wyniki[i,2]/p1[i]
}
```

The plot of cars and the attribute vectors on Figure 2 was obtained using the following R code:

```
xx<-c(-1.5, 1)
yy<-c(-1, 0.5)
```

```
windows()
plot(xx, yy, type="n", las=1, main="", xlab="Dim 1",
ylab="Dim 2")
text(skalowanie$conf, labels=rownames(x))
abline(h=0, v=0)
for (i in 1:ile_modeli)
{
   text(nowe_p[i,1], nowe_p[i,2], labels=colnames(y[i]),
col="navy")
   arrows(0, 0, nowe_p[i,1], nowe_p[i,2], col="navy",
length=0.1)
}
```

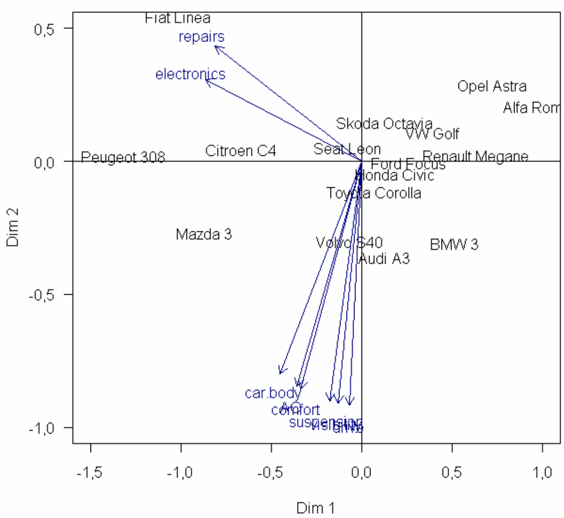

Figure 2. Preference map of cars and attribute vectors

Source: own elaboration.

The distribution of points and vectors in Figure 2 indicates that the evaluation preferences are determined by two groups of variables. Due to the chassis, drive, ride comfort, interior quality, efficiency of ventilation and visibility the most preferred are Mazda 3, Audi A3, BMW 3 and Volvo S40. Due to the reliability, low car repair costs and quality of the electronics top rated are Peugeot 308 and Fiat Linea, and the worst-rated are BMW 3 and Alfa Romeo.

IV. CONCLUSIONS

Presented PROFIT analysis, which is an example of an external vector analysis of preferences mapping may be a practical tool for marketing research. It allows both the identification of preferences from the perspective of specific variables and can be helpful in the interpretation of the dimensions of multidimensional scaling.

PROFIT analysis is a combination of multidimensional scaling and multiple regression analysis and therefore, even if the researcher does not does not have a specialized software, can be successfully implemented with the use of standard analytical tools available in R program.

REFERENCES

Borg I., Groenen P. (2005), *Modern multidimensional scaling. Theory and applications. Second Edition*, Springer-Verlag, New York.

Davison M.L. (1983), *Multidimensional Scaling*, John Wiley and Sons, New York.

Coxon A.P.M. (1982), *The Users Guide to Multidimensional Scaling,* Heinemann Educational Books, London.

Green P.E., Rao V.R. (1972), *Applied Multidimensional Scaling*, Holt, Rinehart and Winston, New York.

Kuhfeld W.F. (2005), *Marketing Research: Uncovering Competitive Advantages*, „Marketing Research Methods in SAS", SAS 9.1 Edition TS-722, 21–34.

Zaborski A. (2008), *Identyfikacja preferencji z wykorzystaniem modeli PREFMAP*, „Prace Naukowe Akademii Ekonomicznej we Wrocławiu" nr 7 (1207), 286–294.

Internet sources: http://opinie.auto.com.pl/

Artur Zaborski, Marcin Pełka

GEOMETRYCZNA PREZENTACJA PREFERENCJI Z WYKORZYSTANIEM ANALIZY PROFIT I PROGRAMU R

PROFIT jest przykładem „zewnętrznej" wektorowej metody map preferencji. Jest ona połączeniem skalowania wielowymiarowego i analizy regresji wielorakiej. Danymi wejściowymi w analizie PROFIT są zarówno współrzędne punktów reprezentujących obiekty na mapie percepcyjnej jak również oceny preferencji obiektów ze względu na wybrane zmienne. Dla konfiguracji punktów reprezentujących obiekty otrzymanej za pomocą skalowania wielowymiarowego przeprowadza się analizę regresji wielorakiej, w której zmiennymi objaśniającymi są współrzędne obiektów na mapie percepcyjnej, a zmiennymi zależnymi oceny marek ze względu na poszczególne cechy. Program dokonuje rozmieszczenia na mapie percepcyjnej zmiennych w postaci wektorów wskazujących kierunek maksymalnej preferencji ze względu na daną zmienną.

Artykuł jest prezentacją analizy PROFIT oraz składni poleceń programu R, pozwalającej na jej realizację. Sposób użycia funkcji zilustrowano przykładem badania preferencji.