Identifying Time Trends in Advertising Expenditure Components: A Simple Regression Approach on Data for 17 European Countries in 1994 to 2007

Katarina Košmelj[1] and Vesna Žabkar[2]

Abstract

The study analyzes the components of advertising spending for a group of European countries with stable total advertising spending over the period 1994-2007. Three components of advertising spending were considered: Electronic, Print, and Online. Our main objective was to study how the components were restructured within the period under study and to find clusters of similar countries. A specific distance for time series, which is expressed as a linear combination of a standard distance measure and the weights, was to be used. The standard distance used for compositional data is the Aitchison distance, however it was found unsuitable due to zero and near zero values in the Online component. As a simple alternative a linear regression model on the components was used, followed by standard cluster analysis on the regression estimates. Results enable a deeper insight into the level of each component and into the structural changes in components for clusters of examined countries.

1   Introduction

1.1   Previous work and motivation for the present work

Košmelj and Žabkar (2008) analyzed the ratio of advertising expenditures (ADSPEND) to gross domestic product (GDP) for 28 European countries during the 1994-2004 time period. Our objective was to reveal different time-trend patterns in the ADSPEND/GDP ratio.

---

[1] Biotehnical Faculty, University of Ljubljana, Slovenia; katarina.kosmelj@bf.uni-lj.si
[2] Faculty of Economics, University of Ljubljana, Slovenia; vesna.zabkar@ef.uni-lj.si

The results showed four clusters of countries with similar trend patterns in the 1994-2004 period: 1) *awakening* countries; 2) *stable* countries; 3) *catching-up* countries; and 4) *leading* countries. Additional data from the Euromonitor database allowed us to prolong the time span for three years up to 2007 and analyze the components of ADSPEND (Euromonitor, 2008).

Advertising expenditure at the country level includes expenses for the following components: press, television, radio, and outdoor. In the last decade, however, a new advertising medium—online—has evolved. This newer medium offers different forms of advertising, including banners, rich media, e-mail campaigns, keyword searches, blogs, and social networks (Chaffey, 2006). All of these media outlets are supported by the Internet. The first country with a reported value for the *online* ADSPEND in the Euromonitor database was Finland in 1996, followed by France, Great Britain, and Sweden in 1997. For some of the countries, *online* advertising expenditures have developed into an important component of ADSPEND. For example, for Sweden, Norway, and Great Britain, its value was around 15 percent in 2007. For some countries, however, the reported values remain very low. The values for Austria, Switzerland, and Portugal are only up to 2 percent (see Table 1).

Our present analysis focuses on the cluster of 17 *stable* countries for the following two reasons:

- In our previous study (Košmelj and Žabkar, 2008), we detected no significant growth in ADSPEND/GDP in this cluster; that is, on average ADSPEND represented about 0.7 percent of GDP. It can be anticipated that no new money was allocated to ADSPEND.

- Online has an important impact in *stable* countries only. For the other clusters, its effect is negligible, even at the end of the observed period.

We define three ADSPEND components: 1) Electronic, which summarizes radio and television; 2) Print, which includes press and outdoor; and 3) Online. For each country, we calculate the proportions (in percentages) reporting the relative magnitude of a particular component of ADSPEND in each of the years studied. The county-year-proportions for the Online component are presented in Table 1. Note that the values at the beginning of the observed period do not exist in the Euromonitor database (Euromonitor, 2008).

Table 1: Proportion (in percent) of the Online component in advertising expenditures by year for 17 European countries during the period 1994-2007. Empty cells indicate that values for Online do not exist in the Euromonitor database (Source: Euromonitor, 2008).

Country	Code	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Austria	AT								0.5	1.2	0.5	1.1	1.3	1.4	1.6
Belgium	BE					0.1	0.4	0.7	0.6	0.6	0.8	1.2	2.3	3.2	4.2
Switz.	CH					0.2	0.3	0.6	0.5	0.5	0.8	1	1.1	1.4	2.0
Germany	DE					0.1	0.4	0.8	1.0	1.4	1.6	1.7	2.0	2.6	3.1
Denmark	DK								3.8	5.4	6.1	6.8	7.6	8.4	9.4
Estonia	EE					0.4	0.6	1.9	2.5	2.5	3.1	2.9	3.5	5.1	6.9
Spain	ES					0.1	0.3	0.9	1.0	1.3	1.4	1.5	1.8	2.2	2.9
Finland	FI			0.1	0.2	0.3	0.6	1.0	1.4	1.4	1.6	2.0	3.0	3.7	4.5
France	FR				0.1	0.2	0.9	1.5	1.1	1.0	1.3	1.6	3.4	4.1	4.5
Gr.Britain	GB				0.1	0.2	0.5	1.3	1.4	1.6	3.5	5.3	9.9	13.7	16.6
Ireland	IE							0.3	0.3	0.4	0.5	0.7	1.1	1.8	2.6
Italy	IT					0.1	0.4	1.7	1.4	1.3	1.3	1.3	1.5	2.1	2.9
Latvia	LV							0.3	0.9	1.2	1.9	1.8	2.5	2.9	3.1
Netherl.	NL						0.6	1.0	0.9	0.9	1.2	1.9	2.7	3.4	4.1
Norway	NO								2.3	1.8	1.9	2.1	2.7	8.2	11.1	14.2
Portugal	PT					0.6	0.5	0.5	0.6	0.6	0.5	0.6	0.5	0.9	1.3
Sweden	SE				0.4	1.3	3.1	5.6	5.5	6.8	7.2	7.7	9.5	11.4	13.1

1.2 Objective of the present work

The objective of the present study is to analyze countries previously clustered as *stable*, taking into account three components of ADSPEND: Electronic, Print, and Online for the period 1994-2007. Our opening objective is to analyze the level of expenditures on each of the three components. Our further aim is to gain deeper insight into how the components were restructured within the period under study. The key research questions are:

- For which countries does Electronic increase on the account of Print?

- For which countries does Print increase on the account of Electronic?

- What is the impact of Online? For which countries is an increase in Online made on the account of Print, on the account of Electronic or on the account of both?

Table 2 presents the compositional data for Denmark. An empty cell in the Online component could be either a structural zero (online ADSPEND did not yet exist); a below-reportable value; or an unreported value to Euromonitor database. We have no information that enables us to distinguish among these three possibilities. Because the year 1994 is considered a starting point for the Internet's commercialization as a marketing and advertising medium, with first advertising

contracts and first e-commerce activities in the previously academic, technical Internet (Cho & Khang, 2006), we shall consider empty cells as zero values.

Table 2: Compositional data for three components (in %) for Denmark.

Component	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Electronic	23.0	23.5	25.2	25.4	25.9	24.9	24.8	24.5	25.1	27.2	28.6	30.4	31.6	32.0
Print	77.0	76.5	74.8	74.6	74.1	75.1	75.2	71.7	69.6	66.8	64.6	62.0	60.0	58.6
Online								3.8	5.4	6.1	6.8	7.6	8.4	9.4

2 Methodology

2.1 Graphical presentation of compositional data

It is common to present three-dimensional compositional data in a ternary graph (simplex), which is a two-dimensional presentation of the plane $x_1 + x_2 + x_3 = 1$, $0 \leq x_1 \leq 1$, $0 \leq x_2 \leq 1$ and $0 \leq x_3 \leq 1$. The ternary graph is an equilateral triangle with a triangular coordinate system. Its vertices correspond to the components, and the values of each component are proportional to the length of the perpendicular segment from the vertex to the opposite side of the triangle. The vertex represents a proportion of 1; the opposite side represents the proportion 0. We include the borders; the simplex is closed (see Figure 1).
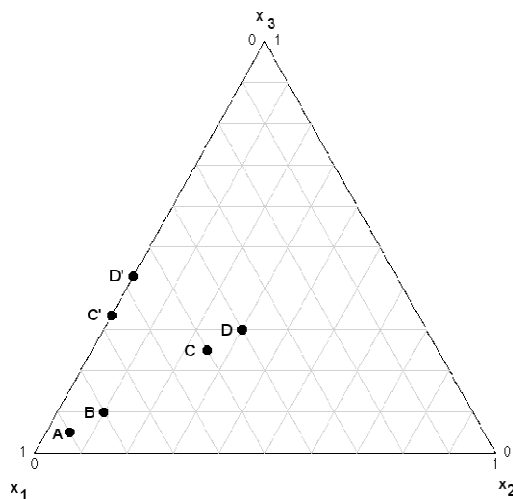

Figure 1: Ternary graph (simplex) and the following points: A(0.9,0.05,0.05), B(0.8,0.1,0.1), C(0.5,0.25,0.25) and D(0.4,0.3,0.3). The point C'(2/3,0,1/3) is a subcomposition of C; D'(4/7,0,3/7) is a subcomposition of D; these two points lie on the simplex border.

2.2 Distance between time series

In this paper, we intended to follow the methodological approach used in Košmelj and Žabkar (2008): cluster analysis and multidimensional scaling. The first step for both approaches is the calculation of a proximity matrix, in our case a dissimilarity matrix between countries, with each being represented by one time series. Standard dissimilarity measures are not appropriate for time series and should be replaced by a measure that takes the time dimension and its ordering property into account. In Košmelj and Žabkar (2008) we present the rationale for the derivation of dissimilarity D between two time series. It takes into account the dissimilarities d_t at successive time points t, $t = 1,...,T$, where d is a standard dissimilarity measure, and the corresponding weights k_t, which assess the impact of an important external characteristic at time points. The weights w_t express the relative importance of the dissimilarities d_t in the calculation of D and incorporate a strong time-ordering condition. To summarize, D can be expressed as the weighted sum of d_t:

$$D = D_T = \sum_{t=1}^{T} w_t \cdot d_t \qquad (2.1)$$

with the weights w_t which are the products of the weights k_t:

$$w_t = \prod_{s=t}^{T-1} k_s, \quad t = 1,...,T-1$$

$$w_T = 1.$$

2.3 Distance between compositional time series

Which distance would be appropriate for d_t in (2.1) for our data? The standard distance used for compositional data is the Aitchison distance (1982, 1986, 2003). The set of positive vectors closed to a closure constant κ (κ is usually 1) is called the simplex of D-part and is denoted as:

$$S^D, S^D = \left\{ \mathbf{x} = [x_1, x_2,...,x_D], x_i > 0; \sum_{i=1}^{D} x_i = \kappa \right\}.$$

The squared Aitchison distance between $\mathbf{x} = [x_1, x_2,...,x_D]$ and $\mathbf{y} = [y_1, y_2,...,y_D]$ is defined as follows:

$$d_a^2(\mathbf{x}, \mathbf{y}) = \frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^{D} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2 = \sum_{i=1}^{D} \left(\ln \frac{x_i}{g(\mathbf{x})} - \ln \frac{y_i}{g(\mathbf{y})} \right)^2, \qquad (2.2)$$

where $g(\mathbf{x}) = \left(\prod_{i=1}^{D} x_i\right)^{1/D}$ is the corresponding geometric mean. This distance satisfies several simple principles important for compositional data analysis (see also Egozcue & Pawlovsky-Glahn, 2006). It should be noted that this distance is suitable only for analyzing positive compositional data. As a result, several approaches based on zero replacement were introduced (e.g., Martin-Fernandez et al., 2000).

Let us illustrate the calculation of the distance D on the trajectories for units X and Y in time, $t = 1,\ldots, 6$, taking into account the Aitchison distance for d_t. Each unit is described by three components. For the illustrative purpose, the change in two successive time points is 0.10 in the first component, on equal account of the other two components; equally for trajectory X and for trajectory Y (Table 3). The last row in the table shows the d_t value. It is evident that the time contribution to D depends on the location of the points within simplex. This effect is particularly evident when a point is very near the simplex border (Table 4).

Table 3: An artificial example of trajectories X and Y with the calculated value of the Aitchison distance d_t, for each time point t.

t	1	2	3	4	5	6
X	0.40,0.30,0.30	0.50,0.25,0.25	0.60,0.20,0.20	0.70,0.15,0.15	0.80,0.10,0.10	0.90,0.05,0.05
Y	0.00,0.50,0.50	0.10,0.45,0.45	0.20,0.40,0.40	0.30,0.35,0.35	0.40,0.30,0.30	0.50,0.25,0.25
d_t	cannot be calculated	1.792	1.463	1.384	1.463	1.794

Table 4: An artificial example of trajectories X and Y with the calculated value of the Aitchison distance, for each time point t. Trajectory X is near the simplex border.

t	1	2	3
X	0.990,0.005,0.005	0.9990,0.0005,0.0005	0.99990,0.09005,0.90005
Y	0.590,0.205,0.205	0.5990,0.2005,0.2005	0.59990,0.20005,0.20005
d_t	3.455	5.312	7.189

To summarize, the use of Aitchison distance d_t in the calculation of the distance D for our data is questionable due to the following reasons:

- Aitchison distance is not defined if any component is zero. In our dataset we have zeros values for the Online component.
- A small time change in a component can have a large impact on the distance D, in particular near the simplex border. In our dataset we have small values and small time changes for Online component.

- The value for Aitchison distance depends on the location of the points within simplex. For example, for the data in Table 3, the same time change in one component on the equal account of the other two components contribute different amount to *D*. To assess the trend, it would be reasonable to expect the same contribution to *D* in each time point.

Therefore we left out this approach and undertook an alternative.

2.4   Regression approach

As an alternative, we used a regression approach. A simple linear regression model is acceptable for all the countries, however for Ireland its use may be questionable (Table 5).

Table 5: Compositional data for three components (in %) for Ireland.

Component	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Electronic	42.6	43.3	42.0	37.7	39.5	39.9	33.5	31.3	27.1	25.1	24.4	27.5	28.1	27.9
Print	57.4	56.7	58.0	62.3	60.5	60.1	66.2	68.3	72.5	74.4	74.9	71.4	70.1	69.6
Online							0.3	0.3	0.4	0.5	0.7	1.1	1.8	2.6

For simplicity reasons, we shifted the time origin to the year 2000, thus the new time variable is *t*, *t* = -6,…,7. For each country, we modelled the Electronic component E and the Print component P as follows:

$$E = \beta_{0E} + \beta_{1E} \cdot t + \varepsilon$$
$$P = \beta_{0P} + \beta_{1P} \cdot t + \varepsilon$$

Similarly, for the Online component:

$$O = \beta_{0O} + \beta_{1O} \cdot t + \varepsilon$$

however only from the time point *t** onwards with the non-zero values for Online component, *t* = *t**,…,7; for example, for Denmark *t**=1, for Ireland *t**=0.

The intercept β_0 is the predicted value in 2000, for Online component, this value can be a meaningless extrapolation. The slope β_1 presents the average year-change (in percent). Results are given in Appendix 1.

3 Results

3.1 Regression

Figure 2 presents the scatterplot for 17 countries in the space of intercept Electronic and intercept Print. These two values present the predicted value for the year 2000 according to the linear regression model. The plot shows that for the majority of countries the predicted valued for Print is higher than for Electronic. The most outstanding example among *Print dominant* countries is Switzerland (Print more than 80 percent), among *Electronic dominant* are Italy and Portugal (Electronic around 60 percent).
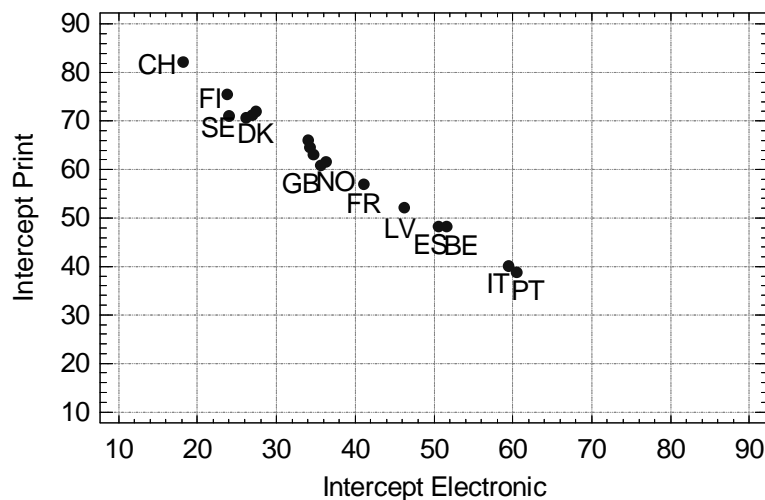

Figure 2: Scatterplot for 17 countries in the space of intercept Print and intercept Electronic. These two values are the predicted values for the year 2000 according to the linear regression model. Some countries are labelled.

Graphical presentation of the two slopes is very informative, too (Figure 3). The negative slope for Print is notable nearly for all countries, the slope for Electronic is mainly positive; hence the majority of countries are *Pro Electronic*. The countries are situated in three quadrants. In the second quadrant we have two countries; Ireland has the highest increase of Print (1.37) and the highest decrease in Electronic (-1.53). In the third quadrant we find four countries with negative slope Print and slope Electronic near zero (Estonia, Finland, Great Britain, and Norway). In the fourth quadrant we find the remaining 11 countries with positive

slope Electronic and negative slope Print. The extreme values are for Switzerland (1.23 Electronic, -1.37 Print). Four countries (Great Britain, Norway, Sweden and Denmark) are considerably below the line $y = -x$ suggesting the highest impact of the Online component. For Great Britain and Norway, Online increases on the account of Print decreasing, however Electronic is nearly constant (Figure 4). For Sweden and Denmark, Print decreases on the account of Electronic and Online decreasing (Figure 5).
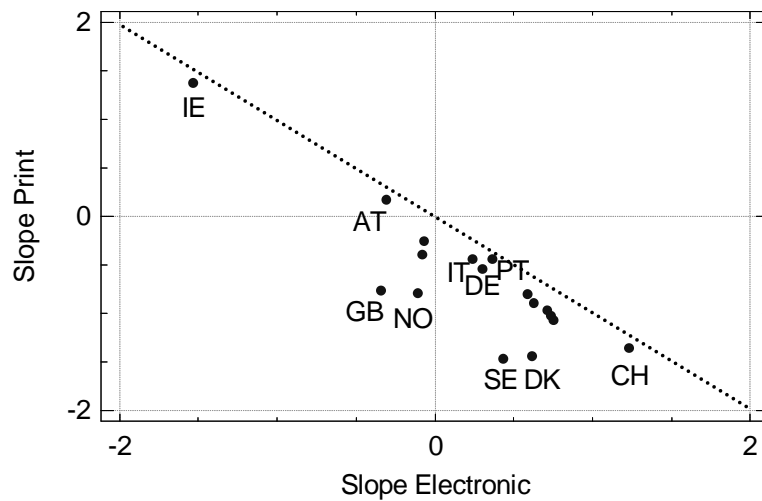

Figure 3: Scatterplot for 17 stable countries in the space of slope Print and slope Electronic. Some countries are labelled.
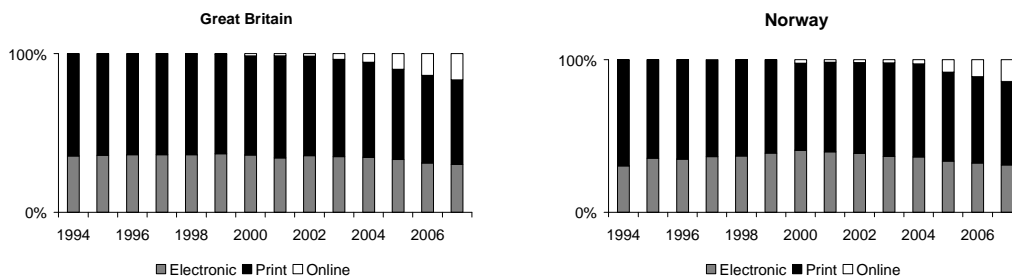

Figure 4: Components for Great Britain and Norway. Increase in Online on the account of Print decreasing.
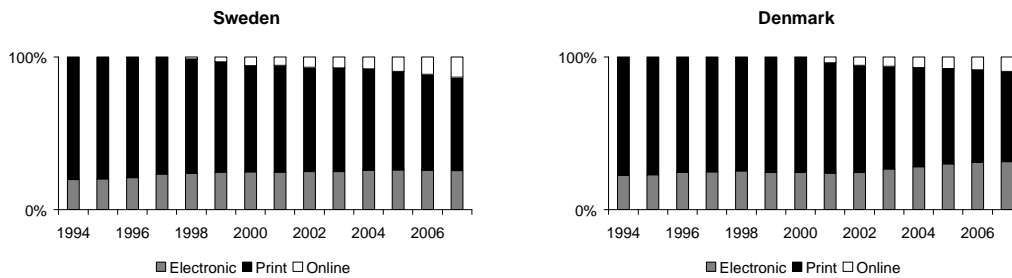
Figure 5: Components for Sweden and Denmark. Increase in Online and Electronic on the account of Print decreasing.

3.2 Clustering

Finally, we undertook cluster analysis on six variables for each country: intercept Electronic, intercept Print, intercept Online, slope Electronic, slope Print, and slope Online (see Appendix 1). We used squared Euclidean distance on standardized variables. The dendrogram for Ward's method is presented in Figure 6.
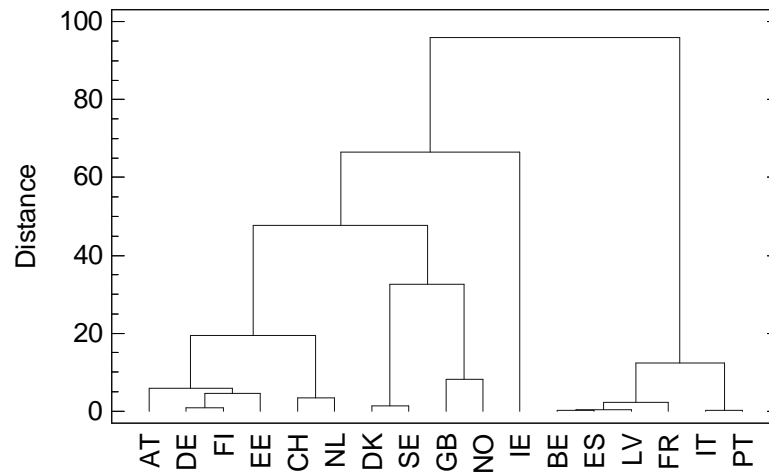
Figure 6: Dendrogram obtained by Ward's method. Variables were standardized, squared Euclidean distance was used.

The results can be summarized as follows: there are two main clusters, however in the first Ireland is separated. As the linear model for Ireland is questionable, we choose three main clusters: the first includes ten countries, the second six, in the last is Ireland only. Table 6 presents the centroids of the obtained clusters. According to the situation in 2000, the first cluster contains *Print dominant countries* (Print around 70%), the second *Electronic dominant countries* (Electronic around 52%). Print is decreasing with a similar rate in the two clusters (0.8 % per year on average). In both clusters Electronic and Online increase, however with a different pace. In the second cluster the effect of Electronic is higher than the effect of Online, however in the first cluster it is the opposite. The subcluster of cluster 1 (Denmark, Sweden, Great Britain and Norway) has an important Online effect (see Table 7).

Table 6: Centroids of the obtained clusters.

Cluster	n	Countries	Name of the cluster	Intercept Electronic	Intercept Print	Intercept Online	Slope Electronic	Slope Print	Slope Online
1	10	Switzerland, Netherlands, Germany, Finland, Austria, Estonia Denmark, Sweden, Great Britain, Norway	*Print dominant*	28.8	**69.3**	1.3	0.2	**-0.8**	0.7
2	6	Belgium, Spain, Latvia, France, Italy, Portugal	*Electronic dominant*	**51.6**	47.5	0.6	0.5	**-0.8**	0.3
3	1	Ireland*		34.3	65.2	-0.1	-1.5	1.4	0.3

*results of the linear regression model are questionable

4   Conclusions

The objective of this study was to analyze the three components of advertising spending (Electronic, Print, Online) for the period 1994-2007 for the selected countries in the terms of their level and trend. The components were analyzed for the *stable* countries because it could be assumed that the spending allocated to advertising is only transferred from one medium to another.

Our first attempt was to use the same methodology as for the ratio ADSPEND/GDP, however using the Aitchison distance in the definition of the distance between two compositional time series. This approach failed. As a simple alternative, we used a simple linear regression model and standard cluster analysis on the estimates obtained in the regression procedure. Results are satisfactory; they enable a deeper insight into the level of each component and into the structural changes in components for these countries.

Our further attempt is focused on a distance for compositional time series that will assess the trend in the components and allow for zero and near zero values. A promising starting point may be a half-taxi metric presented in Miller (2002).

References

[1]   Aitchison, J. (1982): The statistical analysis of compositional data. *Journal of the Royal Statistical Society*. Series B (Methodological), **44**, 139-177.

[2]   Aitchison, J. (1986): *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

[3]   Aitchison, J. (2003): *The Statistical Analysis of Compositional Data*. New Jersey: The Blackburn Press.

[4]   Chaffey, D. (2006): *Internet Marketing: Strategy, Implementation and Practice*. Prentice Hall.

[5]   Cho, C.H. and Khang, H.K. (2006): The State of Internet-related research in communications, marketing, and advertising: 1994-2003. *Journal of Advertising*, **35**, 143-163.

[6]   Egozcue, J.J., and Pawlowsky-Glahn, V. (2006): Simplicial geometry for compositional data. In Antonella Buccianti, G. Mateu-Figueras, Vera Pawlowsky-Glahn (Eds.): *Compositional Data Analysis in the Geosciences: From Theory to Practice, Geological society*, London(265), 145-159.

[7]   Euromonitor (2008): World marketing data and statistics. www.euromonitor.com/womdas

[8]   Košmelj, K., and Žabkar, V. (2008): A methodology for identifying time-trend patterns: An application to the advertising expenditure of 28 European countries in the 1994-2004 Period. *Metodološki zvezki,* 5, 161-171.

[9]   Martín-Fernández, J.A., Barcelo-Vidal, C., and Pawlowsky-Glahn, V. (2000): Zero replacement in compositional data sets. In Kiers, H., Rasson, J., Groenen, P., and Shader, M., (Eds.): *Studies in Classification, Data Analysis and Knowledge Organisation*. Proceedings of 7th Conference of the International Federation of Classification Societies p. 155-160. Berlin: Springer-Verlag.

[10] Miller, W.E. (2002): Revisiting the geometry of a ternary diagram with the half-taxi metric. *Mathematical Geology*, **34**, 275-290.

Appendix 1

Table 7: Intercept Electronic, intercept Print and intercept Onlin; slope Electronic, slope Print and slope Online. Values obtained with the linear regression model.

Country	Code	Int_El	Int_Print	Int_Online	Slope_El	Slope_Print	Slope_Online
Austria	AT	34.14	65.39	0.44	-0.31	0.17	0.16
Belgium	BE	51.69	47.44	0.40	0.71	-0.98	0.40
Switzerland	CH	18.20	81.28	0.41	1.23	-1.37	0.17
Germany	DE	27.43	71.65	0.70	0.31	-0.55	0.31
Denmark	DK	26.27	70.78	3.31	0.62	-1.45	0.87
Estonia	EE	34.81	63.32	1.43	-0.08	-0.40	0.60
Spain	ES	50.67	48.48	0.66	0.59	-0.81	0.27
Finland	FI	23.80	74.95	1.09	-0.06	-0.26	0.38
France	FR	41.22	57.56	0.96	0.75	-1.08	0.41
Gr. Britain	GB	35.74	61.01	1.74	-0.34	-0.77	1.59
Ireland	IE	34.34	65.19	-0.12	-1.53	1.37	0.31
Italy	IT	59.48	39.62	0.86	0.24	-0.44	0.22
Latvia	LV	46.30	52.80	0.43	0.63	-0.90	0.40
Netherlands	NL	26.99	71.98	0.57	0.74	-1.03	0.43
Norway	NO	36.42	60.88	-0.68	-0.11	-0.80	1.78
Portugal	PT	60.52	39.06	0.52	0.37	-0.44	0.05
Sweden	SE	24.09	71.34	4.16	0.44	-1.47	1.17