FLOWS IN AUDIENCE MARKETS

M. Silvia Farina - Alis, S.R.L

Introduction

An interesting way to study,in the course of time, readership variations, affecting editorial and marketing results, could be the observation of movements of readers measured on socio-demographic classes, inside a specific audience market.
Often there is no correspondence between variations in total audience size and internal switching of readers . In fact we can have increase of total audience and loss of loyalty or , on the contrary, we can keep or increase loyalty and obtain lower or stable figures for total audience.
The view angle of the present paper is to observe and model readers switching along the time, inside a specific market, using socio-demographic strata as repeated measures of the same phenomena.
The hypothesis below is that strong modifications in reading behaviour are in general diffused in homogeneous way inside cells, and often, when they are not, it's possible to find homogeneous behaviour if one excludes special 'influent cells' , where movements are incoherent with total sample.
The cases described here come from published data of Italian national readership survey
(AUDIPRESS).
The paper includes the study of three different markets, where reading variables are measured on selected socio-demographic cells. Reading variables measure audience size for disjoint classes of readers, for example exclusive and duplicated, on competitor titles, chosen inside the same market. On these variables we apply multivariate analysis tools to look for the existence of significant internal relations .
The model implementation is substantially composed by two sections :

- audience variables choice and construction
- tailoring and validation of analysis results.

The first section is a very important part of the task, in which we work for validation of data, checking basic constraints :

- variables consistency
- cells comparability
- minimum sample requirements.

From the general building process of variables, direct and indirect results emerge concerning also static composition of markets and internal relations of variables.
In this paper we concentrate our efforts in showing results on dynamic flows, improving concomitant aspects of analysis such as :

- audience data coherence
- optimal number of cells
- influence analysis.

Objectives

This study suggests a different way to go through the data for improving the knowledge of audience variations, focusing internal switching of readers due to occasional events or cyclic phenomena. The evidences we search for often don't take place or the results don't explain the variables completely , but we always find consistent indicators about one part of the phenomena.
In any case our aim is to isolate and recognize contributions of variables to specific audience penetration , inside markets where we voluntary stress comparison between competitors. Some times the absence of clear contributions give us a strong signal to move in depth with analysis, eliminating elements that can disturb analysis or concentrating our study on specific sub-samples.

Methodology

1. Statistical background

The problem of determining fluxes between aggregate phenomena is not new to the statistical community. As early as 1950 W.S. Robinson [1] pointed out the difficulty of interpreting correlations between aggregates as correlations between individuals. This has been referenced as ecological fallacy and stems from the fact that to have meaningful results at individual level one should consider paired samples in a controlled experiment. In this way any individual yields a paired information like in a pre - post analysis.

The researcher, then, can count individual by individual the number of positive and negative answers to questions of categorical nature and determine exactly the number of individuals moving among categories.
In marketing research, on the opposite, we have almost always independent samples. The usual case is to have two independent measures at time 1 and 2 of some categorical variable.

The difference in the response can come from a difference in the level of the phenomenon within categories and from the difference in the composition of sample within categories as well. The wide experience gained in Italy by the analysis of electoral fluxes [2] is in fact a good basis for verifying the applicability of a regression technique to the analysis of fluxes among readers of press titles (dailies and magazines).

This approach, sometimes referenced not quite properly as Goodman model, has been used to estimate the proportion of voters moving among political parties.
The electoral case can be viewed as a lucky case from one side and an unlucky from the other. In fact it is based on data available at the level of electoral sections which are almost constant in composition. But the estimate of the proportion of people moving among parties is done by the method OLS (Ordinary least squares). This means that the proportion of people fluxing is assumed to be constant among sections. This is not likely to be true for the electoral case. Only at local level this can be partly supported.

2. Readership case

In readership there is good confidence that a magazine restyling or some specific marketing campaign will affect readers in a common way among strata and we can take account of fluxes between aggregate phenomena at various sample levels. In the newspaper case we think that the local character of many items will restrict the analysis only to local or regional level. But anyway we expect to measure a constancy of response among some set of people strata.

If we make the hypothesis that the composition of strata remains constant along the time, then we can identify the proportion of people fluxing with the regression coefficients which bind the people categories at time 2 to the categories at time 1.

The interpretation of regression coefficients as flux proportions has to be correctly understood. Because we are at aggregate level, an increase in the readership of a title at the expense of the decrease of another one can come from two sources.

The first is the number of readers actually moving between the titles; the second is a flux between readers and non readers. This flux can have a balance which cannot be distinguished from direct moving of readers between titles. In the extreme case we could even imagine that readers lost by one title stop reading at all while new readers of another title are all coming from the non-readers segment. To avoid this risk, we consider always, in our models, the non readers class.

As you can see from supplied examples, the most common situation is a direct moving between readers classes because the composition of the non readers segment is very stable and comes back on itself. So we can conclude that in the analysis of fluxes in the press market we can identify regression coefficients with proportions of people moving among titles.

A test of the applicability of our approach can be done by simply not imposing conditions of balance of proportions. If the hypothesis of common regression coefficient among strata is true, coefficients will sum to 1 column by column. If this sum significantly deviates from 1 there is room for an hypothesis of a stratum-dependent regression coefficient and the need to check through the influence analysis, which stratum to withdraw from the analysis. In any case a sample reduction suggested by statistical applications or by the interest for a specific analysis, could show new relations significantly different from the ones obtained on the general sample.

Information Support

3. Environment

Since 1990 the Italian readership survey , more recently sponsored by Audipress, takes place two times per year with two waves each working for 12 weeks on a sample that consists at present of 18.000 people for each wave. For each contiguous couples of waves, Audipress publish cumulated results on readership and delivers files containing data for four titles categories : weeklies,monthlies,newspapers and supplements of newspapers.

Since 1990 results on different titles categories lie in the same data base and they are supplied to users at the individual level. In the last 5 years the survey sample design has kept the same scheme and changes took place only on sample size. The sample design respects , inside small geographic areas, socio-demographic strata identified by sex and town size. The age control comes indirectly from the interviewers choice, made on guided selection of names from electoral lists .

Because of the presence of over-sampling, (through which the representativeness of special local areas is reinforced with more interviews than necessary by a Census proportional process), a weighting procedure provides, at the individual level, the adjustment of total sample to Census figures.

Thanks to the fact we are in the presence of a stable context, we have at our disposal useful aspects as

- availability for five different history periods
- elementary strata homogeneity along the time
- stability of market references
- common audience definitions
- 'big' samples and well represented small cells.

4. Cautions for data use

As we said in the previous paragraph, we take a big advantage from the great stability of readership data collection,concerning.above all

- categories and titles covered
- audience definitions
- cells controlled by sample design and weighting procedures.

Nevertheless special attention in model building is required because of :

a- reference market definition
b- elementary cell definition
c- coupling of elementary cells belonging to different periods.

At point a- we emphasise the need to identify the context in which to work, creating disjoint classes of people focused on reading behaviour. Normally we use to proceed in this way

-choosing one or more titles (3/4) to be controlled separately
-split readers classes into exclusive readers and duplicated readers among chosen titles
- accounting for readers of other titles
- accounting for non-readers.

Point b- concerns requirements about the minimum number of individuals belonging to the same cell, without reducing too much the total number of cells. The lower limit we generally set is 25 cases per cell , but normally more than 95% of cells have more than 30 cases.

A good balance must be achieved between

- total number of cells
- number of reading variables investigated.

From our experience when you work with less than 10 reading variables, the number of cells must be about 100 .

With point c- we ensure, for homologous cells , the presence of two different observations along the time. In all cases where we have no correspondent cells, we decide to put together two or more cells.

Finally we would prefer to use disjoint versions of the national readership survey, to avoid the presence of overlapped samples, due to waves contributing to temporally contiguous reports.by
This problem would not take place if we could work by single waves separately and not by cumulated ones.
For the moment this is impossible.

Data Applications

5. Case construction

Each case study we manage with 'fluxes procedure' starts with the **definition of a reference market and at least two observation periods.** Supplied examples analyse only 2 periods, but it's possible and sometimes interesting to consider time series of 3 or 4 periods and measure fluxes between contiguous and extreme points. The reference market is a'reading environment' coherent for two or more periods, in which we define some reading variables.

Taking as example case study nr.1 , we are in the newspapers market of an important Italian region (Lombardia) and we examine following titles :

- Corriere della Sera
- Il Giornale
- Il Giorno
- La Repubblica.

With these titles we build the following new variables , where the reader is the 'average day' reader:

- var.1 : exclusive readers of Corriere della Sera
- var.2 : exclusive readers of Il Giornale
- var.3 : exclusive readers of Il Giorno
- var.4 : exclusive readers of La Repubblica
- var.5 : duplicated readers of one of 4 titles
- var.6 : other Audipress newspaper readers not included in previous groups
- var.7 : non-readers of newspapers.

It's important that variable definition guarantees that **readers classes are disjoint** . Furthermore , to avoid problems connected with changes in the size of reference universe, **we work using penetration** percentages instead of absolute values. This choice allows us on one side to make comparable evaluations referred to different points in time and on the other side to manage cells figures independently from their size.

The same variables are evaluated for the first period (oldest,time 1) and for the second period(most recent,time 2). We use 'X' as prefix for old-period and 'Y' as prefix for new variables. **The periods considered are two different editions of the AUDIPRESS survey (93/II,94/II).** Great care is used to perform homogeneous reading variables, excluding titles not present in both survey editions. The model uses 'Y' variables as dependent variables and 'X' variables as independent ones.

For **cells construction** we use following socio-demographic characters :

- sex
- age (4/5 classes)
- town size
- provinces.

The objective was to build about 100 socio-demographic cells. The examples show results obtained in an 80 cells study and in a 96 cells study. The cells must be 'well ' represented and available for two different editions of the **audipress** survey .

6. Statistical tools

The statistical package we use for present study is SAS® , but also other statistical packages allowing multivariate regression can be used for the estimation of parameters. The model specifies time 2 readers as the dependent vector Y1,...Yh and time 1 readers as the explanatory vector X1,...Xk. The procedure estimates a matrix of coefficients bij, where the columns bear the meaning of proportion of spreading from time 1 categories to time 2 categories. From this matrix a table can be derived to estimate the proportion of a time 2 category coming from a time 1 category.

We outline some important features of the statistical package we use, which should be very much appreciated in this exercise. The first feature is to add a linear constraint to the parameters, forcing a balance condition between proportions. The effect of this is to bind coefficients by row. The second feature is the comparison between coefficients and the test if, column by column, coefficients sum to 1. This is an interesting check , because regression coefficients are also the proportions conveyed from a time 1 category to any time 2 category and they tell us where readers have gone to.

Note that in general coefficients will not sum to 1 by row. In order to know where readers at time 2 are coming from one should first use regression coefficients to get the table of proportions conveyed on the total of readers.

Finally to check the presence of influent cells, we found very useful the 'influence analysis' , able to outline, through statistical indicators, cases (i.e. cells) that disturb the calculation of parameters.

7. Results

After a thorough screening of indicators and tests supplied by statistical analysis, we build for each case study two tables. The first table represents 'FLUX table' , namely the table of relationships between old figures and new figures, where we look, as a quality control check, for a total spread of old phenomena (sum of column values). The second table is the 'COMPOSITION table' where we measure the proportion of new (most recent) penetration figures that we are able to explain through the older variables.

From the examination of 'FLUX table' we emphasise relative movements rather than their size. The observation of 'COMPOSITION table' explores the size of contributions.

From FLUX table we observe :

- var.1 : exclusive readers of Corriere della Sera give contribution to themselves, to var.4 (Repubblica), to var.5 (Duplicated)
- var.2 : exclusive readers of Il Giornale give contribution to var.3 (Il Giorno)
- var.3 : exclusive readers of Il Giorno give contribution to themselves, to var5 (Duplicated).
- var.4 : exclusive readers of La Repubblica give contribution to themselves, to var.6 (Other titles)
- var.5 : duplicated readers of one of 4 titles give contribution to themselves
- var.6 : other newspaper readers not included in previous groups give contribution substantially to themselves
- var.7 : non-readers of newspapers. give contributions substantially to themselves.

The reading **variables with strongest relations with themselves** are var.6 (other readers) and var.7 (non-readers), showing almost null switching with other variables. This is reasonable because of cohesion of classes that var.6 and var.7 represent : i.e. readers of 'local' newspapers and 'non readers of 'AUDIPRESS newspapers'.

Exclusive readers loyalty seems to be strongest for Corriere della Sera. **No loyalty** appears for exclusive readers of Il Giornale , coherently with a change of editorial policy (different political views) that probably affected readers composition . **Internal flows of readers** occur between exclusive readers of Corriere della sera and duplicated readers of the 4 titles considered.

We can improve our knowledge of the phenomenon considering **the absolute balance of the switches, using the 'COMPOSITION table'**. The COMPOSITION table gives us the size of the contributions of old variables to the new ones.
We read , in a different way, the same results as in 'Flux table'.

8. General comments on examples

Cases n. 1,2,3 refer to a newspapers market where we study the same variables over the same time periods. Comparison between case n.1 and case n.2 (reader definition "last 7 days") shows the strength of relations found among audience variables independently from the audience definition used. Comparison between case n.1 and case n.3 shows how we can improve results by increasing the number of elementary cells (=strata) from 80 to 96.

Case n.4 and n.5 refer to a newsweekly market where we study same variables over two different pairs of time periods. Comparison between cases shows how the relationships can change between one title and itself or between one title and other competing titles.

Case n.6 and n.7 refer to a women weeklies market where we study the same variables over the same time periods. Comparison between cases shows how we can improve the statistical quality of results using influence analysis.

9. Cautions

Some cautions must be taken in using results and for improving the process.
In the examples we suggest ways to overcome specific difficulties found during the examination of parameters' estimates. The most frequent difficulties you can find concern:

- presence of 'influent' strata (i.e. dependence of effects among strata)
- poor (unsatisfactory) compliance with the balance equation, which requires the exact distribution of previous classes (time 1 classes) .

10. Limits

The limits existing in the study concern mostly the availability of data which fulfill these conditions:

- observations available for more than 1 period
- same or comparable socio-demographic structure along the time
- same audience definitions
- large enough sample size
- possibility to make markets comparable .

The above requirements are not so common for ad-hoc surveys, but they could be easily found inside national mediasurveys on press or other media.

Finally we don't consider as a limit the possibility that we do not obtain significant results for some case studies, because, following good statistical principles, we take anyway advantage from knowledge of 'not visible effects'. This does not exclude the presence of flows that we can investigate in other ways, building other reading variables, redefining reference markets or restricting the analysis to selected sub-samples.

11. Conclusion on a lighter note

I cannot say whether it was because of these analyses we have just seen, but you may be interested to know that many Italian Publishers for the past year or so have been running all kinds of promotional schemes to capture customers' loyalties and assure repeat purchases. Actually, newstands in Italy nowadays are full of promotional gadgets given free together with publications. For instance, newspapers like Corriere della Sera and Repubblica have been offering free daily instalments of a Geographical Atlas, then of a Universal Pictorial Dictionary: and after 40 weeks or so you can have it bound in special hard covers.

On the part of newmagazines, videocassettes are offered as gifts. And women do not need to go to perfumery shop any more because every issue of woman weekly contains either a bottle of perfume, or a face cream, a pait of fashion earrings....an endless flow of appealing temptations.

And, being in Italy, the last (but not the ultimate, for sure) gimmick has been - guess it! - half a chilo of spaghetti wrapped together with magazine. So you even save a trip to supermatket!

So, boring numbers and tedious analyses may have generated or supported the marketing policies of publishers in the battle engaged to secure reader' precious loyalty.

References.

[1] - Robinson, W.S. (1950). Ecological correlations and the behaviour of the individuals, American Sociological Review, 15, n.3, pp. 351-357.
[2] - P. Corbetta, A.M.L. Parisi, H. Schadee (1988). Elezioni in Italia. Il Mulino - Bologna.
[3] - SAS® STAT User's Guide. SAS Institute Inc., Cary, NC, USA.
[4] - R.D. Cook & S. Weisberg (1992). Residuals and influence in regression. Chapman and Hall - London.

Examples

***Case n. 1**: Six readers classes, built from 'average day' readers of 4 newspapers. Sample size: 80 cells.*

case scheme

a. Model: independent variables XM1-XM7 (time 1)
dependent variables YM1-YM7 (time 2)

b. Time
1. Audipress 93/II
2. Audipress 94/II

c. Sample
1. Italian region (Lombardia)

d. Cell number: 80

e. Character cell definition: age, town size, province

f. Audience measure: 'average day'

g. Reading variables
1. Exclusive Readers of 'Corriere della Sera' (EXCL COR)
2. Exclusive Readers of 'Il Giornale' (EXCL GNL)
3. Exclusive Readers of 'Il Giorno' (EXCL GNR)
4. Exclusive Readers of 'La Repubblica' (EXCL REP)
5. Duplicated Readers of one of previous 4 titles (DUPL)
6. Readers of other Audipress Newspapers not included in previous classes (OTHER)
7. Non readers of Audipress Newspapers (NON READER)

h. Total penetration

reading variables	1 Excl COR	2 Excl GNL	3 Excl GNR	4 Excl REP	5 DUPL	6 Other	7 Non readers
time 1	15.2	2.1	5.5	2.6	5.2	15.5	53.9
time 2	16.8	2.2	5.2	2.5	4.5	15.1	53.7

Model:

Dependent Variable: YM1

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.627507	0.09963914	6.298	0.0001
XM2	1	0.505064	0.35992540	1.403	0.1648
XM3	1	-0.079536	0.17518210	-0.454	0.6512
XM4	1	0.012250	0.25828790	0.047	0.9623
XM5	1	0.642527	0.27105619	2.370	0.0204
XM6	1	-0.001248	0.04676474	-0.027	0.9788
XM7	1	0.036682	0.03149210	1.165	0.2479

Dependent Variable: YM2

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.044571	0.02858038	1.560	0.1232
XM2	1	0.118322	0.10324060	1.146	0.2555
XM3	1	-0.046577	0.05024904	-0.927	0.3570
XM4	1	0.105239	0.07408701	1.420	0.1597
XM5	1	-0.027774	0.07774945	-0.357	0.7220
XM6	1	-0.000138	0.01341395	-0.010	0.9918
XM7	1	0.027752	0.00903316	3.072	0.0030

Dependent Variable: YM3

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.047940	0.03896100	1.230	0.2225
XM2	1	0.244001	0.14073841	1.734	0.0872
XM3	1	0.392687	0.06849989	5.733	0.0001
XM4	1	-0.270654	0.10099601	-2.680	0.0091
XM5	1	-0.103852	0.10598868	-0.980	0.3304
XM6	1	-0.038729	0.01828600	-2.118	0.0376
XM7	1	0.062534	0.01231407	5.078	0.0001

Dependent Variable: YM4

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.074203	0.02916924	2.544	0.0131
XM2	1	0.172909	0.10536771	1.641	0.1051
XM3	1	-0.064516	0.05128434	-1.258	0.2124
XM4	1	0.406667	0.07561346	5.378	0.0001
XM5	1	0.031570	0.07935137	0.398	0.6919
XM6	1	0.031447	0.01369032	2.297	0.0245
XM7	1	-0.004101	0.00921927	-0.445	0.6577

Dependent Variable: YM5

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.289745	0.04673612	6.200	0.0001
XM2	1	0.254706	0.16882437	1.509	0.1357
XM3	1	0.282771	0.08216983	3.441	0.0010
XM4	1	0.133800	0.12115092	1.104	0.2730
XM5	1	-0.182436	0.12713994	-1.435	0.1556
XM6	1	0.018764	0.02193518	0.855	0.3951
XM7	1	-0.031208	0.01477149	-2.113	0.0380

Dependent Variable: YM6

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	-0.141494	0.10195915	-1.388	0.1694
XM2	1	-0.001562	0.36830593	-0.004	0.9966
XM3	1	0.054418	0.17926106	0.304	0.7623
XM4	1	0.791564	0.26430190	2.995	0.0037
XM5	1	0.279062	0.27736749	1.006	0.3177
XM6	1	0.921814	0.04785362	19.263	0.0001
XM7	1	0.001389	0.03222536	0.043	0.9657

Dependent Variable: YM7

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
XM1	1	0.057528	0.15324190	0.375	0.7084
XM2	1	-0.293441	0.55355404	-0.530	0.5976
XM3	1	0.460752	0.26942462	1.710	0.0915
XM4	1	-0.178865	0.39723874	-0.450	0.6538
XM5	1	0.360902	0.41687598	0.866	0.3895
XM6	1	0.068091	0.07192272	0.947	0.3469
XM7	1	0.906953	0.04843387	18.726	0.0001

We can rearrange all this in a table where Y1,...Y7 are rows and X1,..., X7 are columns. We will call this the **flux table** because it gives the proportions of distribution to new reading choices from previous ones. Numbers by column are the percentages of destination of previous readers to the newspaper of the corresponding row.

We give only significant figures, based on T value and associated probability.

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	0.63				0.64		
YM2							0.02
YM3		0.24	0.39	-0.27			0.06
YM4	0.07			0.40		0.03	
YM5	0.29		0.28				
YM6				0.79		0.92	
YM7							0.90
	-----	-----	-----	-----	-----	-----	-----
	0.99	0.24	0.67	0.92	0.64	0.95	0.98

We can see that only XM1, XM6 and XM7 have a good compliance with the required balance equation which states that any previous reader should have a new destination either in a new newspaper or in the non readers class. Note also the negative migration coefficient which means that strata do not comply with the hypothesis of independence of the effect among strata. Applying values of flux table to the original penetration values, we calculate the 'composition table' where each cell represents the penetration contribution, on percentage, to the total of the row.

For all situations in which we have 'excess balance', we proportion results to sum 1.00 along columns. We apply the same procedure to row results. So each row represents the part of readers' penetration explained by the model, split in different contributions expressed trough composition percentages.

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	57.1%				19.6%		
YM2							50.0%
YM3		8.6%	36.2%				55.2%
YM4	34.4%			50.0%		15.6%	
YM5	74.6%		33.3%				
YM6				13.9%		31.8%	
YM7							48.5%

To improve the understanding of the method, we go to the analysis of last seven days readers.

***Case n. 2**: six readers classes, built from 'last seven days' readers of 4 newspapers. Sample size: 80 cells.*

case scheme
Same situation as for case n.1, except for following notes:

a. Model: independent variables XS1-XS7 (time 1)
dependent variables YS1-YS7 (time 2)

f. Audience measure: 'last seven days'

h. Total penetration

reading variables	**1 Excl COR**	**2 Excl GNL**	**3 Excl GNR**	**4 Excl REP**	**5 DUPL**	**6 Other**	**7 Non readers**
time 1	19.8	2.3	7.3	2.7	22.5	19.3	26.1
time 2	22.6	2.0	7.4	2.9	20.3	18.4	26.4

Building again the flux table we get:

	XS1	XS2	XS3	XS4	XS5	XS6	XS7
YS1	0.62				0.33		0.08
YS2						0.03	
YS3	0.09	0.23	0.50				0.09
YS4		0.41		0.38		0.04	
YS5	0.31		0.35		0.43		
YS6	-0.17		0.36	0.40		0.86	
YS7	0.20						0.84
	-----	-----	-----	-----	-----	-----	-----
	1.05	0.64	1.21	0.78	0.76	0.93	1.01.

The composition table is:

	XS1	XS2	XS3	XS4	XS5	XS6	XS7
YS1	54.4%				32.7%		9.3%
YS2						30.0%	
YS3	23.7%	6.3%	37.7%				28.9%
YS4		31.0%		34.5%		27.6%	
YS5	30.0%		10.8%		47.8%		
YS6			11.1%	5.0%		83.8%	
YS7	15.2%						83.0%

Comments.

We can see a better interpretability of the full set. Altough 'last seven days' audience is a weak reading definition if compared with 'average day' audience, from the flux table we find confirmation of strong relationships, seen in case n. 1, between variables :

YS1 vs XS1, XS5
YS3 vs XS2, XS3
YS5 vs XS1, XS3
YS6 vs XS4, XS6
YS7 vs XS7.

A further improvement of case n. 1 could be reached by using a different stratification scheme in order to increase the number of strata (cells). A different stratification scheme was used in order to have a sample of 96 strata.

***Case n. 3**: same as case n. 1 but sample size 96 cells.*

case scheme
Same situation as for case n.1, except for the following :

d. Cell number: 96

The flux table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	0.71				0.50		
YM2		0.24		0.21			0.01
YM3			0.49				0.03
YM4	0.08	0.20		0.19		0.03	
YM5	0.28						
YM6				0.56		0.91	
YM7							0.90
	-----	-----	-----	-----	-----	-----	-----
	1.07	0.44	0.49	0.96	0.50	0.94	0.94

The composition table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	64.3%				15.5%		
YM2		22.7%		22.7%			22.7%
YM3			51.9%				30.8%
YM4	46.0%	15.0%		19.0%		20.0%	
YM5	95.6%						
YM6				9.6%		90.4%	
YM7							90.3%

Comments.

We can see that only few flux coefficients are different from the corresponding ones of case n. 1. Negative values have disappeared and we find a better fit with the balance condition.

***Case n. 4**: Six readers classes built from 'last week' readers of 4 newsweeklies. Sample size :360 cells*

case scheme

a. Model: independent variables XM1-XM7 (time 1)
dependent variables YM1-YM7 (time 2)

b. Time
1. Audipress 93/II
2. Audipress 94/II

c. Sample: Total

d. Cell number: 360

e. Character cell definition: sex, age, town size, region

f. Audience measure: 'last week' readers

g. Reading variables
1. Exclusive Readers of 'Epoca' (EXCL EP)
2. Exclusive Readers of 'L'Espresso' (EXCL ES)
3. Exclusive Readers of 'L'Europeo' (EXCL EU)
4. Exclusive Readers of 'Panorama' (EXCL PA)
5. Duplicated Readers of one of previous 4 titles (DUPL)
6. Readers of other Audipress weeklies not included in previous classes (OTHER W.)
7. Non readers of Audipress weeklies (NON READER)

h. Total penetration

reading variables	1 Excl EP	2 Excl ES	3 Excl EU	4 Excl PA	5 DUPL	6 Other w.	7 Non readers
time 1	2.5	3.1	0.7	5.5	4.4	48.5	35.3
time 2	2.0	3.8	0.6	5.0	3.9	49.2	35.5

The flux table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	0.09	0.07			0.07	0.02	0.01
YM2		0.49	0.26	0.14	0.11	0.01	
YM3		0.08					0.01
YM4				0.30	0.14	0.04	
YM5		0.30	0.71	0.18	0.28		0.02
YM6	0.54			0.35		0.87	0.08
YM7					0.23	0.05	0.88
	-----	-----	-----	-----	-----	-----	-----
	0.63	0.94	0.97	0.97	0.83	0.99	1.00

The composition table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	11.0%	10.5%			14.8%	46.2%	16.7%
YM2		40.0%	4.7%	20.5%	12.6%	12.9%	
YM3		41.7%					58.3%
YM4				33.0%	12.4%	38.8%	
YM5		21.1%	11.4%	22.5%	28%		16.1%
YM6	2.7%			3.9%		85.8%	5.7%
YM7					2.8%	6.8%	87.6%

Comments.

Note that we have no negative flux, that the balance condition is good and that numbers are higly significant. We learn the lesson that sample size should be large enough to avoid the influence of some particular case on the estimate of parameters.

***Case n. 5**: Same as case n. 4 but different time periods.*

case scheme
Same situation as for case n.4, except for following notes:

b. Time

1. Audipress 92/II
2. Audipress 93/II

h. Total penetration

reading variables	**1 Excl EP**	**2 Excl ES**	**3 Excl EU**	**4 Excl PA**	**5 DUPL**	**6 Other w.**	**7 Non readers**
time 1	1.7	3.5	0.8	5.1	3.9	48.3	36.7
time 2	2.5	3.1	0.7	5.5	4.4	48.5	35.3

The flux table is

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	0.15	0.14		0.05		0.03	
YM2		0.21		0.14	0.17	0.02	
YM3				0.03	0.09		
YM4		0.36	0.55	0.28		0.05	
YM5		0.33		0.18	0.40	0.02	
YM6	0.88			0.31		0.91	
YM7				0.11			1.05
	--------	--------	--------	--------	--------	--------	--------
	1.03	1.04	0.55	1.10	0.66	1.03	1.05

The composition table is

	XM1	XM2	XM3	XM4	XM5	XM6	XM7
YM1	11.0%	20.0%		11.0%		58.0%	
YM2		22.6%		19.4%	22.6%	32.3%	
YM3				28.6%	57.1%		
YM4		23.6%	7.3%	21.8%		41.8%	
YM5		26.0%		17.2%	35.0%	22.0%	
YM6	3.1%			2.9%		87.8%	
YM7				13.4%			86.6%

Comments.

Among significant facts we observe that the loyalty of var.2 (exclusive readers of 'Espresso') is better in case 4 compared to the earlier period of case 5, in fact the flux factor YM2-XM2 increases from 0.21 to 0.49. At the same time we observe that the flux from var.2 (exclusive readers of 'Espresso') to var.4 (exclusive readers of 'Panorama') is present only for the past (case 5) and not more for the present (case 4). On the contrary the flux from var.4 (exclusive readers of 'Panorama') to var.2 (exclusive readers of 'Espresso') continues from the past (case 5) to the present (case 4) and it is stable (0.14). Also the loyalty of var.4 (exclusive readers of 'Panorama') is substantially stable (from 0.28 for the earlier case 5 to 0.30 for the more recent case 4).

From the previous analysis the marketing conclusions are very different depending on the title you consider. For Panorama it's necessary not only to improve the loyalty of exclusive readers, but also to stop the flux of exclusive readers of Panorama to Espresso exclusive readers. For Espresso the situation seems satisfactory.

Case n. 6: *seven readers classes, built from 'last week' readers of 5 women weeklies. Sample size: 108 cells.*

Five women weekly magazine and three additional classes are examined in two adjacent time periods. The additional classes account for: readers of more than one title among the in five selected, readers of titles other then the five selected, non readers.

case scheme

a. Model: independent variables XM1-XM8 (time 1)
dependent variables YM1-YM8 (time 2)

b. Time
1. Audipress 93/II
2. Audipress 94/II

c. Sample
Women

d. Cell number: 108

e. Character cell definition: age, town size, region

f. Audience measure: 'last week'

g. Reading variables
1. Exclusive Readers of 'Amica' (EXCL AM)
2. Exclusive Readers of 'Anna' (EXCL AN)
3. Exclusive Readers of 'Gioia' (EXCL GI)
4. Exclusive Readers of 'Grazia' (EXCL GR)
5. Exclusive Readers of 'Donna Moderna'(EXCL DM)
6. Duplicated Readers of two or more of previous 5 titles (DUPL)
7. Readers of other Audipress weeklies not included in previous classes (OTHER W.)
8. Non readers of Audipress weeklies (NON READERS)

h. Total penetration

reading variables	1 Excl AM	2 Excl AN	3 Excl GI	4 Excl GR	5 Excl DM	6 DUPL	7 Other w.	8 Non readers
time 1	1.5	2.0	2.9	3.5	7.2	9.2	40.5	33.2
time 2	1.7	1.9	2.5	3.6	7.2	8.9	41.3	32.9

The flux table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7	XM8
YM1					0.10		0.02	
YM2			0.12					0.01
YM3				0.17		0.08		
YM4			0.24				0.06	
YM5				0.53	0.29		0.06	
YM6		0.32				0.39	0.15	
YM7	0.62				0.28	0.40	0.67	0.16
YM8								0.90
	----	----	----	----	----	----	----	----
	0.62	0.32	0.36	0.70	0.67	0.87	0.96	1.07

The composition table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7	XM8
YM1					41.2%		47.1%	
YM2			15.8%					15.8%
YM3				24.0%		28.0%		
YM4			19.4%				66.7%	
YM5				26.4%	29.2%		33.3%	
YM6						37.1%	62.9%	
YM7	2.2%	1.5%			4.8%	8.9%	65.6%	12.8%
YM8								90.1%

Comments.

We can identify only partly the flux among individual magazines, while there is a good balance for readers of multiple titles both in the group and not in the group (var.6 and var.7). Non readers are strongly consistent between times. Using analysis of 'influential observation' (observations are the strata in our context), we can decide which observations can be removed. From a list of few indicators, where we privilege deviations of residuals and Cook's D indicator, we decide to eliminate the most influential observations (in total 3). Cook's D indicator is a multivariate measure of the change in regression coefficient coming from the removal of the current observation.

Case n. 7 *. Same as case n. 6, but model refit with cleaned data.*

The flux table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7	XM8
YM1						0.05	0.04	
YM2						0.08	0.02	0.01
YM3				0.27	0.10			
YM4			0.17			0.09	0.07	
YM5		0.33			0.30	0.15	0.06	
YM6		0.30				0.37	0.15	
YM7	0.53			0.48	0.24	0.35	0.67	0.14
YM8			0.49		0.31			0.90
	-----	-----	-----	-----	-----	-----	-----	-----
	0.53	0.63	0.66	0.75	0.95	1.09	1.01	1.05

The composition table is:

	XM1	XM2	XM3	XM4	XM5	XM6	XM7	XM8
YM1						20.0%	80.0%	
YM2						31.6%	52.6%	15.8%
YM3				36.0%	28.0%			
YM4			13.9%			19.4%	19.4%	
YM5		9.7%			29.2%	18.1%	31.9%	
YM6		6.3%				32.6%	61.1%	
YM7	1.9%			4.1%	4.1%	7.0%	63.1%	11.1%
YM8			4.2%		6.6%			89.3%

Comments.

From the flux table we see a better balance. We might also exclude new influential observations and iterate the process. The marketing analysis of results shows:

- it's not present any evidence for loyalty of exclusive readers of titles considered, coherently with a market characterized by multiple interest of readers;
- there is a clear switch between exclusive readers of Grazia and Gioia; this is a signal for responsables of both titles to check the common characteristics of their products;
- the class of exclusive readers of Anna is a 'risk class' because we see a flux to exclusive readers of 'Donna Moderna' and to 'Duplicated readers'.