# A Strategy to Interpret Brand Switching Data with a Special Model for Loyal Buyer Entries

***B. G. Mirkin***

## Introduction

The brand switching data (see, for example, Zufryden (1986), Colombo and Morrison (1989), DeSarbo, Manrai and Burke (1990)) can be considered as a special case of contingency tables and processed with known tools like correspondence or log–linear analysis. But the direct use of the methods seems to be improper because, first, the data table contains large main diagonal proportions of "loyal" purchasers, and, second, the row and column items correspond to the same brands. A two–stage strategy is suggested here to overcome each of the problems: 1) using a homogenization model for the diagonal entries "correction"; 2) correspondence–wise clustering of the modified table to reveal the main "switching" flows. The method is illustrated by an example from Zufryden (1986), and, then, is applied to the real data sets.

## Relative Increment of Probability (RIP) Concept and Correspondence Analysis

Correspondence analysis methodology, as it was developed by French researchers in data analysis (see, for example Lebart, and Fénelon (1971), or later a English version (Lebart, Morineau and Warwick (1984)) has proved its efficiency in marketing as well as in more general areas of social research. This is a tool to visualize and to analyze interrelations between items presented by rows or/and columns of a contingency data table having frequencies of pair row–column combinations as its entries.

In Table 1 below an illustrative example of switching data is presented (from Zufryden (1986)): the row and the column items are the same (four basic brand types, 1, 2, 3 and 4, while "brand" 5 indicates the consumers not purchasing the kind of products under consideration). Each entry of the table is the number of consumers who have purchased the column–brand having previously bought the row–brand. Two segments, A and B, of the market are considered.

**Table 1**

| SegmentA | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 11410 | 630 | 434 | 28 | 1484 |
| 2 | 230 | 4035 | 95 | 55 | 560 |
| 3 | 322 | 70 | 217 | 49 | 721 |
| 4 | 390 | 120 | 135 | 3960 | 395 |
| 5 | 1242 | 207 | 621 | 483 | 66447 |

| SegmentB | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 8190 | 550 | 400 | 10 | 840 |
| 2 | 210 | 4050 | 175 | 5 | 560 |
| 3 | 279 | 144 | 7722 | 9 | 846 |
| 4 | 162 | 84 | 56 | 1594 | 102 |
| 5 | 740 | 296 | 814 | 148 | 72002 |

The correspondence analysis visualisation of the homogenized Table 1 data is shown in Figure 1. Such a picture usually shows each of the brands with two points: one for the row (previous purchasing, star on Figure 1) and second for the column (new purchasing, circle on figure 1). Due to this copying effect, it is not easy sometimes to interpret the picture correctly.

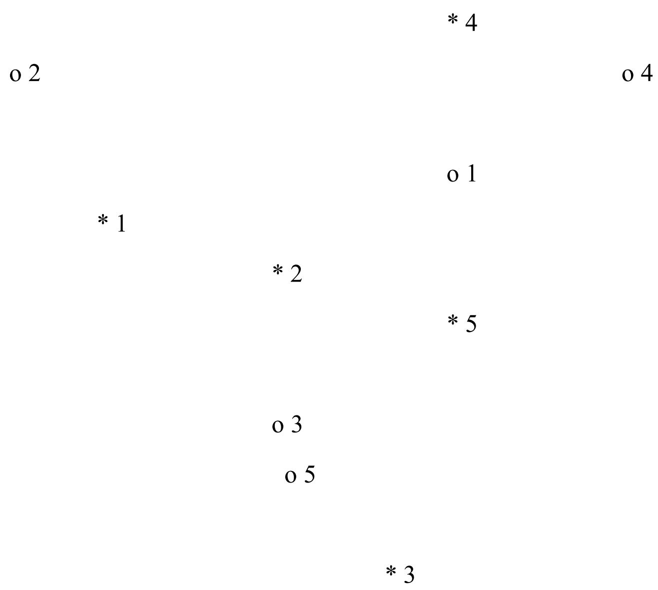
Usually, the correspondence analysis method descriptions are based on so called conditional frequencies (or, probabilities) as a tool to show associations between the row and column items. If, for example, in a sample of 1000 car buyers, 40 buyers have bought car A having bought B before, and 100 of the buyers had cars of brand B previously, we can say that P(A/B) = 40/100 = 40% is conditional frequency (or even probability) of buying A by condition B, because 40% of the previous B–owners have chosen A as their next purchase. In these terms correspondence analysis represents interrelations between the row–profiles consisting of the conditional frequencies of the column–items under so-called chi–square distance (Lebart et Fénelon (1971)). But there exists another characteristic of the co–occurrence data — relative increment of probability (RIP), which is shown to be much more important (Mirkin,1992).

To introduce the RIP concept, let us ask ourselves: this figure, P(A/B) = 40%, is it many or few? The answer cannot be given based only on this. It is necessary to compare this figure with the percentage of car A buyers. If, for example, 800 buyers from the sample purchase A, the percentage of A buyers equals P(A) = 800/1000 = 80%, and P(A/B) is half P(A)! In this case, we could say, the B–owners avoid switching to A. If, however, the number of A buyers equals 100, that is P(A) = 100/1000=10%, we can say that the ratio P(A/B)/P(A) equals 40/10=4 and, so, the B–owners purchase car A four times more often than an average buyer. To express this formally, the Relative Increment of Probability coefficient could be considered:

$$RIP(A/B) = (P(A/B)-P(A))/P(A)$$

It equals, in the first example, (40–80)/80 = -50% and, in the second example, (40–10)/10 = 300%: in the first case the probability (frequency) of A purchasing is 50% decreased for B–owners compared with the average level, and in the second case, this probability is 300% increased compared to the average level.

**Figure 1: The correspondence analysis plane representing interrelations between row items (stars) and column items (circles) by Table 1 (Segment A)**



The examples of this kind show explicitly that the RIP value reflects association between B and A better than the conditional probability P(A/B) does. From the other point, an important property of the RIP is its symmetry (on the contrary, P(A/B), usually, is asymmetric!). In spite of the explicit direction of dependency taken into account here (A–purchasing is considered here as dependent on B–ownership) the following equality holds: RIP(A/B) = RIP(B/A) which could be interpreted as a shortcoming of the RIP value because it does not concern any direction of the influences. In the author's opinion, it is not a shortcoming: one–period statistical data themselves can tell nothing about a causal relationship, which is, specially, reflected in the symmetry of RIP coefficient.

Really, correspondence analysis visualizes the associations between row and column items expressed with the RIP values.

<u>Approximation Model for Correspondence Analysis</u>

Let us consider a contingency table $p_{ij}$ having row and column items $i \in I$ and $j \in J$, respectively, where entries $p_{ij}$ are supposed to be proportions of the individuals corresponding to the row item $i$ and to the column item $j$ simultaneously, with the row and column proportions $p_i$, $i \in I$ and $p_j$, $j \in J$, respectively. The value RIP($j/i$) equals

$$q_{ij} = p_{ij} / \left( p_i p_j \right) - 1$$

The traditional bilinear aggregation model for the values $q_{ij}$ could be written as follows:

$$q_{ij} = \sum_s \mu_s F_s(i) G_s(j) + e_{ij}$$

where $s$ are aggregated unknown factors with the values $\mu_s$, $F_s(i)$ and $G_s(j)$ to be found and $e_{ij}$ are the residuals to be minimized. The values $F_s(i)$ and $G_s(j)$ are interpreted as $s^{th}$ factor scores for row $i$ and column $j$ respectively; and squared $\mu_s$ characterizes the contribution of the factor s into the square scatter of the values $q_{ij}$. The solution of the model with the least square criterion (that is, $\Sigma_{i,j}\, e_{ij}^2$ to be minimized) is presented by the first m principal components of matrix $q_{ij}$ (see for example, Mirkin (1990)). To obtain the correspondence analysis solution for the initial data table $p_{ij}$, the same model equations could be used, but with the least square criterion changed by weighting (Mirkin, 1992):

$$L^2 = \sum_{i,j} p_i p_j \ e_{ij}$$

The proof is a straightforward implication from Lebart's results (see Lebart and Fénelon, 1971, p 239). This form of the correspondence analysis model is appropriate as a tool to generate new data analysis models (like correspondence-wise clustering described in the next section).

The weights $p_i p_j$ in $L^2$ could be interpreted as a comparison of the changes of the initial data imposed by the transformation $p_{ij}$ into $q_{ij}$. This makes the formal difference between the correspondence analysis and the principal component analysis approximation models; the last one doesn't presume any compensation to preliminary data transformations in its criterion.

## Correspondence–wise Clustering

Cluster analysis is a tool to reveal homogeneous groups of multivariate observations and, as such, is used in marketing research widely. However, usually, this tool is applied for one sort of item only: for the set of row items or (not and!) for the set of column items using the table rows (or, respectively, the table columns) as the item representations.

Here, a simultaneous row and column clustering model and procedure are considered, based on the modified correspondence analysis model. The only difference of the clustering model is that the sought factor scores, $F_s(i)$ and $G_s(j)$, are to be 1 or 0 only. In this case any factor could be represented by the sets $V_s$ of row items with $F_s(i) = 1$ and $W_s$ of column items, with $G_s(j) = 1$, and so be referred as cluster "pair" ($V_s, W_s$).

Thus, in this paper, a cluster is a pair of row and column subsets highly connected to each other in the sense of the RIP value. More explicitly, let $V$ and $W$ be subsets of the row and column item sets, respectively. For example, let $V$ consist of the row items corresponding to brands 3 and 5, and $W$ correspond to brands 2 and 4. What is the characteristic showing the switching flow between the subsets? The correspondence-wise clustering model fixes the same RIP coefficient value RIP($V$/$W$) (that is, RIP value computed for $V$ under condition $W$) as optimal value of $\mu$ (Mirkin (1992)). To calculate it we need knowing how many people correspond to the row set $V$, to the column set $W$, and to the cluster ($V$, $W$) itself. Let us use the data from Table 1 (Segment A) with the values from Table 3 substituted instead of the main diagonal values to make the sample more homogeneous. In this case row 3 corresponds to (322+70+217+49+721) = 1379 previous buyers, and row 5 to (1242+207+621+483+2465) = 5018 previous buyers. So, the number of previous buyers corresponding to set $V$ is 1379+5018 = 6397 which equals 54.7% of the homogenized sample. Analogously, using the columns 2 and 4, where 150 and 147 are substituted instead of diagonal entries 4035 and 3960, respectively, we can calculate the amount of new purchasers for brands 2 and 4 (set $W$) as 1177+762 = 1939 which equals 16.6% of the homogenized sample. Lastly, the total of the entries in the rows from $V$ and in the columns from $W$ (simultaneously) equals (70+49+207+483) = 809 = 6.9% of the sample. So

$$\mathrm{RIP}(W/V) = (\mathrm{P}(W/V) - \mathrm{P}(W))/\mathrm{P}(W) = (0.069/0.547 - 0.166)/0.166 = -0.241$$

which means that the conditional probability of buying the brands from $W$ is 24.1% less for previous $V$ purchasers than in all sample. Does this figure reflect important switching flows or not?

The clustering model requires gathering such a set of the row–column pairs that all of them have close values of their RIPs. The step–by–step adding algorithm (Mirkin, 1992) obtains clusters that satisfy the following "contrastness" condition. The RIP coefficient value for the cluster is at least twice more (if positive) or twice less (if negative) than the RIP values of any external row item (in relation to set $W$) and any external column item (in relation to set $V$). For the Table 1 (Segment A) data the algorithm finds the cluster with $V = \{3,5\}$ and $W = \{2\}$ as suboptimal one, having RIP value -57% (see Table 4), which satisfies the contrastness property in the example considered. Adding 4 to $W$ leads to RIP value -24.1%, which is worse than -57%/2 = 28.5%.

Simultaneous row and column clustering could be used as a tool, complementary to traditional correspondence analysis because it allows finding the clusters with positive RIPs, reflecting the continuous fragments of the correspondence analysis spatial representation of the rows and columns as well as with negative RIP, reflecting disconnected points of the spatial representation. The RIP values used for the cluster interpretation seem to be more direct and evident tool to understand than the spatial representation of correspondence analysis itself.

### Brand Switching Table: Loyal and Switching Behaviour

The data are presented with a square brand-by-brand data table (like Table 1) where entries ($i,j$) are the numbers of consumers who have purchased brand $j$ having bought $i$ before (i, $j$ are numbers of brands considered in the survey). Usually, the number of loyal buyers presented by diagonal entries ($i,i$) of the data table is tremendously large in comparison with the number of switchers, represented by off-diagonal entries (as in Table 1). It indicates that the set of consumers is not homogeneous and, so, correspondence-wise analysis cannot be applied to the data directly. There exists so-called Mover-Stayer model which pretends to model this effect considering the set of all consumers as consisting of two classes, stayers and movers, where stayers never change their initial choice and movers' behaviour can be described with a finite-state Markov chain (see Blumen, Kogan and McCarthy, 1955). Recently, Colombo and Morrison (1989) proposed a version of the model with some simpler mover's behaviour assumptions that seem to be more appropriate for this kind of data.

According to their model each of the brands i is characterized with two parameters: proportion $\alpha_i$ of completely loyal users, and proportion $\pi_i$ of potential switchers (the persons who are not completely loyal) who will next buy brand i. The conditional probability p(j/i) that the next purchase of a brand i user will be brand j, equals (by the Colombo and Morrison model) the switching probability $1 - \alpha_i$ multiplied by the next purchasing probability $\pi_j$. For the loyal buyers (i=j) the probability $\alpha_i$ is to be added to the result.

Here another version of the model is considered. The main difference is that this model describes the switching behaviour not for all the consumers, but for real switchers (off–diagonal entries) only. At the same time the proportion of potential switchers among the loyal buyers is presumed here to be independent of brand used, which seems to be more natural from psychological point of view: "loyalty" itself is considered here as a personal parameter, but "forces to switch out and to come in" are brand-dependent. Contrary to previous models, loyal buyer behaviour is not modeled.

Then the correspondence–wise analysis is applied to the data table relating to the potential switchers only; the diagonal entries are reconstructed from the off–diagonal potential switcher model.

## Off–Diagonal Switching Model

The key assumption here is that there are three kinds of consumers: Hard–Core Loyals (HCL), Potential Switchers (PS), and Real Switchers (S), so that every Real Switcher is obligatory a Potential Switcher. The set of Real Switchers is identified as the set of all consumers presented by off–diagonal entries of Brand Switching Table. This set is modeled with two sets of values: $f=(f_i)$, and $g=(g_i)$ where i is brand number (i=1,2,...,n), $f_i$ is the relative "strength out" value forcing a Potential Switcher to go out of the make i, and $g_j$ is the relative "strength in" value forcing a Potential Switcher to join the cell j (i, j = 1,2,...,n). It is possible to consider also the values $f_i$ and $g_j$ as source and target probabilities for the potential switchers, respectively. It is useful to point out that the sense of PS concept here differs from that in Colombo and Morrison (1989).

### A Formal Description of the Model

Let us denote the observed frequency of entry *(i,j)* of the Brand Switching Table through $p_{ij}$, considering it as the probability of the entry (*i, j* =1,...,n). Then the probability of Real Switching behaviour is equal to

$$p_s = \sum_{i \neq j} p_{ij} = 1 - \sum_i p_{ii}$$

Let us denote $\mu$ the probability for any loyal consumer to be a Potential Switcher; then the probability *a*, of Potential Switching behaviour is equal to

$$a = p_s + \mu(1 - p_s)$$

Using this notation, the main model equation can be expressed in the following form:

$$p_{ij} = af_i g_j$$

Formally, these equations are equivalent to the part of Colombo–Morrison's model which concerns the non–diagonal entries: it is possible to identify "strength in" values $g_j$ as their proportions of Potential Switchers who are next $j$'s purchasers, as well as, $af_i$ could be identified as $p_i(1 - \alpha_i)$, where $p_i$ is observed proportion of current $i$'s users, and $1 - \alpha_i$ is the proportion of non–completely loyal current $i$'s users. But the interpretation of those is different, and, more important, no equations for loyal buyers are considered here, contrary to the other models.

Mathematically, the parameters of the model are analogous to those of Colombo and Morrison (1989), but this one contains one supplementary parameter (probability of potential switching behaviour) and so has to fit data better.

## Illustrative Results

The Zufryden (1986) data recalculated from Colombo and Morrison (1989), Tables 4 and 5, are considered here (see Table 1, where frequencies $p_{ij}$, multiplied by 100,000, are presented). The market is defined with four basic brand types, 1, 2, 3 and 4, while "brand" 5 indicates the consumers not purchasing this kind of product.

Estimates of the model parameters, $f$ and $g$, are presented in Table 2. The quality of these estimates can be evaluated with the coefficient of determination, which is equal to the ratio of the totals of the squared estimates and of the squared observed values. The value of the coefficient equals 0.958 and 0.957 for the Segments A and B, respectively.

The probability of potential switching $\mu$ for any loyal purchaser is small in both of the segments and equals 0.037 for Segment A, and 0.027 for Segment B. So does the probability $a$ for the whole set of the purchasers, which equals to 0.117 or 0.089, respectively (and, for the real data case (see below), our approach leads to small potential switching probabilities, as well). This result does not correspond to the results of Colombo and Morrison (1989), where the probabilities of Potential Switchers are much more varied for different brands and reaching 0.44–0.46 for the most popular "brand" 5. To explain this, let us indicate that those in the Colombo and Morrison model present the main part of the "strength out" values f, which, in the present model, have no relation to the problem of potential switching measurement. On the contrary, in this paper's model, the probability of potential switching could be proved to be a monotone function of the real switching probability $p_S$, decreasing when the last is decreased. And for these data the values $p_S$ are rather small being equal to 0.083 for Segment A, or to 0.064, for Segment B. The diagonal elements of the tables estimated with these values of $\mu$, are presented in Table 3.

**Table 2: Source Probabilities to Switch (strength out)**

| | | | | | |
|---|---|---|---|---|---|
| Segment A | 0.2928 | 0.0996 | 0.1266 | 0.0899 | 0.3911 |
| Segment B | 0.2279 | 0.1300 | 0.1999 | 0.0375 | 0.4046 |

**Target probabilities to switch (strength in)**

| | | | | | |
|---|---|---|---|---|---|
| Segment A | 0.2672 | 0.0927 | 0.1304 | 0.0674 | 0.4422 |
| Segment B | 0.1959 | 0.1206 | 0.2130 | 0.0265 | 0.4439 |

Let us point out, however, that this difference does not contradict Colombo and Morrison's main conclusions on the contesting areas for the market of Switching Buyers. The terms of the present model only move these conclusions into area of the strengths "in" and "out". The principal recommendation for the market policy could be formulated here as follows: it is necessary, first, to increase the target values *g* of the brands under control based on the set of potential switchers from the brands with the large values of "strength out" *f*, and, second, decrease the "strength out" values for these brands. But, here the set of potential switchers (where these recommendations are valid) is much more narrow than it is described in Colombo and Morrison's paper. Proportionally, in real market policy for a brand it is necessary to concentrate more on younger generations to provide more proportion of loyal new buyers.

**Table 3: The Number of Potential Switchers Among the Loyal Purchasers**

| Brand | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Segment A | 423 | 150 | 217 | 147 | 2465 |
| Segment B | 219 | 108 | 207 | 43 | 1928 |

The results of correspondence analysis of the Segment A data are shown in Figure 1 (first two factor plane) having rather distant points representing the rows and the columns for the same brands (similar picture holds for Segment B data).

The results of correspondence–wise clustering of the rows and the columns of the potential switching data tables (with estimated diagonal entries) are presented in Table 4. Since the results for both of the market segments are similar, those are represented in the same table with one line representing any cluster for segment A and with the other, beneath, the corresponding cluster for Segment B. The clusters are numbered in decreasing order of their contributions to the scatter of the data (numbers of the clusters for Segment B are presented in parentheses).

**Table 4**

| Cluster | Sources | Targets | RIP % | Contribution % |
|---|---|---|---|---|
| 1 | 1 | 2 | 108.8 | 31.1 |
| (1) | 1 | 2 | 105.9 | 32.8 |
| 2 | 3,5 | 2 | −57.0 | 18.2 |
| (3) | 2,3,5 | 2 | −36.0 | 12.1 |
| 3 | 1 | 1,4 | −47.8 | 17.2 |
| (4) | 1,4 | 1,3,4,5 | −19.7 | 9.1 |
| 4 | 4,5 | 1,4 | 26.6 | 11.0 |
| (2) | 4 | 1,4 | 124.5 | 15.5 |
| 5 | 4,5 | 4 | 29.3 | 3.0 |
| (5) | 4,5 | 4 | 70.8 | 5.8 |

The contents of the table shows that the main switching flow is as follows: brands 4,5 are left for 1 and 4 (clusters 5(5) and 4(2)), and 1 is left for 2 (cluster 1(1)), with the increased intensity

*French data*

Two data tables were considered, 1986 and 1989, concerning the following 14 car brands:

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 Alfa | 2 BMW | 3 Citroen | 4 Fiat | 5 Ford | 6 GM | 7 Lada |
| 8 Mercedes | 9 Peugeot | 10 Renault | 11 Rover | 12 Seat | 13VW | 14 Volvo |

The results are presented in Table 6. [Intermediate steps have been omitted.] The main flows here are much more closed than in Great Britain (see the first 4 clusters; however, those, probably, are generated by non-homogeneous data (to infrequent brands), which is a feature of correspondence analysis model). Clusters 6, 7 of 86 transformed into Clusters 5, 10 of 89 with a small increasing of the flow in 10Re with the flow in 6GM decreased.

**Table 6**

| No | Year | Source set | Target set | RIP % | Contribution % |
|---|---|---|---|---|---|
| 1 | 86 | 2BMW, 8Mer | 2BMW, 8Mer | 710 | 29.1 |
| 1 | 89 | 2BMW, 8Mer | 2BMW, 8Mer | 827 | 32.9 |
| 4 | 89 | 8Mer | 2BMW, 8Mer | 713 | 6.4 |
| 2 | 86 | 1Alfa | 1Alfa | 672 | 5.3 |
| 2 | 89 | 1Alfa | 1Alfa | 828 | 5.6 |
| 3 | 86 | 7Lad | 7Lad | 818 | 5.0 |
| 6 | 89 | 7Lad | 7Lad | 532 | 2.4 |
| 4 | 86 | 14Vol | 14Vol | 917 | 4.4 |
| 3 | | 14Vol | 14Vol | 1040 | 5.6 |
| 5 | 86 | 9Pe | 7Lad, 10Re | 33 | 4.3 |
| 6 | 86 | 3Cit, 5For, 7Lad | 3Cit | 46 | 3.2 |
| 7 | 86 | 6GM, 10Re, 12Sea, 13VW | 6GM, 9Pe, 12Sea | 22 | 4.4 |
| 5 | 89 | 3Cit, 9Pe, 12Se | 3Cit, 10Re | 26 | 5.6 |
| 10 | 89 | 3Cit, 10Re, 13VW | 9Pe, 12Se | 18 | 2.9 |
| 8 | 86 | 1Alfa, 8Mer, 13VW, 14Vol | 1Alfa, 2BMW, 8Mer, 12Sea, 13VW, 14Vol | 71 | 4.9 |

Comparison of the results between the two countries shows the structure of switching flows in Great Britain is more interesting (it might be said that the switching flows in this country are really important); on the other hand, the proportion of potential switchers among loyal buyers in France is much higher.

## Conclusion.

A strategy to analyze switching behavior data is described. It consists of the following two steps: 1) substitution of amounts of loyal "potential switchers" (estimated with a special brand switching model) instead of real numbers of loyal buyers to make the data sample more homogeneous; 2) revealing main flows of switching behaviour using correspondence-wise cluster-analysis method.

The brand switching model proposed here resembles the model of Colombo and Morrison (1989), adding a quantitative parameter reflecting personal "loyalty" without any connection with the brand features which are modeled with so called source (out) and target (in) probabilities. The model shows potential switching behaviour as more rare than Colombo-Morrison's model does.

A new tool, correspondence-wise cluster analysis developed by the author, is proposed to reveal most "contrast" fragments of the switching table. The clusters are presented by pairs "row-subset-column-subset" having close values of relative increment of purchasing probability (RIP) for all entries between. Contrary to the correspondence analysis method, the fragments are discovered both with positive and negative RIP values. The correspondence-wise clustering could be used as a complementary tool to visualisation strategy of correspondence analysis.