Analysis of Car-switching Data

J. C. Gower and B. Zielman

Preliminary Findings

Our general approach with this kind of data would be to try to represent the data graphically in the hope of coming up with a model that could be fitted formally (e.g. by ML). In this instance we already have the model of Colombo and Morrison

$$\begin{aligned} p_{ij} &= (1-a_i) & i \neq j \\ p_{ii} &= (1-a_i)\pi_i + a_i & i = j \end{aligned} \qquad (1)$$

However the diagonal term of this model seems to imply that people may switch to the model they started with - which seems a little strange. The reparameterised form:

$$\begin{aligned} p_{ij} &= (1-b_i)\pi_j/(1-\pi_i) & i \neq j \\ p_{ii} &= b_i & i = j \end{aligned} \qquad (2)$$

seems less plausible but draws attention to the complexities of the "switching" probabilities $\pi_j/(1-\pi_i)$. The simpler form:

$$\begin{aligned} p_{ij} &= b_i\pi_j & i \neq j \\ p_{ii} &= c_i & i = j \end{aligned} \qquad (3)$$

seems worth considering.

The off-diagonal parts of all three models have simple multiplicative form and it follows that if any of these models is acceptable, a plot of the skew symmetric part of the data, as obtained from its singular value decomposition, should have approximate two-dimensional form - this should be so whether we handle the raw data themselves or work with the row conditional probabilities (i.e. divide by the row totals N_i). The singular values duly indicate very good fits in two dimensions and surprisingly, the associated graphical plots show linear structure. To be consistent with the models this linear structure would suggest that there is an approximate linear relationship between $(1-a_i)$ and π_j (model (1)) or between $(1 - b_i)/(1-\pi_i)$ and π_j (model (2)) or between b_i and π_j (model (3)). One way of discriminating graphically between the models is to plot the first term of each model against π_i. These linear relationships did not seem to be justified for the 1989 French data. Neither did the estimated values of a_i and π_i seem to be consistent with data values on the diagonal. We conclude tentatively that models (1) and (2) are unlikely to be acceptable. Nevertheless the linear skew-symmetry seems an adequate approximation for this part of any acceptable model.

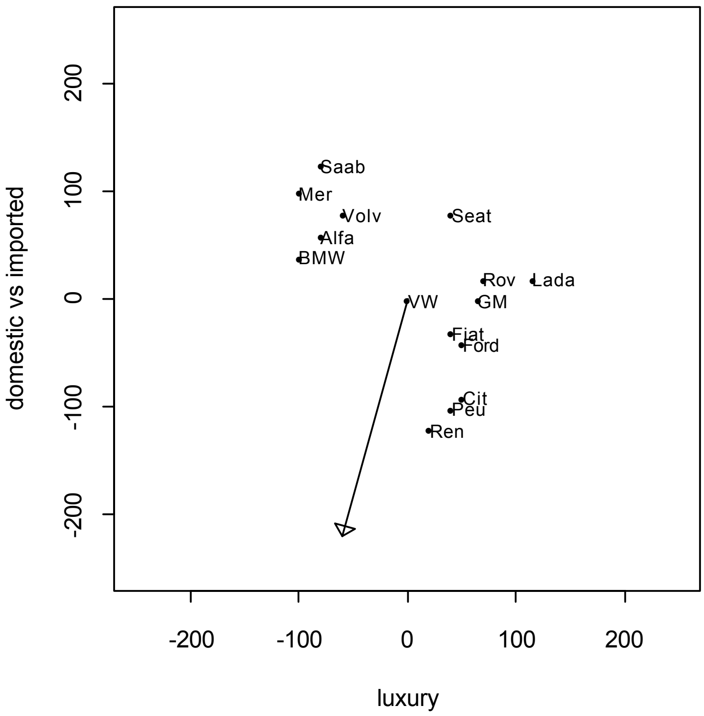
One of the problems with these sets of data is that the diagonal (the loyal non-switchers) tends to dominate. None of these models is log linear, so a preliminary analysis based on log transforms is not directly available. A crude analysis of the complete data verifies that the matrices have high rank so it seems pointless to expect a simple log-linear approximation. One way forward is to concentrate attention on the off-diagonal data which arise solely from the switchers. If one does this, all three models are log-linear with

$$\log(p_{ij}) = r_i + c_j \qquad i \neq j$$

This additive model can be fitted as a Poisson GLM, or crudely by least-squares, to give estimates of a_i and π_j etc. As a check whether or not the models for the full data are consistent with (1) and (2), the estimates just obtained can be used to predict the diagonal values; large residuals would suggest a failure of the model. The log-linear model was fitted by SPSS and fitted very well indeed. This fact is consistent with the observed linearity of the skew-symmetric component of the log-linear model. Again the fitted values were not consistent with the observed diagonals as expected from models (1) and (2). We concluded that of the models considered, model (3) is most likely to fit the 1989 French data. We cannot put any probabilistic interpretations onto this model and preliminary investigations showed no relationships between the parameters. Of course, in its row conditional form, (3) requires that $c_i+b_i \sum_{j\neq i}\pi_j = 1$. We would use this model as a starting point in examining the other data-sets and would hope that it would have some kind of global acceptability. If it did we would be encouraged to make further attempts at a substantive interpretation.

Analysis of Brand switching Among 15 Cars in France 1989 (B. Zielman)

As reported above, the plot of the skew-symmetric part of the data as obtained from singular value decomposition shows a two-dimensional form. The singular values indicate very good fits in two dimensions and the plot shows that skew-symmetry is linear. This is a very good starting



point for the slide-vector model since it requires linear skew-symmetry. The actual data were converted from similarities to dissimilarities by a transformation yielding pseudo-distances d_{ij}=-log n_{ij}/n_{i+} where n_{i+} is the row-total of brand i. The diagonal of the table, indicating brand loyalty was ignored by giving these elements zero weight in the analysis. In the marketing world you may call this an a-priori segmentation into a loyal segment and a switching segment. These pseudo-distances were scaled in two dimensions by the slide-vector model and fitted very well (proportion of sum of squares accounted for: .98). The resulting configuration is depicted in Figure 1.

In Figure 1 the brands are depicted as points in a two-dimensional space where the distances are inversely related to switching between brands. We see two clusters of points. The first cluster with members Saab, Mercedes, Alfa, Volvo and BMW is a cluster of luxury cars, indicating that these makes compete more with members within the cluster than outside this group. The second group of points consists of all other cars, where we see strong competition between the French cars Peugeot, Renault and Citroen. A dimensional interpretation yields the same results; the horizontal dimension can be interpreted as a luxury dimension with typical examples like Mercedes and BMW on the left and plain cars like Lada at the right. The vertical dimension is an asymmetry dimension and makes also a distinction between imported and domestic cars. In France, switching between cars occurs primarily between luxury cars, French cars and other small cars.

Asymmetry is described by the slide-vector, indicated in the figure by an arrow and label z. The slide-vector points to switching in, cars projecting high on the slide vector attract more consumers from the other makes, compared to what they lose. In this example, we see that Renault succeeds best in attracting consumers from the other makes, followed by Peugeot, Citroen, and so on. There are more relative relative switches from Citroen to Renault than from Renault to Citroen, more relative switches from Peugeot to Renault than the other way round, etc. This is also evident from the table since these brands have the largest gain in relative market share. The opposite is true for Saab, which loses consumers to all other brands. Analysis of the other tables showed similar results, with consistent gains in market-share for the French cars.

A possible candidate for the analysis of all tables jointly is an individual differences slide-vector scaling model (Zielman & Heiser, 1992). In summary, our analysis revealed three groups of brands with close competition, where the domestic cars have an advantageous position over the other makes. Analysis of the French automobile market indicated a good fit of the scaling model in two dimensions, and that switching is governed by two dimensions that can be labelled as an import-domestic dimension, and a plain vs. luxury dimension.

Further information on this type of model can be found in: DeLeeuw and Heiser (1982), Krishnaiah and Kanal (?), Zielman (1992), and Zielman and Heiser (1993).