Chapter 5

Nordisk Media Analys (NMA): The brand awareness task.

5.1 Introduction

Daniel Gillblad

5.1.1 The data

NMA analyses the effects of media investments using a method called tracking. This consists of recording all the investments made in most of the media types available, for example different newspapers, TV channels and outdoors advertising, on a daily basis. The impact of these investments are measured using telephone interviews with a statistically representative set of people, who answer questions about what brands they have noticed advertising for lately and which trademarks they prefer. The questions can be categorized into four different classes:

- Top of Mind advertising awareness (TOM)
- In Mind advertising awareness (IM)
- Preferred Brand
- Considered Brand

Top of Mind refers to the first brand mentioned by an interview subject when asked which brands he or she have noticed advertising for lately. In Mind are all the brands that the subject have noticed advertising for. The brand mentioned in the Top of Mind question is included in In Mind. Preferred and Considered refers to which brand that is preferred by the interview subject and which are possible. For example, if the questions concern cars, then Preferred would be which brand of car that most likely would be bought by the interview subject and Considered other possible car brands. The relationship between Preferred and Considered is similar to the relationship between Top of Mind and In Mind.

5.1.2 Preprocessing of the database

Two datasets have been delivered from NMA, one in late 1999 and one early 2001. From both datasets, time series were produced containing investments for a certain brand in eight different media types and the total investment with one sample a week. These media types were:

- Total media investments

- TV
- Radio
- Cinema
- Outdoors Advertising
- Morning Newspapers
- Evening Newspapers
- Popular Press
- Trade/Professional Journals

Time series showing the advertising impact was also constructed, with Top of Mind, In Mind, Preferred and Considered represented as a percentage of the total number of interview subjects on a weekly basis.

The first data delivered from NMA was in the form of a relational database, containing all investments and answers from telephone interviews between approximately August 1996 and August 1999. This database was flattened into two files containing the relevant information for this project, one for the investments and one for the interviews. From these, the relevant time series were produced. These series were used in all the early tests with the NMA data.

The second dataset delivered from NMA covered the period August 1996 to December 2000. The data was delivered as text files containing all media investments with the corresponding date and investor and the value of Top of Mind, In Mind, Preferred and Considered for each brand. Again, time series on the same format as before was produced. This data was used in the final evaluation of the methods on NMA data, using 1996 to 1999 as a training set and the year 2000 as the evaluation set.

5.1.3 The task

Predicting the advertising impact is an important application for NMA since that is essentially their core business. The task is to predict, from the media investments for different brands in previous weeks, the value of Top of Mind, In Mind, Preferred and Considered the next four weeks. Since the exact value of these predictions is not that interesting and since the data contains a fair amount of noise, an average of these four predictions is calculated. To evaluate the results, the smoothed predictions are compared to the same kind of average over the correct values, and a mean square error (MSE) and correlation coefficient are calculated.

It is generally much easier to predict Top of Mind and In Mind from the media investments than Preferred and Considered because of the much more direct influence of advertising. In the case of Preferred and Considered, accurate predictions are probably very hard to make. Therefore it was decided to concentrate on Top of Mind and In Mind in the final evaluation.

5.2 The models used at SICS

Anders Holst

SICS tried two prediction methods, both based on the Naive Bayesian classifier, and estimation of Gaussian distributions. Only the total media investments were used to predict In Mind and Top Of Mind in both methods. This is because a such a simple model as possible was preferred for this task, and the total investment seem to sum up the relevant information in the input pretty well.

The first method uses the four last weeks investments as input. Since these weeks are correlated with each other, a first order Markov model is used to compensate for these dependencies. To be more specific, the complete equation used for calculating the prediction is:

$$
\begin{aligned}
&P(y_t|x_t, x_{t-1}, x_{t-2}, x_{t-3}) \propto \\
&\quad P(y_t)P(x_t \mid x_{t-1}, y_t)P(x_{t-1} \mid x_{t-2}, y_t)P(x_{t-2} \mid x_{t-3}, y_t)P(x_{t-3} \mid y_t) = \\
&\quad P(y_t)\frac{P(x_t, x_{t-1}, y_t)P(x_{t-1}, x_{t-2}, y_t)P(x_{t-2}, x_{t-3}, y_t)}{P(x_{t-1}, y_t)P(x_{t-2}, y_t)}
\end{aligned}
$$

All the distributions in the equation are estimated as Gaussian distributions, and once specific values are inserted on the x:es, the result is a Gaussian distribution for y_t. The mean of this distribution is used as the prediction. The results from this model is labeled SICS 1 below.

The other model is even simpler. The average over the four last weeks is used as a single value from which to predict the output, using a singe Gaussian distribution. There is one twist though: The investment the current week is also used to estimate the difference in the output value. This difference is then added to the current week prediction, to get an adjustment due to this possible change. The motivation for this is that when trying different filters on the inputs and output, the highest correlation to the output was found from the average of the four last week (rather than any specific one of those weeks), and there were also a rather high correlation from the last week to the *difference* in output value between weeks. One would imagine then that to estimate the output for the current week, the estimation of the difference should be added to the estimation of the output for the last week. However, the estimates from week to week vary so much, so that a more consistent and stable approach is to say that the last weeks output is probably the same as this weeks output, except for the estimated difference. Therefore the estimation of the difference is added together with the above prediction of the same week, although the sum is weighted with the certainty (the variance) of the two estimated Gaussian distributions. To use the difference was a kind of experimental idea, but it probably had a very limited effect, since the estimation of the difference usually had a quite high variance, and therefore only affected the other prediction in a very limited way. Anyway, the results from this model is labeled SICS 2 below.

5.3 The model used at Mitthögskolan

Mikael Hall and David Martland

The problem supplied by NMA was treated by using feedforward neural networks. The data set was augmented by integrating and by taking first differences of the inputs.

5.3.1 Treatment

The model space used was given by feedforward artificial neural networks with time delays and converged to networks which had 5 hidden units plus linear output, and delays 0-3. The inputs consisted of the inputs, the first difference and a running four point mean of the given inputs. The data consisted of telephone interview records regarding the visibility in media for certain car brands and for companies selling clothes and patterns of investments in various media types made by these companies. The following inputs were available: Total investment, TV, Morning newspapers, Evening newspapers, Popular magazines and Specialized magazines.

The task was to predict the running four point mean of targets 'Top of Mind' and 'In Mind', which reflect the frequency interviewees report having seen the brand in media (the first brand they recall and at all, respectively), from patterns of investment. The models were tested at NMA on previously unavailable data, consisting of targets from year 2000.

The idea behind using the first difference and the running mean was to give the network a more explicit description of the current state. However this contributes to a certain redundancy of information in the inputs besides increasing the ratio between the number of model parameters and the number of training examples. Therefore steps were taken to reduce the networks ability to overtrain. First some of the available input variables were removed - Radio, Cinema and Outdoors. These variables had usually no or few data entries. Secondly a validation set was used to determine when to stop training. Lastly we used a cost function that besides the mean square error also takes into account the sum of the absolute values of the network weights, thus impairing the ability to fit the training data too closely. This type of addition to the cost function, or regularization, is called weight decay. To see what this does, one have to know how the transfer function used look like. In this case we use a shifted and scaled version of the sigmoid. The sigmoid is given by:

$$f(\omega) = \frac{1}{1+e^{-\omega}},$$

where ω is a linear combination of the inputs, $\omega = \sum w_i x_i$. The sigmoid squash the range of this sum into the range 0 and 1 in a symmetrical way. If ω approach -∞ f will tend to 0, if ω approach ∞ f will tend to 1, and if $\omega \approx 0$, f will keep the sum down almost to the same degree as up, so the behavior is almost linear. When weight decay introduce the bias $\sum {w_i}^2 \approx 0$, the non-linear network must find stronger evidence to develop non-linearities.

The model was run one hundred times for each brand from which the best was chosen according to a test set from the available data.

5.3.2 Results and conclusions

The performance showed a large variance over the brand names. On certain brands good predictions could be made, while other brands were harder to predict. This can be explained by the method of selecting the best network. Instead perhaps we should have formed the prediction of the mean of all (or almost all the runs neglecting the most extreme). This can also depend on changes in media investment behaviour by the harder to predict brands.

5.4 The model used at DSV

Lars Asker

Given the task of predicting the values of the two variables Top of Mind, and In Mind, DSV used the method Bagging-50. After some initial experiments it was decided to use only the total investment for each week for the previous four weeks to represent the data. An additional feature that was also experimented with was the number of different media types for each week, but this additional feature made no difference for the prediction results. The results are presented below in Table 5.1 in Section 5.7.

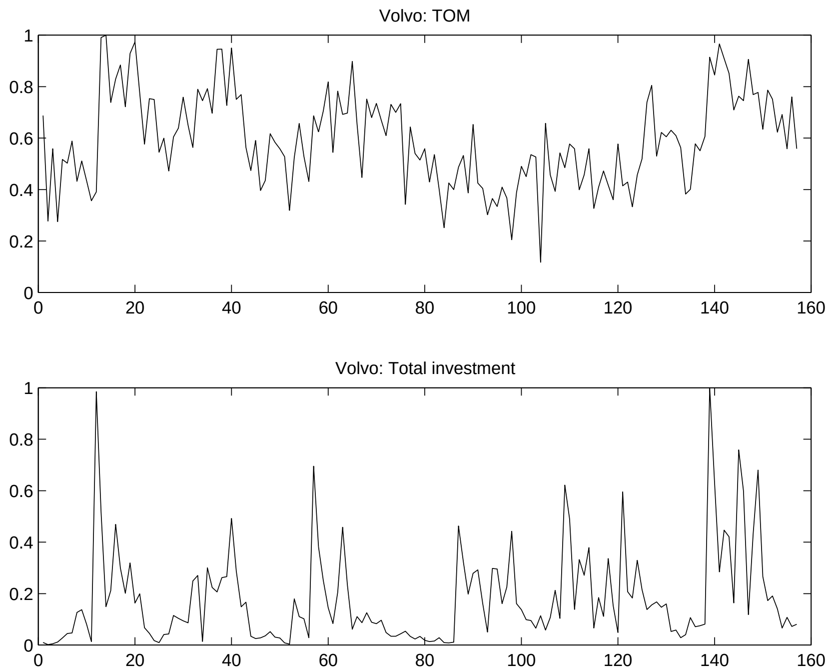

Figure 5.1: Relation between total investments and Top of Mind (scaled). The total data set contains information for 157 weeks.

5.5 The model used at Skövde University
Lars Niklasson

5.5.1 Introduction

Data from three different domains was supplied; the car industry, the travel industry and the clothing industry. For each industry, the investments per week (divided into different media types, *e. g.* TV, outdoor, popular press, etc.) are available. Connected to this, information of the effects of the advertising is also given in the data sets. Each week NMA randomly interviews a number of individuals and rate the percentage of individuals who first mentioned that they had seen an advert for a particular make/brand/company, as the 'Top of Mind' answer for that make/brand/company. In addition, 'In Mind' (*i. e.* if the make/brand/company is mentioned at all), 'Preferred' (*i. e.* if the particular make/brand/company is the preferred choice) and 'Considered' (*i. e.* that the make/brand/company could be a possible choice of purchase) are also given.

The task is to use data from investments and predict different effects of the advertising, *e. g.* rating an alternative as 'Top of Mind', 'In Mind', 'Preferred' or 'Considered'. A typical example can be found in Figure 5.1, where the percentages have been scaled to hide the actual values.

5.5.2 Recurrent networks for time-series prediction

The technique that HS apply is based on simple recurrent artificial neural networks (Elman, 1990), often termed SRNs. An SRN is a network that utilizes implicit representation of time. In this particular case, the investments in advertising (divided into different media types) each week are used as input to a recurrent net. The output is the predicted value of the particular variable describing the effect of the advertising, in terms of 'Top of Mind', 'In Mind', 'Preferred' or 'Considered'.

One network is trained on the data for each make/brand/company. The hypothesis is that different makes/brands/companies have different characteristics, *i. e.* that there is no 'general' model for all (this was also tested in a series of simulations). So far, only data from the car industry has been used.

5.5.3 The data

Input used: TV, Radio, Cinema, Outdoor, Morning press, Evening press, Popular press, Special interest press, Share of voice (week n-1).

One network is trained for each of the four variables, to give a prediction for a particular make. The first 100 weeks are used as training data, and the complete set (about 150 weeks) is used for testing. This means that at least 50 weeks have not been included in the training of the network.

As is often the case with ANN experimentation, a series of simulations was needed to decide on the network topology, *i. e.* training algorithm, representation of data, number of hidden units, etc. Some initial simulations showed that 15, rather than 10 hidden units gave the best result. It should be noted that this is a rather ad hoc number and that no thorough sensitivity analysis has been conducted. The initial simulations also showed improved performance when 'share of voice' was included in the input data. The inclusion of this variable is dependent on its availability during testing. It might to be hard to include in long term predictions, since it requires that the investments for a company/make (which is assumed to be known) is related to the investment of the whole industry (which appears to be harder to predict).

5.5.4 Some assumptions

The particular training environment (Matlab) used demands that a min and max value is given for each variable. The approach adopted here is to scale each value such that it is divided with 3 times the max value (in the training set) for the particular variable in question. The hypothesis is that extrapolation will start linearly rather than squashed (since sigmoidal units are used). This hypothesis can simply be tested by re-running the simulations using another scaling factor.

Since multi-layered ANNs have a tendency to generate different results depending on the initial random weight sets, the approach here is to run each simulation 5 times and take the average output as result. The current model does not have a sensitivity check for each output (*e. g.* standard deviation), but that is trivial to include. If so, it would become obvious if different networks come to very different results. This could be used to identify makes/brands/companies for which the prediction is less reliable. The cause could be related to various situations, *e. g.* errors or noise in the data, availability of training data, etc.

5.5.5 The simulations

Each network (*i. e.* one output variable for one make/company) is trained in 5 independent simulations using 100 weeks for training and the complete set for testing (plotting both network output and target values). The 100 week sequence is trained 2000 times (epochs). Total simulation time for each network is about 1.5 hours. This time can probably be substantially shortened if implemented in a more suitable programming language (*e. g.* C) or the training regime is improved.

5.5.6 Results

The approach taken here is to be objective in the separation of training and test data. The first 100 weeks are always (independent on data quality) used as training set. The results are the average output from the 5 independent networks.

5.5.7 Top of Mind

This variable appears to be predictable for some makes. The problem is that some makes do not have enough data for the training set. Another problem is that the values for some makes is in the region of

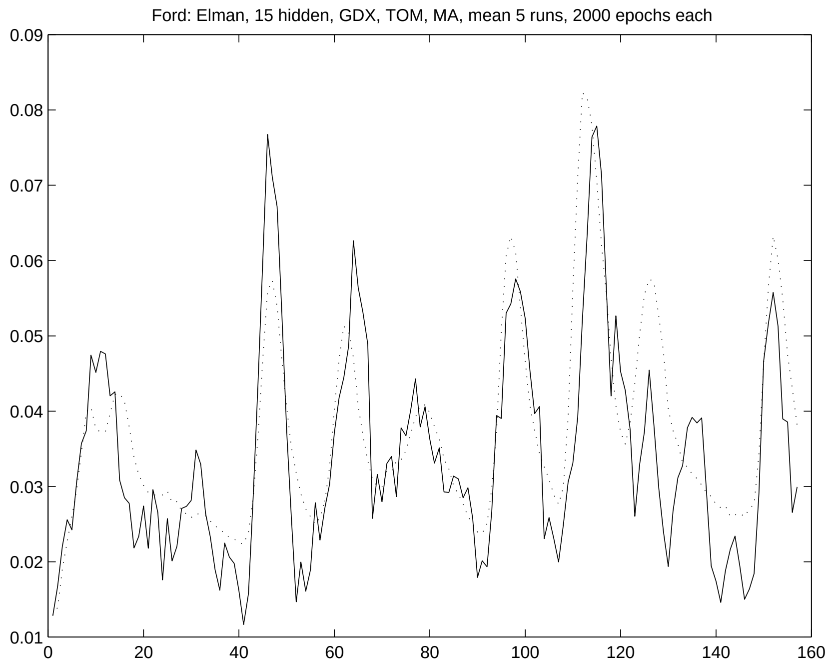

Figure 5.2: The results for Ford.

1/10 of percents. Since the number of people interviewed each week is in the hundreds, the results for these makes are very sensitive for noise. Examples of makes which have any of the problems include, Alfa Romeo, BMW, Chevrolet, Chrysler, Fiat, Honda, Jaguar, Jeep, Mazda, Lada, Peugeot, Renault, Rover, Seat, Skoda, Subaru, Suzuki.

The makes which appear to be possible to predict include, Audi, Ford (the best candidate), Hyundai, Mitsubishi, Nissan, Opel, SAAB, Toyota, Volkswagen and Volvo. Included here is the best (Ford) and worst case (Volkswagen) for these.

Mercedes has a special problem. During the introduction of a new car, it turned over in a test. This had a large effect on the attention it received in the media. The problem occurred around weeks 95-100, and caused the model to extrapolate in such a way that investments for Mercedes were predicted to have an extremely high effect, in the test set. This could possibly also explain some effects for other makes, *e. g.* Volvo, since 'Top of Mind' is a relative variable.

5.5.8 In Mind

The result for this variable is very similar to the results for 'Top of Mind'.

5.5.9 Preferred and Considered

These variables are much harder to predict. The results are not nearly as clear as the ones presented above.

5.5.10 The results

The results of these tests are presented in Table 5.1 in Section 5.7 summing up the results of all partners.

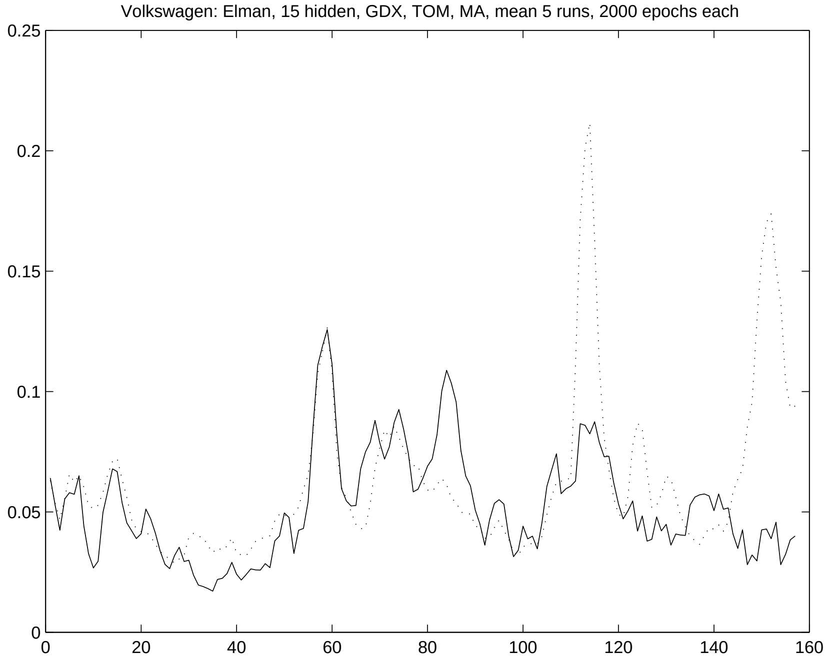

Figure 5.3: The results for Volkswagen.

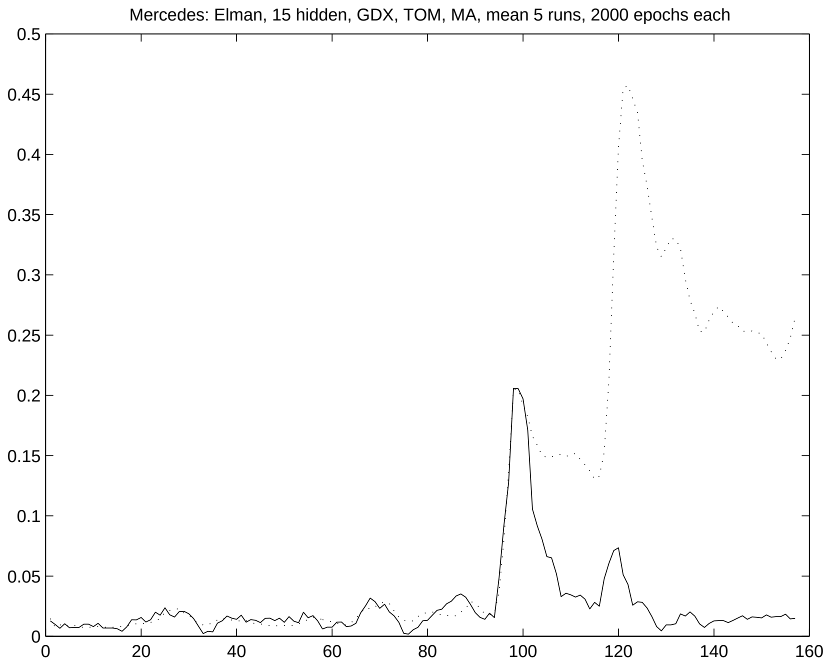

Figure 5.4: The large generalization error for Mercedes, caused by an extreme interest in the car, due to some technical errors.

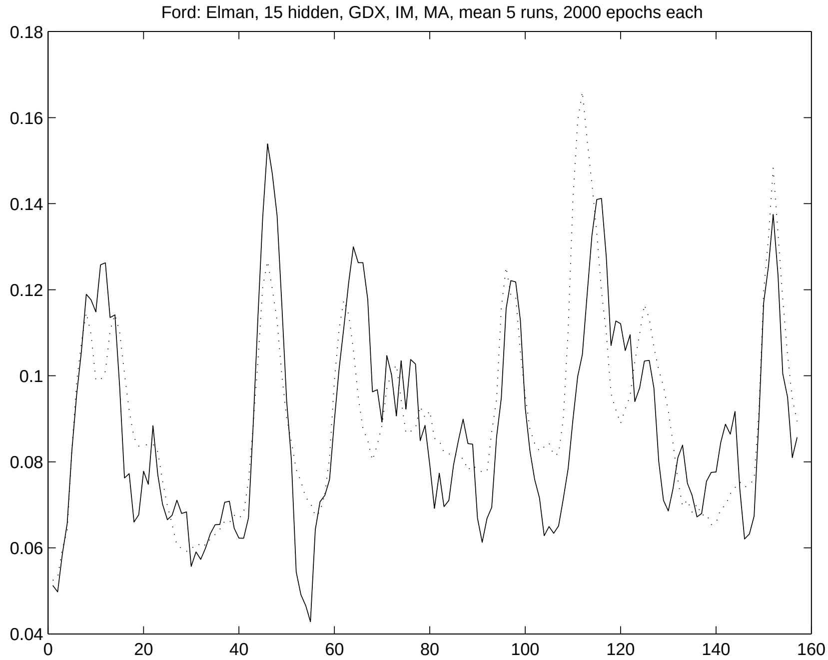

Figure 5.5: 'In Mind' for Ford.

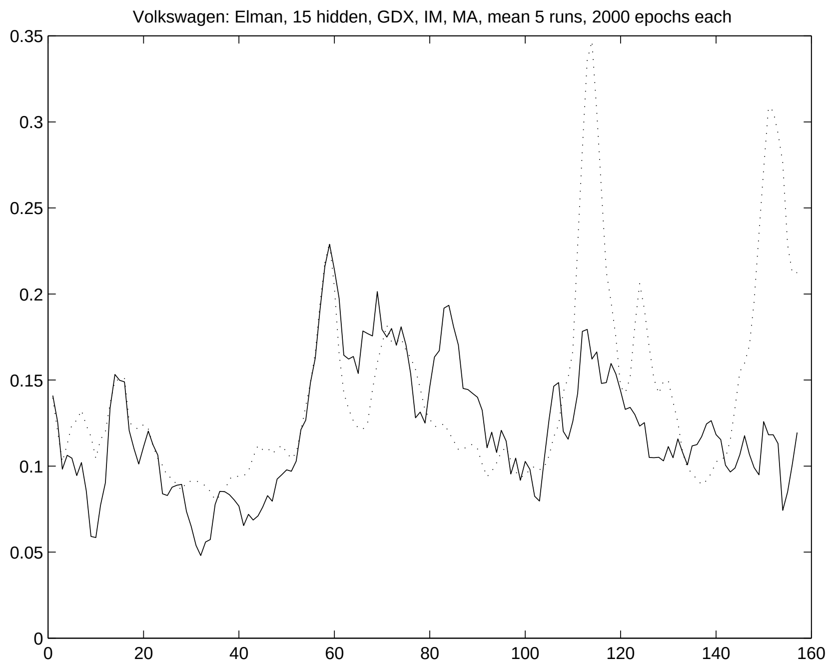

Figure 5.6: 'In Mind' for Volkswagen.

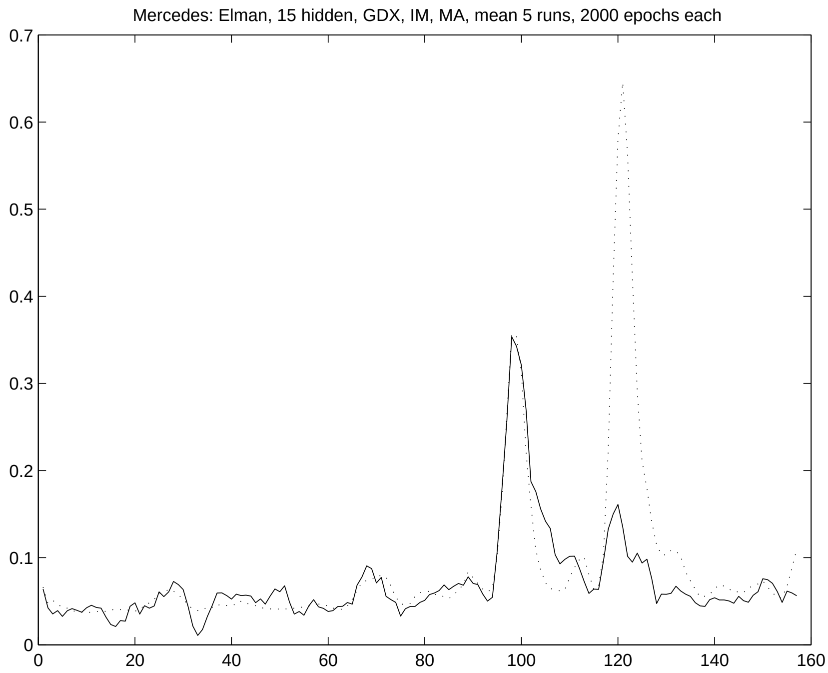

Figure 5.7: 'In Mind' for Mercedes.

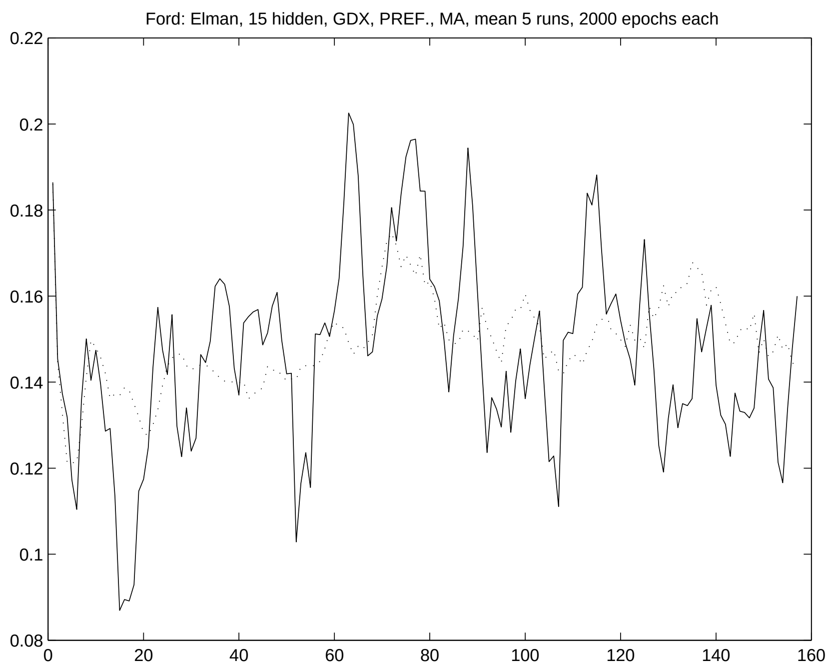

Figure 5.8: 'Considered' for Ford.

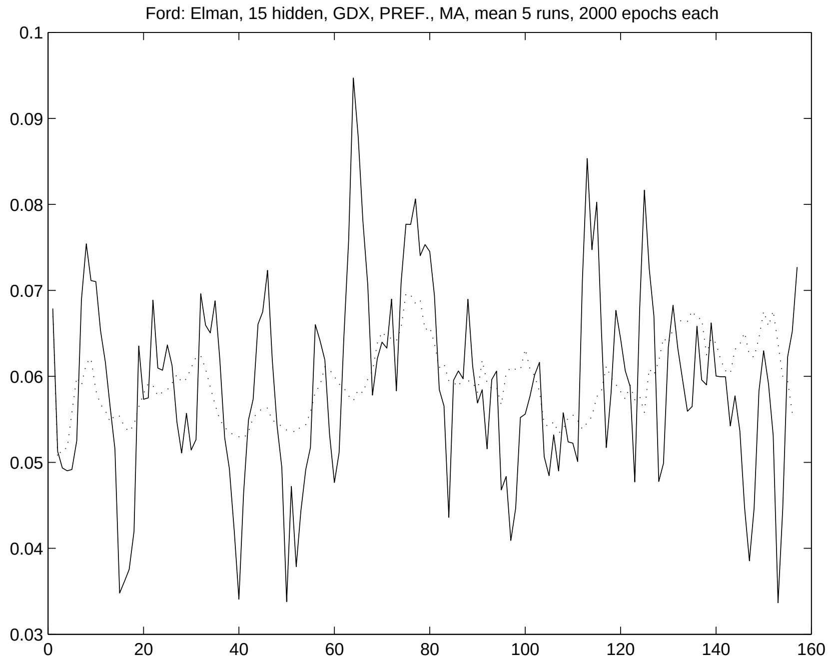

Figure 5.9: 'Preferred' for Ford.

5.6 The model used at Halmstad University

Thorsteinn Rögnvaldsson and Jim Samuelsson

5.6.1 Definitions

Before presenting our results, we introduce some definitions.

We define the following sets of brand names, media categories, and advertisement impact categories:

$$\begin{aligned}
\mathcal{C} &= \{\text{Alfa Romeo}, \text{Audi}, \text{BMW}, \text{Chevrolet}, \text{Chrysler}, \text{Citroën}, \text{Daewoo}, \\
&\quad \text{Daihatsu}, \text{Dodge}, \text{Fiat}, \text{Ford}, \text{Honda}, \text{Hyundai}, \text{Jaguar}, \text{Jeep}, \text{Kia}, \\
&\quad \text{Lada}, \text{Land-Rover}, \text{Lexus}, \text{Mazda}, \text{Mercedes}, \text{Mitsubishi}, \text{Nissan}, \\
&\quad \text{Opel}, \text{Peugeot}, \text{Renault}, \text{Rover}, \text{Saab}, \text{Seat}, \text{Skoda}, \text{Subaru}, \text{Suzuki}, \\
&\quad \text{Toyota}, \text{Volkswagen}, \text{Volvo}\}, \qquad (5.1) \\
\mathcal{G} &= \{\text{tv}, \text{radio}, \text{cinema}, \text{outdoor}, \text{morning papers}, \text{evening papers}, \\
&\quad \text{popular press}, \text{specialist press}\} \qquad (5.2) \\
\mathcal{A} &= \{\text{Top of Mind}, \text{In Mind}, \text{Preferred}, \text{Considered}\} \qquad (5.3)
\end{aligned}$$

We also define the following subsets of $\mathcal{C}$, based on the brand name's geographic origin or the total annual advertisement investment for that brand:

$$\begin{aligned}
\mathcal{C}_{german} &= \{\text{Audi}, \text{BMW}, \text{Mercedes}, \text{Opel}, \text{Volkswagen}\}, \qquad (5.4) \\
\mathcal{C}_{asian} &= \{\text{Daewoo}, \text{Daihatsu}, \text{Honda}, \text{Hyundai}, \text{Kia}, \text{Lexus}, \\
&\quad \text{Mazda}, \text{Mitsubishi}, \text{Nissan}, \text{Subaru}, \text{Suzuki}, \text{Toyota}\}, \qquad (5.5) \\
\mathcal{C}_{swedish} &= \{\text{Saab}, \text{Volvo}\} \qquad (5.6) \\
\mathcal{C}_{others} &= \{\text{Alfa Romeo}, \text{Chevrolet}, \text{Chrysler}, \text{Citroën}, \text{Dodge}, \text{Fiat}, \\
&\quad \text{Ford}, \text{Jaguar}, \text{Jeep}, \text{Lada}, \text{Land-Rover}, \text{Peugeot}, \text{Renault}, \\
&\quad \text{Rover}, \text{Seat}, \text{Skoda}\}, \qquad (5.7) \\
\mathcal{C}_{<50} &= \{\text{Alfa Romeo}, \text{Chevrolet}, \text{Daewoo}, \text{Daihatsu}, \text{Dodge}, \text{Jaguar}, \\
&\quad \text{Jeep}, \text{Kia}, \text{Lada}, \text{Land-Rover}, \text{Lexus}, \text{Subaru}, \text{Suzuki}\}, \qquad (5.8) \\
\mathcal{C}_{50-150} &= \{\text{BMW}, \text{Chrysler}, \text{Citroën}, \text{Fiat}, \text{Honda}, \text{Mazda}, \\
&\quad \text{Mercedes}, \text{Peugeot}, \text{Rover}, \text{Seat}, \text{Skoda}\}, \qquad (5.9) \\
\mathcal{C}_{>150} &= \{\text{Audi}, \text{Ford}, \text{Hyundai}, \text{Mitsubishi}, \text{Nissan}, \text{Opel}, \text{Renault}, \\
&\quad \text{Saab}, \text{Toyota}, \text{Volkswagen}, \text{Volvo}\}. \qquad (5.10)
\end{aligned}$$

For later work, we have also set the following brand names aside to be used as an out-of-sample test set:

$$\begin{aligned}
\mathcal{C}_{test} &= \{\text{Kia}, \text{Lada}, \text{Land-Rover}, \text{Lexus}, \text{Peugeot}, \text{Rover}, \text{Saab}, \text{Seat}, \\
&\quad \text{Skoda}, \text{Subaru}, \text{Suzuki}, \text{Toyota}, \text{Volvo}\}. \qquad (5.11)
\end{aligned}$$

The Swedish brands Volvo and Saab were set aside for the reason that Swedish brand names have a "national" aspect that none of the others show. This fact is illustrated in figure 5.10.

We denote an individual car brand name by c, a media category by g, and w denotes the week number. Then $i_{cg}(w)$ is defined as the advertisement investment, measured in money, made by car maker c in media category g, during week w.

We define $t_c(w)$ as the level of Top of Mind that car maker c has reached during week w.

We have set two tasks for our work so far:

1. For an individual car brand, explore the possibility of constructing a model for how investment affects the Top of Mind. That is, construct prediction models of the form: $\hat{t}_c(w) = f_c(i_{cg}(w), i_{cg}(w-1), ...)$.

2. For all car brands, explore the possibility of constructing a general model for how investment affects Top of Mind. That is, construct prediction models of the form: $\hat{t}(w) = f(i_{cg}(w), i_{cg}(w-1), ...)$.

Here the hat $\hat{t}$ over t indicates that it is an estimate and not the true value. In the first case, we use historic data about car brand name c and try to predict how the mapping from investment to Top of Mind will be in the future. In the second case, we use historic data from a pool of car brand names, and try to find a model that generalizes to another brand name.

All models are evaluated using the normalized prediction error (NPE)

$$\text{NPE} = \frac{\sum_w \left[\hat{t}(w) - t(w)\right]^2}{\sum_w \left[\bar{t} - t(w)\right]^2}, \tag{5.12}$$

where $\bar{t}$ is the mean Top of Mind over the training weeks, and the sum runs over all test weeks (or training weeks if the NPE is computed over the training set). The NPE is a measure of how good our model is compared to the best constant model, which outputs the mean training value. An NPE close to unity indicates a less well performing model.

We throughout use the first 100 weeks for training and the last 50+ weeks for testing.

5.6.2 Data preprocessing

Removing data

It is unusual to find any investment in movie commercials (some car manufacturers invest in movie commercials but most do not). We have therefore excluded the movie media from our investigation.

Transforming data

The investment data is (obviously) not normally distributed. Instead, it is rather skewed since small investments are significantly more frequent than large investments. We have tried to transform the investment data using Box-Cox like transformations, *i. e.*

$$x \to (x + \lambda)^\gamma, \tag{5.13}$$

so that the data is less skewed.

Normalization

It intuitively seems reasonable to normalize the numbers, both investment quantities and impact quantities, before building a model. We have tried two different normalization strategies for the the investment data. These are:

$$\tilde{i^1_{cg}}(w) \equiv \frac{i_{cg}(w)}{\sum_{\mathcal{C},\mathcal{G}} i_{cg}(w)} \tag{5.14}$$

$$\tilde{i^2_{cg}}(w) \equiv \frac{i_{cg}(w)}{\sum_{\mathcal{C}} i_{cg}(w)} \tag{5.15}$$

where the sum in equation (5.15) runs over over all car brands, and the sum in (5.14) runs over the different car brands as well as media categories. Expression (5.15) is a measure of the “share of voice” in media channel g, whereas expression (5.14) is a measure of the "share of voice" in the total signal from all media channels and car brands.

Note: The data contains a variable that is expected to hold the total investment over all media groups in week w. We found that this differed quite a bit from what we get when we sum over car brand names and/or media groups. We therefore did not use the "total" variable in the data set.

For Top of Mind, we normalize according to

$$\tilde{t_c}(w) \equiv \frac{t_c(w)}{\sum_{\mathcal{C}} t_c(w)} \tag{5.16}$$

where the sum runs over the different car makers. This normalized Top of Mind is a measure of what fraction of Top of Mind the car brand name has among people who actually have a Top of Mind opinion at all. The unnormalized Top of Mind typically sums to only about 0.7, when summed over all car brand names, indicating that about 30% of the interviewed subjects have no car name on the top of their mind.

Filtering

The data is quite noisy and it is difficult to discern any clear structure. We therefore filter the data to remove the high frequency components.

We have tried two different methods for filtering the investment data, based on two different assumptions about the process. The first is a moving average, the second is a first order Gamma filter.

The moving average filter works as follows:

$$I_{cg}^{r1}(N; w) \equiv \frac{\sum_{n=0}^{N-1} \tilde{i_{cg}^{r}}(w-n)}{N}; r = 1, 2. \tag{5.17}$$

That is, the investment in week w is replaced by the average investment over the current week and previous $N-1$ weeks ($N = 5$ typically). The "assumption" here is that an investment made in week w will have an impact N weeks into the future, and the fractional impact is the same for all weeks, but no impact after that.

The first order Gamma filter works as follows:

$$I_{cg}^{r2}(\beta, w) \equiv \frac{\sum_{n=0}^{\infty} \tilde{i_{cg}^{r}}(w-n)e^{-\beta n}}{1-e^{-\beta}}; r = 1, 2. \tag{5.18}$$

That is, the investment in week w is replaced by the exponentially averaged investment over the current week and previous weeks. The "assumption" here is that an investment made in week w will have an exponentially decaying impact on "all" weeks into the future, with a decay factor β ($\beta = 0.3$ typically).

The Top of Mind is only filtered with a moving average filter:

$$T_c^1(N; w) \equiv \frac{\sum_{n=0}^{N-1} \tilde{t_c}(w-n)}{N} \tag{5.19}$$

We have tried both building models for the filtered $T_c^1(N; w)$ and for the unfiltered $t_c^1(w)$. In the latter case we filter the predictions and the true Top of Mind afterwards, to be able to compare the results.

5.6.3 Results

The time lag

There are several different methods for estimating the time lag between the "input", *i. e.* the investment, and the "output", *i. e.* the Top of Mind score. We have tried four different methods:

- Computing the Pearson correlation coefficient ρ.
- Computing a correlation between the sign of change for the different series, *i. e.* sign runs.
- Look at the series and subjectively judge what seems to be the lag.
- Construct models with different lags and see which model that gives the best performance.

The "optimal" time lag also depends on the filtering method.

The moving average filter

When using the moving average filter (5.17) we conclude that the "optimal" time lag seems to be of the order of two weeks. This will be illustrated below, for the car maker Ford.

The Gamma filter

We explored the time lag when using the Gamma filter in several ways. First we experimented a little to find a suitable value for the decay constant β. This was set to $\beta = 0.3$. We then did single variable linear regressions for each of the variables and for different groups of car brand names, using different lags.

In the case of large investors, $\mathcal{C}_{>150}$ (excluding the car manufacturers that are in the test set $\mathcal{C}_{test}$), we concluded that the lag for TV commercials is zero, whereas other media seem to have a lag of 2-4 weeks. However, the TV media has the overall strongest effect on the Top of Mind.

In the case of medium size investors, $\mathcal{C}_{50-150}$ (excluding the car manufacturers that are in the test set $\mathcal{C}_{test}$), we concluded that the lag for TV commercials is 2, whereas other media have a lag of zero weeks. In this group the TV is not dominant. Instead movies, popular and evening press seem to have the strongest effect (although the effect is not very prominent).

Because of this observed difference between groups of car brands we therefore chose to work only with large investors. We try both a lag of 2 weeks for all inputs, and to set the lag for TV commercials to zero and other advertisement media to 2 weeks. The exponential decay in the Gamma filter is set to $\beta = 0.3$ (no significant difference is seen for $\beta \in [0.25, 0.45]$).

Linear models for individual car brands

We do most of the model exploration using linear models, to allow more freedom for choosing input variables and time lag. We have tried using both partial least squares (PLS) and ordinary least squares (LS) for the model construction. PLS was used in the case of moving average filter on the inputs, LS was used in the case of the Gamma filter representation.

In the PLS case, we also normalize the input variables to have zero mean and unit variance. In the LS case we do not do this.

Ford benchmark: PLS using moving average filter

To show the impact of changes in the models when varying different model parameters we have presented and discussed above, we use the following model parameters as a benchmark reference:

(i) car maker: $c =$ Ford, denoted F below

(ii) normalization: $r = 1$ (*i. e.* sum over both media group and brand names)

(iii) filtering: $s = 1$ and $N = 5$ weeks (*i. e.* moving average filter)

(iv) time lag: $l = 2$ weeks (for all inputs)

(v) media categories: $\mathcal{G} -$ cinema $=$ {tv, radio, outdoor, morning papers, evening papers, popular press and specialist press } (cinema is excluded because of lack of data)

When training brand name specific models, we use as training set (Ω_{train}) the first 100 weeks out of the total number of weeks that are available (typically 157 weeks). The remaining weeks constitute the test set (Ω_{test}).

The NPE values for the Ford benchmark model are: NPE_{train}=0.36, NPE_{test}=0.48.

PLS using moving average & varying model parameters

Figure 5.11 shows (to the left) the two curves that represent our benchmark reference model defined above; the real data for Top of Mind, $T^1_F(5;w)$ and the corresponding modeled values, $\hat{T}^1_F(5;w)$. To the right in the same figure is shown what happens when we choose a different normalization of the input data; *i. e.* $r = 1 \rightarrow 2$. In the latter case, the investments are normalized over media categories only. In the benchmark case, the normalization is done over both media categories and car brand names.

There seems to be no major difference between the two normalization methods.

In figure 5.12 (left) we have switched back to $r = 1$, but instead turned off the moving average filtering, which mathematically means $N = 1$ week. The predictive power goes down (obviously since our output now has a larger noise content), which is indicated by a higher NPE value. In Figure 5.12 (right) we plot NPE as a function of N.

Going back to $N = 5$ weeks, we now remove the time lag, *i. e.* $l = 0$. Figure 5.13 (left) shows the result, and in the right part of the figure we plot NPE as a function of l. Here we use the same lag for all inputs.

Results for some other car makers

To investigate how well our model does for some other car makers, we pick other cars from the subsets $\mathcal{C}_{<50}, \mathcal{C}_{50-150}, \mathcal{C}_{>150}$ defined in (5.8), (5.9) and (5.10). From $\mathcal{C}_{<50}$ (to which Ford also belongs) we choose Audi and Volvo, from the middle group $\mathcal{C}_{50-150}$ we take BMW, and from the low-level investors $\mathcal{C}_{<50}$ we choose Alfa Romeo. We otherwise use the same parameters as in the benchmark reference set.

In figure 5.14 we see that for Audi our model performs rather well, for the training set as well as the test set. For Volvo however, the model has problems, especially for the data in the test set. Figure 5.15 illustrates the situation for BMW and Alfa Romeo. The NPE values for both these car makers are ≈ 1, confirming the impression of poor performance from the plots. The poor performance is probably due to lack of data for these two car makers.

PLS with moving average & impact of media groups

Having convinced ourselves that our approach makes sense, we can now tackle the probably most interesting problem: Through which media group does an investor reach the highest Top of Mind level? We have in the first phase of the project addressed this question by studying how the NPE value varies with the choice of input; *i. e.* media groups. Can we, for example, make a model with good predictive power by only using a few media groups; *i. e.* a few members of the set $\mathcal{G}$? Once again we choose to study Ford, but instead of the full set $\mathcal{G}$, we now try subsets $\mathcal{G}'$, various combinations of the members of $\mathcal{G}$.

In figure 5.16 we show the result when only using the media group TV as input. The NPE value is basically the same as when using the full set $\mathcal{G}$, indicating that Ford and TV commercials go along well.

In figure 5.17 (left) we have only used *popular press* as input. One sees that the predictive power is gone, and this is confirmed by a higher NPE value. This fact indicates that the impact from *popular press* is very low for Ford. From figure 5.17 (right) we make the same conclusion for *morning press*.

As a matter of fact, when looping over all the different input combinations; *i. e.* all possible different subsets $\mathcal{G}'$, for Ford, Audi and Volvo, we find that the subset with the lowest value for NPE_{pred} contains *tv* as an element, for all three car makers.

LS using Gamma filter

All the LS models are constructed by regressing on the unfiltered output (Top of Mind or In Mind). Both the prediction and the model output are then filtered, with a moving average filter, afterwards.

Ford: We first constructed a LS model for Ford using all available variables (a total of 7). This gave NPE = 0.52 for normalized Top of Mind, and NPE = 0.67 for In Mind, over the test set Ω_{test}. The same numbers for the training set Ω_{train} were 0.38 and 0.38, respectively.

We then tried backward elimination, *i. e.* removing the variables one by one, always removing the one that resulted in the maximum improvement. This resulted in the following set of "optimal" four variables for Ford: *TV, evening press, popular press, specialist press* . They yield NPE = 0.48 for normalized Top of Mind, and NPE = 0.51 for In Mind, over the test set Ω_{test}. The numbers for the training set Ω_{train} were 0.45 and 0.45, respectively. The model's output over training and test sets are shown in Figure 5.18.

Transforming the variables using a transformation of the general form (5.13) gives models with the following performance: NPE = 0.53 for Top of Mind and 0.50 for In Mind (test data) when all variables are used. Backward elimination gives *TV, morning press, popular press* as the "optimal" variable set, and the NPE scores are 0.45 on normalized Top of Mind and 0.37 on In Mind (test set). The result using these variables (transformed) is shown in Figure 5.19.

Note: The performance using only investment in TV commercials as input is essentially as good as the "optimal" variables.

Audi: We also constructed a model using all inputs for Audi. This gave NPE = 3.86 for normalized Top of Mind, and NPE = 2.07 for In Mind, over the test set Ω_{test}. The same numbers for the training set Ω_{train} were 0.82 and 0.75, respectively.

Backward elimination resulted in the "optimal" variable *TV* for Audi. Using only investment in TV commercials as input gave NPE = 0.84 for normalized Top of Mind, and NPE = 0.88 for In Mind (test data). The result is shown in Figure 5.20

Forward selection gave the same result.

Transforming the variables a la (5.13) gives a model with NPE = 1.37 for normalized Top of Mind, and NPE = 0.79 for In Mind, when all input variables are used. Backwards elimination leads to a model with only one variable, investment in TV commercial, and the NPE values 0.71 and 0.78 for normalized Top of Mind and In Mind, respectively. The results using only TV investments is shown in Figure 5.21.

All car brands: We tried optimizing the variables with respect to the median test NPE for Top of Mind over all car brands in $\mathcal{C}_{>150}$. This experiment would reveal investment variables that are generally informative, *i. e.* not only important for a specific car brand name but also for other car brand names. Our conclusion from this experiment was that the median NPE_{test} does not change much when variables are removed. However, the spread (especially upwards) decreases significantly. The best overall model is achieved when only TV commercial investment is used. The result is shown in Figure 5.22.

The same is true if we transform the variables, which is shown in Figure 5.23.

Typically, the test NPE for normalized Top of Mind for a large investor when using transformed variables will be about 0.7. However, the spread is large with values as low as 0.4 and as high as above 1.0.

Linear models across car brands

One interesting question that we see is if it is possible to construct a model based on a few car brands and generalize this model to a new brand name. That is, is the effect of commercials on Top of Mind brand specific or is it a generic effect?

We have built models on a subset of $\mathcal{C}_{>150}$ and then tested on another car brand in $\mathcal{C}_{>150}$ (so-called cross validation). Our conclusion is that only TV commercials have a general effect on Top of Mind (or In Mind for that matter).

The typical test NPE value is 0.7. Again, the spread is considerable.

It seems possible to use only TV commercial investment as input, and train on any set of car brands, and the resulting model will on average be as good as a model constructed for that specific brand name. The typical NPE for normalized Top of Mind will be 0.7.

For some individual brands, like Ford, it seems possible to make a better model than the generic TV-model, if one trains the model on brand specific data.

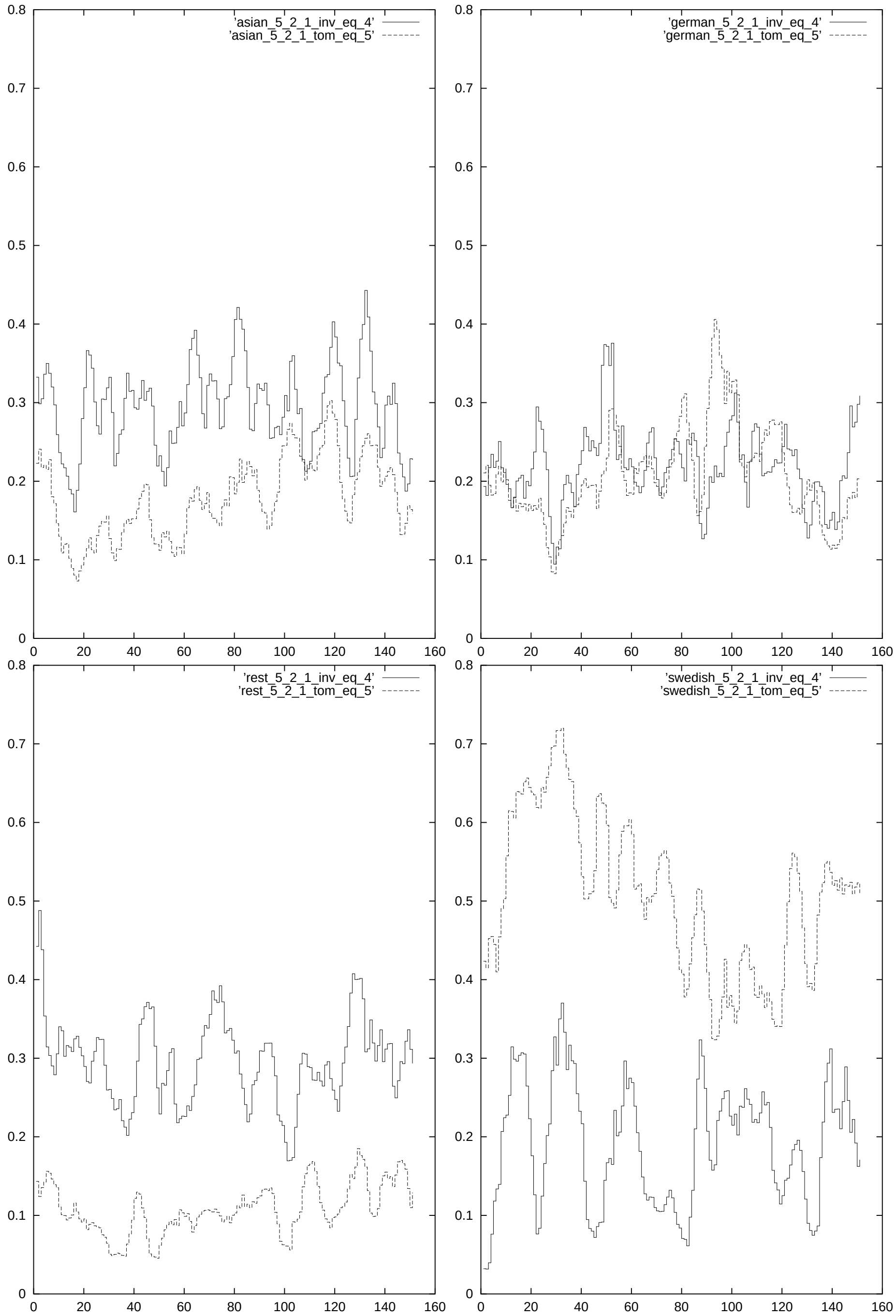

Figure 5.10: Car maker patriotism: Time series for *normalized* levels of investment (solid lines) and Top of Mind (dashed lines). Top left: Car brands from Asia (Toyota, Honda, Hyundai, ...). Top right: Car brands from Germany (Mercedes, BMW, Volkswagen, Audi). Bottom right: Swedish car brands (Volvo and Saab). Bottom left: Other brand names (Chrysler, Jaguar, ...). Swedish car makers get the best Top of Mind returns for invested money, followed by German car makers.

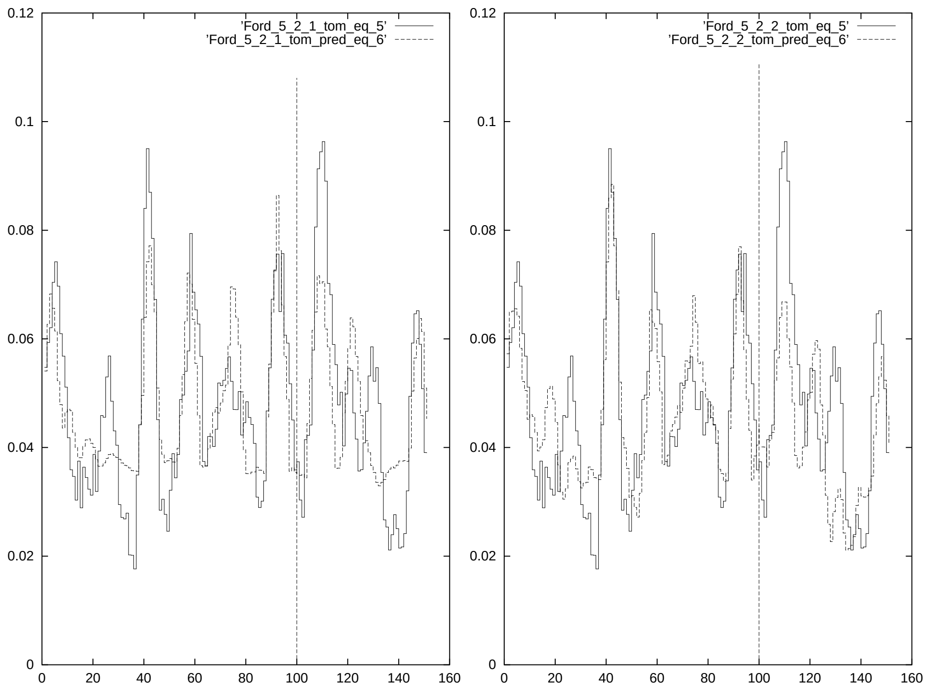

Figure 5.11: Left: Results for the parameters in the benchmark reference (section 5.6.3): NPE$_{train}$=0.36, NPE$_{test}$=0.48. Right: Results when normalizing the investment over media group only: NPE$_{train}$=0.33, NPE$_{test}$=0.62. (Solid lines are true values, and dashed lines are predicted values.)

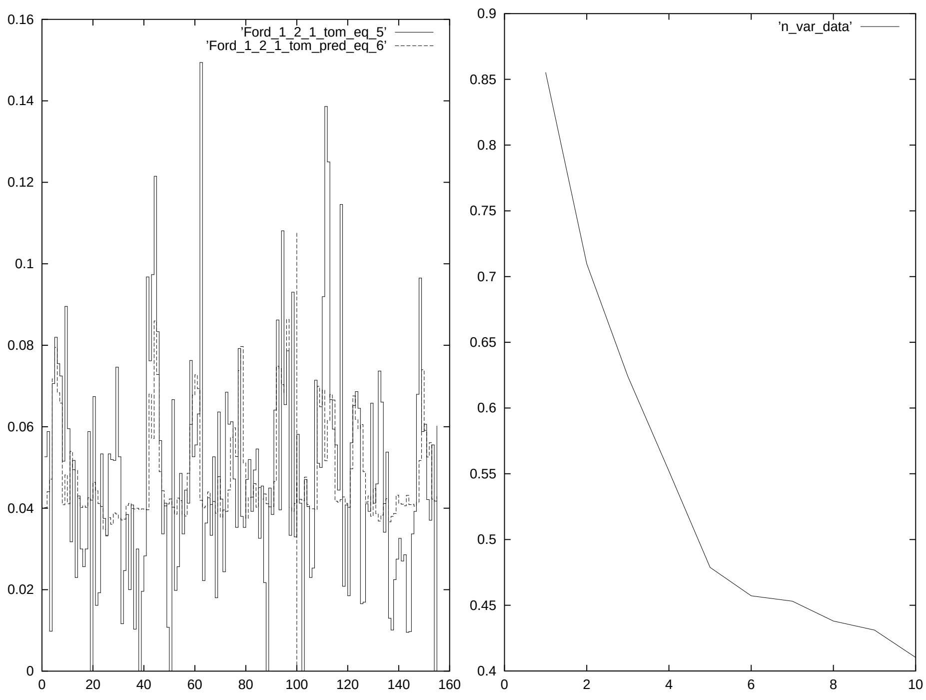

Figure 5.12: Left: Filtering switched off; NPE$_{train}$=077 and NPE$_{test}$=0.86. (Solid line: real data; dashed line: modelled data.) Right: NPE$_{test}$ as a function of the filter length N. Note that the decrease of NPE$_{test}$ is mostly due to the decrease in noise content in the output.

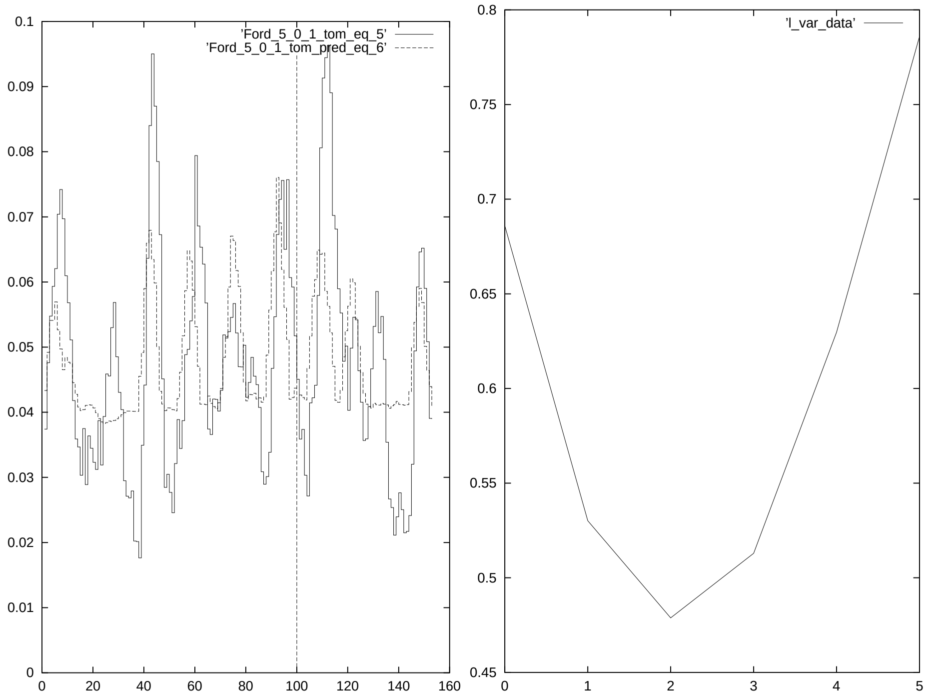

Figure 5.13: Left: No time lag; NPE$_{train}$=0.67, NPE$_{test}$=0.69. (Solid line: real data; dashed line: modelled data.) Right: NPE$_{test}$ as a function of the time lag l.

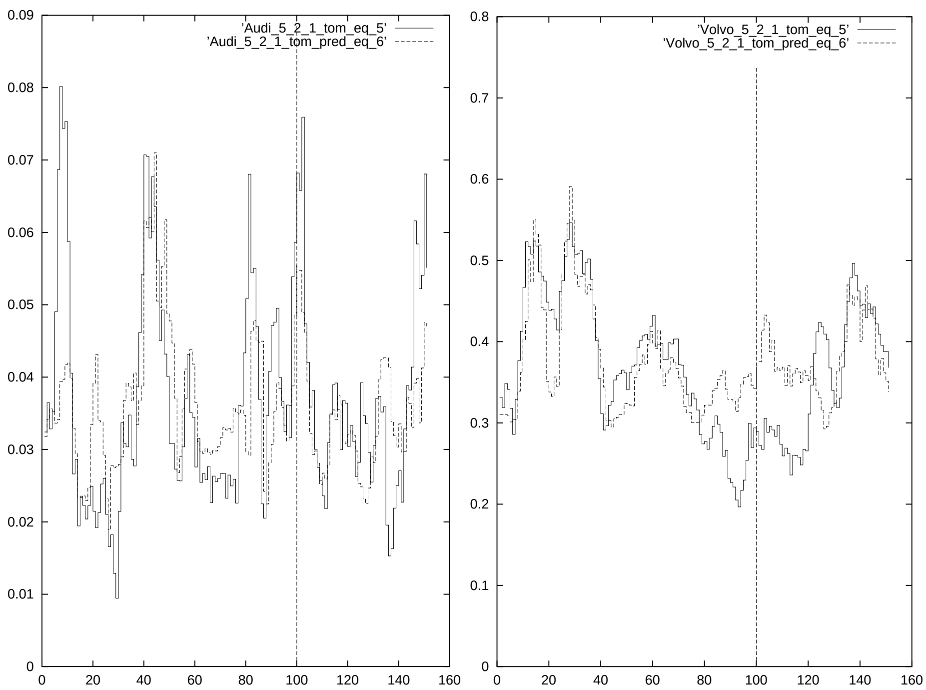

Figure 5.14: Results for Audi (left) and Volvo (right) with parameters otherwise as in the benchmark reference; Audi: NPE$_{train}$=0.60, NPE$_{test}$=0.64; Volvo: NPE$_{train}$=0.37, NPE$_{test}$=0.92. (Solid lines: real data; dashed lines: modelled data.)

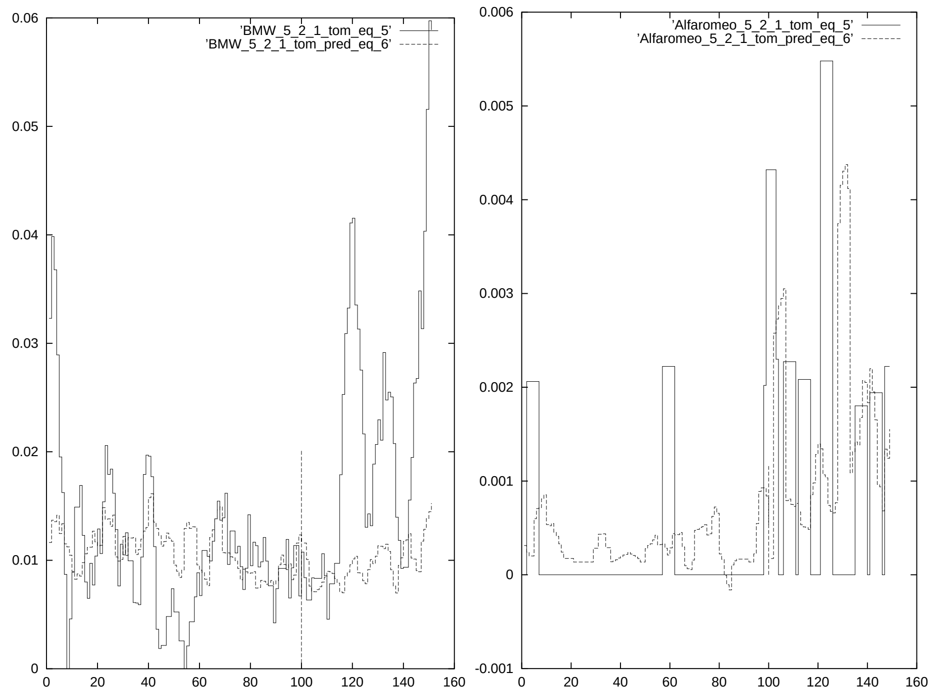

Figure 5.15: Results for BMW (left) and Alfa Romeo (right) with parameters otherwise as in the benchmark reference; BMW: NPE$_{train}$=0.90, NPE$_{test}$=1.00; Alfa Romeo: NPE$_{train}$=0.94, NPE$_{test}$=1.10. (Solid lines: real data; dashed lines: modelled data.)

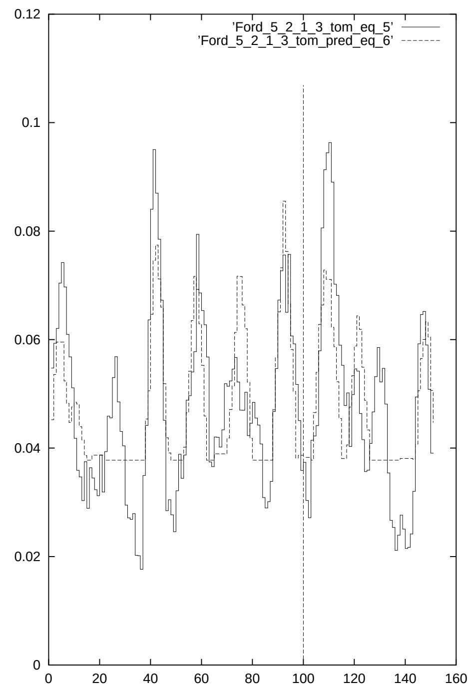

Figure 5.16: The result for Ford when using only TV as input: NPE$_{train}$=0.39, NPE$_{test}$=0.45

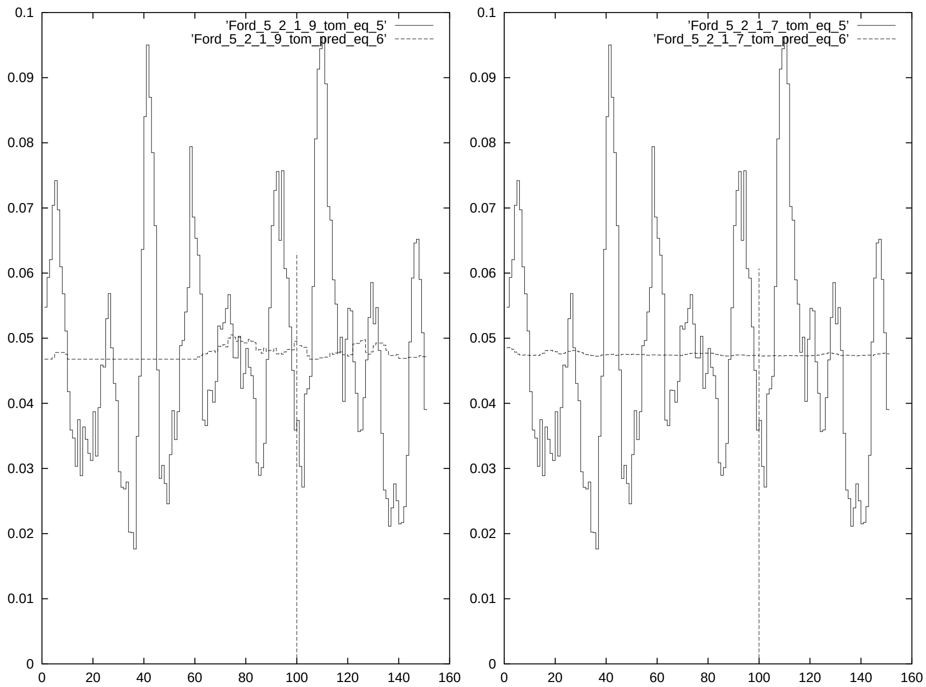

Figure 5.17: Left: The result for Ford when using only *popular press* as input: NPE_{train}=1.00, NPE_{test}=1.03. Right: The result for Ford when using only *morning press* as input; NPE_{train}=1.00, NPE_{test}=1.00 (Solid lines: real data; dashed lines: modelled data.)

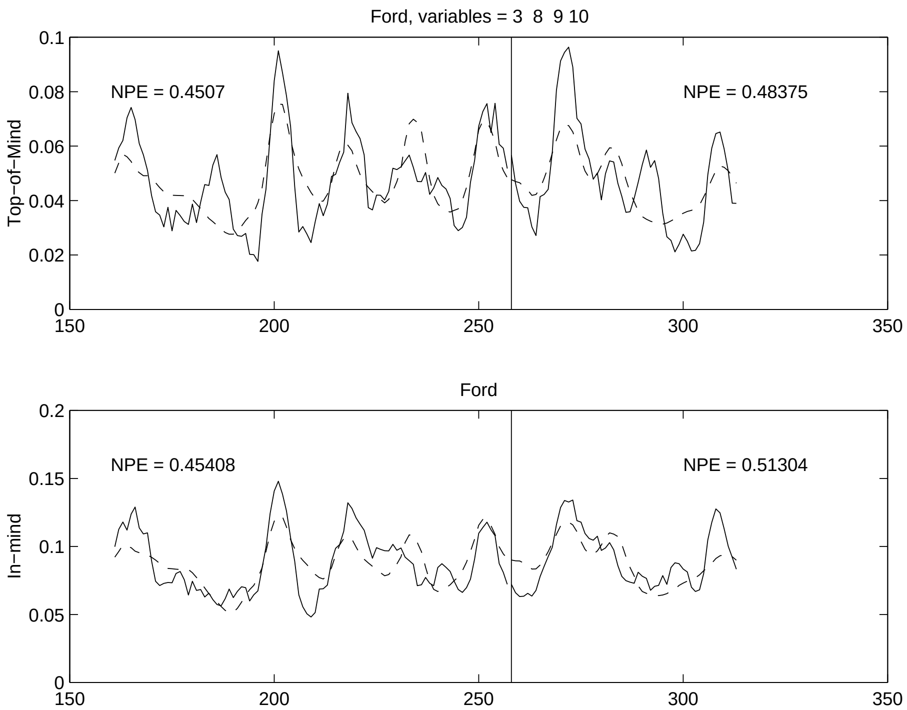

Figure 5.18: Result when using the "optimal" set of variables for Ford (optimal with respect to test set NPE). The optimal set is *TV, evening press, popular press, specialist press.* The Top of Mind NPE values are: NPE_{train} = 0.45 and NPE_{test} = 0.48. (Solid lines are true values, and dashed lines are predicted values.)

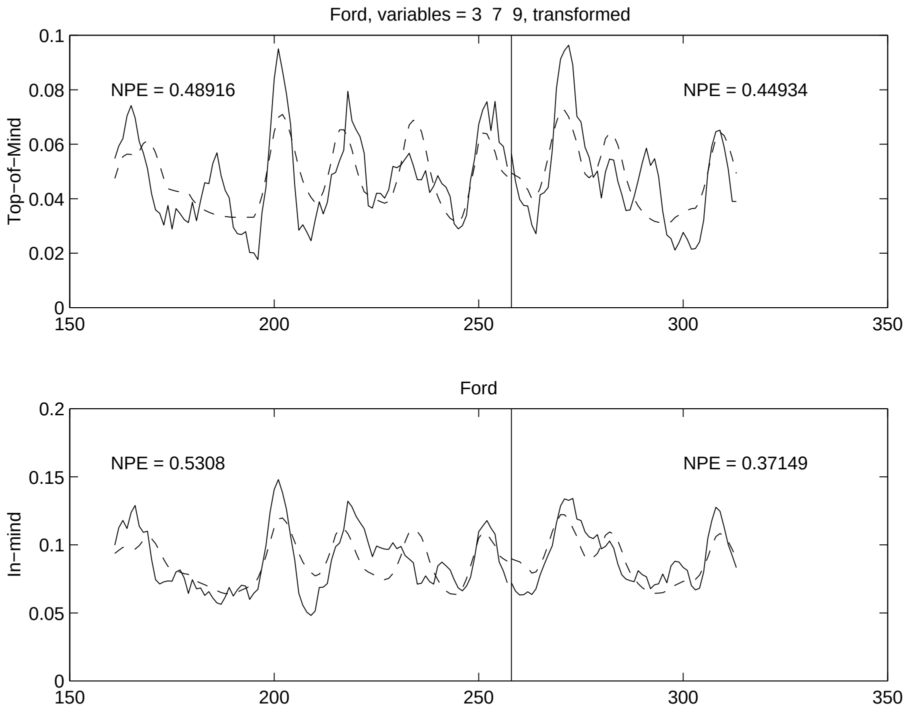

Figure 5.19: Result when using the "optimal" set of transformed variables for Ford (optimal with respect to test set NPE). The optimal set is *TV, morning press, popular press.* The Top of Mind NPE values are: $\text{NPE}_{train} = 0.45$ and $\text{NPE}_{test} = 0.37$. (Solid lines are true values, and dashed lines are predicted values.)

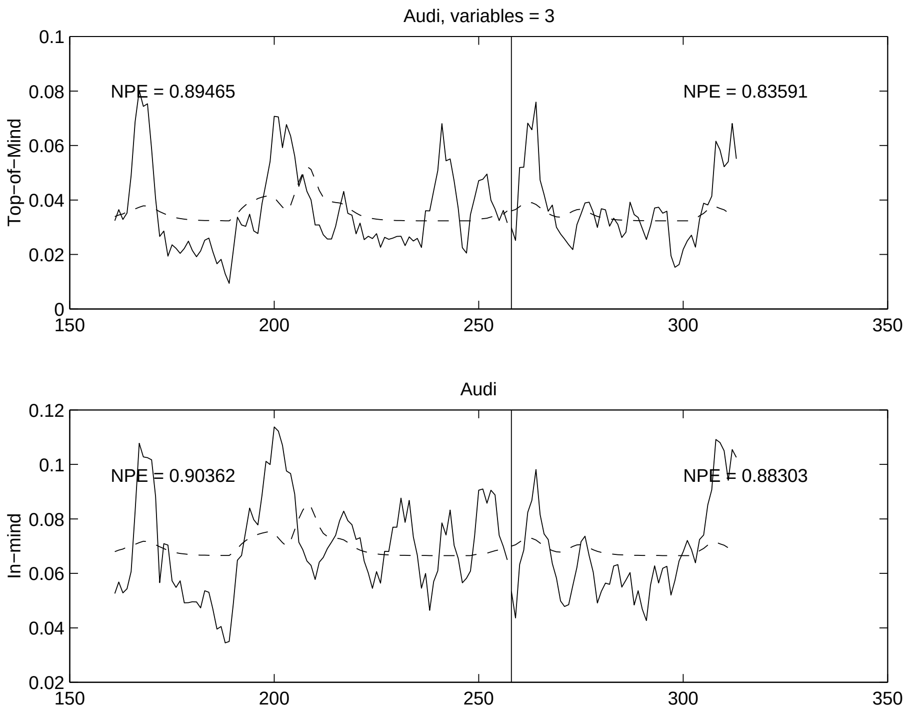

Figure 5.20: Result when using the "optimal" set of variables for Audi (optimal with respect to test set NPE). The optimal set is *TV*. The Top of Mind NPE values are: $\text{NPE}_{train} = 0.84$ and $\text{NPE}_{test} = 0.88$. (Solid lines are true values, and dashed lines are predicted values.)

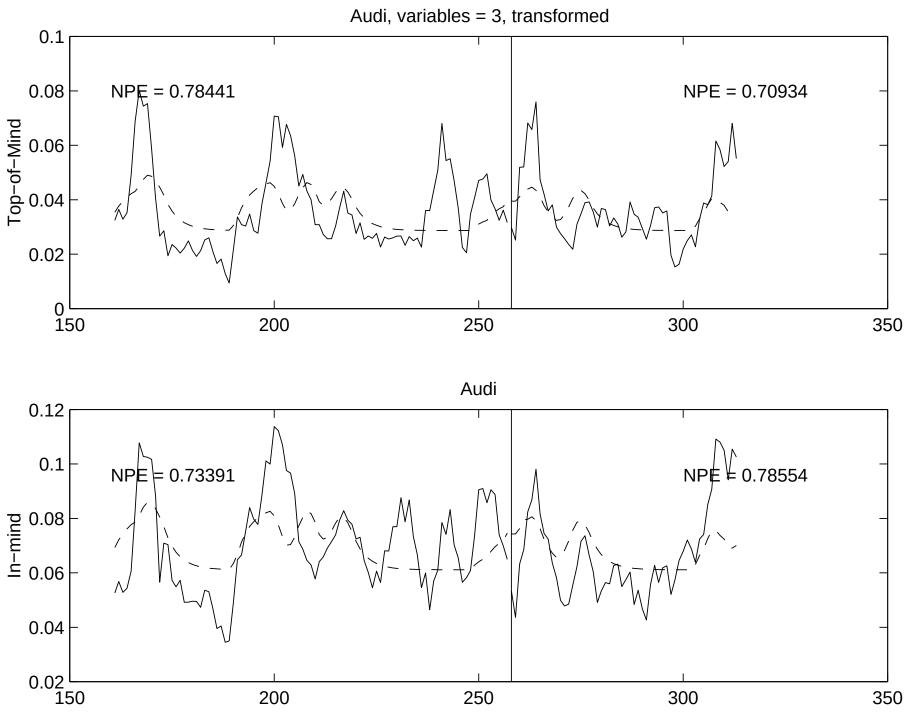

Figure 5.21: Result when using the "optimal" set of transformed variables for Ford (optimal with respect to test set NPE). The optimal set is TV. The Top of Mind NPE values are: $\text{NPE}_{train} = 0.71$ and $\text{NPE}_{test} = 0.78$. (Solid lines are true values, and dashed lines are predicted values.)

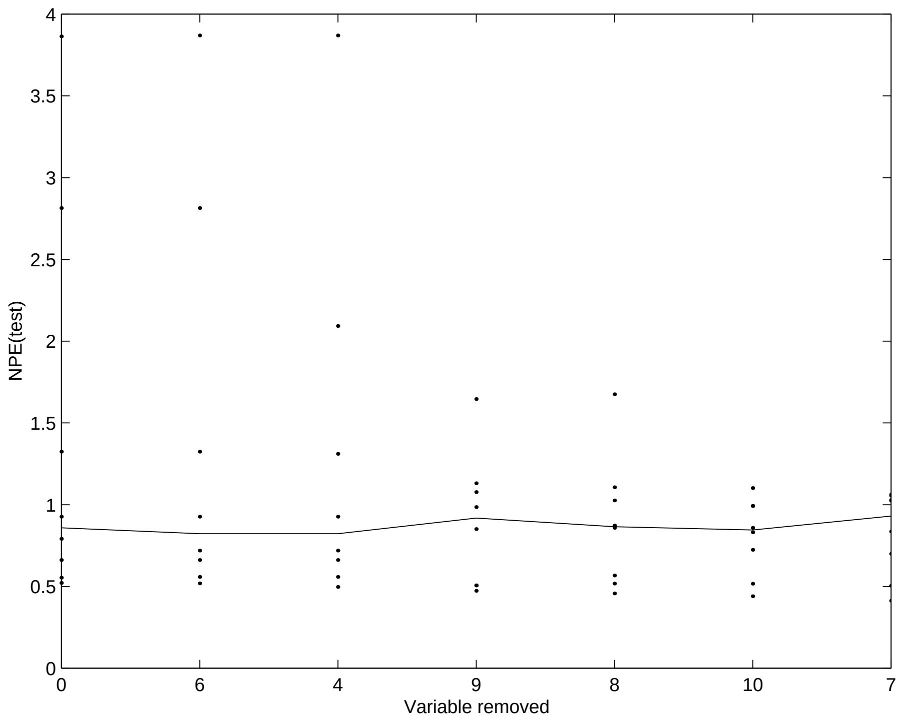

Figure 5.22: Backwards elimination for all models. The median NPE_{test} does not change much when variables are removed, but the spread in values (over brands) decreases significantly.

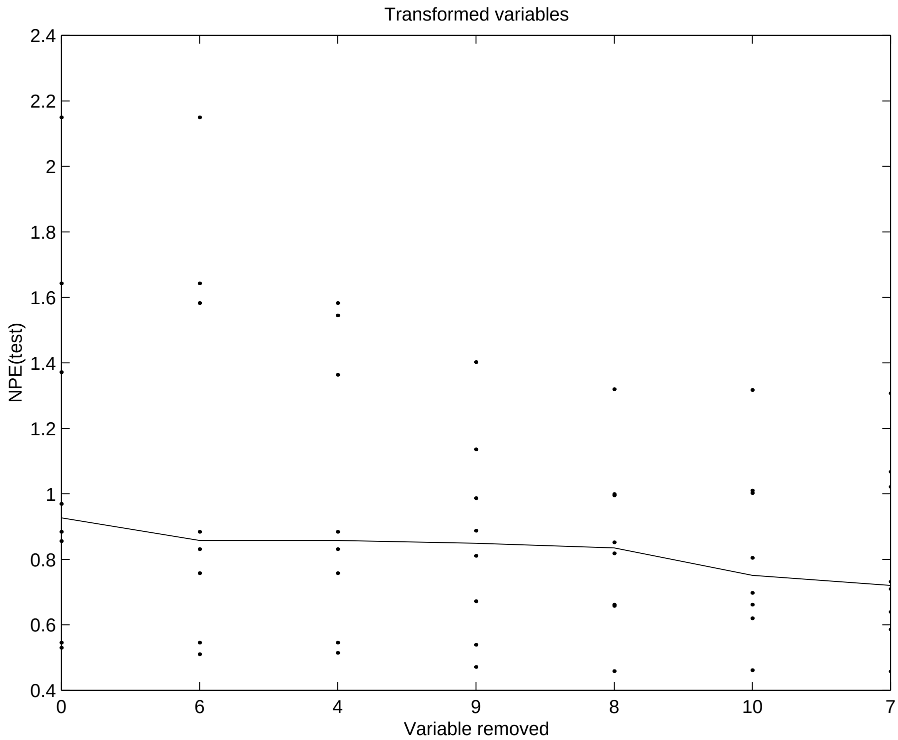

Figure 5.23: Backwards elimination for all models, when the variables have been "Box-Cox" transformed. The median NPE_{test} does decrease when variables are removed, and the spread in values (over brands) decreases significantly. However, the spread is larger in the end than what we find for untransformed variables.

5.7 Results

Anders Holst

This is a summary of the results of the different methods on a "blind test", where the correct answer had not been given to the participants in advance.

	Audi		Ford		Mazda		Nissan	
Method	ToM	IM	ToM	IM	ToM	IM	ToM	IM
SICS 1	-0.27	-0.17	0.29	0.58	0.44	0.88	0.61	0.69
SICS 2	-0.34	-0.31	0.23	0.49	0.31	0.88	0.58	0.61
DSV	0.11	-0.21	0.18	0.36	0.52	0.76	0.60	0.64
MittH	0.10	-0.17	0.36	0.49	-0.38	0.70	0.52	0.62
Skövde	-0.13	-0.10	-0.57	-0.09	0.25	0.85	0.40	0.58
NMA	—	0.05	—	0.12	—	0.20	—	0.08

	Volvo		Dressman		H&M		KappAhl	
Method	ToM	IM	ToM	IM	ToM	IM	ToM	IM
SICS 1	0.32	0.52	0.03	0.31	0.18	0.53	0.36	0.45
SICS 2	0.22	0.40	0.00	0.30	0.52	0.61	0.33	0.39
DSV	0.00	0.24	-0.11	0.34	0.07	0.20	0.28	0.29
MittH	0.64	0.40	0.04	0.30	-0.23	-0.44	0.24	0.28
Skövde	0.63	0.56	0.17	0.31	0.26	0.37	0.33	0.43
NMA	—	0.18	—	0.64	—	—	—	0.40

Table 5.1: The correlation coefficients between the predictions and the true series when the different methods are applied to the different brands.

Most correlations in the table are really low, or even negative indicating that there is a high level of noise, or being correct or wrong by "coincidence" rather than skill. When counting the number of "highest" rates each method gets, they all get a couple each, except SICS 1 which gets six best results (one of which is shared with SICS 2). Also, the method used by NMA seems to produce lower correlations than most of the other methods on many brands (except Dressman where everyone else made worse). Everyone gets a negative correlation on Audi, which may signal some change in advertising behavior. Everyone also get quite high correlation on Mazda, In Mind. Whether this is pure chance, or would be maintained in future tests is hard to tell.

In summary, the results suggest that NMA could gain by introducing some of the techniques above for their brand awareness analysis.