Analyzing the Effect of Offline Media on Online Conversion Actions

Didier Nibbering
Flavius Frasincar*
Damir Vandic

Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738, NL-3000 DR Rotterdam, the Netherlands
E-mail: {nibbering, frasincar, vandic}@ese.eur.nl
Fax: +31 (0)10 408 9162
*Corresponding author

Abstract: Recognizing the Web as one of the most popular mediums for information distribution, online advertising is nowadays a booming business. Arriving to a company website is often done by interacting with search engine advertisements. In this paper, we investigate how offline advertising by means of TV and radio influences the search engine advertisement that leads to users visiting a company marketing website (a conversion action). Our research is based on the search engine-driven conversion actions of the 2012 marketing campaign "Do Us A Flavor" of the chips manufacturer Lays, for which we experimented with several prediction models: linear regression (linear model), Support Vector Regression (non-linear model), and six distributed lag models (linear autoregressive models). Our results show that offline commercials positively influence the online marketing campaign. We have also determined that this influence is higher for TV than for radio, and that general-purpose TV channels have a higher impact on the number of conversion actions than specialized TV channels. In addition, we observed that TV advertisements have the highest influence on conversion actions in the first 50 minutes after the advertisement broadcasting.

Keywords: e-commerce; offline media; online conversion actions; marketing campaign.

Reference to this paper should be made as follows: Nibbering, D., Frasincar, F., and Vandic, D. (xxxx) 'Analyzing the Effect of Offline Media on Online Conversion Actions', *Int. J. Web Engineering and Technology*, Vol. x, No. x, pp. xxx–xxx.

Biographical notes: Didier Nibbering is a PhD candidate at the Econometric Institute of Erasmus University Rotterdam and holds a cum laude master degree in econometrics and operations research. His main research interest is in the fields of time series econometrics and macroeconomic forecasting. His work is presented at major international conferences, such as at meetings from the Econometric Society and the International Association for Applied Econometrics.

Flavius Frasincar obtained the master degree in computer science from Politehnica University Bucharest, Romania, in 1998. In 2000,

he received the professional doctorate degree in software engineering from Eindhoven University of Technology, the Netherlands. He got the PhD degree in computer science from Eindhoven University of Technology, the Netherlands, in 2005. Since 2005, he is assistant professor in information systems at Erasmus University Rotterdam, the Netherlands. He has published in numerous conferences and journals in the areas of databases, Web information systems, personalization, and the Semantic Web. He is a member of the editorial boards of the International Journal of Web Engineering and Technology, and Decision Support Systems.

Damir Vandic received cum laude the master degree in Economics & Informatics from Erasmus University Rotterdam in 2010, specializing in computational economics. For his PhD research, he has obtained the NWO Mosaic grant and is currently a PhD candidate at the same university. The focus of his research is on using the Semantic Web to improve the searching and browsing of products on the Internet. His research interests cover areas such as machine learning, the Semantic Web foundations and applications, knowledge systems, and Web information systems.

1 Introduction

The last few years online advertising has grown rapidly. The Dutch online advertising market was 619 million euro in the first half year of 2013, a year on year uplift of 5.8% (van Rijsewijk et al., 2013). Online advertising are expected to continue to grow at a rate of 6%. The rise of social media networking and the possibility to advertise on websites like Google have created new marketing opportunities for companies (Rishika et al., 2013; Goh et al., 2013). Managers believe that the use of online media costs less than offline media (Barwise & Farley, 2005). Another advantage of online advertising is the possibility of targeting at specialized audience segments (Zeff & Aronson, 1999; Gopal et al., 2011). Nevertheless, the main advantage is perhaps the easiness to track the results of online advertising campaigns. Companies could save and interpret all the data from an online campaign in order to improve the effectiveness and efficiency of the marketing investments.

Besides the advantages of online media, we know that no website can compete with the sheer size and usage volume of the TV audience. An average American family watches 60 hours of television a week. Almost every household has a television with more than 75% having multiple televisions (Baran, 2010). Probably for this reason, we still have offline commercials, both on radio and television, and almost all these offline commercials try to attract their audience to a website, or at least mention the advertiser's website. As a result, offline advertising increases website visits by influencing consumer awareness (Winer & Ilfeld, 2002; Naik & Peters, 2009; Liaukonyte et al., 2015).

A marketing company has facilitated a campaign from the potato chips manufacturer Lays, which consists of an online and offline part. Both parts have the goal that as many people as possible go to the campaign website of Lays. Television commercials about the Lays website incite viewers to use a search engine to search for the campaign, which makes offline advertisement viewers more likely to end up at the campaign website. In this paper we aim to measure this effect in the Lays campaign. Did the commercials actually influenced the number of people visiting the campaign website? The answer to this question gives a marketing company the ability to evaluate a campaign which consists of an online and offline part. When the results of the evaluation are used for setting up new marketing campaigns, the company could achieve a more efficient result by the appropriate use of the offline media resources.

The outline of this paper is as follows. Section 2 discusses related work in the sales-advertising relationships. Section 3 presents the details of the advertisement campaign we study in this paper and formulates the research question. In Section 4 we introduce the data and in Section 5 we show how we answer the research question by means of this data. Finally, Section 6 presents and interprets the results and Section 7 concludes the paper and suggests future work.

2 Related Work

Although there is a lot of research in the sales-advertising relationships, only more recently online advertising increases in popularity as research topic. Kireyev et al. (2015) introduce a multivariate time series model to examine the relation between search engine advertisements and online display ads, such as banners, for a bank. They find that the display adds not only increase sales, but also positively affect the number of people that visit the website of the bank via a search engine.

Dinner et al. (2014) do not only compare online advertising methods to each other, but also make a comparison to offline advertising. In terms of sales, they conclude that search engine advertising is more effective than offline advertising. This raises the question whether we can use the traditional advertisement methods to increase the conversions of search engine advertising. Liaukonyte et al. (2015) show that television advertising increases website traffic and sales.

To examine the relation between the offline advertisements and the online website traffic, we rely on methods applied in research to traditional sales-advertising relationships. Rao (1972) investigated several econometric models (three OLS-models and two Koyck models) on the relationship between advertising and cigarettes sales and tested these models on in-sample and out-of-sample performance. A Koyck model, in which a first order autoregressive series were assumed for the errors, turned out to be the best performing model. When we consider a visit to the campaign website as a sales and we define advertising as the radio and television commercials, it is clear that it could be appropriate to use a (variation of) Koyck model to analyze the relationship between website visits and offline media.

Bass & Clarke (1972) used a variety of estimation and testing procedures on six different distributed lag models, with an example of sales and advertising data for a dietary weight control product. They demonstrated that models of the dynamics of sales and advertising need not be limited to the restrictive Koyck model. Non-monotonic lag distributions appeared in the case of the dietary weight control product to be more consistent with the evidence than the Koyck model. Moreover, they found that marketing studies which posit distributed lag effects, would be more appropriately estimated by maximum likelihood than by OLS.

The duration of the advertising effect on sales has also been investigated by (Clarke, 1976). With the help of a survey of the econometric literature, he determines the duration of cumulative advertising effect on sales. He found that distributed lag models quite broadly obtain statistically significant results. The models may not be perfect, but careful inspection of these shows that the results of 70 studies are in very close agreement on the duration of the cumulative effect of advertising. Another important result of his research entails the fact that the published econometric literature indicates that 90% of the cumulative effect of advertising on sales of mature, frequently purchased, low-priced products occurs within three to nine months of the advertisement. We hypothesize that the relationship between commercials and website visitation is established in a much smaller time interval.

Winer & Ilfeld (2002) demonstrate that offline advertising could increase website visitation. They empirically determine the factors that drive traffic in the Internet space. Offline advertising is one of these factors. Naik & Peters (2009) finds positive effects of offline media on online spendings and Liaukonyte et al. (2015) also argues that television advertising influences online shopping. However, the relationship between the offline part of an advertising campaign and the online website visitation has never been investigated in an econometric way. In this research we use a large dataset of a marketing company, with information about contacts with the website and the data of the offline and online campaigns. Our research hypothesis is that offline media influence the online website visitations.

3 The Advertising Campaign "Do Us A Flavor"

The potato chips manufacturer Lays faced a huge drop in market share in 2009 because supermarkets were expanding their assortment of pretzels from house brands. To stop the loss of market share Lays needed to distinguish from the house brands, which have a similar assortment at lower prices. Lays can differentiate from other brands at product level by offering more flavor variety. At brand level Lays had to create preferences for her brand.

Lays started in 2010 with an advertising campaign called "Maak de Smaak". In English, this is literally translated to "Make the flavor", but in the English campaigns the slogan "Do Us A Flavor" was used. Throughout this paper, we will refer to this campaign with "Do Us A Flavor". The main goal of the campaign was to create brand preference by involving consumers in the product development. Therefore Lays asked the Dutch people to devise a potato chips flavor for the new Lays Limited Edition potato chips, to purchase the trial packs of the final flavors and to vote on their favorite flavor.

In the first phase of the campaign as many Dutch people as possible were encouraged to send new flavors and a brief motivation to the website `http://www.lays.nl`. Television commercials provided coverage of a large target group. In the second phase a jury, led by the famous Dutch cook Joop Braakhekke, selected three finalists. Thereafter Lays produced the three final potato chips flavors. By means of an advertisement in the newspapers, all the participants were thanked for their flavor submissions.

In the third phase the finalists held a campaign for their own flavor. They called the Dutch people to vote on their potato chips flavor. Lays paid attention to this battle with television commercials, online advertising, mobile advertising, and outdoor events. The fact that the action packages of the final flavors were actually for sale in the shops, completed the experience.

Consumers could really buy and taste the final flavors, before they voted on the website of Lays. The interactive elements on the website and the "Do Us A Flavor" fan pages on online social media such as Facebook motivated people to join the campaign. People could share photos and videos with friends to promote their favorite flavor and to generate more votes. In the last phase the winner was honored. The favorite flavor of the Netherlands, "Patatje Joppie" (tastes like French fries with a Dutch sauce), is in the stores for one year now. The creator of the winning flavor got a reward of 25,000 euro plus 1% of the turnover of the new flavor.

The "Do Us A Flavor" campaign in 2010 was a great success. There were more than 675.000 flavor submissions and there were 6 million bags with potato chips of the three final flavors sold. Therefore the campaign started again on January 16, 2012. This year people had more influence on which three potato chips flavors have been sold in the shops in August.

Everyone who had sent a flavor to the website automatically saw two other submitted flavors in a Battle. In this Battle you had to vote on the flavor which you think is the tastiest. It was also possible to play a Battle without sending in your own flavor. All the submitted flavors have been discussed in the online Battles. Battles were randomly chosen and each flavor joined a Battle an equal number of times. After the online Battles, a Top-1000 of flavors was established. A jury selected her Top-8 from this Top-1000. Then the Dutch people chose two finalists and the jury added a third finalist. For all the other aspects the campaign was nearly the same as in 2010 (Pepsico Nederland, 2014). The flavors "Aioli e Aglio", "Whiskey Cocktail", and "Spicy Reggae Chicken" are in the stores at this moment.

3.1 Research Question

The research question investigated in this paper in relation to the previously presented case is: Do the offline media affect the online conversion actions? With the number of conversion actions we mean the number of people visiting the campaign website. We use econometric models to answer this question.

First we describe the dataset and determine which data we use in our experiments. We need the data to make a proper model, which can explain the number of conversion actions that take place at a given time. This model should allow us to assess the effects of various campaign factors on the number of conversion actions in a given time period. In this way we quantify the effects of the offline media on the number of contacts with the campaign website.

After we have defined the variables we first introduce the performance measures for the models and then we start with a simple econometric model: The Ordinary Least Squares (OLS) linear regression model. In this model we can easily estimate the factors which affect the number of conversion actions. So we can deduce a direct relationship between offline and online media. We also use this model to estimate how long a commercial affects the number of contacts with the campaign website.

We also experiment with a Support Vector Regression (SVR) model as a non-linear alternative to the linear regression model. This model is known for its excellent performance in various fields (Collobert & Bengio, 2001), and it has in particular some successful applications in different problems of time series prediction (Lu et al., 2009). We estimate both models on a training set, and then we use the estimated models to forecast on a test set. We use a number of measures, such as the accuracy, efficiency, and unbiasedness of the predictions, to evaluate the out-of-sample performance of the models. We examine whether models which explain the online conversion actions with the aid of the information about the offline media significantly outperform a model in which the offline media are not included. Then we include lags of the offline media variables in models to investigate how long television and radio commercials affect the number of people who are going to the website.

The Koyck model is often used to establish the dynamic relationship between sales and advertising (Franses & Oest, 2004). When we consider a conversion action as a sales and we define radio and television commercials as advertising, it is clear that we could use the Koyck model to analyze the lagged offline media variables more precisely. We test the Koyck model and five alternative distributed lag models on out-of-sample performance. We use the best performing model to estimate the effects of the offline media.

4 Data

In this section we describe the data. First we explain how the data of the campaign is recorded. Then we provide the definitions of the used data variables.

We can distinguish two types of online contacts with the campaign: search engine advertising (SEA) and display advertising. SEA entails that advertisers pay a fee to Internet search engines to be displayed alongside organic web search results (Ghose & Yang, 2009). With display advertising we refer to online banners: plain content-targeted text ads and banner ads including visual features (Goldfarb & Tucker, 2011). An online contact with advertisement, either SEA or a banner, can lead to a conversion action; an actual visit to the website of the campaign.

The online data of the campaign is recorded in conversion paths. These paths consist of the online contacts of the campaign with an individual (by for example banners or advertisement in search engines) before proceeding to a conversion action. There are five different conversion actions:

- Level 1: Go to the homepage of the website;
- Level 2: Start a Battle on the website;
- Level 3: Receive the results of the Battle;
- Level 4: Send a flavor;

- Level 5: Receive a confirmation of the submission;

The higher the level of a conversion, the more value is assigned to this conversion. Figure 1 shows three conversion paths in which the balls represent the online contacts via SEA and display advertising, and the black ones are conversion actions. For each conversion action, the date and time are stored.

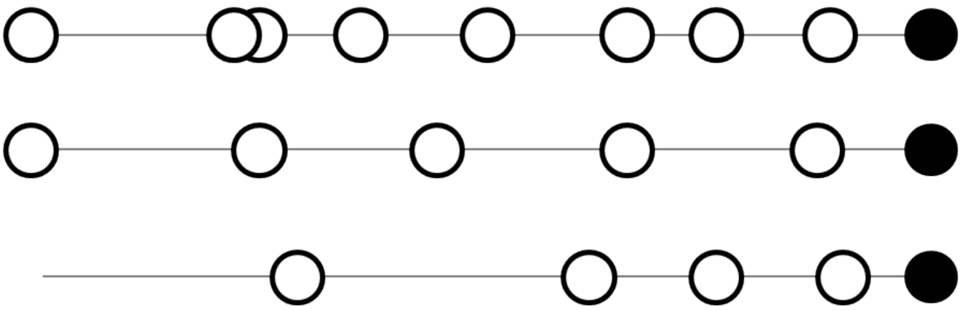

Figure 1 An example of conversion paths of three different individuals. The white balls represent online contacts with search engine advertisement or display advertising over time, and the black balls the resulting conversion action.

We know from each online contact the date and the time it took place and whether it was a SEA or a display contact. A conversion path is linked to multiple conversion actions when a path results in multiple conversion actions. This leads to the formation of identical paths. In the dataset these paths are combined into one unique path with the highest reached conversion level as final conversion action. For each conversion path, up to ten contacts are recorded. The contacts which took place more than 90 days before the conversion action and the contacts which took place more than 30 days before the previous contact, are not included in the conversion paths

The dataset consists of 600,000 unique conversion paths. These paths include 36% of the submitted flavors and 39% of the Battles. The remaining submissions and Battles are from people who have gone directly to the website or through email. They did not have online contacts before they started their activity on the website of Lays and therefore they have no conversion path. The offline data of the campaign consists of information about the television and radio commercials. From every commercial we know at what date and time it was broadcasted, on which channel, which program is broadcasted before and after the commercial, and the audience and listener ratings of the commercial.

4.1 Samples

We have obtained data from 12 January 2012 to 12 March 2012 on the "'Do Us A Flavor" campaign. We deal with a number of outliers in the first week of the campaign, caused by a few events in the campaign. On Sunday 15 January attention is paid to the start of the campaign in the television program "Life4You". On Wednesday 18 January there was a big advertisement in the newspaper "NRC

Handelsblad" and on Thursday 19 January listeners of the "Coen en Sander Show" on radio station "3FM" are told about the campaign. Because we want to measure the effect of the radio and television commercials on the number of conversion actions and we do not know what the effect of these events are on the target group, we decided to take 20 January 2012 as the starting point of the sample to be used for our models. Figure 2 shows the number of conversion actions (y_sea) and the gross rating points of radio (grp_radio) and television commercials (grp_tv) plotted against time. The television commercials are broadcasted from 16 January to 19 February. We have two periods with radio commercials, from 23 January to 29 January and from 13 February to 19 February. We choose 20 February as the end of the sample. So we investigate a period where we have television commercials each day and two periods with radio commercials. To evaluate the performance of our models we have to test the models on a test sample. Our sample consists of 32 days. We take the first 27 days as a training sample and the last 5 days as a test sample.

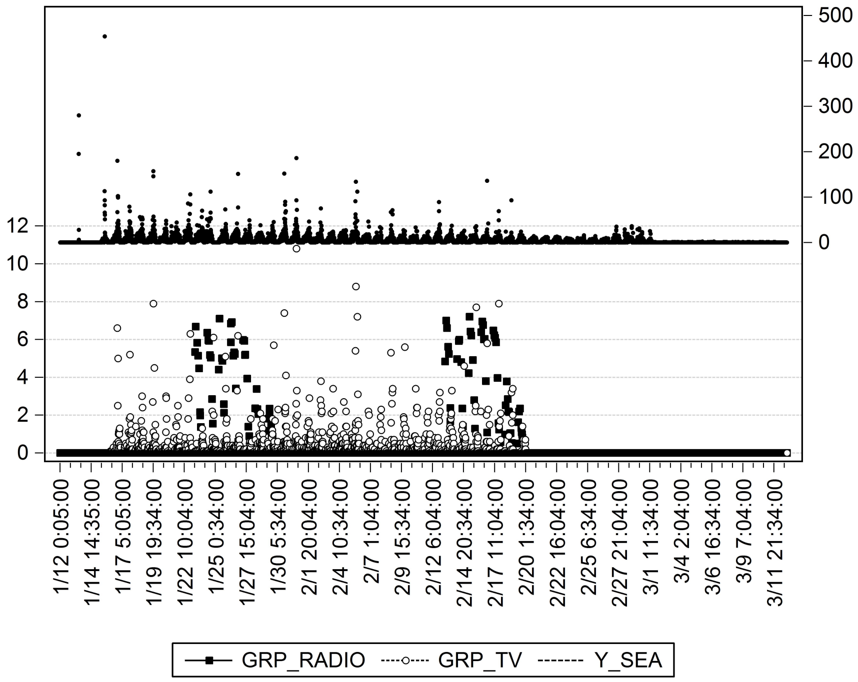

Figure 2 The upper panel shows the conversion actions over time, and the lower panel shows over the same time period the gross rating points of the radio (black squares) and television (white circles) commercials.

4.2 Variables

We want to explain the effect of the number of radio and tv commercials on the number of online conversion actions. To do this properly we have to identify the explanatory variable(s). Since offline commercials can induce viewers to search for the campaign website but have no influence on banner views, we take the SEA conversion actions as the dependent variable, i.e., we consider only the people who came in contact with the campaign website only by SEA contacts.

To create the dependent variable *y_sea*, we divide the sample period in periods of five minutes. For each five minutes we store how many online conversion actions took place without distinguishing conversion levels. So each row in the dependent variable represents the number of conversion actions in five minutes. The number of rows is equal to the number of minutes in the considered time frame divided by five.

As explanatory variable we take the *time*, because it is plausible that people are more likely to use the Internet at certain times of the day. We use a variable which assigns a one to the first five minutes of the day, a two to the second five minutes and counts this way up to 288. With the start of a new day the variable starts again at one and counts up to 288. In addition to *time*, we use other variables to indicate temporal information. We create a dummy variable for each hour of the day, $hour_i$. The dummy variable of hour i has value one when the time interval of five minutes takes place in hour i and otherwise it is equal to zero.

We use two different variables for the radio and television commercials to explain the number of conversion actions. First we take the explanatory variable *nr_radio*, the number of radio commercials which took place in the five minutes interval we are considering. Because it is possible that commercials which are broadcasted some time ago also have influence on the conversion actions in the five minutes we are considering, we have to include lags of this variable.

For the other variable, *grp_radio*, we are not using the number of radio commercials in one period but a measurement of the size of the reached audience, the gross rating points (GRP), in five minutes. For this variable we include the lags as well. We apply the same principles, as we described above, to the variables for the television commercials: *nr_tv* and *grp_tv*.

To distinguish the differences in the influence of commercials on the number of conversion actions for different channels, we include dummy variables in the models. The television commercials are broadcasted on twelve different channels: Animal Planet, Comedy Central, Discovery, National Geographic, Net5, RTL4, RTL5, RTL7, RTL8, SBS6, TLC, and Veronica. The variable ch_i takes a value of one when a television commercial took place on channel i and a value of zero otherwise. Just as with the variables *nr_tv* and *grp_tv*, we add the lags of the dummy variables of the television channels. The radio commercials are only broadcasted on one channel, "Radio 538". With the described dependent variable and explanatory variables, we get the following model:

$$\begin{aligned} y\_sea_t = f(&time_t, hour_t, nr\_radio_t,\\ &grp\_radio_t, nr\_tv_t, grp\_tv_t, ch_{ap,t}, \ldots,\\ &ch_{vn,t}, nr\_radio_{t-1}, grp\_radio_{t-1}, nr\_tv_{t-1},\\ &grp\_tv_{t-1}, ch_{ap,t-1}, ..., ch_{vn,t-1}, \ldots, \end{aligned}$$

$$nr\_radio_{t-k}, grp\_radio_{t-k}, nr\_tv_{t-k},$$
$$grp\_tv_{t-k}, ch_{ap,t-k}, \ldots, ch_{vn,t-k}).$$

where $f(.)$ is a function that has to be learned and k is the number of included lags.

5 Methods

In this section we first provide the measures we use to evaluate the out-of-sample performance of the models. Then we explain how we use the OLS-models and the SVR-models to estimate the effects of the offline media. We explain the method of Ordinary Least Squares and the method of Support Vector Regression. Last, we describe the distributed lag models. All models have been implemented and evaluated in Matlab.

5.1 Forecasting

In this section we explain how we investigate the out-of-sample performances of the models by bias and efficiency and how we compare different models by their relative predictive accuracy (Franses, 1998). The out-of-sample forecasting *accuracy* of the models will be measured by the mean absolute error (MAE):

$$MAE = \left[\frac{1}{n} \sum_{i=1}^{n} |y_i - y_{f\,i}| \right] \tag{1}$$

where y is an $n \times 1$ vector with the observations, y_f is an $n \times 1$ vector with the predictions, and n is the number of observations. We estimate our models on the training set, then we use the estimated models to predict the target values for the test set. The errors have to be as small as possible so we have to find a function that can accurately approximate the target values. Besides measuring the deviations of the predictions to the target values we also evaluate the reliability of the forecasts. First we test wether the mean of the forecast errors differ significantly from zero with the *unbiasedness* test:

$$\frac{\overline{e}}{\frac{s_e}{\sqrt{n}}} \sim t\,(n-1) \tag{2}$$

where n is the number of predictions, $\overline{e}$ the mean of the differences between the target values y and the predicted values y_f, and s_e the standard deviation:

$$s_e = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \overline{e})^2} \tag{3}$$

It is desirable to have forecasts that are unbiased, that is the average forecast error $\overline{e}$ should be close to zero. If this is not the case, the model under- or overestimates the mean of the target values. For the *efficiency* test we use the Mincer-Zarnowitz regression:

$$e_i = \alpha_1 + \alpha_2 \widehat{y}_i + \eta_i \tag{4}$$

where e_i is the difference between the target value, y_i, and the predicted value, $\widehat{y}_i$, and η_i the error. It should hold that α_1 and α_2 are equal to zero. This means that it should not be possible to forecast the forecast error itself with any information available at the moment of forecasting. We could also compare the predictive accuracy of models by the Diebold-Mariano test (DM-test)(Diebold & Mariano, 1995):

$$d_i = \left(y_i - y^1_{f\,i}\right)^2 - \left(y_i - y^2_{f\,i}\right)^2, \overline{d} = \frac{1}{n}\sum_{i=1}^{n} d \tag{5}$$

$$D = \frac{\overline{d}}{\sqrt{\frac{Var(d)}{n}}} \approx N\left(0, 1\right) \tag{6}$$

here $y^1_{f\,i}$ are the values predicted by model one and $y^2_{f\,i}$ are the predicted values of model two.

5.2 Variable selection

First we examine whether a model in which the offline media are included as explanatory variables significantly outperforms a model in which the offline media are not included. We estimate a linear regression model on the training set with only the variables *time* and $hour_i$, $i \in 2, ..., 24$. We use this estimated model to predict the target values of the test sample and then we calculate the MAE.

Thereafter we estimate the same model on the same training set but now we add the variables *nr_radio*, *grp_radio*, *nr_tv*, *grp_tv* and the dummies ch_i, $i \in$ {Comedy Central, Discovery, National Geographic, Net5, RTL4, RTL5, RTL7, RTL8, SBS6, TLC, Veronica}. We estimate this model iteratively and in each iteration we include an additional lag of the variables which we described above. We use each estimated model to predict the target values on the test sample. We calculate the MAE of these predictions and investigate the MAE trend. We use the model which has the best performance according to these criteria, to analyze the effect of the offline media.

When the predictions of the models in which the offline media are included as explanatory variables are significantly better than the model in which these variables are not included, we answer the main question positively. If this is the case, we consider the following question: For how long does the offline media affect the online conversion actions? We look at the model which has the best score according to the MAE. In other words, how many lags of the offline media variables are included in the best-performing model? The number of lags in this model gives a good indication of how long the offline media still influence the online conversion actions.

5.3 Ordinary Least Squares

We start with a linear regression model and the most popular method in regression analysis: Ordinary Least Squares. This method is based on the following equation:

$$y = X\beta + \varepsilon \tag{7}$$

Here y is an $n \times 1$ vector with the target values, in our case the number of conversions actions, X an $n \times k$ matrix with k explanatory variables and n observations, β a $k \times 1$ vector of unknown parameters and ε is an $n \times 1$ vector of unobserved disturbances. The OLS-method estimates the $k \times 1$ vector of estimates of β, b, by minimizing $S(b) = (y - Xb)'(y - Xb)$ (Heij et al., 2004). $S(b)$ is continuously differentiable, and minimizing $S(b)$ is an unconstrained and convex minimization problem, such that b can be calculated by differentiating S(b) with respect to b and equating the differential to zero:

$$\frac{\delta S}{\delta b} = X'(Xb - y) \tag{8}$$

Solving this for b, we obtain:

$$b = (X'X)^{-1}X'y \tag{9}$$

5.4 Support Vector Regression

Besides a linear regression model we also estimate a non-linear regression model. Vapnik (1999) has proposed a method to solve regression problems using Support Vector Machines (SVM). It has yielded excellent performance on many regression and time series prediction problems (Collobert & Bengio, 2001). Therefore we use the method of SVM in addition to the OLS-models. We implemented this method with the software package libsvm (Chang & Lin, 2011). We use the ε-SVR with a linear kernel function and we set the parameters cost and epsilon at 1 and 0.001, respectively. We use the formulation of Vapnik for the standard form of Support Vector Regression, ε-SVR (Vapnik, 1998).

Consider a set of training points, $\{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_k, y_k)\}$, where $\mathbf{x}_i \in \mathbb{R}^n$ is a feature vector and $y_i \in \mathbb{R}^1$ is the target output. Under given parameters $C > 0$ and $\varepsilon > 0$, the standard form of support vector regression is:

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{k}\xi_i + C\sum_{i=1}^{k}\xi_i^* \tag{10}$$

subject to

$$\begin{aligned} \mathbf{w}^T\phi(\mathbf{x}_i) + b - z_i &\leqslant \varepsilon + \xi_i, \\ z_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b &\leqslant \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\geqslant 0, i = 1, ..., k. \end{aligned}$$

Equation 10 shows the primal optimization problem where $\phi(\mathbf{x}_i)$ maps $\mathbf{x}_i$ into a higher-dimensional space and $C > 0$ is the regularization parameter.

5.5 Distributed Lag Models

First we introduce notation for the distributed lag models: y_sea_t is the number of conversion actions in time period t, c a constant value, T an 1×24 vector with the explanatory variables which indicate the time (*time* and $hour_i$, $i \in (2, ..., 24)$) in time period t, and O an 1×15 vector with the offline media explanatory variables, (nr_radio, grp_radio, nr_tv, $grp\_tv \times ch_i$ with the dummies ch_i, $i \in$ (Animal Planet, Comedy Central, Discovery, National Geographic, Net5, RTL4, RTL5, RTL7, RTL8, SBS6, TLC, Veronica)) in time period t. β_1, β_2, and β_3 are respectively 1×24, 1×15, and 1×15 vectors of unknown parameters. λ_i, $i \in (1, ..., 4)$ is an unknown parameter value and ε_t is the error term in time period t.

With this notation we can write a simple linear regression model, as described in Equation 7, as a direct lag model. The lagged variables are included as explanatory variables in this model:

$$y\_sea_t = c + T_t\beta_1 + \sum_{j=1}^{l} O_{t-j}\beta_{j+1} + \varepsilon_t \tag{11}$$

The problem of this simple linear model is the necessity of specifying l, the number of lags. For any finite l a truncation bias results (Clarke, 1976). Therefore we consider geometric distributed lag models. These models make the current number of conversion actions a function of current and past offline media levels, where the lag coefficients have a geometrically decaying pattern (Franses & Oest, 2004). So these models involve an infinite number of lagged variables. The most commonly used distributed lag model is the Koyck model (Koyck, 1954). *Model I* is:

$$y\_sea_t = c + T_t\beta_1 + \lambda_1 y\_sea_{t-1} + O_t\beta_2 + \varepsilon_t + \lambda_2\varepsilon_{t-1} \tag{12}$$

While this model has been the most popular model out of the general class of distributed lag models, it is by no means the only reasonable model of a distributed lag effect. Bass and Clarke tested six different distributed lag models (Bass & Clarke, 1972). These models have the form of an ARMAX model (Bass & Clarke, 1972). We estimate these six models without the condition that the lambda's in the autoregressive part equal the lambda's in the moving average part of the ARMAX models. We use the Matlab function “garchfit” which fits the model specifications by maximum likelihood. Thereafter, we test the six models on out-of-sample performance and use the best performing model to estimate the effects of the offline media.

First, we estimate the Koyck model from Equation 12. An alternative to the Koyck model is one in which the geometric decay in the effect of the offline media on the online conversion actions does not begin with the initial period but is delayed until period two. *Model II* is:

$$\begin{aligned} y\_sea_t = c + T_t\beta_1 + \lambda_1 y\_sea_{t-1} \\ + O_t\beta_2 + O_{t-1}\beta_3 + \varepsilon_t + \lambda_2\varepsilon_{t-1} \end{aligned} \tag{13}$$

In *Model III* we extend the Koyck model to a second-order lag model:

$$\begin{aligned} y\_sea_t = {} & c + T_t\beta_1 \\ & + \lambda_1 y\_sea_{t-1} + \lambda_2 y\_sea_{t-2} + O_t\beta_2 \\ & + \varepsilon_t + \lambda_3\varepsilon_{t-1} + \lambda_4\varepsilon_{t-2} \end{aligned} \tag{14}$$

Model IV shows the second-order lag function with a period lag in O:

$$\begin{aligned} y\_sea_t = {} & c + T_t\beta_1 \\ & + \lambda_1 y\_sea_{t-1} + \lambda_2 y\_sea_{t-2} + O_t\beta_2 + O_{t-1}\beta_3 \\ & + \varepsilon_t + \lambda_3\varepsilon_{t-1} + \lambda_4\varepsilon_{t-2} \end{aligned} \tag{15}$$

In *Model V* we extend the Koyck model to a third-order lag model:

$$\begin{aligned} y\_sea_t = {} & c + T_t\beta_1 \\ & + \lambda_1 y\_sea_{t-1} + \lambda_2 y\_sea_{t-2} + \lambda_3 y\_sea_{t-3} \\ & + O_t\beta_2 + \varepsilon_t + \lambda_4\varepsilon_{t-1} + \lambda_5\varepsilon_{t-2} + \lambda_6\varepsilon_{t-3} \end{aligned} \tag{16}$$

Model VI is the last distributed lag model. It is a third-order lag function in which O is lagged by one period in addition to the lags in the dependent variable. This gives the equation:

$$\begin{aligned} y\_sea_t = {} & c + T_t\beta_1 \\ & + \lambda_1 y\_sea_{t-1} + \lambda_2 y\_sea_{t-2} + \lambda_3 y\_sea_{t-3}\beta_2 \\ & + O_t + O_{t-1}\beta_3 \\ & + \varepsilon_t + \lambda_4\varepsilon_{t-1} + \lambda_5\varepsilon_{t-2} + \lambda_6\varepsilon_{t-3} \end{aligned} \tag{17}$$

6 Results

In this section we evaluate the predictions of the estimated models. First we select two regression models, one based on OLS and one based on SVR. Then we select a distributed lag model to estimate the effects of the offline media more specifically. We discuss the estimated effects for radio and television commercials and distinguish the effects for each television channel.

6.1 Model Selection

When the out-of-sample performances of the models in which the offline media are included as explanatory variables are significantly better than the model in which these variables are not included, we conclude that the offline part of the campaign affect the online conversion actions. First we use the linear regression model with the OLS-method to answer this question. We have to choose the model with the optimal number of lags of the offline media variables included, to compare the forecasts of this model with the forecasts of the OLS-model without any offline media variables included. From Figure 3, we observe that the predictions reach their minimum MAE by a model with ten lags included. This means that the OLS-model in which we take the effect of a commercial for 50 minutes into account, has the best out-of-sample performance. This model is defined as follows:

$$y\_sea = c + Tb_1 + Ob_2 + O\,(-1\ to - 10)\,b_3 + \varepsilon \tag{18}$$

Here y_sea is an $n \times 1$ vector with the target values, c is an $n \times 1$ vector with constants, T an $n \times 24$ matrix with the explanatory variables which indicate the time ($time$ and $hour_i$, $i \in 2, ..., 24$), b_1 a 24×1 vector of estimated coefficients, O an $n \times 15$ matrix with the offline media explanatory variables, (nr_radio, grp_radio, nr_tv, grp_tv and the dummies ch_i, $i \in$ (Comedy Central, Discovery, National Geographic, Net5, RTL4, RTL5, RTL7, RTL8, SBS6, TLC, Veronica)), b_2 a 15×1 vector of estimated coefficients, $O\,(-1\ to - 10)$ is an $n \times 150$ matrix with the ten lags of the offline media explanatory variables, b_3 a 150×1 vector of estimated coefficients, ε an $n \times 1$ vector of disturbances, and n is the number of observations.

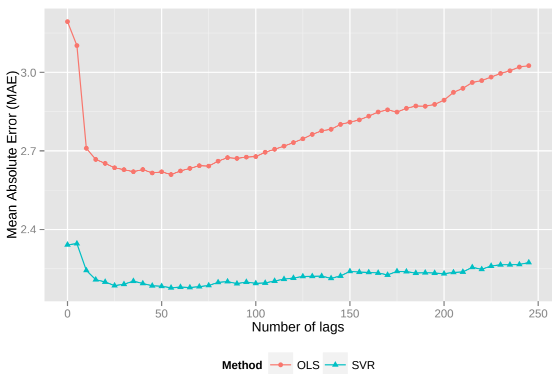

Figure 3 Changes in MAE as a result of the addition of lags in the OLS- and SVR-models.

Now we use the SVR-model to answer the main question. We choose an SVR-model with the optimal number of lags of the offline media variables included, to compare the forecasts of this model with the forecasts of the SVR-model without any offline media variables included. For this choice, we look again at the model with the smallest MAE. This is the model with a MAE of 2.1773 by ten included lags. In Figure 3 we see the changes in MAE as a result of the addition of lags in the SVR-models.

Table 1 shows the out-of-sample performance measurements and the in-sample R-squared for the six estimated distributed lag models. It is not difficult to select the best performing model. *Model IV* has not only high R-squared, which means that it explains a large part of the in-sample variance, it also produces the most accurately and efficient forecasts. Unfortunately, these forecasts are biased.

Table 1 Statistics of the six distributed lag models.

	in-sample R^2	accuracy MAE	efficiency M-Z R^2	biasedness t-value
Model I	0.5497	2.2423	0.1831	-2.1117
Model II	0.6358	2.1782	0.0463	-6.7900
Model III	0.5684	2.5675	0.2969	-2.6136
Model IV	0.6426	2.1160	0.0266	-6.8619
Model V	0.5994	3.3392	0.5190	-1.2267
Model VI	0.6501	2.1205	0.0301	-5.7634

6.2 The Effect of Offline Media on Online Conversion Actions

We compare the out-of-sample performance of the optimal OLS model, which takes the effect of a commercial for 50 minutes into account, with the same model but without the offline media variables. Table 2 shows the out-of-sample performance measurements for these two models. The first model performs better on all the criteria. The t-value of the unbiasedness test of the model with offline media variables included is smaller, so the model without offline media variables underestimates the target values much more. On the basis of the R-squared of the Mincer-Zarnowitz regression we conclude that the forecasts of the first model are more efficient. We use the DM-test to determine whether the differences in predictive accuracy are statistically significant. This test gives us a p-value of 0.0410, so we can reject the null-hypothesis that there is no difference in predictive accuracy between the models on a significance level of ten percent.

Table 2 Forecast measurements of the OLS-model with offline media variables up to ten lags and of an OLS-model without any offline media variables.

OLS	with offline media	without offline media
t-value	-12.7104	-14.3442
M-Z R^2	0.0679	0.0842
MAE	2.6096	3.1935

We compare also the out-of-sample performance of the SVR-model with the same model but without the offline media variables. Table 3 shows that the first model performs better on all the criteria except on the unbiasedness test. The model without offline media variables included underestimates the target values just a little less. The DM-test gives us a p-value of 0.0812, so we can reject the null-hypothesis that there is no difference in predictive accuracy between the models on a significance level of ten percent.

Table 3 Forecast measurements of the SVR-model with offline media variables up to ten lags and of a SVR-model without any offline media variables.

	with offline media	without offline media
t-value	-7.6172	-5.8273
M-Z R^2	0.0002	0.0167
MAE	2.1773	2.3415

Both the DM-test for the SVR-models as the DM-test for the OLS models show that a model with the offline media variables included is significantly better than a model without these variables included. This means that the information about the offline media deployment explains a part of the number of conversion actions. So we conclude that the offline media affect the online conversion actions.

We could deduce the total effect of one independent variable on the dependent variable from the distributed lag model. When we increase an offline media variable i with value one in time period t, it is clear from Equation 16 that the number of online conversion actions in period t increases with $\beta_{2,i}$. The number of conversion actions in period $t+1$ increases with $\lambda_1\beta_{2,i} + \beta_{3,i}$. The total effect, that is the total increase of the number of conversion actions over all the time periods:

$$\begin{aligned}&\beta_{2,i} + \lambda_1\beta_{2,i} + \beta_{3,i} + \lambda_2\beta_{2,i} + \lambda_1\left(\beta_{3,i} + \lambda_1\beta_{2,i}\right)\\ &\quad +2\lambda_1\lambda_2\beta_{2,i} + \lambda_1^2\left(\beta_{3,i} + \lambda_1\beta_{2,i}\right) + \lambda_2\beta_{3,i} + ...\end{aligned}$$

If we rearrange terms, we get:

$$\begin{aligned}(\beta_{2,i} + \beta_{3,i}) &+ (\lambda_1 + \lambda_2) \times (\beta_{2,i} + \beta_{3,i})\\ &+(\lambda_1 + \lambda_2)^2 \times (\beta_{2,i} + \beta_{3,i})\\ &+(\lambda_1 + \lambda_2)^3 \times (\beta_{2,i} + \beta_{3,i}) + ...\end{aligned}$$

In equation 6.2 we recognize a geometric series. So the total effect is equal to:

$$\frac{\beta_{2,i} + \beta_{3,i}}{1 - (\lambda_1 + \lambda_2)}$$

We use *Model IV* to interpret the variables $grp\_tv \times ch_i$. We estimate the total effect of the increase in GRP of a commercial broadcasted on ch_i with Equation 19:

$$\frac{(\beta_{2,grp\_tv \times ch_i} + \beta_{3,grp\_tv \times ch_i}) \bigtriangleup grp}{1 - (\lambda_1 + \lambda_2)} \tag{19}$$

The total increase in the number of conversion actions as a result of an increase in the GRP of a commercial on ch_i with one, is shown in Table 4. RTL4 generates the most conversion actions per GRP and a commercial on Animal Planet or RTL8 has the smallest effect. We could interpret the values in Table 4 as follows: When we broadcast for example five additional GRP's of a television commercial on RTL4, this results in $5 \times 22 = 110$ extra conversion actions.

The maximum GRP of a commercial measured in the campaign period is 10.5. The number of additional conversions per GRP after 50 minutes, or after ten lags,

Table 4 Estimated yield of conversion actions for each channel as the GRP increases by one.

Channel	Total conversions
RTL4	22
Comedy Central	21
TLC	17
RTL5	16
Net5	14
Discovery Channel	13
RTL7	12
SBS6	12
Veronica	9
National Geographic	1
RTL8	< 1
Animal Planet	< 1

Table 5 Maximum Likelihood Estimates Model IV (part 1).

Parameters	**Value**	**SE**
$\mathbf{c}$	-0.5501	0.4312
$\boldsymbol{\lambda}_1$	0.3011	0.0056
$\boldsymbol{\lambda}_2$	0.1382	0.0068
$\boldsymbol{\lambda}_3$	-0.0110	0.0074
$\boldsymbol{\lambda}_4$	-0.0256	0.0083

is at most 0.07. This means that in this period a commercial supplies no additional conversions actions anymore. This is in agreement with the results of the OLS- and SVR-models, which indicate a model with ten lags included as the best performing model.

Tables 7 and 8 show that the number of radio commercials has a negative effect on the number of conversion actions. When we increase the number of radio commercials in a time period and the total GRP remains constant, the number of conversion actions will decrease. However, this effect is not significant because of the large standard error (SE). The number of television commercials has the opposite effect and is significant. When we increase the number of television commercials, the additional conversion actions per GRP increase. Radio commercials supply not so much conversion actions as television commercials. The total effect of an extra GRP is only two conversion actions. The effect is small, occurs immediately and disappears quickly.

7 Conclusion

This paper began by asking whether offline media affect online conversion actions. OLS-models and SVR-models of the number of SEA conversion actions were estimated from the data about the marketing campaign “Do Us A Flavor”. The models showed that the offline media affect the online conversion actions. In addition, both the OLS-models and the SVR-models showed that a model with ten lags of the offline media variables included has the best performance. This is

Table 6 Maximum Likelihood Estimates Model IV (part 2).

Parameters $\mathbf{T}_t$	**Value**	**SE**
time	0.0079	0.0182
hour_2	0.4591	0.8368
hour_3	0.5171	1.4391
hour_4	0.3499	1.8626
hour_5	0.2847	1.4593
hour_6	0.1421	1.9852
hour_7	-0.0922	1.5157
hour_8	0.2741	1.7236
hour_9	1.0117	1.7991
hour_10	1.6933	2.0529
hour_11	2.5410	2.2603
hour_12	2.7609	2.4703
hour_13	3.3028	2.6966
hour_14	3.3177	2.8828
hour_15	2.9882	3.1097
hour_16	4.1575	3.2692
hour_17	3.5960	3.5355
hour_18	3.8375	3.7578
hour_19	3.2207	3.9359
hour_20	4.5021	4.1494
hour_21	3.2374	4.3924
hour_22	3.3603	5.5902
hour_23	1.5525	5.8064
hour_24	-0.2909	5.0454

Table 7 Maximum Likelihood Estimates Model IV (part 3).

Parameters $\mathbf{O}_t$	**Value**	**SE**
nr_tv	0.7406	0.1809
nr_radio	-0.6400	0.7774
grp_radio	1.0152	0.1889
Animal	3.1459	1.5563
Comedy	5.2575	0.1027
Discovery	-0.8173	0.3843
Geographic	0.9814	0.4414
Net5	1.9567	0.4585
RTL4	8.6707	0.0490
RTL5	2.4608	0.5674
RTL7	1.1022	0.2877
RTL8	0.2527	0.6127
SBS6	3.4427	0.0545
TLC	-3.3899	0.2382
Veronica	-0.6406	0.0826

an indication that a commercial influences the campaign website visitation up to 50 minutes after broadcasting.

For a more detailed examination we estimated six distributed lag models. Based upon the out-of-sample performances we selected the best model which we used to divide the estimated effect of the offline media, radio and television commercials, on the online conversion actions. This analysis showed that radio commercials achieve poor results in comparison with television commercials. Second, we found that the number of television commercials has a positive effect on the number of conversion actions.

Table 8 Maximum Likelihood Estimates Model IV (part 4).

Parameters $\mathbf{O}_{t-1}$	**Value**	**SE**
nr_tv	1.2409	0.1538
nr_radio	-0.3397	0.8527
grp_radio	0.2516	0.2300
Animal	-3.6411	3.5378
Comedy	6.3369	0.1677
Discovery	8.0543	0.1098
Geographic	-0.3107	0.3468
Net5	6.1690	0.2421
RTL4	3.6666	0.0773
RTL5	6.2692	0.2497
RTL7	5.5689	0.1491
RTL8	-0.6447	0.6305
SBS6	3.2248	0.0777
TLC	12.9997	0.1240
Veronica	5.5209	0.0811

Furthermore we distinguished the effects per television channel. There are large differences in the reactions of people to a commercial per channel. There are channels on which commercials have no significant effect on the website visitation, such as National Geographic and Animal Planet. RTL4 broadcasted the most productive commercials. So general-purpose TV channels have a higher impact than specialized TV channels on the number of conversion actions. We concluded that the offline media affect the online conversion actions. The conversion actions represent the people visiting the campaign website after a contact with search engine advertising. Our results show that, people are going to use a search engine to search for the campaign website after seeing a commercial about the website.

Now that we found the effects of the offline media on the online conversion actions we recommend the involved marketing company to use these results for further research. The first point for further research is the comparison of the target group of the campaign with the viewer audiences of the television channels. Because the audience of each channel reacts totally different to a commercial, the campaign would be much more efficient when the commercials are broadcasted on the channels which reach the right audience. In the analyzed campaign we saw that the commercials on Animal Planet and National Geographic made no sense, probably because their audience is not interested in potato chips.

The second point of future research is the inclusion of the cost of the offline media resources in the research. When we combine the effects of the offline media we have found in this paper with the costs of the offline media resources, we can provide the most cost efficient allocation of the offline media resources. The third point that needs further work is next to the inclusion of the costs, the addition of the information about the display advertising. In this way it is possible to make a model which estimates the optimal combination of offline and online media to achieve the best cost-efficient result.

Acknowledgment

Damir Vandic is supported by an NWO Mosaic scholarship for project 017.007.142: *Semantic Web Enhanced Product Search (SWEPS).*

References

Baran, S. (2010), *Introduction to Mass Communication: Media Literacy and Culture-6/e.*

Barwise, P. & Farley, J. (2005), 'The State of Interactive Marketing in Seven Countries: Interactive Marketing Comes of Age', *Journal of Interactive Marketing* **19**(3), 67–80.

Bass, F. & Clarke, D. (1972), 'Testing Distributed Lag Models of Advertising Effect', *Journal of Marketing Research* pp. 298–308.

Chang, C.-C. & Lin, C.-J. (2011), 'LIBSVM: a Library for Support Vector Machines', *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Clarke, D. (1976), 'Econometric Measurement of the Duration of Advertising Effect On Sales', *Journal of Marketing Research* pp. 345–357.

Collobert, R. & Bengio, S. (2001), 'Svmtorch: Support Vector Machines for Large-scale Regression Problems', *The Journal of Machine Learning Research* **1**, 143–160.

Diebold, F. & Mariano, R. (1995), 'Comparing Predictive Accuracy', *Journal of Business and Economic Statistics* **13**(3), 253–263.

Dinner, I. M., Van Heerde, H. J. & Neslin, S. A. (2014), 'Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising', *Journal of Marketing Research* **51**(5), 527–545.

Franses, P. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press.

Franses, P. & Oest, R. (2004), On the Econometrics of the Koyck Model, Technical report, Erasmus School of Economics (ESE). `http://repub.eur.nl/res/pub/1190/ei200407.pdf`.

Ghose, A. & Yang, S. (2009), 'An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets', *Management Science* **55**(10), 1605–1622.

Goh, K., Heng, C. & Lin, Z. (2013), 'Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content', *Information Systems Research* **24**(1), 88–107.

Goldfarb, A. & Tucker, C. (2011), 'Online display advertising: Targeting and obtrusiveness', *Marketing Science* **30**(3), 389–404.

Gopal, R., Li, X. & Sankaranarayanan, R. (2011), 'Online Keyword Based Advertising: Impact of Ad Impressions on Own-channel and Cross-Channel Click-through Rates', *Decision Support Systems* **52**(1), 1–8.

Heij, C., Boer, P. d., Franses, P., Koek, T. & van Dijk, H. (2004), *Econometric Methods with Applications in Business and Economics*, Oxford University Press.

Kireyev, P., Pauwels, K. & Gupta, S. (2015), 'Do display ads influence search? attribution and dynamics in online advertising', *International Journal of Research in Marketing*.

Koyck, L. (1954), *Distributed Lags and Investment Analysis*, North-Holland Publishing Company Amsterdam.

Liaukonyte, J., Teixeira, T. & Wilbur, K. C. (2015), 'Television advertising and online shopping', *Marketing Science* **34**(3), 311–330.

Lu, C.-J., Lee, T.-S. & Chiu, C.-C. (2009), 'Financial time series forecasting using independent component analysis and support vector regression', *Decision Support Systems* **47**(2), 115–125.

Naik, P. A. & Peters, K. (2009), 'A hierarchical marketing communications model of online and offline media synergies', *Journal of Interactive Marketing* **23**(4), 288–299.

Pepsico Nederland (2014), 'Pepsico nederland', From: `http://www.pepsico.nl/media/releases/the_battle.php`.

Rao, V. (1972), 'Alternative Econometric Models of Sales-advertising Relationships', *Journal of Marketing Research* pp. 177–181.

Rishika, R., Kumar, A., Janakiraman, R. & Bezawada, R. (2013), 'The effect of customers' social media participation on customer visit frequency and profitability: An empirical investigation', *Information Systems Research* **24**(1), 108–127.

van Rijsewijk, R., van der Heijden, L. & Gagandeep, S. (2013), 'Iab Report On Online Ad Spend H1 2013'. `http://goo.gl/MO9AGS`.

Vapnik, V. (1998), *Statistical Learning Theory*, Wiley, New York.

Vapnik, V. (1999), *The Nature of Statistical Learning Theory*, 2 edn, Springer.

Winer, R. & Ilfeld, J. (2002), 'Generating Website Traffic', *Journal of Advertising Research* **42**(5), 49–61.

Zeff, R. & Aronson, B. (1999), *Advertising On the Internet*, John Wiley & Sons, Inc.