Attribution modeling

de Haan, Evert

Published in:
Handbook of Market Research

DOI:
10.1007/978-3-319-05542-8_39-1

Attribution Modeling

Evert de Haan

Contents

Introduction 2
Dataset 3
Problems with Basic Attribution Methods 7
  Touch-Based Attribution 8
  Correlations and Regression Models 10
  Shapley Value–Based Attribution 14
Attribution Modeling Process with Experimental Data 16
Additional Topics for Individual-Level Attribution 30
  Attribution with Individual-Level Quasi-Experimental Data (Propensity Score Matching) 30
  Exploring Paths in Conversion (Markov Chain) 39
  Further Methods of Individual-Level Attribution 45
Attribution with Aggregate-Level (Quasi-)Experimental Data 45
  Before-After Analysis 46
  Before-During-aAfter Analysis 51
  Difference-in-Differences Analysis 54
  Further Methods of Aggregate-Level Attribution 61
Conclusion 63
Cross-References 63
References 64

Abstract

Marketing attribution is the process of allocating appropriate credit to each marketing touchpoint a customer has encountered before conducting the desired customer action, e.g., a purchase. Ideally, this credit should be capturing the incremental effect of the touchpoint on the customer action. Finding this incremental effect is relevant for marketers to decide on budget allocations and to

E. de Haan (✉)
Departement of Marketing, Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
e-mail: evert.de.haan@rug.nl

C. Homburg et al. (eds), *Handbook of Market Research*,
https://doi.org/10.1007/978-3-319-05542-8_39-1

decide how, when, and where to target which customer. This chapter introduces and discusses various marketing attribution techniques. The techniques range from basic attribution techniques, like touch-based attribution and Shapley values, to advanced attribution techniques, like randomized field experiments and Markov chains. The chapter discusses the up- and downsides of each attribution technique, discusses alternative methods if one method is inappropriate, and links this to the concept of incrementality and causality, i.e., to which degree the technique gives proper credits to the different channels or touchpoints the customer has encountered. This chapter is accompanied by the necessary R-scripts to generate the datasets and estimate the attribution techniques, which can also be downloaded at http://www.evertdehaan.com.

Keywords

Marketing attribution · Advertising effectiveness · Credit allocation · Attribution modeling · Last click · First click · Last touch · First touch · Shapley values · Causality · Incremental effects · Field experiments · Propensity scores · Matching · Markov models · R-code · R-script · Model estimation

Introduction

Due to digitalization and the rise of the Internet, it has become easier to track individual consumers in their (online) customer journey. For marketers, this means that they can get insights into which touchpoints a consumer has encountered. For example, how many times did the consumer see a banner advertisement, what did (s)he search for on a search engine, if (s)he used a price comparison site, and when and how (s)he has visited a firm's website. Furthermore, with online retailing, it is easily possible to observe which consumers have purchased and link the touchpoints to this purchase. The touchpoints and the moment of purchase together form a customer's *path to purchase*. A question that can arise when looking at such a path to purchase is which touchpoints have influenced the purchase decision. This question is the core of (online) marketing attribution.

Attribution is defined as the process to "allocate appropriate credit for a desired customer action to each marketing touchpoint across all online and off-line channels" (Kannan et al. 2016). The desired customer action is typically a conversion or a purchase. In other words, with attribution, one wants to find out:

- To what extent a (combination of) touchpoint(s) has/have impacted the likelihood to purchase (or the likelihood of another desired outcome) for an individual customer. This is individual-level attribution.
- How a (combination of) marketing channel(s) influence(s) the overall sales (or another desired outcome) for the firm. This is aggregate-level attribution.

As Hanssens (2021) has put it, “the key challenge in digital attribution is to estimate the *incremental* purchase probability achieved by a certain media intervention.” The word “incremental” is crucial here, i.e., the difference in the purchase probability, because of a specific media intervention (i.e., touchpoint or marketing outing). Finding this incremental effect and thus conducting attribution is relevant for marketers to decide on budget allocations and to decide how, when, and where to target which customer. The importance of this is clear from some practical examples. When eBay stopped using search engine advertising (SEA), they saw in many cases no change in traffic to eBay’s website because of the substitution coming from organic (i.e., nonpaid) search engine traffic (Blake et al. 2015). Procter and Gamble cut $200 million in digital ad spend and reallocated this to other channels, including television and radio, which increased its reach by 10% (Johnson 2018). When Uber cut two-third of their ad spending, which saved $100 million in costs, they saw almost no change in app installations (WARC 2021). Knowing which touchpoints contribute to the desired outcome is thus crucial for a firm’s bottom-line performance and can help to improve marketing effectiveness.

This chapter will discuss techniques for both individual- and aggregate-level attribution. The chapter furthermore provides R-scripts for the attribution techniques and to generate the datasets. The datasets are easy to adjust, e.g., the scripts can easily be changed to include additional consumers and additional marketing channels or to adjust the model’s assumptions and effect sizes. With the script to estimate the attribution models, it is also easy to apply the models to other, e.g., real-life datasets. All of this makes the datasets and R-scripts, and hence this chapter, useful for teaching purposes and marketing practitioners.

The next section of this chapter introduces the dataset for the individual-level attribution. Hereafter, some basic attribution methods are discussed, including touch-based attribution, regression-based models, and Shapley values. Section “Attribution Modeling Process with Experimental Data,” introduces a kind of golden standard for attribution, namely a randomized field experiment at the individual customer level. In section “Additional Topics for Individual-Level Attribution,” attribution methods are discussed when the ideal data is not available or when additional insights are needed, including propensity score matching and Markov models. In section six, some methods for aggregate-level attribution are discussed. The final section concludes this chapter.

Dataset

We will use the same dataset throughout the next three sections, covering individual consumer-level attribution. The dataset can be generated with the R-script or be downloaded directly, together with all other datasets and R-scripts used in this chapter, at http://www.evertdehaan.com. The dataset contains 50,000 customer journeys, with ~25% of the journeys resulting in a purchase and each journey having between 1 and 50 touchpoints in total. We have in total eight unique touchpoints, ranging from banner impressions and clicks to SEA for brand- and product-related keywords.

The first step in attribution is thus to have the right data available to conduct attribution modeling. For managers and researchers, who want to apply the techniques from this chapter on their own data, we recommend to collect the right data and structure it in the right way. The right data would (at an individual customer level) include the different touchpoints a customer has encountered over time, and can be further enriched with customer-specific information (e.g., demographic and customer relationship data) and other browsing behavior (e.g., clickstream data which provides information on what a customer does on the website, see for instance also Moe 2003). The structure of these data can be similar to the example data, discussed in this section and used in the following three sections. If individual-level data is not available (in general, or for some channels), another option would be to use time series data; more on the structure of this type of data and how to model attribution with these data are discussed in the section "Attribution with Aggregate-Level (Quasi-)Experimental Data."

We can get some descriptive statistics of the journeys with the following R-script.

```
# Get descriptive statistics
library(psych)
describe(consumers)
```

Table 1 shows some of the descriptive results, together with a description of what each variable measures. The 7th ("Banner_no_click") up until the 14th ("Direct_visit") variable measure the number of occurrences for the eight different touchpoints in each path to purchase. With attribution, we want to see how these touchpoints influence the purchase, i.e., the 19th (last) variable. We furthermore have information on the length of the customer relationship and the customer lifetime value. The dataset contains data from two field experiments, captured with the "Firm_banner" and "Flyer_region" variables. In the section "Attribution Modeling Process with Experimental Data," we will explain these experiments further.

An example of a path to purchase, which will come back throughout this chapter, is shown in Fig. 1. This path of purchase is from customer #134 from the dataset. We see that this customer is first coming into contact with a banner by the firm but does not click on it. After this, the customer uses a search engine where (s)he uses a product-related keyword and clicks on a sponsored search link to visit the website. The two subsequent visits are direct visits to the website, which occur by typing in the website's URL. Hereafter are again two banner impressions; the customer clicks on the second banner impression, leading to another website visit. Next are another direct visit and a banner impression. Finally, the customer uses a search engine to search for the company (i.e., branded search), clicks on the sponsored search link, and conducts a purchase.

When looking at Fig. 1, we might ask which of the nine touchpoints has made sure that this customer has conducted the purchase. Was this because of the branded SEA in the end? Or did the banner impression which in the beginning started this

Table 1 Variable descriptions and descriptive statistics

Variable name	Description	Mean	sd	Median	Min	Max
1. Consumer_ID	A unique customer ID				1	50,000
2. Existing_customer	Dummy indicating if the customer made a purchase before	0.50			0	1
3. Relation_length	Amount of months active at the firm (measured on January 1)	15.25	19.53	1	0	60
4. CLV	Customer lifetime value in dollars (measured on January 1)	904.43	1197.41	55.64	0	7341
5. Firm_banner	Dummy variable indicating if the customer was in the firm's banner (1) or charity banner (0) group	0.80			0	1
6. Email_group	Dummy variable indicating if the customer signed up for the email newsletter (1) or not (0) (introduced on January 1)	0.27			0	1
7. Banner_no_click	Count variable, indicating the number of banner impressions without a click (from the firm or charity)	2.04	2.49	1	0	29
8. Banner_click	Count variable, indicating the number of banner clicks (for the firm or charity banner)	0.12	0.39	0	0	6
9. SEA_product_click	Count variable for the amount of website visits through sponsored search results using product-related keywords	0.50	0.89	0	0	10

(continued)

Table 1 (continued)

Variable name	Description	Mean	sd	Median	Min	Max
10. SEA_brand_click	Count variable for the amount of website visits through sponsored search results using firm−/brand-related keywords	0.81	1.21	0	0	14
11. Price_comp_click	Count variable, indicating the number of website visits through a price comparison site	0.44	0.84	0	0	12
12. Email_no_click	Count variable, indicating the number of emails received without clicking on a link and visiting the website	0.10	0.45	0	0	7
13. Email_click	Count variable, indicating the amount of email links clicked on and, i.e., visiting the website	0.05	0.26	0	0	7
14. Direct_visit	Count variable, indicating the number of direct website visits (e.g., by typing in the URL)	2.49	2.82	2	0	32
15. First_channel*	String variable, naming the first channel in the path to purchase					
16. Last_channel*	String variable, naming the last channel in the path to purchase					
17. Amount_touchpoints	Count variable, summing up all touchpoints in the path to purchase	6.53	5.99	5	1	50

(continued)

Table 1 (continued)

Variable name	Description	Mean	sd	Median	Min	Max
18. Flyer_region	Dummy variable, indicating if the consumer lives in a region where the firm distributed their flyers (1) or not (0)	0.50			0	1
19. Purchase	Dummy variable, indicating if the path to purchase ended with a purchase (1) or not (0)	0.26			0	1

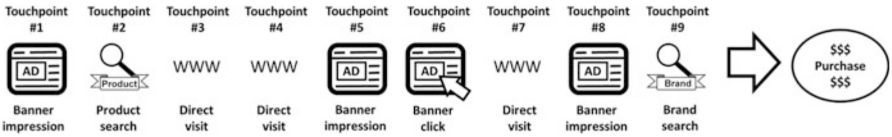

Fig. 1 Example customer path to purchase

path to purchase cause the end result? Or should all touchpoints get (equal or different) credits for the purchase? Or would the customer also have conducted the purchase without any of the (advertisement-based) touchpoints?

We want to determine what would have happened if certain (combinations of) touchpoints were not there. If the conversion still happens without a specific touchpoint, then that touchpoint should not get any credit for the conversion. If the conversion would not happen without the touchpoint, that touchpoint should get credit. A challenge is that we can only observe the path to purchase as it was, i.e., we do not observe if the outcome would be different if the path to purchase would be different. Luckily, attribution techniques can provide insights into this, as discussed in the following sections of this chapter.

Problems with Basic Attribution Methods

This section discusses some basic attribution methods, which have been used a lot in practice, but have the downside that they do not answer the attribution question, namely what the incremental effect of a channel is. We discuss these methods, to show how they work, what insights they bring, and to what extent they relate to the attribution question and are thus useful for practice. We will start with touch- or click-based attribution methods, followed by correlations and regression models, and finally, we discuss the more advanced Shapley value–based attribution method.

Touch-Based Attribution

An advantage of online attribution is that we can observe all consumers' online touchpoints, website visits, and conversions. Based on these data, we can observe the customers' path to purchase and perform *touch-based attribution*, sometimes also called *click-based attribution*. Historically, the most popular form of touch-based attribution is last-touch attribution, i.e., the last touchpoint a customer comes into contact with before a purchase gets all credit for that purchase. In the example from Fig. 1, last-touch attribution would thus give branded SEA full credits for the purchase. There are also other touch-based attribution methods, e.g., first-touch-based attribution gives full credit to the first touchpoint, average (or linear) touch attribution gives all touchpoints equal credit, and time decay attribution gives the lowest credit to the first touchpoint and the highest credit to the last touchpoint. Figure 2 visualizes the different touch-based attribution methods and also indicates what this means for the journey from Fig. 1.

To see to what extent touch-based attribution can come to different conclusions, let us use the dataset introduced in the previous section. The following R-script conducts three different touch-based attribution techniques. Table 2 presents the outcomes of the R-script.

```
# Number of occurrences of touchpoint
consumers_journey$n <- 1
aggregate(consumers_journey$n, by=list(consumers_journey
$Channel_name), FUN=sum)
consumers_journey$n <- NULL
# Last-touch, total amount conversions
aggregate(consumers$Purchase, by=list(consumers$Last_channel),
FUN=sum)
# First-touch, total amount conversions
aggregate(consumers$Purchase, by=list(consumers$First_channel),
FUN=sum)
# Average-touch, total amount conversions
sum(consumers$Banner_click/consumers$Amount_touchpoints*consumers
$Purchase)
```

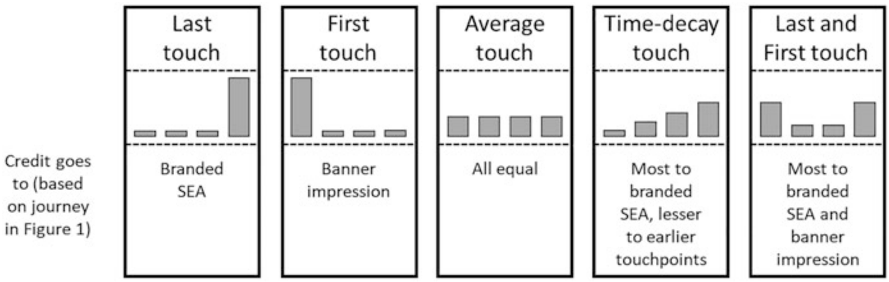

Fig. 2 Touch-based attribution examples

Table 2 Credit allocation using touch-based attribution

		Conversions attributed to channel			Conversion percent per channel		
	n	Last touch	First touch	Average touch	Last touch	First touch	Average touch
Banner click	6,029	154	185	217.6	2.55%	3.07%	3.61%
Banner impression	101,930	0	3,979	3,309.4	0.00%	3.90%	3.25%
Direct	124,486	7,838	4,716	5,426.5	6.30%	3.79%	4.36%
Email click	2,265	166	159	157.1	7.33%	7.02%	6.94%
Email received	5,141	0	354	253.8	0.00%	6.89%	4.94%
Price comparison	21,872	1,468	1,212	992.4	6.71%	5.54%	4.54%
SEA branded	40,253	2,473	1,166	1,633.4	6.14%	2.90%	4.06%
SEA product	24,773	692	1,020	800.8	2.79%	4.12%	3.23%
Total	**326,749**	**12,791**	**12,791**	**12,791**	**3.91%**	**3.91%**	**3.91%**

```
sum(consumers$Banner_no_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$Direct_visit/consumers$Amount_touchpoints*consumers
$Purchase)
sum(consumers$Email_click/consumers$Amount_touchpoints*consumers
$Purchase)
sum(consumers$Email_no_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$Price_comp_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$SEA_brand_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$SEA_product_click/consumers
$Amount_touchpoints*consumers$Purchase)
```

As shown in Table 2, the three different touch-based attribution methods come to different conclusions. First of all, with last touch, a sale can only be attributed to a channel that leads directly to a website visit; this is why a mere banner impression or a received email that does not receive a click do not get credit for a conversion. With first touch and average touch, also mere exposures can get credit. For these latter two attribution methods, banner impressions get a relatively high amount of credit. Sometimes for first-touch attribution and average-touch attribution also only actually clicks are being used (i.e., click-based instead of touch-based attribution), which would change the numbers since the banner impression and received email would then drop out.

Direct visits and branded SEA get much credit with last-touch attribution but less with the other two attribution methods. This difference might be because at the end of a (longer) customer journey, the consumer already knows where to buy and goes directly to the website (or types in the brand name at a search engine). Product-related SEA, i.e., people clicking on a sponsored search link when using a product-related keyword, gets relatively more credit with first-touch and average-touch attribution. This difference might also be due to the stage in which the customer uses product-related SEA. Product-related SEA is mainly used at the start when the customer looks for broader information and is not sure yet where to buy the product. These differences are also in line with findings from Rutz and Bucklin (2011).

A downside of all the touch-based attribution methods is that none of these methods answer the attribution question, namely “to what an extent a (combination of) touchpoint(s) has/have impacted the likelihood to purchase.” We do, namely, not observe if the conversion would not have happened without a specific (combination of) touchpoint(s). The touch-based attribution methods do thus not work well for attribution purposes, as scientific studies have also shown (e.g., De Haan et al. 2016; Li and Kannan 2014). Instead, we have to use alternative methods that tell us if the effects are causal, i.e., without the touchpoint the outcome (e.g., purchase) would have been different. The criteria for causality, and hence suitable attribution, are:

1. **Covariation:** A shock in the independent variable (i.e., the exposure to a touchpoint) correlates with a shock in the dependent variable (i.e., the desired customer outcome).
2. **Order in time:** The shock in the independent variable has to occur before the shock in the dependent variable.
3. **No third variable:** There are no other variables or reasons that explain the effect (e.g., confounds like seasonality or self-selection bias). In other words, the change in the dependent variable is due to the change in the independent variable (i.e., the touchpoint occurring in the path to purchase).

Touch-based attribution methods do not meet the third criterion since the exposure to a touchpoint and purchase might be driven by, for instance, seasonality, e.g., during peak seasons, the number of purchases is higher, and firms spend more on marketing activities. Another explanation might be the self-selection by consumers, e.g., consumers who sign up for a newsletter are already more likely to purchase in the future, even without receiving the newsletter. We thus have to find alternative methods that can give more certainty about the causality to conduct accurate attribution, especially in terms of excluding all potential third variables.

More information on causality is also provided in other chapters in this book, e.g., Artz and Doering (2021), Bornemann and Hattula (2018), Ebbes et al. (2016), and Valli et al. (2017).

Correlations and Regression Models

Another, more algorithmic, way of doing basic attribution is looking at correlations or estimating a regression model. In such a way, one can relate the touchpoints a customer has come into contact with to the dependent variables of interest, e.g., a purchase. Furthermore, it is possible to see how the touchpoints correlate to each other and other variables like the consumer's characteristics. Let us investigate with the following R-script some correlations in our dataset.

```
# Make correlation plot
library(corrplot)
my_data <- consumers[, c(2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,19)]
res <- cor(my_data)
corrplot(res, type = "upper", tl.col = "black", tl.srt =
45, sig.level = 0.01, insig = "blank")
rm(my_data)
rm(res)
```

Figure 3 visualizes the output of the correlation analysis. We can observe that the "purchase" variable correlates positively, but weakly, with most other variables. "Purchase" is strongest correlated with the "direct visit" and "amount touchpoints" variables. These correlations mean that the direct visits and the path to purchase length are the strongest indicators of a purchase. This could make sense since consumers who visit the website directly might be more likely to know the online retailer already. Indeed, we can observe that direct visits also positively correlate with the "existing customer," "relation length," and "CLV" (customer lifetime value, i.e., how much the customer's future transactions are worth in terms of net present value) variables, indicating that these consumers might already be loyal to the online

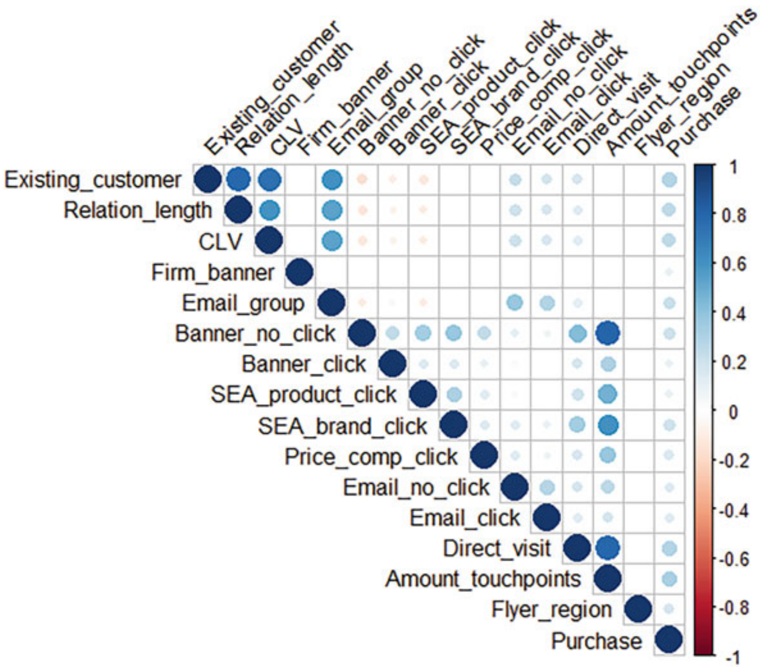

Fig. 3 Correlation matrix

retailer. In this case, it might thus not be the channel that is driving the purchase, but the type of consumer that uses the channel determines the chance of a conversion.

Furthermore, longer paths to purchase might relate to more informed and interested consumers and are more likely to convert. This indicates that the amount of touchpoints is not per se driving purchase, but it is the consumer's underlying "degree of interest." Correlations thus have to be interpreted with caution since they only tell us something about the association between variables but do not give us causal information.

When looking at the correlations from Fig. 3, we can furthermore see that the banner variables and the SEA product variable are positively related to each other, which might relate to the phases in the path to purchase; consumers who first see a banner might later be interested in clicking on it, and later search for product information. Correlation tables can thus be convenient to understand patterns in the data, which could be input for more complicated methods to find out causal effects.

A more elaborate form of investigating the associations between variables is by estimating a regression model (see for more details Skiera et al. [2018]). An advantage of a regression model over correlations is that we can control for third variables and thus can get a step closer to finding causal relations. Since the dependent variable "purchase" is binary, a logistic regression model is most appropriate in our case (see for more details Tillmanns and Krafft [2017]). With the following R-script, we estimate some logistic regression models.

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Logistic regression models to predict the purchase likelihood
model1 <- glm(Purchase ~ as.factor(Last_channel), data=consumers,
family=binomial)
model2 <- glm(Purchase ~ as.factor(Last_channel) +
Existing_customer + Relation_length + log(CLV+1) +
Amount_touchpoints + Flyer_region, data=consumers,
family=binomial)
tab_model(model1, model2, transform = NULL, collapse.ci = TRUE,
p.style = "stars")
```

Table 3 shows the output of the models. Model 1 only looks at the last channel used and is thus very similar to the last-touch attribution procedure. "Banner_click" is here the reference category, meaning that the interpretation of the parameters is relative to this touchpoint. The variables "Banner_impression" and "Email_received" have a strong negative parameter estimate, which makes sense because these variables do not directly relate to a conversion, as discussed in the previous section. The parameter for "Direct_visit" is positive and significant, meaning that there is a significantly higher chance of conversion when this channel is the last touchpoint than when "Banner_click" is the last touchpoint. These parameters are all in line with what we saw in Table 2. When we control for some customer-

Table 3 Logistic regression output using touch-based attribution

	Model 1	Model 2
	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*
(Intercept)	−1.05 *** (−1.24 – −0.87)	−3.47 *** (−3.70 – −3.25)
Last_channel [Banner_impression]	−17.52(−99.01 – −124.06)	−18.43(−87.53 – −111.02)
Last_channel [Direct_visit]	0.28 ** (0.10–0.47)	0.08(−0.13–0.30)
Last_channel [Email_click]	2.12 *** (1.77–2.48)	1.76 *** (1.37–2.15)
Last_channel [Email_received]	−17.52(−498.46 – −646.32)	−18.71(−566.97 – −613.56)
Last_channel [Price_comp_click]	0.22 * (0.03–0.42)	0.26 * (0.04–0.48)
Last_channel [SEA_brand_click]	0.24 * (0.05–0.43)	0.13(−0.09–0.35)
Last_channel [SEA_product_click]	−0.25 *(−0.45 – −0.05)	−0.20(−0.43–0.04)
Existing_customer		−0.42 **(−0.72 – −0.12)
Relation_length		0.01 *** (0.01–0.01)
CLV + 1 [log]		0.23 *** (0.19–0.27)
Amount_touchpoints		0.15 *** (0.15–0.16)
Flyer_region		1.05 *** (1.00–1.10)
Observations	50,000	50,000
R^2	0.077	0.294

* *p < 0.05* ** *p < 0.01* *** *p < 0.001*

specific variables, as is done in Model 2 in Table 3, we can see that “Direct_visit” becomes statistically insignificant. This change in significance might be because the direct channel is used more by existing and loyal customers, who also have a higher chance of conversion, i.e., these variables are confounding variables. Indeed, if the consumer is an existing customer of the firm and has a long relationship and a high CLV, the likelihood of conversion is higher, i.e., the type of customer using the channel (partly) explains the effectiveness of the channel. With this, we demonstrate the advantage of a regression model compared to merely looking at correlations, namely that we can correct for some confounding variables.

Although the correlation table nicely shows how the different channels relate to each other, which can give us an idea about what touchpoints belong together to target specific customers, and how the channels relate to purchase, it does not give us information on the causal effects. As was the case with touch-based attribution, potential third variables are, however, not excluded. The logistic regression model can be more appropriate for attribution than the touch-based attribution since we can incorporate third variables as control variables, like the strength of the relation with the customer. It still does not provide us with causal information since we cannot say with certainty that we control for all other factors and what would have happened if a particular channel or touchpoint did not occur. Hence, also correlations and regression models do not address the underlying question of attribution.

Shapley Value–Based Attribution

As discussed, with attribution, we want to determine what would have happened to the outcome variable of interest provided that a channel was not there. Both the touch-based attribution and attribution based on correlations and a regression model do not answer the attribution question. A method that comes closer, and is also more popular in recent years, is Shapley value–based attribution, which has its roots in cooperative game theory (Shapley 1953). This method compares similar paths to purchase, with the only difference that some paths contain a specific touchpoint, but the other paths do not contain this touchpoint.

To give an example of this method, let us imagine a simplified path to purchase of Fig. 1. Assume we have consumers who first comes into contact with a banner, and the consumer does not click on it. After this, these consumers search for information using a product-specific keyword and visit the website by clicking on a SEA link, and finally the consumers search for information using a brand-specific keyword and revisit the website by clicking on a SEA link. If we have a large enough group of these consumers, we can calculate the conversion probability of such a path. Similarly, we can investigate another group of consumers, but with the difference that the first touch point (i.e., the banner impression) is not included in the path to purchase. If we compare the first path to purchase's conversion probability with the second path to purchase, we can calculate the increase in conversion probability if an extra touchpoint was there. This procedure is what we do with the following R-script.

```
# Get all paths to purchase that are "Banner impression" -> "SEAR
product click" ->  "SEA brand click"
consumers_shapley <- consumers[consumers$Amount_touchpoints==3,]
consumers_shapley <- consumers_shapley[consumers_shapley
$First_channel=="Banner_impression",]
consumers_shapley <- consumers_shapley[consumers_shapley
$Last_channel=="SEA_brand_click",]
consumers_shapley <- consumers_shapley[consumers_shapley
$SEA_product_click==1,]
mean(consumers_shapley$Purchase)
# Get similar paths to purchase excluding "Banner impression"
consumers_shapley_ex1 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex1 <- consumers_shapley_ex1
[consumers_shapley_ex1$First_channel=="SEA_product_click",]
consumers_shapley_ex1 <- consumers_shapley_ex1
[consumers_shapley_ex1$Last_channel=="SEA_brand_click",]
mean(consumers_shapley_ex1$Purchase)
# Get similar paths to purchase excluding "SEAR product click"
consumers_shapley_ex2 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex2 <- consumers_shapley_ex2
[consumers_shapley_ex2$First_channel=="Banner_impression",]
consumers_shapley_ex2 <- consumers_shapley_ex2
[consumers_shapley_ex2$Last_channel=="SEA_brand_click",]
```

```
mean(consumers_shapley_ex2$Purchase)
# Get similar paths to purchase excluding "SEA brand click"
consumers_shapley_ex3 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex3 <- consumers_shapley_ex3
[consumers_shapley_ex3$First_channel=="Banner_impression",]
consumers_shapley_ex3 <- consumers_shapley_ex3
[consumers_shapley_ex3$Last_channel=="SEA_product_click",]
mean(consumers_shapley_ex3$Purchase)
# Impact channel 1
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex1
$Purchase)
# Impact channel 2
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex2
$Purchase)
# Impact channel 3
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex3
$Purchase)
```

Figure 4 visualizes the output of the R-script above. We can see that the full path to purchase discussed above occurs 86 times in our dataset, and in 16.28% of the cases result in a conversion. When we leave out the banner impression, the reduced path to purchase occurs 267 times in our dataset, and in 7.49% of the cases result in a conversion, meaning that when the banner impression is left out, the conversion

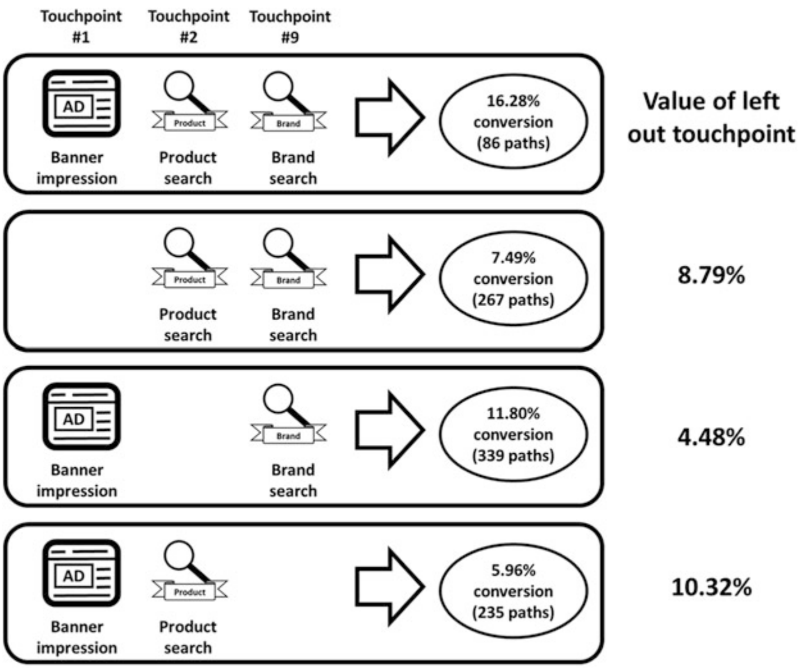

Fig. 4 Shapley value–based attribution example

probability drops 8.79% points. This difference is the credit the banner impression should get. Similarly, Fig. 4 provides the value of the other two touchpoints, with branded SEA getting the highest credit of a 10.32% points increase in the conversion probability. It is important to note that we get a higher conversion rate when adding up the value of each touchpoint than the observed conversion of 16.28%. There can be multiple reasons for this, e.g., there could be a synergy effect, meaning that the combination of touch points is responsible for the purchase, and dropping one thus has a substantial effect. There could also be a time effect, e.g., a path to purchase with three touchpoints generally has a higher conversion since a longer path to purchase could indicate more interested consumers.

Of course, the example given can easily be applied to all kinds of paths to purchase with different combinations of touchpoints. Doing so can provide information in what situation which touchpoint has the largest impact on conversion. Furthermore, the Shapley value procedure is appropriate to investigate the impact of the order of channels, achievable by comparing similar paths to purchase, with a difference that the order of the touch points differs. A challenge is that specific paths to purchase might only occur a limited number of times in the dataset, making the calculated conversion probability less reliable.

In general, the Shapley value approach is a more appropriate way of attribution than touch-based or regression-based attribution. This Shapley value procedure comes close to answer the attribution question "what would have happened if certain (combinations of) touchpoints were not there," i.e., "estimate the *incremental* purchase probability achieved by a certain media intervention" from Hanssens (2021). It does, however, not fully answer this question. First of all, the paths to purchase we are comparing come from different consumers, and the characteristics of these consumers might also (partly) explain the differences in the conversion probabilities. Secondly, touchpoints do not occur at random, e.g., the branded SEA only occurs if the consumer is actively looking for a brand using a search engine. If branded SEA would have been off, the consumer would probably still have visited the website but instead would have clicked on an organic search link to do so or would visit the website directly. Shapely values thus still have some limitations, and we cannot be sure the findings are causal. Because of that, we have to use more advanced techniques and make use of certain features in the data (e.g., a field experiment or other form of randomization), which can provide us with causal insights. The following two sections discusses some of these techniques.

Attribution Modeling Process with Experimental Data

An ideal form of attribution is when there is data available, coming from a field experiment where one (or multiple) channels are turned off for some groups of customers. This "golden standard" is especially useful for individual-level attribution. To find the impact of a channel, consumers are randomly placed into two groups: one group that can encounter the channel of interest (i.e., the treatment group) and one group that cannot encounter this channel (i.e., the control group).

Due to the random allocation of consumers in these two groups, we can assume that consumers of the two groups only differ in terms of being targeted or not being targeted via one specific channel (see for more details Landwehr (2019)). The dataset generated in the "Dataset" section of this chapter includes a randomized field experiment, similar to the experiments by Hoban and Bucklin (2015) and Li et al. (2021). In this experiment, 80% of the consumers are in the group which can encounter the banner ad from the firm (i.e., our treatment group). The other 20% of the consumers are in the group which can encounter the banner ad from an unrelated charity organization (i.e., our control group). The control group uses an unrelated charity ad since this makes it possible to observe how many ads from the firm the consumer would have seen if the consumer was in the treatment group. The actual treatment (i.e., the firm ad) was thus not provided to the control group, but the charity ad captures for the consumers in the control group if they would have seen the firm ad (and also, how many times) or if they would have been in the treatment group. This method assumes that the targeting of the charity ad in the control group is set up the same as the targeting of the firm's ad in the treatment group. For more details on this method, see Hoban and Bucklin (2015).

Since the charity ad is unrelated to the firm, the causal effect of this ad on firm performance and customer behavior should be zero. The number of exposures to the charity ad might, however, correlate with the conversion probability. This correlation might occur since someone who sees the ad more often might (1) be online more often and (2) visit the website where the campaign is running more often, which both relate to the conversion probability. The charity ad captures these confounding effects, and the difference between the treatment group (containing both the causal and confounding effects) and the control group (containing only the confounding effects) thus captures the causal effect of the firm's ad.

Next to banner advertising, the dataset also contains a second random experiment. The firm distributed a flyer in some randomly selected regions (e.g., based on postal code), while this was not the case in other regions. This setup leads to a random 50% of the consumers in the dataset receiving a flyer from the firm. Since we have information in which region the consumers live (e.g., based on sign-up information from the customer or their IP address information), we know which individual consumers did get the flyer. This setup thus allows us to investigate the effectiveness of this offline advertising form at the individual consumer level. We can furthermore investigate if there are synergy effects between the banner ads and the flyer, i.e., if being exposed to both advertising forms increases the purchase likelihood beyond the two individual effects (i.e., positive synergy) or if they weaken each other since they might be substitutes (i.e., negative synergy). This setup of distributing flyers is somewhat in line with the study by Wiesel et al. (2011), although they have conducted and analyzed their experiment at a higher level of aggregation, namely at the regional level instead of the individual consumer level.

A good thing to inspect first is if consumers in the firm's banner group and who have received a flyer indeed have a higher likelihood of purchasing. Furthermore, we can explore if there is a synergy effect between the two channels.

```
# Plot of banner ad and purchase likelihood
library(ggplot2)
myData <- aggregate(consumers$Purchase,
                    by = list(Firm_banner = consumers$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Firm_banner), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                width = 0.25) +
  labs(x = "Firm_banner", y = "Conversion rate") +
  ggtitle("Conversion rate by firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%",
sep = "")), size = 4, vjust = 5)
# Plot of flyer and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Flyer_region = consumers
$Flyer_region),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Flyer_region", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Flyer_region, "Flyer_region"))
p <- ggplot(data = myData, aes(x = factor(Flyer_region), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
               width = 0.25) +
 labs(x = "Flyer_region", y = "Conversion rate") +
  ggtitle("Conversion rate by Flyer_region") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%",
sep = "")), size = 4, vjust = 5)
# Plot of banner ad + flyer and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Firm_banner = consumers$Firm_banner,
Flyer_region = consumers$Flyer_region),
                     FUN = function(x) c(mean = mean(x), sd = sd(x),
                                         n = length(x)))
myData <- do.call(data.frame, myData)
```

```
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "Flyer_region", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner /",
                        myData$Flyer_region, " Flyer_region"))
p <- ggplot(data = myData, aes(x = factor(Firm_banner), y = mean,
                               fill = factor(Flyer_region)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
               width = 0.25) +
 labs(x = "Firm_banner", y = "Conversion rate") +
  ggtitle("Conversion rate by firm_banner and flyer_region") +
  scale_fill_discrete(name = "Flyer region") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
```

The left plot of Fig. 5 shows for banner advertising the difference between consumers in the treatment group (i.e., those in the firm ad group) versus the control group (i.e., those in the charity ad group). We can see that the conversion rate is 27.77% for the treatment group, which is significantly higher than the control group's 16.88% conversion rate. Banner advertising does thus seem to be very effective, increasing the purchase likelihood with 10.89% points. This value can be used as an input to perform ROI calculations for banner advertising. If we assume the average consumer gets 2.16 banner impressions (based on the mean number of banner impressions with and without a click, see Table 1), the cost per mile (CPM, i.e., the costs of 1,000 banner impressions) is $10, and the value of a conversion (i.e., profit before marketing costs) is $1, then the ROI would be:

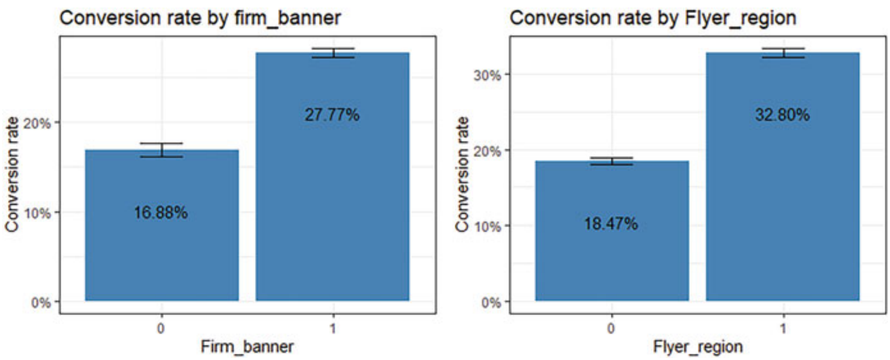

Fig. 5 Banner and flyer effectiveness visualized

$$ROI = \frac{\Delta coversion \cdot profit - impressions \cdot \frac{CPM}{1000}}{impressions \cdot \frac{CPM}{1000}} \cdot 100\%$$
$$= 0.1089 \cdot 1 - 2.16 \cdot \frac{\frac{10}{1000}}{2.16 \cdot \frac{10}{1000 \cdot 100\% = 404.17\%}} \quad (1)$$

Where "$impressions \cdot \frac{CPM}{1000}$" are the costs per consumer for targeting him or her with banner advertising, while "$\Delta coversion \cdot profit$" is the improvement in profitability before taking the marketing costs into account. In this case, we can see that the ROI is 404.17%, i.e., banner advertising is very profitable on average. Note that this is a generated dataset and that in reality the effectiveness of banner advertising is usually much smaller compared to this example. Hoban and Bucklin (2015) have found an uplift of between 0.065 and 0.985% points in the purchase likelihood, depending on the user segment.

The right-hand graph of Fig. 5 shows that receiving a flyer also significantly increases the purchase likelihood from 18.47% to 32.80%. The effect of receiving a flyer is thus even stronger compared to the effectiveness of banner advertising, although a flyer might also be much more expensive than banner impressions. When we know the cost of distributing a flyer, we can again use this to calculate the ROI of flyers and decide to invest in it.

Figure 6 shows the synergy effect of banner advertising and the flyer. When the firm does not distribute a flyer, being in the firm's banner group increases the purchase likelihood from 13.61–20.26%, i.e., an increase in conversion of 6.65% points. When the firm distributes a flyer, being in the firm's banner group increases the purchase likelihood from 19.70–35.91%, i.e., an increase in conversion of 16.21% points. In other words, when the firm distributes a flyer, the banner becomes more effective. Such positive synergy between on- and offline advertising is in line with some studies, e.g., Lesscher et al. (2021) and Pauwels et al. (2016a).

To formally test the direct and synergy effects, we can also use a logistic regression model, as done with the following R-script.

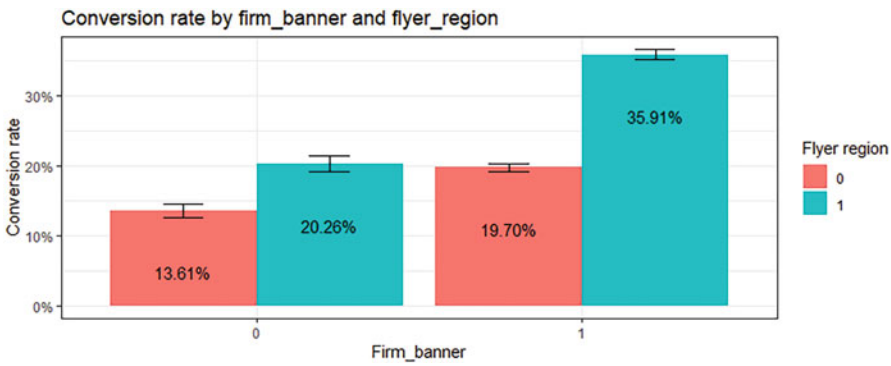

Fig. 6 Banner and flyer synergy visualized

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Logistic regression model of banner ad treatment + flyer and
purchase likelihood
model1 <- glm(Purchase ~ Firm_banner + Flyer_region,
data=consumers, family=binomial)
model2 <- glm(Purchase ~ Firm_banner*Flyer_region, data=consumers,
family=binomial)
tab_model(model1, model2, transform = NULL, collapse.ci = TRUE,
p.style = "stars")
```

The first two columns of Table 4, i.e., Model 3 and 4, show the output of these models. Model 3 shows that being in the firm's banner group and living in a flyer region significantly impact the purchase likelihood. We can also see that the parameter for "flyer region" is larger in value, i.e., we can conclude that the flyer is most effective in increasing the purchase likelihood. These findings are all in line with what we could see in Fig. 5.

When we look at Model 4 from Table 4, we can also see the interaction effect. The interaction, i.e., the parameter for "Firm_banner * Flyer_region," is statistically significant and positive, meaning that there is indeed a positive synergy, i.e., the combined effect of the two channels is larger than the sum of the two channels' individual effects. This finding is in line with what we have observed in Fig. 6.

To interpret these parameters and the synergy effect, let us take an example. By looking at Model 4 in Table 4, we can observe that being in the firm banner (i.e., treatment) group positively impacts purchase. The parameter of 0.44 indicates that the odds of purchasing are ([exp(0.44)-1]*100%) 55.3% higher for consumers in the firm ad group than consumers in the charity ad group, provided that they did not receive a flyer. If they did receive a flyer, then the log-odds increases with 0.35 (i.e., the parameter for the interaction effect), meaning that the odds of purchasing are ([exp(0.44 + 0.35)-1]*100%) 120.3% higher for consumers in the firm ad group than consumers in the charity ad group, provided that they do receive a flyer. For more details on interpreting the parameters of a logistic regression model and how to recalculate this to probabilities, please check Chapter 8 of Leeflang et al. (2015).

We can easily extend the estimated models, for instance, to find out in which situations a marketing instrument is more or less effective. Let us, for this part, only focus on banner advertising. One assumption would be that the higher purchase likelihood due to the banner ad will only occur for consumers who have actually seen the banner, i.e., the number of impressions should be above zero to have an effect. For consumers who have not seen the banner, being randomly allocated in the firm ad group or charity ad group should not make a difference in the purchase likelihood. We can test this assumption with the following R-script.

```
# Create some new variables
consumers$Banner_exposures <- consumers$Banner_no_click +
consumers$Banner_click
```

Table 4 Logistic regression output based on randomized field experiments

	Model 3	Model 4	Model 5	Model 6	Model 7
	Purchase	Purchase	Purchase	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*
(Intercept)	−2.02 ***(−2.08 – −1.97)	−1.85 ***(−1.93 – −1.77)	−1.54 ***(−1.64 – −1.45)	−3.79 ***(−4.20 – −3.42)	−1.65 ***(−1.73 – −1.57)
Firm_banner	0.65 *** (0.59–0.71)	0.44 *** (0.36–0.53)	−0.01(−0.12–0.09)	−0.01(−0.43–0.44)	−0.04(−0.13–0.05)
Flyer_region	0.77 *** (0.73–0.82)	0.48 *** (0.37–0.58)			
Firm_banner * Flyer_region		0.35 *** (0.23–0.46)			
Banner_seen			−0.07(−0.19–0.04)	0.29(−0.13–0.74)	
Firm_banner * Banner_seen			0.90 *** (0.77–1.02)	2.12 *** (1.62–2.58)	
Existing_customer				2.80 *** (2.42–3.23)	
Firm_banner * Existing_customer				−0.03(−0.50–0.40)	
Banner_seen * Existing_customer				0.04(−0.43–0.47)	
(Firm_banner * Banner_seen) * Existing_customer				−1.49 ***(−1.97 – −0.98)	
Banner_exposures +1 [log]					0.06(−0.01–0.13)
Firm_banner * Banner_exposures +1 [log]					0.70 *** (0.62–0.78)
Observations	50,000	50,000	50,000	50,000	50,000
R^2 Tjur	0.038	0.039	0.030	0.128	0.062

* *$p < 0.05$* ** *$p < 0.01$* *** *$p < 0.001$*

```
consumers$Banner_seen <- ifelse(consumers$Banner_exposures==0,0,1)
# Plot of banner ad group + banner ad seen and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Banner_seen = consumers$Banner_seen,
Firm_banner = consumers$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                              fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Logistic regression model of banner ad treatment + banner ad seen
and purchase likelihood
model3 <- glm(Purchase ~ Firm_banner + Banner_seen +
Firm_banner*Banner_seen, data=consumers, family=binomial)
tab_model(model1, model2, model3, transform = NULL, collapse.ci =
TRUE, p.style = "stars")
```

As the two bars on the left in Fig. 7 show, there is no substantial difference between the control and treatment groups if there were no banner ad impressions. In both cases, the purchase likelihood is just over 17%. This nonsignificant difference is indeed in line with what we would expect; if there is no ad impression, there should be no difference between the firm and charity ad groups. Investigating this is also a good test if the randomization of the experiment worked; if the difference was significant, this would indicate that something might have gone wrong with the randomization.

The two bars on the right of Fig. 7 show a substantial difference between the two groups, provided that there was at least one banner ad exposure. Consumers in the control group who saw the charity banner at least once have a purchase likelihood of 16.56%. Consumers in the treatment group who saw the firm’s banner at least once have a purchase likelihood of 32.46%. Being exposed to the banner at least once does have a positive effect since it increases the purchase likelihood by 15.90% points.

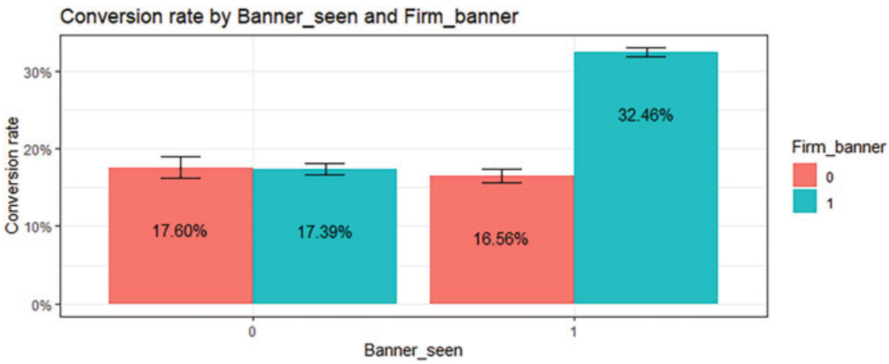

Fig. 7 Banner effectiveness when the banner is seen visualized

The regression output, presented in Model 5 of Table 4, confirms what we see in Fig. 7. The parameter “Firm_banner” is insignificant, which indicates that when the number of ad exposures is zero, the treatment and control groups do not differ in their purchase likelihood. The parameter for “Banner_seen” is also insignificant, indicating no difference in the purchase likelihood between the consumers who did or did not see the charity ad for the control group. This insignificance aligns with what we might expect since the charity ad is irrelevant and should not cause sales at the firm. However, this parameter could, in theory, be significant since it does capture the difference between being more active online (on websites where the campaign is running), i.e., it captures the confounding effects of being exposed to a banner ad.

The parameter for the interaction effect between “Firm_banner” and “Banner_seen” captures the causal effect of seeing the banner ad of the firm. In Model 5 of Table 4, we can see that this interaction is significant and positive, indicating that consumers who saw the firm’s banner ad are more likely to purchase due to the exposure to the firm’s ad.

Now that we know that the banner ad is, on average, effective in generating purchases, we might want to find out how the effects differ between consumers. For this, we can add additional interaction effects to our regression equation. To illustrate this, we can investigate if the banner works better for potential or existing customers by running the following R-script.

```
# Plot of banner ad treatment + banner ad seen for new/potential
customers
myData <- consumers[consumers$Existing_customer==0,]
myData <- aggregate(myData$Purchase,
                    by = list(Banner_seen = myData$Banner_seen,
Firm_banner = myData$Firm_banner),
                   FUN = function(x) c(mean = mean(x), sd = sd(x),
                                       n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
```

```
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                               fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                    width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner for
potential/new customers") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 2)
# Plot of banner ad treatment + banner ad seen for existing
customers
myData <- consumers[consumers$Existing_customer==1,]
myData <- aggregate(myData$Purchase,
                    by = list(Banner_seen = myData$Banner_seen,
Firm_banner = myData$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                               fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
            position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge(0.9),
                width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner for
existing customers") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Logistic regression model of banner ad treatment + banner ad seen
and purchase likelihood
model4 <- glm(Purchase ~ Firm_banner*Banner_seen*Existing_customer,
data=consumers, family=binomial)
```

```
tab_model(model1, model2, model3, model4, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

The top graph of Fig. 8 shows the impact of banner advertising for new (i.e., potential) customers. What can be seen is that when there are no ad impressions (i.e., no treatment, the two left bars), the control and treatment group do not differ. This makes sense, since the groups do not differ from each other. When the ad is seen, i.e., there is at least one ad impression, the two groups differ from each other. When the new/potential customer sees the charity banner, the purchase likelihood is 2.93%, and with the firm banner this is 19.85%. The causal impact of banner advertising is thus 16.92% points increase in purchase likelihood.

The bottom graph of Fig. 8 shows the effects for existing customers. We can see that they have a higher purchase likelihood than new customers, even when they do not see the banner. This makes sense, since existing customers have already made a purchase before and are more likely to come back compared to someone who has not made a purchase before. We also can see that there is again no significant difference between the control group and the treatment group when the banner ad is not seen, which again is what we would expect. If the control group sees the charity banner, the purchase likelihood goes up from 27.16% to 34.11%. This significant increase

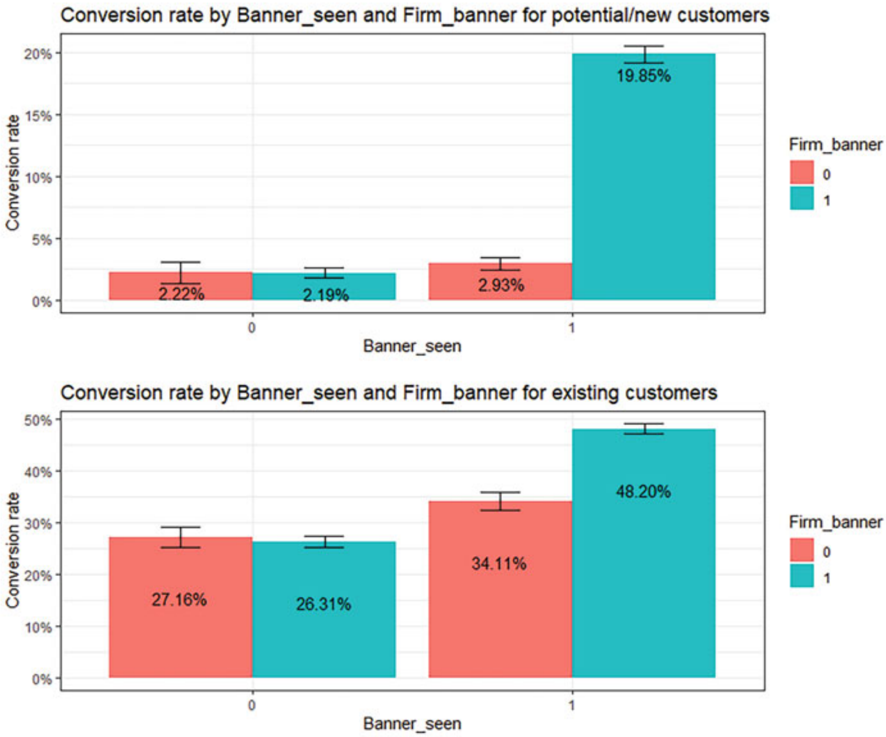

Fig. 8 Banner effectiveness difference between potential and existing customers visualized

can be explained by the fact that the ones seeing the charity ad are more online and they visit the website where the campaign is running, i.e., because they are more active, they might be more likely to purchase. This part of the effect is however not driven by the ad itself. For the treatment group, we do however see that the growth in the purchase likelihood is significantly larger when they see the banner ad at least once. This extra increase of ([48.20–26.31%] – [34.11–27.16%]) 14.94% points is the causal effect of the banner ad for existing customers. We can see that this effect of the banner for existing customers is smaller (a bit smaller in percentage points, and a lot smaller in percentages) than for new customers, where the purchase likelihood jumps up when exposed to the firm's banner.

To test the significance of the effects, we can look at Model 6 in Table 4. The parameter for "Firm_banner" and its interaction with "Existing_customer" are both insignificant, meaning that the randomization has no significant effect on the purchase likelihood provided that there are no ad impressions. The parameter for "Banner_seen" is insignificant, which indicates that seeing the charity banner does not increase the purchase likelihood. The parameter for "Existing_customer" is statistically significant, indicating that existing customers are more likely to conduct a purchase.

The interaction between "Firm_banner" and "Banner_seen" is significant, as expected, which in this case means that for potential customers, seeing the firm's banner at least once has a positive impact on the purchase likelihood. The three-way interaction, which also includes "Existing_customer," is significant and negative; this means that, for existing customers, exposures to the firm's banner are significantly *less* effective than for potential customers. The banner is still effective for existing customers since the two-way interaction minus the three-way interaction remains positive (i.e., 2.12–1.49 = 0.63) (One can also test the significance of this.63, by changing the "Existing_customer" dummy in a "New_customer" dummy in Model 6 (i.e., this dummy is the opposite of the "Existing_customer" dummy).). This positive but smaller impact of banner advertising for existing customers aligns with what we saw in Fig. 8. If the firm wants to target banners, it is thus more effective to target the potential customers instead of the existing customers.

Finally, with the following R-script, we can also test the impact of having more ad exposures.

```
# Plot of banner ad treatment + number of exposures and purchase
likelihood
myData <- consumers
myData$Banner_exposures <- ifelse(myData$Banner_exposures > 6, 6,
myData$Banner_exposures)
myData$Firm_banner <- as.factor(myData$Firm_banner)
myData <- aggregate(myData$Purchase,
                   by = list(Firm_banner = myData$Firm_banner,
Banner_exposures = myData$Banner_exposures),
                   FUN = function(x) c(mean = mean(x), sd = sd(x),
                                       n = length(x)))
myData <- do.call(data.frame, myData)
```

```
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "Banner_exposures", "mean",
"sd", "n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner /",
                    myData$Banner_exposures, " Banner_exposures"))
p <- ggplot(data = myData, aes(x = Banner_exposures, y = mean,
                             group = as.factor(Firm_banner), color =
Firm_banner))
p + geom_line() +
  geom_point() +
  geom_errorbar(aes(ymin=mean - 2*se, ymax=mean + 2*se), width=.2,
                position=position_dodge(0.05)) +
  labs(x = "Number of banner exposures", y = "Conversion rate") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  ggtitle("Conversion rate by number of ad exposures") +
geom_text(position = position_dodge(width= .9), aes(y = mean, label
= paste(format(mean*100, digits = 2, nsmall = 2), "%", sep = "")), size =
4, vjust = -1)
# Logistic regression model of banner ad treatment + number of
exposures and purchase likelihood
model5 <- glm(Purchase ~ Firm_banner*log(Banner_exposures+1),
data=consumers, family=binomial)
tab_model(model1, model2, model3, model4, model5, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

The bottom (red) line in Fig. 9 shows the impact of zero (first bar) up to six-plus (seventh bar, six and more ad impressions are aggregated due to the size of the graph and the small amount of observations with many ad impressions) ad impressions for the control group (i.e., those in the charity ad group). We do not see much change in purchase likelihood here, e.g., the purchase likelihood is 17.60% with zero ad impressions, and this increases to 19.00% with six or more impressions, all falling within the same confidence interval.

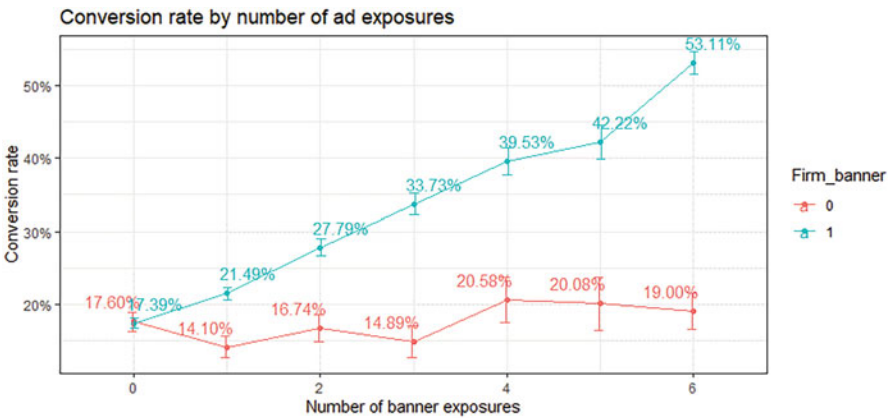

Fig. 9 Banner effectiveness depending on the impression amount visualized

The top (blue) line in Fig. 9 shows the impact of having more banner ad exposures for consumers in the treatment group, i.e., those exposed to the firm's ad. Here we can see a clear positive impact of having more ad exposures on the purchase likelihood. The difference between the treatment group and the control group is the *incremental effect* of the firm's banner ad. The incremental effect seems to increase with the number of banners, indicating that a higher number of banner ads increases conversion further. When the incremental effect would stabilize, this would signal that an additional banner impression is not worthwhile. This kind of analysis can help to investigate the optimal amount of banner impressions. For more details, also see the study by Hoban and Bucklin (2015). The timing of advertising can be determined similarly, as Braun and Moe (2013) demonstrate in their study. Försch and De Haan (2018) study a combination of frequency and timing of advertising.

Model 7 in Table 4 shows the parameter estimates of the frequency effect. Again, the parameter for "Firm_banner" is insignificant, indicating that the treatment and control groups do not differ in their purchase likelihood when there are zero ad exposures, which aligns with what we expect. Furthermore, "log(Banner_exposures + 1)" is insignificant, indicating that a higher number of exposures to the charity ad does not significantly impact the purchase likelihood. This parameter captures the confounding effects. The interaction effect captures the causal impact of the firm's banner ad and is, therefore, the parameter of primary interest. This interaction effect is statistically significant and positive, meaning that the higher number of exposures to the firm's banner causes an increase in the likelihood to purchase. These results are again in line with Fig. 9.

The models discussed in this section are easily adaptable. It is, for instance, possible to test if receiving a flyer also influences the impact of the number of ad exposures or if the effects of the number of ad impressions are nonlinear. If there is information on the websites where the campaign is running, it is also possible to investigate if an ad exposure from website A has a different impact on the purchase likelihood than an ad exposure from website B. Similarly, it is possible to test the impact of different ad creatives by having different groups of people exposed to different ad creatives. Such a setup is also called *A/B testing*. It is furthermore possible to test other channels with a randomized experiment.

An advantage of a randomized field experiment is that we can be sure of the causality, provided there is true randomization, and there are thus no other differences between the two groups. A downside is that a randomized field experiment at the consumer level is not possible for all channels. Luckily this can be overcome by conducting the experiment at a more aggregate level, e.g., varying between regions or conducting changes over time, as explained in the section "Attribution with Aggregate-Level (Quasi-)Experimental Data" of this chapter. Randomization is impossible for some other channels since consumers decide which channel to use, i.e., the channel exposure cannot be randomized. In such cases, other techniques have to be applied, as discussed in the next section. Another downside of conducting field experiments is that it can be challenging and time consuming to do this for all channels.

Furthermore, when investigating the interaction effects between channels, e.g., if online and offline advertising might make each other stronger, a more complicated experimental design is needed. Such a design is doable for two channels (e.g., via a two-by-two experimental design, as done in our example in this section), but this will be practically impossible for a large number of channels. In such cases, different methods are appropriate, as explained in the "Exploring Paths in Conversion (Markov Chain)" section of this chapter.

Additional Topics for Individual-Level Attribution

This section discusses some more advanced attribution methods suitable for individual-level data. In some cases, conducting field experiments is not possible or one is interested in the customer journey that is affected by a certain channel, without all potential channels being the subject of an experiment. In these cases, it is still possible to conduct attribution and find the causal impacts of specific (combinations of) channels, although the methods become more complex and there might be less certainty that all criteria of causality do indeed hold. We will start with propensity score matching, which is suitable when the channel is not subject to an experiment but one still wants to explore the impact as a quasi-experiment. The second method uses Markov chains to explore how different channels affect the journey, and the outcome of a journey, which provides more detailed insights compared to looking at (combinations of) channels in isolation. We will end this section with a discussion of other methods and further procedures for individual-level attribution.

Attribution with Individual-Level Quasi-Experimental Data (Propensity Score Matching)

In some cases, a random experiment is not possible, feasible, or desirable, but we still want to draw causal conclusions. If we, for instance, observe that customers who signed up for a loyalty program are buying more frequently, can we then conclude that this is due to the loyalty program? Or are customers who purchase more also more likely to join a loyalty program? Or is it a combination of the two? With a loyalty program, it is impossible to randomly sign-up people since customers have to give their consent to sign up, and it is undesirable to deny random customers' membership to such a program. A random experiment is thus impossible, infeasible, or undesirable to do in such a situation. For other channels, e.g., email, it is possible to not send out an email to a random group of customers to test the effectiveness of this channel, but this might be undesirable, for instance, because the customers expect to receive the email for which they have signed up.

In the cases described above, a feasible alternative is to conduct a quasi-experiment, which is possible via *propensity score matching* (PSM). With PSM, we link consumers to each other who are equally likely to get exposed to a channel, but due to random chance, one did get exposed, and the other one did not get exposed. If we do this for many pairs of consumers, we get two groups of which both had an equal chance of exposure, but one had the exposure (i.e., the treatment group) and the other did not (i.e., the control group). If we have these two groups, we can then investigate the channel the same way as with a randomized field experiment, as described in the previous section. We can thus test if the two groups differ in terms of their outcome (e.g., purchase likelihood), which would provide us the causal impact of the channel.

Multiple studies use PSM to conduct attribution. One example is the study by De Haan et al. (2018), who have instigated what impacts device switching in an online customer journey (e.g., a person starts looking for information on a smartphone and then switches to a laptop). Since device usage and device switching is not a random decision and is not randomizable with an experiment, De Haan et al. (2018) matched sessions that were equally likely to have a specific device switch, but in some cases, the switch did occur while in other cases this switch did not occur. De Haan et al. (2018) find that switching from a mobile device to a nonmobile device increases the conversion probability, especially if the product the customer is investigating is riskier (e.g., the product category has a higher perceived risk, the customer has less experience buying in the category, or the product is relatively expensive).

To conduct PSM, we have to go through the following five steps (see Chapter 1 of Pan and Bai 2015 for more details):

1. Estimate the likelihood of being treated (e.g., subscribing to an email newsletter)
2. Check for the overlap in the propensity scores (e.g., likelihood of subscribing) of treatment and control group
3. Match observations from the treatment group with observations from the control group who have a similar propensity score
4. Verify the balance of covariates to check if the matching was successful
5. Conduct multivariate analysis based on the matched sample, similar as we have done in the "Attribution Modeling Process with Experimental Data" section

A logistic regression model can estimate the treatment likelihood, with treatment (e.g., receiving an email newsletter) as a dependent variable and the drivers of treatment as independent variables. The independent variables should be exogenous, which means that the treatment does not influence these variables. The number of website visits might be a good predictor for email subscription, but the email subscription might also drive website visits. Rubin (2001) recommends selecting the independent variables based on theory and prior research.

Demographic variables might work in our case since this might be related to the probability of signing up for the email service, and the email does not influence people's demographics. Unfortunately, however, we do not have demographic variables in our dataset. Since the email service started on the first of January (see Table 1), and we have the relationship length and the CLV up until the first of January, we can use these latter two variables as independent variables; since these two variables provide information before the email service's introduction, the email cannot have caused changes in the relationship length or CLV.

For the PSM, we will be using the "MatchIt" package in R (Ho et al. 2021). The following R-script estimates the propensity scores and matches the consumers who have a similar propensity to sign up for the email service.

```
# Load packages
library(MatchIt)
library(cobalt)
# Only include existing customers, since they are the only ones
signing up to the email (for potential customers, email sign up is
zero in our dataset)
existing_customers <- consumers[consumers$Existing_customer==1,]
# Log transform CLV, since it is highly skewed
existing_customers$Log_CLV <- log(existing_customers$CLV)
# Conduct matching of consumers who did and did not subscribe to the
email (i.e. step 1 and 3 of the PSM process). In this example, the
Email_group membership is predicted using the Relation_length and
the log of CLV. The matching is automatically performed with the
code.
m.out <- matchit(Email_group ~ Relation_length + Log_CLV,
                 data=existing_customers, caliper=0.05)
# Investigate the PSM outcome
summary(m.out, standardize = TRUE)
love.plot(m.out)
```

In the R-script, the term "caliper" indicates the number of standard deviations the matched consumers' propensity score can maximum be apart. If we set this value higher, this would result in consumers who being less similar to each other can be matched. This higher value does result in more matched consumers, but the difference between the matched consumers can get larger, which reduces the appropriateness of PSM since we want to create comparable groups. If we set the value of "caliper" to zero, only consumers with identical propensity scores are matched, resulting in more similar samples, but this can substantially reduce the sample size since there might be few (or sometimes even non) exact matches. Determining the appropriate value for "caliper" can be achieved by trial and error, e.g., by checking what happens to the outcome of step 4; if the samples are unbalanced, the value of "caliper" should decrease; if the sample is very well balanced, but the sample sizes are too small, one can test if a higher value of "caliper" also still works.

Figure 10 shows the output of the results of the matching. As shown at the bottom of this Figure, we have 11,727 observations in the control group (i.e., those who do not receive an email) and 13,304 consumers in the treatment group (i.e., those who do receive an email). After matching, we have 10,257 observations in both groups. At the top of Fig. 10, we see some descriptive statistics. The mean relationship lengths before matching are 32.61 months and 28.04 months for the treatment and control group, respectively, which is equal to a 0.27 standard deviation difference. After matching, this difference drops to 0.02 standard deviations. A similar pattern is visible for the logarithmic of CLV and the "distance" (i.e., the polarity score). As the

```
Call:
matchit(formula = Email_group ~ Relation_length + Log_CLV, data = existing_customers,
    caliper = 0.05)

Summary of Balance for All Data:
                Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance               0.5528        0.5074          0.4538     0.8578    0.1210   0.1822
Relation_length       32.6097       28.0416          0.2669     1.0013    0.0761   0.1192
Log_CLV                7.3907        7.1379          0.3684     0.7584    0.0956   0.1366


Summary of Balance for Matched Data:
                Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance               0.5309        0.5273          0.0356     1.0395    0.0110   0.0235
Relation_length       29.9863       29.5953          0.0228     0.9991    0.0065   0.0169
Log_CLV                7.2835        7.2649          0.0272     1.0224    0.0096   0.0246
                Std. Pair Dist.
distance                 0.0356
Relation_length          0.8579
Log_CLV                  0.7120

Percent Balance Improvement:
                Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance                   92.2       74.7      90.9     87.1
Relation_length            91.4       34.1      91.4     85.9
Log_CLV                    92.6       92.0      90.0     82.0

Sample Sizes:
          Control Treated
All         11727   13304
Matched     10257   10257
Unmatched    1470    3047
Discarded       0       0
```

Fig. 10 R-output of “MatchIt”

“Percent Balance Improvement” section shows, the differences between the means for distance, relation length, and the log of CLV of the two groups drop by 92.2%, 91.4%, and 92.6%, respectively. In other words, after matching, the groups are much more balanced.

Figure 11 visualizes the standard mean differences, as presented in Fig. 10, which clarifies that the matched (adjusted) sample is much more balanced than the unmatched (unadjusted sample). As can be seen, the mean difference between the two groups drops substantially when going from the unmatched sample (red dots on the right-hand side in Fig. 11) to the matched sample (blue dots on the left-hand side in Fig. 11). In the matched sample, the differences are close to zero, which is again what we want to achieve.

Next, we want to investigate steps two and four of the PSM procedure, which we can do with the following R-script.

```
# Plot balance diagnostics (i.e., step 4 of the PSM process)
plot(m.out, type = "jitter", interactive = FALSE)
plot(m.out, type = "qq", interactive = FALSE)
bal.plot(m.out, which = "both")
bal.plot(m.out, var.name = "Relation_length", which = "both")
bal.plot(m.out, var.name = "Log_CLV", which = "both")
```

Figure 12 shows the distribution of the propensity scores for the matched treatment and control groups (i.e., the middle two groups) and the propensity scores of the unmatched consumers (i.e., the top and bottom groups). By looking at this Figure, there seems to be a very similar distribution and a good overlap in the propensity scores of the two matched groups, i.e., we seem to meet the criterion mentioned in the second step of the PSM procedure ("Check for the Overlap in the Propensity Score").

Figure 13 shows the Q-Q plot of the two variables which we used to match the consumers. In this particular plot, we want the observations close to the diagonal line, which indicates that the matched consumers in the treatment and control groups are the same. As can be seen, if we use all consumers (i.e., matched and unmatched), there is quite some deviation from the diagonal line, i.e., the consumers are not very similar. After matching, the observations almost perfectly follow this diagonal line, indicating that the matching worked, and we meet the fourth step of the PSM procedure ("Verify the balance of covariates to check if the matching was successful").

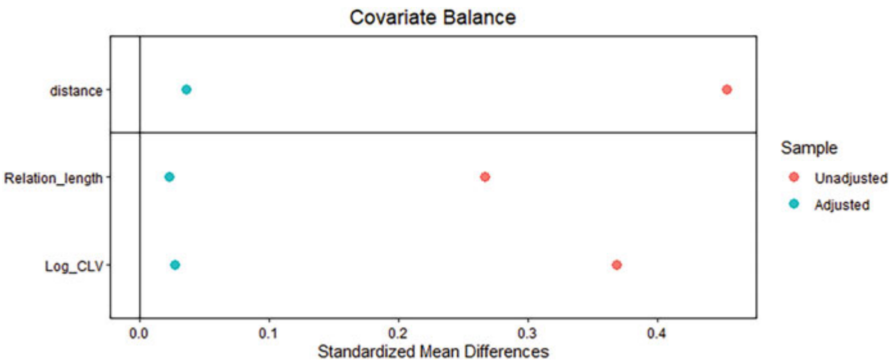

Fig. 11 R-output of covariate balance

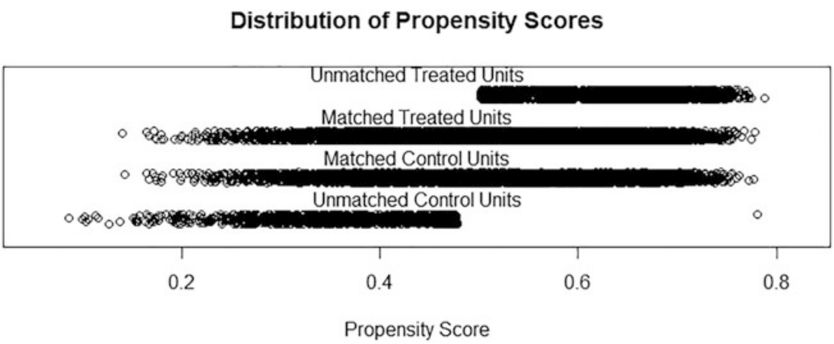

Fig. 12 R-output of propensity scores distribution for matched and unmatched groups

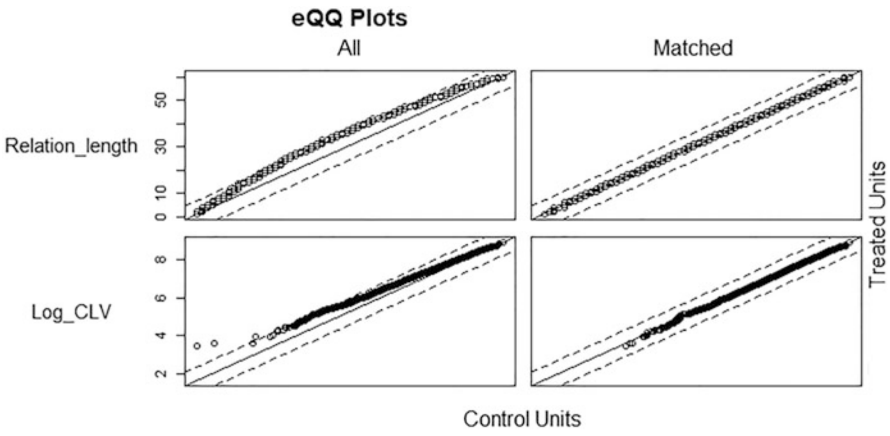

Fig. 13 R-output of QQ-plots for matched and all samples

Figure 14 is somewhat similar to the previous two figures but visualizes the overlap in a density plot. Again, the distributions of the "distance" (i.e., the propensity scores), relation length, and the logarithmic of CLV are not very well overlapping before matching (i.e., the graphs on the left), while after matching, they are very nicely overlapping (i.e., the graphs on the right). Again, we thus seem to meet the criteria of steps two and four of the PSM procedure.

Since the matching seems to have worked, we can use this new matched sample similar to the randomized field experiment discussed in "Attribution Modeling Process with Experimental Data" section of this chapter. For email, we can, for instance, check if (after matching) there is a significant effect on the likelihood to purchase and if this effect differs for different customers, e.g., those who were in the firm's ad group or the charity ad group.

```
# Get a dataframe with the matched data
matched_data <- match.data(m.out)
# Plot of email and purchase likelihood (unmatched)
myData <- aggregate(existing_customers$Purchase,
                   by = list(Email_group = existing_customers
$Email_group),
                   FUN = function(x) c(mean = mean(x), sd = sd(x),
                                      n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Email_group", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Email_group, "Email_group"))
p <- ggplot(data = myData, aes(x = factor(Email_group), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                   ymin = mean - 2*se), position = position_dodge
                   (0.9),
```

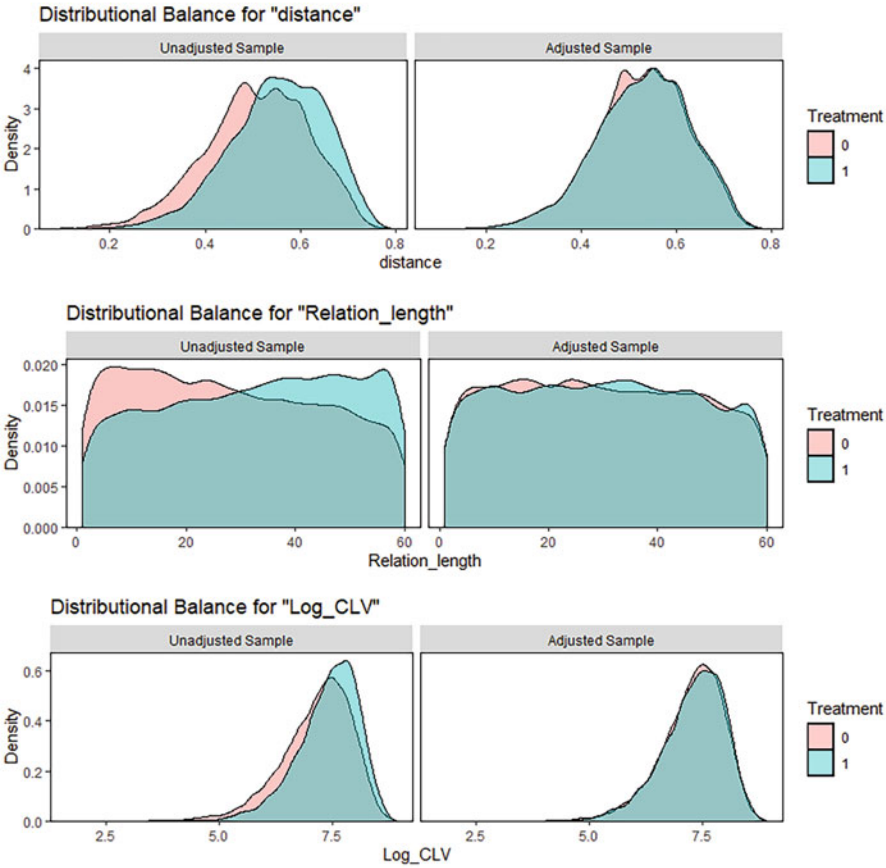

Fig. 14 Distribution balance of “distance” (top), “Relation_length” (middle), and “Log-CLV” (bottom)

```
                    width = 0.25) +
  labs(x = "Email_group", y = "Conversion rate") +
  ggtitle("Conversion rate by Email_group (full sample)") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Plot of email and purchase likelihood (matched)
myData <- aggregate(matched_data$Purchase,
                         by = list(Email_group =
                         matched_data$Email_group),
                         FUN = function(x) c(mean = mean(x), sd = sd(x),
                                              n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Email_group", "mean", "sd", "n", "se")
```

```
myData$names <- c(paste(myData$Email_group, "Email_group"))
p <- ggplot(data = myData, aes(x = factor(Email_group), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
               ymin = mean - 2*se), position = position_dodge(0.9),
               width = 0.25) +
  labs(x = "Email_group", y = "Conversion rate") +
  ggtitle("Conversion rate by Email_group (matched sample)") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Estimate logistic regression models
model1 <- glm(Purchase ~ Email_group, data=matched_data,
family=binomial)
model2 <- glm(Purchase ~ Email_group, data=existing_customers,
family=binomial)
model3 <- glm(Purchase ~ Email_group + Firm_banner +
Email_group*Firm_banner, data=matched_data, family=binomial)
model4 <- glm(Purchase ~ Email_group + Firm_banner +
Email_group*Firm_banner, data=existing_customers, family=binomial)
tab_model(model1, model2, model3, model4, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

Figure 15 shows the purchase likelihood for the full sample (left graph) and the matched sample (right graph). Before matching, the consumers who receive an email have a purchase likelihood of 41.27%, while those who do not receive the email have a purchase likelihood of 34.36%. There is thus a 6.91% point higher purchase likelihood for the consumers who do receive an email. This 6.91% point difference is a combination of the actual causal effect of the email and the confounding effects (e.g., those signing up for an email are already more loyal and have a higher purchase likelihood to begin with). If we look at the matched sample, the difference between

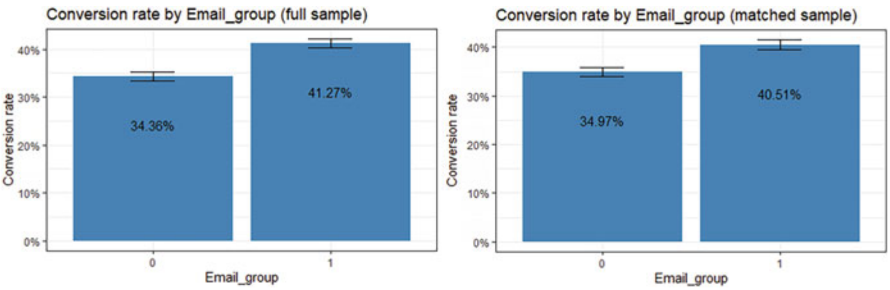

Fig. 15 Email effectiveness for unmatched (left) and matched (right) samples visualized

the consumers who do and those who do not receive the email drops to 5.54% points. This 5.54% point increase in the conversion likelihood is thus the true causal impact of the email.

The output of the four logistic regression models, which replicate Fig. 15, can be seen in Table 5. Model 8 and Model 10 in Table 5 are estimated on the matched sample, while Model 9 and Model 11 use the full sample. Model 9 shows the effect of email is significant and 0.29, meaning that people who receive an email have a ([exp(0.29)-1]*100%) 33.6% higher odds of purchasing than consumers who did not receive an email. Since this second model uses the full sample, it does include some confounds due to the self-selection to subscribe to the email; those who receive an email might be more loyal customers and have a higher likelihood of purchasing. Comparing the two groups is thus a bit comparing apple and oranges, and we cannot interpret the parameter as a causal effect of email on purchase likelihood.

Model 8 in Table 5 uses the matched sample. The two groups are in this sample comparable to each other and only differ in the subscription to the email service. The parameter estimate is now still significant and has a value of 0.24, meaning that people who receive an email have a ([exp(0.24)-1]*100%) 27.1% higher odds of purchasing consumers who did not receive an email; due to the matching we can assume that this effect is causal. Not controlling for the self-selection does thus somewhat overestimate the impact of email on the purchase likelihood, which is in line with what we could see in Fig. 15.

Model 10 and 11 from Table 5 test if there is a synergy effect between email and exposure to the firm banner. Since the interaction effect is insignificant in both models, we cannot conclude that there is a synergy effect.

PSM is a good way to find the effectiveness of a channel in situations where conducting a randomized field experiment is not an option. A challenge here is that the variables used for matching should be exogenous, i.e., the variables should

Table 5 Logistic regression output of matched (columns 1 and 3) and full (columns 2 and 4) samples

	Model 8	Model 9	Model 10	Model 11
	Purchase	Purchase	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*
(Intercept)	−0.62 *** (−0.66 – −0.58)	−0.65 *** (−0.69 – −0.61)	−0.97 *** (−1.06 – −0.87)	−0.99 *** (−1.08 – −0.90)
Email_group	0.24 *** (0.18–0.29)	0.29 *** (0.24–0.35)	0.31 *** (0.18–0.45)	0.38 *** (0.26–0.50)
Firm_banner			0.42 *** (0.32–0.53)	0.42 *** (0.32–0.52)
Email_group * Firm_banner			−0.09 (−0.24–0.06)	−0.10 (−0.24–0.03)
Observations	20,514	25,031	20,514	25,031
R^2 Tjur	0.003	0.005	0.008	0.010

* *$p < 0.05$* ** *$p < 0.01$* *** *$p < 0.001$*

correlate with, but not be influenced by, the channel choice. For more details on PSM, including alternative matching procedures, check the book by Pan and Bai (2015).

Exploring Paths in Conversion (Markov Chain)

Another method used for attribution is the Markov chain. With a Markov chain, we can observe through which stages (i.e., channels or touchpoints) a consumer moves in the path to purchase and what the final stage (i.e., conversion or not) is. We can use this Markov chain to estimate the likelihood of encountering a specific touchpoint and the likelihood that a particular path will convert. Anderl et al. (2016) used a Markov approach to map the purchase journey and compare this method to basic attribution methods like last-touch attribution.

Figure 16 provides a simple example of such a graphical Markov chain with three channels. We can see that at the start of the path to purchase, the customer is most likely to use channel 2 or 3, then switches between the three channels, and channel 2 is most likely to lead to a purchase (based on last touch). Such a visualization nicely illustrates how the path to purchase looks like, i.e., how consumers switch between touchpoints and how this leads to an outcome. Instead of visualizing, such a Markov chain can also be represented with a transition matrix, as we will demonstrate later in this section.

For attribution, we can turn off a channel in the Markov chain and simulate what would happen to the path to purchase in terms of touchpoints that the consumer encounters and the outcome of the path to purchase. This is a more sophisticated procedure than touch-based attribution since we consider that the entire path to purchase can change when a channel drops out and that there might be alternative channels that pick up the role of the turned-off channel. Fig. 17 visualizes an example of what happens when we turn off a channel, e.g., the traffic to the other channels changes, and the impact on purchase is also affected.

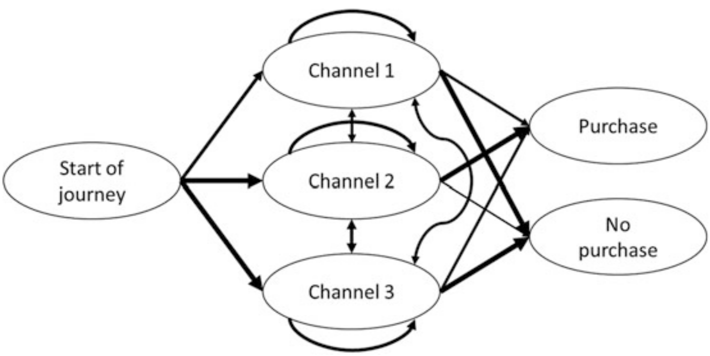

Fig. 16 Simple example Markov chain with three channels

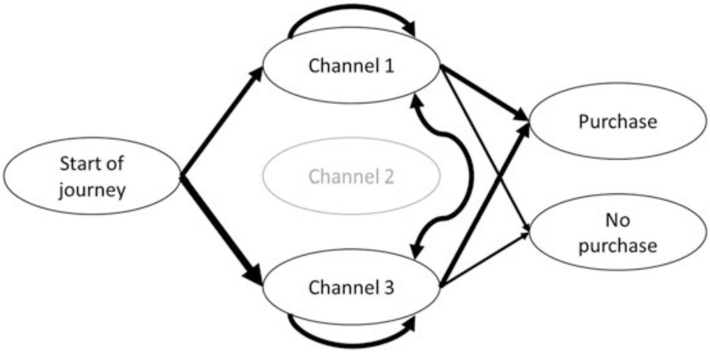

Fig. 17 Simple example Markov chain with one channel turned off

Let us with the following R-script estimate the Markov chain on our dataset. For this, we only use consumers who are in the group that could see the firm's banner. For estimating the Markov chain, we use the R-packages "msm" (Jackson 2019) and "markovchain" (Spedicato 2021).

```
# Add start channel
consumers_journey2 <- consumers_journey[consumers_journey
$Channel_number==1,]
consumers_journey2$Channel_number <- 0
consumers_journey2$Channel_name <- "Start"
consumers_journey2$Channel <- 1
consumers_journey2 <- rbind(consumers_journey, consumers_journey2)
# Add final channel (i.e., conversion or not)
consumers_journey3 <- consumers_journey[consumers_journey
$Channel_number==1,]
touchpoints <- aggregate(consumers_journey2$Channel_number,by=list
(consumers_journey2$Consumer_ID),FUN=max)
consumers_journey3$Channel_number <- touchpoints$x + 1
consumers_journey3$Channel_name <- ifelse(consumers$Purchase==1,
"Purchase", "No Purchase")
consumers_journey3$Channel <- ifelse(consumers$Purchase==1,
10, 11)
consumers_journey2 <- rbind(consumers_journey2,
consumers_journey3)
consumers_journey2 <- consumers_journey2[order(consumers_journey2
$Consumer_ID, consumers_journey2$Channel_number),]
rm(consumers_journey3)
rm(touchpoints)
# Split data up in firm and charity banner group
consumers_journey2_firm <-
 consumers_journey2[consumers_journey2$Firm_banner==1,]
consumers_journey2_charity <-
 consumers_journey2[consumers_journey2$Firm_banner==0,]
# Estimate transition matrix
library(msm)
```

```
markov_chain <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_firm)
markov_chain <- prop.table(markov_chain, margin=1)
markov_chain <- as.data.frame(t(markov_chain))
markov_chain <- markov_chain[,3]
markov_chain[100:121] <- c(rep(0,9),1,0,rep(0,10),1)
library(markovchain)
mcjourney <- new("markovchain", states = c
("Start","Banner_impression","Banner_click",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                  "Email_received", "Email_click",
"Direct_visit",
                                  "Purchase", "No purchase"),
                 transitionMatrix = matrix(data=markov_chain,
                 byrow= TRUE, nrow = 11))
show(mcjourney)
plot(mcjourney, edge.arrow.size=0.25)
```

Figure 18 visualizes the Markov chain, and Fig. 19 provides the transition matrix; both show the same information in a different format. We can observe that there is a 34.72% chance the consumer will first encounter a banner impression. When the consumer in the current state sees a banner impression, there is a 20.74% chance that the next phase will be a direct visit. A direct visit has a 6.80% chance to be followed by a purchase (i.e., conversion) and a 13.04% chance of an unsuccessful end of the path to purchase (and an 80.16% chance of continuing with one of the eight touchpoints).

Such a transition matrix can thus show what the likely next touchpoint the customer will encounter is and how likely the path to purchase will end successfully or unsuccessfully. We can make a simulation based on the transition matrix to investigate what happens after a certain number of stages. The following R-script calculates what the path to purchase looks like after 5 and 50 stages.

```
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney ^ 50)
after50touchpoints
```

When running this R-script, we can see that after five stages, i.e., five touchpoints (including the start), there is a 13.28% chance of a conversion, a 35.04% chance of the path to purchase ending with no purchase, and a 51.68% chance that the path to purchase will continue. After 50 stages, there is a 27.76% chance of a conversion, a 72.21% chance of no conversion, and a 0.03% chance of the path to purchase to continue. The 27.76% conversion is similar to the 27.77% conversion we saw for the firm banner group in Fig. 5, i.e., this transition matrix nicely matches reality.

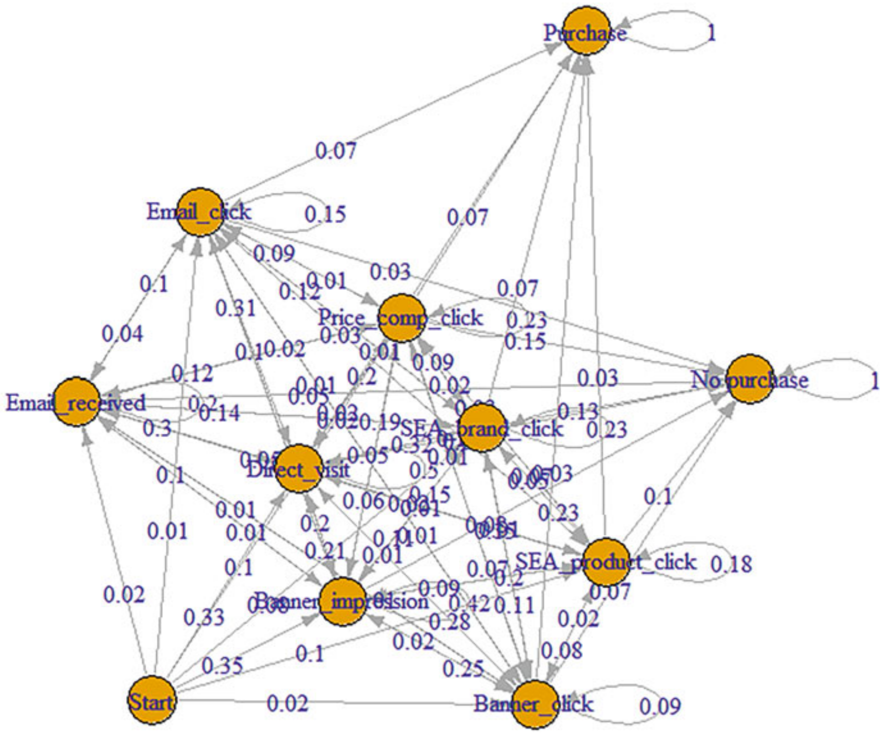

Fig. 18 Markov chain plot path to purchase

We can estimate what would happen if banner advertising is off by setting off this channel in the transition matrix. We can do this with the following R-script.

```
# Estimate transition matrix with banner impressions and clicks
dropped out
markov_chain_banner_off <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_firm)
markov_chain_banner_off <- markov_chain_banner_off[c(1,4:9),c
(1,4:11)]
markov_chain_banner_off <- prop.table(markov_chain_banner_off,
margin=1)
markov_chain_banner_off <- as.data.frame(t
(markov_chain_banner_off))
markov_chain_banner_off <- markov_chain_banner_off[,3]
markov_chain_banner_off[64:81] <- c(rep(0,7),1,0,rep(0,8),1)
mcjourney_banner_off <- new("markovchain", states = c("Start",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                     "Email_received",
"Email_click", "Direct_visit",
                                     "Purchase", "No purchase"),
                       transitionMatrix = matrix
(data=markov_chain_banner_off,
```

	Start	Banner_imp.	Banner_click	SEA_prod._click	SEA_brand_click	Price_comp_click	Email_received	Email_click	Direct_visit	Purchase	No purchase
Start	0.000	0.347	0.015	0.102	0.083	0.099	0.018	0.008	0.328	0.000	0.000
Banner_imp.	0.000	0.415	0.022	0.090	0.110	0.055	0.015	0.007	0.207	0.000	0.080
Banner_click	0.000	0.254	0.091	0.077	0.108	0.067	0.012	0.007	0.285	0.028	0.071
SEA_prod._click	0.000	0.200	0.019	0.182	0.233	0.067	0.012	0.006	0.148	0.032	0.100
SEA_brand_click	0.000	0.148	0.010	0.050	0.229	0.019	0.018	0.008	0.319	0.067	0.131
Price_comp_click	0.000	0.189	0.011	0.032	0.094	0.230	0.018	0.007	0.198	0.072	0.149
Email_received	0.000	0.096	0.009	0.047	0.140	0.122	0.199	0.042	0.300	0.000	0.047
Email_click	0.000	0.097	0.013	0.029	0.123	0.086	0.102	0.148	0.306	0.072	0.025
Direct_visit	0.000	0.200	0.010	0.021	0.050	0.021	0.000	0.000	0.500	0.068	0.130
Purchase	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
No purchase	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Fig. 19 Transition matrix path to purchase

```
                                              byrow= TRUE, nrow = 9))
show(mcjourney_banner_off)
plot(mcjourney_banner_off, edge.arrow.size=0.25)
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney_banner_off ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney_banner_off ^ 50)
after50touchpoints
```

When running this R-script, we can observe that when the banner channel is off, the conversion rate after 50 stages is 33.39%, i.e., a substantial increase compared to the 27.76% we found when banner advertising is on. We do, however, need to remember that 34.72% of the journeys started with a banner impression and 1.51% start with a banner click (see Fig. 19), and these journeys would now not have occurred, leaving us with 63.77% of the journeys. If we take 63.77% of 33.39%, we get to a conversion rate of 21.29%, which is substantially lower than the 27.76% we have found earlier.

We can validate if this 21.29% conversion rate is close to the truth by investigating the transition matrix of the control group, i.e., the group where the firm's banner advertising was off in reality. We do this with the following R-script.

```
# Estimate transition matrix for the charity banner group
markov_chain <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_charity)
markov_chain <- prop.table(markov_chain, margin=1)
markov_chain <- as.data.frame(t(markov_chain))
markov_chain <- markov_chain[,3]
markov_chain[100:121] <- c(rep(0,9),1,0,rep(0,10),1)
mcjourney <- new("markovchain", states = c
("Start","Banner_impression","Banner_click",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                      "Email_received", "Email_click",
"Direct_visit",
                                      "Purchase", "No purchase"),
                         transitionMatrix = matrix(data=markov_chain,
                                      byrow= TRUE, nrow = 11))
show(mcjourney)
plot(mcjourney, edge.arrow.size=0.25)
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney ^ 50)
after50touchpoints
```

Running this R-script, we find that the conversion after 50 stages is 16.88%, i.e., still somewhat lower than the 21.29% that we found in our previous analysis, but it is perfectly in line with what we observed in Fig. 5. Turning off one channel in the Markov chain does thus seem to provide us a somewhat accurate estimate of the conversion, but it is far from perfect. This is because turning off one channel will also affect the transition matrix since some channels might be good alternatives. Combining the Markov chain with experimental data, as we have done here, does provide good insights. An advantage of this Markov chain, e.g., by comparing the Markov chain of the firm's banner group and the charity's banner group, is that we can explore how turning off one channel also impacts the usage of other channels in the path to purchase. For more details, also have a look at Spedicato et al. (2016).

Further Methods of Individual-Level Attribution

In this section, we discussed some methods for individual-level attribution. The provided R-scripts can be adopted and adjusted to be suitable for other, e.g., real-life datasets. The discussed models can also be used to investigate other outcome variables. Retention, customer acquisition, and other forms of customer behavior are examples of variables that can be valuable to investigate (e.g., Gupta et al. 2004; Gupta and Zeithaml 2006).

Next to other outcome variables, also other techniques can be used to analyze the data. Alternatives for the logistic regression model are causal trees and causal forests, which are machine learning techniques that try to estimate the treatment effect (i.e., the difference between the treatment and control group) and explain in which situations this treatment effect is larger or smaller. Such techniques can be convenient to determine which consumers should get a treatment (i.e., be targeted). For details on such techniques, see Hitsch and Misra (2018).

More advanced attribution techniques are available to investigate the impact of the order in which touchpoints occur. This can be done by including carryover and spillover effects in the models. Furthermore, the time between touchpoints can be important, which can be captured by including parameters that capture decay and restoration effects. Braun and Moe (2013) and Li and Kannan (2014) provide information on models that incorporate such effects.

Attribution with Aggregate-Level (Quasi-)Experimental Data

In some cases, it is impossible to do attribution at the individual customer level, for instance, when one wants to find out the impact of mass media advertising on conversion, but it is unknown which individual consumers have come in contact with this form of advertising. Furthermore, with stricter privacy laws, consumers blocking and deleting cookies, and consumers switching between different browsers and devices, tracking consumers across their path to purchase can become more complicated and sometimes even impossible, not allowed, or undesired. In such cases, attribution using experimental data is still achievable at a more aggregate level, e.g., instead of at the individual consumer level, one can conduct aggregate-level experiments or look at variations in the data across regions or over time.

Similarly, as with attribution at the individual level, randomized field experiments can be helpful for attribution at a more aggregate level. A challenge, in this case, is that it is not possible to randomly allocate consumers in different (treatment and control) groups. Instead, the randomization can, for instance, take place at the product, category, or regional level, over time, or at a combination of these levels. Two examples of studies conducting such experiments are Blake et al. (2015) and Wiesel et al. (2011). This section discusses three examples of analysis coming from data from such aggregate-level experiments. The first two forms of analysis, namely the *before-after analysis* and the *before-during-after analysis*, focus on variations over time. The third form of analysis focuses on variations over time and regions and is called the *difference-in-differences analysis*. In the end, we will also discuss some further methods when conducting aggregate-level attribution.

Before-After Analysis

To test the impact of a channel on the traffic to a website or a store, or any other outcome variable, and an experiment at the individual level is not possible, not desirable, or just impractical, an alternative solution is to turn off the channel for some time to see what happens to the desired outcome variable. If turning off a channel is too risky, an alternative is to decrease the expenditures for this channel temporally or, if the expectation is that the channel has a positive impact on the outcome variable, the expenditure can temporally be increased.

To give a simple example of this procedure, let us generate one dataset in line with the data used by Blake et al. (2015). In this example, we look at the search engine channel, and a firm decides to stop using SEA after 20 weekly observations. As one might assume, when stopping SEA, the traffic to the website will decrease. However, it might also be possible that some people who go to the website by clicking on a SEA link would still visit the website when the firm does not conduct SEA; instead of visiting the website through SEA, the consumer might in such case use an organic (nonpaid) link on the search engine, or they might have visited the website directly. Indeed, Blake et al. (2015) have shown that for eBay conducting SEA for branded keywords (i.e., keywords or search phrases containing “eBay”) is not profitable; when eBay stops bidding for these keywords, they still show up high on the organic search results, and users still visit eBay’s website. In such a case, organic search is a perfect alternative for paid search, and paid search should not get credit for the visits and the resulting sale. If after turning off SEA the traffic decreases, i.e., it is not (entirely) substituted by other channels, SEA should get credit for the lost traffic.

To investigate this, let us generate a dataset of 40 weekly observations. The firm turns off SEA after week 20, and we have data on the traffic coming to a website via SEA and organic search and the total traffic coming in via the search engine.

```
# Load required packages -----------------------------------------
library(ggplot2)
library(reshape2)
#Turn off scientific notation
options(scipen = 999)
# Generate dataset before-after analysis -------------------------
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Paid_volume <- ifelse(search_data$SEA_on==1,
                                  25000+sample(0:5000, n),
                               0)
search_data$Organic_volume <- ifelse(search_data$SEA_on==1,
                                  search_data$Paid_volume*1.5 +
10000                             + sample(0:5000, n),
                                  67500 + sample(5000))
# Create total traffic variable
search_data$Total_volume <- search_data$Paid_volume + search_data
$Organic_volume
# Create plot of total, organic and paid traffic
dd = search_data[,c(1,3:5)]
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","green","blue"))
rm(dd)
rm(n)
```

After running this R-script, we have a dataset with 40 weekly observations. Figure 20 shows the plotted data. As can be seen, in the first 20 weeks, both the organic traffic and the paid traffic are relatively stable, but after week 20, the paid traffic drops to zero, which makes sense since SEA is off for this period. We can also see that after week 20, there is a substantial increase in organic traffic to the website. The total traffic does go down somewhat, but not as much as the lost SEA traffic, i.e., it seems that organic search is partly capturing the lost traffic of SEA. Assuming that everything else is stable over time, the explanation for the increase in organic search is that some visitors who would otherwise have clicked on a SEA link now click on the organic search link. This example nicely shows that last-click attribution would have overestimated the impact of SEA since there is an alternative channel available which browsers would use to go to the website if SEA is not available.

To see if SEA has a significant impact on the total traffic to the website, we can run the following R-script to estimate some regression models.

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
```

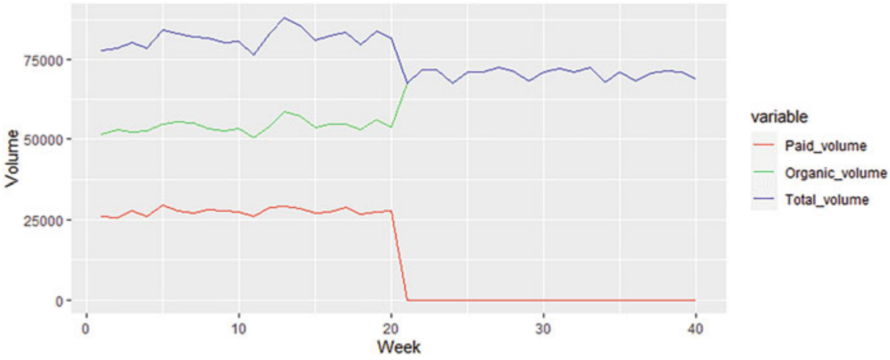

Fig. 20 Data of before-after analysis visualized

```
#regression models before-after analysis
before_after_model1 <- lm(Total_volume ~ SEA_on, data=search_data)
before_after_model2 <- lm(Total_volume ~ SEA_on + Week,
data=search_data)
before_after_model3 <- lm(Organic_volume ~ SEA_on,
data=search_data)
before_after_model4 <- lm(Organic_volume ~ SEA_on + Week,
data=search_data)
tab_model(before_after_model1, before_after_model2,
before_after_model3,
        before_after_model4, collapse.ci = TRUE, p.style =
"stars")
```

As we can see from the output in Table 6, Model 12 shows that when SEA is on, the traffic coming in through the search engine is 11,178.78 visitors higher on average (95% CI [9703.67, 12653.88]), which is highly significant. Controlling for the trend variable "week," which could capture a linear upward or downward trend in the number of visitors over the 40 weeks, does not substantially change the estimate and also does not improve the model, as is shown by Model 13. Model 14 shows that when SEA is on, organic search traffic to the website is a significant 16,432.58 units lower, or the other way round; when turning off SEA, organic search traffic goes up 16,432.58 on average, which is the substation effect of this channel for SEA. Controlling for the week again does not bring much change, as Model 15 shows.

We can use the parameter estimates also as a basis to find the return on investment (ROI) of SEA. Let us assume a 10% conversion rate, a gross profit (before marketing costs) of \$20 per conversion, and SEA costs \$0.50 per click on average. What we furthermore need is the total traffic from SEA when SEA is on. For this, we can take the mean amount of SEA clicks when SEA is on, which we can get with the following R-script.

```
# Mean amount of Paid clicks (i.e. variable 3 in our dataset) when
SEA is on mean(search_data[search_data$SEA_on==1,3])
```

Table 6 Regression output of before-after analyses

	Model 12	Model 13	Model 14	Model 15
	Total_volume	Total_volume	Organic_volume	Organic_volume
Predictors	*Estimates*	*Estimates*	*Estimates*	*Estimates*
(Intercept)	70444.05 *** (69400.99–71487.11)	67755.39 *** (63765.61–71745.17)	70444.05 *** (69632.73–71255.37)	68747.27 *** (65615.41–71879.13)
SEA_on	11178.78 *** (9703.67–12653.88)	12941.83 *** (10024.28–15859.39)	−16432.58 ***(−17579.95 – −15285.20)	−15319.93 ***(−17610.13 – −13029.74)
Week		88.15(−38.22–214.53)		55.63(−43.57–154.83)
Observations	40	40	40	40
R^2/R^2 adjusted	0.861/0.857	0.868/0.861	0.957/0.956	0.958/0.956

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

We find that there are 27,611.35 paid clicks on average when SEA is on. We can now fill in the following ROI formula, based on the parameter from Model 12 from Table 6 (The assumption here is that the 11,178.78 additional visitors is indeed caused by SEA, i.e., without SEA we would lose these visitors, in line with what the model shows.):

$$\begin{aligned}&\textbf{\textit{Incremental gross profit}}\ (\textbf{\textit{if SEA is on}}) = \mathbf{11,178.78}\ \textbf{\textit{additional visitors}} \cdot \\ &\mathbf{10\%}\textbf{\textit{conversion}} \cdot \mathbf{20/}\textbf{\textit{conversion}} = \mathbf{22,357.56}\end{aligned} \tag{2}$$

$$\textbf{\textit{Incremental costs}}\ (\textbf{\textit{if SEA is on}}) = \mathbf{27,611.35}\ \textbf{\textit{clicks}} \cdot \mathbf{0.50/}\textbf{\textit{click}} = \mathbf{13,805.68} \tag{3}$$

$$\textbf{\textit{ROI}} = \frac{\textbf{\textit{Gross profit}} - \textbf{\textit{Costs}}}{\textbf{\textit{Costs}}} \cdot \mathbf{100\%} = \mathbf{22,357.56} - \frac{\mathbf{13,805.68}}{\mathbf{13,805.68} \cdot \mathbf{100\%} = \mathbf{61.94\%}} \tag{4}$$

So, the ROI of SEA is, in this case, 61.94%. We can also replace the 11,178.78 in the formula with the upper and lower score of the confidence interval (i.e., the 95% CI [9703.67, 12653.88]). If we do so, we find the 95% confidence interval of the ROI to be 40.58% and 83.31%, which indicates SEA is, on average, an excellent investment in this case. Note that this is a generated dataset. In reality, the effectiveness and ROI of SEA can be much different. Blake et al. (2015) have for instance found a very low effectiveness of SEA in their study, which might be explained by the fact that they used data from a very well-known firm, namely eBay.

One can easily use this before-after analysis for other questions when investigating the impact of SEA:

- For search engine advertising, instead of looking at the visitors via SEA and organic search, one can also directly investigate the total visitors of the website or the total purchases per week. The advantage of this is that turning off SEA might also impact other channels (e.g., direct website visit), and conversion rates might differ per channel (e.g., the 10% conversion we used in the ROI calculation might not hold for every channel), so looking at sales as a dependent variable can be more appropriate to calculate the ROI.
- Instead of looking at the overall impact of SEA, one can also use this procedure for different types of keywords; e.g., what happens when turning off SEA for branded keywords? Or what happens when turning off SEA for keywords for a specific product category?
- Furthermore, this procedure is helpful to investigate the impact of bids; e.g., what happens when lowing the bids for a period by x%? Or what happens if the budget is temporarily increased or decreased with y%? These kinds of experiments can help to find more optimal bids and determine the budget allocation.

By conducting such experiments, one can thus find out if SEA is, on average, a profitable channel, but it is also possible to improve the bids per keyword and the overall budget allocation. Next to SEA, before-after analyses are also helpful for other channels, including offline mass media like TV and radio advertising.

A limitation of a before-after analysis is that we assume that all other factors which influence traffic before and after the change (e.g., turning off the channel), which we do not control for, are stable over time. Let us assume that the company conducting the experiment is selling sunglasses, and the experiment starts in October, and we do observe that the sales in the "after" period are lower than the "before" period. Can we then conclude that this decrease in sales is due to turning off SEA? Or could this decrease be (partly) due to seasonality, e.g., the "after" period was less sunny, and the demand for sunglasses was lower than the "before" period? Alternatively, if a company is growing over time, the after period is affected by the experiment and the company's growth, which might hide part of the effect estimated with the model. Adding control variables in the regression model can (partly) overcome these confounding effects; e.g., one can control for temperature and the average hours of sunshine per day and a trend by adding a trend variable to the model.

In practice, we cannot control for everything that might change over time and impact our dependent variable of interest. With that, we cannot guarantee causality since we cannot exclude all potential third variables. However, there are two ways we can improve upon the before-after analysis. The following two sections discuss these alternative analyses.

Before-During-aAfter Analysis

An extension of the before-after analysis is the before-during-after analysis in which the turned-off channel (or changed in any other way) is turned back on after the experimental period. An advantage with this is that we can now inspect if the outcome variable (e.g., visitors or sales) goes back to the level as it was before; this is what we would expect when the found impact is indeed causal and when there is no long-term (positive or negative) impact of turning off or on a channel (e.g., via carryover effects or other dynamic effects). This approach is also used by Blake et al. (2015) when investigating the impact of SEA at eBay.

To conduct the before-during-after analysis, let us first generate some data to make this clearer. We can do this with the following R-script.

```
# Generate dataset before-during-after analysis attribution ------
n=60 #has to be a multiple of 3 for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n),
                  c(rep(1, times=n/3),rep(0, times=n/3),rep(1,
times=n/3)),
                  c(rep(0, times=n/3),rep(1, times=n/3),rep(0,
times=n/3)),
                  c(rep(0, times=n/3),rep(0, times=n/3),rep(1,
times=n/3)))
```

```
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
names(search_data)[3] <- paste("During")
names(search_data)[4] <- paste("After")
search_data$Paid_volume <- ifelse(search_data$SEA_on==1,
                              25000+sample(0:5000, n),
                              0)
search_data$Organic_volume <- ifelse(search_data$SEA_on==1,
                          search_data$Paid_volume*1.5 + 10000 +
                          sample(0:5000, n),
                          67500 + sample(5000))
# Create total traffic variable
search_data$Total_volume <- search_data$Paid_volume + search_data
$Organic_volume
# Create plot of total, organic and paid traffic
##Subset the necessary columns
dd = search_data[,c(1,5:7)]
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","green","blue"))
rm(dd)
rm(n)
```

After running this R-script, we have a dataset similar to the before-after study in the previous section, but with 20 additional weekly observations, i.e., the period in which SEA is back on. Figure 21 plots these data. For the first 40 weekly observations, we can see a similar pattern as with the before-after study visualized in Fig. 20, and in the last 20 weeks, when SEA is back on, we can see that everything goes back to a similar situation as in the first 20 weeks. This last period is interesting since if there were other changes over time (e.g., seasonality, an overall change in the website traffic), we would expect that the after period is different from the before period. If the before and after periods are similar, and the period during the experiment is different, we have some more certainty that the change is due to turning off SEA and not (also) due to other factors.

To see if turning off SEA has a significant impact, and if this impact is gone when turning SEA back on, we can run the following R-script to estimate the necessary regression models.

```
#regression models before-during-after analysis
before_during_after_model1 <- lm(Total_volume ~ During + After,
data=search_data)
before_during_after_model2 <- lm(Total_volume ~ During + After +
Week, data=search_data)
tab_model(before_during_after_model1, before_during_after_model2,
          collapse.ci = TRUE, p.style = "stars")
```

As we can see from the output in Table 7, in the "during" period (i.e., when SEA is off), the total traffic coming in through the search engine is significantly decreasing compared to the "before" period. This finding is in line with the before-after analysis. The "after" period parameter is insignificant; since the "before" period is our reference case, this means that there is no significant difference in the number of website visits before the experiment took place and after the experiment has finished. When SEA is back on, the situation does thus goes back to normal. If the "after" period would significantly deviate from the "before" period in the before-during-after analysis, it would signal that there is also something else changing over time, or there might be dynamic effects of turning off and on SEA which the model does not fully capture. As a result, in the case of a significant "after" period, we should be cautious when interpreting the "during" period. If we find the reason for a significant "after" period, e.g., seasonality, we can control for these factors in the model and investigate if controlling for this indeed leads to a nonsignificant "after" period.

An even more robust approach to exclude other factors is the difference-in-differences analysis, discussed in the next section.

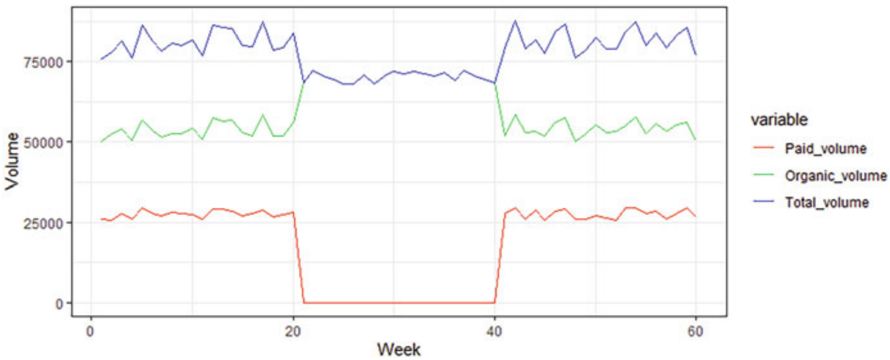

Fig. 21 Data of before-during-after analysis visualized

Table 7 Regression output for before-during-after analyses

	Model 16	Model 17
	Total_volume	Total_volume
Predictors	*Estimates*	*Estimates*
(Intercept)	81233.12 *** (79856.18–82610.07)	80331.76 *** (78343.03–82320.50)
During	−10961.42 ***(−12908.72 – −9014.13)	−12678.30 ***(−16038.48 – −9318.12)
After	471.80(−1475.50–2419.10)	−2961.95(−8783.57–2859.66)
Week		85.84(−51.39–223.08)
Observations	60	60
R^2/R^2 adjusted	0.756/0.748	0.763/0.750

* *p < 0.05* ** *p < 0.01* *** *p < 0.001*

Difference-in-Differences Analysis

With the before-after analysis, there could be confounding factors that change over time and influence the results. The before-during-after analysis is somewhat more robust since one can compare the before and after periods to see if everything goes back to normal when the experimental period is over. There might still have been some other factors influencing the results, e.g., a holiday during the experimental period or activities by a competitor. To be even more confident that the effects are indeed causal, one can make the change (or treatment) only in some regions (the treatment regions) and not make this change in other (similar) regions (the control regions). If the regions are randomly assigned, and there is no difference observed between the regions before the experiment, then the deviations during the experiment show us the impact of the experiment (e.g., turning off a channel in the treatment regions). Multiple studies use this approach of conducting changes in only some regions; examples include Blake et al. (2015) and Wiesel et al. (2011).

To give an example of a difference-in-differences analysis, let us generate some data with the following R-script.

```
# Generate dataset dif-in-dif analysis attribution --------------
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Total_traffic_region1 <- ifelse(search_data$SEA_on==1,
                                  75000+sample(0:3000, n),
                                  70000+sample(0:3000, n))
search_data$Total_traffic_region2 <- 75000+sample(0:3000, n)
# Create plot of traffic per region
dd = search_data[,c(1,3:4)]
library(reshape2)
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","blue"))
# Create some additional variables
dd$treatment_region <- c(rep(1, times=n),rep(0, times=n))
dd$SEA_off <- c(rep(0, times=n/2),rep(1, times=n/2),rep(0,
times=n/2),rep(1, times=n/2))
rm(n)
```

As shown in Fig. 22, the traffic to the website is for the control region (region 2) relatively stable over the 40 weeks. For the treatment region (region 1), the traffic volume is similar to the control region in the first 20 weeks (i.e., when in both regions SEA was on), but the traffic volume starts to deviate when SEA is off in the treatment region. Assuming that the regions are indeed similar, and thus also that

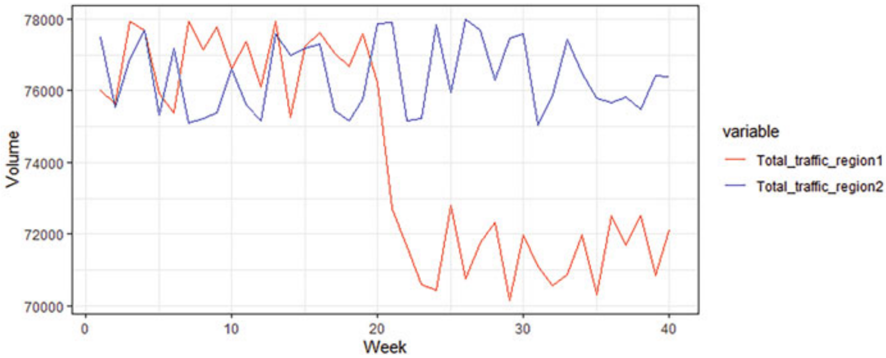

Fig. 22 Data of difference-in-differences analysis visualized

seasonality and trends similarly affect both regions, we can assume that the deviation between the two regions is indeed due to the treatment (i.e., turning off SEA).

To test the significance, we can estimate a difference-in-differences model. This model looks as follows:

$$\boldsymbol{Y_{it} = \beta_0 + \beta_1 \cdot Treatment_i + \beta_2 \cdot After_t + \beta_3 \cdot Treatment_i \cdot After_t + e_{it}} \quad (5)$$

Where $Treatment_i$ indicates if the observation is from the treatment group (1) or control group (0), $After_t$ indicates if the period was after (1) or before (0) the treatment has taken place. The parameter β_1 indicates if the treatment group differs from the control group in the period before the experiment took place; if the treatment region is similar to the control region, this parameter should be insignificant. The parameter β_2 indicates if the control group differs in the after period compared to the before period; if there are no changes over time, this should be insignificant. If there are changes over time, which are not due to the experiment, these changes are captured by β_2. The main parameter of interest is β_3, which captures the deviation of the treatment group from the control group after the treatment; if the two groups are indeed the same over time, with the only difference being the treatment, then β_3 captures the causal effect of the treatment.

To estimate the difference-in-differences model, the following R-script can be used.

```
# Regression model difference-in-differences analysis
dif_dif_model1 <- lm(Volume ~ treatment_region + SEA_off +
treatment_region*SEA_off, data=dd)
dif_dif_model2 <- lm(Volume ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week, data=dd)
tab_model(dif_dif_model1, dif_dif_model2,
          collapse.ci = TRUE, p.style = "stars")
```

As can be seen in Table 8, the parameter for "treatment region" is insignificant in both Models 18 and 19. In line with the explanation of the parameter β_1 of Eq. (5),

Table 8 Regression output for difference-in-differences analyses

	Model 18	Model 19
	Volume	Volume
Predictors	*Estimates*	*Estimates*
(Intercept)	76327.15 *** (75901.70–76752.60)	76381.41 *** (75802.38–76960.43)
treatment_region	532.20(−69.48–1133.88)	532.20(−73.30–1137.70)
SEA_off	150.55(−451.13–752.23)	253.90(−704.20–1211.99)
treatment_region * SEA_off	−5525.40 ***(−6376.30 – −4674.50)	−5525.40 ***(−6381.70 – −4669.10)
Week		−5.17(−42.29–31.96)
Observations	80	80
R^2/R^2 adjusted	0.849/0.843	0.849/0.841

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

this insignificance indicates that the treatment group does not differ significantly from the control group in the period before the experiment took place. If the regions are assigned randomly to the control and treatment condition, an insignificant parameter is thus indeed what we would expect. If the parameter is significant, it signals that there is already a difference between the two groups before the experiment took place. The parameter of "SEA_off," i.e., our treatment, is also insignificant. Given the interaction in the model, this parameter indicates that the traffic volume is not significantly different for the control group in the "after" period. The interaction is highly significant; this parameter is the effect of the treatment (i.e., turning off SEA) in the treatment region and tells us that turning off SEA does significantly decrease the traffic to the website. We can conclude that turning off SEA results in a visitor drop of 5,525.40 (i.e., the difference-in-differences effect).

We can also investigate the difference-in-differences by simply using a cross-table, which can sometimes be easier to understand and communicate the results than a somewhat more complicated regression model.

```
# Create crosstable of difference-in-differences
with(dd, tapply(Volume,
list(treatment_region=treatment_region,SEA_off=SEA_off), mean) )
```

Table 9 shows the cross-table with the difference-in-differences effect, i.e., taking the difference between the before and after periods for the control group and the treatment group and then taking the difference between these two differences (hence "difference-in-differences"), is the same as the parameter estimate from Table 8. An advantage of using a regression model instead of a cross-table is that we can observe if the difference is statistically significant, and we can include control variables in the regression model.

For a difference-in-differences analysis, it is essential that the treatment and the control groups are comparable. In their study, Blake et al. (2015) have assured this

Table 9 Difference-in-differences analyses in a cross-table

	After (treatment period)	Before (control period)	Difference
Treatment group	71,484.50	76,859.35	−5,374.85
Control group	76,477.70	76,327.15	−150.55
Difference	−4,993.20	532.20	**−5,525.40**

by matching regions to each other, which are similar in their historical sales patterns over time. Wiesel et al. (2011) did match regions that were similar in consumer expenditure, the recency, frequency, and monetary value of the purchases, and the number of new and existing customers.

If the control region is different in size compared to the treatment region, but (apart from the treatment) there are no differences over time, this is still not a problem; the size differences will be captured by β_1 of Eq. (5), which would then be statistically significant. Alternatively, weights can make the regions in the before period comparable.

If there are differences in size between the groups, but these differences are not absolute (e.g., region one has on average 1,000 more visitors, which stay similar over time) but relative (e.g., region one has on average 10% more visitors, which stays similar over time) in size, it is better to log transform the dependent variable of interest. By log transforming the dependent variable, the model investigates the relative differences.

To give another example of a difference-in-differences analysis, now with some of the challenges discussed above let us create a new dataset. In this new dataset, the control region is ~75% larger than the treatment region, we have a holiday period in weeks 25 and 26, which causes 10% additional sales in both the treatment and control regions, and the traffic is increasing over time.

```
# Generate second dataset dif-in-dif analysis attribution
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Holidays <- ifelse(search_data$Week>24 & search_data
$Week<27 , 1, 0)
search_data$Total_traffic_region1 <- (ifelse(search_data$SEA_on==1,
                                    75000 + search_data$Week*300 +
sample(0:3000, n),
                                    70000 + search_data$Week*300) +
sample(0:3000, n))*ifelse(search_data$Holidays==1, 1.1, 1)
search_data$Total_traffic_region2 <- (75000 + search_data$Week*300
+ sample(0:3000, n))* ifelse(search_data$Holidays==1, 1.1, 1)*1.75
# Create plot of traffic per region
dd = search_data[,c(1,4:5)]
library(reshape2)
dd = melt(dd, id=c("Week"))
```

```
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","blue"))
# Create some additional variables
dd$treatment_region <- c(rep(1, times=n),rep(0, times=n))
dd$SEA_off <- c(rep(0, times=n/2),rep(1, times=n/2),rep(0,
times=n/2),rep(1, times=n/2))
dd$Holidays <- ifelse(dd$Week>24 & dd$Week<27 , 1, 0)
rm(n)
```

As shown in Fig. 23, the control region (i.e., region 2, indicated with the blue line) is indeed larger than the treatment region (i.e., region 1, indicated with the red line). Furthermore, we can observe a drop in traffic to the website in the treatment region directly after week 20. This drop is not visible in the treatment region. We do observe in both groups the holiday peak in weeks 25 and 26. If we would just take the mean value of traffic before and after turning off SEA in the treatment region, the number of visitors before and after the treatment would be very similar because of the upward trend and the holiday peak in weeks 25 and 26. Using a simple before-after analysis would, in this case, not be appropriate, and a difference-in-differences analysis is likely to show better the causal effect of turning off SEA.

To demonstrate this, let us estimate a series of difference-in-differences models with the following R-script.

```
# Regression model difference-in-differences analysis
dif_dif_model3 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off, data=dd)
dif_dif_model4 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week, data=dd)
dif_dif_model5 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week + Holidays, data=dd)
tab_model(dif_dif_model3, dif_dif_model4, dif_dif_model5,
          collapse.ci = TRUE, p.style = "stars")
```

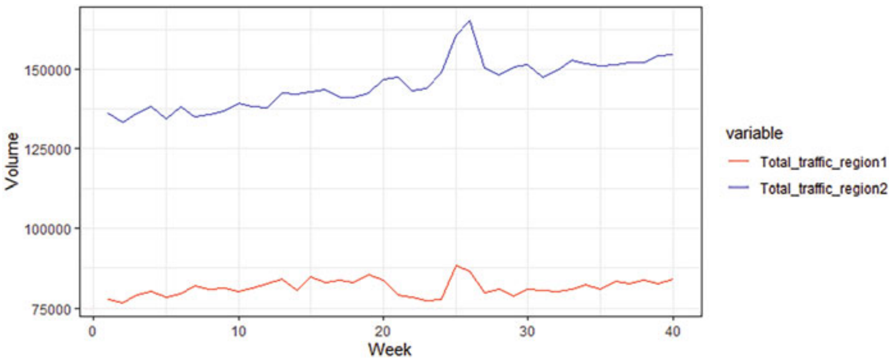

Fig. 23 Data of trending difference-in-differences analysis visualized

Table 10 shows that the parameter for "treatment region" and the difference-in-differences effect (i.e., the interaction parameter) are consistent over the three versions of the model. Since the changes over time (i.e., the trend and the holidays) have the same relative effect size in both regions, the model is robust for these changes. The confidence interval for the difference-in-differences parameter does become smaller if we control for the trend and the holiday weeks because these events result in more noise in the data.

The parameter for "treatment region" is 0.54, which means, due to the log transformation of the dependent variable, the treatment region has about (exp (−0.54)*100% – 1) 41.7% fewer visitors compared to the control region, or in other words, the control region has (exp(0.54)*100% –1) 71.6% more visitors than the treatment regional, comparable to the 75% which we have set it to be. The holiday parameter is significant and has a value of 0.10, meaning that during the holidays, the number of visitors is (exp(0.10)*100% – 1) 10.5% higher, in line with the 10% we have set. Turning SEA off in the treatment region, i.e., the difference-in-differences effect, leads to a (exp(−0.08)*100% – 1) 7.7% decrease in traffic to the website. These figures can be the basis for ROI calculations of the SEA channel. For more information on interpreting the parameters of a regression model when the dependent variable is log transformed, please have a look at Chapter 2 of Leeflang et al. (2015).

In Model 20 of Table 10, the parameter for "SEA_off" is significant. In line with the discussion of eq. (5), a significant β_2 means that there is a difference in the "before" and the "after" period for the control region. This is true since there is an upward trend and a holiday peak in the after period. When including the "week" and "holiday" variables as control variables, as we do in Model 22, this effect disappears, i.e., we have with the full model controlled for the differences over time.

Table 10 Regression output for trending difference-in-differences analyses

	Model 20	Model 21	Model 22
	log(Volume)	log(Volume)	log(Volume)
Predictors	*Estimates*	*Estimates*	*Estimates*
(Intercept)	11.84 *** (11.83–11.86)	11.81 *** (11.80–11.83)	11.80 *** (11.79–11.81)
treatment_region	−0.54 ***(−0.56 – −0.52)	−0.54 ***(−0.55 – −0.52)	−0.54 ***(−0.54 – −0.53)
SEA_off	0.08 *** (0.06–0.10)	0.02(−0.00–0.05)	−0.00(−0.02–0.01)
treatment_region * SEA_off	−0.08 ***(−0.11 – −0.06)	−0.08 ***(−0.11 – −0.06)	−0.08 ***(−0.09 – −0.07)
Week		0.00 *** (0.00–0.00)	0.00 *** (0.00–0.00)
Holidays			0.10 *** (0.09–0.12)
Observations	80	80	80
R^2/R^2 adjusted	0.989/0.989	0.993/0.993	0.998/0.998

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Let us compare the results from Table 10 with what we would find if we would only have information from the treatment region and, based on that, estimate a before-after analysis. We do this with the following R-script.

```
# Regression model before-after analysis on difference-in-
differences data
before_after_model5 <- lm(log(Volume) ~ SEA_off, data=dd,
subset=treatment_region==1)
before_after_model6 <- lm(log(Volume) ~ SEA_off + Week, data=dd,
subset=treatment_region==1)
before_after_model7 <- lm(log(Volume) ~ SEA_off + Week + Holidays,
data=dd, subset=treatment_region==1)
tab_model(before_after_model5, before_after_model6,
before_after_model7,
          collapse.ci = TRUE, p.style = "stars")
```

Indeed, Model 23 in Table 11 shows that when we do a simple before-after analysis, without controlling for the time effects the parameter for turning SEA off is precisely zero and insignificant, i.e., there is no difference between the before and after periods. This is indeed in line with what we see when we look at the total traffic from region 1 in Fig. 23; due to the upward trend, the before and after periods look very similar, although it is clear that the traffic drops immediately when SEA is off, but it recovers due to the (unrelated to SEA) upward trend. When controlling for this upward trend by including a trend variable, as is done in Model 24 in Table 11, we get much closer to the actual effect of SEA as shown by the difference-in-differences analyses in Table 10. We are still somewhat off because the traffic in the after period is higher for two weeks due to the holiday peak, which is unrelated to turning SEA off and thus hides part of the drop in the number of visitors. When also controlling for this holiday peak, as is done in Model 25 in Table 11, we get the same (i.e., accurate) parameter estimate for turning off SEA as in the difference-in-differences models from Table 10. The results from Table 11 do thus show that a before-after analysis is sensitive for other variations over time. When we control for these changes, as we do in Model 25 in Table 11, we get a pretty accurate parameter estimate, but if we do not control for all variables, we might get the wrong estimate (as shown in Model 23 of Table 11). With this, we thus show the superiority of a difference-in-differences analysis over the before-after analysis.

The difference-in-differences analysis still has some drawbacks, namely that it needs an experiment to run this analysis, which is in practice not always possible, or it might be too time consuming or too risky to conduct. Furthermore, with a difference-in-differences analysis and the before(−during)-after analyses, it is hard to capture long-term effects. When, for instance, turning off TV advertising, it might be that this does not directly lead to significantly lower visitors or sales since it might take some time before the effect takes place (i.e., wear-in and wear-out effects). In order to capture this, the model needs to include dynamic effects. Including dynamics is possible with lag terms or stock variables (Hanssens 2021). Other methods

Table 11 Regression output for trending difference-in-differences data using before-after analyses

	Model 23	Model 24	Model 25
	log(Volume)	log(Volume)	log(Volume)
Predictors	*Estimates*	*Estimates*	*Estimates*
(Intercept)	11.31 *** (11.29–11.32)	11.27 *** (11.25–11.29)	11.26 *** (11.25–11.27)
SEA_off	0.00(−0.02–0.02)	-0.07 ***(−0.10 – −0.03)	−0.09 ***(−0.11 – −0.08)
Week		0.00 *** (0.00–0.00)	0.00 *** (0.00–0.00)
Holidays			0.10 *** (0.08–0.12)
Observations	40	40	40
R^2/R^2 adjusted	0.000/−0.026	0.395/0.362	0.851/0.839

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

which capture dynamic effects, and do not rely on experimental data, will be discussed in the next section.

For more information on exploiting data from field experiments and applied time series analysis, see Artz and Doering (2021) and Wang and Yildirim (2021).

Further Methods of Aggregate-Level Attribution

In some cases, it is not possible, or it might be too time consuming or too risky, to conduct an experiment. Because of this, managers might not be willing to set up an experiment or have the time to wait for the results. Firms typically have historical data on their advertising expenditures per channel and data on other marketing mix components like pricing and distribution and performance data like sales. These data are typically available over time, and the changes over time in the marketing variables can thus be related to changes over time in firm performance.

One challenge is that decisions on changes in the marketing mix are not set in isolation; e.g., (past) firm performance might drive current pricing and advertising expenditure. A drop in sales might make managers decide to lower the price to recover market share, while higher sales in the previous period might lead to higher budgets for advertising in the current period, as the advertising budget is often a function of (previous and expected) revenue. A standard regression model does not take these kinds of reversed causality and endogeneity into account.

Another challenge is that the effects can be dynamic, e.g., spending more on advertising in the current period might not only affect current firm performance, but it might also still have an impact in the next period(s). These carryover effects are not taken into account in a standard regression model unless explicitly included via, for instance, (multiple) lag terms or ad stock variables (see Hanssens [2021] for more details on this).

A model which can take reversed causality, (specific forms of) endogeneity, and dynamics into account is the vector autoregressive (VAR) model. One example of an attribution paper that uses a VAR model is Wiesel et al. (2011), who investigate how online and offline advertising affect the online and offline sales channel usage and the profit of a B2B office furniture seller. Another attribution paper that uses a series of VAR models is De Haan et al. (2016). They investigate how various online and offline advertising channels impact website progression and sales revenue at an online retailer for five different product categories. Both Wiesel et al. (2011) and De Haan et al. (2016) demonstrate how the VAR models can be used to improve advertising budget allocation. Other studies related to attribution which have used VAR models include Trusov et al. (2009), Srinivasan et al. (2010), Pauwels et al. (2016b), and Srinivasan et al. (2016), just to name a few. The book chapter from Srinivasan (2021) discusses the VAR model in detail, and for the interested reader on www.evertdehaan.com there is also a R-script available to conduct a simple VAR model for attribution.

There are also more sophisticated methods capable of performing attribution at an aggregate level. One R-package to highlight is "CausalImpact" (Brodersen and Hauser 2021). This package is helpful to explore experiments, i.e., a random event in which there is a sudden change in one marketing channel. The package tries to estimate the value of the outcome variable if the event did not occur, i.e., the so-called counterfactual scenario. This counterfactual scenario is exactly what we want to have when conducting attribution, namely finding out what would have happened if a channel or touchpoint was not active. For more details, see Brodersen and Hauser (2021) and try out the example code, which can be downloaded at www.evertdehaan.com.

The aggregate-level attribution methods can also be used with other dependent variables similar to the individual-level attribution methods. Examples would be the number of offline (i.e., brick-and-mortar) stores visits over time, revenue, profitability, market share, and stock return. These variables indicate another advantage of aggregate-level attribution, namely that one can use dependent variables not measured at the individual customer level or variables that are hard to link to an individual's online behavior or touchpoints. Also, variables related to perceptual outcomes, e.g., customer mindset metrics like customer satisfaction and brand awareness, are interesting as (intermediate) outcome variables. Srinivasan et al. (2010) and De Haan et al. (2021) demonstrate the advantage of using such perceptual outcomes instead of or next to using more transactional (financial and behavioral) outcome variables.

Furthermore, it is possible to combine different techniques, e.g., start with a VAR model, followed by an aggregate-level experiment, as Wiesel et al. (2011) did. Such a setup can help overcome some of the challenges and risks, e.g., the data for a VAR model is easier to collect and is a good way to get first insights, while an experiment provides more certainty of finding causal effects. Also, other techniques to analyze aggregate-level data are possible, including state-space models, structural models, and Bayesian analysis. See the book by Leeflang et al. (2017) for details on these and further methods.

Conclusion

As introduced at the beginning of this chapter, attribution is the process to "allocate appropriate credit for a desired customer action to each marketing touchpoint across all online and off-line channels" (Kannan et al. 2016). With this it is essential to find the incremental effect a specific touchpoint has on the outcome of a path to purchase. Basic attribution methods, based purely on encountering a specific touchpoint (i.e., touch-based attribution), are not suited for this since they do not provide information about what would have happened if the touchpoint was not there. In order to find this out, Shapley values already provide better insights, but the best insights can be retrieved by conducting (individual- or aggregate-level) field experiments.

In situations where individual-level field experiments are not possible, feasible, or desirable, more elaborate attribution methods are available, including PSM, Markov chain, and aggregate-level field experiments. All of these methods have their advantages and disadvantages, as discussed in this chapter. Furthermore, there is a wide range of other (simple and complicated) procedures and models for attribution. However, the basic idea remains the same; finding out what would have happened without a specific touchpoint gives a certain amount of credit to that touchpoint.

Therefore, this chapter introduces attribution modeling, which can help to critically evaluate current attribution methods used within an organization and give directions on how this can be improved. We should also not consider attribution a goal in itself, but it can help decide on budget allocations and decide how, when, and where to target which customer. With this, attribution can help make marketing more accountable and make better advertising and targeting decisions.

Cross-References

► Analysis of Variance
► Applied Time-Series Analysis in Marketing
► Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers
► Experiments in Market Research
► Exploiting Data from Field Experiments
► ExploitingData from Field Experiments
► Field Experiments
► Logistic Regression and Discriminant Analysis
► Modeling Marketing Dynamics Using Vector Autoregressive (VAR) Models
► Panel Data Analysis: A Non-technical Introduction for Marketing Researchers
► Regression Analysis
► Return on Media Models

References

Anderl, E., Becker, I., Von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing, 33*(3), 457–474.

Artz, M., & Doering, H. (2021). Exploiting Data from Field Experiments. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica, 83*(1), 155–174.

Bornemann, T., & Hattula, S. (2018). Experiments in Market Research. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Braun, M., & Moe, W. W. (2013). Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science, 32*(5), 753–767.

Brodersen, K. H., & Hauser, A. (2021). CausalImpact. Available at https://cran.r-project.org/web/packages/CausalImpact/vignettes/CausalImpact.html

De Haan, E., Wiesel, T., & Pauwels, K. (2016). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing, 33*(3), 491–507.

De Haan, E., Kannan, P. K., Verhoef, P. C., & Wiesel, T. (2018). Device switching in online purchasing: Examining the strategic contingencies. *Journal of Marketing, 82*(5), 1–19.

De Haan, E., Verhoef, P. C., & Wiesel, T. (2021). Customer feedback metrics for marketing accountability. *Review of Marketing Research*, forthcoming.

Ebbes, P., Papies, D., & Van Heerde, H. J. (2016). Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Försch, S., & de Haan, E. (2018). Targeting online display ads: Choosing their frequency and spacing. *International Journal of Research in Marketing, 35*(4), 661–672.

Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science, 25*(6), 718–739.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7–18.

Hanssens, D. M. (2021). Return on Media Models. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Hitsch, G. J., & Misra, S. (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. Available at SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111957

Ho, D., Imai, K., King, G., Stuart, E., Whitworth, A., & Greifer, N. (2021). Package 'MatchIt'. Available at https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf

Hoban, P. R., & Bucklin, R. E. (2015). Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research, 52*(3), 375–393.

Jackson, C. (2019). Package 'msm'. Available at https://cran.r-project.org/web/packages/msm/msm.pdf

Johnson, L. (2018). When Procter & Gamble Cut $200 Million in Digital Ad Spend, It Increased Its Reach 10%. Available at https://www.adweek.com/brand-marketing/when-procter-gamble-cut-200-million-in-digital-ad-spend-its-marketing-became-10-more-effective/

Kannan, P. K., Reinartz, W., & Verhoef, P. C. (2016). The path to purchase and attribution modeling: Introduction to special section.

Landwehr, J. R. (2019). Analysis of Variance. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., & Pauwels, K. (2015). *Modeling markets: Analyzing marketing phenomena and improving marketing decision making* (Ser. International series in quantitative marketing). Springer.

Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., & Pauwels, K. (2017). *Advanced methods for modeling markets* (Ser. International series in quantitative marketing). Springer.

Lesscher, L., Lobschat, L., & Verhoef, P. C. (2021). Do offline and online go hand in hand? Cross-channel and synergy effects of direct mailing and display advertising. *International Journal of Research in Marketing*, forthcoming.

Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research, 51*(1), 40–56.

Li, J., Luo, X., Lu, X., & Moriguchi, T. (2021). The double-edged effects of E-commerce cart retargeting: Does retargeting too early backfire? *Journal of Marketing*, forthcoming.

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology, 13*(1–2), 29–39.

Pan, W., & Bai, H. (Eds.). (2015). *Propensity score analysis: Fundamentals and developments*. Guilford Publications.

Pauwels, K., Demirci, C., Yildirim, G., & Srinivasan, S. (2016a). The impact of brand familiarity on online and offline media synergy. *International Journal of Research in Marketing, 33*(4), 739–753.

Pauwels, K., Aksehirli, Z., & Lackman, A. (2016b). Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *International Journal of Research in Marketing, 33*(3), 639–655.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3), 169–188.

Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research, 48*(1), 87–102.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games, 2*(28), 307–317.

Skiera, B., Reiner, J., & Albers, S. (2018). Regression Analysis. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Spedicato, G. A. (2021). Package 'markovchain'. Available at https://cran.r-project.org/web/packages/markovchain/markovchain.pdf

Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., & Cordón, I. (2016). The markovchain package: A package for easily handling Discrete Markov Chains in R. https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf

Srinivasan, S. (2021). Modeling Marketing Dynamics Using Vector Autoregressive (VAR) Models. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research, 47*(4), 672–684.

Srinivasan, S., Rutz, O. J., & Pauwels, K. (2016). Paths to and off purchase: Quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science, 44*(4), 440–453.

Tillmanns, S., & Krafft, M. (2017). Logistic Regression and Discriminant Analysis. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing, 73*(5), 90–102.

WARC. (2021). Uber's former performance chief charts extent of ad fraud. Available at https://www.warc.com/newsandopinion/news/ubers-former-performance-chief-charts-extent-of-ad-fraud/44527

Wiesel, T., Pauwels, K., & Arts, J. (2011). Practice prize paper – Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science, 30*(4), 604–611.

Valli, V., Stahl, F., & Feit, E. M. (2017). Field Experiments. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Wang, W., & Yildirim, G. (2021). Applied Time-Series Analysis in Marketing. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.