BUSA 603

Module 5: Marketing Mix Models

Brian Weikel

brian.weikel@franklin.edu

FRANKLIN UNIVERSITY
Makes It Possible

Business Analytics
Franklin University

Spring 2024

Outline[1]

1. The Concept of Marketing Mix
2. Model Specifications, Diagnostics and Fit
3. Media Planning
4. Patterns of Advertising Response, Wear-In, Wear-Out and Response Curves
5. Synergy Measurement via Moderation in Linear Models

[1] These lecture notes map to chapters 14 and 15 of Davis (2022).

The Concept of Marketing Mix

The Concept of Marketing Mix

The **marketing mix** refers to variables that a marketing manager can control to influence a brand's sales or market share. Traditionally, these variables are a subset of the 4 P's or 7 P's.[2]

The perennial question that managers face is, what level or combination of these variables maximizes sales, market share, or profit?

The answer to this question, in turn, depends on the following question: How do sales or market share respond to past levels of expenditures on these variables? Researchers, industry practitioners and data scientists have focused intently on trying to find answers to this question.

[2] See Module 1 for the definition of each P. As mentioned in Module 4, a particular item's regular price and clearance price are almost always determined by the finance department of a company.

Marketing Mix Modeling

To address the question, they have developed a variety of market response econometric models. Unsurprisingly, they are referred to as **marketing mix models** (MMM).[3]

MMM, the use of statistical analysis to estimate the past impact and predict the future impact of various marketing tactics on sales, can deeply inform marketing plans.

While marketing spend and bottom line results are often perceived as disconnected by marketing professionals, MMM closes the loop and may be used to improve the return on marketing investment (ROMI).

[3] Beyond the textbook of the course, other MMM introductions are Tellis (2006) and Solana and Weikel (2014).

Model Specifications, Diagnostics and Fit

A Linear Model Specification

While from a practitioner perspective there are 5 basic different models used to model the marketing mix, we will focus on two models: the **linear model** and the **multiplicative model**. The linear model takes the form

$$Y_t = \alpha + \sum_{i=1}^{I} \beta_{1i} A_{it} + \sum_{j=1}^{J} \beta_{2j} P_{jt} + \sum_{k=1}^{K} \beta_{3k} R_{kt} + \sum_{l=1}^{L} \beta_{4l} Q_{lt} + \sum_{m=1}^{M} \beta_{5m} X_{mt} + \epsilon_t, \quad (1)$$

where:

1. Y_t is the **dependent variable**, also known as the **outcome variable**, **response variable** or **regressand**. It may be brand gross value sales, item units, or item **equivalized units**.[4]

[4] Equivalized units is a unit of measurement that permits items to be compared across brands and product categories. A few examples include pounds, kilograms, gallons, ounces, liters, doses, uses, and cases.

Independent Variables

2. The other capital Latin letters represent **independent variables** of the marketing mix.

   - I advertising independent variables: $A_{it}, i \in \{1, 2, \ldots, I\}$.
   - J pricing independent variables: $P_{jt}, j \in \{1, 2, \ldots, J\}$.
   - K sales promotion, merchandising, and distribution independent variables: $R_{kt}, k \in \{1, 2, \ldots, K\}$.
   - L assortment quality independent variables: $Q_{lt}, l \in \{1, 2, \ldots, L\}$.
   - M other factors, such as seasonality, competition, and economic conditions beyond the marketers control that influences sales: $X_{mt}, m \in \{1, 2, \ldots, M\}$.

   Independent variables may also be called **explanatory variables** or **regressors**.

Coefficients and Errors

3. The subscript t represents time.[5]

4. The **coefficients**, α, $\beta_{1i}, \beta_{2j}, \beta_{3k}, \beta_{4l}$, and β_{5m} for $i \in \{1, 2, \ldots, I\}, j \in \{1, 2, \ldots, J\}, k \in \{1, 2, \ldots, K\}, l \in \{1, 2, \ldots, L\}$ and $m \in \{1, 2, \ldots, M\}$, are **parameters** that the researcher wants to estimate. They represent the effect of the independent variables on the dependent variable.

5. The ϵ_t are **errors** that are assumed be **independently and identically distributed** (IID). This assumption means that there is no pattern to the errors so that they constitute just random noise, which is referred to as **white noise**.[6]

[5] Implicit in the model specification is that that there is one product, one market and one customer segment, such as brand-national-all customers. In industry practice, a more granular dependent variable and independent variables are modeled. For example, sub-brand or price-promoted group for product. Store, sales channel (e.g., brick-and-mortar or online) or, designated market area (DMA) in lieu of national.

[6] Many times the ϵ_t are also assumed to be normally distributed.

Multiplicative Model Specification

The multiplicative model takes the form

$$Y_t = \exp\{\alpha\} \prod_{i=1}^{I} A_{it}^{\beta_{1i}} \prod_{j=1}^{J} P_{jt}^{\beta_{2j}} \prod_{k=1}^{K} R_{kt}^{\beta_{3k}} + \prod_{l=1}^{L} Q_{lt}^{\beta_{4l}} \prod_{m=1}^{M} X_{mt}^{\beta_{5m}} \exp\{\epsilon_t\}, \quad (2)$$

or after a natural logarithm transformation,

$$\log Y_t = \alpha + \sum_{i=1}^{I} \beta_{1i} \log A_{it} + \sum_{j=1}^{J} \beta_{2j} \log P_{jt} + \sum_{k=1}^{K} \beta_{3k} \log R_{kt} + \sum_{l=1}^{L} \beta_{4l} \log Q_{lt} + \sum_{m=1}^{M} \beta_{5m} \log X_{mt} + \epsilon_t. \quad (3)$$

Multiplicative Model Specification Benefits

After this transformation, the error terms in Equation 3 are assumed to be IID, and usually normally distributed.

The multiplicative model has many benefits.

1. This model implies that the dependent variable is affected by an **interaction** of the variables of the marketing mix. In other words, the independent variables have a **synergistic effect** on the dependent variable. For example, larger advertising combined with a price decrease may enhance sales more than the sum of larger advertising or the price decrease occurring alone.
2. The multiplicative model specification implies that response of sales to any of the independent variables can take on a variety of shapes depending on the value of the coefficient.
3. The coefficients not only estimate the effects of the independent variables on the dependent variables, but they are also **elasticities**.

Multiplicative Model Specification Limitations

The multiplicative model has several limitations.

1. Using the natural logarithm specification implies rigidity in modeling select advertising effects. For example, using the natural logarithm transformation results in inflexibility in modeling the sales-advertising response shape effect.[7]
2. The multiplicative model implies that the elasticity of sales to advertising is constant. That is, the percentage change in sales for a percentage change in advertising is the same indifferent to advertising or sales levels. This result is quite implausible.

The two models are **regression** models, a simple but powerful statistical methodology.

[7] In industry practice the natural logarithm transformation is replaced by a more flexible transformation, such as the Hill function. For more information, see Jin et al. (2017).

Regression coefficients for these these models are frequently estimated by the **principle of least squares**.

An **estimator** is a formula, method, or recipe (e.g., numerical algorithm) for estimating an unknown **population parameter**. An **estimate** is the numerical value obtained when sample data are substituted in the formula.

For simplicity purposes we will consider a bivariate regression specification. Let the **residuals** from a fitted straight line be,

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = \{1, 2, \ldots, n\}. \tag{4}$$

From the definition of $\hat{y}_i$, these residuals are seen to be measured in the vertical (i.e., y) direction. Different values of the intercept ($\hat{\beta}_0$) and slope ($\hat{\beta}_1$) define a different line, and hence a different set of residuals.

Least-Squares Principle

The least squares principle is to identify values for β_0 and β_1 to minimize the **residual sum of squares** (RSS):

$$\text{RSS} = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{5}$$

Upon minimizing RSS with sample data, β_0 is replaced by $\hat{\beta}_0$ and β_1 is replaced by $\hat{\beta}_1$.

There are necessary conditions for a stationary value of RSS. Using calculus, these conditions are:

$$-2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0, \quad \text{and} \tag{6}$$

$$-2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0. \tag{7}$$

Normal Equations

Simplifying Equations (6) and (7) yields the **normal equations**[8]

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i, \quad \text{and} \tag{8}$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = 0. \tag{9}$$

Rewriting Equation (8) and using sample data in the formulas yields $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ and $\hat{\beta}_1 = r\frac{s_y}{s_x}$. That is, the y-intercept and slope, respectively.

[8]Equations (8) and (9) use the word normal due to the geometry of the least-squares principle. For additional information see Johnston and DiNardo (1997).

A Regression Example - Petroleum Distillate Demand

Since the OPEC oil embargo of 1973 there has been a continuing controversy in the U.S. about national and state tax policy for gasoline and other petroleum distillates.

Other major disruptions in the production and/or distribution of petroleum distillates include, but are not limited to:

- Iranian revolution in the late 1970s
- Iran-Iraq war in the early 1980s
- Persian Gulf War in 1990
- Iraqi War in 2003
- Hurricanes Katrina and Rita in 2005
- Hurricanes Gustav and Ike in 2008
- Arab Spring of 2011
- Russia's war with Ukraine starting in 2022

Modeling Energy Demand is An Active Field

A crucial component of any such debate is a reliable model for demand, which by its very nature has a long, and what some may call dubious, history.[9]

As may be expected, energy demand modeling is still an active field.[10] We will review a rather simple petroleum demand function.

It is a multiplicative regression model.

[9] For example see Sims (1980), Malinvaud (1981) and Freedman et al. (1983).

[10] A recent review of hundreds of recent articles on modeling the demand for energy is Verwiebe et al. (2021).

Data Sources

The data was obtained from the Federal Reserve Bank of St. Louis and the Federal Highway Administration of the U.S. Department of Transportation. The observation level is month. Variables sourced from the sites follow.

- Y: per capita personal consumption of gasoline in gallons (at annual rates).
- P: real price of a gallon of gasoline in 2012 prices; that is 1 gallon = \$3.35 at 2012 prices.[11]
- Z: per capita annual personal income in 1000's of 2012 U.S. dollars.
- y: natural logarithm of Y.
- p: natural logarithm of P.
- z: natural logarithm of Z.

[11] For additional information about deflating nominal values to real values see this Federal Reserve Bank of Dallas page.

Gasoline Demand is Highly Seasonal

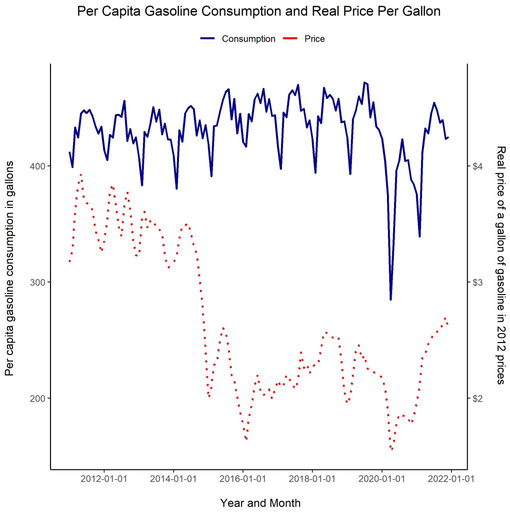

Figure 1: Gasoline Consumption and Real Price per Gallon

COVID-19 Produced Several Structural Breaks

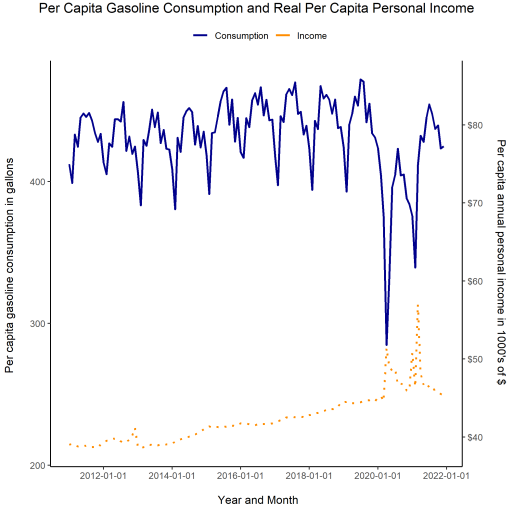

Figure 2: Gasoline Consumption and Real Per Capital Income

Creating a Few Explanatory Variables

For $i \in \{1, 2, \ldots, 12\}$, define the following monthly indicators, which will be used to model monthly demand seasonality.

$$x_{it} = \begin{cases} 1, & \text{if the year of the month of observation } t \text{ is } i, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $x_{1t} = 1$ if the month is January and 0 otherwise for observation t, $x_{2t} = 1$ if the month is February and 0 otherwise for observation t, ...

To control for effects that yield a long-term trend, we will include a trend explanatory variable in our demand specification. Let January 2010 be $t = 1$. The log trend variable is defined as $w_t = \ln(t + 1), t \geq 1$.

Consider the following simple equilibrium model for the demand of gasoline:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 p_t + \gamma_0 w_t + \sum_{i=1}^{11} \gamma_i x_{it} + \epsilon_t. \qquad (10)$$

The errors, $\epsilon_t, t = \{1, 2, \ldots, T\}$, are assumed to be IID.

Since the model is specified as a log-log model in consumption, income and price, β_1 is the long-run income elasticity of demand for gasoline and β_2 is the long-run price elasticity of demand for gasoline for small changes in income and price, respectively.[12]

[12] To illustrate the relationship between the real price of gasoline elasticity and the real price of gasoline coefficient, suppose $\beta_2 = -0.5$. Then a 1% increase in the real price of gasoline is expected to produce a 0.5% decrease in the quantity demanded of gasoline.

Statistical Inference and Model Diagnostics

The standard deviation of the sampling distribution of a coefficient estimate is often referred to as the **standard error** of the estimate.

Recall the smallest significance level at which a null hypothesis can be rejected is called the **probability value**, or **p-value**.[13]

The **F-statistic** is the value of the test statistic for the null hypothesis that all explanatory variable coefficients are zero.

$\mathbf{R^2}$ measures the proportion of the total variation in the dependent variable explained by the linear combination of the explanatory variables. Mathematically, it is 1 less the ratio of the RSS divided by the **total sum of squares** (TSS): $R^2 = 1 - \frac{RSS}{TSS}$.

For n the number of observations and k the number of coefficients, **Adjusted** $\mathbf{R^2} = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$. Thus the statistic takes explicit account of the number of explanatory variables in the regression.

[13] The textbook states, "If the p-value is less than or equal to .05, which is a commonly accepted threshold, then it is acceptable to say the estimate is significantly different than 0, or just significant."

Output from Model Estimation via R `lm` Function

```
Call:
lm(formula = y ~ z + p + trend + M1 + M2 + M3 + M4 + M5
+ M6 +  M7 + M8 + M9 + M10 + M11, data = df2)

Residuals:
      Min        1Q    Median        3Q       Max
-0.026546 -0.008086 -0.001191  0.007757  0.043656

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.514919   0.193360  28.521  < 2e-16 ***
z            0.173673   0.052087   3.334  0.00123 **
p           -0.061654   0.009575  -6.439 5.22e-09 ***
trend       -0.008236   0.002698  -3.053  0.00296 **
M1          -0.043412   0.006104  -7.112 2.32e-10 ***
M2          -0.092616   0.006072 -15.253  < 2e-16 ***
```

Output from Model Estimation Continued

```
              Estimate Std. Error t value Pr(>|t|)
M3            0.009892   0.006061   1.632  0.10605
M4            0.003857   0.006091   0.633  0.52811
M5            0.047504   0.006128   7.752 1.11e-11 ***
M6            0.053482   0.006121   8.738 9.54e-14 ***
M7            0.052569   0.006101   8.617 1.72e-13 ***
M8            0.063026   0.006099  10.333  < 2e-16 ***
M9            0.015509   0.006105   2.540  0.01274 *
M10           0.032990   0.006076   5.430 4.48e-07 ***
M11          -0.007506   0.006049  -1.241  0.21780
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1

Residual standard error: 0.01281 on 93 degrees of freedom
Multiple R-squared:  0.9306,Adjusted R-squared:  0.9201
F-statistic: 89.04 on 14 and 93 DF,  p-value: < 2.2e-16
```

Exponential Attraction and Multinomial Logit Models

Other marketing model specifications include **attraction models** and the **multinomial logit model** (MNL).

- Attraction models are based on the premise that market response is the result of the attractive power of a brand relative to that of other brands with which it competes.
  - The **linear attraction model** specifies a brand's share of market sales as a linear function of its share of total marketing effort.
  - The **exponential attraction model** expresses the market share of the brand as an exponential attraction of the marketing mix.
- The MNL results from a log-centering transformation of the exponential attraction model.[14]

[14] For additional information about attraction models and MNL, see Cooper and Nakanishi (1988).

Other Model Specifications

1. The **Koyck model** may be considered a simple augmentation of the basic linear model, which includes the lagged dependent variable as an independent variable.
2. The **distributed lag model** is a model with multiple lagged values of both the dependent variable and the independent variables.[15]
3. **Hierarchical models** are multistage models in which coefficients (of advertising) estimated in one stage become the dependent variable in the other stage. The second stage contains the characteristics by which advertising is likely to vary in the first stage, such as ad content, medium, or campaign duration.[16]

[15] For additional information on the Koyck and distributed lag models see Johnston and DiNardo (1997).

[16] For additional information on hierarchical models, see Tellis et al. (2000).

Media Planning

Basic Components of Media Plan

Media planning involves essentially three basic activities.[17]

1. **Defining the marketing problem**: Does the firm know where the business is coming from and where the potential for increased business lies? Does it know the markets of greatest importance and greatest opportunity? Does it know who is most likely to buy? Does the firm need to stimulate trial or defend a purchase? Does the firm need to reach everyone or only a selective group of consumers? How much product loyalty exists?
2. **Translating marketing requirements into actionable media objects**: If the marketing objective is to stimulate trial among all potential consumers, then reaching many people is more important than reaching fewer people more frequently. If the product is purchased often, then reaching people more frequently might be a more appropriate tactic.

[17] A reference that has been used for years by marketers to understand the nuances of media planning is Sissors and Baron (2010).

Formulating Media Strategies

3. **Defining a media solution by formulating media strategies**: If reaching people is a primary objective, one should select affordable media that will generate more reach than other media forms. If a specific demographic group is to be reached, then media selection should be based on reaching the group effectively and efficiently.

The objective of any media plan results in defining **media goals**. Goals must be positive, action-oriented statements representing an extension of the **marketing objectives** and hence, must also be marketing-goal oriented. A few example goals follow.

1. Reach at least 80% of the potential market within the first month of advertising, ensuring that the average consumer will be exposed to a minimum of 3 advertising messages.
2. Direct advertising to current and potential purchasers of Product Line Z by weighting current purchasers characteristics by 60% and potential characteristics by 40%.

Media Strategies

Media strategies are the solutions to the media objectives. Strategy statements reflect the specific course to be taken with media.

1. Which media will be used?
2. How often will each be used?
3. How much of each media will be used?
4. During which periods of the year?

Media Strategy Scheduling Objective

Every media plan should have a scheduling objective to guide the planner in allocating media across the year.

Beyond the general timing considerations, the media planner should also consider the strategy of flighting versus continuity.

- **Flighting** refers to periodic waves of advertising interspersed with periods of complete and absolute inactivity.
- **Continuous** advertising is a schedule with little or no variation in activity.
- **Pulsing** is a combination of the above two concepts. A continuous base of support augmented by intermittent bursts of heavy activity.

Flighting

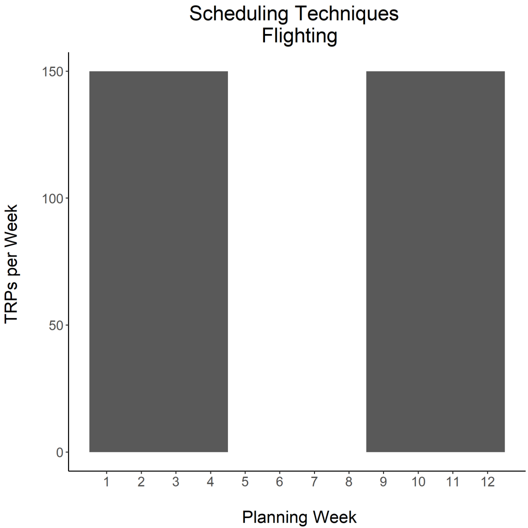

Figure 3: Flighting

Continuity

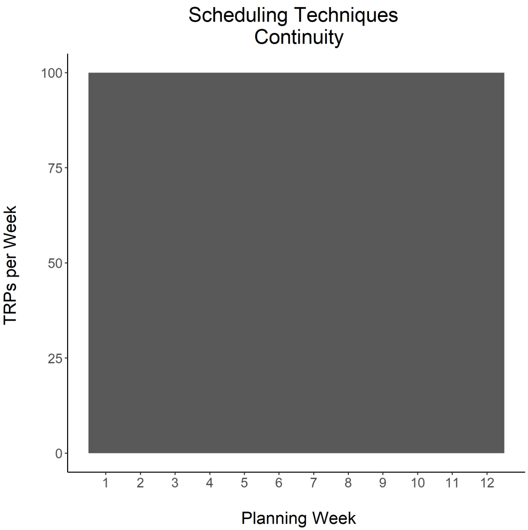

Figure 4: Continuity

Pulsing

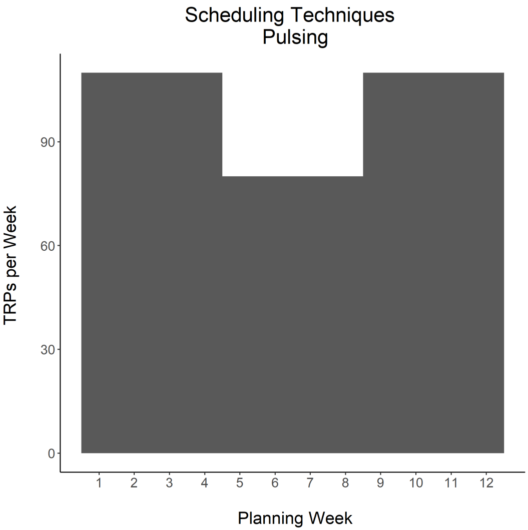

Figure 5: Pulsing

Impressions and Rating

For **measured media** – broadcast network TV, cable network TV, Spanish-language broadcast network TV, Spanish-language cable TV, network radio, local radio, magazines, newspapers, out-of-home and select digital advertising – advertising planning considers expenditures, target impressions, and target rating points.

- **Target audience**, or just audience, is a group of people defined based on their common characteristics, such as demographics and behaviors.
- **Impressions** are the sum of all exposures.
- **Rating** is the percentage of individuals (or homes) exposed to a particular media's program.

TRPs and GRPs

- **Target rating points** (TRPs) are the sum of audience ratings delivered by a given list of media vehicles. For a given audience, it is equal to 100 multiplied by audience reach multiplied by audience average frequency, where **reach** is the number of different of individuals (or homes) exposed to a media schedule within a given period of time.
- **Gross rating points** (GRPs) are the sum of ratings delivered by a given list of media vehicles. The reference to GRPs commonly indicates household (hh) rating points. In contrast, TRPs are commonly refer to people rating points.[18]

While GRPs are almost never used for planning, they are frequently used instead of expenditures to model the marketing mix.

[18] In the advertising industry one will frequently observe and hear people interchanging TRPs and GRPs. This is acceptable as long as the population base is being used in the reference.

Patterns of Advertising Response, Wear-In, Wear-Out and Response Curves

Temporal Effects of Advertising

There are multiple important patterns of response to advertising.

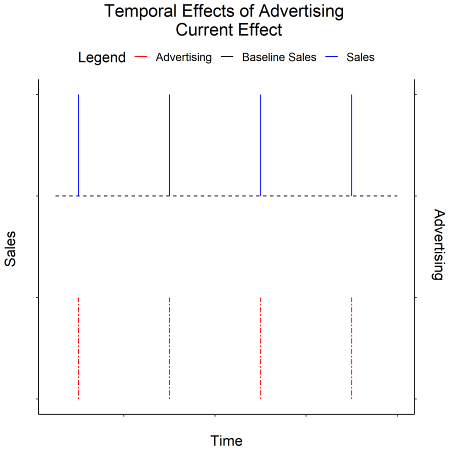

Figure 6: Current Effect

Carryover Effects of Short Duration

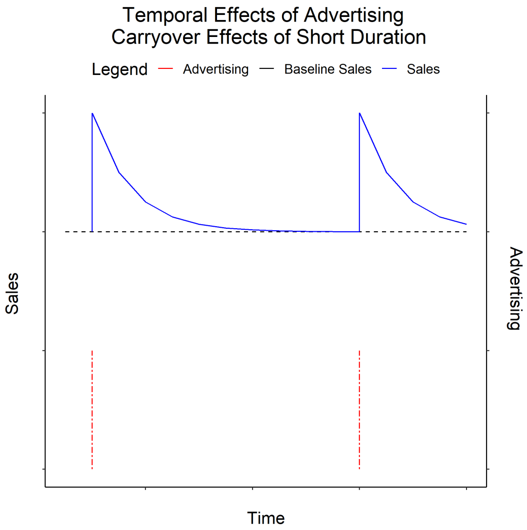

Figure 7: Carryover Effects of Short Duration

Carryover Effects of Long Duration

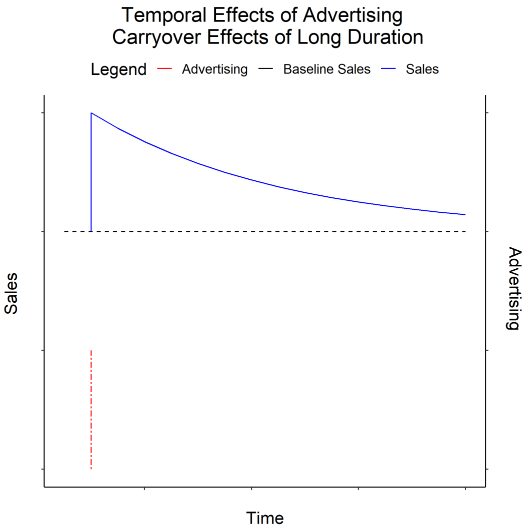

Figure 8: Carryover Effects of Long Duration

Persistent Effect

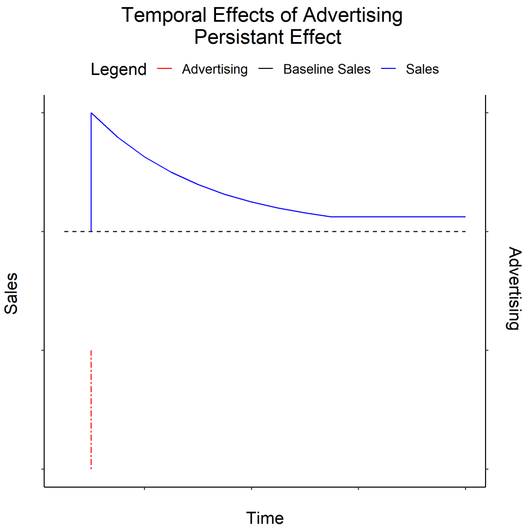

Figure 9: Persistent Effect

Adstock

A simple exponential function specification, referred to as **adstock**, leads to many of the effects given above.[19]

$$\text{Adstock}_t = \text{GRPs}_t + \phi \text{Adstock}_{t-1}, \tag{11}$$

where $\phi = 0.5^{1/\varphi}$. ϕ is referred to as the **retention rate**, and φ is the **half-life**.

A "two-week half-life" means that it takes two weeks for the awareness of advertising to decay to half its present level.

Beyond its use as an input to an adstock explanatory variable in a model to measure advertising effectiveness, half-life enables brand managers to efficiently space advertising schedules to maximize the effect of each advertising exposure.

[19] The idea of adstock originated from Broadbent (1979).

Dynamic Effects

Wear-in is the increase in the response of sales to advertising, from one week to the next of a campaign, even though advertising occurs at the same level each week.

If wear-in occurs, it typically occurs at the start of a campaign. It could occur because repetition of a campaign in subsequent periods enables more people to see the ad, talk about it, think about it, and respond to it than would have done so on the very first period of the campaign.

Wear-out is the decline in sales response of sales to advertising from week to week of a campaign, even though advertising occurs at the same level each week.

Wear-out typically occurs at the end of a campaign because of consumer tedium.

Wear-In and Wear-Out

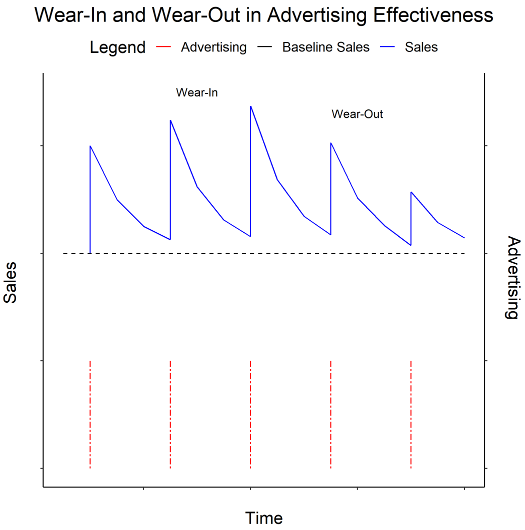

Figure 10: Wear-in and Wear-out

Linear and Nonlinear Response to Advertising

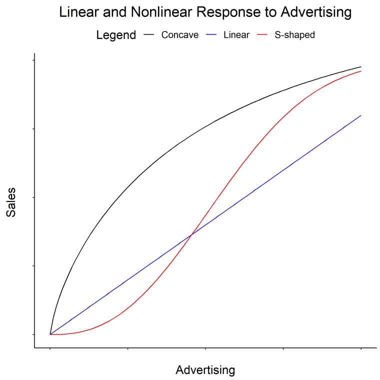

Figure 11: Linear and Nonlinear Response to Advertising

The Linear curve has constant returns to scale (CRTS), the Concave curve has decreasing returns to scale (DRTS), and the S-shaped has sections that have increasing returns to scale (IRTS) and DRTS, and an inflection point with CRTS.

Synergy Measurement via Linear Model Moderation

Integrated marketing is an approach that uses different media channels to tell a story or convey an idea.

- An integrated marketing campaign might start with a TV ad featuring a memorable character and message. The character could then appear in other channels: on billboards, on in-store displays, on coupons, in social media posts, in the original TV ad re-posted to YouTube, even in direct mail and email messages.[20]
- By marketing their character across complementary channels, the company creates strong brand consumer awareness and association. Thus, integrated marketing goes beyond the concurrent use of multiple media.[21]

[20] Coupons are a type of **consumer promotion** (CP). A subdivision of a firm's marketing department typically manages CP and customer shopper marketing tactics (e.g., demonstrations, in-store radio, brand displays on shopping carts, in-store floor graphics . . .).

[21] In the standard multimedia approach the effectiveness of an activity does not depend upon any other activity.

Synergy

In integrated marketing each activity's effectiveness depends on all other brand activities.

Synergy represents the joint effect of two different activities. It emerges when the combined effect of two activities exceeds the sum of their individual effects.

- Synergies may arise from combining different media types, scheduling their in-phase or out-phase timing, using consistent creative designs, and creating cross-media integrated content.
- As synergy increases, the optimal total media budget increases as well. Moreover, incremental outlays should be allocated be allocated in favor of the less effective media.[22]

[22] For additional information see Naik and Raman (2003).

Moderation

Under appropriate assumptions, the expected long-term profit of an advertised brand increases as synergy increases. Thus there are many instances where brand managers allocate a non-zero budget outlay to a catalytic activity even if it is completely ineffective.[23]

For the linear model specification, synergy is measured by **moderation**.

- Moderation is enacted by including an interaction term in the linear model's specification.
- An **interaction** is the multiplication of two explanatory variables.[24]

---

[23] For insights on how this result was derived, see Raman and Naik (2004).

[24] The reference article for the moderation presentation of the textbook is Spiller et al. (2013).

Synergy R&D Continues in the Industry

While synergy measurement is still an active area of marketing research, where much of the focus is on examining different non-linear model specifications, estimation methods, and long-term relationships, some practitioners use moderated regression to measure synergy.

We will illustrate moderated regression by examining the potential synergy between coupons and TV for a consumer packaged goods (CPG) brand.[25]

[25] A paper that attempts to measure the long-term synergy of coupons and other media is Reimer et al. (2014). In this paper they introduce a non-moderated regression modeling approach that enables brand managers to quantify marketing effectiveness based on all available data. Their approach combines existing “best practices” methods of customer segmentation and long-run effects modeling to investigate marketing mix effectiveness.

Defining Numeric Distribution and ACV

For our moderated regression specification, we will need a few definitions.[26]

Numeric Distribution %: The percentage of stores that stock a given item (e.g., UPC, sub-brand, brand ...) within the universe of stores in the relevant market.[27]

All Commodity Volume (ACV): The dollar value of store sales in all categories by stores in the relevant market.

% ACV: The dollar value of store sales in all categories by stores in the relevant market where the item is carried divided by ACV.

[26] For additional details on these definitions see Bendle et al. (2020).

[27] A **relevant market** is a set of stores that have a set of common characteristics. For example, "Club" is a commonly examined relevant market, where the market includes B.J's, Costco, and Sam's Club.

Defining PCV, TDPs and Velocity

Product Commodity Volume (PCV): For a given item's category, the dollar value of category sales by stores in the relevant market.

% PCV: For a given item's category, the dollar value of store category sales by stores in the relevant market where the item is carried divided by PCV.

Total Distribution Points (TDP): For a given item, the sum of % ACV of its UPCs for the relevant market.[28]

Velocity: An item's relevant market equivalized unit sales divided by the relevant market's ACV. This is a measure of how fast an item moves, controlling for differences in distribution.

[28] For a detailed description, see this insightful Nielsen IQ explanation.

Markets Of Interest

As the relevant market to determine if there is synergy between coupons and TV, the CPG brand has chosen the top 20 \$2M+ ACV Grocery Chain Holding Companies. Examples of such companies are Cincinnati-headquartered The Kroger Company and Giant Eagle, Inc.

The brand is from a category that does not have discernible seasonality. The Tissues Category, Laundry Detergent Category, and Men's Deodorant Categories are examples of categories that do not have pronounced seasonal effects.

The dependent variable of interest for the brand analysis is velocity. The denominator, ACV, is in billions of dollars.

Model dimensions are markets indexed by m, $m \in \{1, 2, \ldots 20\}$. Time is measured in 4-week periods indexed by t, $t \in \{1, 2, \ldots, 26\}$. The brand is denoted as b.

Improving the Precision of the Objective Statement

The objective of the analysis is to determine if coupons (i.e., moderator) changes the association between national broadcast TV GRPs (i.e., the primary marketing mix modeling explanatory variable) and velocity. The coupons are market-specific (i.e., retailer-specific) promotional offers distributed through retailer-specific circulars, such as Sunday newspaper retailer inserts.

We will use CPG brand household TV GRPs as the TV explanatory variable.[29]

We will use a **dummy explanatory variable** to model coupons.

[29] This models assumes the adstock retention rate is 0 and all markets receive the same number of national brand hh GRPs for a given t.

Model Specification

Given data available, the linear model specification is

$$\frac{Y_{bmt}}{ACV_{mt}} = \alpha_1 + \beta_1 x_{bt} + \gamma_1 z_{bmt} + \delta_1 x_{bt} z_{bmt} + \epsilon_{bmt}, \tag{12}$$

where Y_{bmt} is equivalized units of b in m during t, ACV_{mt} is all commodity volume of m in t, z_{bmt} is a dummy variable that is 1 if there was a coupon drop for b in m during t and 0 otherwise, and x_{bt} are national hh TV GRPs of b during t.

$x_{bt} z_{bmt}$ is the interaction term.

The errors ϵ_{bmt} are assumed to be IID, where independence is across all m and all t. They are assumed to be normally distributed.

Model Specification Interpretation

To emphasize, notice that the marketing mix variable of interest, brand b TV hh GRPs, and the moderator, the coupon dummy variable, and the iteration term, enter the model specification separately.

In addition, it is not appropriate to include an interaction term in the model without the main effects; x_{bt} and z_{bmt} each entering the model specification separately.

Importance of Moderator *p*-value

The textbook mentions that if we are wondering whether there is evidence in the data that the association between the marketing outcome and the marketing mix variable (i.e., TV GRPs) changes depend on the moderator (i.e., retailer-specific coupons), then we just check for the significance of the interaction by examining its *p*-value.

- If the interaction is significant, then there is evidence that the marketing outcome's association with the marketing mix variable depends on that moderator.
- If it is not significant, then we say that the marketing outcome's association with the marketing mix variable does not depend on that moderator.

Model Estimation Output from the R `lm` Function

```
Moderator Regression - TV GRPs and Coupons
Call:
lm(formula = velocity ~ x + z + xz, data = df3)
Residuals:
     Min       1Q   Median       3Q      Max
-2559.54  -348.74    39.96   439.08  1424.21
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6589.1435   247.7924  26.591   <2e-16 ***
x              1.1861     0.4962   2.390   0.0172 *
z           2682.5474  1163.0017   2.307   0.0215 *
xz             2.0002     2.3106   0.866   0.3871
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 610.7 on 516 degrees of freedom
Multiple R-squared:  0.6203,Adjusted R-squared:  0.6181
F-statistic:   281 on 3 and 516 DF,  p-value: < 2.2e-16
```

The interaction is not statistically significant. This means that that association between the brand's TV hh GRPs and the brand's velocity do not depend on whether coupons are dropped, or not dropped, in a market. That is, there is not a discernible synergy effect from coupons and TV on velocity.

- $\hat{\alpha}_1$, 6589.14, is the estimate of brand velocity when brand hh TV GRPs are 0, $x_{bt} = 0$, and there is not a brand coupon drop, $z_{bmt} = 0$.
- $\hat{\beta}_1$, 1.19, is the simple slope of brand hh TV GRPs on brand velocity when there is not a brand coupon drop.
- $\hat{\gamma}_1$, 2682.55, is the simple effect of a brand coupon drop versus no brand coupon drop when when brand hh TV GRPs are 0.
- $\hat{\delta}_1$, 2.00, is the change in effect of the brand coupon drop versus no brand coupon drop when hh TV GRPs increase by 1 unit. The effect is not statistically significant.

Another Moderator Regression Case Study

For this case study the focus is on the moderating effect of digital radio on national radio (e.g., The Sean Hannity Radio Show). The digital radio platforms are Apple Music, Amazon Music, Pandora and Spotify.

While the company is the same as that of the previous case study, the brand is different. Similar to the brand of the previous case study, this brand's seasonality is not pronounced.

While for more than a decade the brand has advertised on national radio, only during the last 3 years has the brand used digital radio advertising. The business problems to be addressed are:

- To determine if digital radio is moderating the effectiveness of national radio.
- If digital radio is moderating national radio, determine its magnitude.

Specification of the Moderator Regression

Given data available, the linear model specification is

$$\frac{Y_{bt}}{ACV_t} = \theta_0 + \theta_1 x_{1bt} + \theta_2 x_{2bt} + \theta_3 x_{1bt} x_{2bt} + \upsilon_{bt}, \tag{13}$$

where Y_{bt} is national equivalized units of b during t, ACV_t is all commodity volume of all channels in which b is distributed in t, x_{1bt} are digital radio adults aged 18 to 54 (A 18-54) impressions in millions (i.e., the moderator), and x_{2bt} are national radio A 18-54 impressions in millions.[30]

$x_{1bt}x_{2bt}$ is the interaction term.

The errors υ_{bt} are assumed to be IID and normally distributed.

[30] Like the previous case study's model, this simple models assumes the adstock retention rate is 0 for both radio series.

Moderating Regression Estimates

```
Moderator Regression - Spotify, Amazon Music, Apple Music,
and Pandora and National Radio Adult 18-54 Impressions
Call:
lm(formula = velocity ~ x1 + x2 + x1x2, data = df_7)
Residuals:
    Min      1Q  Median      3Q     Max
-4174.9 -1546.3  -152.9  1151.5  4577.3
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.338e+04  9.677e+02  34.492   <2e-16 ***
x1          2.563e+01  1.231e+01   2.082   0.0492 *
x2          3.638e+01  1.719e+01   2.116   0.0459 *
x1x2        2.970e-01  1.442e-01   2.059   0.0515 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 2201 on 22 degrees of freedom
Multiple R-squared:  0.883,Adjusted R-squared:  0.8671
F-statistic: 55.37 on 3 and 22 DF,  p-value: 2.048e-10
```

Interpreting the Moderating Regression Results

The above results imply that national radio A 18-54 impressions are associated with the velocity of brand *b*, where there is dependence on digital radio A 18-54 impressions. While the interaction effect coefficient is positive and statistically significant at a significance level of 0.10, the effect of brand brand *b* national radio on brand *b* velocity many not be constant across all values of the moderating explanatory variable, digital radio A 18-54 impressions.

Spotlight analysis compares marketing mix variable estimates across two or more values of the moderating explanatory variable. For example, we may mean center the digital radio explanatory variable x_{1bt} and re-estimate the model. From the mean-centered value of x_{1bt} we may also add or subtract 1 standard deviation (SD) and re-estimate the model to obtain other spotlights.[31]

[31] The interaction term is a function of the mean centered moderating explanatory variable.

Spotlighting at the Moderator's Mean

```
Moderator Regression - Spotify, Amazon Music, Apple Music,
and Pandora and National Radio Adult 18-54 Impressions
Spotlighting at the Moderator's Mean
Call:
lm(formula = velocity ~ x1_meanc + x2 + x1_meanc_x2, data = df_7)
Residuals:
    Min      1Q  Median      3Q     Max
-4174.9 -1546.3  -152.9  1151.5  4577.3
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.521e+04  7.975e+02  44.147  < 2e-16 ***
x1_meanc    2.563e+01  1.231e+01   2.082   0.0492 *
x2          5.760e+01  1.025e+01   5.621 1.19e-05 ***
x1_meanc_x2 2.970e-01  1.442e-01   2.059   0.0515 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2201 on 22 degrees of freedom
Multiple R-squared:  0.883,Adjusted R-squared:  0.8671
F-statistic: 55.37 on 3 and 22 DF,  p-value: 2.048e-10
```

High-Level Spotlighting

```
Moderator Regression - Spotify, Amazon Music, Apple Music,
and Pandora and National Radio Adult 18-54 Impressions
Spotlight is 1 SD ABOVE the Mean-Centered Moderator
Call:
lm(formula = velocity ~ x1_sdcn1 + x2 + x1_sdcn1_x2, data = df_7)
Residuals:
    Min      1Q  Median      3Q     Max
-4174.9 -1546.3  -152.9  1151.5  4577.3
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.671e+04  1.244e+03  29.512  < 2e-16 ***
x1_sdcn1    2.563e+01  1.231e+01   2.082   0.0492 *
x2          7.504e+01  1.036e+01   7.242 2.96e-07 ***
x1_sdcn1_x2 2.970e-01  1.442e-01   2.059   0.0515 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2201 on 22 degrees of freedom
Multiple R-squared:  0.883,Adjusted R-squared:  0.8671
F-statistic: 55.37 on 3 and 22 DF,  p-value: 2.048e-10
```

Low-Level Spotlighting

```
Moderator Regression - Spotify, Amazon Music, Apple Music,
and Pandora and National Radio Adult 18-54 Impressions
Spotlight is 1 SD BELOW the Mean-Centered Moderator
Call:
lm(formula = velocity ~ x1_sdcp1 + x2 + x1_sdcp1_x2, data = df_7)
Residuals:
    Min      1Q  Median      3Q     Max
-4174.9 -1546.3  -152.9  1151.5  4577.3
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.370e+04  8.773e+02  38.419   <2e-16 ***
x1_sdcp1    2.563e+01  1.231e+01   2.082   0.0492 *
x2          4.016e+01  1.569e+01   2.561   0.0178 *
x1_sdcp1_x2 2.970e-01  1.442e-01   2.059   0.0515 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 2201 on 22 degrees of freedom
Multiple R-squared:  0.883,Adjusted R-squared:  0.8671
F-statistic: 55.37 on 3 and 22 DF,  p-value: 2.048e-10
```

Interpretation of Spotlights

The coefficient estimates of the moderating explanatory variable, $\hat{\theta}_1$, and the interaction term, $\hat{\theta}_3$, are the same for all moderating regression specifications. This also holds for these coefficient estimates' standard errors. This result follows from the mathematics of least squares.

Spotlighting at the Moderator's Mean When digital radio A 18-54 impressions are average, for every 1M increase in network radio A 18-54 impressions velocity increases by 57.6.

High-Level Spotlighting When digital radio A 18-54 impressions are higher, for every 1M increase in network radio A 18-54 impressions velocity increases by 75.04.[32]

[32] "High" is one standard deviation greater than the mean of digital radio A 18-54 impressions.

Floodlight Analysis

Low-Level Spotlighting When digital radio A 18-54 impressions are lower, for every 1M increase in network radio A 18-54 impressions velocity increases by 40.16.[33]

Floodlight analysis is spotlight analysis for many values across a wide range of continuous moderator variable values.

Spotlight analysis and floodlight analysis are useful modeling tools for marketing analytics because they allow one to examine associations between a marketing mix variable and a marketing outcome variable at different moderator variable levels.

[33] "Low" is one standard deviation less than the mean of digital radio A 18-54 impressions.

References

Bendle, N., Farris, P., Pfeifer, P., and Reibstein, D. (2020). *Marketing Metrics - The Manger's Guide to Measuring Marketing Performance*. Pearson FT Press, Upper Saddle River, NJ, 4th edition.

Broadbent, S. (1979). One way TV advertisements work. *Journal of the Market Research Society*, 21(3):139–166.

Cooper, L. G. and Nakanishi, M. (1988). *Market Share Analysis*. Kluwer, Norwell, MA.

Davis, B. (2022). *Marketing Analytics*. Edify Publ., ISBN: 978-1-7346888-4-9.

Freedman, D., Rothenberg, T., and Sutch, R. (1983). On energy policy models. *Journal of Business & Economic Statistics*, 1(1):24–32.

Jin, Y., Wang, Y., Sun, Y., Chan, D., and Koehler, J. (2017). Bayesian methods for media mix modeling with carryover and shape effects.

Johnston, J. and DiNardo, J. (1997). *Econometric Methods*. The McGraw-Hill Companies, New York, NY, 4th edition.

Malinvaud, E. (1981). Econometrics faced with the needs of macroeconomic policy. *Econometrica*, 49(6):1363–1375.

Naik, P. A. and Raman, K. (2003). Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research*, 40(4):375–388.

Raman, K. and Naik, P. A. (2004). Long-term profit impact of integrated marketing communications program. *Review of Marketing Science*, 2(1):0000102202154656161014.

Reimer, K., Rutz, O. J., and Pauwels, K. (2014). How online consumer segments differ in long-term marketing effectiveness. *Journal of Interactive Marketing*, 28(4):271–284.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sissors, J. Z. and Baron, R. B. (2010). *Advertising Media Planning*. McGraw-Hill, New York, NY, 3rd edition.

Solana, J. and Weikel, B. (2014). Marketing mix modeling - What marketing professionals need to know. Product Leadership GRAD Technical Report, Nielsen.

Spiller, S. A., Fitzsimons, G. J., John G. Lynch, J., and Mcclelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, 50(2):277–288.

Tellis, G. J. (2006). Modeling the marketing mix. In Grover, R. and Vriens, M., editors, *The Handbook of Marketing Research : Uses, Misuses, and Future Advances*, pages 506–522. SAGE Publications, Inc., New York, NY.

Tellis, G. J., Chandy, R. K., and Thaivanich, P. (2000). Which ad works, when, where, and how often? modeling the effects of direct television advertising. *Journal of Marketing Research*, 37(1):32–46.

Verwiebe, P. A., Seim, S., Burges, S., Schulz, L., and Müller-Kirchenbauer, J. (2021). Modeling energy demand – A systematic literature review. *Energies*, 14(23):7859.