

Marketing Mix Modeling Guidebook

2023.9 Hakuhodo DY media partners Inc.

CHAPTER 1 Introduction of Marketing Mix Modeling

1. Marketing Effectiveness Measurement Challenges [p.3](#)
2. Marketing Mix Modeling Overview [p.8](#)
3. Background of MMM's Renewed Interest [p.13](#)

CHAPTER 2 Basics of Marketing Mix Modeling

1. Data Selection [p.21](#)
2. Data Cleansing [p.30](#)
3. Creating Model Structure [p.40](#)
4. Parameter Estimation [p.58](#)
5. Model Validation [p.63](#)
6. Model Utilization [p.78](#)
7. Other Topics on MMMs [p.83](#)

CHAPTER 3 Considerations on MMM model structure

1. Model Structure Creation Methods and Model Types [p.88](#)
2. Verification of the Impact of Model Structure on Analysis Results [p.96](#)
3. Simulation Study [p.102](#)
4. Problems with Model Structure and Hints for Solution [p.119](#)
5. HakuHodo DY Group MMM Solutions [p.132](#)

CHAPTER

1

Introduction of Marketing Mix Modeling

1.1. Challenges in Measuring Marketing Effectiveness

Companies are looking for explanations for media and marketing contribution to business results, not reach and awareness metrics.

Indicators that have been measured so far

- Reach / Reach efficiency
- Ad recognition / Recall Rate
- Brand recognition rate
- Brand image
- Purchase intention rate



Indicators that currently require explanation

- Number of new acquisitions
- WAU / MAU
- Sales volume
- Sales / Market share


Decisions must be made based on integrated metrics, such as contribution to business results, to maximize marketing effectiveness.



Cookie* utilization, developed in digital advertising, is a method that can measure business contribution, but with the advent of the cookie-less era, the scope of its use is narrowing.

Expectations for Cookies

Combine TV viewing logs, survey data, and POS as well as digital advertising with cookies



Measuring User Behavior with Cookies

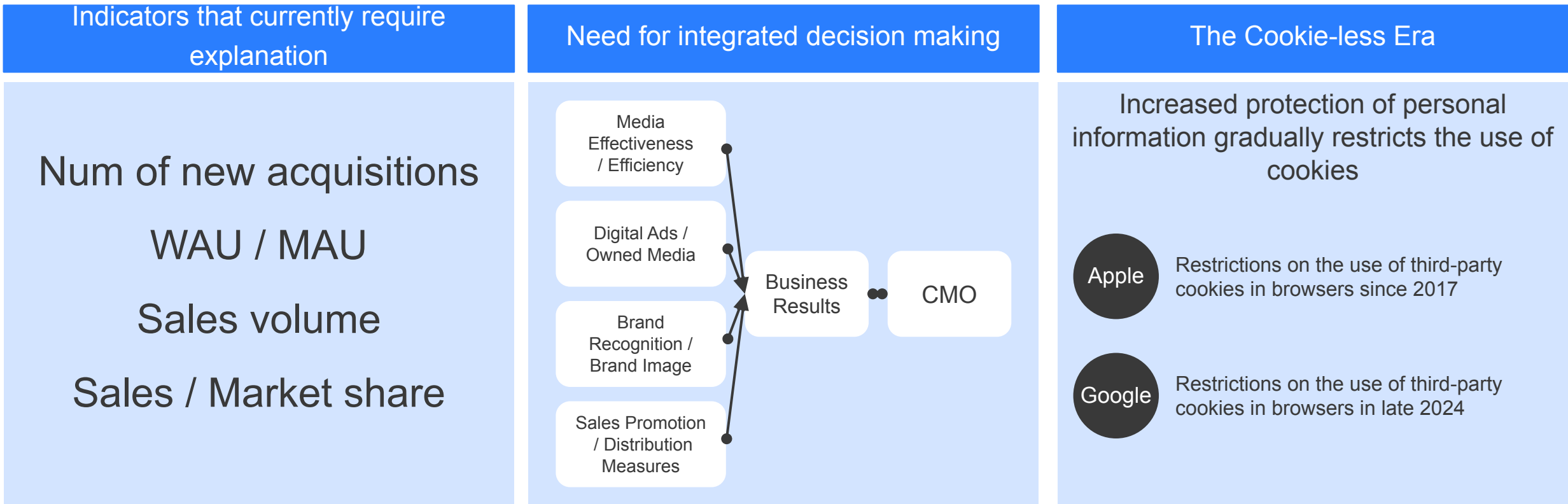
The Cookie-less Era

Increased protection of personal information gradually restricts the use of cookies

- Apple: Restrictions on the use of third-party cookies in browsers since 2017
- Google: Restrictions on the use of third-party cookies in browsers in late 2024

*Cookie=3rd party Cookie

There is a need for a method that can measure the business contribution of media and marketing in a cross-sectional and integrated manner and without relying on cookies.



1.2. Marketing Mix Modeling Overview

MMM is a method for statistically estimating the business contribution of media and marketing activities without relying on cookies.

▼ Conceptual Diagram of MMM



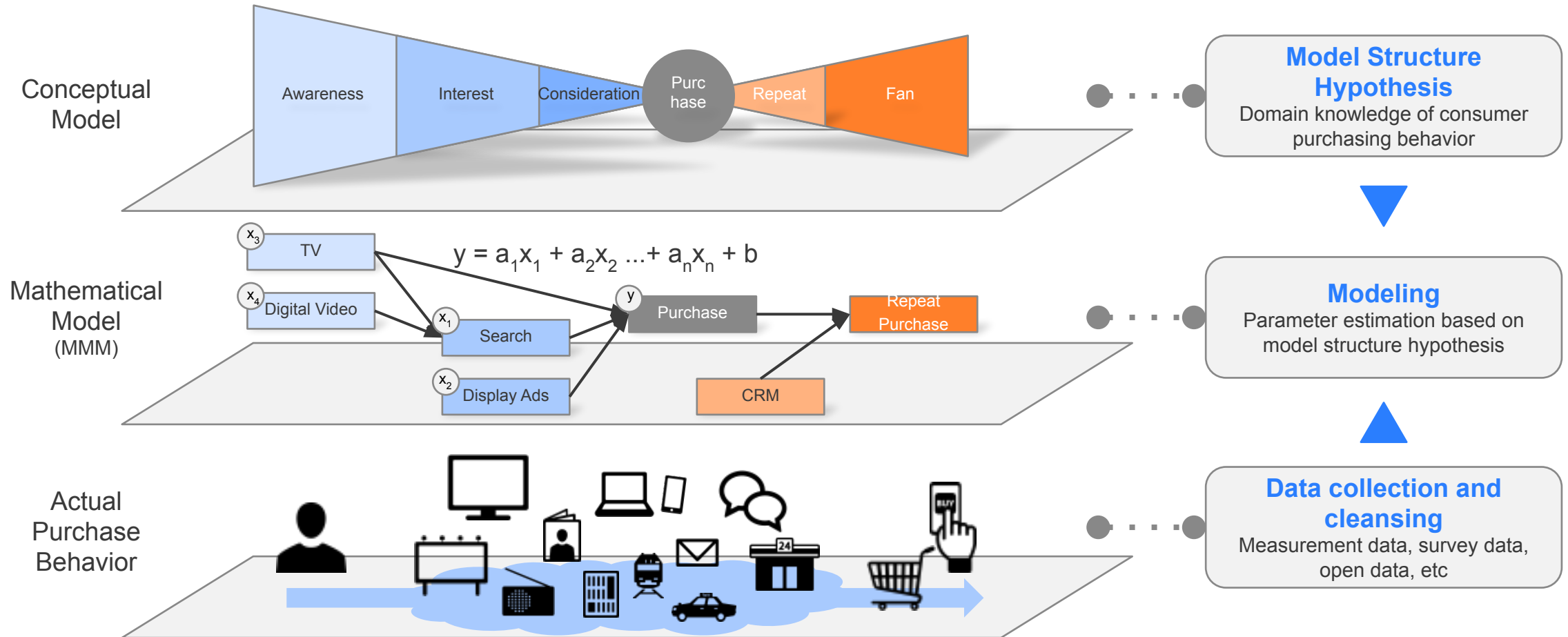
▼ Benefits of using MMM

MMM can **statistically estimate** marketing effectiveness **without relying on cookies**

MMM can estimate **marketing effectiveness in terms of business contribution** such as number of acquisitions, sales volume and sales

MMM can estimate **the effectiveness of various media, marketing activities, and external factors** as well as digital advertising

In MMM, we collect and process data that measures consumer behavior and psychology and expresses the relationship between media/marketing activities and business results.



MMM enables you to diagnose, predict and prescribe media and marketing activities based on business contribution.

1.

What, how much, and how are the media and marketing activities contributing to the business?



Understanding Marketing Effectiveness

We can understand the contribution of each factor to business results across

2.

Are marketing and media budget allocations reasonable?



Budget Allocation Optimization

We can calculate the necessary costs and efficient budget allocation for targeted business results

3.

What evidence supports future marketing strategies and media plans?



Business Results Simulation

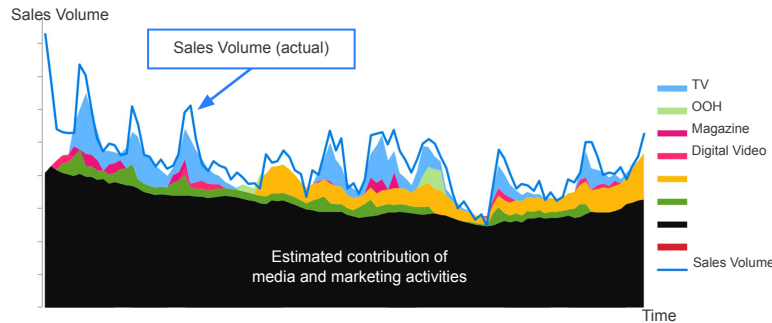
We can simulate the business results of the new marketing strategy and media plan

1. What, how much, and how are the media and marketing activities contributing to the business?



Understanding Marketing Effectiveness
We can understand the contribution of each factor to business results across

▼ Estimated contribution of media and marketing activities



▼ Cost-effectiveness of media and marketing activities

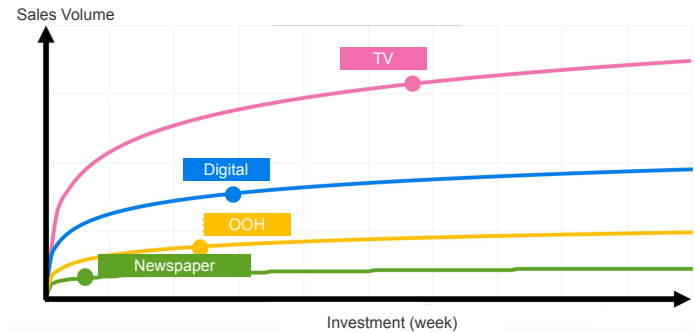
	Cost (MM)	Contribution (Order)	CPA (Cost per Order)
TV (TIME)	2,900	264,481	10,967
TV (SPOT)	623	60,304	10,330
BS	56	26,936	2,075
CS	170	49,140	3,451
Newspaper	559	28,078	19,902
Radio	193	49,009	3,928
OOH	152	26,551	5,732
Magazine	38	4,675	8,021
Digital Ads	764	171,810	4,445

2. Are marketing and media budget allocations reasonable?

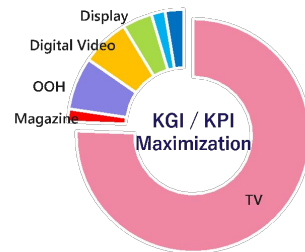


Budget Allocation Optimization
We can calculate the necessary costs and efficient budget allocation for targeted business results

▼ Elasticity of each media investment to business contribution



▼ Optimal budget allocation to targeted goals



[Optimal Budget Allocation]

TV	1億5120万円 (75.6%)
Magazine	355万円(1.8%)
OOH	1,455万円(7.3%)
Digital Video	1,328万円 (6.6%)
Display	832万円(4.2%)
Search Ads	386万円(1.9%)
SNS	524万円(2.6%)

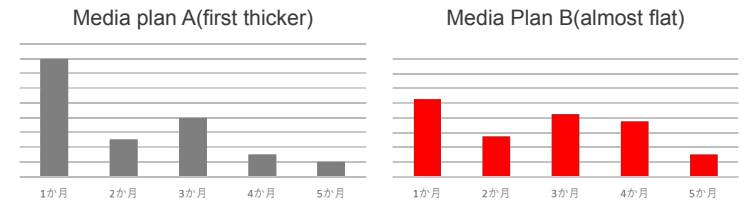
TTL: 2.0億円

3. What evidence supports future marketing strategies and media plans?

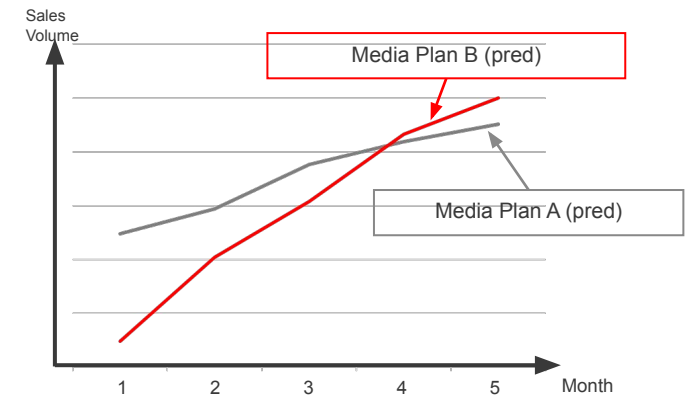


Business Results Simulation
We can simulate the business results of the new marketing strategy and media plan

▼ Media plans and Business Results Simulation

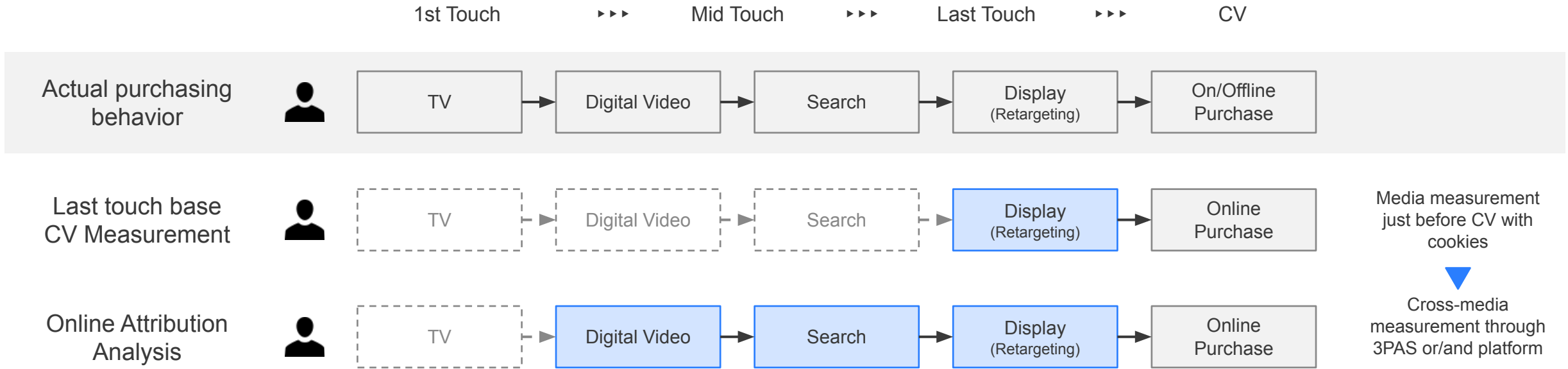


Business Results Simulation



1.3. Background of Renewed Interest in MMM

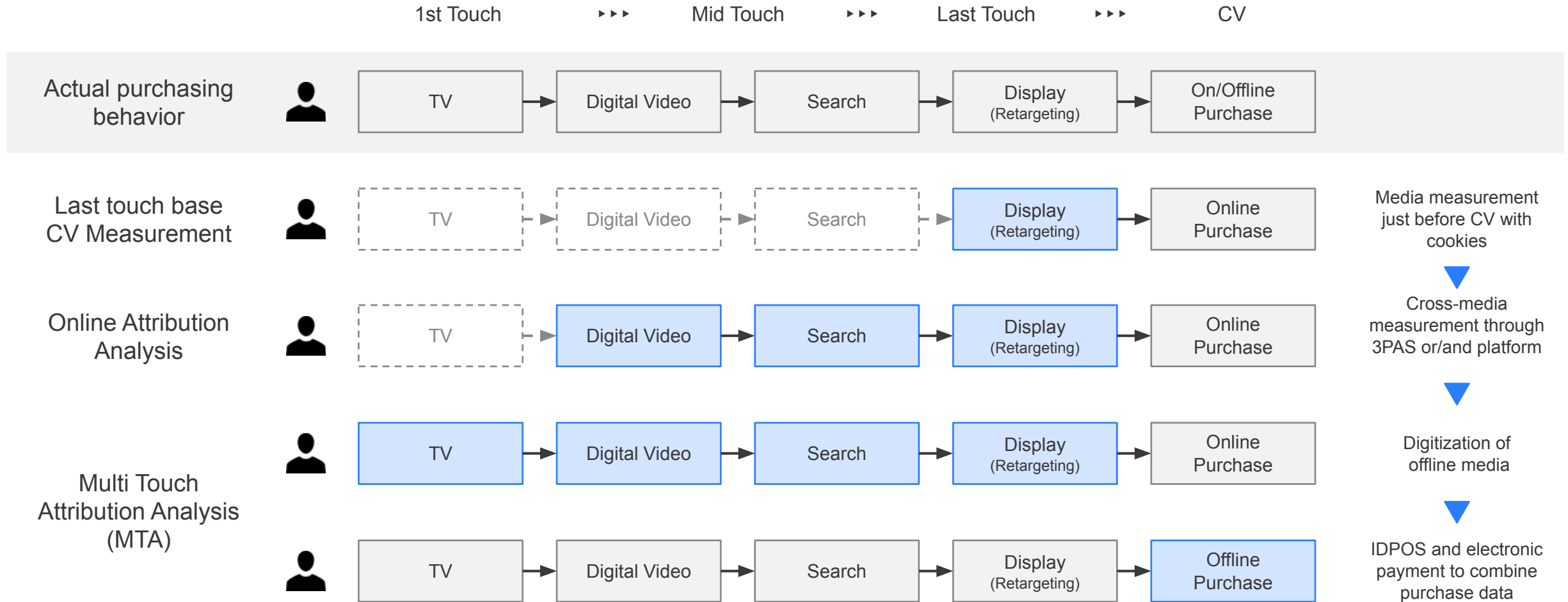
In digital ads, cookies make it possible to measure the path from the 1st contact to the point of purchase and calculate the contribution of each media contact.



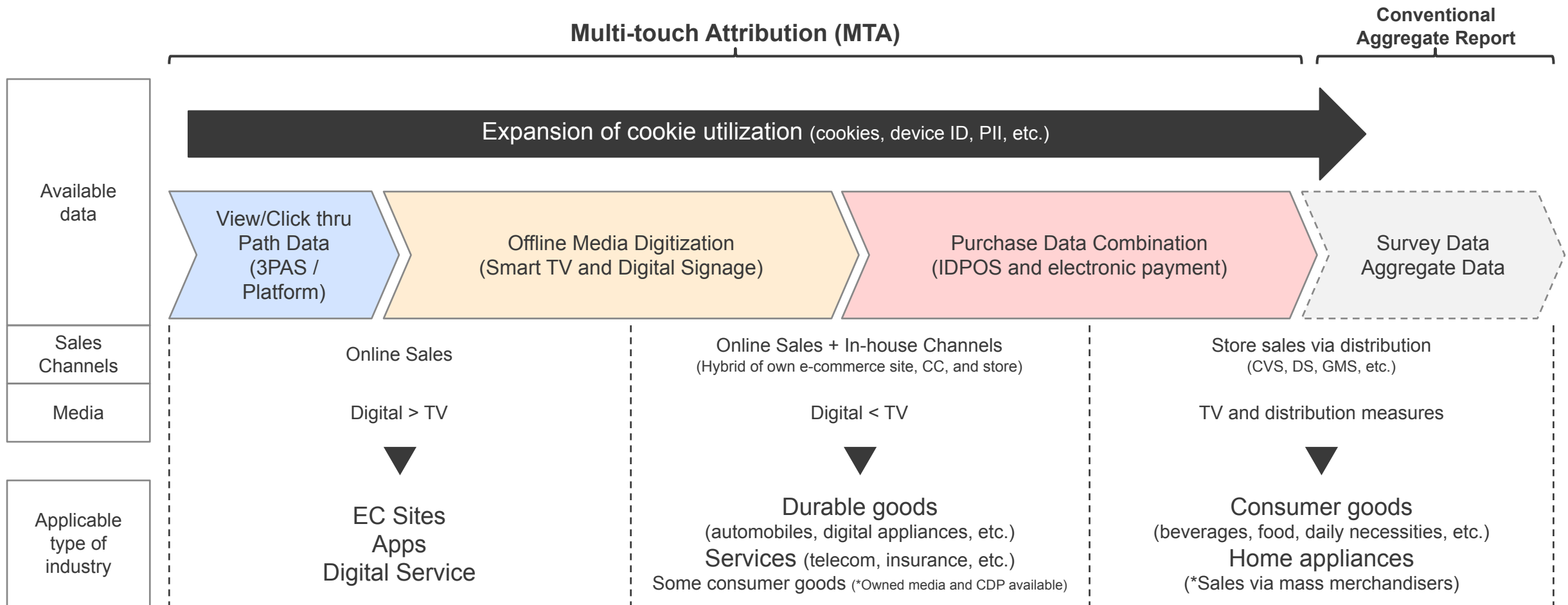
Media measurement just before CV with cookies

▼
Cross-media measurement through 3PAS or/and platform

With the penetration of smart TVs and electronic payment, offline contact points are also combined with data. It is now possible to analyze the path from advertising contact to purchase through on-offline integration.



Thus, there were high expectations that the method of measuring effectiveness using cookies could be extended to products purchased primarily offline.



However, due to causes such as cookie restrictions and low EC rate, data merging has not been fully realized, and the scope of MTA utilization is currently limited.

Cookie Restrictions

Cookie restrictions are already having a large impact, especially in Japan.

▼ Cookie Regulation Trends

Apple Restrictions on the use of third-party cookies in browsers since 2017

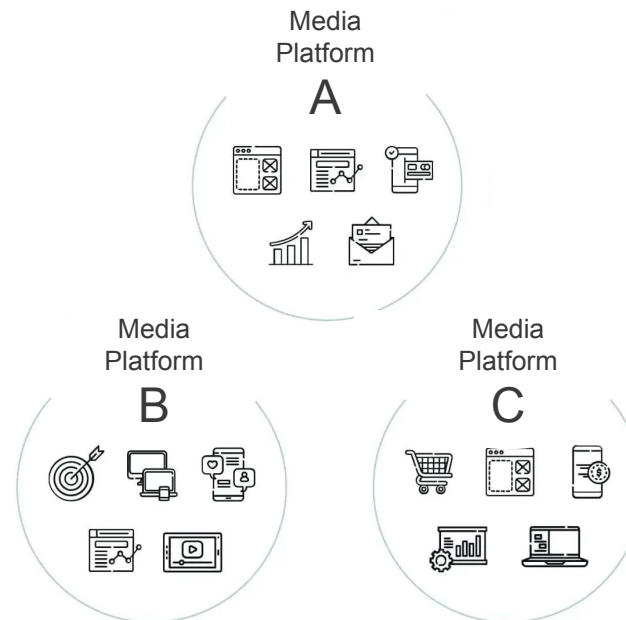
Google Restrictions on the use of third-party cookies in browsers in late 2024

▼ iPhone Market Share in Japan *1



Walled-Gardenization

Closed data utilization environment per Platform development is progressing



Low EC Rate

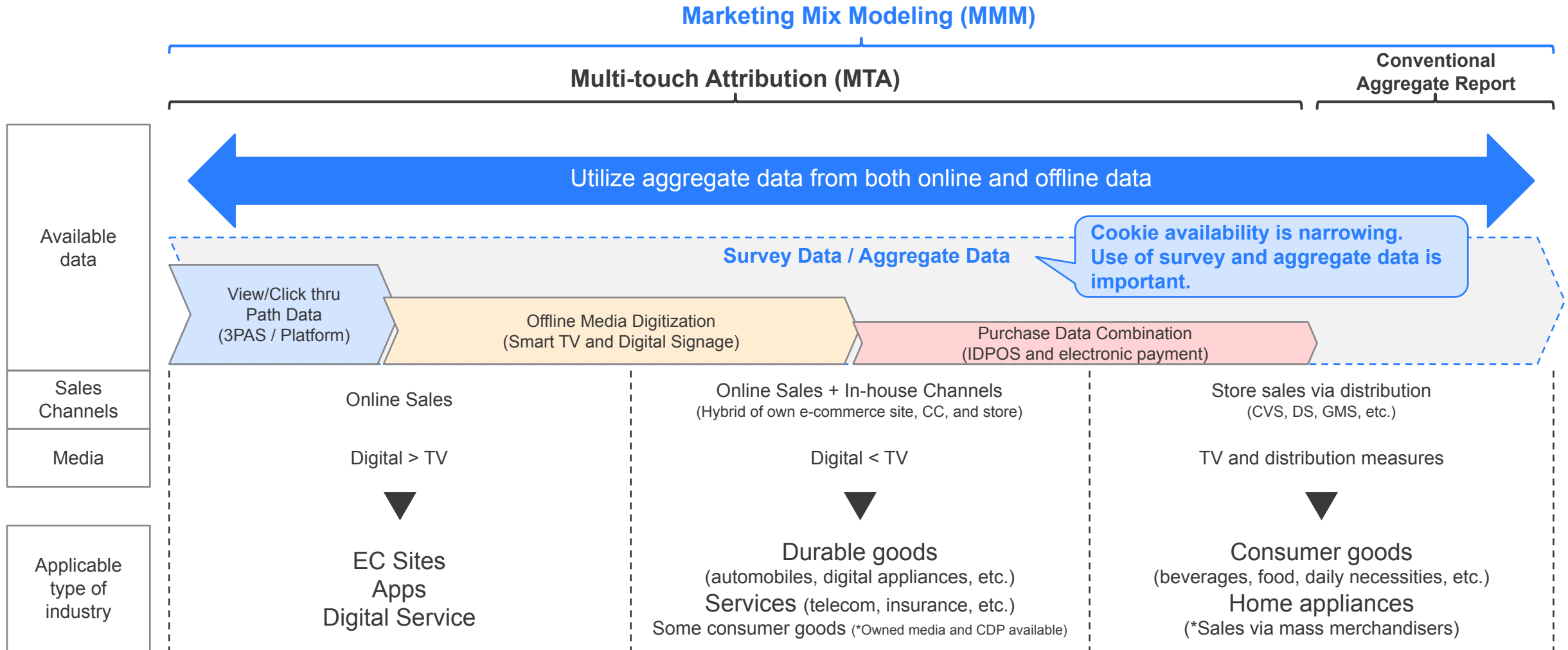
In some categories, the EC rate remains below 10%.

▼ EC rate by category *2

category	2021
Product Sales Average	8.8%
Books, Video, Music	46.2%
Home appliances, AV / PC	38.1%
Lifestyle & Interior	28.3%
Apparel	21.2%
Cosmetics	7.5%
Food, Beverage & Alcohol	3.8%
Automobiles & Motorcycles	3.9%
Other	2.0%

*1: statcounter *2: The Ministry of Economy, Trade and Industry (METI)"FY2021 E-Commerce Market Survey" (2021)

Therefore, MMM, which can estimate the business contribution of marketing through on-off integration using aggregate data without relying on cookies, is once again attracting attention.



CHAPTER

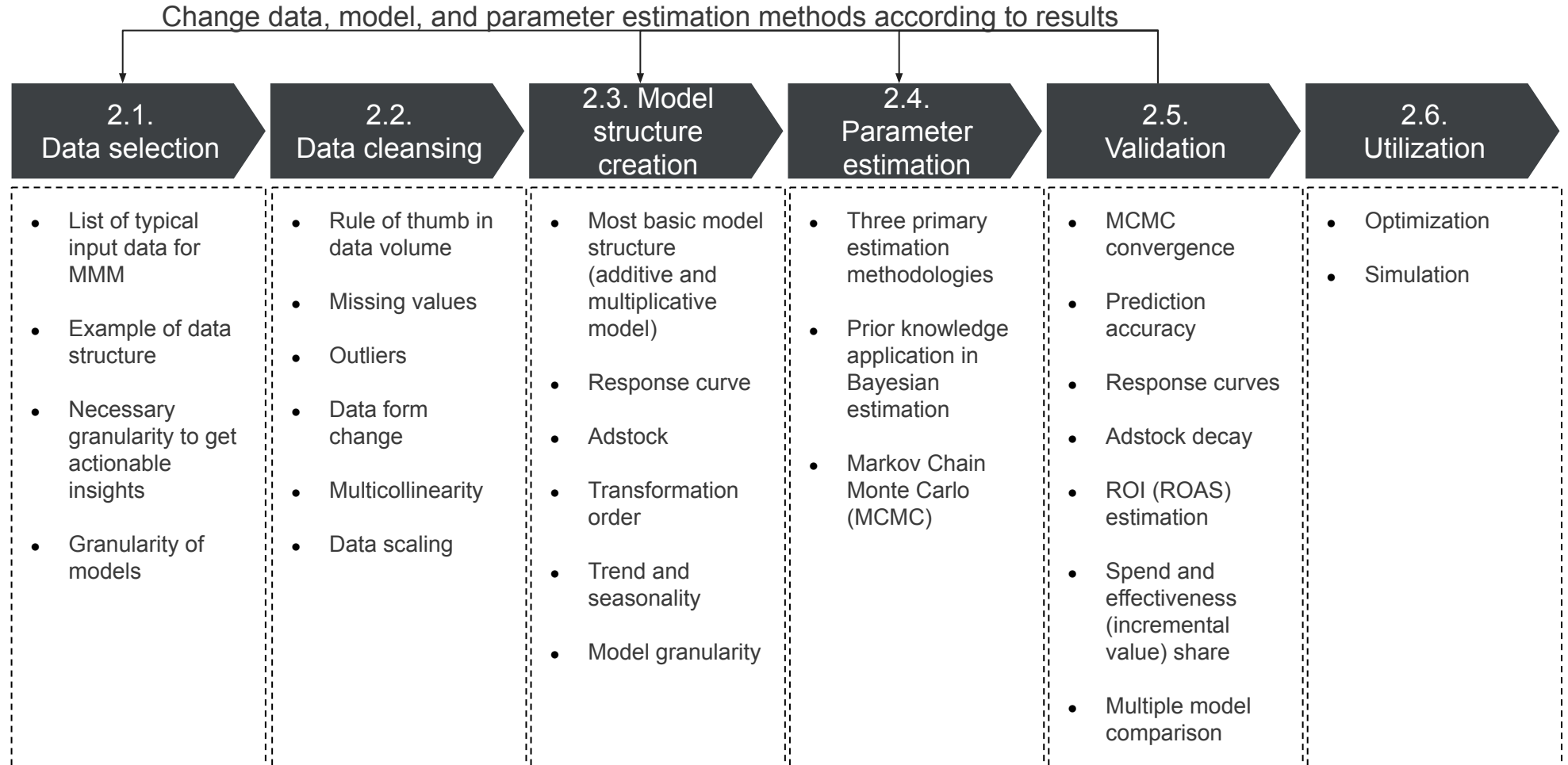
2

Basics of Marketing Mix Modeling

Generally the marketing mix modeling process consists of data selection, data cleansing, model structure creation, model creation, validation and utilization.

While some steps are consistent with general regression models, there are MMM specific considerations such as input data selection, response curve transformations, adstock and the validation of the results.

This section will walk you through step-by-step process and typical pitfalls.



2.1. Data Selection

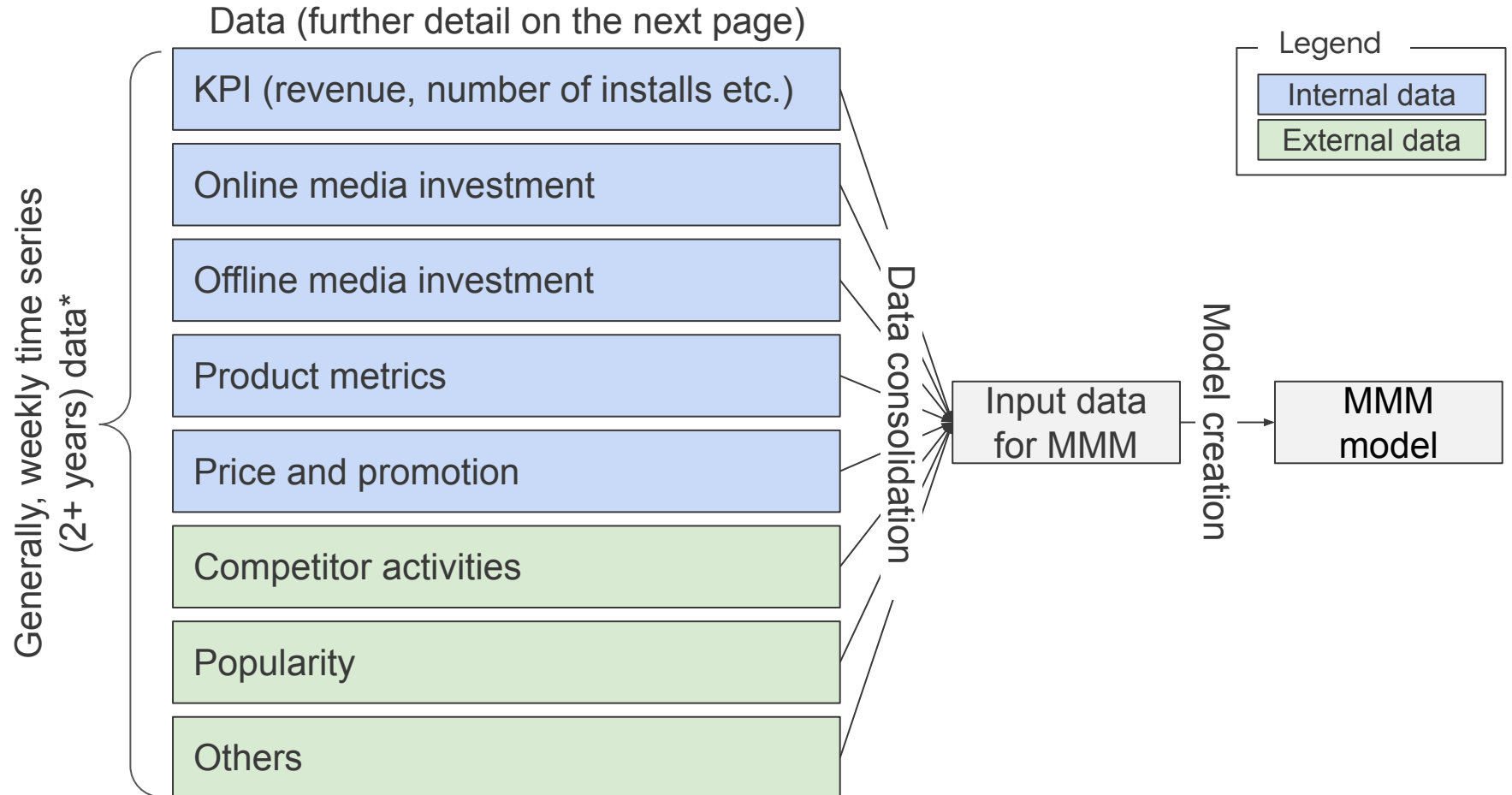
Data selection overview

22

This slide shows an overview of the data selection and consolidation process for marketing mix modelling. Models need to source various data across internal and external, and online and offline sources

All data needs to be time-series data (generally daily or weekly) with the same granularity. For example, if a modeler wants to create a daily level MMM, all the data must be daily level. Thus, input data granularity constrains the model granularity.

Generally data collection and consolidation and its automation (if a modeler intends to use the model continuously) takes several weeks.



*: data volume necessary for MMM depends on business

MMM models need data across KPI data, media data and control variables

23

	Data		Description	Example data source
Dependent variable	KPI (Key Performance Indicator) data		Revenue, number of conversions, number of active users, number of app installs etc.	Internal data
Independent variables	Media variables	Offline media metrics	Advertising spend or GRP of TV, radio, print and OOH	Internal and media agency's database
		Online media metrics	Advertising spend and impressions of digital media by product (YouTube, Google Search, Google App Campaign, Facebook, TikTok etc.)	Each media's API, ads manager etc.
	Control variables	Product metrics	Product metrics which may have an impact on the KPI (product update, app/web user ratings, consumer surveys etc.)	Internal data, app platform site etc.
		Price and promotion	Promotional data (price, promotion amount, promotion type, discount amount, in-store display type, inventory availability etc.) and event/trade show schedule	Internal data, retail data etc.
		Competitors' activities	Data regarding competitors (promotions, new product launches, media activity, user ratings, app rankings etc.)	Competitors' website, app platform sites etc.
		Popularity	Data describing trends of the modeled product (number of hashtags, Google Trends Index, app rankings etc.)	Each media's interface
Others	Macroeconomic data (e.g. GDP growth rate, Covid-19 infections etc.), seasonality (holidays, weather) etc.	Public data sources		

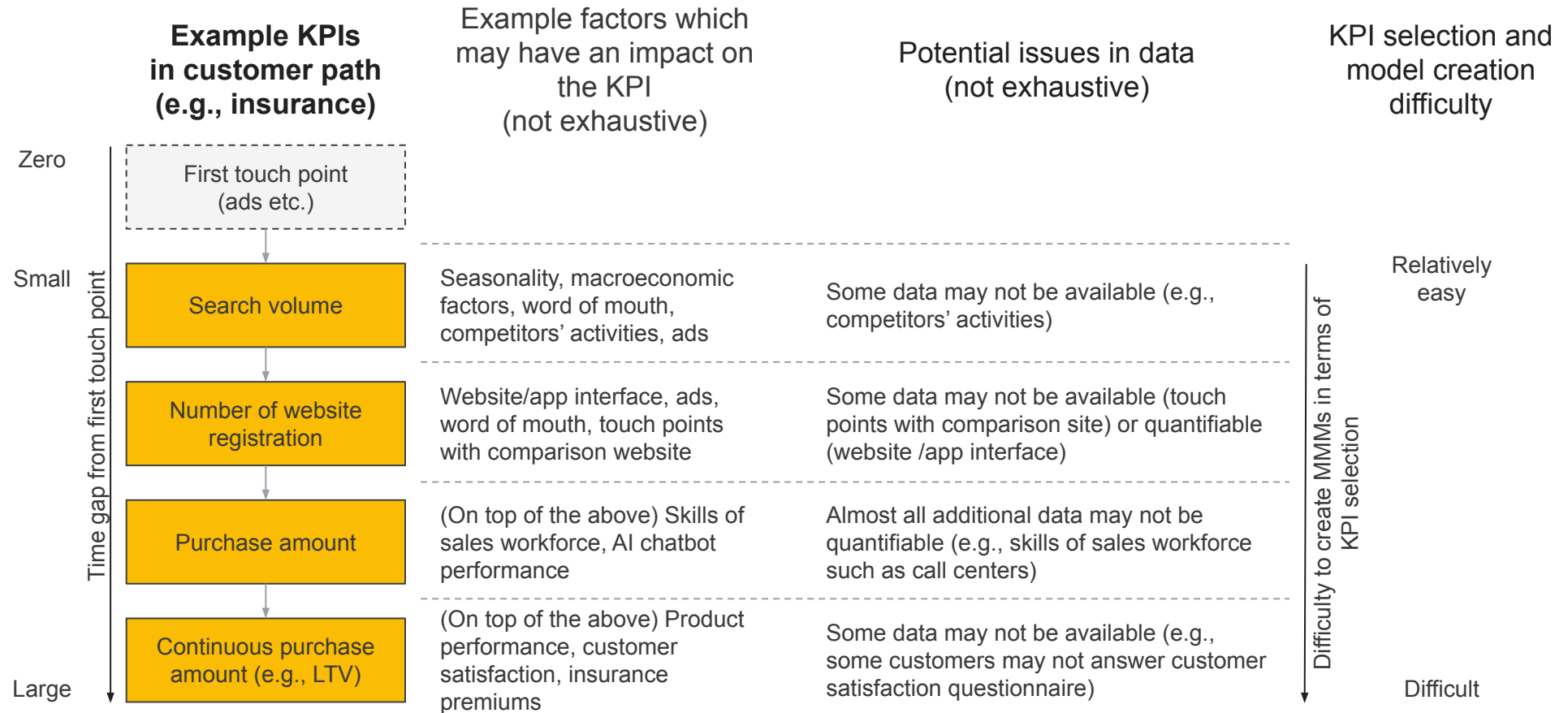
Generally, weekly and 2+ years data*

*: data volume necessary for MMM depends on business

KPI selection is another important factor to determine the most suitable model

24

A modeler needs to select appropriate KPI(s) in the model, generally intermediate variables such as brand search volume, number of purchases or final result KPIs such as purchase amount or LTV are selected. However, if deeper funnel KPIs (e.g., LTV) are selected, it is more difficult to create the MMMs because there are factors not available or quantifiable (e.g., skills of sales workforce). Hence, a modeler may need to compromise the KPI selection (e.g., select an intermediary variable such as search volume) as the KPI to avoid the data availability issue.



Spend, impressions or clicks for media variables?

25

One typical question in the data selection is how to select metrics for media variables.

While media spend amounts have been utilised for conventional MMMs in decades gone by, they are not necessarily appropriate for MMM models which include digital media.

Generally impressions are a valid variable to describe media contribution while the potential use of reach and frequency data is a topic for ongoing research.

Option for media variables	Description	General evaluation
Spend	<ul style="list-style-type: none"> Use each media's spend as the media variable in the MMM model 	<ul style="list-style-type: none"> Not recommended: spend amount does not describe how many ad impressions are exposed to media users.
Impressions	<ul style="list-style-type: none"> Use each media's number of exposed ad impressions as media variable in the MMM model Coefficients of the models are calculated as incremental KPI (e.g., revenue) per impression Example variants: viewable impressions, GRP (for TV) 	<ul style="list-style-type: none"> Recommended: impression is generally a good choice to describe ads exposure regardless of the path (see ads -> offline purchase, see ads -> click -> purchase etc.). For YouTube, co-viewed impression is a consideration. However, the methodology is in development.
Reach and frequency	<ul style="list-style-type: none"> Use the number of people ads reached and the average frequency of exposure Coefficients of the models are calculated as incremental KPI (e.g., revenue) per reach and frequency 	<ul style="list-style-type: none"> Challenging: while reach and frequency of ads is a valid indicator to describe the ads' effectiveness, calculating reach and frequency across different campaigns may not be possible due to lack of individual level data.
Clicks	<ul style="list-style-type: none"> Use each media's attributed number of clicks (number of clicks after seeing ads) as media variable in the MMM model Coefficients of the models are calculated as incremental KPI (e.g., revenue) per acquired click 	<ul style="list-style-type: none"> Recommended in limited situations: clicks cannot describe the effect of views (i.e., people see an ad but do not click on it and go on to purchase the products). Search campaigns to attract users to a branding website may be an exception to use clicks as a variable.
Views	<ul style="list-style-type: none"> Use each media's number of views (e.g., number of views with length more than 3 seconds), as media variable in the MMM model Coefficients of the models are calculated as incremental KPI (e.g., revenue) per acquired view 	<ul style="list-style-type: none"> Not recommended: views may not be appropriate to describe some situations such as short video formats (e.g., < 3 secs). Also, definition of "view" may be different by media. Thus, collecting apples-to-apples data across media may be difficult.

Example MMM input data (consolidated data)

Illustrative

date	revenue	google_search_clicks	google_search_spend	youtube_bumper_imp	youtube_bumper_spend	youtube_unskippable_imp	youtube_unskippable_spend	num_of_emails
2022-05-28	2,629,157	41,762	187,354	122,578,910	245,158	83,013,377	130,671	115,007
2022-05-29	5,455,074	61,772	70,359	250,073,569	500,147	206,689	58,723	7,884
2022-05-30	2,752,447	1,770,945	124,937	33,634,045	67,267	17,164,564	53,446	79,883
2022-05-31	4,495,993	94,222	46,635	214,849,357	429,698	56,422,041	134,518	44,655
2022-06-01	4,594,333	754,469	298,228	80,448,369	160,896	56,155,803	104,845	68,761
2022-06-02	3,198,586	55,192	9,900	32,198,560	64,397	11,901,778	54,564	93,130
2022-06-03	5,345,409	354,542	59,173	205,844,064	411,687	121,191,369	327,351	88,906
2022-06-04	4,634,572	105,363	511,908	70,692,283	141,384	13,144,918	36,897	55,226
2022-06-05	5,179,344	603,654	13,218	93,235,820	186,471	208,994,326	530,661	42,621
2022-06-06	4,605,407	72,608	164,127	115,014,135	230,028	13,623,608	51,037	30,084
2022-06-07	4,014,675	754,671	6,565	138,992,378	277,984	110,588,001	438,674	10,570
2022-06-08	5,862,088	69,290	245,447	12,007,451	24,015	3,994,038	30,242	57,745
2022-06-09	4,891,005	2,528	50,513	226,324,229	452,648	7,194,686	24,375	126,447

...

Date: Daily data for daily level MMM. Weekly data for weekly level MMM.

KPI: Revenue amount, number of app installs, number of active users etc.

Media:

- Impressions (clicks) and spend amount data by media.
- No impression level data (impression, time stamp) is needed.
- Different breakdowns should be considered. (e.g., product breakdown (bumper/instream skippable/Masthead)). Please see the next page for further detail.

Control variables:

- Product metrics
- Price and promotion
- Competitors' activities
- Popularity
- Macroeconomic data etc.

Media granularity best practice

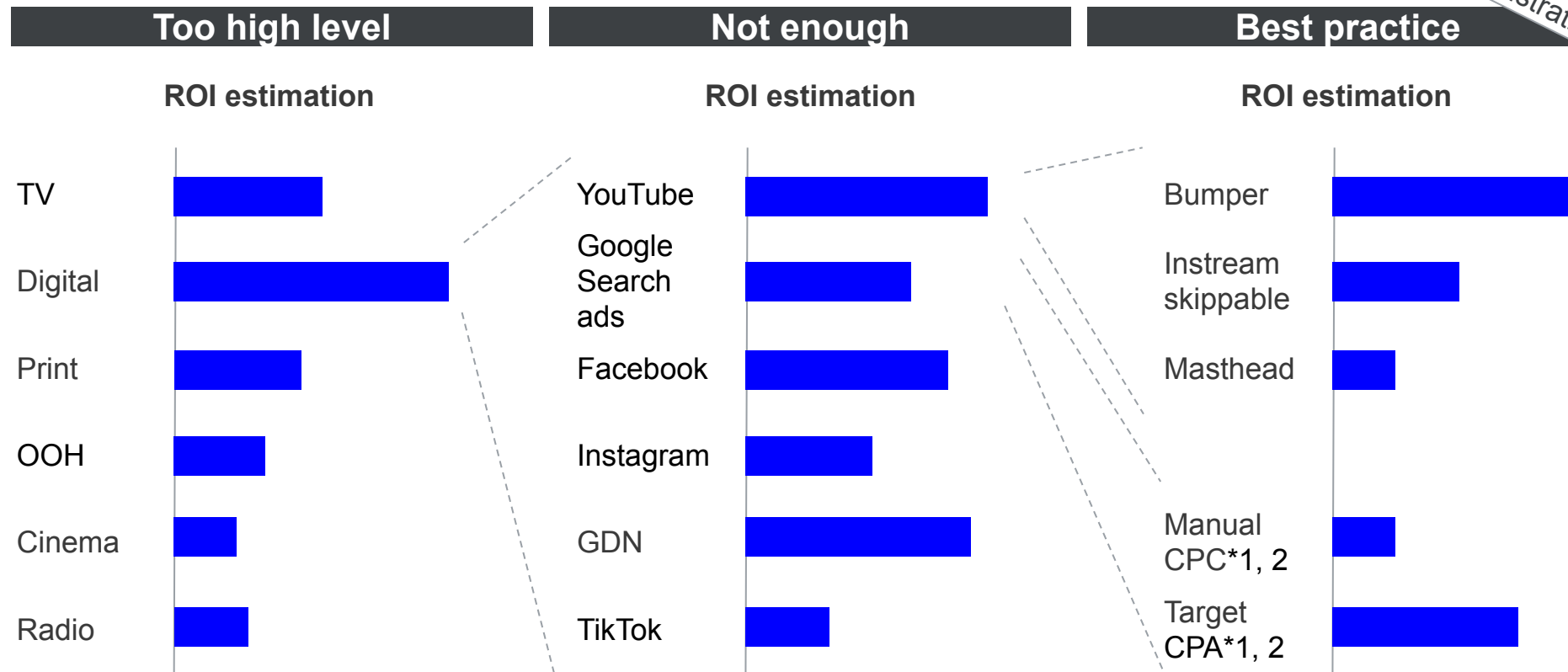
27

To deliver actionable insights, media data in MMM input data needs to be sufficiently granular.

For example, ROI estimation at a high level such as TV and total digital is not enough to provide actionable insights because the model user cannot find out improvement points in TV and digital respectively.

Optimal granularity to deliver actionable insights is placement level (e.g. Bumper, Masthead) or bidding strategy level to analyze performance. However, high result resolution may result in lack of data. Rules of thumb for data volume are described on the next page.

Degree of granularity in MMM output



*1: Automated solution (e.g., Performance Max on Google) can be treated as is because the advertiser cannot control the ads placement.

*2: For performance campaigns (e.g., Google Search, App Campaign), further breakdown may be recommended. For example, word match type (broad match, phrase match and exact match) on Search is worth considering as the consumer behavior depending on the match types may be different. For App Campaign, division by device (iOS, Android, connected TV etc.) may be a consideration as the system efficiency to expose ads is different.

Rules of thumb in total data volume

Number of columns

Illustrative

	date	revenue	google_search_clicks	google_search_spend	youtube_bumper_imp	youtube_bumper_spend	youtube_unskippable_imp	youtube_unskippable_spend	num_of_emails
Number of rows	2022-05-28	2,629,157	41,762	187,354	122,578,910	245,158	83,013,377	130,671	115,007
	2022-05-29	5,455,074	61,772	70,359	250,073,569	500,147	206,689	58,723	7,884
	2022-05-30	2,752,447	1,770,945	124,937	33,634,045	67,267	17,164,564	53,446	79,883
	2022-05-31	4,495,993	94,222	46,635	214,849,357	429,698	56,422,041	134,518	44,655
	2022-06-01	4,594,333	754,469	298,228	80,448,369	160,896	56,155,803	104,845	68,761
	2022-06-02	3,198,586	55,192	9,900	32,198,560	64,397	11,901,778	54,564	93,130
	2022-06-03	5,345,409	354,542	59,173	205,844,064	411,687	121,191,369	327,351	88,906
	2022-06-04	4,634,572	105,363	511,908	70,692,283	141,384	13,144,918	36,897	55,226
	2022-06-05	5,179,344	603,654	13,218	93,235,820	186,471	208,994,326	530,661	42,621
	2022-06-06	4,605,407	72,608	164,127	115,014,135	230,028	13,623,608	51,037	30,084
	2022-06-07	4,014,675	754,671	6,565	138,992,378	277,984	110,588,001	438,674	10,570
	2022-06-08	5,862,088	69,290	245,447	12,007,451	24,015	3,994,038	30,242	57,745

While the necessary data volume for MMM model depends on the situation, number of rows (dates) per parameter in an MMM model is an indicator to sense whether the data volume is enough or not. If a modeler use 20 parameters and 100 days or weeks of data to create an MMM model, number of rows per parameter = $100 / 20 = 5$. More exactly, degrees of freedom is an indicator to consider the sample size. In the above example, the degrees of freedom are $100 - 20 = 80$ and the degrees of freedom per parameter are $80 / 20 = 4$. This effectively means using 4 samples to estimate each parameters. Generally speaking, 4 samples to estimate a parameter may not be enough.

If the data volume is not enough for an MMM model, the modeler needs to reduce number of parameters in the model to ensure enough sample size per parameter, or consider increasing number of days or weeks, or improve the data granularity from national level to geo unit level or sub brand level ([p.56-57](#)).

Rules of thumb in media data volume

29

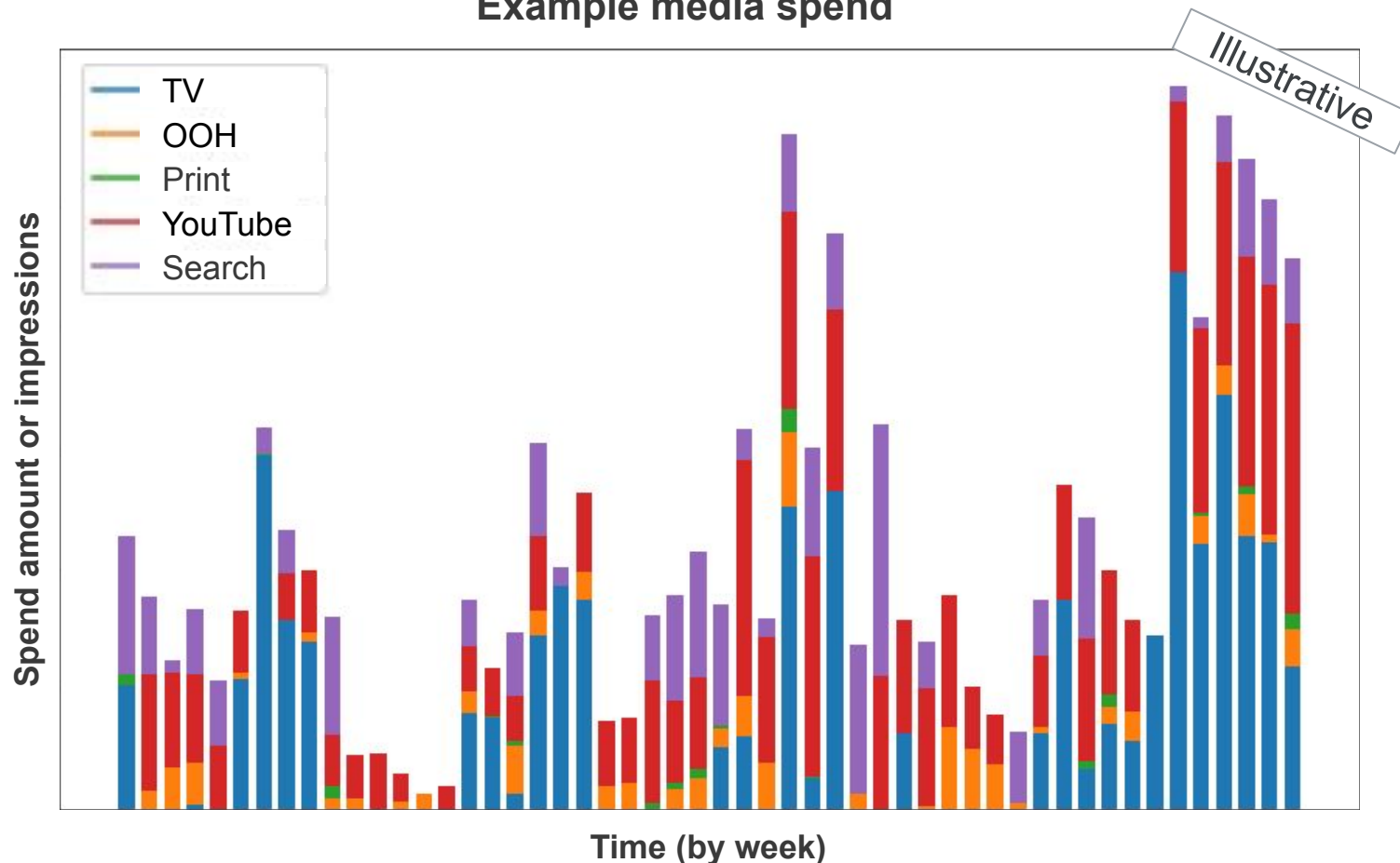
Another consideration in terms of data volume is media data volume.

Flight time skewness: in media investment, some media investments may be skewed to specific seasons such as Christmas. For those media, the number of samples is effectively limited and it may not be easy to estimate model parameters for the media.

Investment volume: total investment volume is also an indicator of media data volume. It is difficult for an MMM model to detect the contribution of media with very small investment. Data consolidation of media with small investment would be recommended (p.39).

Target audience size: KPI (e.g., revenue) increase driven by digital media targeting a narrow audience is also limited and it may not be possible for a MMM model to detect such contribution. Data consolidation of media with only narrow targeting may be needed.

Example media spend



MMM model may not be able to include **OOH (orange)** and **print (green)** in the model due to the small spend size

2.2. Data Cleansing

Missing values in MMMs

After the data selection described in the previous pages, a modeler proceeds with data cleansing.

Firstly, missing values could be a significant problem in MMM. The modeler needs to check whether the missing values exist in the MMM input data and identify potential cause of the missing values.

Generally, missing values need to be imputed before starting the modeling. The next page describes general approach to impute missing values.

Example of missing values

Illustrative

day	Display_cost	Search_imp	Search_cost	organic_imp	Conversions	hc
2021/1/30	149055	549400	35414	2459	28	
2021/1/31	145045	420468	52115	3035	17	
2021/2/1		360250	73056	4275	35	
2021/2/2		613114	66325	4041	35	
2021/2/3		564660	134438	2282	43	
2021/2/4		1003826	103184	3906	80	
2021/2/5		628812	87997	2258	25	
2021/2/6		966091	78677	2318	45	
2021/2/7		670042	68908	4285	50	
2021/2/8		563921	87062	2813	28	
2021/2/9		471092	90951	3179	56	
2021/2/10		690770	81403	2907	34	
2021/2/11		446356	106596	3355	34	
2021/2/12		545286	100453	3640	30	
2021/2/13		344858	98801	3890	36	
2021/2/14		653477	81645	2327	20	
2021/2/15		530004	117828	3722	47	
2021/2/16		787354	135520	2957	56	

Missing values

➔ Missing values are a significant problem in data science because statistical analysis packages **return errors for data that has missing values**

Why do missing values exist? (typical reasons)

<p>Effectively zero</p>	<p>Data does not exist because there were no campaigns or events</p>
<p>Lack of granularity</p>	<p>Data does not exist due to lack of data granularity (e.g., TV has only weekly level data vs. digital which has daily level data)</p>
<p>Systematic or human error</p>	<p>Data does not exist because of systematic or human error (e.g., expiration of the data retention period, system crash etc.)</p>

How to handle missing values

Type	Direction	Typical imputation method*1	Description	Pros	Cons
Effectively zero	Imputation with zero	Simple imputation	Average imputation	Easiest	Underestimation of variance of independent variables by assigning same values or values on a single regression line
Lack of granularity			Normal regression	Easier	
Systematic or human error	Choose imputation method	Regression imputation	Stochastic regression imputation*2	Avoidance of underestimating variance of independent variables	Dependency on the imputation order if there are multiple variables that have missing values
			Multiple imputation*2	On top of the above, relatively long computation time	

Stochastic regression or multiple imputation are realistic options

*1:Takahiro Hoshino. (2016). Statistical science of missing data. *2: Whether those method are successful or not is dependent on whether regression models for imputation are equivalent to or part of the true model.

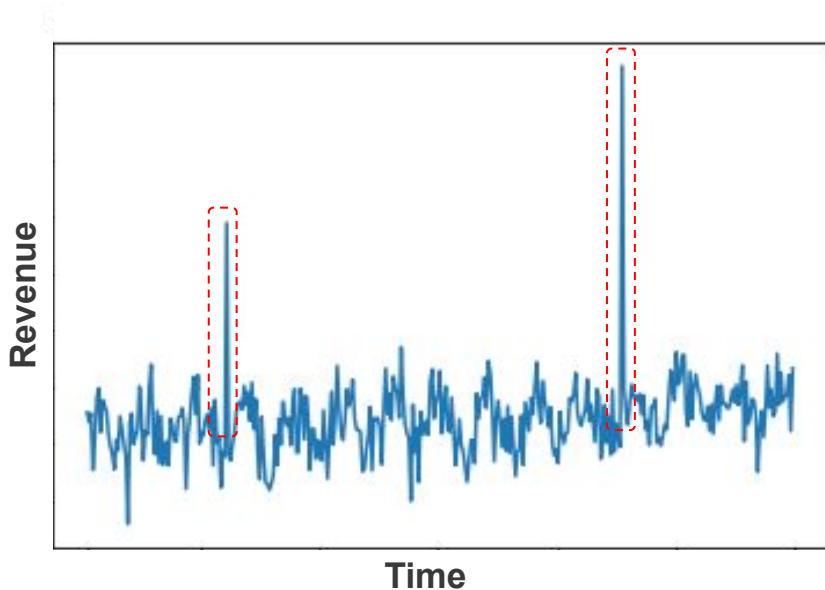
Outliers in MMMs

Secondly, after imputation of missing values, checking outliers is the next step.

While a modeler can choose an appropriate way to detect outliers (e.g., Modified Stahel-Donoho), an important consideration regarding outliers in MMM is identification of reasons for the outliers to consider if other variables should be included in the model.

Typical examples are holiday or event flag which may have an impact on revenue of your business.

Example of outliers in time series data



➔ Outliers of each variable in an MMM should be recognised upfront by MMM modelers and the stakeholders for the contextual check of models in “output check” section.

Outliers check before modeling

Check data input accuracy	If the data set is manually created, there may be errors in the data input process. If so, it should be revised
Check scheduled event existence	Whether primary scheduled events (e.g., marketing events, new brand launches, holidays) are included in the data or not
Check sudden event existence	Whether sudden events (e.g., Covid-19 infection jump, disaster, competitors' actions) are included in the data or not

} If variables to describe scheduled and sudden event are not included in the model, additional variables should be added to the input data.*

*: Increasing number of parameters may affect quality of models.

Data form change

34

Typical data form change in MMMs

Categorical variable -> binary variables

Categorical variable should be transformed to binary variables

day	Event_type	day	Event_1	Event_2
2021/2/1	0	2021/2/1	0	0
2021/2/2	0	2021/2/2	0	0
2021/2/3	0	2021/2/3	0	0
2021/2/4	0	2021/2/4	0	0
2021/2/5	0	2021/2/5	0	0
2021/2/6	1	2021/2/6	1	0
2021/2/7	1	2021/2/7	1	0
2021/2/8	0	2021/2/8	0	0
2021/2/9	0	2021/2/9	0	0
2021/2/10	0	2021/2/10	0	0
2021/2/11	0	2021/2/11	0	0
2021/2/12	0	2021/2/12	0	0
2021/2/13	2	2021/2/13	0	1
2021/2/14	2	2021/2/14	0	1

Typically some control variables such as marketing events, holiday, weather etc. may be categorical variables. Those should be converted into binary variables.

Sometimes categorical variables may be ordinal (e.g., popularity ranking) and should not be changed to binary variables.

Character variable -> binary/numerical variables

Character variable should be transformed to numerical and/or binary variables

day	Promotion	day	Discount_pct	Promo_grocery	Promo_fashion
2021/2/1	10% dicsount, grocery	2021/2/1	10	1	0
2021/2/2	10% dicsount, grocery	2021/2/2	10	1	0
2021/2/3	10% dicsount, grocery	2021/2/3	10	1	0
2021/2/4	10% dicsount, grocery	2021/2/4	10	1	0
2021/2/5	10% dicsount, grocery	2021/2/5	10	1	0
2021/2/6	10% dicsount, grocery	2021/2/6	10	1	0
2021/2/7	10% dicsount, grocery	2021/2/7	10	1	0
2021/2/8	5% discount, fashion	2021/2/8	5	0	1
2021/2/9	5% discount, fashion	2021/2/9	5	0	1
2021/2/10	5% discount, fashion	2021/2/10	5	0	1
2021/2/11	5% discount, fashion	2021/2/11	5	0	1
2021/2/12	5% discount, fashion	2021/2/12	5	0	1
2021/2/13	5% discount, fashion	2021/2/13	5	0	1
2021/2/14	5% discount, fashion	2021/2/14	5	0	1

Typically promotion meta data may be a character variable (text format). The data should be divided into discount amount, item category, form of campaign (e.g., absolute discount, percentage discount, buy 1 get 1 free etc.) and frequency. Also, promotion part should be modeled appropriately by promotion type, category etc.

After checking data volume, missing values and outliers, data form change may be needed for categorical or character variables.

Typically some control variables may be categorical or character form which cannot be handled in an MMM model.

Those should be converted into numerical or binary variables to quantify the impact on the KPI in the MMM model.

Introduction of multicollinearity

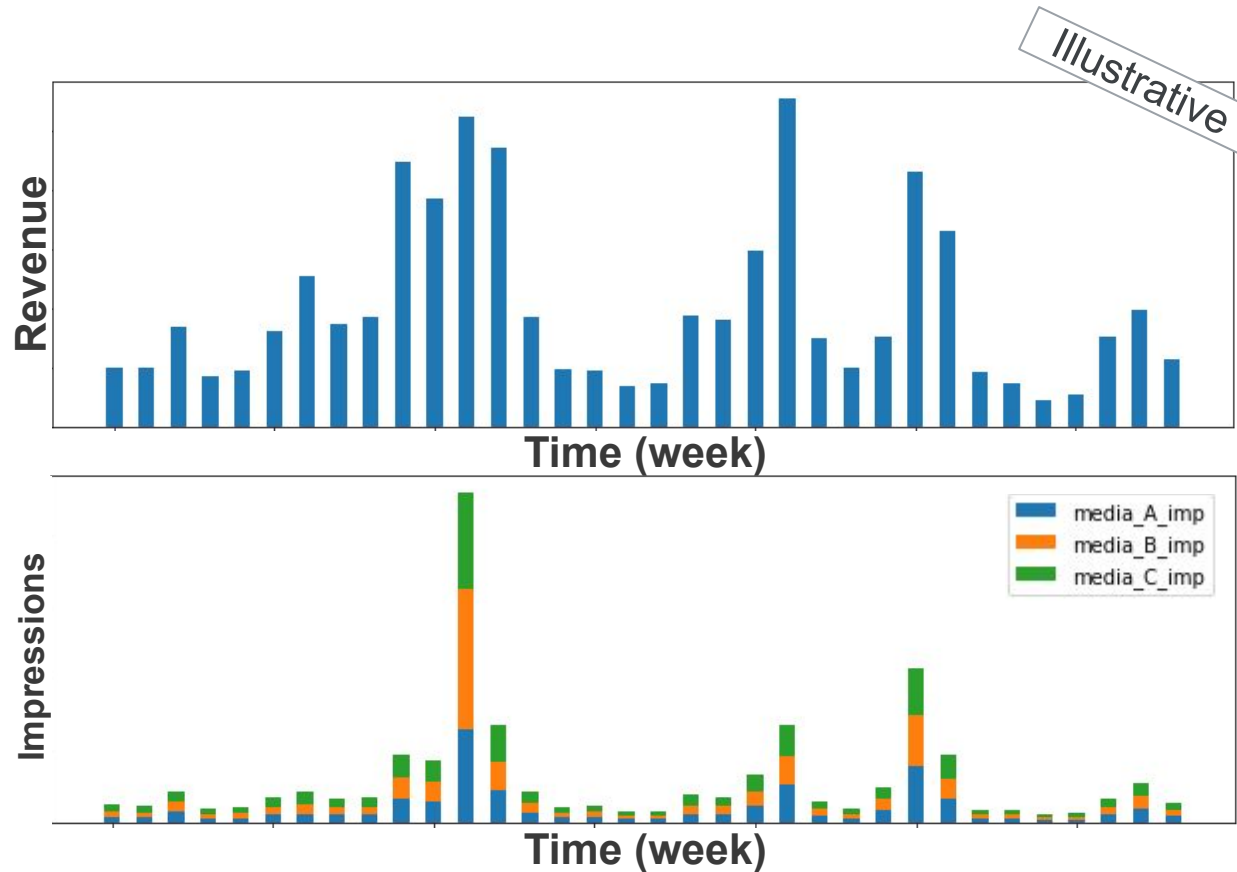
35

Another potential pitfall in MMMs is multicollinearity among independent variables which may lead to in errors in the estimation.

In the example on the right, the regression model has impressions of three media respectively as independent variables.

However, estimation result in coefficient of a media which is confusing. Though generally media investment (advertising) has a positive impact on revenue, the coefficient of media B is negative. The model suggests that more investment on media B leads to less revenue. The model is inconsistent with our understanding of media investment.

A simplified example*1: $Revenue = a_1 * imp_{mediaA} + a_2 * imp_{mediaB} + a_3 * imp_{mediaC} + b$



Regression Result

$$Revenue = 0.16 * imp_{mediaA} \\ - 1.02 * imp_{mediaB} \\ + 1.70 * imp_{mediaC} + 0.164$$

- Coefficient of media B impressions is **negative**
- In other words, the more investment in media B, the less revenue

Is this true?

*1: This model is too simple to describe actual situations and just to explain multicollinearity example.

Mechanism of multicollinearity

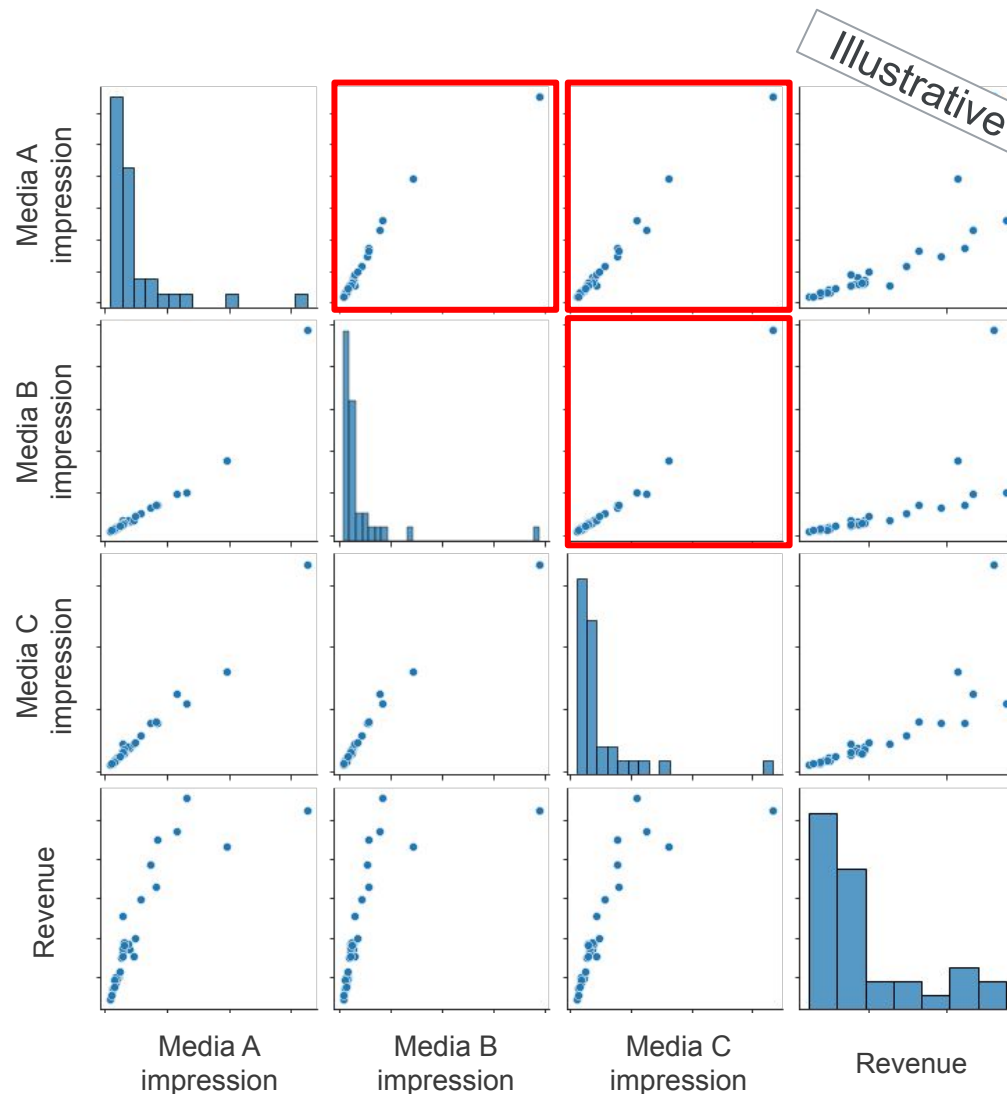
36

The right chart shows scatter plots between media variables (media A, B and C) to understand the mechanism of multicollinearity.

There are strong correlations (e.g., correlation coefficient > 0.9) between media A and B, A and C and B and C respectively.

The regression model cannot detect contribution of media A, B and C separately. The coefficients are determined inappropriately.

In the next page, one of the indicators to detect multicollinearity is introduced.



Issue

There are strong correlations (e.g., correlation coefficient > 0.9) between weekly impressions of media A, B and C

In other words, when spend on media A, B or C increases, spend on the other media increases

As a result, the model cannot identify which media contributed to the revenue increase

Negative coefficient of media B implies mathematical instability of the model

➔ Generally if there is strong correlation (e.g., correlation coefficient > 0.9) among media in weekly/daily investment, the model is not accurate

How to detect multicollinearity

37

To detect multicollinearity, VIF (Variance Inflation Factor) can be utilised. VIF is calculated from a multiple correlation coefficient between one independent variable and the other independent variables.

Generally VIF = 10* is a threshold equivalent to 0.9 or -0.9 in the multiple correlation coefficients.

How to handle the multicollinearity is described in the next page. While some methods (e.g., ridge regression) can avoid overfitting problems due to multicollinearity, data structure change is typically needed for inference problems such as MMM.

Example regression:

$$\text{Revenue} = a_1 * \text{imp}_{\text{mediaA}} + a_2 * \text{imp}_{\text{mediaB}} + a_3 * \text{imp}_{\text{mediaC}} + b$$

Regression among variables	Multiple correlation coefficient	VIF (Variance Inflation Factor)	VIF value
$\text{imp}_{\text{mediaA}} = a_{11} * \text{imp}_{\text{mediaB}} + a_{12} * \text{imp}_{\text{mediaC}} + b_1$	R_A	$\text{VIF}_A = \frac{1}{1 - R_A^2}$	97.3
$\text{imp}_{\text{mediaB}} = a_{21} * \text{imp}_{\text{mediaA}} + a_{22} * \text{imp}_{\text{mediaC}} + b_2$	R_B	$\text{VIF}_B = \frac{1}{1 - R_B^2}$	20.5
$\text{imp}_{\text{mediaC}} = a_{31} * \text{imp}_{\text{mediaA}} + a_{32} * \text{imp}_{\text{mediaB}} + b_3$	R_C	$\text{VIF}_C = \frac{1}{1 - R_C^2}$	131.4



Generally variables which have VIF > 10 need to be “merged”, “removed” or “divided”* (please see the next page)

*: O'Brien, Robert. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity. 41. 673-690. 10.1007/s11135-006-9018-6

How to fix multicollinearity

Option	Description	How it works	Pros	Cons
Merge	Combine variables which have multicollinearity		<ul style="list-style-type: none"> Less data loss compared to “remove” 	<ul style="list-style-type: none"> Some variables difficult to combine (e.g., temperature + media spend) Less data granularity
Remove	Remove variables which have multicollinearity		<ul style="list-style-type: none"> Straightforward 	<ul style="list-style-type: none"> Data loss Exclusion of some variables Misattribution of KPI to remaining variables
Divide	Divide the model by segment (e.g., audience, region etc.)		<ul style="list-style-type: none"> No data loss 	<ul style="list-style-type: none"> Model complexity Data availability at segment level (e.g., TV spend by audience segment is not available) <p>While “divide” is the best option, the feasibility depends on data availability. In some categories (e.g., tech), it may not be possible. Intervention may be also possible depending on the situation (p.127).</p>

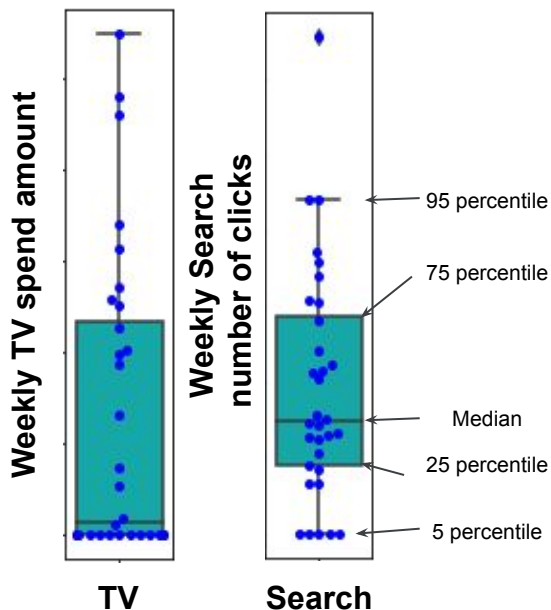
Data Scaling (normalization)

The next step in data cleansing is data scaling.

Data scaling (or data normalisation) is necessary to account for multiple variables which have different scales. To keep variables positive in MMM following normalisation, mean scaling is an effective option because almost variables in MMMs take positive values (e.g., media impressions should be positive value).

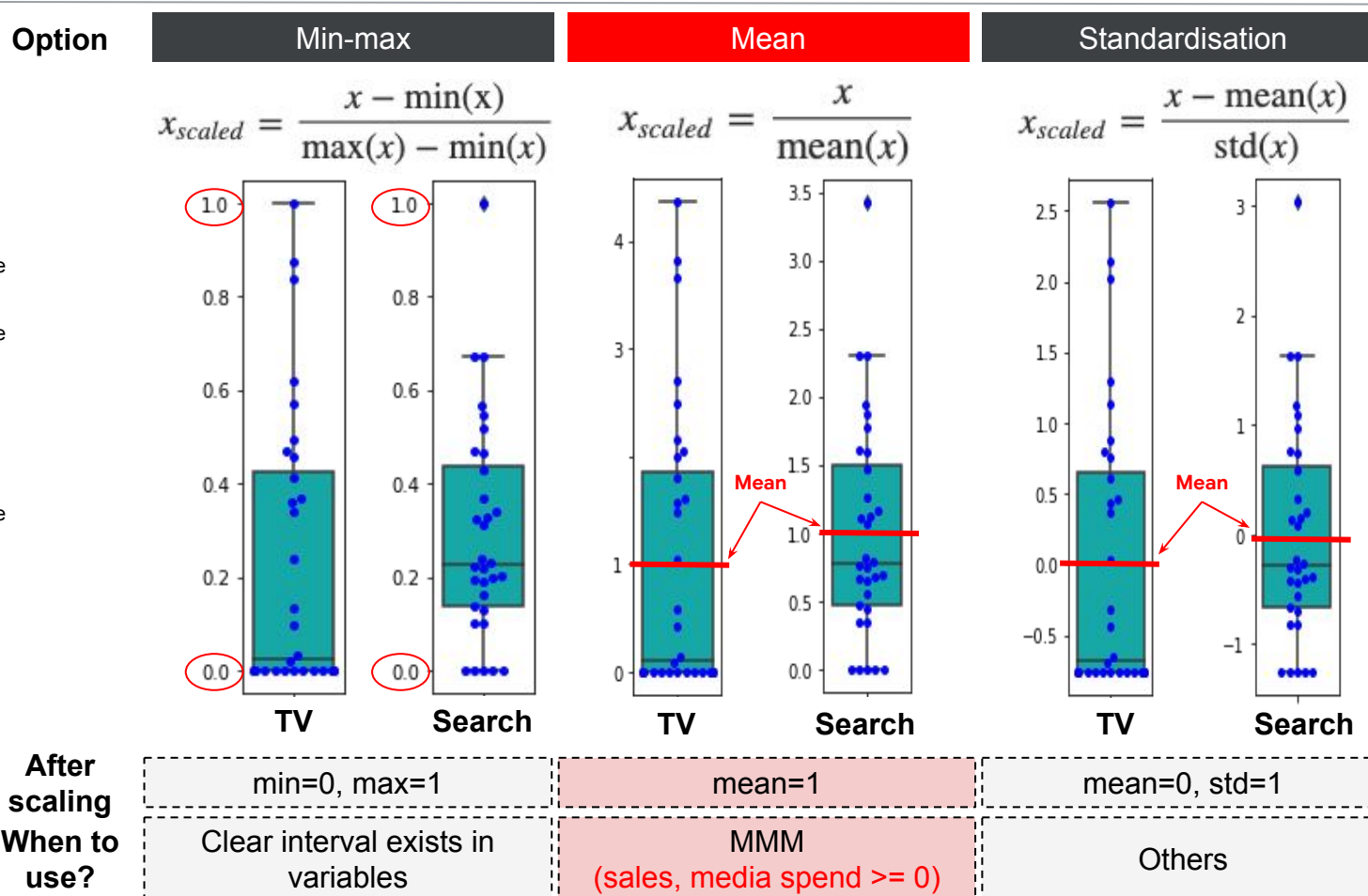
Further data transformations (e.g, logarithm, adstock) are discussed in the next section.

Why is data scaling needed?



- Differently scaled variables are not comparable after MMM model creation
- Variables with large units may adversely affect the estimation of model parameters

Typical data scaling (normalisation) options



2.3. Creating Model Structure

This section focuses on basic model structures may not be suitable for actual situations. Please read chapter 3 for the considerations.

Summary of a MMM model structure

In this section, the primary elements of a basic MMM are explained.

While an MMM is a regression model, there are elements specific to this application. The model consists of A) response curve transformation, B) adstock transformation, and C) trend and seasonality and other variables.

Complex structures such as multi-layer models are described in CHAPTER 3. This section focuses on unique elements of a MMM model compared to linear regression models.

An example of an additive MMM model*1

$$\text{Revenue}_t = b + \sum_m \beta_m * \text{Hill} \left(\text{Adstock}(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m \right) + \text{trend}_t + \text{seas}_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

Intercept (bias) points to b .
Do the transformations for all media then sum points to the summation \sum_m .
(A) Response curve transformation points to the Hill function.
(B) Adstock transformation points to the Adstock function.
(C) Trend model points to trend_t .
(C) Seasonality model points to seas_t .
Other variables (control variables) model*2 points to $\sum_c \gamma_c d_{t,c}$.
Random noise points to ϵ_t .

*1: This model may be too simple to describe actual situations and is to explain basic features in a MMM model.

*2: While the detail is not described in this guidebook, other variables should be modeled appropriately. For example, revenue response to price discount is different depending on the discount percentage, item and timing and the response may not be linear.

Additive or multiplicative model

Before a deep dive into an explanation of MMM elements, a modeler needs to understand the difference between an additive and multiplicative structure.

An additive model separates the effect of each media and other variables. On the other hand, a multiplicative model treats the effects dependently other because of the multiplicative structure.

While a multiplicative model can be converted to an additive model by taking the logarithm on both side of the equation, interpretation of the model result is different. The next two pages explain the difference.

Additive model

Multiplicative model

Comparison between an additive model and multiplicative model

(Response curve and adstock are disregarded for the sake of simplicity*1)

$$Revenue = \underbrace{b}_{\text{Bias (intercept)}} + \underbrace{w_{TV} * x_{TV} + w_{SEM} * x_{SEM} + \dots}_{\text{Effect of media investment}} + \underbrace{w_{prom} * x_{prom} + \dots}_{\text{Effect of other variables}}$$

$$Revenue = \underbrace{b}_{\text{Bias (intercept)}} * \underbrace{x_{TV}^{w_{TV}} * x_{SEM}^{w_{SEM}} * \dots}_{\text{Effect of media investment}} * \underbrace{x_{prom}^{w_{prom}} * \dots}_{\text{Effect of other variables}}$$

$$\Leftrightarrow \log Revenue = \log b + w_{TV} \log x_{TV} + w_{SEM} \log x_{SEM} + \dots + w_{prom} \log x_{prom} + \dots$$

➡ Multiplicative model has a same structure with additive model except for “log” of revenue and variables.

*1: These models may be too simple to describe actual situations and is to explain basic features in a MMM model.

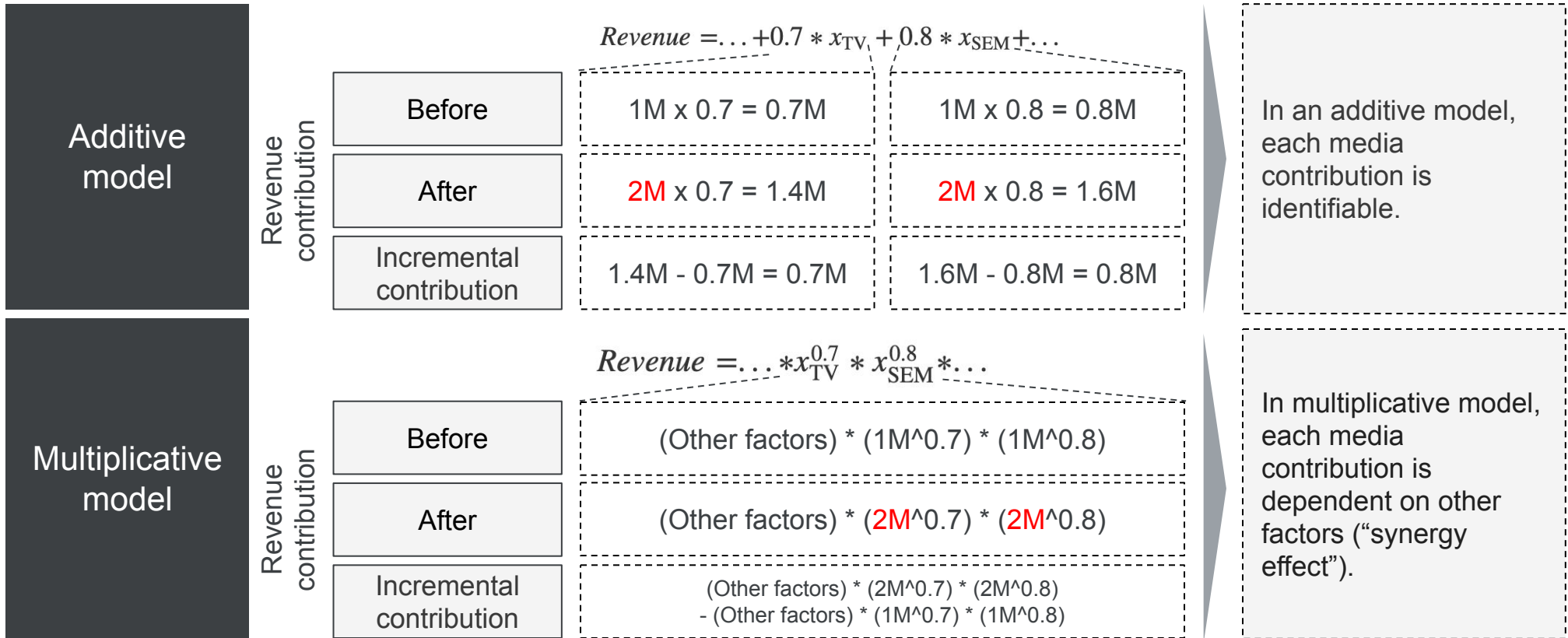
Difference between an additive and multiplicative model

The example on the right shows an interpretation difference between an additive and multiplicative model.

When advertising impressions on media increase from 1 million to 2 million, the incremental revenue can be calculated by media in an additive model.

On the other hand, the effect estimation of each media is inseparable in a multiplicative model. Dependency among media is described in the model.

Example scenario: double TV budget and SEM (Search Engine Marketing) budget respectively (from 1mn to 2mn impressions)*1,2



*1: These models may be too simple to describe actual situations and is to explain basic features in a MMM model. The detail is discussed in CHAPTER 3. *2: In digital advertising, for example, when impressions are doubled, CPM increases and the budget may more than double. Although a model should be created that takes into account changes in CPM according to budget, CPM is usually assumed to be constant due to a combination of the number of samples and the number of parameters.

(A) Overview of response curve transformation

44

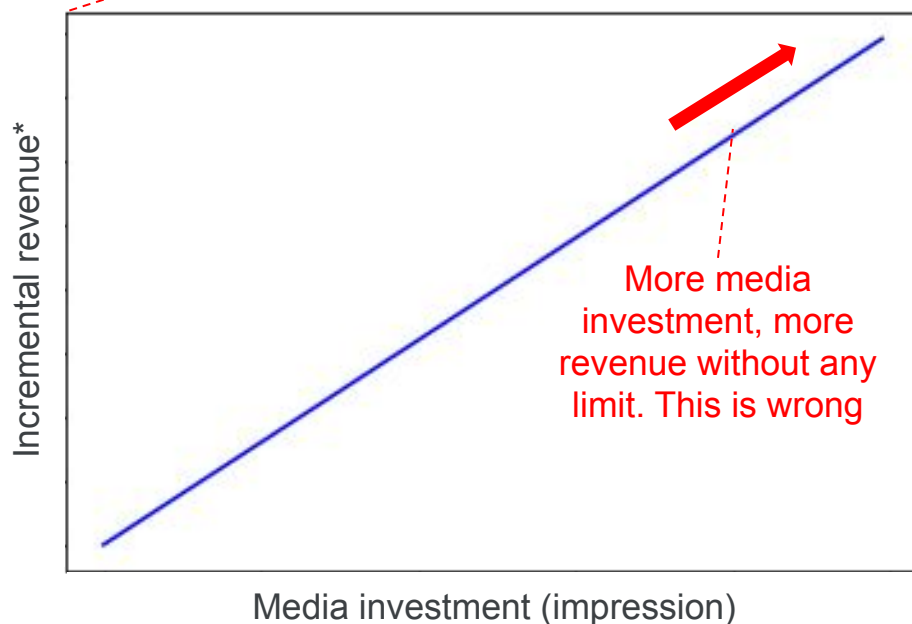
Response curve transformation is one of the important elements in MMM to describe the saturation of media investment.

The linear structure in an ordinary linear regression model assumes infinite revenue growth is possible by increasing media investment. However, in reality the potential effect of media investment is limited due to limits of the number of users, number of impressions and frequency of the ads on a media. To describe this situation, response curve transformation is needed as shown in the next page.

Additive
model*

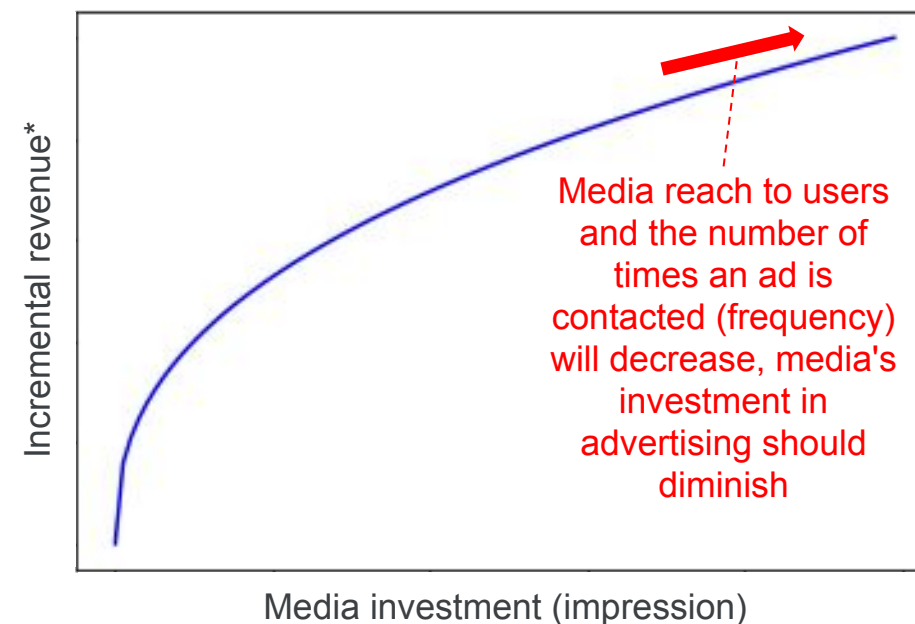
$$Revenue = b + w_{TV} * x_{TV} + w_{SEM} * x_{SEM} + \dots + w_{prom} * x_{prom} + \dots$$

Linear model



Adjustment is needed

Response curve model



*For multiplicative model, while incremental revenue of a media depends on other factors, the diminishing return effect is applied for each media response curve.

Typical response curves

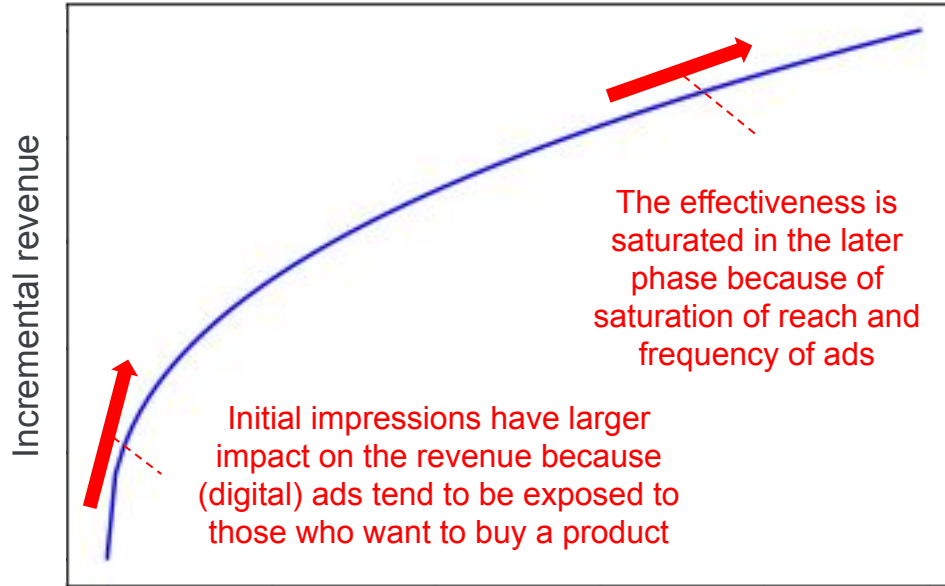
The typical response curve shape is a concave curve or S-shape curve.

For a concave curve, the exponential function is used to describe the saturation of the media investment.

For an S-shaped curve, a hill function is used to show a slow rise in the initial phase of investment.

The next page shows that a hill function has flexible features to model different shapes.

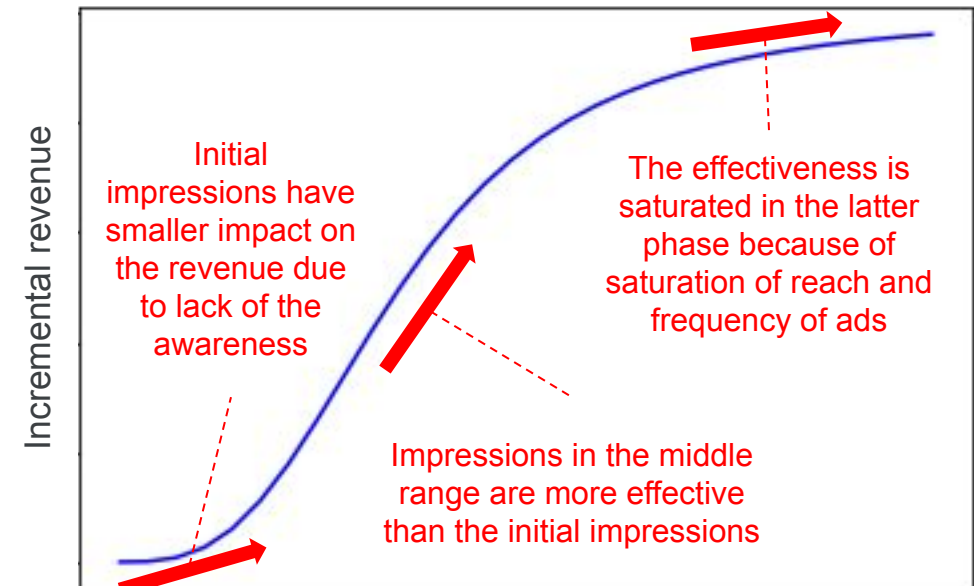
Concave curve



Media investment (impression)

Example: $y = \beta_{\text{media}} * x_{\text{media}}^{r_{\text{media}}}$
 $(\beta_{\text{media}} > 0, r_{\text{media}} < 1)$

S-Shaped curve



Media investment (impression)

Example: $y = \beta_{\text{media}} * \frac{1}{1 + \left(\frac{x_{\text{media}}}{K_{\text{media}}}\right)^{-S_{\text{media}}}}$ **Hill function**
 Scale $(\beta_{\text{media}}, K_{\text{media}}, S_{\text{media}} > 0)$

Characteristics of a hill function

A typical hill function curve has two parameters.

Half saturation: this parameter determines the point of the “half saturation”.

Slope: this parameter determines the pitch of the curve.

By using the above two parameters, primary features can be described.

Convergence: the example curve converges to 1 when x is infinite.

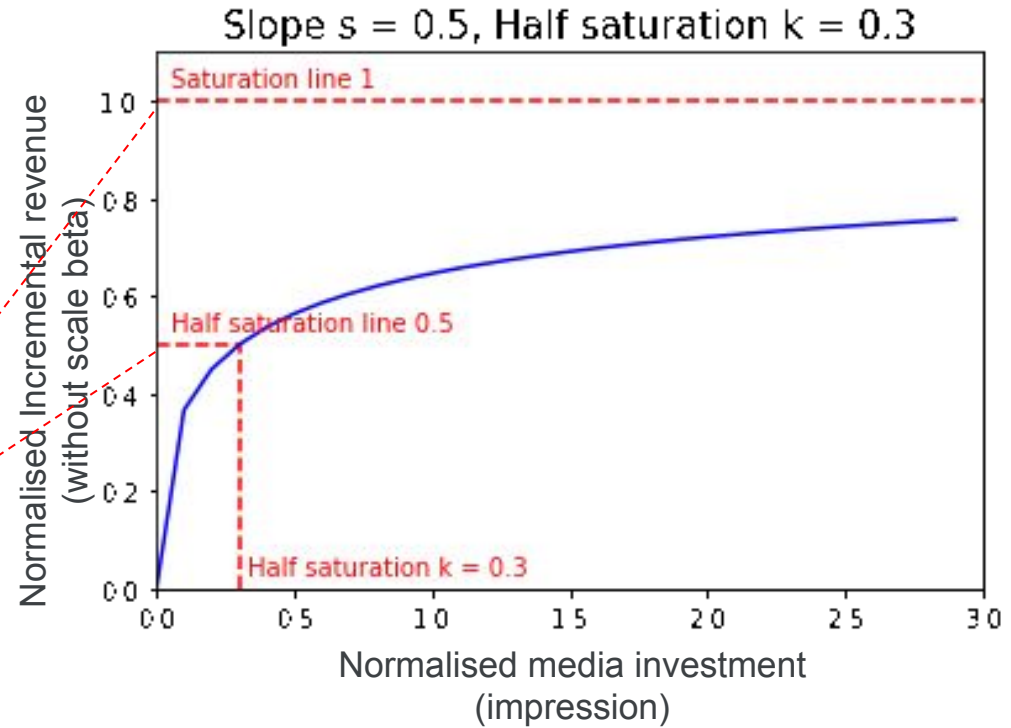
Half saturation: when the media investment x is equal to the half saturation parameter, the function = 1/2.

$$\text{Hill}(x_{\text{media}}; K_{\text{media}}, S_{\text{media}}) = \frac{1}{1 + \left(\frac{x_{\text{media}}}{K_{\text{media}}}\right)^{-S_{\text{media}}}}$$

(x_{media}, β_{media}, K_{media}, S_{media} > 0)

$$\lim_{x_{\text{media}} \rightarrow \infty} \text{Hill}(x_{\text{media}}; K_{\text{media}}, S_{\text{media}}) = 1$$

$$\text{Hill}(x_{\text{media}} = K_{\text{media}}; K_{\text{media}}, S_{\text{media}}) = \frac{1}{1 + 1^{-S_{\text{media}}}} = \frac{1}{2}$$



➡ Higher slope (s), steeper slope
Higher half saturation (k), slower saturation*

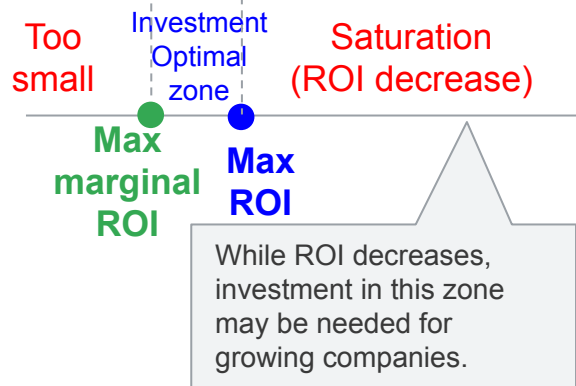
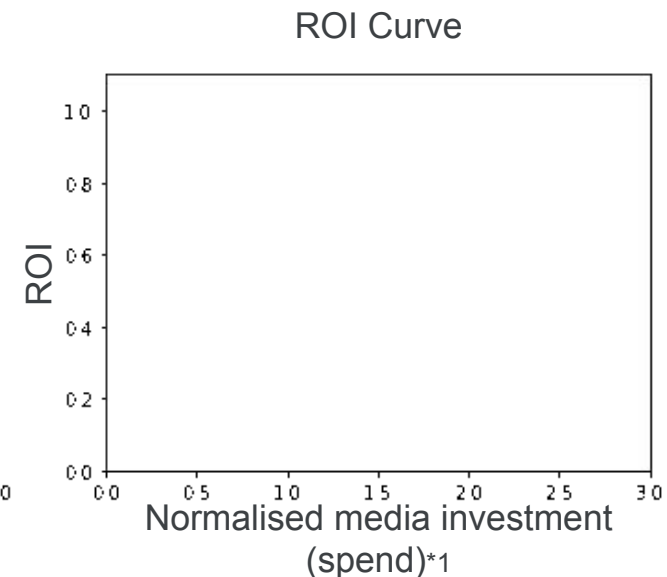
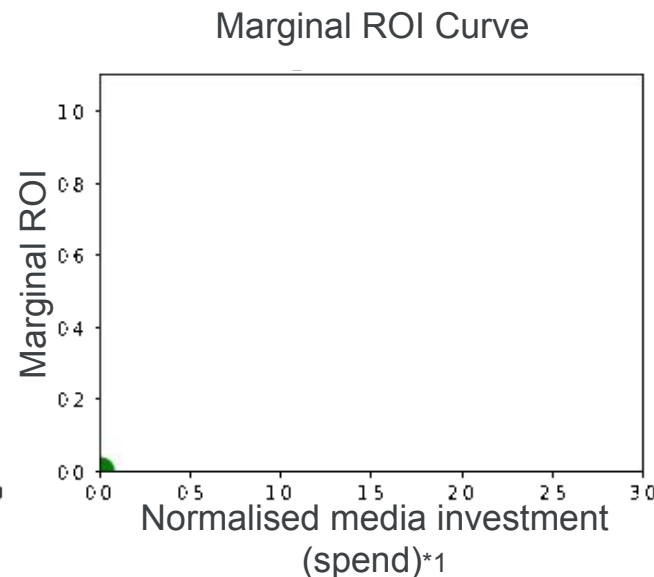
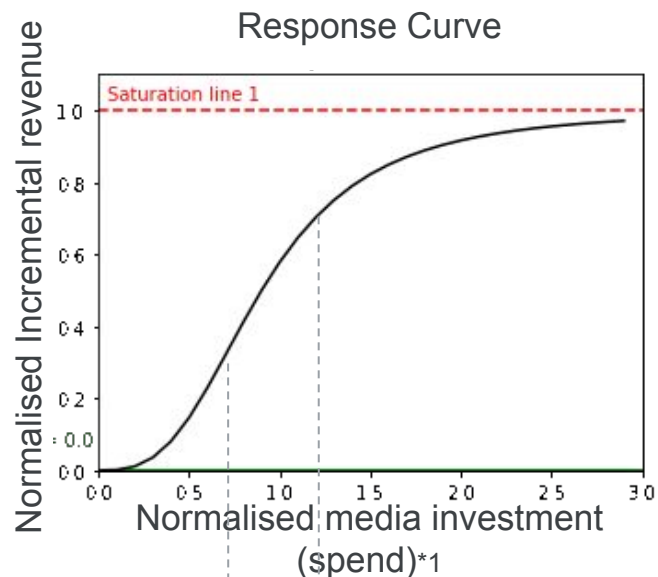
*: In practice, the modeler may introduce constraints on the parameters K and S to avoid extremely steep response curves. This is because if S is too large, the response curve may become too steep and the algorithm for optimization ([p.80](#)) may not work well.

Definition of key metrics on the curve

Based on a response curve, key metrics on the curve can be defined. As the horizontal axis is media investment (or impressions) and the vertical axis is incremental revenue driven by the media investment, two Return On Investment (ROI) related metrics are defined.

Marginal ROI (ROAS): which effectively means “revenue increase from the next impression (or one dollar)”. Mathematically it is the gradient of the tangents of the response curve.

ROI (ROAS): which is the ratio of incremental revenue to media investment.



Marginal ROI

The gradient of the tangent of the response curve. It can be thought of as the “speed of revenue increase by further spend on a media”

ROI (ROAS)

Gradient of a line that takes the origin and a point on the response curve

$$\frac{\text{Incremental revenue}}{\text{media investment (spend)}}$$

*1: To calculate ROI, modelers need to transform input variable from impression to spend by using average cost per impression after estimating response curves.

(B) Overview of adstock

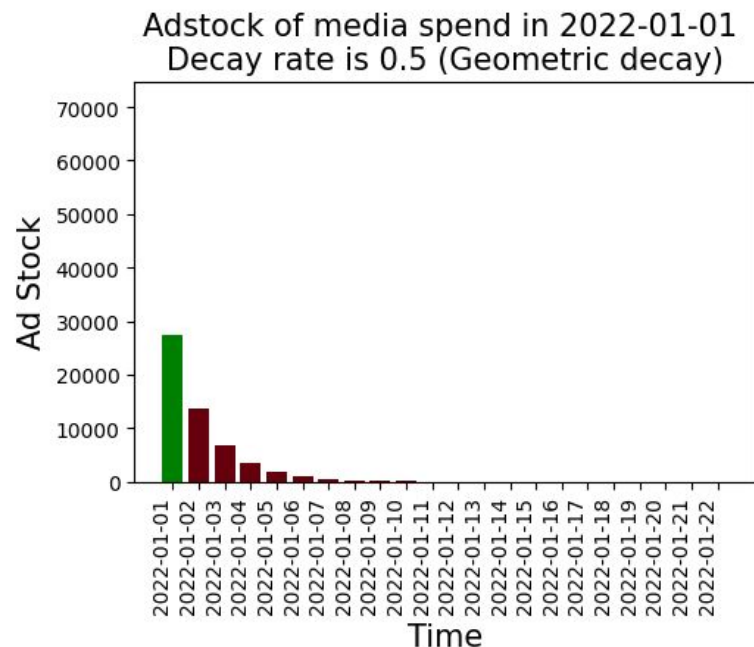
48

Another key aspect of MMMs is the concept of adstock. Adstock is a variable transformation to describe decay effects in media investment.

For example, if YouTube ads are exposed to users today, the effect of the ads may continue for a couple of weeks, depending on the ads' format and creatives. While generally video format ads may have a long effect compared to static images (high decay rate), the placement of the ads may also impact on the duration.

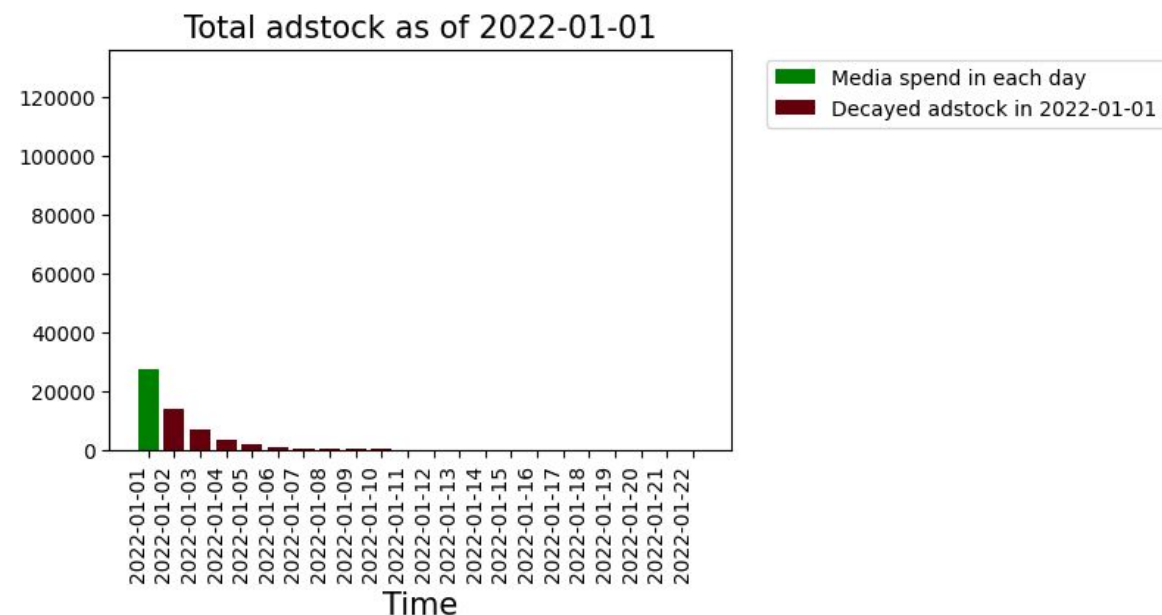
Based on the above assumptions, media variables are transformed into adstock variables in MMM models.

Attenuation effect on media impressions for each week
(Green = media impressions for each week, red = residual effect)



MMMs assume that each media spend has 'decayed effect' which drives (decreasing) impact over time

Sum of the effects listed on the left
(adstock)



Adstock is the accumulated 'decayed effect'. Instead of media spend itself, MMM models use the adstock as an input variable

Typical adstock curves 1/3

Geometric adstock

Example formula

$$\text{adstock}_{t,media} = x_{t,media} + \lambda_{media} * \text{adstock}_{t-1,media}$$

Media impression

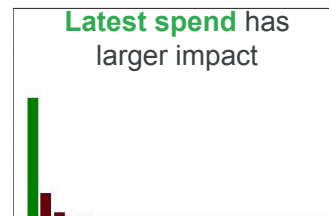
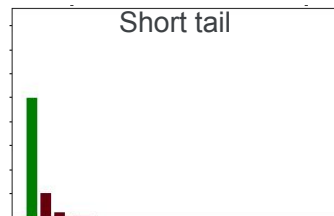
Decay parameter

Media spend and the adstock in each time

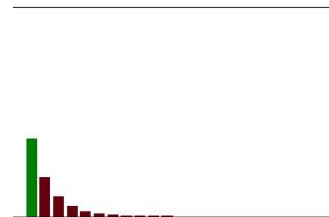
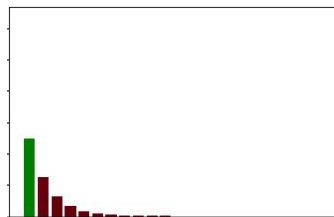
Adstock

$\lambda = 0.2$

Media impact is based on mainly each time's spend (impression)

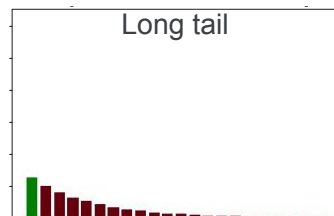


$\lambda = 0.5$



$\lambda = 0.8$

Media impact is based on mainly accumulated effect



Examples

High decay rate

Medium decay rate

Low decay rate

Typical adstock curves 2/3

Weibull PDF adstock*

L: Length of decay parameter, l: lag between ads' exposure and the time period t

$$\text{adstock}_{t,media} = \sum_{l=0}^L \exp\left(-\frac{l}{\lambda}\right)^k * x_{t-l,media}$$

k: scale parameter
("decay speed parameter")

$$\lambda = \frac{l}{(-\ln(0.001))^{\frac{1}{k}}}$$

Lambda: shape parameter x: Media impression

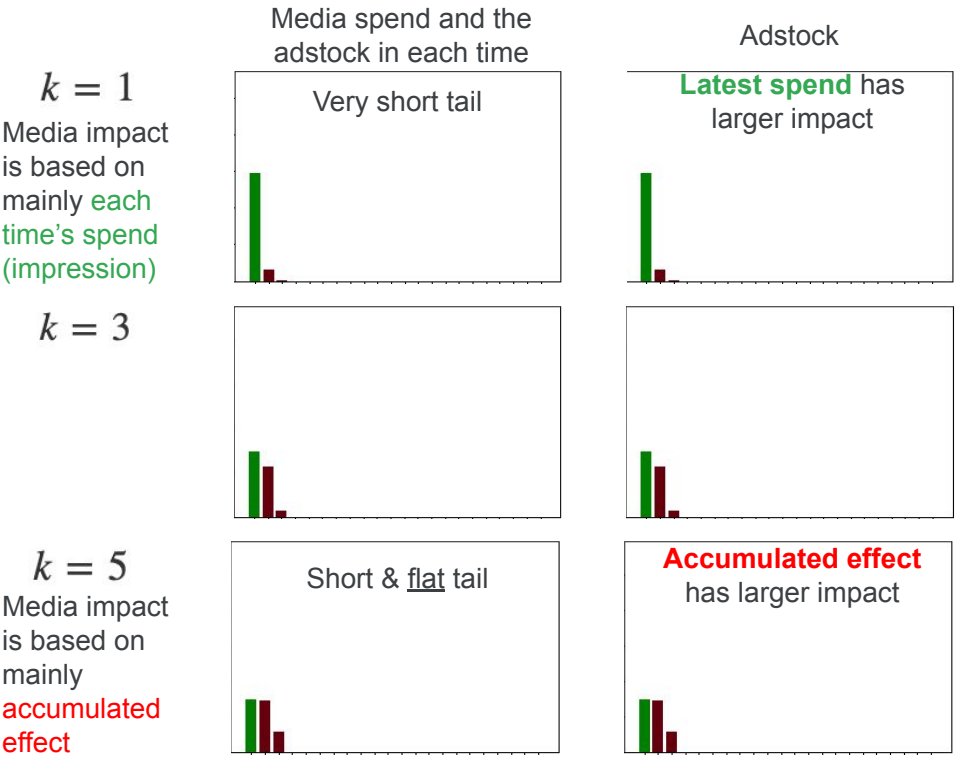
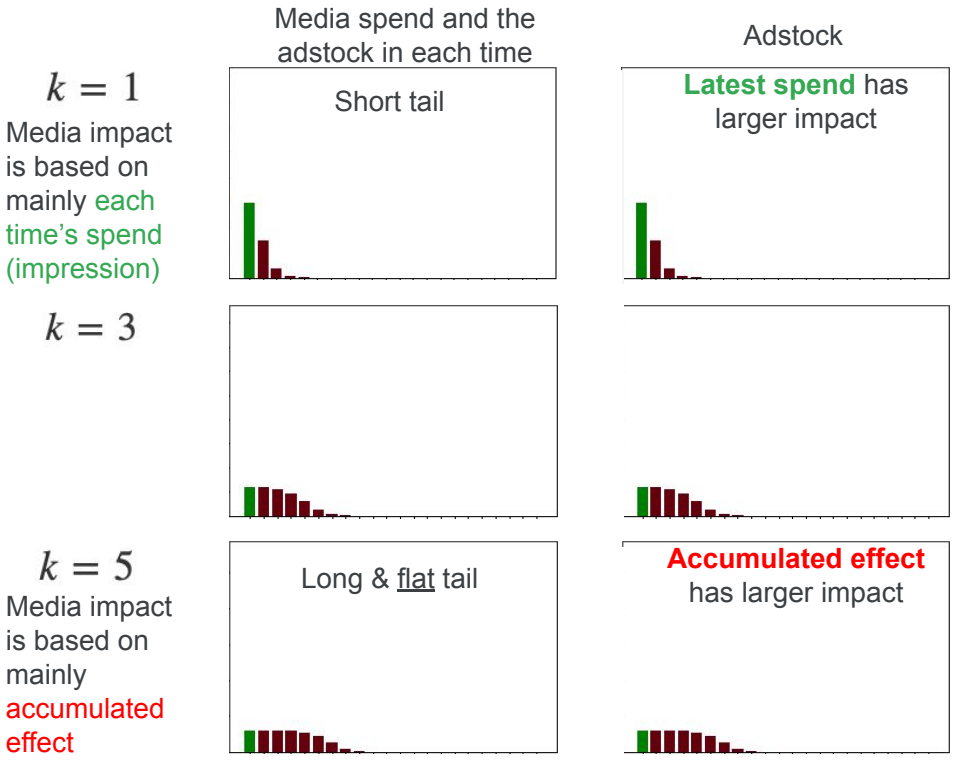
Example*

Large decay length (L=10)

Small decay length (L=3)

Examples

High decay rate
Medium decay rate
Low decay rate



*: <https://github.com/annalectnl/weibull-adstock>. PDF: Probability Distribution Function

Typical adstock curves 3/3

Carry-over effect*1

Example*

w: Weight parameter x: Media impression Theta: Decay delay parameter

$$adstock_{t,media} = \frac{\sum_{l=0}^{L-1} w_{media}(l) * x_{t-l,media}}{\sum_{l=0}^{L-1} w_{media}(l)}$$

L: Length of decay parameter, l: lag between ads' exposure and the time period t

$$w_{media}(l; \alpha_m, \theta_m) = \alpha_m^{(l-\theta_m)^2}$$

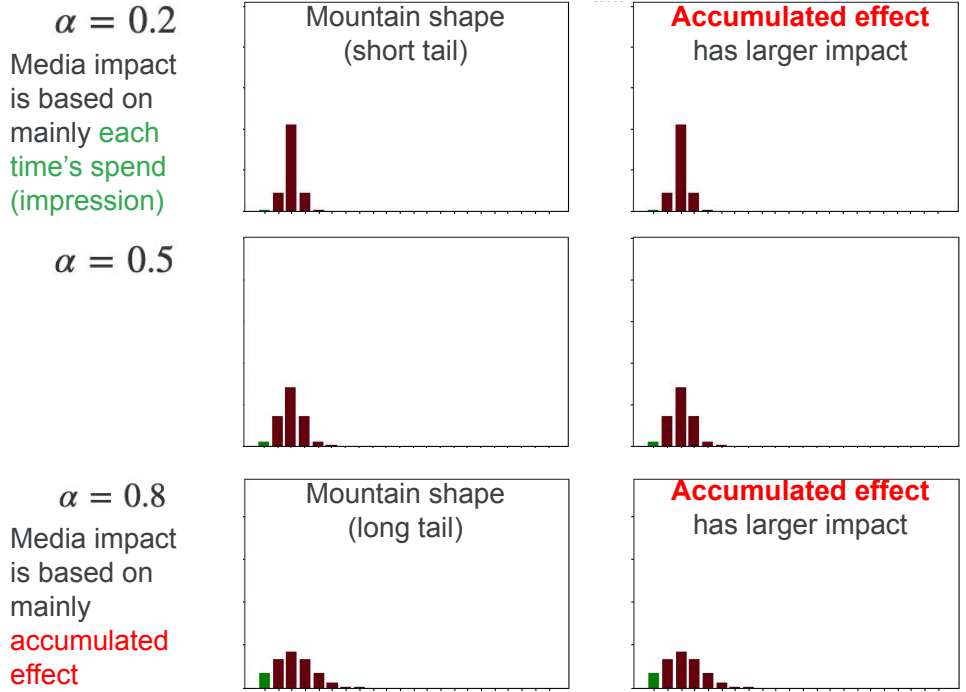
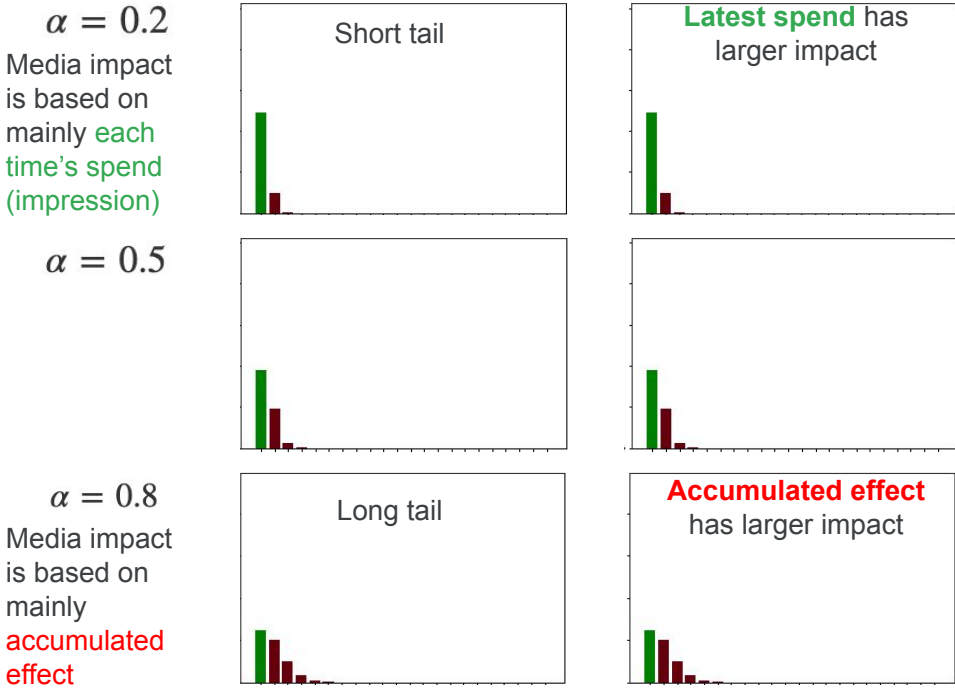
Alpha: Decay rate parameter

Without decay delay (theta = 0)

With decay delay (theta = 2)*2

Examples

High decay rate
Medium decay rate
Low decay rate



*2: the above pattern may happen if an advertiser runs campaigns targeting future sales (e.g., campaign in early December to drive Xmas sales)

*1: Google (2017), Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects

Data transformation order

The two transformations (response curve and adstock transformations) were introduced in the previous pages.

A typical question on the transformations is the order. There are two options: 1) response curve then adstock transformation and 2) adstock then response curve transformation.

Yuxue et al.* gives a rule of thumb in the choice as described on the right.

Response curve then adstock

Adstock then response curve

Example (and when to use)

$$\mathbf{x}_{\text{media,Hill}} = \begin{pmatrix} \text{Hill}(x_{t,\text{media}}; K_{\text{media}}, S_{\text{media}}) \\ \vdots \\ \text{Hill}(x_{t-l,\text{media}}; K_{\text{media}}, S_{\text{media}}) \end{pmatrix}$$

Firstly, transform media investment (impressions) to incremental effect of the media by the response curve

$$y = \beta_{\text{media}} * \text{Adstock}(\mathbf{x}_{\text{media,Hill}}, L, w_{\text{media}}(l; \alpha_{\text{media}}, \theta_{\text{media}}))$$

Secondly, calculate the stock of the effect



“if the media spend is heavily concentrated in some single time periods with an on-and-off pattern”*

Firstly, transform media investment (impressions) to adstock

$$y = \beta_{\text{media}} * \text{Hill}(\text{Adstock}(x_{t,\text{media}}, \dots, x_{t-l,\text{media}}; L, w_{\text{media}}(l; \alpha_{\text{media}}, \theta_{\text{media}})), K_{\text{media}}, S_{\text{media}})$$

Secondly, transform the adstock to incremental effect of the media by the response curve



“If media spend in each time period is relatively small compared to the cumulative spend across multiple time periods, ... we would prefer to apply the shape transformation after the adstock transformation”*

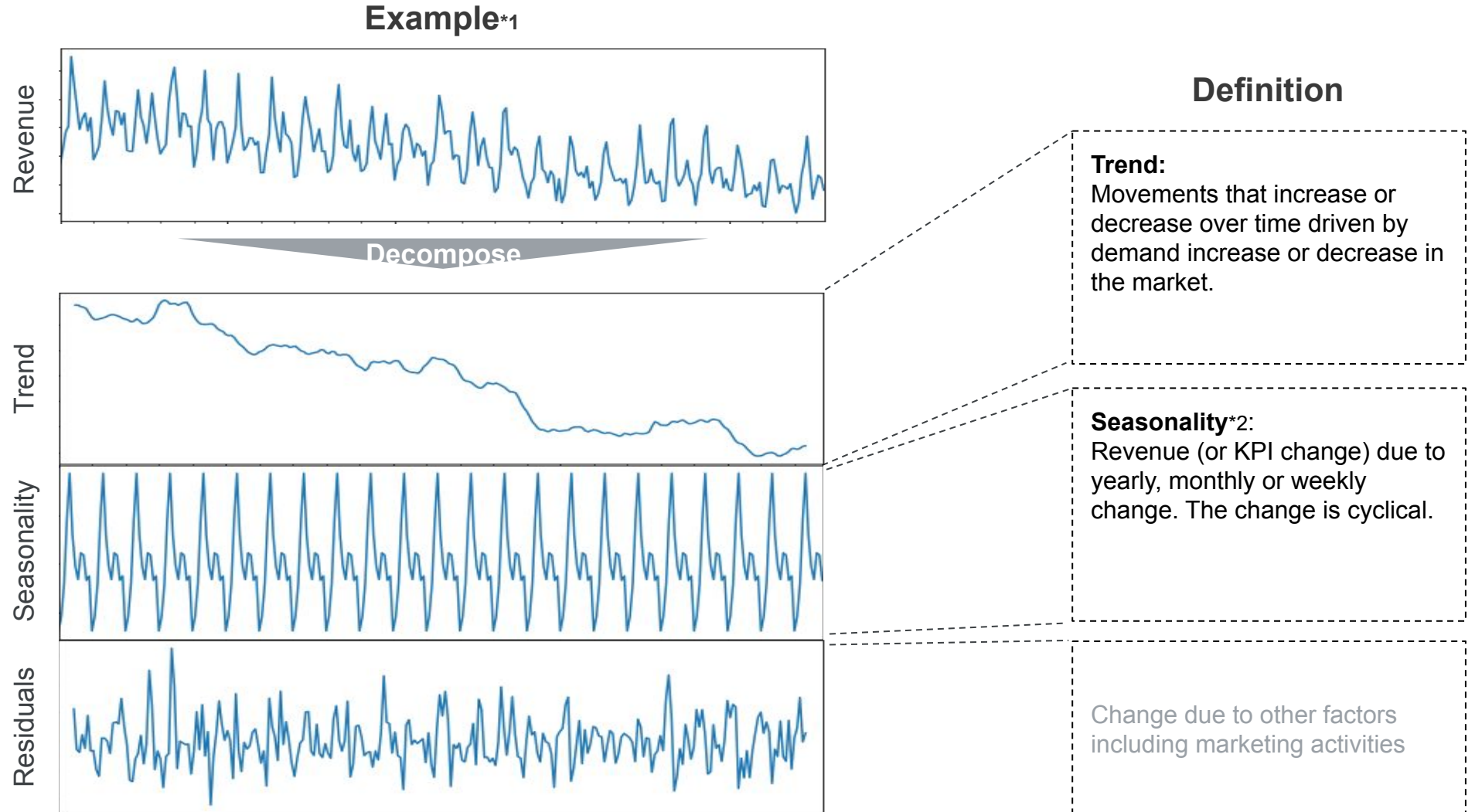
*: Google (2017), Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects

(C) Overview of trend and seasonality

Trend and seasonality are notions in time series models. In a time-series model such as ARIMA, the dependent variable is decomposed into trend, seasonality and residuals. Depending on the time series, other cycles can also be considered.

In MMMs, these elements need to be described appropriately.

Primary options are described on the next page.



*1: This example data is based on arrival volume of onion in Tokyo wholesale market.

*2: Depending on the model, holiday effect may be also decomposed. Also, there may be different cyclic factors (weekly, monthly, quarterly trend etc.), depending on the situation.

Description of trend and seasonality

Choices of trend and seasonality description

There are two options:

Using observed variables: create a trend and seasonality model with observed variables. Market data such as macroeconomic data and industry level reports (e.g., distribution volume of a product category announced by an industry association) may be able to describe the trend.

Describing unobserved variables: assume specific functions for trend and seasonality without observed data. Also, more flexible structure such as Bayesian Structure Time Series (BSTS) can be utilised.*1

Using observed variables

Trend

Variables related to overall demand in the market like macroeconomic factors (GDP growth, wage rate) or distribution volume of a product category announced by an industry association etc.

Seasonality

Variables to capture the spike/dip or cyclical change in KPIs during holidays like Thanksgiving, Christmas, New Year, Back to School, Retailer Promotions days etc.

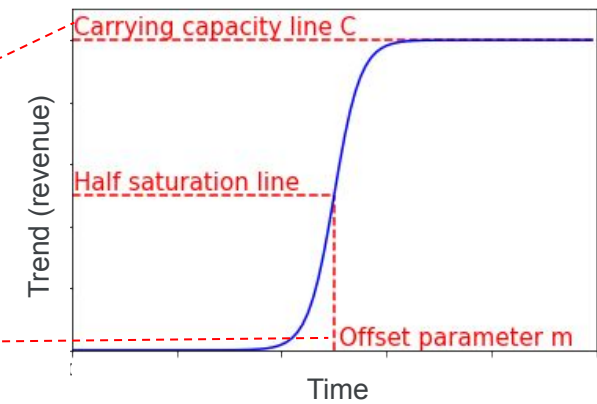
Describing unobserved (latent) variables

Example function

$$\text{trend}_t = \frac{C}{1 + e^{-k(t-m)}}$$

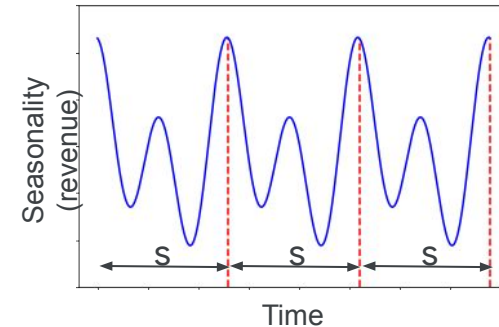
C: Carrying capacity
k: Growth rate
m: offset parameter

Example curve shape



$$\text{seas}_t = \sum_{k=1}^d \left(\gamma_{1,k} \sin \frac{2\pi kt}{s} + \gamma_{2,k} \cos \frac{2\pi kt}{s} \right)$$

d: Degree of seasonality (higher d means more complex curve)
γ: Scale parameter (higher value means larger fluctuation)
s: Seasonal frequency (52 for weekly, 365 for daily etc.)



*1: Google (2015), Inferring causal impact using Bayesian structural time-series models

Recap on a MMM model structure

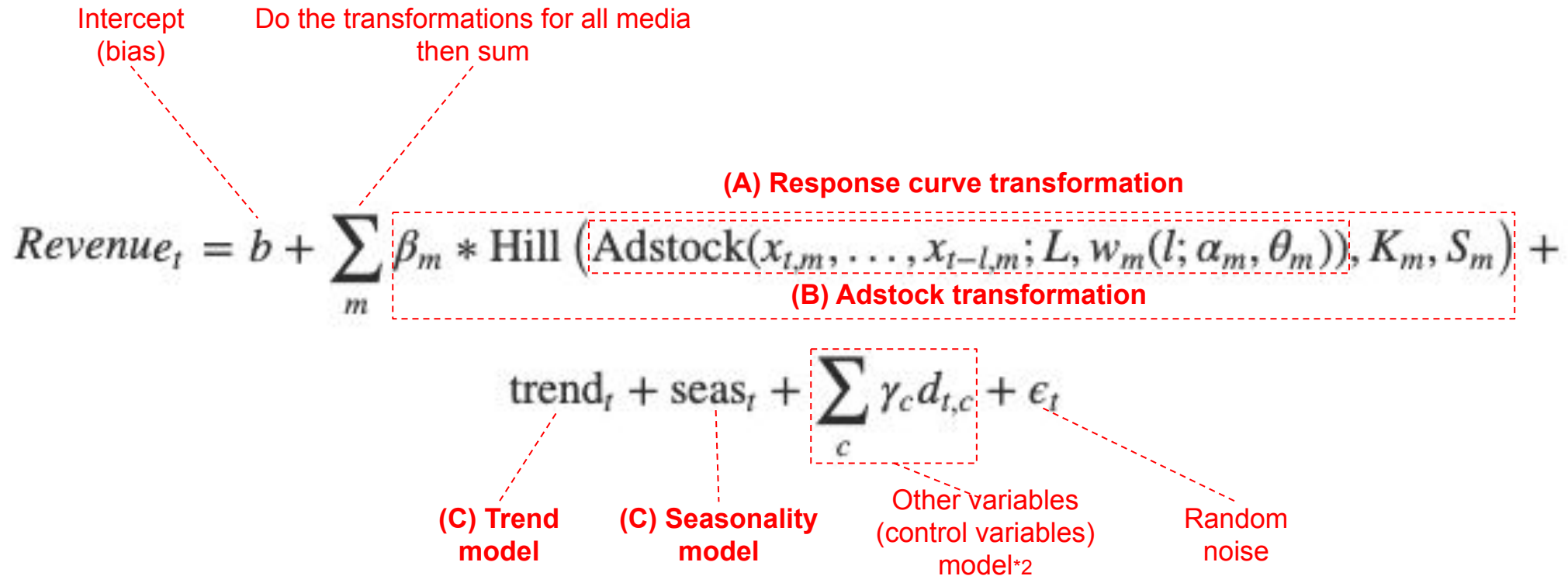
In this section, specific elements in MMM were introduced.

A modeler needs to specify each element and the order of response curve and adstock transformations depending on the data availability and volume in the time series data.

Variant models (geo, brand and audience models) are described on the next page.

Also, further advanced models are discussed in the next chapter.

An example of an additive MMM model*1



*1: This model may be too simple to describe actual situations and is to explain basic features in a MMM model.

*2: While the detail is not described in this guidebook, other variables should be modeled appropriately. For example, revenue response to price discount is different depending on the discount percentage, item and timing and the response may not be linear.

A variant model (geo unit level MMM)

Geo-level MMM

Situation

- Relatively short-term data availability (e.g., only 2 years data = 104 data points for each variable).
- On the other hand, a geographical breakdown (e.g., media investment at zip code level) is available.

Solution

Create an MMM model at the **geographic unit** (e.g., prefecture, zip code) level to have geo-level parameters for media effectiveness (scale), trend, seasonality and other variables.*1

$$\text{Revenue}_{t,g} = b_g + \sum_m \beta_{m,g} * \text{Hill}(\text{Adstock}(x_{t,m,g}, \dots, x_{t-l,m,g}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m) + \text{trend}_{t,g} + \text{seas}_{t,g} + \sum_c \gamma_{c,g} d_{t,c,g} + \epsilon_{t,g}$$

National Model

1 Market × 104 Weeks

104

Observations



DMA Model

208 Market × 104 Weeks

21,840

Observations

Example*3

As a result, **data volume to create response curves will increase**. That could increase the accuracy and stability of the model.

Modelers may need to have different response and adstock curves by areas. For example, estimating response and adstock curve separately between cities and rural areas may be a consideration because consumer behavior is different.

*1: Google (2017), Geo-level Bayesian Hierarchical Media Mix Modeling. *2: While the detail is not described in this guidebook, other variables should be modeled appropriately. For example, revenue response to price discount is highly depending on the discount percentage, item and timing and the response may not be linear.

A variant model (brand or audience level MMM)

Brand or audience level MMM

Situation

- Media investment effectiveness and efficiency could be significantly different depending on the audience or brand.
- In addition, brand or audience level data is available.

Solution

Create an MMM model at the **audience or brand unit** level to have audience or brand parameters for media effectiveness (scale), trend, seasonality and other variables.*1

$$\text{Revenue}_{t,b} = \tau_b + \sum_m \beta_{m,b} * \text{Hill}(\text{Adstock}(x_{t,m,b}, \dots, x_{t-l,m,b}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m)$$

Or **combining geo and brand breakdown models** is also an option.

$$+ \text{trend}_{t,b} + \text{seas}_{t,b} + \sum_c \gamma_{c,b} d_{t,c,b} + \epsilon_{t,b} \quad *2$$

$$\text{Revenue}_{t,g,b} = \tau_{g,b} + \sum_m \beta_{m,g,b} * \text{Hill}(\text{Adstock}(x_{t,m,g,b}, \dots, x_{t-l,m,g,b}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m)$$

$$+ \text{trend}_{t,g,b} + \text{seas}_{t,g,b} + \sum_c \gamma_{c,g,b} d_{t,c,g,b} + \epsilon_{t,g,b} \quad *2$$

*1: Google (2017), Geo-level Bayesian Hierarchical Media Mix Modeling

*2: While the detail is not described in this guidebook, other variables should be modeled appropriately. For example, revenue response to price discount is different depending on the discount percentage, item and timing and the response may not be linear.

2.4. Parameter Estimation

Overview of parameter estimation

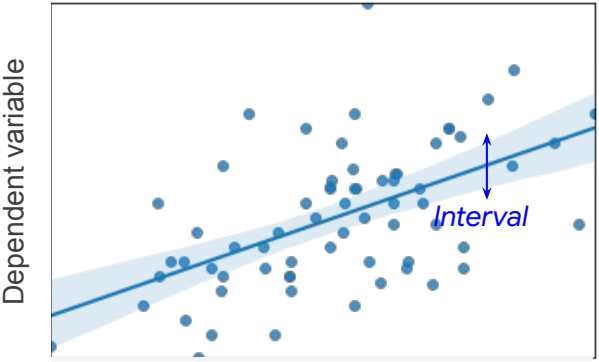
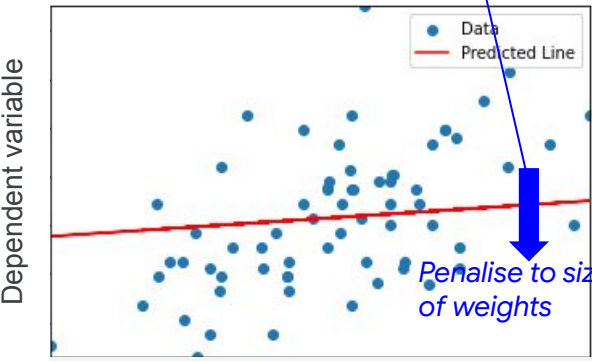
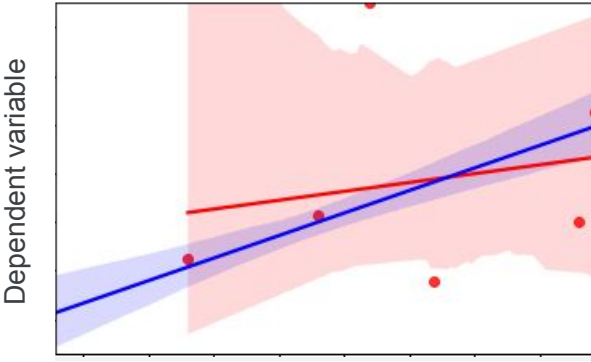
In parameter estimation for MMMs, one of three options is typically used.

Ordinary Least Squares: this option is similar to ordinary multivariate linear regression.

Regularisation: this option prioritises the model's KPI (e.g. revenue) prediction accuracy rather than estimation of incremental revenue driven by each media.

Bayesian estimation: this option prioritises getting insights on estimation of the incremental revenue driven by each media. Detailed information such as credible intervals of parameters and the posterior distributions can also be estimated. In this playbook, this approach is taken for this reason.

Primary options to estimate parameters in MMM

OLS (Ordinary Least Squares)	Regularisation	Bayesian estimation
$y \sim f(x, w)$ $L = \min_w y - f(x, w) ^2$  <p>Dependent variable</p> <p>Independent variable</p>	$y \sim f(x, w)$ $L = \min_w (y - f(x, w) ^2 + \lambda w ^2)$  <p>Dependent variable</p> <p>Independent variable</p>	$y \sim f(x, w)$ $p(w y, x) \sim p(y x, w)p(w)$ $p(w) \sim \text{prior}$  <p>Dependent variable</p> <p>Independent variable</p>
<p>Estimate parameters by minimizing the "distance L" between the data and the model. By assuming normal distribution of residuals in the model, confidence interval of parameters can be calculated in addition to point estimation of parameters.</p>	<p>Estimate parameters by minimizing the "distance L" between the data and the model with penalisation of size of parameters*. To avoid overfitting to the data, cross validation is used to optimise penalisation parameter lambda. Generally the result includes only point estimation of the parameters.*1</p>	<p>Estimate parameters through sampling process based on Bayesian theorem such as MCMC. The sampling process needs a prior distribution as an assumption. The result includes the posterior distribution and credible interval estimation of parameters. The approach gives modelers flexibility to compare assumption (prior distribution) and result (posterior distribution).</p>

*1: There are ways to add confidence interval in regularisation approach such as bootstrap resampling may be computationally intensive.

Overview of Bayesian estimation 1/3

Description of data and parameters in MMM

In Bayesian estimation, Bayes theorem is utilised to estimate parameters in an MMM.

Bayes' theorem consists of likelihood, prior, evidence and posterior.

The MMM estimates the posterior which describes parameter distribution in the model.

The use of Bayes' theorem to estimate the posterior is described on the next page.

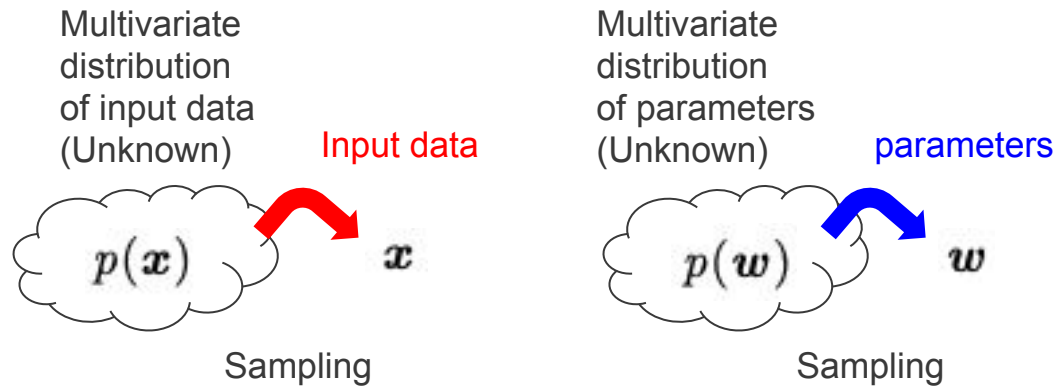
An example of an additive MMM model*1:

$$Revenue_t = b + \sum_m \beta_m * Hill(Adstock(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m^1) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

To simplify, we denote:

Input data: \mathbf{x} **Parameter:** \mathbf{w}

Then **data** and **parameters** are considered to be samples from the probability distributions which are unknown.



Bayes' theorem*2

$$p(\mathbf{w}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{w})p(\mathbf{w})$$

Likelihood:

Probability distribution of the observed data given parameters value

Prior:

Probability distribution of the parameters independently from any observations

$$\Leftrightarrow p(\mathbf{w}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{w})p(\mathbf{w})}{p(\mathbf{x})}$$

Parameter information

Posterior:

Probability distribution of the parameters given observed data

Evidence:

Probability distribution of the observed data independently from any parameter value

➔ By using the Bayes' theorem and assuming Bayesian priors, parameter information in a MMM can be estimated. Please see the next page for the mechanism.

*1: This model may be too simple to describe actual situations and is to explain basic features in a MMM model.

*2: <https://towardsdatascience.com/bayesian-inference-problem-mcmc-and-variational-inference-25a8aa9bce29>

Overview of Bayesian estimation 2/3

To use Bayes' theorem to estimate parameter in an MMM, a modeler needs to understand how to use each part of the theorem.

Disregard the evidence: The denominator of Bayes' theorem can be disregarded because it is independent of the parameters w .

Calculate likelihood from the data: this can be calculated from the data and MMM model assumptions.

Thus, if we also have an assumption of the prior distribution on the right hand side of the equation, the posterior distribution can be calculated. The estimation process (MCMC) is explained on the next page.

Overview of how to use the Bayes theorem

An example of an additive MMM model:

$$Revenue_t = \hat{b} + \sum_m \hat{\beta}_m * Hill(Adstock(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

Bayes' theorem (and how to use it for parameter estimation in MMM)*3:

Probability distribution of difference between predicted and actual revenue

$$\prod_{i=1}^N p_\epsilon(f(x_i, w) - Revenue_t)$$

Likelihood can be calculated from the model and the data

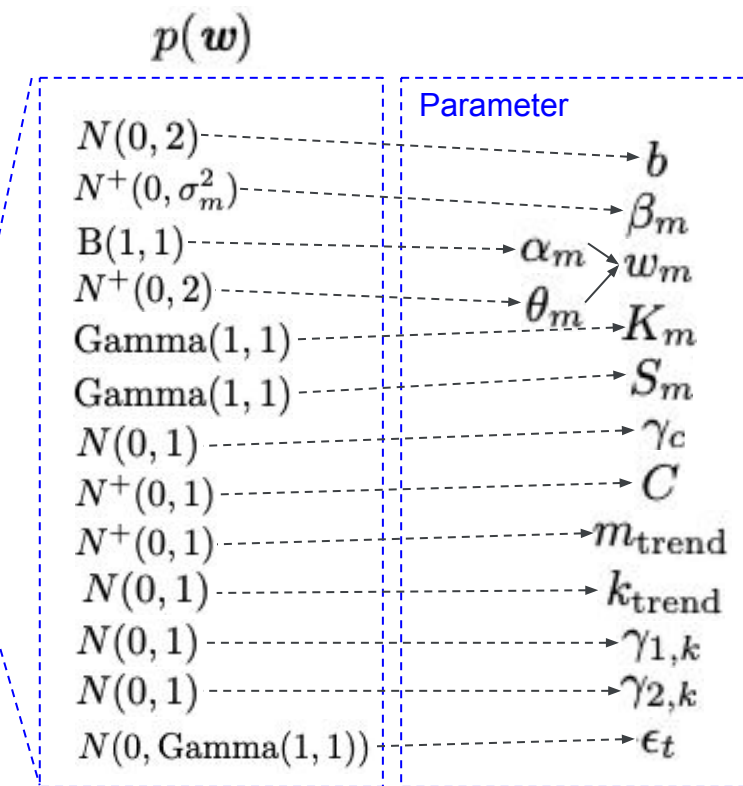
$$p(w|x) = \frac{p(x|w)p(w)}{p(x)}$$

Posterior distribution (parameter information to be estimated)

Disregard evidence because:

1. Fixed value (no dependency on parameter w) in the MMM model
2. Difficult to calculate*1

Example of Bayesian prior distribution*2



*1: <https://towardsdatascience.com/bayesian-inference-problem-mcmc-and-variational-inference-25a8aa9bce29>

*2: The prior distribution can be modified based on prior knowledge, such as past model results for similar businesses. See [LightweightMMM](#) for examples of parameter settings.

*3: Precisely because the data (bold x) are divided into independent variable x and dependent variable y , and the errors of y and the regression model are assumed to be normally distributed under given x , $p(w|x,y) = p(y|x,w)*p(w)/p(y|x)$.

Overview of Bayesian estimation 3/3

The steps to use Markov Chain Monte Carlo to estimate parameters are as follows:

Step 1 parameter sampling: based on the prior distributions, initial samples are extracted.

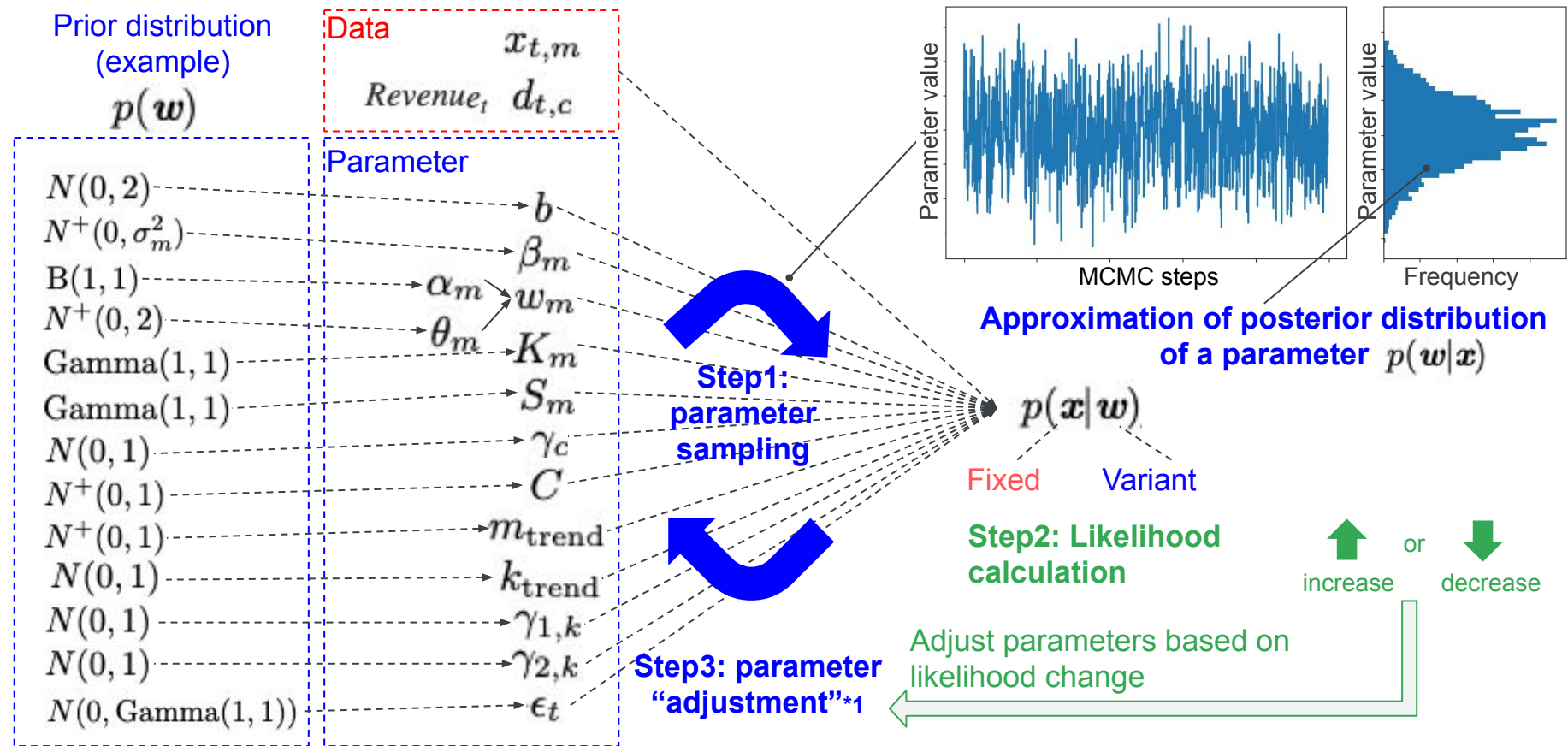
Step 2 likelihood calculation: as explained on the previous page, the likelihood can be calculated.

Step 3 “adjustment” of parameters: based on the likelihood calculation, parameters value are adjusted. A specific algorithm such as Metropolis-Hastings algorithm is used to “adjust” parameters.

Step 4 Repeat Steps 2 and 3 thereafter.

The above process is called an “MCMC step”. After iteration of MCMC steps (e.g., 5000), a modeler can estimate the distribution of parameters based on tracing of the steps. One such set of MCMC steps is referred to as a “chain”. Modelers generally run multiple chains to check the stability of the distributions.

Overview of MCMC (Markov Chain Monte Carlo) to estimate parameters in an MMM



*1: There are several methods of “adjustment” such as Metropolis-Hastings algorithm, Gibbs sampling (a special case of Metropolis-Hasting algorithm).

2.5. Model Validation

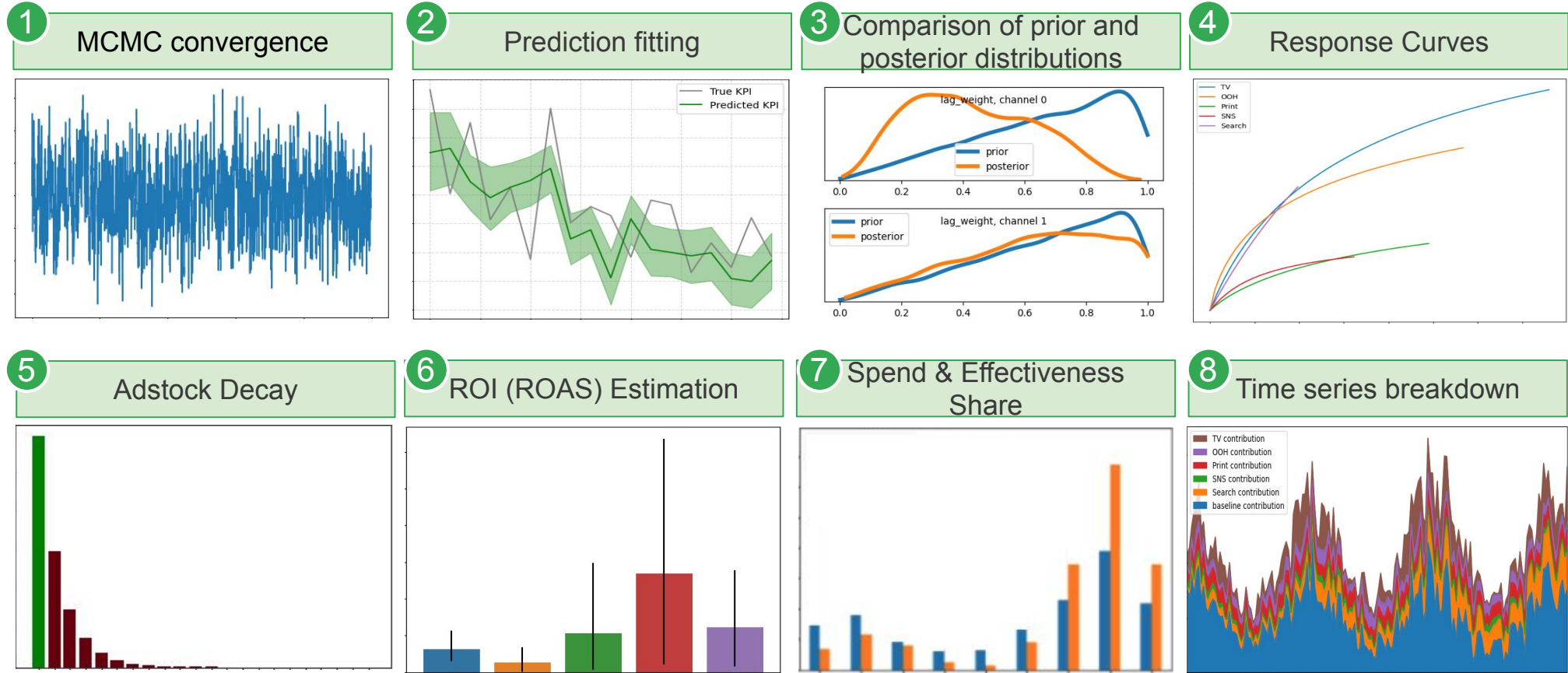
Model evaluation (“9 checkpoints”)

After parameter estimation, the modeler needs to validate the model on several dimensions.

In this section, typical validation points (“9 checkpoints”) are described.

These “9 checkpoints” consist of both objective metric evaluation and subjective evaluation. This is because the MMM model is based on several assumptions such as response curves and adstock.

Primary model evaluation points*1



9 Compare multiple models in the above points and choose the best model

*1: using Lightweight MMM to visualise the charts on 2-8.

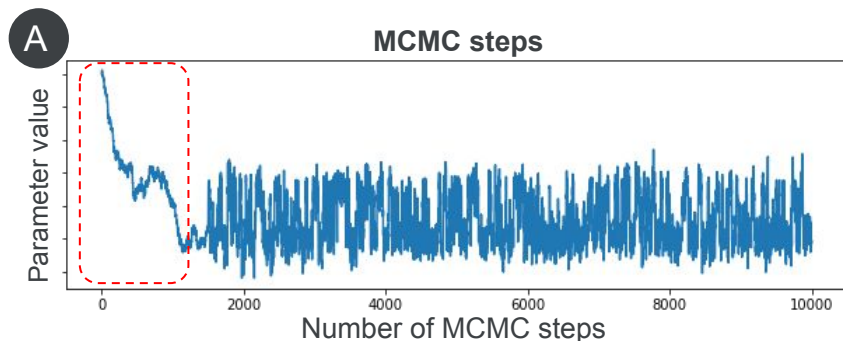
MCMC convergence check 1/2

In parameter estimation by MCMC, a modeler checks whether the parameters are estimated appropriately.

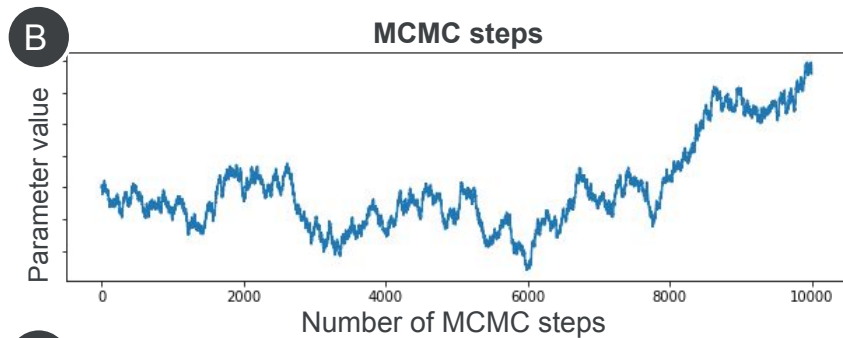
One of the primary indicators is Rhat which is an approximation of the ratio of variance of average of chains (one set of MCMC steps) to average of variance within each chain.

Generally Rhat < 1.1 is considered as convergence of the parameters while there are multiple suggestion on the threshold (such as 1.01). Also, a modeler can check the stability of MCMC steps with the visualisation as shown on the right.

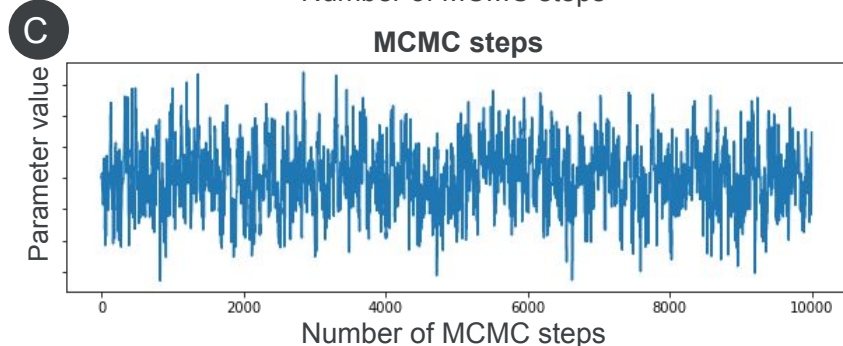
Potential problems in parameter estimation with MCMC (a single chain example)



Not good
First part (t=0-1500) looks very different.



Not good
Parameter value exploration is not consistent.



Good
Steps look consistent and the space is explored many times.

A metric to check MCMC convergence

Increase the number of chains (“attempts of MCMC”) and calculate Rhat (the Gelman-Rubin statistic).

m = number of chains
 n = number of MCMC steps in a chain (excluding burn-in period)
 x_{ij} = j th observation in i th chain

$$\hat{R} = \sqrt{\frac{\hat{V}}{W}}$$

Variance of average of chains

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B$$

$$B = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{x}_{..})^2}{m-1}, \bar{x}_i = \frac{\sum_{j=1}^n x_{ij}}{n}, \bar{x}_{..} = \frac{\sum_{i=1}^m \bar{x}_i}{m}$$

$$W = \frac{\sum_{i=1}^m s_i^2}{m}, s_i^2 = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n-1}$$

Average of variance within each chain

If Rhat > 1.1, MCMC does not converge (not good).

MCMC convergence check 2/2

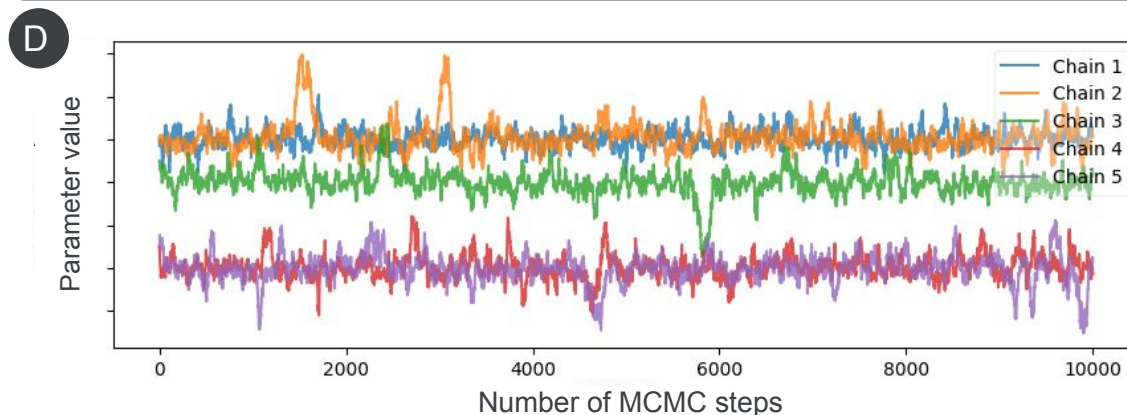
If a modeler see errors in MCMC, there are options they should consider.

Discard burn-in period: discard the early samples (e.g., the first 1,500 steps in an MCMC chain) to reduce dependency on the initial sample value.

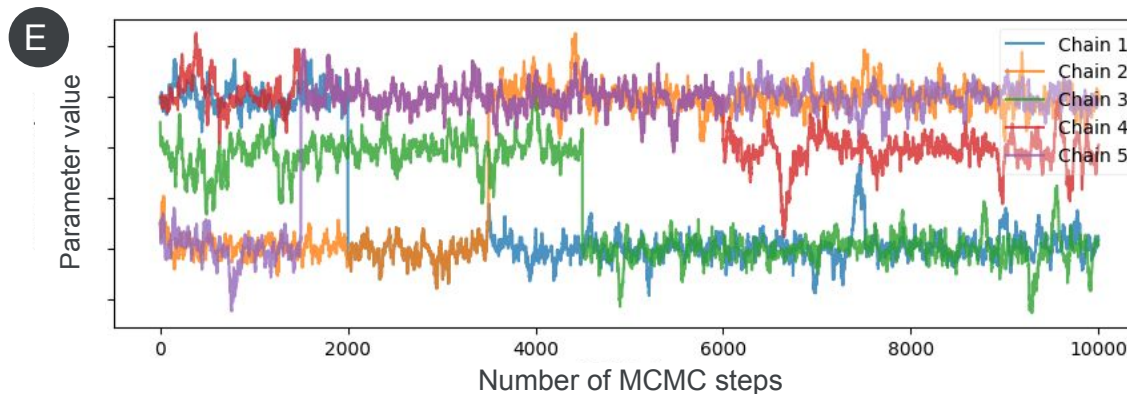
Increase MCMC steps: increase the length of each MCMC chain.

Identify different behaviors in the data: as shown in “E” case, if there is a significant difference across MCMC chains, a modeler needs to consider potential missing variables in the input data.

Potential problems in parameter estimation with MCMC (multiple chains example)



Not good. Some of chains have lower parameter values than others depending on the initial value.



Not good. There are “jumps” in some chains.

What to do? (but not limited to)

	Example	What to do
Discard burn-in period	A	Discard ‘burn-in’ period (e.g., t=0 to t=1500 in the previous page) so that Rhat value decreases.
Increase MCMC steps	B D	Increase number of samples (number of MCMC steps) to ensure convergence across chains.
Identify different behaviors in the data	E	Investigate the root cause of “jumps” in the chain. For example, the tendency to estimate parameter values may differ by region, and it may be necessary to separate the model by region, etc.

Prediction fitting check

67

A prediction fitting check measures a model's prediction accuracy on the KPI.

KPI fitting check with the sample data: check the model prediction to understand whether the model predicts the KPI accurately on the past data.

KPI fitting check with new data (out-of-sample data): do the same thing with new data to understand whether the model is applicable for future prediction as well.

In terms of volume balance between sample data and out-of-sample data, 7:3 or 8:2 is a consideration. However, business nuance is more important than a rule of thumb. For example, using 2 years weekly data to create a model and test the model with a quarter data may be valid if the business model has similar life cycle across the quarters.

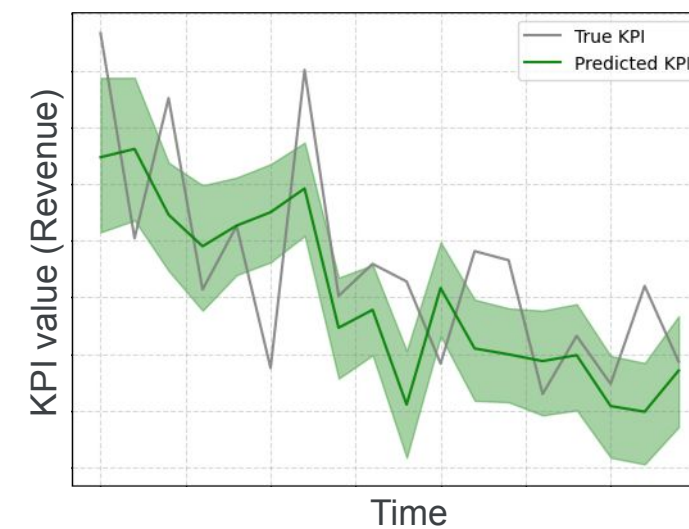
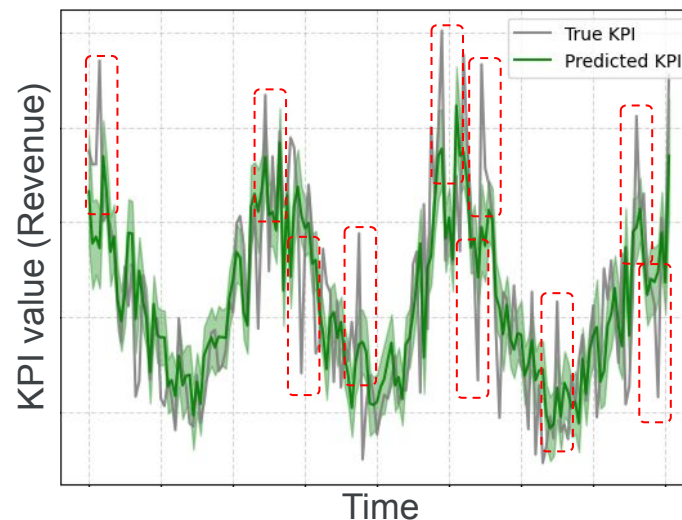
Primary metrics for prediction fitting are introduced on the next page.

KPI fitting check

Sample data

Out-of-sample data

Output example



How to check

Check gap between true the KPI value and the predicted KPI value to consider additional variables.

Apply the model to new data (out-of-sample data) and check the gap between **the predicted** and **the actual** value.

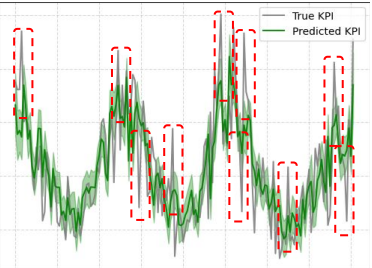
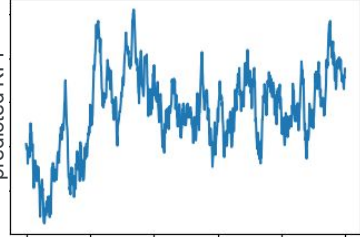
Example case

There are **significant differences**. Probably those are due to some events or holidays. A potential solution is to add events and holiday variables to the data. *1

In terms of model application to the new data, the model describes the actual value well.

*1: Instead of adding new variable(s), adding flexible trend and seasonality structure with latent variables (e.g., BSTS (Bayesian Structure Time Series)) may be a consideration.

Metrics for prediction fitting checks (for both sample data and out-of-sample data)

	R-squared	MAPE	Durbin-Watson statistic
Objective and how to check	<ul style="list-style-type: none"> Evaluation of degree of fitting between the actual KPI (e.g., revenue) and the predicted KPI Low (e.g., R-squared < 0.85) or high (e.g. MAPE > 10) values may imply there are other important variables not included in the model 		<ul style="list-style-type: none"> Detection of autocorrelation of the differences between the actual and predicted KPI DW > 2.5 or DW < 1.5 may imply other important variables not included in the model
Definition	<p>Sum of squared gap between predicted KPI and the mean of the KPI</p> $R^2 = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$ <p>Sum of squared difference between actual KPI and the mean of the KPI</p>	<p>Times 100 for percentage description</p> $MAPE = \frac{1}{T} \sum_{t=1}^T \left \frac{y_t - \hat{y}_t}{y_t} \right \times 100$ <p>Average percentage of difference between actual and predicted KPI</p>	$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$ <p>Sum of squared difference between actual and predicted KPI</p> <p>Sum of squared change of the gap overtime</p>
Example	<p>R-squared = 0.588, MAPE = 29.1%</p>  <p>Green: predicted KPI Gray: actual KPI</p>		<p>High autocorrelation graph example</p>  <p>Difference between actual and predicted KPI</p> <p>Time</p> <p>High correlation implies other important variables the MMM does not include.</p>

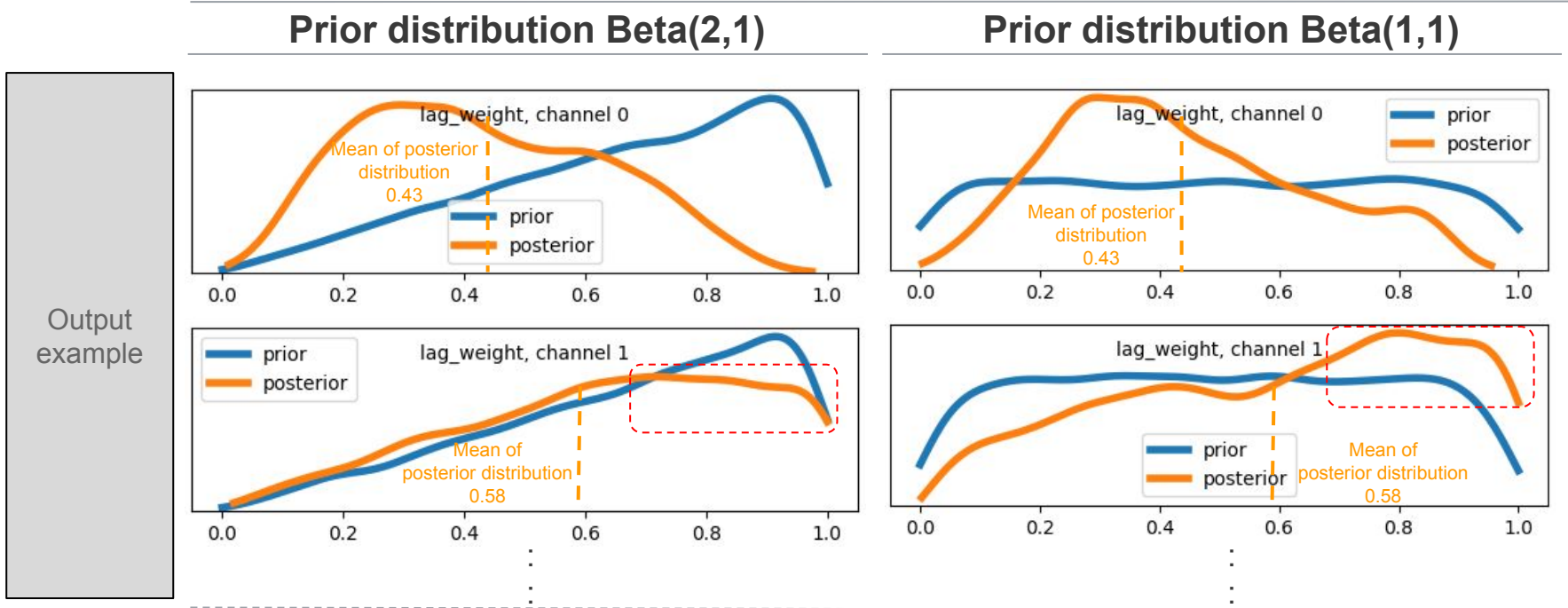
Comparison of prior and posterior distributions

A modeler needs to check consistency of parameter by using different prior distribution options in terms of:

Difference of posterior distribution: by changing prior distribution, posterior distribution may change. However, if there is consistent tendency in the data to the business KPI, the posterior distribution will be consistent. If the posterior distribution varies depending on the prior distribution, a modeler needs to validate rationale to choose a specific prior distribution based on the past experience or industry standards. This point is a benefit of Bayesian approach because a modeler can understand relation between the model's assumption (prior distribution) and the result (posterior distribution).

Difference of mean of posterior distribution*1: also, checking consistency of mean of the posterior distributions is necessary as the mean is used to calculate each media contribution and the ROAS.

Example



Output example

How to check

Example case

Estimate parameters with different prior distribution options to check:

- Whether posterior distributions are significantly different depending on the options
- What to extent means*1 of the parameters is different depending on the options

If parameter estimation/distribution is inconsistent, rationale to choose prior distributions should be validated based on evidences (past experience, industry standards etc.).

In the above example, while there are differences in the posterior distribution of the parameter "lag_weight, channel 1" depending on the prior distribution (Beta(2,1) or Beta(1,1)), the means of the posterior distributions are nearly identical regardless of the prior distribution. However, it is necessary to investigate why the probability density of 0.7 to 1.0 is higher for the prior distribution Beta(1,1).

*1: If means of parameters are used to calculate KPI contribution and ROAS (ROI).

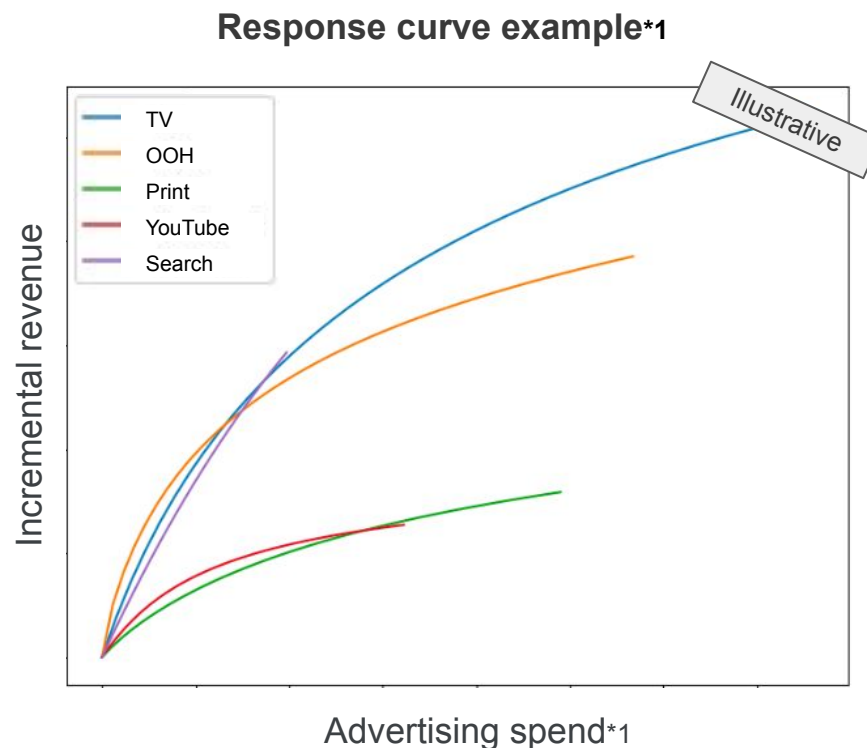
Response curve checks

Due to the presence of several assumptions in an MMM model such as response curves and adstock, the modeler needs to conduct some qualitative and (to some extent subjective) checks.

The first is a comparison of the response curves' shape and gradient. Generally "s-shape" and concave curves are appropriate for traditional and digital media respectively.

Moreover, subjective comparison of the gradient of the curves is needed in conversation with the marketing operations team.

How to check response curves (subjective check)



Curve shape

Does each media's response curve shape make sense given prior knowledge?

On the left, the TV curve should probably be an "S-curve" instead of a concave curve.

Gradient comparison

Do the gradients of the curves make sense given prior knowledge?

On the left, the gradients of TV and OOH curves may be too steep (steeper than digital media)

What to do (example)

If they do not make sense, the modeler needs to adjust the response curve models

(e.g., in the example on the left adding parameter constraints to TV, OOH curves)

*1: To calculate ROI, modelers need to transform input variable from impression to spend by using average cost per impression after estimating response curves.

*2: Control variables (promotion, price etc.) can be analyzed in the same way, if necessary.

Adstock decay check

Adstock decay also needs subjective checks.

A modeler needs to consider the different of ads format in each media and check whether the model describes these differences appropriately.

Generally rich media formats such as video ads have a relatively low decay rate (slow decay) while static formats have a relatively high decay rate (fast decay). The modeler needs to check if the model describes the decays appropriately by comparing adstock parameters across media.

Example and how to check

Illustrative

Example

Assume a Geometric decay

$$\text{adstock}_{t,media} = x_{t,media} + \lambda_{media} * \text{adstock}_{t-1,media}$$

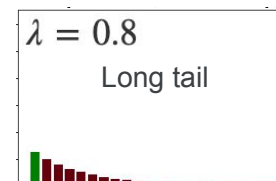
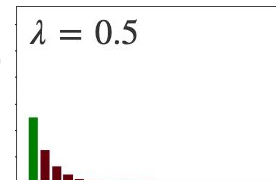
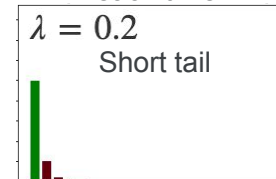
Decay
parameter

Parameter estimation result

	mean	std	median	5.0%	95.0%
TV	0.32	0.14	0.32	0.10	0.55
OOH	0.59	0.18	0.59	0.26	0.88
Print	0.67	0.24	0.70	0.32	1.00
YouTube	0.70	0.22	0.76	0.34	1.00
Search	0.71	0.21	0.75	0.41	1.00

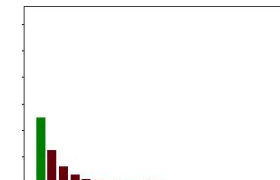
Check adstock implied by the parameter estimates
(As illustrated to the right)

Media spend and
adstock in
each time



Adstock

Latest spend has
larger impact

Relative
similarity

TV (0.32)

OOH (0.59)

Print (0.67)

YouTube (0.70)

Search (0.71)

How to
check

Do the estimated adstock decays make sense compared to accumulated experiences and industry benchmarks?

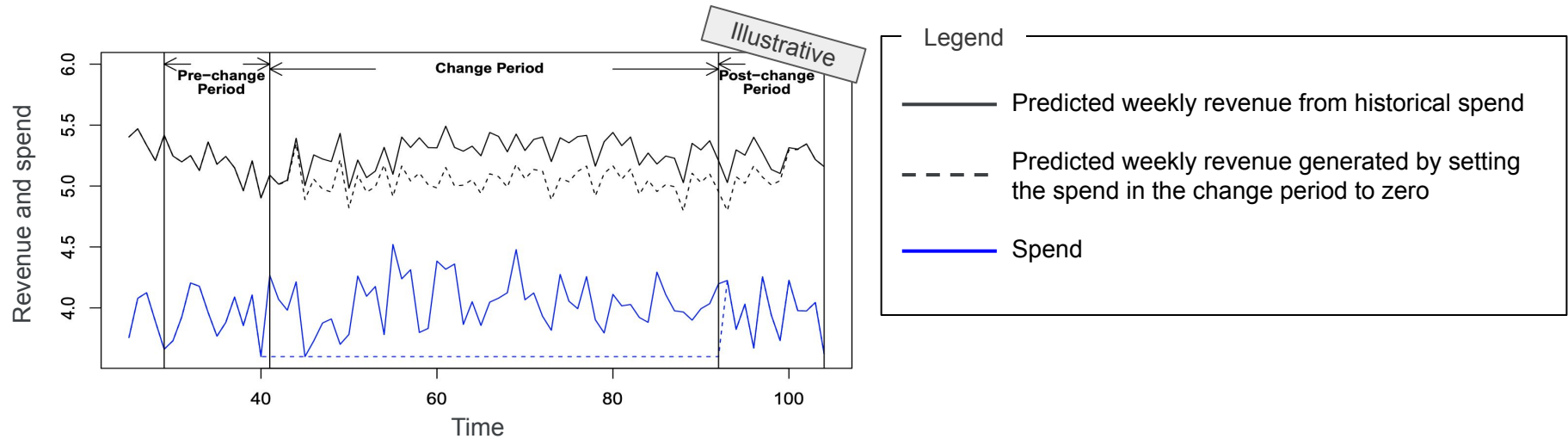
In the above case, TV decays faster than digital media. Taking the formats into consideration (TV commercial vs. newsfeed on Facebook and text ads on Search), this estimation may not make sense as we would expect video ads to drive awareness for a longer duration than newsfeeds and text ads.

How to calculate estimated ROAS (ROI)

After checking response curve shape and adstock decay estimation and concluding they are valid, a modeler calculates the ROAS (Return on Ads Spend) or ROI (Return on Investment) by media.

The denominator is the total spend of the media. The numerator is the difference between the model's predicted KPI with the actual media spend and the model's predicted KPI with an assumption that the media spend is zero. In other words the the numerator represents the incremental KPI driven by the media investment.

Definition of estimated ROAS (ROI) in MMM modeling*1



Predicted value of KPI (sales) by model
(Calculated by substituting past advertising investment into the model)*2

Predicted weekly revenue generated by setting the spend in the change period to zero*2

$$ROAS_m = \frac{\sum_{t_0 \leq t \leq t_1 + L - 1} \hat{Y}_t^m(x_{t-L+1,m}, \dots, x_{t,m}; \Phi) - \hat{Y}_t^m(\tilde{x}_{t-L+1,m}, \dots, \tilde{x}_{t,m}; \Phi)}{\sum_{t_0 \leq t \leq t_1} x_{t,m}}$$

← Numerator: Incremental revenue

← Denominator: Total spend*2

= ROAS (Return on Ad Spend)

*1: Google (2017), Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects

*2: If media impressions are used as a variable in the model, the average cost per impression (CPM, etc.) should be multiplied by the impressions and the units should be adjusted to monetary values.

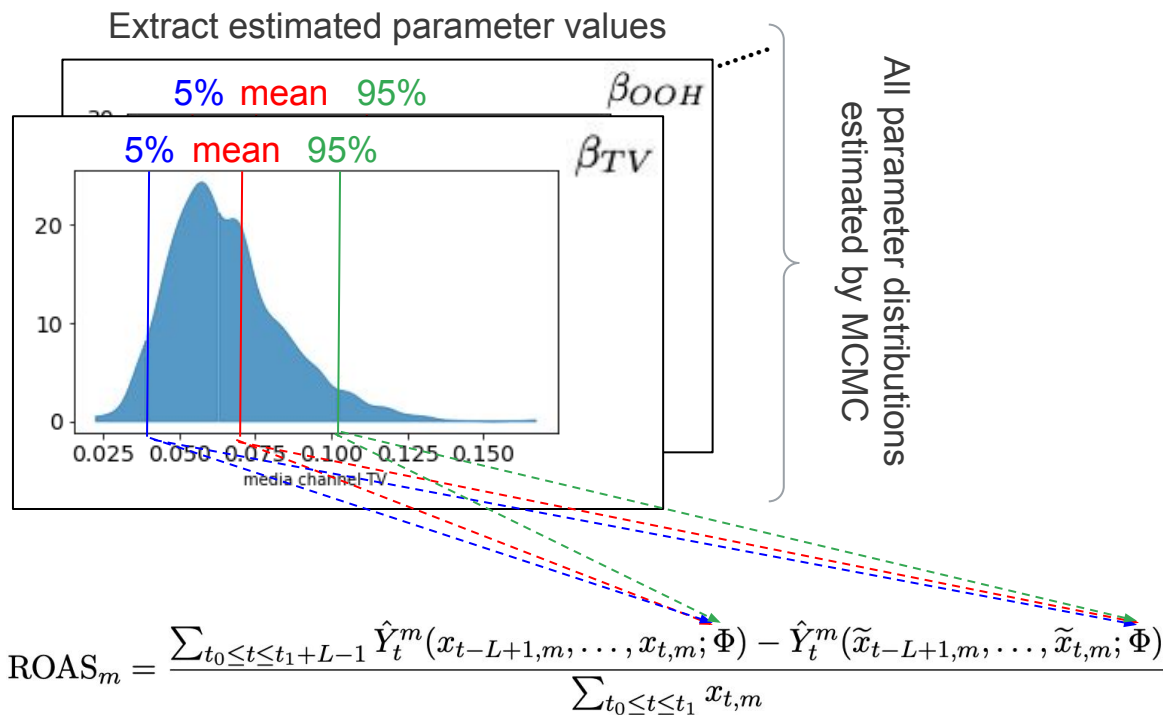
How to calculate estimated ROAS (ROI)

Through MCMC, a modeler has a posterior distribution for each parameter in the model, including statistics such as the mean, median and 5th and 95th percentiles. By using these values, the modeler can calculate intervals and mean value of ROAS (ROI) estimation.

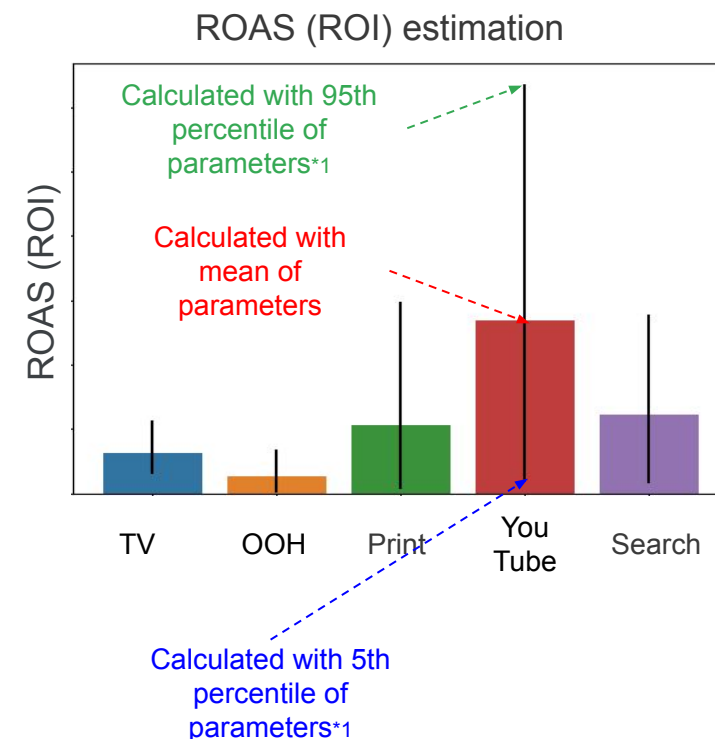
Then the modeler compare the ROI estimation result with intervals (called as “credible intervals” in Bayesian estimation). The next page describes how to validate it subjectively.

How to estimate ROAS (ROI) in MMMs

Illustrative



Assign extracted parameter values (e.g., mean, 5th & 95th percentile) to the model to estimate key statistics of the distribution of ROAS*2



*1: More exactly, calculated with parameters values which give the 95th percentile of ROAS and which give the 5th percentile of ROAS.

*2: If media impressions are used as a variable in the model, the average cost per impression (CPM, etc.) should be multiplied by the impressions and the units should be adjusted to monetary values.

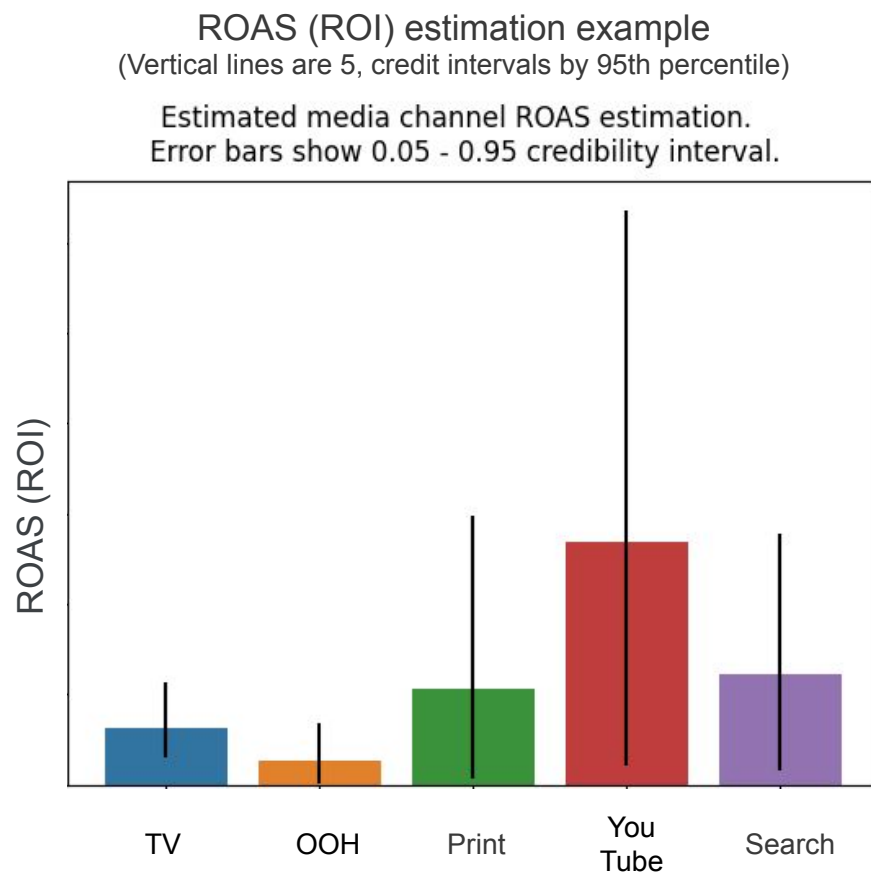
ROAS (ROI) estimation check

74

After ROAS (ROI) calculation, a modeler can check the estimated averages and intervals of these results.

Average (mean) ROAS: if there are outliers (extremely high or low ROAS), the drivers of this result should be investigated. If the reasons cannot be identified, one recommended approach is to compare the outcome of multiple models to see the differences persist.

Intervals of ROAS: wider intervals may imply instability and lack of data (e.g., a media flight is only few weeks). A modeler should avoid conclusive messages in such a situation.

How to check ROAS (ROI) estimation result Illustrative

What to do

Average
check

Do the estimates of mean ROAS make sense given any prior knowledge or industry norms?

Also, are the results consistent when changing model assumptions (e.g., prior distribution)?

Interval
check*1

What are the causes of the width of the credit interval? (e.g., lack of data, instability in performance due to skillful implementation of advertising, etc.)

The model creator checks the performance of each media campaign and considers splitting the model per target customer or per ad delivery surface to eliminate instability.

*1: There are ways to add confidence interval in regularisation approach as well such as bootstrap resampling may be computationally intensive.

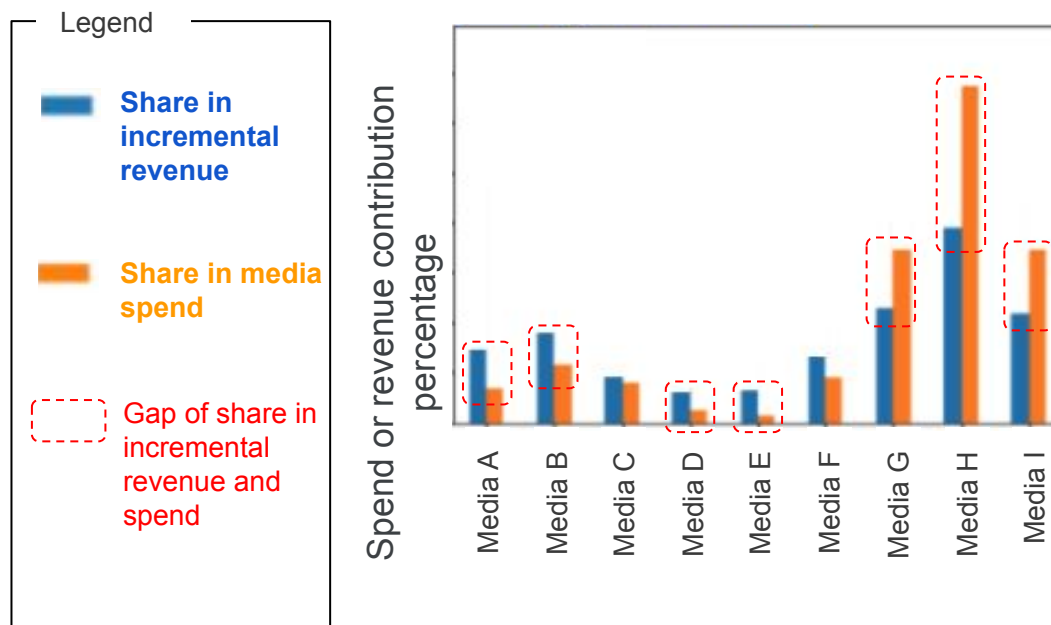
Spend and effectiveness share check

Subjective checks on spend and effectiveness share are needed to ensure that a marketing operations team will consider the model estimation result to be valid.

This check compares the advertising spend share and share of incremental KPI driven by each media. Reasons for any large share gap may need clarification because the existing media allocation by the marketing operations team will be largely driven by their expectations of incremental KPI.

Example output*1

Illustrative



Compare share in **spend** and **incremental revenue** for each media to analyze the gap

What to do

Gap check

If some media show **significant share** gap between the spend and the incremental revenue, will this make sense to stakeholders?

For example, media G, H and I in the left chart show relatively low incremental revenue share compared to their media spend. Will this make sense to the stakeholders?*2

Action

- Compare multiple MMM results and choose the most credible model the stakeholders will accept and/or
- Accept the significant gap shown by the model and consider a budget shift

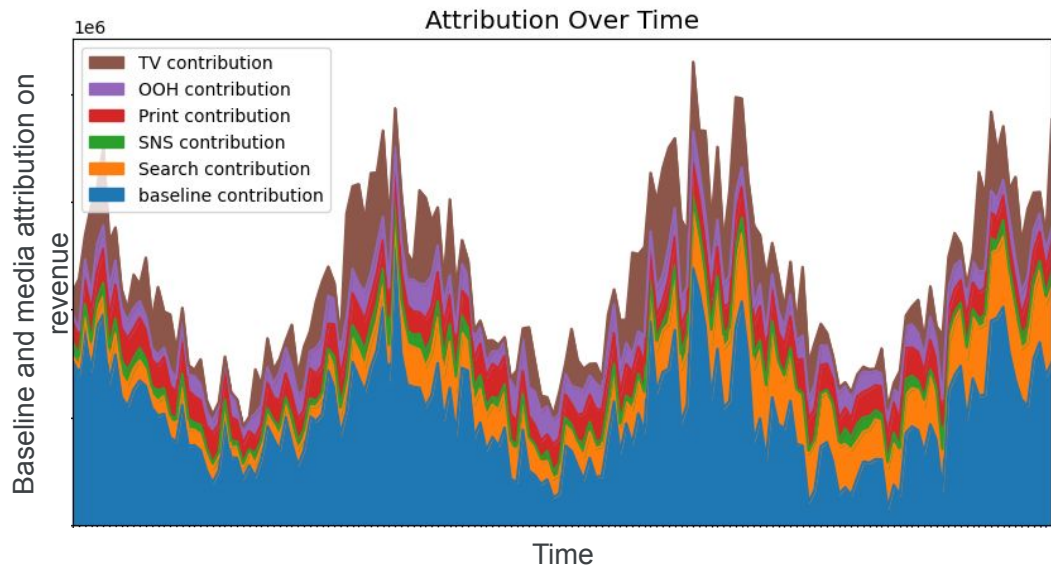
*1: Depending on the audience of the presentation, it is preferable to explain with a credit interval. *2: In raising hypotheses, it is desirable to investigate and consider the facts as objectively as possible. For example, a competitor's large-scale campaign in a particular media may decrease ROAS for that media. In some cases, the competitor's campaign trends by media cannot be included as an explanatory variable due to the physical unavailability of data, and thus the impact must be considered qualitatively.

Analysis of contribution in time series data

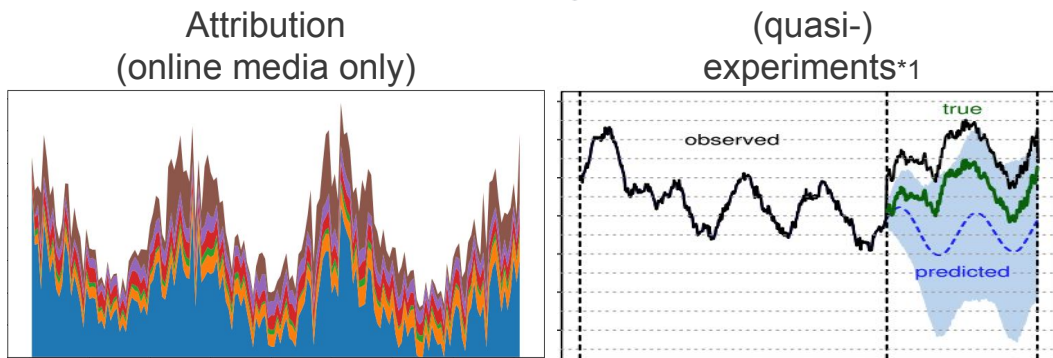
Based on the calculation in [p.72](#), each media's contribution in each time unit (e.g., week or day) is calculated by an MMM. The visualization is shown as a stacked area time series graph on the example chart. By comparing the chart with calendar (holidays), events and some big campaign periods, a modeler can understand whether a model can describe the past data correctly.

Moreover, comparing the time-series breakdown with different measurement sources such as Attribution and experimental tests is also a consideration to understand reasons of the difference.

Example output Illustrative



Comparison



What to do

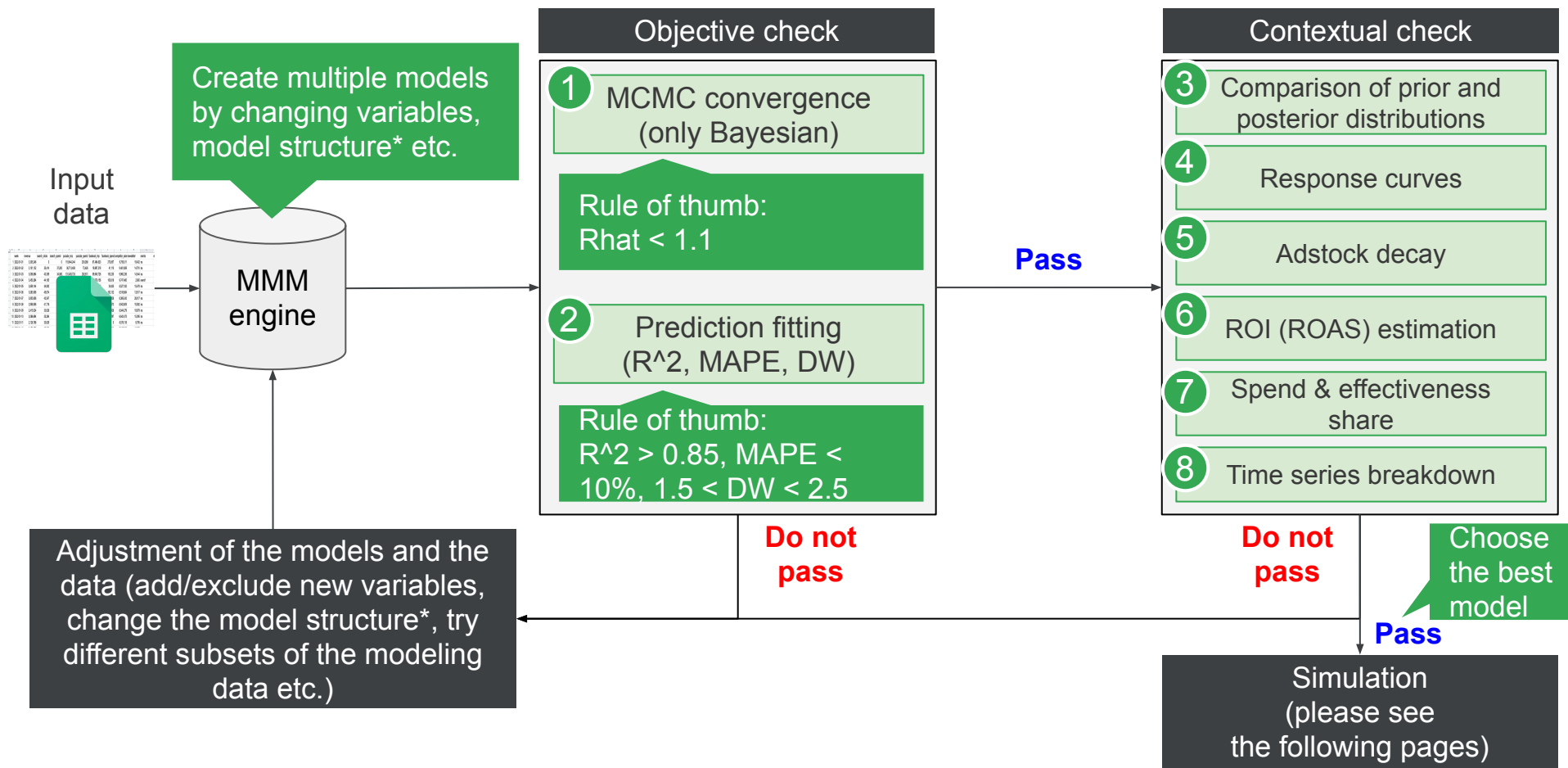
Qualitative validation	<p>Check whether trend of contribution of media and baseline (factors other than media) are correctly captured by the model.</p> <p>For example: Demand due to holidays and events should be allocated to baseline contribution.</p>
Comparison with other measurement results	<p>Check whether the MMM estimation result on the media contribution is consistent with other measurement solutions such as attribution and experimental tests.</p>
Action	<p>Identify primary reasons of the difference across MMM, attribution and experimental tests</p> <p>Adjust MMM parameters where possible (please see p.84, "calibration with experimental tests")</p>

*: Google(2015), Inferring Causal Impact using Bayesian structural time-series models

Multiple model comparison

As a MMM assumes specific structure such as response curves and adstock, the model validation needs both objective and contextual (to some extent subjective) checks by both the modeler and the marketing operations team so that they can have confidence in the model estimation. To achieve that, generally a modeler needs to create multiple models with different variables, bayesian priors and model structures to choose the best one.

While there is often criticism of the objectivity of MMMs, the above process is a practical way to utilise a MMM for business decision making.



Adjustment of the models and the data (add/exclude new variables, change the model structure*, try different subsets of the modeling data etc.)

Choose the best model

Simulation (please see the following pages)

*: The structure of the model is explained in CHAPTER 3 for direction.

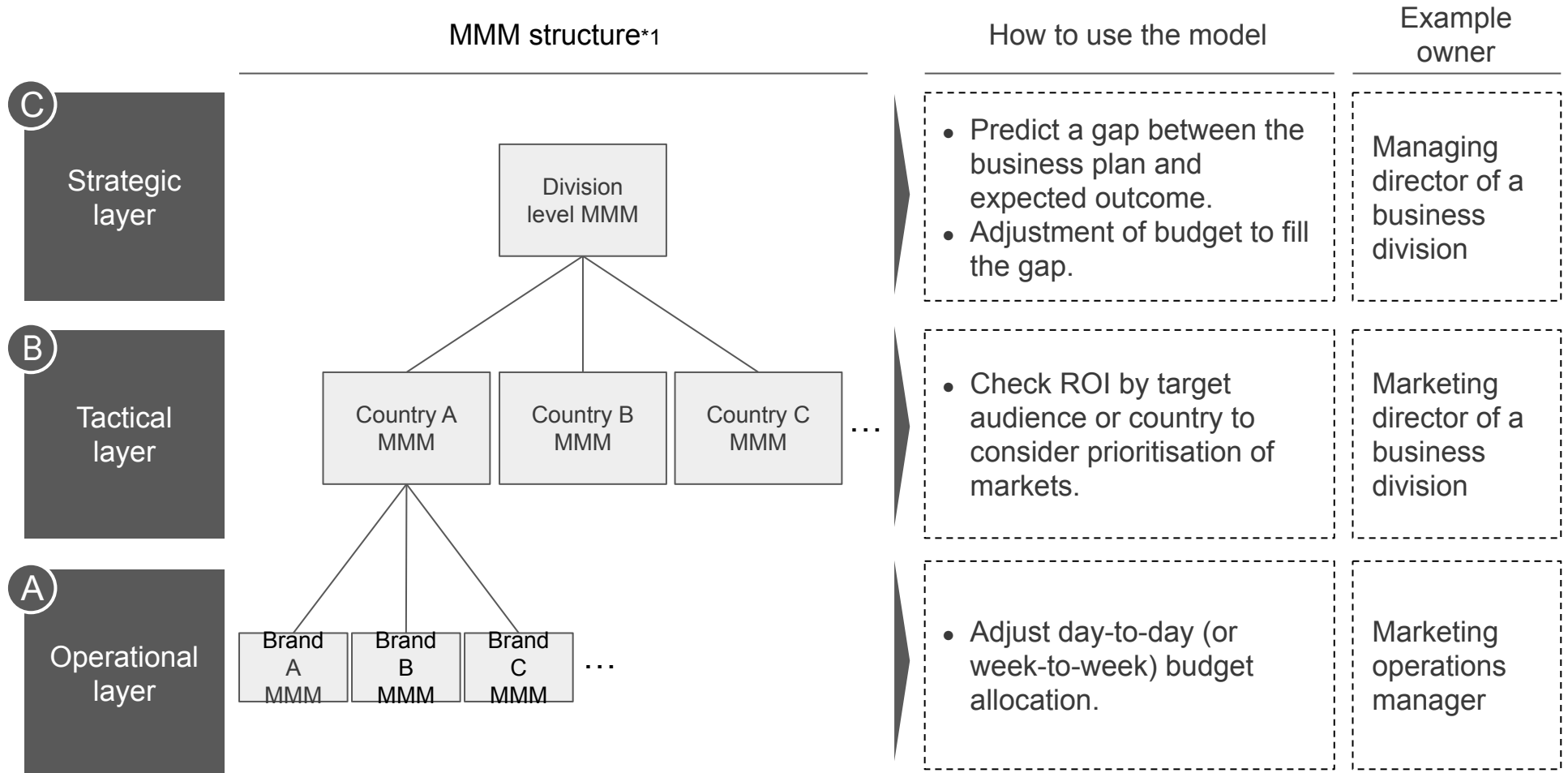
2.6. Model Utilization

Overview of an in-house MMM use case

After model creation, a modeler and the marketing team can utilise the model for several purposes.

The chart on the right shows an example of MMM creation in three layers where the bottom two layers have sub-category level MMMs (e.g., brand A, B and C). Lower layer MMMs are consolidated to a higher layer MMM. *1

In each layer, the objectives and use of the models is different. While the marketing operations layer uses the model for advertising or promotional budget optimisation, the higher layers may use the models to consider investment allocation across categories (e.g., countries).



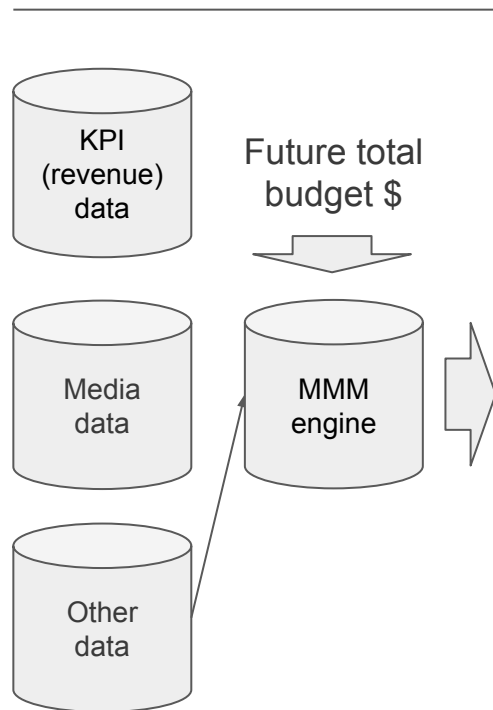
*1: Depending on similarity of the businesses, brands and customers' behavior across markets, models can consider hierarchical models as described in [p.56-57](#). If there is no similarity across businesses, brands and customers, creating each model in the bottom layer (operational layer) and adding those up is a realistic way.

A MMM usage in operational layer (budget allocation)

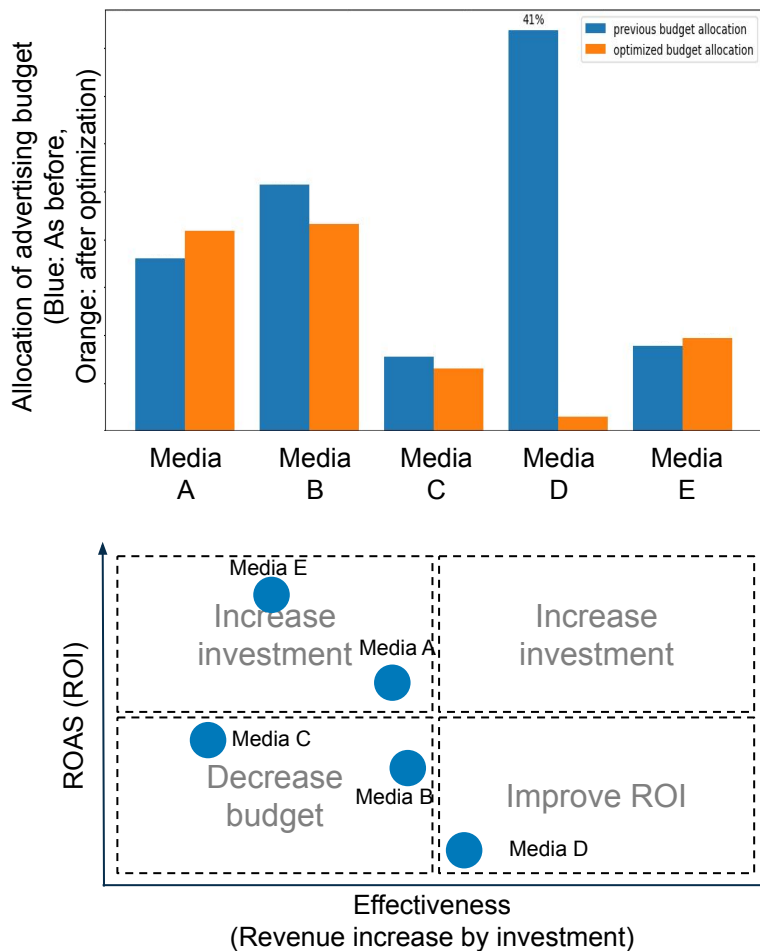
Marketing managers may use an MMM to consider optimal advertising spend across media in the short-term.

An MMM is able to suggest an optimal budget allocation by inputting future total budget (e.g., total advertising budget in the next month) and other data predictions (e.g., weather predictions, competitor activity predictions etc.) in the model. *

If the user does not have predictions of the other data, they need to make assumptions the data (e.g., using last year's data).



Input and output Illustrative



How to use this?

Adjust the next term (week, month, quarter or year) budget allocation based on the recommendation.

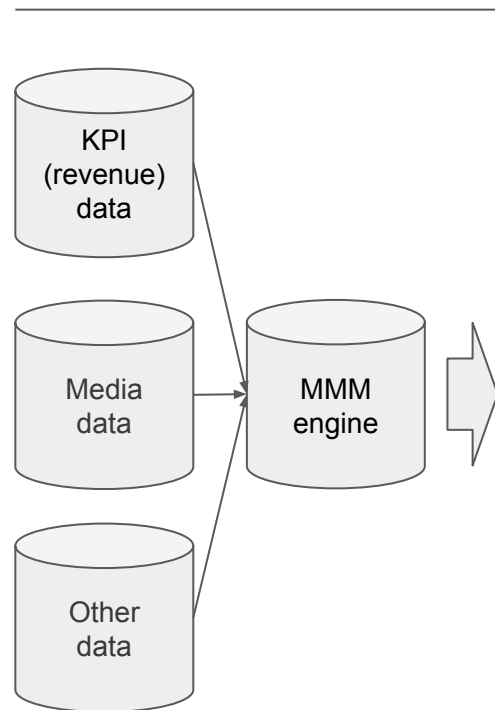
*: In practical terms, limits may be placed on the scope of optimization to avoid the risk of sudden budget changes. For example, optimization calculations are performed with certain limits on the percentage decrease in the TV budget and the percentage increase in the digital budget, etc.

B MMM usage in tactical layer (ROI(ROAS) estimation)

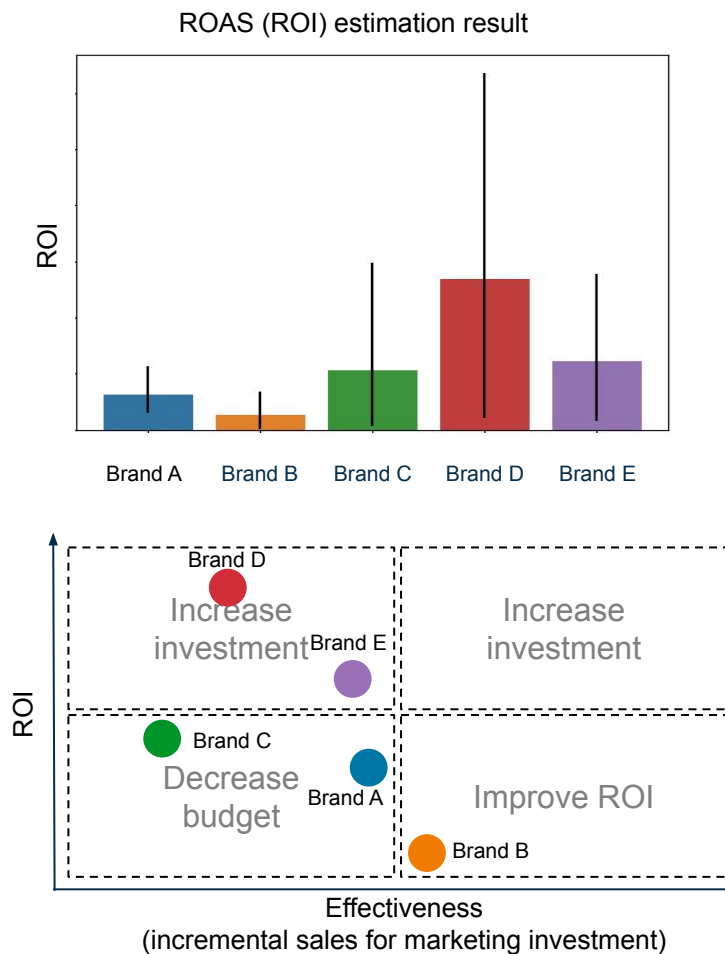
Marketing or business directors may want to estimate ROAS or ROI in the past fiscal periods at a business domain or country level to understand what worked and what did not work.

An MMM is able to estimate the ROAS or ROI across business categories based on MMMs in the operational layer.

Such estimation may contribute to market or business domain prioritisation in the future.



Input and output Illustrative



How to use this?

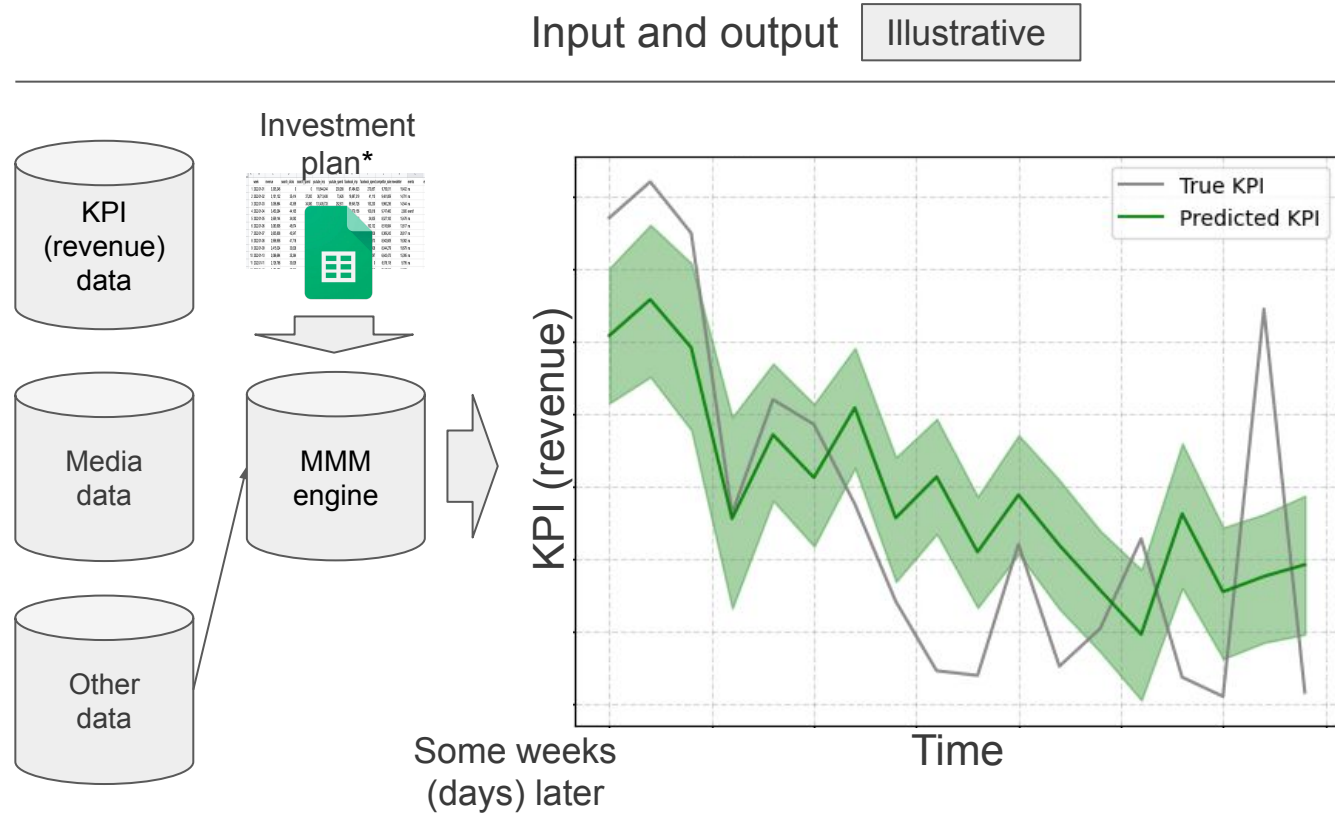
Estimate marketing ROI and ROAS for each business category (e.g., country, brand, target customer) and use as input for prioritizing markets and business areas and tactics.*1

*1: Technically, the allocation of promotional and advertising budgets should refer to marginal ROI (marginal ROI) or response curves, not to ROI (ROAS) (p.47). It is possible to compare marketing ROI across different business categories (countries, brands, target customers, etc.) and consider priorities to maximize overall profit.

C MMM usage in strategic layer (KPI prediction)

A managing director of a business division may want to know the gap between predicted and actual revenue to consider complementary investment to achieve financial targets.

An MMM may be able to predict the future revenue with appropriate input data such as the future planned advertising and promotion investment and the other data predictions (e.g., weather predictions, competitor activity predictions etc.) in the model.



How to use this?

- Track a gap between the revenue predicted by the model and actual revenue (or number of installs).
- Consider addressing any gap by increasing investment.

*1: Data on the external environment (e.g., forecasts of competitor trends, predictions about the economy, etc.) as well as investment plans for sales promotion and advertising are needed or assumptions need to be made.

2.7. Other topics on MMMs

	Calibration with experiments	Coefficients which vary over time	Long-term effectiveness
Description	<ul style="list-style-type: none"> If a company conducts experimental tests (e.g., conversion lift, geo lift) or quasi-experimental tests (e.g., DID, Causal Impact) continuously, it is worth considering calibration of the model with the results of these experiments. Bayesian methodologies can give modelers guidance on how to calibrate the models. 	<ul style="list-style-type: none"> Response curve and adstock of media may vary depending on other factors (e.g., time gap between ads and sales in December due to Xmas). Coefficients which vary over time may be suitable if businesses have frequent and significant change in terms of media response. 	<ul style="list-style-type: none"> For prominent and long-seller brands, long-term effects driven by media is important. However, typical MMMs describe only the short-term effect driven by media investment (generally some weeks). Assuming long-term impact of media is one of the considerations..
Consideration	<ul style="list-style-type: none"> Experimental test data is not always available across markets, audiences and brands. Experimental tests may be accurate during the test periods. However, it may not be accurate outside the periods. To resolve that, an MMM may need coefficients which vary over time. Conversion lift results rely on individual data which can be impacted by privacy protection in place for cookies and device ids. If a company conducts experimental tests as always-on measurement, they may be missing opportunities due to having control groups in the test which will not be exposed to their ads. While several white papers suggest how to calibrate the results of MMMs and experiments, there is no universally agreed approach as of Jan 2023. 	<ul style="list-style-type: none"> While coefficients which vary over time are suitable for dynamic businesses and calibration with experimental design as written on the left, the number of parameters tend to be large and it is very challenging to ensure MCMC convergence. Management and maintenance of the prior and posterior distributions of coefficients which vary over time can be challenging because future data may have an impact on the existing estimates of time varying parameters depending on the constraints. For example, insertion of dependencies on parameters from coefficients in adjacent time points may change the estimation with new data. *1 	<ul style="list-style-type: none"> A modeler needs to identify established metrics such as brand surveys continuously used in the company to measure long-term effects driven by marketing activities. The marketing stakeholders should have same understanding on how such brand metrics can influence their long-term KPI. Additional research in terms of the brand metrics selection and quantitative analysis between revenue and brand metrics may be needed. Moreover, to model a long-term effect on the KPI, a modeler needs to consider several constraints such as data volume (e.g., at least four years), detectable KPI impact size etc.*2

*1: Edwin NG, Zhishi Wang, Athena Dai (2021), Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling

*2: Google (2019), Measuring effectiveness: Three grand challenges

Challenges of MMM as an analytics system (not exhaustive)

MMMs for performance clients (e.g., short-term, always on use, frequent updates) are very different to traditional MMMs (e.g., modeling once a year, 4 years data for a long-seller brand). To use MMMs continuously, modelers need to resolve data related issues as follows:

- **Data pipeline automation:** how to automate data collection across internal and external data sources (e.g., Analytics Hub on Google Cloud). Some existing data sources (e.g., public data) may not have APIs to automate the data collection. A modeler may need to build the data pipeline from scratch.
- **Data granularity issues:** some data sources may not have enough granularity (e.g., TV ads data is weekly while daily data is needed). If so, a modeler may need assumptions to generate granular data or have to compromise the granularity of the model.
- **Input data prediction issues:** To predict the future KPI, future input data (e.g., future weather) should be collected on top of media or promotion scenarios. Some data may not be available and requires assumptions (e.g., using last year's weather data for the prediction).
- **Operational process to use MMMs:** Many marketing practices are still using existing measurement products such as Multi-Touch Attribution (MTA) based on individual level data. How, when and by who to use an MMM on top of existing solutions is an outstanding issue. Also, when and how the model parameters should be updated are still in discussion.

Media specific modeling knowledge (not exhaustive)

In terms of model designs on specific media, following points are in discussion.

- **Bias of estimation on search media:** consumer behavior on search media can be different depending on search word types (e.g., positive/negative words, competitors' product word etc.). Correct modeling of search media is still in discussion. *1
- **Reach and frequency on video media (e.g., YouTube):** existing models typically do not consider the effect of frequency of media exposure per person (e.g., a higher frequency may drive consumers' purchase) rather than total volume of impressions or spend. How to create a suitable model with media reach and frequency is an issue.

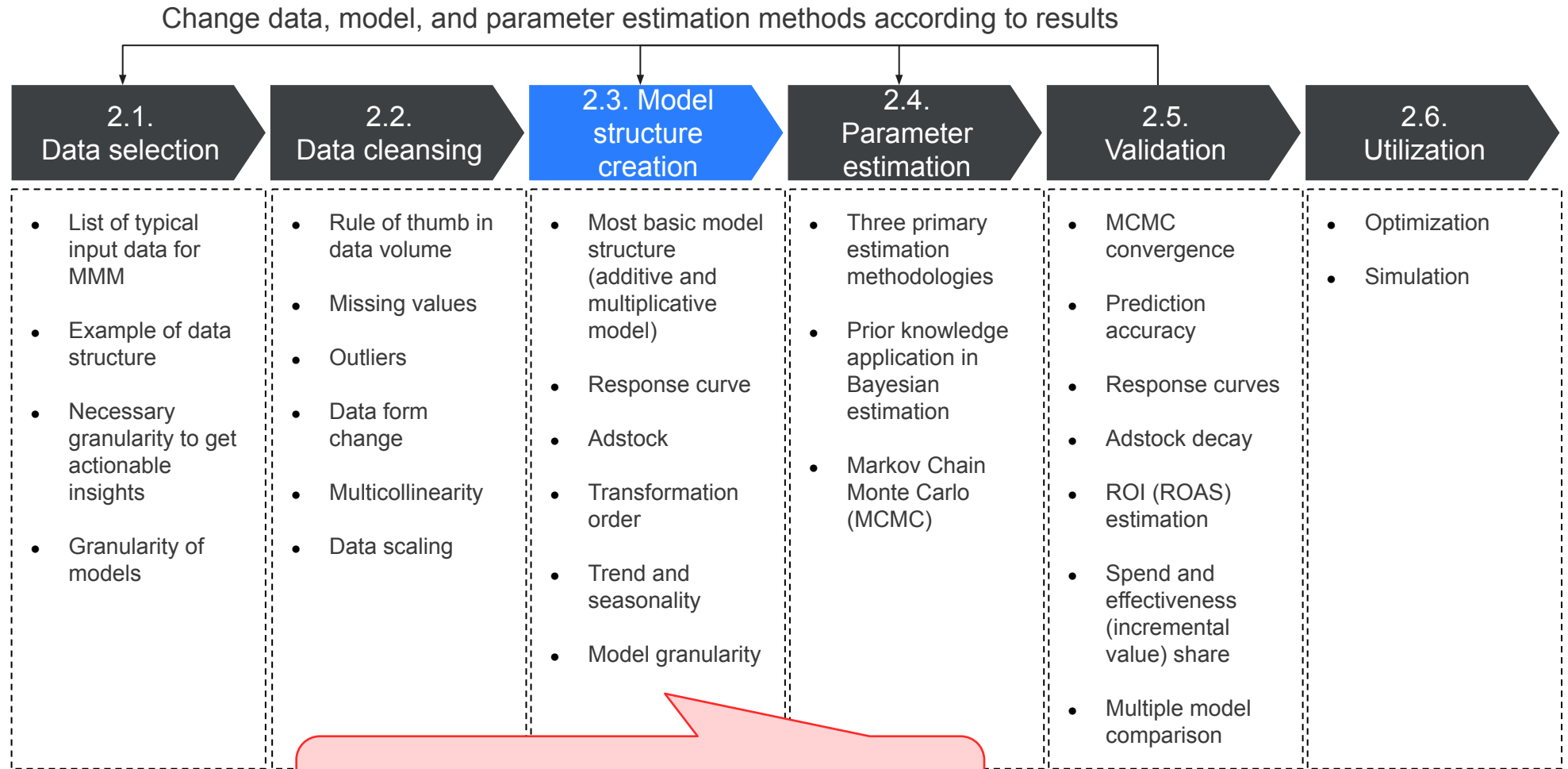
*1: Bias Correction For Paid Search in Media Mix Modeling, Google, 2018

CHAPTER

3

Considerations on MMM Model Structure

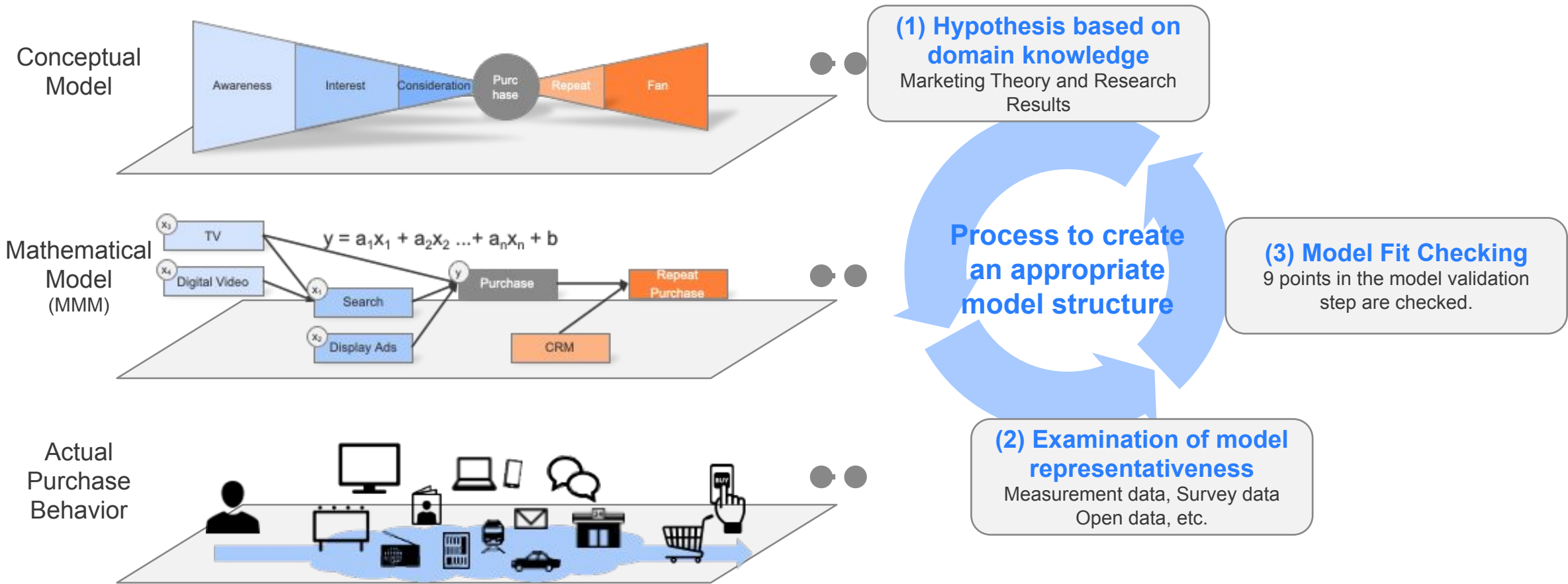
Creating the model structure is a very difficult step in the MMM model building process. It is necessary to consider and determine the appropriate model structure from various angles. In practice, the creation of the model structure is a back-and-forth process with data selection and parameter estimation steps. In this section, we show through simulations how the adequacy of the model structure affects the results of the MMM analysis. Then, tips for finding an appropriate model structure will be presented.






How to create an appropriate model structure?

3.1. Model Structure Creation Methods and Model Types

- In practice, the hypothesis building and analysis work is conducted from three perspectives, and by going back and forth between the steps, the appropriate model structure is determined.
- First, hypotheses are formulated from (1) domain knowledge of the process by which consumers purchase the commercial products and services to be modeled. Next, (2) the representativeness of the model is examined from the viewpoint of the possibility of measuring and collecting data. Finally, a model based on the hypothesis is created and verified, and (3) the fit of the model is confirmed.



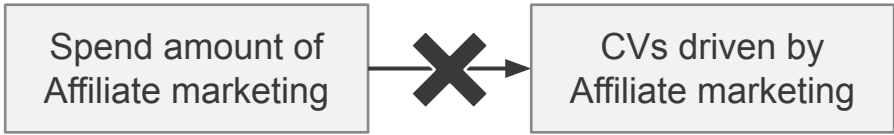
- HDY group believes that an appropriate model structure is one that reflects real-world causal relationships. In order to avoid reflecting incorrect causal relationships in the model structure, we recommend that 3 points are kept in mind.
- In particular, domain knowledge of marketing communications and media is needed to determine the temporal back and forth relationship.
 - Analysts should inquire with marketers and media representatives to gain a better understanding of the question items and survey methodology of the survey indicator used for the variable if it is marketing communications, or the media submission and measurement methods of the media to be measured if it is media.

- 1. Chronological relation** 
 - Causality is always a before/after relationship in time
 - Model structure reflects marketing media causality
- 2. Logical non-inclusion relation** 
 - Variables in an inclusion relationship must not be arranged in the same hierarchy in the model structure
 - For the objective variable, the model structure should be separated
- 3. Fulfillment of identifiability of causal graph*1 and parameters** 
 - Estimating causal graphs from data, assuming distribution of functions and parameters representing relationships among variables*1
 - If necessary, intervene and disconnect the back door
 - Adjust appropriate explanatory variables in light of single door, back door, and front door criteria

*1: There is way of casual discovery to infer causal graph with assumption on parameters and functions to describe relation among variables (e.g., IC algorithm, greedy equivalence search (GES), LiNGAM, NOTEARS algorithm etc.). The methodologies are out of this guidebook scope. Also, in this section, all graphs are DAG (Directed Acyclic Graph, which is directed and not cyclic graph) for the simplicity.

- Causality is always a forward-backward relationship when arranged on a time axis. An event that occurs later in time (i.e., its measured data) is the result of an event that precedes it.
- That is why it is important to be familiar with how to measure data and to use domain knowledge about the position of consumers' purchasing behavior on the time axis and its backward and forward relationship with other data.

Example of wrong causation: opposite of before/after relationship



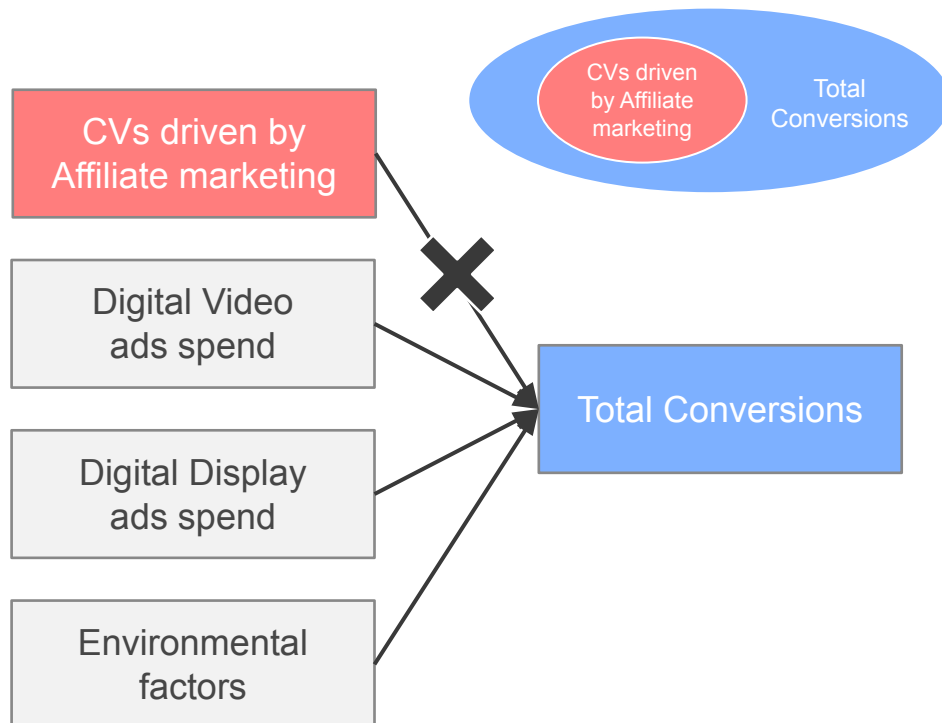
Affiliate marketing is paid on a pay-for-performance basis for conversions, so there is no causal relationship from the amount of money spent to conversions, but rather an inverse relationship.

Example of correct causation: Temporal pre- and post-relationships

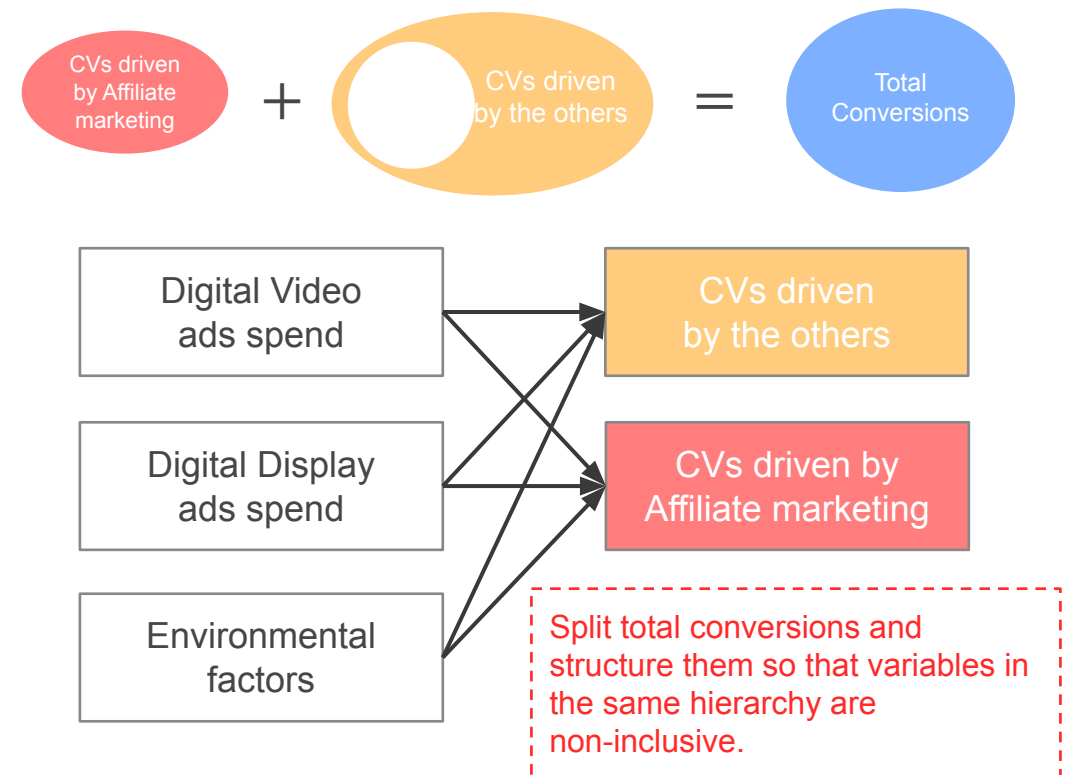


- If you want to include both a variable and a higher-level concept variable that numerically includes that variable in the model structure, you should format the data so that the variables are non-inclusive of each other.
- A possible method would be to create two model structures by logically decomposing a high level concept variable into two variables with non-inclusive relationships.

Example of wrong causation: Inclusive relation



Example of correct causation: Split into non-inclusive relation



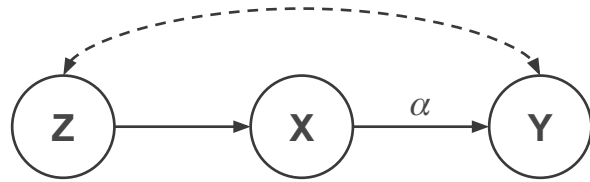
- In order to make the model structure more realistic, the parameter identifiability condition must be satisfied if intermediate indicators such as site visits and search volume are incorporated into the model, in addition to the amount of advertisement placement.
- The effect is estimated by identifying variables that satisfy criterion of single door, back door, or front door and adjusting (adding to the explanatory variables) the appropriate variables.

Legend

- Y: Objective variable
- X: Cause of interest for the objective variable
- Z: Factors other than X that are assumed to be causally related
- U: Unobserved variable
- : indicates the direction of causality
- ↔ : indicates the confounding among variables
- $r_{A|B,C}$: Partial regression coefficient of B when A is multiple regressed on B and C
- r_{AB} : Regression coefficient of A in a simple regression of A with B

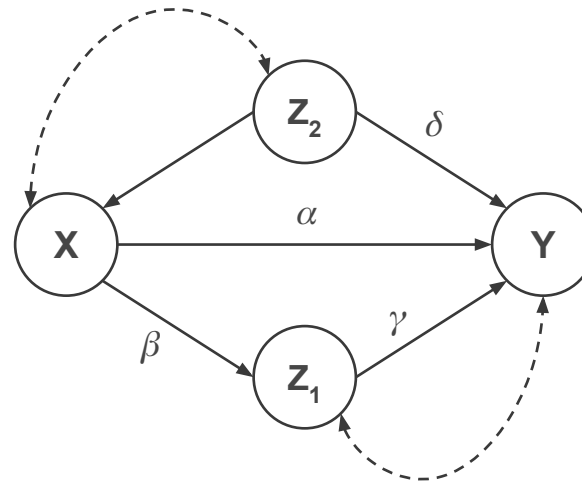
Identifiability of causal graph*1 and parameters

Single door criterion*2



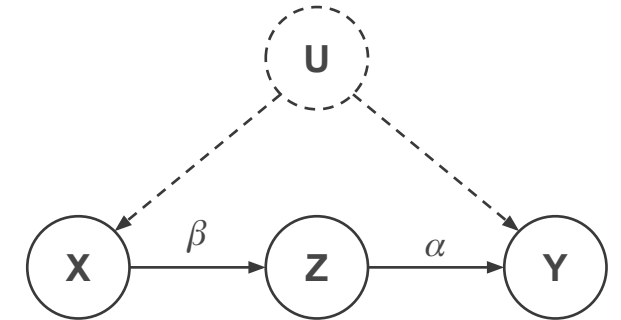
- Z meets single door criterion
- Direct effect α from X to Y is identifiable
- The direct effect is given by $\alpha=r_{YX,Z}$, consistent with the partial regression coefficient α for $Y=\alpha X+\beta Z+\epsilon$

Back door criterion*2



- Z_2 meets backdoor criterion
- The overall effect $\alpha+\beta\gamma$ from X to Y is identifiable
- The overall effect is given by $(\alpha+\beta\gamma)=r_{YX,Z}$, consistent with the partial regression coefficient α for $y=(\alpha+\beta\gamma)X+\delta Z_2+\epsilon$

Front door criterion*2



- Z meets frontdoor criterion
- The overall effect $\alpha\beta$ from X to Y is identifiable
- It is the product of the direct effects α and β , given by $\alpha=r_{YZ,X}$ and $\beta=r_{ZX}$, respectively

*1: There is way of casual discovery to infer causal graph with assumption on parameters and functions to describe relation among variables (e.g., IC algorithm, greedy equivalence search (GES), LiNGAM, NOTEARS algorithm etc.). The methodologies are out of this guidebook scope. Also, in this section, all graphs are DAG (Directed Acyclic Graph, which is directed and not cyclic graph) for the simplicity. *2: Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE"(2009)

Differences in Model Structure by Sales Channel Type

- Based on the checkpoints regarding the appropriateness of the model structure, the general model structure of practical MMM can be categorized into three types: online, hybrid, and offline, depending on the sales route of products and services to consumers.
- Type A (online type) can be subdivided into two types, “click route” and “click route + search route”, depending on whether or not TV and digital video implementation are used. The click route can be treated as a limited version of the click + search route.

Legend
 → indicates the direction of causality

Each type is a DAG (acyclic directed graph) and assumes the following

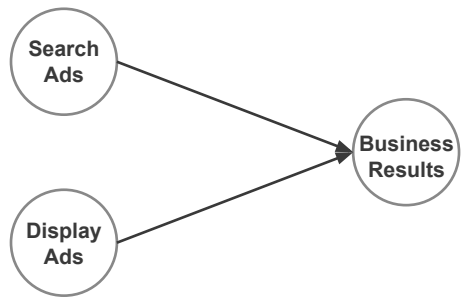
- Each element is a set of variables summarized by media characteristics, which must be decomposed appropriately when conducting MMM.
- It is assumed that there is no time dependence in causality.
- The confounding among unobservables and variables is omitted, and its problems are discussed later.

A. Online type *1,2

B. Hybrid type *1,2

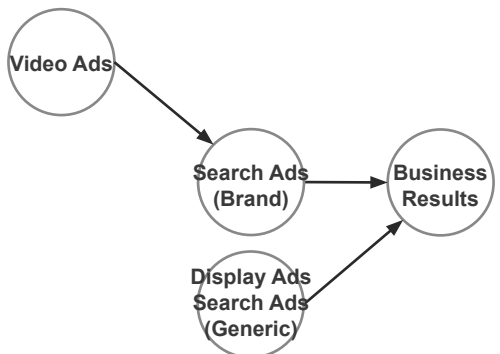
C. Offline type*2

Click route



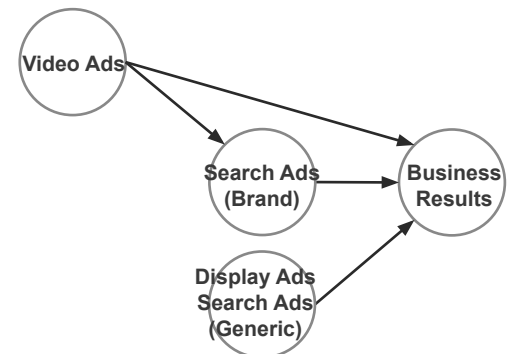
- “Click route” consists only of causal relationships that result in online CVs primarily through pay-per-click performance media placements, such as search and display ads

Click route + Search route

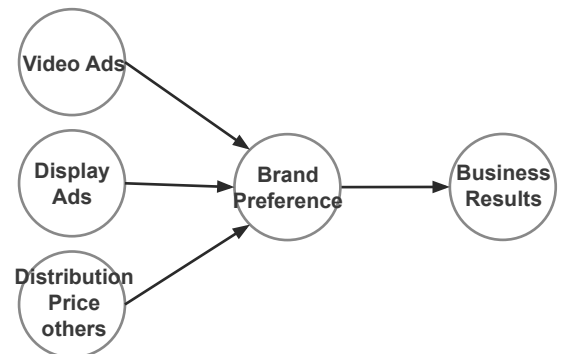


- “Search route” is added to “Click route”
- “Search route” is causal relationships in which the number of searches is changed by cognitive advertising such as TV and digital video

Click + Search + Offline route



- Offline route is added to “Online type”
- “Offline route” indicates a causal relationship in which store visits occur directly from cognitive advertising, such as TV and digital video ads.



- "Offline type" indicates a causal relationship in which sales fluctuate due to changes in brand preference caused by cognitive advertising such as TV and digital video ads

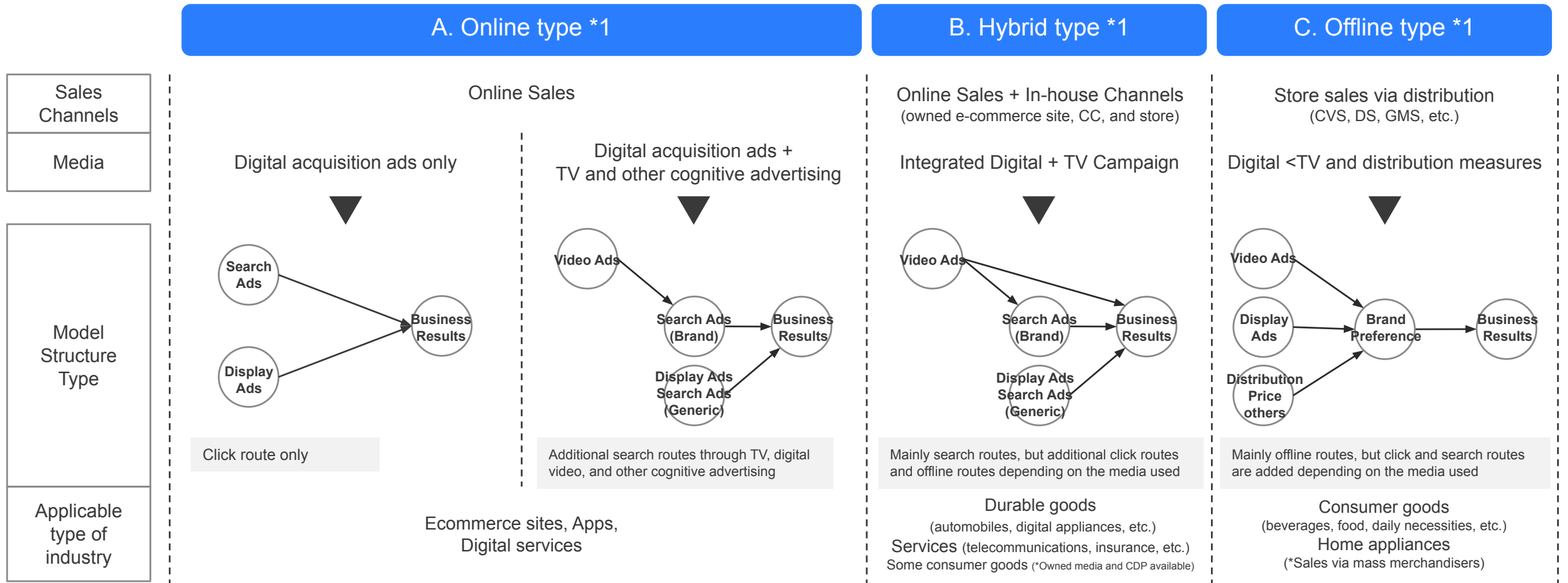
*1: Although it is originally necessary to consider the volume of brand searches, it is omitted here for the sake of simplicity. *2: The confounding (existence of unobserved common causes) indicated by the dotted arrows in both directions is an example in the above figure and should be considered on a case-by-case basis, as it may vary from business to business. For example, there may be a correlation between the amount of investment in search ads and online sales due to seasonality in “Online type”, and between the amount of advertising placement and the delivery rate in “Offline type” due to business practices.

- “Online type” model structure can be applied to cases where sales are made exclusively online, such as e-commerce sites, apps, and digital services. “Hybrid type” is a model structure in which “Offline route” is added to “Online type”, and can be applied to durable goods and service industries. In the case of consumer goods and home appliances that are sold in stores via distribution such as CVS, DS, and GMS, “Offline type” is used.
- When MMM is actually implemented, it is necessary to examine the model structure with reference to the checkpoints regarding its appropriateness.

Legend
 → indicates the direction of causality

Each type is a DAG (acyclic directed graph) and assumes the following

- Each element is a set of variables summarized by media characteristics, which must be decomposed appropriately when conducting MMM.
- It is assumed that there is no time dependence in causality.
- The confounding among unobservables and variables is omitted, and its problems are discussed later.



*1: The model structure shown on this page is an example and may vary from business to business. See note on previous page.

3.2. Verification of the impact of model structure on analysis results

- Model structures can be categorized into three types depending on the sales route of products and services to consumers. However, the creation of model structures is a process in which the analyst's subjectivity and experience play a significant role.
- Therefore, in this section, we would like to confirm the risk of inadequate consideration of model structure by examining the impact of adopting a model structure (MMM's model structure hypothesis) that differs from the true data generation process (real consumer behavior) on the analysis results, using a simple "Online type" as an example.

Legend
 → indicates the direction of causality

Each type is a DAG (acyclic directed graph) and assumes the following

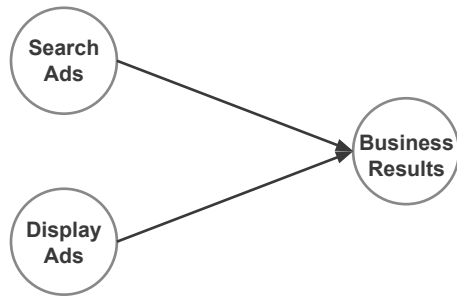
- Each element is a set of variables summarized by media characteristics, which must be decomposed appropriately when conducting MMM.
- It is assumed that there is no time dependence in causality.
- The confounding among unobservables and variables is omitted, and its problems are discussed later.

A. Online type *1,2

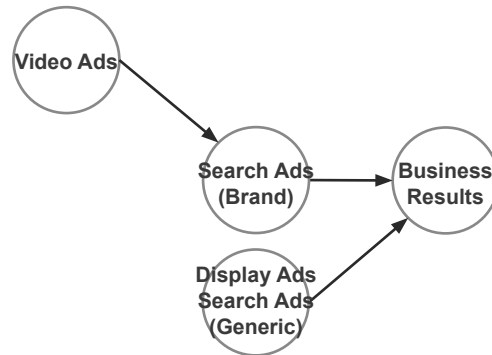
B. Hybrid type *1,2

C. Offline type*2

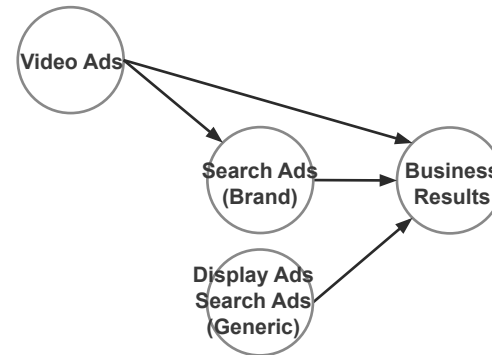
Click route



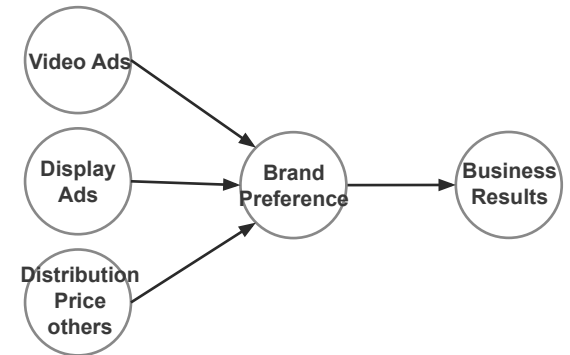
Click route + Search route



Click + Search + Offline route



Offline route*2



- "Click route" consists only of causal relationships that result in online CVs primarily through pay-per-click performance media placements, such as search and display ads

- "Search route" is added to "Click route"
- "Search route" is causal relationships in which the number of searches is changed by cognitive advertising such as TV and digital video

- Offline route is added to "Online type"
- "Offline route" indicates a causal relationship in which store visits occur directly from cognitive advertising, such as TV and digital video ads.

- "Offline type" indicates a causal relationship in which sales fluctuate due to changes in brand preference caused by cognitive advertising such as TV and digital video ads

*1: Although it is originally necessary to consider the volume of brand searches, it is omitted here for the sake of simplicity.

- The simulation study compares the generated simulation data (true data) with the amount of contributions estimated by MMM when the model structure that generates the simulation data and the model structure hypothesis applied to MMM are consistent (scenario 1) and when they are not (scenario 2).
- In Scenario 2, when conducting MMM, a single-layer model structure was adopted, which simply predicts the objective variable by the amount of ad placements, including the search route, which is originally a two-layer structure, unlike the structure used when generating the simulation data.

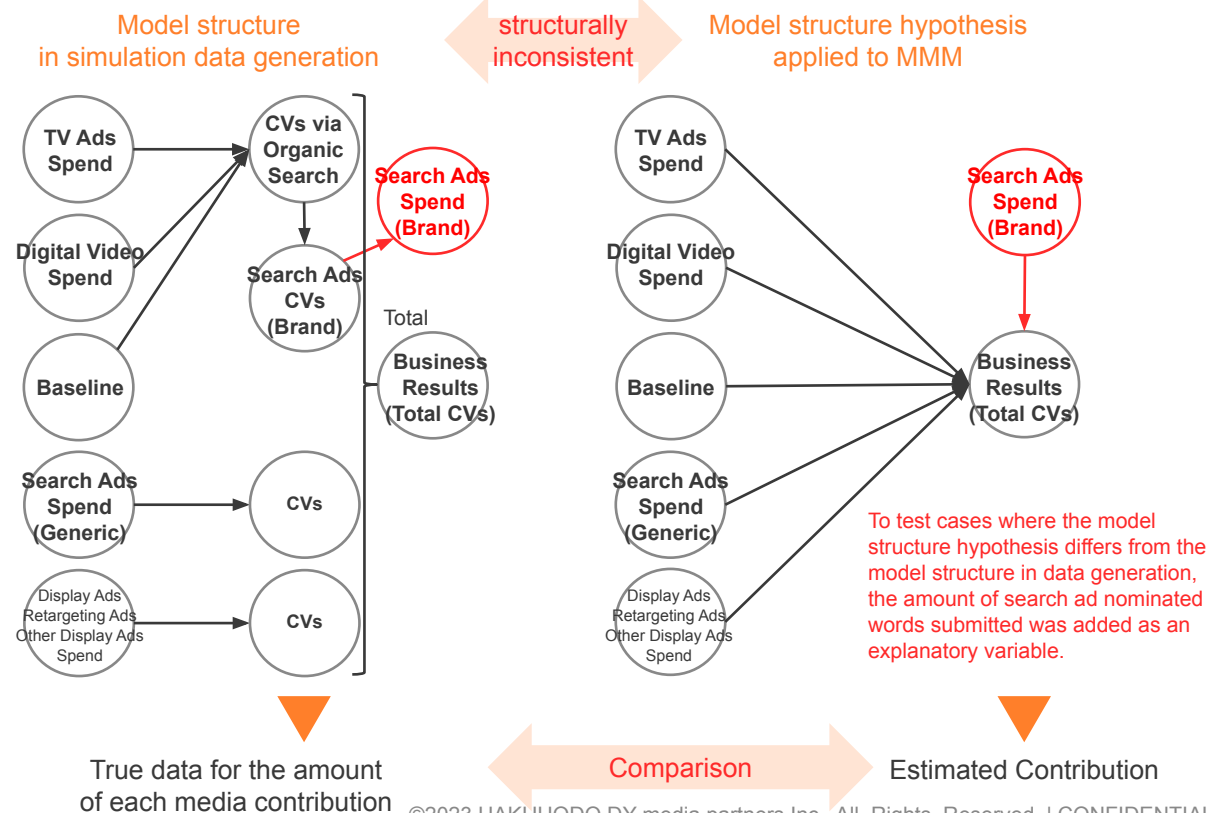
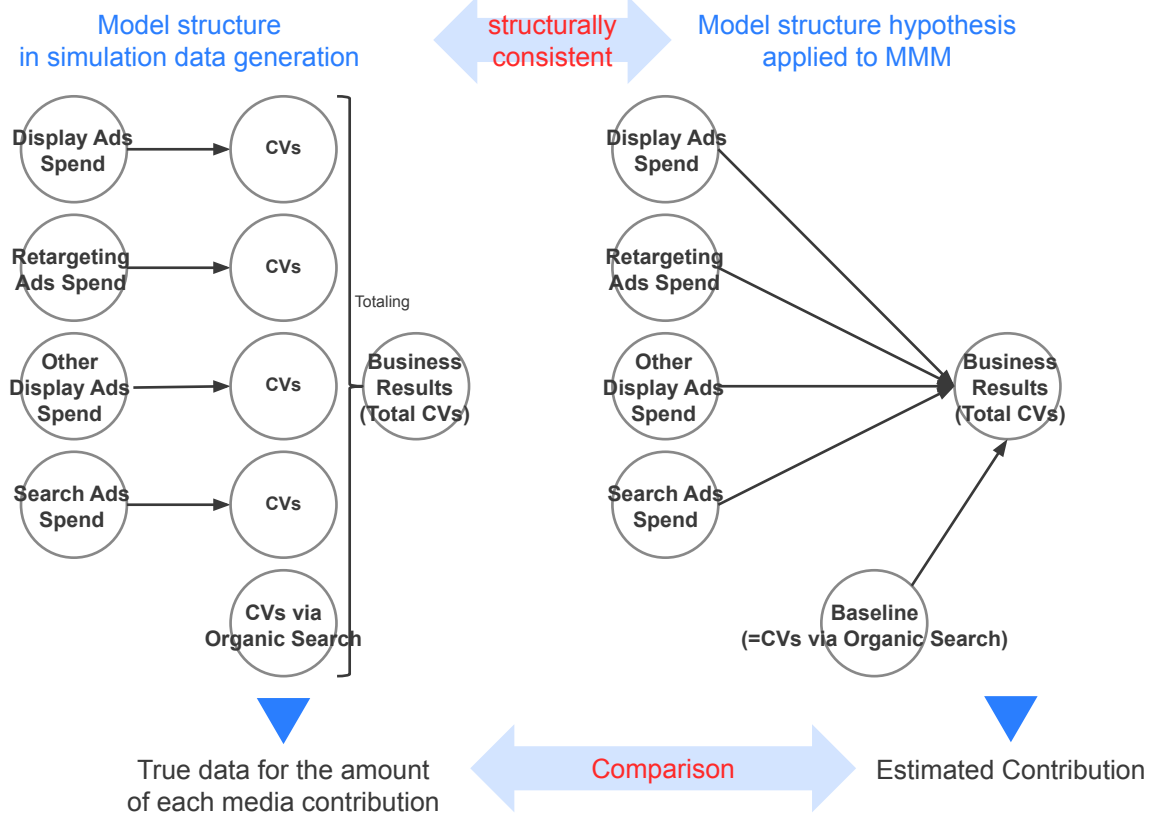
Legend
 → indicates the direction of causality

Each type is a DAG (acyclic directed graph) and assumes the following

- Each element is a set of variables summarized by media characteristics, which must be decomposed appropriately when conducting MMM.
- It is assumed that there is no time dependence in causality.
- The confounding among unobservables and variables is omitted, and its problems are discussed later.

Scenario 1: Verification of click route

Scenario 2: Verification of click + search route



- In the case of “Online-type click route”, a simple data generation process can be assumed in which results such as the number of clicks and CVs vary depending on the amount of ad spend. The relationship between media is assumed to be independent.
- Simulation data on the number of CVs for each media is generated and used as the data for the amount of contribution by each media. The total CVs are then used as the objective variable of MMM. For the number of CVs via organic search, which is independent of ad spend, simulation data that takes into account seasonality and trends was separately prepared.

Scenario 1: Verification of click route

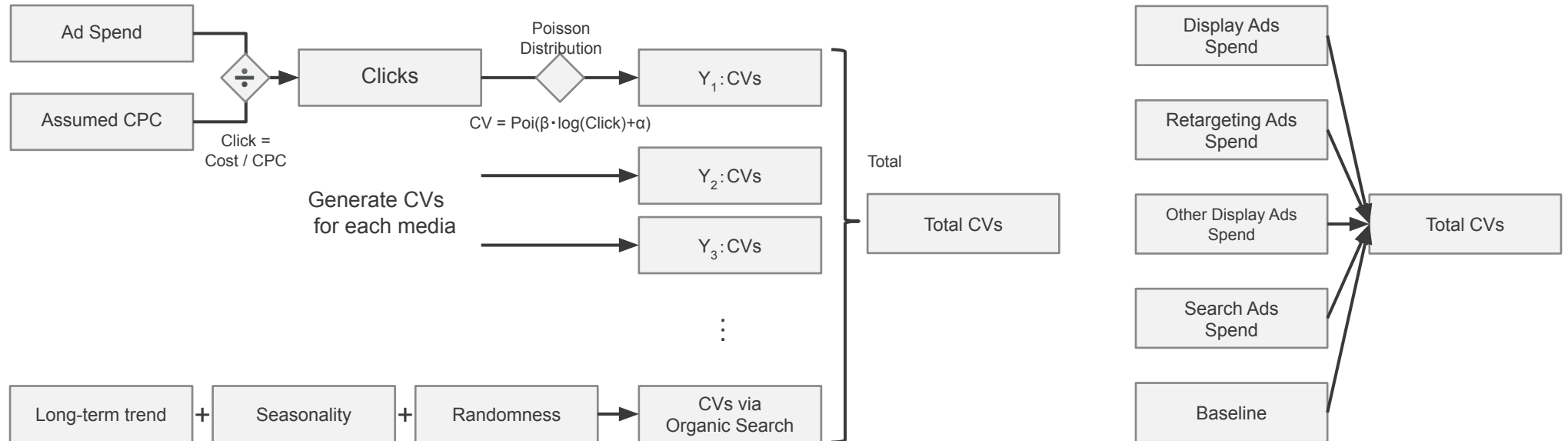
Model structure in simulation data generation

Comparison

Model structure hypothesis applied to MMM

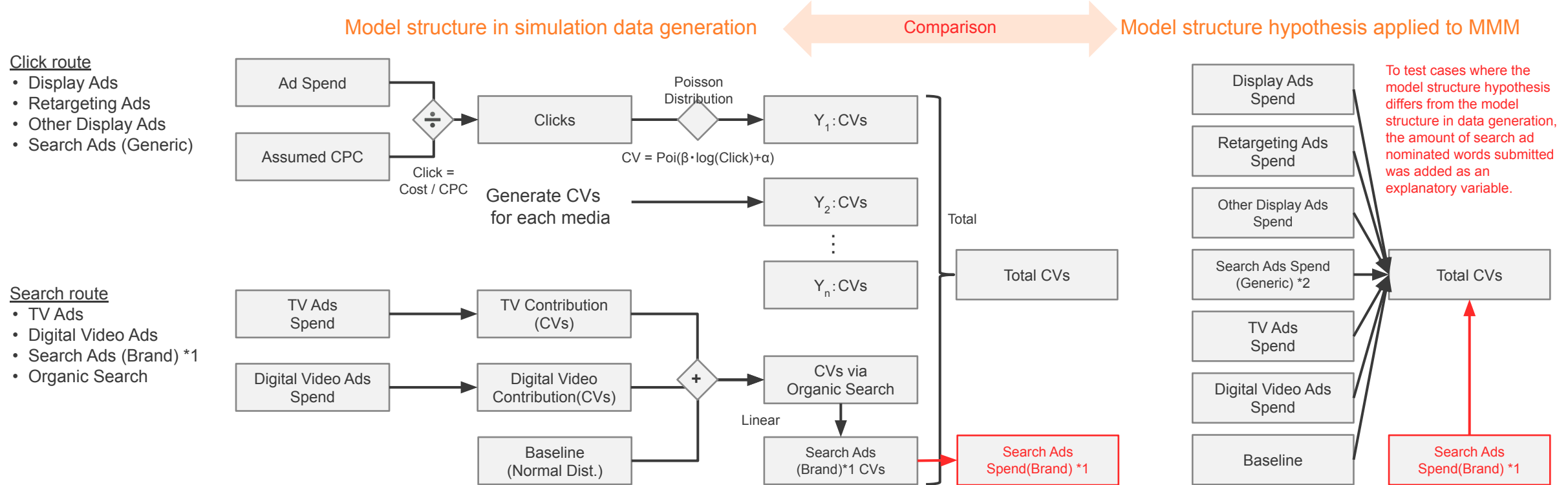
Click route

- Display Ads
- Retargeting Ads
- Other Display Ads
- Search Ads



- “Online-type search route” was assumed to be a two-layer data generation process in which the number of CVs via organic search and the number of Search Ads CVs(Brand) varied depending on the amount of TV ad and digital video ads spends. “Click route” is the same as in scenario 1.
- When conducting MMM, a single-layer model was applied with total CVs as the objective variable and the amount of ads spend for each media as the explanatory variable. In other words, scenario 2 is a case in which the model structures at the time of data generation and MMM verification do not match.
 - In the data generation process, the amount of Search Ads Spend(Brand) is determined dependent on the number of CVs, so it should not be added to the explanatory variables, but it was intentionally added to the explanatory variables in order to verify the impact of the incorrect model structure.

Scenario 2: Verification of click + search route



*1: Search Ads(Brand) refer to branded keyword campaigns among search ads. The number of CVs via Search Ads(Brand) is generated from the amount of TV ads and digital video ads spends, and the amount of placements is dependent on the number of CVs.

*2: Search Ads(Generic) refer to search ads other than brand keyword campaigns such as category keywords.

- In the simulation study, a simple single-layer model (Lightweight MMM) is applied to the simulated data generated in scenarios 1 and 2 to compare the estimated contribution of each media (number of CVs) by MMM with the data from the simulated data.

Single-layer model *1

$$Revenue_t = b + \sum_m \beta_m * Hill(Adstock(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

Assumption on the model structure*1

- As an additive linear structure is assumed in the model, media variables are independent each other. (e.g., no dependency between TV/YouTube and search ads volume)
- Organic brand query volume is not included as a variable
- No non-media variable is utilised in this model (i.e., trend and seasonality are modeled with a concave curve and sinusoidal function respectively.)

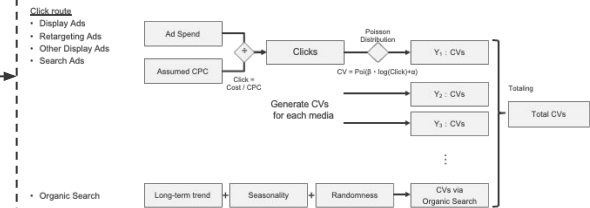
Assumption on the parameters*1

- Hill function (p.46) and geometric adstock (p.49) are assumed for media variables (using "hill_adstock" in Lightweight MMM).
- Use default settings in LightweightMMM for latent unobserved trends and seasonality, and for the prior distribution of each parameter.

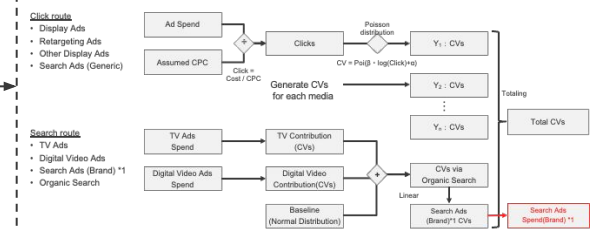
Apply the model

Simulation Data

Simulation scenario 1 "click route"

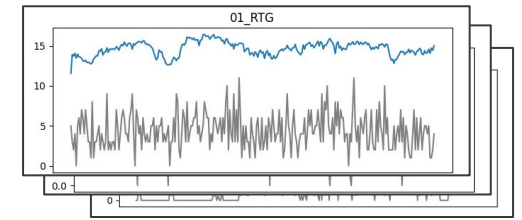


Simulation scenario 2 "click route" + "search rule"

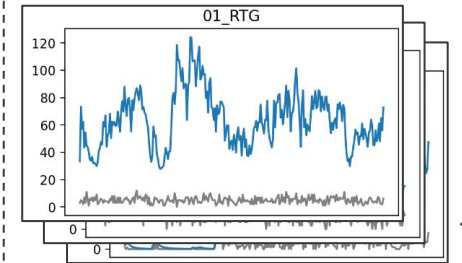


Comparison

True conversion contribution vs. the model's estimation



True conversion contribution vs. the model's estimation



*1: The tested single layer model refers to Lightweight MMM, which is Google unofficial open source as of Sep 2023.

3.3. Simulation Study

- The simulation results of “Click route” are shown below.
- In scenario 1, the true model structure at the time of simulation data generation matches the model structure of the MMM to be applied.

Single-layer model *1

$$Revenue_t = b + \sum_m \beta_m * Hill (Adstock(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

Assumption on the model structure*1

- As an additive linear structure is assumed in the model, media variables are independent each other. (e.g., no dependency between TV/YouTube and search ads volume)
- Organic brand query volume is not included as a variable
- No non-media variable is utilised in this model (i.e., trend and seasonality are modeled with a concave curve and sinusoidal function respectively.)

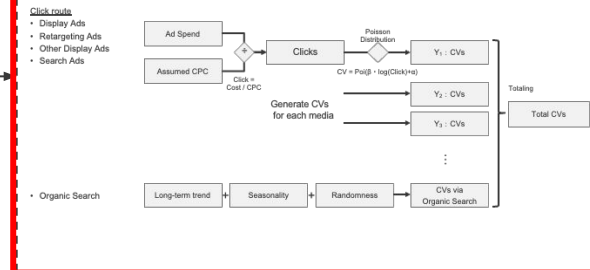
Assumption on the parameters*1

- Hill function (p.46) and geometric adstock (p.49) are assumed for media variables (using “hill_adstock” in Lightweight MMM).
- Use default settings in LightweightMMM for latent unobserved trends and seasonality, and for the prior distribution of each parameter.

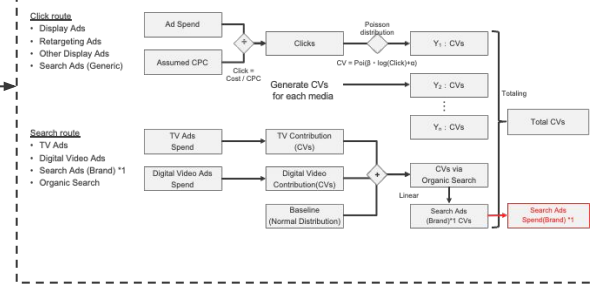
Apply the model

Simulation Data

Simulation scenario 1 “click route”

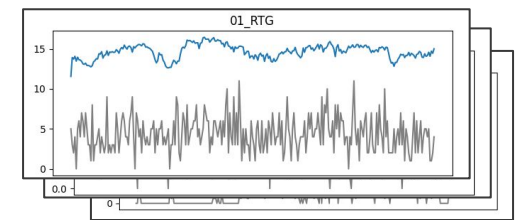


Simulation scenario 2 “click route” + “search rule”

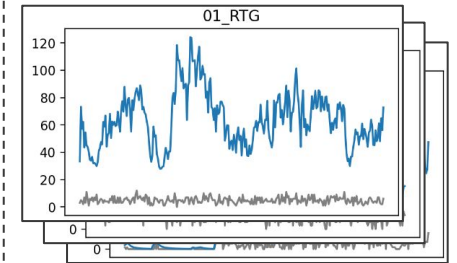


Comparison

True conversion contribution vs. the model's estimation



True conversion contribution vs. the model's estimation

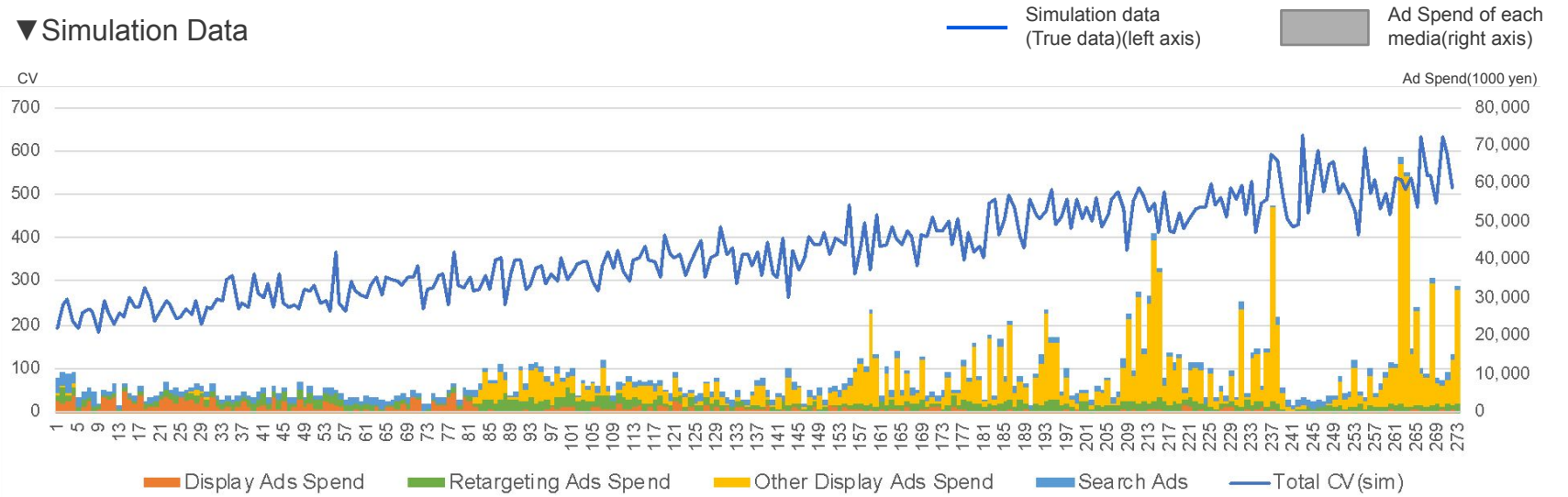


*1: The tested single layer model refers to Lightweight MMM, which is Google unofficial open source as of Sep 2023.

Scenario 1: Verification of Click Route | Pattern A. Baseline 95% Simulation Data

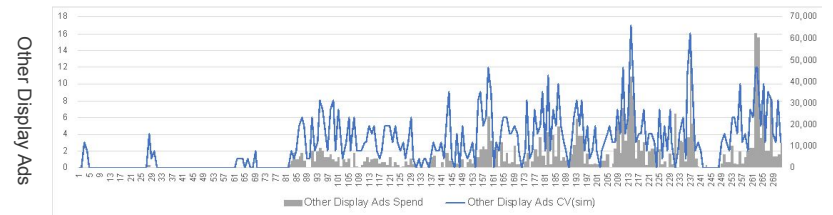
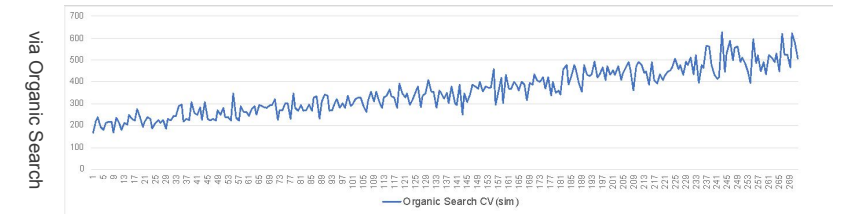
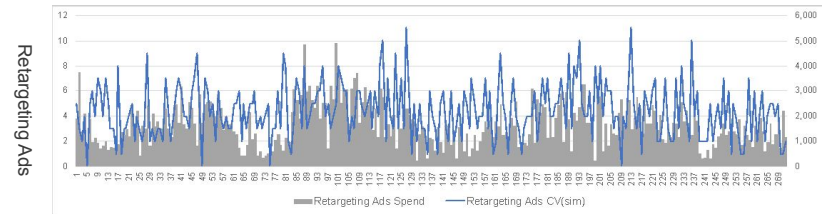
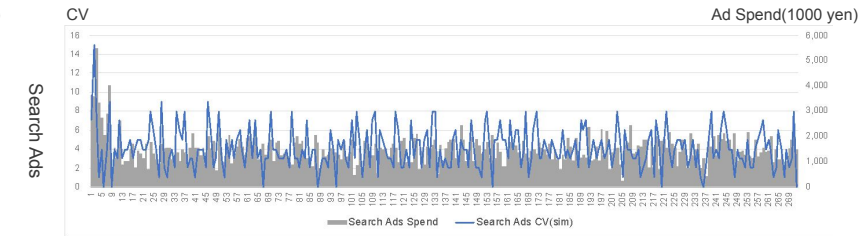
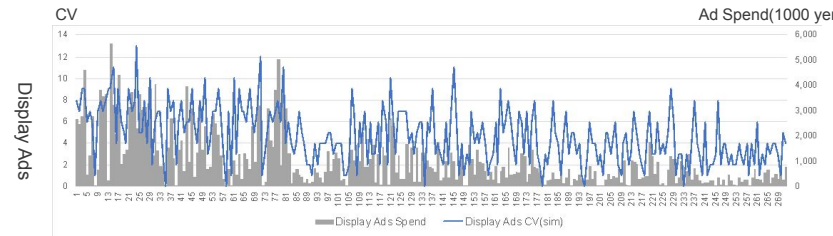
The simulation data generation method described previously is used to generate the true data so that the percentage of the data via organic search (= baseline) is 95% of the objective variable (total CV).

▼ Simulation Data



Legend

- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.

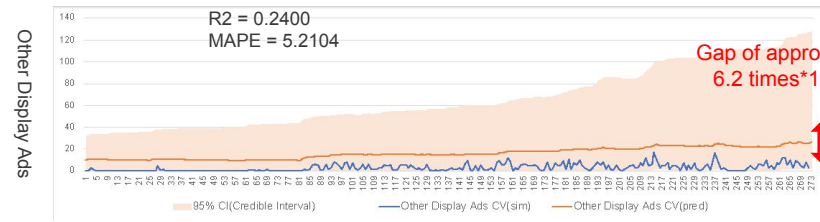
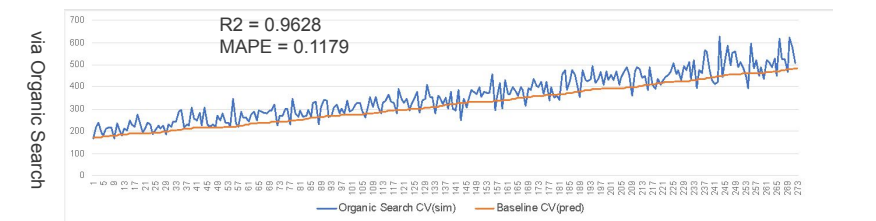
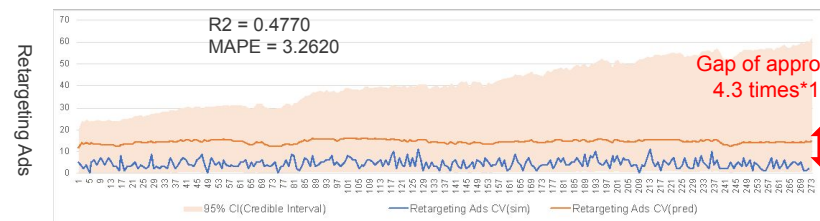
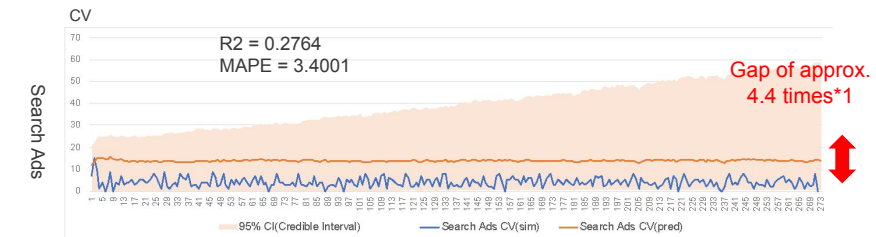
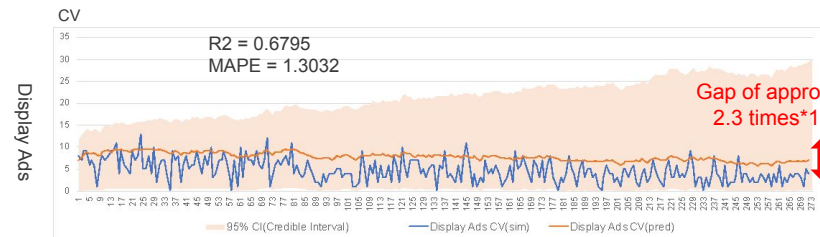
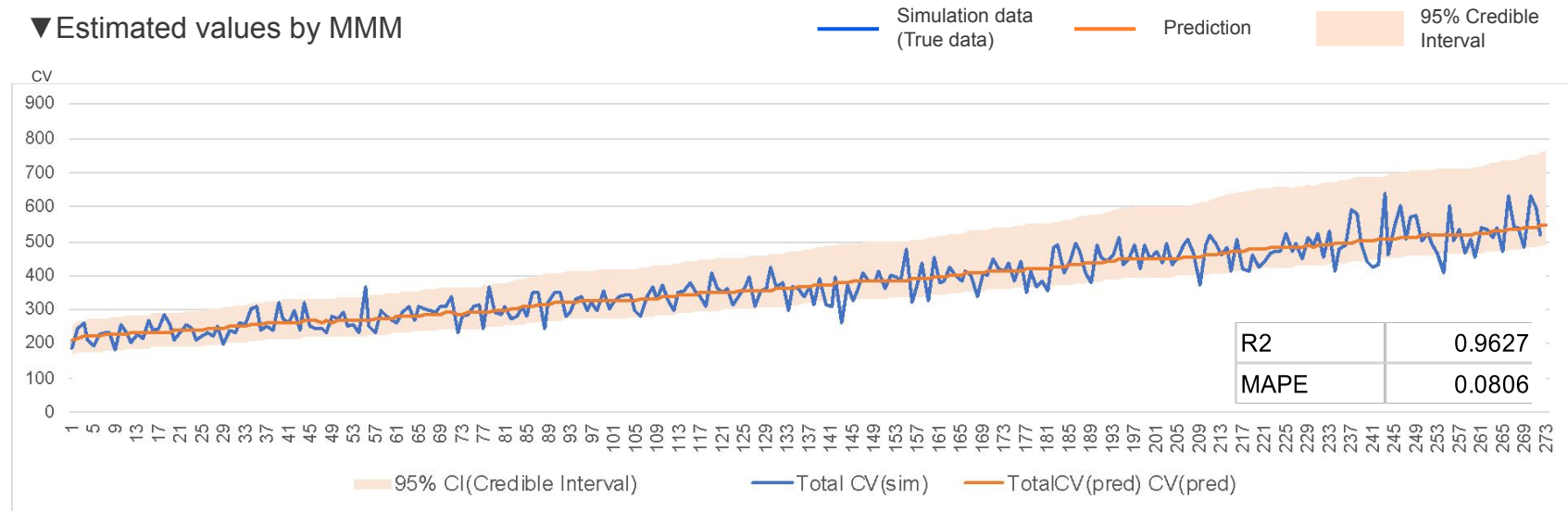


Scenario 1: Verification of Click Route | Pattern A. Baseline 95% Analysis Results

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.96$ and $MAPE = 0.08$. The simulation data for the objective variable and each media fell within the 95% credit interval of the estimates by MMM.

However, the estimated contribution of each media has errors with the simulated data, ranging from about 1.3 to 5.2 (2.3 to 6.2 times) in terms of MAPE (Mean Absolute Percentage Error).

▼ Estimated values by MMM



Legend

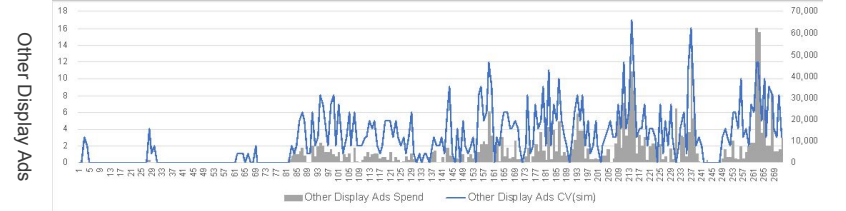
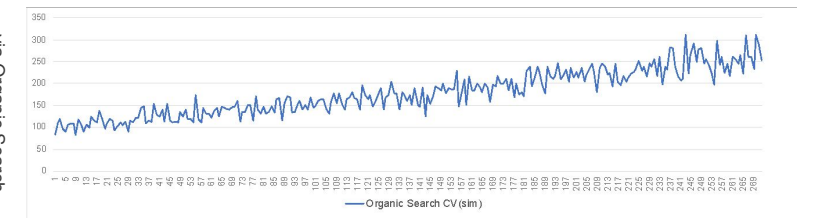
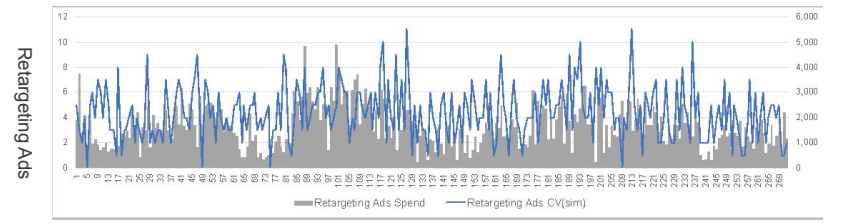
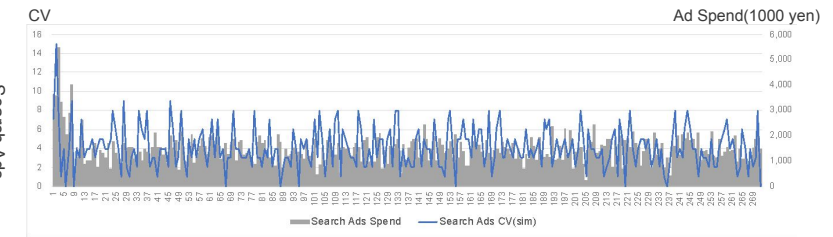
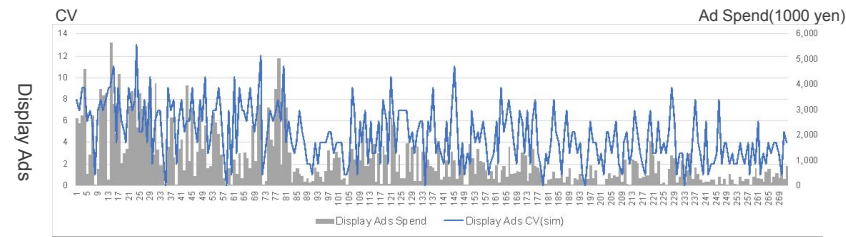
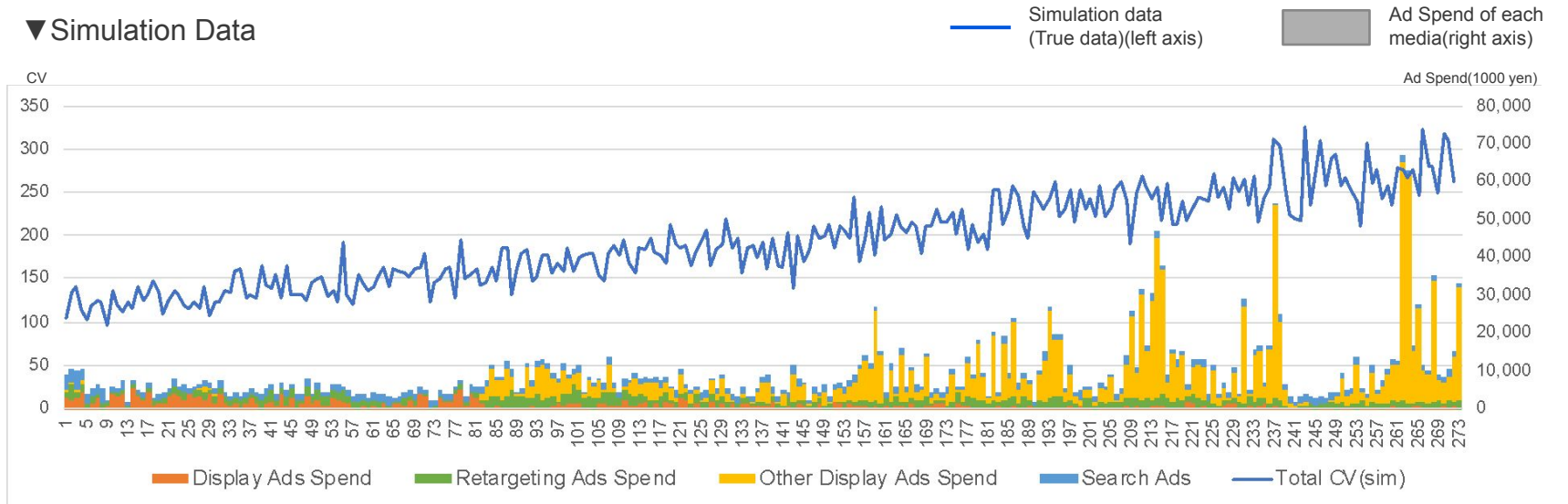
- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.

*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

Scenario 1: Verification of Click Route | Pattern B. Baseline 90% Simulation Data

The percentage via natural search (= baseline) was then reduced to 90% of the objective variable (total CV) to generate the true data.

▼ Simulation Data



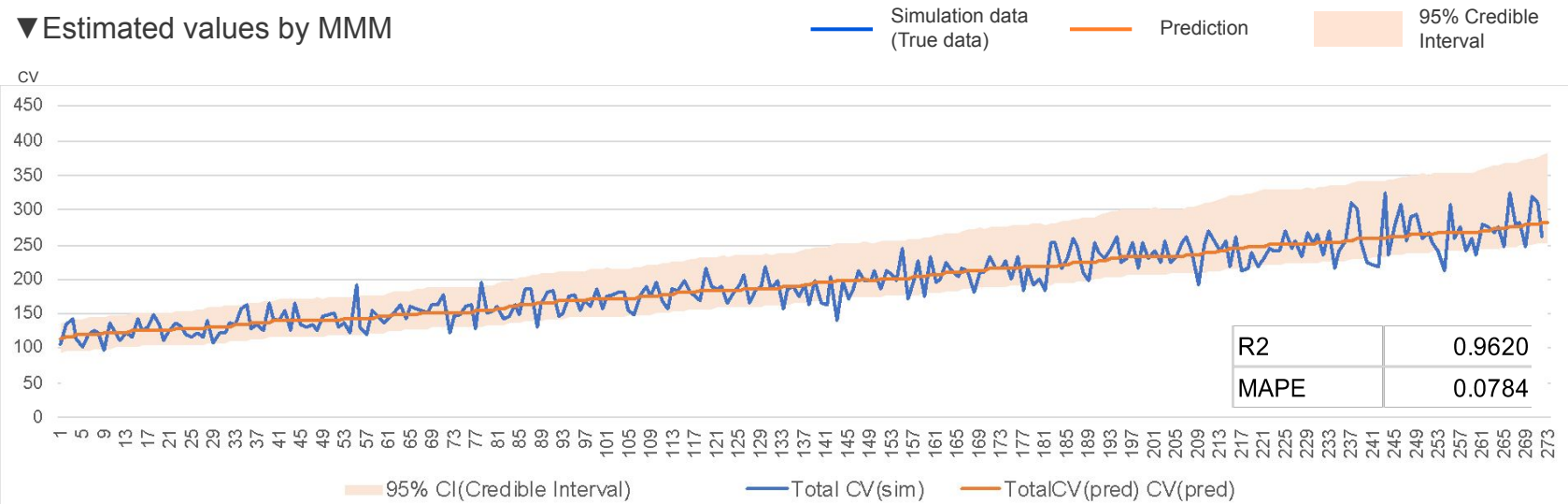
- Legend**
- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
 - Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
 - Via organic search: Number of CVs other than via the above ads.
 - (sim) and (pred) represent simulated data and MMM estimates, respectively.

Scenario 1: Verification of Click Route | Pattern B. Baseline 90% Analysis Results

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.96$ and $MAPE = 0.078$. The simulation data for the objective variable and each media fall within the 95% credit interval of the estimates from the MMM.

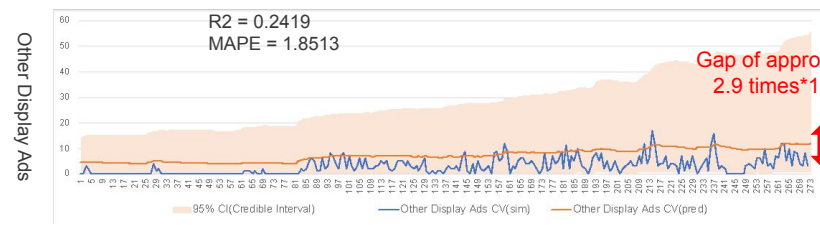
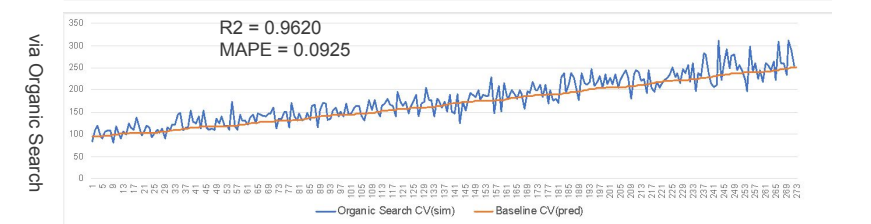
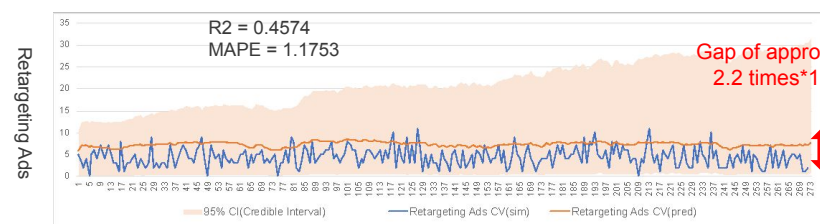
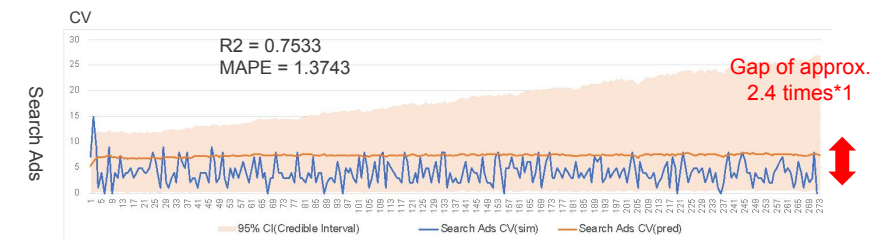
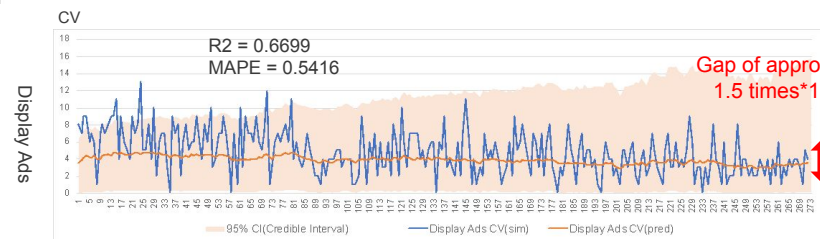
The estimated contribution of each media has errors with the simulated data, ranging from about 0.5 to 1.9 (1.5 to 2.9 times) in terms of MAPE (Mean Absolute Percentage Error). This is a smaller error than the simulation of Pattern A. Baseline 95%.

▼ Estimated values by MMM



Legend

- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.

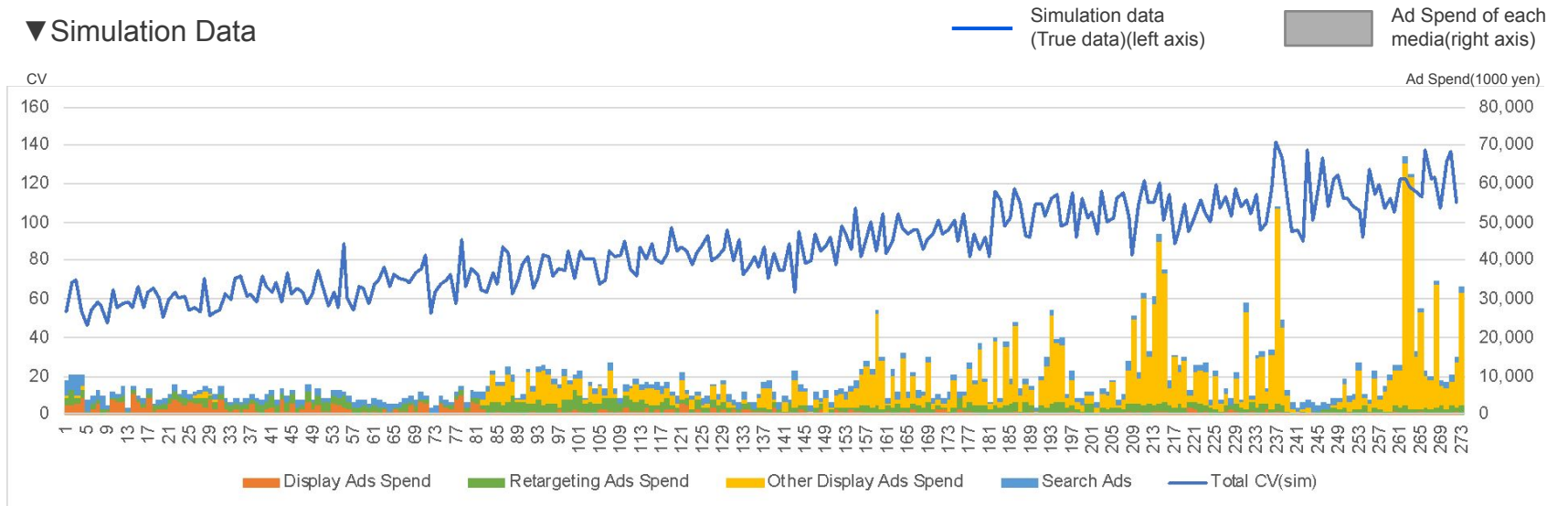


*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

Scenario 1: Verification of Click Route | Pattern C. Baseline 80% Simulation Data

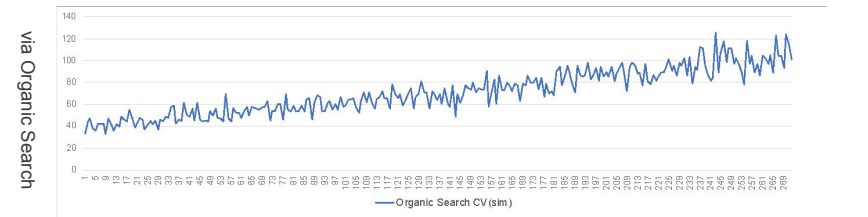
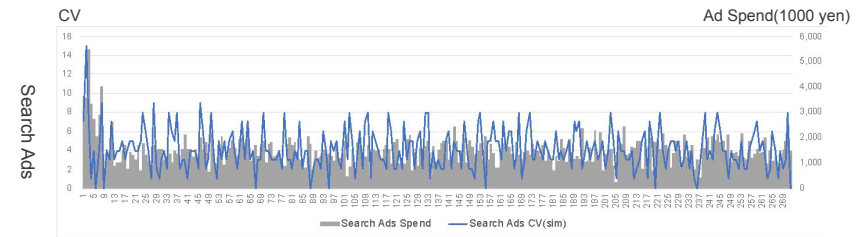
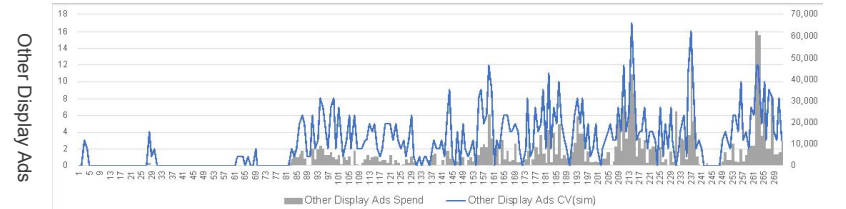
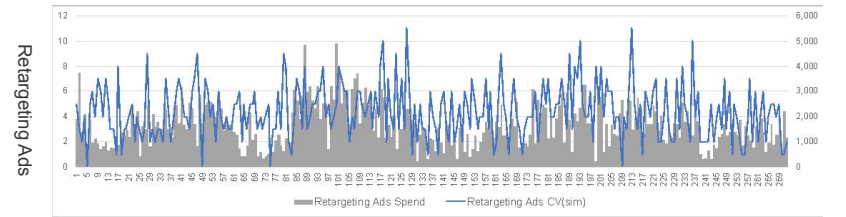
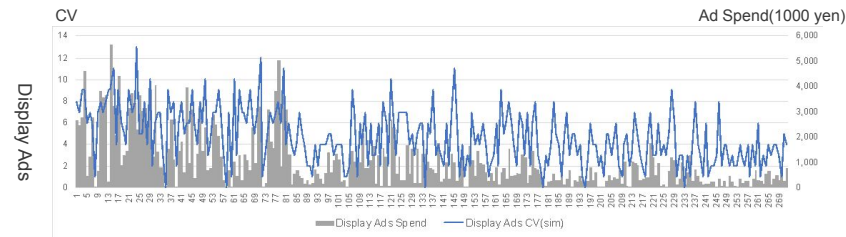
The percentage via organic search (= baseline) was then reduced to 80% of the objective variable (total CV) to generate the true data.

▼ Simulation Data



Legend

- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.

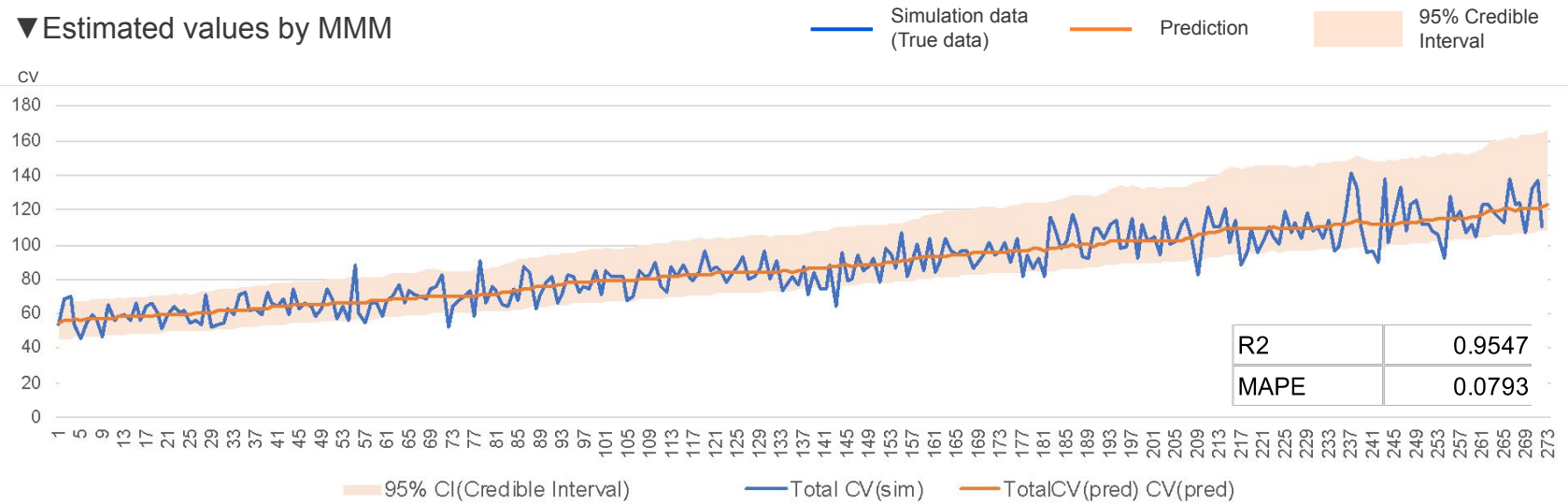


Scenario 1: Verification of Click Route | Pattern C. Baseline 80% Analysis Results

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.95$ and $MAPE = 0.08$. The objective variable and the simulated data for each media other than display ads fall within the 95% credible interval of the estimated values by MMM.

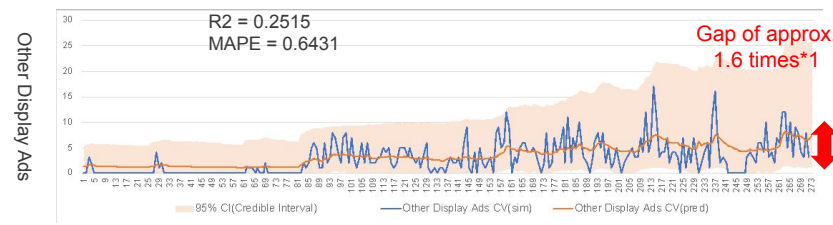
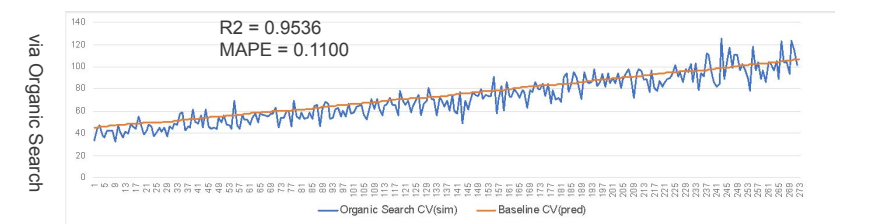
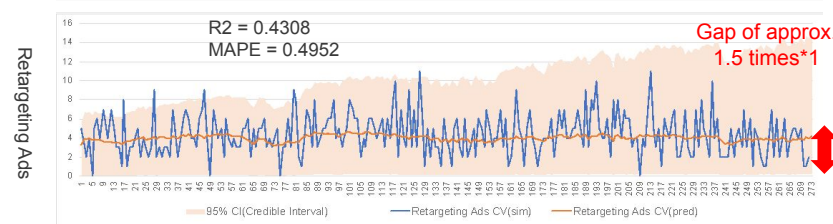
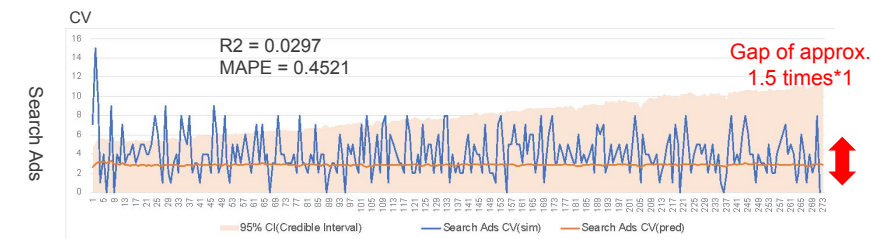
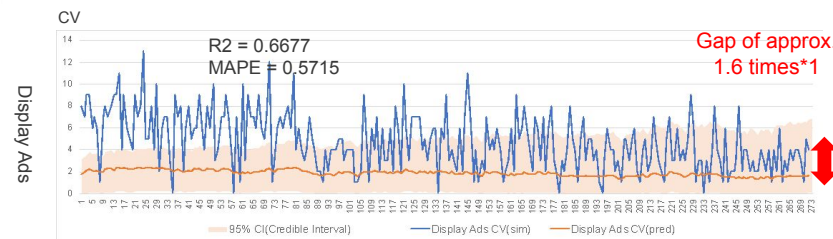
The error between the estimated contribution of each media and the simulated data is smaller than Pattern B. Baseline 90%, and is about 0.5 to 0.6 (1.5 to 1.6 times) in MAPE (Mean Absolute Percentage Error).

▼ Estimated values by MMM



Legend

- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.

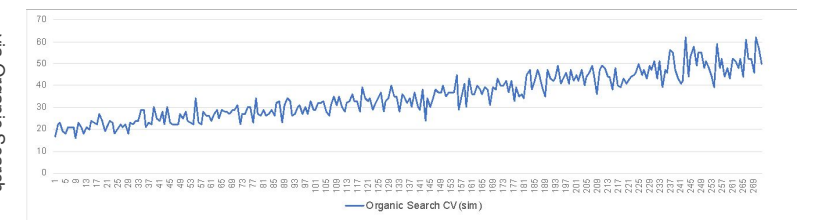
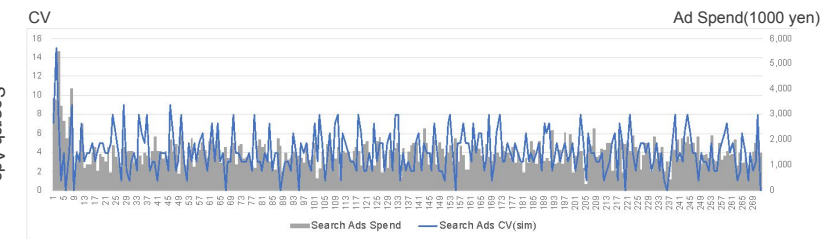
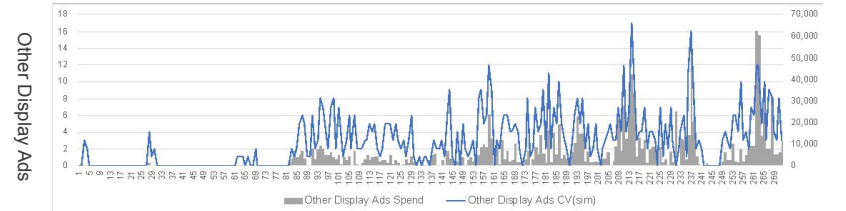
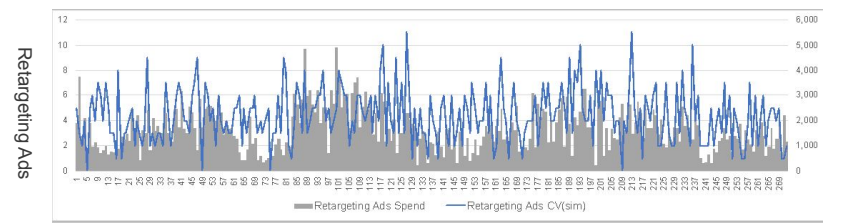
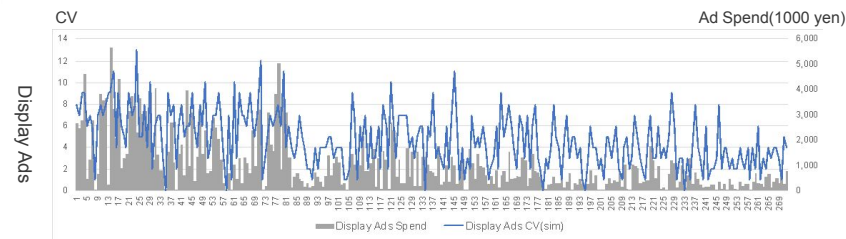
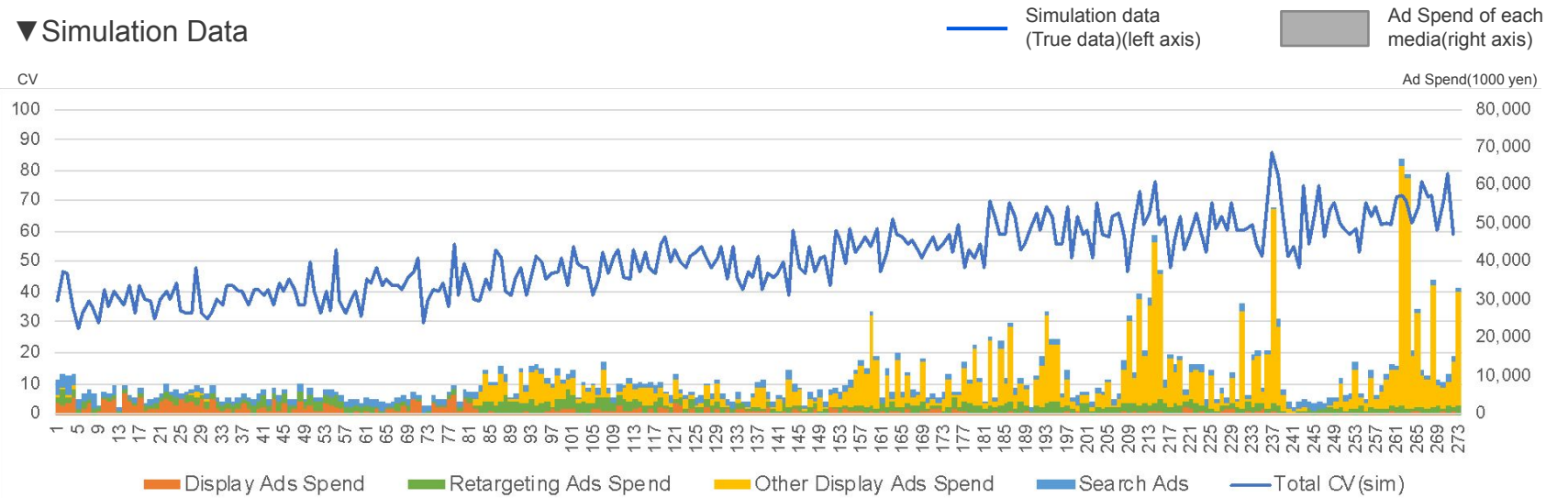


*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

Scenario 1: Verification of click route | Pattern D. Baseline 70% Simulation Data

The percentage via organic search (= baseline) was then reduced to 70% of the objective variable (total CV) to generate the true data.

▼ Simulation Data

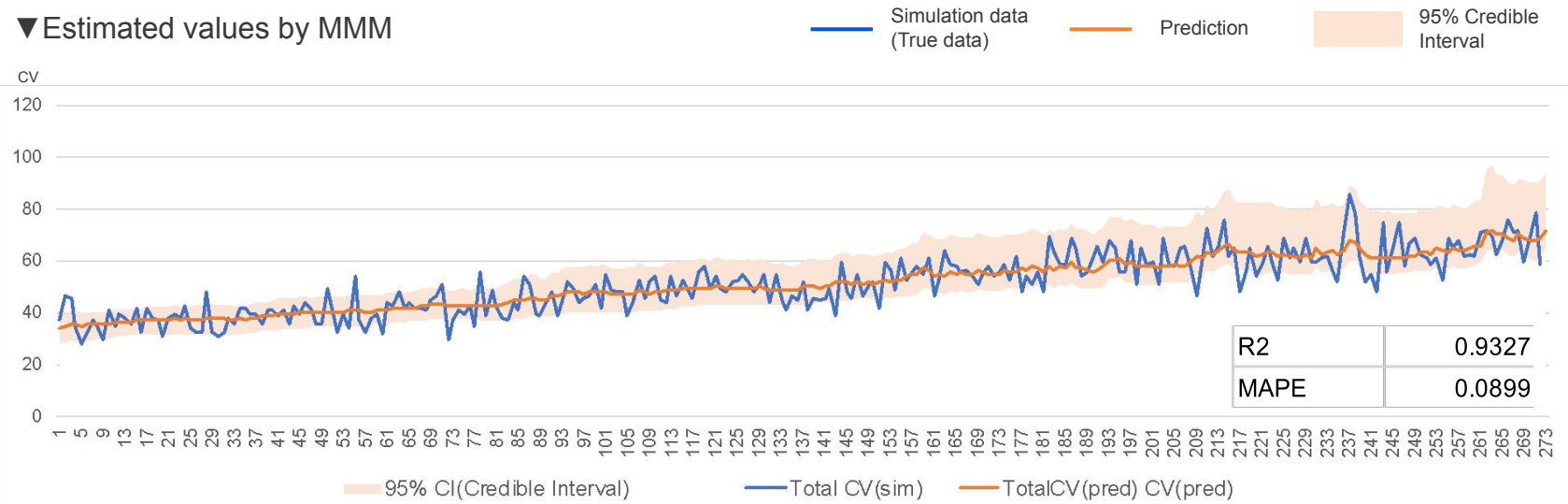


- Legend**
- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
 - Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
 - Via organic search: Number of CVs other than via the above ads.
 - (sim) and (pred) represent simulated data and MMM estimates, respectively.

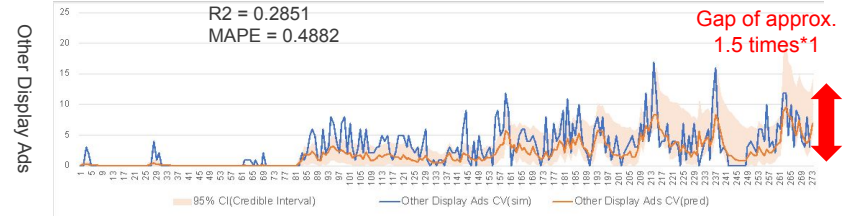
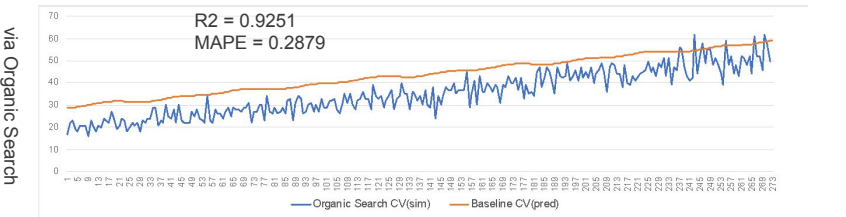
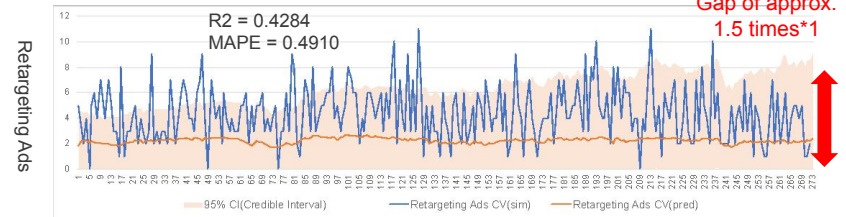
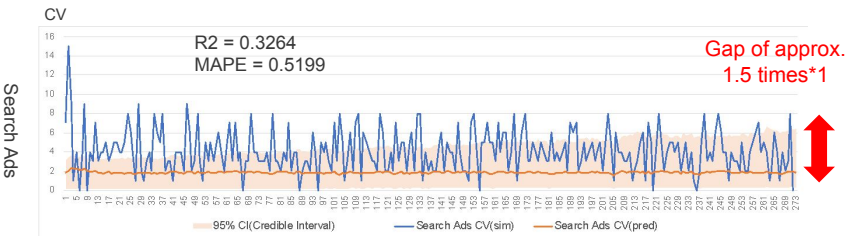
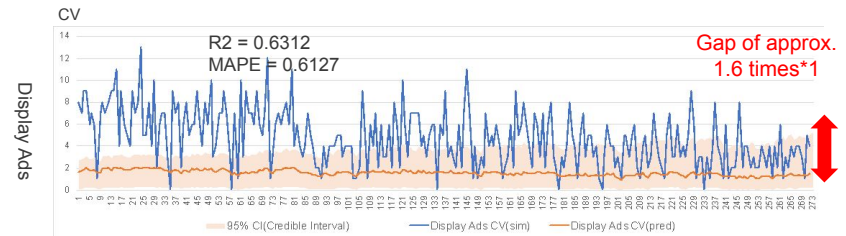
Scenario 1: Verification of Click Route | Pattern D. Baseline 70% Analysis Results

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.93$ and $MAPE = 0.09$. While the simulation data of the objective variable fell within the 95% credible interval of the estimates by MMM, the simulation data of each media did not fall within the 95% credible interval. Only display ads captured the trend. The error between the estimated contribution of each media and the simulated data has narrowed, with MAPE (Mean Absolute Percentage of Error) ranging from 0.5 to 0.6 (1.5 to 1.6 times).

▼ Estimated values by MMM



- Legend**
- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
 - Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
 - Via organic search: Number of CVs other than via the above ads.
 - (sim) and (pred) represent simulated data and MMM estimates, respectively.

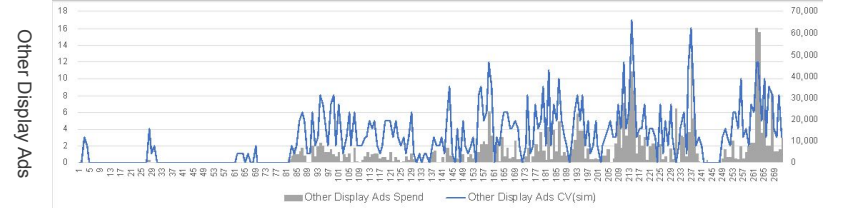
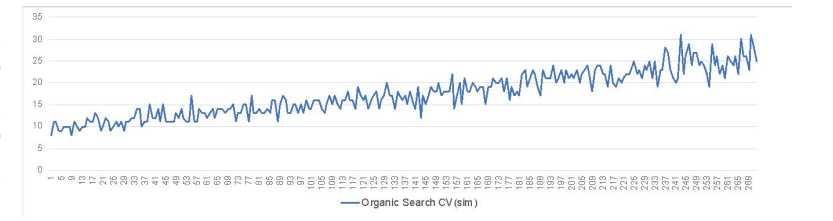
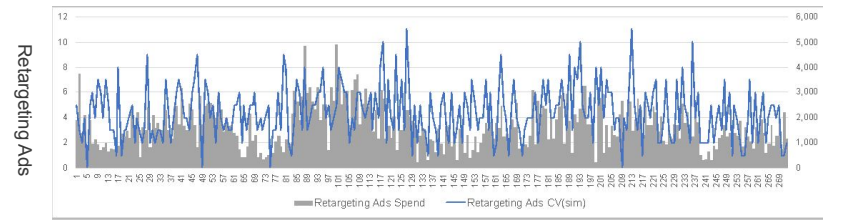
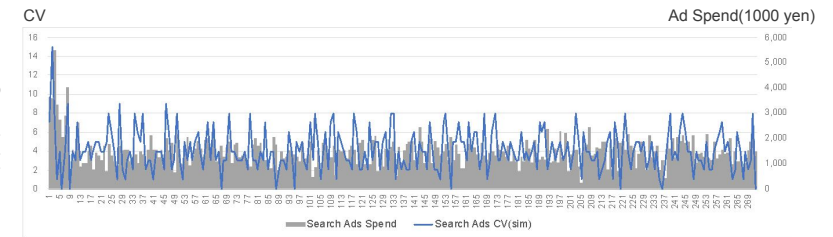
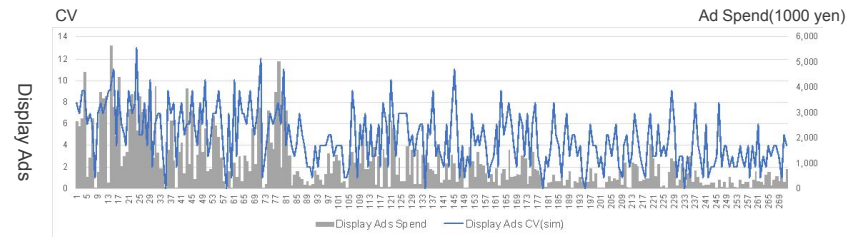
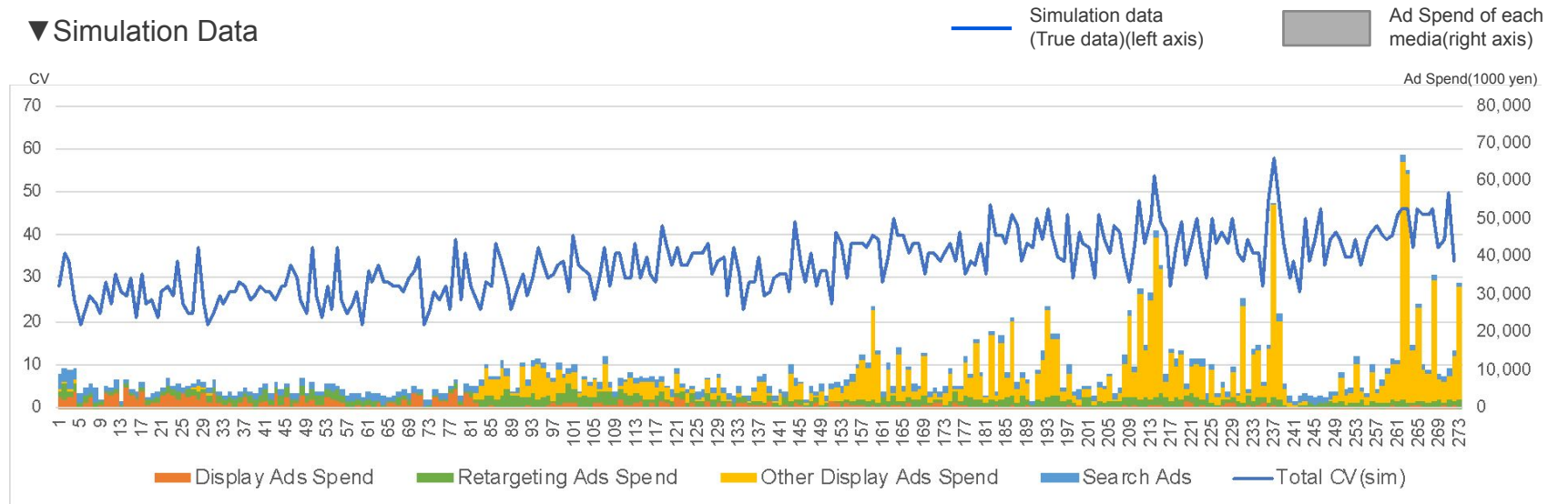


*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

Scenario 1: Verification of click route | Pattern E. Baseline 50% Simulation Data

The percentage via organic search (= baseline) was then reduced to 50% of the objective variable (total CV) to generate the true data.

▼ Simulation Data



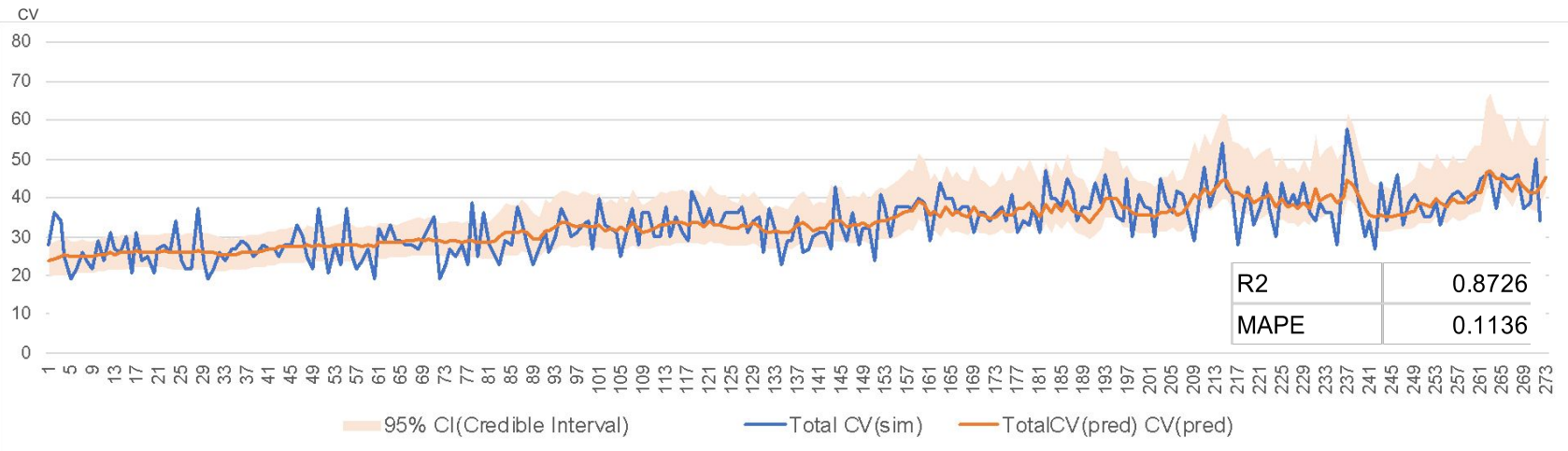
- Legend**
- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
 - Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
 - Via organic search: Number of CVs other than via the above ads.
 - (sim) and (pred) represent simulated data and MMM estimates, respectively.

Scenario 1: Verification of Click Route | Pattern E. Analysis of 50% Baseline

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.87$ and $MAPE = 0.11$. While the simulated data of the objective variable fell within the 95% credible interval of the estimate by MMM, the contribution of each media did not fall within the 95% credible interval. Only display ads captured the trend.

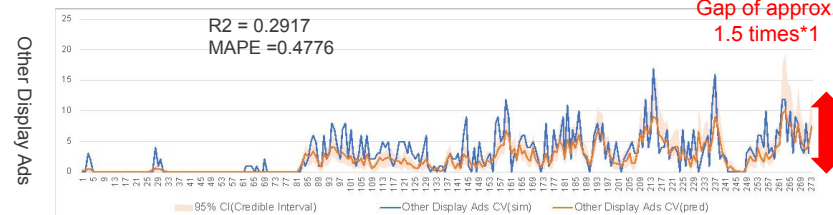
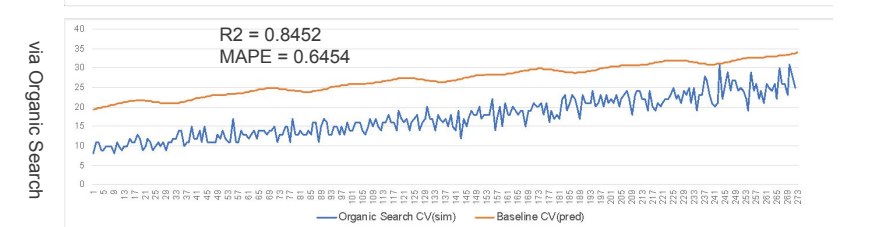
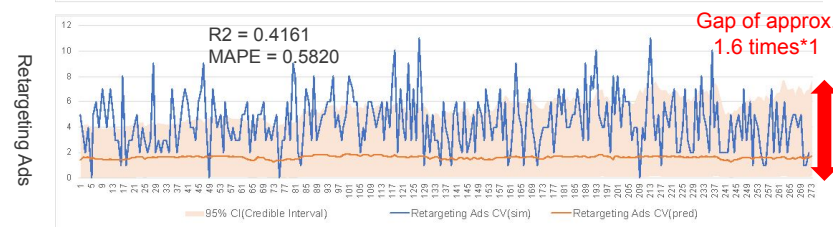
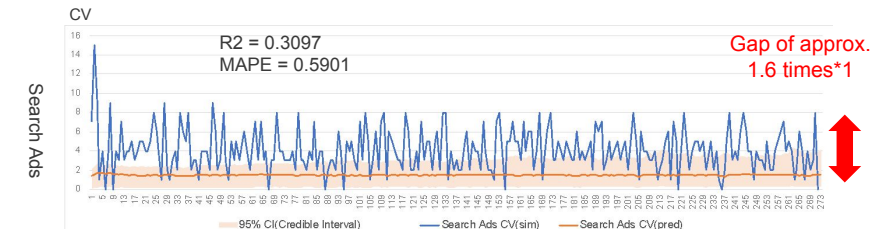
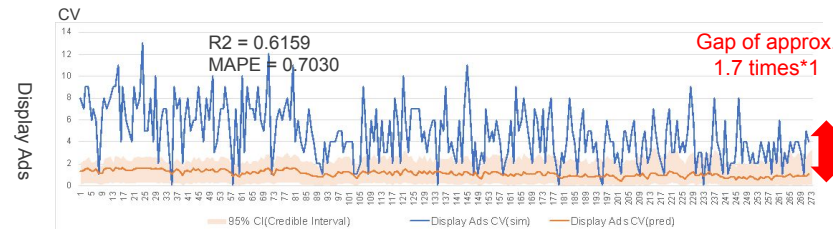
The error between the estimated contribution of each media and the simulated data has narrowed to approximately 0.5 to 0.7 (1.5 to 1.7 times) in terms of MAPE (Mean Absolute Percentage Error).

▼ Estimated values by MMM



Legend

- Display ads: Amount of ad spend and number of CVs for programmatic ads in the still-image format without retargeting
- Retargeting ads: Amount of ad spend and number of CVs for programmatic ads delivered by retargeting in still image format
- Other display ads: Amount of ad spend and number of CVs for premium ads in still image format
- Search ads: Amount of ad spend and number of CVs for search ads. In "Click route", the campaign is assumed to be a general word campaign with no restrictions on search volume, such as brand words.
- Via organic search: Number of CVs other than via the above ads.
- (sim) and (pred) represent simulated data and MMM estimates, respectively.



*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

- The simulation results for the click + search route are shown.
- In scenario 2, the true model structure at the time of simulation data generation does not match the model structure of the MMM to be applied.
 - The search route is generated with a two-layer structure, but the model applied is a single-layer model.

Single-layer model *1

$$Revenue_t = b + \sum_m \beta_m * Hill (Adstock(x_{t,m}, \dots, x_{t-l,m}; L, w_m(l; \alpha_m, \theta_m)), K_m, S_m) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + \epsilon_t$$

Assumption on the model structure*1

- As an additive linear structure is assumed in the model, media variables are independent each other. (e.g., no dependency between TV/YouTube and search ads volume)
- Organic brand query volume is not included as a variable
- No non-media variable is utilised in this model (i.e., trend and seasonality are modeled with a concave curve and sinusoidal function respectively.)

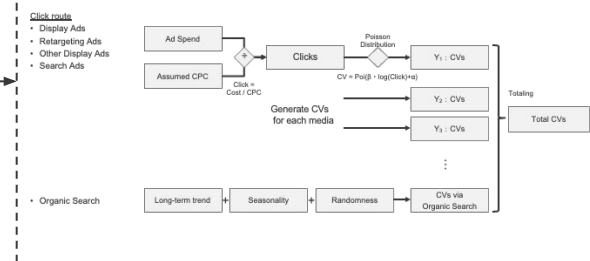
Assumption on the parameters*1

- Hill function (p.46) and geometric adstock (p.49) are assumed for media variables (using "hill_adstock" in Lightweight MMM).
- Use default settings in LightweightMMM for latent unobserved trends and seasonality, and for the prior distribution of each parameter.

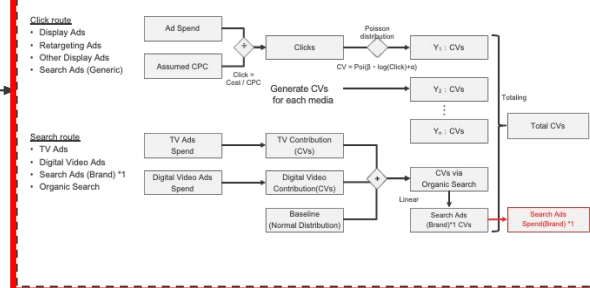
Apply the model

Simulation Data

Simulation scenario 1 "click route"

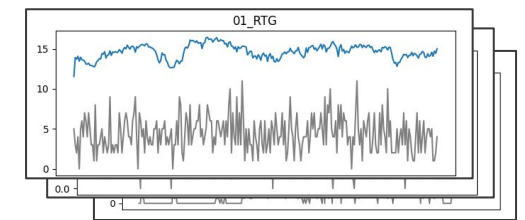


Simulation scenario 2 "click route" + "search rule"

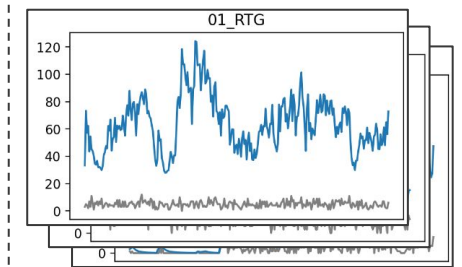


Comparison

True conversion contribution vs. the model's estimation



True conversion contribution vs. the model's estimation

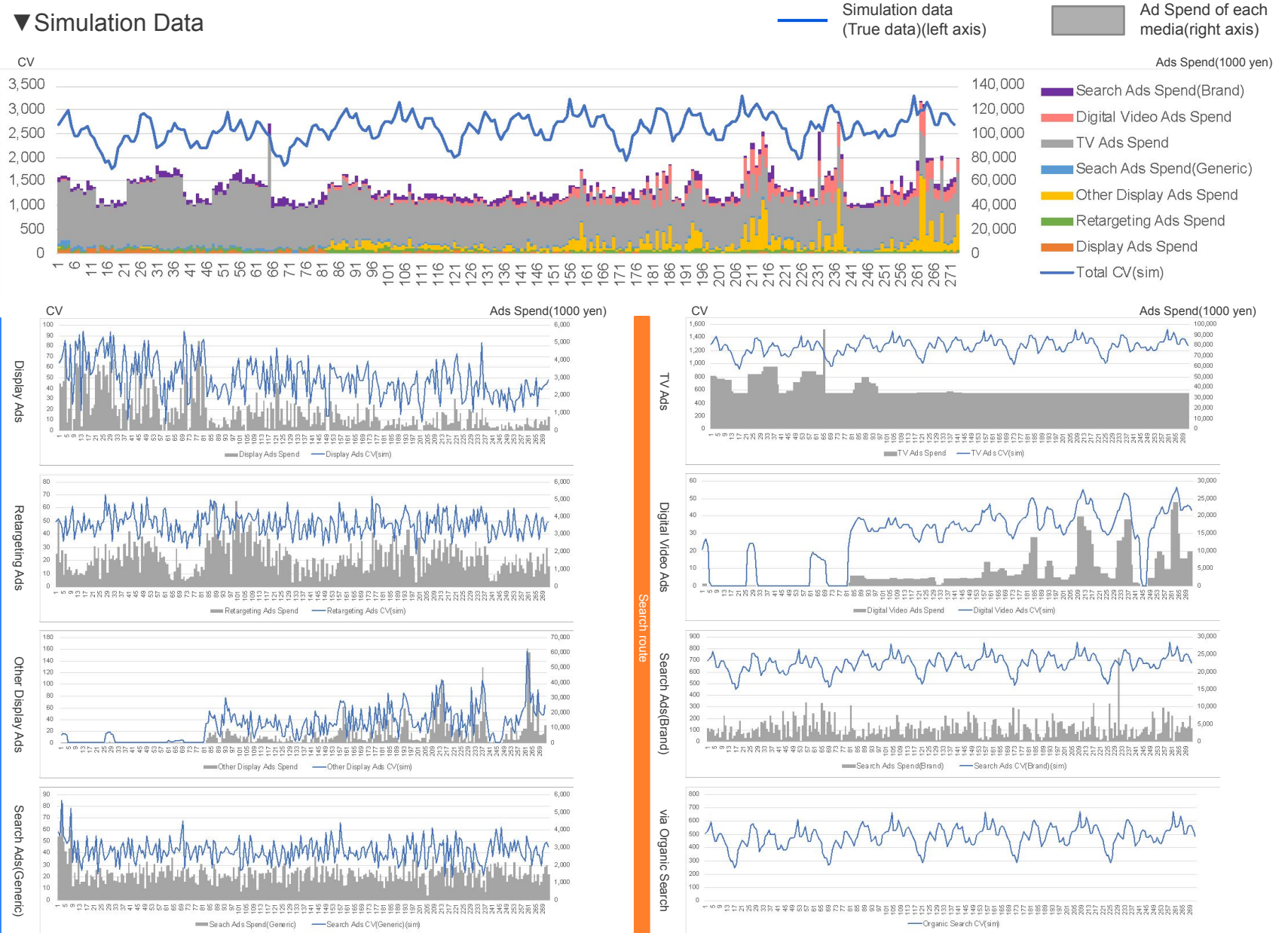


*1: The tested single layer model refers to Lightweight MMM, which is Google unofficial open source as of Sep 2023.

Scenario 2: Verification of click + search route | Simulation data

Using the simulation data generation method for scenario 2 described previously, true data was generated so that search route CVs (via Search Ads(Brand) + Organic Search) accounted for 90% of total CVs. CVs via search ads (brand) and organic search were made to be 50% and 50%. The amount of TV ads spend was simulated as a series of uniform placements depending on the percentage of time placements, with a smaller variance, and a mountain being formed by spot placements. Display, retargeting, other display, and search ads(generic) in the click route were generated independently, as in scenario 1.

- Legend**
- Display Ads: Amount of ads spend and number of CVs for programmatic ads without retargeting in still image format
 - Retargeting Ads: Amount of ads spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other Display Ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search Ads(Generic): Value of ads spend and number of CVs for category keywords and other general keywords campaigns, excluding brand keywords, in search ads.
 - Search Ads(Brand): Among search ads, the number of CVs as a direct effect of brand keyword campaigns. Excluding CVs as indirect effects of TV ads and digital videos.
 - via Organic Search: Number of CVs as a direct effect of organic search. Excluding contributions as indirect effects from TV ads and digital videos.
 - TV ads: The amount of TV ads spend and the sum of CVs contributed as indirect effects to search ads(brand) and organic search
 - Digital video: The amount of digital video ads spend and the sum of CVs contributed as indirect effects to search ads(brand) and organic search.
 - (sim) indicates simulated data = true data, and (pred) indicates the amount estimated by MMM, respectively.



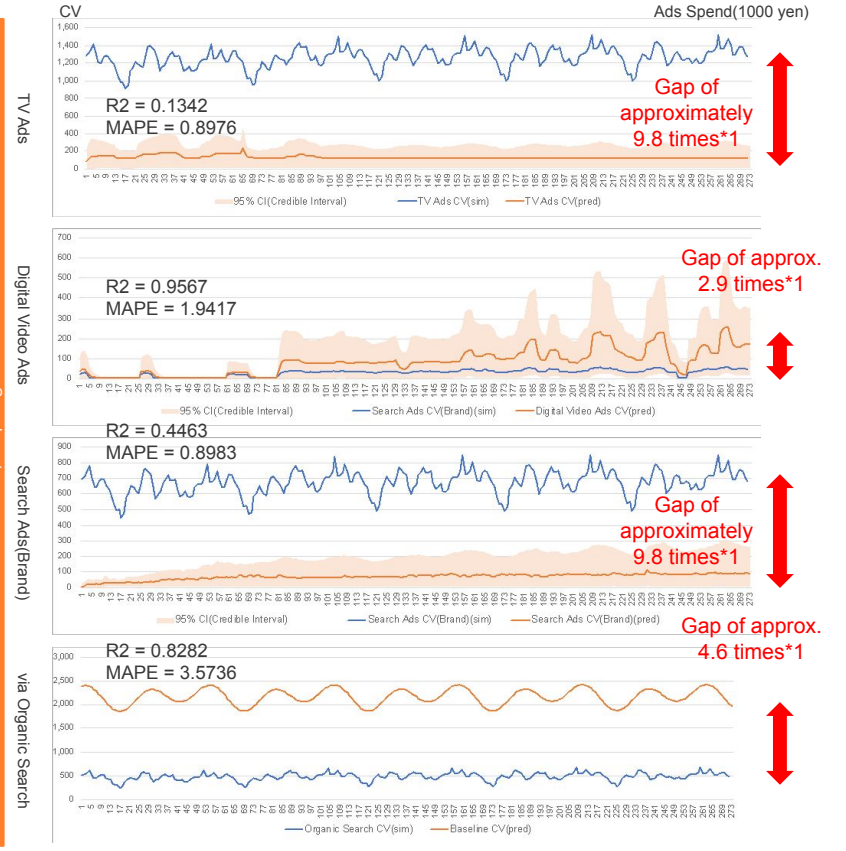
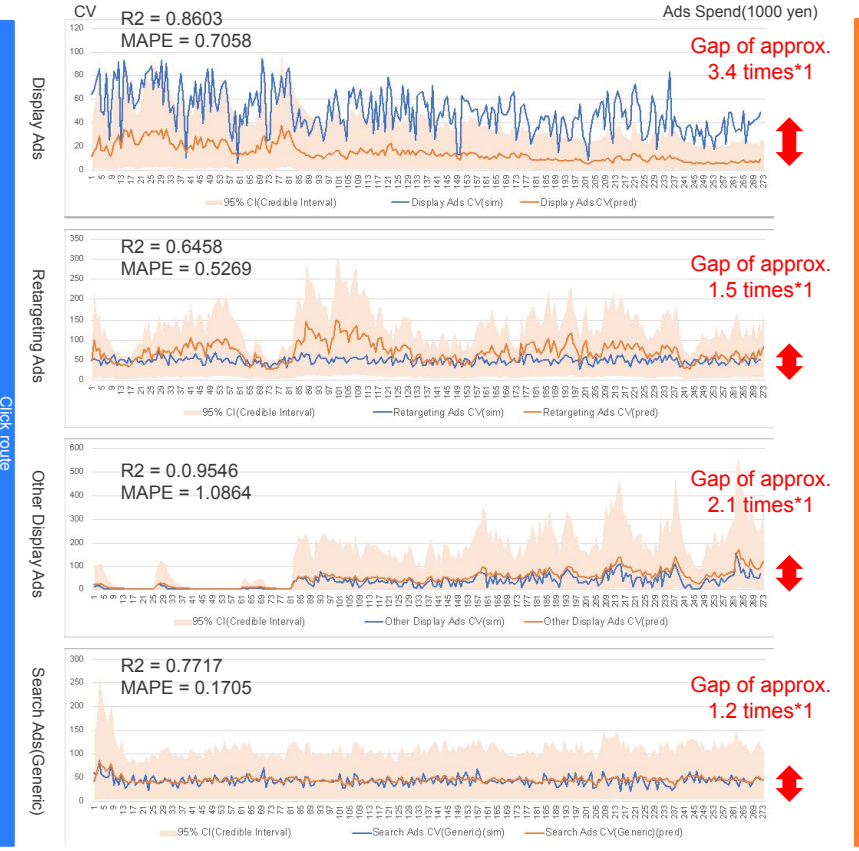
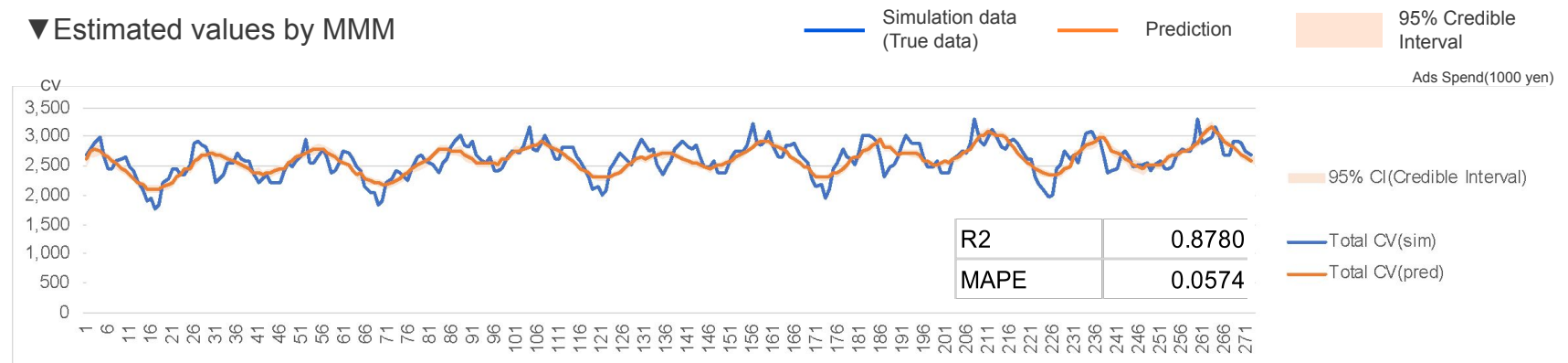
Scenario 2: Click + Search Route Verification | Analysis Results

Applying the single-layer model resulted in a high accuracy of $R^2 = 0.88$ and $MAPE = 0.06$. The simulated data for the objective variable did not fall within the 95% credible interval of the estimate by MMM because the contribution amount of the search route, which accounts for 90% of the total CVs, was not estimated correctly. In particular, the contribution amount of TV ads as an indirect effect and the contribution amount of search ads(brand) as a direct effect were underestimated, resulting in an error of about 9.8 times.

In the click route, media other than display ads fall within the 95% credible interval, and the error is smaller than in the search route.

- Legend**
- Display Ads: Amount of ads spend and number of CVs for programmatic ads without retargeting in still image format
 - Retargeting Ads: Amount of ads spend and number of CVs for programmatic ads delivered by retargeting in still image format
 - Other Display Ads: Amount of ad spend and number of CVs for premium ads in still image format
 - Search Ads(Generic): Value of ads spend and number of CVs for category keywords and other general keywords campaigns, excluding brand keywords, in search ads.
 - Search Ads(Brand): Among search ads, the number of CVs as a direct effect of brand keyword campaigns. Excluding CVs as indirect effects of TV ads and digital videos.
 - via Organic Search: Number of CVs as a direct effect of organic search. Excluding contributions as indirect effects from TV ads and digital videos.
 - TV ads: The amount of TV ads spend and the sum of CVs contributed as indirect effects to search ads(brand) and organic search
 - Digital video: The amount of digital video ads spend and the sum of CVs contributed as indirect effects to search ads(brand) and organic search.
 - (sim) indicates simulated data = true data, and (pred) indicates the amount estimated by MMM, respectively.

▼ Estimated values by MMM



*1: MAPE (Mean Absolute Percentage Error) of simulation data and each media estimator converted to a multiplier

- The simulation study found that the single-layer model generally worked for the Scenario 1 (click route), but not for the Scenario 2(click + search route).
- There is a large risk of making an error in estimating the media contribution amount due to an error in the model structure hypothesis that applies the single-layer model to a data generation process with a hierarchical structure, such as the search route in scenario 2(CVs for search ads(brand) are generated by TV and digital video ads).
- In addition, the error in the estimated contribution amount may be large when the proportion of the objective variable or the variance of the data is extremely small.

Single-layer model *1

$$Revenue_t = b + \sum_m \beta_m * Hill (Adstock(x_{t,m}, \dots, x_{t-L,m}; L, w_m(t; \alpha_m, \theta_m)), K_m, S_m) + trend_t + seas_t + \sum_c \gamma_c d_{t,c} + e_t$$

Assumption on the model structure*1

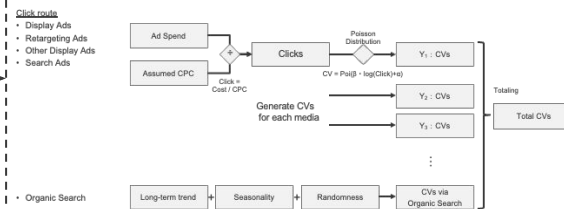
- As an additive linear structure is assumed in the model, media variables are independent each other. (e.g., no dependency between TV/YouTube and search ads volume)
- Organic brand query volume is not included as a variable
- No non-media variable is utilised in this model (i.e., trend and seasonality are modeled with a concave curve and sinusoidal function respectively.)

Assumption on the parameters*1

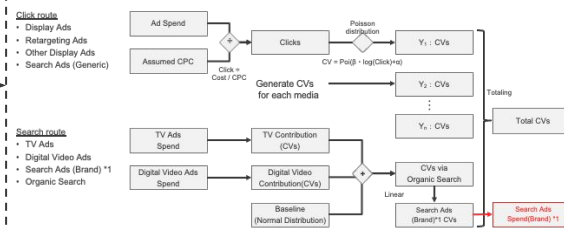
- Hill function (p.46) and geometric adstock (p.49) are assumed for media variables (using "hill adstock" in Lightweight MMM).
- Use default settings in LightweightMMM for latent unobserved trends and seasonality, and for the prior distribution of each parameter.

Simulation Data

Simulation scenario 1 "click route"



Simulation scenario 2 "click route" + "search rule"



Apply the model

Result

- When the baseline proportion is 80-90%, the true data falls within the 95% credit interval of the estimated contribution of each media, as estimated by applying the single-layer model.
 - However, the estimation accuracy of the estimated contribution of each media varies depending on the ratio of the baseline to the total CVs, resulting in a gap of approximately 1.5 to 2 times or more.
 - On the other hand, there are cases where the estimation accuracy is high in cases where there is a sharp variation in the amount of ads spend, such as in the case of premium ads.
-
- Using a model structure hypothesis that differs from the data generation process in the MMM results in large errors. For the search route, the errors ranged from about 2.9 to 9.8 times larger. The underestimation and gap in TV ads can be attributed to the TV media characteristics of time placement (small variance).
 - On the other hand, digital video, which has a peak in placement, has a gap of approximately 2.9 times, although the true data falls within the 95% credible interval.
 - The incorrect setup of the model structure and the degree of variance for each medium make it difficult to analyze consistently the degree of gap and the causes of overestimation and underestimation.

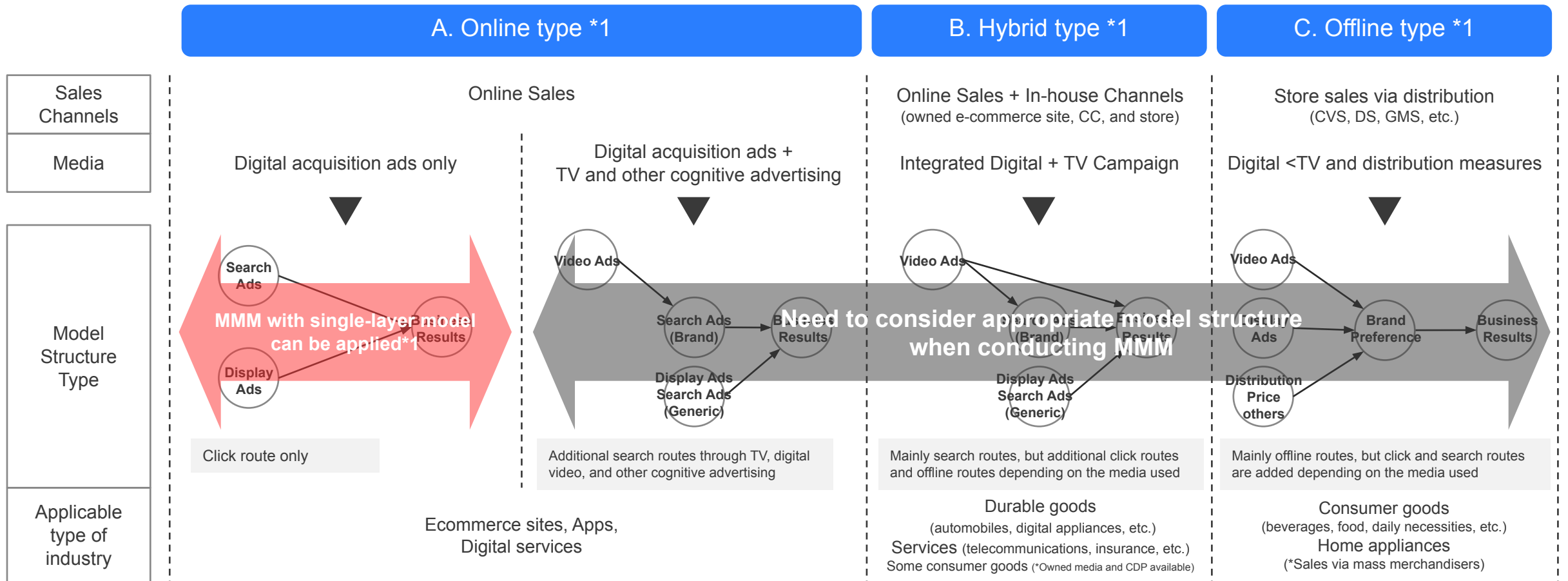
*1: The tested single layer model refers to Lightweight MMM, which is Google unofficial open source as of Sep 2023.

- HDY group believes that the single-layer model can be used for MMM in companies that use only digital acquisition ads and conduct online sales, because the click route works to estimate the amount of contribution. *1
 - If the variance of the explanatory variables is small, the error in the estimated contribution amount may be large.
- However, when branding activities are developed through cognitive ads such as TV and digital video, or when offline sales channels are used, applying a single-layer model may result in large errors in estimating the contribution amount, so it is very important to consider the model structure.

Legend
 → indicates the direction of causality

Each type is a DAG (acyclic directed graph) and assumes the following

- Each element is a set of variables summarized by media characteristics, which must be decomposed appropriately when conducting MMM.
- It is assumed that there is no time dependence in causality.
- The confounding among unobservables and variables is omitted, and its problems are discussed later.



*1: However, there may be cases where a search route exists even if only digital acquisition ads is used, such as when search ads for apps is used, in which case the single-layer model may not work. Verification of the existence of sales routes is necessary for each individual business

3.4. Problems with Model Structure and Hints for Solution

- As the simulation study shows, a single-layer model cannot handle the search route when it is included. Since the search route is a major route in the expansion of online sales through branding and in hybrid online/offline sales channels, we would like to explore ways to apply MMM to the search route.
- Another characteristic of campaign planning that includes the search route is that TV and display ads are often implemented at the same time, and the correlation between the explanatory variables is high.
- These two issues are represented in the causal graph below.

Legend

- Y: Objective variable such as online CV and sales
- S: Search ads(brand), organic search
- C: Click-based ads such as Display ads, search ad (generic words), etc.
- T: Cognitive advertising such as TV and digital video ads
- P: Cognitive index indicating brand selection probability
- Directed arrow indicates the direction of causality
- ↔ Two-way sagittal lines indicate correlation due to unobserved common causes

A. Online type *1

B. Hybrid type *1

Click Route + Search Route

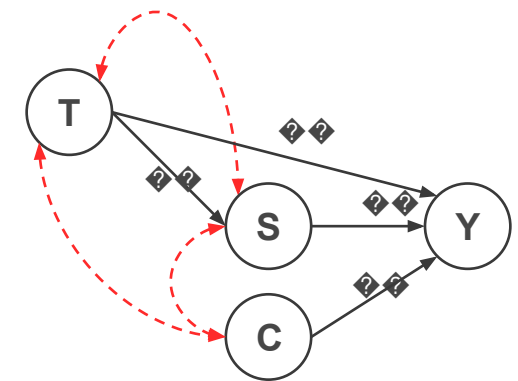
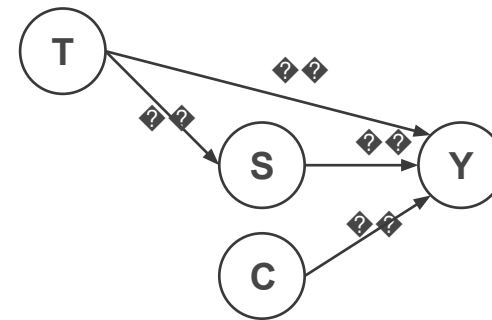
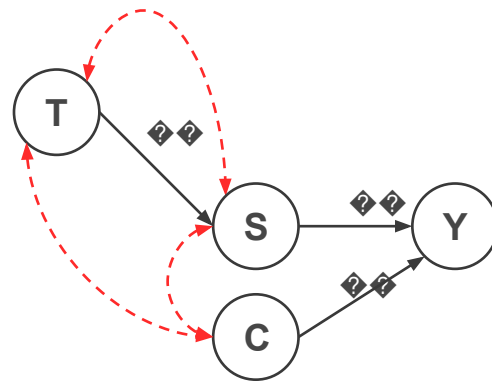
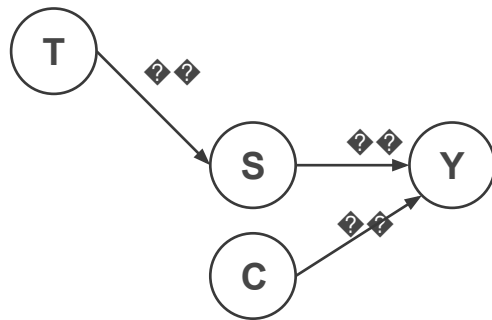
Click + Search + Offline Routes

Ideal: Explanatory variables are independent

Real: Explanatory variables are confounded

Ideal: Explanatory variables are independent

Real: Explanatory variables are confounded



- Directed arrow lines T→S S→Y, etc. indicate the direction of causality
- The bidirectional arrow line T↔S T↔C S↔S indicates correlations due to unobservable common causes such as simultaneity of campaign implementation or coincident seasonality, for example, when TV and display ad flight patterns are identical.
- S and Y may be confounded, and the ratio via search ads may be high for the objective variable

*1: Although it is necessary to consider the volume of nominated searches with respect to search advertisements, for the sake of simplicity, this section omits such consideration.

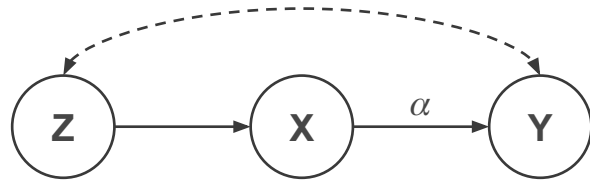
- To apply MMM to the expansion of online sales through branding and hybrid online-offline sales channels, the solution is to adopt a model structure that satisfies the previously mentioned parameter identifiability conditions.
- In this section, we provide hints to solve the problem regarding the model structure by applying the identifiability conditions of the parameters.

Legend

- Y: Objective variable
- X: Cause of interest for the objective variable
- Z: Factors other than X that are assumed to be causally related
- U: Unobserved variable
- : indicates the direction of causality
- ↔ : indicates the confounding among variables
- $r_{AB,C}$: Partial regression coefficient of B when A is multiple regressed on B and C
- r_{AB} : Regression coefficient of A in a simple regression of A with B

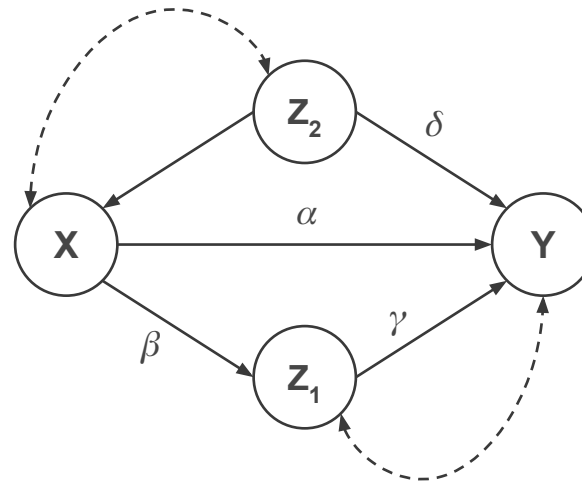
Identifiability of causal graph*1 and parameters

Single door criterion*2



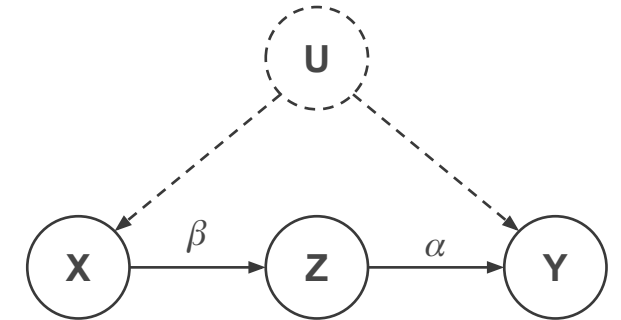
- Z meets single door criterion
- Direct effect α from X to Y is identifiable
- The direct effect is given by $\alpha=r_{YX,Z}$, consistent with the partial regression coefficient α for $Y=\alpha X+\beta Z+\varepsilon$

Back door criterion*2



- Z_2 meets backdoor criterion
- The overall effect $\alpha+\beta\gamma$ from X to Y is identifiable
- The overall effect is given by $(\alpha+\beta\gamma)=r_{YX,Z}$, consistent with the partial regression coefficient α for $y=(\alpha+\beta\gamma)X+\delta Z_2+\varepsilon$

Front door criterion*2



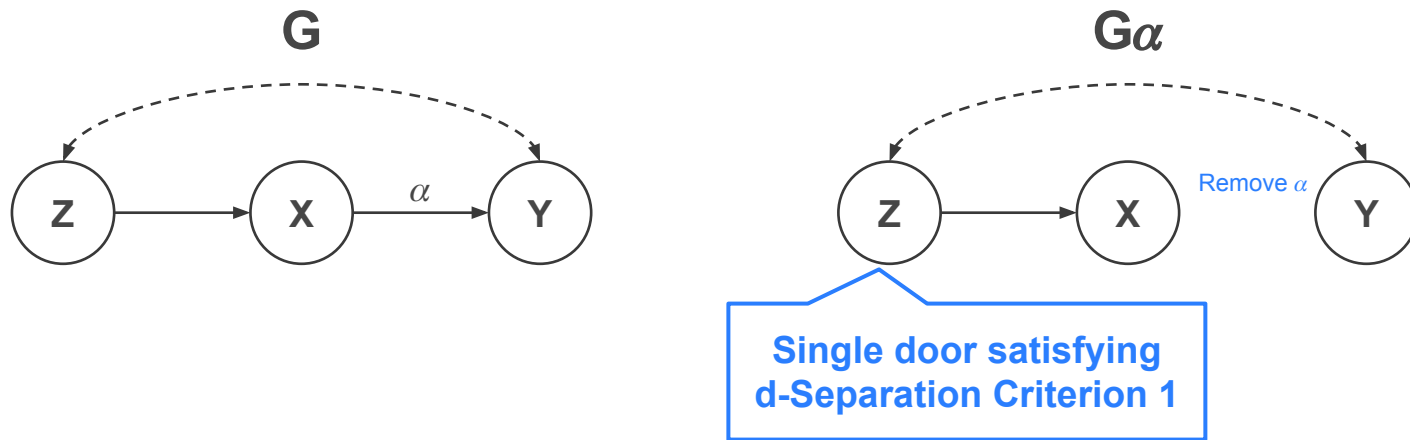
- Z meets frontdoor criterion
- The overall effect $\alpha\beta$ from X to Y is identifiable
- It is the product of the direct effects α and β , given by $\alpha=r_{YZ,X}$ and $\beta=r_{ZX}$, respectively

*1: There is way of casual discovery to infer causal graph with assumption on parameters and functions to describe relation among variables (e.g., IC algorithm, greedy equivalence search (GES), LiNGAM, NOTEARS algorithm etc.). The methodologies are out of this guidebook scope. Also, in this section, all graphs are DAG (Directed Acyclic Graph, which is directed and not cyclic graph) for the simplicity. *2: Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE"(2009)

- The single-door criterion shows how to estimate the direct effect of a variable set Z on the causal relationship of the explanatory variable $X \rightarrow$ objective variable Y , when the variable set Z has an indirect effect on X . Note that Z can be the empty set ϕ , and the simple relationship $X \rightarrow Y$ is considered a special case where Z is the empty set ϕ .
- If the variable set Z satisfies the single door criterion and the d-separation criterion, the direct effect α of $X \rightarrow Y$ is given by the regression coefficient $\alpha = r_{XY.Z}$, which is consistent with the partial regression coefficient α for $Y = \alpha X + \beta Z + \epsilon$.

Legend

- Y: Objective variable
- X: Cause of interest for the objective variable
- Z: Factors other than X that are assumed to be causally related
- U: Unobserved variable
- \rightarrow : indicates the direction of causality
- \leftarrow : indicates the confounding among variables
- $r_{A.B.C}$: Partial regression coefficient of B when A is multiple regressed on B and C
- r_{AB} : Regression coefficient of A in a simple regression of A with B



Single-Door Criterion

Let G be any path diagram in which α is the path coefficient associated with link $X \rightarrow Y$, and let G_α denote the diagram that results when $X \rightarrow Y$ is deleted from G . The coefficient α is identifiable if there exists a set of variables Z such that:

- Any element of Z is not a descendant of Y .
- In G_α , Z d-separates X and Y .

If Z satisfies these two conditions, then α is equal to the regression coefficient $r_{YX.Z}$.

d-Separation Criterion

A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if:

1. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $j \leftarrow m \rightarrow j$ such that the middle node m is in Z ; or
2. p contains an inverted fork (or collider) $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z .

A set Z is said to d-separate X from Y if and only if Z blocks every path from a node in X to a node in Y .

*:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

At $X \rightarrow Y$ in graph G , we can decompose it into the regression coefficient $r_{YX} = \alpha + IYX$. The path between X and Y passes through Z , but Z is not a collider and may have some effect on Y , $IYX \neq 0$.

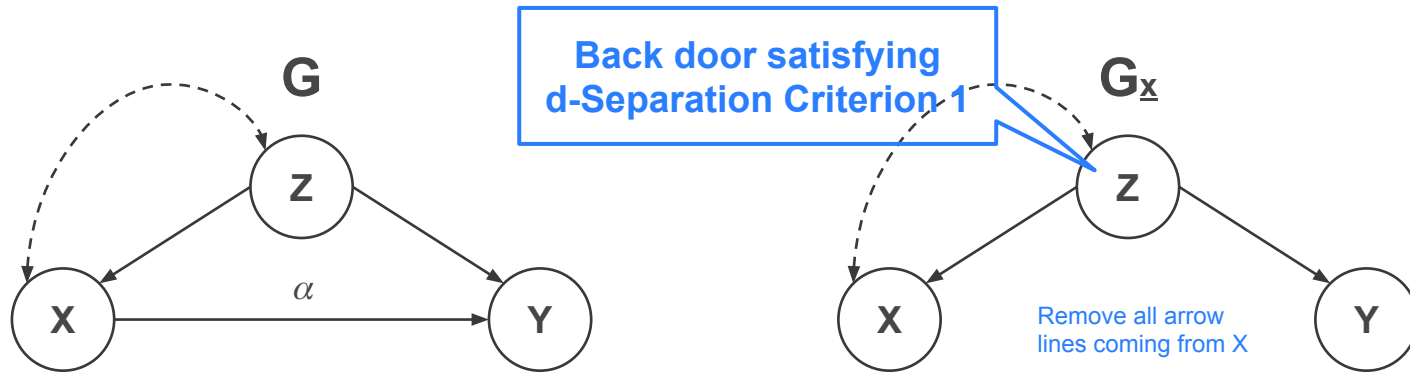
In the subgraph G_α , excluding the direct effect α that we want to identify, the path between X and Y runs through Z . Since Z satisfies the single door criterion, adjusting by Z gives $IYX = 0$ for $r_{YX} = \alpha + IYX$. The regression coefficient for $X \rightarrow Y$ is thus given by $r_{YX.Z} = \alpha + IYX.Z$. This α is consistent with the partial regression coefficient of X in the following regression equation.

$$Y = \alpha X + \beta Z + \epsilon$$

- Causal relationships, called backdoors, tend to creep into complex model structures, making the estimation of causal effects difficult. The method for finding these backdoors is the backdoor criterion.
- In the causal graph G , Z is a backdoor for X . Even in this case, by adjusting Z , the partial regression coefficient α for $Y = \alpha X + \beta Z + \varepsilon$ matches the direct effect of $X \rightarrow Y$ $\alpha = r_{XY.Z}$

Legend

- Y: Objective variable
- X: Cause of interest for the objective variable
- Z: Factors other than X that are assumed to be causally related
- U: Unobserved variable
- \rightarrow : indicates the direction of causality
- \leftrightarrow : indicates the confounding among variables
- $r_{A.B.C}$: Partial regression coefficient of B when A is multiple regressed on B and C
- r_{AB} : Regression coefficient of A in a simple regression of A with B



Back-Door Criterion

For any two variables X and Y in a causal diagram G, the total effect of X on Y is identifiable if there exists a set of measurements Z such that:

- No member of Z is a descendant of X; and
- Z d-separates X from Y in the subgraph G_x formed by deleting from G all arrows emanating from X

Moreover, if the two conditions are satisfied, then the total effect of X on Y is given by $r_{YX.Z}$

d-Separation Criterion

A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if:

1. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $j \leftarrow m \rightarrow j$ such that the middle node m is in Z; or
2. p contains an inverted fork (or collider) $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z.

A set Z is said to d-separated X from Y if and only if Z blocks every path from a node in X to a node in Y.

*:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

At $X \rightarrow Y$ in graph G, we can decompose it into the regression coefficient $r_{YX} = \alpha + IYX$. The path between X and Y runs through Z, but Z is not a confluence. Therefore, Z has an effect on X and Y, and $IYX \neq 0$.

In the subgraph G_x , where all arrow lines coming from X are removed from the graph G, there exist $X \leftarrow Z \rightarrow Y$ and $X \leftarrow \leftarrow Z \rightarrow Y$, which are backdoor paths of X. By adjusting Z, $IYX = 0$ at $r_{YX} = \alpha + IYX$. The regression coefficient for $X \rightarrow Y$ is thus given by $r_{YX.Z} = \alpha + IYX.Z$. This α is consistent with the partial regression coefficient of X in the following regression equation.

$$Y = \alpha X + \beta Z + \varepsilon$$

- The front door criterion is effective for the causal relationship of explanatory variable $X \rightarrow$ objective variable Y when the unobserved variable is the common cause causing the correlation between X and Y .
- Finding a variable Z that satisfies the front door criterion between $X \rightarrow Y$ and estimating the direct effects of $Z \rightarrow Y$ and $X \rightarrow Z$ gives the direct effect $\alpha\beta$ of $X \rightarrow Y$.
 - The direct effect α of $Z \rightarrow Y$ is given by the regression coefficient $\alpha = r_{YZ.X}$, which is consistent with the partial regression coefficient α of $Y = \alpha Z + \beta'X + \varepsilon$.
 - The direct effect β of $X \rightarrow Z$ is given by the regression coefficient $\alpha = r_{ZX}$, consistent with the partial regression coefficient β of $Z = \beta X + \varepsilon$.

Legend
 Y: Objective variable
 X: Cause of interest for the objective variable
 Z: Factors other than X that are assumed to be causally related
 U: Unobserved variable
 \rightarrow : indicates the direction of causality
 \leftarrow : indicates the confounding among variables
 $r_{A.B.C}$: Partial regression coefficient of B when A is multiple regressed on B and C
 r_{AB} : Regression coefficient of A in a simple regression of A with B



At $X \rightarrow Y$ in graph G, we can decompose it into the regression coefficient $r_{YX} = \alpha + \gamma YX$. The path between X and Y passes through the unobserved variable U , but since U is not a collider, it does not satisfy the d-Separation Criterion, so it may have some influence and $\gamma YX \neq 0$.

If there is a variable Z that satisfies the front door criterion in the $X \rightarrow Y$ path, then the overall effect $\alpha\beta$ from $X \rightarrow Y$ can be identified by using Z . The overall effect $\alpha\beta$ is the product of the direct effects α and β , given by $\alpha = r_{YZ.X}$ and $\beta = r_{ZX}$.

Front-Door Criterion

A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if:

Z intercepts all directed paths from X to Y ;

There is no unblocked backdoor path from X to Z ; and

All back-door paths from Z to Y are blocked by X .

d-Separation Criterion

A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if:

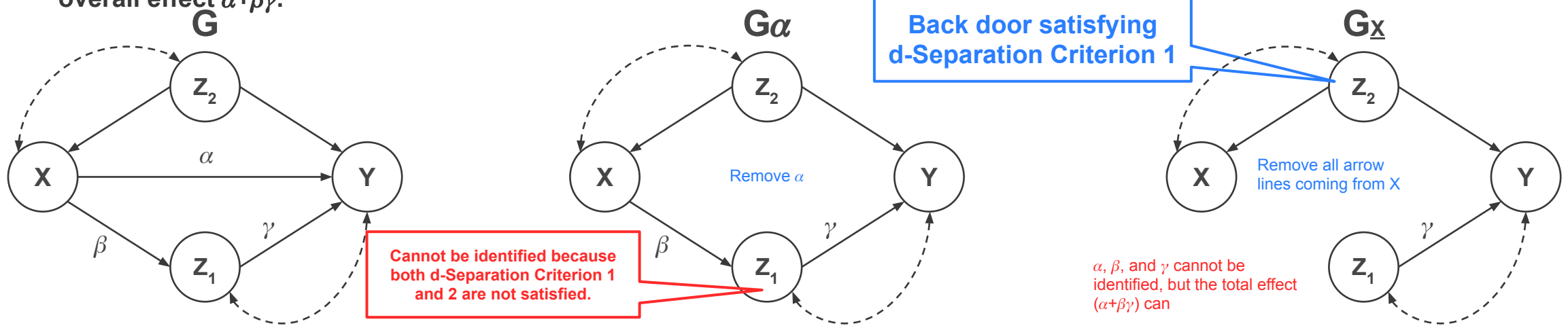
1. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $j \leftarrow m \rightarrow j$ such that the middle node m is in Z , or
2. p contains an inverted fork (or collider) $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and such that no descendent of m is in Z .

A set Z is said to d-separated X from Y if and only if Z blocks every path from a node in X to a node in Y .

*:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

- If a variable that does not meet the d-Separation criterion is related to the effect you wish to identify, you may not be able to identify the direct effect. In such cases, it is necessary to consider whether the effect can be identified as an overall effect.
- In Graph G below, Z1 cannot satisfy the d-Separation Criterion 2 when adjusted and the d-Separation Criterion 1 when not adjusted, so the X→Y effect can only be identified as the overall effect $\alpha+\beta\gamma$.

Legend
 Y: Objective variable
 X: Cause of interest for the objective variable
 Z: Factors other than X that are assumed to be causally related
 U: Unobserved variable
 →: indicates the direction of causality
 ↔: indicates the confounding among variables
 $r_{AB.C}$: Partial regression coefficient of B when A is multiple regressed on B and C
 r_{AB} : Regression coefficient of A in a simple regression of A with B



Cannot be identified because both d-Separation Criterion 1 and 2 are not satisfied.

Back door satisfying d-Separation Criterion 1

α , β , and γ cannot be identified, but the total effect ($\alpha+\beta\gamma$) can

At X→Y in graph G, we can decompose it into the regression coefficient $r_{YX}=\alpha+r_{YX}$. Since the paths between X and Y pass through Z1 and Z2, and since none of them are colliders, they are not directed separable and $r_{YX}\neq 0$.

In the partial graph G_α with the X→Y arrow line removed, the path Z1 between X and Y opens the path $X\rightarrow Z_1\leftarrow Y$ (Z1 is the collider) when adjusted from the d-Separation Criterion 1 and $X\rightarrow Z_1\rightarrow Y$ (the chain path) when not adjusted from the d-Separation Criterion 2. Therefore, it cannot be identified.

So we further consider a subgraph G_x with all arrow lines coming out from X removed. In this case, there exist $X\leftarrow Z_2\rightarrow Y$ and $X\leftarrow Z_1\rightarrow Y$ as backdoor paths of X. If Z2 is adjusted, Z2 satisfies the d-Separation Criterion 1.

d-Separation Criterion
 A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if:
 1. p contains a chain $i\rightarrow m\rightarrow j$ or a fork $j\leftarrow m\rightarrow i$ such that the middle node m is in Z, or
 2. p contains an inverted fork (or collider) $i\rightarrow m\leftarrow j$ such that the middle node m is not in Z and such that no descendent of m is in Z.
 A set Z is said to d-separate X from Y if and only if Z blocks every path from a node in X to a node in Y.

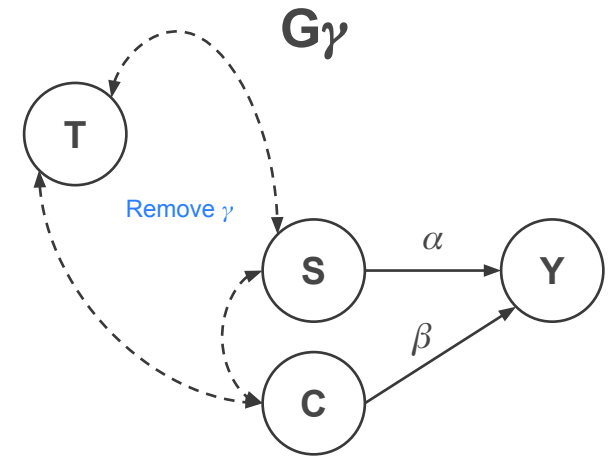
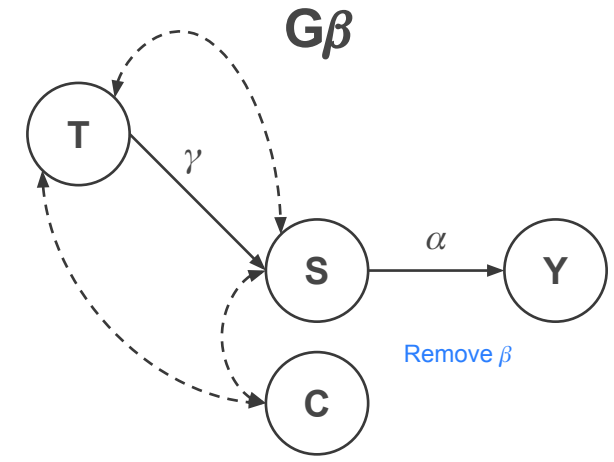
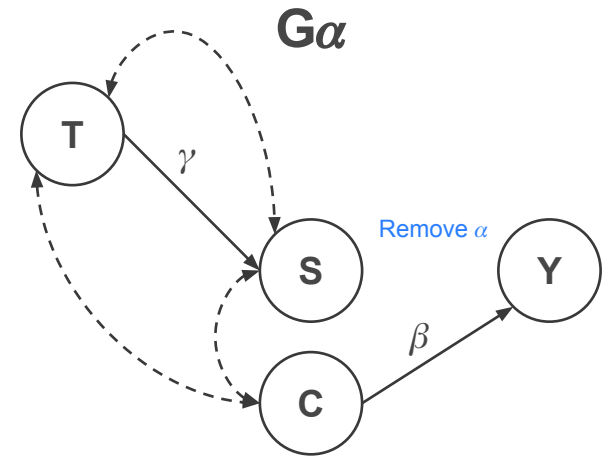
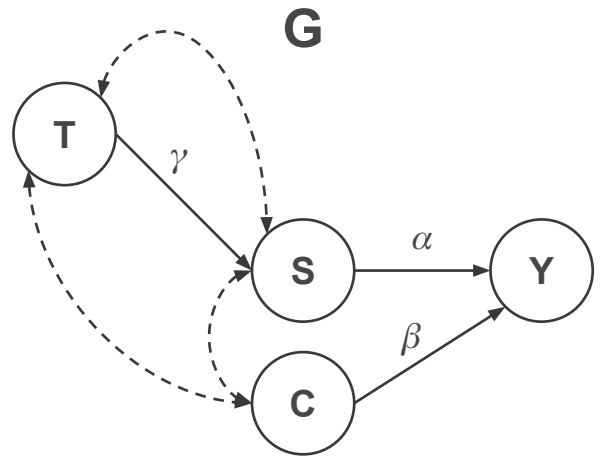
Back-Door Criterion
 For any two variables X and Y in a causal diagram G, the total effect of X on Y is identifiable if there exists a set of measurements Z such that:
 • No member of Z is a descendant of X; and
 • Z d-separates X from Y in the subgraph G_x formed by deleting from G all arrows emanating from X

*:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

- In the online-type model structure, identifiability can be determined in accordance with the identifiability condition*1.
- It is identifiable for the direct effect α of $S \rightarrow Y$ and the direct effect β of $C \rightarrow Y$. In other words, it can be said that modeling is possible for the click route. *2
- The direct effect γ of $T \rightarrow S$ is difficult to identify due to correlation caused by unobservable common causes such as simultaneity of campaign implementation and common periodicity. This is the difficulty in modeling search routes.

Legend

- Y: Objective variable such as online CV and sales
- S: Search ads(brand), organic search
- C: Click-based ads such as Display ads, search ad (generic words), etc.
- T: Cognitive advertising such as TV and digital video ads
- P: Cognitive index indicating brand selection probability
- Directed arrow indicates the direction of causality
- ↔ Two-way sagittal lines indicate correlation due to unobserved common causes



As a characteristic of campaign planning, TV and display ads are often implemented at the same time, and the correlation between explanatory variables is almost always high.

To identify the direct effect α of $S \rightarrow Y$, consider the subgraph G_α . since the paths S and Y are effectively separated by C, the direct effect α is given by $\alpha = r_{Y,S.C}$ and is identifiable.

To identify the direct effect β of $C \rightarrow Y$, consider the subgraph G_β . since the paths S and Y are effectively separated by S, the direct effect β is given by $\beta = r_{Y,C.S}$ and is identifiable.

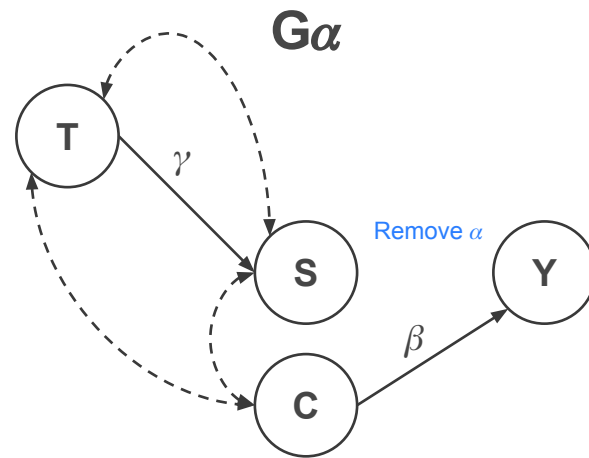
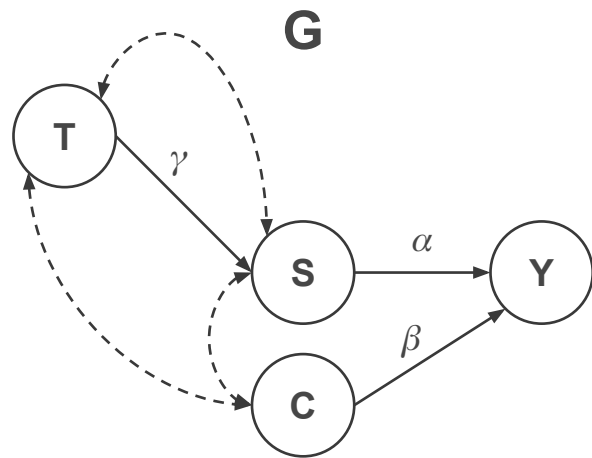
To identify the direct effect γ of $T \rightarrow S$, consider the subgraph G_γ . There are two paths between T and S, $T \leftrightarrow S$ and $T \leftrightarrow C \leftrightarrow S$, which cannot be blocked by adjusting T and cannot be identified.

*1:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).
 *2:Although S and Y may be confounded (i.e., the ratio of CVs via search ads to the objective variable may be high), they are omitted here for the sake of simplicity.

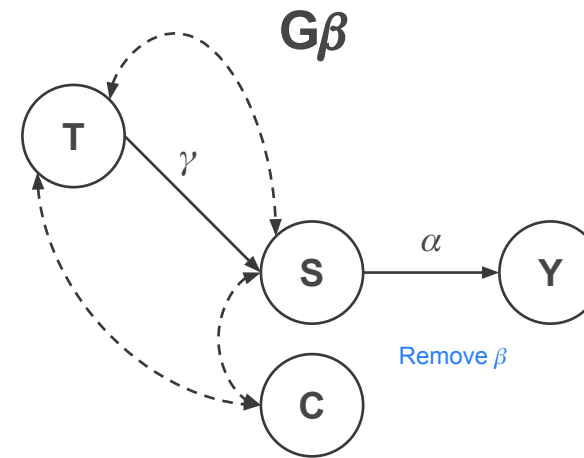
- As mentioned earlier, the direct effect γ of $T \rightarrow S$ was difficult to identify.
- Considering the estimation of the amount of contribution to Y (= business results) via the search route by T (= cognitive ads) here, we would like to eliminate the correlation with C (= click-type ads). Therefore, if we intervene in C (=click-type ads) and eliminate the correlation with T (=cognitive ads), we may at least be able to identify the overall effect $\alpha\gamma$ of $T \rightarrow S \rightarrow Y$. *2
- By intentionally implementing both marketing trial and MMM in this way, the number of patterns of verifiable data will increase, and the possibility of identifying specific effects will increase.

Legend

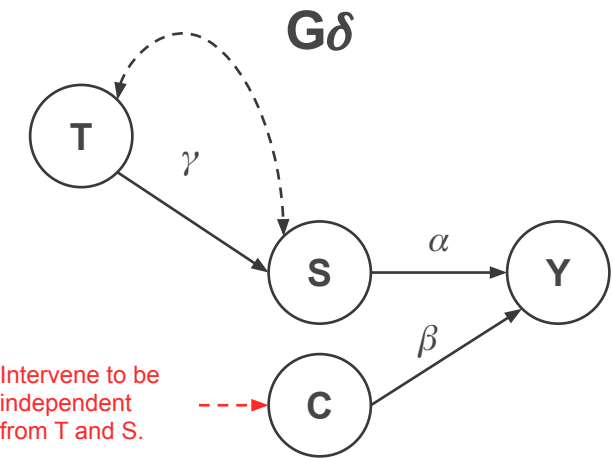
- Y: Objective variable such as online CV and sales
- S: Search ads(brand), organic search
- C: Click-based ads such as Display ads, search ad (generic words), etc.
- T: Cognitive advertising such as TV and digital video ads
- P: Cognitive index indicating brand selection probability
- Directed arrow indicates the direction of causality
- ↔ Two-way sagittal lines indicate correlation due to unobserved common causes



To identify the direct effect α of $S \rightarrow Y$, consider the subgraph G_α . since the paths S and Y are effectively separated by C, the direct effect α is given by $\alpha = r_{YS.C}$ and is identifiable.



To identify the direct effect β of $C \rightarrow Y$, consider the subgraph G_β ; since the paths S and Y are effectively separated by S, the direct effect β is given by $\beta = r_{YC.S}$ and is identifiable.



Intervene to be independent from T and S.

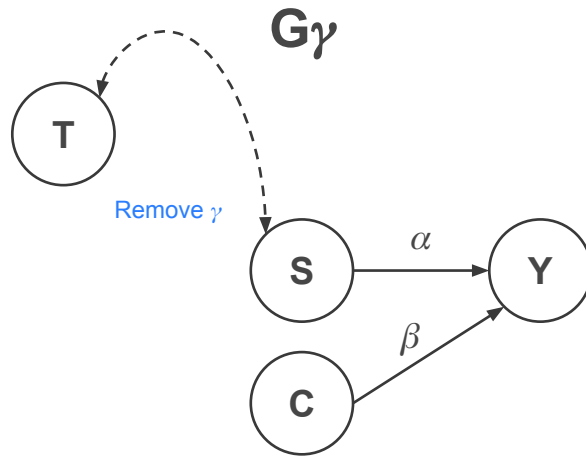
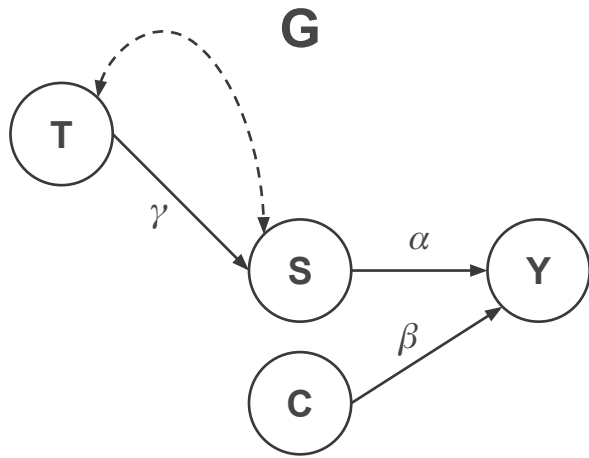
Intervene in C to avoid confounding between C and S, and between C and T. This at least allows us to identify the overall effect $\alpha\gamma$ of $T \rightarrow S \rightarrow Y$ and the direct effect β of $C \rightarrow Y$. given by $\alpha\gamma = r_{YT.C}$ and $\beta = r_{YC.T}$, respectively.

*1:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

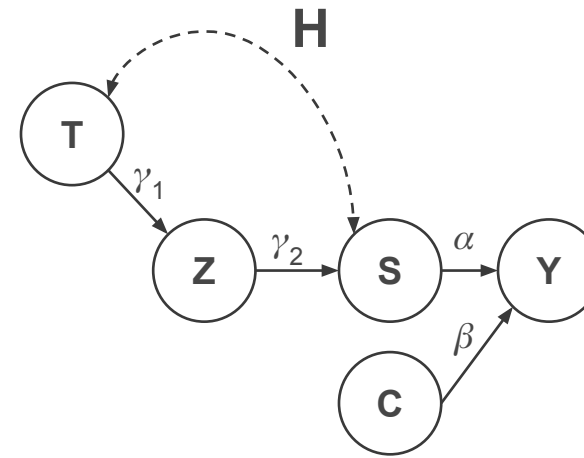
*2:Although S and Y may be confounded (i.e., the ratio of CVs via search ads to the objective variable may be high), they are omitted here for the sake of simplicity.

- Another method is to add a variable Z that satisfies the front door criterion between $T \rightarrow S$. Variable Z is a candidate for adoption of cognitive indicators such as pure brand recall. *2

Legend
 Y: Objective variable such as online CV and sales
 S: Search ads(brand), organic search
 C: Click-based ads such as Display ads, search ad (generic words), etc.
 T: Cognitive advertising such as TV and digital video ads
 P: Cognitive index indicating brand selection probability
 → Directed arrow indicates the direction of causality
 ↔ Two-way sagittal lines indicate correlation due to unobserved common causes

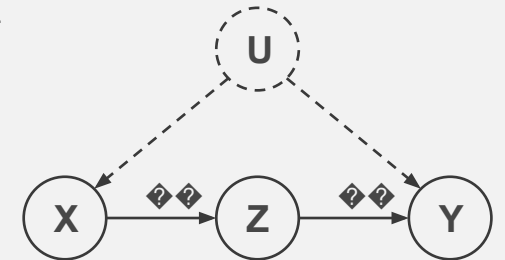


To identify the direct effect γ of $T \rightarrow S$, consider the subgraph G_γ ; the paths of T and S are correlated by an unobserved common cause, so we cannot identify the direct effect γ .



We identify the overall effect $\gamma_1\gamma_2$ by adding a variable Z to the path $T \rightarrow S$ that satisfies the front door criterion. Z blocks $T \rightarrow S$, there is no backdoor from $T \rightarrow Z$, and the backdoor from $Z \rightarrow S$ is blocked by T.

Front-Door Criterion *1
 A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X,Y) if:
 Z intercepts all directed paths from X to Y;
 There is no unblocked backdoor path from X to Z; and
 All back-door paths from Z to Y are blocked by X.



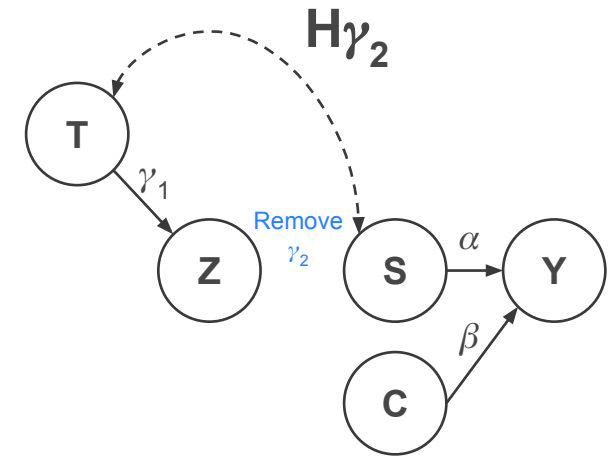
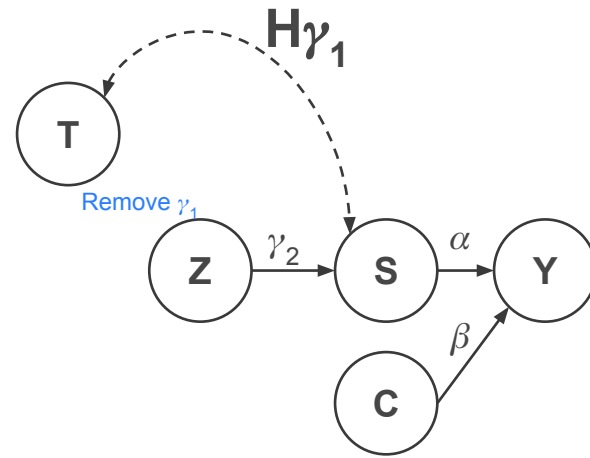
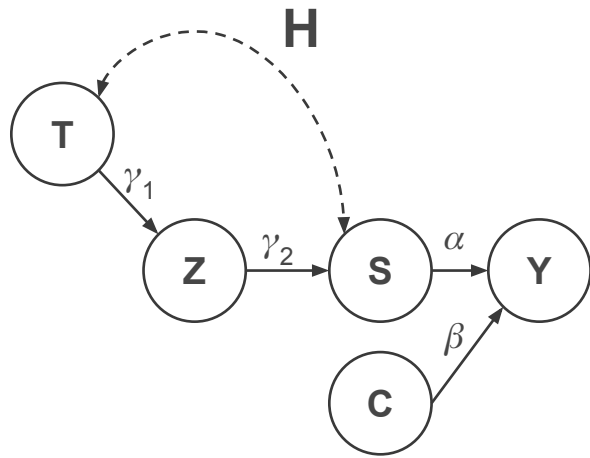
*1:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

*2:Although S and Y may be confounded (i.e., the ratio of CVs via search ads to the objective variable may be high), they are omitted here for the sake of simplicity.

- If Z satisfying the front door criterion can be added to the $T \rightarrow S$ path, then γ_1 and γ_2 become identifiable.
- Apply the identifiability condition to each of γ_1 and γ_2 and calculate γ_1 and γ_2 . The product, $\gamma_1\gamma_2$, is the overall effect of $T \rightarrow S$ in graph H.

Legend

- Y: Objective variable such as online CV and sales
- S: Search ads(brand), organic search
- C: Click-based ads such as Display ads, search ad (generic words), etc.
- T: Cognitive advertising such as TV and digital video ads
- P: Cognitive index indicating brand selection probability
- Directed arrow indicates the direction of causality
- ↔ Two-way sagittal lines indicate correlation due to unobserved common causes



To identify the overall effect $\gamma_1\gamma_2$ of $T \rightarrow S$, consider a graph H with an additional variable Z that satisfies the front door criterion.

Identify the direct effect γ_1 of $T \rightarrow Z$. Removing γ_1 , we can identify the direct effect γ_2 given by $\gamma_2 = r_{SZ}$, since all paths between T and Z can be directly separated by the collider Y.

Identify the direct effect γ_2 of $Z \rightarrow S$. Removing γ_2 , the paths between Z and S can be directed separated by adjusting T, so the direct effect γ_2 is given by $\gamma_2 = r_{SZ.T}$ and can be identified.

*1:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

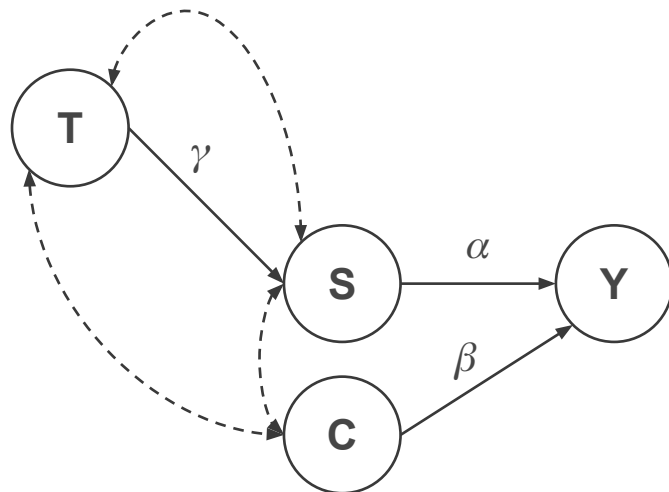
*2:Although S and Y may be confounded (i.e., the ratio of CVs via search ads to the objective variable may be high), they are omitted here for the sake of simplicity.

- The use of the parameter's identifiability condition in this way expands the possibility of identifying the causal effect of the variable of interest, either as a direct effect or as an overall effect.
- It also demonstrates that even when there are unobserved variables for which a causal relationship is assumed but have not been measured or are unavailable, it is possible to identify specific causal effects by adjusting the appropriate variables that satisfy the identifiability condition.

Legend

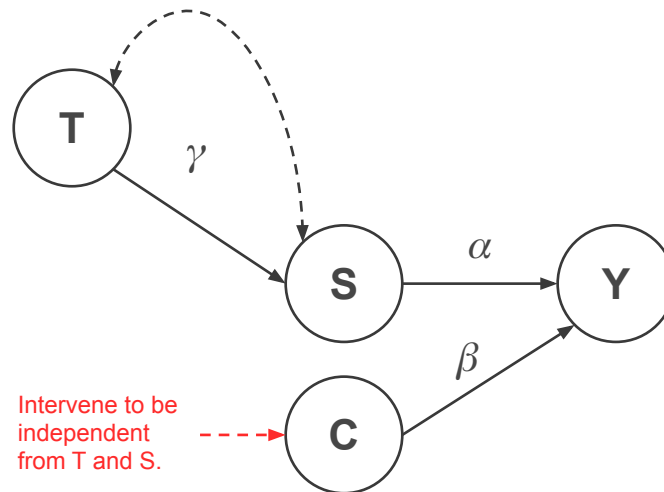
- Y: Objective variable such as online CV and sales
- S: Search ads(brand), organic search
- C: Click-based ads such as Display ads, search ad (generic words), etc.
- T: Cognitive advertising such as TV and digital video ads
- P: Cognitive index indicating brand selection probability
- Directed arrow indicates the direction of causality
- ↔ Two-way sagittal lines indicate correlation due to unobserved common causes

Real: Explanatory variables are confounded

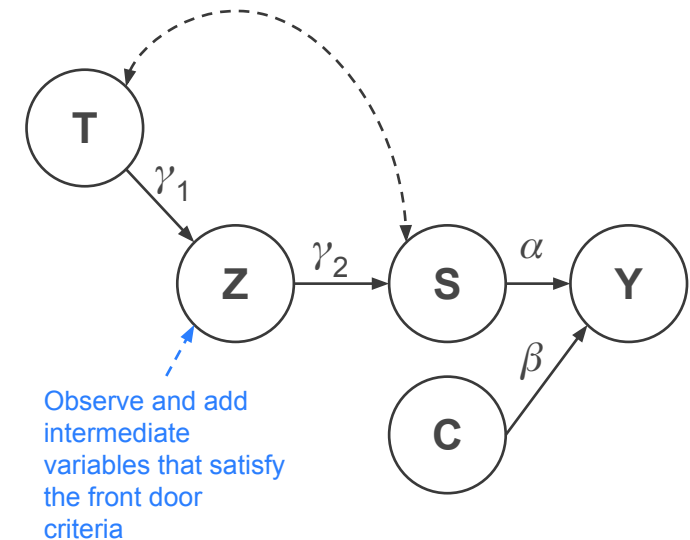


As a characteristic of campaign planning, TV ads and display ads are often implemented at the same time, and the correlation between explanatory variables is almost always high.

a. Identification by Intervention



b. Identification by Front Door



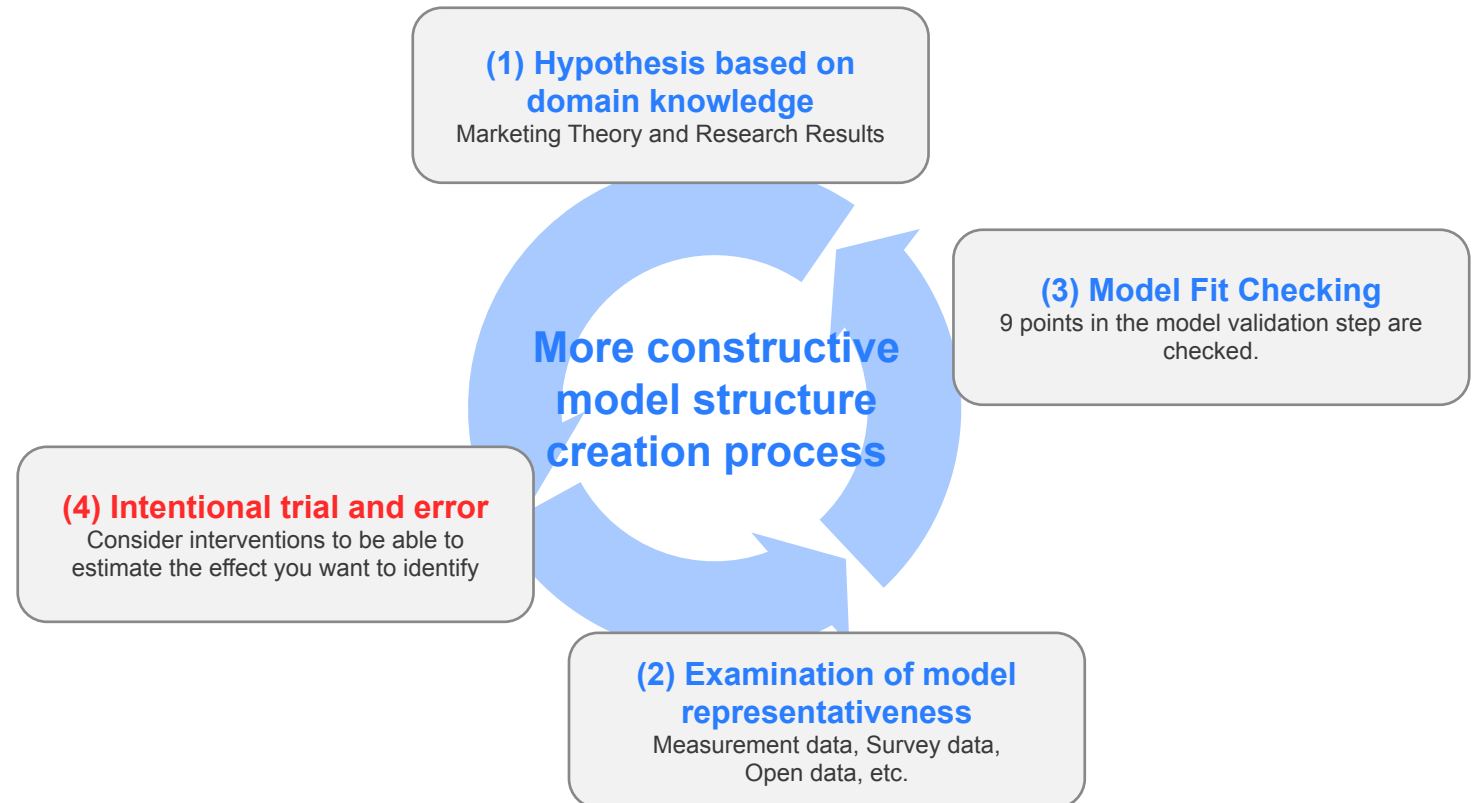
*1:For identifiability conditions, refer to Judea Pearl "CAUSALITY MODELS, REASONING, AND INFERENCE" (2009).

*2:Although S and Y may be confounded (i.e., the ratio of CVs via search ads to the objective variable may be high), they are omitted here for the sake of simplicity.

- In practice, it is stated that the appropriate model structure is determined by conducting hypothesis building and analysis work from the three perspectives and going back and forth between the steps, but to solve the model structure problem, intervention in the media and marketing plan in the first place should also be considered.
- In order to properly estimate the effect of a particular media/marketing plan, intervening in the appropriate elements in accordance with the parameter identifiability condition, increasing the pattern of observed data, and other trial-and-error procedures will increase the accuracy and expand the possibilities of MMM-based effectiveness verification.

Examples of Intentional Trial and Error Methods

- Establish a stand-alone media placement period for the media you wish to identify.
- Conversely, establish a period of time when you do not place the media you wish to identify.
- Establish mutually different flight patterns for the media you wish to identify and shift the timing
- Measure intermediate variables to satisfy the front door criterion and use them for validation



3.5. Hakuhodo DY Group MMM Solutions

This section presents Hakuhodo DY Group's MMM capabilities.
The content in this section is proprietary to Hakuhodo DY Group.

HDY group provides Analytics AaaS which visualizes, diagnoses, predicts and prescribes marketing activities through MMM.

Analytics AaaS

Monitoring:
Visualisation of marketing activities

Marketing Mix Modeling:
Diagnostics, prediction and prescription of marketing activities



Visualisation of media operation and marketing activities including estimation of integrated reach of TV and digital media



Quantify the business contribution of each media and marketing activity, and prescribe KPI forecasts and budget allocation necessary to achieve the business goals

Analytics AaaS

1.

Various mathematical models corresponding to product characteristics

A wide variety of mathematical models in marketing science can be used to address influencing factors and causal relationships that vary depending on the characteristics of the product or service and the position of the product.

2.

**Marketer / Media Planner
×
Data Scientist**

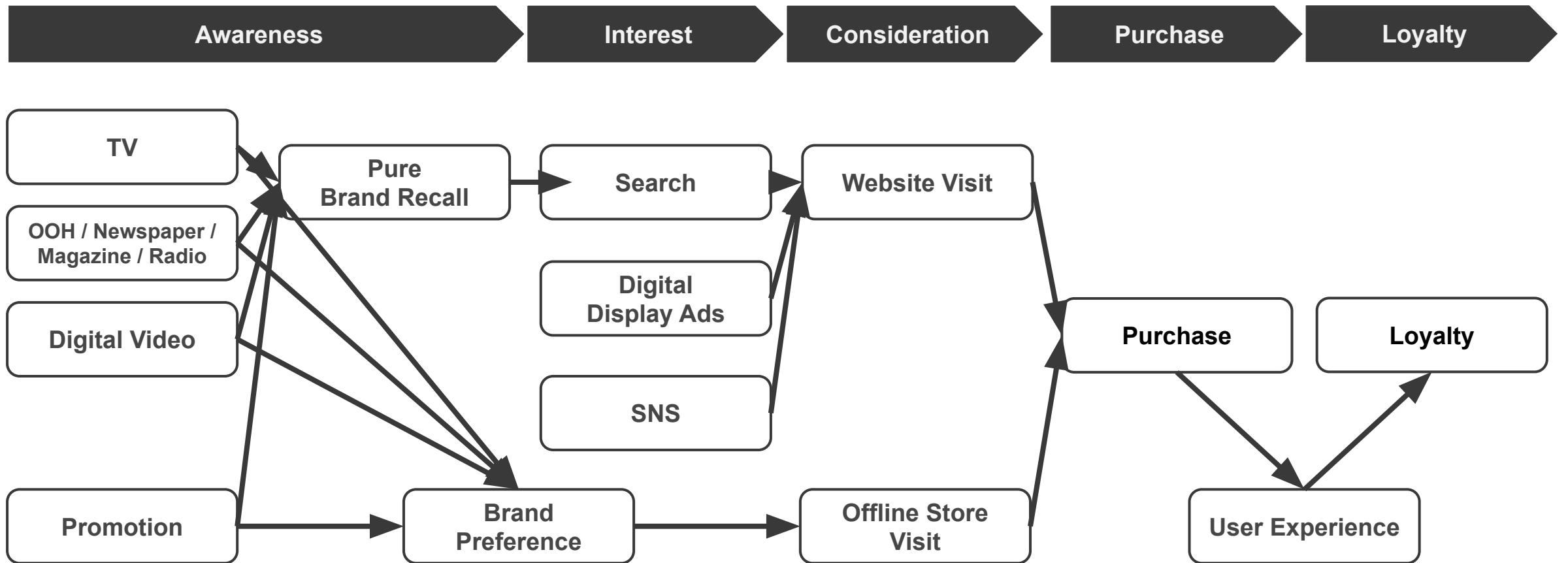
PDCA and consulting by marketers and media planners using MMM model-driven media operations dashboard

3.

Automation of operational flow

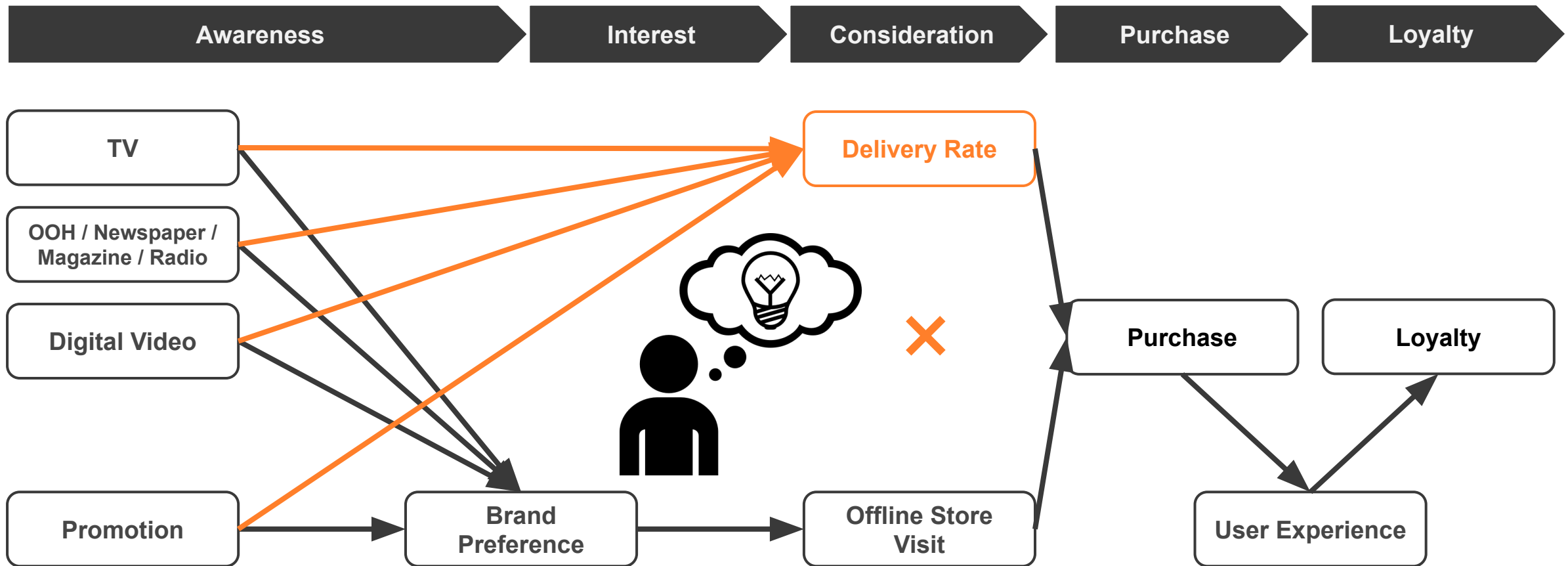
We minimize your time and effort by automating all data preparation, cleansing, and importing to the greatest extent possible.

Analytics AaaS can apply all three model structure types: online, hybrid, and offline type.

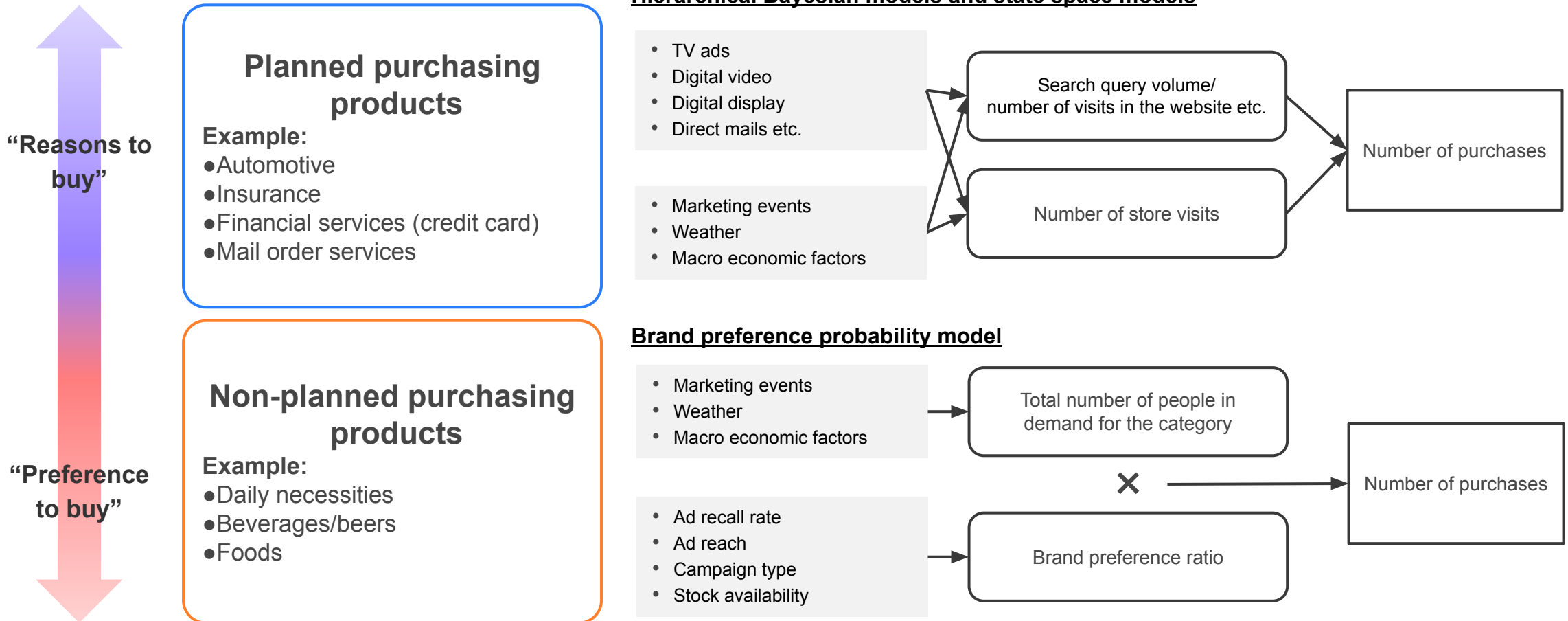


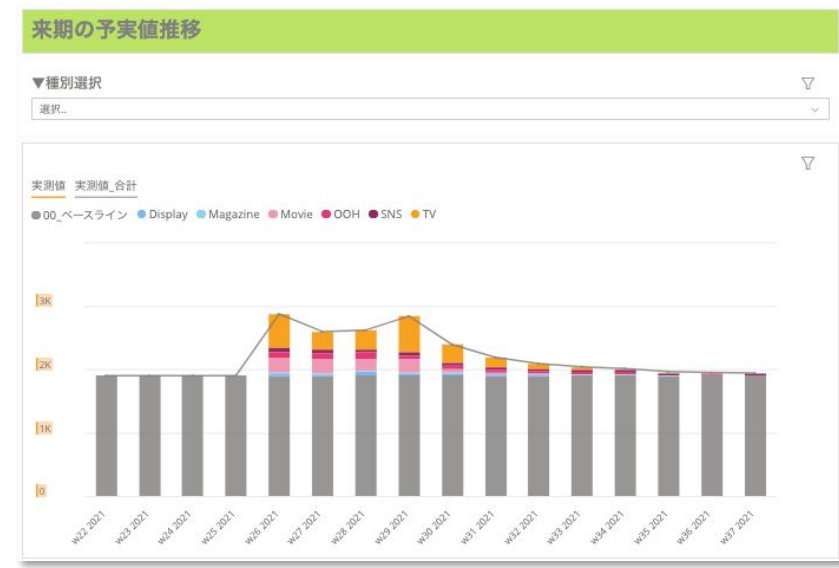
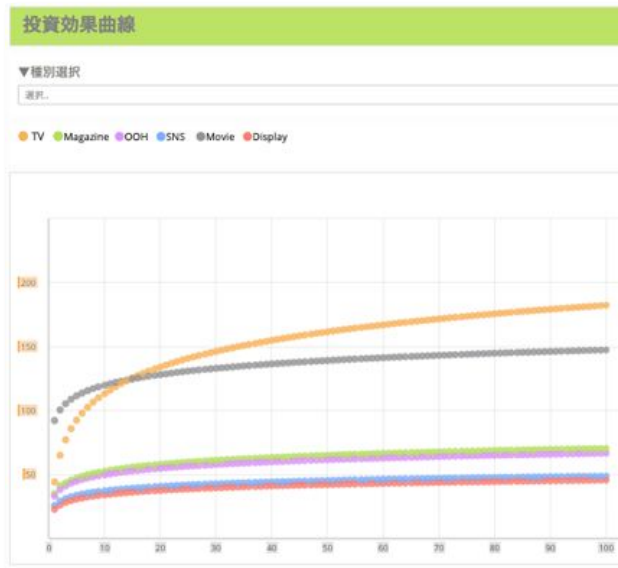
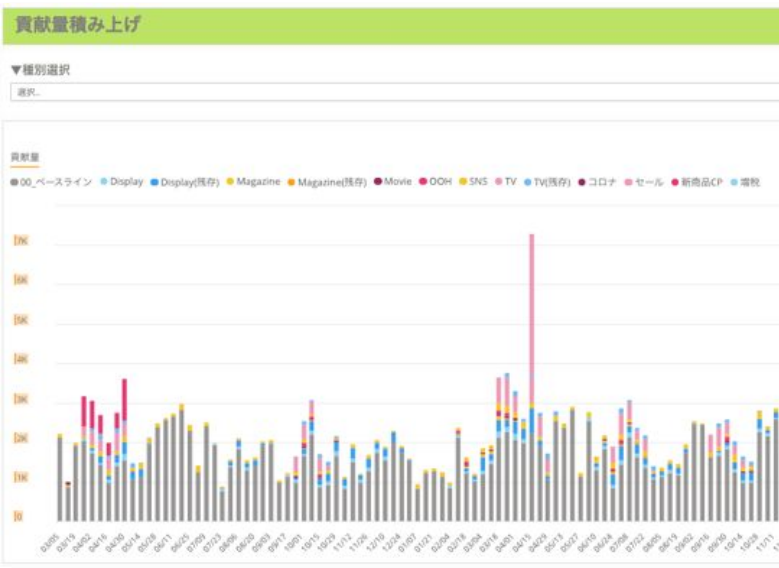
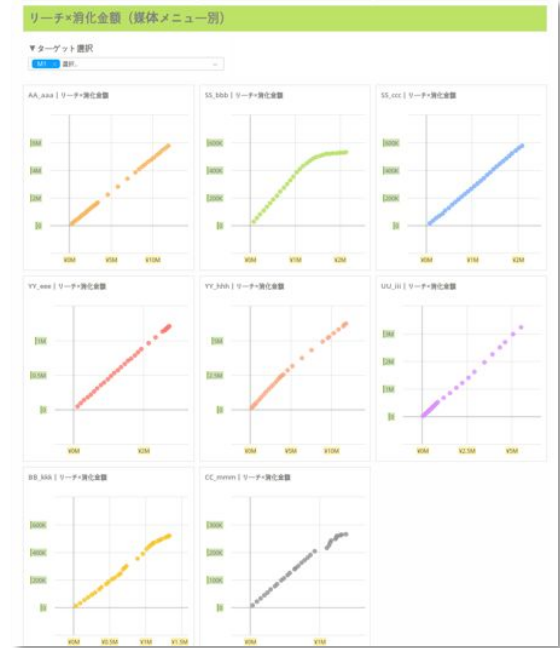
In particular, consumer goods such as beverages, food, and daily necessities, as well as home appliances.

Analytics AaaS for CPG is available for offline model structures.



Analytics AaaS provides MMM that apply a model structure appropriate to the purchasing behavior of consumers for each product or service.





Thank you for reviewing the "Marketing Mix Modeling Guidebook."

While MMM is a globally recognized effective model, it requires a correct understanding for utilization.

There is a risk of making wrong decisions regarding advertising investments if there is a lack of understanding the model or if compatibility with your own business is not considered.

When you need any help of MMM, please contact Hakuholdo DY Media Partners.

Contact information

Hakuholdo DY Media Partners Public Relations Office Yamasaki, Toda

Mail : mp.webmaster@hakuholdody-media.co.jp

End of the guidebook

CHAPTER 1 Introduction of Marketing Mix Modeling

Author:
Data Science Director, Data Science Dept., Data
Driven Planning Div., HAKUHODO INC.
And AaaS Business Strategy Div., Hakuhodo DY
media partners Inc.
Takashi Miyakoshi

CHAPTER 2 Basics of Marketing Mix Modeling

Author:
Senior Marketing Effectiveness Research Manager,
Consumer and Market Insights, Google Asia Pacific
Hirotoshi Nakahara

CHAPTER 3 Considerations on MMM Model Structure

Author:
Data Science Director, Data Science Dept., Data
Driven Planning Div., HAKUHODO INC.
And AaaS Business Strategy Div., Hakuhodo DY
media partners Inc.
Takashi Miyakoshi

Data Scientist, Data Science Dept., Data Driven
Planning Div., HAKUHODO INC.
Daiki Maruo

Technical Support:
Senior Marketing Effectiveness Research Manager,
Consumer and Market Insights, Google Asia Pacific
Hirotoshi Nakahara

Google Reviewers (CHAPTER 2, 3):
Head of Analytics, Consumer and Market Insights, Google Australia
Senior Marketing Research Manager, Consumer and Market Insights, Google Japan
Data Scientist, Strategic Insights & Communication, Google Japan

Rohan Gifford
Minh Nguyen
Takashi J Ozaki