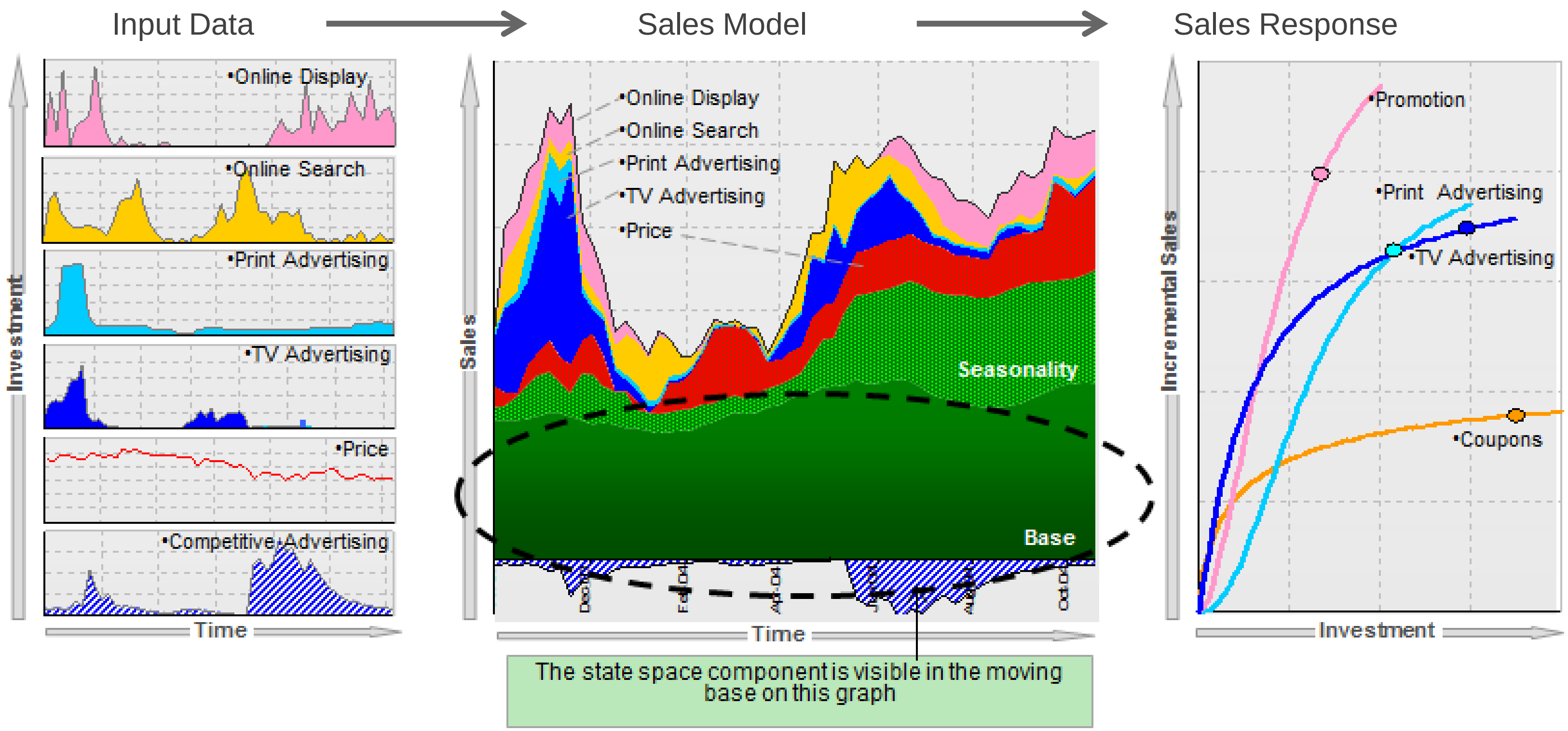
MROI Analytics

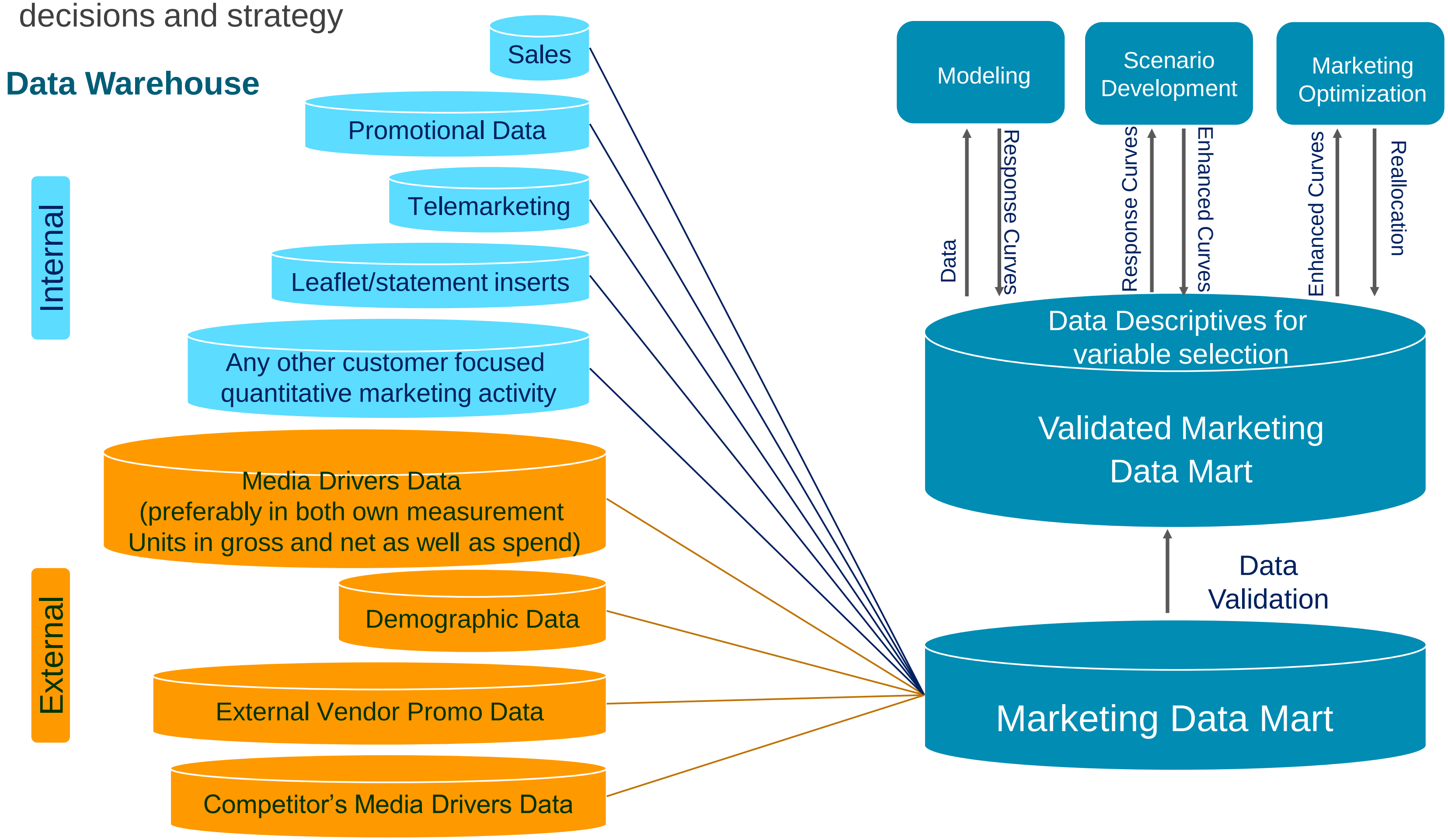


- There are always factors that can influence sales but that we do not have means to measure e.g. competitor Operator Incentives.
- This does not mean that we should ignore those factors because the effects of these unknown parameters may otherwise be misattributed to various other known factors like media etc. and thereby biasing their effectiveness estimate
- To deal with this aspect of the problem we employ the so-called state space modeling technique
- The nice feature of this approach is that it accounts for the parameters that have influenced sales that we are not aware of or we do not have an accurate way to track.

Data management

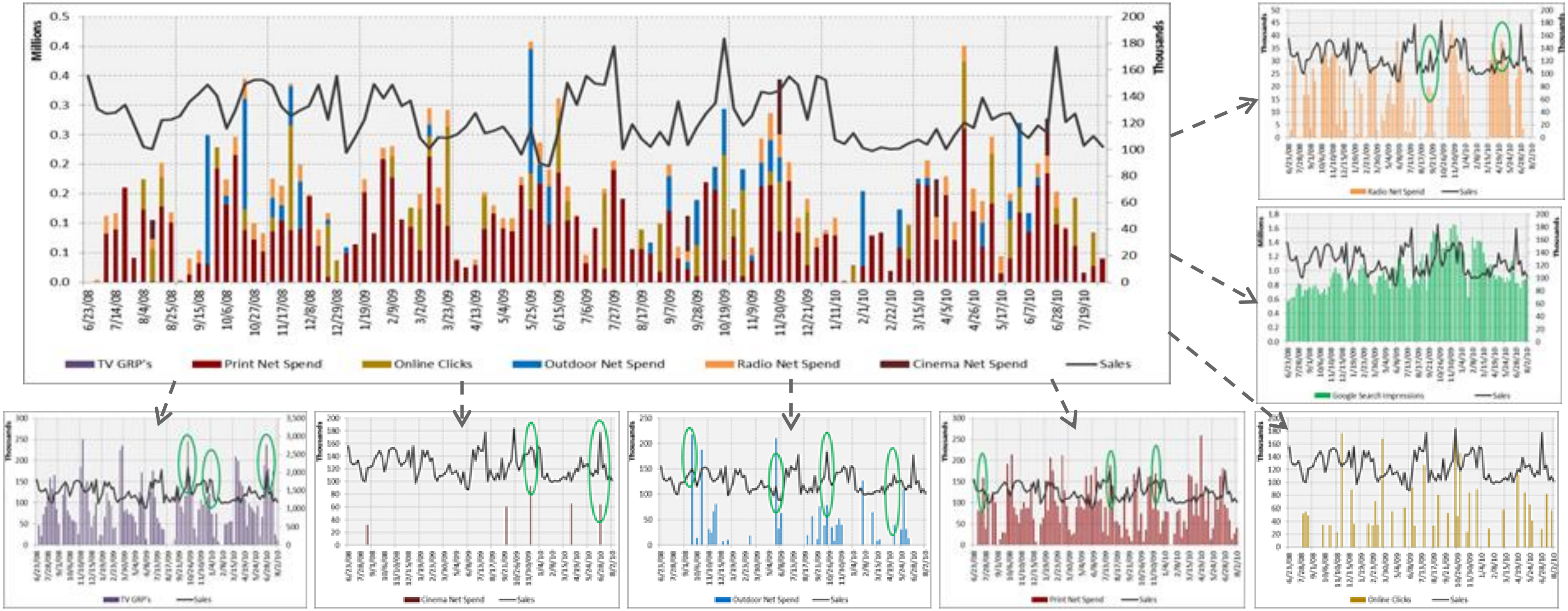
We can isolate and measure the impact of the complex range of marketing levers that use this insight to make your marketing more effective.

This is an ongoing capability that allows you to continuously measure, manage and optimize your marketing decisions and strategy



Descriptive statistics for media variables

We examine the relationship of each explanatory factor with sales



Correlation Matrix	Sales	TV GRP's	Print Net Spend	Online Clicks	Google Search Impressions	Outdoor Net Spend	Radio Net Spend	Cinema Net Spend
Sales	1							
TV GRP's	0,212	1						
Print Net Spend	0,006	0,098	1					
Online Clicks	0,082	0,246	-0,078	1				
Google Search Impressions	-0,016	0,032	0,120	0,094	1			
Outdoor Net Spend	0,025	0,006	0,263	-0,015	0,063	1		
Radio Net Spend	0,134	0,188	0,371	-0,059	0,018	0,028	1	
Cinema Net Spend	-0,052	0,217	-0,021	0,038	0,138	0,044	0,237	1

- The first step of the process consists of descriptive charts between each marketing driver's time series investment information against sales.
- This step is useful for obtaining an understanding of the market situation and diagnose – albeit tentatively – which uplifts in sales are expected to be explained and by which marketing activity.

The correlation matrix provides information about:

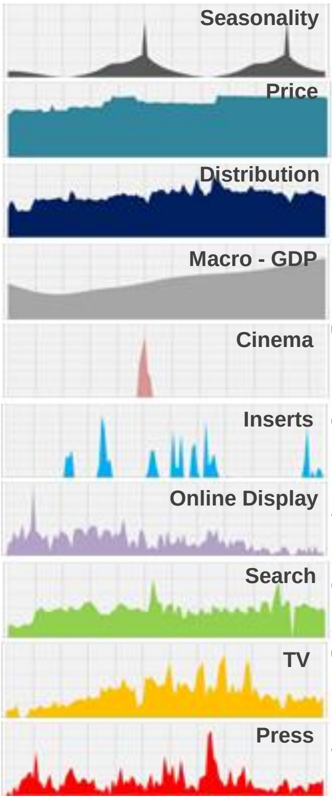
- the correlation between sales and the explanatory factors. This is useful for understanding which variable is expected to come up significant in the econometric model
- the existence of multi-colinearity among the explanatory factors

Data transformations for media variables and control factors

Raw media data are transformed so that they represent the way the media marketing activities work in practice

Raw Data

investment



time

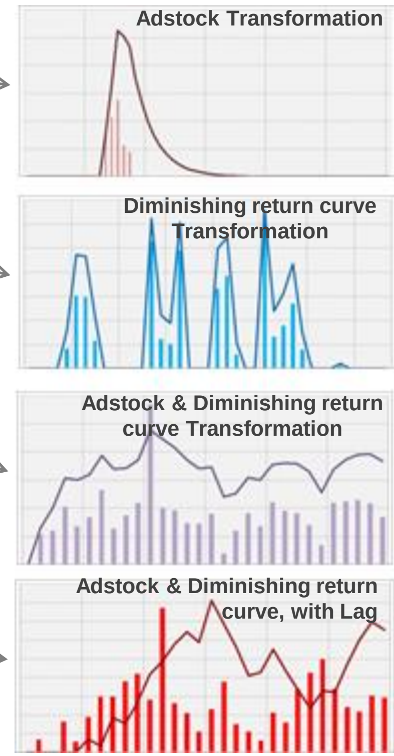
The adstock transformation is given by $A_t = T_t + \lambda A_{t-1}$, where T_t is investment at time *t, and* λ is the decay rate

The diminishing return transformation is given by $D_t = \frac{1}{1+e^{-ax_t}}$ where a is curvature and x is investment

The adstocked diminishing return transformation is given by $D_t = \frac{1}{1+e^{-ax_t}}$ where a is curvature and x is adstocked investment

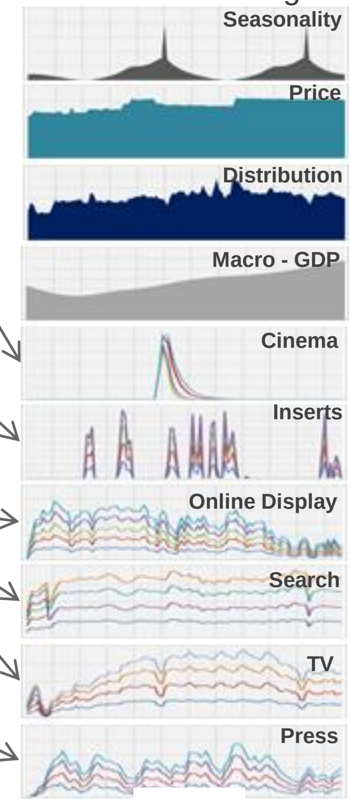
The lagged, adstocked diminishing return transformation is given by $D_{t+\lambda} = \frac{1}{1+e^{-ax_t}}$ where a is curvature and x_t is adstocked investment

Transformation (indicative period)



time

Transformed data used for modeling



time

- Gather granular investment information on marketing activities (preferably in actual metric units).
- Gather supporting media investment information.
- Understand the way each marketing activity works and apply a suitable transformation.

- Apply adstock / diminishing return (DR) / lag transformation as appropriate, so as to reflect the way each marketing driver works in practice.
- Appropriately treat price (smooth promo price or create price gaps and thresholds) so that the way price works on the market is fully reflected.

- The transformations applied will create the time series (explanatory factors) to be used in the econometric models
- The idea is to match the spikes of the transformed variables to uplifts in sales so that the models generate meaningful insight to the highest possible extent.

Multiple versions of the same transformation are produced for each variable. The transformations correspond to different parameters for the lag / curvature and decay rate. This results in many combinations for the variables to be tested – The idea is to test all of them and choose the one which fits the data best.

Model fitting

First pass:

Baselines and incremental sales are estimated via the following state space model:

$$y_t = \mu_{t+} \sigma_\xi^2 + \sum \beta_t x_t + \varepsilon_t$$

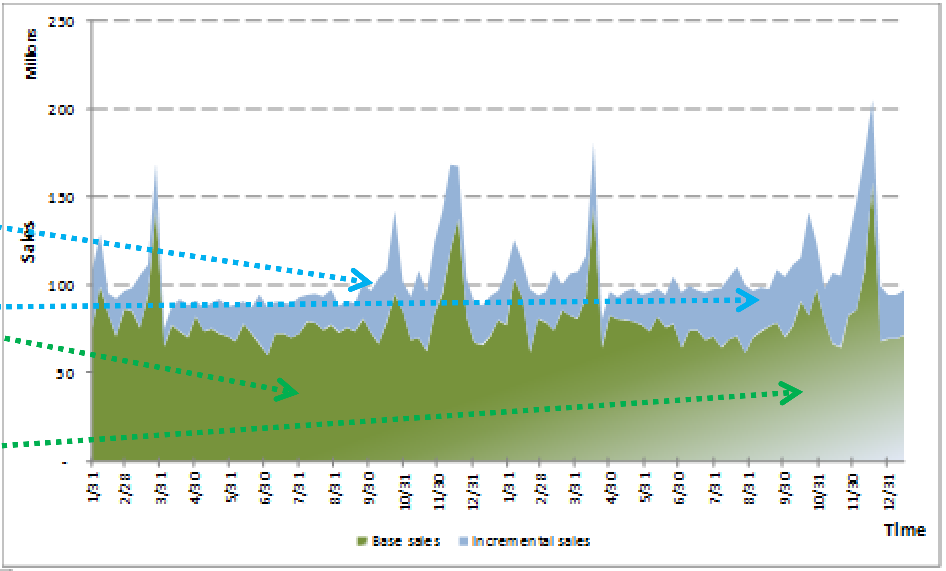
where $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$

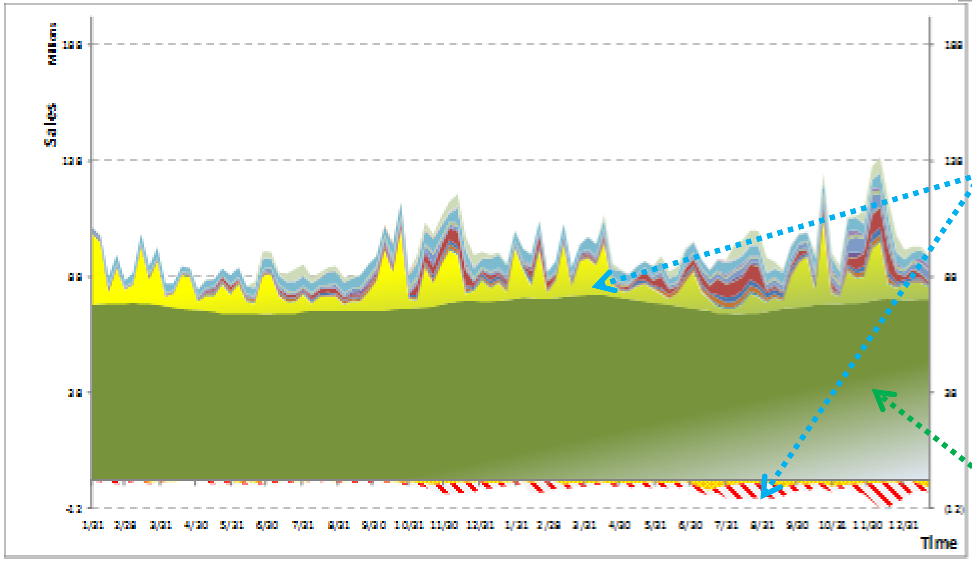
$$a_{t+1} = \begin{pmatrix} \mu_{t+1} + \sigma_\xi^2 \\ \beta_{t+1} \end{pmatrix} = \begin{pmatrix} \mu_t + \xi_t \\ \beta_t \end{pmatrix} \text{ where } \xi_t \sim NID(0, \sigma_\xi^2)$$

Commandeur, Koopman and Ooms (2011), Koopman, Shephard and Doornik (1998) :

where:

- x_t is a $k \times 1$ data matrix,
- β_t (the coefficient vector) s a $k \times 1$ fixed parameter vector,
- σ^2_ξ denotes the variance term (supplied by the user; a potential starting point is the variance of the y_t vector).





The incremental sales are broken down to their corresponding constituent parts as per the corresponding regressor $\sum \beta_t x_t$

Smoothed baseline estimates given by μ_t

Second pass:

- The baseline sales estimate is smoothed via the Kalman filtering recursive algorithm so that $\mu_t + \sigma_\xi^2$ becomes $\mu_t + \xi_t$ from which μ_t is then isolated as the smoothed baseline. The smoothed baseline is exclusively based on past values of the observed time series.
- Smoothing is achieved based on the following sequence of steps:

– A forward pass from $t=1,\ldots,n$

– A backward pass from $t=1,\ldots,n$ using output of the Kalman filter and the state and disturbance smoothers

Model fitting

Create all possible models to be fitted

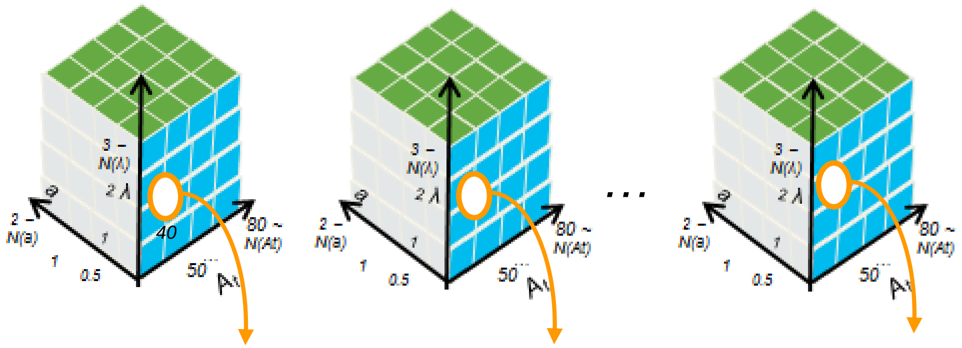
1. We identify the variables that can be fitted

For each of the n variables of our cleansed dataset, all possible combinations of appropriate transformations are created (*):

Variable 1: $N(A_t) \cdot N(a) \cdot N(\lambda) = N_1$

Variable 2: $N(A_t) \cdot N(a) \cdot N(\lambda) = N_2$

Variable n: $N(A_t) \cdot N(a) \cdot N(\lambda) = N_n$



- Any point in the above cube represents a transformation that is applicable to variable n
- In this way each variable is associated with N(n) variants
- A complete model is specified by a unique combination of each of the variables' variants in the dataset
- This results in $N_1 \cdot N_2 \cdot \ldots \cdot N_n$ different candidate models

(*) Exceptions concern:

- ***Competitor type of variables***: Only ad-stock type transformations are applicable
- ***Price, macro-economic factors, control factors and demographics***: Although these variables are used in the model fitting process, the transformations discussed are not applicable. These variables work in a linear form with sales so they are modeled "as is".

2. We group these variables according to their type so that we form all possible models

Dependent	Sales							
Main	Seasonality							
List1	Outdoor 55.2	Outdoor 60.2	Outdoor 65.2	Outdoor 70.2	Outdoor 55.3	Outdoor 60.3	Outdoor 65.3	Outdoor 70.3
List2	TV 60.2	TV 70.2	TV 80.2	TV 90.2	TV 60.3	TV 70.3	TV 80.3	TV 90.3
List3	Radio 50.2	Radio 60.2	Radio 70.2	Radio 80.2	Radio 50.3	Radio 60.3	Radio 70.3	Radio 80.3
List4	Coupons 55.2	Coupons 60.2	Coupons 65.2	Coupons 70.2	Coupons 55.3	Public Relations 50.2	Public Relations 80.2	Public Relations 60.3
List5	Customer Incentives 55.2	Customer Incentives 60.2	Customer Incentives 65.2	Customer Incentives 70.2	Customer Incentives 55.3	Customer Incentives 60.3	Customer Incentives 65.3	Customer Incentives 70.3
List6	Telemarketing 50.2	Telemarketing 60.2	Telemarketing 70.2	Telemarketing 80.2	Telemarketing 50.3	Telemarketing 60.3		
List7	Newspaper Inserts 55.2	Newspaper Inserts 60.2	Newspaper Inserts 65.2	Newspaper Inserts 70.2	Newspaper Inserts 55.3	Newspaper Inserts 60.3	Newspaper Inserts 65.3	Newspaper Inserts 70.3
List8	TV Competitors 60.2	TV Competitors 70.2	TV Competitors 80.2	TV Competitors 90.2	TV Competitors 60.3	TV Competitors 70.3	TV Competitors 80.3	TV Competitors 90.3

- The available variables for modeling are grouped in lists where each variable can be present only once.
- We identify all variables of similar nature so that they can be attributed to the same list
- The number of different driver types determine the number of lists to be used
- We reserve separate room for the variables that need to be present in all models

Model fitting

We choose the best model

5. We comparatively assess all filtered models

6. We choose the final model

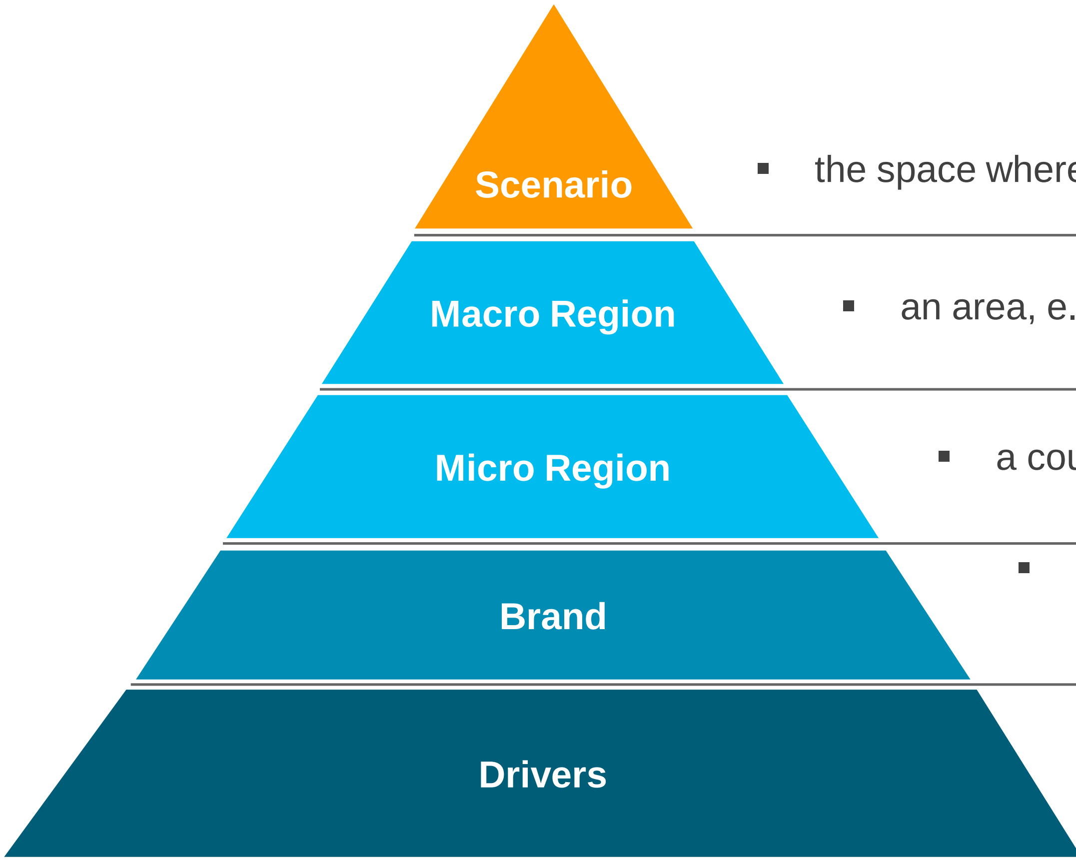
All filtered models are checked for statistical robustness based on a number of different statistical tests:

Statistical Test	*Test Objective*	*Implementation / Test Statistic Used*	*Reasons for selection of this particular test*	*In case the test fails…*
1. R-squared	Provides a measure of the variation that can be explained by the model at the overall level (and not as a summation over all available points)	$1-(\frac{n-1}{n})SSE/RWSSE$	In contrast to the usual OLS *R-squared* it accounts for the seasonal pattern of the dependent variable	In addition to R^2 a objective model selection criterion is to choose a model such that it has the lowest values of AIC and BIC (combined) – as compared to the other models.
2. Durbin-Watson	Detects the presence of autocorrelation in the model residuals	$DW=\frac{\Sigma_{t=2+d}^{T}(\upsilon_t-\upsilon_{t-1}-1)^2}{\Sigma_{t=1+d}^{T}\upsilon_t^2}$	Only the first order serial correlations needs to be tested No lagged dependent is incorporated as an explanatory variable	Since there is serious evidence for *first-order serial correlation,* the coefficient estimates may be inefficient and, thus, the corresponding tests of statistical significance may be inaccurate
3. Q-Stat Test	Checks whether the model residuals are independently distributed	$Q(p,d)=T(T+2)\Sigma_{j=1}^{p}\frac{r_j^2}{T-j}$	A number of simulation studies have proven the superior performance of the Box-Ljung implementation used over the alternative Box-Pierce test	Failing to satisfy a principal modeling assumption, the model is eliminated from further consideration
4. Jarque-Bera Test	Tests whether the assumption that the deriving residuals follow the normal distribution is satisfied (based on comparing the data's kurtosis and skewness with the normal distribution's)	$JB=\frac{n}{6}(s^2)+\frac{1}{4}k^2$	A number of simulation studies have proven the superior performance of the Jarque-Bera test over the alternative Kuiper, Shapiro and Wilk, Kolmogoro-Smirnov and Cramer-von-Mises tests	The following correction actions are undertaken: Investigation of the auxiliary residuals, detection of outliers and correction of the functional form of the dependent or independent variables
5. Two-sided F-test about hetero-skedasticity	Checks presence of unequal variance of model error terms	$H(h)=\sum_{t=T-h+1}^{T}\upsilon_t^2 \Big/ \sum_{t=d+1}^{d+1+h}\upsilon_t^2$	Given the satisfaction of the normality requirement, the *F-test* is robust and avoids the Type I inflations that are generated by the alternative Levene's, Bartlett's and Brown-Forsythe tests	Appropriate transformations of the dependent (*eg*. log-transformations) or independent variables (*i.e*. any non-linear transformation may be applicable), so that any inferences that are made remain accurate. Alternatively the heteroskedasticity-consistent standard errors are used for all inference purposes
6. AIC (i.e. Akaike Information Criterion)	Provides a suitable measure for goodness-of-fit of the estimated model	$AIC=\log(PEV)+2k/T$	Suitable for State Space Models (Cavanaugh and Johnson, 1999; Hurvich, Simonoff and Tsai, 1998)	The AIC test – when considered independently of the BIC can be thought as a replacement of the R^2 measure. In such a case, the model with the lowest AIC is preferable
7. BIC (i.e. Bayesian Information Criterion)	Provides a suitable measure for goodness-of-fit of the estimated model	$BIC=\log(PEV)+\frac{k(\log T)}{T}$	Penalizes the extensive use of free parameters	In case of multiple models with equal value of AIC, among these the model with the lowest BIC is finally chosen

Final selection is made based on: i. AIC, ii. BIC and iii. R-squared

Hierarchy of the optimization – budget allocation

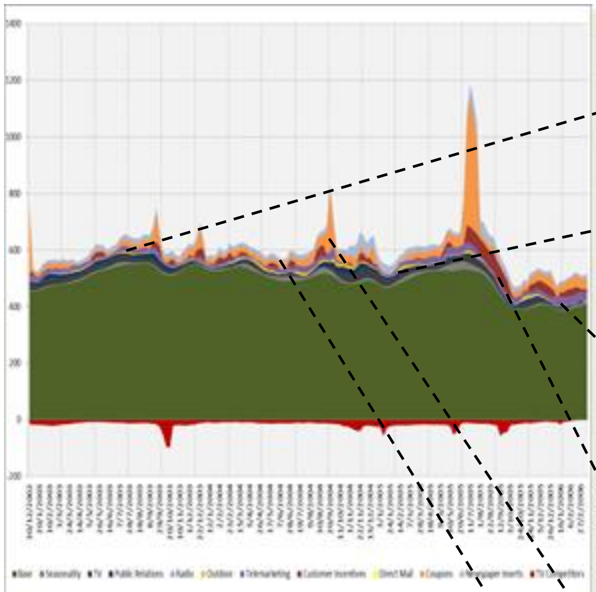
- Budget allocation is conducted along the following data hierarchy (from top to bottom)
- The hierarchy is essentially the order on which to run the optimization and satisfy business needs



- the space where optimization occurs
- an area, e.g. EMEA, or a category, e.g. Tier 1 cities
- a country, e.g. U.K.
- a singular product, e.g. mobile phones, or a collection of products, e.g. under mobile phones: fashion, classic, etc.
- marketing activities such as TV, Press, Outdoor, Radio, etc.

How does the optimization work? – 1

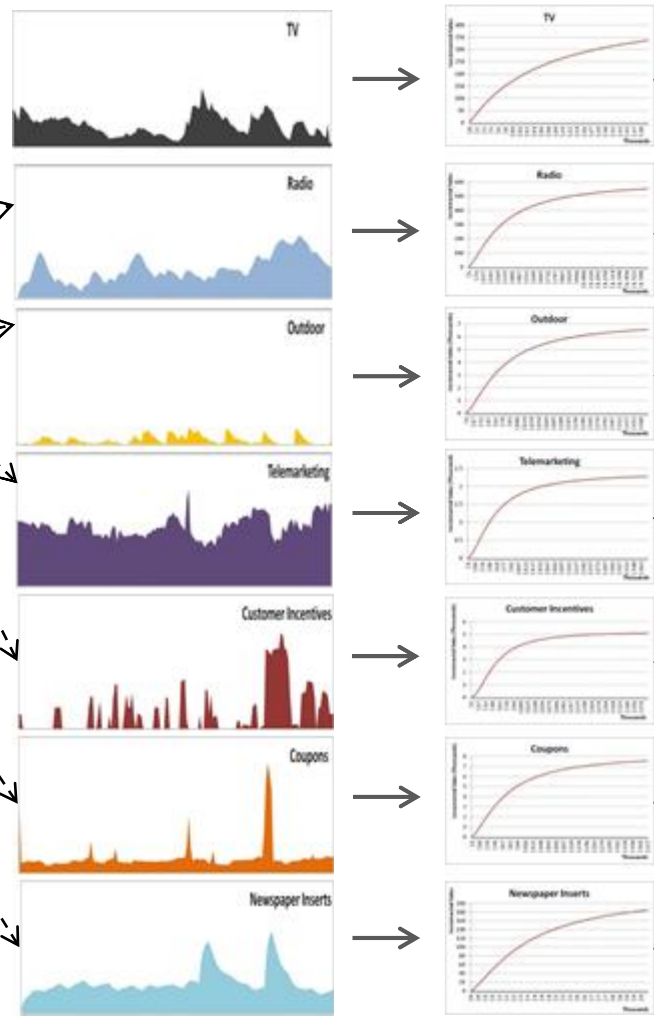
The weekly contributions of media drivers are transformed to the corresponding annual effectiveness curves



The transformation from weekly to annual effectiveness takes special care in repeating the weekly sales uplift pattern, as quantified by the modeling procedure and project that to an annual level.

Weekly effectiveness

Annual Curves



The annual effectiveness curves are subsequently used as an input to the AMAP optimization engine

How does the optimization work? – 2

Our optimization tools help optimize both the portfolio and each campaign

1 Marketing investment is linked to Sales and Profit Contribution

Sales

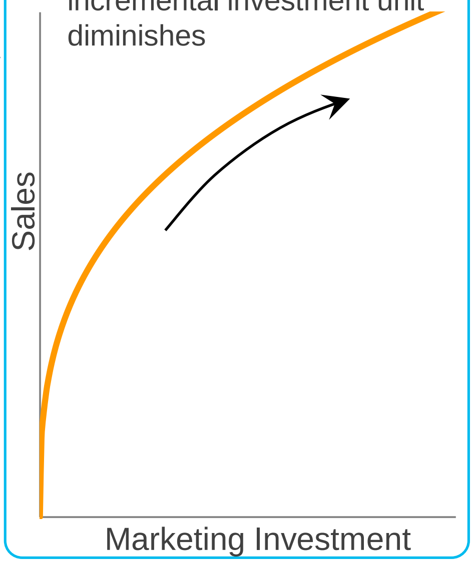
Gross Profit

Marketing Investment

Profit Contribution

2 The link between Marketing and Sales is estimated through response functions

As investment increases sales grows, but return per each incremental investment unit diminishes



3 Value is created by reallocation of marketing resources from investment options with smaller marginal return to options with higher marginal return