Available online at http://www.journalijdr.com

International Journal of DEVELOPMENT RESEARCH

ISSN: 2230-9926

International Journal of Development Research
Vol. 5, Issue, 02, pp. 3454-3466, February, 2015

Full Length Research Article

DERIVING INSIGHT AND INTUITION FROM BUSINESS DATA: REVIEW BY EXAMPLE

***Tesfaye Demissie**

Data Science Statistician, United Kingdom

ARTICLE INFO

Article History:

Received 10th November, 2014
Received in revised form
21st December, 2014
Accepted 02nd January, 2015
Published online 27th February, 2015

Key words:

Predictive Modelling,
Marketing Mix Modelling,
ROI, Business Analysis,
Direct Marketing

ABSTRACT

'Insight' combines several ideas. It includes 'classic' areas, such as knowing who consumers are, what they do, where they are, what they buy, what they would like to buy, what media they are exposed to and what media they choose to view, listen to or read. It also includes more psychological areas – what consumers think and feel, what their objectives and strategies are, and how these influence the way they behave. The list of attributes to examine is long and it should remain in the hands of the researcher or data analyst which variables are retained in the model and which are excluded. Hopefully, these decisions will be based on a thorough knowledge of the phenomenon being modelled. We will often be in the position of either building or interpreting a regression model that has been constructed from a large database consisting of many variables. Having gone through the process of building such a model, we will have a better sense of what sorts of subjective choices are made along the way. Statistical models are an assistant, not a master and this article gives an introduction to the subject by reviewing some of the widely available algorithms and comparing their capabilities, strengths, and weakness in three business area examples: *direct marketing*, *predicting financial indicators,* and *market mix modelling*. In doing so, the article creates 'common language' between analytics personnel (Statisticians, data miners, and computer scientists) and management (personnel with marketing and financial expertise) and narrow the gap that exists between algorithmic reasoning (theory) and practical application such as Return of Investment (ROI) for business decision making.

INTRODUCATION

Data plays a key role and can provide a competitive edge across many different sectors and to many business processes. Practice of business is changing, whether it is traditional or over the internet due to the increasing business data storage. Whether in direct marketing or other medium, there is huge amount of consumer data gathered everyday through for example loyalty card, online auction (such as eBay, uBid, etc.), digital fingerprints (cookies, tracking devices), search engines (google, yahoo, etc.), online retailers (Amazon, Netflix), credit card and credit granting industries (mortgage, bank, insurance). The list goes on and collection of data and analysis is key for the current digitally interconnected business. The term Marketing Mix Modelling (MMM) for instance is widely used and applied indiscriminately to a broad range of marketing models used to evaluate different components of marketing plans, such as advertising, promotion, packaging, media weight levels, sales force numbers, etc. (Hallward, 2008).

****Corresponding author: Tesfaye Demissie***
Data Science Statistician, United Kingdom

These models can be of many types, but multiple regression is the workhorse of most MMM. MMM is an analytical approach that use historic information, such as syndicated point-of-sale (POS) data and companies' internal data, to quantify the sales impact of various marketing activities. Iterated by Wasserman (2006), mathematically, this is done by establishing a simultaneous relation of various marketing activities with the sales, in the form of a linear or a non-linear equation, through the statistical technique of regression or predictive modelling. Regression is based on a number of inputs (or independent variables) and how these relate to an outcome (or dependent variable) such as sales or profits or both. Once the model is built and validated, the input variables (advertising, promotion, etc.) can be manipulated to determine the net effect on a company's sales or profits. Discussed and demonstrated in this paper are the steps in execution of such an analysis, how to evade its major drawbacks, and the benefits that can be derived from the analysis.

METHODS

The main focus here is to emphasise data-driven decision making. In doing so, the mathematical and algorithmic rigors are kept aside from data mining principles to solve business problems in pragmatic way. The approach exposes business decision makers to data-driven decision making concepts and ideas by using data-driven solutions for real-world's practical business scenarios, especially in the areas of direct marketing, market mix modelling and predictive modelling of business data. Each area will be covered with separate case study to stress their unique relevance to business performance. The techniques are beneficial enough to be applicable to sectors such as banking, insurance, investments, retailing, electronic commerce, advertising, and direct marketing.

The most commonly used statistical functions will be discussed together with their practicality. A complete data analytics steps will be exercised starting from numerical and graphical ways of understanding the data through data exploration, commonly known as exploratory data analysis (EDA). This later step helps understanding details of the data such as its pattern, trend, unusual recordings, outliers, etc. Once we grasp the clear picture of the data, we follow to learn about the data using modelling. This stage is ideal to learn the concept of 'model' in statistical term and why we need it. The most popular and widely used statistical technique for estimating and modelling, ordinary least square (OLS), will be covered together with the interpretation and evaluation of the model (statistical significance and practical relevance technique). The concept of diminishing returns (model saturation) versus marketing efforts will be covered including weeding out bad information from the model. Finally we will experience techniques of measuring explanatory power and predictive capability of a model including model's capturing ability of complex relationship in the data.

EDA Techniques: Business Data Perspective

Being familiar with the data is important and one shouldn't be tempted on the availability of models and methods without first performing EDA. EDA signifies the importance of models and methods on capturing the essence of the data not their complexity. Clearly put in (Johannes Ledolter, 2013), data knows more about the modelling and predicting process we are deploying and EDA is about sniffing for the knowledge from data. The most popular EDA techniques include drawing summary statistics for numeric data types (mean, median, mode, standard deviation); exploring distribution of variables using frequency tables for categorical variables, boxplots and histograms for numerical variables; and exploring pairwise relationships or correlations among variables using scatterplots (graphical method of investigating relationship – subjective measure), correlation tables (such as Pearson's correlation for assessing strength and direction of *linear* relationship between variables – objective measure) and cross-tabulations (correlation for categorical data). Correlation between numerical and categorical variables cannot be computed directly (Michael H Kutner *et al*., 2005). There are also advanced graphical EDA techniques such as scatterplot matrices, trellis graphs, density and spine plots, spatial graphs, time series graphs, etc. Both basic and advanced EDA explorations should be used simultaneously, as one informs the other and their joint and simultaneous application leads to a better insight about patterns and abnormalities in the data.

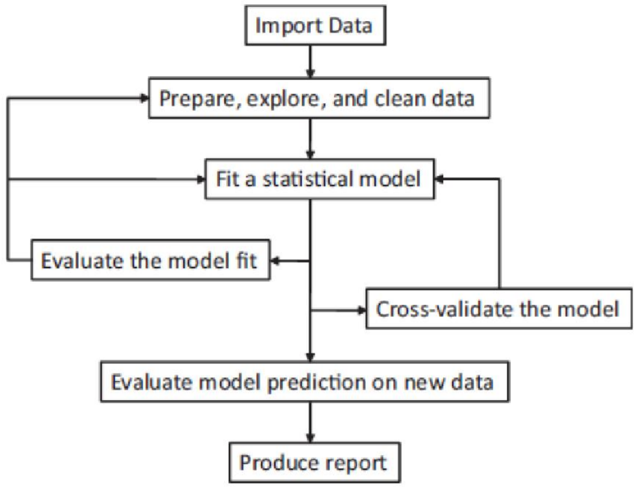

Figure 1. Modelling Process

Data Transformation and Trellis Graphs: The main goal of these tools is to answer question such as why some customers spend more than others. Is there a relationship between direct marketing tool such as catalogues and expenditure? Investigate customer's purchasing behaviour and check if those who spend more in the past will spend more in the future (for customer retention). They are capable of detecting and target 'pockets' of potential customers (application of promotions and coupons). These are advanced and powerful ideas of EDA, especially for mining large datasets. Transformation is carried out to obtain consistent, more linear relationship of patterns. On the other hand, trellis graph allow conditional view about data. They divulge unsuspected relationship in sub-segments (pocket) of data. It allows for a more granular inspection of the data and for the discovery of new segments specific relationship, valuable tool for marketing decision: could either lead to more specific, custom-made marketing or advertising, or it could lead to abandoning entire market segment altogether.

Scatter plot matrix: is a single graph to show relationship between many variables; and nonlinear effects (skewed distributions in histograms or funnel effects in scatterplots) can be smoothed out using suitable transformation to the data (logarithmic, inverse, exponential, quadratic or power function transformation).

Density plots**:** being similar to histogram, mainly used for categorical input variables and represents the data distribution in the most granular form. In a C-2-C relationship, this tool can be used to distinguish between bad and good loans.

Spine plots**:** this are used to visualise cross-tabulation. Categorical responses occur frequently in business data, especially in marketing where we study customer's choices and preferences. While traditional scatterplots are not very useful for exploring data with categorical responses, density plots and spine plots can unveil new, previously unknown knowledge.

Graphs for panel data**:** these are graphs for data that occurs from following a set of customers over time and record their behaviour and performances. They allow analysis to understand purchasing patterns. They are vital in helping business target individual customers with tailored coupons and promotions.

Overlaid time series and aggregating: panel data are essentially a bunch of individual time series – one series for each individual observation. Plotting all those time series into one single graph leads to information overload and clutter. Exploring all the variable data at once leads to more information than the graph can carry, as a result we don't learn much at all. In such cases, we aggregate data for better insight. Aggregate data by either temporal or cross-sectional component, and then graph it. It is standard graph plotted in an innovative way to take advantage of the special structure of panel data.

Aggregate the cross-sectional information*:* to explore trends over time, keeping temporal information intact. Aggregate the temporal information to learn variation from one customer to another, for example. More customers means panel data are challenging and we have to be careful about how best to extract the kind of knowledge that supports one's business goals.

RESULTS AND DISCUSSION

What is a Model?

A model is an abstraction or an approximation of reality. It is important to meet the desire to learn from the past to predict the future, prevalent desire in business for decision making. Good business decision is a culmination of good model, common sense and knowledge expertise. Models allow to separate unpredictable (noise) from predictable (pattern) of data. EDA gives qualitative nature of a relationship but wouldn't give a quantified relationship of patterns in variables, models do. Obtaining a model using data is possible with click of a mouse. Interpretation of model and its parameters is a key for decision making (coefficients and intercept). Least square regression determines the best model by defining a line that minimizes the sum of squared residuals (smallest average distance to all data points). R-squared measures the overall quality of the regression model – it is a measure of how well the model captures the actual data variability.

There are two sources of uncertainty: the variability (uncertainty) left after fitting the model (sum of squared error = SSE) and the overall variability (uncertainty) in the data (sum of squared total = SST). R-squared is the ratio of the difference (SST-SSE = SSR) to the overall variability (SST) (Johannes Ledolter, 2013; Julian J. Faraway, 2000 and Max Kuhn and Kjell Johnson, 2013). R-squared has to be evaluated in context and there is no single bench mark that applies to all situations. In most cases an adjusted R-squared that penalizes the models for inclusion of a nonsense variable is used for comparison of models. The strength of predictor's impact relative to its uncertainty is measured using 'signal-to-noise' ratio = estimate over standard error of the estimate, usually knows as statistical usefulness of the predictor. As clearly put by Lodolter (2013), the significance of the signal-to-noise ratio is tested using statistical significance measures and p-value in order to quantify exactly how good it is. The statistical concept of p-value is that it is a measure of probability that, given a particular set of data, the observed signal could have occurred simply due to random chance. The smaller this probability (<=0.01, 0.025, 0.05), the more confidant we are that the observed signal is 'real' and hence the variable can be a good candidate for the model. Statistical significance or usefulness has to be accompanied by confidence intervals to allow us judge practical importance of a variable in the presence of uncertainty.

Flexible Models: Introduction of dummy variables and interaction terms

Dummy variable are binary variables as a result of recoding data. Incorporating dummy variables into a regression model renders more flexible modelling options. Using a dummy variable together with another numerical predictor, we can model data trends that resemble parallel lines, that is, lines with the same slope but a different intercept (Max Kuhn and Kjell Johnson, 2013). Interaction terms are result of multiplication of two variables. Interaction terms alleviate the assumption of linearity. Relationship between two variables can change when introducing a third variable- Simpson's paradox. Models with a dummy variable, a numerical variable and the interaction between both variables allow us to model data trends that resemble non parallel lines. In other words, such models allow us capture trends with different slopes and intercepts; control for as many variables as we like and thus get a better understanding of the true causal relationship between two variables. If there exist a non-linear relationship in a data, we have to identify suitable data transformation; and apply linear regression method to the transformed data values. While the basic principles of linear regression remain intact, special attention has to be paid to the interpretation of the resulting coefficients. For example, if we deployed logarithmic transformation to earn linearity, the model's intercept value is interpreted as elasticity between input and output values (Jean Paul Isson and Jesse Harriott, 2013 and Johannes Ledolter, 2013).

When modelling is used to arrive at most compelling decisions, we have to apply several different ideas and concepts as discussed such as introduction of dummy and interaction terms, EDA, and transformation. As put by Faraway (2000) and as a thumb rule choosing the best fit model lies on model fit statistics such as the value of R-squared or Adjusted R-Squared, F-statistic and residual standard error, AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion). R-squared measures the proportion of the total uncertainty explained by the model. Adjusted R-squared is similar to R-squared, but penalizes the model for inclusion of too many useless predictors. Residual standard error measures how tightly the data fits around the regression line. AIC measures the overall fit of a model and penalizes it for too many useless predictors. Similar to AIC, BIC also penalizes a model for using too much data. The interpretation of AIC value corresponding to a predictor variable is in such a way that removal of the variable from the model will lead to that specific AIC measure. Smaller AIC or BIC is usually better. Higher R-squared, Adjusted R-squared, F – statistic and smaller residual error are considered better. If models tend to give identical quality measures, choose the simpler model, model with less number of variables (principle of Parsimony). If most of the variables are insignificant, modification of the model by dropping insignificant predictors in level of insignificancy is recommended.

Multicollinearity and Variable selection

In the current digital age, it is becoming increasingly easy to collect data about market and ourselves. However, every piece of information is not useful for modelling and we have to weed out irrelevant information from regression model (Rahul Saxena and Anand Srinivasan, 2013). In the process, it is important to distinguish between information that enters a model and what is left out. It is customary to start from what we observe from R-squared measure and p-value of the regression model. R-squared is interpreted in such a way that, while there may be additional factors that also affect a response variable, variables included in the model captured R-squared times 100% of all the possible variations in the response variable. Similarly, p-value answers the question: are the variables included statistically important? Multicollinearity refers to the existence of two or more correlated independent variables in the model and results in using the same information twice. It is important to include only predictors that are uncorrelated with one another but highly correlated with the response variable (weed for small number of variables using correlation table and scatterplots).

The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in a linear regression. High multicollinearity must be avoided since it produces wrong coefficients and t-statistic. It is calculated for each explanatory variable. To calculate a VIF for an explanatory variable X_i, we can run an ordinary least square regression that has X_i as a function of all the other explanatory variables in the equation. Then, calculate the VIF factor for X_i using the formula $1/(1-R^2)$ where R^2 is the R^2 of the regression. This procedure is repeated for each explanatory factor and a VIF greater than 3 signals high multicollinearity (Michael H Kutner *et al*., 2005). Cut off value for correlation is subjective and should be dealt with care. However, curing for multicollinearity becomes daunting especially when several predictors are involved.

Forward Selection: Start modelling with one predictor variable and successively add predictors according to their relevance. Relevance of predictor variables is established using its correlation with response variable, R-squared measure (inclusion increases R-square if relevant), its statistical significance (lower p-value), and its contribution to alternative measures of model fit (AIC & BIC).

Backward Elimination*:* Start with the largest model possible and gradually remove individual predictors based on their relevance.

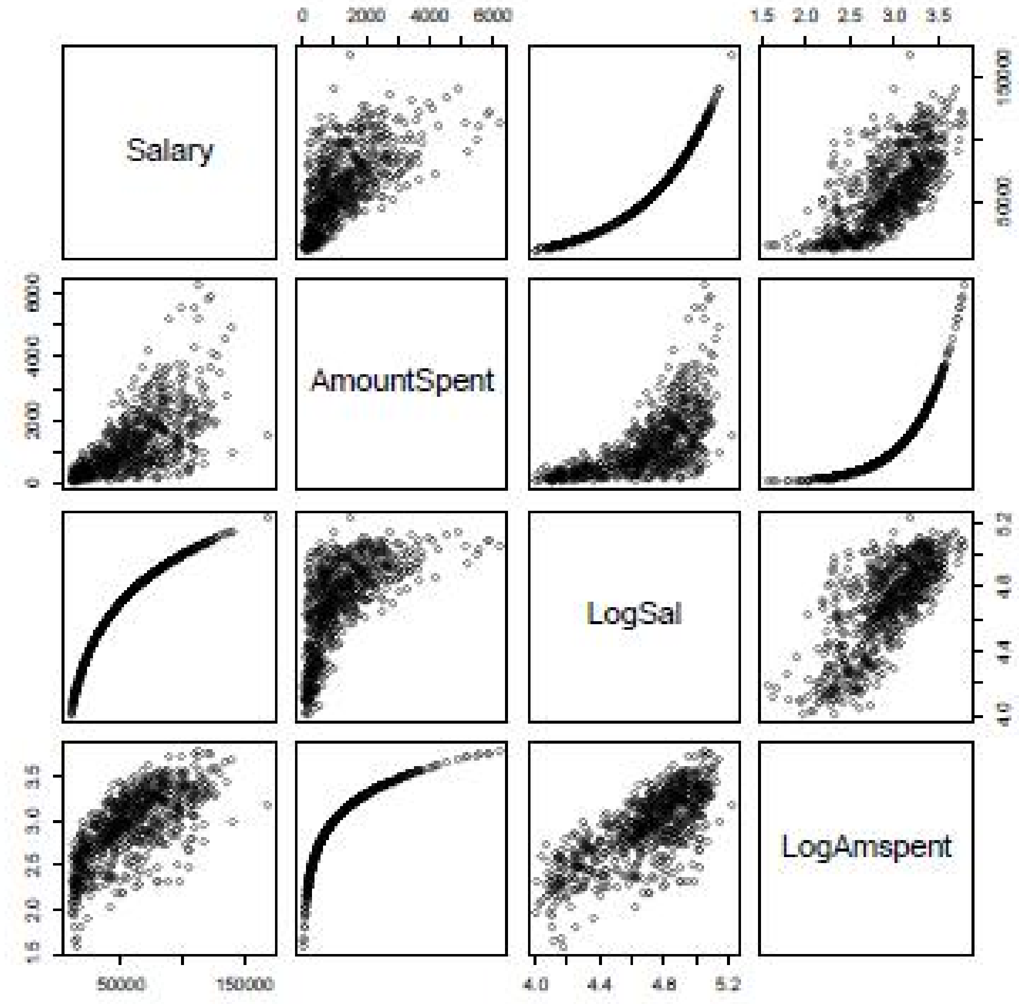

Figure 2. Scatterplot of marketing data

Stepwise Regression: Building regression model in small, careful steps rather than in large, uncontrollable chunks is used to cure multicollinearity through a procedure known as Stepwise regression. Stepwise regression (combination of forward selection and backward elimination) can automatically reduce the number of variables to a smaller, more manageable subset and makes sure variables included in the model do not post problems with respect to multicollinearity. However it does not guarantee optimality in itself and has to be seen in conjunction with EDA and domain knowledge for valuable and reasonable result. Further stepwise regression assumes that the input data follows linear trends and cannot identify patterns of nonlinear relationship. Data transformation will cure the nonlinearity and establishing linearity might be possible on the transformed variables and correlation (Pearson's correlation) measures the strength of the linear relationship between variables (correlation will not tell much about the strength of the shape if variables indicate nonlinear pattern).

Example scenario1: Direct Marketing: We recall the different EDA tools and apply to direct marketing data (1000 records) that contains ten different pieces of information about a customer's (Johannes Ledolter, 2013). These include customer's demographics, customer housing status including distance from a store selling similar items, and the customer's salary, purchase history, number of catalogues received, and the amount of money spent at the store. In particular, we are interested in carefully scrutinizing some of these ten pieces of information: the customer's salary, location, and the impact of both on the amount of money spent at the store. Also, we are interested in finding answers to the following questions:

How does salary relate to a customer's spending? Is the relationship linear? Or does a nonlinear model better capture the impact of salary on a customer's spending pattern? How does a customer's location relative to a store (i.e., close vs. far) impact his spending? Do customers who live *farther* from a store that sells similar items spend *more*? Does the rate of a customer's spending depend on his location? In other words, do customers who live farther away spend their money at a *faster* rate?

The figure shows a scatterplot for the marketing data (salary and amount spent) on the logarithmic scale. We can see that the plot of LogSal vs. LogAmspent seems evenly scattered. In fact, it appears to follow a very nice straight-line pattern. In other words, it is very reasonable to apply a linear model to the transformed rather than untransformed data which experiences funnel (spread out variability).

Table 1. Linear model estimates for direct marketing data

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	-2.160421	0.150415	-14.36	<2e-16 ***
LogSal	1.066792	0.031998	33.34	<2e-16 ***
SalLocInteraction	0.060852	0.004134	14.72	<2e-16 ***

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2331 on 694 degrees of freedom.
Multiple R-squared: 0.6561, Adjusted R-squared: 0.6551
F-statistic: 661.9 on 2 and 694 DF, p-value: < 2.2e-16

From the model, the coefficient of the interaction term equals 0.060852.What does this imply? We start again by considering the case where the customers live close to store. If the customers are close, then the interaction term becomes zero (dummy 0=close, 1= far) and all that remains is the coefficient of LogSal; it equals 1.066792. Recall that both amount spent and salary are log-transformed, so this value needs to be interpreted in terms of its spending elasticity: for customers who live close, every 1%increase in salary results in a 1.066792% increase in spending. The insight above applies only to customers who live close to a store that sells similar items. How about those who live far away? For those customers, the interaction term reduces to 0.060852x LogSal x1, so the salary slope for customers who live far away adds up to: (1.025344+0.060852) = 1.086196. In other words, for customers who live far away, every 1% increase in salary results in a 1.086196% increase in spending which is 0.02% higher than that of customers who live closer.

Assessing the Quality of a Model: Predictive Power versus Model Fit

When it comes to business data, a model is necessary due to two main reasons: to capture patterns and trends of data in the past; and to predict the future. Both are vital ingredients for business decisions (Lawrence S. Maisel and Gary Cokins, 2014). The performance of the model in capturing past patterns is evaluated against what is known as model fit statistics. To get a hierarchical analysis of variance table corresponding to introducing each of the terms in the model one at a time, in the same order as in the model formula, we can use the standard analysis of variance (ANOVA) measure; or alternatively plot the model that produces a set of four plots: residuals versus fitted values, a Q-Q plot of standardized residuals, a scale-location plot (square roots of standardized residuals versus fitted values, and a plot of residuals versus leverage that adds bands corresponding to Cook's distances of 0.5 and 1. On the other hand, the predicting power of a model is evaluated using a prominent data mining technique such as training and test set data that can be measured by root mean square error (RMSE) and mean absolute error (MAE).

Regression Trees and Nonparametric regression

In most cases, business processes do not behave in a linear fashion. Example of such business cases include the law of diminishing or increasing return where additional unit of variable input yields smaller and smaller increase in output. Such examples include the effect of advertising on sales after reaching saturated level (decreasing return) and effect of experience on salaries (increasing return) (Rahul Saxena and Anand Srinivasan, 2013). Learning curve is another example where increase in retention of information is sharpest after some initial attempts and then gradually evens out. Also Moore's Law states that the number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years (grows at an exponential rate). In all the above cases, to earn linearity, the suggested and widely used approach is variable transformation. Plenty of transformation analytical function exist and the approaches often are a trial and error procedure and there is a potential of missing out on some potentially important functional relationship due to simply unawareness of alternative functions. Often, the trial and error approach gets cumbersome when several variables are investigated. Nonparametric regression and Regression trees promise to uncover (and subsequently model) complicated interactions in the data in an automated fashion (Johannes Ledolter, 2013). The methods do not make 'parametric' (normality) and any functional relationship assumptions (linearity) between the response and the predictors. Regression trees are computationally intensive methods that can guide about which variable to include in the model. They are simplistic and excellent choice for initial data inspection. They give a clear picture of the structure of the data and provide highly intuitive insight into the kinds of interactions between variables. Nonparametric regression also assume that there exists some relationship and approximation of the relationship is controlled by the investigator, that is, an investigator chooses between very close approximation and crude representation. The only disadvantage is that both methods do not allow any insightful and intuitive interpretation due to luck of linearity assumption in terms of quantifying the relationship.

Interpretation is subjective, not objective which ideally is what decision makers are looking for. The approaches are powerful at providing accurate prediction and insight into past relationships which outweighs the limitation with respect to objective intuition and insight (Bramer, 2013). The idea of nonparametric regression is that rather than fitting a single function to the entire data, method suggests to split up the data into convenient sub segments and then fit different functions to each segment separately known as piecewise regression modelling. The shortcomings are segments of regression lines are not connected, or couldn't connect smoothly; and flexible functional relationship cannot be established. As discussed in (Wolfgang Jank, 2011), the suggested solution is to use what is known as smoothing splines which implements piecewise polynomial that accommodates polynomial function of any order and assures that pieces are combined in a smoothed and continuous fashion. It contains penalty terms to assure that the resulting function is not too variable.

The original dataset has 4499 stock price records with 73 variable fields. Financial analysts and investors are often interested in understanding the relevance of accounting information in explaining (and predicting) stock returns. In particular, an analyst may be interested in determining how a company's profitability, liquidity, leverage, market share, size, or cash flow affect its market value. The point raised in (Johannes Ledolter, 2013 and Stephen G. Cecchetti, 2015) and one problem with such an analysis is that there is often an abundance of accounting information available. In particular, besides information obtained directly from a company's financial statements, such as debt, cash, or revenues, analysts often compute ratios based on the information found in financial statements.

These ratios could include profitability ratios such as gross margin or profit margin, liquidity ratios such as the current ratio or operating cash flow ratios, activity ratios such as asset turnover or stock turnover ratios, debt ratios such as debt-to-equity ratio or the debt service coverage ratio, market ratios such as earnings per share or enterprise value relative to net sales, and very many more indicators. The point is that financial information is plentiful. Moreover, financial ratios combine individual pieces of information from financial statements into new variables. As a result, many of the financial indicators and ratios are correlated. In the following, we discuss modelling of a set of data containing 24 different financial indicators and ratios. From a statistical point of view, this subset even still brings up problems of multicollinearity and, as a result, challenges with respect to smart variable selection. The business is interested in the future performance of investment and base its business trading decision on anticipation of future changes in stock price. In other words, we would like to find a model of the form:

Stock Price = $\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k$, where X_i are some (or all) of the 23 predictors.

We know we should perform EDA examination on the variables and visualize the distributions before proceeding, but

Fit

N	4499
Equation	Stock Price = 23.3 - 0.0007877 Total Debt + 0.001063 Cash - 0.009203 Reinvestment Rate + 0.6592 ROE - 0.0002402 ROC + 0.2664 Net Margin + 0.0008021 Invested Capital + 0.0001214 BV of Assets + 0.003251 Net Income + 1353 EBIT - 1353 EBITDA - 2.389e-05 FCFF - 0.3014 Cash as % of Firm Value + 0.04706 Cash as % of Revenues - 15.49 Cash as % of Total Assets + 0.002117 Capital Expenditures + 1353 Depreciation + 8.039e-05 Trailing Revenues + 3.944e-05 Trailing Net Income + 1.943 Intangible Assets/Total Assets + 14.73 Fixed Assets/Total Assets + 0.000552 Market D/E - 22.66 Market Debt to Capital
R^2	0.063
R^2 adjusted	0.058
AICc	32174.166
SE of fit (RMSE)	35.6237

***Examples scenario2: Predictive Modelling of financial indictors*:**

The example case study is based on real-world example of complex financial indicator data (Johannes Ledolter, 2013). for the sake of brevity, and avoid repetitions we'll skip it and let the model give us an insight. We can see the adjusted R^2 are very small which triggers suspicious situation towards the relative importance of the variables in the above saturated model. Also the p-values of most of the predictor variables are

Table 2. Parameter estimate values of the linear predictive model

Parameter	Estimate	95% CI	SE	VIF	p-value
Constant	23.30	20.02 to 26.58	1.6741	-	<0.0001
Total Debt	-7.877 E-04	-0.001405 to -1.708 E-04	3.1467 E-04	29.33	0.0123
Cash	0.001063	1.450 E-04 to 0.001981	4.6822 E-04	5.61	0.0232
Reinvestment Rate	-0.009203	-0.1129 to 0.09449	0.052893	10.41	0.8619
ROE	0.6592	0.2610 to 1.057	0.20307	1.04	0.0012
ROC	-2.402 E-04	-0.04345 to 0.04297	0.022039	1.00	0.9913
Net Margin	0.2664	0.02799 to 0.5049	0.12163	1.53	0.0285
Invested Capital	8.021 E-04	1.718 E-04 to 0.001432	3.2152 E-04	45.05	0.0126
BV of Assets	1.214 E-04	2.064 E-06 to 2.408 E-04	6.0895 E-05	8.97	0.0462
Net Income	0.003251	-0.001560 to 0.008062	2.4540 E-03	18.34	0.1853
EBIT	1353	-3590 to 6297	2521.4	1.391809 E+14	0.5915
EBITDA	-1353	-6297 to 3590	2521.4	2.088120 E+14	0.5915
FCFF	-2.389 E-05	-1.385 E-04 to 9.078 E-05	5.8490 E-05	10.56	0.6830
Cash as % of Firm Value	-0.3014	-0.7539 to 0.1512	0.23084	1.01	0.1918
Cash as % of Revenues	0.04706	-0.03810 to 0.1322	0.043436	1.49	0.2787
Cash as % of Total Assets	-15.49	-21.58 to -9.398	3.1075	1.80	<0.0001
Capital Expenditures	0.002117	-0.001872 to 0.006106	2.0349 E-03	8.04	0.2982
Depreciation	1353	-3590 to 6297	2521.4	9.999270 E+12	0.5915
Trailing Revenues	8.039 E-05	-1.019 E-04 to 2.627 E-04	9.2968 E-05	5.92	0.3872
Trailing Net Income	3.944 E-05	-0.002469 to 0.002548	1.2796 E-03	5.73	0.9754
Intangible Assets/Total Assets	1.943	-4.566 to 8.453	3.3204	1.38	0.5584
Fixed Assets/Total Assets	14.73	9.074 to 20.38	2.8838	1.63	<0.0001
Market D/E	5.520 E-04	-0.006926 to 0.008030	3.8145 E-03	1.02	0.8849
Market Debt to Capital	-22.66	-27.94 to -17.39	2.6900	1.30	<0.0001

not significant, which is quite enough reason to consider some sort of transformation to weed for not necessary variables. Now we need to start winnowing down our model to a minimal adequate level. The least significant of the slopes in the above summary table was that for "ROC", so we are going to toss out "ROC" first.

Update (model. ~.-ROC))

The next variable that turned out to be not significant is "Depreciation". We repeat the above process in a stepwise approach until we are satisfied enough for only significant variables to remain in the model. At each step perform the analysis of variance test (ANOVA) for the consecutive models and check for the significance of variable omitted (Anova (model1, model2)). After updating for 14 times, the following variables appear to be significant to remain in the final model.

Table 3. Variables of the final model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NetMargin	1	23803	23803	13.684	0.000234 ***
FCFF	1	698	698	0.401	0.526627
CashasofFirmValue	1	18459	18459	10.612	0.001180 **
CashasofRevenues	1	6887	6887	3.960	0.047007 *
TrailingRevenues	1	10074	10074	5.792	0.016368 *
IntangibleAssetsTotalAssets	1	1512	1512	0.869	0.351519
FixedAssetsTotalAssets	1	11038	11038	6.346	0.011994 *
Residuals	679	1181084	1739		

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

However, the correlation between each pair of variables (abbreviated names) look extremely high. This seems to agree to the scatterplot (not shown here), and is an indication that none of the pairs show a clear linear data pattern. Recall that we have applied stepwise regression to the data above.

Stepwise regression assumes that the input data follows linear trends. Looking at the many non-significant correlation figures, was it really justified to apply stepwise regression to this data? Or, should we rather have first transformed the data appropriately in order to render more linear relationships?

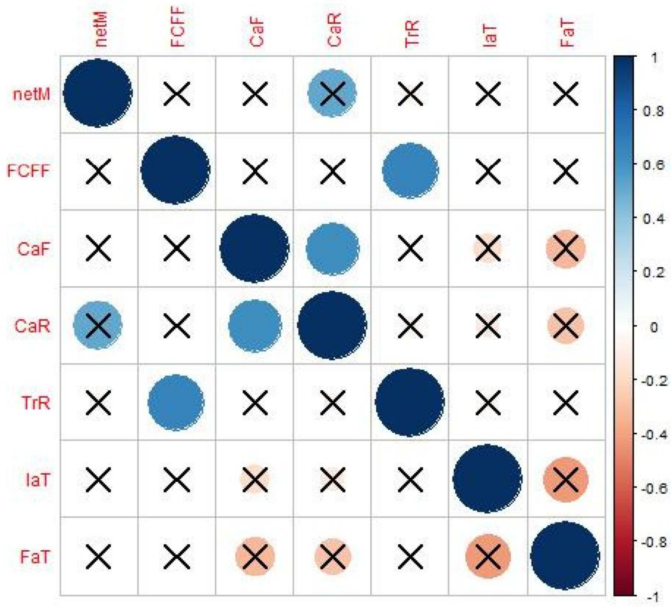

Figure 3. Correlation values of financial indicator data

The stepwise regression procedure is carried out on the log-transformed financial indicator data to have minimal set of variables. Examining residuals is a key part of all statistical modelling. Carefully looking at residuals can tell us whether our assumptions are reasonable and our choice of model is appropriate. Residuals can be thought of as elements of variation unexplained by the fitted model. Since this is a form of error, the same general assumptions apply to the group of residuals that we typically use for errors in general: one expects them to be (roughly) normal and (approximately) independently distributed with a mean of 0 and some constant variance. The plot of the residuals from the transformed data seem to adhere to this assumption (Q-Q Plot assess normality of the residuals, for example). Therefore, in addition to considering other measures, we can comfortably conclude that the model fitted on the transformed data performs better.

The analytics performed included stepwise regression modelling to select predictive variables. Several variables indicated weak correlation to stock price and logarithmic transformation is carried out. Selection of optimal model is made after comparing the performance of stepwise regression, regression tree and nonparametric regression models that captures remaining nonlinearity. Overall performance of models were compared on the original data and transformed data to arrive at the final model. It appears that kitchen sink (saturated) model is loser model with stepwise and nonparametric model estimations indicating some model improvements.

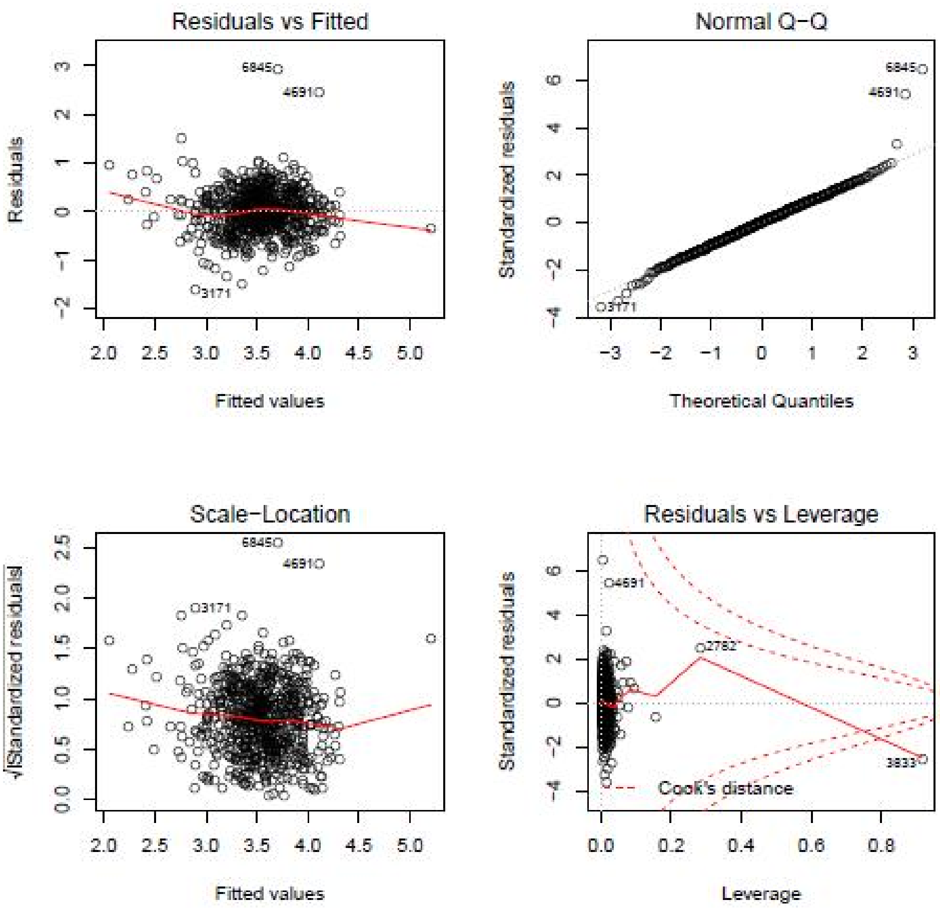

Figure 4. Residual plot

The final estimates of model coefficients therefore becomes:

Table 4. Estimated coefficients of transformed model

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.32828	0.45336	9.547	< 2e-16 ***
log (TotalDebt)	0.04520	0.01095	4.128	4.12e-05 ***
log(RevenuesLastyr)	-1.20264	0.22130	-5.434	7.68e-08 ***
log(NetIncome)	0.21990	0.02385	9.219	< 2e-16 ***
log(FCFF)	-0.11658	0.04566	-2.553	0.0109 *
log(CashasofFirmValue)	-1.62125	0.28984	-5.594	3.23e-08 ***
log(CapitalExpenditures)	-0.08838	0.02027	-4.361	1.50e-05 ***
log(TrailingRevenues)	1.15154	0.22464	5.126	3.86e-07 ***
log(IntangibleAssetsTotalAssets)	-0.52685	0.13067	-4.032	6.16e-05 ***
log(MarketDE)	-0.49646	0.08248	-6.019	2.88e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4561 on 677 degrees of freedom
Multiple R-squared: 0.3592, Adjusted R-squared: 0.3506
F-statistic: 42.16 on 9 and 677 DF, p-value: < 2.2e-16

The winner is the nonparametric regression model below showing significant model performance both in capturing past patterns and predicting the future.

Log (StockPrice) = 4.33 + 0.05log (TotalDebt) – 1.20log (RevenuesLastyr) + 0.22log (NetIncome) -0.12log (FCFF) – 1.62log (CashasofFirmValue) – 0.89log (CapitalExpenditures) + 1.55log (TrailingRevenues) – 0.53log (IntangibleAssetsTotalAssets) .0.5log (MarketDE).

Market Mix Modelling

MMM combines advanced statistical analytic techniques and econometrics principles with marketing science to objectively measure the relative productivity of a complete set of marketing creativities to produce momentary business sales. It defines the effectiveness of each of the marketing elements in terms of its contribution to sales-volume, effectiveness (volume/revenue generated by each unit of effort), efficiency (sales volume generated divided by cost) and ROI (Kormanik, 2010). These findings are then adopted to adjust marketing tactics and strategies, optimize the marketing plan and also to forecast sales while simulating various scenarios. In real terms, the extended 7 P's (Process, Physical evidence, and People included) are factors driving sales although bases sales and external factors also contribute to overall performance. MMM involves understanding a market and its data, data collection and visualization. It involves application of statistical methods to measure the impact of media investments, promotional activities and price tactics on sales or brand awareness (Ataman *et al*., 2010). Further the model is one of the vital tools to assist and implement a marketing strategy by measuring effectiveness (contribution of marketing activities to sales), efficiency (short term and long term Return On Investment-ROI of marketing spend), price elasticity, and impact of competitors.

Figure 5. The 7 P's of MMM

Trade promotion is a key activity in every marketing plan. It is aimed at increasing sales in the short term by employing promotion schemes which effectively increases the customer awareness of the business and its products. The response of consumers to trade promotions is not straight forward and is the subject of much debate. Non-linear models exist to simulate the response. As argued by Pratt, Malabe and Blake (2010), using MMM we can understand the impact of trade promotion at generating incremental volumes. It is possible to obtain an estimate of the volume generated per promotion event in each of the different retail outlets by region. Sponsorship is a cash and/or in-kind fee paid to a property (typically in sports, arts, entertainment or causes) in return for access to the exploitable commercial potential associated with that property. For example, a corporate entity may provide equipment for a famous athlete or sports team in exchange for brand recognition. The sponsor earns popularity this way while the sponsored can earn a lot of money. A particular form of specialized brand sponsorship where a brand sponsors an unusual event or pastime that then becomes synonymous with that brand (to the point where future brands may be excluded from participation) is known as 'about sponsorship'. This provides a strong walled-garden sponsorship relationship between particular events and the brand. In general, price changes of the brand impacts sales negatively. This effect can be captured through modeling the price in MMM. The model provides the price elasticity of the brand which tells us the percentage change in the sales for each percentage change in price.

How MMM works

A statistical model is estimated on historical data with sales as a dependent variable and list of explanatory variables as marketing activities, price, and seasonality and macro factors. The simplest and broadly industrially accepted model is linear regression:

$$Sales_t = \alpha + \beta_1 \times \mathrm{var1}_t + \beta_2 \times \mathrm{var2}_t + ... + \varepsilon_t$$

The output of the model is then used to carry out further analysis like media effectiveness, ROI and price elasticity and to simulate what-if scenarios.

This is accomplished by setting up a model with the sales volume/value as the dependent variable and independent variables created out of the various marketing efforts. Once the variables are created, multiple iterations are carried out to create a model which explains the volume/value trends with a certain accuracy. Further validations are carried out, either by using a validation data, or by the consistency of the business results. The output can be used to analyse the impact of the marketing elements on various dimensions. The contribution of each element as a percentage of the total plotted year on year is a good indicator of how the effectiveness of various elements changes over the years. The yearly change in contribution is also measure by a due-to analysis which shows what percentage of the change in total sales is attributable to each of the elements. For activities like television advertising and trade promotions, further analysis like effectiveness can be carried out.

This analysis tells the marketing personnel the incremental gain in sales that can be obtained by increasing the respective marketing element by one unit (Collins e al., 2010). If detailed spend information per activity is available then it is possible to calculate the ROI of the marketing activity. Not only is this useful for reporting the historical effectiveness of the activity, it also helps in optimizing the marketing budget by identifying the most and least efficient marketing activities. Once the final model is ready, the results from it can be used to simulate marketing scenarios for a 'What-if' analysis. The marketing manager can reallocate this marketing budget in different proportions and see the direct impact on sales/value. He/she can optimize the budget by allocating spends to those activities which give the highest return on investment.

Data Collection

Different variables are identified as possible factors that significantly drive sales. The selection of these variables depends on product type, industry and market characteristics. As an example, ice cream sales definitively show a seasonal pattern and may depend on temperature. In general, the following are some types of data to consider when developing dataset for MMM:

Economic data

Employment and unemployment, discretionary income, inflation rates, gross domestic product, interest rates, energy costs, etc. An understanding of the effects of general economic variables is vital to building sound models.

Industry data

What are the trends in the specific industry? Is the market for the product or service growing? What is the rate of growth? Is international trade affecting the industry? Are important geographic differences evident within the industry?

Product category data

What are the trends in the specific product category? For example, is the refrigerated soy milk category growing? At what rate? How does this growth vary by geographic region? What are the trends by brand?

Product lines and SKUs (Stock Keeping Units)

What is the history of each major brand within the category? What new products or new SKUs have been introduced, and when, for each major brand? What is the history of private label brands and SKUs in the category?

Pricing data

A history of average prices for each SKU in the category is essential. Pricing is almost always an important variable.

Distribution levels

What is the history of distribution levels for each product and SKU? What is the quality of that distribution? Average number of shelf facings per SKU?

Retail depletions

It's essential to have a clean measure of sales to end-users, undistorted by fluctuations in inventories. Factory shipments are worthless for modelling purposes, in most instances. Retail take-away (or retail depletions) in dollars and in units (ounces, pounds, cases, etc.) is the most common measure of sales to consumers. The goal is to accurately measure sales to ultimate users (the people the marketing efforts are focused upon).

Advertising measures

Money spent on media advertising is seldom useful by itself. The media advertising must be translated into television GRP (gross rating point) equivalents, or some other common "currency." That is, the print advertising, the radio advertising, the online advertising, etc. must all be converted into common units of measure (typically, television GRP equivalents). The money spent by specific media type (adjusted for comparative effectiveness) is another way of weighting media inputs as variables. All of this is apt to prove worthless, however, if copy-testing scores are not included for each of the ads. A media plan of 100 GRPs per week might have no effect if a weak commercial is run, but might have great effect if a terrific commercial is aired. Likewise, the exact media schedule is important, and the length of time each commercial is on the air must be considered because of wear-out effects.

Consumer promotion

Consumer (or end-user) promotions come in many forms, but the primary characteristic of these promotions (as compared to advertising) is the immediacy of the effects. Promotions are designed to have powerful, short-term effects on sales. Temporary price reductions, cents-off coupons, and buy one/get one free are examples of common consumer promotions. These promotions must be understood, measured, and incorporated into the models. If not fully comprehended, the promotion effects could easily overwhelm the modelling effort.

Trade promotion

These promotions usually take the form of discounts or allowances given to the trade in order to stage in-store promotions of some type (temporary price reductions, end-aisle displays, in-store signage, local advertising, and so on). Trade promotions must be fully understood and included within the models because of the sales fluctuations they cause. When the manufacturer offers one dollar off the price of each case for 30-days (a typical trade promotion), the retailer is very likely to take actions to increase sales of that brand.

Sales force effects

Every company and industry are different, but the nature and strength of a company's sales force (and how it is organized, managed, and compensated) can create variables for the marketing models. Sales organizations tend to be very expensive, so it's generally worthwhile to try to include sales force variables within the models.

Service effects

If services are an important part of the customer's experience in buying and/or using a product, then this variable must be measured and incorporated into the models. For example, if a new product must be installed by a service technician, then the interaction between customer and technician can be a major variable, and must be tracked with some type of customer satisfaction survey.

Then factors that could drive sales could be modelled according to:

$$Sales_t = \alpha + \beta_1 \times \mathrm{var}_t^1 + \beta_2 \times \mathrm{var}_t^2 + ... + \varepsilon_t$$

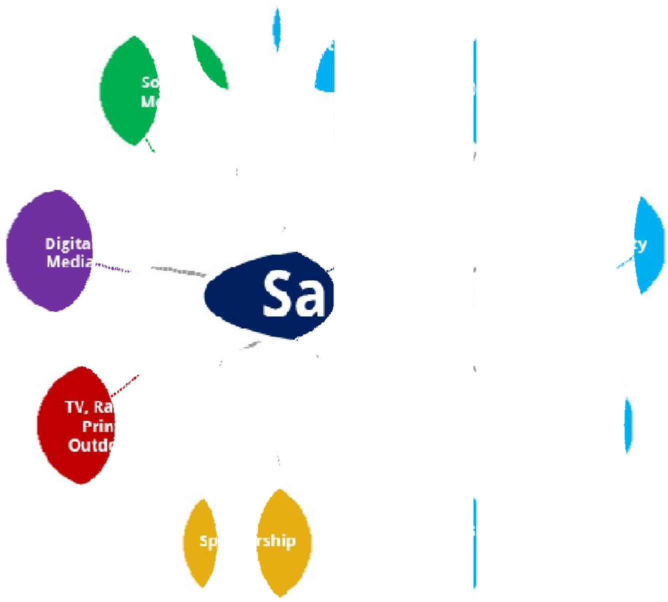

Figure 6. Sales deriving forces

A MMM project starts with a workshop where client's needs are identified and a corresponding solution is proposed by the quant team. Once the project's scope is defined and agreed, the quant team requests the data necessary for the analysis. Usually the client provides restricted data such as sales, price, distribution, margins etc. while media information and external factors are retrieved from research agencies or data providers. Regardless of method, MMM can be successful only if accurate and highly specific data are available upon which the modelling can be based. The greatest barrier to successful modelling is always a lack of relevant, specific, accurate data. So, the first step in any modelling effort is designing the data warehouse that will support the modelling. The next step is collecting and cleaning all of the historical data and entering it into the data warehouse, and then cleaning and entering new data on a continuing basis. Clean, accurate, highly specific data is absolutely essential to successful modelling. The data must be specific to individual brands and product lines, not the company as a whole.

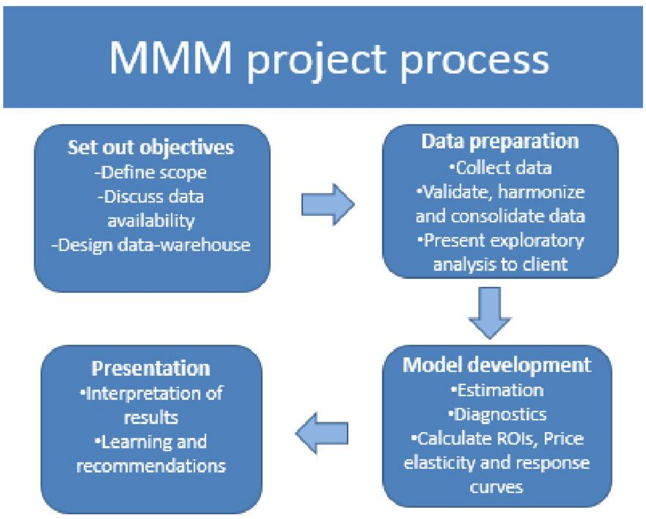

Figure 7. MMM Project process

Once the data warehouse is built, some exploratory analysis is conducted and presented to the client in order to validate the data and avoid possible mistakes. After data validation comes model estimation and calculation of ROIs, price elasticity and response curves. The cycle ends with a final presentation describing and interpreting the output of the model and providing recommendations.

Quantifying the effect of TV and Radio Advertising on Sales

- The performance of TV and radio advertising is expressed in terms of Gross Rating Points (GRPs). A rating point is a percentage of the potential audience and GRPs measure the total of all rating points during an advertising campaign.
– GRPs (%) = Reach * Frequency
– Example: Let's assume a commercial is broadcasted two times on TV
- Gross Rating Points (GRPs) equal Reach 'times' Frequency, expressed as a percentage.
- GRPs measure the total of all Rating Points during an advertising campaign. A Rating Point is one percent of the potential audience. For example, if 20 percent of all targeted televisions are tuned a show that contains a commercial, 20 Rating Points will be gained.
- If, the next time the show is on the air, 32 percent are tuned in, a total of 20 + 32 = 52 Rating Points, and so on through the campaign. The word "gross" reflects that the calculation double-counts (actually multiple-counts) the audience; that is to say, it is possible to reach a percentage higher than 100.
- Gross Rating Points could also be applied to other media besides television: radio, print, billboards, the web, and so on. If we attach a banner to our corporate headquarters building, and 5 percent of our target population drives by the billboard twice every day for 80 days, then GRPs = 5 x 2 x 80= 800.
- Media planners calculate total Reach, average Frequency, and GRPs as part of the planning of a campaign. The goal is to obtain the highest possible GRPs at the lowest possible cost, while remaining focused on the target market. After the campaign, we can calculate actual Reach x Frequency = GRPs to produce a permanent record.
- The exposure to TV advertising builds awareness, resulting in sales.

Equity or Base

The very break-up of sales volume into base (volume that would be generated in absence of any marketing activity) and incremental (volume generated by marketing activities in the short run) across time gain gives wonderful insights (Wolfgang Jank, 2011). The base grows or declines across longer periods of time while the activities generating the incremental volume in the short run also impact the base volume in the long run. The variation in the base volume is a good indicator of the strength of the brand and the loyalty it commands from its users Brand equity refers to the marketing effects or outcomes that accrue to a product with its brand name compared with those that would accrue if the same product did not have the brand name. And, at the root of these marketing effects is consumers' knowledge. In other words, consumers' knowledge about a brand makes manufacturers/

advertisers respond differently or adopt appropriately adept measures for the marketing of the brand. The study of brand equity is increasingly popular as some marketing researchers have concluded that brands are one of the most valuable assets that a company has. Brand equity is one of the factors which can increase the financial value of a brand to the brand owner, although not the only one.

Response curves

The model that is linear in parameter but can account for non-linearity through variable transformations. Advertising for instance can have diminishing returns to scale, i.e., the relationship between advertising and demand can be nonlinear. Typically, each incremental amount of advertising causes a progressively lesser effect on demand increase. This is a result of advertising saturation. Advertising variable can be transformed to an appropriate nonlinear form like the logistic or negative exponential distribution, depending upon the type of diminishing returns or 'saturation' effect the response function is believed to follow. It is this transformed variable that is used in the sales response models. For example if advertising awareness followed a logarithmic distribution, then in a linear sales response model we would have:

$St = Log\ (GRPt) + \alpha\ (t)$

Where St is sales at time t, GRPt is the level of advertising GRPs at time t and α is the random error component. Advertising typically has a lower elasticity than other elements of the marketing-mix. This is considered acceptable by business personnel since advertising is also believed to have a long-term positive effect on Brand Equity, which is usually not captured by most econometric models.

Price Elasticity

- Assumption: constant elasticity across the sample which implies a linear relation between volume and price.
- By using the coefficient of the regression, it is possible to derive an estimate for price elasticity:

$$Elasticity = \frac{Avg\,\mathrm{Pr}\,ice}{AvgSales} * coeff$$

Example scenario3: MMM: - Sales data from 200 point of sales of a typical retailor were gathered to investigate the impact of advertising on sales.

Table 5. MMM parameter estimates

MMM: Log (Sales) = log (TV) + log (Newspaper) + log (Radio)

Coefficients (Elasticity's):

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	0.403322	0.045819	8.802	7.01e-16 ***
Log (TV)	0.349747	0.007662	45.647	< 2e-16 ***
Log (Newspaper)	0.015810	0.008047	1.965	0.0509 .
Log (Radio)	0.171342	0.007551	22.692	< 2e-16 ***

Residual standard error: 0.1085 on 196 degrees of freedom
Multiple R-squared: 0.9325, Adjusted R-squared: 0.9314
F-statistic: 902 on 3 and 196 DF, p-value: < 2.2e-16

Advertising marketing campaign was carried out on television, on radio and on newspapers. EDA and appropriate data analysis were performed on the sales data. The following were the output of the regression model analysis which gives insight to the marketing manager in making subsequent business decisions.

The final model for Sales as a result of the advertising budg*et al*located to the three marketing channels will be (note all the three are (newspaper near) significant to be included in the model):

Log (Sales (j)) = 0.403322 + 0.349747*log (TV (j)) + 0.015810*log (Newspaper (j)) + 0.171342*log (Radio (j)),

where j=1…number of sale points.

Accordingly therefore, the optimum budget for each advertising campaign at each sale point is allocated as follows: Let's rename the elasticity's of TV, Newspaper and Radio as $\beta_{i,}$ for i = 1, 2, 3, respectively.

Log(TV(j)) = log(TV(j)) * β_1; log(Newspaper (j)) = log(Newspaper(j)) * β_2; log(Radio(j)) = log(Radio (j)) * β_3.

The baseline at each sale point, log (**Base(j)) = log(Sales(j))/ log(TV(j))* log(Newspaper (j)) * log(Radio (j))**

Conclusion and Recommendations

The analysis and comprehension of business data are fundamental parts of all business organizations. Monitoring national economies and retail sales tendencies depend on data analysis. Also, customer behaviour has become more complex due to the diversity of applications that compete in the marketplace. On the other hand, computer systems have swamped us with large volumes of data and information, much of which is irrelevant for a specific analysis objective. Analyzing business data has become easier due to data management infrastructures that separate the operational data from the analytical data, and from Internet applications and cloud computing, which facilitate the gathering of large-volume historical data logs. Performing data analysis on business data involves extracting insight and intuition so as to make data-driven decision making through the use of statistical thinking and data mining algorithms and models to business problems; and thus the objective of data analysis should be that of discovering useful and meaningful knowledge and separating the relevant from the irrelevant.

As we have seen in this article using three different case scenarios, collinearity complicates our ability to interpret the fitted model since the coefficients change as variables are added or removed from the fit. There are a collection to methods for attempting to validate a regression model. We have to deploy reasonable model testing and validation techniques used to test the fitted model, producing an independent measure of its predictive ability, for example. Fitting a collection of so many predictors requires an automated routine and we might have to use an automated, "greedy" tool (stepwise regression) that adds variables to a regression, one at a time, to make the R^2 statistic grow as rapidly as possible. It can also be configured to remove variables from a model, starting from something complex and reducing it to a simpler, more parsimonious model.

REFERENCES

Ataman, *et al*. 2010. The Long-Term Effect of Marketing Strategy on Brand Sales. *Journal of Marketing Research*, 47 (5), 866-882.

Collins, J., *et al*. 2010. How precise magazine inputs can improve media mix modelling: The impact of more behavioural metrics on ROI. *Journal of Advertising Research*, 50(1), 10-15.

Hallward, J. 2008. Make measurable what is not so: Consumer mix modelling for the evolving media world. *Journal of Advertising Research*, 48(3), 339-351.

Jean Paul Isson, Jesse Harriott 2013. Win with advanced business analytics: creating business value from your data, John Wiley & Sons, Inc., Hoboken, New Jersey.

Johannes Ledolter 2013. Data mining and business analytics with R, John Wiley & Sons, Inc,

Julian J. Faraway 2000. Practical Regression and Anova using R, Julian Faraway.

Kormanik, B. 2010. The ROI of marketing: Here's how to figure out if your spending is really paying off. *Hotel Interactive*. 19/11/2010, accessed 20/01/2015: http://www.hotelinteractive.com/article.aspx?articleid=18833

Larry Wasserman 2006. All of Nonparametric Statistics, Springer.

Lawrence S. Maisel and Gary Cokins 2014. Predictive Business Analytics: *Forward-Looking Capabilities to Improve Business Performance,* John Wiley & Sons, Inc., Hoboken, New Jersey.

Max Kuhn and Kjell Johnson 2013. Applied Predictive Modelling, Springer.

Michael H Kutner *et al* 2005. Applied linear statistical models.-5th ed., McGraw-Hill/Irwin.

Bramer, M. 2013. *Principles of Data Mining*, Undergraduate Topics in Computer Science, DOI 10.1007/978-1-4471-4884-5 3, Springer-Verlag London.

Pratt, S., S. McCabe, I. Jimenez and A. Blake 2010. Measuring the effectiveness of destination marketing campaigns: Comparative analysis of conversion studies. *Journal of Travel Research*, 49(2), 179-190.

Rubinson, J. 2009. Empirical evidence of TV advertising effectiveness. *Journal of Advertising Research*, 49(2), 220-226.

Rahul Saxena and Anand Srinivasan 2013. Business Analytics: A Practical Guide, Springer, New York.

Stephen G. Cecchetti 2015. Money, Banking, and Financial Markets, 4th Edition McGraw-Hill/Irwin.

Wolfgang Jank 2011. Business Analytics for Managers, Springer, London.

****