Performance **pointers**

By Rajat Kohli

RETAIL AND CONSUMER GOODS COMPANIES RUN MULTIPLE CAMPAIGNS, PROMOTIONS, AND INCENTIVES TO LURE CUSTOMERS TO BUY MORE. A GREAT DEAL OF TIME, ENERGY, AND RESOURCES ARE DEPLOYED TO EXECUTE THE PROMOTION PROGRAMS. IN MOST ORGANISATIONS THE ALLOCATION OF FUNDS TO VARIOUS PROGRAMS IS BASED ON GUT FEEL OR PAST EXPERIENCE. IF THE PROMOTIONS DO NOT PAN OUT THE WAY THEY WERE INTENDED OR PERFORM BETTER THAN EXPECTED, THE DECISION MAKER IS UNABLE TO EXPLAIN THE PHENOMENON OR REPEAT THE PERFORMANCE. ALSO THE PROMOTIONS MAY BE TARGETED AT A MACRO SEGMENT WHILE THEY MAY BE EFFECTIVE ONLY AT A MICRO SEGMENT THEREBY REDUCING THE OVERALL EFFECTIVENESS OF THE PROGRAM.

→ Using Analytics the gut-based decision can be supported by facts. It helps the business make a better business decision and stay ahead of its competitors who may rely purely on gut feel or past experience.

We have chosen an MLM (Multi Level Marketing) company as an example as the problem of allocating funds to various incentives is further amplified due to the large sales force engaged in the MLM companies. Retail and consumer goods industry can draw parallel between incentives and promotions that are run through the year targeted at various segments to improve sales.

BACKGROUND

According to Philip Kotler, one of the distribution channels through which marketers deliver products and services to their target market is Direct Selling where companies use their own sales force to identify prospects and develop them into customers, and grow the business. Most Direct Selling companies employ a multi-level compensation plan where the sales personnel, who are not company employees, are paid not only for sales they personally generate, but also for the sales of other promoters they introduce to the company, creating a down line of distributors and a hierarchy of multiple levels of compensation in the form of a pyramid. This is what we commonly refer to as Multi-Level Marketing (MLM) or network marketing. Myriad companies like Amway, Oriflame, Herbalife, etc. have successfully centered their selling operations on it. As part of Sales Promotion activities MLM companies run Incentive programs for their Sales Representatives who are rewarded for their superior sales performance and introducing other people to the company as Sales Representatives.

BUSINESS CHALLENGE

Although Incentives play a major role in sales lift, there are many other environmental factors such as advertisement spend, economic cycle, seasonality, company policies, and competitor policies also affect sales. It becomes increasingly difficult to isolate the impact of incentives on the sales.

Usually MLM companies run multiple and overlapping incentive programs i.e. at any given time more than one incentive programs run simultaneously. See figure below. Rewards could be monetary or include non-monetary items like jewelry, electronic items, travel, cars etc. and are offered on a market-by-market basis. A key question that arises is – "how do we understand the effectiveness of these multiple incentive programs?"

The success of any MLM company is largely dependent on the performance of its Sales Representatives, Incentive programs are of paramount importance for realising the company's marketing objectives and hence form a vital component of its marketing mix. Some of the large Direct Selling Corporations end up spending millions of dollars on Incentive programs in every market. It is important for the incentive managers to understand the effectiveness of the incentive programs, often measured in terms of Lift in Sales, in order to drive higher Return on Incentive Investment as against making gut-based investment decisions.
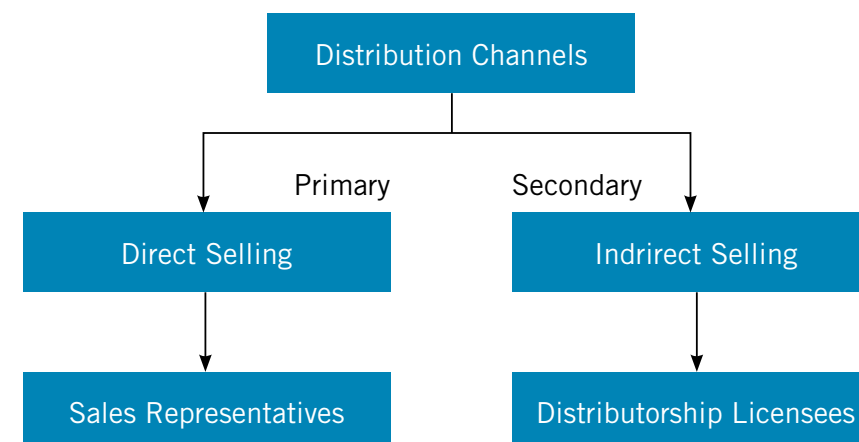
To summarise, it is not easy to measure the ROI on Incentive programs for two broad reasons. First, it is difficult to separate the Lift in Sales due to Incentive programs from that due to other concurrent communication and marketing mix actions. Secondly, in most cases there may be an absence of "Silence Period" i.e. one or more incentive programs are functional at all points of time. This makes the task of baseline sales estimation virtually impossible by conventional methods. Usage of Analytics – the science of making data-driven decisions – becomes indispensible in order to address the above constraints while at the same time statistically quantify the individual Incentive ROI and make sales forecasts with sound accuracy. While doing so a systematic approach using best practices is followed in order to obtain reliable results in a consistent and predictable manner.

APPROACH

It is very tempting to jump straight into the data exploration exercise. However, a structured approach ensures the outcome will be aligned with the business objectives and the process is repeatable.

OBJECTIVE SETTING

The first step towards building an analytics based solution is to list down the desired outcomes of the endeavor prior to analysing data. This means to thoroughly understand, from a business perspective, what the company really wants to accomplish. For instance, it may be important for one MLM company to evaluate the relative effectiveness of various

Distribution Channels

Primary — Direct Selling → Sales Representatives

Secondary — Indirect Selling → Distributorship Licensees

components of its marketing mix which includes Incentives along with pricing, distribution, and advertising etc. while some other company may be interested in tracking the ROI of its past incentive programs in order to plan future incentive programs.

In addition to the primary business objectives, there are typically other related business questions that the Incentive Manager would like to address. For example, while the primary goal of the MLM Company could be ROI estimation of Incentive programs, the incentive manager may also want to know which segment of Representatives are more responsive to a particular type of Incentive or to find out if the incentives are more effective in driving the sales of a particular product category.

Moreover, it may also be prudent to design the process in a way that it could be repeatedly deployed across multiple countries rapidly and cost effectively. This is feasible where a direct selling company has uniform data encapsulation practices across all markets. A good practice while setting the objectives is to identify

the potential challenges at the outset. The biggest challenge is the volume of Operational Data. In the case of retail and direct sellers it runs into billions of observations over a period of few years.

DATA STUDY

Most Direct Sellers maintain sales data at granular levels and aggregated over a period of time and across categories along with Incentive attributes, measures and performance indicators. Hence, we can safely assume that most companies will have industry specific attributes like a multi-level compensation system. However, every MLM company will also have its own specific set of attributes which differentiate it from its competitors. It is therefore vital to develop a sound understanding of historical data in the given business context before using it for model building.

This exercise also entails accurately understanding the semantics of various data fields. For instance, every MLM company will associate a leadership title with

its Representatives. However, the meaning of a title and the business logic used to arrive at the leadership status of a Representative will vary from company to company. Good understanding of other Representative Population attributes like Age, duration of association with company, activity levels, down line counts etc. also leads to robust population segmentation.

Depending on business objectives other data related to media spend, competitor activity, macro economic variables etc. should also be used. A potential issue of data fragmentation might arise here as voluminous data is broken up into smaller parts for ease of storage, which needs to be recombined logically and accurately using special techniques while reading raw data.

Moreover, raw data coming from a datawarehouse usually contains errors and missing values. Hence, it becomes important to identify them in the data using a comprehensive data review exercise so that they may be suitably corrected once the findings have been validated by the

client. In extreme cases the errors and inconsistencies might warrant a fresh extraction of data. This may lead to an iterative data review exercise which is also used to validate the entire data understanding process. Any lapse in data understanding before preparing data for modeling might lead to bias and errors in estimation in future. A data report card helps clients understand the gaps in their data and establish procedures to fill the gaps. A sample scorecard is shown for reference.

DATA PREPARATION

Modeling phase requires clean data with specific information in a suitable format. Data received from the company cannot be used as input for modeling as is. Data preparation

is needed to transform input data from various sources into the desired shape. Not all information available from raw data may be needed. For example, variables like name of Incentive program, source of placing orders, educational qualifications of Representatives etc. may not be needed for building a model. Key variables that must go in model building are identified and redundancies removed from the data. Observations with incorrect data are deleted and missing values may be ignored or suitably estimated. Using the derived Representative attributes the population is segmented into logically separate strata. Data from different sources is combined to form

a single table and new variables are derived to add more relevant variables. In the final step of the data preparation exercise measures are aggregated across periods, segments, geographies and product categories. The aggregated data is used as an input for modeling.
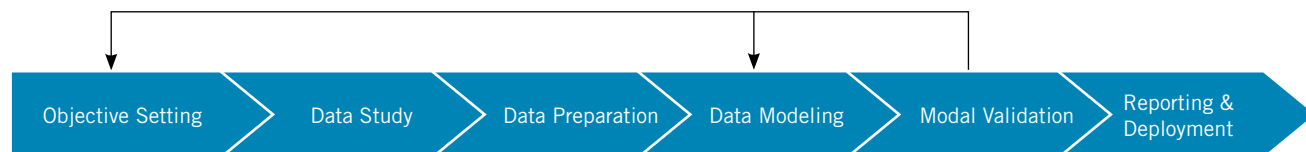
DATA MODELING

A statistical model is a set of equations and assumptions that form a conceptual representation of a real-world situation. In the case of an MLM company a model could be a relationship between the Sales and other variables like Incentive costs, number of Representatives, media spend, Incentive attributes
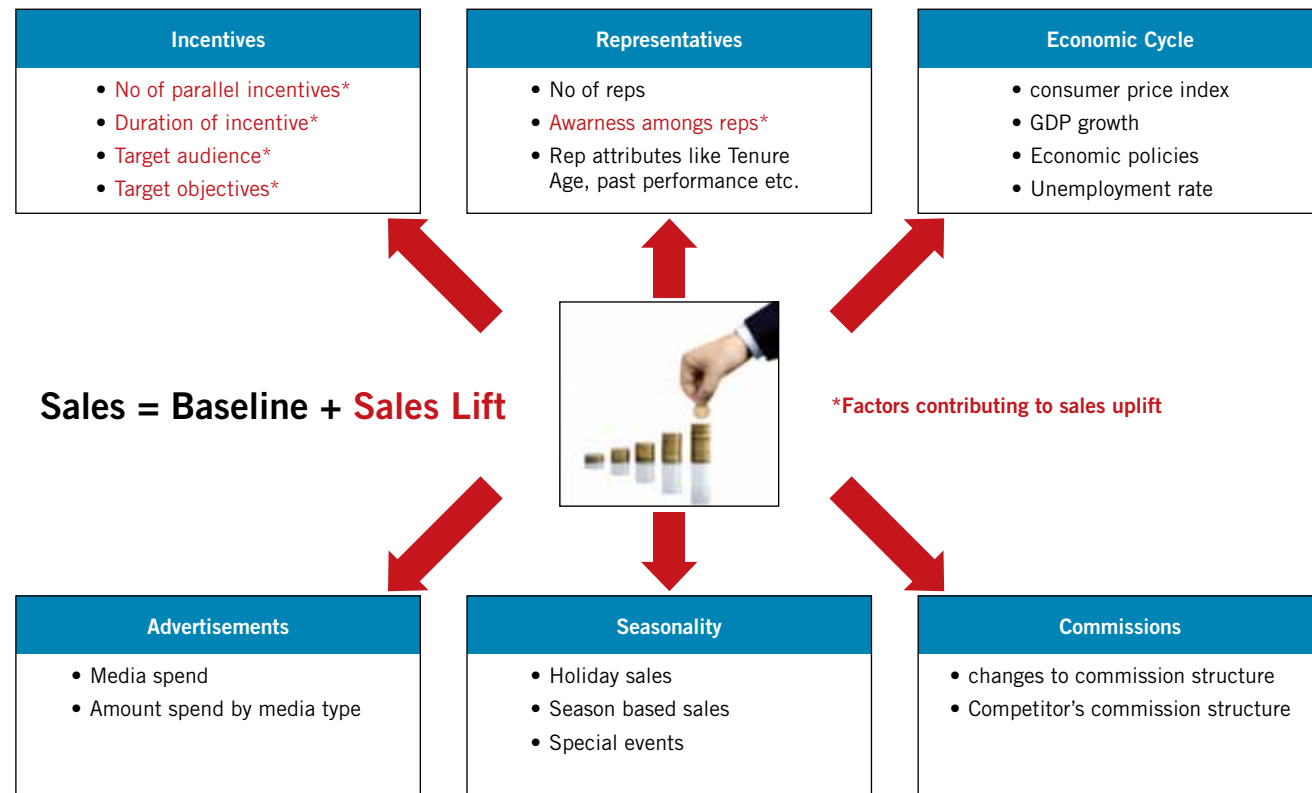
KPIs	Score	Comments
Skewed	★★	Negative value and others in measures
Variety of Data	★★★	Will require additional data like promotion spend, competitor data etc.
Depth of Data	★★★★	High Level of granularity
Breadth of Data	★★★	Lomited no of measures available
Transformations required	★★	Linearization required for promotional lag effect
Duplicates	★★★	Duplicity observed in few data objects
Reference Integrity	★★	Contradiction between KPI metrics & Operational data
Completeness	★★	Many of the data fields have blank values
Ambiguity	★★	Many data fields in Sales data are redundant

Incentives
• No of parallel incentives*
• Duration of incentive*
• Target audience*
• Target objectives*

Representatives
• No of reps
• Awarness amongs reps*
• Rep attributes like Tenure Age, past performance etc.

Economic Cycle
• consumer price index
• GDP growth
• Economic policies
• Unemployment rate

Sales = Baseline + Sales Lift

*Factors contributing to sales uplift

Advertisements
• Media spend
• Amount spend by media type

Seasonality
• Holiday sales
• Season based sales
• Special events

Commissions
• changes to commission structure
• Competitor's commission structure

January	February	March	April	May	June	July	August	September	October	November	December
Incentive A								Incentive A			
	Incentive H				Incentive H						
										Incentive G	
	Incentive B										
		Incentive C									
		Incentive D									
Incentive E						Incentive B				Incentive E	
							Incentive F				Incentive F

| Objective Setting | Data Study | Data Preparation | Data Modeling | Modal Validation | Reporting & Deployment |

and Representative attributes. Before commencing with the modeling exercise the level at which the model should be built needs to be ascertained. A Top Down approach builds the model at the highest level and the results are proportionally disseminated down at the population segment level and then on to individual level. The resultant model may not be able to properly account for the variation in the dependent variable and introduce bias in estimates as the existence of separate strata in the population while model building is ignored. Choosing a Bottom Up approach on the other hand builds the model at the individual level and the results are aggregated up to the segment level and then on to the top level. This is exhaustive, but at the same time, very tedious, as Sales data usually

runs into millions of observations and not all Representatives may be actively contributing to the Sales at all times. Moreover, if the project objective revolves around estimating national figures this exercise may become redundant. The Middle Out approach may be the best approach to model the data. The model is built at the Segment level and depending on the requirement the results may be aggregated up to the top level or proportionately disseminated down at the individual level.

The first step of model building exercise is to specify the model equation. This requires the determination of the Dependent variable, Independent variables and the Control Variables. Control variables are those variables that determine the relationship between the dependent variable and

independent variables. In a baseline estimation scenario, the Sales measure is the dependent variable; Incentive cost and other Incentive attributes form the independent variables; segmentation variables, time series, geography, inflation and other variables like media spend act as Control variables. Usually the model is non-linear i.e. the dependent variable is not directly proportional to one or more independent variables. A non-linear model may be transformed to a linear model by use of appropriate data transformations. For example, the relationship between Sales and Incentives is non-linear. Representative Incentives behave like consumer coupons where there is an initial spike in Sales at the start of the Incentive followed by a rapid decline, but the impact returns at the end of the incentive as

Representatives try to beat the deadline. Application of coupon transformation to Incentive variables therefore produces a linear relationship between Sales and Incentives.

Model coefficients are then estimated using advanced statistical techniques like Factor Analysis, Regression and Unobserved Component Modeling. The common practice followed across industry is to use Regression Analysis for explaining the relationship between the dependent variable and the independent variables and separately employ time series ARIMA (Auto Regressive Integrated Moving Average) models for forecasting as the data invariably has a time component. To solve the Regression models with all incentive attributes accounted for, they are first condensed into a few underlying factors accounting for most of the variance. These factors are then part of the regression along with the control variables. Final coefficients are a combination of factor loadings and model coefficients. This Regression model equation allows us to understand how the expected value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

The time series model is developed by reducing the non-stationary data to stationary data, removing the Seasonality and Cyclic components from it and estimating the coefficients of the ARIMA model. This approach often leads to in

concordant answers from Regression and ARIMA models as Regression analysis will miss the trend and ARIMA forecasting may fail to account for causal effects.

Unobserved Component Modeling may be employed if a very high accuracy is desired from the model. It leverages the concepts of typical time series analysis where observations close together tend to behave similarly while patterns of correlation (and regression errors) breaks down as observations get farther apart in time. Hence, regression coefficients are allowed to vary over time. Usual observed components representing the regression variables are estimated alongside the unobserved components such as trend, seasonality, and cycles. These components capture the salient features of the data series that are useful in both explaining and predicting series behavior. Once the model coefficients are determined it is essential to validate the model before using it for forecasting.

MODEL VALIDATION

The validity of the model is contingent on certain assumptions that must be met. Firstl, the prediction errors should be Normally Distributed about the predicted values with a mean of zero. If the errors have unequal variances, a condition called heteroscedasticity, Weighted Least Squares method should be used in place of Ordinary Least Squares Regression. A plot of residuals against the predicted values of the dependent variable, any independent variable or time

can detect the violation of the above assumption.

Another assumption that is made in time series data is that prediction errors should not be correlated through time i.e. errors should not be auto-correlated. This may be checked using the Durbin-
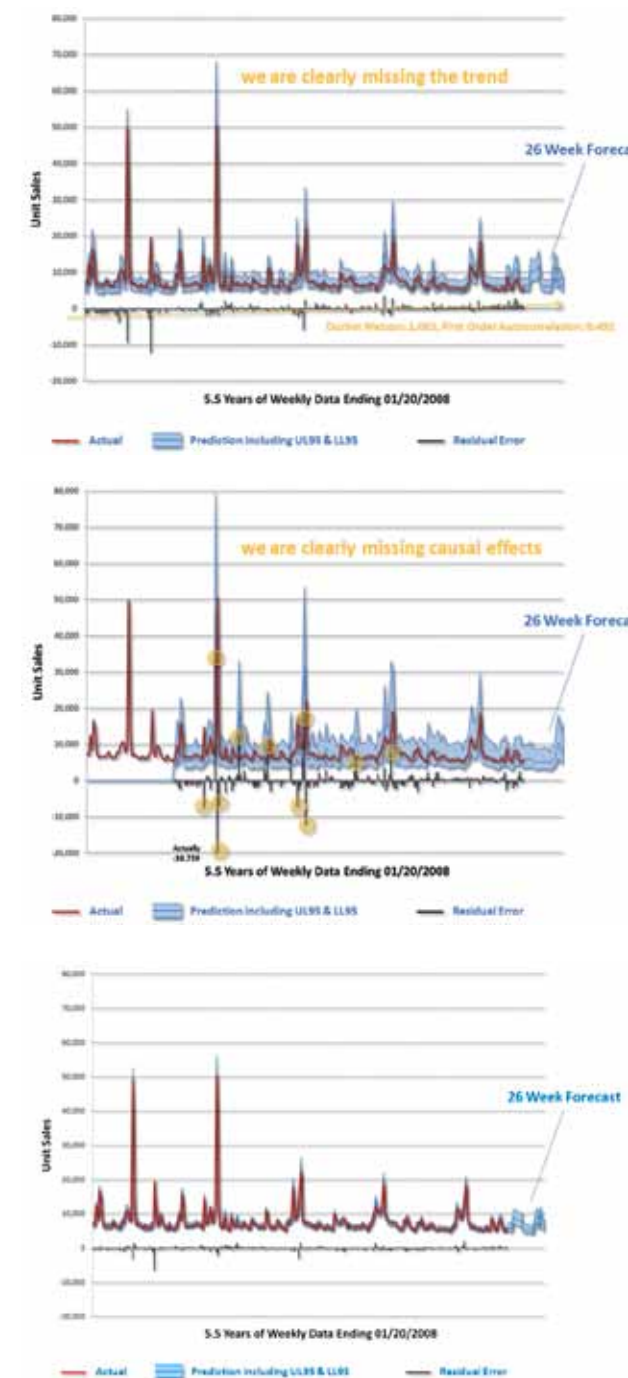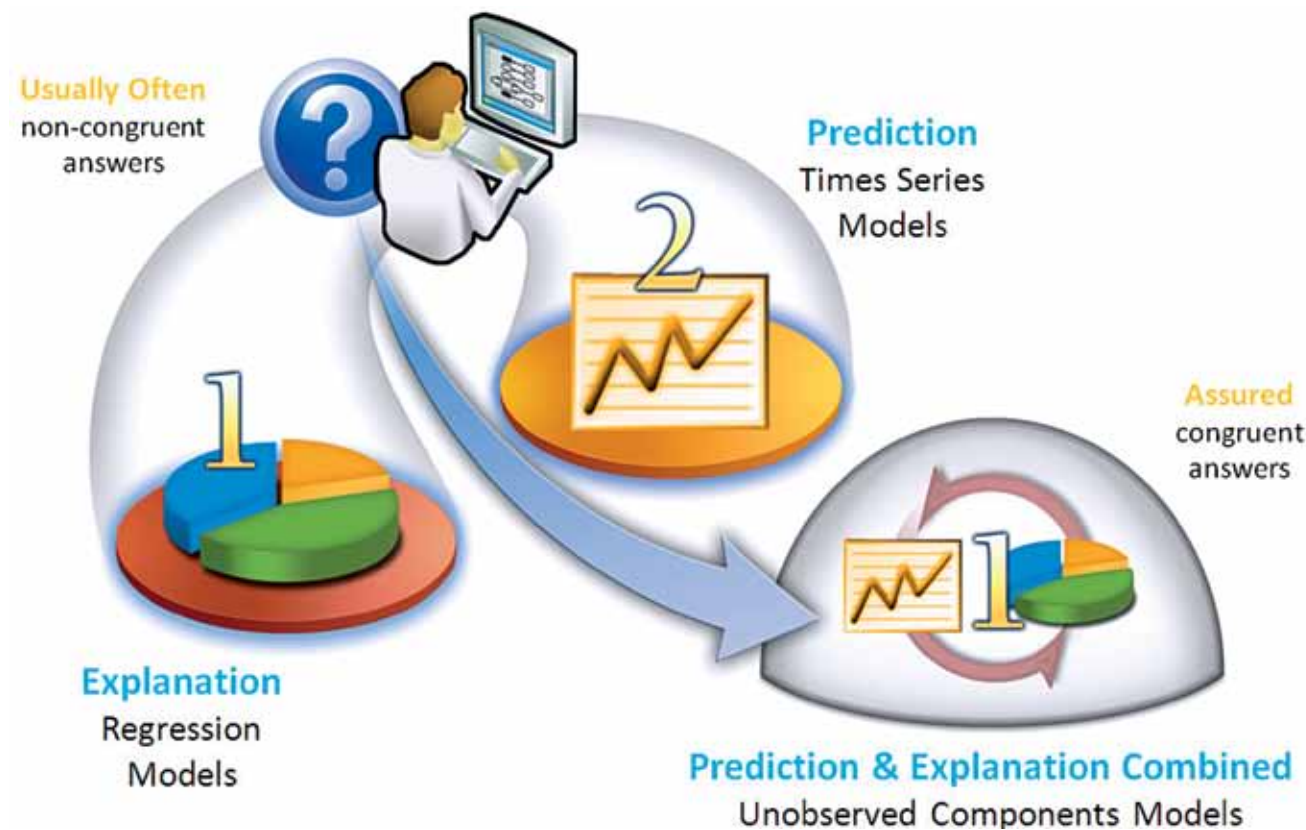
Watson test. If errors are found to be auto-correlated then Generalized Least Squares Regression should be used. It is also important to check for correlation among the independent variables, a condition called multi-collinearity. It can induce errors in coefficient estimates and inflate their observed variances indicated by a variable's Variation Inflation Factor (VIF). Multi-collinearity can be easily detected in a multiple regression model using a correlation analysis matrix for all the independent variables. High values of correlation coefficients indicate multi-collinearity.
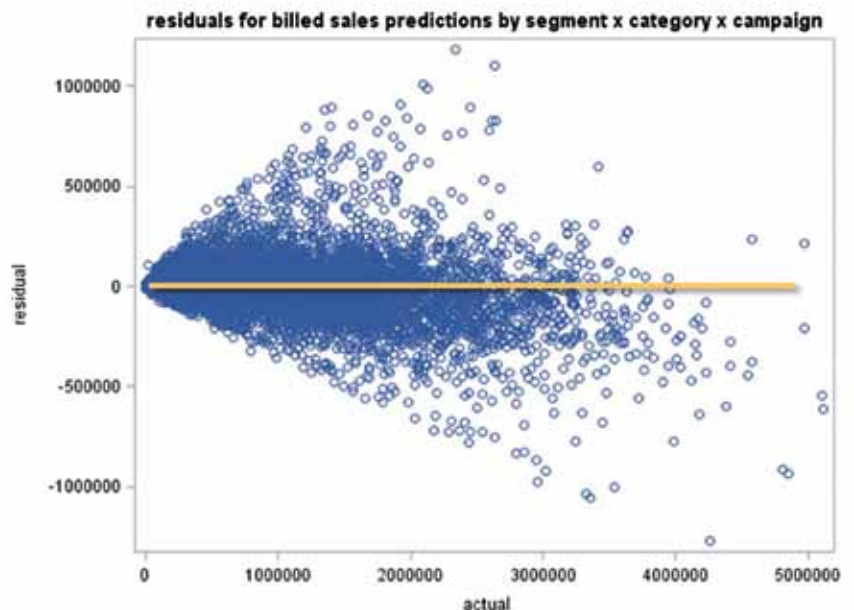
The simplest way to solve this problem is to remove collinear variables from the model equation. However, it may not always be feasible to remove variables from the equation. For example, the cost of an Incentive program is an important variable that cannot be removed if found to have a high Variation Inflation Factor. In such cases Ridge regression may be used in place of OLS Regression. However, some bias may sneak in the coefficient estimates. The goodness of model fit may be adjudged by the values of R^2, which is the model coefficient of determination. Its value ranges between 0 and 1. A good fit will have an R^2 a value of greater than 0.9. But any value of R^2 close to 1 must be bewared as it could be causing over-fitting. Such a model would give inaccurate forecasts. A low value of Mean Average Percentage Error (MAPE) of predicted value is also indicative of a good fit. Once the model assumptions are validated and goodness of fit established the model equation can be used for reporting and deployment purposes.
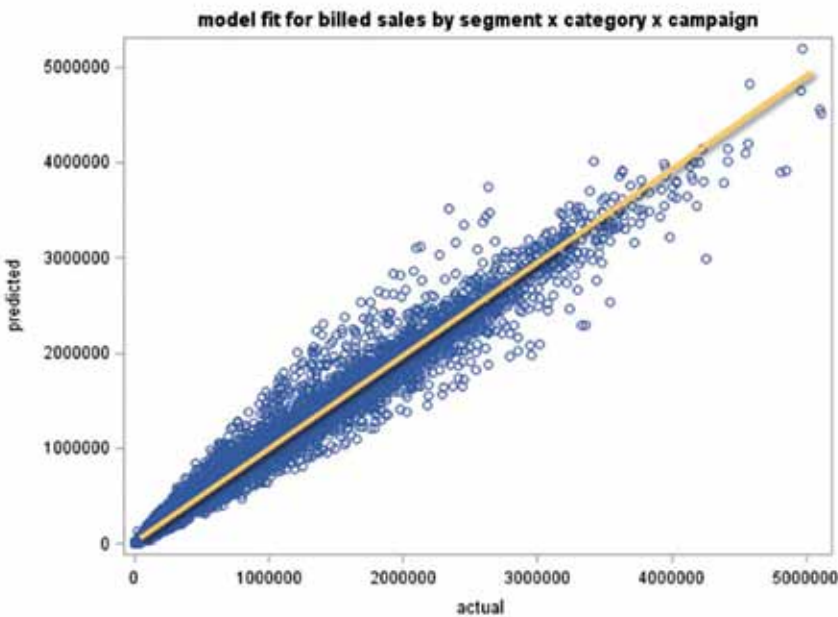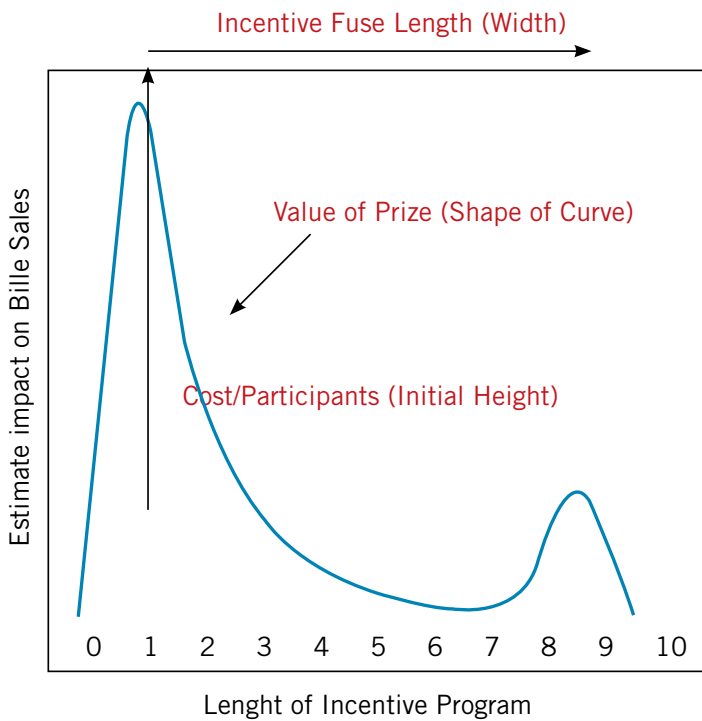
REPORTING & DEPLOYMENT

Depending on the chosen dependent variable based on the scope of the incentive modeling exercise the baseline measures like Sales, Volume, Representative count etc can be estimated using the model equation. These estimates along with other variables and derived values can be used to obtain insights about Incentive performance through dashboards with

KPIs and other pre defined reports like annual Lift in Sales vs. Incentive Cost, Baseline Sales vs. Sales Representative Count, etc. The key to realizing the business objectives & deriving value from the modeling outcomes is to capture & present the findings in the most suitable form which will enable the end user to understand the business implications as well as to flexibly slice & dice data in any way in a convenient fashion without having to make any costly investments in acquiring and maintaining system resources. For example, an incentive manager could look at the average ROI of a particular type of Incentive

program as a pre-built report and be given the flexibility of being able to compare the cost of that Incentive with that of another type of Incentive over an online hosted analytics platform which presents pre-canned reports along with user customizable reports and multi-dimensional Data Analysis capability. Such a system can give the end user the freedom to access the reports and analyse the data anytime anywhere using an internet browser. Once deployed, it may be refreshed with additional data in future and may also be used for multiple markets with minor region specific customisations.

residuals for billed sales predictions by segment x category x campaign

Incentive Fuse Length (Width)

Value of Prize (Shape of Curve)

Cost/Participants (Initial Height)

Estimate impact on Bille Sales

0 1 2 3 4 5 6 7 8 9 10

Lenght of Incentive Program

model fit for billed sales by segment x category x campaign

predicted

5000000

4000000

3000000

2000000

1000000

0

0   1000000   2000000   3000000   4000000   5000000

actual

CONCLUSION

The Insight-based approach will significantly increase the confidence level of incentive managers while planning the incentive programs for MLM activities. They will be able to identify the Incentive programs which deliver high, medium and low paybacks, and hence optimise investment in them. It will also help the Direct Seller to check if any product categories are more responsive to Incentives than others. The endeavor can make significant impact where counter intuitive facts surface. For example, any particular event or holiday, which might be an influencing factor in designing incentive programs during a particular time of the year, may actually turn out to be an insignificant contributor to company sales. Incentive Managers can simulate various scenarios by assigning different values to the contributors and macro economic variables and forecast the ROI of near future incentive programs. This will enable regional incentive managers to drive efficiency and effectiveness in incentive planning and realise the company objective of enhanced Sales and ROI. The share of Incentive programs in marketing budget of most Direct Sellers has been progressively increasing and the expenditure incurred is steadily going up in face of competition in emerging markets like India and China which are fast becoming the engines of growth for global Direct Sellers. Investment in analytics based decision support systems will prove to be the difference maker for Direct Sellers. ✖

	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8
Var1	1	0.09	0.31	0.01	-0.09	-0.38	-0.69	0.82
Var2	0.09	1	-0.15	-0.43	-0.56	0.87	-0.08	0.78
Var3	0.31	-0.15	1	0.42	-0.3	-0.26	0.31	0.2
Var4	0.01	-0.43	0.42	1	0.76	0.23	-0.58	0.12
Var5	-0.09	-0.56	-0.3	0.76	1	0.72	-0.15	-0.84
Var6	-0.38	0.87	-0.26	0.23	0.72	1	0.68	-0.36
Var7	-0.69	-0.08	0.31	-0.58	-0.15	0.68	1	-0.58
Var8	0.82	0.78	0.2	0.12	-0.84	-0.36	-0.58	1

ABOUT THE AUTHOR
Rajat Kohli is a Business Analyst with Enterprise Genie Business Solutions, a Bangalore-based Analytics company that partners with businesses for understanding their short-term and long-term goals and equips them to achieve these goals by leveraging data as a weapon of their competitive advantage.