BUSINESS ANALYTICS AND STRATEGY

Project 1

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

SUMMARY

1) Diagnosis: We have erratic sub-optimal returns on our investments made on in-house projects. We will look at and analyse a particular instance and hope to replicate the methodology across the organization.
2) Forecast: While maximizing returns, we would like the risk of our investments to stay below a particular value (usually an industry benchmark).
3) Add value through search: We need to figure out a method to allocate our investments better - we could do this in any number of ways. Firstly, we will use a logit model to explicate the relationship between our decision variables and customer response. Secondly, we will pick the optimal way of allocating our resources across these decision variables subject to nonlinear constraints.
4) Choose: Use the estimated regression coefficients for the decision variables to formulate a nonlinear optimization problem. In particular, maximise the return on investments subject to a limiting value for the risk. Operationalise this by using the canonical Karush-Kuhn-Tucker conditions alongside the Lagrange Multiplier method.
5) Commit: Allocate resources according to the decision rule set by our strategy.
6) Balanced Scorecard Strategy Map: We are strengthening the firm's capabilities and internal processes. This should help us acquire more customers for the same level of investment effort as a result of which our per customer acquisition cost will go down. Our modeling effort will also result in less erratic rates of acquisition which will in turn ensure a more stable internal supply chain and an improved customer experience.

MODEL

We will derive the logit model using the principle of utility maximization. We will later modify it and use it as a classifier. So, let the decision maker, n, face J alternatives to choose from. The utility that the decision maker obtains from alternative j is decomposed into (1) a part labeled V_{nj} that is known up to some parameters, and (2) an unknown part ξ_{nj} that is treated as random. Thus,

$$U_{nj} = V_{nj} + \xi_{nj}$$

The logit model is obtained by assuming that each ξ_{nj} is independently, identically distributed extreme value. The density function for the unobserved component of utility is thus,

$$f(\xi_{nj}) = e^{-\xi_{nj}} e^{-e^{-\xi_{nj}}}$$

And the cumulative distribution function is,

$$F(\xi_{nj}) = e^{-e^{-\xi_{nj}}}$$

The probability that decision maker, n, chooses alternative i over all other j alternatives is

$$P_{ni} = Prob(V_{ni} + \xi_{ni} > V_{nj} + \xi_{nj})$$

$$P_{ni} = Prob(\xi_{nj} < \xi_{ni} + V_{ni} - V_{nj})$$

$$P_{ni} | \xi_{ni} = \prod_{j\ not\ equal\ to\ i} e^{-e^{-(\xi_{ni}+V_{ni}-V_{nj})}}$$

But ξ_{ni} is not given, so

$$P_{ni} = \int_{-\infty}^{\infty} \left(\prod_{j\ not\ equal\ to\ i} e^{-e^{-(\xi_{ni}+V_{ni}-V_{nj})}} \right) e^{-\xi_{ni}} e^{-e^{-\xi_{ni}}}\, d\xi_{ni}$$

Let,

$$\xi_{ni} = s$$

$$P_{ni} = \int_{-\infty}^{\infty} (\prod_{j\ not\ equal\ to\ i} e^{-e^{-(s+V_{ni}-V_{nj})}})\, e^{-s} e^{-e^{-s}}\, ds$$

$$P_{ni} = \int_{-\infty}^{\infty} (\prod_{j} e^{-e^{-(s+V_{ni}-V_{nj})}})\, e^{-s}\, ds$$

$$P_{ni} = \int_{-\infty}^{\infty} (e^{-\sum_j (e^{-(s+V_{ni}-V_{nj})})})\, e^{-s}\, ds$$

$$P_{ni} = \int_{-\infty}^{\infty} (e^{-e^{-s} \sum_j (e^{-(V_{ni}-V_{nj})})})\, e^{-s}\, ds$$

Let,

$$t = e^{-s}$$

Thus,

$$P_{ni} = \int_{\infty}^{0} -(e^{-t \sum_j (e^{-(V_{ni}-V_{nj})})})\, dt$$

$$P_{ni} = \int_{0}^{\infty} (e^{-t \sum_j (e^{-(V_{ni}-V_{nj})})})\, dt$$

The evaluation of which results in the logit choice probability,

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

An important property of such a model is worth mentioning. For any two alternatives i and k the ratio doesn't depend upon any other alternatives.

$$\frac{P_{ni}}{P_{nk}} = \frac{\frac{e^{V_{ni}}}{\sum_j V_{nj}}}{\frac{e^{V_{nk}}}{\sum_j V_{nj}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}}$$

That is, the relative odds of choosing i over k are the same no matter what other alternatives are available or what the attributes of the other alternatives are. The logit model exhibits this independence from irrelevant alternatives.

For a binary logit model,

$$P_{n1} = \frac{1}{1 + e^{V_{n2} - V_{n1}}}$$

We will use this model specification as a classifier to identify the customers who are likely to subscribe to the firm's product. Thus,

$$P_{n1} = \frac{1}{1 + e^{-z_i}}$$

Where,

$$z_i = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

Now, a generalised linear model has three components:

1) Response variables which are assumed to share the same distribution from the exponential family where the exponential family is represented by the following density function:

$$f(y_i|\theta_i) = \exp(y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i))$$

2) A set of parameters β and explanatory variables

$$X = \begin{bmatrix} {x_1}^T \\ . \\ {x_n}^T \end{bmatrix}$$

Where,

$${x_1}^T = [x_{11} \quad \cdots \quad x_{1p}]$$

3) A monotone link function $g(\mu_i) = {x_1}^T\beta$

For logistic regression the monotone function is the logit which is given by,

$$z_i = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

Where,

$$P(Response\ Variable = 1\ for\ i^{th}\ response) = p_i$$

Such a model is estimated using the likelihood function which is given by,

$$L = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$$

Where y_i takes the value $0\ or\ 1$

$$L = \prod_{i=1}^{N} (\frac{1}{1+e^{-z_i}})^{y_i}\,(1 - \frac{1}{1+e^{-z_i}})^{(1-y_i)}$$

$$L = \prod_{i=1}^{N} (\frac{1}{1+e^{-z_i}})^{y_i}\,(\frac{1}{1+e^{z_i}})^{(1-y_i)}$$

$$l = ln\ L = -\sum_{i=1}^{N} y_i \ln(1+e^{-z_i}) + (1-y_i)\ln(1+e^{z_i})$$

Now, the Score statistic is

$$S = \frac{dl}{d\beta}$$

This statistic is **0** at the maximum likelihood estimate.

The Jacobian is given by,

$$J = \frac{d(\frac{dl}{d\beta})}{d\beta}$$

The Newton-Raphson method involves the following iterative scheme:

$$b^m = b^{m-1} - J^{-1}S^{m-1}$$

Where the superscript m represents the m^{th} approximation and b is the vector of estimates.

MODEL ADEQUACY

To obtain the asymptotic distributions for a variety of statistics, we will use the Taylor series approximation. For the log likelihood function for a vector parameter β we have,

$$l(\beta) = l(b) + (\beta - b)^T S(b) + \frac{1}{2}(\beta - b)^T S'(b)(\beta - b)$$

Where, b is the estimate of β, $S(b)$ is the score statistic

$$l(\beta) = l(b) + (\beta - b)^T S(b) - \frac{1}{2}(\beta - b)^T J(b)(\beta - b)$$

If b is the maximum likelihood estimate of the log likelihood function, we have

$$l(\beta) - l(b) = -\frac{1}{2}(\beta - b)^T J(b)(\beta - b)$$

$$2\big(l(b) - l(\beta)\big) = (\beta - b)^T J(b)(\beta - b)$$

(1)

Now using a similar approach for the score statistic,

$$S(\beta) = S(b) + (\beta - b)^T S'(b)$$

$$S(\beta) = S(b) - J(b)(\beta - b)^T$$

Since b is the maximum likelihood statistic,

$$S(b) = 0$$

$$S(\beta) = -J(b)(\beta - b)^T$$

$$(b - \beta) = J(b)^{-1} S(\beta)$$

Now,

$$E[(b - \beta)(b - \beta)^T] = J(b)^{-1} E[S(\beta) S(\beta)^T] J(b)^{-1}$$

But,

$$E[S(\beta) S(\beta)^T] = J(b)$$

Thus,

$$E[(b-\beta)(b-\beta)^T] = J(b)^{-1}$$

Also,

$$(b-\beta)J(b)(b-\beta)^T \sim \chi^2{}_p \qquad (2)$$

So,

$$b \sim N(\beta, J(b)^{-1})$$

Define the likelihood ratio,

$$\lambda = \frac{L(b_{max}, y)}{L(b, y)}$$

$$2\log\lambda = 2[l(b_{max}, y) - l(b, y)] = D$$

Where D is the deviance.

Now,

$$D = 2[l(b_{max}, y) - l(\beta_{max}, y)] - 2[l(b, y) - l(\beta, y)] + 2[l(\beta_{max}, y) - l(\beta, y)]$$

From (1) and (2), the first term in the square brackets has the distribution $\chi^2{}_m$ where m is the number of parameters in the saturated model. The second term has the distribution $\chi^2{}_p$ where p is the number of parameters in the model of interest. The third term, $v = 2[l(\beta_{max}, y) - l(\beta, y)]$ is a positive constant which will be near zero if the model of interest fits the data almost as well as the saturated model fits.

$$D \sim \chi^2(m-p, v)$$

Where v is the non-centrality parameter.

CANONICAL INPUT FOR THE CODE

Input variables:

1 - age (numeric)

2 - job : type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur"," student","blue-collar","self-employed","retired","technician","services")

3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means divorced or widowed)

4 - education (categorical: "unknown","secondary","primary","tertiary")

5 - default: has credit in default? (binary: "yes","no")

6 - balance: average yearly balance, in euros (numeric)

7 - housing: has housing loan? (binary: "yes","no")

8 - loan: has personal loan? (binary: "yes","no")

related with the last contact of the current campaign:

9 - contact: contact communication type (categorical: "unknown","telephone","cellular")

10 - day: last contact day of the month (numeric)

11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

12 - duration: last contact duration, in seconds (numeric)

other attributes:

13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

Output variable (desired target):

17 - y - has the client subscribed a term deposit? (binary: "yes","no")

MINIMAL WORKING CODE

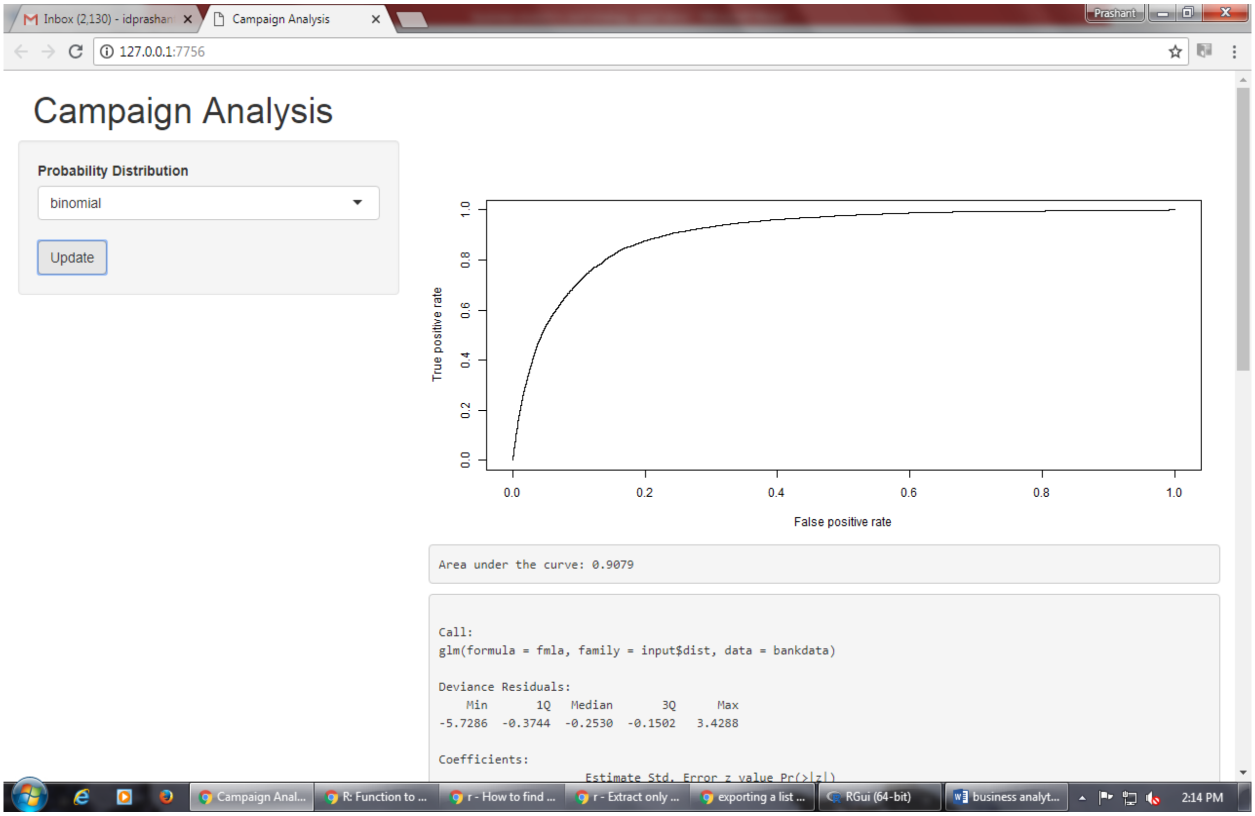
```
# we will load the shiny package that we have installed in the call below
library(shiny)


#we will call the user interface function by passing our arguments
ui <- fluidPage(
#we will be analysing a tele-calling promotional campaign


 headerPanel('Campaign Analysis'),
#the user will have multiple options to pick from
#We will pick the 'binomial' logit distribution to analyse the customers who have been
contacted to subscribe to term deposits
sidebarLayout(


 sidebarPanel(
   selectInput('dist', 'Probability Distribution',
c("binomial","gaussian","Gamma","inverse.gaussian","poisson","quasi","quasibinomial",
"quasipoisson")),
#this will be a call to update our analysis
 actionButton(inputId = "go",
   label = "Update")
),
mainPanel(
```

```
plotOutput("plt"),

verbatimTextOutput("p1"),

verbatimTextOutput("p2")

)


)

)



server <- function(input, output) {


#the model and its estimation will be based on the distribution selected by the user


#we will load the data here

bankdata<-read.csv(file="E:/bank-full.csv", header=TRUE, sep=",")


#this section is used to create our modeling formula

bankdata1<-bankdata

library(dplyr)

bankdata1<-select(bankdata1,-y)

fmla <- as.formula(paste("y ~ ", paste(names(bankdata1), collapse= "+")))
```

```
#for the bank marketing campaign data we will use the logistic regression as our model
of choice
bank1 <- eventReactive(input$go,{
glm(formula=fmla,family=input$dist,data=bankdata)
})

#we will load the library ROCR so that we can plot the ROC curve
library(ROCR)

output$plt<-renderPlot({
pred<-reactive(prediction(predict(bank1()),bankdata$y))
perf <- reactive(performance(pred(),"tpr","fpr"))
plot(perf())
})

#we will load the library pROC so that we can compute the area under the ROC curve
#this represents the probability that the higher outcome will be predicted higher by our
model than the lower outcome
library(pROC)
output$p1<-renderPrint(
auc(bankdata$y,predict(bank1()))
)
```

```
#finally we will print out the regression coefficients so that our campaign expenditure can
be optimised using subsequent analysis
output$p2<-renderPrint(
summary(bank1())
)

}

shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE



ANALYSIS

The area under the curve (AUC) statistic is 0.9079 which represents the probability that the higher coded dependent variable will be predicted higher by the classifier than the lower coded dependent variable. In particular, we will analyse the summary part of the output of our code here, so as to provide context for strategy formulation. There are a few independent variables in our regression model which represent the decision variables that the firm can control. The corresponding regression coefficient is a measure of the return on the investment made by the firm in those specific areas. Each of these estimated regression coefficients is a statistic with a standard error that represents the risk that comes with the return on investment.

We will assume the decision makers has a budget that needs to be allocated across these independent decision variables like say for instance, the number of contacts to be made or the duration of campaign calls. These variables come with a cost associated in executing them say a 1 unit increase in the duration of phone contacts costs the firm x dollars. So we will weigh the corresponding regression coefficient by dividing it by x. This metric represents the return on investing 1 dollar in that specific decision variable.

Let $R_i$denote such a return on investment for the i^{th} decision variable. Let $Var\ (R_i)$ denote the variance of this statistic. Let w_i be the proportion of investments allocated to the i^{th} decision variable. The proportion of allocations made by the firm sum to 1. Let us assume the goal of the decision maker is to limit the total risk to a certain value for the allocated investments. We will formulate this problem in the space of constrained optimization.

$$Max \sum_{i=1}^{n} w_i R_i$$

Subject to,

$$\sum_{i=1}^{n} w_i = 1$$

$$\sum_{i=1}^{n} Var\ (w_i R_i) < c$$

This problem can be solved using the Karush-Kuhn-Tucker conditions (for the inequality constraint) and the method of Lagrange Multipliers (for the equality constraint).

We formulate the Lagrangian as follows:

$$L(x, \lambda) = f(X) + \lambda h(x)$$

Where,

$$f(X) = \sum_{i=1}^{n} w_i R_i - \mu(\sum_{i=1}^{n} Var\ (w_i R_i) - c)$$

$$h(x) = \sum_{i=1}^{n} w_i - 1$$

The root of the gradient of this loss function is the optimal solution to our problem.

$$\nabla L(x, \lambda) = \nabla f(X) + \lambda \nabla h(x) = 0$$

$$\nabla f(X) = -\lambda \nabla h(x)$$

Thus the extremum point of our function subject to the constraints is where the gradient of $f(X)$ and the gradient of $h(x)$ lie in the same direction.

ALTERNATE COMPETING MODEL – THE PROBIT

Utility is decomposed into observed and unobserved parts:

$$U_{nj} = V_{nj} + \xi_{nj} \; for \; all \; j$$

Consider the vector composed of each ξ_{nj} , labeled $\xi_n{}' = (\xi_{n1}, \dots, \xi_{nj})$. We assume that ξ_n is distributed normal with a mean vector of zero and covariance matrix Ω. The density of $\xi_n$is

$$\emptyset(\xi_n) = \frac{1}{2\pi^{J/2}|\Omega|^{1/2}} e^{-\frac{\xi_n{}'\Omega^{-1}\xi_n}{2}}$$

The choice probability is,

$$P_{ni} = Prob\left(V_{ni} + \xi_{ni} > V_{nj} + \xi_{nj}\right) for \; all \; j \; not \; equal \; to \; i$$

$$P_{ni} = \int I(V_{ni} + \xi_{ni} > V_{nj} + \xi_{nj} for \; all \; j \; not \; equal \; to \; i) \, \emptyset(\xi_n) d\xi_n$$

Where $I(.)$ is an indicator of whether the statement in parentheses holds, and the integral is over all values of ξ_n. This integral does not have a closed form. It must be evaluated numerically through simulation.

We will look at two important properties of a probit model – taste variation and the repercussion on the independence from irrelevant alternatives.

Taste Variation: Assume that representative utility is linear in parameters and that the coefficients vary randomly over decision makers instead of being fixed. The utility is

$$U_{nj} = \beta_n{}' x_{nj} + \xi_{nj}$$

Suppose the β_n is normally distributed in the population with mean b and covariance W: $\beta_n \sim$ N(b,W). The goal of the research is to estimate the parameters b and W. The utility can be rewritten with β_n decomposed into its mean and deviations from its mean.

$$U_{nj} = b'x_{nj} + \widetilde{\beta_n}'x_{nj} + \xi_{nj}$$

The last two terms are random and denote their sum as η_{nj}.

$$E(\eta_{nj}) = E\left(\widetilde{\beta_n}'x_{nj} + \xi_{nj}\right) = 0$$

Assume that β_n is normally distributed with mean b and variance σ_B. Assume that $\xi_{nj}s$ are independently identically distributed with variance σ_ξ. Thus for a two-alternative model,

$$Var(\eta_{nj}) = {x_{nj}}^2\sigma_B + \sigma_\xi$$

$$Cov(\eta_{n1}, \eta_{n2}) = E\left(\widetilde{\beta_n}'x_{n1} + \xi_{n1}\right)\left(\widetilde{\beta_n}'x_{n2} + \xi_{n2}\right)$$

$$Cov(\eta_{n1}, \eta_{n2}) = x_{n1}x_{n2}\sigma_B$$

The covariance matrix is

$$\Omega = \begin{pmatrix} {x_{n1}}^2\sigma_B + \sigma_\xi & x_{n1}x_{n2}\sigma_B \\ x_{n1}x_{n2}\sigma_B & {x_{n2}}^2\sigma_B + \sigma_\xi \end{pmatrix}$$

$$\Omega = \sigma_B\begin{pmatrix} {x_{n1}}^2 & x_{n1}x_{n2} \\ x_{n1}x_{n2} & {x_{n2}}^2 \end{pmatrix} + \sigma_\xi\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Now economic behavior is unaffected by multiplicative transformation of utility. We therefore need to set the scale. This is done by setting $\sigma_\xi = 1$. Under this normalization,

$$\Omega = \sigma_B\begin{pmatrix} {x_{n1}}^2 & x_{n1}x_{n2} \\ x_{n1}x_{n2} & {x_{n2}}^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Repercussion on the Independence from Irrelevant Alternatives property: We will consider the famous red bus-blue bus problem. A traveler has a choice of going to work by car or taking a blue bus. For simplicity assume that the representative utility of the two modes are the same, such that the choice probabilities are equal: $P_c = P_{bb} = 1/2$, where c is car and bb is blue bus. In this case, the ratio of probabilities is one:

$$P_c/P_{bb} = 1$$

Now suppose that a red bus is introduced and that the traveler considers the red bus to be exactly like the blue bus. The probability that the traveler will take the red bus is therefore the same as for the blue bus, so that the ratio of their probabilities is one:

$$P_{rb}/P_{bb} = 1$$

However, in the logit model the ratio P_c/P_{bb} is the same whether or not another alternative, in this case the red bus, exists. This ratio therefore remains at one. The only probabilities for which $P_c/P_{bb} = 1$ and $P_{rb}/P_{bb} = 1$ are $P_c = P_{bb} = P_{rb} = 1/3$, which are the probabilities that the logit model predicts. In real life, however, we would expect the probability of taking a car to remain the same when a new bus is introduced that is exactly the same as the old bus.We would also expect the original probability of taking bus to be split between the two buses after the second one is introduced. That is, we would expect $P_c = 1/2$ and $P_{bb} = P_{rb} = 1/4$. In this case, the logit model, because of its IIA property, overestimates the probability of taking either of the buses and underestimates the probability of taking a car. The ratio of probabilities of car and blue bus, P_c/P_{bb}, actually changes with the introduction of the red bus, rather than remaining constant as required by the logit model.

We will use the method of maximum simulated likelihood wherein the procedure is the same as maximum likelihood except that simulated probabilities are used instead of exact probabilities. The log likelihood function is given by:

$$LL(\theta) = \sum_n \ln P_n(\theta)$$

Where θ is a vector of parameters, $P_n(\theta)$ is the exact probability of the observed choice of observation n and the summation is over a sample of N observations. The maximum likelihood (ML) estimator is the value of θ that maximizes $LL(\theta)$. Since the gradient of $LL(\theta)$ is zero at the maximum, the ML estimator can also be defined as the value of θ at which

$\sum_n S_n(\theta) = 0$ where $S_n(\theta) = \frac{\partial \ln P_n(\theta)}{\partial\theta}$ is the score for observation n.

Let $\widetilde{P_n(\theta)}$ be a simulated approximation to $P_n(\theta)$. The simulated log likelihood function is $SLL(\theta) = \sum_n \ln \widetilde{P_n(\theta)}$, and the maximum simulated likelihood (MSL) estimator is the value of θ that maximizes $SLL(\theta)$.

Let the score be denoted by $g_n(\theta^*)$ which varies over the decision makers in the population. When taking a sample the researcher is drawing values of $g_n(\theta^*)$ from its distribution in the population. The distribution has zero mean by assumption and variance denoted by W. The researcher calculates the sample mean of these draws $g(\theta^*)$. By the central limit theorem,

$$\sqrt{N}(g(\theta^*) - 0) \xrightarrow{d} N(0, W) \qquad (1)$$

such that $\quad g(\theta^*) \sim N(0, \frac{W}{N})$ asymptotically.

We will now derive the properties of the estimator $g(\hat{\theta})$ from the distribution of $g(\theta^*)$. We will take the Taylor's first order expansion of $g(\hat{\theta})$ around $g(\theta^*)$.

$g(\hat{\theta}) = g(\theta^*) + \mathrm{D}(\hat{\theta} - \theta^*)$ where D = $\frac{\partial g(\theta^*)}{\partial\theta}$. Now by definition,

$g(\hat{\theta}) = 0$ for maximum likelihood estimation.

$$(\hat{\theta} - \theta^*) = -D^{-1}g(\theta^*)$$

$$\sqrt{N}(\hat{\theta} - \theta^*) = \sqrt{N}(-D^{-1})g(\theta^*)$$

Now from (1),

$\sqrt{N}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, D^{-1}WD^{-1})$ such that $\hat{\theta} \sim N\left(\theta^*, \frac{D^{-1}WD^{-1}}{N}\right)$ asymptotically

We will express the simulated value, $\widetilde{g(\theta)}$ as

$$\widetilde{g(\theta^*)} = \widetilde{g(\theta^*)} + g(\theta^*) - g(\theta^*) + E_r\widetilde{g(\theta^*)} - E_r\widetilde{g(\theta^*)}$$

$$\widetilde{g(\theta)} = g(\theta^*) + [E_r\widetilde{g(\theta^*)} - g(\theta^*)] + [\widetilde{g(\theta^*)} - E_r\widetilde{g(\theta^*)}]$$

$$\widetilde{g(\theta)} = A + B + C$$

Where A is the non-simulated estimator, B is the simulation bias and C is simulation noise. We will first consider the simulation noise, C.

$$\mathrm{C} = \widetilde{g(\theta^*)} - E_r\widetilde{g(\theta^*)}$$

$$\mathrm{C} = \frac{1}{N}\sum_n \widetilde{g_n(\theta^*)} - E_r\widetilde{g_n}(\theta^*)$$

$$\mathrm{C} = \frac{1}{N}\sum_n d_n$$

There is a distribution of values of d_n over the possible realizations of the draws used in simulation. The distribution has zero mean, since the expectation over draws is subtracted out when creating d_n. Label the variance of the distribution as S_n/R, where S_n is the variance when one draw is used in simulation.

$\sqrt{N}\mathrm{C} \xrightarrow{d} N(0, S/R)$such that $C \sim N\left(0, \frac{S}{RN}\right)$ where S is the population mean of S_n

We will now consider the bias term. Usually, the defining term $g_n(\theta^*)$ is a function of a statistic, l_n, that can be simulated without bias. In our case this is the choice probability. We will now re-express $\widetilde{g_n(\theta^*)}$ by taking a taylor series expansion around the unsimulated value.

$$\widetilde{g_n(\theta^*)} = g_n(\theta^*) + g_n{}'\left[\widetilde{l_n(\theta^*)} - l_n(\theta^*)\right] + \frac{1}{2} g_n{}''\left[\widetilde{l_n(\theta^*)} - l_n(\theta^*)\right]^2$$

$$E_r \widetilde{g_n(\theta^*)} - g_n(\theta^*) = \frac{1}{2} g_n{}'' Var_r(\widetilde{l_n(\theta^*)})$$

The simulation bias is then,

$$E_r \widetilde{g(\theta^*)} - g(\theta^*) = \frac{1}{N}\sum_n E_r \widetilde{g_n(\theta^*)} - g_n(\theta^*)$$

$$E_r \widetilde{g(\theta^*)} - g(\theta^*) = \frac{1}{N}\sum_n g_n{}'' \frac{Q_n}{2R}$$

Where $Var_r(\widetilde{l_n(\theta^*)}) = \frac{Q_n}{R}$

$$E_r \widetilde{g(\theta^*)} - g(\theta^*) = \frac{Z}{R}$$

Where Z is the sample average of $g_n{}'' \frac{Q_n}{2}$

Thus,

$$\sqrt{N}B = \sqrt{N}\frac{Z}{R}$$

Now using Taylor's expansion,

$$\sqrt{N}\left(\hat{\theta} - \theta^*\right) = \sqrt{N}(-D^{-1})\widetilde{g_n(\theta^*)}$$

$$\sqrt{N}\left(\hat{\theta} - \theta^*\right) = \sqrt{N}(-D^{-1})(\mathrm{A} + \mathrm{B} + \mathrm{C})$$

And the estimator can be expressed as

$$\hat{\theta} = \theta^* - (D^{-1})(\mathrm{A} + \mathrm{B} + \mathrm{C})$$

We can now examine the properties of our estimators. For the bias term, as R increases faster than N,

$$\sqrt{N}B = \sqrt{N}\frac{Z}{R} \xrightarrow{p} 0$$

Similarly the simulation noise term also vanishes,

$$S/R \xrightarrow{p} 0$$

Thus we have from (1),

$$\sqrt{N}\left(\hat{\theta} - \theta^*\right) = \sqrt{N}(-D^{-1})\mathrm{A} \xrightarrow{d} N(0, D^{-1}WD^{-1})$$

Now $D = H = E(\frac{\partial^2(\ln P_n(\theta^*))}{\partial\theta^2})$ and by the information identity,

$$W = -H$$

Thus,

$$\sqrt{N}\left(\hat{\theta} - \theta^*\right) = \sqrt{N}(-D^{-1})\mathrm{A} \xrightarrow{d} N(0, H^{-1}(-H)H^{-1})$$

$$\sqrt{N}\left(\hat{\theta} - \theta^*\right) = \sqrt{N}(-D^{-1})\mathrm{A} \xrightarrow{d} N(0, -H^{-1})$$

$\hat{\theta} \sim N(0, -H^{-1}/N)$ asymptotically

ALTERNATE COMPETING MINIMAL WORKING ALGORITHMS

The Accept-Reject simulator when the maximum simulated likelihood estimation is used is calculated as follows:

1. Draw $\widetilde{\xi_{ni}}^{r} = L_i \eta^r$ as follows
    a. Draw J − 1 values from a standard normal density using a random number generator. Stack these values into a vector, and label the vector, η^r.
    b. Calculate $\widetilde{\xi_{ni}}^{r} = L_i \eta^r$ where L_i is the choleski factor of Ω, the covariance matrix.
2. Using these values of the errors, calculate the utility difference for each alternative, differenced against the utility of alternative i. That is, calculate

$$\widetilde{U_{nji}}^{r} = V_{nj} - V_{ni} + \widetilde{\xi_{nji}}^{r}$$

3. Determine whether each utility difference is negative. Thus, calculate $I^r = 1$indicating an accept, and $I^r = 0$ otherwise, indicating a reject.

4. Repeat steps 1–3 many times. Label the number of repetitions (including the first) as R, so that r takes values of 1 through R.

5. The simulated probability is the number of accepts divided by the number of repetitions. Thus,

$$\widetilde{P_{ni}} = \frac{1}{R} \sum_{r=1}^{R} I^r$$

The Accept-Reject simulator when the method of simulated scores is used is as follows:

1. Take a draw of the random terms from their density.

2. Calculate the utility of each alternative with this draw.
3. Determine whether alternative j has the highest utility.
4. If so, call the draw an accept. If not, then call the draw a reject and repeat steps 1 to 3 with a new draw. Define B^r as the number of draws that are taken until the first accept is obtained.

Perform steps 1 to 4 R times, obtaining B^r for $r = 1, \dots, R$. The simulator of $\frac{1}{P_{nj}(\theta)}$ is $\frac{1}{R}\sum_{r=1}^{R} B^r$

BUSINESS ANALYTICS AND STRATEGY

Project 2

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

HIGH LEVEL TAKE-AWAY

1) Diagnosis: Our profitability is low and we aren’t making enough money. But we cannot afford to raise prices across the board.
2) Forecast: We would like an x% increase in our profits.
3) Add value through search: If the prices cannot be raised across the board, we restrict our search to a set of consumers where we can afford to do this. This set would typically comprise of consumers who aren’t in line with the amount of business that we do with them.
4) Choose: Select the consumers who are receiving disproportionate discounts. We will use the customer lifetime value metric in our analysis to look at the value they bring.
5) Commit: Renegotiate our commercial terms with this set of consumers.
6) Balanced Scorecard Strategy Map: We will get our customers in line with our targeted discount rate offerings. This will increase our overall profitability. Such a reduction in costs, in part, can be used to fuel growth and retention programs. This will increase our overall revenues.

MODEL

Hazard models are used to compute the probability that the customer is still with the company and hasn't gone to the competition. Let t be a random variable representing the time the customer attrites (dies) with probability density function $f(t)$. Let $F(t)$ be the cumulative distribution function such that $F(t) = \int_0^t f(t)dt$. The survivor function is given by

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(t)dt$$

Now hazard function is the probability that the customer attrites in an instantaneous time period Δt given that the customer has remained with the firm till time t. This is given by

$$h(t) = \frac{f(t)}{S(t)}$$

For an exponential probability density function for time t, we have

$$f(t) = \lambda e^{-\lambda t}$$

$$S(t) = e^{-\lambda t}$$

$$h(t) = \lambda$$

Thus the hazard function for an exponential distribution is a constant.

Thus the retention rate,

$$r = 1 - h(t)$$

The value of a customer upto period τ is

$$\sum_{t=1}^{\tau} \frac{r^{t-1}(R_t - C_t)}{(1+\delta)^{t-1}}$$

Where R_t is revenue, δ is the discount rate and C_t is cost in period t.

We will assume a constant hazard rate and profit contribution to demonstrate a canonical methodology for computation. Thus

$$\sum_{t=1}^{\tau}\frac{r^{t-1}(R_t - C_t)}{(1+\delta)^{t-1}} = (R-C) + \frac{(R-C)r}{(1+\delta)} + \frac{r^2(R-C)}{(1+\delta)^2} + \cdots$$

$$\sum_{t=1}^{\tau}\frac{r^{t-1}(R_t - C_t)}{(1+\delta)^{t-1}} = (R-C)[1 + \frac{r}{(1+\delta)} + \frac{r^2}{(1+\delta)^2} + \cdots]$$

$$\sum_{t=1}^{\tau}\frac{r^{t-1}(R_t - C_t)}{(1+\delta)^{t-1}} = (R-C)(\frac{1}{1-\frac{r}{(1+\delta)}})$$

$$\sum_{t=1}^{\tau}\frac{r^{t-1}(R_t - C_t)}{(1+\delta)^{t-1}} = (R-C)(\frac{1+\delta}{1+\delta-r})$$

Alternatively, we can model a weibull distribution for our data where

$$f(t) = (\frac{a}{b})(\frac{t}{b})^{a-1}e^{-(\frac{t}{b})^a}$$

$$F\ (t) = 1 - e^{-(\frac{t}{b})^a}$$

$$S(t) = \ e^{-(\frac{t}{b})^a}$$

$$h(t) = (\frac{a}{b})(\frac{t}{b})^{a-1}$$

The way the hazard function is parameterized above is how the rweibull in the software R is also modeled. So,

$$h(t) = b^{-a}at^{a-1}$$

This is the baseline hazard; we will further adopt the proportional hazard formulation to model covariates. Thus,

$$h(t) = b^{-a}at^{a-1}e^{\beta^T z}$$

Where, z are the covariates to be modeled

Integrating with respect to t we have the cumulative hazard function,

$$H(t) = (\frac{t}{b})^{a} e^{\beta^{T} z}$$

Multiplying by -1 and exponentiating we get the survivor function,

$$S(t) = \exp(-(\frac{t}{b})^{a} e^{\beta^{T} z})$$

We will make the following transformation $X = Log(T)$

$$P(X > x|z) = \exp(-(\frac{e^{x}}{b})^{a} e^{\beta^{T} z})$$

$$P(X > x|z) = \exp(-\exp[-alogb + \ ax + \ \beta^{T} z \,])$$

Now from the help file on the survreg function in R we have,

survreg's scale = 1/(rweibull shape)

survreg's intercept = log(rweibull scale)

So let μ and σ denote the intercept and scale of survreg. In addition, let us define

γ as the

regression coefficients of survreg. Further, the substitutions, $b = \ e^{\mu}$, $a = \ \frac{1}{\sigma}$

, and $\beta = \ \frac{-\gamma}{\sigma}$ give

$$P(X > x|z) = \exp(-\exp \frac{1}{\sigma} [x - \mu - \ \gamma^{T} z \,])$$

This is equivalent to

$X = Log\ (T) = \ \mu + \ \gamma^{T} z + \ \sigma W$

Where W is an extreme value random variable.

Such a model is estimated by maximizing the likelihood function,

$$L(.|t) = \prod_{i=1}^{n} f(t)^{v_i} S(t)^{1-v_i}$$

$$L(.|t) = \prod_{i=1}^{n} (\frac{a}{b})(\frac{t}{b})^{a-1} e^{-(\frac{t}{b})^a}{}^{v_i} e^{-(\frac{t}{b})^a}{}^{1-v_i}$$

With a covariate vector z we have,

$$L(.|t) = \prod_{i=1}^{n} (\frac{a}{b_1})(\frac{t}{b_1})^{a-1} e^{-(\frac{t}{b_1})^a}{}^{v_i} e^{-(\frac{t}{b_1})^a}{}^{1-v_i}$$

Where,

$$\frac{1}{b_1} = \frac{1}{b} e^{\frac{\beta^T z}{a}}$$

CANONICAL INPUT FOR THE CODE

Data required to compute the customer level hazard function:

1) Observation time for the customer (time)

2) Whether the customer is still observed to be with the firm (status)

3) Average price of the product and service availed by the customer (price)

4) Number of promotions received by the customer (promotions)

5) Number of Sales calls in any given period of time on an average (calls)

6) Size of the customer (size)

7) Type of product availed by the customer (product)

8) Customer segment (segment)

Data required to plot customer lifetime value (CLV) vs. discount:

2) Revenue stream expected in the coming years (revenue)

3) Cost stream incurred for a customer by the firm in the coming years (cost)

4) Discount negotiated by the customer (discount)

5) Annual discount factor (factor)

MINIMAL WORKING CODE

```
# app1

# we will load the shiny package that we have installed in the call below

library(shiny)


#we will call the user interface function by passing our arguments

ui <- fluidPage(

#we will be conducting a pool analysis wherein we will plot the customer lifetime value

# vs discounts offered by the firm

 headerPanel('Pool Analysis'),

#the user will have two options to pick from

#either an exponential hazard which is constant

#or a weibull hazard where the scale is not equal to one that either monotonically

decreases or increases with time


 sidebarPanel(

   selectInput('dist', 'Probability Distribution', c("exponential","weibull"))

 ),

#this will be a call to update our pool

 actionButton(inputId = "go",

   label = "Update"),

plotOutput("plt")
```

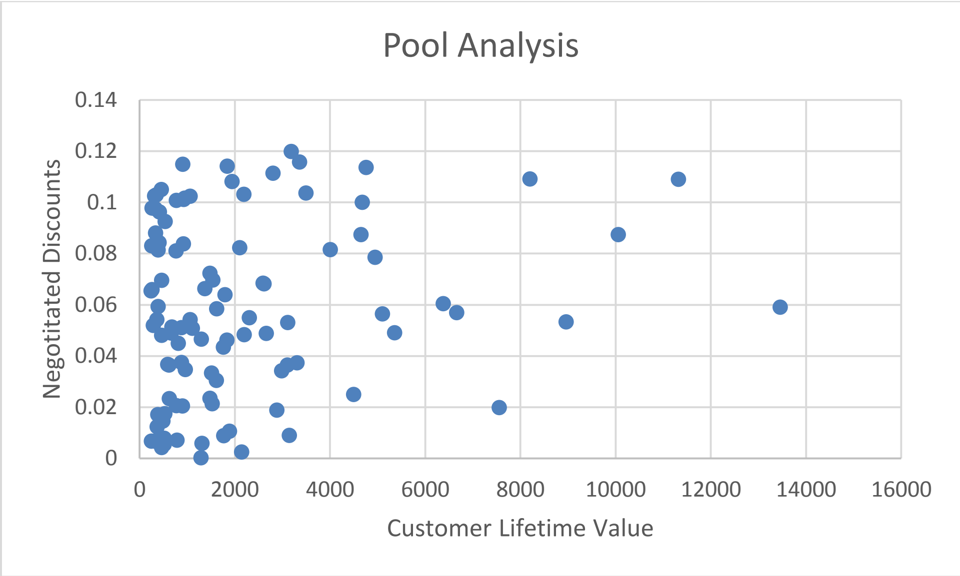
```
)

#the server code will build the output required for subsequent analysis

server <- function(input, output) {

#we will call the library 'survival' so that our data can be modeled using survreg()
library(survival)

#the model and its estimation will be based on the distribution selected by the user
#the covariates modeled here are price, promotions, sales calls, size of the firm, product
type, and customer segment

#we will attach a pre-transformed and loaded file here
attach(survdata1)
#I understand transforming and loading of data is one of the key stages in our end-to-end
process

data1 <- eventReactive(input$go,{
   survreg(Surv(time, status) ~ price + promotions + calls + size + product + segment,
dist='input$dist')
  })
```

```
#we will save the summary of our model output in the variable summ1
summ1<-reactive({
summary(data1())
})

#we will convert the estimated accelerated failure time surveg() parameters to rweibull
proportional hazard parameters

rweibullscale<-reactive({(exp(summ1()$coeff[1]))})
rweibullshape<-reactive({1/summ1()$scale})

diff<-reactive({abs(rweibullshape()-1)})

#we will make a distinction between the exponential and weibull distribution here
output$plt<-renderPlot({
if(diff()<0.0001)
#we will model a constant hazard
#and we will compute the customer lifetime value (clv)
{
ehazard<-reactive((1/rweibullscale())*exp(summ1()$coeff[2]*price +
summ1()$coeff[3]*promotions + summ1()$coeff[4]*calls + summ1()$coeff[5]*size +
summ1()$coeff[6]*product + summ1()$coeff[7]*segment))
retention<-reactive(1-ehazard())
```

```
clv<-(revenue-cost)*(1+factor)/(1+factor-retention())

plot(clv,discount)

}

#else we will model a varying hazard and discretise the function

#this is because we see revenue streams as being discrete-either on a monthly or an
annual basis

{

clv<-(revenue-cost)

#we will calculate the clv based on 5 yrs prediction of revenue and cost

for(t in 1:5)

{

#whazard<-matrix(nrow=1,ncol=5)

#retention<-matrix(nrow=1,ncol=5)


#we will attempt to calculate the average hazard rate by averaging out the values over the
discretised time interval

whazard1<-
reactive((rweibullshape()/rweibullscale())*((t/rweibullscale())^(rweibullshape()-
1))*exp(summ1()$coeff[2]*price + summ1()$coeff[3]*promotions +
summ1()$coeff[4]*calls + summ1()$coeff[5]*size + summ1()$coeff[6]*product +
summ1()$coeff[7]*segment))

whazard2<-
reactive((rweibullshape()/rweibullscale())*(((t+0.5)/rweibullscale())^(rweibullshape()-
```

```
1))*exp(summ1()$coeff[2]*price + summ1()$coeff[3]*promotions +
summ1()$coeff[4]*calls + summ1()$coeff[5]*size + summ1()$coeff[6]*product +
summ1()$coeff[7]*segment))
whazard3<-
reactive((rweibullshape()/rweibullscale())*(((t+1)/rweibullscale())^(rweibullshape()-
1))*exp(summ1()$coeff[2]*price + summ1()$coeff[3]*promotions +
summ1()$coeff[4]*calls + summ1()$coeff[5]*size + summ1()$coeff[6]*product +
summ1()$coeff[7]*segment))
whazard<-reactive((whazard1+whazard2+whazard3)/3)
retention<-reactive(1-whazard())
clv<-clv+retention()^t*(revenue-cost)/(1+factor)^t
}
plot(clv,discount)
}

})




#server function ends here
}

shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE



ANALYSIS

A typical price waterfall begins with a list price. This is where the negotiations for discounts happen. Such discounts happen for a brief period of time and attempt to raise volume and market share. Other types of discounts include offers extended to the client via growth programs, promotion and inventory rebates among a host of other activities carried out by the firm.

Fig. 1. Pricing Waterfall

We will attach the output of our code so as to provide context for analysis.

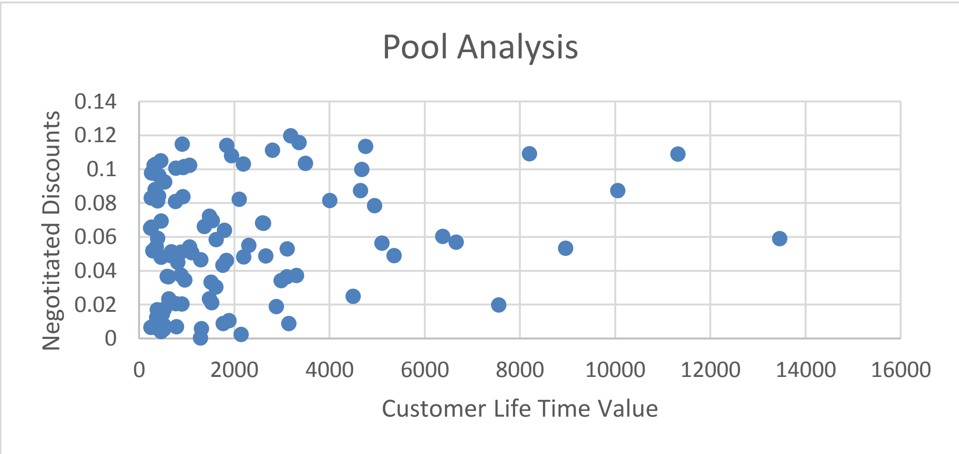


Fig. 2. CLV Pool Analysis

Firms typically carry out a profit pool analysis. But the profits generated are a snapshot at any given point of time. Instead of price and quantity, we will use the customer lifetime value and discounts to inform our analysis. So, we have plotted negotiated discounts offered to the customer vs the customer lifetime value (CLV). Such a method of analysis, we believe is more forward looking as it takes into account future revenue streams as well.

Pricing has to be one of the least understood facets of doing business and is usually driven by gut and competition. There are likely to be customers that receive high discounts despite not contributing to high CLV levels. To think of raising prices may seem disastrous at first; but we will not be carrying out such an exercise across the board but will engage in a discriminatory methodology wherein we will target only low performing customers to renegotiate discounts. Likewise, there are likely to be a number of customers who contribute to high CLV levels but are not offered substantial discounts. We will also target these customers to retain and increase their sales share. Thus, after analyzing say, for instance, negotiated discounts, we will carry out this mode of looking at CLV and discounts for each of the discount types that the firm offers like growth programs, promotions, and inventory rebates. From fig.1, even a 1% increase in the pocket price will lead to a 5% increase in pocket margin.

Also, concurrently, we will attack the problem from 2 other directions. Because of our renegotiations with the customers it is likely we will lose a percentage to the competition. We will put in place a growth program to acquire new customers. We will also deploy a loyalty program to increase customer retention rates. This should see us serving a better pool of clientele.

BUSINESS ANALYTICS AND STRATEGY

Project 3

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

HIGH LEVEL TAKE-AWAY

1) Diagnosis: We have a regime of rapidly changing technology and it is becoming difficult to keep a track of our portfolio mix.
2) Forecast: We have a typical portfolio of business units in mind which is at variance with our current offerings.
3) Add value through search: We need to figure out a method of going from the current A to the forecasted B for which we will need access to the relationship between the market shares for our business units, competitor data, and the decision variables that the firm controls.
4) Choose: We will through the process of modeling arrive at the drivers of our market share. These levers should set us up for the road ahead.
5) Commit: Allocate resources according to the decision rule set by our strategy.
6) Balanced Scorecard Strategy Map: Our investments in the decision variables will let us manipulate our business unit level market shares. This should see us strengthening our cash cows and high growth stars. Concurrently, we will also look to retire or upgrade our other units gradually. We will be serving a more profitable portfolio mix.

MODEL

From Kotler (1984), we have

$$S_i = kM_i$$

Where, the firm's market share is proportional to its marketing effort.

$$M_i = \text{the marketing effort of the product of firm } i$$

$$k = \text{the propotionality constant}$$

But,

$$\sum_{i=1}^{m} S_i = 1$$

Thus,

$$\sum_{i=1}^{m} kM_i = 1$$

$$k = \frac{1}{\sum_{i=1}^{m} M_i}$$

So,

$$S_i = \frac{M_i}{\sum_{i=1}^{m} M_i}$$

We will assume that the firms vary in the effectiveness of their marketing efforts.

Thus,

$$S_i = \frac{\alpha_i M_i}{\sum_{i=1}^{m} \alpha_i M_i}$$

Now, the Multiplicative Competitive Interaction (MCI) model is given by,

$$M_i = P_i^{p_i}. A_i^{a_i} D_i^{d_i}$$

Where,

$p_i, a_i, and\ d_i$ are the parameters to be estimated.

The Multinomial Logit (MNL) model is given by,

$$M_i = e^{p_i P_i + a_i A_i + d_i D_i}$$

We will use a combination of MCI and MNL models. Thus,

$$S_i = \frac{A_i}{\sum_{J=1}^{m} A_j}$$

Where A_i is the attraction of brand i given by,

$$A_i = e^{(\alpha_i + \xi_i)} \prod_{k=1}^{K} f_k(X_{ki})^{\beta_k}$$

The market share is assumed to increase as X_{ki} increases.

So for a variable like say price at very low levels the point elasticity is likely to be high, we will have,

$$e_{S_i} = \beta_k(1 - S_i)$$

And,

$$f_k = X_{ki}$$

Whereas for a variable like say advertising expenditure at very low levels the point elasticity is likely to be negligible. So we will have,

$$e_{S_i} = \beta_k(1 - S_i)\, X_{ki}$$

And,

$$f_k = e^{X_{ki}}$$

Where,

$$e_{S_i} = the\ point\ elasticity\ at\ S_i$$

Now, we will provide a method to compute the market share for each brand. Let us assume that the purchase frequency that is the number of purchases per period by an individual buyer is a random variable which has a statistical distribution. We will also assume that the buyer does not purchase the same brand always but does so with a probability. This will be called the individuals choice probability. So we will compute the brands market share for a specified period from the individual level buying behavior as,

$$Market\ Share_i = \frac{Average\ number\ of\ Units\ Purchased\ for\ Brand\ i}{Average\ number\ of\ Units\ Purchased\ for\ All\ Brands}$$

$$E(S_i) = \frac{1}{\bar{\mu}}\int_0^{\infty}\int_0^{1} \mu\pi_i g(\mu,\pi_i)\, d\pi_i d\mu$$

Where,

$E(S_i) = Expected\ Market\ Share\ for\ brand\ i$

$\mu = the\ mean\ purchase\ frequency\ per\ period\ (per\ individual)$

$\bar{\mu} = the\ population\ mean\ of\ \mu$

$\pi_i = the\ individual\ choice\ probability\ for\ brand\ i$

$g(\mu,\pi_i) = the\ joint\ density\ function\ for\ \mu\ and\ \pi_i$

Now,

$$\int_0^{\infty}\int_0^{1} \mu\pi_i g(\mu,\pi_i)\, d\pi_i d\mu = \bar{\mu}\bar{\pi}_i + cov(\mu,\pi_i)$$

Where,

$$\bar{\pi}_i = the\ population\ mean\ of\ \pi_i$$

$$cov(\mu,\pi_i) = the\ covariance\ between\ \mu\ and\ \pi_i$$

Thus,

$$E(S_i) = \frac{\bar{\mu}\bar{\pi}_i + cov(\mu,\pi_i)}{\bar{\mu}}$$

$$E(S_i) = \bar{\pi}_i + \frac{cov(\mu,\pi_i)}{\bar{\mu}}$$

Finally, will use the log-linear regression methodology to model and analyse our market shares. Specifically for the canonical MNL model,

$$\log(S_i) = \alpha_i + \sum_{k=1}^{K} \beta_k X_{ki} + \xi_i - \log[\sum_{j=1}^{m} e^{(\alpha_i + \sum_{k=1}^{K} \beta_k X_{ki} + \xi_i)}]$$

If we sum the above equation over i (i $= 1, 2, \ldots,$ m) and divide by m, we have

$$\log(\overline{S_\iota}) = \overline{\alpha_\iota} + \sum_{k=1}^{K} \beta_k \overline{X_{k\iota}} + \overline{\xi_\iota} - \log[\sum_{j=1}^{m} e^{(\alpha_i + \sum_{k=1}^{K} \beta_k X_{ki} + \xi_i)}]$$

Where, $\overline{S_\iota}$ is the geometric mean and $\overline{\alpha_\iota}, \overline{X_{k\iota}}, \ and \overline{\xi_\iota}$ are the arithmetic means

$$\log\left(\frac{S_i}{\overline{\overline{S_\iota}}}\right) = (\alpha_i - \overline{\alpha_\iota}) + \sum_{k=1}^{K} \beta_k (X_{ki} - \overline{X_{k\iota}}) + (\xi_i - \overline{\xi_\iota})$$

This specification is linear in the parameters and can be estimated by linear regression and so we will be using the lm () procedure in R.

We can indeed extend and enrich our model further by,

$$A_i = e^{(\alpha_i + \xi_i)} \prod_{k=1}^{K} \prod_{j=1}^{m} f_k(X_{kj})^{\beta_k ij}$$

$$S_i = \frac{A_i}{\sum_{j=1}^{m} A_j}$$

Where for $\mathrm{i} = \mathrm{j}$, we arrive at the direct effect of the focal firm's marketing effort on market share; whereas for i different than j, we get the indirect effect due to the competitors' marketing efforts such that β_{kij} is the parameter for the cross-competitive effect of variable X_{kj} on brand i.

Also where we have the problem of heteroskedasticity in estimation, we can use the method of Generalized Least Squares (GLS) to estimate our model parameters. So let

V be the variance-covariance matrix of the error term in our regression model. Here, $V$is a symmetric non-singular matrix. Therefore,

$$V = K'K = KK$$

Where K is the squared root matrix and equal to $V^{1/2}$. We define,

$Z = K^{-1}Y, B = K^{-1}X, and\ G = K^{-1}\xi$ such that $Z = B\beta + G$, then

$$E(G) = K^{-1}E(\xi) = 0$$

$$Var\ (G) = K^{-1}Var(\xi)K^{-1}$$

$$Var\ (G) = K^{-1}VK^{-1}$$

$$Var\ (G) = K^{-1}(KK)K^{-1}$$

$$Var\ (G) = I$$

The steps of an iterative Generalised Least Squares procedure are as follows.

1. The Ordinary Least Square procedure is used to estimate the parameters in one of the regression models, and V is estimated from the residuals.
2. The data are re-weighted by the estimated $V^{-1/2}$
3. The first two steps are repeated until the estimated values of the regression parameters converge.

CANONICAL INPUT FOR THE CODE

Input variables:

1 - log centered price for the business unit

2 - centered advertising numbers for the business unit

3 - dummy variable for the brands with brand 1 being the reference brand

4 - competitor data on price and advertising

5 - aggregate market share data for the business unit and the competition

MINIMAL WORKING CODE

```
# app3

# we will load the shiny package that we have installed in the call below

library(shiny)


#we will call the user interface function by passing our arguments

ui <- fluidPage(


 headerPanel('Market Share Analysis'),

#the user will have multiple options to pick from

#We will pick the 'MCI+MNL' specification to model our business unit level market shares

sidebarLayout(


 sidebarPanel(

   selectInput('dist', 'Model', c("MCI","MNL","MCI+MNL")),

#this will be a call to update our analysis

 actionButton(inputId = "go",

   label = "Update")

),

mainPanel(

verbatimTextOutput("p1")

)
```

```
)
)

server <- function(input, output) {

#we will use the linear regression lm() specification to model our data

#we will attach the pre-loaded input data here
attach(mshare)

#in the absence of competitor information we will use data on the focal firm's advertisement and price
mshare1 <- eventReactive(input$go,{
lm(lshare ~ bdummy+cadvert+lcprice,data=mshare)
})

output$p1<-renderPrint(
summary(mshare1())
)
```

```
}

shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE

For each of the business unit, the firm will have access to the decision variables like price and advertising that it controls along with competitor data. We will be able to predict business unit level market shares. But these will be logged market shares from our model. For each of the n brands that will include our competitors, we will be able to arrive at this value, $\log\left(\frac{S_i}{\bar{S}_i}\right)$.

To find the actual market shares we will have to solve for m such simultaneous equations in m parameters. Finally we will compute the ratio of market share of our firm to the market share of the 3rd placed firm. This subsequent processing is required to carry out our portfolio mix analysis. So we will have a forecast of the mix - where we want to be - in terms of market shares and growth rates for each of the business units so that a balance in our portfolio can be achieved.

To achieve such a balanced mix, we need to arrive at the required value of the decision variables that the firm controls. This will point to the investment calls to be made for each of these decision variables. The exercise will typically be carried out keeping the risk of investments being made at the forefront of our decision making process.

ANALYSIS

We will use the predicted market shares from our model to inform our analysis. We will take that as our basis for subsequent analysis. In particular, we will use the Boston Consulting Group’s growth share matrix to evaluate our investments. This matrix helps companies decide which markets and business units to invest in based on two factors – company competitiveness and market attractiveness. These factors are operationalized by the market share and the growth rate respectively.

Thus, the high share-low growth rate units should be milked for cash (cash cows), high share-high growth rate units should be invested in (stars), low share-low growth rate should be divested (dogs) and low share-high growth rate units (question marks) should either be discarded or invested based on their individual potential. This traditional framework provides the firm the necessary input to exploit mature units and explore new growth areas.

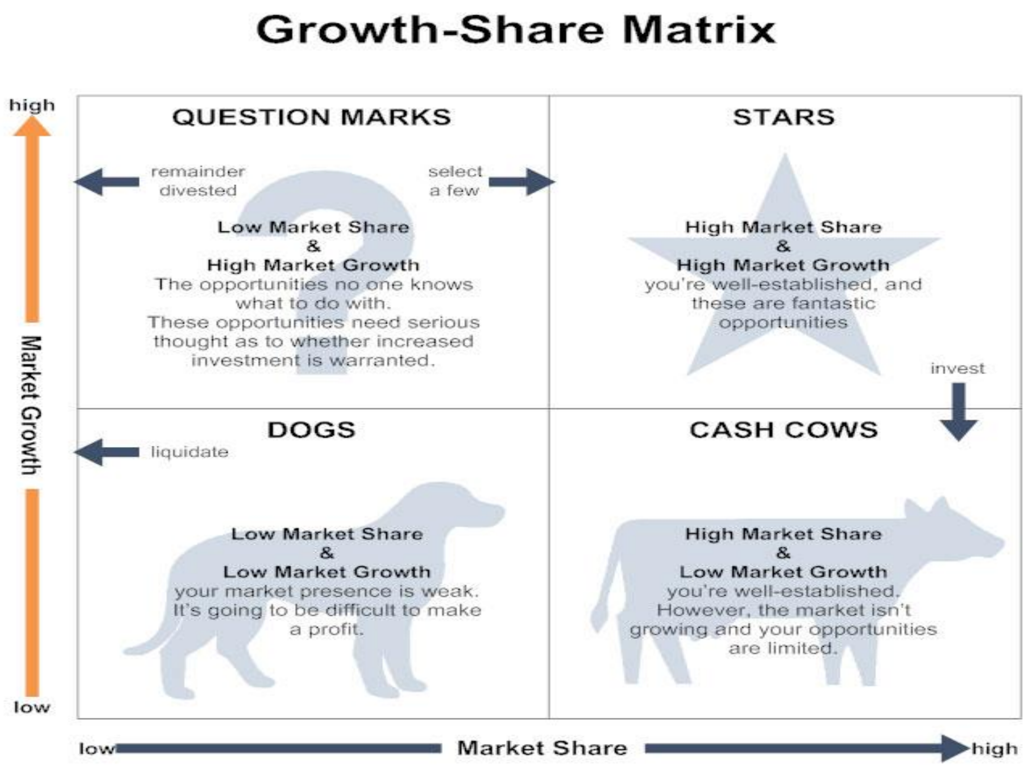


Fig.1 Growth Share Matrix

Now, with rapid technological change, this matrix needs to be seen in a new light.

"We keep speed in mind with each new product we release…. And we continue to work on making it all go even faster…. We're always looking for new places where we can make a difference."

—Google's company-philosophy statement.

Thus, such a portfolio matrix analysis should be carried out more frequently. With speed comes experimentation and with it comes failure. So it is important that firms avoid wasting resources. Companies need to respond to changes in the market place by cashing out the stars, retiring the cows more quickly and maximizing the value of dogs. For the few question marks that get selected for investment, the decision is likely to be made based on rigorous analytics.

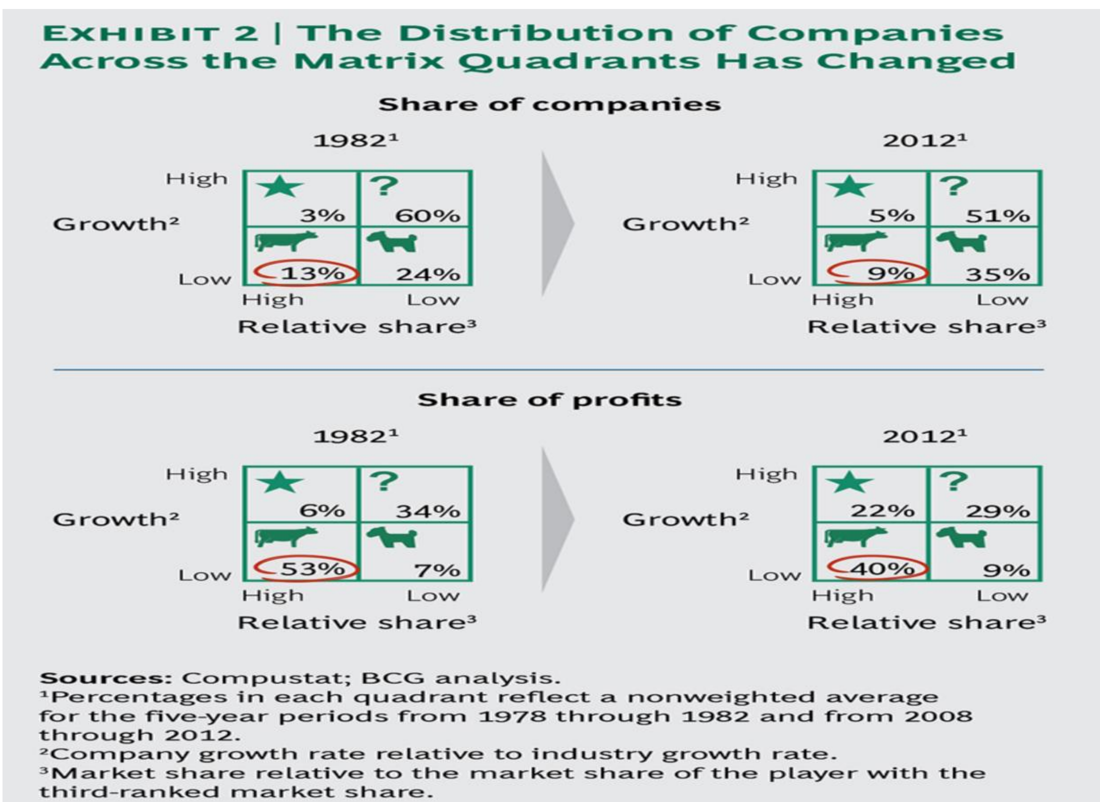


Fig.2 The Changing Times

As can be seen from fig.2, the percentage of stars has gone up and that of the cows has gone down from 1982 - 2012. Traditionally, cash earned from the cows was funneled into the cash burning high growth stars. But now clearly there is less discretion in the manner in which resources can be allocated. Under such conditions, it has become all the more crucial to maintain a balance in the portfolio that the firm holds. This is where our market share modeling exercise carried out frequently will help with analysis that is demanded and considered de rigueur in today’s fast changing environment.

BUSINESS ANALYTICS AND STRATEGY

Project 4

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

HIGH LEVEL TAKE-AWAY

1) Diagnosis: The firm is facing a lack of data maturity to forecast metrics for its innovative initiatives.
2) Forecast: We should be able to track the return on investments (ROIs) made on our innovative initiatives.
3) Add value through search: The firm has access to ROI numbers of its innovative initiatives. No or limited access to other data on decision variables like product features, promotion, pricing and distribution. So we will put a tracking mechanism in place that uses only the ROI metric for a rigourous analysis.
4) Choose: We will use a time series model formulation to track our ROIs.
5) Commit: Allocate resources based on a rigourous understanding of our ROIs.
6) Balanced Scorecard Strategy Map: Having a rigorous investment tracking mechanism in place will see us invest more in innovative initiatives that are beginning to pay well. At the same time we can cut down on other initiatives which aren’t doing that well. We can also appropriately phase our investments in line with their growth cycles. This will result in a more profitable use of our financial resources.

MODEL

Let us consider a standard autoregressive AR(1) process.

$$y_t = \mu + \emptyset_1 y_{t-1} + u_t$$

Where u_t is white noise and uncorrelated with y_{t-1} and which has a variance ${\sigma_u}^2$

Using the lag operator,

$$(1 - \emptyset_1 L) y_t = \mu + u_t$$

Since the series is assumed to be stationary, we have $\emptyset_1 < 1$, So

$$y_t = \left(1 + \emptyset_1 L + {\emptyset_1}^2 L + \cdots\right)(\mu + u_t)$$

Taking expectations,

$$E(y_t) = E\left(1 + \emptyset_1 L + {\emptyset_1}^2 L + \cdots\right)\mu + E\left(1 + \emptyset_1 L + {\emptyset_1}^2 L + \cdots\right)u_t$$

$$E(y_t) = \left(1 + \emptyset_1 + {\emptyset_1}^2 + \cdots\right)\mu$$

$$E(y_t) = \frac{\mu}{(1 - \emptyset_1)}$$

If we remove the intercept term,

$$E(y_t) = 0$$

Thus we will consider a revised AR(1) model,

$$y_t = \emptyset_1 y_{t-1} + u_t \qquad (1)$$

We will multiply both sides by y_{t-1},

$$y_t y_{t-1} = \emptyset_1 {y_{t-1}}^2 + y_{t-1} u_t$$

Taking expectations,

$$E(y_t y_{t-1}) = E(\emptyset_1 {y_{t-1}}^2) + E(y_{t-1} u_t)$$

$$\gamma_1 = \emptyset_1 \gamma_0$$

Where γ_1 is the auto-covariance and γ_0 is the variance

Multiplying (1) by y_{t-2}, we get

$$y_t y_{t-2} = \emptyset_1 y_{t-1} y_{t-2} + y_{t-2} u_t$$

Taking the expectation

$$E(y_t y_{t-2}) = \emptyset_1 E(y_{t-1} y_{t-2}) + E(y_{t-2} u_t)$$

$$\gamma_2 = \emptyset_1 \gamma_1$$

$$\gamma_2 = {\emptyset_1}^2 \gamma_0$$

In general,

$$\gamma_s = {\emptyset_1}^s \gamma_0$$

We can write the autocorrelation coefficients as,

$$\tau_s = \frac{\gamma_s}{\gamma_0}$$

Thus they can be written as

$$\tau_1 = \emptyset_1$$

$$\tau_2 = {\emptyset_1}^2$$

And in general,

$$\tau_s = {\emptyset_1}^s$$

These are the Yule-Walker equations.

Thus for $\emptyset_1 > 0$ the auto correlation function (ACF) decays exponentially towards zero and for $\emptyset_1 < 0$ decays with oscillations. In general for an AR(p) process, the partial autocorrelation function (PACF) spikes through lag p.

Now let us consider an MA(1) process with no intercept.

$$y_t = \theta_1 u_{t-1} + u_t \qquad (2)$$

Squaring both sides and taking expectations,

$$E({y_t}^2) = E({u_t}^2) + {\theta_1}^2 E({u_{t-1}}^2) + E(u_t u_{t-1})$$

$$\sigma^2 = (1 + {\theta_1}^2){\sigma_u}^2$$

Now multiplying (2) by y_{t-1}, and taking expectations

$$E(y_t y_{t-1}) = \theta_1 E(u_{t-1} y_{t-1}) + E(y_{t-1} u_t)$$

$$E(y_t y_{t-1}) = \theta_1 E(u_{t-1}(\theta_1 u_{t-2} + u_{t-1})) + E(y_{t-1} u_t)$$

$$E(y_t y_{t-1}) = \theta_1 E({u_{t-1}}^2)$$

$$\gamma_1 = \theta_1 {\sigma_u}^2$$

Now the auto-correlation coefficient at lag 1 is given by

$$\frac{\gamma_1}{\sigma^2} = \frac{\theta_1}{1 + {\theta_1}^2}$$

The auto correlation coefficients at all the other lags is equal to 0

$$\gamma_2 = E(y_t y_{t-2})$$

$$\gamma_2 = E(\theta_1 u_{t-1} + u_t)(\theta_1 u_{t-2} + u_{t-3})$$

$$\gamma_2 = 0$$

In general for $s > 1$,

$$\gamma_s = 0$$

Whereas the autocorrelation function (ACF) spikes at lag 1, the partial autocorrelation function (PACF) decays exponentially. Now for an MA (q) process, unlike the AR process, the forecast point collapses to zero or the constant term for $s > q$. This limits the utility of an MA process in forecasting.

For an ARMA process, the ACF is identical to that of an AR process after lag q that is it will decay after lag q. Also unlike the AR process, the PACF will not cut off but dampen out. This feature should help us distinguish between a pure AR process and an ARMA process.

The AR autoregressive models can be estimated using the method of ordinary least squares. They contain only the lagged values of the dependent variable and can be used like any other regressors. However, in the context of MA and mixed ARMA models, the error terms are not known and so are usually replaced by residuals but even these aren't known until after the model has been estimated. So we use the method of maximum likelihood. The method computes the coefficients and the residuals at the same time.

Now let us consider a zero mean white noise process. We will also assume that the $y_t s$ are normally distributed. The sample auto-correlation coefficients in this case are also approximately normally distributed (Brooks 2008, p.209).

$$\tau_s \sim approx\ N(0,1/T)$$

where T is the sample size and τ_s denotes the auto correlation coefficient at lag s estimated from a sample.

A 95% non-rejection region would be given by $\pm 1.96 * \frac{1}{\sqrt{T}}$ for s not equal to 0. If the sample autocorrelation coefficient τ_s falls outside this non-rejection region for a given value of s, then the null hypothesis that the true value of the coefficient at that lag s is zero is rejected.

We can test the joint hypothesis that all m of the τ_k correlation coefficients are simultaneously equal to zero using the Q-statistic developed by Box and Pierce (1970). This is the Box-Pierce statistic.

$$Q = T \sum_{k=1}^{m} \tau_k^2$$

Where T is the sample size and m is the maximum lag length. The correlation coefficients are squared so that the positive and negative coefficients do not cancel each other out. This sum of squares of independent standard normal variable is itself a χ^2 variate with m degrees of freedom.

The Box--Pierce test has poor small sample properties. A variant of the Box--Pierce test, having better small sample properties, has been developed and is known as the Ljung--Box (1978) statistic.

$$Q^k = T(T+2)\sum_{k=1}^{m}\frac{\tau_k^2}{T-k} \sim \chi_m^2$$

As the sample size increases this Ljung-Box statistic approaches the Box-Pierce statistic as the (T-k) in the denominator and the (T+2) in the numerator cancel out.

CANONICAL INPUT FOR THE CODE

Input variable:

1 – A time series measure of the return on investment (ROI) metric for each of the firm’s innovation initiatives.

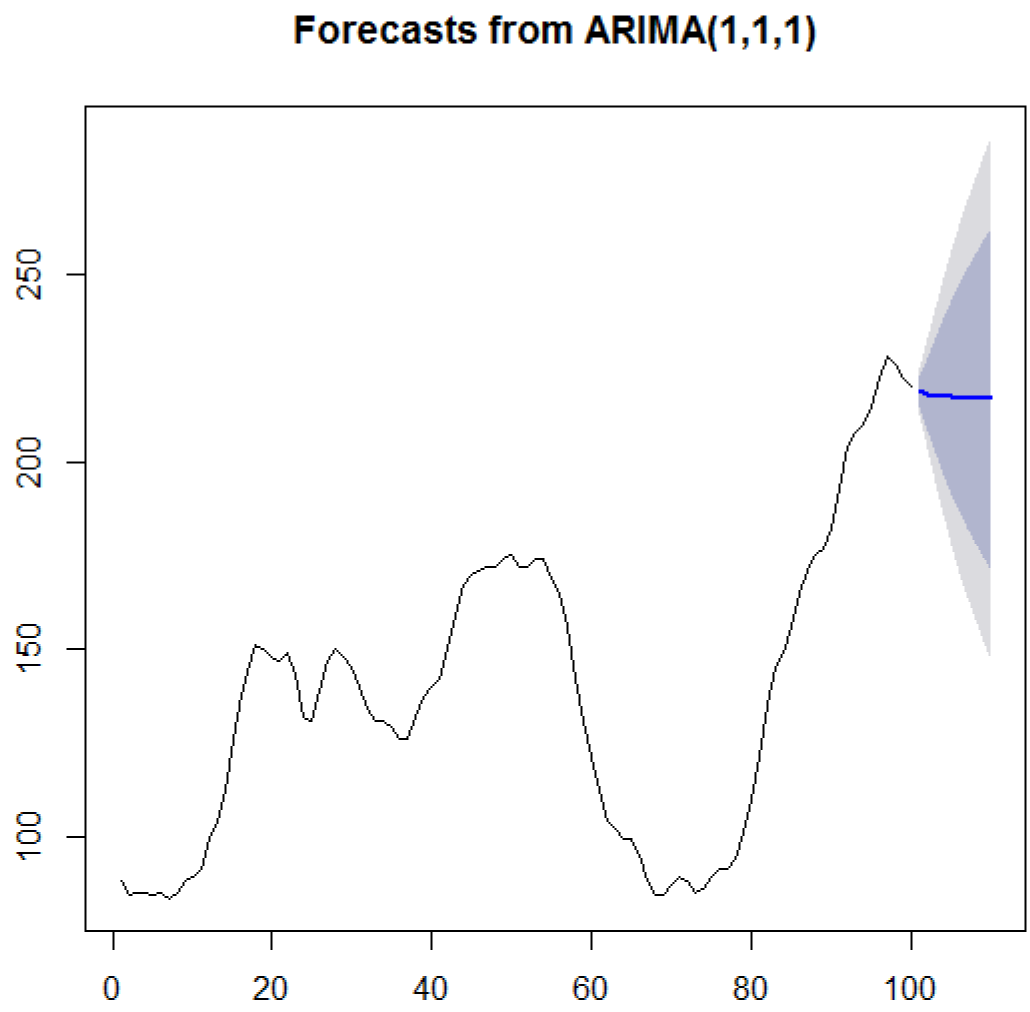
MINIMAL WORKING CODE

```
# app4

# we will load the shiny package that we have installed in the call below

library(shiny)


#we will call the user interface function by passing our arguments

ui <- fluidPage(


 headerPanel('Time Series Analysis'),

#the user will have multiple options to pick from

#We will pick the 'ARIMA' specification to model our time series data

sidebarLayout(


 sidebarPanel(

    selectInput('dist', 'Model', c("AR","MA","ARIMA")),

#this will be a call to update our analysis

 actionButton(inputId = "go",

    label = "Update")

),

mainPanel(

plotOutput("plt")

)
```

```
)
)

server <- function(input, output) {

#we will use the auto.arima function to model our data
#based on the aic criteria this function returns the best arima specification after conducting a detailed search

#we will attach the library forecast here so that we can use the auto.arima function
library(forecast)

#we will attach the pre-loaded input data here
attach(times1)

tseries1 <- eventReactive(input$go,{
if(input$dist=="ARIMA")
auto.arima(times1)
})

#we will forecast data upto 20 time units ahead using the forecast() function
output$plt<-renderPlot(
plot(forecast(tseries1(),h=20))
```

```
)

}

shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE



ANALYSIS

Market intelligence and a survey of one's own strenggth will yield a selection of innovation opportunities. The firm has to accordingly invest in R&D, capability-building and appropriate business development. A typical innovation initiative can be analysed through the lens of the probability of success over a duration of time and the inherent risk element in resourcing such initiatives.

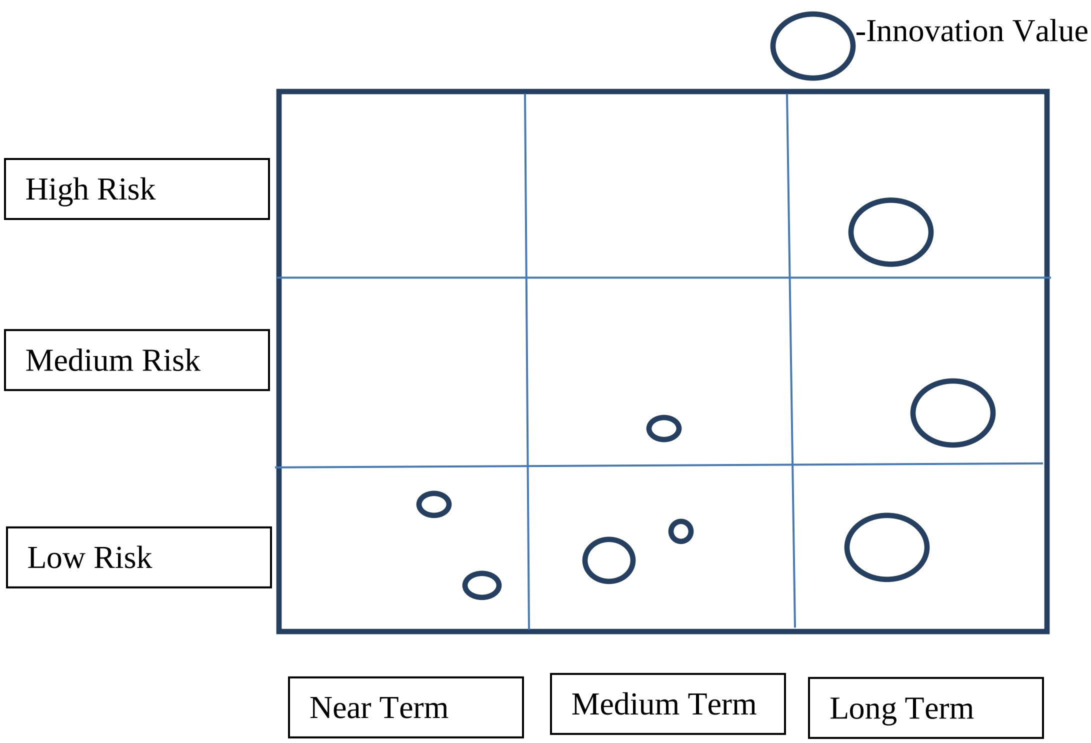


Fig.1 Innovation Initiatives Analysis

Initially due to lack of data maturity, the firm is likely to face problems in putting a forecasting mechanism in place. This is where our time series model will be used. The model will not require additional input decision variables like product features, promotions, pricing and physical distribution but the sole top level metric, the return on investment (ROI), can be tracked. We will track ROI measures for a variety of innovation initiatives that the firm has undertaken. A rigourous exercise to evaluate our choices and allocations can be carried out frequently.

BUSINESS ANALYTICS AND STRATEGY

Project 5

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

HIGH LEVEL TAKE-AWAY

1) Diagnosis: Our test marketing efforts cost us a lot. We have a long product development cycle time.
2) Forecast: We would like an x% decrease in our test marketing costs.
3) Add value through search: Either we renegotiate our commercial terms with our trade partners or bring down our go-to-market cycle time.
4) Choose: We predict customer trial rates instead of letting the test marketing phase run its full course. This will bring down the effective test marketing cycle time.
5) Commit: Commit to go/no go calls based on our predicted trial rates.
6) Balanced Scorecard Strategy Map: Lower test marketing cycle time results in lower new product development costs. This also means a quicker go-to-market which means lesser trade allowances in distribution. Less costs mean increased profitability for our projects.

MODEL

We will use probability models to predict new product trial. In doing so I will use Fader and Hardie (2008).

Many researchers attempt to describe/predict the behavior using observed variables. However they still use random components to model factors that are not included in the model. The authors treat behavior as if it were random/stochastic and propose the model of individual-level behavior that is summed across individuals to obtain a model of aggregate behavior. They, in particular identify 7 steps from problem formulation to its solution.

1) Determine the marketing decision problem/information needed
2) Identify the observable individual-level behavior of interest – denote this by x
3) Select a probability distribution that characterizes this individual-level behavior
    a. This is denoted by $f(x|\theta)$
    b. The parameters of this distribution are individual-level latent characteristics
4) Specify a distribution to characterize the distribution of latent characteristic variables across the population
    a. Denote this by $g(\theta)$
    b. This is often called the mixing distribution
5) Derive the corresponding aggregate or observed distribution for the behavior of interest

$$f(x) = \int f(x|\theta)\ g(\theta)d\theta$$

6) Estimate the parameters of the mixing distribution by fitting the aggregate distribution to the observed data
7) Use the model to solve the marketing decision problem

So, Ace Snackfoods, Inc. has developed a new shelf-stable juice product called Kiwi Bubbles. Before deciding whether or not to "go national" with the new product, the marketing manager for Kiwi Bubbles has decided to commission a year-long test market using IRI's BehaviorScan service, with a view to getting a clearer picture of the product's potential. The product has now been under test for 24 weeks.

On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.) The marketing manager for Kiwi Bubbles would like a forecast of the product's year-end performance in the test market. First, she wants a forecast of the percentage of households that will have made a trial purchase by week 52.

We will treat the time of trial purchase as a random variable. Let T denote this random variable of interest and t denote a particular realization. We will assume the time to trial is characterized by the exponential distribution with parameter λ. Thus, the probability that the customer has tried by time t is given by,

$$F(t|\lambda) = P(T \leq t|\lambda) = 1 - e^{-\lambda t}$$

We will assume the trial rates are distributed across the population according to the gamma distribution.

$$g(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{Gamma(r)}$$

Where r is the shape parameter and α is the scale parameter. The gamma distribution is a flexible unimodal distribution and is mathematically convenient.

The cumulative distribution of time to trial at the market level is given by,

$$P(T \leq t|r, \alpha) = \int_0^{\infty} P(T < t|\lambda)\, g(\lambda|r, \alpha) d\lambda$$

$$P(T \leq t|r, \alpha) = 1 - (\frac{\alpha}{\alpha + t})^r$$

We call this the exponential-gamma distribution

The log-likelihood function is defined as,

$$LL(r, \alpha|data) = 8 * lnP(0 < T \leq 1) + 6 * lnP(1 < t \leq 2) + \cdots + 4 * lnP(23 < T \leq 24) + (1499 - 101) * lnP(T > 24)$$

The model parameters are estimated using the nlminb() function in the R software. Subsequently, we will predict cumulative trials from the panel upto week 52.

CANONICAL INPUT FOR THE CODE

Week	No. of households that bought the product (Cumulative)	Week	No. of households that bought the product (Cumulative)
1	8	13	68
2	14	14	72
3	16	15	75
4	32	16	81
5	40	17	90
6	47	18	94
7	50	19	96
8	52	20	96
9	57	21	96
10	60	22	97
11	65	23	97
12	67	24	101

Table.1 Kiwi Bubbles Cumulative Trial

MINIMAL WORKING CODE

```
# app5

# we will load the shiny package that we have installed in the call below

library(shiny)


#we will call the user interface function by passing our arguments

ui <- fluidPage(


 headerPanel('New Product Trial'),

#the user will have multiple options to pick from

#We will pick the 'exponential-gamma' specification to model our new product trial data

sidebarLayout(


 sidebarPanel(

    selectInput('dist', 'Model', c("exponential-gamma","pareto","nbd")),

#this will be a call to update our analysis

 actionButton(inputId = "go",

    label = "Update")

),

mainPanel(

verbatimTextOutput("p1")

)
```

```
)
)

server <- function(input, output) {

#we will use the nlminb() function to maximise our log likelihood function

trial <- eventReactive(input$go,{
if(input$dist=="exponenial-gamma")
{

#we will construct the log likelihood here and pass its parameters as arguments
#this is for the 24 week trial data that we have access to
#we will predict the cumulative trials from the panel upto week 52
ll1<-function(x)
{
a<-x[1]
b<-x[2]
return(
-(
8*log((a/(a+0))^b-(a/(a+1))^b)
+
6*log((a/(a+1))^b-(a/(a+2))^b)
```

+

2*log((a/(a+2))^b-(a/(a+3))^b)

+

16*log((a/(a+3))^b-(a/(a+4))^b)

+

8*log((a/(a+4))^b-(a/(a+5))^b)

+

7*log((a/(a+5))^b-(a/(a+6))^b)

+

3*log((a/(a+6))^b-(a/(a+7))^b)

+

2*log((a/(a+7))^b-(a/(a+8))^b)

+

5*log((a/(a+8))^b-(a/(a+9))^b)

+

3*log((a/(a+9))^b-(a/(a+10))^b)

+

5*log((a/(a+10))^b-(a/(a+11))^b)

+

2*log((a/(a+11))^b-(a/(a+12))^b)

+

1*log((a/(a+12))^b-(a/(a+13))^b)

+

4*log((a/(a+13))^b-(a/(a+14))^b)

+

3*log((a/(a+14))^b-(a/(a+15))^b)

+

6*log((a/(a+15))^b-(a/(a+16))^b)

+

9*log((a/(a+16))^b-(a/(a+17))^b)

+

4*log((a/(a+17))^b-(a/(a+18))^b)

+

2*log((a/(a+18))^b-(a/(a+19))^b)

+

0*log((a/(a+19))^b-(a/(a+20))^b)

+

0*log((a/(a+20))^b-(a/(a+21))^b)

+

1*log((a/(a+21))^b-(a/(a+22))^b)

+

0*log((a/(a+22))^b-(a/(a+23))^b)

+

4*log((a/(a+23))^b-(a/(a+24))^b)

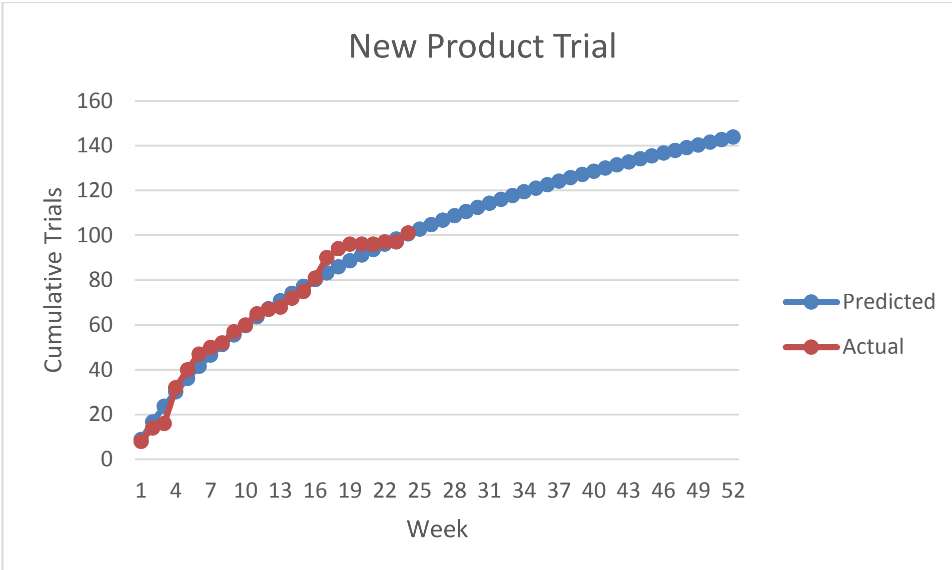
+

1398*log((a/(a+24))^b)

```
)
)
}

}

#we will use bounds on the function call so that invalid parameters aren't selected during the search
#a starting vector of parameters is supplied as seed

nlminb(c(6,2),ll1,lower=c(0.01, 0.01), upper=c(1000,1000))

})

#we will output the maximum likelihood estimates here which will aide in subsequent analysis
output$p1<-renderPrint(
trial()$par
)



}
```

```
shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE



ANALYSIS

Now the direct costs of test marketing includes pilot plant to build the product, commercials, an ad agency, point of sale material produced in small quantities, higher media expenses due to low volumes, couponing, sampling and higher trade allowances to gain distribution. So an appropriate selection criteria for a product as a go/no go is crucial since what usually follows is a nation level roll-out. This criteria is decided by rule-of-thumb. Say a less than 20% trial rate during test marketing means the product needs to be shelved. One of the key objectives of such a criteria is to decide whether the product will meet its minimum financial target and be classified a success. Test marketing can tell you if you have a loser, a medium success, or a giant success. Our modeling exercise will help us predict the value of this metric for evaluation. This will drastically reduce the costs associated with the pilot phase.

A key objective of test marketing is to decide how to market. So for a totally unfamiliar product the firm may not be sure how the sales force would react to innovation. There are estimates though that say introducing your sales force to the product during the test phase adds 15-17% to the eventual national level roll-out. This will let you know the kind of distribution your sales force is capable of getting. Also, being first with a new product is extremely important in, for instance say, the grocery-product category. When Gillette successfully introduced “The Dry Look,” a hair spray for men, several other companies tried to cash in on the success by introducing hair sprays of their own. Since hair sprays came in much larger containers than most hair preparations for men, more shelf space was required to stock the same number of competing brands. And because retailers tend to devote a fixed amount of shelf space to a particular product category, many of the companies with me-too products had difficulty getting trade acceptance for copied items. Our modeling exercise will help us reduce our cycle time and bring the product to market quicker.

OPTIMISATION AND STRATEGY

Project 6

Prashant Prakash Deshpande

Contact: idprashantd@gmail.com

Date: 15th April, 2019

HIGH LEVEL TAKE-AWAY

1) Diagnosis: Our portfolio mix has been generating output below our targeted benchmark return and overshooting our risk appetite.
2) Forecast: We need to re-allocate our investments so that they meet the set benchmark and minimize risk.
3) Add value through search: This is a problem in nonlinear programming and we need to set up a formulation to solve such a system of portfolio allocation.
4) Choose: We will specifically use the interior point method that leverages a modified Newton's method and a search direction that is aligned with the critical path.
5) Commit: Allocate funds to the portfolio according to a set of rules arrived at by applying principles from mathematical programming.
6) Balanced Scorecard Strategy Map: We will be able to better utilize our financial resources through the application of an optimization formulation. This will result in higher growth, better profitability and lower risk.

MODEL

We will consider a barrier formulation of a primal linear programming problem.

Min

$$c^T x - \mu \sum_{i=1}^{n} \ln(x_i)$$

Subject To

$$Ax = b$$

$$x \geq 0$$

If we consider the Lagrangian we can deduce the following system:

$$\nabla_x L(x, y) = c - \mu X^{-1} e - A^T y = 0$$

$$\nabla_y L(x, y) = b - Ax = 0$$

Where we let X denote the square matrix whose ith diagonal entry is just the ith entry of x. This is where we can introduce a slack vector $s = \mu X^{-1} e$ where e is an $n * 1$ vector of 1's and reformulate in the following form:

$$A^T y + s = c$$

$$Ax = b, x \geq 0$$

$$Xs = \mu e, s \geq 0$$

For a quadratic programming problem we can carry out a similar exercise and arrive at the following optimality conditions:

$$F(x, y, s) = \begin{bmatrix} A^T y - Qx + s - c \\ Ax - b \\ XSe \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \; ; \; x, s \geq 0$$

Where S is a diagonal matrix like X.

This system has $n + m + n$ variables and exactly the same number of constraints i.e. it is a square system. Because of the nonlinear equations $x_i s_i = 0$ we cannot solve this system using linear system solutions methods but we can apply the modified Newton's method. Interior-point approaches use the following strategy to handle the inequality constraints: One first identifies an initial solution $(\mathrm{x0}, \mathrm{y0}, \mathrm{s0})$. that satisfies the first two (linear) blocks of equations in $\mathrm{F}(\mathrm{x}, \mathrm{y}, \mathrm{s}) = 0$ but not necessarily the third block $\mathrm{XSe} = 0$, and also satisfies the non-negativity constraints strictly, i.e., $\mathrm{x0} > 0$ and $\mathrm{s0} > 0$: Notice that a point satisfying some inequality constraints strictly lies in the interior of the region defined by these inequalities{rather than being on the boundary. This is the reason why the method we are discussing is called an interior-point method.

Assume that we have a current estimate (x^k, y^k, s^k) of the optimal solution to the problem. The Newton step from this point is determined by solving the following system of linear equations:

$$J(x^k, y^k, s^k) \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = -F(x^k, y^k, s^k)$$

Where $J(x^k, y^k, s^k)$ is the jacobian such that

$$J(x^k, y^k, s^k) = \begin{bmatrix} -Q & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix}$$

Thus the Newton equation reduces to:

$$\begin{bmatrix} -Q & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -X^k S^k e \end{bmatrix}$$

The central path is characterized by a scalar $\tau > 0$ and the points on the central path are obtained as solutions of the following system:

$$F(x_\tau, y_\tau, s_\tau) = \begin{bmatrix} 0 \\ 0 \\ \tau e \end{bmatrix}, x_\tau, s_\tau \geq 0$$

Where $(x_\tau)_i (s_\tau)_i = \tau$

Thus a centered direction is obtained by applying the Newton's method to the following system:

$$F(x_\tau, y_\tau, s_\tau) = \begin{bmatrix} 0 \\ 0 \\ XSe - \tau e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The modified Newton equation can be written as

$$\begin{bmatrix} -Q & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau e - X^k S^k e \end{bmatrix}$$

Which can be further written as:

$$\begin{bmatrix} -Q & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma^k \mu^k e - X^k S^k e \end{bmatrix}$$

Where,

$$\mu^k = \mu(x^k, s^k) = \frac{(x^k)^T s^k}{n}$$

And

$0 \leq \sigma^k \leq 1$ which is a user defined quantity.

CANONICAL INPUT FOR THE CODE

1. The returns on the instruments that make up our portfolio

2. The variance-covariance matrix for the return on investments

3. Constraint conditions

MINIMAL WORKING CODE

```
# app6

# we will load the shiny package that we have installed in the call below

library(shiny)


#we will call the user interface function by passing our arguments

ui <- fluidPage(


 headerPanel('Portfolio Allocation'),

#the user will have multiple options to pick from: Quadratic Programming (QP), Linear
Programming (LP), Integer Programming (IP)

#We will pick the 'QP' specification to model our system

sidebarLayout(


 sidebarPanel(

   selectInput('form', 'Model', c("QP","LP","IP")),

#this will be a call to update our analysis

 actionButton(inputId = "go",

   label = "Update")

),

mainPanel(

verbatimTextOutput("p1")

)
```

```
)
)

server <- function(input, output) {



#we will attach the library 'quadprog' here so that we can use the solve.QP function
library(quadprog)



qp1 <- eventReactive(input$go,{
if(input$form=="QP")
{
#we will consider a portfolio consisting of 3 instruments for simplicity
#the working though can be generalised to accommodate a bigger portfolio size

#we will load the values for the returns made on the instruments that comprise our
portfolio
r1<-a
r2<-b
r3<-c
#we will also load the target portfolio return and the variance covariance matrix
```

```
#we will assume the covariances to be zero

r<-d

var1<-e

var2<-f

var3<-g


Dmat<- matrix(0,3,3)

diag(Dmat)<- c(2*var1,2*var2,2*var3)

dvec<- c(0,0,0)

Amat<- matrix(c(1,r1,1,0,0,-1,0,0,1,r2,0,1,0,0,-1,0,1,r3,0,0,1,0,0,-1),8,3)

bvec<- c(1,r,0,0,0,-1,-1,-1)

solve.QP(Dmat,dvec,Amat,bvec=bvec,meq=1)

}

})


output$p1<-renderPrint(qp1()$solution)



}


shinyApp(ui = ui, server = server)
```

CANONICAL OUTPUT OF THE CODE

1. The output of the solve.QP() function in R.
   a. This will yield optimal allocations across our portfolio instruments where the sum of allocated proportions equals 1.

ANALYSIS

Across industries, senior executives know that managing capital investments wisely means better cash flow, faster growth, and competitive advantage. Many organizations, however, struggle to manage spending on hundreds or even thousands of capital projects and miss substantial growth and profitability opportunities as a result. Small capital projects, typically defined as projects valued at less than $50 million though often less than $5 million, can account for up to 50 percent of an organization's total capital-expenditure (capex) spend. The need for such projects is often driven by a complex diversity of factors, from growth to maintenance to health, safety, and environmental compliance. However, when looked at collectively, this large and complex spend category typically receives insufficient management attention or control, causing inefficient capital allocation and poorly prioritized project portfolios.

We will use the Markovitz mean-variance portfolio optimization technique that seeks to balance risk and return of our portfolio. Let $R_i$denote the return on investment for the i^{th} instrument. Let $Var\ (R_i)$ denote the variance or risk of such returns. Let w_i be the proportion of investments allocated to the i^{th} instrument. The proportion of allocations made by the firm sum to 1. Let us assume the goal of the decision maker is to minimise risk and achieve a minimum targeted rate of return on the portfolio. We can formulate this problem in the space of constrained optimization.

$$Min \sum_{i=1}^{n} Var\ (w_i R_i)$$

Subject to,

$$\sum_{i=1}^{n} w_i = 1$$

$$\sum_{i=1}^{n} w_i R_i \geq c$$

We will have access to the variance-covariance matrix and the numbers for returns for each of our investment instruments that make up the portfolio. Using our method of interior point algorithm we will be able to solve such a quadratic system for optimal values of w_i. Thus, portfolio allocation can happen according to a set of rules that meet our investment appetite.

EXTENSION REFERENCES

Blattberg, R.C., Kim, B., Neslin, S.A. 2008. *Database Marketing: Analyzing and Managing Customers*. Springer, New York, NY.

Box, GEP and Pierce, DA 1970, Distributions of Residual Autocorrelations in Autoregressive Integrated Moving Average Models, *Journal of the American Statistical Association* 65, 1509—26

Brooks, C 2008, *Introductory Econometrics for Finance*, 2nd edn, Cambridge University Press, New York.

Cooper, L.G., Nakanishi, M. 1988. *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*. Kluwer Academic Publishers, Norwell, MA.

Fader, P.S., Hardie, B.G.S. 2008. *Applied Probability Models in Marketing Research: Introduction.* 19th Annual Advanced Research Techniques Forum.

https://hbr.org/1976/05/test-marketing-in-new-product-development

Ljung, GM and Box, GEP 1978, On a Measure of Lack of Fit in Time Series Models, *Biometrika* 65(2), 297--303

https://www.mckinsey.com/industries/capital-projects-and-infrastructure/how-we-help-clients/portfolio-optimization

https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/nine-practices-for-better-capital-investment-management

https://www.mckinsey.com/~/media/McKinsey/dotcom/client_service/Strategy/PDFs/The_Eight_Essentials_of_Innovation_Performance.ashx

https://www.vendavo.com/blog/pricing/pricing-waterfalls-101/