Team6_Ex3

0.1 Set-up

```r
# Clear All Variables & Clear the Screen
rm(list=ls())
cat("\014")
```

```r
# Read in the Data
data.adv = read.csv("Ex3_Data_R.csv")

# Explore the data
str(data.adv)

## 'data.frame':    66 obs. of  3 variables:
##  $ Day        : chr  "5/1/2019" "5/2/2019" "5/3/2019" "5/4/2019" ...
##  $ Visitors   : int  530 530 530 511 511 511 514 514 514 545 ...
##  $ Total.Spent: int  365 365 365 435 435 435 303 303 303 188 ...

summary(data.adv)

##      Day                Visitors        Total.Spent
##  Length:66          Min.   : 511.0    Min.   : 188.0
##  Class :character   1st Qu.: 615.0    1st Qu.: 430.0
##  Mode  :character   Median : 918.5    Median : 786.5
##                     Mean   : 862.1    Mean   : 714.1
##                     3rd Qu.:1044.0    3rd Qu.: 884.0
##                     Max.   :1197.0    Max.   :1400.0
```

0.2 Data Preparation

In our case we have to create new variables: log(Total.Spent), Total.Spent^2, and Lag(Total.Spent)

```r
# Create log Total.Spent
data.adv$logTotal.Spent <- log(data.adv$Total.Spent)

# Create Total.Spent^2
data.adv$Total.Spent2 <- (data.adv$Total.Spent)^2

# Create LAG Visitors - V_(t-1)
data.adv$lagVisitors <- c(NA, head(data.adv$Visitors, -1))

# Display the data
summary(data.adv)

##      Day                Visitors        Total.Spent      logTotal.Spent
##  Length:66          Min.   : 511.0    Min.   : 188.0    Min.   :5.236
##  Class :character   1st Qu.: 615.0    1st Qu.: 430.0    1st Qu.:6.064
##  Mode  :character   Median : 918.5    Median : 786.5    Median :6.667
##                     Mean   : 862.1    Mean   : 714.1    Mean   :6.453
##                     3rd Qu.:1044.0    3rd Qu.: 884.0    3rd Qu.:6.784
##                     Max.   :1197.0    Max.   :1400.0    Max.   :7.244
##
##   Total.Spent2      lagVisitors
##  Min.   :  35344   Min.   : 511.0
##  1st Qu.: 184900   1st Qu.: 615.0
##  Median : 618822   Median : 872.0
```

```
## Mean    : 610178    Mean    : 860.5
## 3rd Qu.: 781456    3rd Qu.:1044.0
## Max.    :1960000    Max.    :1197.0
##                      NA's    :1
```

1 Short-run Response. Analyze the short-run response of clicks (i.e., visitors) to advertising (i.e., paid search spending).

1.a Fit the following three models of advertising-click response (at this stage we will not add other information to the model; keep things simple and stick to advertising), report your results and comment, briefly.

1.a.i Simple linear

```
# Run the simple linear regression on Visitors and Total.Spent on paid search
lm.model1 <- lm(Visitors ~ Total.Spent, data = data.adv)
summary(lm.model1)

##
## Call:
## lm(formula = Visitors ~ Total.Spent, data = data.adv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -224.92 -158.68  -32.95  120.84  410.81
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 533.06454   52.36345  10.180  5.2e-15 ***
## Total.Spent   0.46076    0.06703   6.873  3.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 172.4 on 64 degrees of freedom
## Multiple R-squared:  0.4247, Adjusted R-squared:  0.4157
## F-statistic: 47.24 on 1 and 64 DF,  p-value: 3.1e-09
```

1.a.ii Concave logarithmic

```
# Run the Concave logarithmic regression on Visitors and Total.Spent on paid
search
lm.model2 <- lm(Visitors ~ logTotal.Spent, data = data.adv)
summary(lm.model2)
```

```
## 
## Call:
## lm(formula = Visitors ~ logTotal.Spent, data = data.adv)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -237.60 -165.78   -8.76   97.31  396.87
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1075.92     255.38  -4.213 8.05e-05 ***
## logTotal.Spent   300.32      39.45   7.613 1.55e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 164.7 on 64 degrees of freedom
## Multiple R-squared:  0.4752, Adjusted R-squared:  0.467
## F-statistic: 57.95 on 1 and 64 DF,  p-value: 1.554e-10
```

1.a.iii Concave quadratic

```
# Run the concave quadratic regression on Visitors and Total.Spent on paid
search
lm.model3 <- lm(Visitors ~ Total.Spent + Total.Spent2, data = data.adv)
summary(lm.model3)

## 
## Call:
## lm(formula = Visitors ~ Total.Spent + Total.Spent2, data = data.adv)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -226.79 -148.75   23.52   72.67  408.24
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.816e+02  9.675e+01   2.911  0.00498 **
## Total.Spent    1.293e+00  2.827e-01   4.574 2.29e-05 ***
## Total.Spent2  -5.621e-04  1.861e-04  -3.021  0.00364 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 162.4 on 63 degrees of freedom
## Multiple R-squared:  0.4975, Adjusted R-squared:  0.4815
## F-statistic: 31.18 on 2 and 63 DF,  p-value: 3.857e-10
```

In all three models, 'Total.Spent' is significant. The 'logTotal.Spent' in model2 is significant. The 'Total.Spent2' in model3 is also significant.

The R-squared is 0.42 for simple linear (model1), 0.48 for concave log (model2), and 0.50 for concave quadratic (model3). Model3 has the highest R-squared.

1.b Compute the advertising elasticity implied by each model (use July's monthly clicks and advertising spent to scale the elasticities)

```r
# Create a boolean list that returns TRUE for data in July
is_july <- substring(data.adv$Day,1,1)==7

# Calculate mean visitors in July
mean_visitors <- mean(data.adv$Visitors[is_july])
# Calculate mean ads spent amount in July
mean_spent <- mean(data.adv$Total.Spent[is_july])

# Compute the advertising elasticity for the simple linear model
model1.b1 <- unname(lm.model1$coefficients[2])
adv.elas.1 <- model1.b1*(mean_spent/mean_visitors)
print(paste("advertising elasticity for the simple linear model is",
adv.elas.1))

## [1] "advertising elasticity for the simple linear model is
0.264177941565051"

# Compute the advertising elasticity for the concave log model
model2.b1 <- unname(lm.model2$coefficients[2])
adv.elas.2 <- model2.b1*(1/mean_visitors)
print(paste("advertising elasticity for the concave log model
is",adv.elas.2))

## [1] "advertising elasticity for the concave log model is
0.301340745109823"

# Compute the advertising elasticity for the concave quadratic model
model3.b1 <- unname(lm.model3$coefficients[2])
model3.b2 <- unname(lm.model3$coefficients[3])
adv.elas.3 <- (model3.b1 + 2*model3.b2*mean_spent)*(mean_spent/mean_visitors)
print(paste("advertising elasticity for the concave quadratic model
is",adv.elas.3))

## [1] "advertising elasticity for the concave quadratic model is
0.373144831506776"
```

1.c Which model is best? Which, if any, would you reject?

The concave quadratic model (model3) seems to be the best. It has the highest R-squared of 0.5. Both the Total.Spent and the square Total.Spent are significant. Its quadratic nature also assumes diminishing marginal returns to advertising spending, which accounts for saturation.

We'd like to reject the simple linear model (model1). It has the lowest R-squared. Furthermore, it assumes constant marginal returns, which can be problematic during extrapolation.

2 Long-run Response. Analyze the long-run response of clicks to advertising using the same data.

2 Fit the same three models of advertising response but incorporate an exponentially decaying lag effect for advertising. Report your results and briefly comment.

2.a.i Simple linear with lag effect

```
# Run the simple linear regression with exponentially decaying lag effect for
Total.Spent on paid search
lm.model4 <- lm(Visitors ~ Total.Spent + lagVisitors, data = data.adv)
summary(lm.model4)

##
## Call:
## lm(formula = Visitors ~ Total.Spent + lagVisitors, data = data.adv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.371 -15.012  -6.624   7.324 159.681
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.50704   21.63573   1.780   0.0800 .
## Total.Spent  0.04992    0.02190   2.279   0.0261 *
## lagVisitors  0.92129    0.03074  29.973   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.16 on 62 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9622, Adjusted R-squared:  0.9609
## F-statistic: 788.5 on 2 and 62 DF,  p-value: < 2.2e-16
```

2.a.ii Concave logarithmic with lag effect

```
# Run the Concave logarithmic regression with exponentially decaying lag
effect for Total.Spent on paid search
lm.model5 <- lm(Visitors ~ logTotal.Spent + lagVisitors, data = data.adv)
summary(lm.model5)

##
## Call:
## lm(formula = Visitors ~ logTotal.Spent + lagVisitors, data = data.adv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -98.087 -17.810  -6.218   4.808 162.816
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -120.65169   76.61903  -1.575   0.1204
## logTotal.Spent  30.56469   14.19507   2.153   0.0352 *
## lagVisitors      0.91847    0.03234  28.397   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.34 on 62 degrees of freedom
##    (1 observation deleted due to missingness)
## Multiple R-squared:  0.9619, Adjusted R-squared:  0.9606
## F-statistic: 781.7 on 2 and 62 DF,  p-value: < 2.2e-16
```

2.a.iii Concave quadratic with lag effect

```
# Run the concave quadratic regression with exponentially decaying lag effect
for Total.Spent on paid search
lm.model6 <- lm(Visitors ~ Total.Spent + Total.Spent2 + lagVisitors, data =
data.adv)
summary(lm.model6)

##
## Call:
## lm(formula = Visitors ~ Total.Spent + Total.Spent2 + lagVisitors,
##     data = data.adv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.582 -15.634  -6.140   6.599 159.999
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.541e+01  2.831e+01   1.251    0.216
## Total.Spent   6.474e-02  8.911e-02   0.727    0.470
## Total.Spent2 -9.391e-06  5.470e-05  -0.172    0.864
## lagVisitors   9.192e-01  3.320e-02  27.686   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.51 on 61 degrees of freedom
##    (1 observation deleted due to missingness)
## Multiple R-squared:  0.9622, Adjusted R-squared:  0.9603
## F-statistic: 517.4 on 3 and 61 DF,  p-value: < 2.2e-16
```

All three models have an R-squared of around 0.96.

For model4 (simple linear model with lag effect), both the Total.Spent and the lagVisitors are significant.

For model5 (concave logarithmic model with lag effect), both the logTotal.Spent and the lagVisitors are significant.

For model6, (concave quadratic model with lag effect), only the lagVisitors is significant.

2.b Compute the long-run advertising elasticity implied by each model (again use July's monthly figures for scaling).

```
# Examine the coefficients
lm.model4$coefficients

## (Intercept) Total.Spent lagVisitors
## 38.50703925  0.04991942  0.92129085

lm.model5$coefficients

##     (Intercept) logTotal.Spent    lagVisitors
##   -120.6516933     30.5646919      0.9184698

lm.model6$coefficients

##    (Intercept)    Total.Spent   Total.Spent2    lagVisitors
##   3.540698e+01   6.473992e-02  -9.390582e-06   9.192410e-01

# Compute the long-run advertising elasticity for the simple linear model
with lag effect
model4.b1 <- unname(lm.model4$coefficients[2])
model4.lambda <- unname(lm.model4$coefficients[3])
adv.elas.4 <- model4.b1/(1-model4.lambda)*(mean_spent/mean_visitors)
print(paste("long-run advertising elasticity for the simple linear model with
lag effect is", adv.elas.4))

## [1] "long-run advertising elasticity for the simple linear model with lag
effect is 0.363633278634914"

# Compute the long-run advertising elasticity for the concave log model with
lag effect
model5.b1 <- unname(lm.model5$coefficients[2])
model5.lamdba <-unname(lm.model5$coefficients[3])
adv.elas.5 <- model5.b1/(1-model5.lamdba)*(1/mean_visitors)
print(paste("long-run advertising elasticity for the concave log model with
lag effect is",adv.elas.5))

## [1] "long-run advertising elasticity for the concave log model with lag
effect is 0.376166812213122"

# Compute the long-run advertising elasticity for the concave quadratic model
with lag effect
model6.b1 <-unname(lm.model6$coefficients[2])
model6.b2 <- unname(lm.model6$coefficients[3])
model6.lambda <- unname(lm.model6$coefficients[4])
adv.elas.6 <- (model6.b1+2*model6.b2*mean_spent)/(1-
model6.lambda)*(mean_spent/mean_visitors)
print(paste("long-run advertising elasticity for the concave quadratic model
with lag effect is",adv.elas.6))
```

```
## [1] "long-run advertising elasticity for the concave quadratic model with
lag effect is 0.383433021857056"
```

2.c Which model is best? Which, if any, would you reject?

Model5 (Concave logarithmic with lag effect) might be the best model. Given a similar R-squared of 0.96, it incorporates the effect of both current spending on paid search and exponentially decaying lag effect for spending on paid search, while model6 only incorporates the lag effect and leaves out the present effect. The concave shape of this model accounts for the diminishing effect on visitor numbers by increased ads spending.

We want to reject the linear model again because its assumption of constant marginal returns and the problem when extrapolation. We would also want to reject the Concave quadratic mode because it does not take into account the effect of present advertisement spend as significant.

3 Saturation. Drawing on the models that you fitted above, compute the saturation level for advertising spending. Report results in terms of daily advertising spending.

```
# Calculate the saturation level for advertising spending in model3 (concave-
quadratic)
adv.saturation.3 <--model3.b1/(2*model3.b2)
adv.saturation.3
```

```
## [1] 1150.322
```

```
# Calculate the saturation level for advertising spending in model6 (concave-
quadratic with lag effects)
adv.saturation.6 <--model6.b1/(2*model6.b2)
adv.saturation.6
```

```
## [1] 3447.066
```

Model1,2,3 and 5 don't have a saturation level.

The saturation level for model 3 is 1150. This indicates the level of daily ads spending until it has negative effect on number of visitors.

Combining the spillover effect of previous advertisement spends, the saturation level as per model 6 for daily advertising spending will be around $3447 This is the maximum point after which advertisement negatively affects number of visitors.