Homework-2

```
##Load Libraries
library(readxl)
library(Hmisc)
library(MASS)
library(dplyr)
library(ggplot2)
library(skimr)
```

Evaluating the Focal Model

In the first part, we are evaluating the validity of the focal model as described in the homework.

The focal model is given as :

$$Y_t = \lambda Y_{t-1} + \beta_1 Z_{1t} + \beta_2 Z_{2t} + \ldots + intercept + \epsilon_t$$

The focal model will include :

- Lagged Sales
- Intercept
- Square rooted Variables (to model diminishing returns). The choice of variables will depend on further analysis, as shown below.

```
# Import Data
multdata <- read_excel("HW2_MultimediaHW.xlsx")
```

```
#Plotting a correlation plot between all the variables in the dataset
library(psych)
```

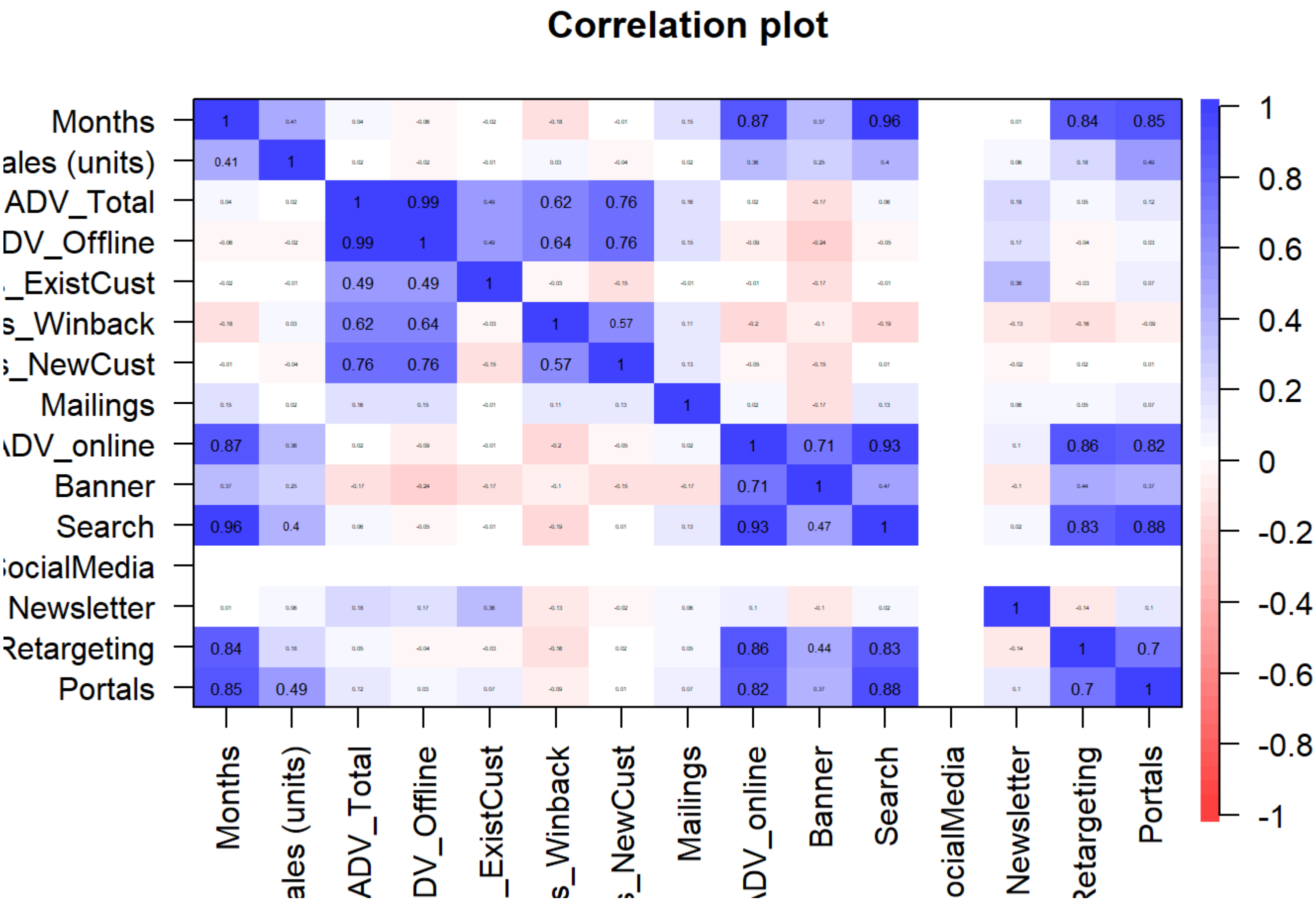
```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:Hmisc':
##
##     describe
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
corPlot(multdata, cex = 0.36,xlas=2,ylas=1,MAR=2)
```

```
## Warning in cor(x, use = use, method = method): the standard deviation is zero
```



From the above plot, we see that in the online ads, search and portals are highly correlated with a value of 0.88

```
# Extract Vectors of Dependent and Independent Variables

Sales <-multdata$`Sales (units)`
Stm1<-Lag(Sales,shift=1) #creating the lag of sales value
Stm1 <- Stm1[-1]
Sales <- Sales[-1]

Catlg.Exist <-multdata$`Catalogs_ExistCust`
Catlg.Winback <-multdata$`Catalogs_Winback`
Catlg.NewCust <-multdata$`Catalogs_NewCust`
Mailings <-multdata$`Mailings`
Banner <-multdata$`Banner`
Search <-multdata$`Search`
Newsl <-multdata$`Newsletter`
Retarg <-multdata$`Retargeting`
Portal <-multdata$`Portals`
```

```
#Diminishing Returns
SCatlg.Exist <- sqrt(Catlg.Exist)[-1]
SCatlg.Winback <-sqrt(Catlg.Winback)[-1]
SCatlg.NewCust <-sqrt(Catlg.NewCust)[-1]
SMailings <-sqrt(Mailings)[-1]
SBanner <- sqrt(Banner)[-1]
SSearch <- sqrt(Search)[-1]
SNewsl <- sqrt(Newsl)[-1]
SRetarg <- sqrt(Retarg)[-1]
SPortal <- sqrt(Portal)[-1]
```

Final Model: Running a focal model with intercept

```
focal_model <- lm(Sales ~ Stm1 +
                    SCatlg.Exist +
                  SCatlg.Winback +
                  SCatlg.NewCust +
                  SMailings +
                  SNewsl +
                  SPortal)
summary(focal_model)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SMailings + SNewsl + SPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1154.79  -422.49    79.34   366.83  1757.57
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2122.2149  1100.0980   1.929   0.0624 .
## Stm1              0.1433     0.1946   0.736   0.4667
## SCatlg.Exist    -24.4168    16.5790  -1.473   0.1503
## SCatlg.Winback   53.7332    25.2653   2.127   0.0410 *
## SCatlg.NewCust  -26.7490    14.3242  -1.867   0.0708 .
## SMailings        -9.6636    42.8265  -0.226   0.8229
## SNewsl          168.6910   131.9458   1.278   0.2100
## SPortal         819.1793   300.6776   2.724   0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 685.8 on 33 degrees of freedom
## Multiple R-squared:  0.3874, Adjusted R-squared:  0.2574
## F-statistic: 2.981 on 7 and 33 DF,  p-value: 0.01545
```

```
AIC(focal_model)
```

```
## [1] 660.9629
```

```
# Elasticity ---

coeff.focal <- focal_model$coefficients
eta.focal <- c()

for (i in 3:(length(coeff.focal))) {

  xbar <- mean(get(names(coeff.focal)[i]))
  print(coeff.focal[i])
  placeholder <- (coeff.focal[i]*sqrt(xbar))/(2*(coeff.focal[1] + coeff.focal[i]*sqrt(xbar)))
  eta.focal <- append(eta.focal, placeholder)

}
```

```
## SCatlg.Exist
##    -24.41677
## SCatlg.Winback
##       53.73318
## SCatlg.NewCust
##      -26.74895
## SMailings
## -9.663627
##  SNewsl
## 168.691
##  SPortal
## 819.1793
```

```
eta.focal
```

```
##   SCatlg.Exist SCatlg.Winback SCatlg.NewCust      SMailings         SNewsl
##   -0.028858530    0.028208513   -0.022139207   -0.003327277    0.071789914
##        SPortal
##    0.183654691
```

Variables such as Banner, Social Media have been neglected because they consist of more than 70% null values.

We have also observed that ‘Search’ is highly correlated with ‘Retargetting’ with a value of 0.83 and highly correlated with ‘Portals’ with a value of 0.88. ‘Retargetting’ and ‘Portals’ are inter-correlated with a value of 0.7.

We are neglecting the Search and Retargetting in our final focal model since this model provides us with a better AIC value.

Model Iteration 1 : Full Model (All variables) with Intercept

```
#Running Full Model with intercept
model.1 <- lm(Sales~
                Stm1 +
                SCatlg.Exist +
                SCatlg.Winback +
                SCatlg.NewCust +
                SMailings +
                SBanner +
                SSearch +
                SNewsl +
                SRetarg +
                SPortal
              )

summary(model.1)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SMailings + SBanner + SSearch + SNewsl + SRetarg + SPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1166.64  -481.27    45.89   368.91  1688.63
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1560.11816 1305.48737   1.195   0.2414
## Stm1              0.07296    0.21486   0.340   0.7365
## SCatlg.Exist    -25.42388   17.45955  -1.456   0.1557
## SCatlg.Winback   49.71340   28.51097   1.744   0.0915 .
## SCatlg.NewCust  -25.29200   15.65407  -1.616   0.1166
## SMailings       -17.72565   46.26717  -0.383   0.7043
## SBanner         -13.94797   62.94786  -0.222   0.8261
## SSearch         147.32006  222.56314   0.662   0.5131
## SNewsl          125.69160  145.20000   0.866   0.3936
## SRetarg         -86.53916   95.53703  -0.906   0.3723
## SPortal         866.98152  631.84484   1.372   0.1802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 707.9 on 30 degrees of freedom
## Multiple R-squared:  0.4065, Adjusted R-squared:  0.2087
## F-statistic: 2.055 on 10 and 30 DF,  p-value: 0.06231
```

```
AIC(model.1)
```

```
## [1] 665.6599
```

```
BIC(model.1)
```

```
## [1] 686.2228
```

As we can see above Adj R-squared is 0.2087, which is pretty low. Most importantly, F-statistic p-value = 0.06231, which means the overall model is insignificant at 5% significance level.

Checking to see if removal of the intercept makes any difference:

Model Iteration 2 : Full Model (All variables) without Intercept

```
# Running Full Model Without Intercept
model.2 <- lm(Sales~
                0 +
                Stm1 +
                SCatlg.Exist +
                SCatlg.Winback +
                SCatlg.NewCust +
                SMailings +
                SBanner +
                SSearch +
                SNewsl +
                SRetarg +
                SPortal
)

summary(model.2)
```

```
##
## Call:
## lm(formula = Sales ~ 0 + Stm1 + SCatlg.Exist + SCatlg.Winback +
##     SCatlg.NewCust + SMailings + SBanner + SSearch + SNewsl +
##     SRetarg + SPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1129.31  -468.21    74.31   362.50  1532.86
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## Stm1              0.1868     0.1939   0.964    0.343
## SCatlg.Exist    -17.4265    16.2371  -1.073    0.291
## SCatlg.Winback   61.6088    26.9008   2.290    0.029 *
## SCatlg.NewCust  -29.5819    15.3418  -1.928    0.063 .
## SMailings       -13.9054    46.4742  -0.299    0.767
## SBanner         -20.4477    63.1440  -0.324    0.748
## SSearch         270.3785   198.6666   1.361    0.183
## SNewsl          148.9346   144.8817   1.028    0.312
## SRetarg        -131.4936    88.4238  -1.487    0.147
## SPortal         766.1708   630.4975   1.215    0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 712.8 on 31 degrees of freedom
## Multiple R-squared:  0.984,  Adjusted R-squared:  0.9789
## F-statistic: 191.1 on 10 and 31 DF,  p-value: < 2.2e-16
```

```
AIC(model.2)
```

```
## [1] 665.5667
```

```
BIC(model.2)
```

```
## [1] 684.416
```

Adjusted R-squared shoots up to **0.9789**!

Multiple R-squared of 0.984 says that 98.4% of the variation in Sales is already explained by the model.

But this is misleading, since the removal of intercept or constant term often increases the R-Sq of the model as sum of squares regression may increase relatively more than Sum of squares residuals by forcing the regression line to go through origin [1], and may also bias the other coefficients. We are not reasonably certain in this case, that zero advertising should lead to zero sales, therefore there is no justification to remove the intercept.

For example, we can see that if we only consider the lagged Sales variable and remove the intercept (model `t1` below), the R-squared still remains exceedingly high and seems to say that 96% of the variation in Sales is explained by lagged Sales itself, which is not true.

```
t1 <- lm(formula = Sales ~ Stm1 - 1)
summary(t1)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2681.2  -398.8   106.5   742.6  2275.4
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## Stm1  0.98677    0.03043   32.42   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 950.9 on 40 degrees of freedom
## Multiple R-squared:  0.9633, Adjusted R-squared:  0.9624
## F-statistic:  1051 on 1 and 40 DF,  p-value: < 2.2e-16
```

```
AIC(t1)
```

```
## [1] 681.6485
```

Therefore, we will discard this model (***model.2***) and see if we can improve on ***model.1***

Model Iteration 3 : Reduced Model (Using Stepwise Selection via least AIC method) with Intercept

Let us use Stepwise Selection Method on the first Full Model `model.1` to make our model more parsimonious and also retain only the variables that contribute to independently explain the variation in sales. We are using the least AIC for the stepwise selection criteria.

To this end, we will use the `stepAIC()` function in the `MASS` R package, that iteratively handles removal of independent variables in a stepwise fashion and applies the least AIC rule to select the most parsimonious model.

```
model.3 <- stepAIC(model.1, k=2, trace=FALSE)
summary(model.3)
```

```
##
## Call:
## lm(formula = Sales ~ SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SRetarg + SPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1387.09  -446.67    23.65   357.29  1761.02
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2490.95     848.17   2.937  0.00583 **
## SCatlg.Exist     -23.94      13.25  -1.807  0.07937 .
## SCatlg.Winback    34.83      22.32   1.560  0.12768
## SCatlg.NewCust   -19.38      13.29  -1.458  0.15373
## SRetarg          -85.39      63.23  -1.351  0.18551
## SPortal         1367.01     388.90   3.515  0.00124 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 672.9 on 35 degrees of freedom
## Multiple R-squared:  0.3744, Adjusted R-squared:  0.2851
## F-statistic:  4.19 on 5 and 35 DF,  p-value: 0.004345
```

```
AIC(model.3)
```

```
## [1] 657.8205
```

As we can see above, the overall model is statistically significant now (F Statistic p-value=0.004345, which is less than 0.05). The Adj R-squared is higher than `model.1` .

The above can be considered the final Focal Model. However, below is a major the concern regarding the focal model -

The coefficients of `SCatlg.Exist` (-19.38), `SCatlg.NewCust` (-19.38), `SRetarg` (-85.39) indicate that increase in advertising catalogs on Existing or New Customers as well as Retargeting customers is associated with a decrease in Sales, which is unreasonable (Can also be seen in the trends plot provided separately that existing customer catalog spending has similar trends to Sales).

Reasoning

The negative trend of current advertising spend to Sales is due to the fact that the advertising done in previous months have typically increased the Sales in current month. However current month's advertising spend might have been reduced exactly for that reason, to take advantage of 'market memory'.

If we check our line plot of the variation of Existing Customer Catalog Spending and Variation in Sales, we notice that, following trends in general marketing prices, the advertising was increased when sales came down, in the following month the Sales went up, consequently there was no need to spend as much on that channel, therefore advertising spend was lowered.

Based on the above reasoning, we could try a regression model based on lagged variables. In our extended analysis, we have used 1 month lagged data as the regressors on current month's sales.

$$Y_t = \lambda Y_{t-1} + \beta_1 Z_{1(t-1)} + \beta_2 Z_{2(t-1)} + \ldots + intercept + \epsilon_t$$

```
# Create Lagged Independent Variables
LCatlg.Exist <-Lag(Catlg.Exist,shift=1)
LCatlg.Winback <-Lag(Catlg.Winback,shift=1)
LCatlg.NewCust <-Lag(Catlg.NewCust,shift=1)
LMailings <-Lag(Mailings,shift=1)
LBanner <- Lag(Banner,shift=1)
LSearch <-Lag(Search,shift=1)
LNewsl <- Lag(Newsl,shift=1)
LRetarg <- Lag(Retarg,shift=1)
LPortal <- Lag(Portal,shift=1)
```

```
# Model Diminishing Returns on Lagged variables
SLCatlg.Exist <- sqrt(LCatlg.Exist)[-1]
SLCatlg.Winback <-sqrt(LCatlg.Winback)[-1]
SLCatlg.NewCust <-sqrt(LCatlg.NewCust)[-1]
SLMailings <-sqrt(LMailings)[-1]
SLBanner <- sqrt(LBanner)[-1]
SLSearch <- sqrt(LSearch)[-1]
SLNewsl <- sqrt(LNewsl)[-1]
SLRetarg <- sqrt(LRetarg)[-1]
SLPortal <- sqrt(LPortal)[-1]
```

Model Iteration 4 : Full Model (with all lagged variables) with Intercept

```
#Running Full Model with intercept and all lagged variables
model.L <- lm(Sales~
                Stm1 +
                SLCatlg.Exist +
                SCatlg.Winback +
                SLCatlg.NewCust +
                SLMailings +
                SLBanner +
                SLSearch +
                SLNewsl +
                SLRetarg +
                SLPortal
              )
summary(model.L)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SLCatlg.Exist + SCatlg.Winback +
##     SLCatlg.NewCust + SLMailings + SLBanner + SLSearch + SLNewsl +
##     SLRetarg + SLPortal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1026.3  -257.7    -2.0   290.6  1416.7
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      573.9576  1038.7100   0.553   0.5847
## Stm1               0.3022     0.1638   1.845   0.0749 .
## SLCatlg.Exist     34.6814    13.9340   2.489   0.0186 *
## SCatlg.Winback    22.6572    15.8311   1.431   0.1627
## SLCatlg.NewCust   19.7861     8.7831   2.253   0.0317 *
## SLMailings        51.7004    39.5954   1.306   0.2016
## SLBanner          -9.3880    56.8943  -0.165   0.8700
## SLSearch         -24.0549   183.7386  -0.131   0.8967
## SLNewsl           29.7046   122.1825   0.243   0.8096
## SLRetarg         -27.4948    83.1452  -0.331   0.7432
## SLPortal         759.3162   559.9818   1.356   0.1852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 621.1 on 30 degrees of freedom
## Multiple R-squared:  0.5432, Adjusted R-squared:  0.391
## F-statistic: 3.568 on 10 and 30 DF,  p-value: 0.003303
```

```
AIC(model.L)
```

```
## [1] 654.9277
```

```
BIC(model.L)
```

```
## [1] 675.4906
```

This model is overall significant (F-statistic p-value: 0.003303) and has a higher R-squared = 0.5432, which means it explains 54.32% of the variation in Sales. However, the model is not parsimonious, there are multiple variables which are contributing to the model.

In the next step, lets do a Stepwise Selection process on the above model to drop variables one at a time iteratively and select the model with the least AIC. As done before, we use the `stepAIC()` function to accomplish this.

Model Iteration 5 : Step-wise Reduced Model

```
model.AICL <- stepAIC(model.L, k=2, trace=FALSE)
summary(model.AICL)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SLCatlg.Exist + SCatlg.Winback +
##     SLCatlg.NewCust + SLMailings + SLPortal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -978.71 -346.30   21.61  349.47 1394.83
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     769.0711   757.7378   1.015  0.31729
## Stm1              0.3375     0.1456   2.318  0.02659 *
## SLCatlg.Exist    36.5817    12.0828   3.028  0.00468 **
## SCatlg.Winback   25.4233    14.2505   1.784  0.08335 .
## SLCatlg.NewCust  20.8053     8.2073   2.535  0.01602 *
## SLMailings       53.7314    35.2657   1.524  0.13685
## SLPortal        507.9294   273.6269   1.856  0.07210 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 588.7 on 34 degrees of freedom
## Multiple R-squared:  0.5349, Adjusted R-squared:  0.4528
## F-statistic: 6.517 on 6 and 34 DF,  p-value: 0.0001201
```

```
AIC(model.AICL)
```

```
## [1] 647.6673
```

Comparison of RMSE of All Models

RMSE of the 3 non-lagged models (with an without intercepts, full, reduced and lagged step-wise reduced):

```
model.1.resid <- Sales - model.1$fitted.values
model.1.RMSE <- sqrt(mean(model.1.resid^2))
model.1.RMSE
```

```
## [1] 605.5776
```

```
model.2.resid <- Sales - model.2$fitted.values
model.2.RMSE <- sqrt(mean(model.2.resid^2))
model.2.RMSE
```

```
## [1] 619.8242
```

```
model.3.resid <- Sales - model.3$fitted.values
model.3.RMSE <- sqrt(mean(model.3.resid^2))
model.3.RMSE
```

```
## [1] 621.7452
```

```
model.AICL.resid <- Sales - model.AICL$fitted.values
model.AICL.RMSE <- sqrt(mean(model.AICL.resid^2))
model.AICL.RMSE
```

```
## [1] 536.1004
```

As we can see our step-wise reduced lagged model has the least RMSE.

Justifications for the Reduced Lag Model

Observations from `model.AICL` :

1. The model has the least RMSE (536.1) among all models full or reduced, with or without intercept
2. Most of the variables in the model are now statistically significant.
3. Unlike the non-lagged model, the Coefficients of the model reflect the positive relationship of different advertising channels on Sales (e.g. `SLCatlg.Exist` : 36.5817).
4. The model also has the least AIC of all models computed before : 647.6673

"**Functional Forms**"

Log-Log Model

So far we have tried all models with square-rooted variables. Let us try different other functional forms such as log-log, lin-log:

```
# Model Diminishing Returns on Lagged variables
LLCatlg.Exist <- log(1 + LCatlg.Exist)[-1]
LLCatlg.Winback <-log(1 + LCatlg.Winback)[-1]
LLCatlg.NewCust <-log(1 + LCatlg.NewCust)[-1]
LLMailings <-log(1 + LMailings)[-1]
LLBanner <- log(1 + LBanner)[-1]
LLSearch <- log(1 + LSearch)[-1]
LLNewsl <- log(1 + LNewsl)[-1]
LLRetarg <- log(1 + LRetarg)[-1]
LLPortal <- log(1 + LPortal)[-1]

# Transform the response variable to create a log-log model
LSales <- log(1+Sales)
```

```
#Running Full Model without intercept and all lagged variables
model.loglog <- lm(LSales ~
                Stm1 +
                LLCatlg.Exist +
                LLCatlg.Winback +
                LLCatlg.NewCust +
                LLMailings +
                LLBanner +
                LLSearch +
                LLNewsl +
                LLRetarg +
                LLPortal
              )
summary(model.loglog)
```

```
##
## Call:
## lm(formula = LSales ~ Stm1 + LLCatlg.Exist + LLCatlg.Winback +
##     LLCatlg.NewCust + LLMailings + LLBanner + LLSearch + LLNewsl +
##     LLRetarg + LLPortal)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.256657 -0.076418  0.008198  0.051246  0.224949
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.226e+00  4.806e-01  15.035 1.65e-15 ***
## Stm1             2.182e-05  4.039e-05   0.540   0.5930
## LLCatlg.Exist    2.763e-02  1.558e-02   1.773   0.0864 .
## LLCatlg.Winback  2.673e-02  1.685e-02   1.586   0.1231
## LLCatlg.NewCust -9.908e-03  1.317e-02  -0.752   0.4576
## LLMailings       1.142e-02  1.416e-02   0.806   0.4264
## LLBanner        -4.477e-03  2.303e-02  -0.194   0.8472
## LLSearch         1.008e-01  1.533e-01   0.657   0.5160
## LLNewsl          7.916e-02  6.133e-02   1.291   0.2066
## LLRetarg        -1.855e-02  2.602e-02  -0.713   0.4814
## LLPortal         1.725e-01  1.576e-01   1.095   0.2823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.13 on 30 degrees of freedom
## Multiple R-squared:  0.5115, Adjusted R-squared:  0.3487
## F-statistic: 3.142 on 10 and 30 DF,  p-value: 0.007327
```

```
AIC(model.loglog)
```

```
## [1] -39.76705
```

```
BIC(model.loglog)
```

```
## [1] -19.20419
```

Running stepwise selection on the full log-log model:

```
model.loglog1 <- stepAIC(model.loglog, k=2, trace=FALSE)
summary(model.loglog1)
```

```
##
## Call:
## lm(formula = LSales ~ LLCatlg.Exist + LLCatlg.Winback + LLNewsl +
##     LLPortal)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.287037 -0.068705 -0.003515  0.073410  0.236014
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     7.626518   0.184644  41.304  < 2e-16 ***
## LLCatlg.Exist   0.021824   0.012067   1.809  0.07887 .
## LLCatlg.Winback 0.020081   0.007687   2.612  0.01304 *
## LLNewsl         0.095007   0.053053   1.791  0.08174 .
## LLPortal        0.220900   0.067016   3.296  0.00221 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1253 on 36 degrees of freedom
## Multiple R-squared:  0.4554, Adjusted R-squared:  0.3949
## F-statistic: 7.525 on 4 and 36 DF,  p-value: 0.0001634
```

```
AIC(model.loglog1)
```

```
## [1] -47.30705
```

RMSE of Log-Log Model:

```
#RMSE of log-log model
model.loglog1.resid <- Sales - exp(model.loglog1$fitted.values)
model.loglog1.RMSE <- sqrt(mean(model.loglog1.resid^2))
model.loglog1.RMSE
```

```
## [1] 559.2984
```

```
#RMSE of step-wise log-log model
model.loglog.resid <- Sales - exp(model.loglog$fitted.values)
model.loglog.RMSE <- sqrt(mean(model.loglog.resid^2))
model.loglog.RMSE
```

```
## [1] 531.1001
```

Since, ‘Search’ and ‘Portals’ are highly correlated, wanted to see their effect on the model individually:

```
#Running the log-log model without search
model.loglog_withoutsearch <- lm(LSales ~
                Stm1 +
                LLCatlg.Exist +
                LLCatlg.Winback +
                LLCatlg.NewCust +
                LLMailings +
                LLBanner +
                LLNewsl +
                LLRetarg +
                LLPortal
              )
summary(model.loglog_withoutsearch)
```

```
##
## Call:
## lm(formula = LSales ~ Stm1 + LLCatlg.Exist + LLCatlg.Winback +
##     LLCatlg.NewCust + LLMailings + LLBanner + LLNewsl + LLRetarg +
##     LLPortal)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.247234 -0.069897  0.004292  0.059440  0.229056
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.502e+00  2.320e-01  32.331   <2e-16 ***
## Stm1             2.760e-05  3.906e-05   0.706   0.4852
## LLCatlg.Exist    2.802e-02  1.543e-02   1.816   0.0790 .
## LLCatlg.Winback  2.365e-02  1.604e-02   1.475   0.1504
## LLCatlg.NewCust -7.966e-03  1.271e-02  -0.627   0.5355
## LLMailings       1.256e-02  1.393e-02   0.902   0.3741
## LLBanner        -2.928e-03  2.270e-02  -0.129   0.8982
## LLNewsl          7.964e-02  6.076e-02   1.311   0.1995
## LLRetarg        -9.595e-03  2.196e-02  -0.437   0.6652
## LLPortal         2.297e-01  1.302e-01   1.764   0.0876 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1288 on 31 degrees of freedom
## Multiple R-squared:  0.5045, Adjusted R-squared:  0.3606
## F-statistic: 3.507 on 9 and 31 DF,  p-value: 0.004276
```

```
AIC(model.loglog_withoutsearch)
```

```
## [1] -41.1807
```

```
BIC(model.loglog_withoutsearch)
```

```
## [1] -22.3314
```

```
model.loglog_withoutsearch.resid <- Sales - exp(model.loglog_withoutsearch$fitted.values)
model.loglog_withoutsearch.RMSE <- sqrt(mean(model.loglog_withoutsearch.resid^2))
model.loglog_withoutsearch.RMSE
```

```
## [1] 533.7469
```

```
#Running the log-log model without portals
model.loglog_withoutportals <- lm(LSales ~
                Stm1 +
                LLCatlg.Exist +
                LLCatlg.Winback +
                LLCatlg.NewCust +
                LLMailings +
                LLSearch +
                LLBanner +
                LLNewsl +
                LLRetarg
              )
summary(model.loglog_withoutportals)
```

```
##
## Call:
## lm(formula = LSales ~ Stm1 + LLCatlg.Exist + LLCatlg.Winback +
##     LLCatlg.NewCust + LLMailings + LLSearch + LLBanner + LLNewsl +
##     LLRetarg)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.262287 -0.075675  0.008873  0.057111  0.239816
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.046e+00  4.531e-01  15.550 3.45e-16 ***
## Stm1             3.429e-05  3.888e-05   0.882   0.3845
## LLCatlg.Exist    2.890e-02  1.559e-02   1.854   0.0733 .
## LLCatlg.Winback  2.650e-02  1.690e-02   1.568   0.1270
## LLCatlg.NewCust -8.859e-03  1.317e-02  -0.672   0.5063
## LLMailings       1.247e-02  1.418e-02   0.880   0.3858
## LLSearch         1.934e-01  1.283e-01   1.508   0.1418
## LLBanner        -1.947e-03  2.299e-02  -0.085   0.9330
## LLNewsl          8.572e-02  6.123e-02   1.400   0.1714
## LLRetarg        -1.392e-02  2.575e-02  -0.541   0.5926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1304 on 31 degrees of freedom
## Multiple R-squared:  0.492,  Adjusted R-squared:  0.3445
## F-statistic: 3.336 on 9 and 31 DF,  p-value: 0.005825
```

```
AIC(model.loglog_withoutportals)
```

```
## [1] -40.16082
```

```
BIC(model.loglog_withoutportals)
```

```
## [1] -21.31152
```

```
model.loglog_withoutportals.resid <- Sales - exp(model.loglog_withoutportals$fitted.values)
model.loglog_withoutportals.RMSE <- sqrt(mean(model.loglog_withoutportals.resid^2))
model.loglog_withoutportals.RMSE
```

```
## [1] 549.1674
```

Lin-Log Model

Running Full Model with intercept and all lagged variables

```
model.linlog <- lm(Sales ~
                Stm1 +
                LLCatlg.Exist +
                LLCatlg.Winback +
                LLCatlg.NewCust +
                LLMailings +
                LLBanner +
                LLSearch +
                LLNewsl +
                LLRetarg +
                LLPortal
              )
summary(model.linlog)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + LLCatlg.Exist + LLCatlg.Winback +
##     LLCatlg.NewCust + LLMailings + LLBanner + LLSearch + LLNewsl +
##     LLRetarg + LLPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1288.73  -373.73    57.29   267.38  1338.14
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1124.5530  2322.3329  -0.484   0.6317
## Stm1               0.1038     0.1952   0.532   0.5988
## LLCatlg.Exist    127.2598    75.3041   1.690   0.1014
## LLCatlg.Winback  139.3961    81.4075   1.712   0.0972 .
## LLCatlg.NewCust  -45.4552    63.6207  -0.714   0.4805
## LLMailings        51.2554    68.4395   0.749   0.4597
## LLBanner         -37.7494   111.2775  -0.339   0.7368
## LLSearch         401.9472   740.7440   0.543   0.5914
## LLNewsl          360.2804   296.3170   1.216   0.2335
## LLRetarg        -109.2678   125.7146  -0.869   0.3917
## LLPortal        1069.3319   761.3406   1.405   0.1704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 628 on 30 degrees of freedom
## Multiple R-squared:  0.533,  Adjusted R-squared:  0.3773
## F-statistic: 3.424 on 10 and 30 DF,  p-value: 0.00431
```

```
AIC(model.linlog)
```

```
## [1] 655.8376
```

```
BIC(model.linlog)
```

```
## [1] 676.4004
```

```
model.linlog1 <- stepAIC(model.linlog, k=2, trace=FALSE)
summary(model.linlog1)
```

```
##
## Call:
## lm(formula = Sales ~ LLCatlg.Exist + LLCatlg.Winback + LLNewsl +
##     LLPortal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1120.65  -424.05   -44.51   354.16  1382.04
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       706.39     895.84   0.789  0.43555
## LLCatlg.Exist      98.68      58.55   1.685  0.10055
## LLCatlg.Winback   111.99      37.30   3.003  0.00484 **
## LLNewsl           458.76     257.40   1.782  0.08314 .
## LLPortal         1100.05     325.14   3.383  0.00174 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 607.8 on 36 degrees of freedom
## Multiple R-squared:  0.475,  Adjusted R-squared:  0.4167
## F-statistic: 8.143 on 4 and 36 DF,  p-value: 8.767e-05
```

```
AIC(model.linlog1)
```

```
## [1] 648.6345
```

RMSE of Lin-Log Model:

```
model.linlog1.resid <- Sales - model.linlog1$fitted.values
model.linlog1.RMSE <- sqrt(mean(model.linlog1.resid^2))
model.linlog1.RMSE
```

```
## [1] 569.5787
```

“**Synergy**”

Calculating Synergy for the final model:

```
# Create dataframes
reduced_data_df <- as.data.frame(cbind(SCatlg.Exist, SCatlg.Winback, SCatlg.NewCust, SMailings,
 SSearch, SNewsl, SRetarg, SPortal))
full.model <- lm(Sales ~ Stm1 + ., data = reduced_data_df)
step.model <- stepAIC(full.model, direction = "both",
                      trace = FALSE)
coeff <- step.model$coefficients
```

```r
# Elasticity --- Incremental sales unit for increase in sqrt(variable)
eta <- c()
for (i in 2:length(coeff)) {

  xbar <- mean(get(names(coeff)[i]))
  print(coeff[i])
  placeholder <- (coeff[i]*sqrt(xbar))/(2*(coeff[1] + coeff[i]*sqrt(xbar)))
  eta <- append(eta, placeholder)

}
```

```
## SCatlg.Exist
##    -23.94153
## SCatlg.Winback
##       34.83049
## SCatlg.NewCust
##      -19.37807
##   SRetarg
## -85.38985
##  SPortal
## 1367.006
```

```r
# Calculate RMSE

model.residuals <-  Sales - full.model$fitted.values
model.RMSE <- sqrt(mean(model.residuals^2))
model.RMSE
```

```
## [1] 606.073
```

```r
# Output
summary(full.model)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + ., data = reduced_data_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1165.26  -464.88    50.09   394.28  1708.99
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1585.11179 1280.50233   1.238   0.2251
## Stm1              0.07143    0.21143   0.338   0.7378
## SCatlg.Exist    -24.66577   16.85641  -1.463   0.1535
## SCatlg.Winback   48.60959   27.63849   1.759   0.0885 .
## SCatlg.NewCust  -24.62428   15.12385  -1.628   0.1136
## SMailings       -14.99707   43.90899  -0.342   0.7350
## SSearch         139.96466  216.67209   0.646   0.5230
## SNewsl          124.48199  142.85463   0.871   0.3902
## SRetarg         -90.12538   92.70080  -0.972   0.3385
## SPortal         876.87731  620.52278   1.413   0.1676
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 697 on 31 degrees of freedom
## Multiple R-squared:  0.4056, Adjusted R-squared:  0.233
## F-statistic:  2.35 on 9 and 31 DF,  p-value: 0.03734
```

```
AIC(full.model)
```

```
## [1] 663.727
```

```
BIC(full.model)
```

```
## [1] 682.5763
```

```
eta
```

```
##   SCatlg.Exist SCatlg.Winback SCatlg.NewCust        SRetarg        SPortal
##    -0.02388119     0.01598208    -0.01343665    -0.02596063     0.22608528
```

Synergy - All Variables, excl. Social Media

```
data_df_with_banner <- as.data.frame(cbind(reduced_data_df, SBanner))
full.model.synergy <- lm(Sales ~ Stm1 + . , data = data_df_with_banner)
step.model.synergy <- stepAIC(full.model.synergy, scope = . ~ .^2, direction = "both",
                        trace = FALSE)
coeff.synergy <- step.model.synergy$coefficients

# Elasticity --- Lambda calculated

eta.synergy <- c()

#MANUAL****
non_interaction_var <- 6

for (i in 3:(non_interaction_var+2)) {

  xbar <- mean(get(names(coeff.synergy)[i]))
  print(coeff.synergy[i])
  placeholder <- (coeff.synergy[i]*sqrt(xbar))/(2*(coeff.synergy[1] + coeff.synergy[i]*sqrt(xba
r)))
  eta.synergy <- append(eta.synergy, placeholder)

}
```

```
## SCatlg.Exist
##    -197.3159
## SCatlg.Winback
##      -283.3183
## SCatlg.NewCust
##       -20.0531
##  SNewsl
## 70.2209
##   SPortal
## -634.1612
##   SBanner
## -2525.552
```

```
summary(step.model.synergy)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SNewsl + SPortal + SBanner + SCatlg.Exist:SCatlg.Winback +
##     SCatlg.Winback:SNewsl + SCatlg.Exist:SPortal + SPortal:SBanner,
##     data = data_df_with_banner)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -807.62 -218.07  -29.76  265.39  962.96
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 8196.0279  1525.7809   5.372 9.03e-06 ***
## Stm1                          -0.1706     0.1389  -1.228 0.229206
## SCatlg.Exist                -197.3159    54.7971  -3.601 0.001168 **
## SCatlg.Winback              -283.3183    69.4617  -4.079 0.000323 ***
## SCatlg.NewCust               -20.0531     9.9651  -2.012 0.053554 .
## SNewsl                        70.2209    97.1768   0.723 0.475707
## SPortal                     -634.1612   639.8939  -0.991 0.329863
## SBanner                    -2525.5517  1171.1654  -2.156 0.039477 *
## SCatlg.Exist:SCatlg.Winback    5.8934     1.8661   3.158 0.003693 **
## SCatlg.Winback:SNewsl         43.5594    16.2392   2.682 0.011942 *
## SCatlg.Exist:SPortal          59.1390    23.7237   2.493 0.018631 *
## SPortal:SBanner              908.5971   420.6751   2.160 0.039188 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 464.3 on 29 degrees of freedom
## Multiple R-squared:  0.7533, Adjusted R-squared:  0.6597
## F-statistic: 8.049 on 11 and 29 DF,  p-value: 3.288e-06
```

```
AIC(step.model.synergy)
```

```
## [1] 631.6738
```

```
BIC(step.model.synergy)
```

```
## [1] 653.9503
```

```
eta.synergy
```

```
##   SCatlg.Exist SCatlg.Winback SCatlg.NewCust         SNewsl        SPortal
##    -0.06444950    -0.04444268    -0.00414950     0.00887481    -0.06584876
##        SBanner
##    -0.17002027
```

```
model.synergy.residuals <-  Sales - full.model.synergy$fitted.values
model.synergy.RMSE <- sqrt(mean(model.synergy.residuals^2))
model.synergy.RMSE
```

```
## [1] 605.5776
```

Synergy - All Variables, excl. Social Media & Banner

```
full.model.synergy.2 <- lm(Sales ~ Stm1 + . , data = reduced_data_df)
step.model.synergy.2 <- stepAIC(full.model.synergy.2, scope = . ~ .^2, direction = "both",
                        trace = FALSE)
coeff.synergy.2 <- step.model.synergy.2$coefficients

# Elasticity --- Lambda calculated

eta.synergy.2 <- c()

# MANUAL ****
non_interaction_var <- 6


for (i in 3:(non_interaction_var+2)) {

  xbar <- mean(get(names(coeff.synergy.2)[i]))
  print(coeff.synergy.2[i])
  placeholder <- (coeff.synergy.2[i]*sqrt(xbar))/(2*(coeff.synergy.2[1] + coeff.synergy.2[i]*sqr
t(xbar)))
  eta.synergy.2 <- append(eta.synergy.2, placeholder)

}
```

```
## SCatlg.Exist
##    -199.1124
## SCatlg.Winback
##      -307.4464
## SCatlg.NewCust
##      -22.27744
##   SNewsl
## 59.28215
##   SRetarg
## -401.1279
##   SPortal
## -479.9302
```

```
summary(step.model.synergy.2)
```

```
##
## Call:
## lm(formula = Sales ~ Stm1 + SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SNewsl + SRetarg + SPortal + SCatlg.Exist:SCatlg.Winback +
##     SCatlg.Winback:SNewsl + SCatlg.Exist:SPortal + Stm1:SRetarg,
##     data = reduced_data_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -797.68 -272.70  -50.79  228.95 1001.17
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                8601.66709 1734.01504   4.961 2.83e-05 ***
## Stm1                         -0.27898    0.16299  -1.712 0.097642 .
## SCatlg.Exist               -199.11236   57.99221  -3.433 0.001815 **
## SCatlg.Winback             -307.44645   71.70328  -4.288 0.000182 ***
## SCatlg.NewCust              -22.27744   10.96937  -2.031 0.051524 .
## SNewsl                       59.28215  104.99886   0.565 0.576689
## SRetarg                    -401.12791  227.77994  -1.761 0.088776 .
## SPortal                    -479.93021  744.50117  -0.645 0.524232
## SCatlg.Exist:SCatlg.Winback    6.32227    1.88445   3.355 0.002227 **
## SCatlg.Winback:SNewsl        46.55850   16.97961   2.742 0.010352 *
## SCatlg.Exist:SPortal         57.27342   24.39051   2.348 0.025896 *
## Stm1:SRetarg                  0.07842    0.04618   1.698 0.100179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 475.4 on 29 degrees of freedom
## Multiple R-squared:  0.7413, Adjusted R-squared:  0.6431
## F-statistic: 7.554 on 11 and 29 DF,  p-value: 6.13e-06
```

```
AIC(step.model.synergy.2)
```

```
## [1] 633.6213
```

```
BIC(step.model.synergy.2)
```

```
## [1] 655.8977
```

```
eta.synergy.2
```

```
##   SCatlg.Exist SCatlg.Winback SCatlg.NewCust         SNewsl        SRetarg
##   -0.061663410   -0.046092466   -0.004394518    0.007163870   -0.035989593
##        SPortal
##   -0.045801673
```

```
model.synergy.2.residuals <-  Sales - full.model.synergy.2$fitted.values
model.synergy.2.RMSE <- sqrt(mean(model.synergy.2.residuals^2))
model.synergy.2.RMSE
```

```
## [1] 606.073
```

```
focal_data_df <- as.data.frame(cbind(SCatlg.Exist, SCatlg.Winback, SCatlg.NewCust, SMailings, SN
ewsl, SPortal))

focal.model.synergy <- lm(Sales ~ Stm1 + . , data = focal_data_df)
step.model.focal.synergy <- stepAIC(focal.model.synergy, scope = . ~ .^2, direction = "both",
                        trace = FALSE)

summary(step.model.focal.synergy)
```

```
##
## Call:
## lm(formula = Sales ~ SCatlg.Exist + SCatlg.Winback + SCatlg.NewCust +
##     SNewsl + SPortal + SCatlg.Exist:SCatlg.Winback + SCatlg.Winback:SNewsl +
##     SCatlg.Exist:SPortal, data = focal_data_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -875.43 -282.84  -37.96  291.47  988.91
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6709.738   1308.739   5.127 1.37e-05 ***
## SCatlg.Exist                -159.329     51.928  -3.068 0.004361 **
## SCatlg.Winback              -260.729     65.853  -3.959 0.000393 ***
## SCatlg.NewCust               -22.829      9.800  -2.329 0.026308 *
## SNewsl                        56.873    100.081   0.568 0.573817
## SPortal                     -345.645    546.429  -0.633 0.531520
## SCatlg.Exist:SCatlg.Winback    5.728      1.803   3.177 0.003291 **
## SCatlg.Winback:SNewsl         40.055     16.749   2.391 0.022832 *
## SCatlg.Exist:SPortal          44.143     21.706   2.034 0.050342 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 483.8 on 32 degrees of freedom
## Multiple R-squared:  0.7044, Adjusted R-squared:  0.6304
## F-statistic:  9.53 on 8 and 32 DF,  p-value: 1.243e-06
```

```
AIC(step.model.focal.synergy)
```

```
## [1] 633.0911
```