Analysis of Marketing Mix

Yuping Qi, Breno Albuquerque, Jacqueline Huang

Data clearning

Import data

```r
data <- read_xlsx(path  = "Multimedia_Data.xlsx",sheet = 1)
```

Basic Exploration

```r
head(data)
```

```
## # A tibble: 6 x 15
##   Months `Sales (units)` ADV_Total ADV_Offline Catalogs_ExistC~ Catalogs_Winback
##    <dbl>           <dbl>     <dbl>       <dbl>            <dbl>            <dbl>
## 1      1           3445.      661.        591.             504.               0
## 2      2           3355.     1249.       1199.             306.             315.
## 3      3           3980.     1409.       1333.            1299.               0
## 4      4           4816.     1720.       1660.             324.             200.
## 5      5           4294.      671.        621.             621.               0
## 6      6           4134.      687.        618.             618.               0
## # ... with 9 more variables: Catalogs_NewCust <dbl>, Mailings <dbl>,
## #   ADV_online <dbl>, Banner <dbl>, Search <dbl>, SocialMedia <dbl>,
## #   Newsletter <dbl>, Retargeting <dbl>, Portals <dbl>
```

```r
glimpse(data)
```

```
## Rows: 42
## Columns: 15
## $ Months               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...
## $ `Sales (units)`      <dbl> 3444.523, 3354.753, 3979.682, 4816.464, 4294.312...
## $ ADV_Total            <dbl> 661.31944, 1249.01816, 1408.83775, 1719.92887, 6...
## $ ADV_Offline          <dbl> 591.1781, 1198.6046, 1333.1326, 1659.9652, 620.8...
## $ Catalogs_ExistCust <dbl> 503.9151, 306.3622, 1298.6937, 323.9706, 620.810...
## $ Catalogs_Winback     <dbl> 0.0000, 314.6137, 0.0000, 200.1860, 0.0000, 0.00...
## $ Catalogs_NewCust     <dbl> 87.26294, 577.62865, 0.00000, 1131.56606, 0.0000...
## $ Mailings             <dbl> 0.000000, 0.000000, 34.438880, 4.242577, 0.00000...
## $ ADV_online           <dbl> 70.14136, 50.41357, 75.70517, 59.96365, 50.41357...
## $ Banner               <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Search               <dbl> 39.86268, 38.16640, 38.16640, 38.16640, 38.16640...
## $ SocialMedia          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Newsletter           <dbl> 26.886107, 8.854604, 34.146204, 18.404685, 8.854...
## $ Retargeting          <dbl> 0.000000, 0.000000, 0.000000, 0.000000, 0.000000...
## $ Portals              <dbl> 3.392569, 3.392569, 3.392569, 3.392569, 3.392569...
```

```r
summary(data)
```

```
##      Months       Sales (units)     ADV_Total         ADV_Offline
##  Min.   : 1.00   Min.   :3355   Min.   :  59.61   Min.   :   0.0
##  1st Qu.:11.25   1st Qu.:4406   1st Qu.: 709.14   1st Qu.: 617.2
##  Median :21.50   Median :4690   Median : 924.22   Median : 771.9
##  Mean   :21.50   Mean   :4809   Mean   :1047.16   Mean   : 935.3
##  3rd Qu.:31.75   3rd Qu.:5195   3rd Qu.:1408.27   3rd Qu.:1294.8
##  Max.   :42.00   Max.   :6976   Max.   :1971.53   Max.   :1815.1
##  Catalogs_ExistCust Catalogs_Winback Catalogs_NewCust     Mailings
##  Min.   :   0.0     Min.   :  0.00   Min.   :   0.00   Min.   : 0.00
##  1st Qu.: 328.7     1st Qu.:  0.00   1st Qu.:   0.00   1st Qu.: 0.00
##  Median : 598.0     Median :  0.00   Median :  43.63   Median : 0.00
##  Mean   : 567.6     Mean   : 83.42   Mean   : 272.87   Mean   :11.42
##  3rd Qu.: 625.6     3rd Qu.:174.15   3rd Qu.: 487.42   3rd Qu.:19.24
##  Max.   :1298.7     Max.   :438.54   Max.   :1131.57   Max.   :84.47
##    ADV_online          Banner            Search         SocialMedia
##  Min.   : 50.41   Min.   : 0.000   Min.   : 38.17   Min.   :0
##  1st Qu.: 70.14   1st Qu.: 0.000   1st Qu.: 45.38   1st Qu.:0
##  Median : 99.97   Median : 0.000   Median : 66.11   Median :0
##  Mean   :111.84   Mean   : 5.179   Mean   : 69.83   Mean   :0
##  3rd Qu.:136.61   3rd Qu.: 0.000   3rd Qu.: 88.19   3rd Qu.:0
##  Max.   :295.21   Max.   :87.611   Max.   :134.87   Max.   :0
##    Newsletter      Retargeting       Portals
##  Min.   : 7.057   Min.   : 0.00   Min.   :2.544
##  1st Qu.:16.691   1st Qu.: 0.00   1st Qu.:3.393
##  Median :19.779   Median : 0.00   Median :4.707
##  Mean   :20.734   Mean   :10.85   Mean   :5.246
##  3rd Qu.:25.139   3rd Qu.:18.56   3rd Qu.:6.867
##  Max.   :53.609   Max.   :49.30   Max.   :9.303
```

```r
sapply(data, function(x) sum(x != 0))
```

```
##             Months      Sales (units)          ADV_Total        ADV_Offline
##                 42                 42                 42                 39
## Catalogs_ExistCust   Catalogs_Winback   Catalogs_NewCust           Mailings
##                 39                 16                 21                 17
##         ADV_online             Banner             Search        SocialMedia
##                 42                  4                 42                  0
##         Newsletter        Retargeting            Portals
##                 42                 17                 42
```

Rename variables for easier referece

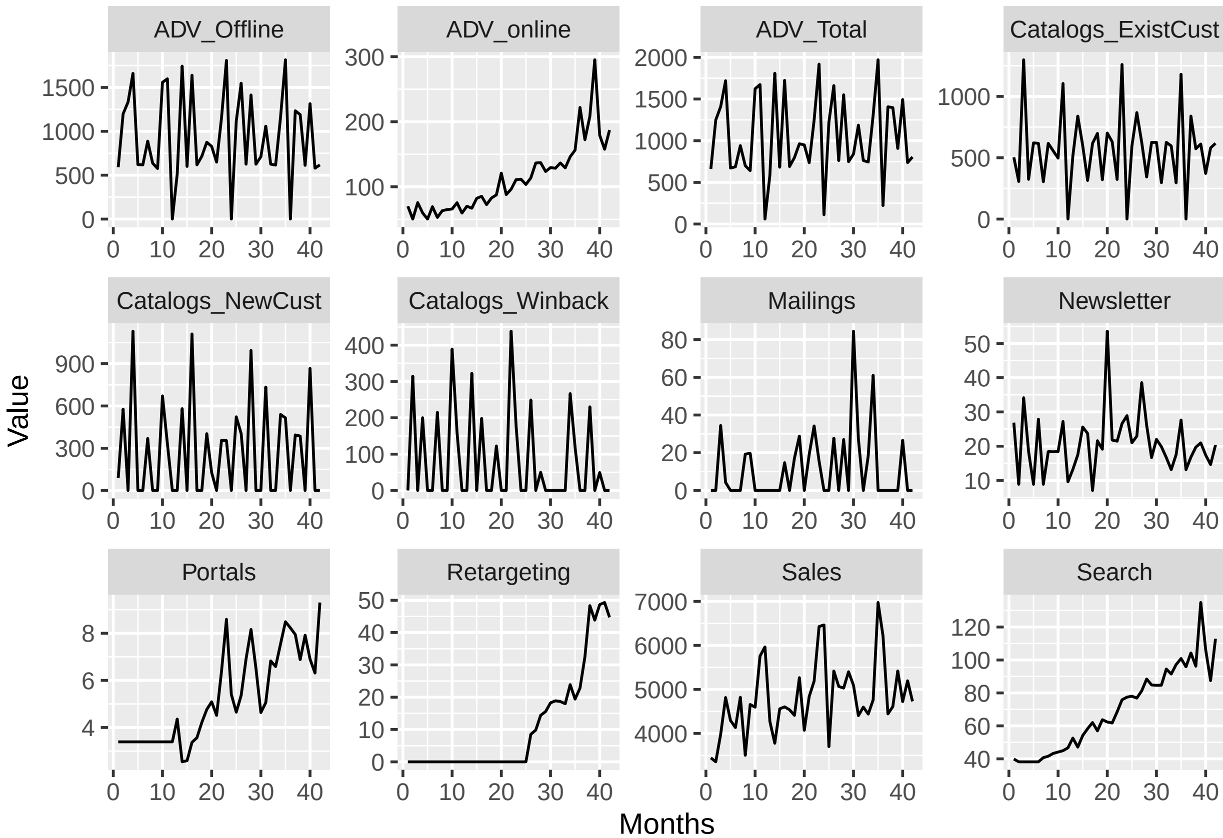
```r
data1 <- data %>%
  rename(Sales = `Sales (units)`)
```

Drop the variables that shows useless data, such as SocialMedia and Banner that has 0 mostly

```r
data2 <- data1 %>%
  select(-Banner, -SocialMedia)
```

Plot the explanatory variables against the response

```r
ggplot(data2 %>% gather(key = "Series", value = "Value", -Months), aes(x=Months, y=Value)) +geom_line()
```



Transform data and create the lag variable for seasonality

```r
#Auxiliary functions
logx <- function(x) log(x+1)
lag2 <- function(x) lag(x,2)

data3 <- data2 %>%
  mutate_at(c("Catalogs_ExistCust", "Catalogs_Winback", "Catalogs_NewCust", "Mailings", "Search", "News
  mutate(lag_Sales = dplyr::lag(Sales))
```

Glimpse of remainning data

```r
head(data3)
```

```
## # A tibble: 6 x 14
##   Months Sales ADV_Total ADV_Offline Catalogs_ExistC~ Catalogs_Winback
##    <dbl> <dbl>     <dbl>       <dbl>            <dbl>            <dbl>
## 1      1 3445.      25.7        24.3             22.4              0
## 2      2 3355.      35.3        34.6             17.5             17.7
## 3      3 3980.      37.5        36.5             36.0              0
## 4      4 4816.      41.5        40.7             18.0             14.1
## 5      5 4294.      25.9        24.9             24.9              0
```

```
## 6      6 4134.      26.2        24.9             24.9              0
## # ... with 8 more variables: Catalogs_NewCust <dbl>, Mailings <dbl>,
## #   ADV_online <dbl>, Search <dbl>, Newsletter <dbl>, Retargeting <dbl>,
## #   Portals <dbl>, lag_Sales <dbl>
```

```r
cor_mat <- stats::cor(data3 %>% as.data.frame(), use = "pairwise.complete.obs")
ggcorrplot(cor_mat, method = "circle", hc.order = T)
```

There seems to be significant correlations between multiple variables, so let's examine them further

```r
fit1<- lm(Sales~Catalogs_ExistCust+Catalogs_Winback+Catalogs_NewCust+Mailings+Search+Newsletter+Retarge
summary(fit1)
```

```
##
## Call:
## lm(formula = Sales ~ Catalogs_ExistCust + Catalogs_Winback +
##     Catalogs_NewCust + Mailings + Search + Newsletter + Retargeting +
##     Portals + lag_Sales + ADV_online + ADV_Total + ADV_Offline,
##     data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1110.73  -435.00    91.04   357.16   997.25
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)        2717.0064  1473.1328   1.844  0.07573 .
## Catalogs_ExistCust  106.9059    63.6198   1.680  0.10401
## Catalogs_Winback     48.9580    27.9581   1.751  0.09087 .
## Catalogs_NewCust     44.0535    42.6728   1.032  0.31074
## Mailings             19.3044    44.9482   0.429  0.67086
## Search              446.4306   275.1753   1.622  0.11594
## Newsletter          176.4349   153.8161   1.147  0.26107
## Retargeting          28.7204   102.7676   0.279  0.78194
## Portals             720.7726   540.6423   1.333  0.19322
## lag_Sales            -0.2115     0.2001  -1.057  0.29950
## ADV_online         -405.1425   217.5311  -1.862  0.07306 .
## ADV_Total           274.6469   124.1858   2.212  0.03533 *
## ADV_Offline        -368.6177   102.8344  -3.585  0.00126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 604.3 on 28 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.5964, Adjusted R-squared:  0.4234
## F-statistic: 3.448 on 12 and 28 DF,  p-value: 0.003388
```

The results show a high R squared and low P value, indicating that the explanatory variables are significant.However, there are individual variables that are not independently significant, and some siginificant at only siginificant at 10%

We suspect that there is multicollinearity and run the following test

```r
#pairwise correlation
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.6.2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
explan_var <- data3[,3:14]
ggpairs(explan_var)
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

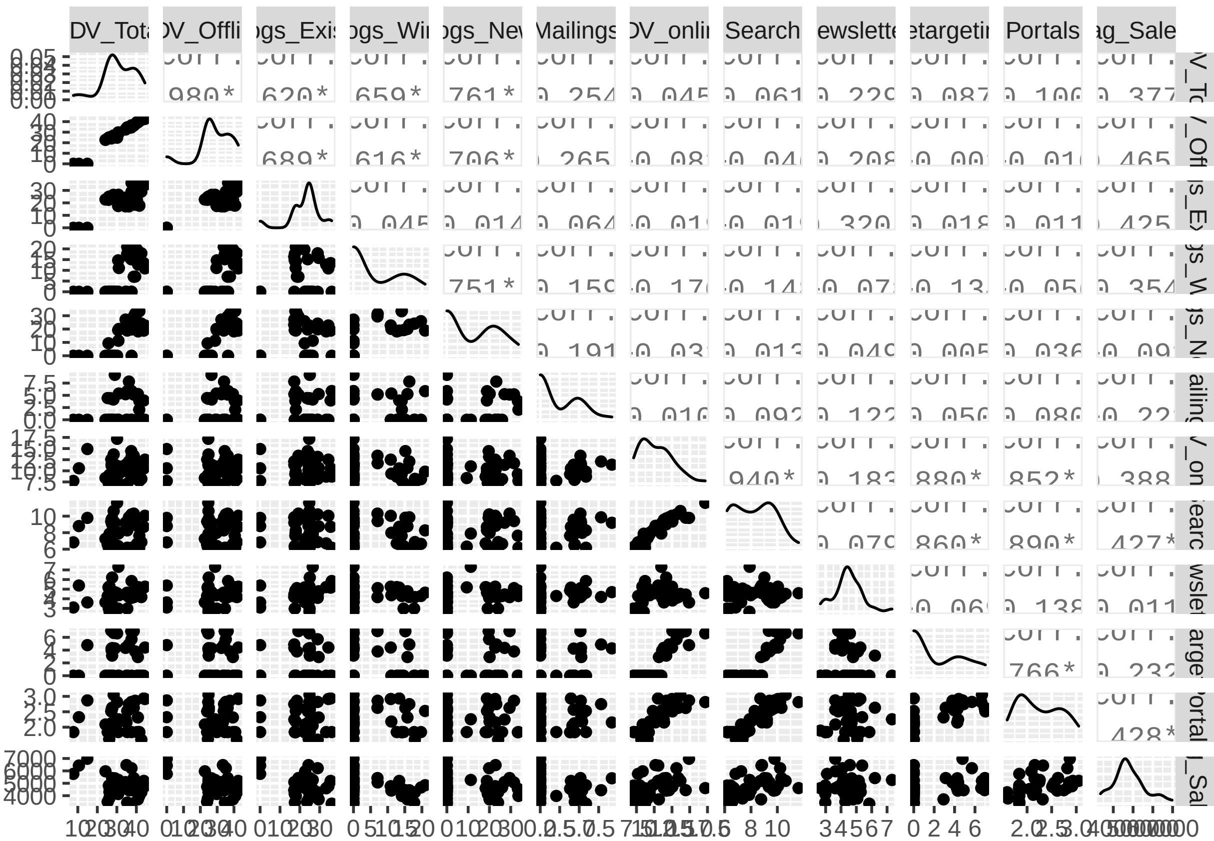
```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
```

```
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).
```



```r
cor(explan_var)
```

```
##                       ADV_Total  ADV_Offline Catalogs_ExistCust Catalogs_Winback
## ADV_Total            1.00000000  0.979765178         0.62035806       0.65912450
## ADV_Offline          0.97976518  1.000000000         0.68881456       0.61617228
## Catalogs_ExistCust   0.62035806  0.688814556         1.00000000       0.04468071
## Catalogs_Winback     0.65912450  0.616172277         0.04468071       1.00000000
## Catalogs_NewCust     0.76077635  0.705871302         0.01404391       0.75060019
## Mailings             0.25424173  0.264686862         0.06445301       0.15870667
## ADV_online           0.04500137 -0.082896013        -0.01930585      -0.17001498
## Search               0.06111566 -0.046489298        -0.01880519      -0.14849819
## Newsletter           0.22918396  0.208035690         0.32042146      -0.07827340
## Retargeting          0.08707959 -0.002050281         0.01769831      -0.13436513
## Portals              0.10013539 -0.010352677         0.01073198      -0.05554444
## lag_Sales                    NA           NA                 NA               NA
##                      Catalogs_NewCust   Mailings  ADV_online      Search
## ADV_Total                 0.760776347 0.25424173  0.04500137  0.06111566
## ADV_Offline               0.705871302 0.26468686 -0.08289601 -0.04648930
## Catalogs_ExistCust        0.014043910 0.06445301 -0.01930585 -0.01880519
## Catalogs_Winback          0.750600188 0.15870667 -0.17001498 -0.14849819
## Catalogs_NewCust          1.000000000 0.19149890 -0.03230760  0.01260171
## Mailings                  0.191498896 1.00000000  0.01012050  0.09155569
## ADV_online               -0.032307596 0.01012050  1.00000000  0.94000253
## Search                    0.012601711 0.09155569  0.94000253  1.00000000
## Newsletter                0.049368391 0.12201415  0.18273650  0.07896467
## Retargeting               0.005277629 0.04966158  0.88000014  0.86009887
```

```
## Portals                 0.035988795 0.07991258  0.85205834  0.89030759
## lag_Sales                        NA         NA          NA          NA
##                    Newsletter  Retargeting     Portals lag_Sales
## ADV_Total           0.22918396  0.087079591  0.10013539        NA
## ADV_Offline         0.20803569 -0.002050281 -0.01035268        NA
## Catalogs_ExistCust  0.32042146  0.017698310  0.01073198        NA
## Catalogs_Winback   -0.07827340 -0.134365131 -0.05554444        NA
## Catalogs_NewCust    0.04936839  0.005277629  0.03598879        NA
## Mailings            0.12201415  0.049661580  0.07991258        NA
## ADV_online          0.18273650  0.880000143  0.85205834        NA
## Search              0.07896467  0.860098866  0.89030759        NA
## Newsletter          1.00000000 -0.069215106  0.13780991        NA
## Retargeting        -0.06921511  1.000000000  0.76552182        NA
## Portals             0.13780991  0.765521818  1.00000000        NA
## lag_Sales                   NA           NA          NA         1
```

It is evident that even after variable transformation, there is still significant correlation between ADV_Offline and Catalog, ADV_Offline and Mailng, ADV_Online and Search, ADV_Online and Newsletter, ADV_Online and Retargeting, ADV_Online and Portals, and ADV_Total and ADV_Online and Offline

The above findings make sense. ADV_Online emncompasses Search, Retargeting, Portals, and possibly Newsletters, and ADV_Offline includes Mailing, Newsletters, and Catalogs. Therefore, our team made the decision to drop the ADV_Total, ADV_Offline, ADV_Online variable

```
data4 <- data3 %>%
  select(-ADV_Total, -ADV_Offline, -ADV_online)
head(data4)
```

```
## # A tibble: 6 x 11
##   Months Sales Catalogs_ExistC~ Catalogs_Winback Catalogs_NewCust Mailings
##    <dbl> <dbl>            <dbl>            <dbl>            <dbl>    <dbl>
## 1      1 3445.             22.4              0               9.34     0
## 2      2 3355.             17.5             17.7            24.0      0
## 3      3 3980.             36.0              0               0        5.87
## 4      4 4816.             18.0             14.1            33.6      2.06
## 5      5 4294.             24.9              0               0        0
## 6      6 4134.             24.9              0               0        0
## # ... with 5 more variables: Search <dbl>, Newsletter <dbl>, Retargeting <dbl>,
## #   Portals <dbl>, lag_Sales <dbl>
```

Model Development

Try fitting the regression model again

```
fit2<- lm(Sales ~. -Months, data = data4)
summary(fit2)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months, data = data4)
##
## 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
## Catalogs_ExistCust   -24.66577   16.85641  -1.463   0.1535
## Catalogs_Winback      48.60959   27.63849   1.759   0.0885 .
## Catalogs_NewCust     -24.62428   15.12385  -1.628   0.1136
## Mailings             -14.99707   43.90899  -0.342   0.7350
## Search               139.96466  216.67209   0.646   0.5230
## Newsletter           124.48199  142.85463   0.871   0.3902
## Retargeting          -90.12538   92.70080  -0.972   0.3385
## Portals              876.87731  620.52278   1.413   0.1676
## lag_Sales              0.07143    0.21143   0.338   0.7378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 697 on 31 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4056, Adjusted R-squared:  0.233
## F-statistic:  2.35 on 9 and 31 DF,  p-value: 0.03734
```

```r
AIC(fit2)
```

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

```r
BIC(fit2)
```

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

Then we use backward stepwise regression to eliminate variables

```r
#eliminate Mailings which have the highest P value aside from lag variable
fit3 <- lm(Sales ~. -Months -Mailings, data = data4)
summary(fit3)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings, data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1121.12  -482.07    34.87   376.06  1756.01
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1543.17193 1256.88566   1.228   0.2285
## Catalogs_ExistCust   -23.90819   16.47759  -1.451   0.1565
## Catalogs_Winback      49.61384   27.09968   1.831   0.0765 .
## Catalogs_NewCust     -25.53557   14.67973  -1.740   0.0916 .
## Search               131.46640  212.24715   0.619   0.5400
```

```
## Newsletter          120.07482  140.29328   0.856   0.3984
## Retargeting         -86.26705   90.73109  -0.951   0.3488
## Portals             863.62522  610.70088   1.414   0.1670
## lag_Sales             0.09426    0.19780   0.477   0.6369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 687.3 on 32 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4033, Adjusted R-squared:  0.2542
## F-statistic: 2.704 on 8 and 32 DF,  p-value: 0.02145
```

```r
AIC(fit3)
```

```
## [1] 661.881
```

```r
BIC(fit3)
```

```
## [1] 679.0167
```

```r
#eliminate Search which have the highest P value aside from lag variable
fit4 <- lm(Sales ~. -Months -Mailings -Search, data = data4)
summary(fit4)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings - Search, data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1173.91  -456.93    38.27   390.23  1758.38
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1911.3220  1097.0810   1.742   0.0908 .
## Catalogs_ExistCust  -24.1471    16.3185  -1.480   0.1484
## Catalogs_Winback    47.5163    26.6350   1.784   0.0836 .
## Catalogs_NewCust   -24.6400    14.4713  -1.703   0.0980 .
## Newsletter         130.7045   137.9333   0.948   0.3502
## Retargeting        -51.9304    71.1528  -0.730   0.4706
## Portals           1089.2393   485.5859   2.243   0.0317 *
## lag_Sales            0.1144     0.1933   0.592   0.5578
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 680.9 on 33 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3962, Adjusted R-squared:  0.2681
## F-statistic: 3.093 on 7 and 33 DF,  p-value: 0.01276
```

```r
AIC(fit4)
```

```
## [1] 660.3696
```

```r
BIC(fit4)
```

```
## [1] 675.7918
```

```r
#eliminate Retargeting which have the highest P value aside from lag variable
fit5 <- lm(Sales ~. -Months -Mailings -Search -Retargeting, data = data4)
summary(fit5)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings - Search - Retargeting,
##     data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1125.14  -439.39    84.34   373.91  1787.18
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2076.4968  1066.0810   1.948  0.05974 .
## Catalogs_ExistCust  -23.9232    16.2031  -1.476  0.14902
## Catalogs_Winback     54.3291    24.7737   2.193  0.03524 *
## Catalogs_NewCust    -27.3190    13.9015  -1.965  0.05761 .
## Newsletter          164.7168   128.9270   1.278  0.21005
## Portals             806.9819   291.6213   2.767  0.00908 **
## lag_Sales             0.1563     0.1833   0.853  0.39957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 676.2 on 34 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3864, Adjusted R-squared:  0.2782
## F-statistic: 3.569 on 6 and 34 DF,  p-value: 0.00753
```

```r
AIC(fit5)
```

```
## [1] 659.0261
```

```r
BIC(fit5)
```

```
## [1] 672.7347
```

```r
#eliminate Newsletter which have the highest P value aside from lag variable
fit6 <- lm(Sales ~. -Months -Mailings -Search -Retargeting -Newsletter, data = data4)
summary(fit6)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings - Search - Retargeting -
##     Newsletter, data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1285.36  -500.01     0.92   319.87  1812.26
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2516.1414  1018.0879   2.471  0.01847 *
## Catalogs_ExistCust  -17.3606    15.5055  -1.120  0.27049
## Catalogs_Winback     50.2769    24.7907   2.028  0.05022 .
## Catalogs_NewCust    -24.8726    13.8928  -1.790  0.08206 .
## Portals             844.8852   292.7168   2.886  0.00664 **
## lag_Sales             0.1677     0.1847   0.908  0.37009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 682.3 on 35 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.357,  Adjusted R-squared:  0.2651
## F-statistic: 3.886 on 5 and 35 DF,  p-value: 0.006612
```

```r
AIC(fit6)
```

```
## [1] 658.9487
```

```r
BIC(fit6)
```

```
## [1] 670.9437
```

```r
#eliminate Catalogs_ExistCust which have the highest P value aside from lag variable
fit7 <- lm(Sales ~. -Months -Mailings -Search -Retargeting -Newsletter -Catalogs_ExistCust, data = data4
summary(fit7)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings - Search - Retargeting -
##     Newsletter - Catalogs_ExistCust, data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1130.68  -569.93    43.59   332.50  1657.15
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1804.5441   798.1380   2.261   0.0299 *
## Catalogs_Winback   55.2247    24.4795   2.256   0.0302 *
## Catalogs_NewCust  -26.5445    13.8609  -1.915   0.0635 .
## Portals           763.1066   284.4544   2.683   0.0110 *
```

```
## lag_Sales           0.2711     0.1605   1.689   0.0999 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 684.7 on 36 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3339, Adjusted R-squared:  0.2599
## F-statistic: 4.512 on 4 and 36 DF,  p-value: 0.004666
```

```r
AIC(fit7)
```

```
## [1] 658.3915
```

```r
BIC(fit7)
```

```
## [1] 668.6729
```

```r
#eliminate Catalogs_NewCust which have the highest P value aside from lag variable
fit8 <- lm(Sales ~. -Months -Mailings -Search -Retargeting -Newsletter -Catalogs_ExistCust -Catalogs_Ne
summary(fit8)
```

```
##
## Call:
## lm(formula = Sales ~ . - Months - Mailings - Search - Retargeting -
##     Newsletter - Catalogs_ExistCust - Catalogs_NewCust, data = data4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1278.3  -541.1  -114.7   404.2  1549.6
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2095.3639   811.3117   2.583   0.0139 *
## Catalogs_Winback   19.0518    16.1229   1.182   0.2449
## Portals           754.2801   294.4911   2.561   0.0146 *
## lag_Sales           0.1943     0.1609   1.207   0.2350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 708.9 on 37 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2661, Adjusted R-squared:  0.2066
## F-statistic: 4.472 on 3 and 37 DF,  p-value: 0.008905
```

```r
AIC(fit8)
```

```
## [1] 660.369
```

```r
BIC(fit8)
```

```
## [1] 668.9369
```

According to the AIC/BIC development, we see that fit7 achieves the lowest AIC and BIC as well. Therefore, our focal model will be

```
final <- lm(Sales ~ Catalogs_Winback + Catalogs_NewCust  + Portals + lag_Sales, data = data4)
summary(final)
```

```
##
## Call:
## lm(formula = Sales ~ Catalogs_Winback + Catalogs_NewCust + Portals +
##     lag_Sales, data = data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1130.68  -569.93    43.59   332.50  1657.15
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1804.5441   798.1380   2.261   0.0299 *
## Catalogs_Winback   55.2247    24.4795   2.256   0.0302 *
## Catalogs_NewCust  -26.5445    13.8609  -1.915   0.0635 .
## Portals           763.1066   284.4544   2.683   0.0110 *
## lag_Sales           0.2711     0.1605   1.689   0.0999 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 684.7 on 36 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3339, Adjusted R-squared:  0.2599
## F-statistic: 4.512 on 4 and 36 DF,  p-value: 0.004666
```

Model Extension

```
library(dplyr)
library(tidyr)
library(readxl)
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 3.6.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## Registered S3 methods overwritten by 'forecast':
##   method       from
##   fitted.Arima TSA
##   plot.Arima   TSA
```

```
library(lmtest)
```

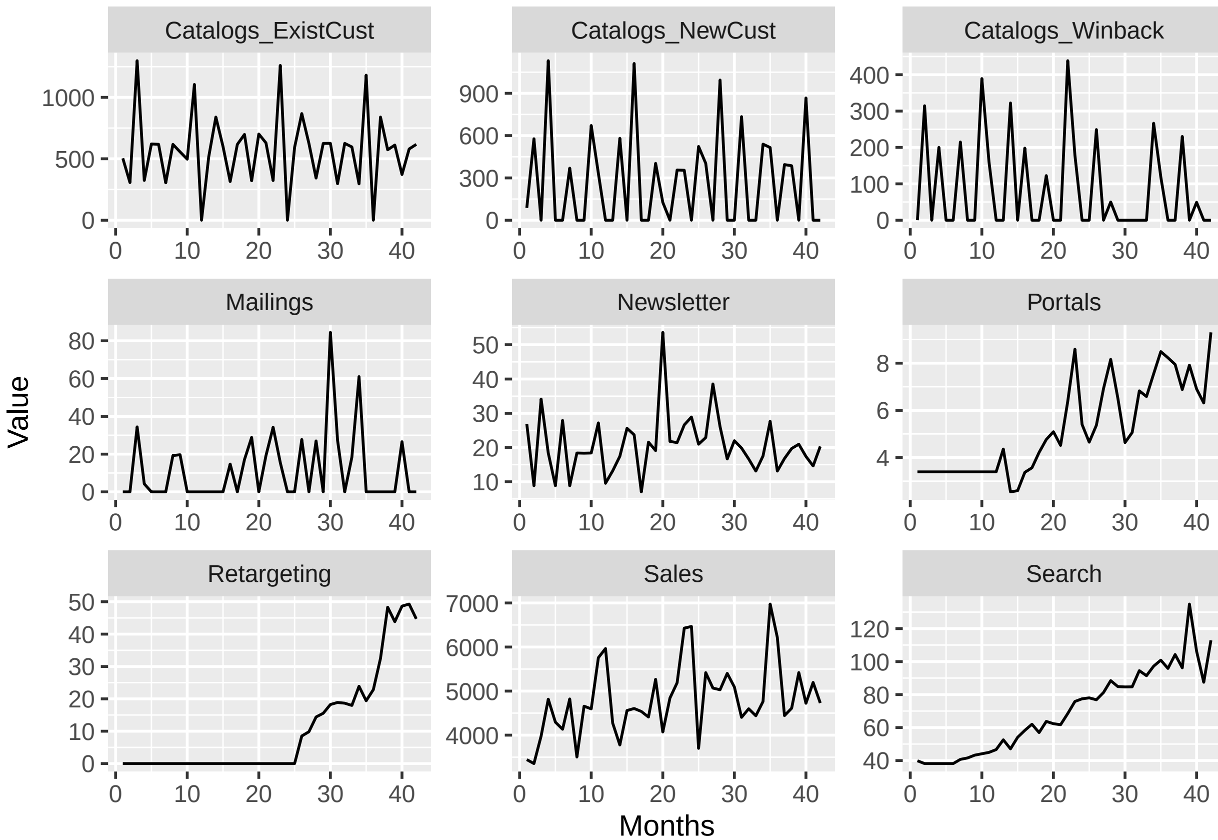
```
## Warning: package 'lmtest' was built under R version 3.6.2
```

```
## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.2

##
## Attaching package: 'zoo'

## The following object is masked from 'package:tsibble':
##
##     index

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

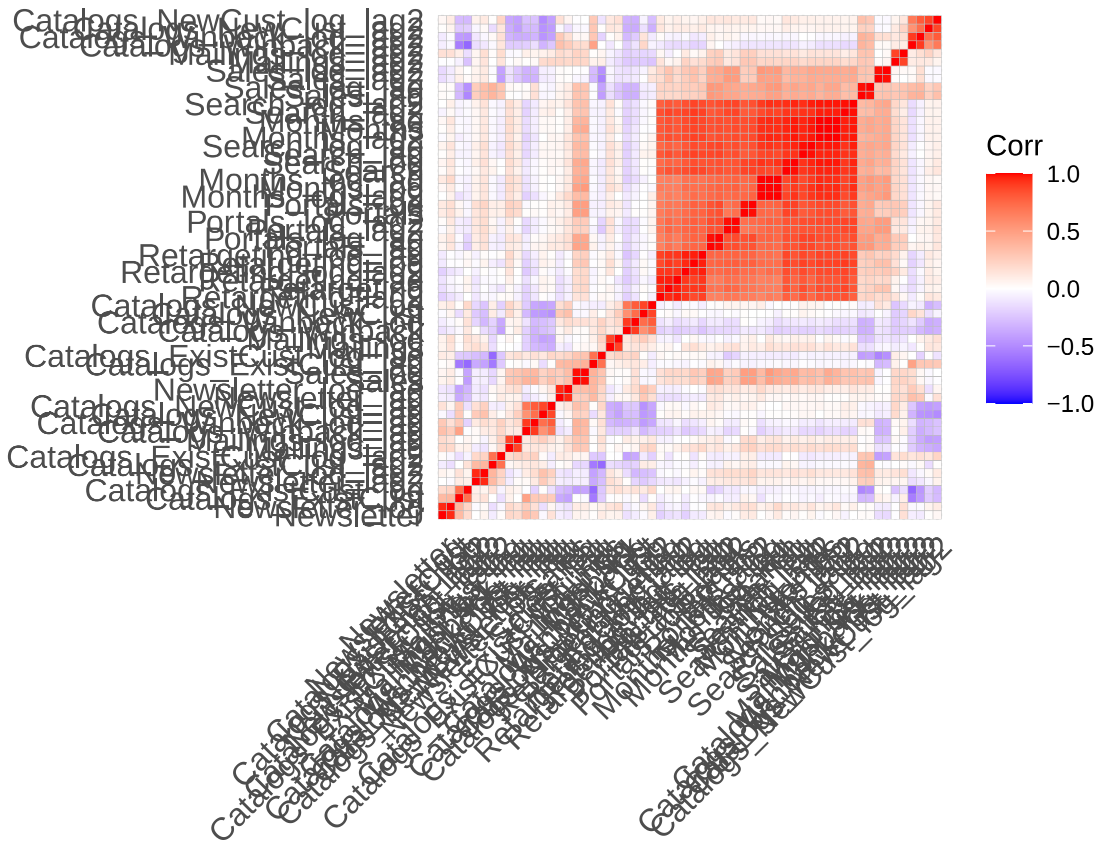
```
library(ggplot2)
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
#Auxiliary functions
logx <- function(x) log(x+1)
lag2 <- function(x) lag(x,2)

data <- read_xlsx(path  = "Multimedia_Data.xlsx",sheet = 1)%>%
  rename(Sales = `Sales (units)`) %>%
  select( -Banner, -SocialMedia, -ADV_Total, -ADV_Offline, -ADV_online)

ggplot(data %>% gather(key = "Series", value = "Value", -Months), aes(x=Months, y=Value)) +geom_line() 
```

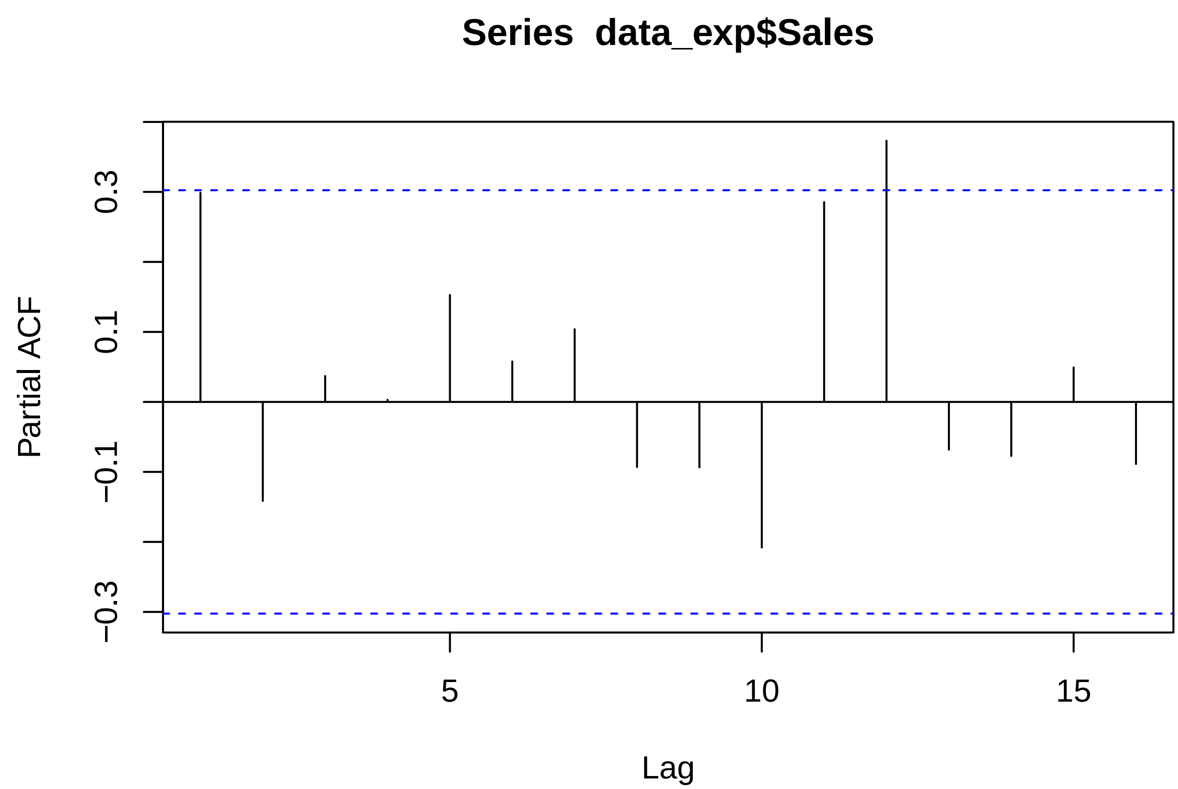



```
budget <- data %>% tail(12)

data_exp <- data %>%
  mutate_all(.funs = list(log = logx)) %>%
  mutate_all(.funs = list(lag = lag, lag2 = lag2))

# Correlation
require(ggcorrplot)
cor_mat2 <- cor(data_exp, use = "pairwise.complete.obs")
ggcorrplot(cor_mat2, hc.order = T)
```

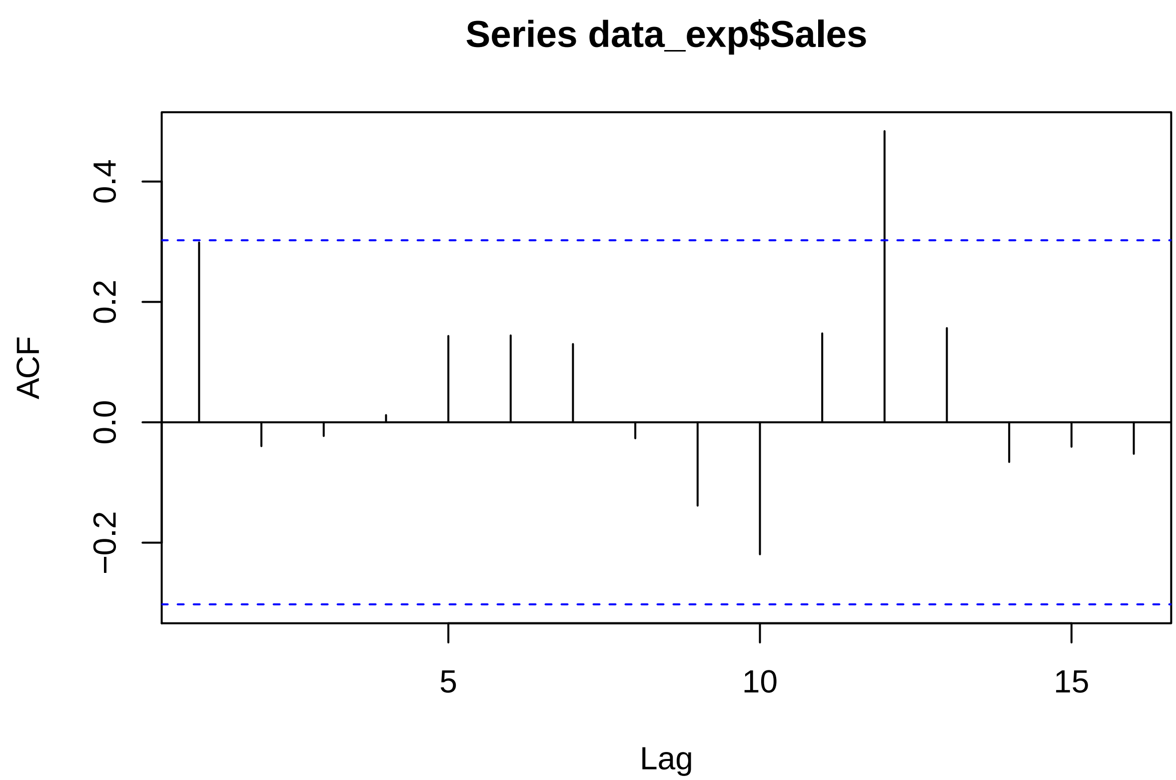


We have autocorrelation and seasonality issues. Let's deal with them. But first, lets get our data ready .

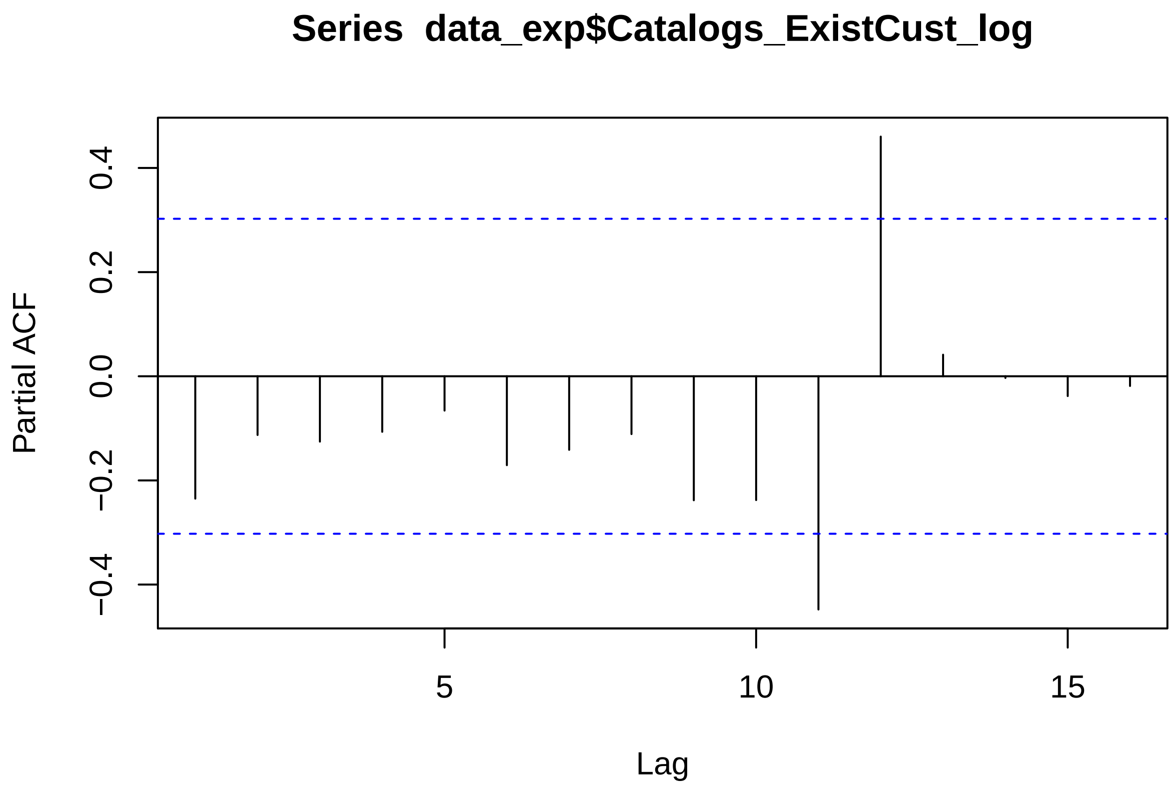
```r
pacf(data_exp$Sales)
```



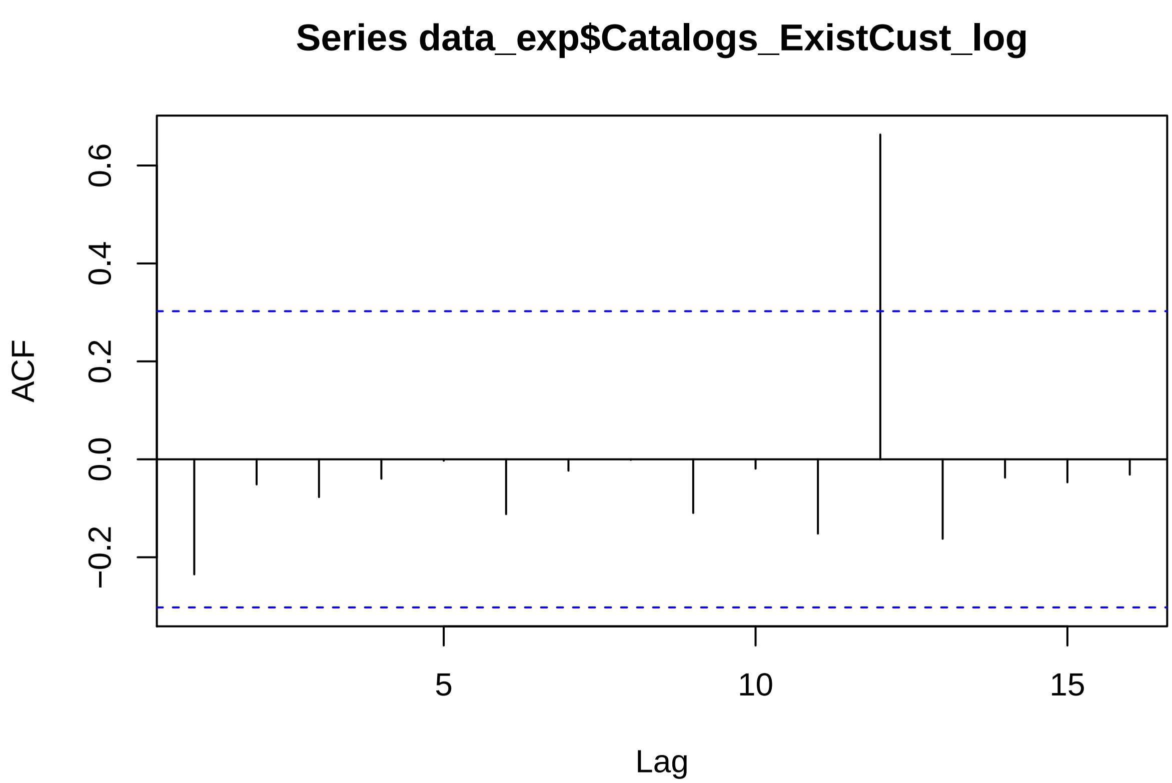

```r
acf(data_exp$Sales)
```



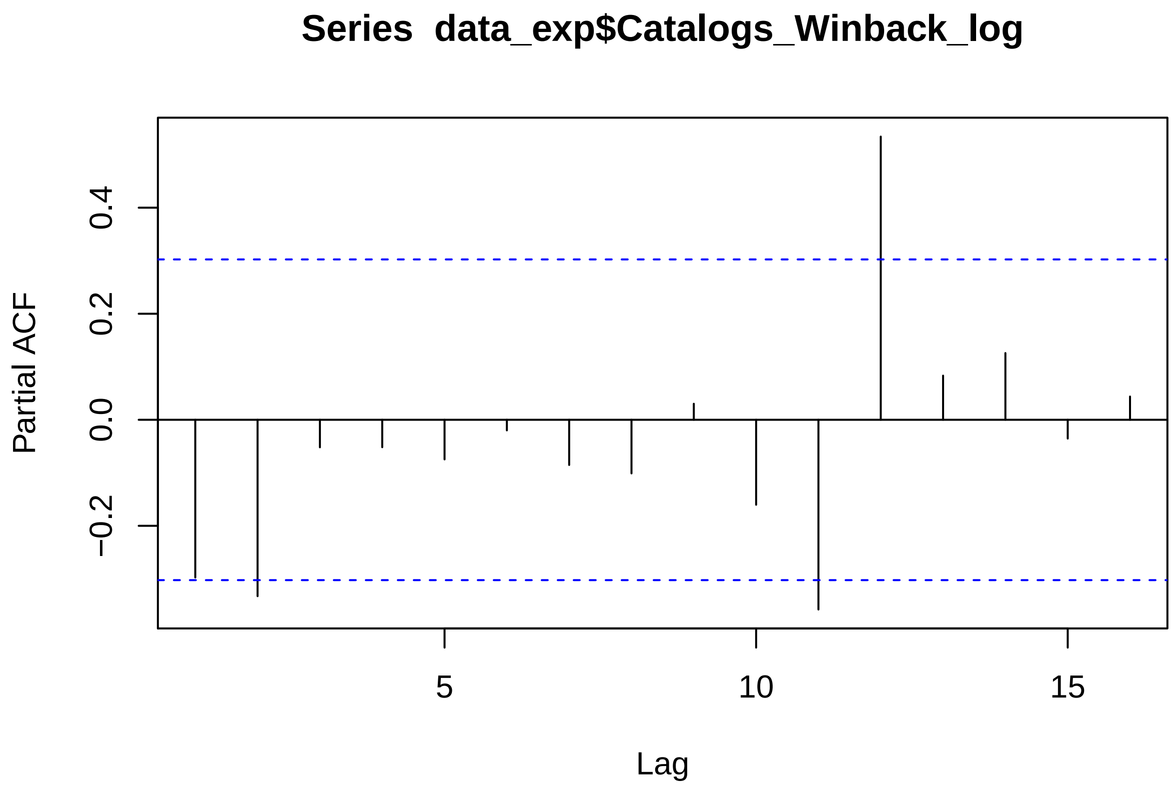
```r
pacf(data_exp$Catalogs_ExistCust_log)
```



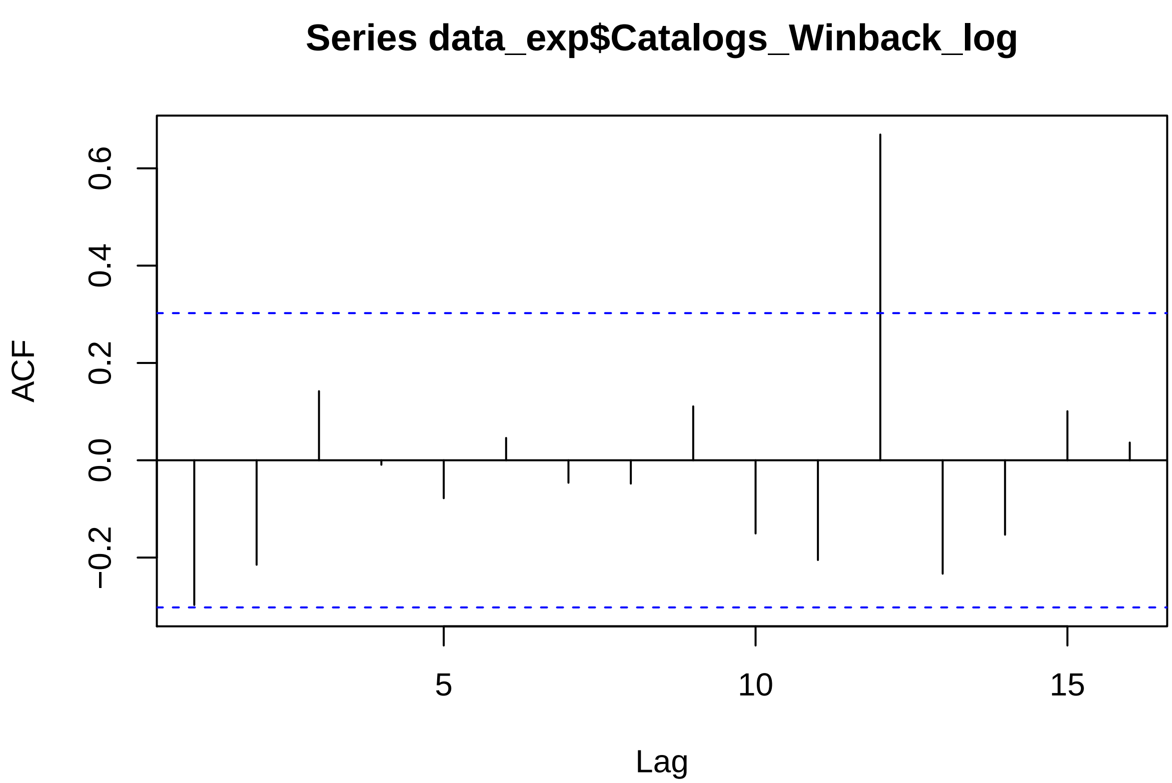

```r
acf(data_exp$Catalogs_ExistCust_log)
```



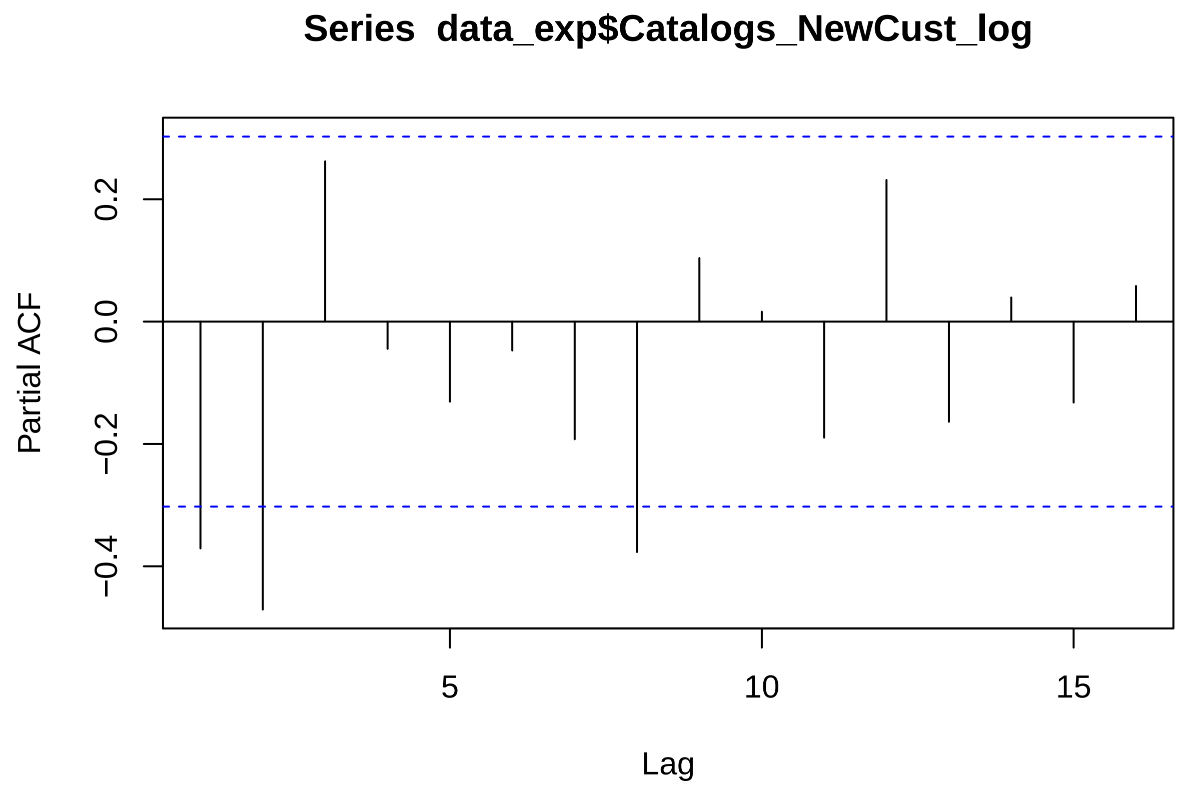
```r
pacf(data_exp$Catalogs_Winback_log)
```



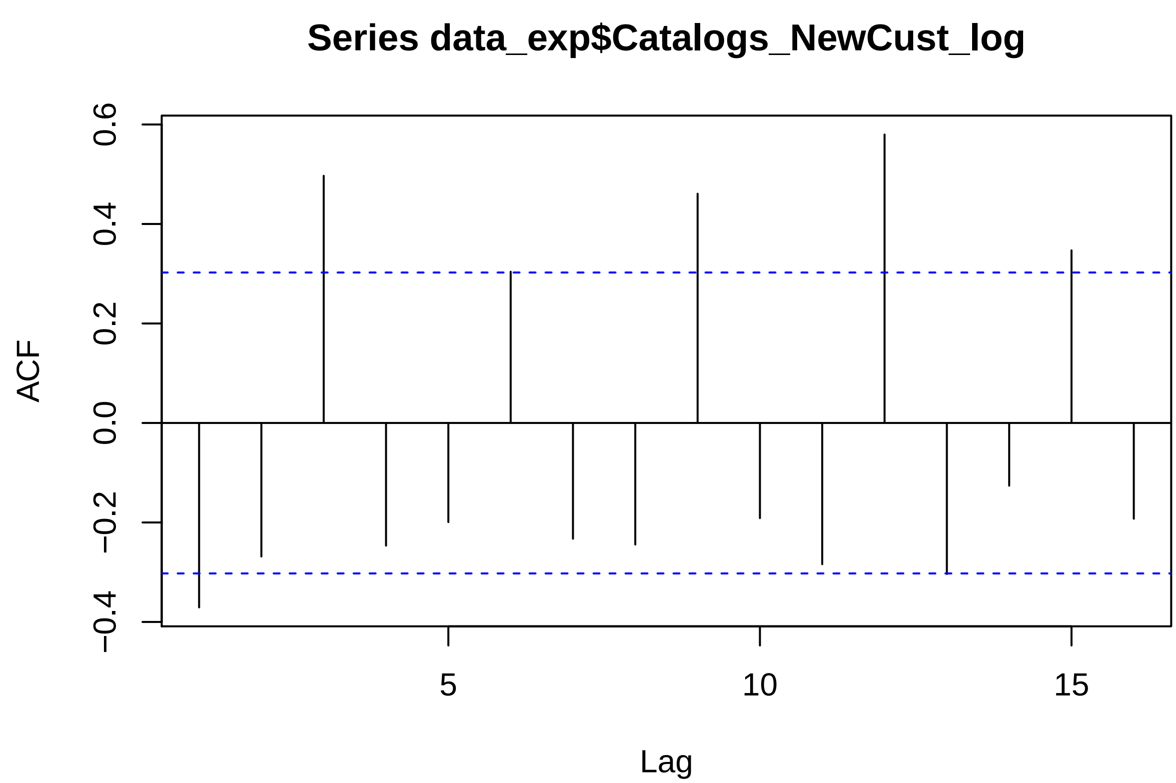

```r
acf(data_exp$Catalogs_Winback_log)
```



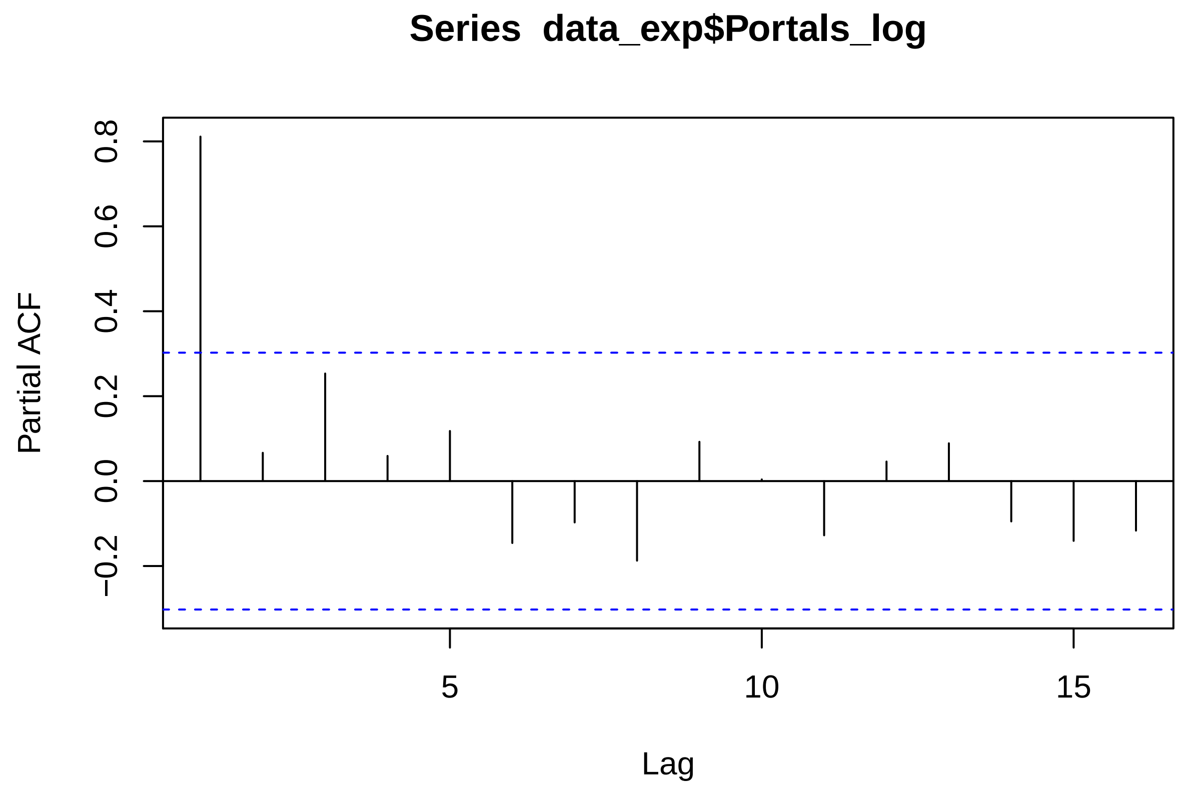
```r
pacf(data_exp$Catalogs_NewCust_log)
```



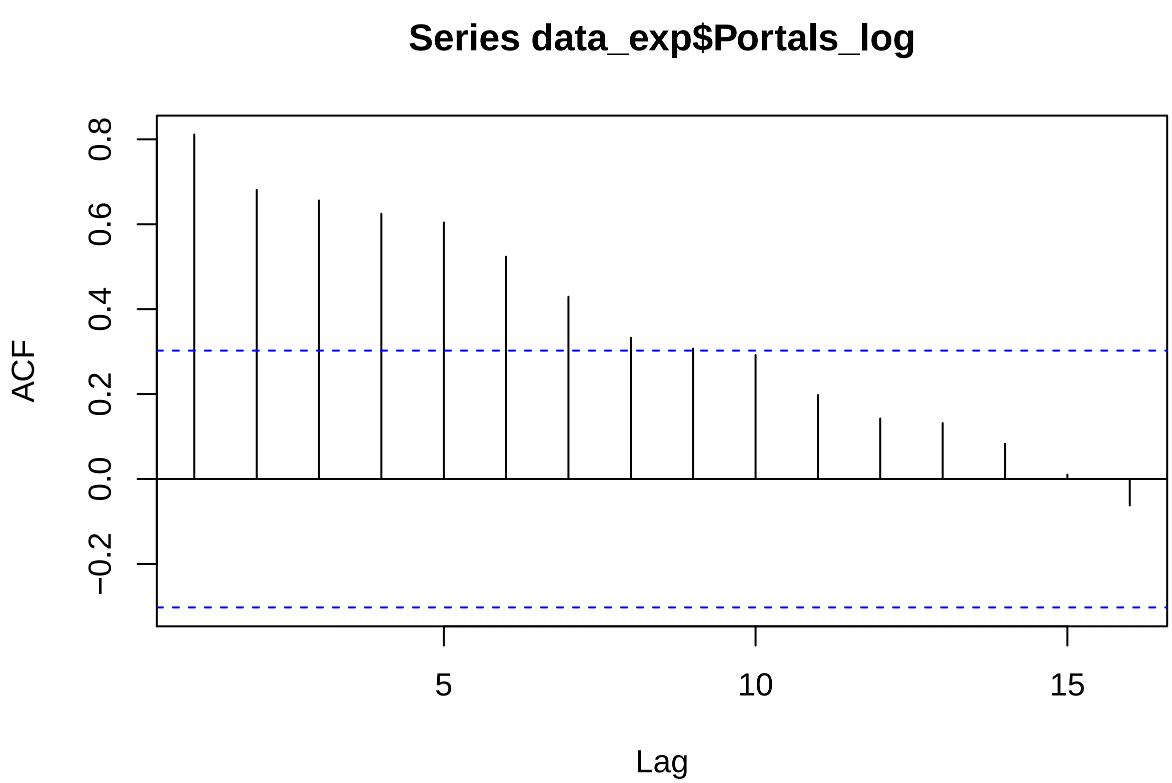

```r
acf(data_exp$Catalogs_NewCust_log)
```



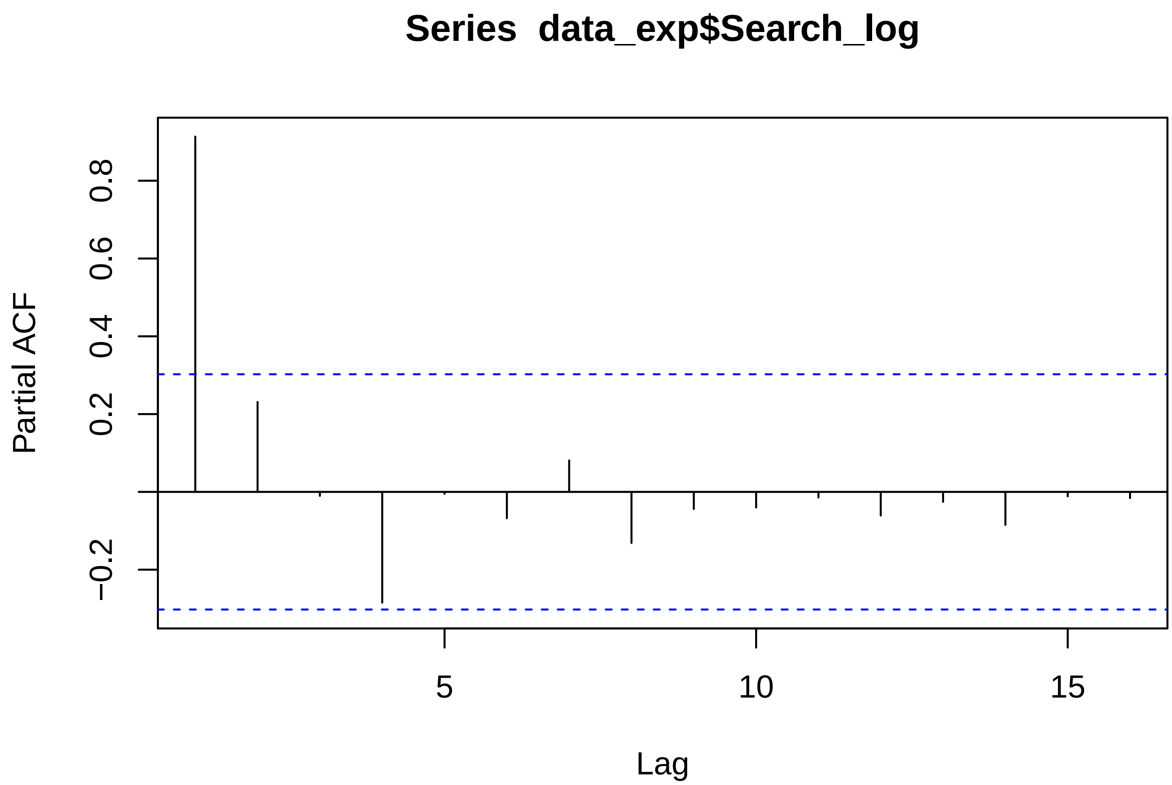
```r
pacf(data_exp$Portals_log)
```



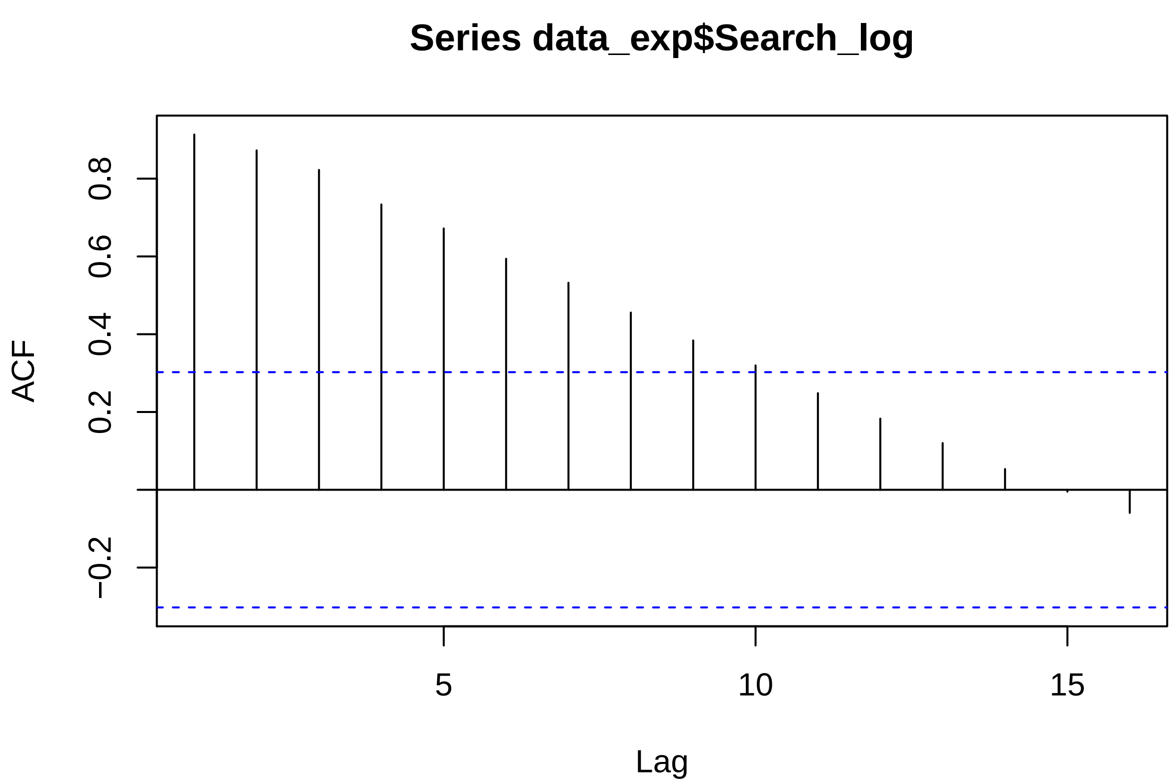

```r
acf(data_exp$Portals_log)
```



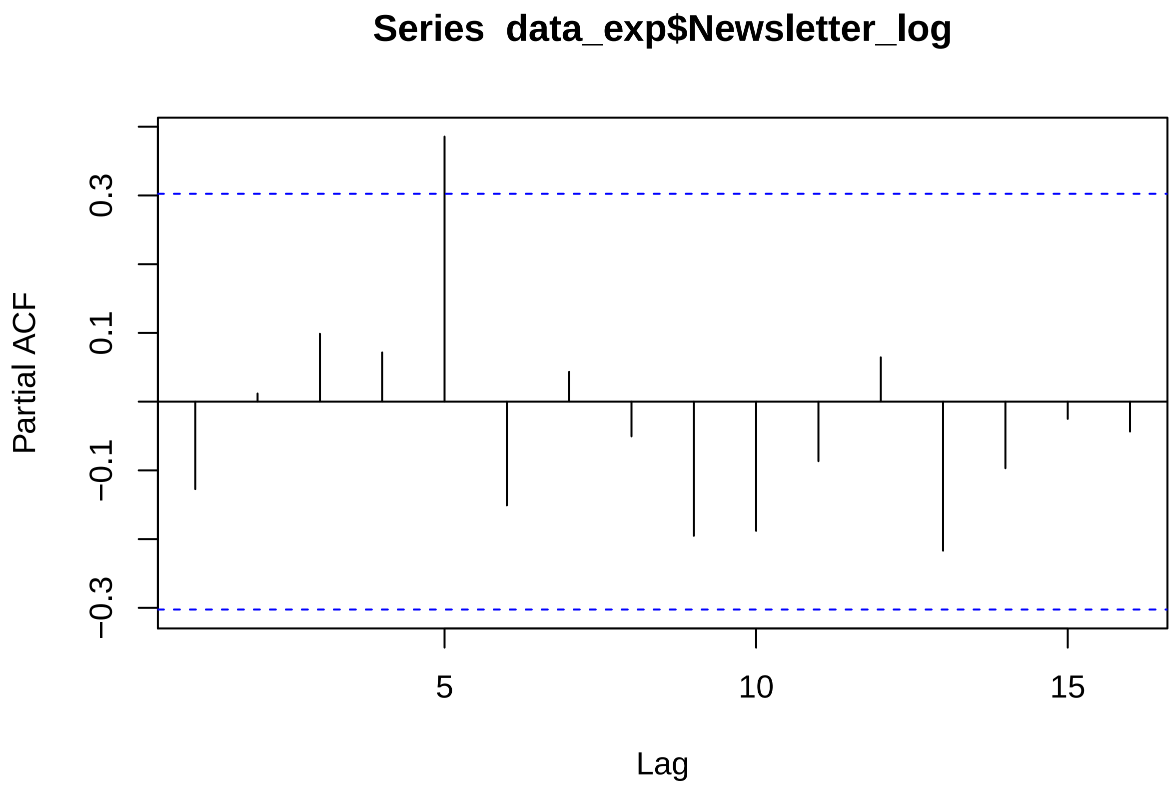
```r
pacf(data_exp$Search_log)
```



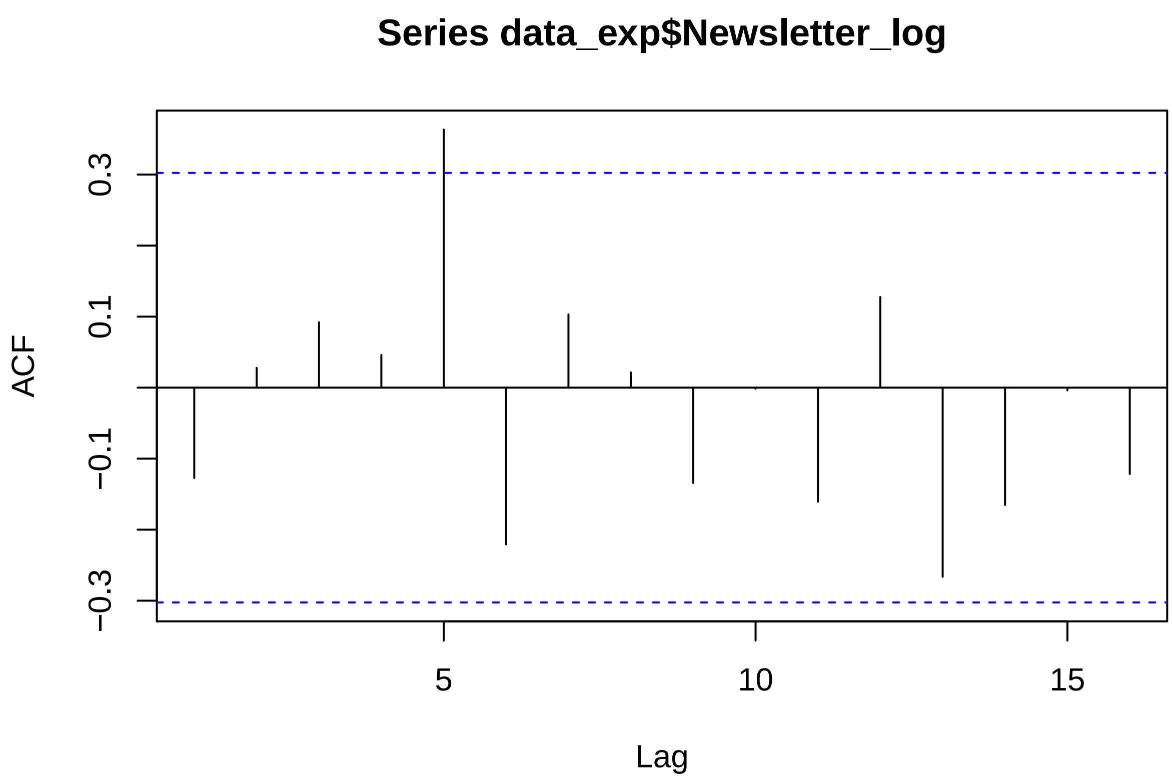

```r
acf(data_exp$Search_log)
```



```
pacf(data_exp$Newsletter_log)
```




```
acf(data_exp$Newsletter_log)
```



Building the extension model

```r
# Xreg

Covars <- cbind(
  Month_log = data_exp$Months_log,
  Portals_log = data_exp$Portals_log,
  Winback_Newsletter_log_lag = data_exp$Catalogs_Winback_log_lag * data_exp$Newsletter_log_lag ,
  ExistCust_Newsletter_log_lag = data_exp$Catalogs_ExistCust_log_lag * data_exp$Newsletter_log_lag
)

modfin <- arima(data_exp$Sales, xreg = Covars, order = c(0,0,0), seasonal = list(order = c(1,0,0), peri

summary(modfin)
```

```
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Winback_Newsletter_log_lag
##       0.5737  2034.0966   229.7782     584.6757                     35.1215
## s.e.  0.1477   528.9101   137.2745     294.2048                     15.5505
##       ExistCust_Newsletter_log_lag
##                             48.850
## s.e.                        17.789
##
## sigma^2 estimated as 191444:  log likelihood = -309.9,  aic = 631.8
##
## Training set error measures:
```

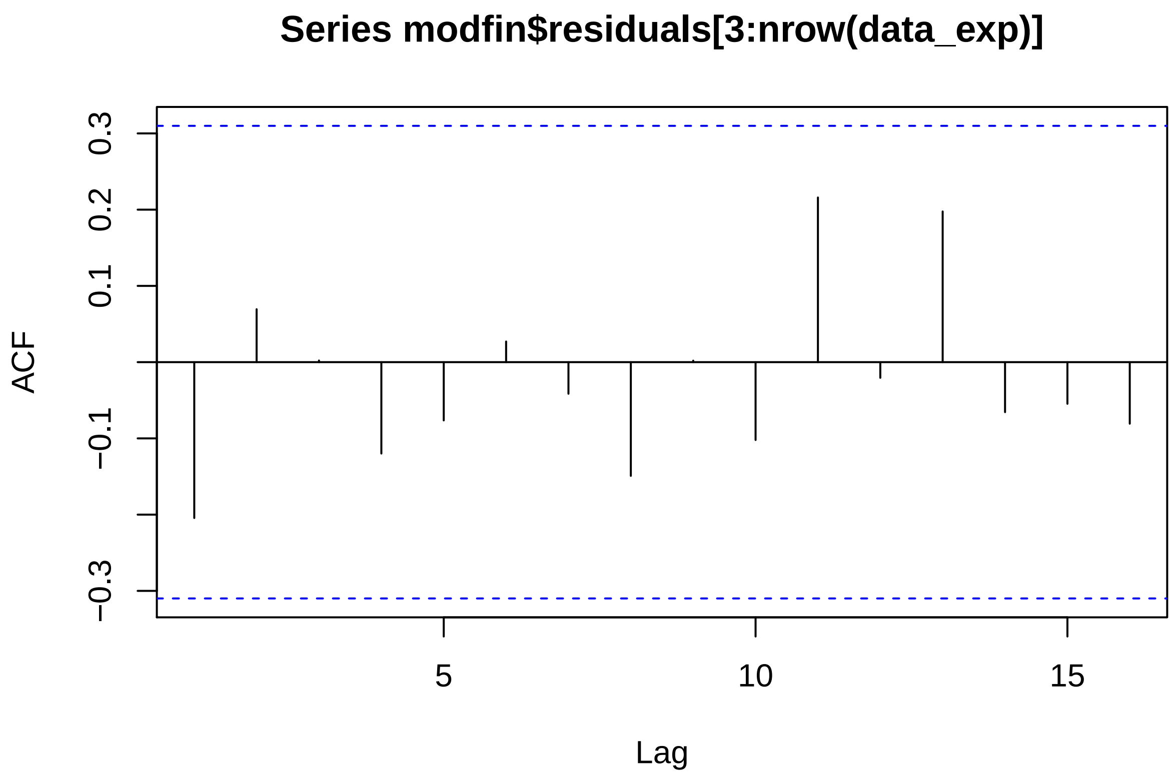
```
## Warning in trainingaccuracy(object, test, d, D): test elements must be within
## sample
```

```
##                 ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

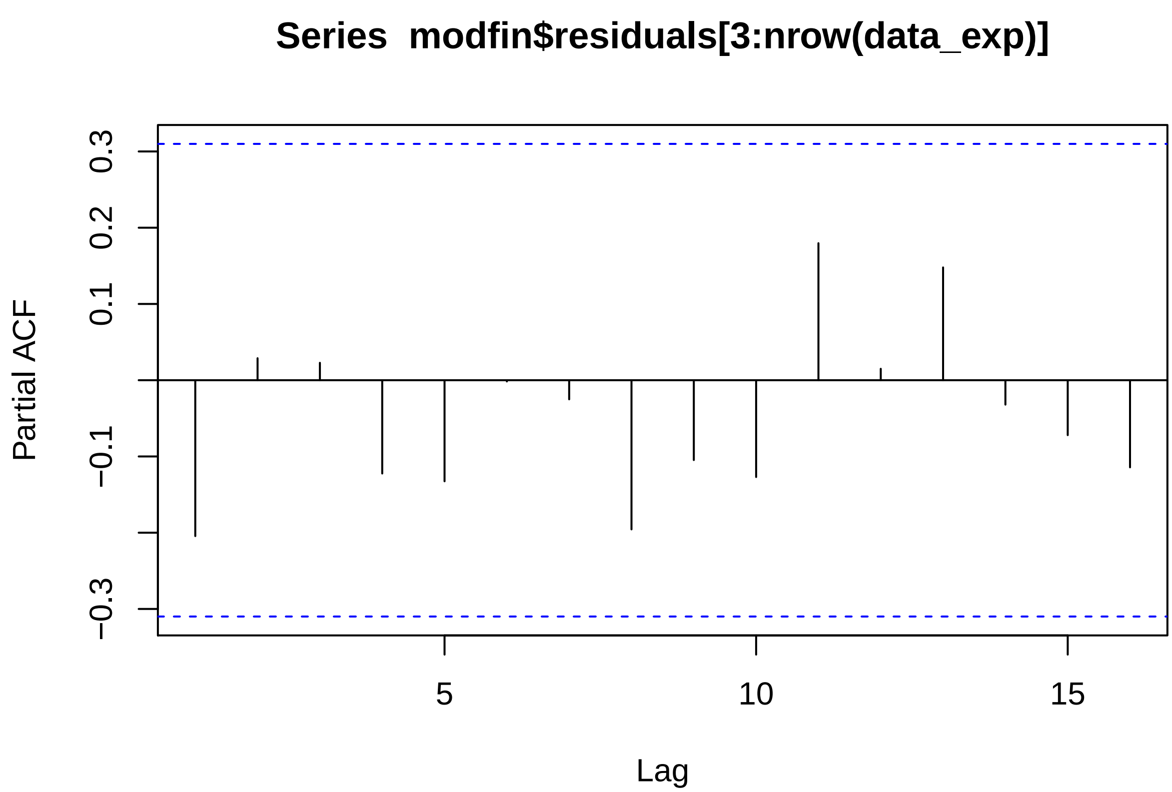
```r
coeftest(modfin)
```

```
##
## z test of coefficients:
##
##                                Estimate Std. Error z value  Pr(>|z|)
## sar1                            0.57374    0.14771  3.8843 0.0001026 ***
## intercept                    2034.09664  528.91011  3.8458 0.0001201 ***
## Month_log                     229.77816  137.27450  1.6739 0.0941583 .
## Portals_log                   584.67567  294.20476  1.9873 0.0468882 *
## Winback_Newsletter_log_lag     35.12152   15.55053  2.2585 0.0239120 *
## ExistCust_Newsletter_log_lag   48.84995   17.78895  2.7461 0.0060311 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

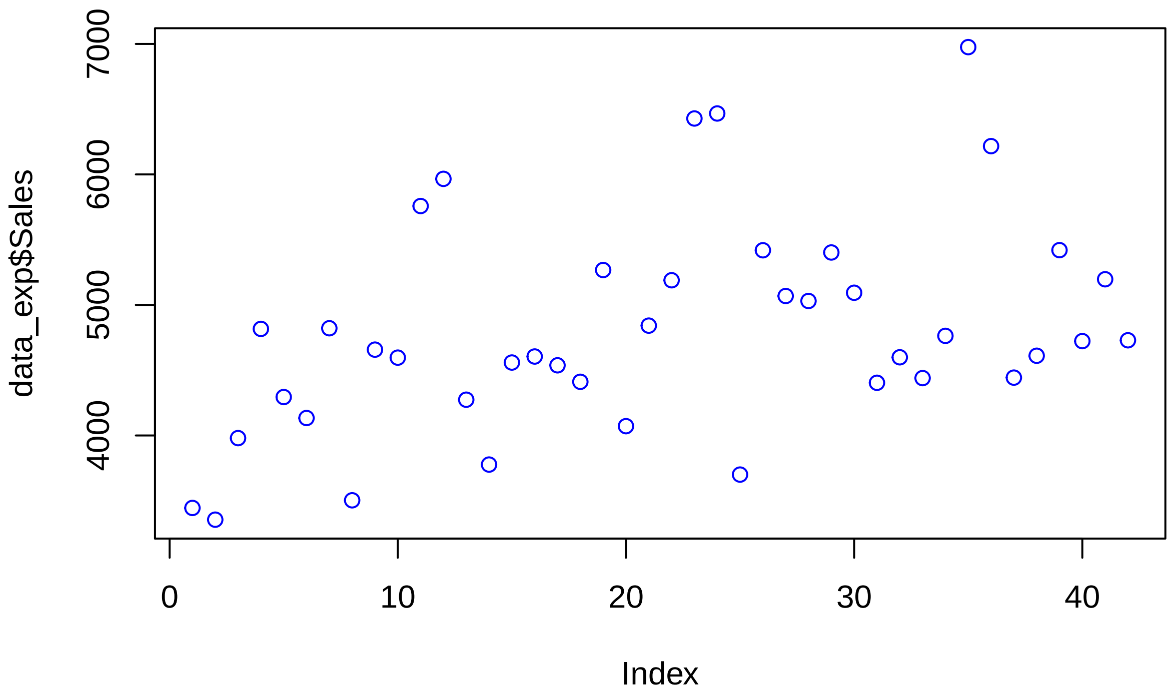
```r
acf(modfin$residuals[3:nrow(data_exp)])
```




```r
pacf(modfin$residuals[3:nrow(data_exp)])
```



```r
plot(data_exp$Sales, col = "blue")
lines(fitted(modfin))
```



Extension model For the extension model, our team dedicated to asses for the presence of several marketing industry expected effects. [https://www.accenture.com/_acnmedia/pdf-92/accenture-market-mix-optimization.pdf]

Seasonality Sales exhibits a strong seasonal behavior. This means that cosmetics tend to sell more during certain periods of the year. Units sold during the current month is correlated to units sold twelve months ago. This patterns explains a great amount of sales variation in the sample. Accordingly to our extended model Sales (in log) on a specific month departs from almost 50% of what was sold (in log) in the same month last year.

```r
mod1 <- arima(data_exp$Sales, order = c(0,0,0), seasonal = list(order = c(1,0,0), period = 12))
```

Intercept and Deterministic Time Trend Does Sales follows a time trend? Our analysis concluded that, despite any investiments in advertising, units sold tend to grow in time, on a decreasing rate. This result may be due to product earned media, organic aceptance rate or market and/or economic growth during the period. We didn't control for macroeconomic or industry specific variables. Jointly with baseline revenue (intercept was always signficant), this effect accounts for the expected sales in absence of any advertising expenditure.

```r
Covars2 <- cbind(
  Month = data_exp$Months
)
mod2 <- arima(data_exp$Sales, xreg = Covars2, order = c(0,0,0), seasonal = list(order = c(1,0,0), perio

# Covars2 <- cbind(
#   Month_log = data_exp$Months_log,
#   Portals_log = data_exp$Portals_log,
#   Winback_Newsletter_log_lag = data_exp$Catalogs_Winback_log_lag * data_exp$Newsletter_log_lag ,
```

```r
#   ExistCust_Newsletter_log_lag = data_exp$Catalogs_ExistCust_log_lag * data_exp$Newsletter_log_lag
# )
# mod2 <- arima(data_exp$Sales_log, xreg = Covars2, order = c(0,0,0), seasonal = list(order = c(1,0,0),


# Carryover effect (Adstock)
# When controled to seasonal effect, sales demonstrated few or no signal of carryover effects. It seams


Covars3 <- cbind(
  Month = data_exp$Months
)
mod3 <- arima(data_exp$Sales, xreg = Covars2, order = c(1,0,0), seasonal = list(order = c(1,0,0), perio

stargazer(mod1, mod2, mod3)
```

```
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12))
##
## Coefficients:
##         sar1  intercept
##       0.7243  4769.8043
## s.e.  0.1015   200.4022
##
## sigma^2 estimated as 320106:  log likelihood = -330.26,  aic = 664.53
##
## Training set error measures:
##                ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars2)
##
## Coefficients:
##         sar1  intercept    Month
##       0.7196  4303.5401  21.7063
## s.e.  0.1021   243.1887   7.5550
##
## sigma^2 estimated as 268247:  log likelihood = -326.47,  aic = 658.93
##
## Training set error measures:
##                ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## Call:
## arima(x = data_exp$Sales, order = c(1, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars2)
##
## Coefficients:
##           ar1    sar1  intercept    Month
```

```
##       -0.0125  0.7223  4304.3161  21.6851
## s.e.   0.1673  0.1074   240.9802   7.4655
##
## sigma^2 estimated as 267599:  log likelihood = -326.46,  aic = 660.92
##
## Training set error measures:
##               ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Dec 08, 2020 - 20:23:12
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{3}{c}{\textit{Dependent variable:}} \\
## \cline{2-4}
## \\[-1.8ex] &  &  &  \\
## \\[-1.8ex] & (1) & (2) & (3)\\
## \hline \\[-1.8ex]
##  ar1 &  &  & $-$0.013 \\
##   &  &  & (0.167) \\
##   & & & \\
##  sar1 & 0.724$^{***}$ & 0.720$^{***}$ & 0.722$^{***}$ \\
##   & (0.101) & (0.102) & (0.107) \\
##   & & & \\
##  intercept & 4,769.804$^{***}$ & 4,303.540$^{***}$ & 4,304.316$^{***}$ \\
##   & (200.402) & (243.189) & (240.980) \\
##   & & & \\
##  Month &  & 21.706$^{***}$ & 21.685$^{***}$ \\
##   &  & (7.555) & (7.466) \\
##   & & & \\
## \hline \\[-1.8ex]
## Observations & 42 & 42 & 42 \\
## Log Likelihood & $-$330.263 & $-$326.465 & $-$326.462 \\
## $\sigma^{2}$ & 320,106.300 & 268,247.400 & 267,599.100 \\
## Akaike Inf. Crit. & 666.525 & 660.930 & 662.925 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{3}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
# Covars <- cbind(
#   Month_log = data_exp$Months_log,
#   Portals_log = data_exp$Portals_log,
#   Winback_Newsletter_log_lag = data_exp$Catalogs_Winback_log_lag * data_exp$Newsletter_log_lag ,
#   ExistCust_Newsletter_log_lag = data_exp$Catalogs_ExistCust_log_lag * data_exp$Newsletter_log_lag
# )
# mod3 <- arima(data_exp$Sales_log, xreg = Covars, order = c(1,0,0), seasonal = list(order = c(1,0,0), p
```

Diminushing returns (Adstock) The extended model confirmed for diminushing return (saturation) on advertising expenditure. Log specification (for all covariates) better fits Sales when compared to the alternative (square root). We didn't try for other especifications.

```
Covars4a <- cbind(
  Month_sqrt = sqrt(data_exp$Months),
  Portals_sqrt = sqrt(data_exp$Portals),
  Catalog_ExistCust_sqrt = sqrt(data_exp$Catalogs_ExistCust),
  Catalog_Winback_sqrt = sqrt(data_exp$Catalogs_NewCust)
)
mod4a <- arima(data_exp$Sales, xreg = Covars4a, order = c(0,0,0), seasonal = list(order = c(1,0,0), peri


Covars4b <- cbind(
  Month_log = (data_exp$Months_log),
  Portals_log = (data_exp$Portals_log),
  Catalog_ExistCust_log = (data_exp$Catalogs_ExistCust_log),
  Catalog_Winback_log = (data_exp$Catalogs_NewCust_log)
)

mod4b <- arima(data_exp$Sales, xreg = Covars4b, order = c(0,0,0), seasonal = list(order = c(1,0,0), peri


summary(mod4b)
```

```
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars4b)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_ExistCust_log
##       0.6954  4025.6948   258.9134     543.1642              -152.3399
## s.e.  0.1197   682.9736   125.3163     288.6456                88.9325
##       Catalog_Winback_log
##                  -21.4352
## s.e.              39.1976
##
## sigma^2 estimated as 212860:  log likelihood = -321.2,  aic = 654.39
##
## Training set error measures:

## Warning in trainingaccuracy(object, test, d, D): test elements must be within
## sample

##                ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

```
# Covars4 <- cbind(
#   Month_sqrt = sqrt(data_exp$Months),
#   Portals_sqrt = sqrt(data_exp$Portals),
#   Winback_Newsletter_sqrt_lag = sqrt(data_exp$Catalogs_Winback_lag) * sqrt(data_exp$Newsletter_lag) ,
#   ExistCust_Newsletter_sqrt_lag = sqrt(data_exp$Catalogs_ExistCust_lag) * sqrt(data_exp$Newsletter_la
# )
# mod4 <- arima(data_exp$Sales_log, xreg = Covars4, order = c(0,0,0), seasonal = list(order = c(1,0,0),
```

Lagged effect We wanted to test if any advertising activitity would take longer than a month to impact sales. For instance, we expected that investiments in Catalog advertising would require more time than online advertising before exerting influence on sales. Since Catalogs demands a long time to print and deliver, returns on spends in this type of media may not be as immediate as other medias. Out team found that catalog advertising actually has always a one-month-lagged impact. After accounting for the proper lag, catalog expenditure coefficient turned to positive - the expected sign.

```
Covars5 <- cbind(
  Month_log = data_exp$Months_log,
  Portals_log = data_exp$Portals_log,
  Catalog_Winback_log = data_exp$Catalogs_Winback_log,
  Catalog_Winback_log_lag1 = data_exp$Catalogs_Winback_log_lag  ,
  Catalog_ExistCust_log_lag1 = data_exp$Catalogs_ExistCust_log_lag
)
mod5 <- arima(data_exp$Sales, xreg = Covars5, order = c(0,0,0), seasonal = list(order = c(1,0,0), perio
summary(mod5)
```

```
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars5)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_Winback_log
##       0.5544  1937.6341   283.1114     634.5609              93.7414
## s.e.  0.1541   630.7149   145.9333     305.5960              47.6094
##       Catalog_Winback_log_lag1  Catalog_ExistCust_log_lag1
##                       125.6633                     81.4510
## s.e.                   48.8331                     72.1066
##
## sigma^2 estimated as 205280:  log likelihood = -311.14,  aic = 636.28
##
## Training set error measures:
```

```
## Warning in trainingaccuracy(object, test, d, D): test elements must be within
## sample
```

```
##                ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

```
coeftest(mod5)
```

```
##
## z test of coefficients:
##
##                             Estimate Std. Error z value Pr(>|z|)
## sar1                         0.55443    0.15411  3.5977 0.000321 ***
## intercept                 1937.63413  630.71486  3.0721 0.002125 **
## Month_log                  283.11135  145.93325  1.9400 0.052379 .
## Portals_log                634.56087  305.59597  2.0765 0.037851 *
## Catalog_Winback_log         93.74140   47.60941  1.9690 0.048957 *
## Catalog_Winback_log_lag1   125.66329   48.83312  2.5733 0.010073 *
```

```
## Catalog_ExistCust_log_lag1   81.45096   72.10659  1.1296 0.258649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Synergy We tested for synergy in medias. Apparently, newsletter plays a complementary role to catalogs. The magnitude of the impacts on sales caused by investments in catalog (Existing Customers and Winback) firmly depends on the value spend on newsletters.

```
Covars <- cbind(
  Month_log = data_exp$Months_log,
  Portals_log = data_exp$Portals_log,
  Catalog_Winback_log= data_exp$Catalogs_Winback_log,
  Winback_Newsletter_log_lag = data_exp$Catalogs_Winback_log_lag * data_exp$Newslettr_log_lag ,
  ExistCust_Newsletter_log_lag = data_exp$Catalogs_ExistCust_log_lag * data_exp$Newsletter_log_lag
)
```

```
## Warning: Unknown or uninitialised column: `Newslettr_log_lag`.
```

```
modfin <- arima(data_exp$Sales, xreg = Covars, order = c(0,0,0), seasonal = list(order = c(1,0,0), peri

summary(modfin)
```

```
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_Winback_log
##       0.7686  2086.9337   243.4285     527.8441              71.6431
## s.e.  0.1004   543.9303   126.6104     258.8950              58.4412
##       ExistCust_Newsletter_log_lag
##                            53.0194
## s.e.                       20.0398
##
## sigma^2 estimated as 174025:  log likelihood = -310.91,  aic = 633.82
##
## Training set error measures:
```

```
## Warning in trainingaccuracy(object, test, d, D): test elements must be within
## sample
```

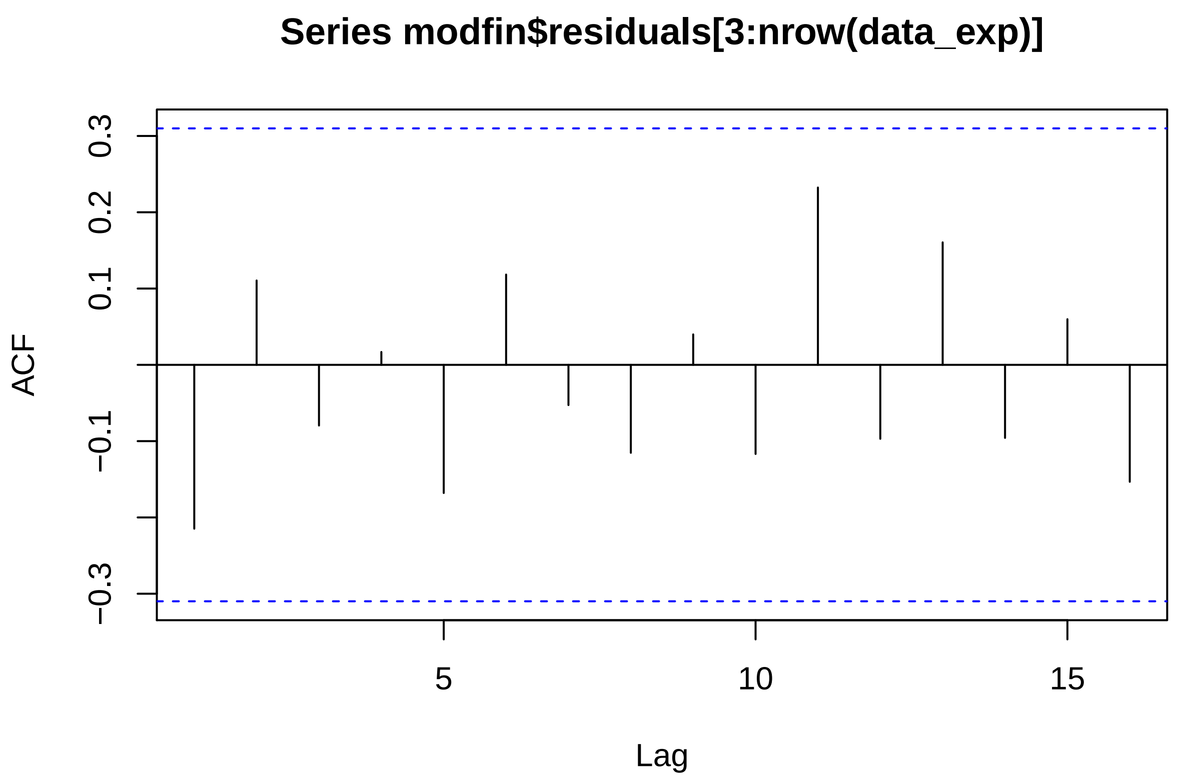
```
##               ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

```
coeftest(modfin)
```

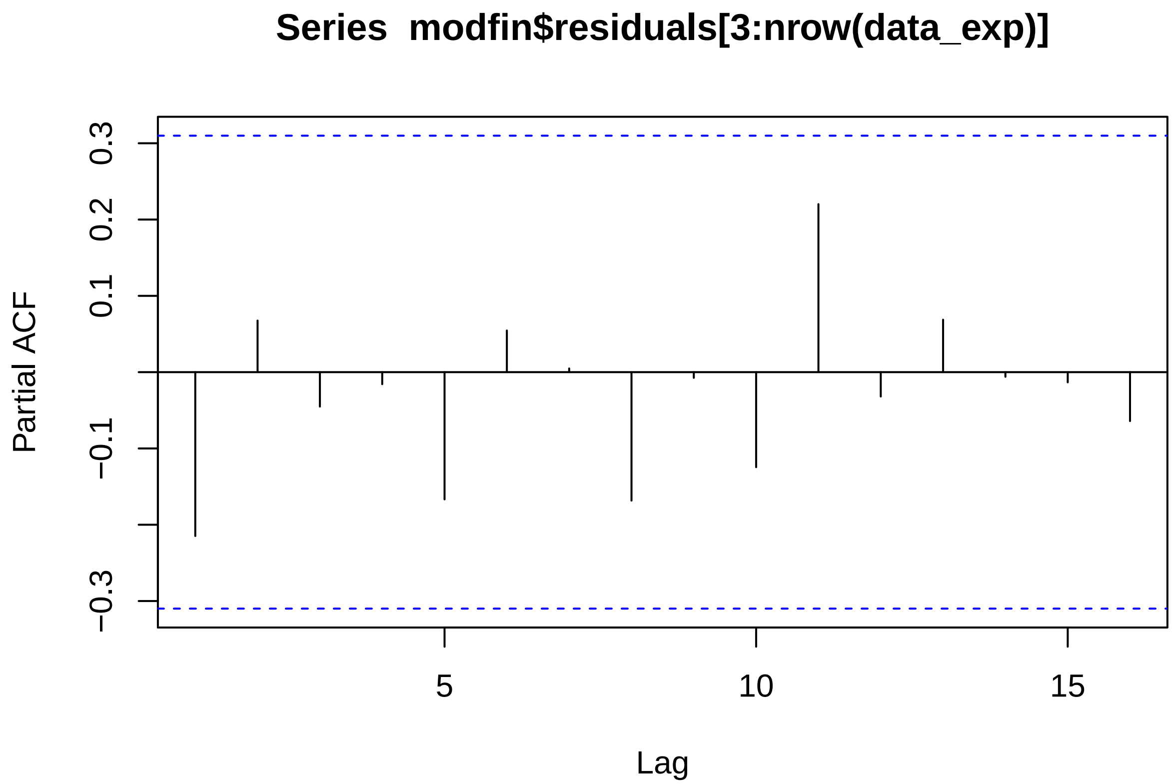
```
##
## z test of coefficients:
##
##                                Estimate Std. Error z value  Pr(>|z|)
```

```
## sar1                          0.76856    0.10042  7.6535 1.956e-14 ***
## intercept                  2086.93371  543.93029  3.8368 0.0001247 ***
## Month_log                   243.42846  126.61043  1.9227 0.0545231 .
## Portals_log                 527.84405  258.89505  2.0388 0.0414666 *
## Catalog_Winback_log          71.64307   58.44123  1.2259 0.2202365
## ExistCust_Newsletter_log_lag  53.01936   20.03981  2.6457 0.0081522 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

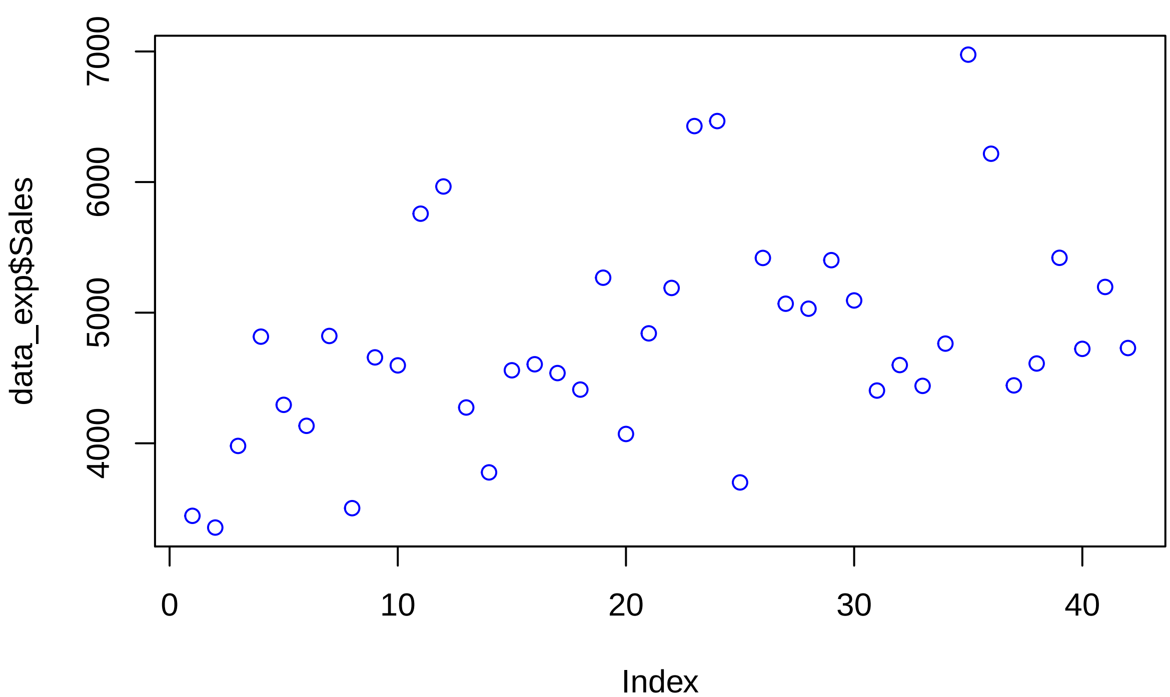
```r
acf(modfin$residuals[3:nrow(data_exp)])
```




```r
pacf(modfin$residuals[3:nrow(data_exp)])
```




```r
plot(data_exp$Sales, col = "blue")
lines(fitted(modfin))
```




```r
stg2 <- stargazer(mod4a,mod4b, mod5, modfin)
```

```
##
```

```
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars4a)
##
## Coefficients:
##         sar1  intercept  Month_sqrt  Portals_sqrt  Catalog_ExistCust_sqrt
##       0.7791  3726.8747    127.3000      341.4032                 -4.4755
## s.e.  0.1020   592.5257     72.8715      232.1215                 18.8201
##       Catalog_Winback_sqrt
##                   -16.0028
## s.e.               12.7277
##
## sigma^2 estimated as 215468:  log likelihood = -323.09,  aic = 658.18
##
## Training set error measures:
##              ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars4b)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_ExistCust_log
##       0.6954  4025.6948   258.9134     543.1642              -152.3399
## s.e.  0.1197   682.9736   125.3163     288.6456                88.9325
##       Catalog_Winback_log
##                  -21.4352
## s.e.              39.1976
##
## sigma^2 estimated as 212860:  log likelihood = -321.2,  aic = 654.39
##
## Training set error measures:
##              ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars5)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_Winback_log
##       0.5544  1937.6341   283.1114     634.5609              93.7414
## s.e.  0.1541   630.7149   145.9333     305.5960              47.6094
##       Catalog_Winback_log_lag1  Catalog_ExistCust_log_lag1
##                       125.6633                     81.4510
## s.e.                   48.8331                     72.1066
##
## sigma^2 estimated as 205280:  log likelihood = -311.14,  aic = 636.28
##
## Training set error measures:
##              ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
```

```
## Call:
## arima(x = data_exp$Sales, order = c(0, 0, 0), seasonal = list(order = c(1, 0,
##     0), period = 12), xreg = Covars)
##
## Coefficients:
##         sar1  intercept  Month_log  Portals_log  Catalog_Winback_log
##       0.7686  2086.9337   243.4285     527.8441              71.6431
## s.e.  0.1004   543.9303   126.6104     258.8950              58.4412
##       ExistCust_Newsletter_log_lag
##                            53.0194
## s.e.                       20.0398
##
## sigma^2 estimated as 174025:  log likelihood = -310.91,  aic = 633.82
##
## Training set error measures:
##               ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Dec 08, 2020 - 20:23:13
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{4}{c}{\textit{Dependent variable:}} \\
## \cline{2-5}
## \\[-1.8ex] &  &  &  &  \\
## \\[-1.8ex] & (1) & (2) & (3) & (4)\\
## \hline \\[-1.8ex]
##  sar1 & 0.779$^{***}$ & 0.695$^{***}$ & 0.554$^{***}$ & 0.769$^{***}$ \\
##   & (0.102) & (0.120) & (0.154) & (0.100) \\
##   & & & & \\
##  intercept & 3,726.875$^{***}$ & 4,025.695$^{***}$ & 1,937.634$^{***}$ & 2,086.934$^{***}$ \\
##   & (592.526) & (682.974) & (630.715) & (543.930) \\
##   & & & & \\
##  Month\_sqrt & 127.300$^{*}$ &  &  &  \\
##   & (72.872) &  &  &  \\
##   & & & & \\
##  Portals\_sqrt & 341.403 &  &  &  \\
##   & (232.122) &  &  &  \\
##   & & & & \\
##  Catalog\_ExistCust\_sqrt & $-$4.476 &  &  &  \\
##   & (18.820) &  &  &  \\
##   & & & & \\
##  Catalog\_Winback\_sqrt & $-$16.003 &  &  &  \\
##   & (12.728) &  &  &  \\
##   & & & & \\
##  Month\_log &  & 258.913$^{**}$ & 283.111$^{*}$ & 243.428$^{*}$ \\
##   &  & (125.316) & (145.933) & (126.610) \\
##   & & & & \\
##  Portals\_log &  & 543.164$^{*}$ & 634.561$^{**}$ & 527.844$^{**}$ \\
##   &  & (288.646) & (305.596) & (258.895) \\
```

```
##   & & & & \\
##  Catalog\_ExistCust\_log &  & $-$152.340$^{*}$ &  &  \\
##   &  & (88.932) &  &  \\
##   & & & & \\
##  Catalog\_Winback\_log &  & $-$21.435 & 93.741$^{**}$ & 71.643 \\
##   &  & (39.198) & (47.609) & (58.441) \\
##   & & & & \\
##  Catalog\_Winback\_log\_lag1 &  &  & 125.663$^{**}$ &  \\
##   &  &  & (48.833) &  \\
##   & & & & \\
##  Catalog\_ExistCust\_log\_lag1 &  &  & 81.451 &  \\
##   &  &  & (72.107) &  \\
##   & & & & \\
##  ExistCust\_Newsletter\_log\_lag &  &  &  & 53.019$^{***}$ \\
##   &  &  &  & (20.040) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Observations & 42 & 42 & 41 & 41 \\
## Log Likelihood & $-$323.090 & $-$321.196 & $-$311.139 & $-$310.909 \\
## $\sigma^{2}$ & 215,468.000 & 212,859.700 & 205,280.500 & 174,024.600 \\
## Akaike Inf. Crit. & 660.180 & 656.393 & 638.278 & 635.817 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{4}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
# Long run Elasticity
Data_means <- data %>% summarise_all(mean, na.rm=T)
Coeffic <- coef(modfin)

LR_elastic <- cbind(
  Winback = (Coeffic["Catalog_Winback_log"]+ Coeffic["Winback_Newsletter_log_lag"]*log(Data_means$Newsle
  Catalog_ExistCust = Coeffic["ExistCust_Newsletter_log_lag"]*log(Data_means$Newsletter)/Data_means$Sale
  Portals = Coeffic["Portals_log"]/Data_means$Sales,
  Newsletter = (Coeffic["ExistCust_Newsletter_log_lag"]*log(Data_means$Catalogs_ExistCust) + Coeffic["W
  )
```