Retail_auto

By Lily

same model, new data (2014-2019) but without transforming Google Trend data

```
new_merged <- read.csv("merged_new.csv", header = TRUE)
head(new_merged)
```

```
##     date  sales suv_index insurance_index
## 1 3/2/14 90,488        65              76
## 2 4/6/14 87,959        62              70
## 3 5/4/14 93,239        60              69
## 4 6/1/14 86,715        59              72
## 5 7/6/14 91,275        63              70
## 6 8/3/14 92,624        65              73
```

```
sapply(new_merged, mode)
```

```
##            date           sales       suv_index insurance_index
##       "numeric"       "numeric"       "numeric"       "numeric"
```

```
#transform date to string then to date format
#new_merged <- transform(new_merged, sales=as.numeric(sales))
date<- as.Date(new_merged$date, format = "%m/%d/%y")
new_merged$date <- date
sapply(new_merged, mode)
```

```
##            date           sales       suv_index insurance_index
##       "numeric"       "numeric"       "numeric"       "numeric"
```

```
#transform data using Jake's method; for later use
new_merged$suv_ <- log(new_merged$suv_index/100)
new_merged$insurance_ <- log(new_merged$insurance_index/100)


class(new_merged$sales)
```

```
## [1] "factor"
```

```
sales <- as.character(new_merged$sales) #unfactor sales
sales <- as.numeric(gsub(",","", sales)) #remove commas
#log sales
new_merged$sales <- log(sales)

len <- length(new_merged$sales)
#lag -1
t1 <- new_merged$sales[2:len]
#lag -12
t12 <- new_merged$sales[13:len]

#new merged with lag-1, lag-12
```

```r
merged_with_lag <- new_merged[1: length(t12),]
merged_with_lag$lag_1 = t1[1: length(t12)]
merged_with_lag$lag_12 =  t12[1:length(t12)]

#tail(new_merged,20)
#tail(merged_with_lag, 12)
```

```r
reg0_new <- lm(sales~lag_1 + lag_12, data=merged_with_lag)
summary(reg0_new)
```

```
##
## Call:
## lm(formula = sales ~ lag_1 + lag_12, data = merged_with_lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.078754 -0.013794 -0.002936  0.011594  0.064926
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.10652    0.66673  -1.660   0.1039
## lag_1        0.11657    0.04984   2.339   0.0238 *
## lag_12       0.97638    0.05733  17.030   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02618 on 45 degrees of freedom
## Multiple R-squared:  0.8986, Adjusted R-squared:  0.8941
## F-statistic: 199.4 on 2 and 45 DF,  p-value: < 2.2e-16
```

```r
reg1_new<- lm(sales~lag_1 + lag_12 + suv_index + insurance_index, data=merged_with_lag)
summary(reg1_new)
```

```
##
## Call:
## lm(formula = sales ~ lag_1 + lag_12 + suv_index + insurance_index,
##     data = merged_with_lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.076019 -0.015460 -0.002533  0.014746  0.059252
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0924483  0.8207703   0.113   0.9108
## lag_1            0.0504584  0.0558860   0.903   0.3716
## lag_12           0.9338392  0.0580400  16.090   <2e-16 ***
## suv_index        0.0016089  0.0008640   1.862   0.0694 .
## insurance_index -0.0009476  0.0009830  -0.964   0.3404
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02517 on 43 degrees of freedom
## Multiple R-squared:  0.9104, Adjusted R-squared:  0.9021
## F-statistic: 109.3 on 4 and 43 DF,  p-value: < 2.2e-16
```

```
source("oosf.R")
```

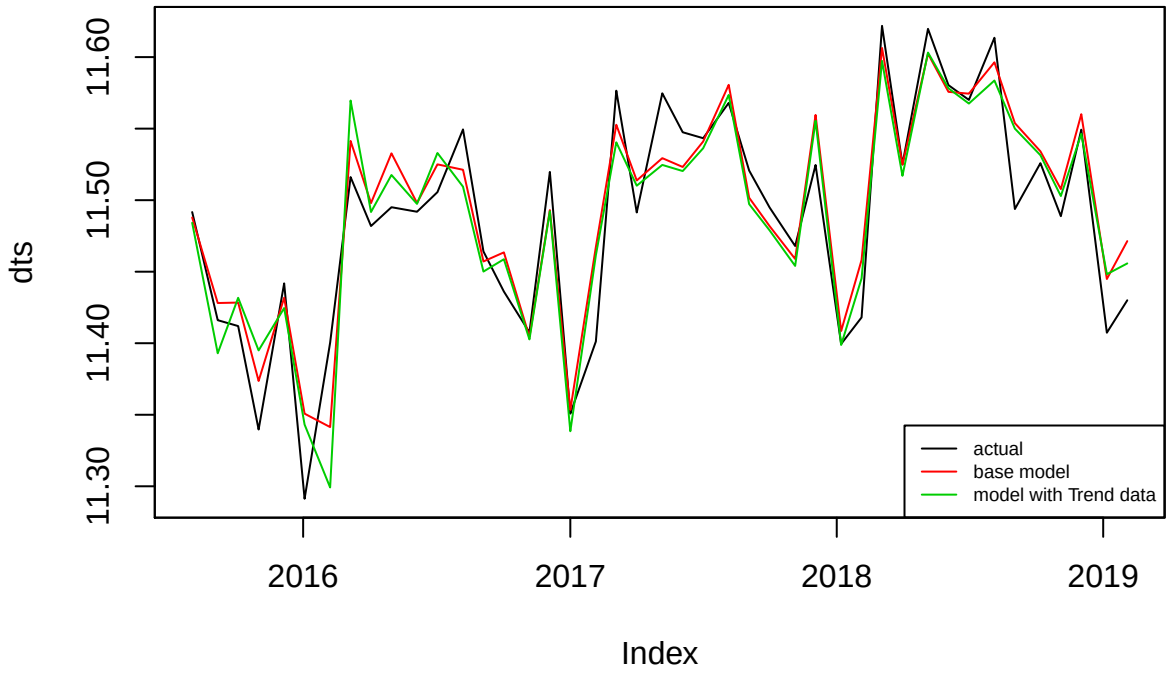
```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

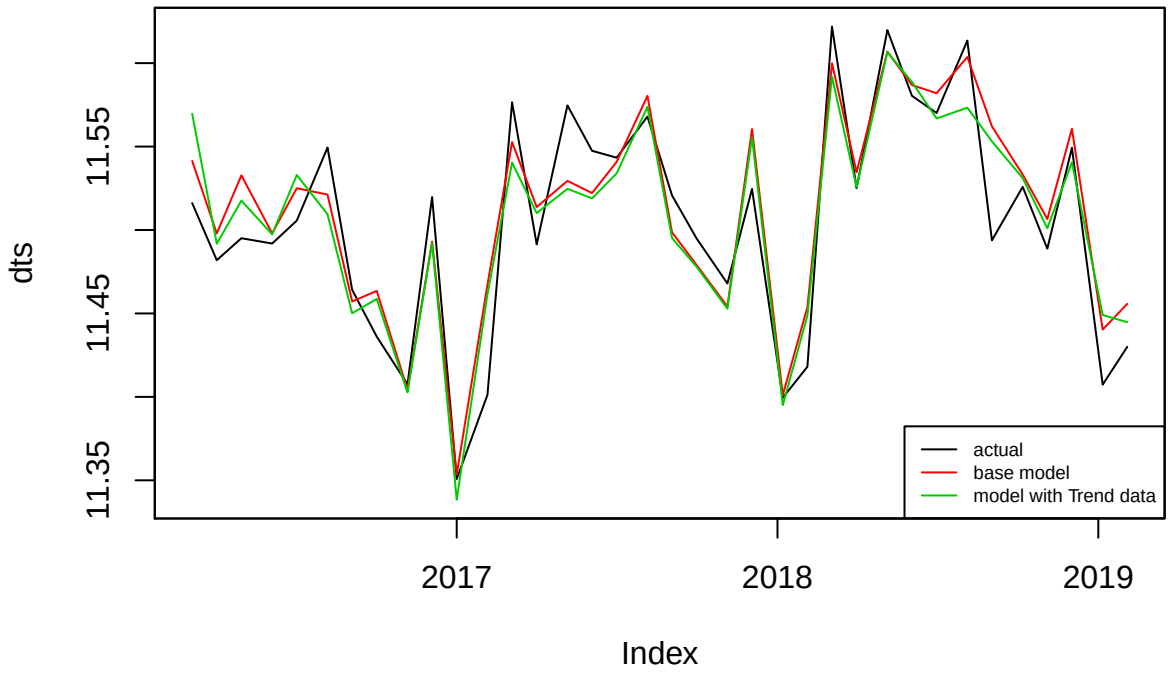
```
merged_zoo <- zoo(new_merged[,-1], as.Date(new_merged[,1]))
y_new <- merged_zoo$sales
x_new <- merged_zoo[,c(2,3)]
z_new <- OutOfSampleForecast12(y_new,x_new,17) #returns prediction from week18 on, rolling window
z_new_fixed <- OutOfSampleForecast12_fixed(y_new,x_new,24) #returns prediction based on 24 week fixed w

# overall fit
MaeReport(z_new)
```




```
##    mae.base  mae.trends   mae.delta
##  0.02215164  0.02470196 -0.11513004
```

```
MaeReport(z_new_fixed)
```




```
##    mae.base  mae.trends   mae.delta
##  0.02106448  0.02249844 -0.06807451
```

```
#MaeReport(z_new,"2007-12-01","2009-06-30")
```

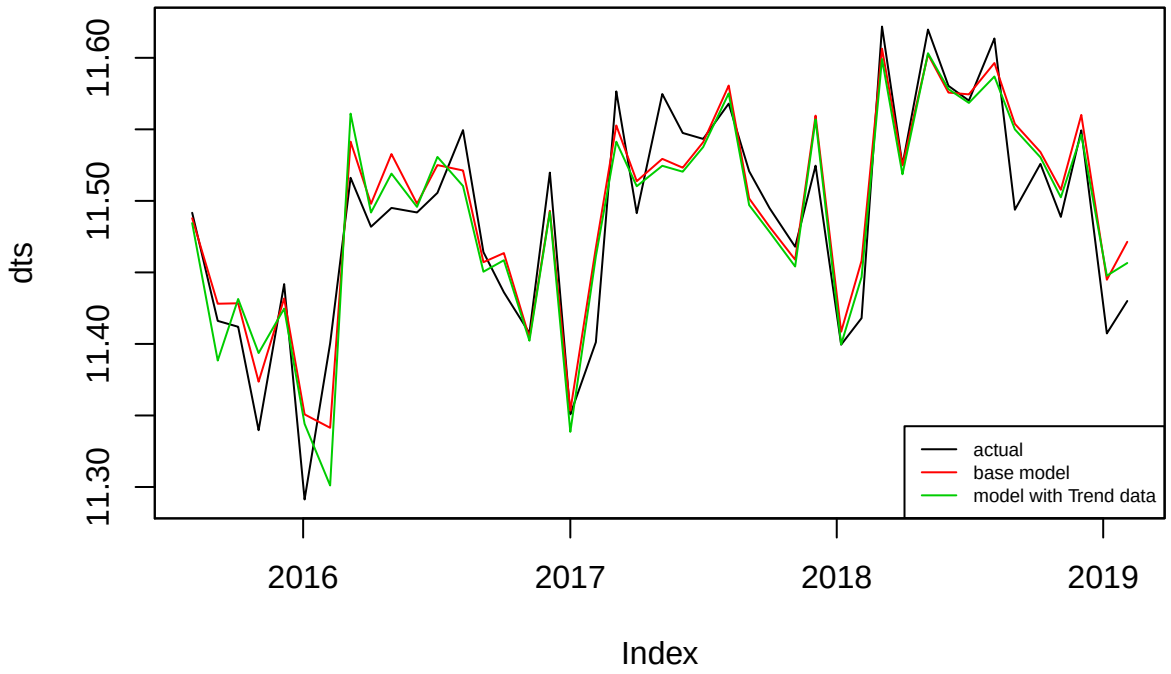
Transform Trend Data based on Jake's method

```
reg1_new<- lm(sales~lag_1 + lag_12 + suv_ + insurance_, data=merged_with_lag)
summary(reg1_new)
```

```
##
## Call:
## lm(formula = sales ~ lag_1 + lag_12 + suv_ + insurance_, data = merged_with_lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.076882 -0.015392 -0.002311  0.013975  0.059562
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.11327    0.84421   0.134   0.8939
## lag_1        0.05141    0.05640   0.912   0.3671
## lag_12       0.93655    0.05805  16.133   <2e-16 ***
## suv_         0.11364    0.06193   1.835   0.0734 .
## insurance_  -0.07014    0.07085  -0.990   0.3277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02527 on 43 degrees of freedom
## Multiple R-squared:  0.9097, Adjusted R-squared:  0.9013
## F-statistic: 108.3 on 4 and 43 DF,  p-value: < 2.2e-16
```

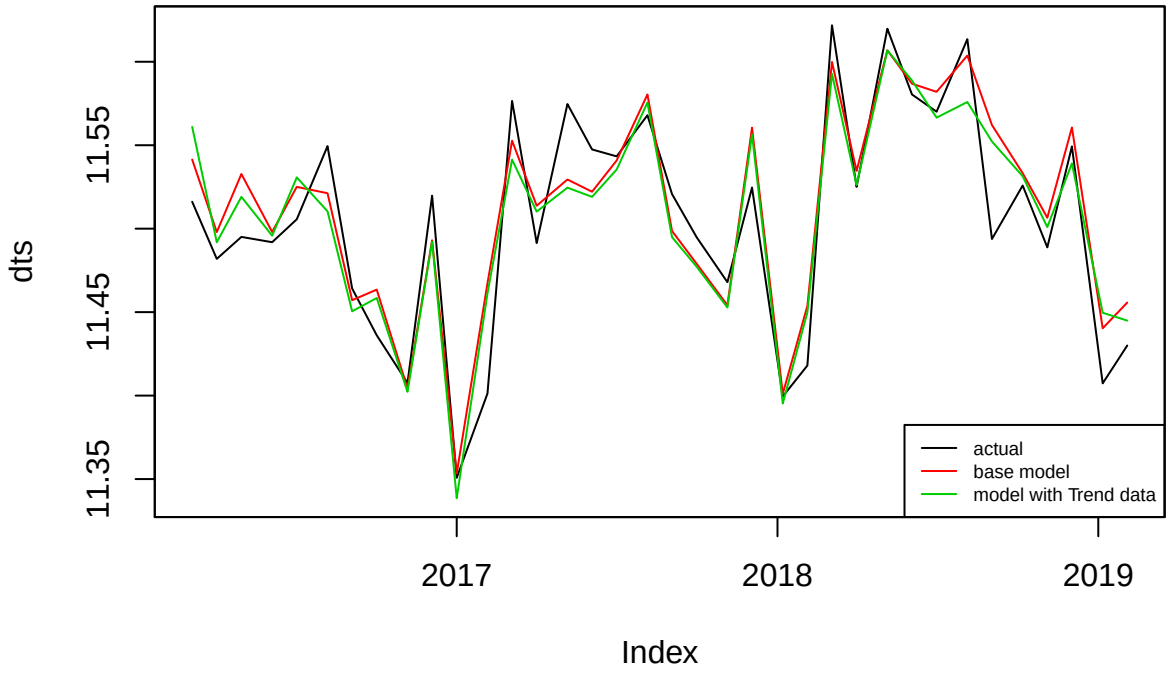
```
x_new_trans <- merged_zoo[,c(4,5)]
z_new_trans <- OutOfSampleForecast12(y_new,x_new_trans,17) #returns prediction from week18 on, rolling
z_new_trans_fixed <- OutOfSampleForecast12_fixed(y_new,x_new_trans,24) #returns prediction based on 24-

MaeReport(z_new_trans)
```




```
##    mae.base  mae.trends   mae.delta
##  0.02215164  0.02431243 -0.09754551
```

```
MaeReport(z_new_trans_fixed)
```




```
##    mae.base  mae.trends   mae.delta
```

```
##  0.02106448  0.02218196 -0.05305014
```

Paper's method

```
library(dyn)
dat <- read.csv("merged.csv")
d <- zoo(dat[,-1],as.Date(dat[,1]))
y <- log(d$sales)

# baseline model
reg0 <- dyn$lm(y~lag(y,-1)+lag(y,-12))
summary(reg0)
```

```
##
## Call:
## lm(formula = dyn(y ~ lag(y, -1) + lag(y, -12)))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.209554 -0.034684  0.002482  0.040477  0.220976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.67266    0.76355   0.881 0.381117
## lag(y, -1)   0.64345    0.07332   8.776 3.59e-13 ***
## lag(y, -12)  0.29565    0.07282   4.060 0.000118 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07985 on 76 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.7185, Adjusted R-squared:  0.7111
## F-statistic:    97 on 2 and 76 DF,  p-value: < 2.2e-16
```

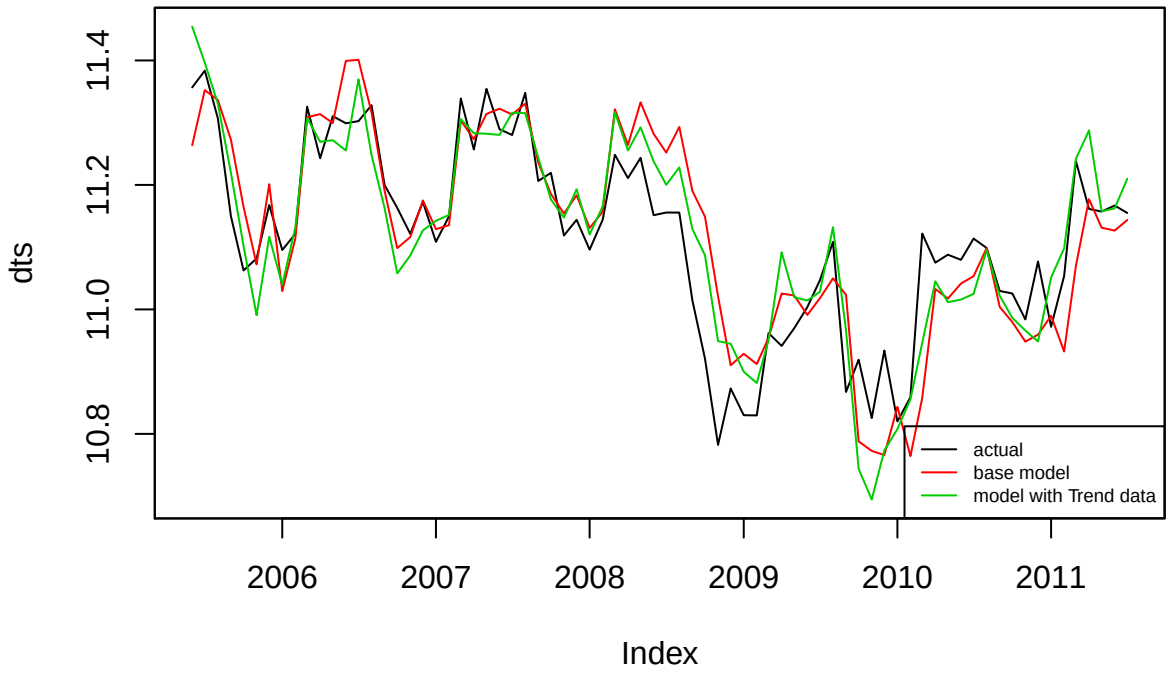
```
reg1 <- dyn$lm(y~lag(y,-1)+lag(y,-12)+suvs+insurance,data=dat)
summary(reg1)
```

```
##
## Call:
## lm(formula = dyn(y ~ lag(y, -1) + lag(y, -12) + suvs + insurance),
##     data = dat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.161327 -0.043774  0.002998  0.036651  0.159219
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.45798    0.78438  -0.584 0.561081
## lag(y, -1)   0.61947    0.06318   9.805 5.09e-15 ***
## lag(y, -12)  0.42865    0.06535   6.559 6.45e-09 ***
## suvs         1.05721    0.16686   6.336 1.66e-08 ***
## insurance   -0.52966    0.15206  -3.483 0.000835 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06509 on 74 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.8179, Adjusted R-squared:  0.808
## F-statistic: 83.08 on 4 and 74 DF,  p-value: < 2.2e-16
```

```
y <- log(d$sales)
x <- d[,c(2,3)]
z <- OutOfSampleForecast12(y,x,17)
z_fixed <- OutOfSampleForecast12_fixed(y,x,24)

# overall fit
MaeReport(z)
```




```
##   mae.base mae.trends  mae.delta
## 0.06343984 0.05667658 0.10660890
```

```
MaeReport(z,"2007-12-01","2009-06-30")
```


```
##   mae.base mae.trends  mae.delta
## 0.08869325 0.06965812 0.21461753
```

```
MaeReport(z_fixed)
```


```
##   mae.base mae.trends  mae.delta
## 0.06351543 0.05581327 0.12126443
```

```
MaeReport(z_fixed,"2007-12-01","2009-06-30")
```


```
##   mae.base mae.trends  mae.delta
## 0.07788803 0.06328151 0.18753225
```

```
plot(z,plot.type="sin",col=c(1,1,"gray40"),lty=c(1,2,1),main="Motor Vehicles and Parts",ylab="log(mvp)"
legend("topright",c("actual","base","trends"),lty=c(1,2,1),col=c(1,1,"gray40"),lwd=c(1.5,1,1))
legend("bottomleft",c("MAE improvement","Overall = 10.5%","During recession = 21.5%"))
```