Auto different data same (similar) model

By Lily

replicate: same model, same data but without transforming Google Trend data

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

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

head(new_merged)
```

```
##         date sales suv_index insurance_index
## 1 2003-02-14    36        65              76
## 2 2004-06-14    29        62              70
## 3 2005-04-14    44        60              69
## 4 2006-01-14    28        59              72
## 5 2007-06-14    39        63              70
## 6 2008-03-14    42        65              73
```

```
new_merged$suv_ <- log(new_merged$suv_index/100)
new_merged$insurance_ <- log(new_merged$insurance_index/100)

#log sales
new_merged$sales <- log(new_merged$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
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)
```

```
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
## -3.1852 -0.2290  0.2754  0.5737  1.0429
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.7526     0.6553   4.201 0.000124 ***
## lag_1         0.0346     0.1480   0.234 0.816216
## lag_12        0.1126     0.1370   0.822 0.415480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9326 on 45 degrees of freedom
## Multiple R-squared:  0.01582,    Adjusted R-squared:  -0.02793
## F-statistic: 0.3616 on 2 and 45 DF,  p-value: 0.6986
```

```
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
## -2.9167 -0.4010  0.1072  0.5144  1.1968
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.370416   1.356944   3.958 0.000279 ***
## lag_1           -0.089251   0.150185  -0.594 0.555444
## lag_12          -0.005522   0.145536  -0.038 0.969908
## suv_index       -0.040079   0.031302  -1.280 0.207275
## insurance_index  0.014131   0.036792   0.384 0.702809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8919 on 43 degrees of freedom
## Multiple R-squared:   0.14,  Adjusted R-squared:  0.05995
## F-statistic: 1.749 on 4 and 43 DF,  p-value: 0.1568
```

```
library(here)
```

```
## here() starts at /Users/linatian/Desktop/msd final project/Sales
```

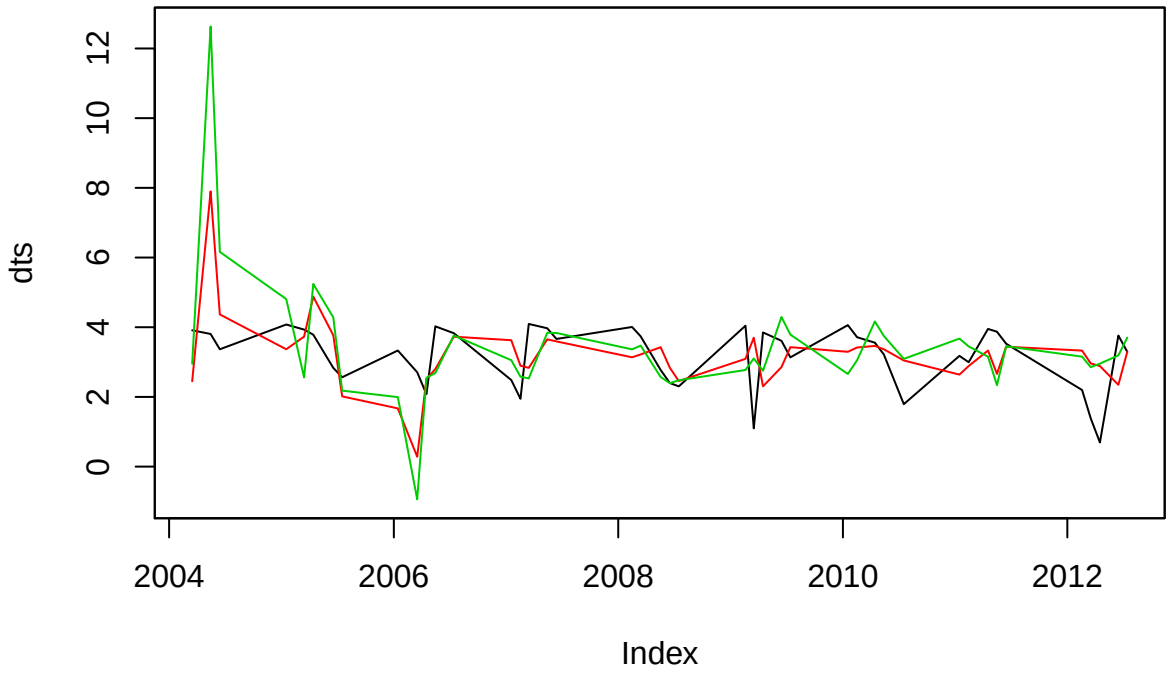
```
source("oosf.R")
```

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

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

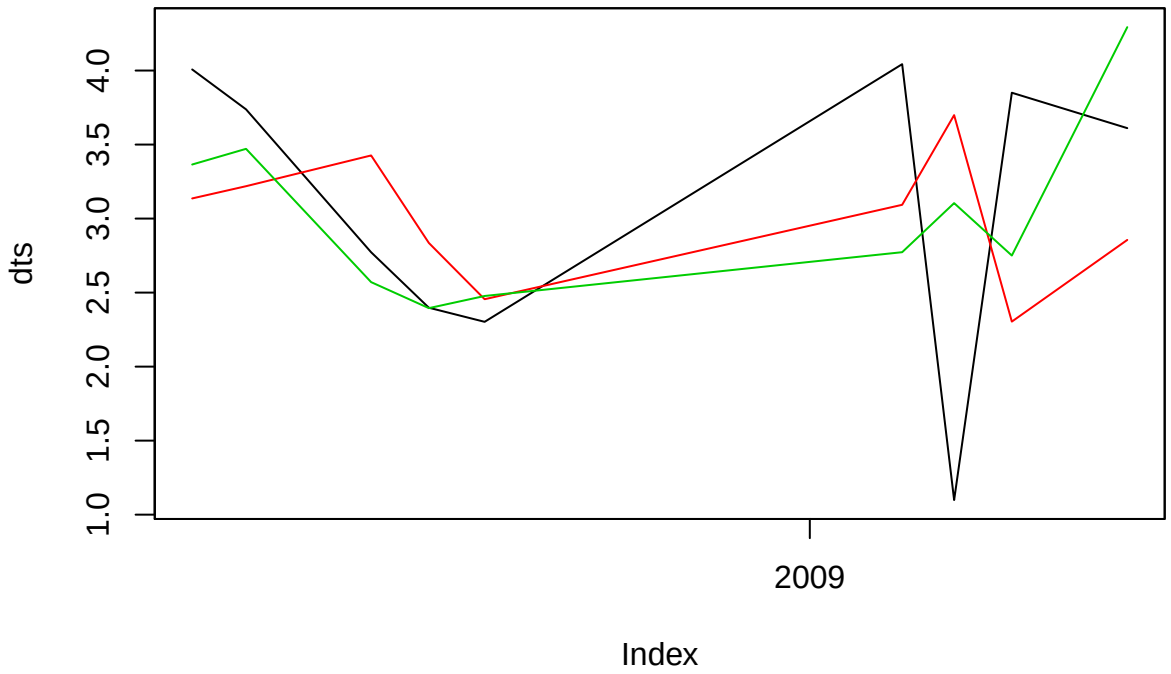
```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
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)
# overall fit
MaeReport(z_new)
```




```
##   mae.base mae.trends  mae.delta
##  0.9266071  1.1261051 -0.2152994
```

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



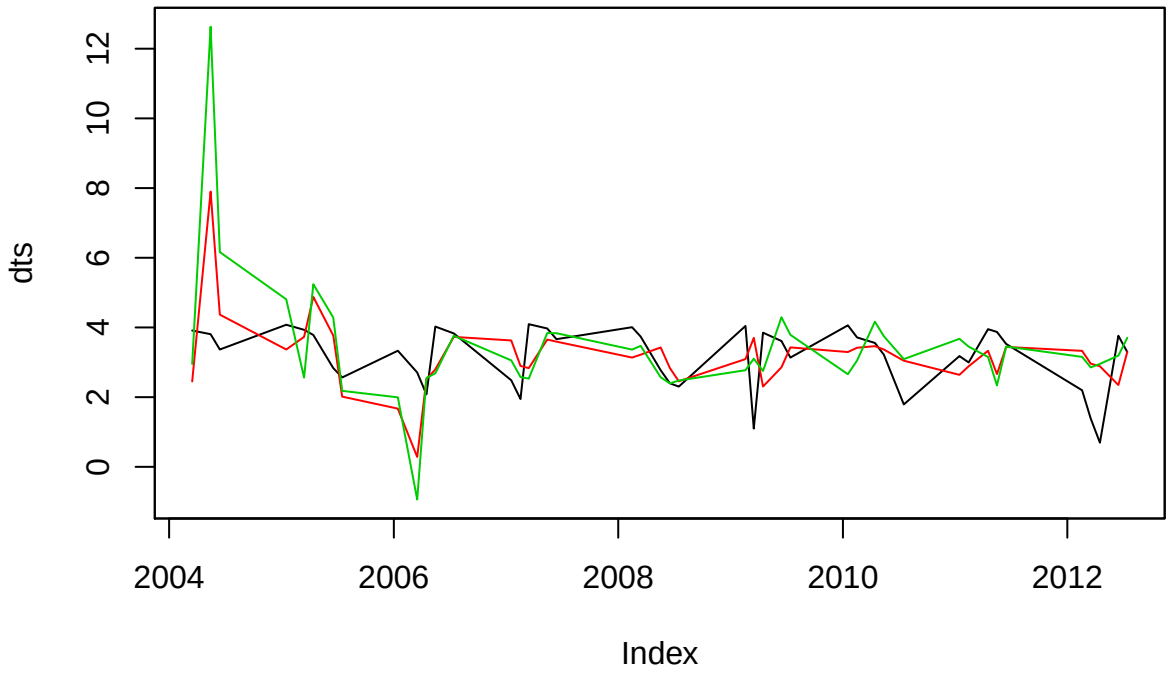
```
##   mae.base mae.trends  mae.delta
##  0.9428646  0.7050709  0.2522035
```

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
## -2.9132 -0.3983  0.1390  0.5321  1.2042
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.831382   0.711575   3.979 0.000261 ***
## lag_1       -0.081357   0.150394  -0.541 0.591327
## lag_12       0.002947   0.146102   0.020 0.984000
## suv_        -2.646104   2.231765  -1.186 0.242269
## insurance_   0.840720   2.652086   0.317 0.752775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8959 on 43 degrees of freedom
## Multiple R-squared:  0.1321, Adjusted R-squared:  0.05142
## F-statistic: 1.637 on 4 and 43 DF,  p-value: 0.1824
```

```
source("oosf.R")
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)
# overall fit
MaeReport(z_new)
```




```
##   mae.base mae.trends  mae.delta
##  0.9266071  1.1261051 -0.2152994
```

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


```
##   mae.base mae.trends  mae.delta
##  0.9428646  0.7050709  0.2522035
```

Paper's method

```
library(dyn)
dat <- read.csv("merged_paper.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)
```

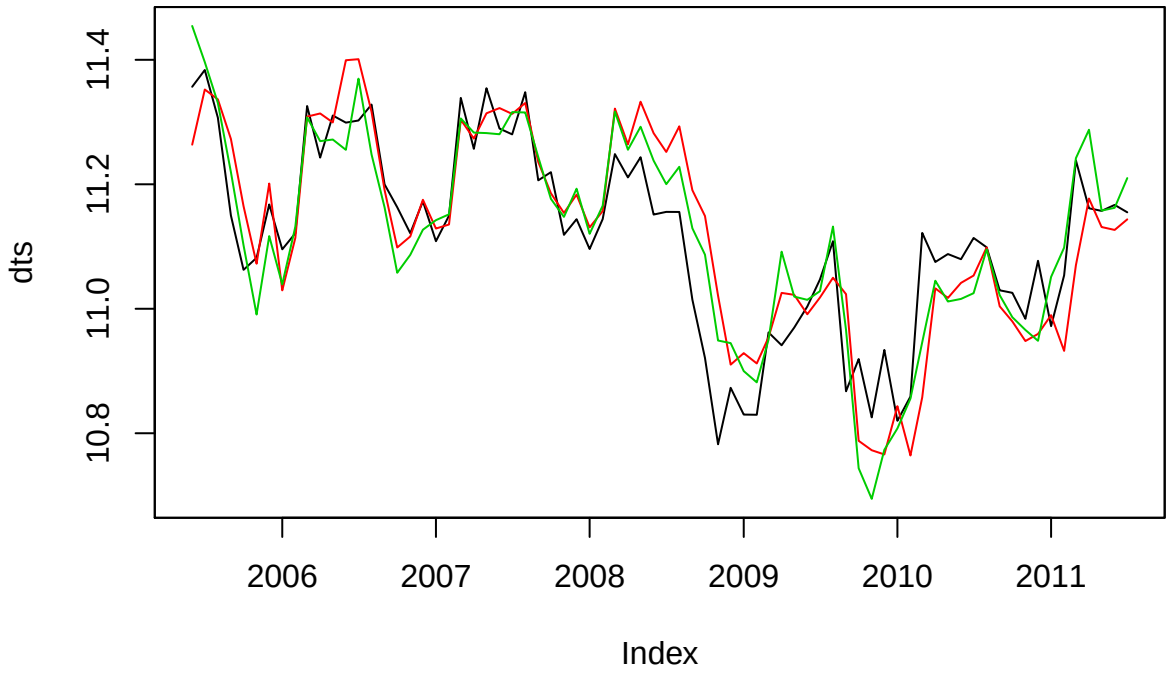
```
##
## Call:
## lm(formula = dyn(y ~ lag(y, -1) + lag(y, -12)))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.209554 -0.034684  0.002482  0.040477  0.220976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.67266    0.76355   0.881 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)
# 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
```

```
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%"))
```