

Corr Coef

Stephanie Yang jy2777

12/3/2017

```
#Function ccf computes the cross-correlation or cross-covariance of two univariate series
#finds absolute max of CCF function
Find_Abs_Max_CCF<- function(a,b,lag.max = lag.max)
{
  #stores output of function ccf in object f
  f <- ccf(a, b, plot = FALSE, na.action = na.omit)
  #correlation coefficient
  cor <- f$acf[,,1]
  #absolute value of correlation coefficient
  abscor <- abs(f$acf[,,1])
  lag <- f$lag[,,1]
  #result is cor and lag, in which to search the absolute max corr
  res <- data.frame(cor,lag)
  #absolute result is abscor and lag
  absres <- data.frame(abscor,lag)
  #in cor+lag which is the absolute max cor+lag
  absres_max <- res[which.max(absres$abscor),]
  return(absres_max)
}
```

We found highest correlation coefficients for all companies. Here is the code for Chipotle. For others, we just changed the variable names and file sources.

CHIPOTLE

```
setwd("~/Desktop/Homework/Statistical Methods/Project/datasets")
#transpose t()
chipotle.metrics <- t(read.csv("chipotle_metrics22.csv", header=FALSE))

#labels columns
colnames(chipotle.metrics)<- c("Date","Likes (Total) FB","Comments (Total) FB","Shares (Total) FB","Rea

#removes duplicate row
chipotle.metrics1 <- chipotle.metrics[-1,]

##Cleaning the metrics sheet
#1) removes space in column titles
colnames(chipotle.metrics1) <- gsub(" ","",colnames(chipotle.metrics1))

#2) removes % symbol of column 10
chipotle.metrics1[,c(11,15,19,20,21,26,27,28,31,33,38,39,43,44,52,59,60,61,69,71,72,73)] <- as.numeric(
chipotle.metrics1 <- as.data.frame(chipotle.metrics1)
class(chipotle.metrics1)

## [1] "data.frame"

#3) removes comma separator for thousands, except for date column which is type character not numeric
#gsub to replace "," with "", and then convert the string to numeric using as.numeric
chipotle.metrics1[,2:73] <- lapply(chipotle.metrics1[,2:73], function(x) as.numeric(gsub(",","", as.cha
```

```

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
##Transforms Monthly to Quarterly Data:
library("lubridate")

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##      date

#creates a year and quarter column per row
chipotle.metrics1$Date <- ymd(chipotle.metrics1$Date)

## Warning in as.POSIXlt.POSIXct(x, tz): unknown timezone 'zone/tz/2017c.1.0/'
## zoneinfo/America/New_York'

chipotle.metrics1$year = year(chipotle.metrics1$Date)
chipotle.metrics1$quarter = quarter(chipotle.metrics1$Date)

#aggregates quarters of same year and takes their sum (sales are also sums) : HOW TO DO IT WITH AGGREGA
library("reshape2")
chipotle.metrics2 <- melt(chipotle.metrics1[,2:75], id=c("quarter", "year"))
chipotle.metrics2 <- dcast(chipotle.metrics2, year + quarter ~ variable, fun.aggregate = sum)

#aggregate(chipotle.metrics1$year ~.,chipotle.metrics1, FUN = mean, na.action = na.omit)
write.csv(chipotle.metrics2,file="chipotle_vizmetrics.csv")
*****
#quarterly sales data, data points from CapitalIQ over 2 years
chipotle.sales <- read.csv("chipotle_sales.csv")

#cleaning sales sheet: subsets,transposes and reformats data
chipotle.sales2 <- chipotle.sales[c(10,13),46:54]
chipotle.sales2 <- t(chipotle.sales2)
colnames(chipotle.sales2)<- c("Date","Sales")
typeof(chipotle.sales2[,2])

## [1] "character"

chipotle.sales2[,2] <- as.numeric(gsub(",","",chipotle.sales2[,2]))
chipotle.sales2 <- as.data.frame(chipotle.sales2)

#binds rows of Pearson cor and lag coefficients per metric
values_chipotle <- rbind(a=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Likes(Total)FB`),
b=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Comments(Total)FB`),
c=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Shares(Total)FB`),
d=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ReactionsPerPostFB`),
e=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Impressions(Total)FB`),
f=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Impressions-viral(Total)FB`),
g=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$ImpressionsperActionFB),
h=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$ImpressionsperPostFB),
i=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Actions(Total)FB`),
j=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsperImpression(%)FB`),
k=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$FansFB),
l=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$PageScoreFB),
m=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Posts(Total)FB`),

```

```

n=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`AverageEngagementPerPost(%)FB`),
o=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerEngagerFB`),
p=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerPostFB`),
q=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`UniqueEngagedAudienceFB`),
r=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`AverageClickEngagementPerPost(%)FB`),
s=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%EngagedAudienceFB`),
t=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%NewEngagersFB`),
u=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Posts(CategoryTotal)FB`),
v=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Clicks(Total)FB`),
w=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`NativeVideoViews(LifetimeTotal)FB`),
x=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Video(TotalActions)FB`),
y=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%SharesFB`),
z=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%CommentsFB`),
aa=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%LikesFB`),
bb=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Likes(Total)TW`),
cc=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Retweets(Total)TW`),
dd=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`TotalActions(%Followers)TW`),
ee=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Actions(Total)TW`),
ff=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`AverageEngagementPerTweet(%)TW`),
gg=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`FollowersTW`),
hh=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Tweets(Total)TW`),
ii=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Clicks(Total)TW`),
jj=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`RetweeterFollowers(Average)TW`),
kk=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%RetweetsTW`),
ll=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%LikesTW`),
mm=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerTweetTW`),
nn=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`UniqueEngagedAudienceTW`),
oo=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`UniqueRetweetersTW`),
pp=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%EngagedAudienceTW`),
qq=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%NewEngagersTW`),
rr=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Reactions(Total)FB`),
ss=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerEngagerTW`),
tt=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`FollowersIG`),
uu=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Likes(Total)IG`),
vv=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Comments(Total)IG`),
ww=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Actions(Total)IG`),
xx=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Media(Total)IG`),
yy=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`AverageEngagementPerMedia(%)IG`),
zz=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerMediaIG`),
ab=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Photo(Total)IG`),
cd=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Video(Total)IG`),
ef=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsPerEngagerIG`),
gh=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`UniqueEngagedAudienceIG`),
ij=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`LikesPerMediaIG`),
kl=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%LikesIG`),
mn=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%EngagedAudienceIG`),
op=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%NewEngagersIG`),
qr=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Videos(Total)YT`),
st=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Likes(Total)YT`),
uv=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Dislikes(Total)YT`),
wx=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Comments(Total)YT`),
yz=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Views(Total)YT`),
aaa=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`Actions(Total)YT`),

```

```

bbb=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$SubscribersYT),
ccc=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`AverageEngagementPerVideo(%)YT`),
ddd=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$ActionsPerVideoYT),
eee=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%LikesYT`),
fff=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%CommentsYT`),
ggg=Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`%ViewsYT`)

#binds rows of Metrics into a dataframe
MetricNames <- rbind(a="Likes (Total) FB",b="Comments (Total) FB",c="Shares (Total) FB",
  d="Reactions Per Post FB",e="Impressions (Total)FB",
  f="Impressions-viral (Total) FB",g="Impressions per Action FB",
  h="Impressions per Post FB",i="Actions (Total) FB",
  j="Actions per Impression (%) FB",k="Fans FB",l="Page Score FB",
  m="Posts (Total) FB",n="Average Engagement Per Post (%) FB",
  o="Actions Per Engager FB",p="Actions Per Post FB",
  q="Unique Engaged Audience FB",
  r="Average Click Engagement Per Post (%) FB",
  s="% Engaged Audience FB",t="% New Engagers FB",
  u="Posts (Category Total) FB",v="Clicks (Total) FB",
  w="Native Video Views (Lifetime Total) FB",x="Video (Total Actions) FB",
  y="% Shares FB",z="% Comments FB",aa="% Likes FB",
  bb="Likes (Total) TW",cc="Retweets (Total) TW",
  dd="Total Actions (%) Followers TW",ee="Actions (Total) TW",
  ff="Average Engagement Per Tweet (%) TW",gg="Followers TW",
  hh="Tweets (Total) TW",ii="Clicks (Total) TW",
  jj="Retweeter Followers (Average) TW",kk="% Retweets TW",
  ll="% Likes TW",mm="Actions Per Tweet TW",
  nn="Unique Engaged Audience TW",oo="Unique Retweeters TW",
  pp="% Engaged Audience TW",qq="% New Engagers TW",
  rr="Reactions (Total) FB",ss="Actions Per Engager TW",
  tt="Followers IG",uu="Likes (Total) IG",vv="Comments (Total) IG",
  ww="Actions (Total) IG",xx="Media (Total) IG",
  yy="Average Engagement Per Media (%) IG",zz="Actions Per Media IG",
  ab="Photo (Total) IG",cd="Video (Total) IG",
  ef="Actions Per Engager IG",gh="Unique Engaged Audience IG",
  ij="Likes Per Media IG",kl="% Likes IG",mn="% Engaged Audience IG",
  op="% New Engagers IG",qr="Videos (Total) YT",
  st="Likes (Total) YT",uv="Dislikes (Total) YT",
  wx="Comments (Total) YT",yz="Views (Total) YT",
  aaa="Actions (Total) YT",bbb="Subscribers YT",
  ccc="Average Engagement Per Video (%) YT",
  ddd="Actions Per Video YT",eee="% Likes YT",fff="% Comments YT",
  ggg="% Views YT")

Chipotle3 <- cbind(MetricNames,values_chipotle)

#prints all rows for which cor >= 0.7 or <= -0.7
Chipotle4 <- Chipotle3[Chipotle3$cor>=0.7 | Chipotle3$cor <= -0.7,]
#orders by highest correlation
Chipotle4[order(abs(-Chipotle4[,2]),decreasing=TRUE),]

##
## y
## c
MetricNames
% Shares FB
Shares (Total) FB
cor lag
-0.8907804 1
-0.8881715 1

```

```
## p          Actions Per Post FB -0.8718518 1
## h          Impressions per Post FB -0.8580475 1
## dd         Total Actions (% Followers) TW -0.8564076 2
## n          Average Engagement Per Post (%) FB -0.8533316 1
## pp         % Engaged Audience TW -0.8496211 2
## cc         Retweets (Total) TW -0.8479910 2
## ee         Actions (Total) TW -0.8453189 2
## bb         Likes (Total) TW -0.8408358 2
## oo         Unique Retweeters TW -0.8407487 2
## nn         Unique Engaged Audience TW -0.8401400 2
## f          Impressions-viral (Total) FB -0.8350345 1
## j          Actions per Impression (%) FB 0.8332941 1
## e          Impressions (Total) FB -0.8289440 1
## w          Native Video Views (Lifetime Total) FB -0.8175949 1
## hh         Tweets (Total) TW -0.8038749 2
## b          Comments (Total) FB -0.7811377 1
## mn         % Engaged Audience IG 0.7408814 3
## fff        % Comments YT -0.7349069 -1
## v          Clicks (Total) FB 0.7312923 1
## l          Page Score FB -0.7223310 2
## d          Reactions Per Post FB -0.7176640 1
## uu         Likes (Total) IG -0.7164535 0
## ww         Actions (Total) IG -0.7082228 0
## ss         Actions Per Engager TW 0.7008883 5
```

```
#checks for level of significance of correlation by computing pvalue
```

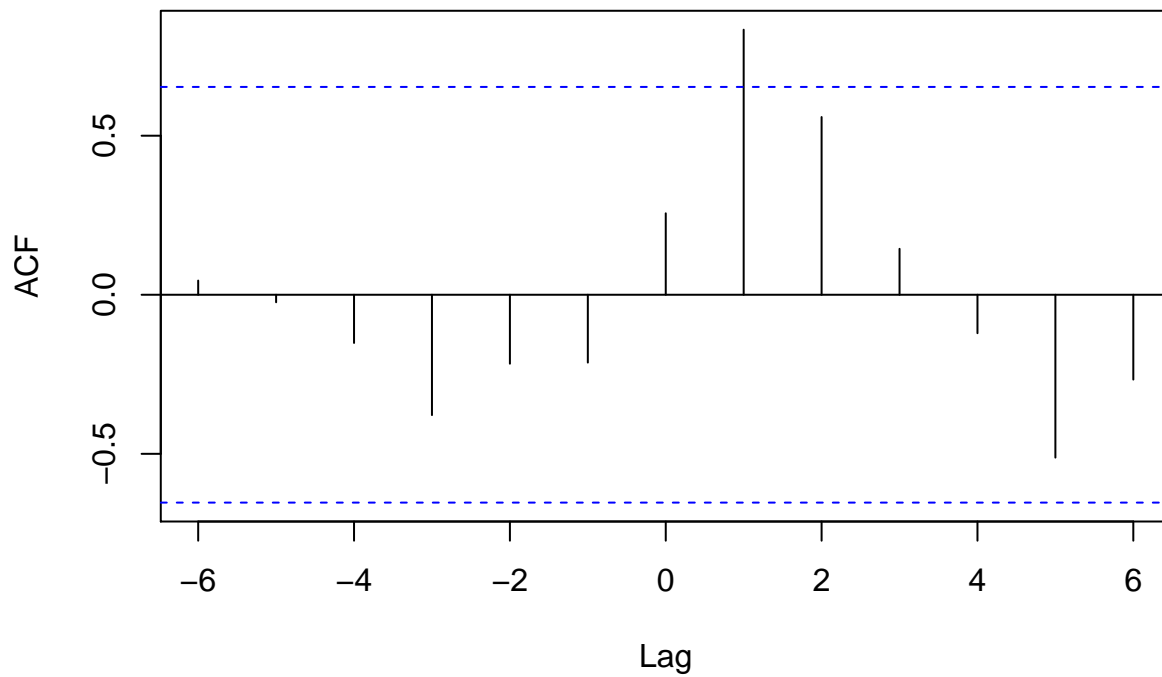
```
#n.used is number of observations in the time series
```

```
#pnorm gives the distribution function
```

```
j <- Find_Abs_Max_CCF(chipotle.sales2$Sales,chipotle.metrics2$`ActionsperImpression(%)FB`)
```

```
2 * (1 - pnorm(abs(j$cor), mean = 0, sd = 1/sqrt( ccf(chipotle.sales2$Sales,chipotle.metrics2$`Actionsp
```

chipotle.sales2\$Sales & chipotle.metrics2\$'ActionsperImpression(%)'



```
## [1] 0.01242346
```

```
# The p-value is smaller than 0.05.
```

```
# For the metrics with highest corr coef, the p-value indicates the relationship is significant
```

```
""
```