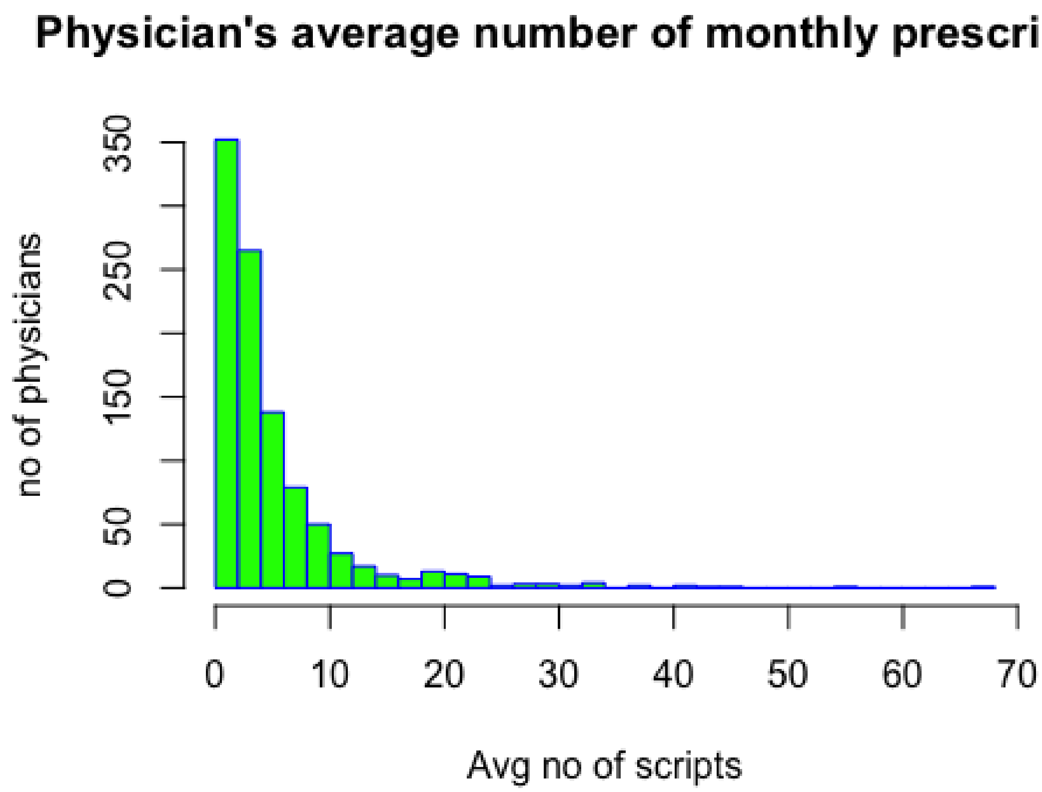
Marketing Analytics Assignment 2

Q1. What are the average number of prescriptions for generalists? For specialists? Plot a histogram of each physician's average number of monthly prescriptions.

```
# Creating a merged dataset:
merged=merge(det_count,det_demo,by="id")

# Calculating average number of prescriptions:
avgscript_gen = merged[generalphys==1,mean(scripts),id][,mean(V1)]
avgscript_spec=merged[specialist==1,mean(scripts),id][,mean(V1)]

#Histogram for each physician's avg number of monthly prescriptions:
avg_pres=merged[,mean(scripts),id]
hist(avg_pres$V1,
     main="Physician's average number of monthly prescriptions",
     xlab="Avg no of scripts",
     ylab = 'no of physicians',
     border="blue",
     col="green",
     las=0.25,
     breaks=40)
```




```
#Result :
#a. The avg prescriptions for generalists is 3.73 approximately.
#b. The avg prescriptions for specialists is 12.55 approximately
#c. The histogram appears to be positively skewed as majority of the
physicians (approx 350) have written 0-2 prescriptions on an average.
```

Q2. What is the correlation between current detailing and scripts written by a physician? Start with the specification: scriptsit = ??0 + ??1detailingit + .it

```
#Running a linear regression model with detailing as the key variable of
interest :
model2.0=lm(scripts~detailing,data=merged,na.action = na.omit)
stargazer(model2.0,type="text", median=TRUE, iqr=TRUE,digits=5,
title="Descriptive Statistics")

##
## Descriptive Statistics
## ====================================================
##                                 Dependent variable:
##                             ------------------------
##                                       scripts
```

```
## ---------------------------------------------------
## detailing                      0.93977***
##                               (0.02780)
##
## Constant                       3.29142***
##                               (0.07081)
##
## ---------------------------------------------------
## Observations                     23,000
## R2                              0.04734
## Adjusted R2                     0.04730
## Residual Std. Error      7.23213 (df = 22998)
## F Statistic         1,142.73900*** (df = 1; 22998)
## ===================================================
## Note:                   *p<0.1; **p<0.05; ***p<0.01

#Result : The model suggests that one additional detailing (visit) is
associated with 0.94 additional scripts written. This indicates a high
correlation between detailing and number of scripts written.
```

3. Does past detailing appear to have an effect on current scripts? What about past script writing? Consider up to three lags. How do you interpret these results?

```
# Part A: Evaluating impact of past detailing on current scripts:

#Creating lagged variables for detailing for upto 3 months
merged[, detail_lag1 := shift(detailing, n=1L, "lag"),id][, detail_lag2 :=
shift(detailing, n=2L, "lag"),id][, detail_lag3 := shift(detailing, n=3L,
"lag"),id]

# Running regression including the lagged variables:
model3a.0 =lm(scripts~detailing+detail_lag1,data=merged,na.action = na.omit)
model3a.1 =lm(scripts~detailing+detail_lag1 +
detail_lag2,data=merged,na.action = na.omit)
model3a.2
=lm(scripts~detailing+detail_lag1+detail_lag2+detail_lag3,data=merged,na.acti
on = na.omit)

stargazer(model2.0,model3a.0,model3a.1,model3a.2, type="text", median=TRUE,
iqr=TRUE,digits=4, title="Descriptive Statistics")

##
## Descriptive Statistics
##
=============================================================================
========================================================
##
Dependent variable:
```

```
##                       -------------------------------------------------------------------------------------------------------------------------------
##
scripts
##                                  (1)                           (2)
(3)                           (4)
## ----------------------------------------------------------------------------------------------------------------------------------------
## detailing                      0.9398***                     0.5717***
0.3865***                     0.3195***
##                                (0.0278)                      (0.0350)
(0.0375)                      (0.0389)
##
## detail_lag1                                                  0.6041***
0.3982***                     0.2735***
##                                                              (0.0357)
(0.0392)                      (0.0412)
##
## detail_lag2
0.5293***                     0.3879***
##
(0.0389)                      (0.0418)
##
## detail_lag3
0.4234***
##
(0.0408)
##
## Constant                       3.2914***                     2.8373***
2.5712***                     2.4120***
##                                (0.0708)                      (0.0767)
(0.0807)                      (0.0844)
##
## ----------------------------------------------------------------------------------------------------------------------------------------
## Observations                    23,000                        22,000
21,000                        20,000
## R2                              0.0473                        0.0591
0.0660                        0.0700
## Adjusted R2                     0.0473                        0.0590
0.0659                        0.0698
## Residual Std. Error      7.2321 (df = 22998)          7.1704 (df = 21997)
7.1412 (df = 20996)         7.1429 (df = 19995)
## F Statistic         1,142.7390*** (df = 1; 22998) 690.4620*** (df = 2;
21997) 494.8960*** (df = 3; 20996) 376.3730*** (df = 4; 19995)
##
=============================================================================================================================================
## Note:
*p<0.1; **p<0.05; ***p<0.01
```

```
#Interpretation:
#1. Compared to the base model, including 1 month lagged variable for
detailing reduces the impact of current detailing from 0.94 to 0.58 on
current scripts. The detaling of the previous month has a higher impact than
the current detailing (0.604 as compared to 0.57 for detailing)
#2. The 2nd and 3rd month lagged variables are also significant and have a
positive impact. However, it turns out to be more impactful as compared to
the 1st lagged month, which contradicts with the business assumption that the
recent months will have more impact.

# Part 2: Evaluating impact of past script writing behaviour on current
scripts:

#Creating lagged variables for scripts upto 3 months
merged[, script_lag1 := shift(scripts, n=1L, "lag"),id][, script_lag2 :=
shift(scripts, n=2L, "lag"),id][, script_lag3 := shift(scripts, n=3L,
"lag"),id]

#Considering only the lagged scripts with current detailing
model3b.1=lm(scripts~detailing+script_lag1,data=merged,na.action = na.omit)
model3b.2=lm(scripts~detailing+script_lag1+script_lag2,data=merged,na.action
= na.omit)
model3b.3=lm(scripts~detailing+script_lag1 +script_lag2
+script_lag3,data=merged,na.action = na.omit)

#Intuitively we feel that the detail lags should have an impact. Hence we are
including it in the model.
model3b.4=lm(scripts~detailing+script_lag1+script_lag2 + script_lag3 +
detail_lag1 + detail_lag2 + detail_lag3 ,data=merged,na.action = na.omit)

model3b.5=lm(scripts~detailing+script_lag1+detail_lag1,data=merged,na.action
= na.omit)

stargazer(model3b.1,model3b.2,model3b.3,type="text", median=TRUE,
iqr=TRUE,digits=4, title="Descriptive Statistics")

##
## Descriptive Statistics
##
=============================================================================
===================================
##                                                          Dependent
variable:
##                       -----------------------------------------------------
--------------------------------------
##                                                                scripts
##                                  (1)                            (2)
(3)
## ---------------------------------------------------------------------------
```

```
--------------------------------------
## detailing                        0.1675***                           0.1022***
0.0827***
##                                   (0.0155)                            (0.0144)
(0.0141)
##
## script_lag1                      0.8353***                           0.4949***
0.3889***
##                                   (0.0036)                            (0.0062)
(0.0066)
##
## script_lag2                                                          0.4015***
0.2608***
##                                                                       (0.0061)
(0.0069)
##
## script_lag3
0.2724***
##
(0.0066)
##
## Constant                         0.4911***                           0.2849***
0.1756***
##                                   (0.0405)                            (0.0380)
(0.0376)
##
## -----------------------------------------------------------------------------
--------------------------------------
## Observations                       23,000                              22,000
21,000
## R2                                 0.7161                              0.7620
0.7795
## Adjusted R2                        0.7160                              0.7620
0.7794
## Residual Std. Error      3.9483 (df = 22997)            3.6064 (df =
21996)            3.4703 (df = 20995)
## F Statistic         28,999.1000*** (df = 2; 22997) 23,472.6100*** (df = 3;
21996) 18,550.7200*** (df = 4; 20995)
##
=============================================================================
===================================
## Note:
*p<0.1; **p<0.05; ***p<0.01
```

```
stargazer(model3b.4,model3b.5,type="text", median=TRUE, iqr=TRUE,digits=4,
title="Descriptive Statistics")
```

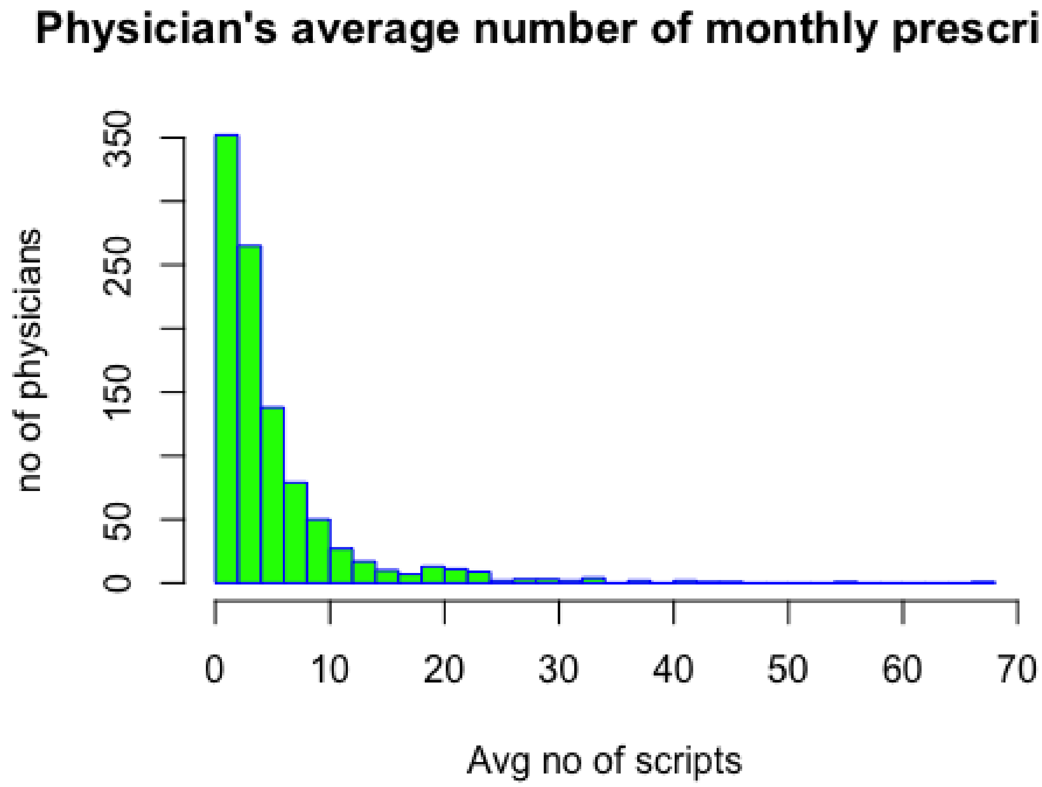
```
##
## Descriptive Statistics
##
```

```
==============================================================================
====
##                                           Dependent variable:
##                      -----------------------------------------------------
-------
##                                                 scripts
##                                (1)                            (2)
## ---------------------------------------------------------------------------
-------
## detailing                    0.0411**                       0.1105***
##                             (0.0189)                       (0.0193)
##
## script_lag1                 0.3886***                      0.8317***
##                             (0.0068)                       (0.0037)
##
## script_lag2                 0.2623***
##                             (0.0071)
##
## script_lag3                 0.2732***
##                             (0.0068)
##
## detail_lag1                  0.0206                        0.0935***
##                             (0.0200)                       (0.0198)
##
## detail_lag2                 0.0794***
##                             (0.0203)
##
## detail_lag3                 -0.0246
##                             (0.0199)
##
## Constant                    0.1204***                      0.4397***
##                             (0.0419)                       (0.0434)
##
## ---------------------------------------------------------------------------
-------
## Observations                 20,000                         22,000
## R2                           0.7811                         0.7160
## Adjusted R2                  0.7810                         0.7159
## Residual Std. Error      3.4660 (df = 19992)          3.9397 (df =
21996)
## F Statistic          10,188.5000*** (df = 7; 19992) 18,481.5700*** (df = 3;
21996)
##
==============================================================================
====
## Note:                                                 *p<0.1; **p<0.05;
***p<0.01

#Result :
#1.Compared to the base model, including lagged scripts variables have an
```

```
impact on current detailing. However the 3rd script lagged variable
contradicts the business assumption that the scripts written in recent months
will have more impact.
#2.Including the script lagged variables have reduced the impact of current
as well as past detailing significantly which means that past script writing
is capturing some variation in the number of scripts written for the current
month.
```

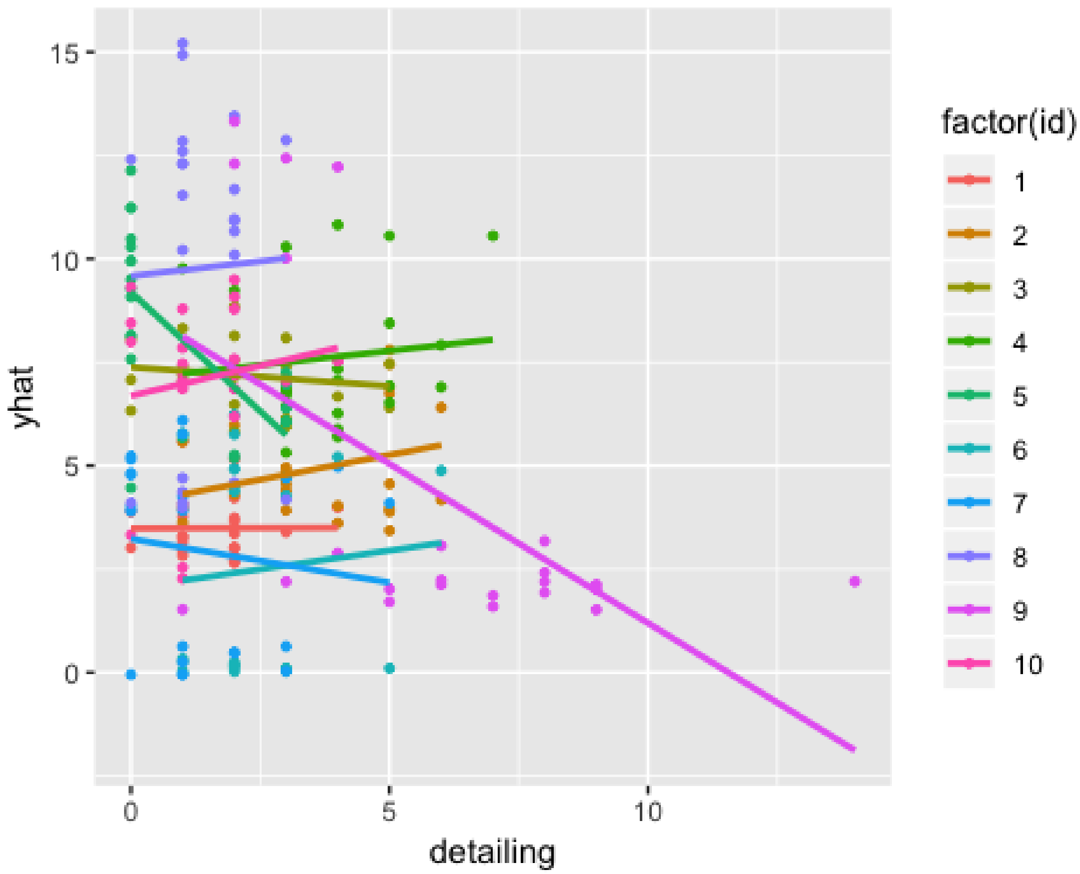
4. Are there large differences in the average prescribing activity across physicians? Can you account for these differences in the regression? How does this affect the estimated influence of past scripts and detailing? Why?

```
summary(avg_pres$V1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.500   2.958   5.076   5.927  66.333

range(avg_pres$V1)

## [1]  0.00000 66.33333

var(avg_pres$V1)

## [1] 43.03256

hist(avg_pres$V1,
     main="Physician's average number of monthly prescriptions",
     xlab="Avg no of scripts",
     ylab = 'no of physicians',
     border="blue",
     col="green",
     las=0.25,
     breaks=40)
```




```
# Result : Yes, there seem to be large differences in the avg prescribing
activity across physicians. The avg prescriptions vary from 0 to 66 with a
mean of 5. There is a large variation in the number of prescriptions written
as indicated by the high variance of 43.
#This is also evident from the positively skewed histogram plotted for
physicians where the right tail indiciates large values for prescribing
activity. These values are further away from the mean resulting in a high
variation.

#These large variations can be accounted in the regression model by taking
fixed effects for each physician.
model4.0 = felm(formula = scripts ~ detailing+detail_lag1+script_lag1|
id,data = merged)
stargazer(model3b.5,model4.0,type="text", median=TRUE, iqr=TRUE,digits=6,
title="Descriptive Statistics")

##
## Descriptive Statistics
## ============================================================================
##                                      Dependent variable:
##                      -------------------------------------------------------
##                                            scripts
##                                    OLS                         felm
```

```
##                                   (1)                           (2)
## ----------------------------------------------------------------------------
## detailing                     0.110460***                  0.078075***
##                                (0.019315)                  (0.018300)
##
## script_lag1                   0.831734***                  0.263946***
##                                (0.003688)                  (0.006648)
##
## detail_lag1                   0.093505***                  0.072393***
##                                (0.019769)                  (0.018636)
##
## Constant                      0.439671***
##                                (0.043436)
##
## ----------------------------------------------------------------------------
## Observations                     22,000                       22,000
## R2                              0.715963                     0.804791
## Adjusted R2                     0.715925                     0.795475
## Residual Std. Error      3.939661 (df = 21996)       3.342834 (df = 20997)
## F Statistic         18,481.570000*** (df = 3; 21996)
## ============================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01

#Intrepretation:
#1. Including the fixed effects for each physician have accounted for the
object bias in the model. The scenarios like targeting physicians due to
popularity, specialities, low-prescribing activity, personal characteristics
etc. have been accounted for.
#2. Including physician fixed effect reduced the impact of current detailing
on the current scripts. Also, it captured the upward bias prevalent in the
script lagged variable.
#3. The slight change in detailing and the larger change in script lagged
variable imply that major variation in the model was being captured by script
lagged variable and so including physician fixed effects didnt change
detailing to that extent.

# Graph:
merged[, yhat := model4.0$fitted]
ggplot( data = merged[id<=10, .(id, detailing, yhat)],
        aes(x = detailing, y = yhat, color=factor(id) ) ) +
  geom_point(size=1) +
  geom_smooth(method = "lm", se = FALSE)
```



5. How does including a linear time trend affect your estimates? How does including time fixed effects affect your estimates? (Note that for technical reasons you should not include the lagged Y variables in a regression that includes time fixed effects.)

```
#Creating time variable which accounts for different 24 months in the
dataset.
merged=merged[, month := rep(1:24, each =1), id]

#Running regression:
model5.0=felm(formula = scripts ~ detailing+detail_lag1| id,data = merged)
#Reference base model.
model5.1=felm(formula = scripts ~ detailing+detail_lag1+month|id,data =
merged) #Model with linear time trend.
model5.2=felm(formula = scripts ~ detailing+detail_lag1|id + month,data =
merged) #Model with fixed time effect.

stargazer(model5.0, model5.1, model5.2,type="text", median=TRUE,
iqr=TRUE,digits=7, title="Descriptive Statistics")

##
## Descriptive Statistics
```

```
## 
==================================================================================
===========
##                                              Dependent variable:               
##                        -----------------------------------------------------------
--------------
##                                                    scripts                     
##                                (1)                    (2)                    
(3)           
## -------------------------------------------------------------------------------
--------------
## detailing                  0.1018847***           0.1260746***           
0.1265448***      
##                            (0.0189640)            (0.0190747)            
(0.0193004)       
##                                                                              
## detail_lag1                0.1025340***           0.1259622***           
0.1305661***      
##                            (0.0193060)            (0.0194038)            
(0.0196305)       
##                                                                              
## month                                            -0.0374154***                          
##                                                   (0.0037538)                           
##                                                                              
## -------------------------------------------------------------------------------
--------------
## Observations                  22,000                 22,000                 
22,000          
## R2                          0.7901347              0.7911230              
0.7918168       
## Adjusted R2                 0.7801302              0.7811552              
0.7816741       
## Residual Std. Error 3.4659710 (df = 20998) 3.4578820 (df = 20997) 
3.4537800 (df = 20977)
## 
==================================================================================
===========
## Note:                                                               *p<0.1; 
**p<0.05; ***p<0.01

#Intrepretation:
#1. Including linear time trend increases the impact of detailing on the
current script. However, in general as time progresses, the number of scripts
written by the physicians reduces.
#2. Including time fixed effects increased the impact of detailing compared
to the base model. This can imply that the downward bias due to seasonality,
low prescribing months,etc. in detailing was accounted for.
```

6. For a regression that includes both physician and time fixed effects, what sort of plausible story can you tell that would cause the detailing coefficients to be biased? What is/are the unobservable variables that result from your story? Would these unobservables bias the detailing coefficient upward or downward?

```
#The physician and time fixed effect doesnt account for targeting of
physicians due to competitor's targeting behaviour.
#For Example: Imagine a physician 'A' in our sample, is being targeted by the
competitor selling a similar drug. Our salesperson would then schedule a
visit to the same physician, persuading him to write the prescription for
Drug X. Also, any prior knowledge about the scheduling of our competitors
would influence our targeting. As a result, the impact of detailing observed
in the model on the number of scripts wriiten is not the true impact.

#The unobservable variable resulting from the above scenario is
"Targeted_by_competition" (A dummy variable indicating if the physician has
been detailed by the competitor). This variable impacts the number of scripts
written by the physician and also the detailing done by the salesperson.
Owing, to the above reasons, it leads to omitted variable bias.

#Absence of variable 'Targeted_by_competition' will introduce a bias which
can be upward/downward based on the scenarios:
#Upward Bias: Aggressive detailing to physicians targeted by our competitors.
#Downward Bias: If effect of our detailing is diluted by our competitors, it
will result in downward bias

#The bias due to this targeting cannot be accounted for by the physician and
the time fixed effects.
```

7. Instead of using scriptsit as the dependent variable, what happens if you use its “first difference”, i.e., scriptsit ??? scriptsit???1? What is the estimated effect of detailing on this outcome? How do you interpret these estimates?

```
#Creating  a variable indicating the first difference in scripts
merged[, sript_diff:= (scripts - script_lag1)]

model7.1=felm(formula = sript_diff ~ detailing+detail_lag1|id + month ,data =
merged)
stargazer(model7.1,type="text", median=TRUE, iqr=TRUE,digits=4,
title="Descriptive Statistics")

##
## Descriptive Statistics
## ===============================================
##                          Dependent variable:
```

```
##                          --------------------------
##                                  sript_diff
## -----------------------------------------------
## detailing                          0.0136
##                                   (0.0235)
##
## detail_lag1                        -0.0077
##                                   (0.0239)
##
## -----------------------------------------------
## Observations                       22,000
## R2                                 0.0108
## Adjusted R2                        -0.0373
## Residual Std. Error     4.1979 (df = 20977)
## ===============================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

```
#Intrepretation:
#As the coefficients are statistically insignificant, there is no impact of
detailing on the change in the number of scripts written in the current month
as compared to the previous month. This can imply that the change in the
scripts is not impacted by the detailing and might be due to variety of other
factors such as increase in the number of customers in the current month.
```

8. Based on the work you have done so far, pick 2-4 model specifications that you think could plausibly approximate the true model. Why did you pick these models? Report their estimates in a single table (using stargazer).What are the range of estimates for detailing across these models? Is that range larger or smaller than the standard errors estimated for each model?

```
#Model 1:
model4.0 = felm(formula = scripts ~ detailing+detail_lag1+script_lag1|
id,data = merged)

#Model 2:
model5.2=felm(formula = scripts ~ detailing+detail_lag1|id + month,data =
merged)

stargazer(model4.0,model5.2,type="text",
          column.labels = c("Model-1","Model-2"),median=TRUE,
iqr=TRUE,digits=4, title="Descriptive Statistics")

##
## Descriptive Statistics
## ===========================================================
##                                Dependent variable:
##                          -------------------------------------
```

```
##                                        scripts
##                              Model-1             Model-2
##                                (1)                 (2)
## ----------------------------------------------------------
## detailing                  0.0781***           0.1265***
##                            (0.0183)            (0.0193)
##
## detail_lag1                0.0724***           0.1306***
##                            (0.0186)            (0.0196)
##
## script_lag1                0.2639***
##                            (0.0066)
##
## ----------------------------------------------------------
## Observations                22,000              22,000
## R2                          0.8048              0.7918
## Adjusted R2                 0.7955              0.7817
## Residual Std. Error 3.3428 (df = 20997) 3.4538 (df = 20977)
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01

#Reason to chose models:
#Model 1 accounts for the lagged variables for both detailing and scripts as
well as it controls for all the attributes specific to the physician.
#Model 2 accounts for the detail lagged variables but majorly controls for
both attributes specific to the physician and any seasonailty or time related
trends. These controls also compensate for the loss of removing the script
lagged variable.
#Hence, Model 2 is our preferred model

#Range of estimates for detailing:
# The coefficient of detailing ranges from 0.078 to 0.126 : 0.0484

#Range of standard errors: 0.001
#The range for detailing coefficients is larger as compared to the range for
standard errors.
```

9. Pick one of these models that you like the best. We'll call this your "preferred" model going forward. Alter it so that it estimates the advertising elasticity. What do you estimate?

```
#Preferred Model:
model5.2=felm(formula = scripts ~ detailing+detail_lag1 | id + month,data =
merged) #model5.2

#Created a physician type variable (phytype). Phytype 1 is for specialist, 2
is for general practitioners and 3 is for  others.
merged=merged[(specialist==1 & generalphys==0),phytype:= 1][(specialist==0 &
generalphys==1),phytype:= 2][(specialist==0 & generalphys==0),phytype:= 3]
```

```r
#Altered Model: The model has been altered to account for advertising
elasticity by including mean_samples. However, since mean_samples and id were
perfectly correlated, we included physician type to control for certain
physician characteristics.
model9.1=felm(formula = log(scripts + 1) ~ log(detailing + 1)+log(detail_lag1
+ 1)+ log(mean_samples + 1) | month,data = merged)
model9.2=felm(formula = log (scripts + 1) ~ log(detailing +
1)+log(detail_lag1 + 1) + log(mean_samples + 1) | phytype + month, data =
merged)

stargazer(model9.1,model9.2, type="text",column.labels = c( "Model-1",
"Model-2"), median=TRUE,
          iqr=TRUE,digits=4, title="Descriptive Statistics")

##
## Descriptive Statistics
## ==============================================================
##                                  Dependent variable:
##                         ---------------------------------------
##                                    log(scripts + 1)
##                              Model-1             Model-2
##                                (1)                 (2)
## --------------------------------------------------------------
## log(detailing + 1)         0.1423***           0.1530***
##                            (0.0129)            (0.0119)
##
## log(detail_lag1 + 1)       0.1420***           0.1443***
##                            (0.0130)            (0.0120)
##
## log(mean_samples + 1)      1.1163***           0.9188***
##                            (0.0203)            (0.0192)
##
## --------------------------------------------------------------
## Observations                22,000              22,000
## R2                          0.2200              0.3356
## Adjusted R2                 0.2191              0.3348
## Residual Std. Error   0.8716 (df = 21975) 0.8045 (df = 21973)
## ==============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01

#Intrepretation:
# The advertising elasticity seems to be positively correlated with the
current scripts. When accounting only for fixed time effects, 1% increase in
the number of average samples provided to physicians per month is associated
to 1.11 % increase in the number of current scripts.
# However, including the physician type in the model reduced the impact of
advertising elasticity to 0.918%. This might be because mean_samples did not
capture the physician characteristics and thus had an upward bias in Model-1
```

10. What happens to the standard errors of your preferred model if you adjust for heteroskedasticity? What happens if you cluster your standard errors by time? What happens if you cluster your standard errors by physician? Which adjustment for the standard errors do you think is most appropriate for this analysis and why?

```
#Preferred Model:
model5.2=felm(formula = scripts ~ detailing+detail_lag1|id + month,data =
merged)

#Adjusting for heteroskedasticity:
summary(model5.2)

##
## Call:
##    felm(formula = scripts ~ detailing + detail_lag1 | id + month,
data = merged)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.642  -1.513  -0.312   1.241  51.007
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## detailing     0.12654    0.01930   6.557 5.63e-11 ***
## detail_lag1  0.13057    0.01963   6.651 2.98e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.454 on 20977 degrees of freedom
##   (2000 observations deleted due to missingness)
## Multiple R-squared(full model): 0.7918   Adjusted R-squared: 0.7817
## Multiple R-squared(proj model): 0.005623   Adjusted R-squared: -0.04282
## F-statistic(full model):78.07 on 1022 and 20977 DF, p-value: < 2.2e-16
## F-statistic(proj model): 59.31 on 2 and 20977 DF, p-value: < 2.2e-16

summary(model5.2, robust = TRUE)

##
## Call:
##    felm(formula = scripts ~ detailing + detail_lag1 | id + month,
data = merged)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.642  -1.513  -0.312   1.241  51.007
##
## Coefficients:
##               Estimate Robust s.e t value Pr(>|t|)
```

```
## detailing    0.12654    0.02242   5.645 1.67e-08 ***
## detail_lag1  0.13057    0.02298   5.681 1.36e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.454 on 20977 degrees of freedom
##   (2000 observations deleted due to missingness)
## Multiple R-squared(full model): 0.7918   Adjusted R-squared: 0.7817
## Multiple R-squared(proj model): 0.005623   Adjusted R-squared: -0.04282
## F-statistic(full model, *iid*):78.07 on 1022 and 20977 DF, p-value: <
2.2e-16
## F-statistic(proj model): 42.77 on 2 and 20977 DF, p-value: < 2.2e-16

#Clustering standard errors by time
model10.1=felm(formula = scripts ~ detailing+detail_lag1|id +
month|0|month,data = merged)

#Clustering standard errors by physician id:
model10.2=felm(formula = scripts ~ detailing+detail_lag1|id + month|0|id,data
= merged)

stargazer(model5.2,model10.1, model10.2, type="text",column.labels =
c("Preferred Model","CL1","CL2","CL3"),
median=TRUE,iqr=TRUE,digits=4,title="Descriptive Statistics")

##
## Descriptive Statistics
## ====================================================================
##                                          Dependent variable:
##                                   ----------------------------------
##                                                scripts
##                                   Preferred Model    CL1       CL2
##                                         (1)          (2)       (3)
## --------------------------------------------------------------------
## detailing                            0.1265***    0.1265*** 0.1265***
##                                      (0.0193)     (0.0202)  (0.0283)
##
## detail_lag1                          0.1306***    0.1306*** 0.1306***
##                                      (0.0196)     (0.0204)  (0.0295)
##
## --------------------------------------------------------------------
## Observations                          22,000       22,000    22,000
## R2                                    0.7918       0.7918    0.7918
## Adjusted R2                           0.7817       0.7817    0.7817
## Residual Std. Error (df = 20977)      3.4538       3.4538    3.4538
## ====================================================================
## Note:                                   *p<0.1; **p<0.05; ***p<0.01

#Intrepretation:
#Adjusting for heteroskedasticity increases the standard error from 0.019 to
```

```
0.022. This might be owing to the fact that there might be variation within specific time or object groups in our dataset.
#Clustering standard errors by time increases the error as compared to the preferred model. However, the standard error reduced as compared to model adjusted with robust standard errors
#Clustering standard errors by physician overall increases the standard error.

#Based on the standard errors clustering with month seems to be a better option. However, intuitively, there seems to be more variation among the physicians.So, physician id as a clustering variable is more appropriate.
```

11. Using your preferred specification, test for different effects of detailing between generalists and specialists. Do they appear to have different responses? Are there any other physician characteristics that are related to differential responses to detailing?

```
#Model to test effects of detailing for specialists:
merged_spec = merged[phytype==1]
model11.1=felm(scripts ~ detailing+detail_lag1|month + id ,data = merged_spec)

#Model to test effects of detailing for generalists:
merged_gen = merged[phytype==2]
model11.2=felm(scripts ~ detailing+detail_lag1|month + id ,data = merged_gen)

stargazer(model11.1,model11.2, type="text",column.labels = c("Spec","Gen"), median=TRUE, iqr=TRUE,digits=4, title="Descriptive Statistics")

##
## Descriptive Statistics
## ==========================================================
##                               Dependent variable:
##                         --------------------------------------
##                                       scripts
##                             Spec                  Gen
##                             (1)                   (2)
## ----------------------------------------------------------
## detailing               0.3583***            0.0475**
##                          (0.0695)            (0.0198)
##
## detail_lag1             0.3381***            0.0560***
##                          (0.0709)            (0.0201)
##
## ----------------------------------------------------------
## Observations              4,070               13,222
## R2                        0.7945              0.5929
## Adjusted R2               0.7835              0.5728
## Residual Std. Error 5.8532 (df = 3862) 2.7240 (df = 12598)
```

```
## =============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01

# Intrepretation:
#There appears to be differences in the responses to detailing for
specialists and generalists. The impact of detailing for specialists is
approximately 7.5 times more as compared to that for generalists.

#There are other physician characteristics that are related to differential
responses. One of the characteristics available in the dataset is
mean_samples.

#Model to test for mean_samples for specialists:
model11b.1=felm(scripts ~
detailing+mean_samples+detailing*mean_samples+detail_lag1|month ,data =
merged_spec)

#Model to test for mean_samples for generalists:
model11b.2=felm(scripts ~
detailing+mean_samples+detailing*mean_samples+detail_lag1|month ,data =
merged_gen)

stargazer(model11b.1,model11b.2, type="text", column.labels =
c("Spec","Gen"),median=TRUE, iqr=TRUE,digits=4, title="Descriptive
Statistics")

##
## Descriptive Statistics
## =================================================================
##                                     Dependent variable:
##                         ----------------------------------------
##                                           scripts
##                                 Spec                    Gen
##                                  (1)                    (2)
## -----------------------------------------------------------------
## detailing                    1.1038***              0.3040***
##                              (0.1523)               (0.0299)
##
## mean_samples                 7.6503***              2.6836***
##                              (0.3838)               (0.0926)
##
## detail_lag1                  0.6446***              0.1842***
##                              (0.1216)               (0.0253)
##
## detailing:mean_samples      -0.4986***             -0.1506***
##                              (0.0991)               (0.0240)
##
## -----------------------------------------------------------------
## Observations                   4,070                  13,222
## R2                            0.2317                 0.1886
```

```
## Adjusted R2                     0.2269              0.1871
## Residual Std. Error    11.0615 (df = 4044) 3.7575 (df = 13196)
## =================================================================
## Note:                                 *p<0.1; **p<0.05; ***p<0.01

#Intrepretation:
#If free samples are provided, the impact of detailing is reduced.
```

12. What are your main takeaways from this investigation of the effect of detailing on physician behavior? Present these takeaways such that a non-technical person could understand them. Discuss whether you think detailing is effective, to what degree, the degree of confidence in your conclusions and explain why that is your level of confidence (i.e, what are your sources of doubt).

```
# Takeaways:
#Effectiveness of Detailing on current scripts:
#1. From our analysis we find that the impact of current and past detailing
on physician script writing behaviour is high. This impact accounts for any
seasonality, prevalence of epidemic, other time related trends in prescribing
activity. It also accounts for popularity, script writing behaviour or any
other characteristics specific to the physician.
#2. For every 10 additional detailings done, the number of scripts written
increases by 1 approximately.
#3. We also observed that the response to detailing is different for
different physicians type. For every 10 additional detailings done,
specialists prescribe 3-4 additional number of scripts whereas generalists
prescribe around 0-1 additional scripts approx.

# We are fairly confident that the detailing has a significant impact on the
physician prescribing behaviour.However, there are certain limitation to our
analysis:
#1. The biases caused by another omitted variables like targeting by
competition couldn't be accounted for.
#2. Limitation in capturing the variations in the samples sent to the
physicians per month owing to the lack of exact number of samples sent.
#3. Loss of data due to NAs generated owing to creation of lagged variable
(systematic sampling) as it is a panel data.
```