Marketing Mix Panel Data Homework

Chu, Hisham, Tamsir, Su

February 7, 2019

Counts Information

- id – identifies the physician
- scripts – the number of new prescriptions ordered by the physician for the drug detailed
- detailing – the number of sales calls made to each physician per month

Demo Information

- id – identifies the physician
- generalphys – dummy for if doctor is a "general practitioner"
- specialist – dummy for if the physician is a specialist in the theraputic class for which the drug is intended
- mean_samples – the mean number of free drug samples given to the doctor over the sample period

```
uniqueN(counts[, .(id, scripts, detailing)] ) / counts[,.N]
```

```
## [1] 0.6690417
```

```
uniqueN(demo[, .(id)] ) / demo[,.N]
```

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

```
# Set key for demo table
setkey(demo, id)

counts[,months := rep(c(6:12,1:12,1:5), 2000)] # from June
counts[,year := rep(c(1999,2000,2001), c(7,12,5))] # from 1999

# Combine the month and the year to create a unique date column.
counts[, yrmn := year*100 + months]
uniqueN(counts[, .(id, yrmn)] ) / counts[,.N]
```

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

```
# Set key for counts table
setkey(counts, id, yrmn)

sum(demo$generalphys == 1)
```

```
## [1] 601
```

```
sum(demo$specialist == 1)
```

```
## [1] 185
```

```
sum(demo$generalphys == 1 & demo$specialist == 1)
```

```
## [1] 0
```

```
sum(demo$generalphys == 0 & demo$specialist == 0)
```

```
## [1] 214
```

```r
# There are doctors who are neither gen or spec.
```

Question 1

```r
# Merge counts onto demo dataset with aggregate columns
doctors <- merge(demo, counts[, .(sumScripts = sum(scripts, na.rm = TRUE),
                                  sumDetailing = sum(detailing, na.rm = TRUE),
                                  avgScripts = mean(scripts, na.rm = TRUE),
                                  avgDetailing = mean(detailing, na.rm = TRUE)),
                              by = id])

uniqueN(doctors[, .(id)] ) / doctors[,.N]
```

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

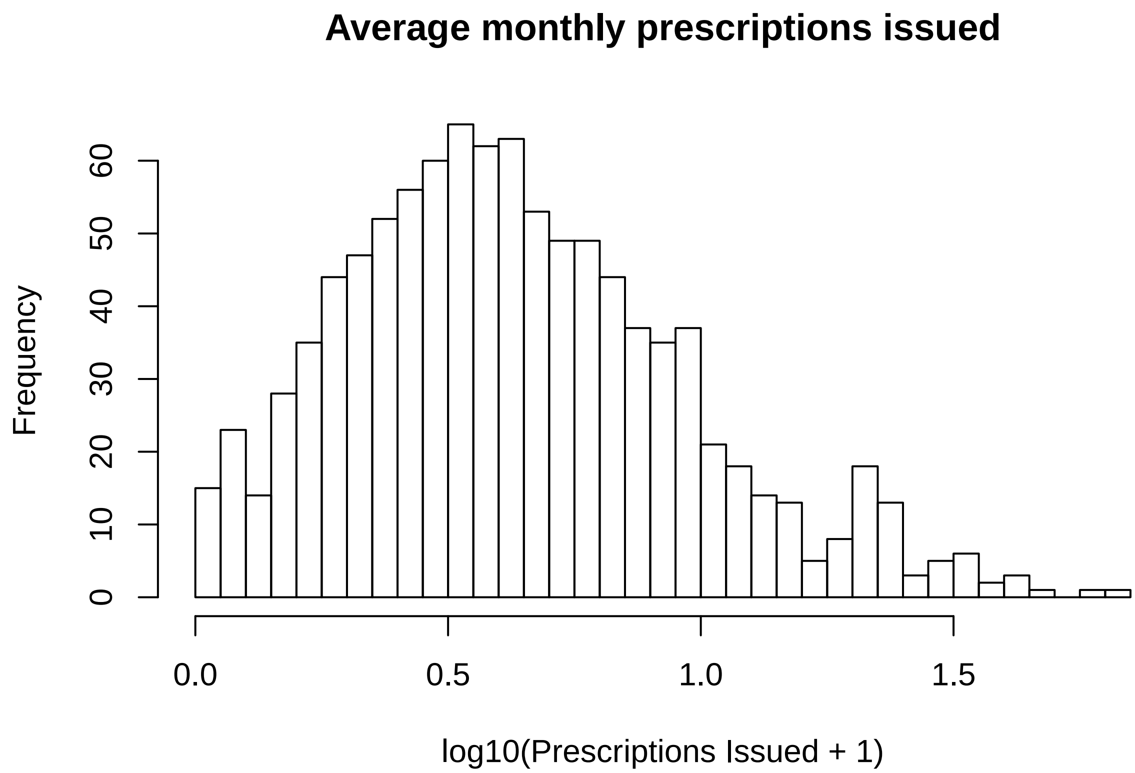
```r
setkey(doctors, id)

# Generalists
genphys <- doctors[doctors$generalphys == 1, mean(avgScripts)] * 24

# Specialists
spec <- doctors[doctors$specialist == 1, mean(avgScripts)] * 24
```

Average scripts per general physician: 89.5757072

Average scripts per specialist: 301.4378378

Histogram:

```r
# Histogram for mean prescriptions issued monthly by each doctor
hist(doctors[, log10(avgScripts + 1)],
     breaks = 50,
     main = "Average monthly prescriptions issued",
     xlab = "log10(Prescriptions Issued + 1)")
```



Question 2

```r
lm1 = lm(scripts ~ detailing, data = counts)
summary(lm1)
```

```
##
## Call:
## lm(formula = scripts ~ detailing, data = counts)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.448  -3.990  -2.231   0.889  90.829
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.29142    0.07081   46.48   <2e-16 ***
## detailing    0.93977    0.02780   33.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.232 on 22998 degrees of freedom
##   (1000 observations deleted due to missingness)
## Multiple R-squared:  0.04734,    Adjusted R-squared:  0.0473
## F-statistic:  1143 on 1 and 22998 DF,  p-value: < 2.2e-16
```

A one unit increase in detailing is associated with a 0.93977 positive increase in scripts.

Question 3

```r
counts[, details1 := shift(detailing, n=1L, "lag")]
counts[, details2 := shift(detailing, n=2L, "lag")]
counts[, details3 := shift(detailing, n=3L, "lag")]
counts[, script1 := shift(scripts, n=1L, "lag")]
counts[, script2 := shift(scripts, n=2L, "lag")]
counts[, script3 := shift(scripts, n=3L, "lag")]
```

```r
lm2 = lm(scripts ~ detailing + details1, data = counts)
lm3 = lm(scripts ~ detailing + details1 + details2, data = counts)
lm4 = lm(scripts ~ detailing + details1 + details2 + details3, data = counts)

stargazer(lm1, lm2, lm3, lm4,
          title = "Lag Detailing", type = "text",
          column.labels = c( "Current", "Lag1", "Lag2", "Lag3"),
          df = FALSE, digits = 2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Lag Detailing
## =============================================================
##                                 Dependent variable:
##                     -----------------------------------------
##                                       scripts
##                       Current      Lag1       Lag2       Lag3
##                         (1)        (2)        (3)        (4)
## -------------------------------------------------------------
## detailing             0.94***    0.57***    0.39***    0.32***
##                       (0.03)     (0.03)     (0.04)     (0.04)
##
## details1                         0.60***    0.40***    0.27***
##                                  (0.04)     (0.04)     (0.04)
##
## details2                                    0.53***    0.39***
##                                             (0.04)     (0.04)
##
## details3                                               0.42***
##                                                        (0.04)
##
## Constant              3.29***    2.84***    2.57***    2.41***
##                       (0.07)     (0.08)     (0.08)     (0.08)
##
## -------------------------------------------------------------
## Observations          23,000     22,000     21,000     20,000
## R2                     0.05       0.06       0.07       0.07
## Adjusted R2            0.05       0.06       0.07       0.07
## Residual Std. Error    7.23       7.17       7.14       7.14
## F Statistic         1,142.74*** 690.46*** 494.90*** 376.37***
## =============================================================
## Note:                            *p<0.05; **p<0.01; ***p<0.001
```

```r
lm6 = lm(scripts~script1, data=counts)
lm7 = lm(scripts~script1 + script2, data = counts)
lm8 = lm(scripts~script1 + script2 + script3, data = counts)
```

```r
stargazer(lm6, lm7, lm8,
          title="Lag Scripts", type="text",
          column.labels=c( "Lag1", "Lag2", "Lag3"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Lag Scripts
## ==========================================================
##                               Dependent variable:
##                     --------------------------------------
##                                    scripts
##                         Lag1          Lag2          Lag3
##                         (1)           (2)           (3)
## ----------------------------------------------------------
## script1               0.81***       0.55***       0.50***
##                      (0.004)        (0.01)        (0.01)
##
## script2                             0.32***       0.23***
##                                     (0.01)        (0.01)
##
## script3                                           0.16***
##                                                   (0.01)
##
## Constant              0.94***       0.64***       0.53***
##                        (0.03)        (0.03)        (0.03)
##
## ----------------------------------------------------------
## Observations           23,999        23,998        23,997
## R2                      0.66          0.70          0.71
## Adjusted R2             0.66          0.70          0.71
## Residual Std. Error     4.30          4.07          4.01
## F Statistic         47,432.03*** 27,907.37*** 19,337.54***
## ==========================================================
## Note:                        *p<0.05; **p<0.01; ***p<0.001
```

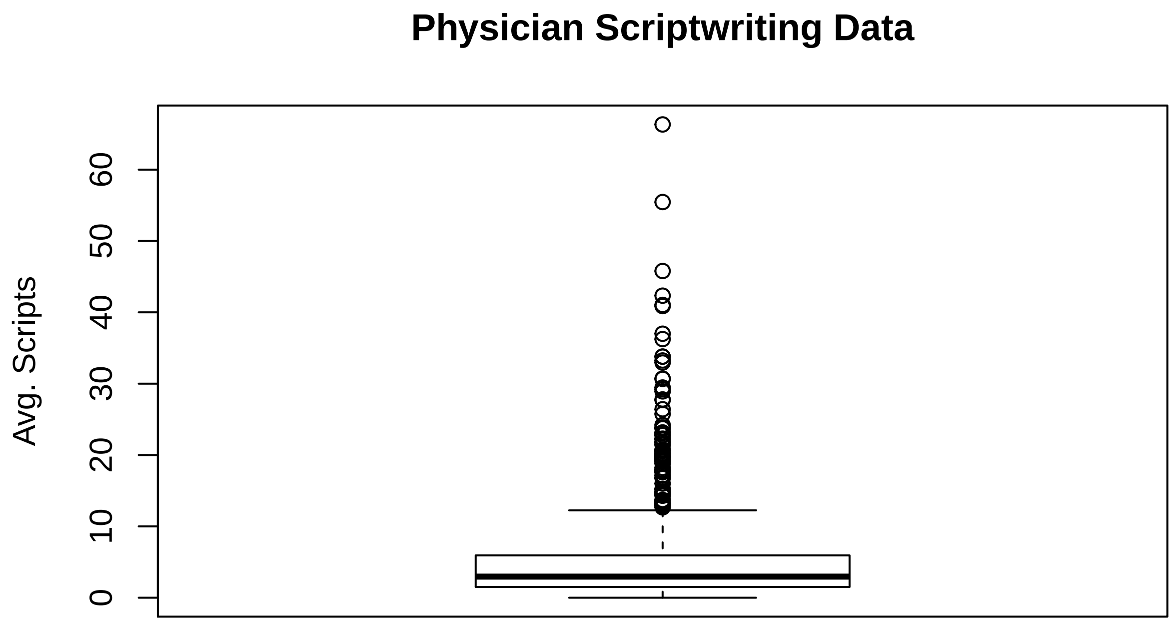
Past detailing has a positive and significant association in prescribed scripts taking into account both individual and combined lag models.

Past scripts has a positive and significant association in current prescribed scripts taking into account both individual and combined lag models.

Question 4

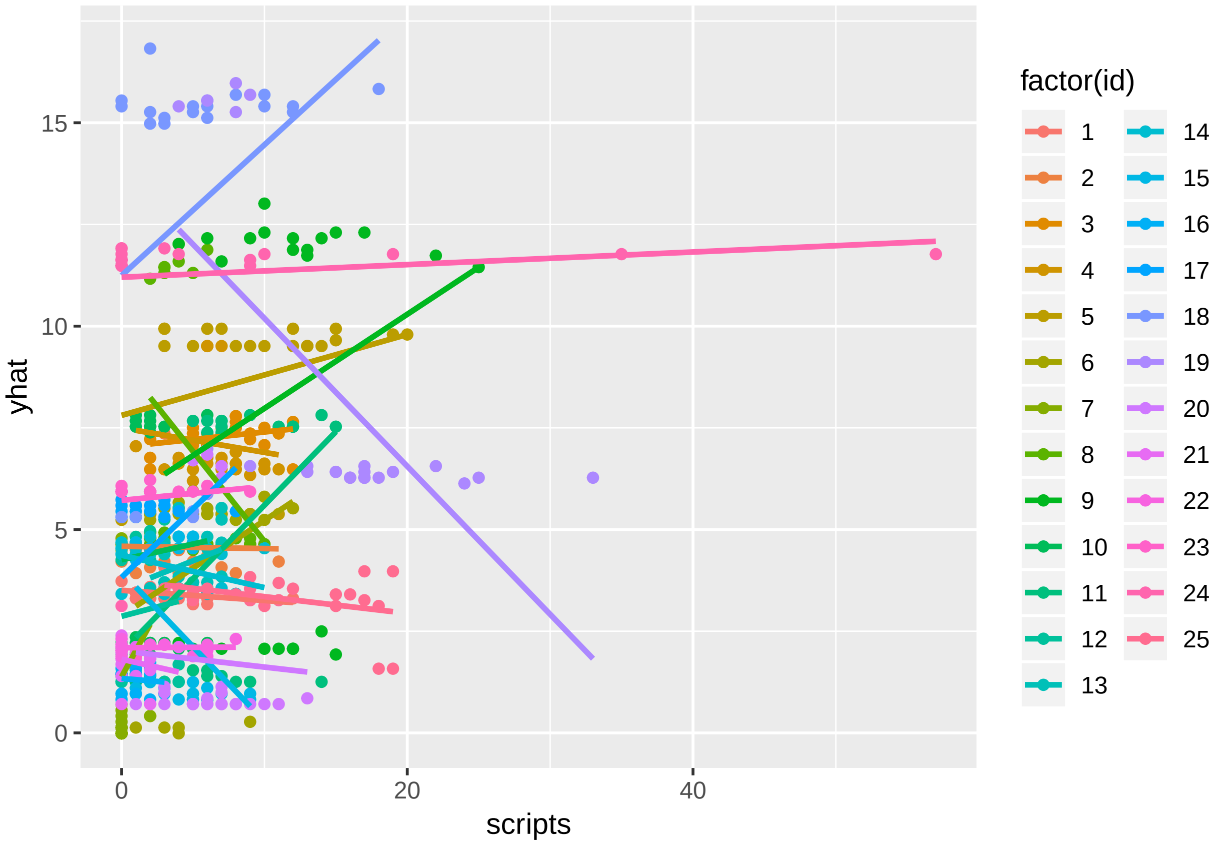
Boxplot of average scripts for all physicians:

```r
boxplot(doctors$avgScripts, data = doctors,
        main = "Physician Scriptwriting Data",
        ylab = "Avg. Scripts")
```




```
ind_fix <- lm(scripts ~ factor(id) + detailing - 1, data = counts)
counts[, yhat := ind_fix$fitted]
```

```
ggplot(data = counts[id <= 25, .(id, scripts, yhat)],
       aes(x = scripts, y = yhat, color = factor(id) ) ) +
  geom_point() + geom_smooth(method = "lm", se = FALSE)
```




```r
summary(doctors$avgScripts)
```

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

Yes, we see a large difference and variation in average presciribing activity across physicians. We can take this into account by creating a factor variable for each physician resulting in their own intercept.

```r
fixedid1 = felm(scripts ~ script1 | id, data = counts)
fixedid2 =felm(scripts~ script1 + script2 | id, data=counts)
fixedid3 =felm(scripts~ script1 + script2 + script3 | id, data=counts )
normalid1 =lm(scripts~ script1, data=counts )
normalid2 =lm(scripts~ script1 + script2, data=counts )
normalid3 =lm(scripts~ script1 + script2 + script3, data=counts )

fixeddetail1 =felm(scripts~ detailing | id, data=counts )
fixeddetail2 =felm(scripts~ detailing + details1 | id, data=counts )
fixeddetail3 =felm(scripts~ detailing + details1 + details2 | id, data=counts )
fixeddetail4 =felm(scripts~ detailing + details1 +
                     details2 + details3| id, data=counts )

normaldetail1 =lm(scripts~ detailing, data=counts )
normaldetail2 =lm(scripts~ detailing + details1, data=counts )
normaldetail3 =lm(scripts~ detailing + details1 + details2, data=counts )
normaldetail4 =lm(scripts~ detailing + details1 +
                    details2 + details3, data=counts )
```

```r
stargazer(fixedid1, fixedid2, fixedid3, normalid1, normalid2, normalid3,
          title="Fixed ID VS. Normal", type="text",
          column.labels=c( "Fixed ID1", "Fixed ID2", "Fixed ID3",
                           "NormalID1", "NormalID3", "NormalID2"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Fixed ID VS. Normal
## ===========================================================================================
##                                                 Dependent variable:
##                      ----------------------------------------------------------------------
##                                                      scripts
##                                   felm                               OLS
##                      Fixed ID1 Fixed ID2 Fixed ID3  NormalID1    NormalID3    NormalID2
##                         (1)       (2)       (3)        (4)          (5)          (6)
## -------------------------------------------------------------------------------------------
## script1               0.22***   0.19***   0.19***    0.81***      0.55***      0.50***
##                       (0.01)    (0.01)    (0.01)     (0.004)       (0.01)       (0.01)
##
## script2                         0.06***   0.05***                 0.32***      0.23***
##                                 (0.01)    (0.01)                   (0.01)       (0.01)
##
## script3                                   0.03***                              0.16***
##                                           (0.01)                                (0.01)
##
## Constant                                             0.94***      0.64***      0.53***
##                                                       (0.03)       (0.03)       (0.03)
##
## -------------------------------------------------------------------------------------------
## Observations          23,999    23,998    23,997     23,999       23,998       23,997
## R2                     0.79      0.80      0.80       0.66         0.70         0.71
## Adjusted R2            0.79      0.79      0.79       0.66         0.70         0.71
## Residual Std. Error    3.43      3.42      3.42       4.30         4.07         4.01
## F Statistic                                       47,432.03*** 27,907.37*** 19,337.54***
## ===========================================================================================
## Note:                                                   *p<0.05; **p<0.01; ***p<0.001
```

```r
stargazer(fixeddetail1, fixeddetail2, fixeddetail3,
          fixeddetail4, normaldetail1, normaldetail2,
          normaldetail3, normaldetail4,
          title="Fixed ID VS. Normal", type="text",
          column.labels=c( "Fixed Detail2", "Fixed Detail2", "Fixed Detail3",
                           "Fixed Detail4", "Normal Detail1", "NormalID2",
                           "Normal Detail3", "Normal Detail4"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Fixed ID VS. Normal
## ==================================================================================================================
##                                                                       Dependent variable:
##                      ---------------------------------------------------------------------------------------------
##                                                                             scripts
##                                                  felm
##                      Fixed Detail2 Fixed Detail2 Fixed Detail3 Fixed Detail4 Normal Detail1 NormalID2
```

```
##                          (1)          (2)          (3)          (4)          (5)          (6)
## ------------------------------------------------------------------------------------------------
## detailing               0.14***      0.10***      0.08***      0.07***      0.94***      0.57***
##                         (0.02)       (0.02)       (0.02)       (0.02)       (0.03)       (0.03)
##
## details1                             0.10***      0.07***      0.06**                    0.60***
##                                      (0.02)       (0.02)       (0.02)                    (0.04)
##
## details2                                          0.14***      0.12***
##                                                   (0.02)       (0.02)
##
## details3                                                       0.06**
##                                                                (0.02)
##
## Constant                                                                    3.29***      2.84***
##                                                                             (0.07)       (0.08)
##
## ------------------------------------------------------------------------------------------------
## Observations            23,000       22,000       21,000       20,000       23,000       22,000
## R2                      0.79         0.79         0.79         0.80         0.05         0.06
## Adjusted R2             0.78         0.78         0.78         0.79         0.05         0.06
## Residual Std. Error     3.51         3.47         3.44         3.43         7.23         7.17
## F Statistic                                                               1,142.74***   690.46***
## ================================================================================================
## Note:
```

As a result of fixing each physician, the coefficients are still positively significant but less than the coefficients in models without fixed effects. This absorbs some of the variation and can account for differences in prescribing activities in physicians. Furthermore, the R-squared stat for the fixed effect models are much higher than the OLS models, indicating that the fixed effect model fits the data better.

Question 5

```
# Fixed time effects vs non time FE
lm5.1f =felm(scripts ~ detailing + details1 + details2
             + details3 | yrmn, data=counts )
lm5.1n =lm(scripts ~ detailing + details1 +
             details2 + details3, data=counts )

stargazer(lm5.1f, lm5.1n,
          title="Fixed Time VS. Normal", type="text",
          column.labels=c( "Fixed Time", "Normal"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Fixed Time VS. Normal
## ==================================================
##                           Dependent variable:
##                      ------------------------------
##                                 scripts
##                           felm             OLS
##                        Fixed Time        Normal
##                           (1)              (2)
## --------------------------------------------------
```

```
## detailing                0.34***         0.32***
##                          (0.04)           (0.04)
##
## details1                 0.28***         0.27***
##                          (0.04)           (0.04)
##
## details2                 0.38***         0.39***
##                          (0.04)           (0.04)
##
## details3                 0.45***         0.42***
##                          (0.04)           (0.04)
##
## Constant                                 2.41***
##                                           (0.08)
##
## ---------------------------------------------------
## Observations             20,000           20,000
## R2                        0.07             0.07
## Adjusted R2               0.07             0.07
## Residual Std. Error       7.13             7.14
## F Statistic                              376.37***
## ===================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

```r
# Physician FE vs Time AND Physician FE
lm5.2f = felm(scripts ~ detailing + details1 +
                details2 + details3 | id + yrmn, data = counts)
lm5.2n = felm(scripts ~ detailing + details1 +
                details2 + details3 | id , data = counts)

stargazer(lm5.2f, lm5.2n,
          title="Fixed Time VS. Normal", type="text",
          column.labels=c( "Fixed Time", "Normal"),
          df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Fixed Time VS. Normal
## ===================================================
##                          Dependent variable:
##                     -------------------------------
##                                scripts
##                        Fixed Time       Normal
##                           (1)             (2)
## ---------------------------------------------------
## detailing                0.091***        0.071***
##                          (0.020)          (0.020)
##
## details1                 0.081***         0.062**
##                          (0.021)          (0.021)
##
## details2                 0.129***        0.124***
##                          (0.021)          (0.021)
##
## details3                 0.086***         0.057**
##                          (0.021)          (0.021)
```

```
##
## ---------------------------------------------------
## Observations              20,000        20,000
## R2                        0.798          0.796
## Adjusted R2               0.787          0.785
## Residual Std. Error       3.420          3.433
## ===================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

When we include fixed time effects, we see that the coefficients are larger than an OLS regression. This is also controlling for seasonalities and other time related events.

Question 6

The current model doesn't take into account the effects of providing free samples to individual physicians on prescription writing. If more free samples are given out to a physician and they also detailed the drug, omission of free samples would bias the coefficient of detailing and overestimate its effect on scripts.

Question 7

```r
counts$newdata = counts$scripts - counts$script1 #Create first difference
did = lm(newdata ~ detailing, data = counts)

NoDiD =felm(scripts ~ detailing + details1 +
              details2 +details3 | id + yrmn, data=counts ) #significant
fixedDiD =felm(newdata ~ detailing + details1 +
                 details2 +details3 | id + yrmn, data=counts ) #Insig

stargazer(NoDiD, fixedDiD,
          title="Scripts vs Scripts Growth", type="text",
          column.labels=c( "Scripts", "Scripts Grwoth"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Scripts vs Scripts Growth
## ===================================================
##                          Dependent variable:
##                     -------------------------------
##                         scripts        newdata
##                         Scripts     Scripts Grwoth
##                           (1)            (2)
## ---------------------------------------------------
## detailing               0.09***          0.02
##                         (0.02)          (0.02)
##
## details1                0.08***         0.003
##                         (0.02)          (0.03)
##
## details2                0.13***          0.05
##                         (0.02)          (0.03)
##
## details3                0.09***         -0.05*
##                         (0.02)          (0.03)
##
## ---------------------------------------------------
```

```
## Observations            20,000         20,000
## R2                       0.80           0.01
## Adjusted R2              0.79          -0.04
## Residual Std. Error      3.42           4.20
## ====================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

In this dataset, using the difference in current scripts and 1 lagged scripts results in a change in number of scripts prescribed over time. We ran both regressions using the same model but with current and current-lagged1 scripts and found that the current-lagged1 resulted in insignificant coefficients. The problem is also that we are answering a different question where one is answering the rate of change over the last and the current scripts is the effects of detailing over the whole dataset/period of time.

Question 8

```r
# Created 9 different models to test
modela = felm(scripts ~ detailing + script1 | id, data=counts )
modelb = felm(scripts ~ detailing + script1 + script2 | id, data=counts)
modelc = felm(scripts ~ detailing + script1 +
                script2 + script3 | id, data=counts)

modeld = felm(scripts ~ detailing + details1 | id, data=counts)
modele = felm(scripts ~ detailing + details1 + details2 | id, data=counts)
modelf = felm(scripts ~ detailing + details1 +
                details2 + details3 | id, data=counts)

modelg = felm(scripts ~ detailing + script1 + details1 | id, data=counts)
modelh = felm(scripts ~ detailing + script1 +
                script2 + details1 + details2 | id, data=counts)
modeli = felm(scripts ~ detailing + script1 + script2 + script3 +
                details1 + details2 + details3 | id, data=counts)

stargazer(modela, modelb, modelc,
          title="Model Comparison", type="text",
          column.labels=c( "ModelA", "ModelB", "ModelC"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Model Comparison
## =====================================================
##                           Dependent variable:
##                       -------------------------------
##                                   scripts
##                         ModelA     ModelB     ModelC
##                          (1)        (2)        (3)
## -----------------------------------------------------
## detailing              0.11***    0.10***    0.10***
##                         (0.02)     (0.02)     (0.02)
##
## script1                0.28***    0.25***    0.25***
##                         (0.01)     (0.01)     (0.01)
##
## script2                           0.10***    0.08***
##                                    (0.01)     (0.01)
##
```

```
## script3                                        0.06***
##                                                (0.01)
##
## -----------------------------------------------------
## Observations          23,000     22,999    22,998
## R2                     0.80       0.80      0.81
## Adjusted R2            0.79       0.80      0.80
## Residual Std. Error    3.37       3.35      3.35
## =====================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

```r
stargazer(modeld, modele, modelf,
          title="Model Comparison", type="text",
          column.labels=c( "ModelD", "ModelE", "Model F"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Model Comparison
## =====================================================
##                              Dependent variable:
##                        ------------------------------
##                                   scripts
##                         ModelD     ModelE    Model F
##                          (1)        (2)        (3)
## -----------------------------------------------------
## detailing             0.10***    0.08***    0.07***
##                        (0.02)     (0.02)    (0.02)
##
## details1              0.10***    0.07***    0.06**
##                        (0.02)     (0.02)    (0.02)
##
## details2                         0.14***    0.12***
##                                   (0.02)    (0.02)
##
## details3                                    0.06**
##                                             (0.02)
##
## -----------------------------------------------------
## Observations          22,000     21,000    20,000
## R2                     0.79       0.79      0.80
## Adjusted R2            0.78       0.78      0.79
## Residual Std. Error    3.47       3.44      3.43
## =====================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

```r
stargazer(modelg, modelh, modeli,
          title="Model Comparison", type="text",
          column.labels=c( "ModelG", "ModelH", "Model I"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Model Comparison
## =====================================================
##                              Dependent variable:
##                        ------------------------------
```

```
##                                 scripts
##                       ModelG     ModelH    Model I
##                        (1)        (2)        (3)
## ---------------------------------------------------
## detailing            0.08***     0.05**     0.05*
##                       (0.02)     (0.02)    (0.02)
##
## script1              0.26***    0.22***    0.20***
##                       (0.01)     (0.01)    (0.01)
##
## script2                         0.13***    0.10***
##                                  (0.01)    (0.01)
##
## script3                                    0.11***
##                                            (0.01)
##
## details1             0.07***     0.05*      0.03
##                       (0.02)     (0.02)    (0.02)
##
## details2                        0.10***    0.10***
##                                  (0.02)    (0.02)
##
## details3                                    0.01
##                                            (0.02)
##
## ---------------------------------------------------
## Observations          22,000     21,000    20,000
## R2                     0.80       0.81      0.81
## Adjusted R2            0.80       0.80      0.80
## Residual Std. Error    3.34       3.30      3.29
## ===================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

Moving forward, we will use models lm5.2f and modelh as we think these models approximate the true model the closest. Below, we will compare these two models side-by-side:

```r
lm5.2f = felm(scripts~ detailing + details1 +
                details2 + details3 | id + yrmn, data = counts )

stargazer(lm5.2f, modelh,
          title="Fixed Time VS. Lag Scripts", type="text",
          column.labels=c( "Fixed Time", "Lag Scripts"),
          df=FALSE, digits=3, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Fixed Time VS. Lag Scripts
## ==================================================
##                            Dependent variable:
##                       -----------------------------
##                                  scripts
##                       Fixed Time     Lag Scripts
##                           (1)            (2)
## --------------------------------------------------
## detailing              0.091***        0.049**
##                         (0.020)        (0.019)
```

```
##
## script1                                0.220***
##                                        (0.007)
##
## script2                                0.130***
##                                        (0.007)
##
## details1               0.081***         0.046*
##                         (0.021)        (0.019)
##
## details2               0.129***        0.103***
##                         (0.021)        (0.019)
##
## details3               0.086***
##                         (0.021)
##
## ----------------------------------------------------
## Observations             20,000          21,000
## R2                        0.798           0.810
## Adjusted R2               0.787           0.800
## Residual Std. Error      3.420           3.304
## ====================================================
## Note:                  *p<0.05; **p<0.01; ***p<0.001
```

We chose these two models because:

- the data shows that past scriptwriting might have an influence on current scriptwriting
- because we think scriptwriting over time should be controlled for using fixed effects, and
- the data shows there is an optimal mix of lagged scripts and lagged detailing. The range of estimates for detailing is 0.05 - 0.09. This range of estimates is larger than the standard errors reported for both models (0.02)

Question 9

```r
# We choose Modelh = Preferred and we interactived mean_samples with current detailing and also lagged

prefmodel <- felm(log(scripts + 1) ~ log(detailing + 1) + log(details1 + 1) +
                log(details2 + 1) + log(details3 + 1) | id + yrmn, data = counts)

stargazer(prefmodel,
          title = "Regression Results", type = "text",
          column.labels = c("Preferred Model"),
          df = FALSE, digits = 2, star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## Regression Results
## ===================================================
##                           Dependent variable:
##                       -----------------------------
##                             log(scripts + 1)
##                              Preferred Model
## ---------------------------------------------------
## log(detailing + 1)              0.06***
##                                 (0.01)
##
```

```
## log(details1 + 1)              0.05***
##                                (0.01)
##
## log(details2 + 1)              0.06***
##                                (0.01)
##
## log(details3 + 1)              0.04***
##                                (0.01)
##
## -------------------------------------------------
## Observations                   20,000
## R2                              0.67
## Adjusted R2                     0.65
## Residual Std. Error             0.58
## =================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

Our model estimates that a 1% increase in detailing is associated with a 6% increase in current scripts. Past detailing also is positively associated with current scripts, ranging between 4% - 6%, depending on the lagged period.

Question 10

```
time.cluster = felm(scripts ~ detailing + details1 +
                details2 + details3 | id + yrmn | 0 | yrmn, data=counts)

id.cluster = felm(scripts ~ detailing + details1 +
                details2 + details3 | id + yrmn | 0 | id, data=counts)

summary(lm5.2f)
```

```
##
## Call:
##    felm(formula = scripts ~ detailing + details1 + details2 + details3 |      id + yrmn, data = coun
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.789  -1.507  -0.298   1.249  50.412
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## detailing  0.09099    0.02019   4.506 6.65e-06 ***
## details1   0.08145    0.02092   3.894 9.90e-05 ***
## details2   0.12903    0.02125   6.072 1.29e-09 ***
## details3   0.08602    0.02116   4.065 4.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.42 on 18977 degrees of freedom
##   (4000 observations deleted due to missingness)
## Multiple R-squared(full model): 0.7977   Adjusted R-squared: 0.7868
## Multiple R-squared(proj model): 0.008231   Adjusted R-squared: -0.04518
## F-statistic(full model):73.22 on 1022 and 18977 DF, p-value: < 2.2e-16
## F-statistic(proj model): 39.37 on 4 and 18977 DF, p-value: < 2.2e-16
```

```r
summary(lm5.2f, robust = TRUE)
```

```
##
## Call:
##    felm(formula = scripts ~ detailing + details1 + details2 + details3 |      id + yrmn, data = coun
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.789  -1.507  -0.298   1.249  50.412
##
## Coefficients:
##           Estimate Robust s.e t value Pr(>|t|)
## detailing  0.09099    0.02330   3.905 9.45e-05 ***
## details1   0.08145    0.02429   3.354 0.000798 ***
## details2   0.12903    0.02409   5.357 8.56e-08 ***
## details3   0.08602    0.02521   3.412 0.000646 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.42 on 18977 degrees of freedom
##   (4000 observations deleted due to missingness)
## Multiple R-squared(full model): 0.7977   Adjusted R-squared: 0.7868
## Multiple R-squared(proj model): 0.008231   Adjusted R-squared: -0.04518
## F-statistic(full model, *iid*):73.22 on 1022 and 18977 DF, p-value: < 2.2e-16
## F-statistic(proj model): 27.82 on 4 and 18977 DF, p-value: < 2.2e-16
```

```r
stargazer(time.cluster, id.cluster,
          title="Model Comparison", type="text",
          column.labels = c("Cluster Time", "Cluster ID"),
          df=FALSE, digits = 3, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Model Comparison
## ==================================================
##                         Dependent variable:
##                    ------------------------------
##                               scripts
##                      Cluster Time    Cluster ID
##                          (1)             (2)
## --------------------------------------------------
## detailing              0.091***       0.091***
##                        (0.022)         (0.027)
##
## details1               0.081***        0.081**
##                        (0.018)         (0.026)
##
## details2               0.129***       0.129***
##                        (0.025)         (0.028)
##
## details3               0.086***        0.086**
##                        (0.025)         (0.031)
##
## --------------------------------------------------
## Observations            20,000          20,000
```

```
## R2                          0.798           0.798
## Adjusted R2                 0.787           0.787
## Residual Std. Error         3.420           3.420
## =====================================================
## Note:                  *p<0.05; **p<0.01; ***p<0.001
```

- When we use robust standard errors, we observe an increase in standard errors (vs the normal SE), while the coefficients are still significant.
- After clustering SE by time and ID, we observe that the SE clustered by time is smaller relative to the SE clustered by ID.

  Even though the SE clustered by time is smaller, we know that there are different groups of physicians within the given dataset. Therefore, we will cluster by ID rather than time.

Question 11

```r
dt = merge(counts, demo, by = 'id')

general = felm(scripts ~ detailing + details1 + details2
              + details3 | id + yrmn | 0 | id, data=subset(dt, generalphys == 1))

spec = felm(scripts ~ detailing + details1 + details2
              + details3 | id + yrmn | 0 | id, data=subset(dt, specialist == 1))

stargazer(general, spec,
          title="Model Comparison", type="text",
          column.labels=c( "GenPhys", "Specialists"),
          df=FALSE, digits=2, star.cutoffs = c(0.05,0.01,0.001))
```

```
##
## Model Comparison
## =====================================================
##                             Dependent variable:
##                        ------------------------------
##                                   scripts
##                           GenPhys       Specialists
##                             (1)             (2)
## -----------------------------------------------------
## detailing                  0.03           0.29**
##                           (0.03)          (0.10)
##
## details1                   0.02           0.28**
##                           (0.02)          (0.10)
##
## details2                   0.06*          0.36**
##                           (0.02)          (0.12)
##
## details3                   0.02            0.22
##                           (0.03)          (0.12)
##
## -----------------------------------------------------
## Observations              12,020          3,700
## R2                         0.60            0.80
## Adjusted R2                0.58            0.79
## Residual Std. Error        2.69            5.79
```

```
## ====================================================
## Note:                *p<0.05; **p<0.01; ***p<0.001
```

- After subsetting the data into general physicians and specialists, we observe a significant and positive coefficient of detailing for specialists, but an insignificant coefficient for general physicians. We conclude that there is a positive association between detailing and scripts for specialists, but not for genral physicians.
- Other physician characteristics given in the dataset include mean samples given to physicians. However, the mean sample does not give us any relevant information w.r.t. time (of when the samples were given). Therefore, we did not include mean samples in our model.

Question 12

General physicians have a lower number of average and total scripts compared to specialists whom have higher numbers of scripts. As a result, a marketer should focus their detailing and marketing efforts on specialists.

Detailing is an effective marketing tool. However, it is much more effective when you target specialists but less effective on generalists. The impact has a positive association with specialists but less so on generalists based on our dataset. We are confident that targeting specialists will have a bigger impact than generalists.

Some sources of doubt may include unobserved variables that might affect the impact of detailing. These could include free samples given over time, salesperson competence and activity, competitive environment… The list is endless.

> Since specialists have a positive response to detailing, we would recommend the sales team to target specialists before targeting general physicians. However, due to the limitations of the given dataset, we would have to dive deeper into other forces that could influence physician's prescription writing behavior for this particular drug before investing further in this detailing marketing campaign.