

```
library(readxl)
library(reshape2)
library(ggplot2)
library(car)
```

```
## Loading required package: carData
```

```
library(stargazer)
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(sandwich)
library(tinytex)
library(rmarkdown)
library(lmtest)
```

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

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

Purpose: Familiarity with Economic Data sources and using Simple Linear Regression.

I obtained my data using the FRED tool in Excel and exported it at the quarterly level from 2001 to 2022.

Getting our data

```
mort_rate <- read_excel('MortgageRateData.xlsx', sheet = 'data')
mort_rate
```

```
## # A tibble: 88 x 10
```

```
##   date                MORTGAGE30US fedfunds CPILFESL  GDPC1 tb3ms  gs20 USSTHPI
##   <dtm>                <dbl>    <dbl>    <dbl>  <dbl> <dbl>  <dbl>  <dbl>
## 1 2001-01-01 00:00:00      7.01      5.59     184. 13219.  4.82  5.59   248.
## 2 2001-04-01 00:00:00      7.13      4.33     185. 13301.  3.66  5.84   252.
## 3 2001-07-01 00:00:00      6.96      3.5      187. 13248.  3.17  5.62   256.
## 4 2001-10-01 00:00:00      6.77      2.13     188. 13285.  1.91  5.48   259.
## 5 2002-01-01 00:00:00      6.97      1.73     189. 13395.  1.72  5.74   263.
## 6 2002-04-01 00:00:00      6.81      1.75     190. 13477.  1.71  5.77   267.
## 7 2002-07-01 00:00:00      6.29      1.74     191. 13532.  1.64  5.19   273.
## 8 2002-10-01 00:00:00      6.08      1.44     192. 13549.  1.33  5.02   276.
## 9 2003-01-01 00:00:00      5.84      1.25     192. 13619.  1.16  4.9    280.
## 10 2003-04-01 00:00:00      5.51      1.25     193. 13741.  1.04  4.59   283.
```

```
## # i 78 more rows
```

```
## # i 2 more variables: CP <dbl>, JTSLDL <dbl>
```

```
table_desc <- read_excel('MortgageRateData.xlsx', sheet = 'background')
```

```
## New names:
## * '' -> '...2'
## * '' -> '...3'
## * '' -> '...4'
## * '' -> '...5'
## * '' -> '...6'
## * '' -> '...7'
## * '' -> '...8'
```

```
table_desc
```

```
## # A tibble: 11 x 8
##   descriptions ...2 ...3 ...4 ...5 ...6 ...7 ...8
##   <chr> <chr> <chr> <chr> <chr> <dtm> <chr> <chr>
## 1 <NA> <NA> <NA> <NA> <NA> NA <NA> <NA>
## 2 date <NA> <NA> <NA> <NA> NA <NA> <NA>
## 3 MORTGAGE30US lin Percent, Not ~ q, a~ Quar~ 2001-01-01 00:00:00 30-Y~ Fred~
## 4 fedfunds lin Percent, Not ~ q, a~ Quar~ 2001-01-01 00:00:00 Fede~ Boar~
## 5 CPILFESL lin Index 1982-19~ q, a~ Quar~ 2001-01-01 00:00:00 Cons~ U.S.~
## 6 GDPC1 lin Billions of C~ q Quar~ 2001-01-01 00:00:00 Real~ U.S.~
## 7 tb3ms lin Percent, Not ~ q, a~ Quar~ 2001-01-01 00:00:00 3-Mo~ Boar~
## 8 gs20 lin Percent, Not ~ q, a~ Quar~ 2001-01-01 00:00:00 Mark~ Boar~
## 9 USSTHPI lin Index 1980:Q1~ q Quar~ 2001-01-01 00:00:00 All~ U.S.~
## 10 CP lin Billions of D~ q Quar~ 2001-01-01 00:00:00 Corp~ U.S.~
## 11 JTSLDL lin Level in Thou~ q, a~ Quar~ 2001-01-01 00:00:00 Layo~ U.S.~
```

3 - It think federal funds will increase mortgage rates, same with inflation and the housing price index. Using things like bonds may decrease when mortgage rates are increasing makes sense, and same with GDP going down. I brought in corporate profits and layoffs to see how all of this is correlated with our daily lives.

4 - OLS

```
raw_model <- lm(MORTGAGE30US ~ ., data = mort_rate)
summary(raw_model)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ ., data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4069 -0.1503  0.0022  0.1043  0.6205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.013e+00  8.811e-01  -1.150  0.253812
## date        -3.583e-09  1.976e-09  -1.814  0.073569 .
## fedfunds     1.492e-01  1.124e-01   1.327  0.188252
## CPILFESL     4.356e-02  1.069e-02   4.077  0.000109 ***
## GDPC1       -1.038e-04  1.130e-04  -0.919  0.361123
```

```
## tb3ms      3.052e-02  1.211e-01   0.252 0.801652
## gs20       7.750e-01  5.864e-02  13.215 < 2e-16 ***
## USSTHPI    6.308e-04  1.221e-03   0.517 0.606853
## CP        -6.675e-04  1.375e-04  -4.855 6.07e-06 ***
## JTSLDL     2.518e-05  5.557e-05   0.453 0.651794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2036 on 78 degrees of freedom
## Multiple R-squared:  0.9749, Adjusted R-squared:  0.972
## F-statistic: 336.1 on 9 and 78 DF,  p-value: < 2.2e-16
```

In short, we only yield 4 meaningful variables, 3 being usable (CPILFESL, gs20, CP). Another important note is our intercept not returning as significant. This can be problematic. We will ignore this issue until we continue to refine our model.

We can see our model is significant and covers a lot of variability. We shouldn't mistake this for over-fitting. Let's see how well our model reduces and its performance as we continue this journey.

5 - Multicollinearity using VIF

```
vif(raw_model)
```

```
##      date fedfunds CPILFESL   GDPC1    tb3ms    gs20  USSTHPI      CP
## 332.74016 69.91669 208.54764 100.58206 69.77715 11.28959 21.14989 14.07216
##      JTSLDL
##      1.80264
```

VIF score are very high, however, we are going to continue forward with: CPILFESL, gs20, CP

```
model <- lm(MORTGAGE30US ~ CPILFESL+gs20+CP, data = mort_rate)
summary(model)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ CPILFESL + gs20 + CP, data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71636 -0.19100 -0.01068  0.22969  0.57592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.4462269   0.7847829  -4.391 3.26e-05 ***
## CPILFESL      0.0226033   0.0038034   5.943 6.16e-08 ***
## gs20          1.1340980   0.0478757  23.688 < 2e-16 ***
## CP          -0.0006357   0.0001668  -3.810 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3088 on 84 degrees of freedom
## Multiple R-squared:  0.9377, Adjusted R-squared:  0.9355
## F-statistic: 421.8 on 3 and 84 DF,  p-value: < 2.2e-16
```

Immediately, we can see how well our model performs. High adjusted and standard R-squared. All variables and intercept are significant. Let's see how our residuals shake out against our variables.

```
vif(model)
```

```
## CPILFESL      gs20      CP
## 11.488862  3.271857  9.009518
```

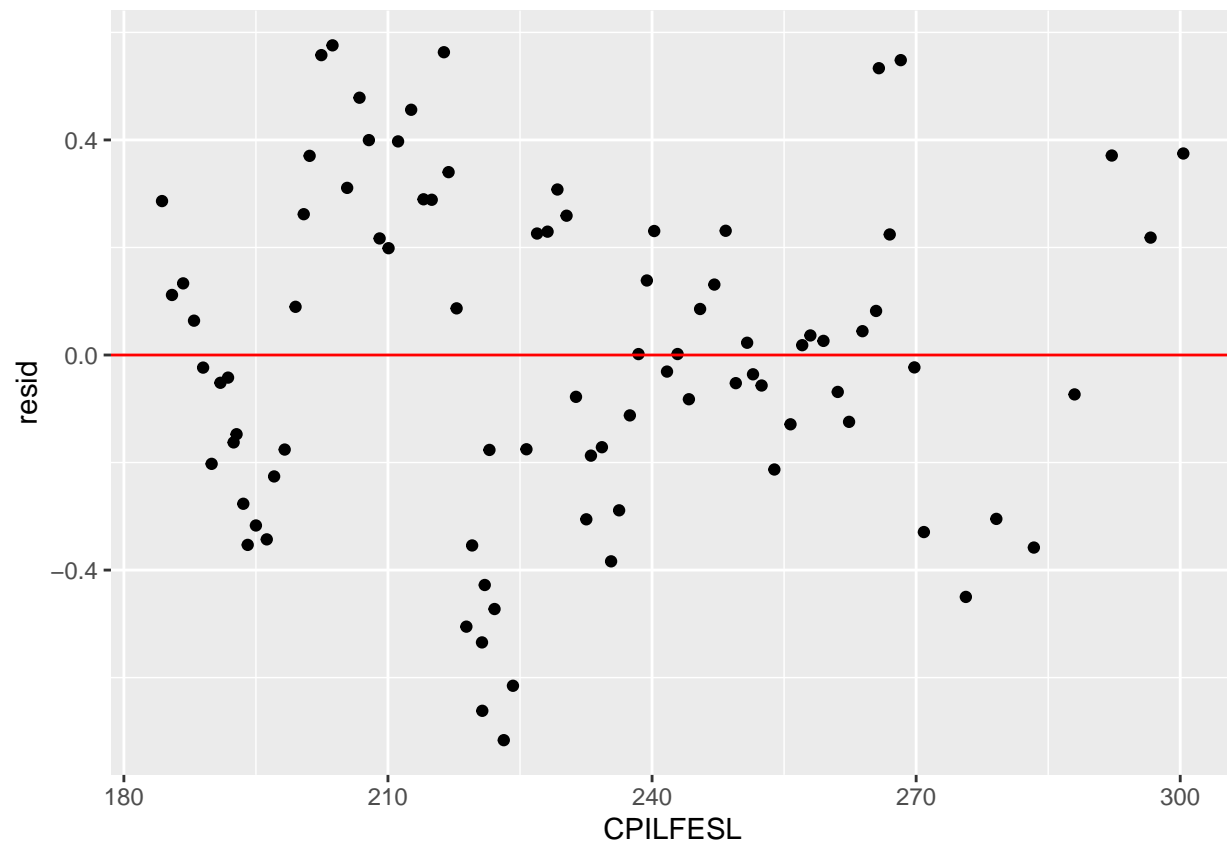
We can see how much we have reduced multicollinearity across our main variables.

```
mort_rate$resid <- resid(model)
mort_rate
```

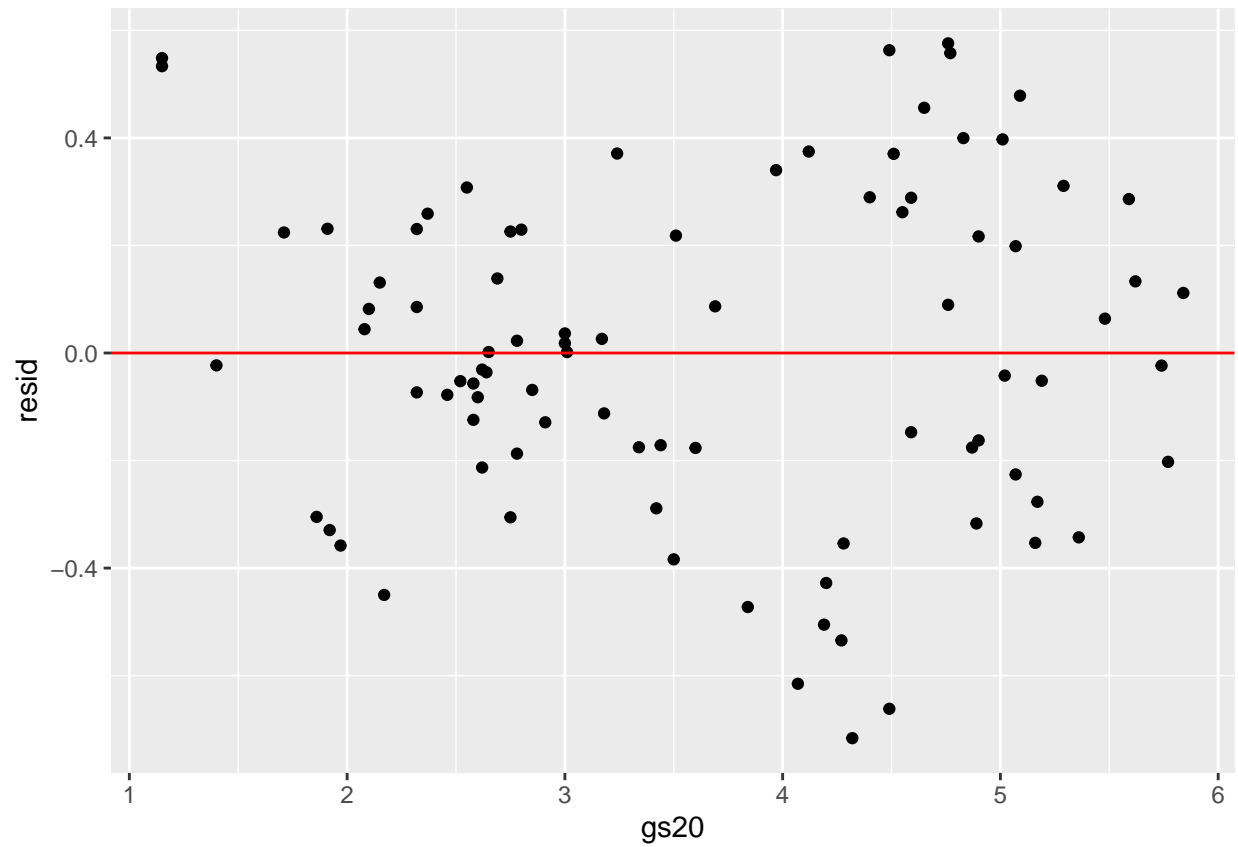
```
## # A tibble: 88 x 11
##   date                MORTGAGE30US fedfunds CPILFESL  GDPC1  tb3ms  gs20  USSTHPI
##   <dtm>                <dbl>      <dbl>    <dbl>  <dbl> <dbl> <dbl> <dbl>
## 1 2001-01-01 00:00:00      7.01      5.59     184. 13219.  4.82  5.59   248.
## 2 2001-04-01 00:00:00      7.13      4.33     185. 13301.  3.66  5.84   252.
## 3 2001-07-01 00:00:00      6.96      3.5      187. 13248.  3.17  5.62   256.
## 4 2001-10-01 00:00:00      6.77      2.13     188. 13285.  1.91  5.48   259.
## 5 2002-01-01 00:00:00      6.97      1.73     189. 13395.  1.72  5.74   263.
## 6 2002-04-01 00:00:00      6.81      1.75     190. 13477.  1.71  5.77   267.
## 7 2002-07-01 00:00:00      6.29      1.74     191. 13532.  1.64  5.19   273.
## 8 2002-10-01 00:00:00      6.08      1.44     192. 13549.  1.33  5.02   276.
## 9 2003-01-01 00:00:00      5.84      1.25     192. 13619.  1.16  4.9    280.
## 10 2003-04-01 00:00:00      5.51      1.25     193. 13741.  1.04  4.59   283.
## # i 78 more rows
## # i 3 more variables: CP <dbl>, JTSLDL <dbl>, resid <dbl>
```

6) Analyze the residuals against each of the x-variables using textplot or scatter. What do they say?

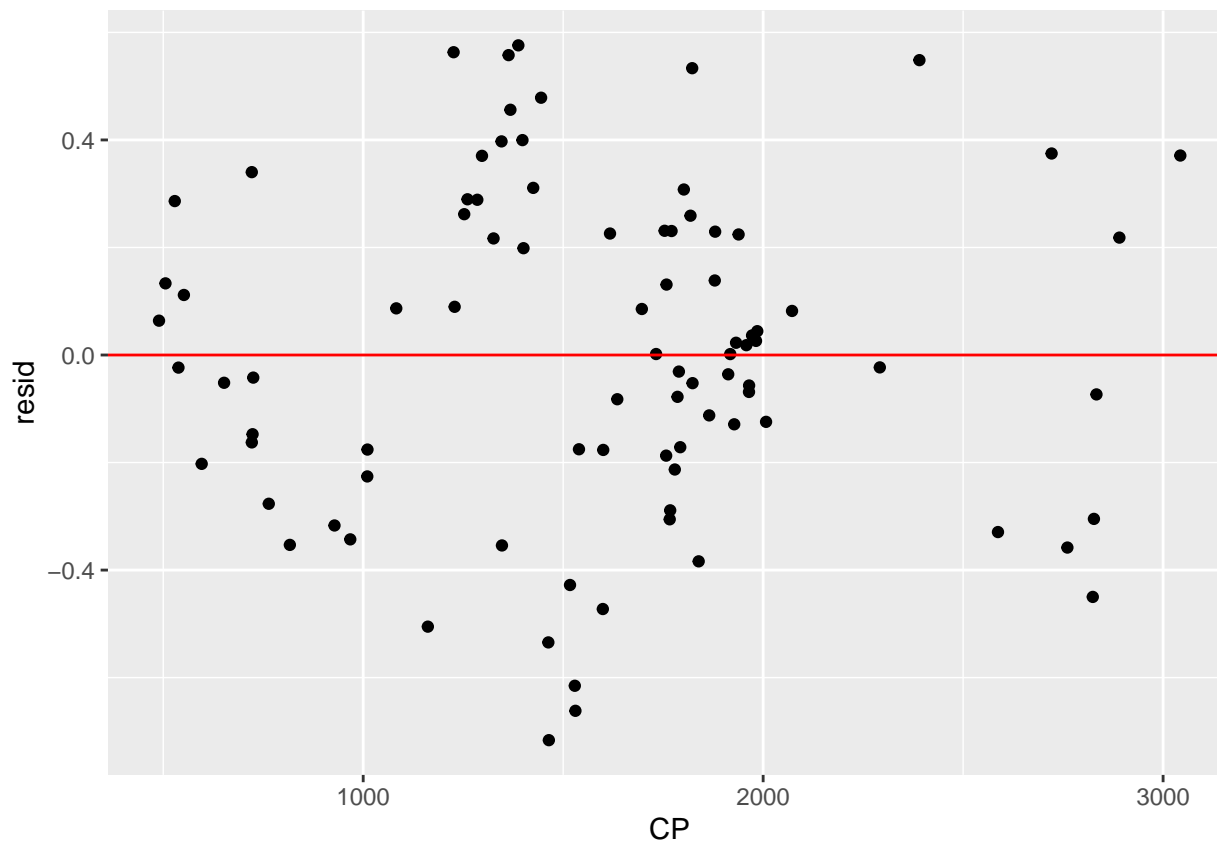
```
ggplot(data = mort_rate) +
  geom_point(mapping = aes(x = CPILFESL, y = resid)) +
  geom_hline(yintercept=0,col="red") #hline is for horizontal line y=0
```



```
ggplot(data = mort_rate) +  
  geom_point(mapping = aes(x = gs20, y = resid)) +  
  geom_hline(yintercept=0,col="red") #hline is for horizontal line y=0
```



```
ggplot(data = mort_rate) +  
  geom_point(mapping = aes(x = CP, y = resid)) +  
  geom_hline(yintercept=0,col="red") #hline is for horizontal line y=0
```



- 7) Change one or two variables into different format (logs, squares, etc.); if there's dummy vars, create interactive terms; then include them into your model.

```
mort_rate$sq_cp <- mort_rate$CP**2
mort_rate$log_CPIFESL <- log(mort_rate$CPIFESL)

model_adj <- lm(MORTGAGE30US ~ CPIFESL+gs20+CP+sq_cp+log_CPIFESL, data = mort_rate)
summary(model_adj)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ CPIFESL + gs20 + CP + sq_cp + log_CPIFESL,
##     data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6045 -0.2024 -0.0275  0.1987  0.7320
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.258e+02  4.135e+01   3.041 0.003165 **
## CPIFESL      1.441e-01  3.893e-02   3.701 0.000388 ***
## gs20         1.060e+00  5.892e-02  17.998 < 2e-16 ***
## CP           7.470e-04  5.083e-04   1.470 0.145500
## sq_cp        -4.028e-07  1.452e-07  -2.774 0.006850 **
```

```
## log_CPIFESL -2.908e+01  9.296e+00  -3.128 0.002435 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2953 on 82 degrees of freedom
## Multiple R-squared:  0.9444, Adjusted R-squared:  0.941
## F-statistic: 278.6 on 5 and 82 DF,  p-value: < 2.2e-16
```

- 8) Carry out HSK test, make conclusion; if there's a case of HSK, carry out WLS (choose what variable as the weight), obtain the estimates. Eliminate some variables as necessary. Modify your model accordingly. Check t, F, p-values for significance.

```
bptest(model, ~ fitted(model) + I(fitted(model)^2) )
```

```
##
## studentized Breusch-Pagan test
##
## data:  model
## BP = 2.0683, df = 2, p-value = 0.3555
```

- 9) As you start with the original model, checked for significance of x-variables, omit one at a time as necessary, each time re-estimate and justify each omission; Use the stargazer commands (in R) to gather your regression outputs later, giving you a more professional output comparison.

We are going to reduce by removing our highest p-value variable. After our raw model, we will remove JTSLDL

```
model_red1 <- lm(MORTGAGE30US~date+fedfunds+CPILFESL+GDPC1+tb3ms+gs20+USSTHPI+CP, data=mort_rate)
summary(model_red1)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ date + fedfunds + CPILFESL + GDPC1 +
##      tb3ms + gs20 + USSTHPI + CP, data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40863 -0.14636  0.00724  0.10334  0.62643
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.328e-01  7.823e-01  -1.065 0.290331
## date        -3.271e-09  1.842e-09  -1.776 0.079629 .
## fedfunds     1.526e-01  1.116e-01   1.368 0.175220
## CPILFESL     4.274e-02  1.048e-02   4.079 0.000107 ***
## GDPC1       -1.263e-04  1.009e-04  -1.252 0.214429
## tb3ms        2.944e-02  1.204e-01   0.244 0.807510
## gs20         7.691e-01  5.689e-02  13.519 < 2e-16 ***
## USSTHPI      8.527e-04  1.113e-03   0.766 0.445720
## CP          -6.973e-04  1.201e-04  -5.804 1.29e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.2026 on 79 degrees of freedom
## Multiple R-squared: 0.9748, Adjusted R-squared: 0.9722
## F-statistic: 381.9 on 8 and 79 DF, p-value: < 2.2e-16
```

Now we remove USSTHPI at p-value = .446

```
model_red2 <- lm(MORTGAGE30US~date+fedfunds+CPILFESL+GDPC1+tb3ms+gs20+CP, data=mort_rate)
summary(model_red2)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ date + fedfunds + CPILFESL + GDPC1 +
##      tb3ms + gs20 + CP, data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40326 -0.15001  0.00872  0.11565  0.63044
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.159e+00  6.548e-01  -1.770  0.08055 .
## date        -4.197e-09  1.386e-09  -3.027  0.00332 **
## fedfunds     1.507e-01  1.113e-01   1.354  0.17950
## CPILFESL     4.876e-02  6.913e-03   7.053 5.62e-10 ***
## GDPC1       -9.919e-05  9.428e-05  -1.052  0.29592
## tb3ms        3.776e-02  1.196e-01   0.316  0.75311
## gs20         7.644e-01  5.641e-02  13.550 < 2e-16 ***
## CP          -6.722e-04  1.153e-04  -5.830 1.12e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2021 on 80 degrees of freedom
## Multiple R-squared: 0.9746, Adjusted R-squared: 0.9724
## F-statistic: 438.7 on 7 and 80 DF, p-value: < 2.2e-16
```

We remove tb3ms

```
model_red3 <- lm(MORTGAGE30US~date+fedfunds+CPILFESL+GDPC1+gs20+CP, data=mort_rate)
summary(model_red3)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ date + fedfunds + CPILFESL + GDPC1 +
##      gs20 + CP, data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41023 -0.14948  0.00942  0.10821  0.61928
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -1.217e+00  6.251e-01  -1.946  0.05508 .
## date        -4.200e-09  1.379e-09  -3.047  0.00312 **
## fedfunds     1.851e-01  2.207e-02   8.385  1.30e-12 ***
## CPILFESL     4.909e-02  6.793e-03   7.226  2.46e-10 ***
## GDPC1        -1.013e-04  9.353e-05  -1.083  0.28216
## gs20         7.687e-01  5.441e-02  14.128  < 2e-16 ***
## CP          -6.694e-04  1.143e-04  -5.856  9.76e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2009 on 81 degrees of freedom
## Multiple R-squared:  0.9746, Adjusted R-squared:  0.9727
## F-statistic: 517.5 on 6 and 81 DF,  p-value: < 2.2e-16
```

Now we remove GDPC1

```
model_red4 <- lm(MORTGAGE30US~date+fedfunds+CPILFESL+gs20+CP, data=mort_rate)
summary(model_red4)
```

```
##
## Call:
## lm(formula = MORTGAGE30US ~ date + fedfunds + CPILFESL + gs20 +
##      CP, data = mort_rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45857 -0.15095  0.00122  0.11747  0.59451
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.517e+00  5.608e-01  -2.705   0.0083 **
## date        -5.027e-09  1.149e-09  -4.375  3.55e-05 ***
## fedfunds     1.700e-01  1.711e-02   9.932  1.02e-15 ***
## CPILFESL     4.834e-02  6.765e-03   7.146  3.34e-10 ***
## gs20         7.634e-01  5.424e-02  14.073  < 2e-16 ***
## CP          -7.066e-04  1.091e-04  -6.474  6.57e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2011 on 82 degrees of freedom
## Multiple R-squared:  0.9742, Adjusted R-squared:  0.9726
## F-statistic: 619.5 on 5 and 82 DF,  p-value: < 2.2e-16
```

```
model.lst = list(raw_model, model)
```

```
stargazer(model_red1,
  model_red2,
  model_red3,
  model_red4,
  title="Displaying results for multiple response variables",
  type = "latex",
  float = FALSE,
  report = "vcs*",
```

```
# se=lapply(model.lst, function(x) sqrt(diag(vcovHC(x, type = "HC1")))),
no.space = FALSE,
header=FALSE,
font.size = "small",
intercept.bottom = F,
column.labels = c("30 year Mortgage Rates", "30 year Mortgage Rates"),
column.separate = c(1, 4),
digits = 2,
t.auto = F,
p.auto = F,
notes.align = "l",
notes = c("lm() function with Robust SE"),
notes.append = TRUE
)
```

	<i>Dependent variable:</i>			
	MORTGAGE30US			
	30 year Mortgage Rates	30 year Mortgage Rates		
	(1)	(2)	(3)	(4)
Constant	−0.83 (0.78)	−1.16 (0.65)*	−1.22 (0.63)*	−1.52 (0.56)***
date	−0.00 (0.00)*	−0.00 (0.00)***	−0.00 (0.00)***	−0.00 (0.00)***
fedfunds	0.15 (0.11)	0.15 (0.11)	0.19 (0.02)***	0.17 (0.02)***
CPILFESL	0.04 (0.01)***	0.05 (0.01)***	0.05 (0.01)***	0.05 (0.01)***
GDPC1	−0.0001 (0.0001)	−0.0001 (0.0001)	−0.0001 (0.0001)	
tb3ms	0.03 (0.12)	0.04 (0.12)		
gs20	0.77 (0.06)***	0.76 (0.06)***	0.77 (0.05)***	0.76 (0.05)***
USSTHPI	0.001 (0.001)			
CP	−0.001 (0.0001)***	−0.001 (0.0001)***	−0.001 (0.0001)***	−0.001 (0.0001)***
Observations	88	88	88	88
R ²	0.97	0.97	0.97	0.97
Adjusted R ²	0.97	0.97	0.97	0.97
Residual Std. Error	0.20 (df = 79)	0.20 (df = 80)	0.20 (df = 81)	0.20 (df = 82)
F Statistic	381.94*** (df = 8; 79)	438.68*** (df = 7; 80)	517.53*** (df = 6; 81)	619.50*** (df = 5; 82)

Note: *p<0.1; **p<0.05; ***p<0.01
lm() function with Robust SE

- 10) What is your final model? Do the signs or the (non)linearity in the final model agree with your intuition?

Our final model is our 4th reduction. All variables are significant and we have high coverage of total variability.

Based on my initial feeling, I was right about federal fundings leading to increases related to the price index. Corporate profits leading to decreases in mortgage rates didn't align with my expectations, but it makes sense. Banks typically lower rates when they earn more money.

- 11) Write up a cover page for your assessment and conclusion, and attach your summary of outputs from stargazer; you can put the other regression outputs on Appendix at the very last page. Make it organized and presentable. Submit your report here on Canvas.