- Autocorrellation issues: I want to find the differences between closures/revenue based on income!
  - 2 different attempts at correcting
    - Using the base state-level data after the initial dip and applying cochrane-orcutt.
      - Performed initial diagnostic via Durbin-Watson, found statistically significant presence of first-order autocorrelation
      - Applied Cochrane-Orcutt correction, for our revenue analysis we go from highly positively correlated to mildly negatively correlated. We can't reject the null for the DW test after applying the Cochrane-Orcutt correction.
      - The estimates I care about, primarily the difference in outcomes between high and low income brackets, remain statistically significant and pretty similar to our first set of plain lm estimates.
      - However, upon visualization of residuals in both the cochrane-orcutt and non-transformed regressions, I'm worried that this isn't enough to actually deal with my problem, as the residuals show a clear pattern stil
    - Rebuilding the county level data with income brackets (think this is my winner)
      - First, I took the county level data, which only includes the change in revenue/closures by county, with no information about income catagory.
      - Then, I pulled median income on a per-county basis from the US census data for 2019, and merged that with our county-level small business data.
      - Then, I reconstructed income quartiles and catagories, and assigned each county a high, medium, or low dummy variable.
      - This let me create panel data! After that, I cut the data to a point after the big drop. Upon further analysis, and applying autocorrelation diagnostics like DW again, I found no statistically significant evidence of autocorrelation, and the residuals look significantly more random. Progress!
    - Plus, I found a neat tool that lets me run R code inside LyX. The future is in fact, now.

```
library(tidyverse)

## - Attaching packages --------------------- tidyverse 1.3.0 -
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
```

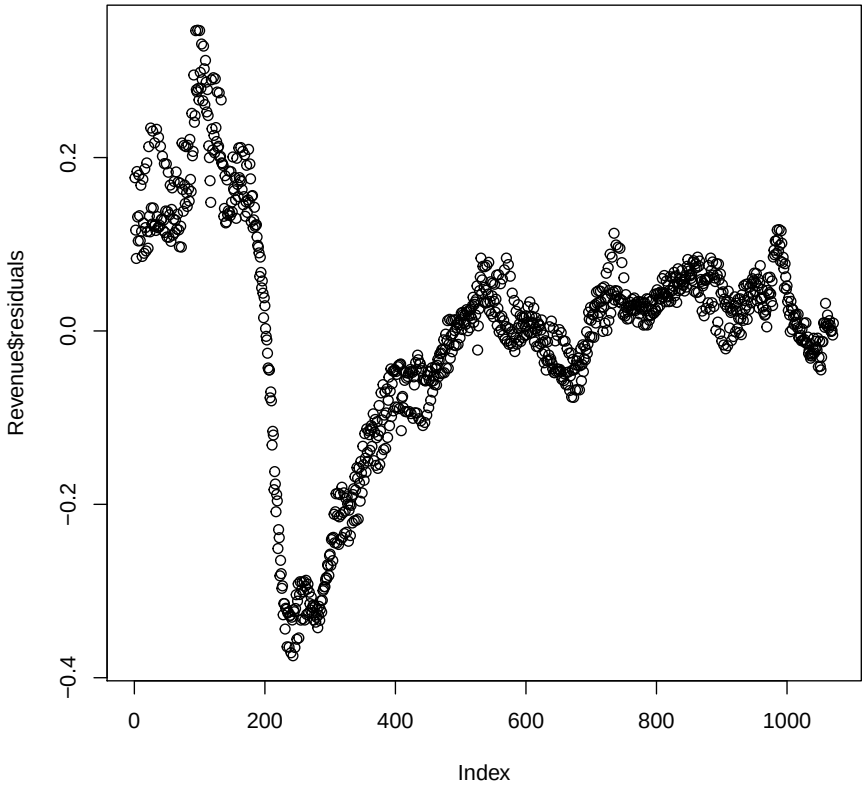
```
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## - Conflicts ---------------------- tidyverse_conflicts() -
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
setwd("~/Documents/Thesis/")
sbTest<-read.csv("~/Documents/Thesis/data/EconomicTracker-main/data/Womply - State - Daily.c
sbTest<- (unite(sbTest, "date", c("year","month","day"), sep = "/"))
sbTest$date <- as.Date(sbTest$date)
sbTestCA <- filter(sbTest, statefips == 6)
sbTestCA$post <- 0
sbTestCA$post <- as.factor(ifelse(sbTestCA$date >= "2020-03-19",sbTestCA$post+1,sbTestCA$pos
sbTestbyclass <- pivot_longer(sbTestCA, 4:6, names_to = "merchantsAll")
sbTestRevenue<- dplyr::select(sbTestCA, date,revenue_inchigh, revenue_incmiddle, revenue_inc
Revenuebyclass<- pivot_longer(sbTestRevenue, 2:4, names_to="MerchantClass", values_to="Perce
Revenue<-(lm(PercentDelta~MerchantClass+post+date,Revenuebyclass))

RevenueCut<- dplyr::filter(Revenuebyclass, date>="2020-06-20")
ClosuresCut<- dplyr::filter(Revenuebyclass, date>="2020-06-20")

Revenue<-(lm(PercentDelta~MerchantClass+date,Revenuebyclass))
RegCutRevenue<-(lm(PercentDelta~MerchantClass+date,RevenueCut))
library(orcutt)
```

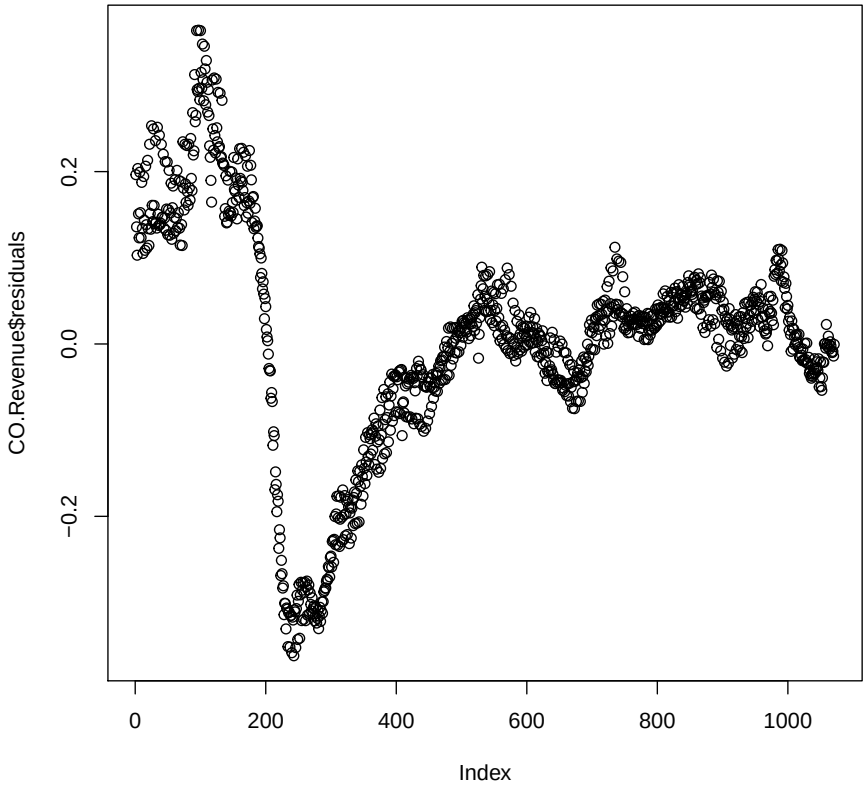
```
## Loading required package:  lmtest
## Loading required package:  zoo
##
## Attaching package:  'zoo'
## The following objects are masked from 'package:base':
##
##    as.Date, as.Date.numeric
```

```
dwtest(Revenue)
```

```
CO.Revenue<-cochrane.orcutt(Revenue)
CO.RevenueCut<-cochrane.orcutt(RegCutRevenue)
plot(Revenue$residuals)
```




```
plot(CO.Revenue$residuals)
```




```
summary(CO.Revenue)

## Call:
## lm(formula = PercentDelta ~ MerchantClass + date, data = Revenuebyclass)
##
##                                   Estimate  Std. Error t value Pr(>|t|)
## (Intercept)                     8.15345072  5.35995704   1.521   0.1285
## MerchantClassrevenue_inclow     0.07684272  0.00155866  49.300   <2e-16 ***
## MerchantClassrevenue_incmiddle  0.07334534  0.00154867  47.360   <2e-16 ***
## date                           -0.00045750  0.00029038  -1.576   0.1154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0351 on 1067 degrees of freedom
## Multiple R-squared:  0.7485 ,  Adjusted R-squared:  0.748
## F-statistic: 1057.3 on 2 and 1067 DF,  p-value: < 7.503e-319
```

```
##
## Durbin-Watson statistic
## (original):    0.07018 , p-value: 5.912e-219
## (transformed): 2.89545 , p-value: 1e+00
```

```
summary(CO.RevenueCut)
```

```
## Call:
## lm(formula = PercentDelta ~ MerchantClass + date, data = RevenueCut)
##
##                                  Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)                     7.4376e+00  1.1479e+00   6.479 1.967e-10 ***
## MerchantClassrevenue_inclow     8.5497e-02  1.7082e-03  50.051 < 2.2e-16 ***
## MerchantClassrevenue_incmiddle  7.8850e-02  1.7060e-03  46.220 < 2.2e-16 ***
## date                           -4.1799e-04  6.1942e-05  -6.748 3.635e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0249 on 581 degrees of freedom
## Multiple R-squared:  0.8452 ,  Adjusted R-squared:  0.8446
## F-statistic: 1055.4 on 2 and 581 DF,  p-value: < 2.008e-234
##
## Durbin-Watson statistic
## (original):    0.59274 , p-value: 3.811e-65
## (transformed): 2.53693 , p-value: 1e+00
```

```
#Next up is the county-level data, reconstruction, residual analysis etc.
```

```
library(plm)
```

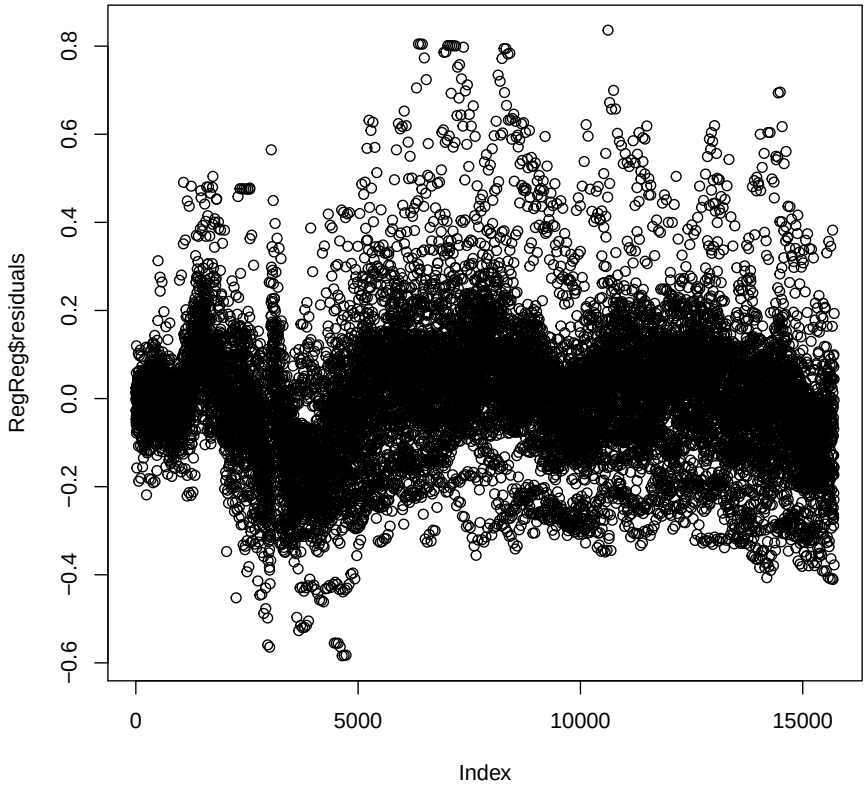
```
##
## Attaching package:  'plm'
## The following objects are masked from 'package:dplyr':
##
##    between, lead
```

```
library(readxl)
library(miceadds)
```

```
## Loading required package:  mice
##
## Attaching package:  'mice'
## The following object is masked from 'package:stats':
##
##    filter
## The following objects are masked from 'package:base':
##
##    cbind, rbind
```

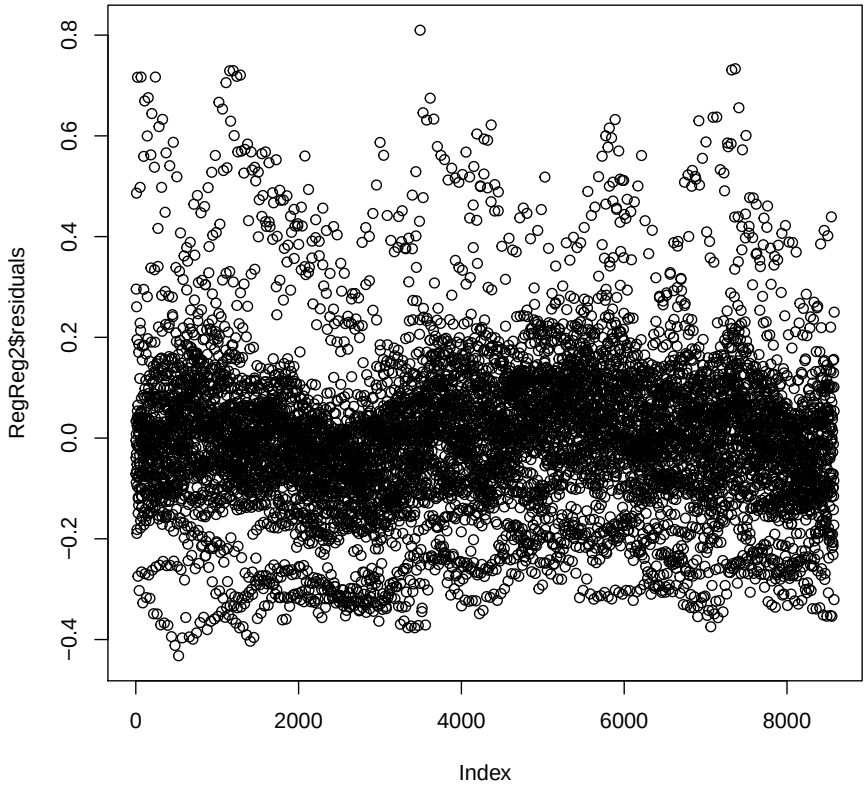
```
## * miceadds 3.11-6 (2021-01-21 11:48:47)

SBcounty<-read.csv("~/Documents/Thesis/data/EconomicTracker-main/data/Womply - County - Dail
countylevelincome <- read_excel("data/countylevelincome.xlsx", skip = 1)
CAincome<- select(countylevelincome,1,2,12)
CAincome$FIPS<-as.numeric(CAincome$FIPS)

## Warning:  NAs introduced by coercion

SBcountyCA<-filter(SBcounty, grepl("^6", countyfips))
SBcountyCA$FIPS<-SBcountyCA$countyfips
CAincome$quartile<-ntile(CAincome$`Median Household Income (2019)`,4)
FullData<-full_join(SBcountyCA,CAincome,by="FIPS")
DataCut<-select(FullData,1:6,10)
DataCut<-(unite(DataCut, "date", c("year","month","day"), sep = "/"))
DataCut$date <- as.Date(DataCut$date)

DataCut$lowInc<-0
DataCut$middleInc<-0
DataCut$highInc<-0
DataCut$highInc<-ifelse(DataCut$quartile == 4, DataCut$highInc<-1, DataCut$highInc<-0)
DataCut$lowInc<-ifelse(DataCut$quartile == 1, DataCut$lowInc<-1, DataCut$lowInc<-0)
DataCut$middleInc<-ifelse(DataCut$lowInc == 0 & DataCut$highInc == 0, DataCut$middleInc<-1,
DataCut$post <- 0
DataCut$post <-(ifelse(DataCut$date >= "2020-03-19",DataCut$post+1,DataCut$post+0))

ClusterReg<-lm.cluster(DataCut,revenue_all~lowInc+middleInc+highInc+post,DataCut$countyfips)
RegReg<-lm(revenue_all~lowInc+middleInc+highInc+post+date,DataCut)
summary(ClusterReg)

## R^2= 0.37628
##
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) -0.01754201 0.02052287  -0.8547546 3.926870e-01
## lowInc       0.05671597 0.03964876   1.4304601 1.525850e-01
## middleInc    0.05395429 0.03297114   1.6364099 1.017538e-01
## post        -0.30534532 0.01780072 -17.1535424 5.913517e-66

summary(RegReg)

##
## Call:
## lm(formula = revenue_all ~ lowInc + middleInc + highInc + post +
##     date, data = DataCut)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.58391 -0.09235 -0.00082  0.08281  0.83622
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.302e+00  3.064e-01  -14.04   <2e-16 ***
## lowInc       5.672e-02  3.864e-03   14.68   <2e-16 ***
## middleInc    5.395e-02  2.829e-03   19.07   <2e-16 ***
## highInc             NA         NA      NA       NA
## post        -3.471e-01  4.368e-03  -79.47   <2e-16 ***
## date         2.340e-04  1.674e-05   13.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1577 on 15703 degrees of freedom
##   (17 observations deleted due to missingness)
## Multiple R-squared:  0.384,Adjusted R-squared:  0.3838
## F-statistic:  2447 on 4 and 15703 DF,  p-value: < 2.2e-16

library(orcutt)
dwtest(RegReg)

##
##  Durbin-Watson test
##
## data:  RegReg
## DW = 1.6469, p-value = 0.7603
## alternative hypothesis: true autocorrelation is greater than 0

plot(RegReg$residuals)
```




```
plmtest<-(plm(revenue_all~lowInc+middleInc+highInc+post+date,DataCut))

## at least one couple (id-time) has NA in at least one index dimensionin resulting pdata.fr
##  to find out which, use e.g., table(index(your_pdataframe), useNA = "ifany")

summary(plmtest)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = revenue_all ~ lowInc + middleInc + highInc + post +
##     date, data = DataCut)
##
## Balanced Panel: n = 357, T = 44, N = 15708
##
## Residuals:
```

```
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.51485087 -0.07696476 -0.00090759  0.06836299  0.80488501
##
## Coefficients: (1 dropped because of singularities)
##             Estimate Std. Error t-value  Pr(>|t|)
## lowInc    0.0567160  0.0035105  16.156 < 2.2e-16 ***
## middleInc 0.0539543  0.0025707  20.988 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    325.31
## Residual Sum of Squares: 315.13
## R-Squared:      0.031272
## Adj. R-Squared: 0.0086776
## F-statistic: 247.746 on 2 and 15349 DF, p-value: < 2.22e-16

DataCut2<- dplyr::filter(DataCut, date>="2020-06-20")
RegReg2<-lm(revenue_all~lowInc+middleInc+date,DataCut2)
plot(RegReg2$residuals)
```




```
dwtest(RegReg2)

##
##  Durbin-Watson test
##
## data:  RegReg2
## DW = 1.9725, p-value = 0.09983
## alternative hypothesis: true autocorrelation is greater than 0

summary(RegReg2)

##
## Call:
## lm(formula = revenue_all ~ lowInc + middleInc + date, data = DataCut2)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.43225 -0.08896 -0.00198  0.07673  0.80980
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.862e+00  5.538e-01  16.004  < 2e-16 ***
## lowInc       4.195e-02  5.165e-03   8.121 5.25e-16 ***
## middleInc    5.444e-02  3.782e-03  14.392  < 2e-16 ***
## date        -4.944e-04  2.988e-05 -16.544  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1558 on 8576 degrees of freedom
## Multiple R-squared:  0.05342,Adjusted R-squared:  0.05308
## F-statistic: 161.3 on 3 and 8576 DF,  p-value: < 2.2e-16
```