Tingwei Adeck

Econometrics Portfolio

ECON 6320

West Texas A&M University

Instructor: Ryan Mattson, PhD

Date: November, 2016

A majority of R platform figures have descriptions of the code in them.

CIDM 6320- HOMEWORK ONE

PROOF EXERCISES

<u>Proof One</u>

Show that $\frac{\sum_i Y_i Z_i}{\sum_i Z_i^2} = \hat{y}_i$

Solution

We start the solution with a simplified linear regression model, where we do not include the y-intercept of the straight line:

$$Y_i = y_i Z_i + \varepsilon_i \qquad (1)$$

Where Y_i= dependent variable, y_i=gradient, Z_i= independent variable, ε_i=random error term

Remember that $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$……………. (1.5) (equation of the residual)

In this case:

$\hat{Y}_i = \hat{y}_i Z_i$ (equation of the fitted line)

(2)

This implies that:

$$\hat{\varepsilon}_i = Y_i - \hat{y}_i Z_i \qquad (3)$$

Remember that, if $a = (b + c)$ then, $a^2 = (b + c)^2$………………... (3.5)

This implies that:

$$\hat{\varepsilon}_i^2 = (Y_i - \hat{y}_i Z_i)^2 \qquad (4)$$

Remember

$\sum_{i=1}^{i} a^2 = \sum_{i=1}^{i} (c + b)^2$ ……………… (4.5)

This implies that:

$$\sum \hat{\varepsilon}_i^2 = \sum (Y_i - \hat{y}_i Z_i)^2 \qquad (5)$$

Take the derivative of (5):

$$\frac{d \sum \hat{\varepsilon}_i^2}{d\hat{y}_i} = \sum 2(Y_i - \hat{y}_i Z_i) Z_i \qquad (6)$$

A majority of R platform figures have descriptions of the code in them.

Methodology for the derivative (Chain rule):

$Let\ w = (Y_i - \hat{y}_i Z_i)^2$

$Let\ u = (Y_i - \hat{y}_i Z_i)^2$

$Then\ w = u^2$ (7)

$Now\ \frac{dw}{du} = 2u$…………… *(7.5)*

$Then\ \frac{du}{d\hat{y}_i} = Z_i$ ……………….. *(8)*

Multiply (7.5) by (8)

$$\frac{dw}{d\hat{y}_i} = \frac{dw}{du} * \frac{du}{d\hat{y}_i} = 2u * Z_i = 2(Y_i - \hat{y}_i Z_i) * Z_i \quad (9)$$

We differentiate in order to find out what values of $\hat{y}_i$ minimize the sum of squared residuals. By differentiating and setting it to 0, we can find the minimum or maximum turning point. By taking a derivative of the first derivative, we can also determine whether it is a maximum or minimum turning point.

We set the derivative in (6) above to 0, like we were looking for the turning points:

$$\sum 2(Y_i - \hat{y}_i Z_i)\, Z_i = 0 \quad (10)$$

We divide both sides by 2:

$$\sum (Y_i - \hat{y}_i Z_i)\, Z_i = 0 \quad (11)$$

We expand equation (11):

$$\sum Y_i Z_i - \hat{y}_i Z_i^2 = 0 \quad (12)$$

Remember that:

$\sum_{i=1}^{i}(a_i - b_i) = \sum_{i=1}^{i} a_i - \sum_{i=1}^{i} b_i$………………. (12.5)

This implies that (12) can be written as:

$$\sum Y_i Z_i - \sum \hat{y}_i {Z_i}^2 = 0 \quad (13)$$

In equation (13), the summation law of differences as illustrated in equation (12.5) is used to break up the summation across the difference.

A majority of R platform figures have descriptions of the code in them.

We carry the terms to opposite sides:

$$\sum Y_i Z_i = \sum \hat{y}_i Z_i^2 \qquad (14)$$

Remember that $\hat{y}_i$ is a constant***

This implies that:

$$\sum Y_i Z_i = \hat{y}_i \sum Z_i^2 \qquad (15)$$

Therefore:

$$\hat{y}_i = \frac{\sum Y_i Z_i}{\sum Z_i^2} \qquad (16)$$

QED

<u>Proof Two</u>

Show that: $E[\hat{y}_i] = y_i$ ………. (16.5)

Where $E[\hat{y}_i]$= Expected value of our fitted value and y_i= dependent variable

We assume that y_i and $\hat{y}_i$ are constants at value $i = 1$ ****

$\hat{y}_i$ has components of Y_i in its equation, which makes it a random variable. Because values of Y_i are dependent on the random variable ε_i , by association $\hat{y}_i$ is a random variable.

Because we are trying to prove equation (16.5), it is only logical to arrange (16) in a way that shows a relationship between $\hat{y}_i$ and y_i . To achieve this, we substitute Y_i in the $\hat{y}_i$ equation (16).

Remember Y_i is given in equation (1):

Therefore:

$$\hat{y}_i = \frac{\sum (y_i Z_i + \varepsilon_i) Z_i}{\sum Z_i^2} \qquad (17)$$

We distribute Z_i in the numerator, that means we multiply Z_i through the bracket:

$$\hat{y}_i = \frac{\sum (y_i Z_i^2 + \varepsilon_i Z_i)}{\sum Z_i^2} \qquad (18)$$

By the law of addition, we can separate the variables above into additive pairs

A majority of R platform figures have descriptions of the code in them.

$$\hat{y}_i = \frac{\sum y_i Z_i^2}{\sum Z_i^2} + \frac{\sum \varepsilon_i Z_i}{\sum Z_i^2} \tag{19}$$

$\sum Z_i^2$ is a common factor and so it can be placed separately under both sides of the additive pairs:

$$\hat{y}_i = y_i \frac{\sum Z_i^2}{\sum Z_i^2} + \frac{\sum \varepsilon_i Z_i}{\sum Z_i^2} \tag{20}$$

The distributive property is used above to factor out the constant y_i from $y_i \frac{\sum Z_i^2}{\sum Z_i^2}$

We assume that y_i and $\hat{y}_i$ are constants at value $i = 1$ ****

The equation becomes:

$$\hat{y}_i = y_i + \frac{\sum \varepsilon_i Z_i}{\sum Z_i^2} \tag{21}$$

We use the statistical concept of expected value (the average value of large number of realizations of a random variable) to show that $\hat{y}_i$ is unbiased.

$$E[\hat{y}_i] = E[y_i] + E\left[\frac{\sum \varepsilon_i Z_i}{\sum Z_i^2}\right] \tag{22}$$

The expected value of a fixed number is that number, meaning $E[y_i] = y_i$

The expected value of an expected number remains an expectation, meaning $E[\hat{y}_i] = E[\hat{y}_i]$

This implies that:

$$E[\hat{y}_i] = y_i + E\left[\frac{\sum \varepsilon_i Z_i}{\sum Z_i^2}\right] \tag{23}$$

Remember that:

$E[k * g(\varepsilon)] = k * E[g(\varepsilon)]$ for constant k and random function $g(\varepsilon)$.

From (22), $\frac{1}{\sum Z_i^2}$ is our constant and $\sum \varepsilon_i Z_i$ is a function of random variables.

$$E[\hat{y}_i] = y_i + \frac{1}{\sum Z_i^2} E[\sum \varepsilon_i Z_i] \tag{24}$$

The expectation of a sum is the sum of its expectations. Based on this statement, we move the expectation sign into the summation:

$$E[\hat{y}_i] = y_i + \frac{1}{\sum Z_i^2} \sum E[\varepsilon_i Z_i] \tag{25}$$

A majority of R platform figures have descriptions of the code in them.

For the proof to occur, ε_i and Z_i should be uncorrelated. We would show that this condition is equivalent to $E[\varepsilon_i Z_i] = 0$

If ε_i and Z_i are uncorrelated then the covariance of ε_i and Z_i equals 0.

$$Cor(Z_i, \varepsilon_i) = \frac{cov(Z_i,\varepsilon_i)}{\sqrt{var(Z_i)var(\varepsilon_i)}} \quad (26)$$

We use the definition of covariance and set it to 0:

$$cov(Z_i, \varepsilon_i) = E[(Z_i - \mu_z)(\varepsilon_i - \mu_\varepsilon)] = 0 \quad (27)$$

We multiply out the covariance equation:

$$E[Z_i\varepsilon_i - Z_i\mu_\varepsilon - \mu_z\varepsilon_i + \mu_z\mu_\varepsilon)] = 0 \quad (28)$$

We use the fact that the expectation of a sum is the sum of its expectations:

$$E[Z_i\varepsilon_i] - E[Z_i\mu_\varepsilon] - E[\mu_z\varepsilon_i] + E[\mu_z\mu_\varepsilon] = 0 \quad (29)$$

We will use the fact that μ_ε and μ_z are fixed numbers and pull them out of the expectations:

$$E[Z_i\varepsilon_i] - \mu_\varepsilon E[Z_i] - \mu_z E[\varepsilon_i] + \mu_\varepsilon \mu_z = 0 \quad (30)$$

We assume that $\mu_\varepsilon = 0$,which simply means that the mean of our error term is simply not relevant. This allows us to cancel any terms with μ_ε and $E[\varepsilon_i]$:

This implies that:

$$E[Z_i\varepsilon_i] = 0 \quad (31)$$

We arrived at (30) by substituting 0 for any variables that had μ_ε and $E[\varepsilon_i]$:

We substitute 0 back in (25) and it becomes:

$$E[\hat{y}_i] = y_i + \frac{1}{\sum Z_i^2} E[\sum \varepsilon_i Z_i] \quad \ldots\ldots (25)$$

$$E[\hat{y}_i] = y_i \quad (32)$$

Therefore:

$E[\hat{y}_i] = y_i$ (Proven)

QED

A majority of R platform figures have descriptions of the code in them.

Replication Exercises

Chapter 6, Question 2

Part A

TABLE 6.11 Variables for Monetary Policy Data

Variable	Description
FEDFUNDS	Effective federal funds rate (in percent)
lag_FEDFUNDS	Lagged effective federal funds rate (in percent)
Democrat	Democrat = 1, Republican = 0
Election	Quarters since previous election (0–15)
Inflation	Annualized inflation rate (1 percent inflation = 1.00)
DATE	Date

Figure 1: Codebook for Exercise.

Democrats

Here we attempt to do a scatter plot of Federal Funds rate (dependent Variable-Y-axis) to the quarters since the previous election for Democrats (independent Variable-X-axis).

The scatter plots are performed in Microsoft Excel & the R platform and the results are presented below.

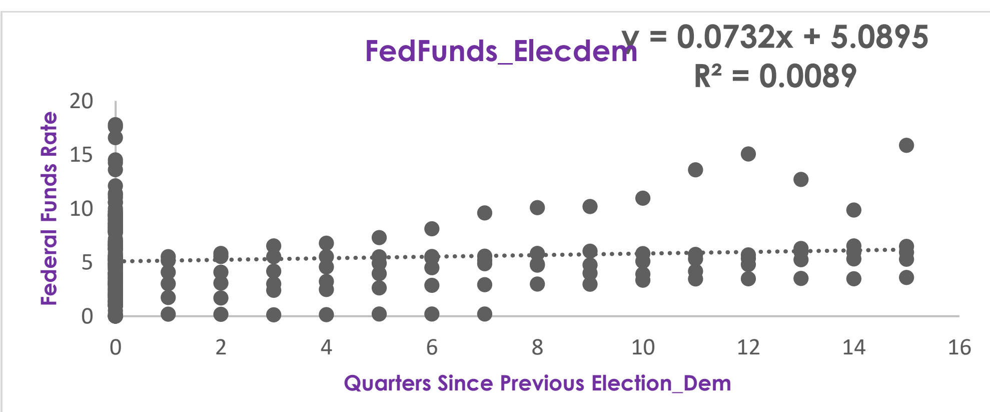


Figure 2: Fed Funds Rate (FFR) to Quarters since Previous Elections for Democrats.

A majority of R platform figures have descriptions of the code in them.

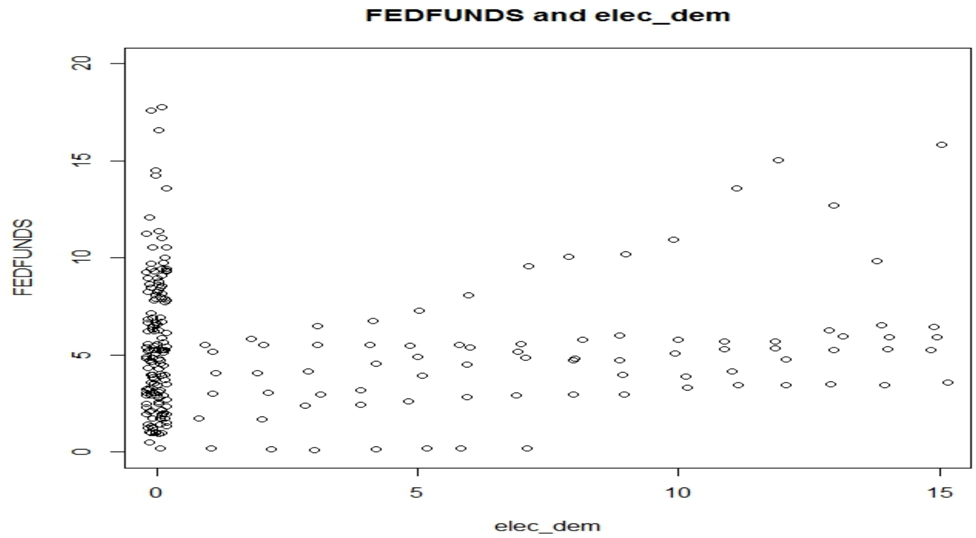


Figure 3: Fed Funds Rate (FFR) to Quarters since Previous Elections for Democrats (R Platform).

Untitled - R Editor

```
> library(XLConnect)
#call for XLConnect package
Loading required package: XLConnectJars
XLConnect 0.2-12 by Mirai Solutions GmbH [aut],
  Martin Studer [cre],
  The Apache Software Foundation [ctb, cph] (Apache POI, Apache Commons
    Codec),
  Stephen Colebourne [ctb, cph] (Joda-Time Java library),
  Graph Builder [ctb, cph] (Curvesapi Java library)
http://www.mirai-solutions.com ,
http://miraisolutions.wordpress.com
> fed = read.csv("fed.csv", head=TRUE)
#read the file "fed.cssv" and call it fed
> summary(fed)
#summarize the data fed and ensure integrity of data
> plot(jitter(fed$elec_dem), jitter(fed$FEDFUNDS), main="FEDFUNDS and elec_dem", xlab="elec_dem",
ylab="FEDFUNDS", xlim=c(0,15), ylim=c(0,20))
#to plot FEDFUNDS (y-axis) against elec_dem (x-axis)with jitter effect
```

Figure 4: Code for Scatterplot above (R Platform).

A majority of R platform figures have descriptions of the code in them.

Republicans

Here we attempt to do a scatter plot of Federal Funds rate (dependent Variable-Y-axis) to the quarters since the previous election for Republicans (independent Variable-X-axis).

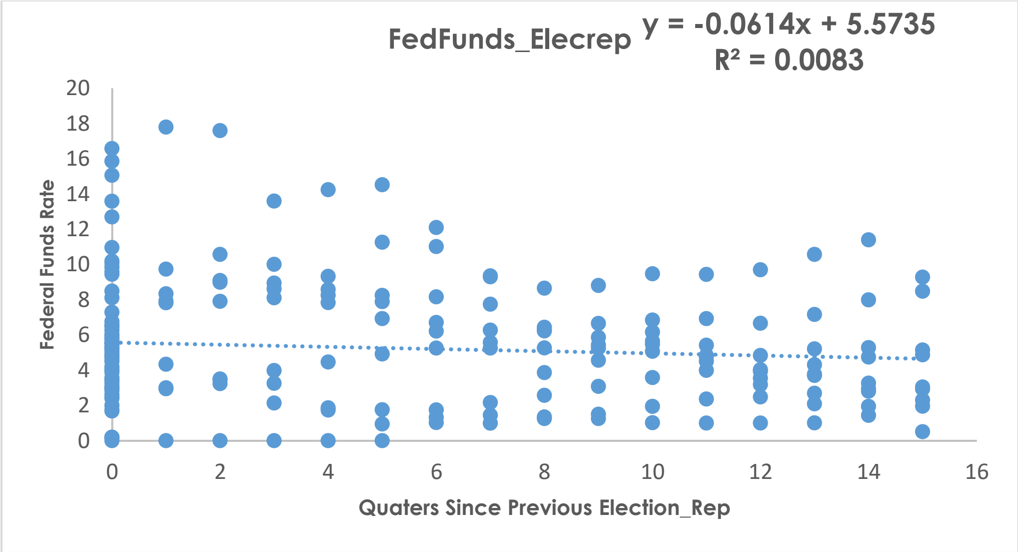


Figure 5: Fed Funds Rate (FFR) to Quarters since Previous Elections for Republicans.

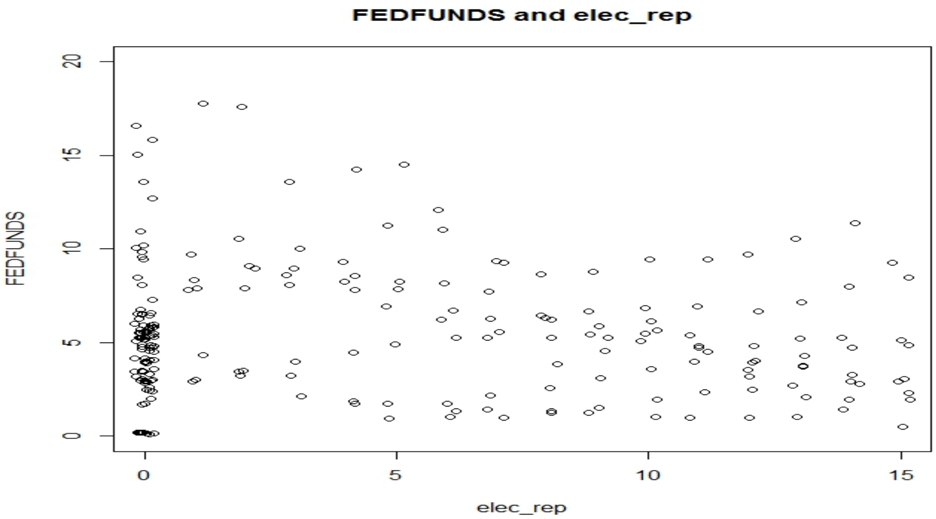


Figure 6: Fed Funds Rate (FFR) to Quarters since Previous Elections for Republicans (R Platform).

A majority of R platform figures have descriptions of the code in them.

Untitled - R Editor

```
> library(XLConnect)
#call for XLConnect package
Loading required package: XLConnectJars
XLConnect 0.2-12 by Mirai Solutions GmbH [aut],
  Martin Studer [cre],
  The Apache Software Foundation [ctb, cph] (Apache POI, Apache Commons
    Codec),
  Stephen Colebourne [ctb, cph] (Joda-Time Java library),
  Graph Builder [ctb, cph] (Curvesapi Java library)
http://www.mirai-solutions.com ,
http://miraisolutions.wordpress.com
> fed = read.csv("fed.csv", head=TRUE)
#read the file "fed.cssv" and call it fed
> summary(fed)
#summarize the data fed and ensure integrity of data
> plot(jitter(fed$elec_rep), jitter(fed$FEDFUNDS), main="FEDFUNDS and elec_rep", xlab="elec_rep",
ylab="FEDFUNDS", xlim=c(0,15), ylim=c(0,20))
#to plot FEDFUNDS (y-axis) against elec_rep (x-axis)with jitter effect
```

Figure 7: Code for Scatterplot above (R Platform).

<u>Comment on the differences in the Relationships</u>

The data for Democrats indicates a positive gradient for the regression line while the data for republicans indicates a negative gradient. In other words, Democrats have a positive relationship between the FFR and the Quarters since the previous election while Republicans have a negative relationship between same variables.

Qualitatively, the Democrats have the highest FFR (~18) in the first quarter containing the election (election=0) and the highest FFR (~16) in the quarter before the next election (election=15) when compared to the Republicans.

A majority of R platform figures have descriptions of the code in them.

Part B

Interaction variable = Party Affiliation (Party_Aff)

New Equation:

$$FFR_i = \beta_0 + \beta_1 ElecVar_i + \beta_2 PartyAff_i + \beta_3 ElecVarPartyAff_i + \epsilon_i \tag{1}$$

Where $FFR_i = \ Federal\ Funds\ Rate, \beta_0 = y - Intercept, \beta_1 = coefficient\ of\ \ Election\ Variable\ , \beta_2 =$

$coefficient\ of\ party\ affiliation\ accounted\ for, \beta_3 =$

$Coefficient\ of\ Election\ Variable\ multiplied\ by\ the\ party\ affiliation\ and\ \epsilon_i = error\ term$

Our interaction variable party affiliation can only take on 2 numbers which are 1 (Democratic president) and 0 (Republican president).

```
> OLS = lm(fed$FEDFUNDS ~ fed$elec_dem)
## run a bivariate OLS model of FEDFUNDS vs elec_dem
> summary(OLS)
## summarize the OLS model ran above
Call:
lm(formula = fed$FEDFUNDS ~ fed$elec_dem)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4690 -2.4074 -0.5075  1.6240 12.5040

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.27597    0.26591   19.84   <2e-16 ***
fed$elec_dem  0.05471    0.05021    1.09    0.277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.403 on 224 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.005274,	Adjusted R-squared:  0.0008336
F-statistic: 1.188 on 1 and 224 DF,  p-value: 0.277


> OLS1 = lm(fed$FEDFUNDS ~ fed$elec_rep)
## run a bivariate OLS model of FEDFUNDS vs elec_rep
> summary(OLS1)
## summarize the OLS1 model ran above
Call:
lm(formula = fed$FEDFUNDS ~ fed$elec_rep)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6539 -2.3910 -0.4639  1.7094 12.0795

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.77394    0.30466  18.952   <2e-16 ***
fed$elec_rep -0.07341    0.04346  -1.689   0.0926 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.391 on 224 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.01258,	Adjusted R-squared:  0.008168
F-statistic: 2.853 on 1 and 224 DF,  p-value: 0.0926
```

Figure 8: OLS performed in R platform.

A majority of R platform figures have descriptions of the code in them.

The model above is given by the equation:

$$FFR_i = \beta_0 + \beta_1 elec\_dem_i + \epsilon_i \qquad (2)$$

$\beta_1 = coefficent\ for\ elec\_dem,\quad \beta_0 = y - intercept,\quad \epsilon_i = error\ term,\quad elec\_dem_i = democrat\ election\ variable,$

$FFR_i = Federal\ funds\ rate$

Table 1: Summary of Bivariate Regression Results for Monetary Policy Data (R platform)

	Democrat	**Republican**
Elec_dem	0.055	
	(0.05)	
	[t = 1.09]	
Elec_rep		-0.073
		(0.04)
		[t=-1.69]
Intercept	5.28*	5.77*
	(0.27)	(0.30)
	[t = 19.84]	[t=18.95]
N	225	225
$\mathbf{R^2}$	0.0053	0.013

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

A majority of R platform figures have descriptions of the code in them.

$$FFR_i = 5.28 + 0.055elec\_dem_i + \epsilon_i \tag{3}$$

(i) A one-unit increase in the elec_rep will yield a 0.073 change (0.073 decrease) in the federal funds rate (FFR) (based on data from Table 1).

(ii) A one-unit increase in the elec_dem will yield 0.055 change (0.055 increase) in the federal funds rate (FFR) (based on data from Table 1).

Methodology: We used a simple gradient approach to find the effect of an increase in the x-variable on the y-variable. The coefficient of any x-variable is given by a change in the y-variable divided by a unit increase in the x-variable.

$$coefficeint\ of\ x - variable = \frac{dy}{dx} \tag{4}$$

<u>Part C</u>

The effect of the election is statistically insignificant under Republicans. The output from R above shows that under Republicans the p-value is 0.093 meaning that we fail to reject the Null hypothesis. A p-value greater than 0.05 means that we do not have enough evidence to reject the Null hypothesis and so our coefficient in this case is insignificant. In other words, we have a weak negative relationship (or correlation) between elec_rep and FEDFUNDS.

The effect of the election is statistically insignificant under Democrats. The output from R above shows that under Democrats the p-value is 0.28 meaning that we fail to reject the Null hypothesis. A p-value greater than 0.05 means that we do not have enough evidence to reject the Null hypothesis and so our coefficient in this case is insignificant. In other words, we have a weak positive relationship (or correlation) between elec_dem and FEDFUNDS.

A majority of R platform figures have descriptions of the code in them.

Methodology for Results Above

We compare the p-values for both coefficients above in order to determine which variable is statistically significant. The R platform also indicates significance by using asterisks as seen for the intercept above, which we can use in determining significance. A p-value below 0.05 usually shows a significant coefficient and a p-value above 0.05 usually denotes an insignificant coefficient.

Additional Test

We can also take a look at the t-stat absolute value alongside the p-value in order to determine significance. A high t-stat absolute value usually indicates statistical significance and a very low value usually shows statistical insignificance.

Part D

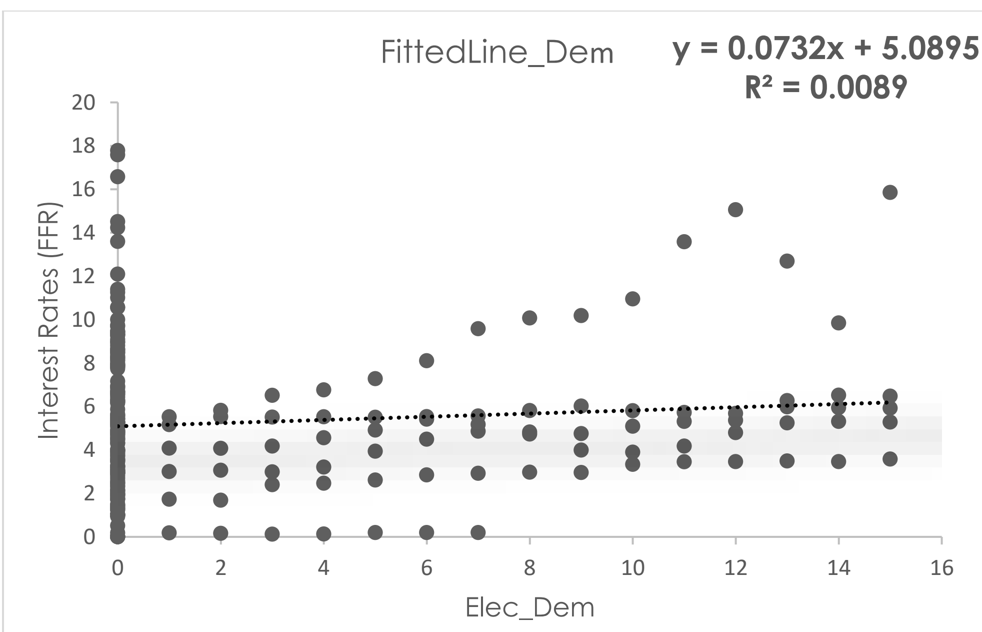


Figure 9: Fitted Line_Democrats (Excel).

A majority of R platform figures have descriptions of the code in them.

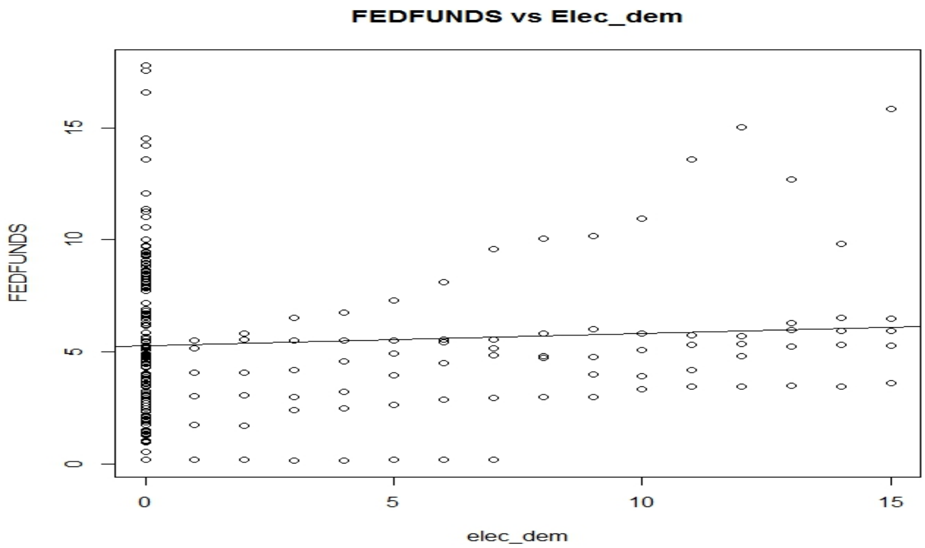


Figure 10: Fitted Line_Democrats (R platform).

```
> OLS = lm(fed$FEDFUNDS ~ fed$elec_dem)
# Create a linear model of FEDFUNDS versus elec_dem
> summary(OLS)
#summarize OLS results
Call:
lm(formula = fed$FEDFUNDS ~ fed$elec_dem)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4690 -2.4074 -0.5075  1.6240 12.5040

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.27597    0.26591   19.84   <2e-16 ***
fed$elec_dem  0.05471    0.05021    1.09    0.277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.403 on 224 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.005274,	Adjusted R-squared:  0.0008336
F-statistic: 1.188 on 1 and 224 DF,  p-value: 0.277

> abline(OLS)
#Create a fitted line for elec_dem OLS results
```

Figure 11: Fitted Line elec_dem Performed in R platform.

A majority of R platform figures have descriptions of the code in them.

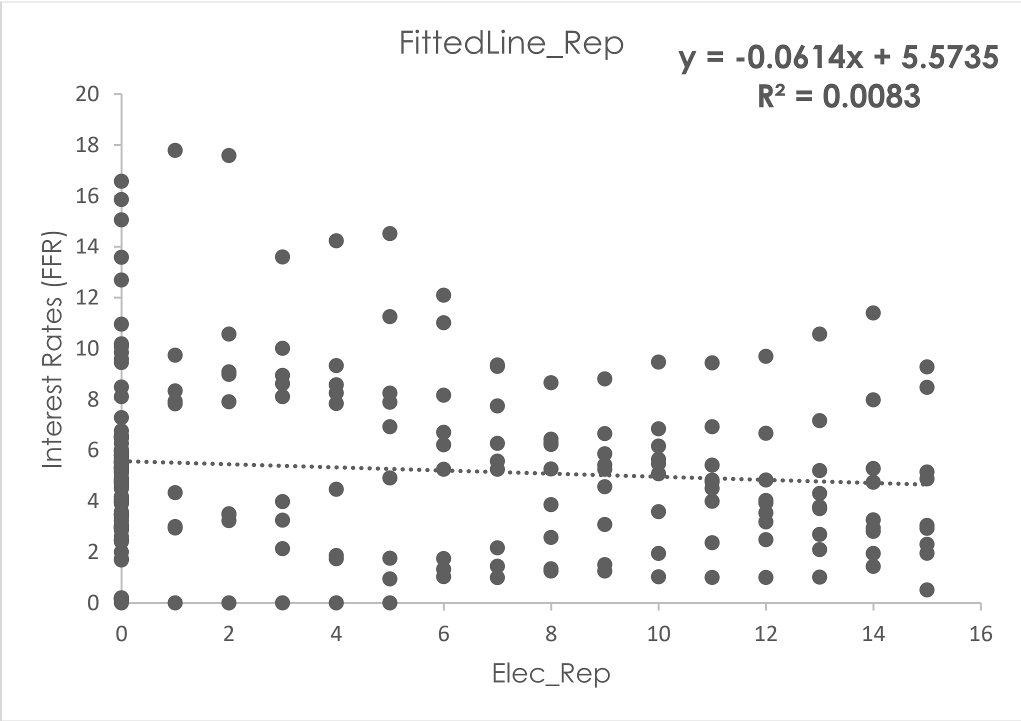


Figure 12: Fitted Line_Republicans (Excel).

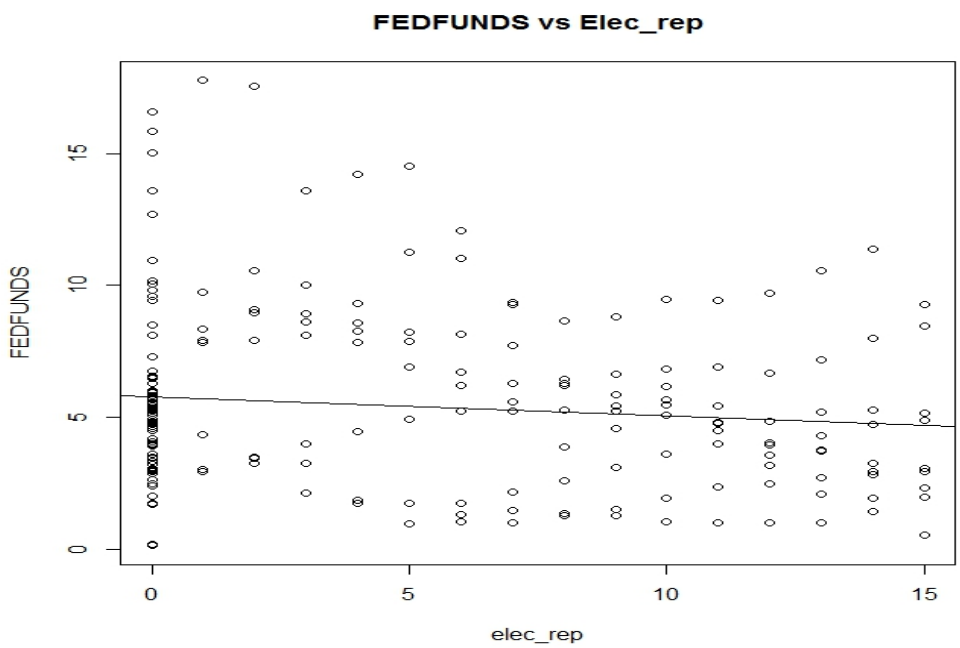


Figure 13: Fitted Line_Republicans (R platform).

A majority of R platform figures have descriptions of the code in them.

```
> plot(fed$elec_rep, fed$FEDFUNDS, main="FEDFUNDS and elec_rep", xlab="elec_rep", ylab="FEDFUNDS")
#Scatterplot of FEDFUNDS vs elec_rep
> OLS = lm(fed$FEDFUNDS ~ fed$elec_rep)
# Create a linear model of FEDFUNDS versus elec_dem
> summary(OLS)
#summarize OLS results
Call:
lm(formula = fed$FEDFUNDS ~ fed$elec_rep)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.6539 -2.3910 -0.4639  1.7094 12.0795 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.77394    0.30466  18.952   <2e-16 ***
fed$elec_rep -0.07341    0.04346  -1.689   0.0926 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.391 on 224 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.01258,	Adjusted R-squared:  0.008168 
F-statistic: 2.853 on 1 and 224 DF,  p-value: 0.0926

> abline(OLS)
#Create a fitted line for elec_rep OLS results
```

Figure 14: Fitted Line elec_rep Performed in R platform.

The fitted line for Democrats (Fig. 9 & 10) shows a relationship with a positive gradient meaning that every one-unit increase in the elec_dem variable yields a 0.055 increase in the FFR according to data from R-platform.

The fitted line for Republicans (Fig. 12 & 13) shows a relationship with a negative gradient meaning that every one-unit increase in the elec_rep variable yields a 0.073 decrease in the FFR according to data from R-platform.

Part E

$$FFR_i = \beta_0 + \beta_1 inflation_i + \beta_2 lag\_FEDFUNDS_i + \beta_3 elec\_dem_i + \epsilon_i \quad (5)$$

Where $\beta_0 = y - Intercept,\ \beta_1 = coefficient\ of\ \ inflation_i\ , \beta_2 = coefficient\ of\ lag\_FEDFUNDS_i, \beta_3 = Coefficient\ of\ elec\_dem_i,$

$and\ \epsilon_i = error\ term$

A majority of R platform figures have descriptions of the code in them.

```
> OLSMV = lm(FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem, data=fed)
## run a multivariate OLSMV model of FEDFUNDS (dependent variable) vs inflation
,lag_FEDFUNDS and elec_dem
> summary(OLSMV)
## we summarize the OLSMV model above
Call:
lm(formula = FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem,
    data = fed)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1096 -0.3424  0.0359  0.3986  5.2496

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.05274    0.11469   0.460    0.646
inflation     0.11690    0.02468   4.736  3.9e-06 ***
lag_FEDFUNDS  0.89385    0.02249  39.736  < 2e-16 ***
elec_dem      0.02940    0.01285   2.289    0.023 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8651 on 221 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.9361,    Adjusted R-squared:  0.9352
F-statistic:  1079 on 3 and 221 DF,  p-value: < 2.2e-16
```

Figure 15: Multivariate OLS accounting for Inflation and lag_FEDFUNDS.

We also run a robust regression to look for heteroscedastic standard-errors (HC1):

```
> library(AER)
##we load the AER package
Loading required package: car
Loading required package: lmtest
Loading required package: zoo

Attaching package: 'zoo'

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

    as.Date, as.Date.numeric

Loading required package: sandwich
Loading required package: survival
> coeftest(OLSMV, vcov=vcovHC (OLSMV, type="HC1"))
##we run a robust regression to test for heteroscedastic errors

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.052741   0.148943  0.3541 0.7235981
inflation    0.116897   0.033963  3.4419 0.0006908 ***
lag_FEDFUNDS 0.893852   0.033686 26.5350 < 2.2e-16 ***
elec_dem     0.029399   0.018044  1.6293 0.1046672
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16: Robust regression testing for heteroscedasticity.

A majority of R platform figures have descriptions of the code in them.

Table 2: Multivariate Regression Model in R platform (Figure 16).

	Variables (Multivariate Regression Results for Monetary Policy Data)
Inflation	0.12*
	(0.034)
	[t=3.44]
Lag_FEDFUNDS	0.89*
	(0.034)
	[t=26.54]
Elec_dem	0.029
	(0.018)
	[t=1.63]
Intercept	0.053
	(0.15)
	[t=0.35]
N	222
$\mathbf{R^2}$	0.94

Standard errors in parentheses.

(*) indicates significance at $p<0.05$, two tailed.

A majority of R platform figures have descriptions of the code in them.

Answers to these questions are derived from the R platform outputs above:

(i) The coefficient for the effect of election for Republicans (-0.027*) is statistically significant with a p-value of 0.0042 and a t-value absolute of 2.89. A p-value below 0.05 gives us enough evidence to reject the Null Hypothesis, which makes our coefficient statistically significant. There is a strong negative relationship between elec_rep and FFR when inflation and lag_FEDFUNDS are brought into the picture.

(ii) The coefficient for the effect of election for Democrats (0.029) is statistically insignificant with a p-value of 0.1047 and a t-value absolute of 1.63. A p-value above 0.05 does not give us enough evidence to reject the Null Hypothesis, which makes our coefficient statistically insignificant. There is a weak positive relationship between elec_dem and FFR.

(iii) The coefficient for lag_FEDFUNDS (0.89*) is statistically significant with a p-value of 2.2e-16 and a significantly high t-value of 26.54. The p-value indicates that we have enough evidence to reject the null hypothesis and so statistically significant. There is a strong positive relationship between the lag_FEDFUNDS and FFR.

(iv) The coefficient for Inflation (0.12*) is statistically significant with a p-value of 0.0007 and a t-value of 3.44. From the p-value, we have enough evidence to reject the null hypothesis and so the coefficient is statistically significant. There is a strong positive relationship between the inflation and FFR.

A majority of R platform figures have descriptions of the code in them.

Chapter 8 Question Six

TABLE 8.13 Variables in the Cell Phones and Traffic Deaths Data

Variable name	Description
year	Year
State	State name
state_numeric	State name (numeric representation of state)
population	Population within a state
DeathsPerBillionMiles	Deaths per billion miles driven in state
cell_ban	Coded 1 if handheld cell phone while driving ban is in effect; 0 otherwise
text_ban	Coded 1 if texting while driving ban is in effect; 0 otherwise
cell_per10thous_pop	Number of cell phone subscriptions per 10,000 people in state
urban_percent	Percent of state residents living in urban areas

Figure 17: Codebook for cell phone panel data.

Part A

$$DPB_{i,t} = \beta_0 + \beta_1 cell\_ban_{i,t} + \beta_2 text\_ban_{i,t} + \epsilon_{i,t} \qquad (1)$$

$DPB_{i,t} = Deaths\ per\ Billion\ Miles,\ \beta_0 = y - intercept,\ \beta_1 = coefficient\ of\ cell\_ban_{i,t}\ ,\ \beta_2 = coefficient\ of\ text\_ban_{i,t}\ ,\ \epsilon_{i,t} = error\ term$

```
> POLS = lm(DeathsPerBillionMiles ~ cell_ban + text_ban, data=cell)
#run a pooled OLS model of DeathsPerBillion vs cell_ban & text_ban
> summary(POLS)
#summarize POLS
Call:
lm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban, data = cell)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.6047 -2.0160 -0.1978  1.7024 11.1173 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  13.3808     0.2094  63.899  < 2e-16 ***
cell_ban     -2.8386     0.5364  -5.292 2.33e-07 ***
text_ban     -2.0616     0.4067  -5.069 6.96e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.979 on 303 degrees of freedom
Multiple R-squared:  0.2467,	Adjusted R-squared:  0.2418 
F-statistic: 49.62 on 2 and 303 DF,  p-value: < 2.2e-16
```

Figure 18: Pooled OLS model R platform.

A majority of R platform figures have descriptions of the code in them.

```
> library(AER)
#we call for the package AER
Loading required package: car
Loading required package: lmtest
Loading required package: zoo

Attaching package: 'zoo'

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

    as.Date, as.Date.numeric

Loading required package: sandwich
Loading required package: survival
> coeftest(POLS, vcov=vcovHC(POLS, type="HC1"))
#We run a robust regression to find heteroscedastic errors
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 13.38078    0.22380 59.7897 < 2.2e-16 ***
cell_ban    -2.83861    0.37188 -7.6331 3.007e-13 ***
text_ban    -2.06157    0.40071 -5.1448 4.820e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 19: Robust Pooled OLS model (HC1 type error) R platform.

The figure above shows a pooled OLS model as demanded and a brief description of this data is presented below:

$$DPB_{i,t} = 13.38 - 2.84 cell\_ban_{i,t} - 2.06 text\_ban_{i,t} + \epsilon_{i,t} \qquad (2)$$

Remember that in this pooled OLS model we don't account for states separately but we treat them as a pool.

Here we see a strong negative relationship between the cell_ban and DPB meaning that a one-unit increase in cell_ban will lead to a 2.84 reduction in the deaths per billion miles. The coefficient for cell_ban is statistically significant according to our output from the R platform above with a p-value of 2.33e-07. The p-value above (Figure 18) means that we have enough evidence to reject the Null Hypothesis, which makes it statistically significant.

Here we see a strong negative relationship between the text_ban and DPB meaning that a one-unit increase in text_ban will lead to a 2.06 reduction in the deaths per billion miles. The coefficient for text_ban is statistically significant according to our output from the R platform above with a p-value of 6.96e-07. The p-value above means that we have enough evidence to reject the Null Hypothesis, which makes it statistically significant.

A majority of R platform figures have descriptions of the code in them.

Part B

The number of road banners or signs installed per state annually, that say “No Driving and texting or calling”. We can call it “Number of banners”. This variable will describe the number of banners or signs installed in each state annually, that advocate no texting or calling while driving.

Part C

We run a one-way fixed model where we account for the “state numeric” and “time”.

```
> pan1 = plm(DeathsPerBillionMiles ~ cell_ban + text_ban, data=cell, index=c("state_numeric","year"), model="within")
#We run a one-way fixed model with the state_numeric and year
> summary(pan1)
#we summarize the panel model
Oneway (individual) effect Within Model

Call:
plm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban, data = cell,
    model = "within", index = c("state_numeric", "year"))

Balanced Panel: n=51, T=6, N=306

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -3.820  -0.822  -0.113   0.565   4.860

Coefficients :
         Estimate Std. Error t-value  Pr(>|t|)
cell_ban -0.81863    0.51957 -1.5756    0.1164
text_ban -1.12564    0.21939 -5.1308 5.749e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    524.94
Residual Sum of Squares: 453.03
R-Squared:      0.13698
Adj. R-Squared: 0.11326
F-statistic: 20.0789 on 2 and 253 DF, p-value: 8.0611e-09
```

Figure 20: One-way fixed model Panel regression results.

A majority of R platform figures have descriptions of the code in them.

```
> pan2 = plm(DeathsPerBillionMiles ~ cell_ban + text_ban, data=cell, index=c("state_numeric","year"), model="random")
#we run a panel random effects model
> summary(pan2)
#we summarize the random effects model
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)
Call:
plm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban, data = cell,
    model = "random", index = c("state_numeric", "year"))

Balanced Panel: n=51, T=6, N=306
Effects:
                var std.dev share
idiosyncratic 1.791   1.338 0.211
individual    6.701   2.589 0.789
theta:  0.7935
Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -4.310  -0.890  -0.233   0.693   6.260
Coefficients :
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 12.84381    0.38509 33.3531 < 2.2e-16 ***
cell_ban    -1.27103    0.48758 -2.6068  0.009591 **
text_ban    -1.15759    0.22043 -5.2516 2.845e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    654.79
Residual Sum of Squares: 559.44
R-Squared:      0.14562
Adj. R-Squared: 0.14419
F-statistic: 25.8213 on 2 and 303 DF, p-value: 4.4182e-11
```

Figure 21: Random effects model Panel regression results.

```
> phtest(pan1,pan2)
#We run this test to determine which model 9fixed or random) is bias
        Hausman Test

data:  DeathsPerBillionMiles ~ cell_ban + text_ban
chisq = 0.18235, df = 2, p-value = 0.9129
alternative hypothesis: one model is inconsistent
```

Figure 22: Hausman test.

The coefficient for cell_ban in the one-way fixed effects model changes and it becomes insignificant (Figure 20). My expectation was that the coefficient for cell_ban will remain significant as it was in Part A but it is not the case in the one-way fixed effects model. In the one-way fixed-effects model, the fixed effects variable ($\alpha_{i,t}$= year and state numeric) makes the cell_ban coefficient insignificant. In this case we can postulate that the one-way fixed effects model produces unbiased results, which show that our fixed effects variable ($\alpha_{i,t}$) is correlated with the independent variable ($DPB_{i,t}$). Intuitively, I expected a negative coefficient for text_ban based on

A majority of R platform figures have descriptions of the code in them.

real-life thinking but the statistics proved me wrong. It is my thinking that an increase in text_ban should lead to a decrease in deaths per billion miles but results show other factors in the mix.

Part D

A possible year-fixed effect will include the number of road construction projects executed per year. The number of roads that have one or more lanes "unused" as a result of road construction projects taking place can have an effect on the number of deaths per billion miles.

Part E

```
> pandm = plm(DeathsPerBillionMiles ~ cell_ban + text_ban, data=cell, index=c("state_numeric","year"), model="within", effect="twoways")
> summary(pandm)
Twoways effects Within Model

Call:
plm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban, data = cell,
    effect = "twoways", model = "within", index = c("state_numeric",
        "year"))

Balanced Panel: n=51, T=6, N=306

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-3.14000 -0.52500  0.00567  0.50100  3.07000

Coefficients :
         Estimate Std. Error t-value Pr(>|t|)
cell_ban -0.60087    0.40193 -1.4950   0.1362
text_ban  0.29135    0.22014  1.3235   0.1869

Total Sum of Squares:    267.89
Residual Sum of Squares: 264.49
R-Squared:      0.0127
Adj. R-Squared: 0.010293
F-statistic: 1.59504 on 2 and 248 DF, p-value: 0.20498
> pandm = plm(DeathsPerBillionMiles ~ cell_ban + text_ban, data=cell, index=c("state_numeric","year"), model="within", effect="twoways")
```

Figure 23: Two-way fixed model Panel regression results.

The results derived were as expected when we took into account a fixed effect variable ($\alpha_{i,t}$) that could cause endogeneity, as well as, the time (τ_t) by running a two-way fixed effects model. Although the state-numeric and year might have an effect on the independent variable as seen in the one-way fixed effects model, some years might also be characterized by unexpected events that trigger more deaths per billion miles like massive road construction projects. In this case we give "year" a specific term in our equation in order to account for specific-year events. I wasn't surprise that adding time (years) to the equation will make cell_ban and text_ban insignificant especially after the results from Part C (one-way fixed effects model). Shockingly, I still see that

A majority of R platform figures have descriptions of the code in them.

text_ban has a positive coefficient, which to me is the most likely reason why we have more deaths on the road when compared to cell_ban.

Part F

```
> panf = plm(DeathsPerBillionMiles ~ cell_ban + text_ban + cell_per10thous_pop + urban_percent, data=cell,
index=c("state_numeric","year"), model="within", effect="twoways")
##Estimation of a two-way fixed effects model while controlling for cell_per10thous_pop and urban_percent
> summary(panf)
##summarize estimated model
Twoways effects Within Model

Call:
plm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban + cell_per10thous_pop +
    urban_percent, data = cell, effect = "twoways", model = "within",
    index = c("state_numeric", "year"))

Balanced Panel: n=51, T=6, N=306

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-2.7400 -0.5130 -0.0135  0.4770  3.1400

Coefficients :
                       Estimate  Std. Error t-value Pr(>|t|)
cell_ban            -0.67979576  0.40294913 -1.6871  0.09286 .
text_ban             0.25592620  0.22219231  1.1518  0.25051
cell_per10thous_pop -0.00034037  0.00017294 -1.9682  0.05017 .
urban_percent        0.01313477  0.01119861  1.1729  0.24197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    267.89
Residual Sum of Squares: 259.07
R-Squared:      0.032939
Adj. R-Squared: 0.02648
F-statistic: 2.09472 on 4 and 246 DF, p-value: 0.082085
```

Figure 24: Two-way fixed effects model while controlling for cell_per10thous_pop and urban_percent.

A one-unit increase in cell_ban will lead to a 0.68 decrease in the number of deaths per billion miles. Also we see that the cell_ban coefficient is statistically insignificant although it comes close to being significant.

A one-unit increase in text_ban will lead to a 0.26 increase in the number of deaths per billion miles. Also we see that the text_ban coefficient is statistically insignificant although it comes close to being significant.

A majority of R platform figures have descriptions of the code in them.

Part G

```
> pang = plm(DeathsPerBillionMiles ~ cell_ban + text_ban + cell_per10thous_pop + urban_percent
 + factor(state_numeric), data=cell, index=c("year"), model="within")
##I ran an LSDV two-way model with dummy variable "state_numeric" as a function of time
> summary(pang)
##summary of LSDV model
Oneway (individual) effect Within Model

Call:
plm(formula = DeathsPerBillionMiles ~ cell_ban + text_ban + cell_per10thous_pop +
    urban_percent + factor(state_numeric), data = cell, model = "within",
    index = c("year"))

Balanced Panel: n=6, T=51, N=306

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-2.7400 -0.5130 -0.0135  0.4770  3.1400

Coefficients :
                           Estimate  Std. Error  t-value  Pr(>|t|)
cell_ban                -0.67979576  0.40294913  -1.6871 0.0928611 .
text_ban                 0.25592620  0.22219231   1.1518 0.2505120
cell_per10thous_pop     -0.00034037  0.00017294  -1.9682 0.0501730 .
urban_percent            0.01313477  0.01119861   1.1729 0.2419736
factor(state_numeric)2  -1.73680147  0.63955303  -2.7156 0.0070836 **
factor(state_numeric)3  -0.94793575  0.62026310  -1.5283 0.1277278
factor(state_numeric)4   2.93185626  0.62991766   4.6543 5.316e-06 ***
factor(state_numeric)5  -5.01920078  0.67455675  -7.4407 1.672e-12 ***
factor(state_numeric)6  -4.58795790  0.60304023  -7.6080 5.905e-13 ***
factor(state_numeric)7  -6.78495439  0.85935136  -7.8954 9.577e-14 ***
factor(state_numeric)8  -2.25885698  0.61496981  -3.6731 0.0002940 ***
factor(state_numeric)9  -3.50869075  2.16932391  -1.6174 0.1070709
factor(state_numeric)10 -1.43739566  0.63127865  -2.2770 0.0236480 *
factor(state_numeric)11 -2.62906899  0.60481086  -4.3469 2.022e-05 ***
factor(state_numeric)12 -3.39722327  0.60494729  -5.6157 5.275e-08 ***
factor(state_numeric)13 -1.40973860  0.64703688  -2.1788 0.0302990 *
factor(state_numeric)14 -5.62664997  0.63573251  -8.8507 < 2.2e-16 ***
factor(state_numeric)15 -4.65502592  0.60770915  -7.6600 4.263e-13 ***
factor(state_numeric)16 -2.40275243  0.64674599  -3.7151 0.0002514 ***
factor(state_numeric)17 -1.25780163  0.63371331  -1.9848 0.0482763 *
```

A majority of R platform figures have descriptions of the code in them.

```
factor(state_numeric)17 -1.25780163  0.63371331  -1.9848 0.0482763 *
factor(state_numeric)18  1.60699144  0.62866782   2.5562 0.0111846 *
factor(state_numeric)19  2.86576701  0.63189049   4.5352 8.995e-06 ***
factor(state_numeric)20 -3.68349829  0.69266415  -5.3179 2.358e-07 ***
factor(state_numeric)21 -5.10187656  0.67508341  -7.5574 8.103e-13 ***
factor(state_numeric)22 -9.00046717  0.84194013 -10.6902 < 2.2e-16 ***
factor(state_numeric)23 -5.42441252  0.60868074  -8.9118 < 2.2e-16 ***
factor(state_numeric)24 -7.44298116  0.62532816 -11.9025 < 2.2e-16 ***
factor(state_numeric)25  2.29537681  0.61661017   3.7226 0.0002445 ***
factor(state_numeric)26 -2.14637960  0.59308396  -3.6190 0.0003589 ***
factor(state_numeric)27  4.97528604  0.67915584   7.3257 3.392e-12 ***
factor(state_numeric)28 -3.82081105  0.63557504  -6.0116 6.593e-09 ***
factor(state_numeric)29 -2.11161506  0.62648354  -3.3706 0.0008708 ***
factor(state_numeric)30 -5.16614896  0.72218871  -7.1535 9.661e-12 ***
factor(state_numeric)31 -6.47244557  0.90711801  -7.1352 1.079e-11 ***
factor(state_numeric)32 -0.74138583  0.63483860  -1.1678 0.2440045
factor(state_numeric)33 -4.89437893  0.73017910  -6.7030 1.385e-10 ***
factor(state_numeric)34 -1.47250039  0.61485329  -2.3949 0.0173747 *
factor(state_numeric)35  0.67467787  0.66850640   1.0092 0.3138552
factor(state_numeric)36 -4.71368173  0.59262077  -7.9540 6.581e-14 ***
factor(state_numeric)37  0.28126771  0.59988197   0.4689 0.6395766
factor(state_numeric)38 -3.52820539  0.64803159  -5.4445 1.257e-07 ***
factor(state_numeric)39 -1.83499660  0.60256202  -3.0453 0.0025771 **
factor(state_numeric)40 -6.98497790  0.70145015  -9.9579 < 2.2e-16 ***
factor(state_numeric)41  3.42284756  0.65721498   5.2081 4.032e-07 ***
factor(state_numeric)42 -0.12944013  0.66014898  -0.1961 0.8447117
factor(state_numeric)43 -0.09984371  0.60985968  -0.1637 0.8700892
factor(state_numeric)44 -1.32924751  0.59894188  -2.2193 0.0273772 *
factor(state_numeric)45 -5.79734113  0.63331972  -9.1539 < 2.2e-16 ***
factor(state_numeric)46 -5.73348303  0.69048569  -8.3036 6.781e-15 ***
factor(state_numeric)47 -4.85839898  0.60418029  -8.0413 3.749e-14 ***
factor(state_numeric)48 -5.90198341  0.65745799  -8.9770 < 2.2e-16 ***
factor(state_numeric)49  3.16607149  0.62937843   5.0305 9.443e-07 ***
factor(state_numeric)50 -4.65238738  0.61750668  -7.5341 9.366e-13 ***
factor(state_numeric)51  0.68630485  0.63973749   1.0728 0.2844160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 25: Least squares dummy variable two-way fixed model.

The coefficients are the same as in Part F above. The same value of -0.68. This should be expected because fixed-effects like the state_numeric will likely have a very little effect on the coefficients for cell_ban and text_ban. The state_numeric isn't expected to change the results especially when inserted as a factor.

A majority of R platform figures have descriptions of the code in them.

Table 3: Summary of Regressions for Cell Phone Data.

Variable	Pooled OLS	One-Way Effects Fixed Model	Random-Effects Model	Two-Way Fixed De-meaned	LSDV
Deaths Per Mile	**13.38***		**12.84***		
	(0.21)		**(0.39)**		
	[t = 63.89]		**[t = 33.35]**		
Cell_ban	**-2.84***	**-0.82**	**-1.27***	**-0.68**	**-0.68**
	(0.54)	**(0.52)**	**(0.49)**	**(0.40)**	**(0.40)**
	[t=-5.29]	**[t = -1.58]**	**[t = -2.61]**	**[t = -1.69]**	**[t = -1.69]**
Text_ban	**-2.06***	**-1.13***	**-1.158***	**0.26**	**0.26**
	(0.41)	**(0.22)**	**(0.22)**	**(0.22)**	**(0.222)**
	[t =-5.07]	**[t = -5.13]**	**[t = -5.25]**	**[t = 1.15]**	**[t = 1.15]**
Cell_per10thous_pop				**-0.0003**	**-0.0003**
				(0.00017)	**(0.00017)**
				[t = -1.97]	**[t = -1.97]**
Urban_percent				**0.013**	**0.013**
				(0.0111)	**(0.0111)**
				[t = 1.173]	**[t = 1.173]**
State_numeric					**50 Observations**
Intercept					
N			**306**		
R^2			**0.92**		

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

A majority of R platform figures have descriptions of the code in them.

Part H

Large Positive Fixed Effects

- Arkansas
- Kentucky
- Louisiana
- Mississippi
- Montana
- South Carolina
- West Virginia

Large positive fixed effects are determined by looking at states that have a positive coefficient and a p-value that denotes significance. The interpretation here means that a one-unit change in the state_numeric will lead to an increase in the deaths per billion miles.

Large Negative Fixed Effects

- Alaska
- California
- Colorado
- Connecticut
- Delaware
- Florida
- Georgia
- Hawaii
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Maine
- Maryland
- Massachusetts
- Michigan
- Minnesota
- Missouri
- Nebraska
- Nevada
- New Hampshire
- New jersey
- Ohio
- Oregon
- Pennsylvania
- Rhode Island
- Texas
- Utah
- Vermont
- Virginia
- Washington
- Wisconsin

A majority of R platform figures have descriptions of the code in them.

Large negative fixed effects are determined by looking at states that have a negative coefficient and a p-value that denotes significance. The interpretation here means that a one-unit change in the state_numeric will lead to a decrease in the deaths per billion miles.

The excluded category are states that have a negative or positive coefficient for the state_numeric but they do not have any significance (high p-value).

The positive and negative effect states vary by a number of speculative state specific reasons:

The difficulty of acquiring a driver’s license.

The average number of lanes per road in each state.

The number of traffic police officers on the road.

The number of banners or signs that advocate against cell phone usage while driving.

The number of road construction projects per state.

A majority of R platform figures have descriptions of the code in them.

Extension Exercices

TABLE 8.13 Variables in the Cell Phones and Traffic Deaths Data

Variable name	Description
year	Year
State	State name
state_numeric	State name (numeric representation of state)
population	Population within a state
DeathsPerBillionMiles	Deaths per billion miles driven in state
cell_ban	Coded 1 if handheld cell phone while driving ban is in effect; 0 otherwise
text_ban	Coded 1 if texting while driving ban is in effect; 0 otherwise
cell_per10thous_pop	Number of cell phone subscriptions per 10,000 people in state
urban_percent	Percent of state residents living in urban areas

Figure 26: Codebook for cell phone panel data.

DM4: The aggregate M4 is a broad aggregate including negotiable money-market securities, such as commercial paper, negotiable CDs, and T-bills. It measures the supply of various forms or type of money into the economy.[1]

Lag_DM4: Is a measure of M4 at time (t-1).

Extension 1

Data Fine-Tune

The data for DM4 was modified for the time period 1968-2010. The original data set was Jan-67 to Aug-16 but figures were provided from Jan-68 to Aug-16. In order to match the data in the "Federal kkedits file", I took data from Jan-68 to Oct-2010.

Only data for the months January, April, July and October were imported into the Federal data file.

The data was converted from percentage to numbers in order to maintain the integrity of the csv file and to allow R code to function properly.

[1] http://www.centerforfinancialstability.org/amfm_data.php?startc=2004&startt=1967#summary

A majority of R platform figures have descriptions of the code in them.

Model from Part E Modified

$$FFR_i = \beta_0 + \beta_1 Inflation_i + \beta_2 lag\_FEDFUNDS_i + \beta_3 elec\_dem_i + \beta_4 DM4_i + \epsilon_i \quad (1)$$

Where $FFR_i = Federal\ Funds\ Rate,\ \beta_0 = y - Intercept, \beta_1 = coefficient\ of\ \ Inflation_i\ , \beta_2 =$ $coefficient\ of\ lag\_FEDFUNDS_i, \beta_3 = Coefficient\ of\ elec\_dem_i, \beta_4 = Coefficient\ of\ DM4_i, \epsilon_i = error\ term$

```
> OLSE = lm(FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem + DM4, data=fed)
##we run a multivariate OLS model controlling for 4 variables
> summary(OLSE)
##we summarize the OLSE model
Call:
lm(formula = FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem +
    DM4, data = fed)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0915 -0.4099  0.0250  0.4495  5.1086

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.31698    0.22278  -1.423   0.1566
inflation     0.12367    0.02851   4.338 2.48e-05 ***
lag_FEDFUNDS  0.89837    0.02603  34.513  < 2e-16 ***
elec_dem      0.04278    0.01675   2.554   0.0115 *
DM4           0.04657    0.02403   1.938   0.0543 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9515 on 167 degrees of freedom
  (61 observations deleted due to missingness)
Multiple R-squared:  0.9293,    Adjusted R-squared:  0.9276
F-statistic: 548.6 on 4 and 167 DF,  p-value: < 2.2e-16
```

```
> coeftest(OLSE, vcov=vcovHC (OLSE, type="HC1"))
##we run a robust regression to find heteroscedastic errors
t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)  -0.316985   0.159422 -1.9883 0.0484085 *
inflation     0.123670   0.035192  3.5142 0.0005678 ***
lag_FEDFUNDS  0.898366   0.034184 26.2803 < 2.2e-16 ***
elec_dem      0.042784   0.023121  1.8504 0.0660171 .
DM4           0.046567   0.018932  2.4597 0.0149242 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 27: OLS model controlling for DM4.

A majority of R platform figures have descriptions of the code in them.

```
> library(AER)
##we load the AER package
Loading required package: car
Loading required package: lmtest
Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

    as.Date, as.Date.numeric

Loading required package: sandwich
Loading required package: survival
> coeftest(OLSMV, vcov=vcovHC (OLSMV, type="HC1"))
##we run a robust regression to test for heteroscedastic errors

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.052741   0.148943  0.3541 0.7235981
inflation    0.116897   0.033963  3.4419 0.0006908 ***
lag_FEDFUNDS 0.893852   0.033686 26.5350 < 2.2e-16 ***
elec_dem     0.029399   0.018044  1.6293 0.1046672
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 28: OLS model controlling from Part E.

A majority of R platform figures have descriptions of the code in them.

Table 4: Summary of OLS with lag_DM4 accounted for in the model.

	OLS Part E (Fig. 28)	**OLS from (Figure 27)**
Inflation	0.12*	0.12*
	(0.03)	(0.03)
	[t=3.44]	[t = 3.51]
Lag_FEDFUNDS	0.89*	0.89*
	(0.03)	(0.03)
	[t=26.54]	[t = 26.28]
Elec_dem	0.029	0.043
	(0.02)	(0.02)
	[t=1.63]	[t = 1.85]
DM4		0.05*
		(0.018)
		[t = 2.46]
Intercept	0.053	-0.32*
	(0.15)	(0.16)
	[t=0.35]	[t = -1.99]
N	222	168
$\mathbf{R^2}$	0.94	0.93

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

A majority of R platform figures have descriptions of the code in them.

Comparison

There is a change in the magnitude (coefficients) of elec_dem in the model with DM4 when compared to the model without DM4. A one-unit change in elec_dem yields a larger increase (0.043) in the FFR when we account for DM4 when compared to a 0.029 increase in the FFR when we do not account for DM4.

The addition of DM4 (plus T-bill) produces a significant effect. A one-unit increase in DM4 leads to 0.047 increase in the FFR. There is a strong positive relationship between DM4 and the FFR.

We work with lag_DM4

```
> OLSE1 = lm(FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem + lag_DM4, data=fed)
##we run a model controlling for lag_DM4 alongside 3 other variables
> summary(OLSE1)
##we summarize our OLSE1 model
Call:
lm(formula = FEDFUNDS ~ inflation + lag_FEDFUNDS + elec_dem +
    lag_DM4, data = fed)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1064 -0.4317  0.0049  0.4195  5.2204

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.19487    0.22611  -0.862   0.3900
inflation     0.12055    0.02893   4.168 4.94e-05 ***
lag_FEDFUNDS  0.89861    0.02644  33.984  < 2e-16 ***
elec_dem      0.04021    0.01723   2.334   0.0208 *
lag_DM4       0.02859    0.02479   1.153   0.2505
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9611 on 166 degrees of freedom
  (62 observations deleted due to missingness)
Multiple R-squared:  0.9282,    Adjusted R-squared:  0.9265
F-statistic: 536.6 on 4 and 166 DF,  p-value: < 2.2e-16
```

```
> coeftest(OLSE1, vcov=vcovHC (OLSE1, type="HC1"))
##we run a robust regression to test for HC1 type errors
t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)  -0.194866   0.141386 -1.3783 0.1699803
inflation     0.120547   0.035764  3.3706 0.0009329 ***
lag_FEDFUNDS  0.898607   0.035080 25.6162 < 2.2e-16 ***
elec_dem      0.040209   0.022794  1.7640 0.0795712 .
lag_DM4       0.028588   0.020818  1.3732 0.1715277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 29: OLS model controlling for lag_DM4.

A majority of R platform figures have descriptions of the code in them.

Comparison

Lag_DM4 and lag_DM4- has no significance in the model when compared to model with DM4. Lag_DM4 has a positive relationship to the federal funds rate but it has no significance.

Extension 2

Q-1

Is there a correlation between U.*S. GDP, labor participation rate, inflation, number of people 25 years and older with 4 or more years of college* (x-variables) and the *percentage of people below the poverty level* (y-variable)? YES or NO!

Potential Data Sources

U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements.

The World Bank, Data Bank.

Federal Reserve Economic Data, Economic Research Division-Federal Reserve Bank of St. Louis; U.S. Bureau of Labor Statistics.

Earliest and Latest Dates

Percentage of people below the poverty level = 1981 to 2013

College years = 1940 to 2015

GDP = 1961 to 2015

Labor participation rate = 1948 to 2016

Code Book for Variables

A majority of R platform figures have descriptions of the code in them.

The GDP which is a measure of all the finished goods and services produced in a country (World Bank, 2016).

The percentage of people below poverty is derived from the number of people that fall below the poverty threshold as established by the U.S. Census Bureau.[2]

APPENDIX B. ESTIMATES OF POVERTY

How Poverty Is Calculated

Following the Office of Management and Budget's (OMB) Statistical Policy Directive 14, the U.S. Census Bureau uses a set of dollar value thresholds that vary by family size and composition to determine who is in poverty (see the matrix below).

Poverty Thresholds for 2013 by Size of Family and Number of Related Children Under 18 Years

(Dollars)

Size of family unit	Related children under 18 years								
	None	One	Two	Three	Four	Five	Six	Seven	Eight or more
One person (unrelated individual):									
Under age 65	12,119								
Aged 65 and older	11,173								
Two people:									
Householder under age 65	15,600	16,057							
Householder aged 65 and older	14,081	15,996							
Three people	18,222	18,751	18,769						
Four people	24,028	24,421	23,624	23,707					
Five people	28,977	29,398	28,498	27,801	27,376				
Six people	33,329	33,461	32,771	32,110	31,128	30,545			
Seven people	38,349	38,588	37,763	37,187	36,115	34,865	33,493		
Eight people	42,890	43,269	42,490	41,807	40,839	39,610	38,331	38,006	
Nine people or more	51,594	51,844	51,154	50,575	49,625	48,317	47,134	46,842	45,037

Source: U.S. Census Bureau.

Weighted Average Poverty Thresholds in 2013 by Size of Family

(Dollars)

One person	11,888
Two people	15,142
Three people	18,552
Four people	23,834
Five people	28,265
Six people	31,925
Seven people	36,384
Eight people	40,484
Nine people or more	48,065

Source: U.S. Census Bureau.

The labor participation rate is a measure of the labor force that participate in work or have a job.[3]

---

2 http://www.census.gov/library/publications/2014/demo/p60-249.html

3 https://fred.stlouisfed.org/series/CIVPART/

A majority of R platform figures have descriptions of the code in them.

Inflation (consumer price index) as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used (World Bank, 2016).

Instructions on Accessing Data

Data can be accessed by following using potential data sources provided above. Data from the world bank can be derived by tweaking the parameters on the left-hand side of the page.

Data for the labor participation was provided by month and so it had to “averaged” in order to get an annual value.

Data Analysis Methodology

Simple polynomial variables regression or multi-variate regression.

Robust regression

Hypothesis

Null: The percentage of people below poverty level is strongly related to GDP, labor participation rate and the number of people of 4 or more years of college.

A majority of R platform figures have descriptions of the code in them.

Works Cited

Bailey, Michael A. *Real Econometrics: The Right Tools to Answer Important Questions*. First ed. New York: Oxford UP, 2016. Print.

DeNavas-Walt, Carmen, and Bernadette D. Proctor. "Income and Poverty in the United States: 2013." *US Census Bureau*. N.p., 16 Sept. 2014. Web. 09 Oct. 2016.

The Center for Financial Stability. "Advances in Monetary and Financial Measurement (AMFM) - Divisia and Fisher-Ideal Monetary Aggregates - The Center for Financial Stability." *Advances in Monetary and Financial Measurement (AMFM) - Divisia and Fisher-Ideal Monetary Aggregates - The Center for Financial Stability*. N.p., 2016. Web. 09 Oct. 2016.

U.S. Bureau of Labor Statistics. "Civilian Labor Force Participation Rate [CIVPART*]." FRED, Federal Reserve Bank of St. Louis*. N.p., 9 October, 2016. Web. 9 October, 2016.

World Bank. "World Data Bank- World Development Indicators." *The World Bank*. N.p., 2016. Web. 10 Oct. 2016.

A majority of R platform figures have descriptions of the code in them.

Forecasting Techniques Beer Data

Table 1: Codebook for Beer Data.

Variable	Description
trend	trend is the independent variable that shows the compounded change in mega liters of beer per quarter.
beer2	beer2 is quarterly beer production derived from beer (monthly beer production). It is the dependent variable.
season2	Change (increase or decrease) in beer production when compared to first quarter.
season3	Change (increase or decrease) in beer production when compared to first quarter.
season4	Change (increase or decrease) in beer production when compared to first quarter.

s2- s4 are dummy variables*

```
> beer
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1991 164 148 152 144 155 125 153 146 138 190 192 192
1992 147 133 163 150 129 131 145 137 138 168 176 188
1993 139 143 150 154 137 129 128 140 143 151 177 184
1994 151 134 164 126 131 125 127 143 143 160 190 182
1995 138 136 152 127 151 130 119 153
> beer2 <- window(ausbeer,start=1992,end=2006-.1)
> beer2
     Qtr1 Qtr2 Qtr3 Qtr4
1992  443  410  420  532
1993  433  421  410  512
1994  449  381  423  531
1995  426  408  416  520
1996  409  398  398  507
1997  432  398  406  526
1998  428  397  403  517
1999  435  383  424  521
2000  421  402  414  500
2001  451  380  416  492
2002  428  408  406  506
2003  435  380  421  490
2004  435  390  412  454
2005  416  403  408  482
```

Figure 1: Code to convert monthly data to quarterly data.

Beer Model Equation

$$beer2_i = \beta_0 + \beta_1 trend_i + \beta_2 season2_i + \beta_3 season3_i + \beta_4 season4_i + \epsilon_i$$

Where $\beta_0 = y - intercept$, $\beta_1 = coefficient\ for\ trend_i$, $\beta_2 = coefficient\ for\ season2_i$, $\beta_3 = coefficeint\ for\ season3_i$, $\beta_4 = coefficient\ for\ season4_i$, $\epsilon_i = error\ term$

A majority of R platform figures have descriptions of the code in them.

R platform code

```
> fit <- tslm(beer2~trend+season)
> summary(fit)

Call:
tslm(formula = beer2 ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max
-44.024  -8.390   0.249   8.619  23.320

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 441.8141     4.5338  97.449  < 2e-16 ***
trend        -0.3820     0.1078  -3.544 0.000854 ***
season2     -34.0466     4.9174  -6.924 7.18e-09 ***
season3     -18.0931     4.9209  -3.677 0.000568 ***
season4      76.0746     4.9268  15.441  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.01 on 51 degrees of freedom
Multiple R-squared:  0.921,     Adjusted R-squared:  0.9149
F-statistic: 148.7 on 4 and 51 DF,  p-value: < 2.2e-16
```

Multivariate regression results for Beer data

trend	-0.38*
	(0.11)
	[t=-3.54]
season2	-34.05*
	(4.92)
	[t=-6.92]
season3	-18.09
	(4.92)
	[t=-3.68]
season4	76.07
	(4.92)
	[t=15.44]
Intercept	441.81
	(4.53)
	[t=97.45]
N	52
R^2	0.92

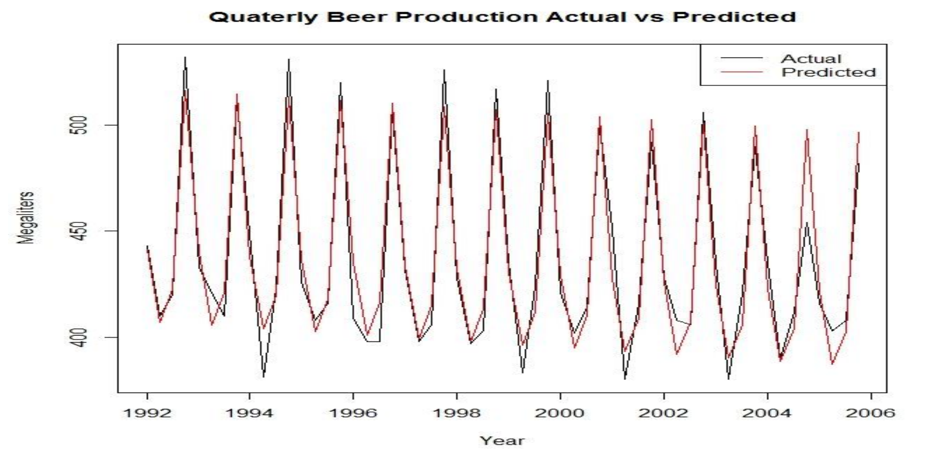
Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

A majority of R platform figures have descriptions of the code in them.

Graph 1 (code and graph)

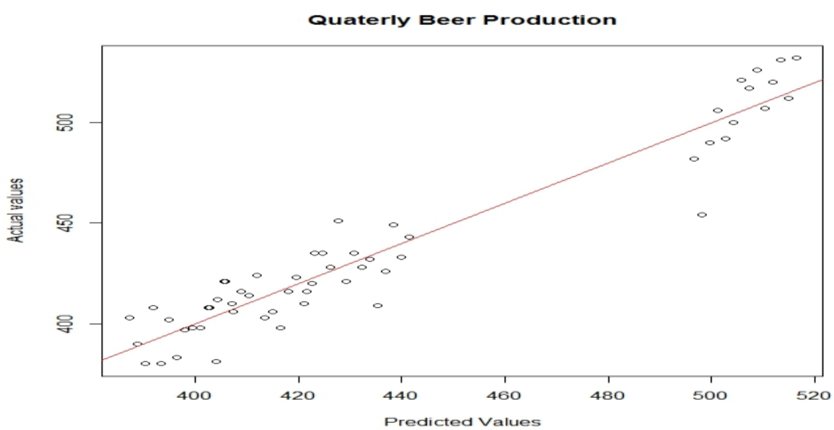
```
> plot(beer2,xlab="Year",ylab="Megaliters",main="Quaterly Beer Production Actual vs Predicted")
> lines(fitted(fit),col=2)
> legend("topright",lty=1,col=c(1,2),legend=c("Actual","Predicted"))
```



This shows the predicted value versus the actual value.

Code and Graph 2

```
> plot(beer2,xlab="Year",ylab="Megaliters",main="Quaterly Beer Production Actual vs Predicted")
> lines(fitted(fit),col=2)
> legend("topright",lty=1,col=c(1,2),legend=c("Actual","Predicted"))
> plot(fitted(fit),beer2,xy.lines=FALSE,xy.labels=FALSE,xlab="Predicted Values",ylab="Actual values",main="Quaterly Beer Production")
> abline(0,1,col="red")
```

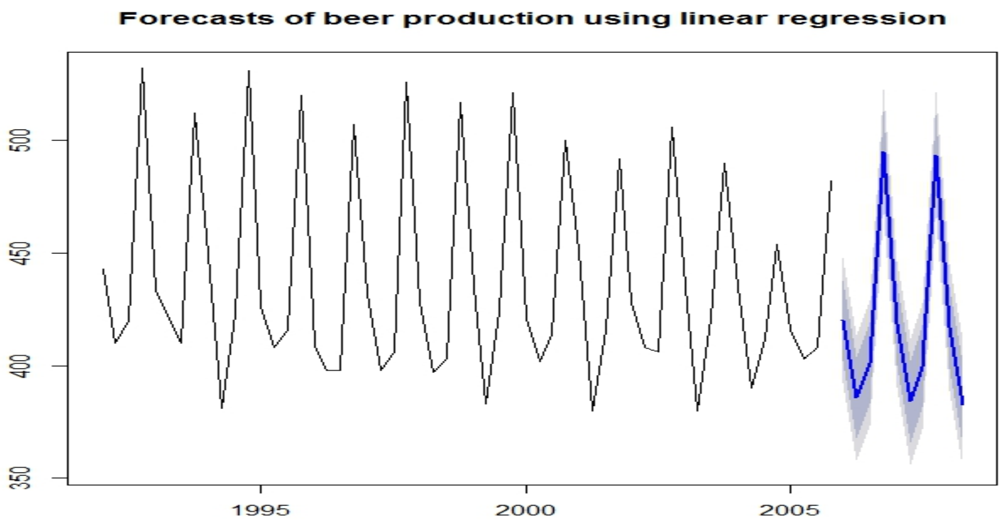


The fitted is used to show how close the predicted value is to the actual value. The line passes through a majority of the points showing that the actual and the predicted values are not so far apart. The R square value of 0.92 also shows that most of the values are close to the fitted line.

A majority of R platform figures have descriptions of the code in them.

Code and Graph 3

```
> fcast <- forecast(fit)
> plot(fcast,main="Forecasts of beer production using linear regression")
```



The thick or dark blue line shows an 80% confidence interval that the prediction will fall into that range and a 95% confidence interval is going to show a 95% confidence interval that the prediction will fall into that range. The graph produces results that can help firms determine their level of beer production needed to gain competitive advantage.

A majority of R platform figures have descriptions of the code in them.

Forecasting Techniques (Philips Curve)

Variables for Philips Curve Replication

cpi, cpi ts	Inflation-consumer price index, cpi time series
u3, u3 ts	Unemployment rate quarterly, u3 time series
mets (me time series)	cpi time series object (4*ts)
mets2 (me time series 2)	u3 time series object (ts)

Preliminary analysis

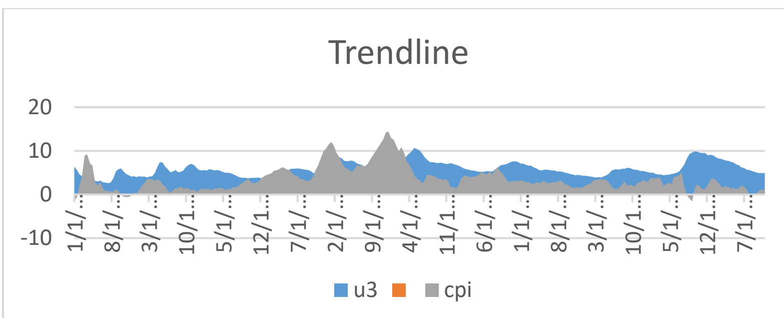


Figure 2: Trend line of cpi and u3.

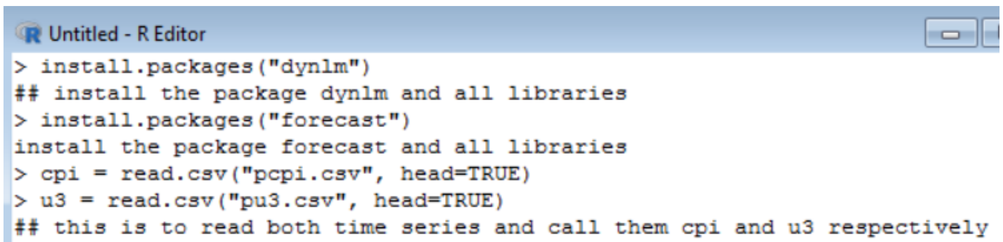
It is important to notice that there is a difference in the trend for cpi and u3 and noticing this provides some comfort as we move forward in the forecasting process.

A majority of R platform figures have descriptions of the code in them.

Preliminary Steps

<u>Step 1</u>

The first step of the replication process is to load the data into R. My working directory contains all my files and so it is advisable to change the working directory into the folder with all relevant files.



```
Untitled - R Editor
> install.packages("dynlm")
## install the package dynlm and all libraries
> install.packages("forecast")
install the package forecast and all libraries
> cpi = read.csv("pcpi.csv", head=TRUE)
> u3 = read.csv("pu3.csv", head=TRUE)
## this is to read both time series and call them cpi and u3 respectively
```

Figure 3: Code to read cpi and u3 time series.

A majority of R platform figures have descriptions of the code in them.

Step 2

Our next step is to convert the cpi and u3 data into a time series object with each year containing four quarters.

mets = me time series (cpi ts)

mets2 = me time series 2 (u3 ts)

```
> mets <- 4*ts(cpi, start=c(1950,1 ), frequency=4)
##this is to convert cpi into a time series object with 4 quaters
> mets
          Qtr1      Qtr2      Qtr3      Qtr4
1950 -5.94744 -2.50868  8.19664 17.46408
1951 35.78244 36.68952 28.59148 26.30704
1952 11.20752  8.32468 11.32148  5.77572
1953  3.07376  3.51872  2.63996  2.89172
##this is the result of mets(me time series)
> mets2 <- ts(u3, start=c(1950,1 ), frequency=4)
##this is to rearrange u3 into 4 quaters horizontally
> mets2
     Qtr1 Qtr2 Qtr3 Qtr4
1950  6.4  5.6  4.6  4.2
1951  3.5  3.1  3.2  3.4
1952  3.1  3.0  3.2  2.8
1953  2.7  2.6  2.7  3.7
##this is the result of mets2 (me time series 2)
```

Figure 4: Code to create a time series object for u3 and cpi.

Step 3

The next step will be to plot both time series and see what they look like. This is to have a feel of the time series data.

```
> > plot.ts(mets, xlim=c(1950,2018), main="cpi time series")
## A plot of our cpi ts (mets)
> plot.ts(mets2, xlim=c(1950,2018), main="u3 time series")
## A plot of our u3 ts (mets2)
```

Figure 5: Code to plot u3 and cpi time series.

A majority of R platform figures have descriptions of the code in them.

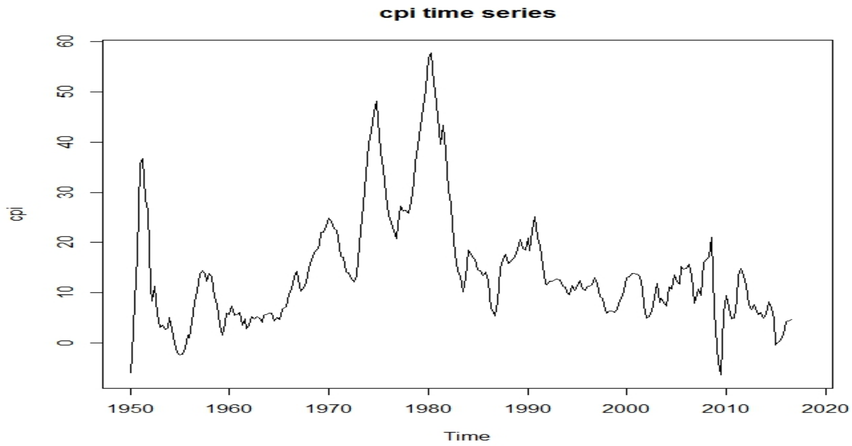


Figure 6: Cpi time series (mets) plot.

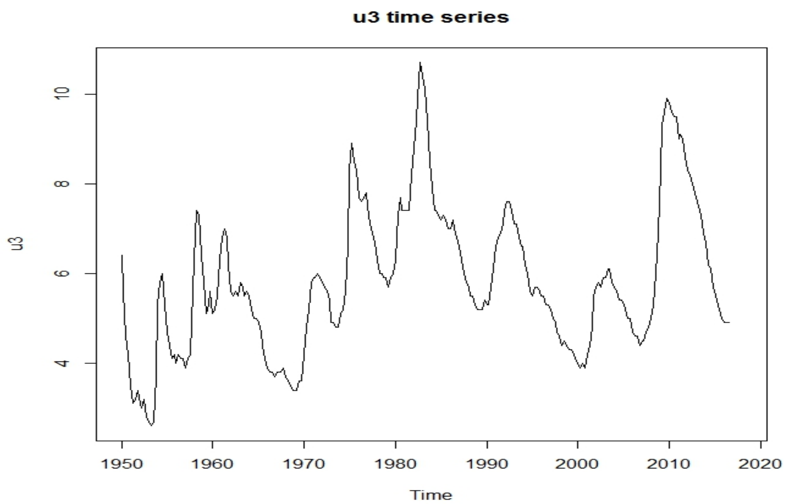


Figure 7: U3 time series (mets2) plot.

Step 4

One of the things I found particularly interesting was to decompose the data in order to see the trend of the data for both time series. This was done by smoothing the data using the Simple Moving Average (SMA) method. We assume the mets and mets2 time series as non-seasonal (a trend and irregular component) and so all we need to do is get a better picture of the trends. A SMA model with an order (n=20) was used after several trials with lower orders.

A majority of R platform figures have descriptions of the code in them.

```
> install.packages ("TTR")
##install the TTR package with SMA function
> library(TTR)
## load or call the TTR package for use
> SMAmets <- SMA(mets,n=20)
## use the SMA fucntion with order 20 on mets (cpi ts)
> plot(SMAmets)
## plot the SMAmets to look at clear trend
> SMAmets2 <- SMA(mets2,n=20)
## use the SMA fucntion with order 20 on mets2 (u3 ts)
> plot(SMAmets2)
## plot the SMAmets2 to look at clear trend
```

Figure 8: SMA Code using "TTR" package.

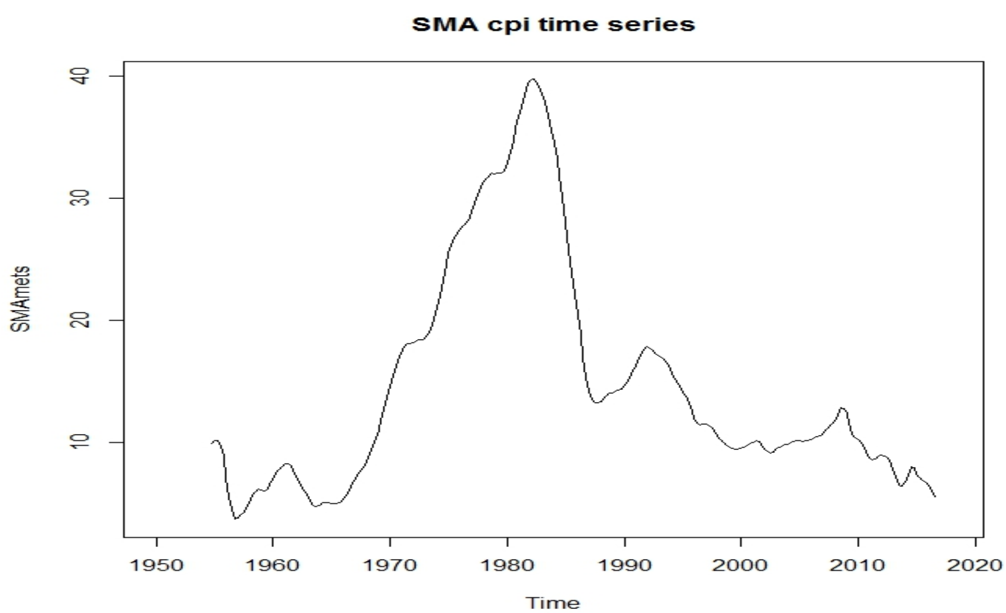


Figure 9: SMA cpi time series.

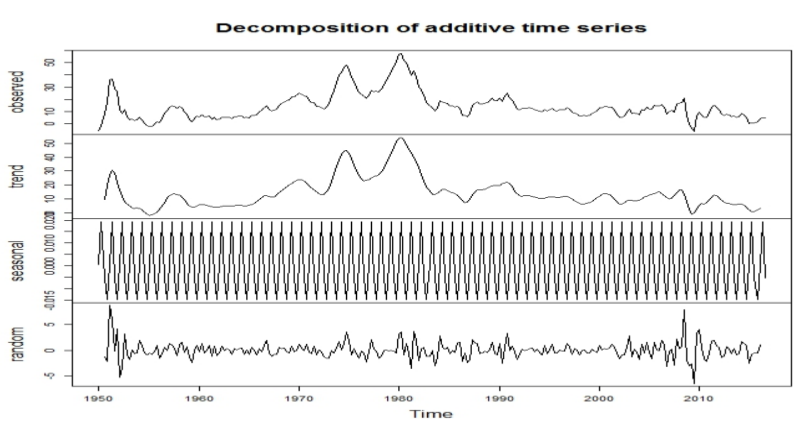


Figure 10: Cpi decomposed series.

A majority of R platform figures have descriptions of the code in them.

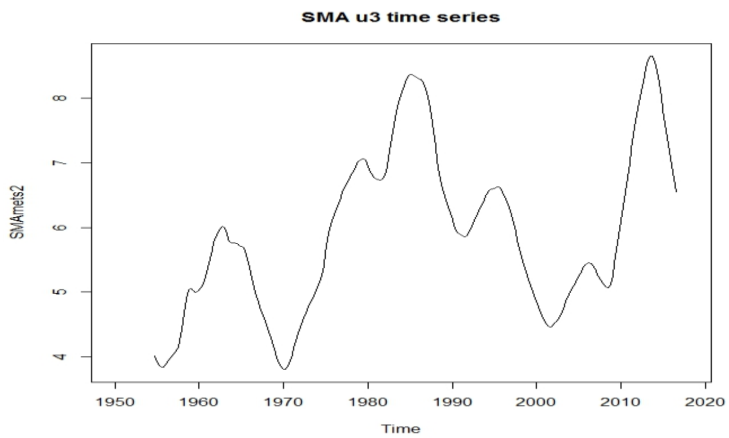


Figure 11: SMA u3 time series.

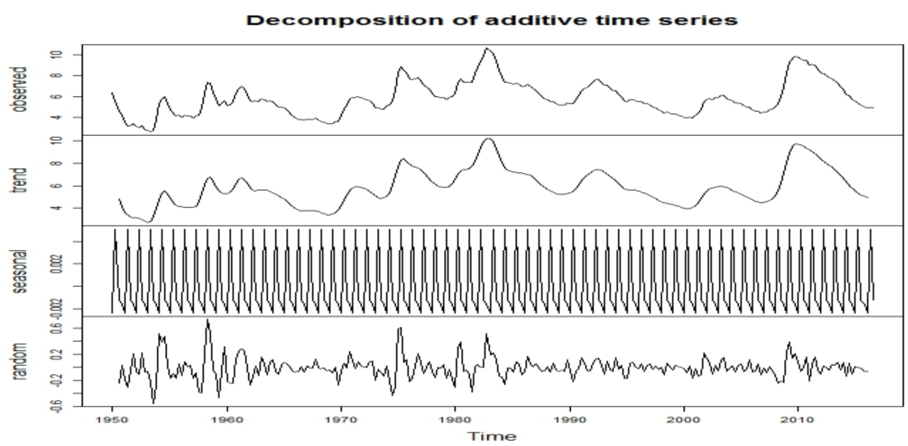


Figure 12: u3 decomposed series.

Our figures above show that 1980 - 1990 (no seasonal considerations) experienced high cpi and high unemployment rate. Both figures also show a slightly similar pattern with some exceptions.

REPLICATION FOR PHILIPS CURVE (HW)

The preliminary analysis has already created the time series for Inflation-cpi (mets) and for unemployment rate (mets2).

Part One (four lags for cpi and four lags for Unemployment)

Our objective is to forecast cpi with 4 lags of inflation-cpi (cpi) and 4 lags of unemployment (u3).
A majority of R platform figures have descriptions of the code in them.

```
> philips.fit <- dynlm((mets) ~ L(mets, c(1,2,3,4)) +
L(mets2, c(1,2,3,4)), start=c(1950,1), end=c(2016,3))
## Create a fit model that can be used to predict cpi and u3
> summary(philips.fit)
## make a summary of the philips model in order to ensure 4+4 lags cpi&u3
Time series regression with "ts" data:
Start = 1951(1), End = 2016(3)

Call:
dynlm(formula = (mets) ~ L(mets, c(1, 2, 3, 4)) + L(mets2, c(1,
    2, 3, 4)), start = c(1950, 1), end = c(2016, 3))

Residuals:
     Min       1Q   Median       3Q      Max
-14.4627  -1.4615   0.0729   1.3565  14.1880

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.84004    0.71696   1.172  0.24243
L(mets, c(1, 2, 3, 4))1   1.28014    0.06220  20.580  < 2e-16 ***
L(mets, c(1, 2, 3, 4))2  -0.32258    0.10134  -3.183  0.00164 **
L(mets, c(1, 2, 3, 4))3   0.09922    0.10017   0.990  0.32288
L(mets, c(1, 2, 3, 4))4  -0.10700    0.06107  -1.752  0.08097 .
L(mets2, c(1, 2, 3, 4))1 -1.48620    0.63728  -2.332  0.02048 *
L(mets2, c(1, 2, 3, 4))2  1.25890    1.20756   1.043  0.29817
L(mets2, c(1, 2, 3, 4))3  0.21441    1.20838   0.177  0.85931
L(mets2, c(1, 2, 3, 4))4 -0.01483    0.63411  -0.023  0.98136
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.878 on 254 degrees of freedom
Multiple R-squared:  0.9389,    Adjusted R-squared:  0.937
F-statistic: 488.3 on 8 and 254 DF,  p-value: < 2.2e-16
> forecast <- forecast(philips.fit$fitted,5)
## create a forecast model by 5 years for cpi
> plot(forecast,xlim=c(2013,2018),ylim=c(-4,12),
main="Mets", xlab="time", ylab="cpi", col="red")
## plot the cpi forecast model from above
```

Figure 13: Code for first cpi forecast figure.

A majority of R platform figures have descriptions of the code in them.

Table 1: Dynamic linear model results (Philips curve).

	Dynamic Regression Results (4 cpi lags and 4 u3 lags)
Lag1 mets	**1.28***
	(0.06)
	[t = 20.58]
Lag2 mets	**-0.32***
	(0.10)
	[t = -3.18]
Lag3 mets	**0.09**
	(0.10)
	[t = 0.99]
Lag4 mets	**-0.11**
	(0.061)
	[t = -1.75]
Lag1 mets2	**-1.49***
	(0.64)
	[t = -2.33]
Lag2 mets2	**1.26**
	(1.21)
	[t = 1.04]
Lag3 mets2	**0.21**
	(1.21)
	[t = 0.18]
Lag4 mets2	**-0.01**
	(0.63)
	[t = -0.023]
Constant	**0.84**
	(0.72)
	[t = 1.17]
N	**255**
R^2	**0.94**

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

mets is the cpi time series object; mets2 is the unemployment (u3) time series object.

A majority of R platform figures have descriptions of the code in them.

The first and second lag of mets (cpi ts) are significant in the "dynlm", as well as, the first lag of mets2 (u3 ts).

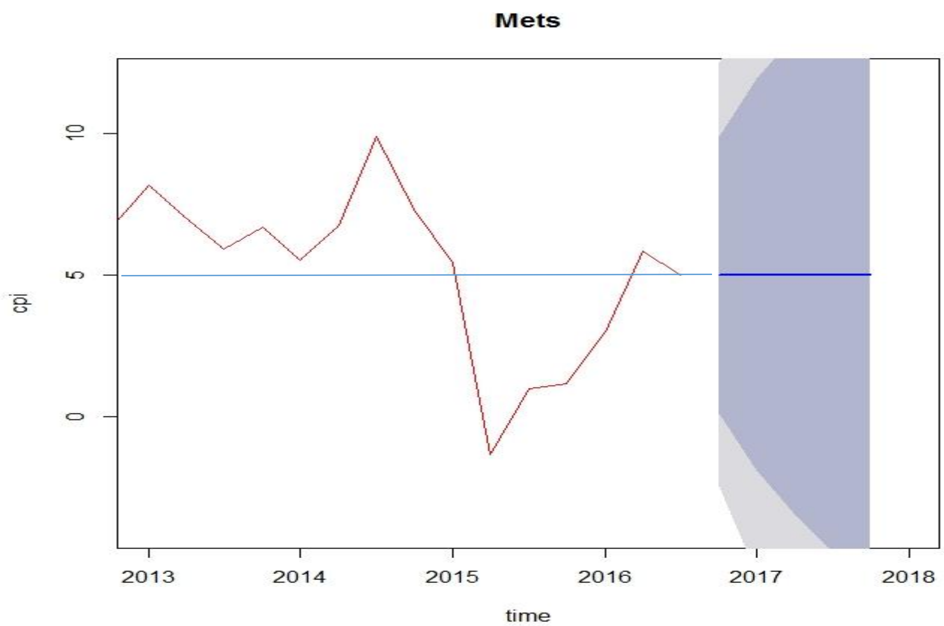


Figure 14: Inflation-cpi 5-year forecast.

Figure 14 shows a flat forecast line at about 5 for the consumer price index.

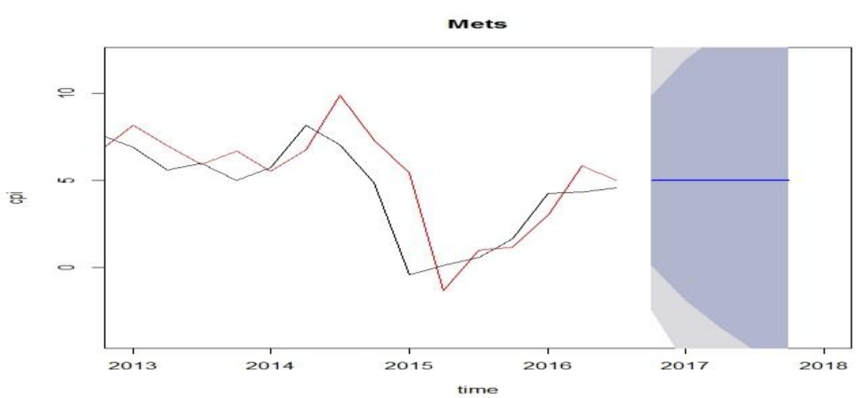


Figure 15: cpi forecast and mets (cpi ts) time series together.

Figure 15 & 16 shows a small difference in the trend for the predicted cpi and the actual cpi.

A majority of R platform figures have descriptions of the code in them.

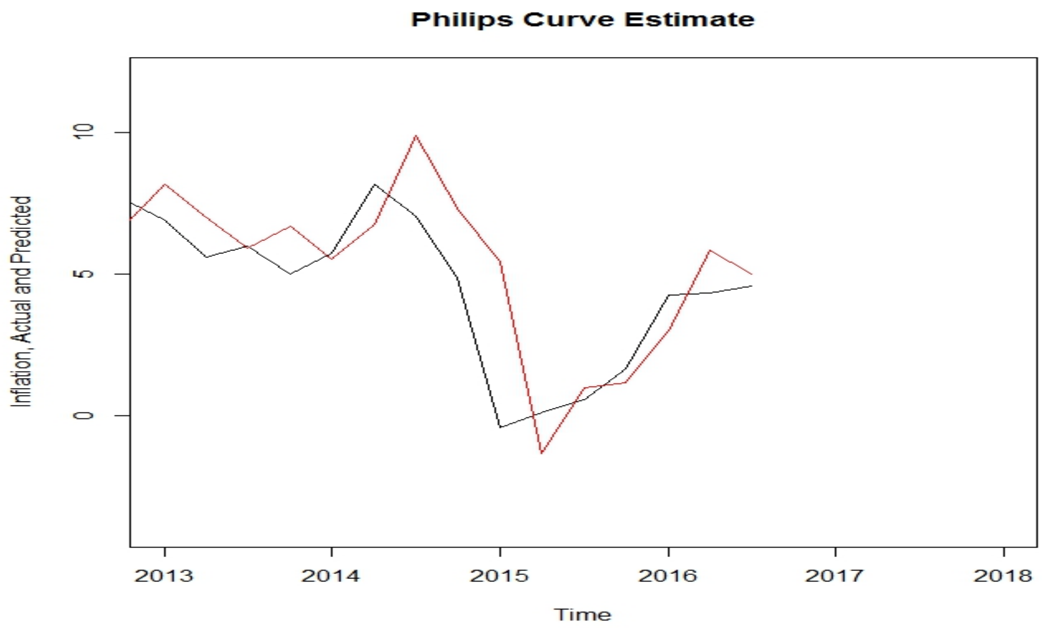


Figure 16: Philips curve estimate.

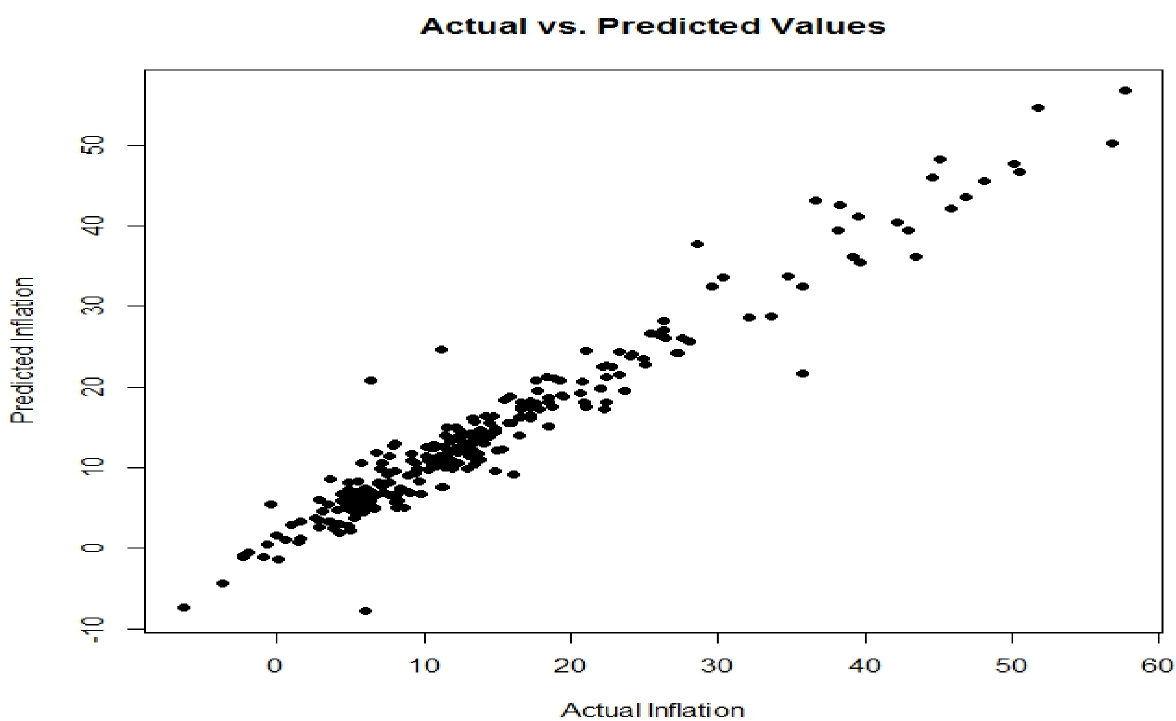


Figure 17: Actual Versus Predicted Inflation Values.

A majority of R platform figures have descriptions of the code in them.

Figure 16 shows a comparison between the predicted inflation value and the actual inflation value. It can be seen from our figure above that most of our predicted values are not significantly far-off from the actual inflation values. I can see a few off points but overall I can guess that our R-square value will be good.

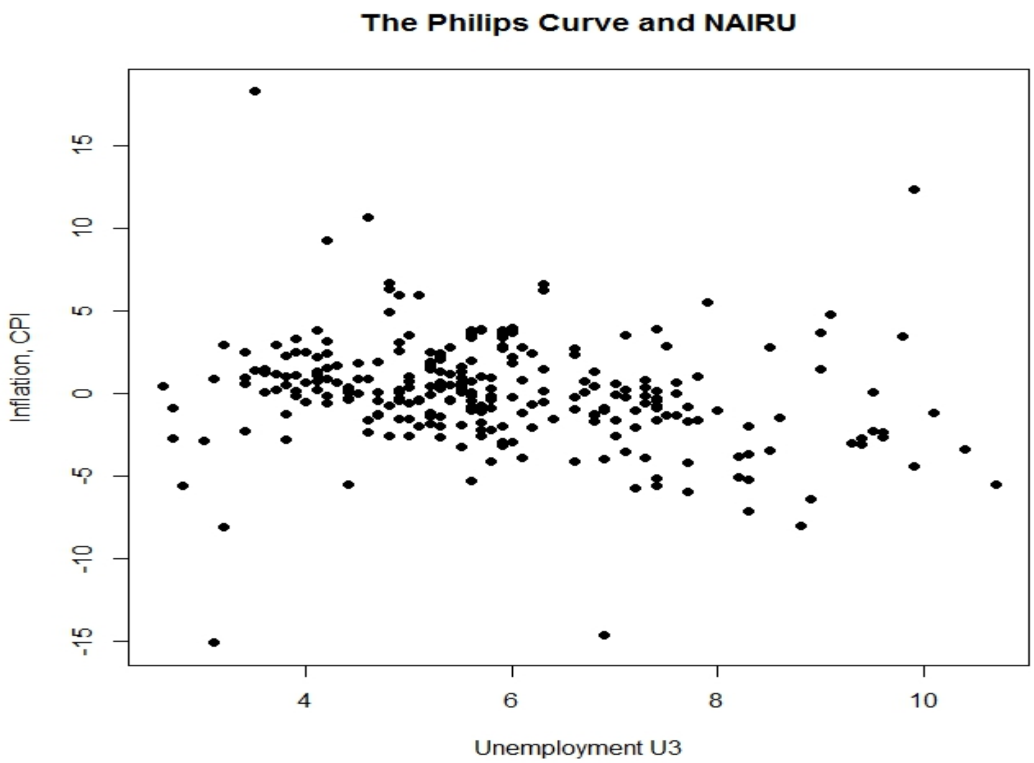


Figure 18: Philips curve and NAIRU (Non-accelerated inflation rate of unemployment).

There is no significant difference in the Philips curve when we lag the cpi time series (NAIRU) and when we keep the cpi time series the same.

EXTENSION (HW)

This time around we run the same Philips curve as above but with two lags for inflation and one lag for unemployment.

A majority of R platform figures have descriptions of the code in them.

```
> philips1.fit <- dynlm((mets) ~ L(mets, c(1,2)) + L(mets2, c(1)), start=c(1950,1), end=c(2016,3))
## Create a fit model that can be used to predict cpi and u3
> summary(philips1.fit)
## make a summary of the philips1 model
Time series regression with "ts" data:
Start = 1950(3), End = 2016(3)

Call:
dynlm(formula = (mets) ~ L(mets, c(1, 2)) + L(mets2, c(1)), start = c(1950,
    1), end = c(2016, 3))

Residuals:
     Min       1Q   Median       3Q      Max
-15.7779  -1.5765  -0.0393   1.3417  14.5327

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.41948    0.70174   2.023   0.0441 *
L(mets, c(1, 2))1   1.35148    0.05687  23.765  < 2e-16 ***
L(mets, c(1, 2))2  -0.40922    0.05718  -7.157 8.34e-12 ***
L(mets2, c(1))     -0.09950    0.11567  -0.860   0.3905
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.977 on 261 degrees of freedom
Multiple R-squared:  0.933,     Adjusted R-squared:  0.9322
F-statistic:  1211 on 3 and 261 DF,  p-value: < 2.2e-16
```

Figure 19: Results for 2 cpi lags and 1 u3 lag.

Table 2: Dynamic linear model results (Philips curve)

	Dynamic Regression Results (2 cpi lags and 1 u3 lag)
Lag1 mets	**1.35***
	(0.06)
	[t = 23.77]
Lag2 mets	**-0.41***
	(0.06)
	[t = -7.16]
Lag1 mets2	**-0.099**
	(0.12)
	[t = -0.86]
Constant	**1.42***
	(0.70)
	[t = 2.023]
N	**262**
R^2	**0.93**

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

mets is the cpi time series object; mets2 is the unemployment (u3) time series object.

A majority of R platform figures have descriptions of the code in them.

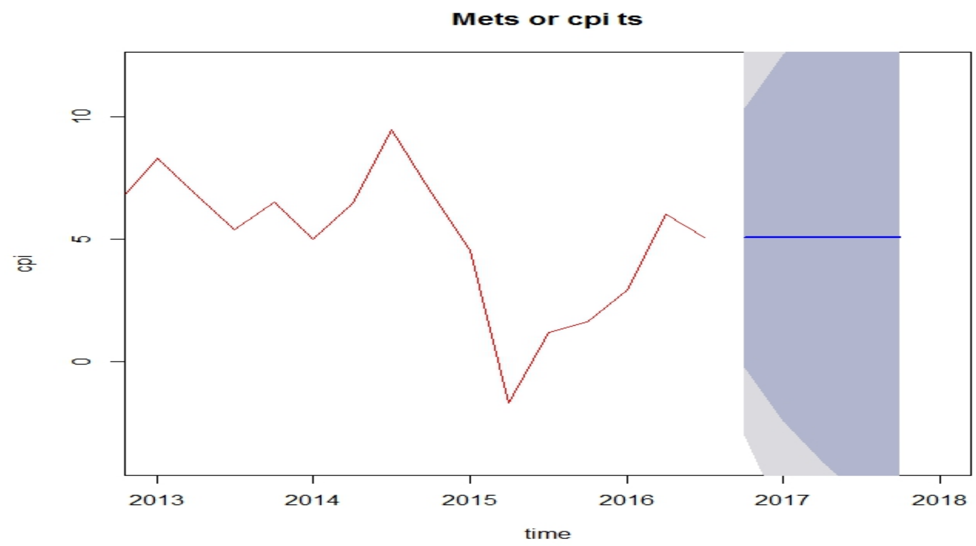


Figure 20: Inflation-cpi 5-year forecast.

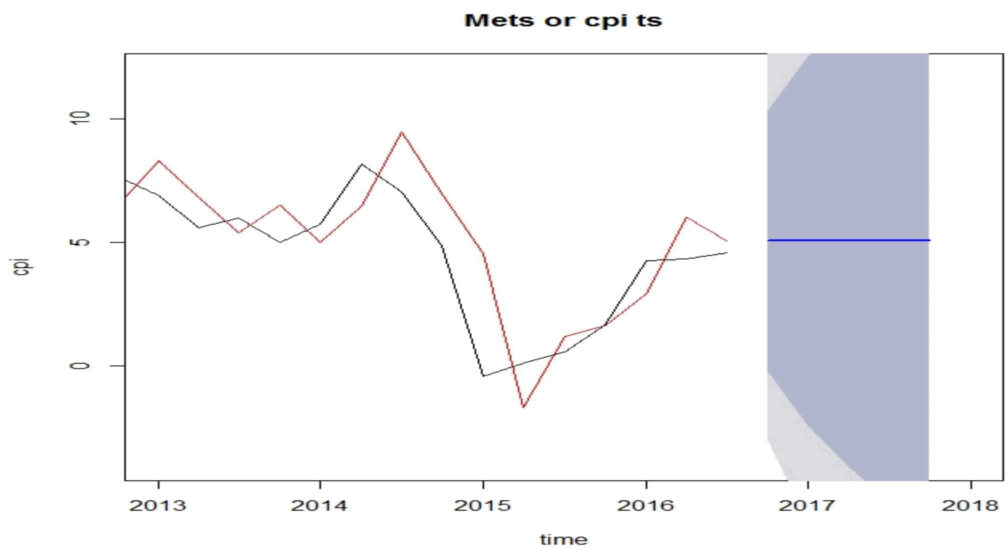


Figure 21: cpi forecast and mets (cpi ts) time series together.

A majority of R platform figures have descriptions of the code in them.

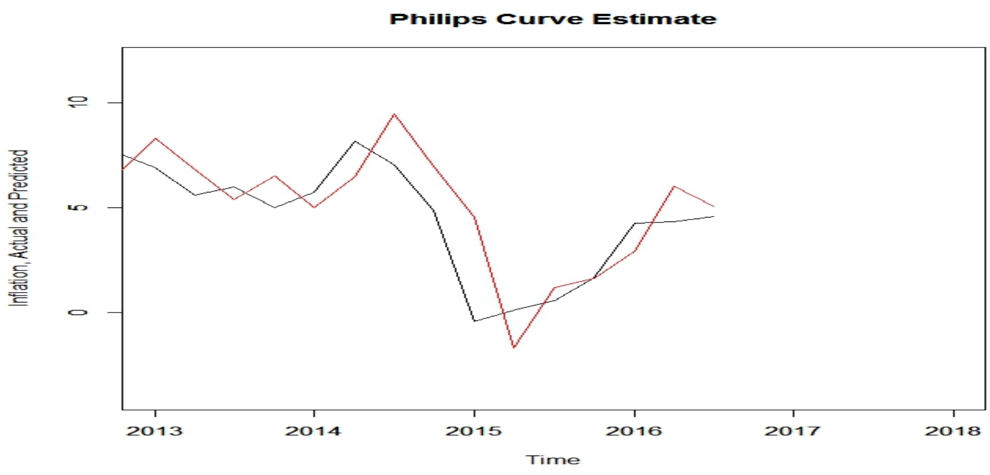


Figure 22: Philips curve estimate.

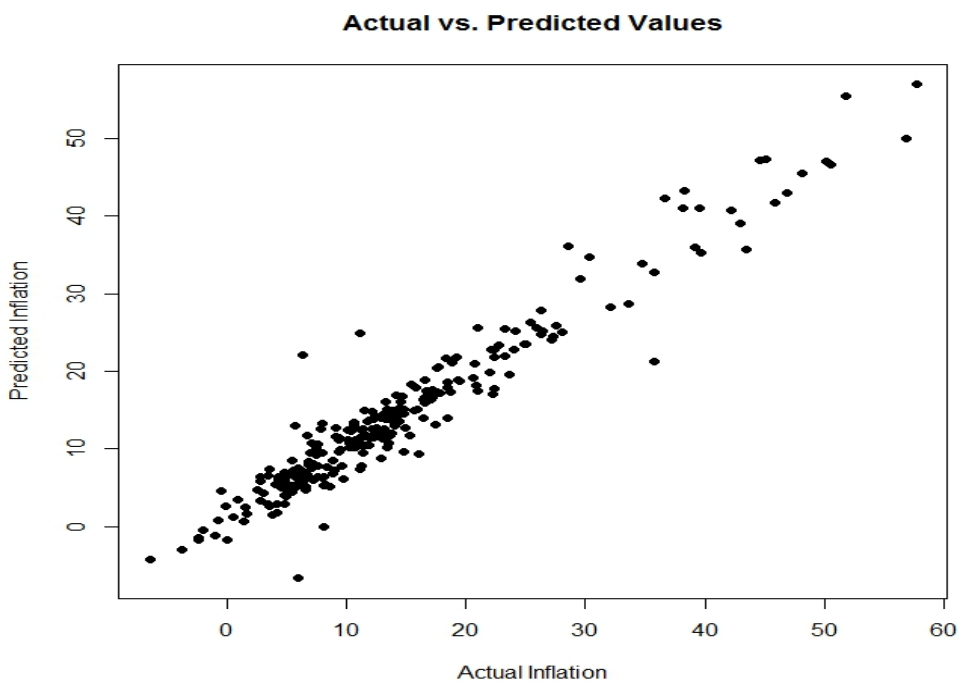


Figure 23: Actual Versus Predicted Inflation Values.

A majority of R platform figures have descriptions of the code in them.

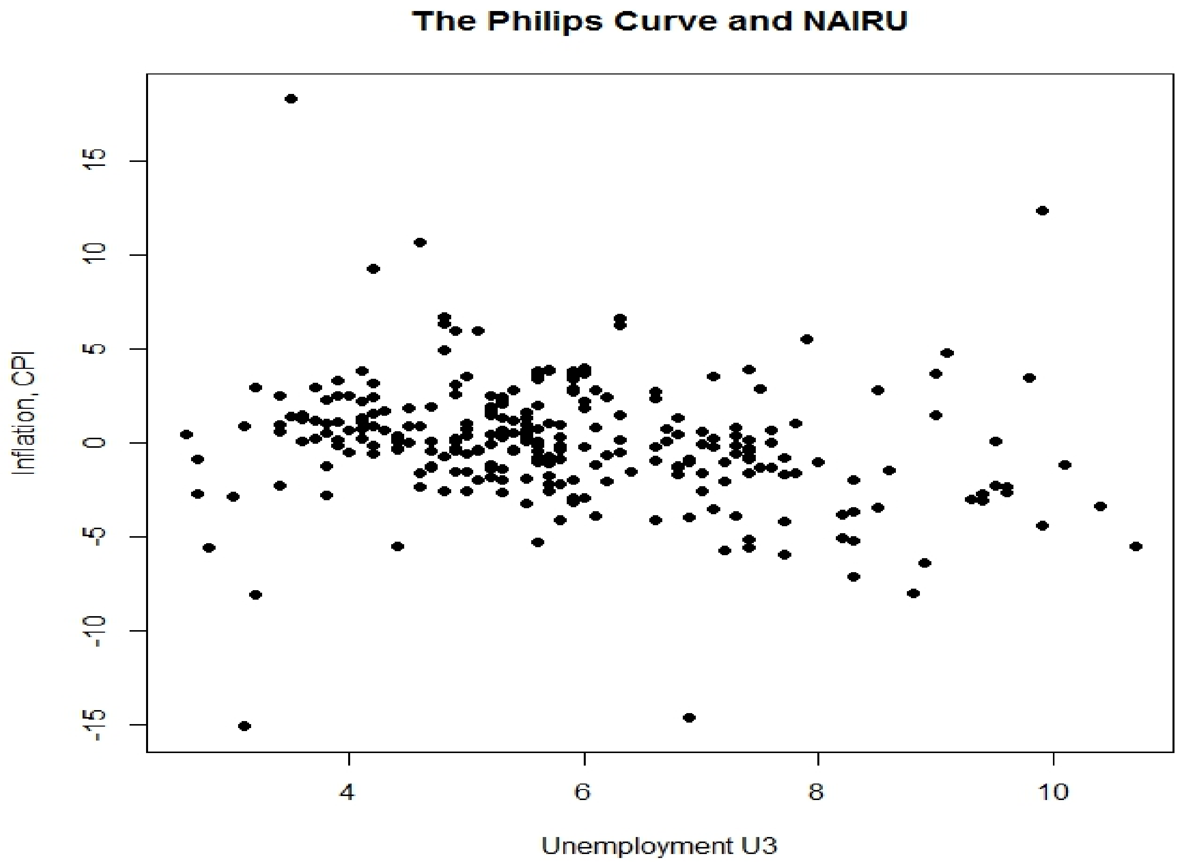


Figure 24: Philips curve and NAIRU- jitter effect (Non-accelerated inflation rate of unemployment).

Conclusion

There is no difference (no significant change) in the forecast from the replication with 4 cpi lags & 4 u3 lags versus 2 lags of cpi & 1 lag of u3. The dynamic regression results (replication and extension) show a slight difference in the significance of some values but no difference in the forecast.

A majority of R platform figures have descriptions of the code in them.

Works Cited

Bailey, Michael A. *Real Econometrics: The Right Tools to Answer Important Questions*. First ed. New York: Oxford UP, 2016. Print.

"Civilian Unemployment Rate." - *FRED*. N.p., 04 Nov. 2016. Web. 08 Nov. 2016.

A majority of R platform figures have descriptions of the code in them.

Appendix

Code for Forecasting Techniques

Install necessary libraries and packages here. Get "dynlm" and "forecast".

Change Working Dir to my [1] "C:/Users/Einstein666/Documents/WTAMU MBA FILE/ECON 6320_01"

I have all my data in this location and so no need for complex code when reading my data

> u3 = read.csv ("pu3.csv", head=TRUE)

> cpi = read.csv ("pcpi.csv), head = TRUE)

mets <- 4*ts(cpi, start=c(1950,1), frequency=4)

mets stands for me time series

mets2 <- ts(u3, start=c(1950,1), frequency=4)

mets2 stands for me time series 2

> philips.fit <- dynlm((mets) ~ L(mets, c(1,2,3,4)) + L(mets2, c(1,2,3,4)), start=c(1950,1), end=c(2016,3))

> summary(philips.fit)

> forecast <- forecast(philips.fit$fitted,5)

5 PERIOD AHEAD FORECAST

> plot(forecast,xlim=c(2013,2018),ylim=c(-4,12), main="Mets", xlab="time", ylab="cpi", col="red")

GRAPH STARTS 2013 TO 2018

A majority of R platform figures have descriptions of the code in them.

```
> lines(mets)

## create a line of original mets time series to compare time series to prediction

> plot(mets, xlim=c(2013,2018), ylim=c(-4,12), main="Philips Curve Estimate", xlab="Time", ylab="Inflation, Actual and Predicted")

## LIMITS THE X AXIS TO BE BETWEEN 2013 AND 2018

lines(philips.fit$fitted, col="red")

## create a line of the philips.fit to compare actual vs predicted inflation

> plot(jitter(mets), jitter(philips.fit$fitted), main="Actual vs. Predicted Values", xlab="Actual Inflation", ylab="Predicted Inflation", pch=19)

## Scatter plot of predicted vs actual inflation values

> plot(mets2,diff(mets,lag=1), main="The Philips Curve and NAIRU", xlab="Unemployment U3", ylab="Inflation, CPI", pch=19)
```

EXTENSION CODE

```
> philips1.fit <- dynlm((mets) ~ L(mets, c(1,2)) + L(mets2, c(1)), start=c(1950,1), end=c(2016,3))

> summary(philips1.fit)

> forecast2 <- forecast(philips1.fit$fitted,5)

## 5 PERIOD AHEAD FORECAST

> plot(forecast2,xlim=c(2013,2018),ylim=c(-4,12), main="Mets", xlab="time", ylab="cpi", col="red")

## GRAPH STARTS 2013 TO 2018
```

A majority of R platform figures have descriptions of the code in them.

> lines(mets)

create a line of original mets time series to compare time series to prediction

> plot(mets, xlim=c(2013,2018), ylim=c(-4,12), main="Philips Curve Estimate", xlab="Time", ylab="Inflation, Actual and Predicted")

LIMITS THE X AXIS TO BE BETWEEN 2013 AND 2018

lines(philips1.fit$fitted, col="red")

create a line of the philips.fit to compare actual vs predicted inflation

> plot(jitter(mets), jitter(philips1.fit$fitted), main="Actual vs. Predicted Values", xlab="Actual Inflation", ylab="Predicted Inflation", pch=19)

Scatter plot of predicted vs actual inflation values

> plot(mets2,diff(mets,lag=1), main="The Philips Curve and NAIRU", xlab="Unemployment U3", ylab="Inflation, CPI", pch=19)

A majority of R platform figures have descriptions of the code in them.

A Critique- "The fatal conceit of foreign intervention" by Dr. Christopher Coyne

Executive Summary

This paper attempts to provide a thorough critique of a *Free Market Institute (FMI)* presentation by Doctor Christopher Coyne at *Texas Tech University* on the topic provided above (the fatal conceit of foreign intervention). This paper will provide an analysis of Doctor Coyne's work on this topic by proposing a method to statistically measure this topic.

Doctor Coyne focuses on two main points in this topic, which are knowledge constraints and incentives. His claim is that our understanding of the idea "fatal conceit of foreign intervention" is grounded in these two main points. He claims that realizing that we have knowledge constraints should be an incentive for home economies or countries to abstain from intervention in the affairs of foreign countries or economies. This critique will delve into his work from a quantitative or econometric standpoint by finding a way to measure his claim that central economic planning or foreign intervention can be detrimental to the well-being of our human society.

Background & Summary

It is relevant to provide a brief introduction of Doctor Christopher Coyne in this critique. Doctor Christopher Coyne (who was the presenter at Texas Tech University) is the F.A. Harper Professor of economics at George Mason University and the associate director of the F.A. Hayek program for advanced study in philosophy, politics and economics at the Mercatus center at George Mason university. I find it relevant to mention his work at the Hayek program because his presentation on the "fatal conceit of foreign intervention" has its roots in Hayek's work.

Doctor Coyne places his idea of the "fatal conceit of foreign intervention" in line or parallel to Hayek's (1988) "conceit of socialism" (Coyne & Mathers 3). I am a little familiar with Hayek's work from my macroeconomic class with Doctor Rex Pjesky in which we talked about the “rational economic order” and “pareto efficiency”. Hayek mentions that the unusual nature of the problem of a “rational economic order” is that the information or knowledge needed to solve societal concerns is not given to a single mind (a centralized system) but this knowledge is dispersed (decentralized) with individuals possessing parts of the whole that are often incomplete and contradictory (Hayek 519). A blunt version of Hayek’s statement might imply anarchy where there is no need for a central planner or any form of law & order but this is not what Hayek means and neither is this what Doctor Coyne advocates.

In summary the presentation talked about the fact that intentions do not always equal good results, a reliance on top-down planning might not be the ideal way to find solutions, and individualism will be preferable to collectivism. Dr. Coyne talks about the fact that knowledge constraints should be a determinant when developing foreign intervention plans; secondly, he talks about the fact that institutions have the ability to offer incentives that shape human behavior (Coyne & Mathers 225). These incentives also determine the way foreign intervention plans are carried out.

A Brief Criticism

Individuals often say that politicians sit down in Washington coming up with fancy numbers and terms that do not really provide any solutions (this is the classic aggregates problem proposed by Hayek). While I am an advocate for individualism over collectivism, it is important to realize that we live in a planet dominated by humans and humans by default are social beings. While Hayek disputes the fact that knowledge is consolidated in a central authority, the constructs of human society have shown that some people are predisposed to more information or knowledge than others. The very essence of human design shows that we are all equals in an abstract sense but genetically and physically discrete (hence a paradox). While it might not be acceptable for a country like the United States to intervene in the Middle East, it is important to notice that the

A majority of R platform figures have descriptions of the code in them.

international environment is like a cob web whereby a crisis in the Middle-East can have an effect on gas prices worldwide. The high level of inter-connectivity between nation states has placed humans in an environment where we must try to be “our neighbor’s keeper”. We must realize that the constructs of human society are enormously complex and it is very difficult to talk about the “fatal conceit of foreign intervention”. I often tell people to realize that when you live in an apartment complex and your neighbor is under attack, it might be a smart move to go out and intervene in the crisis before a gun-shot because you never know where the bullet might be going. Let’s say for an instance that the bullet travels to your destination because you failed to call the cops or intervene before the gun was fired (this provides an incentive for human action and hence foreign intervention). In this light, I seriously think that foreign intervention might not be conceited because countries have the obligation to protect their neighbors before the crisis escalates and destroys everyone.

The programmatic structure of foreign intervention reveals that most countries like the USA who decide to intervene do not think it through especially in regards to finances or resources needed to sustain foreign intervention (validating Hayek’s claim that men know very little about what they imagine they can design). The main idea here should be parallel to the famous environmental concept of sustainable development. The main idea I propose is called sustainable foreign intervention. Then again, I realize that Dr. Coyne makes a good point when he says that human desires cannot really shape the world because a policy of foreign intervention by democrats could become unsustainable once a republican president comes into office. The dilemma here lies in the fact that human desires are contradictory and so cannot be trusted to shape our world but at the same time human positive desires if executed in a sustainable/continuous manner can be a tool to shape our world. The discontinuity of human ideas should not be a reason to think of foreign intervention as conceited, any more as we can’t think of saving our neighbor from a robber or a

A majority of R platform figures have descriptions of the code in them.

sexual predator as conceited. The cost of individualism, which can be likened to customization is extremely high and this makes collectivism a more viable but not a better option. In my case, the main problem of central planning is the divisiveness that exist within the central planning institution itself and so I advocate for united central economic planning.

Statistical Measure

I based my statistical method on his statement paraphrased as such "There is empirical evidence to show that countries that experienced American foreign military intervention tend to revert to their initial state when the intervention ends". A good way to measure the impact of foreign intervention on the well-being of the country being helped (or the country giving help) will be to run a simple bivariate regression model of "USAID" as the x-variable versus "the number of people below the poverty level" or "GDP growth" in the country receiving or giving help as the y-variable (Bailey 45). It is interesting to realize that foreign aid does not only affect the people receiving aid but it also affects those giving away aid.

Variables	**Description**
Constant dollars in aid[4] (it can be country specific)	US dollars spent in aid (USAID)
GDP growth[5] (it can be country specific)	Growth of Gross domestic product for foreign or host country
People below poverty level in USA[6]	People below the poverty level as determined by the US Census Bureau
People below poverty level in host country (alternative measure)	People below the poverty level as determined by the US Census Bureau
ppp= people below the poverty level ***Gdpg = GDP growth*** ***Usaid= aid provided to foreign country by USA***	

[4] https://explorer.usaid.gov/aid-trends.html

[5] http://databank.worldbank.org/data/reports.aspx?Code=NY.GDP.MKTP.KD.ZG&id=af3ce82b&report_name=Popular_indicators&populartype=series&ispopular=y

[6] http://www.census.gov/library/publications/2016/demo/p60-256.html

A majority of R platform figures have descriptions of the code in them.

APPENDIX B. ESTIMATES OF POVERTY

How Poverty Is Calculated

Following the Office of Management and Budget's (OMB) Statistical Policy Directive 14, the U.S. Census Bureau uses a set of dollar value thresholds that vary by family size and composition to determine who is in poverty (see the matrix below).

Poverty Thresholds for 2013 by Size of Family and Number of Related Children Under 18 Years

(Dollars)

Size of family unit	Related children under 18 years								
	None	One	Two	Three	Four	Five	Six	Seven	Eight or more
One person (unrelated individual):									
Under age 65	12,119								
Aged 65 and older	11,173								
Two people:									
Householder under age 65	15,600	16,057							
Householder aged 65 and older	14,081	15,996							
Three people	18,222	18,751	18,769						
Four people	24,028	24,421	23,624	23,707					
Five people	28,977	29,398	28,498	27,801	27,376				
Six people	33,329	33,461	32,771	32,110	31,128	30,545			
Seven people	38,349	38,588	37,763	37,187	36,115	34,865	33,493		
Eight people	42,890	43,269	42,490	41,807	40,839	39,610	38,331	38,006	
Nine people or more	51,594	51,844	51,154	50,575	49,625	48,317	47,134	46,842	45,037

Source: U.S. Census Bureau.

Weighted Average Poverty Thresholds in 2013 by Size of Family

(Dollars)

One person	11,888
Two people	15,142
Three people	18,552
Four people	23,834
Five people	28,265
Six people	31,925
Seven people	36,384
Eight people	40,484
Nine people or more	48,065

Source: U.S. Census Bureau.

Figure 1: Source - US Census Bureau.

Two Models

$$Gdpg_i = \beta_0 + \beta_1 Usaid_i + \epsilon_i \quad (1)$$

$Gdpg_i =$ US GDP growth rate , $Usaid_i = USA\ aid\ to\ foreign\ countries$ $\beta_0 = the\ y - intercept,$ $\beta_1 = slope\ of\ the\ equation, \epsilon_i = error\ term$

$$ppp_i = \beta_0 + \beta_1 Usaid + \epsilon_i \quad (2)$$

$ppp_i = percentage\ of\ people\ below\ the\ poverty\ level\ in\ USA, Usaid_i = USA\ aid\ to\ foreign\ countries$

$\beta_0 = the\ y - intercept, \beta_1 = slope\ of\ the\ equation, \epsilon_i = error\ term$

A majority of R platform figures have descriptions of the code in them.

Looking at the relationship between foreign aid dollars and the level of poverty in the domestic or host country can tell us whether or not “intentions equal results”. Also, looking at the relationship between “USAID” and the GDP of the foreign country or the USA is a good way to gauge the effect of foreign intervention. It would be intuitive to think that high levels of financial foreign aid by the USA can lead to higher levels of poverty in the USA but one must also remember that foreign aid also comes with opportunities that can boost the economy of the giving country. One of these opportunities include boosting development in the host country and by so doing improving home businesses that sell to the host country (this can also lead to an increase in employment for the giving country).

The proposal above can shed some light on Dr. Coyne’s work specifically pertaining to the subject of intentions versus results. This means that foreign aid can be a good intention but the results can be catastrophic or good (depending on the situation) to the country giving it or to the country receiving it. Similarly, foreign aid can be given to collective groups as a good gesture but varying individual needs can make it futile and push more people below the poverty line. In summary, we must be skeptical about foreign intervention according to Dr. Coyne.

The Borgen Project is an NGO that specializes in lobbying for bills that promote poverty alleviation worldwide. The Borgen Project advocates for USA foreign intervention (diplomatic) in countries that suffer from poverty and poverty related issues like lack of food, absence of electricity and other issues. Pundits say that diplomatic foreign intervention by sending aid or finances to developing countries is usually hindered by corruption and diversion of funds by high government officials in those countries and the results end up being sub-par to the intentions of those sending foreign aid. In the eyes of Americans and those who carry out foreign intervention, “poor results” as those “exposed” by foreign intervention pundits means that foreign intervention is conceited or self-important because “it claims or thinks to be what it is” but it is actually “not

A majority of R platform figures have descriptions of the code in them.

what it is". Actual versus intended results is the main determinant in knowing if foreign intervention is conceited. When the actual results do not live up to the intended results, then the mass thinks it is conceited (diplomatic or military). Several years ago, the USA led by President George W. Bush intervened in a Middle East dictatorship regime by Saddam Hussein all in the name of preemptive war and with the intention of stabilizing Iraq and leading them to a diplomatic regime but today the effects of that war show that things haven't really gotten better as intended by the USA some years ago. Circumstances like the story of Iraq & the USA and several other cases in the world have made people feel that foreign intervention is conceited and they cannot be wrong for thinking that way. However, we must realize that some individuals in the collective group have benefited from such intervention and some radical terrorists like Saddam Hussein and Osama Bin Laden have been eliminated and the threat of terror is not as imminent as it used to be. The Borgen Project mentions that foreign aid provides economic growth for the USA, improves USA national security and it solidifies the USA as a global leader and peacemaker (hence not so conceited). The example the USA has shown around the world as a leader in democracy and diplomacy might not make sense in the eyes of several people today (seem conceited) but as we approach the future I strongly think that the dots will be connected and the world will follow in the U.S. way of life.

Conclusion

This paper postulates (as seen in my criticism) that while central planning has its flaws, it shouldn't be rejected by society because it also has advantages. Running the statistical model proposed above can be useful in determining whether foreign intervention is conceited or not. In my opinion, foreign intervention is not conceited or self-important because it has its merits and several people can benefit from a helping hand.

A majority of R platform figures have descriptions of the code in them.

Works Cited

Bailey, Michael A. *Real Econometrics: The Right Tools to Answer Important Questions*. First ed. New York: Oxford UP, 2016. Print.

Coyne, Christopher J., and Rachel L. Mathers. "The Fatal Conceit of Foreign Intervention." *What Is so Austrian about Austrian Economics? Advances in Austrian Economics* (2010): 225-50. Web. 31 Oct. 2016.

Hayek, F.A. “The Fatal Conceit: The Errors of Socialism.” Chicago: The University of Chicago Press. N.p., 1988.

Hayek, Friedrich A., "The Use of Knowledge in Society." 1945. *Library of Economics and Liberty*. 519-30. Web. 31 October 2016. <http://www.econlib.org/library/Essays/hykKnw1.html>.

A majority of R platform figures have descriptions of the code in them.

Unemployment Duration and the last puzzle

Introduction

The continuous rise in the duration of unemployment despite the irregular economic recovery since mid-2009 has been a major point of concern for economists (Valletta & Kuang N.p).

This paper will develop a multivariate regression model to measure the effect of specific factors (unemployment rate percent change, unemployment population demographics, inflation, labor participation rate and GDP) on unemployment duration (UD) (Bailey 128). My thoughts and questions are found in the methodology section of this paper.

Arranz & Serrano in their analysis of unemployment in the Spanish labor market discovered that there was an increase in the unemployment rate from 8% to 25% in the period 2007 to 2012; while half of this rate corresponded to long-term unemployment, worker turnover was an important factor in the Spanish labor market (272). Arranz & Serrano mentioned the importance of recurrent employment in the Spanish labor market because of the huge volume of short-term contracts signed in that labor market (272). Seasonal and recurrent factors that affect the level of unemployment can have a significant effect on the duration of unemployment. There are several factors that can affect unemployment duration (UD) as listed in the executive summary above. Unemployment duration is a good measure of job destruction versus job creation and it also provides a good measure of labor market performance. Because the labor market is a significant factor in the performance of an economy, it is important to study this using econometric techniques in order to provide evidence-based conclusions. Solving the puzzle of understanding the reason behind the trend in UD despite the gradual recovery of the USA economy since the 2008 recession is my main concern and I will try to develop an econometric model that can explain this.

A majority of R platform figures have descriptions of the code in them.

Data

The six independent variables factored-in at the same time can provide good insight that we will use to explain unemployment duration. I have attempted to take away so many terms from the error term in order to provide a better explanation for the UD trend.

Unemployment duration - pud (1949 – 2015): This measures the length of time unemployed people stay unemployed. Unemployment duration is measured in weeks of unemployment. The data was derived from the Federal Reserve Economic Data (FRED) St. Louis as the secondary source but the original source of this data was the Bureau of Labor Statistics (BLS). The paper used this data set as the dependent variable because it was the variable we intended to understand by studying the effect of six independent variables.

Civilian unemployment rate – pu3c (1949-2015): This is the ratio of the number of unemployed people to the total labor force. The secondary source of this data was FRED St. Louis and the primary source was the BLS. My aim of choosing this data was to find out how the civilian unemployment rate could affect the duration of unemployment.

Black or African-American unemployment rate - bur (1973-2015): This is the ratio of the number of Blacks or African-Americans unemployed to the total labor force. The secondary source of this data was FRED St. Louis and the primary source was the BLS. My aim of choosing this data was to find out how the Black or African-American unemployment rate could affect the duration of unemployment. The effect of race on the duration of unemployment is a plausible theory.

White unemployment rate - wur (1955-2015): This is the ratio of the number of Whites unemployed to the total labor force. The secondary source of this data was FRED St. Louis and the primary source was the BLS. My aim of choosing this data was to find out how the White unemployment rate could affect the duration of unemployment. The effect of race on the duration

A majority of R platform figures have descriptions of the code in them.

of unemployment is a plausible theory. Intuitively, our society views Whites as the racial majority and this can have an effect on the length of time they stay unemployed. Logically, one would expect the dominant race to spend less time in the unemployment zone while the minority race will have a longer UD. It is interesting to understand the effect of unemployment population demographics on unemployment duration (UD).

Gross Domestic Product – pgdp (1949-2015): The real GDP is a measure of the value given to all the goods and services produced by a country. Real GDP is a good measure of a country's economic and productivity strength. The data was collected from FRED St. Louis. GDP is a good indicator of the number of firms and the level of production occurring in a country, which can also have an impact on the level of job creation and destruction (job creation or destruction has an impact on UD).

Labor Participation Rate – plpr (1949-2015): This is the ratio of the participating labor force to the total adult population eligible to work. The data was collected from FRED St. Louis as the secondary source and the BLS as the primary source. The number of Americans participating in the labor force has an effect on the number of jobs left for the unemployed (spill-over effect) and this in-turn has an effect on unemployment duration (UD).

Consumer Price Index – pcpi (1949-2015): The consumer price index is a measure of the price paid by urban consumers for a basket of goods and services. The data was collected from FRED St. Louis. Inflation is a good measure of the level of prices in the economy and it can be an incentive for people to stay employed or unemployed and so a good independent variable.

A majority of R platform figures have descriptions of the code in them.

Table 1: Variables for Unemployment Duration Data.

Date	*Variables*
pud	*Annual percent change of unemployment duration*
pu3c	*Annual percent change of civilian unemployment rate*
bur	*Unemployment rate - Black or African-American (annual percent change)*
wur	*Unemployment rate – White (annual percent change)*
pgdp	*Annual percent change in real GDP*
plpr	*Annual percent change in labor participation rate*
pcpi	*Annual percent change in inflation*

Econometric Methodology

My paper will provide a multi-variate regression model to explain the effect of six variables on the duration of unemployment (Bailey 128). The model ran is below:

$$pud_i = \beta_0 + \beta_1 pu3c_i + \beta_2 bur_i + \beta_3 wur_i + \beta_4 pgdp_i + \beta_5 plpr_i + \beta_6 pcpi_i + \epsilon_i \qquad (1)$$

Where $\beta_0 = y - intercept,\ \beta_1 = coefficeint\ for\ pu3c_i,\ \beta_2 = coefficient\ for\ bur_i,\ \beta_3 = coefficeint\ for\ wur_i,\ \beta_4 = coefficient\ for\ pgdp_i, \beta_5 = coefficient\ for\ plpr_i, \beta_6 = coefficeint\ for\ pcpi_i\ \epsilon_i = error\ term$

Null Hypothesis: An increase in pu3c, bur, wur, plpr and pcpi will lead to an increase in pud.

$H_0: \beta_1, \beta_2, \beta_3, \beta_5, \beta_6 = +ve$

Null Hypothesis: An increase in pgdp will lead to a decrease in pud. $H_0: \beta_4 = -ve$

Questions

Is unemployment duration (pud) affected by pu3c, bur, wur, pgdp, plpr, and pcpi?

Is unemployment duration longer for the Black or African American population?

Is unemployment duration longer for the White population?

A majority of R platform figures have descriptions of the code in them.

Does the unemployment rate for Whites or Black/African-American have an effect on the duration of unemployment?

What is the relationship between inflation-consumer price index and unemployment duration (UD)?

What is the relationship between labor participation rate and UD?

What is the relationship between real GDP and UD?

What is the relationship between unemployment rate and UD?

An Ignored Variable-Unemployment benefits duration

The 2009-2011 authorized extension of unemployment insurance (UI) benefits from 26 weeks to a maximum of 99 weeks can be a working theory in understanding unemployment duration (Valletta & Kuang N.p). Estimates from other researchers propose that the impact of prolonging the period of UI benefits has been modest (Valletta & Kuang N.p). Unfortunately, my paper will not be able to answer this question with my econometric model because I do not have the skills best suited to answer this question and because the depth of information needed to produce this econometric model is beyond my reach. A question one can ask here: Is there a significant relationship between length of time for UI benefits and unemployment duration?

R Platform Methodology

This paper makes use of the R platform. The steps involved include:

- Reading the data in the R-platform
- Running a multi-variate model regression
- Running a robust regression to test for Heteroscedastic errors (HC1-type errors using the “AER” package)

A majority of R platform figures have descriptions of the code in them.

Disclaimer: The data presented in Table 1. indicates that all the data presented is a percent change and not the raw data; it is important to reiterate this point (every value in the table shows an increase or a decrease in the variable). The questions above talk about the percent change in the variables mentioned and not in the raw values. The calculation of percent change is provided below:

$$percent\ change = \left(\frac{P_1 - P_0}{P_0}\right) * 100 \tag{2}$$

Where $P_1 = present\ value$, and $P_0 = previous\ value$

```
Untitled - R Editor
> ols = lm(ud$pud ~ ud$pu3c + ud$bur + ud$wur + ud$pgdp + ud$plpr + ud$pcpi)
## We run a multi-variate regression model with 6 independent variables
> summary(ols)
## We summarize the ols model
Call:
lm(formula = ud$pud ~ ud$pu3c + ud$bur + ud$wur + ud$pgdp + ud$plpr +
    ud$pcpi)

Residuals:
    Min      1Q  Median      3Q     Max
-13.698  -5.697  -1.418   4.418  16.446

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -9.9584     4.8155  -2.068  0.04588 *
ud$pu3c       0.5600     1.3877   0.404  0.68894
ud$bur        1.0197     0.4445   2.294  0.02773 *
ud$wur       -0.5006     1.1227  -0.446  0.65836
ud$pgdp       3.8503     1.2847   2.997  0.00491 **
ud$plpr     -15.5537     3.3667  -4.620 4.77e-05 ***
ud$pcpi       0.7649     0.5750   1.330  0.19175
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.044 on 36 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.7744,    Adjusted R-squared:  0.7368
F-statistic: 20.59 on 6 and 36 DF,  p-value: 2.696e-10
```

Figure 1: Regression model ran on R platform.

```
> coeftest(ols, vcov=vcovHC (ols, type="HC1"))
## Robust regression looking for heteroscedastic errors
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  -9.95840    4.42577 -2.2501  0.030649 *
ud$pu3c       0.56000    1.48924  0.3760  0.709102
ud$bur        1.01970    0.48582  2.0989  0.042899 *
ud$wur       -0.50060    1.19211 -0.4199  0.677034
ud$pgdp       3.85030    1.19044  3.2344  0.002613 **
ud$plpr     -15.55367    3.07635 -5.0559 1.266e-05 ***
ud$pcpi       0.76495    0.48302  1.5837  0.122015
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: Robust regression checking for heteroscedastic errors.

A majority of R platform figures have descriptions of the code in them.

Results and Discussion

Table 2: Table of Results for Multi-Variate Model

	Multi-variate Regression Results (six independent variables)- Figure 2
Pu3c	0.56 (1.49) [t = 0.38]
Bur	1.02* (0.49) [t = 2.09]
Wur	-0.50 (1.19) [t = -0.42]
Pgdp	3.85* (1.19) [t = 3.23]
Plpr	-15.55* (3.08) [t = -5.06]
Pcpi	0.76 (0.48) [t = 1.58]
Intercept	-9.96* (4.43) [t = -2.25]
N	37
R^2	0.77

Standard errors in parentheses.

(*) indicates significance at p<0.05, two tailed.

<u>Description of results</u>

A one-unit increase in pu3c (unemployment rate) will lead to a 0.56 increase in pud (a p-value of 0.71>0.05 hence insignificant). Intuitively, this was my thinking because as more people become unemployed, it becomes even more difficult to get a job and this leads to higher unemployment duration (we do not have enough evidence to reject the null hypothesis b/c p-value>0.05).

A majority of R platform figures have descriptions of the code in them.

A one-unit increase in bur (black or African-American (AA) unemployment rate) will lead to 1.02 increase in pud (a p-value of 0.043<0.05 hence significant). We have enough evidence to reject the null hypothesis b/c p-value<0.05. Intuitively, this was not my thinking because as more black people become unemployed, it becomes even more difficult to get a job and this leads to higher unemployment duration but my results show that my intuition is not right.

A one-unit increase in wur (white unemployment rate) will lead to 0.50 decrease in pud (a p-value of 0.68>0.05 hence insignificant). We do not have enough evidence to reject the null hypothesis even though the sign of the coefficient is different from what was hypothesized. Intuitively, this was not my thinking because as more white people become unemployed, it becomes even more difficult to get a job and this leads to higher unemployment duration but my results show that my intuition is not right. I postulated that whites were better suited to get jobs than blacks or AAs, which does not make the result very surprising.

A one-unit increase in pgdp (real GDP percent change) will lead to 3.85 increase in pud (a p-value of 0.003<0.05 hence significant). Intuitively, this was not my thinking because I expected positive GDP change to signify a better economy with more jobs and this should lead to lower unemployment duration (we have enough evidence to reject the null hypothesis because we have enough evidence). Some concerns here lie in the fact that much of USA domestic goods are produced overseas because jobs are outsourced (just one possible reason).

A one-unit increase in plpr (labor participation rate) will lead to 15.55 decrease in pud (a p-value of 1.27 e-05<0.05 hence significant). Intuitively, this was not my thinking because I expected +ve labor participation rate change to mean less jobs for the unemployed and hence higher unemployment duration (we reject the null hypothesis because we have enough evidence). In reflection, the results are not so surprising because one can expect that if more people are participating in the labor force then less people are unemployed and unemployment duration is going to decline.

A one-unit increase in pcpi (consumer price index) will lead to 0.76 increase in pud (a p-value of 0.12>0.05 hence insignificant). Intuitively, this was my thinking because I expected an increase in consumer price index to mean a demand for higher wages by employees and possibly an increase in unemployment because employers will be unwilling to pay very high wages. High levels of unemployment usually lead to longer unemployment duration as proven by my model (we accept the null hypothesis because we do not have enough evidence to reject the null hypothesis).

Conclusion

The methodology used was a multi-variate regression model that tried to explain the effect of six variables on unemployment duration as seen in my model above. The Initial reaction towards this model was that too many variables were included in the model and there might be some complicated results that couldn't be explained. The model worked (the robust regression did not show any HC1-type errors) and some actual results were in-line with the hypothesis proposed in the econometric methodology section. The model explained 77% of the trend in UD, which is well above half. Moreover, the p-value of the model 2.69e-10, is statistically significant meaning that we have enough evidence to reject the null hypothesis. The model also worked because it provided the answers to all the questions that were asked in this project. The results were clear and provided some clues as to where to look for more information that can be used to understand unemployment duration.

A majority of R platform figures have descriptions of the code in them.

Works Cited

Arranz, José M., and Carlos García-Serrano. "Duration and Recurrence of Unemployment Benefits." *Journal of Labor Research* 35.3 (2014): 271-95. Web. 13 Nov. 2016.

Bailey, Michael A. *Real Econometrics: The Right Tools to Answer Important Questions*. First ed. New York: Oxford UP, 2016. Print.

“Federal Reserve Economic Data St. Louis.” N.p., n.d. 16 Nov. 2016.

Pérez, Esther, and Yao Yao. "Can Institutional Reform Reduce Job Destruction and Unemployment Duration? Yes, It Can." *Empir Econ Empirical Economics* 49.3 (2015): 961-83. Web. 13 Nov. 2016.

Valletta, Rob, and Katherine Kuang. "FRBSF ECONOMIC LETTER." *Federal Reserve Bank of San Francisco*. N.p., 30 Jan. 2012. Web. 16 Nov. 2016.

A majority of R platform figures have descriptions of the code in them.

Appendix

Data for Unemployment Duration

(All data is in percent change)

date	pud	pu3c	bur	wur	pgdp	plpr	aegpi	pcpi
1/1/1949	16.3	60.5			-0.5	0.3	-6.3	-1
1/1/1950	20	-14.8			8.7	0.3	5.4	1.1
1/1/1951	-19.2	-36.5			8.1	0.2	7.8	7.9
1/1/1952	-13.4	-9.1			4.1	-0.3	1.2	2.3
1/1/1953	-4.8	-3.3			4.7	-0.3	4.2	0.8
1/1/1954	47.5	93.1			-0.6	-0.2	-6.2	0.4
1/1/1955	10.2	-21.4		-23.53	7.1	0.7	3.9	-0.3
1/1/1956	-13.1	-6.8		-7.69	2.1	1.4	2.9	1.5
1/1/1957	-7.1	4.9		8.33	2.1	-0.7	-0.6	3.4
1/1/1958	33.3	58.1		56.41	-0.7	-0.2	-6.9	2.7
1/1/1959	2.9	-19.1		-21.31	6.9	-0.3	4.6	0.9
1/1/1960	-11.1	0		4.17	2.6	0.2	0.1	1.5
1/1/1961	22.7	21.8		20	2.6	-0.2	-2.8	1.1
1/1/1962	-6.4	-16.4		-18.33	6.1	-0.8	3	1.2
1/1/1963	-4.8	0		2.04	4.4	-0.2	1	1.3
1/1/1964	-5	-7.1		-8	5.8	0	1.8	1.3
1/1/1965	-11.3	-13.5		-10.87	6.5	0.2	4.4	1.6
1/1/1966	-12.7	-15.6		-19.51	6.6	0.7	5.6	3
1/1/1967	-14.6	0		3.03	2.7	0.7	0.6	2.8
1/1/1968	-4.5	-5.3		-5.88	4.9	0	1.9	4.2
1/1/1969	-6	-2.8		-3.13	3.1	0.8	2.7	5.4
1/1/1970	10.1	42.9		48.39	0.2	0.5	-3.1	5.9
1/1/1971	31	20		19.57	3.3	-0.3	-2.6	4.2
1/1/1972	5.3	-6.7		-7.27	5.3	0.3	3.2	3.3
1/1/1973	-16.7	-12.5	-9.62	-15.69	5.6	0.7	5.2	6.3
1/1/1974	-3	14.3	11.7	18.6	-0.5	0.8	-0.4	11
1/1/1975	47.4	51.8	40.95	52.94	-0.2	-0.2	-8.7	9.1
1/1/1976	10.5	-9.4	-5.41	-10.26	5.4	0.7	3.3	5.8
1/1/1977	-9.5	-7.8	0	-11.43	4.6	1	4.3	6.5
1/1/1978	-16.8	-14.1	-9.29	-16.13	5.6	1.6	5.1	7.6
1/1/1979	-9.2	-3.3	-3.15	-1.92	3.2	0.8	3.5	11.3
1/1/1980	10.2	22	16.26	23.53	-0.2	0.2	-2.9	13.5
1/1/1981	16	5.6	9.09	6.35	2.6	0.2	-0.6	10.4
1/1/1982	13	27.6	21.15	28.36	-1.9	0.2	-6.5	6.2
1/1/1983	27.6	-1	3.17	-2.33	4.6	0	-2	3.2
1/1/1984	-9	-21.9	-18.46	-22.62	7.3	0.6	6	4.4
1/1/1985	-13.8	-4	-5.03	-4.62	4.2	0.6	0.6	3.5
1/1/1986	-3.8	-2.8	-3.31	-3.23	3.5	0.6	-1.1	1.9

A majority of R platform figures have descriptions of the code in them.

1/1/1987	-3.3	-11.4	-10.96	-11.67	3.5	0.6	0.7	3.6
1/1/1988	-6.9	-11.3	-10	-11.32	4.2	0.5	1.9	4.1
1/1/1989	-11.9	-3.6	-1.71	-4.26	3.7	0.8	0.6	4.8
1/1/1990	0.8	5.7	-0.87	6.67	1.9	0.2	-1.3	5.4
1/1/1991	14.2	23.2	9.65	27.08	-0.1	-0.5	-4.8	4.2
1/1/1992	29.9	8.7	13.6	8.2	3.6	0.3	-2.2	3
1/1/1993	1.1	-8	-8.45	-7.58	2.7	-0.2	0.6	3
1/1/1994	4.4	-11.6	-11.54	-13.11	4	0.5	2.5	2.6
1/1/1995	-11.7	-8.2	-9.57	-7.55	2.7	0	1.7	2.8
1/1/1996	0.6	-3.6	0.96	-4.08	3.8	0.3	1.1	2.9
1/1/1997	-5.4	-9.3	-3.81	-10.64	4.5	0.4	2	2.3
1/1/1998	-8.2	-8.2	-11.88	-7.14	4.4	0	2	1.5
1/1/1999	-7.6	-6.7	-10.11	-5.13	4.7	0	0.5	2.2
1/1/2000	-5.2	-4.8	-5	-5.41	4.1	0	0.8	3.4
1/1/2001	3.1	17.5	14.47	20	1	-0.4	-3.2	2.8
1/1/2002	27.5	23.4	17.24	21.43	1.8	-0.3	-5.5	1.6
1/1/2003	15	3.4	5.88	3.92	2.8	-0.6	-3.3	2.3
1/1/2004	2.1	-8.3	-3.7	-9.43	3.8	-0.3	0.3	2.7
1/1/2005	-6.1	-7.3	-3.85	-8.33	3.3	0	1.4	3.4
1/1/2006	-8.7	-9.8	-10	-9.09	2.7	0.3	1.6	3.2
1/1/2007	0.6	0	-7.78	2.5	1.8	-0.3	-1.3	2.9
1/1/2008	5.3	26.1	21.69	26.83	-0.3	0	-4	3.8
1/1/2009	36.5	60.3	46.53	63.46	-2.8	-0.9	-13	-0.3
1/1/2010	36.2	3.2	8.11	2.35	2.5	-1.1	-4.3	1.6
1/1/2011	19	-7.3	-1.25	-9.2	1.6	-0.9	1.7	3.1
1/1/2012	0	-9	-12.66	-8.86	2.2	-0.6	2.1	2.1
1/1/2013	-7.1	-8.6	-5.07	-9.72	1.7	-0.6	1.7	1.5
1/1/2014	-7.9	-16.2	-13.74	-18.46	2.4	-0.6	2.6	1.6
1/1/2015	-13.6	-14.5	-15.93	-13.21	2.6	-0.5	1.9	0.1

A majority of R platform figures have descriptions of the code in them.