10/10 Questions Answered
2 questions with unsaved changes

IE 522 HW09

Q1
2 Points

For time series data, we often do a differencing and analyze changes (differences) instead of the original variables. This problem shows why. Let

$$X_i = 10 + t_i + \epsilon_i^X, \quad Y_{1i} = 20 + 2t_i + \epsilon_i^{Y_1}, \quad Y_{2i} = 20 + 2t_i + 5X_i + \epsilon_i^{Y_2},$$

where the error terms are i.i.d. with a $N(0,1)$ distribution. The following generates 100 observations for X_i, Y_{1i} and Y_{2i}. Differencing leads to 99 observations for $\Delta X_i := X_{i+1} - X_i, \Delta Y_{1i} = Y_{1,i+1} - Y_{1i}, \Delta Y_{2i} = Y_{2,i+1} - Y_{2i}$. Assume $t_{i+1} - t_i = 1$.

Choose Files No file chosen

```
set.seed(1)
t=seq(1,100,by=1) #time
X  =10+t+rnorm(100)
Y1 =20+2*t+rnorm(100)
Y2 =20+2*t+5*X+rnorm(100)
data = data.frame(X,Y1,Y2)
ddata=data.frame(apply(data,2,diff))
```

Q1.1
1 Point

When Y_1 and Y_2 are regressed on X, respectively, what R^2's do you get? Other than a common time trend, are Y_1 and X related?

Call:
lm(formula = data$Y1 ~ data$X)
Multiple R-squared: 0.9988,	Adjusted R-squared: 0.9988

Call:
lm(formula = data$Y2 ~ data$X)
Multiple R-squared: 0.9999,	Adjusted R-squared: 0.9999

R_square for Y1~X = 0.9988
R_square for Y2~X = 0.9999
Without common time trend, X is 10+rnorm, and Y1 is 20+rnorm, which is independent (rnorm and rnorm is IID for each other).

CURRENTLY UPLOADED FILES

Q1-1.R

Download ✖ Remove

Choose Files No file chosen

```
#Q1-1

#Data Preparation
set.seed(1)
t=seq(1,100,by=1) #time
X  =10+t+rnorm(100)
Y1 =20+2*t+rnorm(100)
Y2 =20+2*t+5*X+rnorm(100)
data = data.frame(X,Y1,Y2)
ddata=data.frame(apply(data,2,diff))

#Regression
summary(lm(data$Y1 ~ data$X))
summary(lm(data$Y2 ~ data$X))
```

Please select file(s) Select file(s)

Save Answer *Unsaved Changes

Q1.2

1 Point

Write down the equations for $\Delta X_i, \Delta Y_{1i}, \Delta Y_{2i}$. When ΔY_1 and ΔY_2 are regressed on ΔX, respectively, what R^2's do you get? Why do the R^2's change this way?

Call:

lm(formula = ddata$Y1 ~ ddata$X)

Multiple R-squared: 0.0001495,	Adjusted R-squared: -0.01016

Call:

lm(formula = ddata$Y2 ~ ddata$X)

Multiple R-squared: 0.9465,	Adjusted R-squared: 0.946

Choose Files No file chosen

There is linear correlation between Yi and X, however not between delta_Y1 and delta_X, since the term that lefted (as the pdf shown) is iid for each other, so the R_square of Y1~X fitting is reduced dramatically.

There is linear correlation between Y2 and X, and as well between delta_Y2 and delta_X, since delta_Y2 contains the term delta_X in its equation. So the R_square of Y2~X fitting is not change dramatically.

CURRENTLY UPLOADED FILES

A9-Q1-2.pdf

Download Remove

1 / 1

Choose Files No file chosen

Q1-2.R

Download Remove

```r
#Q1-2

#Data Preparation
set.seed(1)
t=seq(1,100,by=1) #time
X  =10+t+rnorm(100)
Y1 =20+2*t+rnorm(100)
```

```
Y2 =20+2*t+5*X+rnorm(100)
data = data.frame(X,Y1,Y2)
ddata=data.frame(apply(data,2,diff))

#Regression
summary(lm(ddata$Y1 ~ ddata$X))
summary(lm(ddata$Y2 ~ ddata$X))

```

Please select file(s) Select file(s)

Save Answer *Unsaved Changes

Q2

8 Points

The dataset USMacroG in the AER package contains quarterly macroeconomic variables from 1950 to 2000. Click **HERE** for details. We want to study the relationship between consumption (dependent variable) and other variables (independent variables). In all of the regression fitting below, it's assumed that the constant regressor is included, significance level is 1% for tests, and the regression models are normally distributed.

Choose Files No file chosen

```
library(AER)
data("USMacroG")

dim(USMacroG)

USMacroG=na.omit(USMacroG)
dim(USMacroG)

Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
```

```
sum(is.na(USMacroG))
sum(is.na(Macro))
```

`USMacroG` contains the original variables before differencing, `Macro` contains the variables after differencing (that is, quarterly changes).

Q2.1

1 Point

What's the R^2 when the consumption (`USMacroG[,"consumption"]`) is regressed on the cpi (`USMacroG[,"cpi"]`) (the variables here are before differencing)? What's the R^2 when `Macro$consumption` is regressed on `Macro$cpi` (the variables here are after differencing, that is, they are quarterly changes)? What could have caused the big difference in the values of the R^2's? Does the cpi quarterly change seem to be useful in explaining the consumption quarterly change?

Call:

lm(formula = USMacroG[, "consumption"] ~ USMacroG[, "cpi"])

Multiple R-squared: 0.9593,   Adjusted R-squared: 0.9591

Choose Files No file chosen

Call:

lm(formula = Macro$consumption ~ Macro$cpi)

Multiple R-squared: 0.003662,   Adjusted R-squared: -0.001319

It is possible that by differentiating, the correlations(the Xs in Y = BX+error model) are being reduced, so that only residuel error is left, which is random and not correlated.

From the R_square achieved, we do not conclude that cpi quarterly change is useful in explaning the consumption quarterly change.

CURRENTLY UPLOADED FILES

Q2-1.R

Download Remove

```
#Q2-1

#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)

sum(is.na(USMacroG))
sum(is.na(Macro))

#Regression for USMacroG
summary(lm(USMacroG[, "consumption"] ~ USMacroG[, "cpi"]))

#Regression for Macro
summary(lm(Macro$consumption ~ Macro$cpi))


```

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 6:28 PM**

Q2.2

1 Point

From now on, we work with quarterly changes only. Consumption refers to quarterly change in consumption, gdp refers to quarterly change in gdp, etc. Construct a scatter plot for consumption and gdp. Does it show some linear relatioship between the two variables? Let consumption (dependent variable) be regressed on gdp. Give the equation for the estimated regression line, the estimate for σ and its degrees of freedom.

It shows on linear correlation between consumption and gdp, yee it is still scattered.

```
Call:
lm(formula = Macro$consumption ~ Macro$gdp)

Residuals:
    Min      1Q  Median      3Q     Max 
-56.713  -9.064  -2.733   9.491  56.189 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.06909    1.72775   6.407 1.04e-09 ***
Macro$gdp    0.39623    0.03038  13.042  < 2e-16 ***
---
S 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Choose Files No file chosen

```
Residual standard error: 18.33 on 200 degrees of freedom
Multiple R-squared:  0.4596,	Adjusted R-squared:  0.4569 
F-statistic: 170.1 on 1 and 200 DF,  p-value: < 2.2e-16
```

The estimated regression line is consumption_hat = 11.06909 + 0.39623*gdp_hat, since it is estimated regression line, we use estimated value, consumption_hat and dgp_hat.

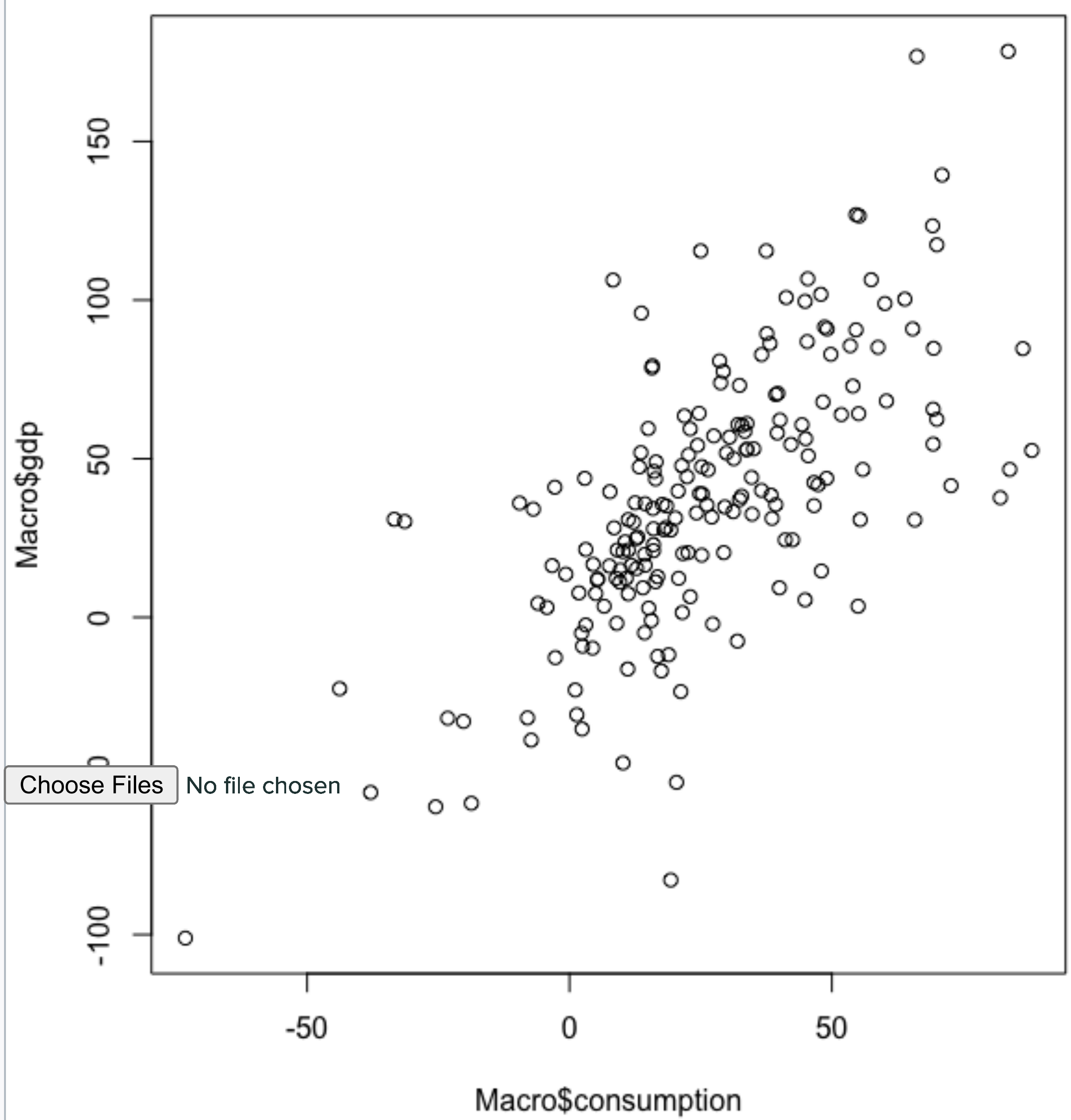
The estimation for sigma is 18.33, and the degree of freedom is 200.

CURRENTLY UPLOADED FILES

Q2-2_Scatter_Plot.png

Download Remove

Choose Files No file chosen



Choose Files No file chosen

▾ Q2-2.R Download ✖ Remove

```r
#Q2-2

#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))

#Scatter plot (consumption, gdp)
plot(Macro$consumption, Macro$gdp)

#Regression (consumption, gdp)
fit2_2 = lm(Macro$consumption ~ Macro$gdp)
summary(fit2_2)
```

Choose Files No file chosen

Select file(s)

Save Answer

Last saved on **Nov 04 at 6:28 PM**

Q2.3

1 Point

Let β_{gdp} be the coefficient for gdp and $\hat{\beta}_{gdp}$ the corresponding OLS estimator. What's the estimated standard error for $\hat{\beta}_{gdp}$? Construct a 99% confidence interval for β_{gdp}. Is there strong enough statistical evidence that $\beta_{gdp} \neq 0$?

```
Call:
lm(formula = Macro$consumption ~ Macro$gdp)

Residuals:
    Min      1Q  Median      3Q     Max
-56.713  -9.064  -2.733   9.491  56.189

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.06909    1.72775   6.407 1.04e-09 ***
Macro$gdp    0.39623    0.03038  13.042  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.33 on 200 degrees of freedom
Multiple R-squared:  0.4596,	Adjusted R-squared:  0.4569
Choose Files No file chosen 70.1 on 1 and 200 DF,  p-value: < 2.2e-16
```

The estinated standard error of Beta_hat_gdp is 0.03038.

The 99% CI of Beta_gdp is [0.317222725828501, 0.475237274171499]

Yes, since the p-value of Beta_gdp is smaller than 2e-16, which is smaller enough for H0:Beta_gdp = 0 to be rejected. Thus, Beta_gdp is unlikely to be zero.

CURRENTLY UPLOADED FILES

Q2-3.R

Download ✖ Remove

```r
#Q2-3

#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))


#Regression (consumption, gdp)
fit2_3 = lm(Macro$consumption ~ Macro$gdp)
summary(fit2_3)

#99% CI of Beta_gdp
Beta_gdp_hat = 0.39623
print(paste(Beta_gdp_hat - qt(p = 0.995, df = 200) * 0.03038 , Beta_gdp_hat + qt(p = 0.995, df = 200) * 0.03038))
```

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 6:29 PM**

Q2.4

1 Point

What's the SST, SSR, SSE, R^2 of the above simple linear regression model? What's the percentage of the variation in consumption that could be explained by gdp in this model.

```
> anova(fit2_4)
Analysis of Variance Table

Response: Macro$consumption
          Df Sum Sq Mean Sq F value    Pr(>F)
Macro$gdp   1  57139   57139   170.1 < 2.2e-16 ***
Residuals 200  67185     336
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

SSR = 57139

SSE = 67186

SST = SSR + SSE = 124134

The percentage of the variation in consumption that can be explained by gdp is 45.95975%

Choose Files No file chosen

UPLOADED FILES

Q2-4.R

Download ✖ Remove

```r
#Q2-4

#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
```

```
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))

#Regression
fit2_4 = lm(Macro$consumption ~ Macro$gdp)
summary(fit2_4)

#Defination
anova(fit2_4)
57139/(57139+67185)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 6:29 PM**

Q2.5

1 Point

Let n be the sample size of the above regression model. Regress consumption[1:(n-1)] on gdp[1:(Choose Files No file chosen consumption using gdp[n]. What's the 95% prediction interval? Compare with the actual value consumption[n]: what's the prediction error?

Y_star_hat = 28.31972

95% prediction interval is [-7.890101, 64.52954]

The true consumption[202] is 49, and the predicted consumption is 28.31972 with 95% PI [-7.890101, 64.52954].

The prediction error is 20.68028.

CURRENTLY UPLOADED FILES

Q2-5.R Download Remove

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 6:29 PM**

Q2.6

1 Point

Suppose you regress consumption on tbill, inflation and cpi. What's the null hypothesis of the F-test? Is the null hypotnesis rejected? What does it mean?

The null hypo of F-test (H0) is H0: None of regressor is useful : B_tbill = B_inflation = B_cpi = 0.

Choose Files No file chosen ---

Call:

lm(formula = consumption ~ tbill + inflation + cpi)

Residuals:

Min 1Q Median 3Q Max

-81.723 -15.179 -3.492 16.394 59.113

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 25.75793 2.61828 9.838 <2e-16 ***
tbill 4.30044 2.44569 1.758 0.0802 .
inflation 0.82269 0.68256 1.205 0.2295
cpi 0.09873 0.87997 0.112 0.9108

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 24.65 on 198 degrees of freedom
Multiple R-squared: 0.03264, Adjusted R-squared: 0.01799
F-statistic: 2.227 on 3 and 198 DF, p-value: 0.08627

Under the confidence level of 95%, since p = 0.08627 > alpha = 0.05, null hypo H0 is not rejected, means that we maintain our hypo that none of the regressor is useful.

CURRENTLY UPLOADED FILES

Q2-6.R Download ✖ Remove

```
#Q2-6
#Data Preparation
library(AER)
data("USMacroG")
[illegible](USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))

#Regression
fit2_6 = lm(Macro$consumption ~ Macro$tbill + Macro$inflation + Macro$cpi)
summary(fit2_6)

```

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 3:37 PM**

Q2.7

1 Point

When consumption is regressed on m1 only, is m1 significant (i.e., is $H_0 : \beta_{m1} = 0$ rejected)? Does adding a second order term improve the fitting? Construct a scatter plot of consumption (y axis) and m1 (x axis), add the estimated regression line resulting from the simple linear regression model, and the quadratic curve resulting from the fitting with a second order term in addition to m1 itself.

```
Call:
lm(formula = Macro$consumption ~ Macro$m1)

Residuals:
   Min     1Q Median     3Q    Max
-97.31 -13.58  -1.63  15.70  70.84
```

Choose Files No file chosen

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.3945     1.9608  11.931  < 2e-16 ***
Macro$m1      0.5526     0.1948   2.837  0.00503 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.45 on 200 degrees of freedom
Multiple R-squared:  0.03868,     Adjusted R-squared:  0.03387
```

```
F-statistic: 8.046 on 1 and 200 DF,  p-value: 0.005029
```

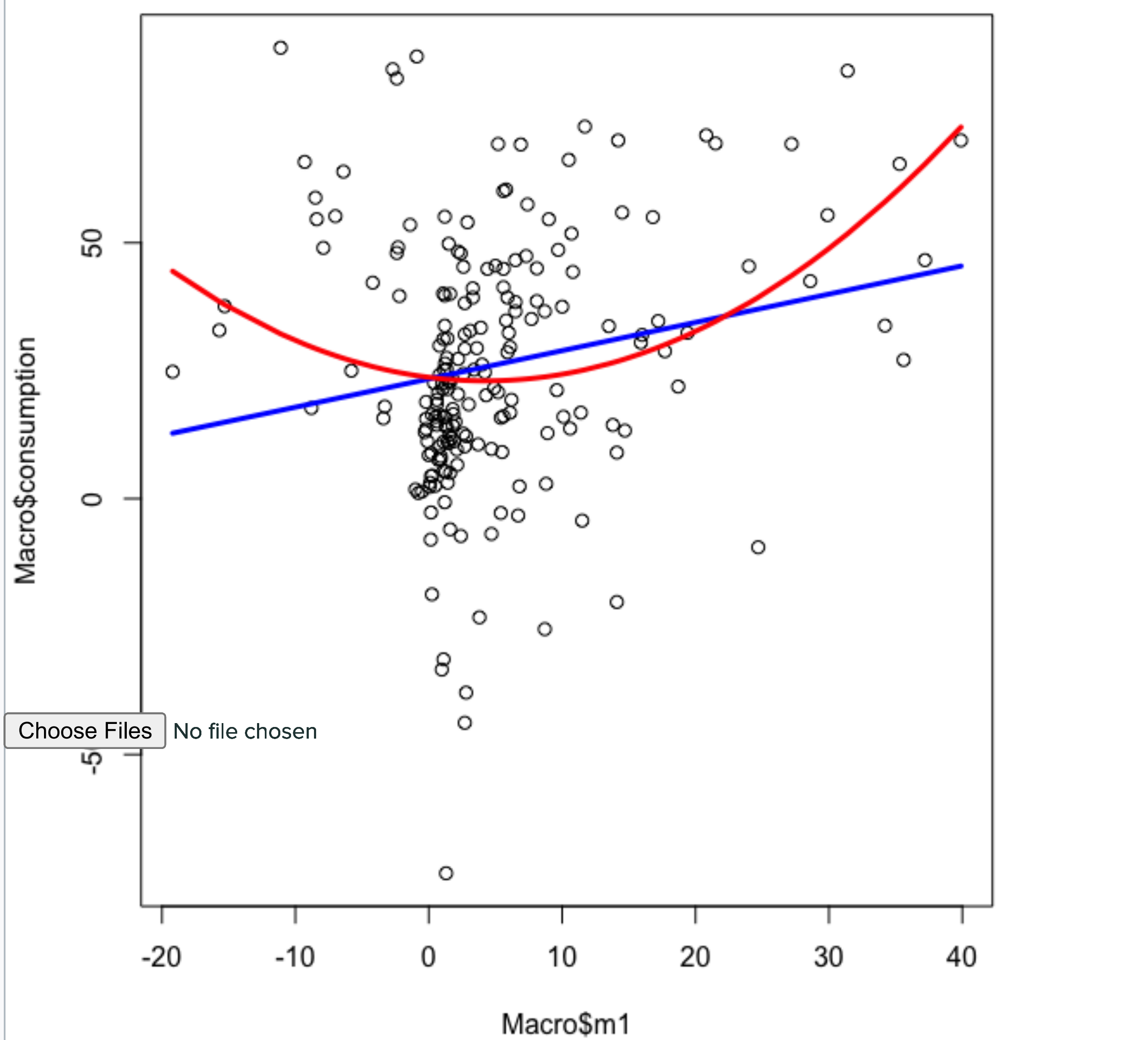
Since p = 0.00503 (simple regression so F-test = t-test), so H0 is rejected, which means that m1 is significant under 95% confidence level.

```
Call:
lm(formula = Macro$consumption ~ Macro$m1 + I(Macro$m1^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-96.549 -12.536  -2.095  15.933  62.358 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   23.71325    1.91247  12.399  < 2e-16 ***
Macro$m1      -0.33071    0.32011  -1.033 0.302804    
I(Macro$m1^2)  0.03899    0.01138   3.426 0.000743 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Choose Files  No file chosen
ndard error: 23.81 on 199 degrees of freedom
Multiple R-squared:  0.09223,	Adjusted R-squared:  0.08311 
F-statistic: 10.11 on 2 and 199 DF,  p-value: 6.586e-05
```

The 2nd_order model is better since its Adjusted R_square is better than the 1st order model.

CURRENTLY UPLOADED FILES

▾ Q2-7_Scatter_Plot_with_Lines.png Download ✖ Remove



Choose Files No file chosen

Q2-7.R

Download Remove

```
#Q2-7
#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))

#Regression
fit2_7 =lm(Macro$consumption ~ Macro$m1)
summary(fit2_7)

#Regression 2nd order
fit2_7.2 =lm(Macro$consumption ~ Macro$m1 + I(Macro$m1^2))
summary(fit2_7.2)

#Plotting
plot(Macro$m1, Macro$consumption)
lines(Macro$m1, fit2_7$fitted.values, lwd=3, col = 'blue')
lines(sort(Macro$m1), fit2_7.2$fitted.values[order(Macro$m1)], lwd = 3, col =

```

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 3:37 PM**

Q2.8

1 Point

Regress consumption on all other 11 variables. What's the adjusted R^2 now? Is m1 significant? From results in #7 and #8, what can you say about whether m1 is useful in explaining consumption?

```
Call:
lm(formula = Macro$consumption ~ ., data = Macro)

Residuals:
    Min      1Q  Median      3Q     Max
-34.287  -6.922  -1.108   6.922  38.723

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.14859    3.07949  -1.022  0.30787
gdp           0.77061    0.04636  16.624  < 2e-16 ***
invest       -0.74133    0.05781 -12.822  < 2e-16 ***
government   -0.58702    0.08414  -6.976  4.9e-11 ***
dpi           0.10500    0.03550   2.958  0.00349 **
cpi          -0.39117    0.46705  -0.838  0.40334
m1            0.14953    0.11079   1.350  0.17873
[illegible]04615  245.71289   0.061  0.95124
unemp        -3.82193    3.05721  -1.250  0.21279
population   10.22284    3.99745   2.557  0.01133 *
inflation   -16.05374  245.71204  -0.065  0.94798
interest    -16.77342  245.72689  -0.068  0.94565
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.71 on 190 degrees of freedom
Multiple R-squared:  0.7533, Adjusted R-squared:  0.739
F-statistic: 52.74 on 11 and 190 DF,  p-value: < 2.2e-16
```

Choose Files No file chosen

Multiple R-squared: 0.7533

The p-value of m1 is 0.17873, which we consider it as not significant.

With only use m1 as the regressor, m1 can be the significant However, with other regressors , m1;s significance seemed to be dimmed, use makes m1 not significant in explaining the consumption anymore.

CURRENTLY UPLOADED FILES

Q2-8.R

Download Remove

```
#Q2-8
#Data Preparation
library(AER)
data("USMacroG")
dim(USMacroG)
USMacroG=na.omit(USMacroG)
dim(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
sum(is.na(USMacroG))
sum(is.na(Macro))

#Regression
fit2_8 = lm(Macro$consumption ~., Macro)
summary(fit2_8)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 04 at 3:37 PM**

Save All Answers

Submit & View Submission