Stat 208 HW 3

Jordan Berninger

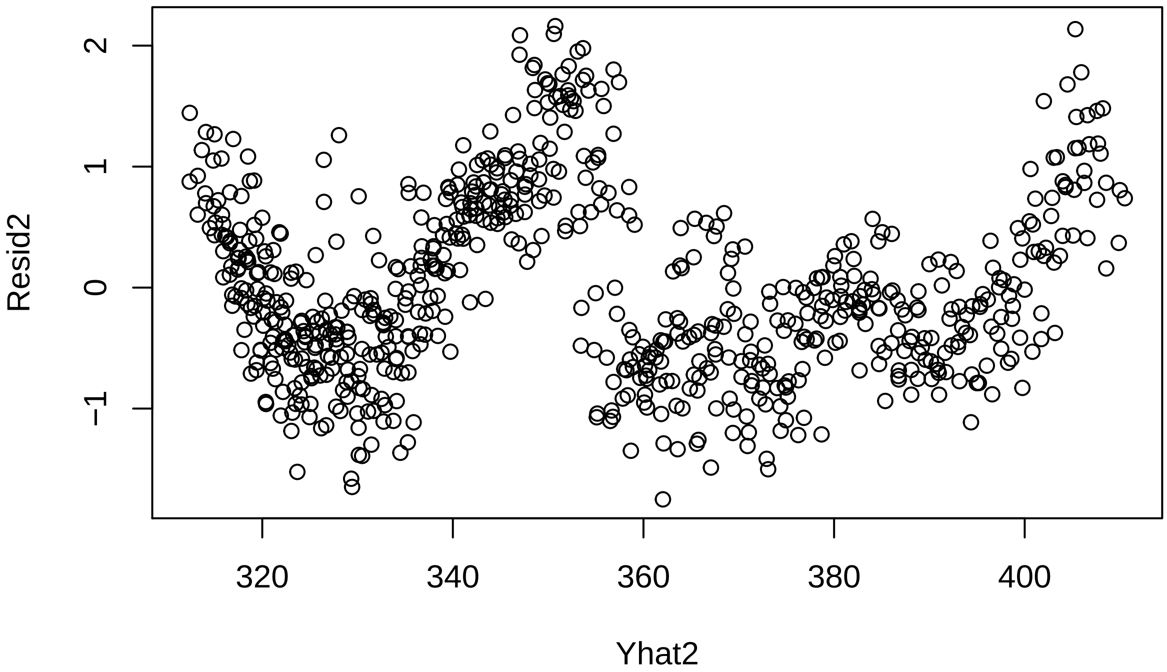
5/19/2020

Question 1

Residual plot for the Mauna Loa seasonal model with monthly term - $\vec{Y} = \mu + \alpha t + \gamma t^2 + S_t + \epsilon_t$. We plot the residuals against the fitted values.

```
d=60#2018-1959+1
T=12
N=d*T
p2=14 # updated this

# Define Observations and Times
# Initialize time (tax) and observation (Y) arrays
tax=1:N
tyr=1959+tax/12

Y=loa$v3
#Plot the data
# Seasonal effect matrix
S=matrix(0,12,11)
diag(S)=rep(1,11)
S[12,]=rep(-1,11)
# Define Design Matrix
D2=matrix(0,N,p2)
for(i in 1:N){
D2[i,1]=1
D2[i,2]=i
D2[i,3]=i^2
if(mod(i-1,12) == 0) {
  D2[i:(i+11),4:14] = S
}
}

# Compute the OLS estimators
H1.2=t(D2) %*% D2
H2.2=solve(H1.2)
H3.2= H2.2 %*% t(D2)
theta_hat2= H3.2 %*% Y

# Compute Estimated Y
Yhat2= D2 %*% theta_hat2

# Compute Residuals
Resid2=Y-Yhat2

# Compute Parameter Standard Errors
se2=matrix(0,p2,1)
SSE2=t(Resid2) %*% Resid2
```

```r
sighat2=SSE2/(N-p2)
for(i in 1:p2){
se2[i]=sighat2^(1/2)*H2.2[i,i]^(1/2)
}

R.sq2 <- 1-SSE2/sum((Y-mean(Y))^2)
plot(Yhat2,Resid2)
```



The strong sinusoidal pattern indicates that the residuals are not random.

Question 2

A very low p-value indicates that we should accept the alternative hypothesis, and conlcude that the residuals are not random.

```r
library(randtests)
turning.point.test(Resid2,alternative="two.sided")
```

```
##
##  Turning Point Test
##
## data:  Resid2
## statistic = -7.4045, n = 720, p-value = 1.317e-13
## alternative hypothesis: non randomness
```

Question 3

The sample correlations for the first 5 lags.

```r
resid.acf=acf(Resid2,lag.max=5,type="correlation",plot=F)
resid.acf
```

```
##
## Autocorrelations of series 'Resid2', by lag
##
##     0     1     2     3     4     5
## 1.000 0.908 0.858 0.810 0.775 0.755
```

From the GLM Disgnostics slide 6, we have $\hat{\rho}_{\hat{\epsilon}} \pm z_{\alpha/2}\sqrt{1/n}$. THis gives us the following 95% confidence intervals.

```r
alpha=0.05
low<-resid.acf$acf - qnorm(1-alpha)*sqrt(1/N) # lower bound for 95% confidence interval
upp<-resid.acf$acf + qnorm(1-alpha)*sqrt(1/N) # upper bound for 95% confidence interval
lag<-as.character(c(0:5))
rbind(lag,low,upp)
```

```
##     [,1]                [,2]                [,3]
## lag "0"                 "1"                 "2"
## low "0.938699924618324" "0.846472897154148" "0.79678592725484"
## upp "1.06130007538168"  "0.969073047917501" "0.919386078018192"
##     [,4]                [,5]                [,6]
## lag "3"                 "4"                 "5"
## low "0.748817525458598" "0.71355367948432"  "0.694034494414054"
## upp "0.87141767622195"  "0.836153830247673" "0.816634645177406"
```

Question 4

Portmanteau test for the first 5 lags. Following GLM Diagnostics slide 6

```r
Q=N*sum(resid.acf$acf^2)
Q
```

```
## [1] 3159.059
```

```r
dchisq(1-alpha,df=5)
```

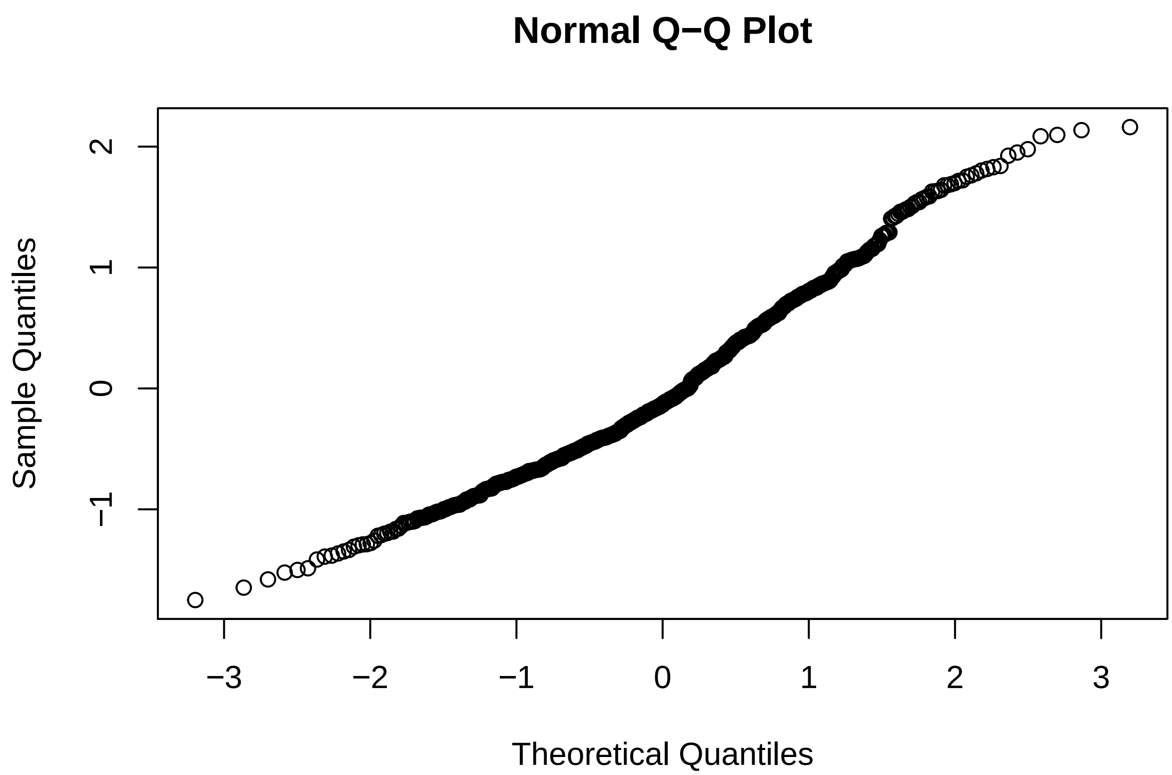
```
## [1] 0.07657453
```

This Protmanteau test decisively rejects independence of the residuals.

Question 5

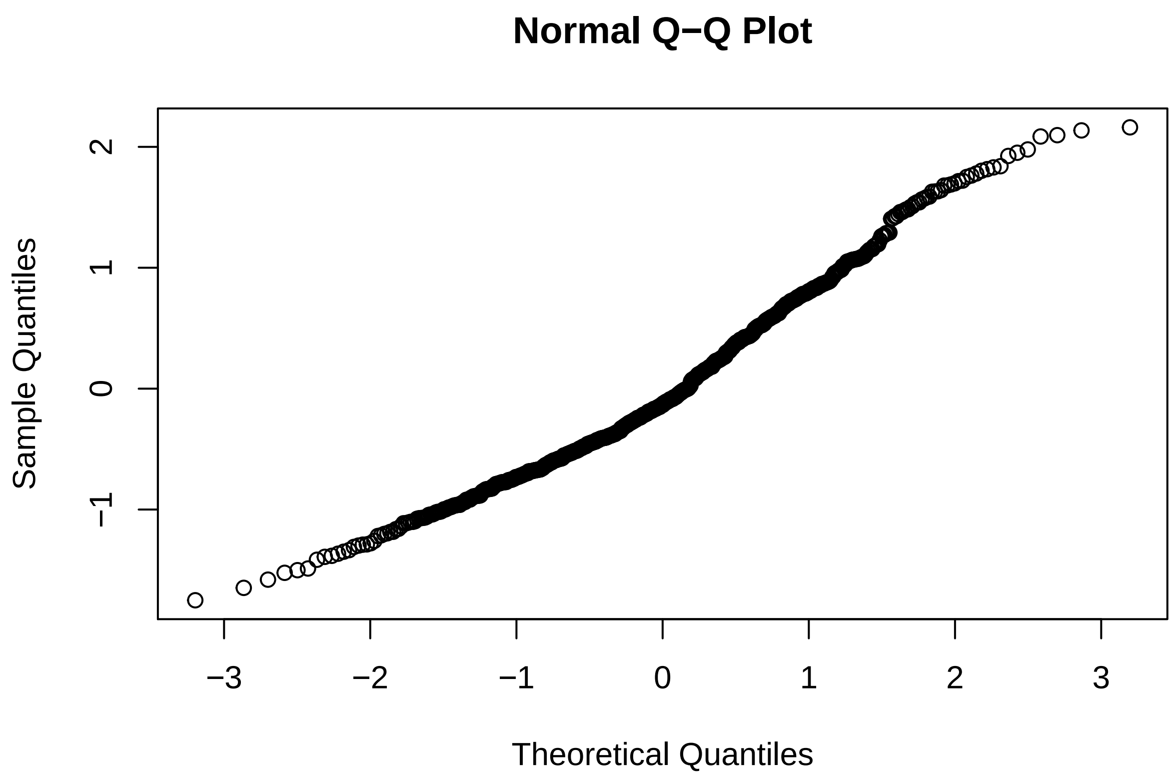
The QQ plot is easy and it shows the residuals strongly match the corresponding theoretical quantiles from a normal distribution.

```r
qqnorm(y=Resid2)
```



The sample squared correlation is extremely high.

```r
cor(Resid2,qqnorm(y=Resid2)$x)^2
```




```
##            [,1]
## [1,] 0.9775961
```

Question 6

Two iid variables A_1, A_2. We know that $P(A_1 < A_2) + P(A_2 < A_1) + P(A_1 = A_2) = 1$ by the law of total probability. We also know that $P(A_1 = A_2) = 0$, since they are assumed to follow a continuous probability distribution. Now, we know $P(A_1 < A_2) = P(A_2 < A_1)$ because of symmetry and the fact that $A_1 \perp A_2$. Thus $P(A_1 < A_2) = P(A_2 < A_1) = 1/2$.

Question 7

Mauna Loa Model 1 ($S_1, \ldots, S_{12}$ model). Compute $R^2_{4,\ldots,14|1,2,3}$. The formula for this is in the GLM Misc. Notes page 1.

```
d=60#2018-1959+1
T=12
N=d*T
p2=14 # updated this

# Define Observations and Times
# Initialize time (tax) and observation (Y) arrays
tax=1:N
tyr=1959+tax/12

Y=loa$v3
#Plot the data
```

```r
# Seasonal effect matrix
S=matrix(0,12,11)
diag(S)=rep(1,11)
S[12,]=rep(-1,11)
# Define Design Matrix
D2=matrix(0,N,p2)
for(i in 1:N){
D2[i,1]=1
D2[i,2]=i
D2[i,3]=i^2
if(mod(i-1,12) == 0) {
  D2[i:(i+11),4:14] = S
}
}

# Compute the OLS estimators
H1.2=t(D2) %*% D2
H2.2=solve(H1.2)
H3.2= H2.2 %*% t(D2)
theta_hat2= H3.2 %*% Y

# Compute Estimated Y
Yhat2= D2 %*% theta_hat2

# Compute Residuals
Resid2=Y-Yhat2


# New Design Matrices
split=4
H.1_3=D2[,1:split]
M.1_3=H.1_3%*%solve(t(H.1_3)%*%H.1_3)%*%t(H.1_3)
Y.1_3=M.1_3%*%Y
Resid.1_3=Y-Y.1_3

H.4_14=D2[,split:ncol(D2)]
M.4_14=H.4_14%*%solve(t(H.4_14)%*%H.4_14)%*%t(H.4_14)
Y.4_14=M.4_14%*%Y
Resid.4_14=Y-Y.4_14

R.sq.conditional=((t(Resid.1_3)%*%Resid.1_3)-(t(Resid2)%*%Resid2))/(t(Resid.1_3)%*%Resid.1_3)
R.sq.conditional
```

```
##           [,1]
## [1,] 0.8821734
```

Question 8

Show that $0 \leq R^2_{1,\ldots,k|k+1,\ldots,p} \leq 1$, where $k < p$.

In class, we saw that

$$
\begin{aligned}
R^2_{1,\ldots,k|k+1,\ldots,p} &= \frac{(Y-\hat{Y}_p)^2-(Y-\hat{Y}_k)^2}{(Y-\hat{Y}_p)^2} \\
&= \frac{R_k^2-R_p^2}{1-R_p^2}
\end{aligned}
$$

Now, we know that $R_k^2 < 1$, so it follows that $\frac{R_k^2-R_p^2}{1-R_p^2} < \frac{1-R_p^2}{1-R_p^2} = 1$.

Also, we know that $R_p^2 > R_k^2$, so it follows that $\frac{R_k^2-R_p^2}{1-R_p^2} < \frac{R_k^2-R_k^2}{1-R_p^2} = 0$.

Thus, we know that $0 \leq R^2_{1,\ldots,k|k+1,\ldots,p} \leq 1$.

Question 9

I wrote this one out to avoid writing matrices in LaTeX, see the screenshots that follow.

Question 10

We can write this model as $y_{m,L} = \mu_m + \tau_m * L + \epsilon_{M,L}$, where $y_{m,L}$ is the termperature recorded in month m and year L, μ_m is the monthly location parameter and τ_m is the monthly slope parameter for $m \in (1,\ldots,12)$, L is the year index, and $\epsilon_{M,L}$ is the error term (one error term for each temperature reading).

$$Y_t = \mu + \Delta \mathbb{1}_{(t>\tau)} + \varepsilon_i \qquad 1 \le t \le n \quad \ldots \quad \tau \text{ known}$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0_\tau \\ \vdots & 1 \\ \vdots & 1 \\ 1 & \vdots \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} \tau \text{ rows} \\ \\ \\ \end{matrix} \qquad B = \begin{pmatrix} \mu \\ \Delta \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\hat{B} = (X^T X)^{-1} X^T Y$$

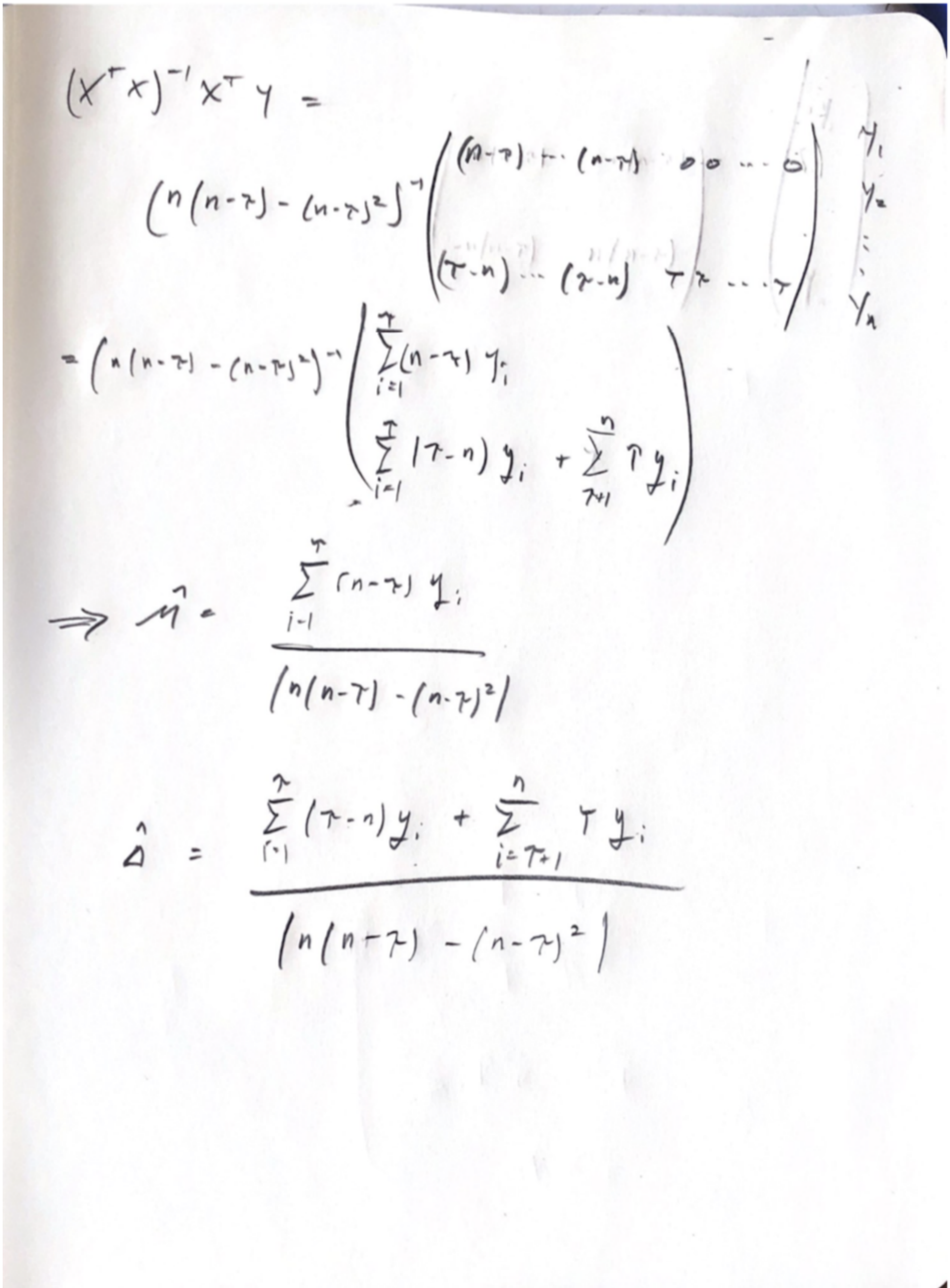
$$X^T X = \begin{pmatrix} 1 & 0 \\ 1 & \vdots \\ \vdots & \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1\,1 & \cdots & 1\,1 \\ 0\,0 & \cdots & 1\,1 \end{pmatrix} = \begin{pmatrix} n & (n-\tau) \\ (n-\tau) & (n-\tau) \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} (n-\tau) & -(n-\tau) \\ -(n-\tau) & n \end{pmatrix} \frac{1}{n(n-\tau) - (n-\tau)^2}$$

$$(X^T X)^{-1} X^T = \frac{1}{n(n-\tau) - (n-\tau)^2} \begin{pmatrix} (n-\tau) & -(n-\tau) \\ -(n-\tau) & n \end{pmatrix} \begin{pmatrix} 1\,1 & \cdots & 1 \\ 0\,0 & \cdots & 1 \end{pmatrix}$$

$$= \left[n(n-\tau) - (n-\tau)^2 \right]^{-1} \begin{pmatrix} \underbrace{\begin{matrix} (n-\tau) & \cdots & (n-\tau) \\ -(n-\tau) & \cdots & -(n-\tau) \end{matrix}}_{\tau} & \underbrace{\begin{matrix} 0 & 0 & \cdots & 0 \\ \tau & \tau & \cdots & \tau \end{matrix}}_{n-\tau} \end{pmatrix}$$

Figure 1: Problem 1



$$(X^T X)^{-1} X^T Y =$$

$$(n(n-\tau) - (n-\tau)^2)^{-1} \begin{pmatrix} (n-\tau) & \cdots & (n-\tau) & 0 & 0 & \cdots & 0 \\ & & & & & & \\ (\tau-n) & \cdots & (\tau-n) & \tau & \tau & \cdots & \tau \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$= (n(n-\tau) - (n-\tau)^2)^{-1} \begin{pmatrix} \sum_{i=1}^{\tau} (n-\tau) y_i \\ \sum_{i=1}^{\tau} (\tau-n) y_i + \sum_{\tau+1}^{n} \tau y_i \end{pmatrix}$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{\tau} (n-\tau) y_i}{(n(n-\tau) - (n-\tau)^2)}$$

$$\hat{\Delta} = \frac{\sum_{i=1}^{\tau} (\tau-n) y_i + \sum_{i=\tau+1}^{n} \tau y_i}{(n(n-\tau) - (n-\tau)^2)}$$

Figure 2: Problem 1