Assignment 2-1

Xinyu Guo, Javed Hossain, Sriramjee Singh

September 29, 2017

```
#Clear memory
rm(list = ls())

#Read the data
dta = read_csv("rice.csv")

#Calculate cost
dta <-  dta %>%
  mutate(COST = AREA*AREAP + LABOR*LABORP + NPKP*NPK + OTHERP*OTHER)

#Normalize with respect to land price
dta <- dta %>%
  mutate(C = COST/AREAP,
         W2 = LABORP/AREAP,
         W3 = NPKP/AREAP,
         W4 = OTHERP/AREAP,
         Y  =PROD,
         W2W2=W2*W2,
         W3W3=W3*W3,
         W4W4=W4*W4,
         W2W3=W2*W3,
         W2W4=W2*W4,
         W3W4=W3*W4,
         W2Y=W2*Y,
         W3Y=W3*Y,
         W4Y=W4*Y,
         W2T=W2*YEARDUM,
         W3T=W3*YEARDUM,
         W4T=W4*YEARDUM,
         YY=Y*Y,
         TT=YEARDUM*YEARDUM
         )

#Create dummy for each farm
for(t in unique(dta$FMERCODE)) {
  dta[,paste("farm",t,sep="_")] <- as.numeric(dta$FMERCODE==t)
}

#Use factors in linear regressions instead of dummies
dta <- dta %>%
  mutate(FMERCODE = as.factor(FMERCODE))
```

Question 1

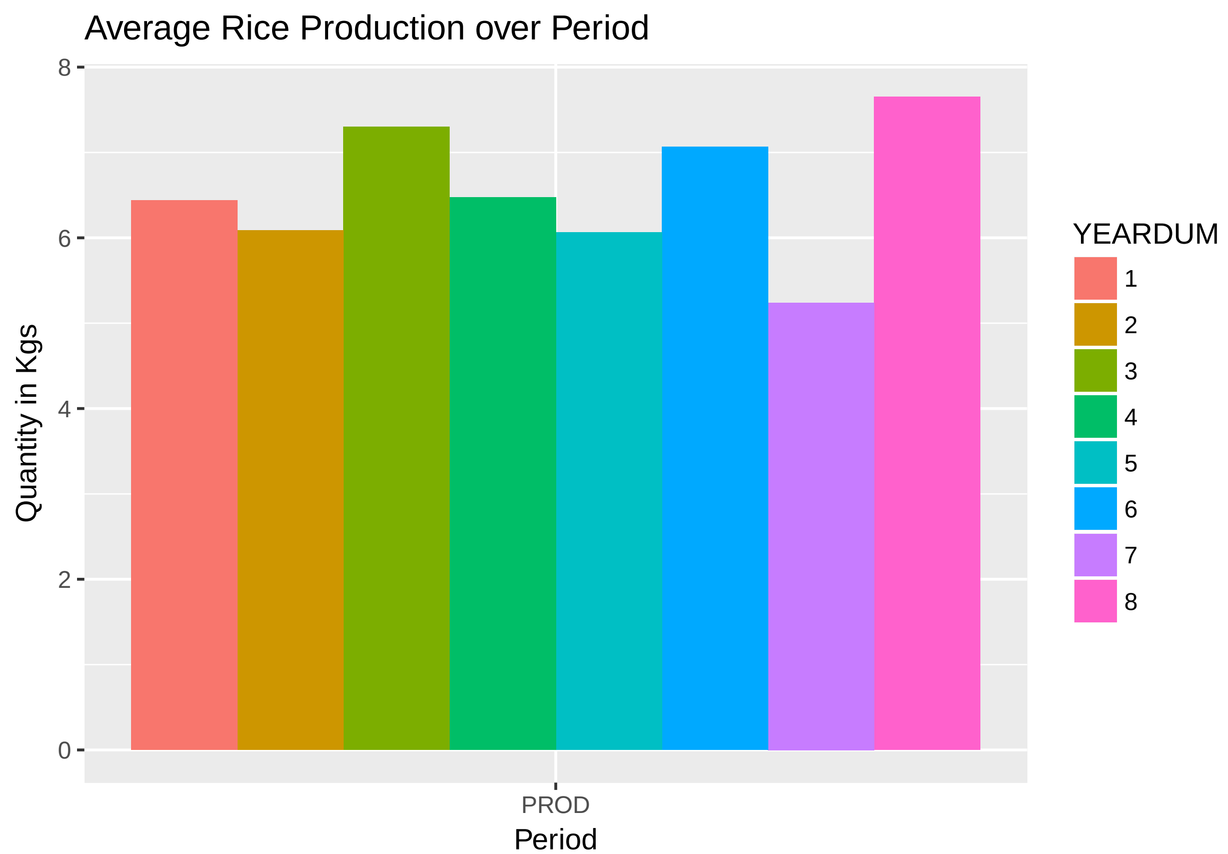
Provide a summary of the data for the output (rice), and input prices (labor, fertilizer, land, other) and the year variable. Do this in any way you want (tables and/or figures). Discuss.

```r
dta_new <- subset(dta, select=c("YEARDUM", "PROD", "AREA", "LABOR", "NPK", "OTHER"))
summary(dta_new)
```

```
##     YEARDUM          PROD             AREA            LABOR
##  Min.   :1.00   Min.   : 0.09   Min.   :0.200   Min.   :  8.00
##  1st Qu.:2.75   1st Qu.: 2.79   1st Qu.:1.000   1st Qu.: 49.75
##  Median :4.50   Median : 5.11   Median :1.750   Median : 87.50
##  Mean   :4.50   Mean   : 6.54   Mean   :2.144   Mean   :108.34
##  3rd Qu.:6.25   3rd Qu.: 8.92   3rd Qu.:3.400   3rd Qu.:152.00
##  Max.   :8.00   Max.   :31.10   Max.   :7.000   Max.   :437.00
##       NPK              OTHER
##  Min.   :  10.00   Min.   :   1.459
##  1st Qu.:  67.72   1st Qu.:  21.352
##  Median : 142.20   Median :  70.556
##  Mean   : 189.23   Mean   : 125.345
##  3rd Qu.: 254.35   3rd Qu.: 169.578
##  Max.   :1030.90   Max.   :1083.379
```

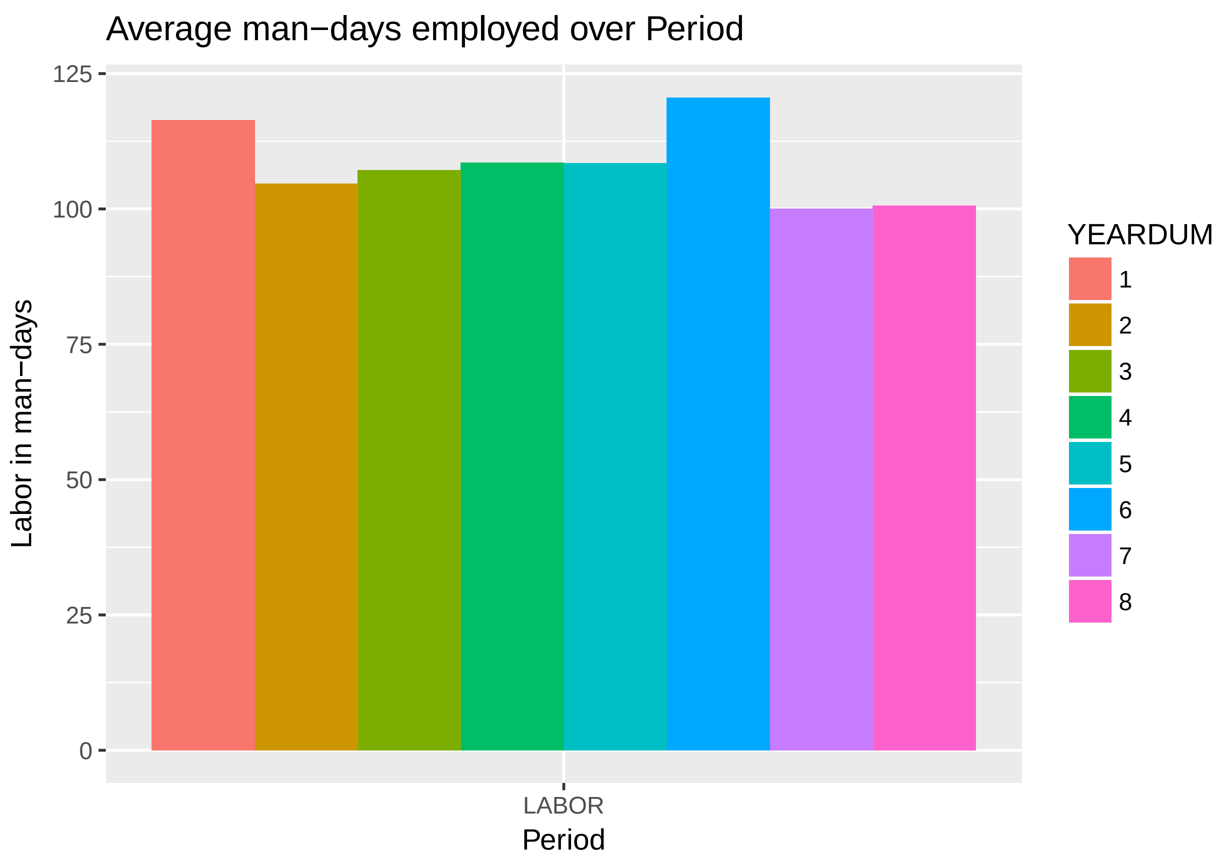
```r
dta_prod <- subset(dta, select=c("YEARDUM", "PROD"))
plot_prod <- dta_prod %>%
  group_by(YEARDUM=factor(YEARDUM)) %>%
  summarise_all(funs(mean)) %>%
  gather(Var, Val, -YEARDUM)%>%
  ggplot(., aes(x=Var, y=Val, fill=YEARDUM)) + geom_bar(stat='identity', position='dodge')

print(plot_prod + labs(y="Quantity in Kgs", x="Period") + ggtitle("Average Rice Production over Period")
```



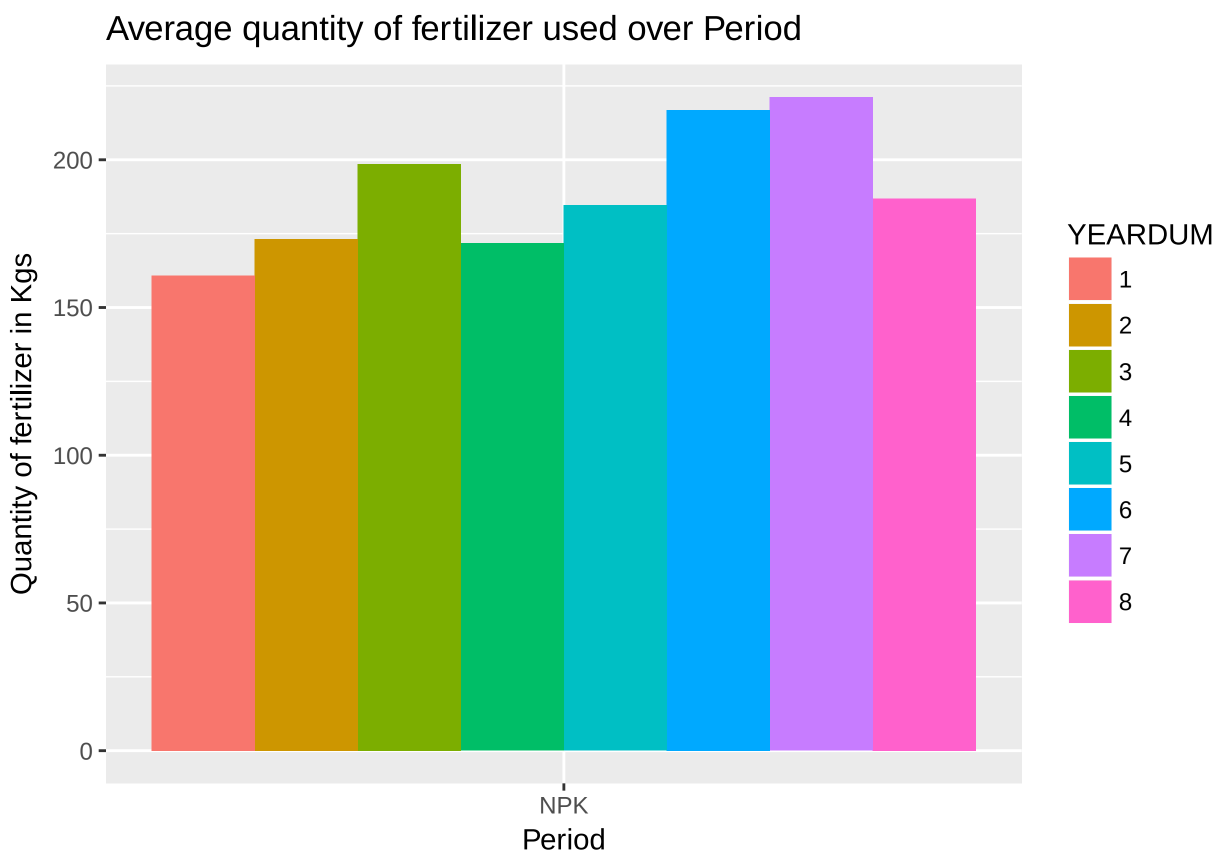
```r
dta_labor <- subset(dta, select=c("YEARDUM", "LABOR"))
plot_labor <- dta_labor %>%
  group_by(YEARDUM=factor(YEARDUM)) %>%
  summarise_all(funs(mean)) %>%
  gather(Var, Val, -YEARDUM)%>%
  ggplot(., aes(x=Var, y=Val, fill=YEARDUM)) + geom_bar(stat='identity', position='dodge')

print(plot_labor + labs(y="Labor in man-days", x="Period") + ggtitle("Average man-days employed over Pe
```



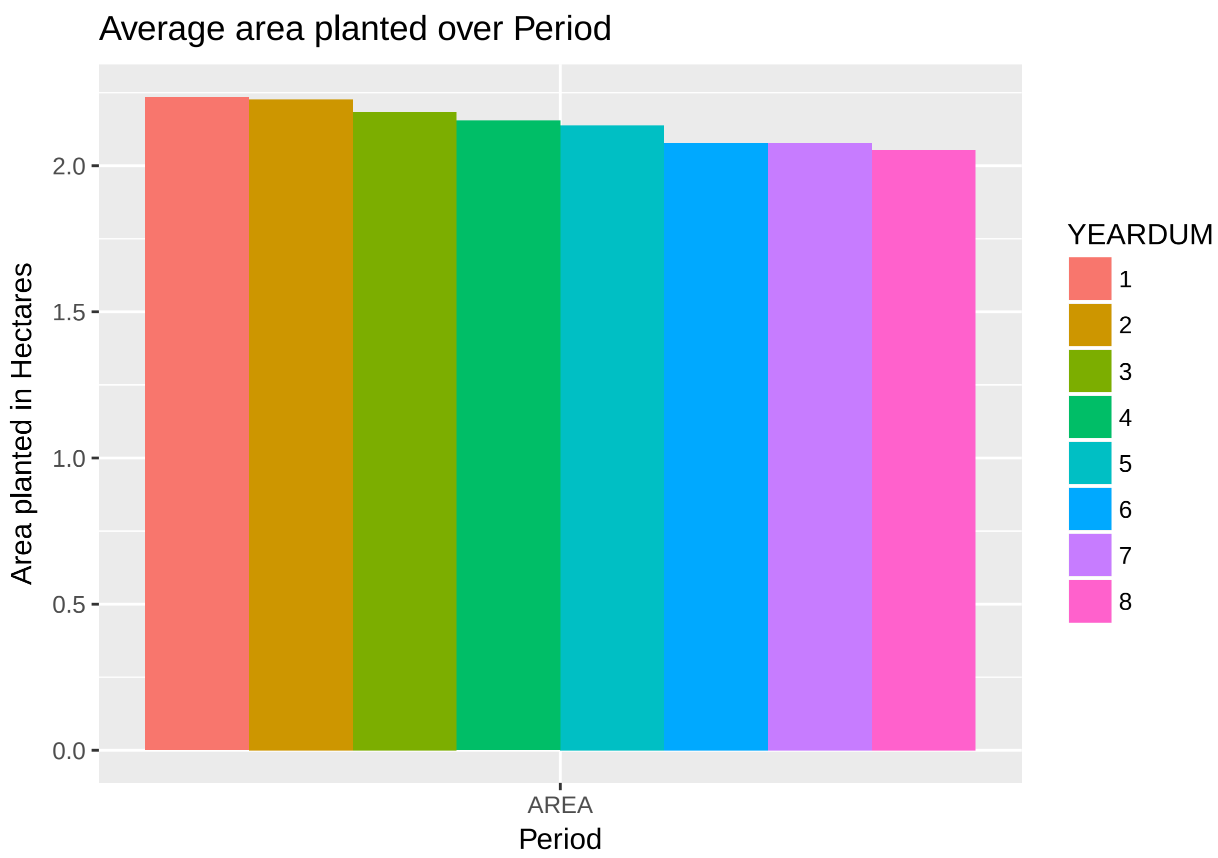

```r
dta_npk <- subset(dta, select=c("YEARDUM", "NPK"))
plot_npk <- dta_npk %>%
  group_by(YEARDUM=factor(YEARDUM)) %>%
  summarise_all(funs(mean)) %>%
  gather(Var, Val, -YEARDUM)%>%
  ggplot(., aes(x=Var, y=Val, fill=YEARDUM)) + geom_bar(stat='identity', position='dodge')

print(plot_npk + labs(y="Quantity of fertilizer in Kgs", x="Period") + ggtitle("Average quantity of fer
```



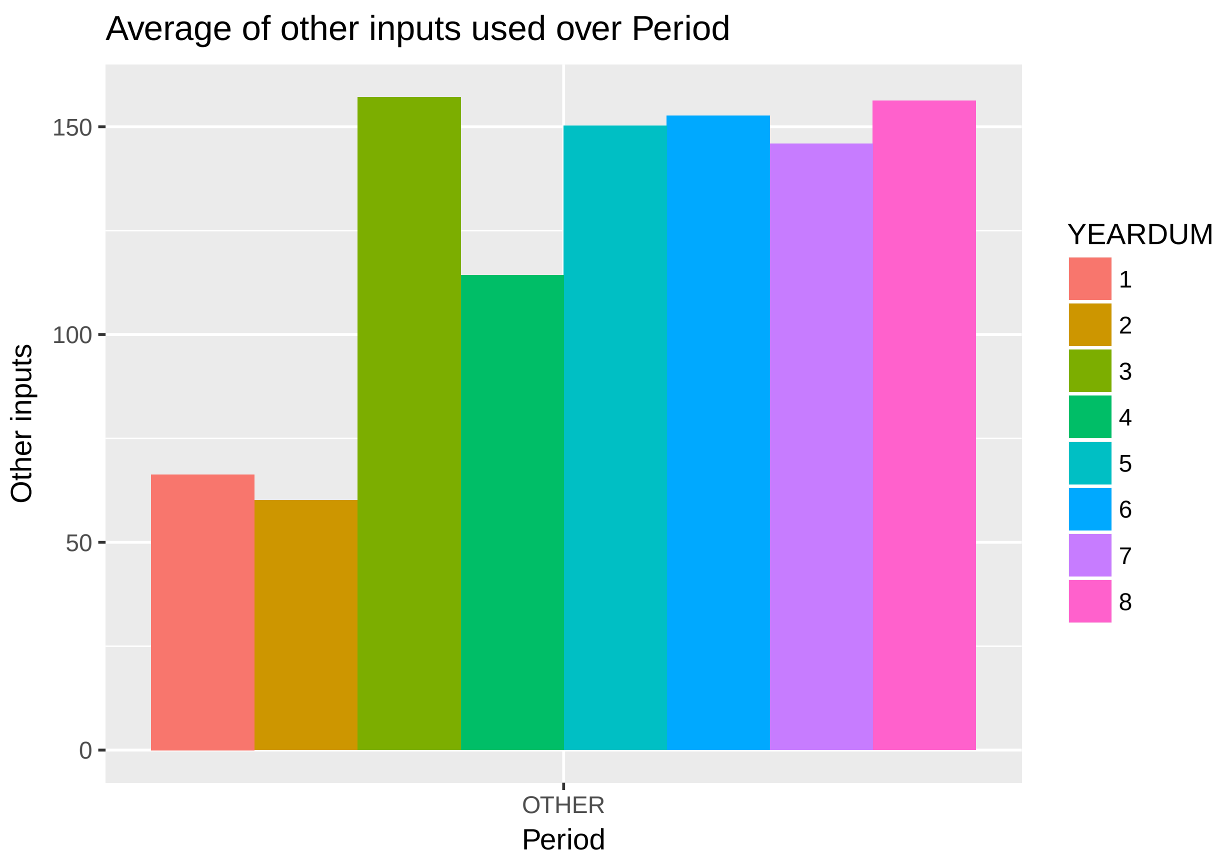

```
dta_land <- subset(dta, select=c("YEARDUM", "AREA"))
plot_land <- dta_land %>%
  group_by(YEARDUM=factor(YEARDUM)) %>%
  summarise_all(funs(mean)) %>%
  gather(Var, Val, -YEARDUM)%>%
  ggplot(., aes(x=Var, y=Val, fill=YEARDUM)) + geom_bar(stat='identity', position='dodge')

print(plot_land + labs(y="Area planted in Hectares", x="Period") + ggtitle("Average area planted over Pe
```




```
dta_other <- subset(dta, select=c("YEARDUM", "OTHER"))
plot_other <- dta_other %>%
  group_by(YEARDUM=factor(YEARDUM)) %>%
  summarise_all(funs(mean)) %>%
  gather(Var, Val, -YEARDUM)%>%
  ggplot(., aes(x=Var, y=Val, fill=YEARDUM)) + geom_bar(stat='identity', position='dodge')

print(plot_other + labs(y="Other inputs", x="Period") + ggtitle("Average of other inputs used over Peri
```



Discussion:

A bar graph is plotted for all variables in consideration to analyse their behavior of over the period. The data set is panel, having 8 periods. Mean value of the variable in a period is taken to see the trend over the entire period.

Rice production is quite good in 3rd, 6th and 8th period, where as 7th period turns out to be the minimum production period. The mean production of the entire period is in range of 6-8 kgs except the 7th period.

Labor employed in the production is quite stable over the period and in the range of 100-125 man-days. 1st and the 6th period labor employed are higher than the other periods.

Fertilizer used in the production is increasing from 1st to 3rd period, dropped in the 4th period and again rising until 7th period and then dropped in 8th period.

Area under plantation shows a very consistent trend. It is decreasing over the period.

Other inputs were very low in the first 2 periods and after that it picked up.

Question 2

Construct a correlation table the output (rice), the prices of inputs (labor, fertilizer, land, other) and the year variable. Discuss.

```
#Construct correlation table
cor(dta_new)
```

```
##            YEARDUM       PROD        AREA       LABOR        NPK
## YEARDUM  1.00000000 0.01685926 -0.04389731 -0.03293881 0.07889002
## PROD     0.01685926 1.00000000  0.88746187  0.88979277 0.81829032
```

```
## AREA    -0.04389731 0.88746187  1.00000000  0.91847583 0.84016972
## LABOR   -0.03293881 0.88979277  0.91847583  1.00000000 0.83507836
## NPK      0.07889002 0.81829032  0.84016972  0.83507836 1.00000000
## OTHER    0.18659963 0.72704520  0.72342149  0.74770188 0.71431166
##             OTHER
## YEARDUM 0.1865996
## PROD    0.7270452
## AREA    0.7234215
## LABOR   0.7477019
## NPK     0.7143117
## OTHER   1.0000000
```

Discussion:

Year has very low correlation with other variables, which means that there is no major change over the period.

Production has very high correlation with the input variables. As expected high inputs give high output.

Area under cultivation is also very highly correlated with all variables except Year.

Similarly, Labor, Fertilizer and Other input variables are highly correlated with all variables except Year.

The variables are complementary in nature.

Question 3

Estimate a normalized quadratic cost function for the production of rice, without the input demands. Normalize the cost and the prices using the price of land. Comment on your results.

```
attach(dta)
E <- cbind(W2, W3, W4, Y, YEARDUM)


# Regression formula
eqcost <- as.formula(C ~ (W2 + W3 + W4 + Y + YEARDUM)^2 + W2W2 + W3W3 + W4W4 +
    TT + YY + FMERCODE)

# Estimate single equation - no restictions
model_1 <- lm(eqcost, data = dta, singular.ok = TRUE)

texreg(model_1, caption = "Summary of regression results", omit.coef = "FMERCODE",
    caption.above = TRUE, label = "tab.OLS")
```

By using normalization, we impose homogenity condition. From this paticular specification of equition, we also impose symetry constrains.

Interestingly, from regression result above, the effect of Y alone on cost is negative and significant. However, when we consider the cross term (the interaction effect of Y and W2,W2,W3), the total effect of Y on cost became possitive and intuitively correct.

Question 4

Estimate a normalized quadratic cost function for the production of rice, simultaneously with the input demands. Normalize the cost and the prices using the price of land. How many constraints are there in the model that you estimate? Comment on your results.

Table 1: Summary of regression results

	Model 1
(Intercept)	4.38^{***}
	(0.93)
W2	-5.26
	(37.18)
W3	-16.05
	(190.35)
W4	-21.84
	(46.38)
Y	-0.48^{***}
	(0.10)
YEARDUM	-0.02
	(0.19)
W2W2	-958.91
	(495.81)
W3W3	-8509.83
	(7697.32)
W4W4	-1046.62
	(666.82)
TT	-0.01
	(0.02)
YY	0.00
	(0.00)
W2:W3	3624.70
	(3412.05)
W2:W4	1962.47^{*}
	(895.57)
W2:Y	7.63^{*}
	(3.29)
W2:YEARDUM	9.16
	(4.97)
W3:W4	1007.21
	(3366.00)
W3:Y	122.27^{***}
	(20.03)
W3:YEARDUM	-2.57
	(25.43)
W4:Y	14.17^{**}
	(4.28)
W4:YEARDUM	-5.61
	(4.64)
Y:YEARDUM	0.03^{***}
	(0.01)
R^2	0.94
Adj. R^2	0.92
Num. obs.	344
RMSE	1.19

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

```
# attach(dta)

eq_X2 <- as.formula(LABOR ~ E)
eq_X3 <- as.formula(NPK ~ E)
eq_X4 <- as.formula(OTHER ~ E)


k1 = length(model_1$coef)
k2 = ncol(E) + 1
Rrestr <- matrix(0, 3 * k2, (k1 + 3 * k2))

# First 3 constrain beta2 ~ beta4

# beta2
Rrestr[1, k1 + 1] <- 1
Rrestr[1, 2] <- -1

# beta22
model_1$coefficients[7]
```

```
W2W2
```

-958.9112

```
Rrestr[2, k1 + 2] <- 1
Rrestr[2, 7] <- -2  # (-2)* model_1$coefficients[7]+ b22=0

# beta23
model_1$coefficients[54]
```

W2:W3 3624.702

```
Rrestr[3, k1 + 3] <- 1
Rrestr[3, 54] <- -1

# beta24
model_1$coefficients[55]
```

W2:W4 1962.473

```
Rrestr[4, k1 + 4] <- 1
Rrestr[4, 55] <- -1

# gama21
model_1$coefficients[56]
```

```
W2:Y
```

7.631417

```
Rrestr[5, k1 + 5] <- 1
Rrestr[5, 56] <- -1

# rho2T
model_1$coefficients[57]
```

W2:YEARDUM 9.160386

```
Rrestr[6, k1 + 6] <- 1
Rrestr[6, 57] <- -1


# beta3
model_1$coefficients[3]
```

```
       W3
-16.05236
```

```
Rrestr[7, k1 + 7] <- 1
Rrestr[7, 3] <- -1

# beta32
model_1$coefficients[54]
```

W2:W3 3624.702

```
Rrestr[8, k1 + 8] <- 1
Rrestr[8, 54] <- -1

# beta33
model_1$coefficients[8]
```

```
     W3W3
-8509.833
```

```
Rrestr[9, k1 + 9] <- 1
Rrestr[9, 8] <- -2

# beta34
model_1$coefficients[58]
```

W3:W4 1007.208

```
Rrestr[10, k1 + 10] <- 1
Rrestr[10, 58] <- -1

# gama31
model_1$coefficients[59]
```

```
W3:Y
```

122.2731

```
Rrestr[11, k1 + 11] <- 1
Rrestr[11, 59] <- -1

# rho2T
model_1$coefficients[60]
```

W3:YEARDUM -2.57274

```
Rrestr[12, k1 + 12] <- 1
Rrestr[12, 60] <- -1
```

```r
# beta4
model_1$coefficients[4]
```

```
  W4
```

-21.8434

```r
Rrestr[2 * k2 + 1, k1 + 2 * k2 + 1] <- 1
Rrestr[2 * k2 + 1, 4] <- -1

# beta42
model_1$coefficients[55]
```

W2:W4 1962.473

```r
Rrestr[14, k1 + 14] <- 1
Rrestr[14, 55] <- -1

# beta43
model_1$coefficients[58]
```

W3:W4 1007.208

```r
Rrestr[15, k1 + 15] <- 1
Rrestr[15, 58] <- -1

# beta44
model_1$coefficients[9]
```

```
W4W4
```

-1046.62

```r
Rrestr[16, k1 + 16] <- 1
Rrestr[16, 9] <- -2

# gama41
model_1$coefficients[61]
```

```
W4:Y
```

14.16626

```r
Rrestr[17, k1 + 17] <- 1
Rrestr[17, 61] <- -1

# rho2T
model_1$coefficients[62]
```

W4:YEARDUM -5.613745

```r
Rrestr[18, k1 + 18] <- 1
Rrestr[18, 62] <- -1


qrestr <- seq(0, 0, length.out = nrow(Rrestr))

# Estimate system of equations
model2 <- list(eq1 = eqcost, eq2 = eq_X2, eq3 = eq_X3, eq4 = eq_X4)  #
```

```
model_2 <- systemfit(model2, method = "SUR", data = dta, restrict.matrix = Rrestr,
    restrict.rhs = qrestr, maxit = 5000)  #

texreg(model_2, omit.coef = "FMERCODE", caption = "System of cost function",
    caption.above = TRUE, label = "tab.system")
```

For this question alone we impose 6*3=18 constrains. For Question 3 we impose symetry constrains by specify model in a paticulary way.(5*5-5=20 in total)

Beacuse we add more information by simultaneously estimating the input demands, the precision of estimates shall increase.

Question 5

Verify whether the cost function in question 4 is concave with respect to the input prices. Discuss.

```
#Extract Hessian matrix of Cost Function
B <- matrix( ,ncol = 3,nrow =  3)
for(i in 1:3){
  for(j in 1:3){
    element_ij <- paste("eq",i+1,"_","EW",j+1,sep = "")
    B[i,j] <- model_2$coefficients[element_ij]
  }
}

#Calculate pricipal minor
B1 <- B[1,1]
B1
```

```
## [1] -557.733
```

```
B2 <- B[c(1,2),c(1,2)]
det(B2)
```

```
## [1] 2888818
```

```
B3 <- B
det(B)
```

```
## [1] 2788847619
```

Since the sign of principal minor does not alter, the cost function is not concave in W. (The reason might be that, during this time period, the production function has changed, like productivity shock or technology improvement. Thus the estimated cost function is no longer the estimation of same underlying input-output relation.)

Table 2: System of cost function

	Model 1
eq1: (Intercept)	1.23*
	(0.52)
eq1: W2	30.17***
	(4.33)
eq1: W3	14.30
	(14.36)
eq1: W4	−40.20**
	(13.07)
eq1: Y	0.11
	(0.06)
eq1: YEARDUM	−0.25*
	(0.11)
eq1: W2W2	−278.87***
	(51.69)
eq1: W3W3	−4075.58***
	(860.56)
eq1: W4W4	−600.59**
	(185.40)
eq1: TT	0.02
	(0.01)
eq1: YY	0.00
	(0.00)
eq1: W2:W3	1287.38***
	(359.58)
eq1: W2:W4	686.62***
	(130.64)
eq1: W2:Y	13.72***
	(0.35)
eq1: W2:YEARDUM	−2.26***
	(0.61)
eq1: W3:W4	1031.17*
	(472.28)
eq1: W3:Y	27.06***
	(1.04)
eq1: W3:YEARDUM	1.21
	(2.01)
eq1: W4:Y	21.82***
	(1.00)
eq1: W4:YEARDUM	5.79***
	(1.67)
eq1: Y:YEARDUM	0.00
	(0.01)
eq2: (Intercept)	30.17***
	(4.33)
eq2: EW2	−557.73***
	(103.37)
eq2: EW3	1287.38***
	(359.58)
eq2: EW4	686.62***
	(130.64)
eq2: EY	13.72***
	(0.35)
eq2: EYEARDUM	−2.26***
	(0.61)
eq3: (Intercept)	14.30
	(14.36)
eq3: EW2	1287.38***
	(359.58)