UNIVERSITY OF NEVADA LAS VEGAS

LEE BUSINESS SCHOOL

ECO 772- ECONOMETRICS II

ASSIGNMENT #01

STUDENT NAME: DANIEL EMAASIT

MAJOR: CIVIL ENGINEERING

INSTRUCTOR NAME: Prof. Ian McDonough

DATE SUBMITTED: 02/13/2015

SPRING 2015

Problem Set One

Daniel Emaasit

February 13, 2015

Problem C17.1

Use the data in PNTSPRD.RAW for this exercise:

Section (i)

The variable favwin is a binary variable if the team favored by the Las Vegas point spread wins. A linear probability model to estimate the probability that the favored team wins is: P(favwin=1|spread)=beta0 + beta1(spread)

Explain why, if the spread incorporates all relevant information, we expect beta0=0.5

<u>If spread is zero, there is no favorite, and the probability that the team we (arbitrarily) label the favorite should have a 50% chance of winning.</u>

subsection (ii)

Estimate the model from part (i) by OLS. Test H0:0.5 against a two-sided alternative. Use both the usual and heteroskedasticity-robust standard errors.

```
## read the data file into R
data1<-read_dta("PNTSPRD.dta")


## estimate the model from part(i) by OLS using usual standard erros
lm1<-lm(favwin~spread, data=data1)

library(AER)

coeftest(lm1)

##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.5769492  0.0282345 20.4342 < 2.2e-16 ***
## spread      0.0193655  0.0023386  8.2806 9.324e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



##testing the hypothesis that H0=0.5
calculated.t.value1=(0.5769492-0.5)/0.0282345
calculated.t.value1
```

```
## [1] 2.725361
```

The rule: Reject the null hypothesis if |calculated t value|>|critical t value|,

The decision: since the |calculated.t.value1|>1.96, we reject the null hypothesis

The inference: beta0 is not equal to 0.5.

```
##estimate the model from part(i) by OLS using heteroskedasticity-robust
standard errors
coeftest(lm1, vcov = vcovHC)

##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.5769492  0.0317187   18.19 < 2.2e-16 ***
## spread      0.0193655  0.0019289   10.04 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##testing the hypothesis that H0=0.5
calculated.t.value2=(0.5769492-0.5)/0.0317187
calculated.t.value2

## [1] 2.425988
```

The rule: Reject the null hypothesis if |calculated t value|>|critical t value|,

The decision: since the |calculated.t.value2|>1.96, we reject the null hypothesis

The inference: beta0 is not equal to 0.5

Subsection (iii)

Is spread statistically significant? What is the estimated probability that the favored team wins when spread=10?

Yes, spread is statistically significant.

```
## estimate the Probaility
newdata1<-data.frame(spread=10)
predict(lm1, type="response", newdata1)

##         1
## 0.7706044
```

The estimated probability is 0.7706042

Subsection (iv)

Now, estimate a probit model for P(favwin=1|spread). Interpret and test the null hypothesis that the intercept is zero. [Hint: Remember that theta= 0.5.]

```r
data1$favwin<-as.factor(data1$favwin)
glm1<-glm(favwin~spread, data=data1, family=binomial(link="probit"))
stargazer(glm1, type = "text")
```

```
##
## =============================================
##                       Dependent variable:
##                   ---------------------------
##                             favwin
## ---------------------------------------------
## spread                     0.092***
##                             (0.012)
##
## Constant                    -0.011
##                             (0.103)
##
## ---------------------------------------------
## Observations                  553
## Log Likelihood             -263.562
## Akaike Inf. Crit.           531.124
## =============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

Subsection (v)

Use the probit model to estimate the probability that the favored team wins when spread 10. Compare this with the LPM estimate from part (iii).

```r
## estimated Probability using the probit model
newdata2<-data.frame(spread=10)
predict(glm1, type="response", newdata2)
```

```
##         1
## 0.8196512
```

The estimated probability is 0.8196512. This probability is higher than the one estimated using the LPM.

Subsection (vi)

Add the variables favhome, fav25, and und25 to the probit model and test joint significance of these variables using the likelihood ratio test. (How many df are in the chi-square distribution?) Interpret this result, focusing on the question of whether the spread incorporates all observable information prior to a game.

```r
## convert the new variables from character types to factor types
data1$favhome<-as.factor(data1$favhome)
data1$fav25<-as.factor(data1$fav25)
data1$und25<-as.factor(data1$und25)

## add the variables to the probit model
```

```
glm2<-glm(favwin~spread+favhome+fav25+und25, data=data1,
family=binomial(link="probit"))
stargazer(glm2, type = "text")
```

```
##
## =============================================
##                       Dependent variable:
##                   ---------------------------
##                             favwin
## ---------------------------------------------
## spread                     0.088***
##                             (0.013)
##
## favhome1                    0.149
##                             (0.137)
##
## fav251                      0.003
##                             (0.159)
##
## und251                      -0.220
##                             (0.251)
##
## Constant                    -0.055
##                             (0.129)
##
## ---------------------------------------------
## Observations                  553
## Log Likelihood             -262.642
## Akaike Inf. Crit.           535.284
## =============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

The value of the Log Likelihood is -262.642

```
## calculate the value of the likelihood ratio statistic
2*(-262.642 - (-263.562))
```

```
## [1] 1.84
```

Since the likelihood ratio statistic of 1.84 is greater than the p-value from the chi-square distribution of 0.61, the added variables are jointly not significant. This can be infered to mean that these added variables have no additional power for predicting the outcome.

Problem C17.2

Use the data in LOANAPP.RAW for this exercise

Subsection (i)

Estimate a probit model of approve on white. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability estimates?

```
## start by reading the data into R
data2<-read_dta("LOANAPP.dta")

## convert the variables from character types to factors
data2$white<-as.factor(data2$white)
data2$approve<-as.factor(data2$approve)

## estimate a probit model of approve on white
glm3<-glm(approve~white, data2, family=binomial(link="probit"))
stargazer(glm3, type="text")

##
## ============================================
##                       Dependent variable:
##                   ---------------------------
##                             approve
## --------------------------------------------
## white1                     0.784***
##                             (0.087)
##
## Constant                   0.547***
##                             (0.075)
##
## --------------------------------------------
## Observations                 1,989
## Log Likelihood              -700.877
## Akaike Inf. Crit.          1,405.755
## ============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01

## find the estimated probability for whites
newdata3<-data.frame(white=1)
newdata3$white<-as.factor(newdata3$white)
predict(glm3, type="response", newdata3)

##         1
## 0.9083879

# find the estimated probability for non-whites
newdata4<-data.frame(white=0)
newdata4$white<-as.factor(newdata4$white)
predict(glm3, type="response", newdata4)

##         1
## 0.7077922
```

The predicted probability for whites and non-whites is 0.9083879 and 0.7077922 respectively.

Subsection (ii)

Now, add the variables hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, and vr to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

```
## convert some of variables from characters to factors
data2$male<-as.factor(data2$male)
data2$married<-as.factor(data2$married)
data2$dep<-as.factor(data2$dep)
data2$sch<-as.factor(data2$sch)
data2$cosign<-as.factor(data2$cosign)
data2$chist<-as.factor(data2$chist)
data2$pubrec<-as.factor(data2$pubrec)
data2$mortlat1<-as.factor(data2$mortlat1)
data2$mortlat2<-as.factor(data2$mortlat2)
data2$vr<-as.factor(data2$vr)


## estimate the new probit model
glm4<-
glm(approve~white+hrat+obrat+loanprc+unem+male+married+dep+sch+cosign+chist+p
ubrec+mortlat2+vr, data2, family=binomial(link="probit"))
stargazer(glm4, type="text")

##
## =============================================
##                       Dependent variable:
##                   ---------------------------
##                             approve
## ---------------------------------------------
## white1                     0.525***
##                             (0.097)
##
## hrat                         0.009
##                             (0.007)
##
## obrat                      -0.029***
##                             (0.006)
##
## loanprc                    -1.007***
##                             (0.238)
##
## unem                       -0.037**
##                             (0.018)
##
## male1                       -0.056
```

```
##                                (0.109)
##
## married1                      0.266***
##                                (0.096)
##
## dep1                           -0.017
##                                (0.119)
##
## dep2                           -0.132
##                                (0.121)
##
## dep3                           -0.115
##                                (0.170)
##
## dep4                          -0.553**
##                                (0.280)
##
## dep5                            4.218
##                               (95.018)
##
## dep6                            3.999
##                               (130.569)
##
## dep7                            2.984
##                               (235.034)
##
## dep8                            3.079
##                               (235.034)
##
## sch1                            0.016
##                                (0.096)
##
## cosign1                         0.055
##                                (0.241)
##
## chist1                        0.589***
##                                (0.096)
##
## pubrec1                       -0.759***
##                                (0.127)
##
## mortlat21                      -0.491
##                                (0.338)
##
## vr1                           -0.192**
##                                (0.082)
##
## Constant                      2.080***
##                                (0.315)
##
```

```
## ---------------------------------------------
## Observations                    1,989
## Log Likelihood                -600.528
## Akaike Inf. Crit.             1,245.055
## =============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

```
## the calculated t value is
(0.525-0/0.097)
```

```
## [1] 0.525
```

The point estimate for white = 0.525 with standard errors of 0.097. Since the calculated t-value of 5.412 is greater than the critical t-value of 1.96, we reject the null hypothesis that whites is equal to zero. The inference is that white is significant. Hence there is stilll strong evidence of discrimination against non whites.

Subsction (iii)

Estimate the model from part (ii) by logit. Compare the coefficient on white to the probit estimate.

```
## estimate the model by logit
glm5<-
glm(approve~white+hrat+obrat+loanprc+unem+male+married+dep+sch+cosign+chist+p
ubrec+mortlat2+vr, data2, family=binomial(link="logit"))
stargazer(glm5, type="text")
```

```
##
## =============================================
##                       Dependent variable:
##                    ---------------------------
##                             approve
## ---------------------------------------------
## white1                     0.946***
##                             (0.173)
##
## hrat                        0.016
##                             (0.013)
##
## obrat                      -0.055***
##                             (0.011)
##
## loanprc                    -1.899***
##                             (0.456)
##
## unem                       -0.067**
##                             (0.033)
##
## male1                       -0.094
##                             (0.205)
```

```
##
## married1                          0.503***
##                                   (0.181)
##
## dep1                              -0.008
##                                   (0.225)
##
## dep2                              -0.258
##                                   (0.229)
##
## dep3                              -0.235
##                                   (0.319)
##
## dep4                              -0.970*
##                                   (0.503)
##
## dep5                              13.662
##                                  (599.419)
##
## dep6                              13.671
##                                  (814.654)
##
## dep7                              11.913
##                                 (1,455.398)
##
## dep8                              12.091
##                                 (1,455.398)
##
## sch1                               0.047
##                                   (0.178)
##
## cosign1                            0.068
##                                   (0.446)
##
## chist1                            1.072***
##                                   (0.172)
##
## pubrec1                          -1.308***
##                                   (0.218)
##
## mortlat21                         -0.900
##                                   (0.594)
##
## vr1                              -0.341**
##                                   (0.154)
##
## Constant                          3.821***
##                                   (0.593)
##
## --------------------------------------------
```

```
## Observations                     1,989
## Log Likelihood                 -600.776
## Akaike Inf. Crit.              1,245.552
## =============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

The coefficient of white is 0.946 with a standard error of 0.173.

Subsection (iv)

Use equation (17.17) to estimate the sizes of the discrimination effects for probit and logit.

```
## multiply the coefficient from the logit model by 0.625
0.625*0.946

## [1] 0.59125
```

The scaled cofficient from the logit model=0.59125, which is close to the one from the probit model.

Problem C17.3

Use the data in FRINGE.RAW for this exercise

Subsection (i)

For what percentage of the workers in the sample is pension equal to zero? What is the range of pension for workers with nonzero pension benefits? Why is a Tobit model appropriate for modeling pension?

```
## start by reading the data into R
data3<-read_dta("FRINGE.dta")

## calculate the percentage of the workers whose pension is equal to zero
(172/616)*100

## [1] 27.92208
```

The range is $7.28 to $2880.27. The tobit model is appropriate because the sample consists of 28% workers with zero pension benefits, and the range of positive benefits is wide.

Subsection (ii)

Estimate a Tobit model explaining pension in terms of exper, age, tenure, educ, depends, married, white, and male. Do whites and males have statistically significant higher expected pension benefits?

```
## estimate a tobit model
data3$exper<-as.numeric(data3$exper)
data3$age<-as.numeric(data3$age)
data3$tenure<-as.numeric(data3$tenure)
```

```
data3$educ<-as.numeric(data3$educ)
data3$depends<-as.numeric(data3$depends)
data3$married<-as.factor(data3$married)
data3$white<-as.factor(data3$white)
data3$male<-as.factor(data3$male)

tobit.model<-tobit(pension~exper+age+tenure+educ+depends+married+white+male,
data=data3)
summary(tobit.model)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##            616            172            444              0
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.252e+03  2.191e+02  -5.717 1.09e-08 ***
## exper        5.203e+00  6.010e+00   0.866    0.387
## age         -4.639e+00  5.711e+00  -0.812    0.417
## tenure       3.602e+01  4.565e+00   7.892 2.97e-15 ***
## educ         9.321e+01  1.089e+01   8.558  < 2e-16 ***
## depends      3.528e+01  2.192e+01   1.610    0.107
## married1     5.369e+01  7.174e+01   0.748    0.454
## white1       1.441e+02  1.021e+02   1.412    0.158
## male1        3.082e+02  6.989e+01   4.409 1.04e-05 ***
## Log(scale)   6.519e+00  3.562e-02 183.014  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Scale: 677.7
##
## Gaussian distribution
## Number of Newton-Raphson Iterations: 4
## Log-likelihood: -3673 on 10 Df
## Wald-statistic: 204.5 on 8 Df, p-value: < 2.22e-16
```

Male is significant while White is not significant.

Subsection (iii)

Use the results from part (ii) to estimate the difference in expected pension benefits for a white male and a nonwhite female, both of whom are 35 years old, are single with no dependents, have 16 years of education, and have 10 years of experience.

```
## create new data frames with the new data points
newdata5<-
data.frame(white="1",male="1",age=35,married="0",depends=0,educ=16,exper=10,
tenure=10)
newdata6<-
data.frame(white="0",male="0",age=35,married="0",depends=0,educ=16,exper=10,
```

```
tenure=10)

## use the predicted values to calculate the difference
predict(tobit.model, newdata5) - predict(tobit.model, newdata6)
## 452.236
```

The difference is $ 452.236

Subsection (iv)

Add union to the Tobit model and comment on its significance.

```
## add union to the tobit model and comment on its significance
data3$union<-as.numeric(data3$union)

tobit.model2<-
tobit(pension~exper+age+tenure+educ+depends+married+white+male+union,
data=data3)
summary(tobit.model2)


## Observations:
##          Total  Left-censored     Uncensored Right-censored
##            616            172            444              0
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.572e+03  2.185e+02  -7.191 6.44e-13 ***
## exper        4.394e+00  5.831e+00   0.753 0.451159
## age         -1.654e+00  5.556e+00  -0.298 0.765987
## tenure       2.878e+01  4.505e+00   6.388 1.68e-10 ***
## educ         1.068e+02  1.077e+01   9.916  < 2e-16 ***
## depends      4.147e+01  2.121e+01   1.955 0.050624 .
## married1     1.975e+01  6.950e+01   0.284 0.776329
## white1       1.593e+02  9.897e+01   1.610 0.107487
## male1        2.572e+02  6.802e+01   3.782 0.000156 ***
## union        4.390e+02  6.249e+01   7.026 2.12e-12 ***
## Log(scale)   6.481e+00  3.548e-02 182.693  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Scale: 652.9
##
## Gaussian distribution
## Number of Newton-Raphson Iterations: 4
## Log-likelihood: -3649 on 11 Df
## Wald-statistic: 260.3 on 9 Df, p-value: < 2.22e-16
```

Union is significant in the model

Subsection (v)

Apply the Tobit model from part (iv) but with peratio, the pension-earnings ratio, as the dependent variable. (Notice that this is a fraction between zero and one, but,though it often takes on the value zero, it never gets close to being unity. Thus, a Tobit model is fine as an approximation.) Does gender or race have an effect on the pension-earnings ratio?

```
## using peratio as the dependent variable
tobit.model3<-
tobit(peratio~exper+age+tenure+educ+depends+married+white+male+union,
data=data3)
summary(tobit.model3)
```

```
##
## Call:
## tobit(formula = peratio ~ exper + age + tenure + educ + depends +
##     married + white + male + union, data = data3)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##            616            172            444              0
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.0550630  0.0144896  -3.800 0.000145 ***
## exper        0.0001697  0.0003861   0.440 0.660230
## age         -0.0002176  0.0003669  -0.593 0.553081
## tenure       0.0017605  0.0003019   5.832 5.48e-09 ***
## educ         0.0053478  0.0007172   7.457 8.88e-14 ***
## depends      0.0008265  0.0014185   0.583 0.560140
## married1     0.0032941  0.0046339   0.711 0.477163
## white1       0.0031793  0.0065656   0.484 0.628215
## male1        0.0025937  0.0045309   0.572 0.567021
## union        0.0300458  0.0041859   7.178 7.08e-13 ***
## Log(scale)  -3.1270435  0.0359053 -87.091  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Scale: 0.04385
##
## Gaussian distribution
## Number of Newton-Raphson Iterations: 4
## Log-likelihood: 607.6 on 11 Df
## Wald-statistic: 163.9 on 9 Df, p-value: < 2.22e-16
```

<u>Gender and race dont have significants effects the pension-earnings ratio.</u>