ECO 5350 T. Fomby
Intro. Econometrics Fall 2016

EXERCISE 9
KEY

Purpose: To learn more about analyzing multiple regression models (including the LPM) with heteroskedastic errors. **This exercise is due on Tuesday, November 22.**

Work **word problem exercises 1, 2, and 4, page 268** in your textbook. Work **computer exercises C4, C6, and C7 on page 271** in your textbook. Each of these exercises will count 10 points each. I want you to provide "clean" answer sheets in the sense that I want your answers to be "typed up" using Microsoft Word or some equivalent document using a Mac word processor. I will be taking up all of the exercises.

1.

(i) False

(ii) True

(iii) True

2.

$$beer^* = \beta_0 onestar + \beta_1 inc^* + \beta_2 price^* + \beta_3 educ^* + \beta_4 female^* + u^*$$

where $beer^* = beer * \frac{1}{inc}, onestar = 1 * \frac{1}{inc}, inc^* = \frac{inc}{inc} = 1, price^* = price * \frac{1}{inc}, educ^* = educ * \frac{1}{inc}, female^* = female * \frac{1}{inc}, u^* = u * \frac{1}{inc}$.

4.

(i) These variables' coefficients have the expected signs if you believe that, in general, students become more accustomed to college work and, therefore, improve their grades semester by semester.

(ii) Yes, this hypothesis makes sense because the next term's GPA should mirror the previous terms' weighted cumulative GPA. The appropriate t-statistic for this test is $t = (\hat{\beta}_1 - 1)/se(\hat{\beta}_1)$. Using the OLS standard error we have t = -0.57. If we use the heteroscedasticity standard error we have t= -0.60. Using the two-sigma rule, we accept the null hypothesis that the population coefficient, β_1, irrespective of which standard error we use.

(iii) The t-statistics are t = -0.157/0.098 = -1.60 and t = -0.157/0.080 = -1.96, using the OLS and heteroscedasticity-robust standard errors, respectively. The two-tailed probability error of the second t-statistic is 0.05 while the first t-statistic's two-tailed probability value is greater than 0.05. Yes, the p-values are dependent upon the standard error used.

C4. For the STATA key see Exercise_C4_Ch8_Key.do
The code is

```
* Use VOTE1.dta
regress voteA prtystrA democA lexpendA lexpendB
* Obtain OLS residuals
predict res, residual
* This just shows that the residuals are, by the Least Squares procedure,
* always orthogonal (independent of) the explanatory variables, X'uhat = 0
regress res prtystrA democA lexpendA lexpendB
generate res2 = res^2
* The Breusch-Pagan Test for Heteroskedasticity (See equation (8.19), p. 253)
* Use the overall F-statistic from this equation
regress res2 prtystrA democA lexpendA lexpendB
* The special case of White's test (See equation (8.20), p. 253)
regress voteA prtystrA democA lexpendA lexpendB
* Obtain predicted values of voteA = voteA_hat
predict voteA_hat
generate voteA_hat2 = voteA_hat^2
* Use the overall F-statistic from this equation
regress res2 voteA_hat voteA_hat2
```

(i) It is the case that the OLS solution provides the result $X'\hat{u} = 0$. In other words, by construction, the OLS the residuals are orthogonal to the explanatory variables in an OLS regression. That is why a regression of the OLS residuals on the explanatory variables of the regression have an R^2 = 0.

(ii) Using Equation (8.19), p. 253, the overall F-statistic of the B-P test equation is 2.33 with a p-value of 0.0581 which is not significant at the 5% level but is at the 10% level. We might say that the incidence of heteroskedasticity in this equation is "marginal."

(iii) Using Equation (8.20), p. 253, the overall F-statistic of the special case of the White test is 2.79 with p-value of 0.0645. Again, the incidence of heteroscedasticity is "marginal."

C6. For the STATA key see Exercise_C6_Ch8_Key.do

The code is

```
* Use Crime1.dta
generate arr86 = (narr86>0)
regress arr86 pcnv avgsen tottime ptime86 qemp86
```

predict arr86hat
summarize arr86hat
*** The weight for each observation using WLS is 1/sqrt[arr86hat*(1 - arr86hat)]**
generate weight = 1/[arr86hat*(1 - arr86hat)]^0.5
generate arr86star = arr86*weight
generate onestar = 1*weight
generate pcnvstar = pcnv*weight
generate avgsenstar = avgsen*weight
generate tottimestar = tottime*weight
generate ptime86star = ptime86*weight
generate qemp86star = qemp86*weight
*** WLS estimation of the LPM**
regress arr86star onestar pcnvstar avgsenstar tottimestar ptime86star qemp86star, noconstant
test avgsen tottime

(i) The smallest fitted value is 0.0066431. The largest fitted value is 0.5576897. Therefore, all of the fitted values are between 0 and 1.

(ii)

```
. * WLS estimation of the LPM
. regress arr86star onestar pcnvstar avgsenstar tottimestar ptime86star qemp86star, noconstant

      Source |       SS           df       MS      Number of obs   =     2,725
-------------+----------------------------------   F(6, 2719)      =    188.31
       Model |  1109.44323         6  184.907205   Prob > F        =    0.0000
    Residual |  2669.84459     2,719   .98192151   R-squared       =    0.2936
-------------+----------------------------------   Adj R-squared   =    0.2920
       Total |  3779.28782     2,725  1.38689461   Root MSE        =    .99092

------------------------------------------------------------------------------
   arr86star |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     onestar |   .4475965   .0179922    24.88   0.000     .4123167    .4828763
    pcnvstar |  -.1678436   .0189122    -8.87   0.000    -.2049272     -.13076
  avgsenstar |   .0053665   .0051146     1.05   0.294    -.0046624    .0153954
 tottimestar |  -.0017615   .0032514    -0.54   0.588     -.008137     .004614
 ptime86star |  -.0246188   .0030451    -8.08   0.000    -.0305898   -.0186479
  qemp86star |  -.0451885   .0054225    -8.33   0.000    -.0558212   -.0345558
------------------------------------------------------------------------------
```

(iii)

The *avgsen* and *tottime* variables are jointly insignificant since the joint F-test statistic has a p-value 0.4129 which is greater than 0.05.

```
. test avgsen tottime

 ( 1)  avgsenstar = 0
 ( 2)  tottimestar = 0

       F(  2,  2719) =    0.88
            Prob > F =    0.4129
```

C7. For the STATA key see Exercise_C7_Ch8_Key.do

The code is

*** Use Loanapp.dta**
*** LPM with OLS standard errors**
regress approve white hrat obrat loanprc unem male married dep sch cosign chist pubrec mortlat1 mortlat2 vr
*** LPM with heteroskedasticity-robust standard errors**
regress approve white hrat obrat loanprc unem male married dep sch cosign chist pubrec mortlat1 mortlat2 vr, vce(robust)
predict yhat
summarize yhat

(i)

```
. * LPM with OLS standard errors
. regress approve white hrat obrat loanprc unem male married dep sch cosign chist pubrec mortlat1 mortlat2 vr

      Source |       SS           df       MS      Number of obs   =     1,971
-------------+----------------------------------   F(15, 1955)     =     25.86
       Model |  35.4004787        15  2.36003192   Prob > F        =    0.0000
    Residual |  178.393534     1,955   .09124989   R-squared       =    0.1656
-------------+----------------------------------   Adj R-squared   =    0.1592
       Total |  213.794013     1,970   .10852488   Root MSE        =    .30208

------------------------------------------------------------------------------
     approve |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .1288196   .0197317     6.53   0.000     .0901223    .1675169
        hrat |    .001833   .0012632     1.45   0.147    -.0006444    .0043104
       obrat |  -.0054318   .0011018    -4.93   0.000    -.0075926    -.003271
     loanprc |  -.1473001   .0375159    -3.93   0.000    -.2208755   -.0737246
        unem |  -.0072989    .003198    -2.28   0.023    -.0135708   -.0010271
        male |  -.0041441   .0188644    -0.22   0.826    -.0411405    .0328523
     married |   .0458241   .0163077     2.81   0.005     .0138418    .0778064
         dep |  -.0068274   .0067013    -1.02   0.308    -.0199699    .0063151
         sch |   .0017525   .0166498     0.11   0.916    -.0309006    .0344057
      cosign |   .0097722   .0411394     0.24   0.812    -.0709094    .0904538
       chist |   .1330267   .0192627     6.91   0.000     .0952492    .1708043
      pubrec |  -.2419268   .0282274    -8.57   0.000    -.2972858   -.1865677
    mortlat1 |  -.0572511    .050012    -1.14   0.252    -.1553336    .0408314
    mortlat2 |  -.1137234   .0669838    -1.70   0.090    -.2450905    .0176438
          vr |  -.0314408   .0140313    -2.24   0.025    -.0589586   -.0039229
       _cons |   .9367312   .0527354    17.76   0.000     .8333077    1.040155
------------------------------------------------------------------------------
```

```
. * LPM with heteroskedasticity-robust standard errors
. regress approve white hrat obrat loanprc unem male married dep sch cosign chist pubrec mortlat1 mortlat2 vr, vce(robust)

Linear regression                               Number of obs     =      1,971
                                                F(15, 1955)       =      14.98
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1656
                                                Root MSE          =     .30208
```

approve	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
white	.1288196	.0258693	4.98	0.000	.0780852	.179554
hrat	.001833	.001467	1.25	0.212	-.0010441	.0047101
obrat	-.0054318	.001331	-4.08	0.000	-.0080421	-.0028215
loanprc	-.1473001	.0378351	-3.89	0.000	-.2215013	-.0730988
unem	-.0072989	.0037122	-1.97	0.049	-.0145792	-.0000187
male	-.0041441	.0193044	-0.21	0.830	-.0420035	.0337152
married	.0458241	.0172374	2.66	0.008	.0120186	.0796296
dep	-.0068274	.0069038	-0.99	0.323	-.0203669	.0067122
sch	.0017525	.017146	0.10	0.919	-.0318739	.0353789
cosign	.0097722	.0395825	0.25	0.805	-.0678561	.0874005
chist	.1330267	.0246202	5.40	0.000	.0847421	.1813114
pubrec	-.2419268	.0427922	-5.65	0.000	-.3258498	-.1580037
mortlat1	-.0572511	.0662234	-0.86	0.387	-.1871269	.0726247
mortlat2	-.1137234	.0910697	-1.25	0.212	-.2923274	.0648806
vr	-.0314408	.0144855	-2.17	0.030	-.0598493	-.0030322
_cons	.9367312	.0593886	15.77	0.000	.8202595	1.053203

The robust confidence interval for β_{white} is [0.078, 0.179].
The non-robust confidence interval for β_{white} is [0.090, 0.167].
The non-robust confidence interval is narrower than the robust confidence interval.

(ii)

```
. summarize yhat
```

Variable	Obs	Mean	Std. Dev.	Min	Max
yhat	1,971	.876205	.1340514	.2273447	1.172988

The minimum fitted value is greater than 0 but the maximum fitted value is greater than 1.0 making the application of WLS to this LPM problematic. Either corresponding observations have to be dropped or the fitted values that are outside of the unit interval have to be “trimmed” to be inside the unit interval. For example, if a yhat is less than zero, make it 0.001. If a yhat is greater than one, make it 0.999. After the trimming then apply WLS.