Second Exam: Economics 388, Econometrics Winter 2002 in R. Butler's course

YOUR NAME:____

Section I (30 points) Questions 1-10 (3 points each)
Section II (40 points) Questions 11-14 (10 points each)
Section III (30 points) Questions 15-16 (15 points each)

Section I. Define or explain the following terms (3 points each)

1.heteroskedasticity robust standard error-

2. weighted least squares estimator--

3. program evaluation-

4. logit-

5. Breusch-Pagan test—

6. $\binom{n}{x}$, or "n choose x" (combination)--

7. dummy variable trap--

8. standardized beta coefficients--

9. Goldfeld-Quandt test--

10. log-likelihood ratio test--

Some Fun Stuff: True, False or Uncertain Questions (11-12) and some regression questions (13 and 14). You are graded on your explanations for the True or False questions, and not on whether you listed T or F correctly….

11. a. T, F or U: "If heteroskedasticity is due to an omitted variable, the OLS estimator $\hat{\beta}$ will be unbiased but the estimated covariance matrix is biased (and so are the associated t-statistics)."

b. T, F or U: "The linearity or boundedness problem with the linear probability model is that the errors are sometimes heteroskedastic."

12. a T,F or U: "To do a LM test of a restriction in the standard OLS model, you follow this procedure: 1) regress Y on the restricted (say "q" restrictions, or q-coefficients forced to be zero, for example) set of regressors, then 2) take the residuals from step 1 and regress them on all of the regressors (including those restricted from the first stage), then 3) take n (the sample size) and multiply it by the adjusted R-square from the second regression. This last statistic (n*adj. R-square) will be distributed as chi-square distribution with q-degrees of freedom under the null hypothesis."

12. b. T, F or U: "Plotting the residuals against the independent variables would never be a useful way to find heteroskedasticity since heteroskedasticity concerns the variances of the error terms and not the means of the error terms."

13. Suppose that wages are regressed on age (AGE), educational attainment (EDUC) and whether or not a worker's occupation is professional or manager (PROF_MAN), indicated in the following Shazam code:

```
ols wage age educ prof_man
```

Suppose that you wanted to test whether males and females had the same regression function (i.e., had the same slopes and intercepts), and you had a dummy variable for females (FEMALE).

a. Write the Shazam code (including any other variables that you need for your analysis) that 1) uses only one regression and one or more test statements (and creating any other variables that you need for your analysis), and 2) formally tests whether there is a difference between the male and female wage function.

b. If you ran your program, how would you know when to reject the null hypothesis that the regression function was the same for males and females?

14. When the log of weekly wages are regressed on a list of the usual suspects (important suspects include AGE=age, AGE_SQ=age-squared, AGE_ED=age*level of educational attainment, and EDUC_ATT= educational attainment), we get the following regression:

```
R-SQUARE =   0.3044     R-SQUARE ADJUSTED =   0.2542

 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR     180 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 AGE       0.80313E-01 0.2936E-01   2.735     0.007 0.200      1.3102     0.4960
 AGE_SQ   -0.12041E-02 0.3224E-03  -3.735     0.000-0.268     -1.5558    -0.3107
 AGE_ED    0.18692E-02 0.1845E-02   1.013     0.312 0.075      0.5342     0.1591
 EDUC_ATT  0.15325E-03 0.7656E-01  0.2002E-02 0.998 0.000      0.0005     0.0003
 MALE      0.22922     0.1090       2.102     0.037 0.155      0.1447     0.0209
 EXEC      0.27740     0.2891      0.9595     0.339 0.071      0.1685     0.0163
 TECH_SAL -0.23893E-01 0.3050     -0.7834E-01 0.938-0.006     -0.0096    -0.0004
 SERV_OCC -0.80815E-01 0.2913     -0.2774     0.782-0.021     -0.0439    -0.0032
 OPER_OCC  0.16959     0.2814      0.6026     0.548 0.045      0.0929     0.0069
 AG_CNSTR  0.30084     0.1955       1.539     0.126 0.114      0.1135     0.0049
 MANUF     0.17270     0.1531       1.128     0.261 0.084      0.0782     0.0043
 TRADE     0.25213     0.1567       1.609     0.109 0.119      0.1257     0.0080
 PUB_ADMN  0.25337     0.2213       1.145     0.254 0.085      0.0744     0.0024
 CONSTANT   3.6003      1.031       3.492     0.001 0.252      0.0000     0.5953
 |_test
 |_ test age=0
 |_ test age_sq=0
 |_ test age_ed=0
 |_end
 F STATISTIC =   6.9617316     WITH    3 AND  180 D.F.  P-VALUE= 0.00019
```

What does this indicate about the impact of AGE on Weekly wages?

15. a. Using matrix algebra, show that if all the slope variables are scaled through by the same constant (say, α , where alpha may be 2, for example), then the associated slope coefficients would be scaled through by one divided by that constant ($1/\alpha$, for example if all the values of the slope variables are doubled, the associated coefficients would be halved in least squares estimation). REALLY BIG HINT: This is very much like, but not quite identical to, the scaling through that we did using the T-matrix in the last class and in the last lecture on heteroskedasticity.

b. Notice that the proof ("show") you did above in part (a) is like, but not quite identical to, the scaling through that we did using the T-matrix in the last class and in the last lecture on heteroskedasticity. What exactly are the differences? (ie, the differences between the scaling matrix you used above in part a and the scaling matrix, T, we used in the class notes)?