Take Home Final Exam

Part I

Each question of the following 6 questions is worth 1 point. Where asked to choose of true/false/uncertain indicate your choice.

Q1: If the errors in the classical linear regression (CLR) model are not normally distributed, although the LSE estimator is no longer BLUE, it is still unbiased. True/false/uncertain.

Q2: The process of multiplying the dependent variable by a contant, adding a constant to the depend variable, or both, does not change the regression R^2. True/false/uncertain.

Q3: If an extra explanatory variable is added to a regression, the estimator of σ^2 will remain the same or fall. True/false/uncertain.

Q4: Since x^2 is an exact function of x, we will be faced with exact multicollinearity if we attempt to use both x and x^2 as regressors. True/false/uncertain.

Q5: Suppose the classical normal linear regression (CNLR) model applies to $y = \alpha + \beta x + \varepsilon$ with $\sigma^2 = 40$. A sample of size 10 yields $\sum x = 20$ and $\sum x^2 = 50$. You plan to test the hypothesis that $\beta = 1$, at the 5% significance level, against the alternative $\beta > 1$. If the true value of β is 4.0, what is the probability that you will correctly reject your null hypothesis?

Q6: If the errors in CLR have so-called Laplace (or double exponential) distribution then the maximum likelihood estimation of the regression coefficients is equivalent to minimizing the sum of absolute values of residuals (LAD estimator). True/false/uncertain.
(The probability density function for Laplace distribution is $f(x) = \frac{1}{2\psi} e^{|x/\psi|}$, $\psi > 0$.)

Part II

Each of the following tasks is worth 2 points.

Suppose we believe that the data on Y (file y.dat) are generated by a gamma distribution. Then, the probability density function for Y is

$$f(y) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda y} y^{P-1}, P > 0, \lambda > 0, y > 0.$$

We are going to estimate the parameters P and λ. We begin by writing $\lambda =$

$e^{-\alpha}$. (This simplifies estimation.) The following table gives the maximum likelihood estimates of P and α:

```
+---------------------------------------------+
| Gamma (Loglinear) Regression Model          |
| Dependent variable                   Y      |
| Number of observations              30      |
| Log likelihood function      -135.6994      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Alpha        2.982974253       .27324762   10.917   .0000
          Scale parameter for gamma model
 P            1.899707559       .45401699    4.184   .0000

Estimated Asymptotic Covariance Matrix
 .074664260122   -.10850692601
-.10850692601     .20613142755
```

(a) The table gives the estimates of α and P. Use these results to estimate λ. Estimate the asymptotic variance of this estimator of λ. What is the asymptotic distribution?

(b) The expected value of the random variable, Y is $\mu = P/\lambda = Pe^{a}$. Estimate μ using your maximum likelihood estimates. Estimate the asymptotic standard error of this estimator. Present a 95% confidence interval for the parameter μ based on your results.

(c) Since $\mu = E[Y]$ is P/λ, you should be able to estimate μ with the sample mean of the observations on Y. Do so, and describe your finding. Using the familiar formula for the variance of the mean, estimate the standard error of this estimator, and compare your result to that in (b).

(d) The variance of this random variable is $\sigma^2 = P/\lambda^2$. You should be able to estimate σ^2 with the sample variance of the observations on Y. Do so, and compare your estimate to the one you get by using the MLEs in the table.

Now, we also have data on another variable X (file x.dat).
We will now formulate a kind of regression model. We believe that $Y|X$ has the gamma distribution specified earlier, but now,

$$\lambda = e^{-(\alpha+\beta x)}$$

The table below presents the maximum likelihood estimates of the parameters of this model.

```
+---------------------------------------------+
| Gamma (Loglinear) Regression Model          |
| Dependent variable                  Y       |
| Number of observations             30       |
| Log likelihood function      -95.96569      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
         Parameters in conditional mean function
 Alpha       1.352410457       .26425908    5.118   .0000
 Beta       -.5148571650   .23630535E-01  -21.788   .0000     2.1833333
         Scale parameter for gamma model
 P           22.93380099       5.8791785    3.901   .0001

 Matrix Cov.Mat. has  3 rows and  3 columns.
              1             2             3
        +-----------------------------------------
       1|      .0698        -.0012       -1.5072
       2|     -.0012          .0006     .1402168D-10
       3|    -1.5072     .1402168D-10    34.5647
```

(e) We are interested in the expected value of $Y|X$. As before, this is P/λ which is now

$$E[Y|X] = Pe^{\alpha+\beta X}$$

Using the results above, form a confidence interval for your estimate of $E(Y|X = \bar{x})$.

(f) Linearly regress Y on a constant and X. What is the slope in this regression? Compare this slope to the maximum likelihood estimates.

(g) Test the hypothesis that β equals 0 using a Wald test and using a likelihood ratio test. The estimates for the restricted model are

```
+---------------------------------------------+
| Gamma (Loglinear) Regression Model          |
| Maximum Likelihood Estimates                |
| Dependent variable                  Y       |
| Weighting variable                ONE       |
| Number of observations             30       |
| Iterations completed                8       |
| Log likelihood function      -135.6994      |
| Restricted log likelihood    -138.7402      |
| Chi-squared                   6.081682      |
| Degrees of freedom                  1       |
| Significance level          .1365908E-01  |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
```

```
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant     2.982974253       .27324762   10.917   .0000
          Scale parameter for gamma model
 P_scale      1.899707559       .45401699    4.184   .0000
```

(h) The following questions are based on the regression model:

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 XD + \beta_5 D + \varepsilon$$

ε is assumed to be zero mean, homoscedastic, and nonautocorrelated. (file reg.dat) Estimate the parameters of the model using ordinary least squares. We are interested in examining the marginal effect of changes in X on $E(Y|X, D)$. What is $\partial E(y|X, D)/\partial X$? Compute this effect with X held at its mean value and both with $D = 1$ and 0. What are the two estimates? How would you compute a standard error for each of these estimates of this effect? How would you test the hypothesis that these two effects are equal?

The data in file reg.dat are now assumed to come from a process in which there is a linear regression model, but possibly a heteroscedastic disturbance. The regression equation is:

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 D + \varepsilon$$

ε has mean 0, but may be heteroscedastic.

(i) Suppose that the true variance of ε is

$$Var(\varepsilon) = \sigma^2 e^{XD}$$

(remember, XD is X times D in the data above). If you estimate the betas using ordinary least squares, what are the properties of the estimator? (Bias, consistency, efficiency, true covariance matrix.)

(j) Suppose you believe that the variance of epsilon is $\sigma^2 e^{XD}$, but, in fact, the true variance is just σ^2. (I.e., your belief is mistaken.) Suppose you fit the model by GLS in spite of the true variance. What are the properties of your estimator? (Note, you can use true GLS here, since there are no free parameters in the variance function.)

(k) Compute the two estimators you described in parts (i) and (j), and report all results. (Note, in part (j), there are no parameters in the variance part, so you can compute the true GLS estimator.) Compare the variances of the OLS and GLS estimator, both true and estimated.

(l) Using the least squares results, compute the White estimator for the variance of the OLS estimator. Describe why you would do this computation.

(m) Suppose the true model is, in fact

$$Var(\varepsilon) = \sigma^2 e^{\alpha XD}$$

where α is a parameter to be estimated. How would you test the hypothesis that alpha equals 1.0 against the alternative hypothesis that alpha is not equal to 1.0?