COLORADO STATE UNIVERSITY
DEPARTMENT OF AGRICULTURAL & RESOURCE ECONOMICS

Problem Set 3b **Agricultural & Resource Economics / Economics 535** *S.R. Koontz*
Fall 2023 **Applied Econometrics**

Regression, Serial Correlation, and Heteroskedasticity

Provide answers to the following questions in a word-processed document. This problem set is to be a team effort between two individuals. Each team must submit a unique assignment. However, you may work with other students. Show the work necessary to communicate clear answers each question. Good answers always communicate. This complete problem set is worth 50 points. Round all reported statistics to the fourth decimal.

The objective of this problem set is to address the problem of serial correlation and heteroskedasticity in applied regression analysis. You will need to estimate regressions, test for serial correlation or heteroskedasticity, and re-estimate regressions correcting for each problem using econometric software. You will then compare the estimates and conclusions of hypotheses tests across the methods.

B. The second part of this problem set examines a model of consumer credit card expenditures and addresses heteroskedasticity. Answer the following questions. (Because estimates of this model are available in textbooks, please attach a printout from your software of your regression results to the end of this Problem Set. Be efficient.)

1. Report the OLS regression results for the following model of average monthly consumer credit card expenditures. Estimate

$$\text{expenditure}_i = \beta_0 + \beta_1\, \text{age}_i + \beta_2\, \text{rent}_i + \beta_3\, \text{income}_i + \beta_4\, \text{income}^2_i + e_t$$

where expenditure_i = expenditure for the ith person, age_i = age of the ith person, rent_i = a dummy variable which indicates if the ith person rents their home, and income_i = the income for the ith person. Report the estimates in equation form or within a table. Report the standard error and p-value of each coefficient under the estimate, the regression F-statistic, its p-value, model R^2, and root error variance.

2. Graph the expenditure data, the predicted values from the model, and residuals against income on a single plot. Ignore potential outliers in the data. Describe the variance of the residuals. Does the error variance appear constant across the sample?

3. Use the Park test, Goldfeld-Quandt test, Breusch-Pagan-Godfrey test, and the White test to examine the model for heteroskedasticity. Report each test statistic and p-value. Discuss how you conduct these tests but it is not necessary to report all the results from the various models. Report enough so that you communicate you understand each test. (Do not make any transformations of the explanatory variables for the Park test. Also, don't use the rent dummy variable within any created variables – use the rent dummy variable but do not multiply it by anything. And do not drop any observations from the data set when conducting the Goldfeld-Quant test. Finally, figure out how to do the Goldfeld-Quandt test using a dummy variable.)

4. Since you appear to have heteroskedasticity, report the regression results with White's heteroskedasticity consistent standard errors and p-values. Discuss if the coefficients change and

discussion if the conclusions drawn from the individual t-tests change. Does heteroskedasticity appear to be an important problem with this sample?

5. Estimate the model which corrects for heteroskedasticity through a transformation of the data. Assume the heteroskedasticity is a function of only the income variable. Assume you have standard error model heterskedasticity. You can estimate this model using ordinary least squares. (See Gujarati page 392-393.) Or you can use weighted least squares if you figure out how to define the weight variable. Do both but report one. Discuss how you do both. And identify which method you are reporting. Report the coefficient estimates, standard errors and p-values, the regression F-statistic, its p-value, the model R^2, and the root error variance estimate.

6. Estimate the model which corrects for multiplicative heteroskedasticity by jointly estimating the mean equation and a variance equation. The variance equation should include one income variable and a dummy variable for individuals that have income higher than the median income. The model is written as follows:

$$\text{expenditure}_i = \beta_0 + \beta_1\ \text{age}_i + \beta_2\ \text{rent}_i + \beta_3\ \text{income}_i + \beta_4\ \text{income}^2_i + e_t$$

$$\sigma^2_i = \exp(\alpha_0 + \alpha_1\ \text{income}_i + \alpha_2\ D_i)$$

Report the coefficient estimates, standard errors and p-values, and report other summary statistics provided by your regression package. Use hypotheses tests to determine if the error variance has increases with income or if there is a jump in the average variance. Conduct a likelihood ratio test of the significance of the heteroskedasticity model. The null hypothesis for this test is H_0: $\alpha_1=\alpha_2=0$.

Alternatively, estimate the model which corrects for heteroskedasticity by iteratively estimating the mean equation and a variance equation. You can estimate this final mean equation model using weighted least squares.

$$\text{expenditure}_i = \beta_0 + \beta_1\ \text{age}_i + \beta_2\ \text{rent}_i + \beta_3\ \text{income}_i + \beta_4\ \text{income}^2_i + e_t$$

$$\ln(\sigma^2_t) = \alpha_0 + \alpha_1\ \text{income}_t + \alpha_2\ D_i + u_t$$

First, estimate the mean equation using OLS. Second, use the mean equation residuals squared and logged in the variance equation as the dependent variable. Use OLS. Third, use the square root exponentiated of the predicted values from the variance equation as weights and re-estimate the mean equation using weighted least squares. (Technically, more iterations are needed but this can be ignored. What you are doing is called a one-step correction.)

Report the coefficient estimates, standard errors and p-values, and other summary statistics. Use hypotheses tests to determine if the error variance increases with income or if there has been an increase in the average error variance. Conduct a joint test of the significance of the heteroskedasticity model. (And attach a printout of your software results.)

7. Discuss the statistical results across the regressions. Do the coefficient estimates change? What about the standard errors and p-values? Be concise.

8. (Optional.) Graph the predicted values from your models against income. Plot predictions from the mean equation and the square root of predictions from the variance equation. Hold the age constant at the mean age and assume your individuals own their home.