Problem 5

Part 1 (No computer)

Consider a linear model to explain monthly beer consumption:

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

$$\mathrm{E}(u|inc, price, educ, female) = 0$$
$$\mathrm{Var}(u|inc, price, educ, female) = \sigma^2 inc^2.$$

Write the transformed equation that has a homoskedastic error term.

Part 2 (Use SAS)

Use the vote dataset.

(i) Estimate a model with *voteA* as the dependent variable and *prtystrA*, *democA*, log(*expendA*), and log(*expendB*) as independent variables. Obtain the OLS residuals, $\hat{u}_i$, and regress these on all of the independent variables. Explain why you obtain $R^2 = 0$.

(ii) Now, compute the Breusch-Pagan test for heteroskedasticity. Use the F statistic version and report the p-value.

Dataset vote

1. state	state postal code
2. district	congressional district
3. democA	=1 if A is democrat
4. voteA	percent vote for A
5. expendA	campaign expends. by A, $1000s
6. expendB	campaign expends. by B, $1000s
7. prtystrA	% vote for president
8. lexpendA	log(expendA)
9. lexpendB	log(expendB)
10. shareA	100*(expendA/(expendA+expendB))

Part 3 (Use SAS in order to verify that the regression equation and robust standard errors I wrote are correct. Print your output table with both standard errors from proc model.)

The variable *smokes* is a binary variable equal to one if a person smokes, and zero otherwise. Using the data in SMOKE, we estimate a linear probability model for *smokes*:

$$\widehat{smokes} = .656 - .069\log(cigpric) + .012\log(income) - .029\,educ$$
$$(.855)\ (.204)\qquad (.026)\qquad (.006)$$
$$[.856]\ [.207]\qquad [.026]\qquad [.006]$$
$$+ .020\,age - .00026\,age^2 - .101\,restaurn - .026\,white$$
$$(.006)\qquad (.00006)\qquad (.039)\qquad (.052)$$
$$[.005]\qquad [.00006]\qquad [.038]\qquad [.050]$$
$$n = 807,\ R^2 = .062.$$

(i) Are there any important differences between the two sets of standard errors?

(ii) Holding other factors fixed, if education increases by four years, what happens to the estimated probability of smoking?

(iii) At what point does another year of age reduce the probability of smoking?

(iv) Interpret the coefficient on the binary variable *restaurn* (a dummy variable equal to one if the person lives in a state with restaurant smoking restrictions).

(v) Person number 206 in the data set has the following characteristics: $cigpric = 67.44$, $income = 6{,}500$, $educ = 16$, $age = 77$, $restaurn = 0$, $white = 0$, and $smokes = 0$. Compute the predicted probability of smoking for this person and comment on the result.

Dataset smoke

1. educ	years of schooling
2. cigpric	state cigarette price, cents per pack
3. white	=1 if white
4. age	in years
5. income	annual income, $
6. cigs	cigs. smoked per day
7. restaurn	=1 if state restaurant smoking restrictions
8. lincome	log(income)
9. agesq	age^2
10. lcigpric	log(cigprice)