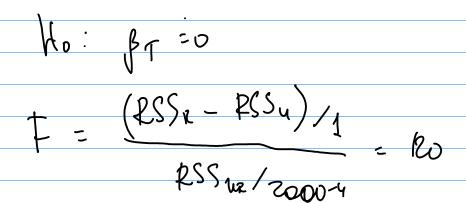


(a) [8 marks] Based on the data characterizing 2000 sellers, the researcher calculates the following equations

 $\hat{S}_i = -20.4 + 5.2E_i + 1.8IQ_i$ RSS = 2735345

$$\hat{S}_i = -22.8 + 5.3E_i + 1.7IQ_i + 19.0T_i RSS = 2579858$$

$$\begin{split} \hat{S}_i &= -22.8 + 5.3E_i + 1.7IQ_i + 19.0T_i \ RSS = 2579858 \\ \hat{S}_i &= -18.1 + 4.8E_i + 1.7IQ_i + 14.3T_i + 1.6E_i * T_i + 0.1IQ * T_i \ RSS = 2577011 \end{split}$$



(b) [7 marks] In addition, the researcher decided to estimate regressions in the specification of equation (1) separately for sellers who did not participate in the training (Equation 4, 1373 observations) and for sellers who received professional training (Equation 5, 627 observations):

$$\hat{S}_i = -18.1 + 4.8E_i + 1.7IQ_i \quad RSS = 1622106 \tag{4}$$

$$\hat{S}_i = -18.1 + 4.8E_i + 1.7IQ_i \quad RSS = 1622106$$

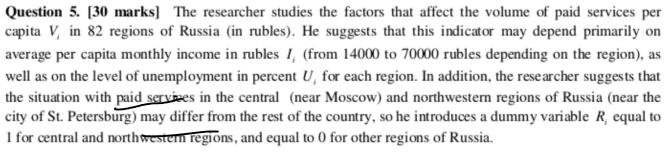
$$\hat{S}_i = -14.4 + 6.4E_i + 1.8IQ_i \quad RSS = 954905$$
(4)

(RSS* - (RSS, + R(S2)) (FSS, + RS) /(2000-6)

(c) [8 marks] The researcher decided to additionally take into account the factor M of the store location (M equal to one for sellers working in Moscow stores and zero for the region), for which he estimated the following equations.

$$\hat{S}_i = -24.3 + 5.3E_i + 1.7IQ_i + 18.9T_i + 3.3M_i RSS = 2574433$$
 (6)

$$\hat{S}_i = -23.2 + 5.3E_i + 1.7IQ_i + 16.5T_i + 1.9M_i + 4.6T_i * M_{ii} RSS = 2572170$$
 (7)



(a) [7 marks] To assess the impact of income on paid services, the researcher first runs a simple linear regression $V_i = -2448.2 + 2.05I_i \quad R^2 = 0.78$ $(3546.8) \quad (0.12)$ The researcher then rank all regions in order of income of the regions of Russia.

The researcher then rank all regions in order of increasing per capita income, and then regresses first for the 20 regions with the lowest income (getting RSS value equal $4.81 \cdot 10^8$), and then for the 30 regions with the highest income (getting RSS value equal $5.87 \cdot 10^9$). How can this information be used to check the data for heteroscedasticity? Carry out the necessary calculations, explaining your actions, and make the conclusion.

Se(Ei) prop I Vor(Ei) = 6 4. Ii WLS with I

 $60 = \frac{252/30-2}{9.551/20}$

Question 1. (17 marks)

A researcher, using a sample of 626 individuals from the National Longitudinal Survey of Youth, is investigating how the probability of a respondent obtaining a bachelor's degree from a four-year college is related to the respondent's score on a test of cognitive ability, $SCORE_i$ ranged from 0 to 100 (22 to 65 in the sample, with mean value 50.2, and most scores lying in the range 40 to 60). $ETHBLACK_i = 1$ and $ETHHISP_i = 1$ for the ethnic black and ethnic hispanic persons correspondingly and zero for the others.

26.7 percent of the respondents earned bachelor's degrees. Defining a variable BA_i to be equal to one if the respondent has a bachelor's degree (or higher degree) and zero otherwise, the researcher fitted the OLS regression (1) (standard errors in parentheses): and logit regressions (2) μ (3) (asymptotic standard errors in parentheses):

$$BA_i = -0.54 + 0.015 \cdot SCORE_i$$
 $R^2 = 0.10$ (1)

$$A_i = -8.54 + 0.13 \cdot SCORE_i$$
 McFadden $R^2 = 0.13$ (2)
(0.99) (0.018) LR stat. = 72.49

$$A_i = -9.05 + 0.14 \cdot SCORE_i$$
 (1.26) ETHBLACK_i $-0.34 \cdot ETHHISP_i$ McFadden $R^2 = 0.14$ (1.09) (0.019) (0.45) (0.66) LR stat. = 80.24

(a) \square Evaluate marginal effect of *ETHBLACK*, for the ethnic black person with $SCORE_i = 70$? (Use the method of direct comparison of two probabilities, explaining why it is most sutable here and presenting and explaining all steps of your calculations).

2 Sesse