SØK3001 H2021 Assessment guidelines

a)The interpretation of coefficients in regressions with only binary explanatory variable and constant term is explained in chapter 2-7 in textbook. Thus, the coefficient in front of cathhs in the simple regression in column (1) is interpreted as the difference in average test score between students attending catholic and public schools (3.46), while the constant term in column (1) represents the average test score for students attending public schools (51.9230).

b) The coefficient in front of cathhs in col (2) is the expected test score difference between catholic and public school students holding family income, mother and father education and gender constant. The estimated effect of log (family income) is ≈1.8 and 1.8/100=0.018 can be interpreted as the change in test score per percent increase in family income. The interpretation of the other coefficients should be straight forward as they are effects of dummy variables. The 90% confidence interval around the population parameter in front of cathhs is $1.4162 \pm t_{0.1} \cdot 0.4171$, where $t_{0.1}$ denotes the critical value in the t-distribution which is 1.645 in our case. Thus a 90% confidence interval around the population parameter is approximately $(0.730, 2.102)$

c) The null hypothesis that attending catholic schools has equal effect for boys and girls is simple to test by a t-test of the coefficient in front of cathhs·female in table 2. The t-test statistic is -0.1602/0.8305 which is lower than the critical value in the t-distribution for all convenient significance levels and hence we cannot reject the null hypothesis that attending catholic school has equal effect for boys and girls.

d) The problem is that students selects themselves to catholic schools. Thus, whether or not a student attends catholic schools is not random but the result of an active choice made by the student and his/her parents which may be partially correlated with other characteristics that affect student's test score. Thus, cathhs can be endogeneous and correlated with the error term in the econometric model due to the self-selection problem, see ch 15 p. 503 and also the discussion of self-selection in chapter 7 in textbook. This would imply that the OLS estimators used in col (1) and col (2) are inconsistent and thus the estimated effects of cathhs there cannot be given a causal interpretation.

e) The self-selection problem disussed above is an argument for using the instrumental variable method to estimate the causal effect. Students should formulate the first stage equation as well as the structural equation in the IV-2SLS framework. The effect in as estimated in col (6) is that a student attending catholic school has 3.95 points higher test score than a student attending public schools, holding parental income, education and gender constant.

f) Students should realize that the regression in col (4) represents the reduced form equation for student test scores. The coefficient in front of parcath in equation (4) is thus the reduced form effect on math test scores from having parents attending catholic school (the instrumental variable). The estimated coefficient of 0.56 implies that students with parents attending catholic schools have approximately 0.56 points higher test score than student with parents not attending catholic schools.

They should realize that the regression in col (5) represents the first stage equation for cathhs where the dummy for whether one of student's parents entered catholic school (parcath) is the instrumental variable. The dependent variable in col (5) is a dummy, parcath, and is to be interpreted as the probability to attend a catholic high school. Thus, the coefficient in front of parcath in col (5) can be interpreted as the difference in the probability to attend a catholic school between a student with and without parents attending catholic school. The coefficient estimate of 0.14 means that the probability to attend catholic highschool is 14% higher for students with parents attending catholic schools than for students with parents not attending.

g)The students should explain the relevance and exclusion requirements for the IV-2SLS method to be valid. The relevance requirement is that parcath predicts cathhs, and can be tested by testing the null hypothesis that the coefficient in front of the instrument parcath in the first stage regression in col (5) is zero. If this hypothesis is rejected with sufficiently clear margin, the relevance requirement appears to be fulfilled. Here, the t-test statistic is 0.1429/0.056=25.5 which implies an F-test statistic (the square of t-statistic) well above the rule of thumb level of 10. Thus parcath is a strong instrument. Further, using the IV-method as in col (6) build on the assumption that parcath only affects test scores through the effect on cathhs (the exclusion restriction). The question of causal interpretation of the IV estimator depends crucially on whether the exclusion restriction is fulfilled or plausible. The exclusion restriction is basically untestable and students might argue that parcath possibly affects student achievement directly and not only through the selection of a catholic high school. In that case, the IV estimator will also be inconsistent. Good discussions of the plausibility of the exclusion restriction should be appreciated.

h) The translation of the estimated effect of cathhs in for example col (6), into math score standard deviation is found by dividing the math score coefficient (3.95) by the standard deviation in math score from table 1, (9.45) which is approximately 0.42. Thus, according to the IV-2SLS results, students attending catholic high schools have 42% of a standard deviation higher test scores in math than students attending public schools. The corresponding number based on the OLS estimate in col (2) is 1.42/9.45= which amounts to 15% of a standard deviation.

i) Students should emphasize that the intuition is that under the hypothesis that cathhs is exogeneous and if a valid instrumental variable is available, the IV-2SLS and OLS estimators should be equal, because both estimators are then consistent. The hypothesis that cathhs is exogeneous can be tested along the lines described in Textbook section 15-5a.