This assignment counts for 20% of the module mark. Please read the guidelines for the assignment.

Use the dataset healthcare_b.dta, available from the module Canvas site. The dataset contains a balanced panel of German individuals and their health care use. The dataset has the following variables

ID = person - identification number
FEMALE = female = 1; male = 0
YEAR = calendar year of the observation
AGE = age in years
AGESQ = age squared.
HSAT = health satisfaction, coded 0 (low) - 10 (high)
HANDDUM = handicapped = 1; otherwise = 0
HANDPER = degree of handicap in percent (0 - 100)
HHNINC = household nominal monthly net income in German marks / 10000
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
MARRIED = married = 1; otherwise = 0
WORKING = employed = 1; otherwise = 0
BLUEC = blue collar employee = 1; otherwise = 0
WHITEC = white collar employee = 1; otherwise = 0
SELF = self employed = 1; otherwise = 0
BEAMT = civil servant = 1; otherwise = 0
DOCVIS = number of doctor visits in last three months
HOSPVIS = number of hospital visits in last calendar year
DOCTOR = dummy variable = 1 if *DOCVIS* > 0, 0 otherwise.
HOSPITAL = dummy variable = 1 if *HOSPVIS* > 0, 0 otherwise.
PUBLIC = insured in public health insurance = 1; otherwise = 0
ADDON = insured by add-on insurance = 1; otherwise = 0

a) Use the dataset to assess how much visits to the doctor over the last three months is affected by having children under 16 in the household (variable HHKIDS). Estimate a simple model using *DOCVIS* as the dependent variable and apply individual fixed effects. Include the following as controls: *AGE*, *AGESQ*, *MARRIED*, *WORKING*, *LINC* and *ADDON*. Generate LINC by taking a natural logarithm of HHNINC. Interpret your results for HHKIDS. [Notes: Some people visit the doctor a lot, but do not remove outliers. For fixed effects estimation, use either 'areg' or 'xtreg' command] [30 marks]

b) Estimate the same model as in a), but use pooled OLS instead of individual fixed effects. Interpret the estimated coefficient for HHKIDS in this model. Give an explanation why the results differ between OLS and fixed effects, and state which one should be preferred, and why. [40 marks]

c) Modify the estimation in a) to test whether women's visits to doctor are more or less affected by children in the household, than that of men. Interpret the result. [30 marks]