Assignment 1
Mohammad Manzoor Hassan Mirza

Question 1

Load the dataset HW-2.csv provided for this homework. This dataset is originated from the World Health Organization Panel Data on Health Care Attainment. It contains information on health care related variables in different countries, namely:

COUNTRY = number assigned to country;
YEAR = year associated to the observation;
LCOMP = log of a composite measure of health care attainment;
LDALE = log of the disability adjusted life expectancy;
LHEXP = log per capita health expenditure;
LHC = log educational attainment;
LGDPC = log normalized per capita GDP;
OECD = dummy variable for OECD country;
POPDEN = population density.

a) how many countries and how many years are there in this dataset? [5 points]

There are 140 countries in the dataset and 5 years.

b) use all these data and Ordinary Least Squares (OLS) to find how the per capita expenditure on health affects the composite measure of health care attainment. Run models with and without time dummies and interpret the magnitude of the coefficients obtained [15 points]

The magnitude of the coefficient for our explanatory variable, LHEXP i.e. per capita expenditure on health, is 0.029, which is statistically significant at the 0.1 significance level. The coefficient in the regression without time dummies indicates that on average, controlling for other factors, a 1% increase in per capita expenditure on health corresponds to a 0.029% increase in LCOMP i.e. composite measure of health care attainment, for non-OECD countries. Interpretation of other coefficients for control variables in the first column below remains the same, except that the magnitude of their correlation with the response variable differs. The coefficients for LGDPC and POPDEN are statistically insignificant!

The coefficient of LHEXP in the regression with time dummies indicates that on average, controlling for other factors, a 1% increase in per capita expenditure on health corresponds to a 0.029% increase in LCOMP i.e. composite measure of health care attainment in the year 1993, for non-OECD countries.

It is essential to note how inclusion of time dummies to capture time trends does not result in any change in the coefficient for our explanatory variable, LHEXP. This could possibly be due to the absence of any unobservable time trends in the error term, which are correlated with LHEXP and impact LCOMP. Alternatively, it is possible that the control variables included in the regression already account for factors which change the same way for all individuals over time, leaving behind no unobservable time trends in the error term.

Panel OLS

	Dependent variable:	
	LCOMP	
	no time dummies	time dummies
LHEXP	0.029*	0.029*
	(0.015)	(0.016)
LDALE	0.408***	0.408***
	(0.034)	(0.034)
LHC	0.030**	0.030**
	(0.013)	(0.013)
LGDPC	0.014	0.013
	(0.017)	(0.018)
POPDEN	-0.00000	-0.00000
	(0.00000)	(0.00000)
Constant	2.331***	2.332***
	(0.122)	(0.123)
OECD Dummies	Yes	Yes
Year Dummies	No	Yes
LDALE Control	Yes	Yes
LHC Control	Yes	Yes
LGDPC Control	Yes	Yes
POPDEN Control	Yes	Yes
Observations	700	700
R2	0.913	0.913
Adjusted R2	0.912	0.912
F Statistic	1,213.604*** (df = 6; 693)	724.023*** (df = 10; 689)
Note:		*p<0.1; **p<0.05; ***p<0.01

c) the coefficients obtained above do not represent the causal effect of health expenditure on the measure of health care attainment. Why? Can you suggest, and show the results, of two alternatives to improve these results? How do the results change when you follow these suggestions? [50 points]

The coefficients above do not represent the causal effect of health expenditure (X) on the measure of health care attainment (y) since the variable of interest i.e. X is not randomized here. This means that the β coefficient does not solely reflect how changes in the input variable i.e. X would cause a change in the outcome variable i.e. y, but it also captures how X affects y through unobservable factors, defined as the "back door" channel. This problem occurs when these unobservable factors, which impact y, are correlated with X. Moreover, this is an observational study where there is no control group to compare to.

The two alternatives that we could use to improve the results and get closer to the causal effect are: first differences regression or fixed effects regression, given time trends have already been dealt with using time dummies in the regression model. Both these approaches serve the same purpose i.e. to address the problem of omitted variables in panel data by removing unobserved unit level effects. However, even after accounting for time fixed effects i.e. factors which change the same way over time across individuals, and individual fixed effects i.e. factors which vary across individuals but do not change over time, we still do not reach the causal effect below. This is because we do not and cannot account for factors which change differently for different individuals over time.

When we follow these suggestions, we see that the coefficient for LHEXP drops from 0.029 to 0.0004 and 0.001 in the first-differences and fixed-effects model, respectively. It also becomes statistically insignificant at the 0.1 significance level. This implies that once we account for individual fixed effects, an increase in

healthcare expenditure no longer corresponds to an increase in the measure of health care attainment. In other words, there exists no correlation between the input and outcome variable here, as was mistakenly seen when we had only accounted for time trends above.

Panel OLS

	Dependent variable:			
	LCOMP			
	no time dummies	time dummies	first-differences	fixed-effects
LHEXP	0.029*	0.029*	0.0004	0.001
	(0.015)	(0.016)	(0.0004)	(0.001)
LDALE	0.408***	0.408***	0.580***	0.573***
	(0.034)	(0.034)	(0.031)	(0.035)
LHC	0.030**	0.030**	0.018**	0.022**
	(0.013)	(0.013)	(0.007)	(0.009)
LGDPC	0.014	0.013		
	(0.017)	(0.018)		
POPDEN	-0.00000	-0.00000		
	(0.00000)	(0.00000)		
Constant	2.331***	2.332***	0.001***	
	(0.122)	(0.123)	(0.0001)	
OECD Dummies	Yes	Yes	Yes	Yes
Year Dummies	No	Yes	Yes	Yes
LDALE Control	Yes	Yes	Yes	Yes
LHC Control	Yes	Yes	Yes	Yes
LGDPC Control	Yes	Yes	Yes	Yes
POPDEN Control	Yes	Yes	Yes	Yes
Observations	700	700	560	700
R2	0.913	0.913	0.906	0.930
Adjusted R2	0.912	0.912	0.905	0.911
F Statistic	1,213.604*** (df = 6; 693)	724.023*** (df = 10; 689)	893.106*** (df = 6; 553)	1,044.673*** (df = 7; 553)

Note: *p<0.1; **p<0.05; ***p<0.01