**Impact of social demographics on household earnings in the US**

**Name: Imisi Raphael Aiyetan**

**Course: Econometrics 512**

**Final Examination**

The dataset contains the following variables for a sample of individuals within the age of 37-45:

* Treat-1 if Year of Schooling is 15, 0 otherwise- This is the **treatment variable.** It reflects the number of years the individual spent in the school. The year falls between 12 and 19 inclusive.
* 1 if female, 0 if male. **Covariate**
* Ethnicity- This also distinguishes between black and Hispanic race. 1 if black, 0 otherwise and 1 If Hispanic, 0 otherwise. **Covariate**
* Total out-of-school work experience. **Covariate**
* Married-1 if married, 0 otherwise. **Covariate**
* Ssf-1 if father schooling is 15, 0 otherwise. **Instrument** variables
* Ssm-1 if mother schooling 15, 0 otherwise. **Instrument** variables
* Sib-1 if siblings are 3, 0 otherwise. I**nstrument** variables
* Earnings- the wage of the individual. This is the main **dependent variable** in this analysis

Data source: Earnings function of US for 2000.

Link to dataset: <https://www.macmillanihe.com/companion/gujarati-econometrics-by-example-2e/learning-resources/Data-sets/>

**Question 1:**



Where:

**Lnearnings**: the wage of the individual.

**Treatment**: Treat-1 if Year of Schooling is 15, 0 otherwise- This is the **treatment variable.** It reflects the number of years the individual spent in the school.

**Wexp**: Total out-of-school work experience. **Covariate**

**Female**: 1 if female, 0 if male. **Covariate**

**Ethblack:** 1 if black, 0 otherwise. **Covariate**

**Ethhisp:** 1 if Hispanic, 0 otherwise. **Covariate**

**Married:** 1 if married, 0 otherwise. **Covariate**

SS**f:** 1 if father schooling is 15, 0 otherwise. **Instrument** variable

**SSm**: if mother schooling 15, 0 otherwise. **Instrument** variable

**Sib**: 1 if siblings are 3, 0 otherwise. I**nstrument** variable

**Code:**

clear all

use "/Users/imisiaiyetan/Downloads/Table19\_4\_data.dta"

\* To generate the treatment variable

gen treatment = 0

replace treatment = 1 if s >=15

\* To generate the instrumental variables

gen ssf = 0

replace ssf = 1 if sf>=15

gen ssm = 0

replace ssm = 1 if sm>=15

gen sib = 1

replace sib = 0 if siblings > 3

**Question 2:**

This study considers Instrumental Variables (IV) analysis to estimate the effect of the treatment defined in question 1 on the outcome. Before proceeding to IV regression, the study starts by estimating the effect of treatment on outcome using Ordinary Least Square (OLS).

**Table 1: OLS without robust**



**Table 2: OLS with robust**

****

It is evident from the two results that there isn’t different between OLS with and OLS without robust. However, if we assumed that the treated variable is not randomly assigned, then the treatment and the error term are correlated (i.e. there is unobservables that correlate with the treatment). Therefore, to check if this is true, the study proceeds with IV regression analysis.

**Question 3:**

The study starts by estimating the IV-regression using Two Stages Least Square (TSLS) and alternatively estimate IV-regression using Generalized Method of Moment (GMM).

**Table 3: IV-Regression (2SLS)**



**Table 4: IV-Regression (GMM)**



To estimate the standard errors, I considered bootstrap method as the most appropriate method given the IV regression. First of all, the study estimates bootstrap IV-regression before it proceeds to estimate the standard errors.

**Table 5: Bootstrap IV-regression**



**Table 6: Standard Error Estimation**



**The reasons why the standard errors generated by bootstraps approach are the correct standard errors**

**Answer:**

Inference using conventional standard errors in 2SLS is based on an estimate of a moment that in finite samples often does not exist (as the coefficient has no finite variance when exactly identified). The bootstrap solved this problem by using resampling technique to estimate the percentiles of distributions, which always exist, and does much better. In that case, while asymptotic theory favors the resampling of the t-statistic, then avoiding finite sample 2SLS standard estimate altogether and focusing on the bootstrap resampling of the coefficient distribution alone provides the best performance, with tail rejection probabilities on IV coefficients that are very close to nominal size in iid, non-iid, low and high leverage settings.

**Interpretation of the treatment effect estimate**

**Answer:**

In this study, the treatment effect, which is the average causal effect of individuals receiving 15years free education on their earnings. The results as shown in Table 3 or Table 4 indicates that the treatment when account for endogeneity (e.g. siblings, father schooling and mother schooling) has a significant positive on earnings. Intuitively, the result shows that when an individual is incentivized to stay in school for 15 year or more than and he/she is from an educated background with not more than three siblings, individual earnings increase by 24 percent. In addition, the covariates (i.e. total work experience and female) are highly significant to explain earning behaviors. In terms of Average Treatment Estimation (ATE), the difference between the mean of the outcome for controls and the mean of the outcome for the treatment gives 30.3 (note: this value is calculated as the 24.7-0.7+7). In this case, it is plausible since I assumed the treatment is homogeneous and the whole population is affected by the instrumental variable.

**Table 7 IV-regression bias**

