

Problem Set 2 - Econometrics

Giacomo Lo Conte, Kun Wu, Francesca Eustacchi, Neeharika Kakunuri, Noor Ahmed Khoso

February 2, 2022

1. Theory

a. Consider a case where Z is a binary instrument, T is a discrete multi-valued treatment ($T \in 0, 1, 2, \dots, J$), $T = ZT_1 + (1-Z)T_0$ is the observed level of treatment, and Y is the outcome. Assume that the random variables $T_0, T_1, Y_0, Y_1, \dots, Y_J$ are jointly independent of Z and that the level of treatment is higher when $Z = 1$ than when $Z = 0$, i.e. $T_1 \geq T_0$. Further, assume that $Pr(T_1 \geq j \geq T_0) > 0$ for at least one $j \in 0, 1, 2, \dots, J$, which means that the instrument on average affects the level of treatment. Show that under these assumptions the Wald estimator is a weighted average of the causal response of a unit change in treatment,

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[T|Z=1] - E[T|Z=0]} = \sum_{j=1}^J w_j E[Y_j - Y_0 | T_1 \geq j > T_0], \quad (1)$$

where the weights are

$$w_j = \frac{Pr(T_1 \geq j > T_0)}{\sum_{i=1}^J Pr(T_1 \geq i > T_0)} \quad (2)$$

b) Provide an interpretation of both factors on the RHS of Equation (1), i.e. $E[Y_j - Y_0 | T_1 \geq j > T_0]$ and w_j . What does the presence of both factors teach us about the local average treatment effect?

c) In the lecture we have proven that in absence of defiers the IV estimate for an outcome Y_i , a binary treatment D_i and a binary instrument Z_i is

$$\beta^IV = E(Y_i | D_i = 1) - E(Y_i | D_i = 0) = E(Y_{i1} - Y_{i0} | complier) \quad (3)$$

Now suppose the share of defiers is $0 < \phi_D < 1$. i) Derive the IV estimate for this case. ii) Assume that $\phi_D = a\phi_C$ with $0 < a < 1$ and $\beta^IV > 0$. Discuss whether additional, plausible assumptions on $E(Y_{i1} - Y_{i0} | defier)$ allow you to recover a lower bound for β^IV .

2. Simulation Exercise

An important property of the IV estimator is that it is biased in small samples but consistent. For this reason one should never write that an IV estimator is used to obtain unbiased estimates. We want to better understand the small sample properties through simulations of the sampling distribution of IV and OLS estimators. In all simulations, let x , y , z , u and ϵ be random variables and assume the data-generating process:

$$y = \alpha + \beta x + \epsilon \quad (4)$$

$$x = \gamma_0 + \gamma_1 z + u \quad (5)$$

Set the parameter $\beta = 1$ and, unless required otherwise, $\alpha = \gamma_0 = 0$. Moreover, construct ϵ such that it's Pearson correlation with x is 0.4.

a) Sampling Distribution under strong instruments

Construct the instrument z such that its Pearson correlation with x is 0.5 while its correlation with ϵ is zero. Consider four different sample sizes: $N = 50$, $N = 100$, $N = 250$ and $N = 1000$. For each sample size, run at least 10,000 simulations whereby you estimate β^{OLS} and β^{IV} (choose fewer replications if you have problems with computing power). In the same graph, plot the sampling distributions for OLS and IV for all four sample sizes. Discuss the difference in shape of the sampling distributions between OLS and IV.

We constructed x , z and ϵ through a random extraction of 10,000 observations. The three variable are correlated as required. Notice that, once we defined these three, the value of γ_1 can be defined as $\rho_{xz} \frac{\sigma_x}{\sigma_z}$. Its value then is defined by the ratio of the standard deviations, that are $\sigma_x = \sigma_z = \sigma_\epsilon = 1$. In the second step, we derive $y = x + \epsilon$.

Among our population of 10,000 of units, we run a bootstrap (10,000 iterations) of our OLS and IV estimates over four different samples. Graphics in figure ?? show clearly how the distributions of the the IV and OLS estimates change when the sample size increases.

How we can see, as long as the sample size increases we see that the distribution variance decreases both for the OLS and IV estimator. This is the graphic representation of the consistency property. Moreover, we can see graphically also the demonstration of the efficiency property of the OLS. OLS indeed shows a consistent lower variance and the frequency of the median values rises faster. In other words, IV estimator distribution is flatter than the OLS one. However, it is evident that OLS distribution biased, contrary to the IV. Indeed, while the IV estimator distribution is centered on 1, the OLS distribution is centered

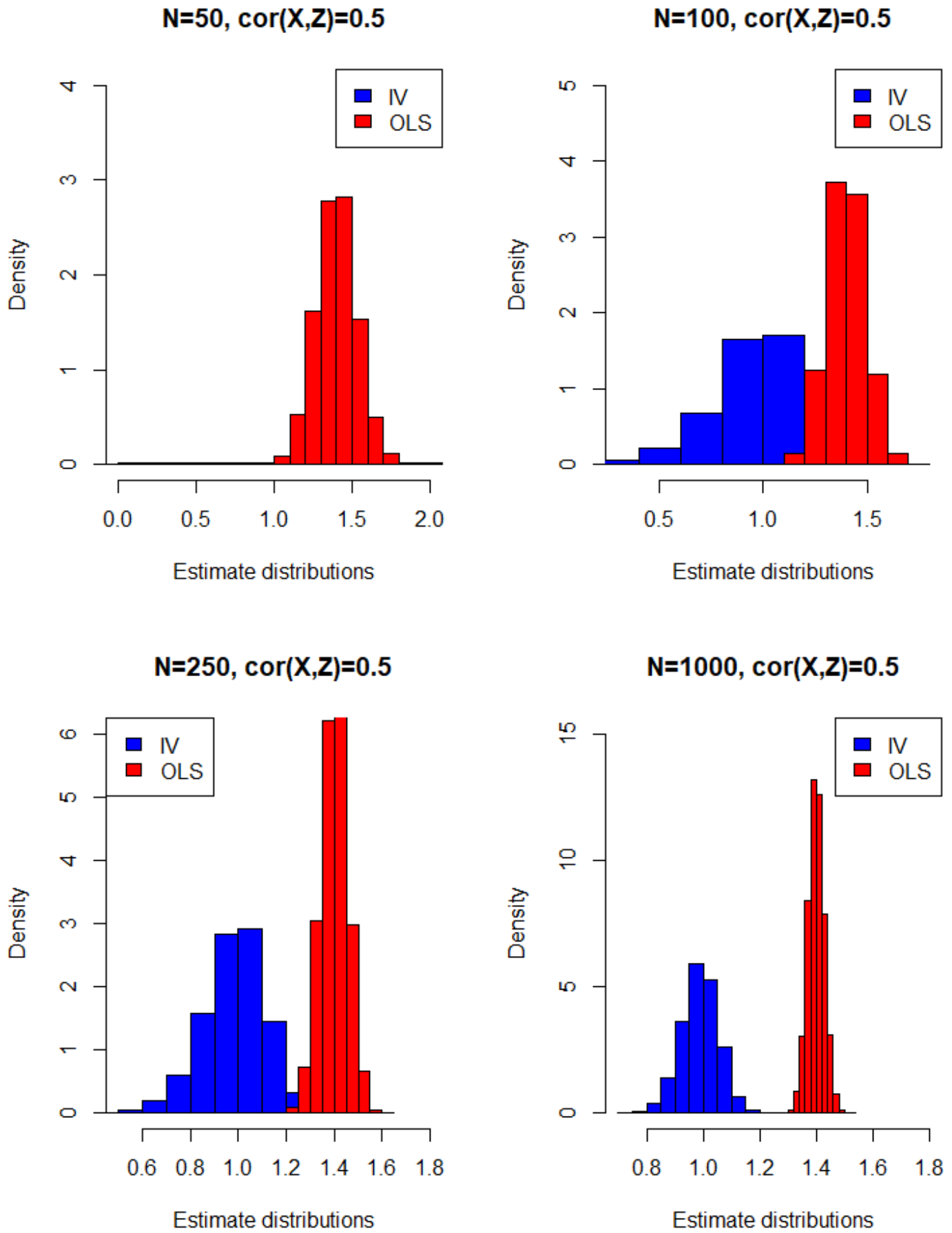


Figure 1: Distributions of IV and OLS estimates when $\rho_{XZ} = 0.5$

around 1.4. The bias is positive and is exactly equal to $\frac{Cov(x, \epsilon)}{\sigma_x^2}$. Since $Cov(x, \epsilon) = \rho_{x\epsilon}\sigma_x$, the bias is equal to $\frac{\rho_{x\epsilon}}{\sigma_x} \approx 0.4$. From these graphs, then it is clear that, in a non-biased setting the IV estimates may have more problems with the statistical inference. The IV errors are indeed larger than the OLS.

b) Weaker instruments:

Now repeat the analysis from a), but instead assume that the correlation between the instrument z and the regressor x is 0.15. Discuss the difference in sampling distributions between OLS and IV and, in addition, discuss the difference between sampling distributions based on weak and strong IVs.

In figure ?? we can hardly observe the OLS distribution for a problem of scale, however it remains basically unchanged. In the IV case, we can notice a very different scenario. The IV distribution is much flatter than the previous case. To fully grasp the difference we can observe the scale of the axes. While in figure ??, the top-right graph presents an interval between 0.3 and 1.7, in figure ?? the same graph the interval is between -1000 and 600. This is due to the formula of the variance for the IV estimator. The asymptotic variance of the IV is:

$$\widehat{Avar}(\beta^{IV} - \beta | X, Z) = E[(Z'X)^{-1}Z'uu'Z(Z'X)^{-1} | X, Z]$$

$$\widehat{Avar}(\beta^{IV} - \beta) = E[uu'](Z'X)^{-1} = \Sigma_u^2(Z'X)^{-1}$$

The asymptotic variance of the IV estimator can be written as the variance matrix of the error times the inverse of the covariance of Z and X. In our simple univariate case $Z'X = Cov(x, z) = \rho_{xz}\sigma_z\sigma_x$. The variance of the estimator, then, is inversely related to ρ_{xz} . For this reason, the weaker is the instrument and the higher is the variance of the IV estimator. This leads to problems for the statistical inference since the results are not significant even if unbiased.

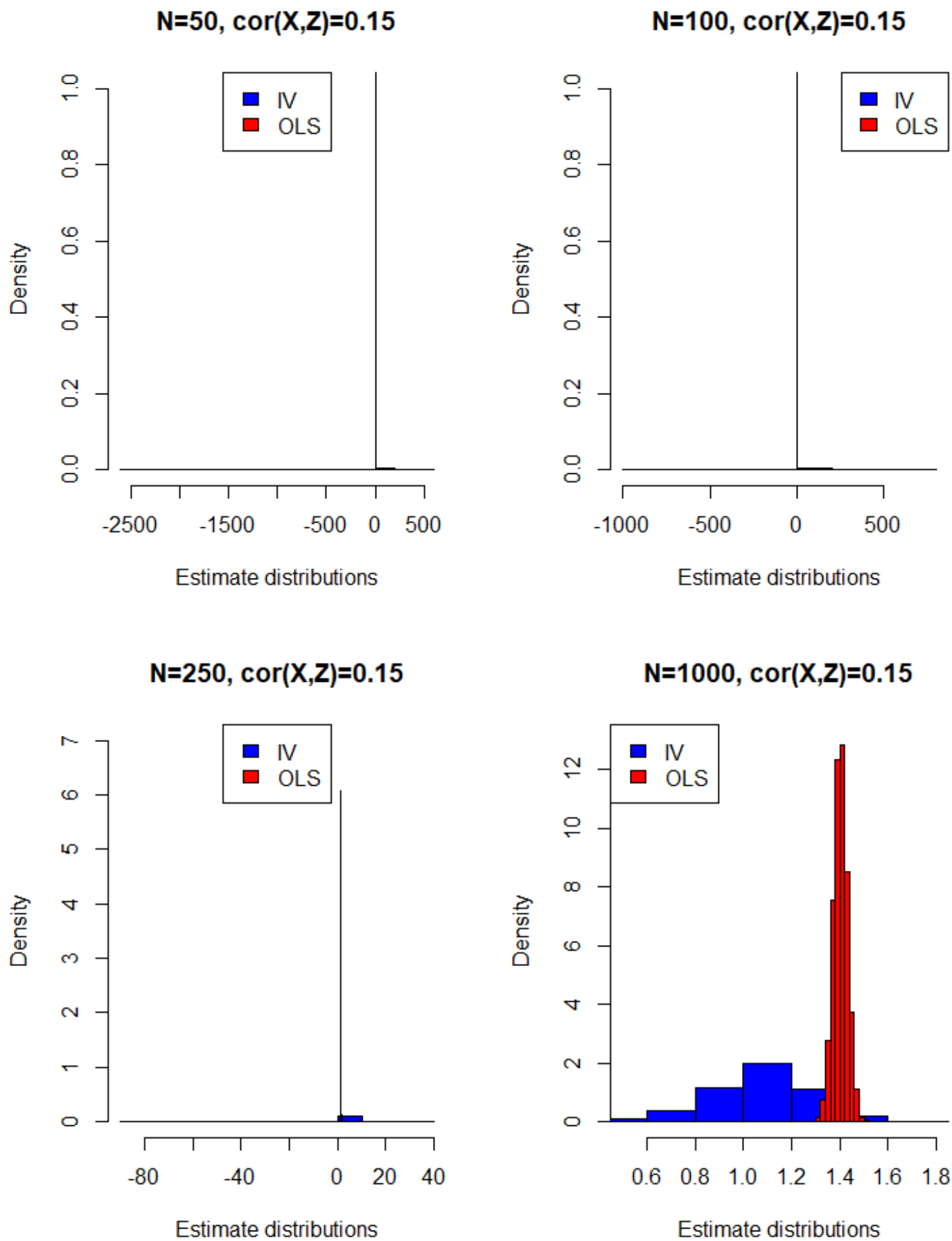


Figure 2: Distributions of IV and OLS estimates when $\rho_{XZ} = 0.15$

Empirical Application

The empirical application is based on a cross-sectional dataset `assign2.dta`, which contains the following variables:

- *age*: age of surveyed individual
- *logearn*: log annual earnings
- *yob*: year of birth
- *schooling*: age at which the person left school.

a) The goal is to estimate the returns to education. For this purpose, estimate an OLS regression of *logearn* on *schooling*, controlling for fourth-order polynomials in age and year of birth. Interpret the coefficient of *schooling*. A common way to obtain causal estimates is to use changes in compulsory schooling laws for identification. In this case, birth cohorts born before 1933 ($yob < 33$) had to go to school until they were 14 years old, whereas compulsory schooling age was raised to 15 years for all cohorts born from 1933 onwards. This change in compulsory schooling laws can be used as an instrumental variable for the actual duration of schooling. The instrument is a dummy *LAW* that equals unity if a person is born 1933 or later and zero otherwise.

Benchmark regression :

b) Discuss this instrument in theory, assuming that schooling S_i is related to the instrument Z_i by the latent assignment mechanism

$$S_i = 1(\gamma_0 + \gamma_1 Z_i > \eta_i), \text{ with } E(Z_i \eta_i) = 0$$

The random variable η_i represents the individual resistance to treatment. Why could there be a first stage? Under what condition is this instrument valid? What are potential threats to validity? 2 Furthermore, explain who are the compliers, always-takers and never-takers in this case. Would the IV estimate correspond to the average treatment effect (why or why not)? c) As in any good empirical project, begin with a graphical inspection of the relationships of interest. This is best done through *binscatters*. Produce and discuss the graphs listed below. In all graphs, include a vertical line at $yob = 33$.

- Plot the probability that a person leaves school before age 15 against the year of birth.
- Binscatter of *schooling* and year of birth
- Binscatter of log earnings and year of birth.

Table 1: Benchmark Regression - tbd

	<i>Dependent variable:</i>
	logearn
age	0.005*** (0.001)
yob	0.005*** (0.001)
schooling	0.157*** (0.002)
Constant	2.964*** (0.055)
Observations	30,801
R ²	0.140
Adjusted R ²	0.140
Residual Std. Error	0.496 (df = 30797)
F Statistic	1,673.896*** (df = 3; 30797)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression with the 4th order polynomial for age and year of birth

	<i>Dependent variable:</i>
	logearn
age	0.005*** (0.001)
yob	0.005*** (0.001)
schooling	0.157*** (0.002)
Constant	2.964*** (0.055)
Observations	30,801
R ²	0.140
Adjusted R ²	0.140
Residual Std. Error	0.496 (df = 30797)
F Statistic	1,673.896*** (df = 3; 30797)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

d) Calculate the Wald estimator (without controls) "by hand", i.e. based on conditional averages. Compare your results to those of a 2SLS estimation based on an inbuilt command (e.g. *ivregress* in Stata or *ivreg* in R). Interpret your results and compare them to the OLS results in a).

Table 3: Regression of D on Z

	<i>Dependent variable:</i>
	schooling
LAW	0.990*** (0.014)
Constant	14.624*** (0.012)
Observations	30,801
R ²	0.137
Adjusted R ²	0.136
Residual Std. Error	1.146 (df = 30799)
F Statistic	4,868.939*** (df = 1; 30799)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression of Y on Z

	<i>Dependent variable:</i>
	logearn
LAW	0.162*** (0.007)
Constant	5.671*** (0.005)
Observations	30,801
R ²	0.019
Adjusted R ²	0.019
Residual Std. Error	0.530 (df = 30799)
F Statistic	607.678*** (df = 1; 30799)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

The manual Wald estimator is 0.1633267.

Table 5: IVREG with R code

	<i>Dependent variable:</i>
	logearn
schooling	0.163*** (0.006)
Constant	3.283*** (0.095)
Observations	30,801
R ²	0.138
Adjusted R ²	0.138
Residual Std. Error	0.497 (df = 30799)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01