**2.1**

This study to investigate the effect of property rights on economic outcomes. Here loggpgp95 is the log of GDP per capita in 1995, which is economic outcome or dependent variable.

Correlation between dependent and independent variables.

R console : correlation coefficients among variables used in study

|  |
| --- |
| > cor(ps5[,3:6], use="complete.obs", method="kendall")  euro1900 avexpr logpgp95 logem4  euro1900 1.0000000  avexpr 0.2606533 1.0000000  logpgp95 0.5646801 0.5370499 1.0000000  logem4 -0.4850327 -0.3138626 -0.4886933 1.0000000 |

Here it is seen that the variables avexpr and logem4 are negatively correlated

Estimated regression equations are

Full model r1

Omitted model r2

Where

R console : full and omitted linear model

|  |
| --- |
| > r1<-lm(logpgp95~avexpr+logem4, data=ps5)  > summary(r1)  Call:  lm(formula = logpgp95 ~ avexpr + logem4, data = ps5)  Residuals:  Min 1Q Median 3Q Max  -2.11173 -0.30105 0.05473 0.40377 1.13805  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 7.19487 0.58769 12.243 < 2e-16 \*\*\*  avexpr 0.36131 0.05579 6.476 1.30e-08 \*\*\*  logem4 -0.32019 0.06382 -5.017 4.11e-06 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.6 on 67 degrees of freedom  (93 observations deleted due to missingness)  Multiple R-squared: 0.6885, Adjusted R-squared: 0.6792  F-statistic: 74.05 on 2 and 67 DF, p-value: < 2.2e-16  > r2<-lm(logpgp95~avexpr, data=ps5)  > summary(r2)  Call:  lm(formula = logpgp95 ~ avexpr, data = ps5)  Residuals:  Min 1Q Median 3Q Max  -1.9020 -0.3160 0.1380 0.4225 1.4406  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 4.62609 0.30058 15.39 <2e-16 \*\*\*  avexpr 0.53187 0.04062 13.09 <2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.7179 on 109 degrees of freedom  (52 observations deleted due to missingness)  Multiple R-squared: 0.6113, Adjusted R-squared: 0.6078  F-statistic: 171.4 on 1 and 109 DF, p-value: < 2.2e-16 |

These above 2 models say that the both are significant and parameters also significant. So the model will be look like as given below

Fitted Full model r1

Fitted Omitted model r2

We see when *logem4* omitted from the model, the effect of *avexpr* is over-estimate.

Now see the impact of instrumental variable on regressed while independent variable avexpr is exist in the model.

R console : regression model with instrument variable

|  |
| --- |
| > iv1<-ivreg(logpgp95~avexpr|logem4, data=ps5)  > summary(iv1, vcov = sandwich, diagnostics = TRUE)  Call:  ivreg(formula = logpgp95 ~ avexpr | logem4, data = ps5)  Residuals:  Min 1Q Median 3Q Max  -2.28175 -0.55059 0.03401 0.62273 1.57418  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 2.3702 0.9283 2.553 0.0129 \*  avexpr 0.8684 0.1365 6.361 1.97e-08 \*\*\*  Diagnostic tests:  df1 df2 statistic p-value  Weak instruments 1 68 26.56 2.38e-06 \*\*\*  Wu-Hausman 1 67 24.09 6.20e-06 \*\*\*  Sargan 0 NA NA NA  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.8899 on 68 degrees of freedom  Multiple R-Squared: 0.3045, Adjusted R-squared: 0.2942  Wald test: 40.46 on 1 and 68 DF, p-value: 1.972e-08 |

Weak instruments: This is an F-test on the instruments in the first stage. The null hypothesis is essentially that we have weak instruments, so a rejection means our instruments are not weak, which is good.

Again see the impact of instrument variable while a control variable also exists in the model. Here lat\_abst is the control variable.

R console : regression model with instrument and control variable

|  |
| --- |
| > iv2<-ivreg(logpgp95~avexpr+lat\_abst|lat\_abst+logem4, data=ps5)  > summary(iv2, vcov = sandwich, diagnostics = TRUE)  Call:  ivreg(formula = logpgp95 ~ avexpr + lat\_abst | lat\_abst + logem4,  data = ps5)  Residuals:  Min 1Q Median 3Q Max  -2.62939 -0.76718 0.01379 0.82551 2.03605  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 1.652 1.515 1.090 0.279497  avexpr 1.029 0.263 3.913 0.000215 \*\*\*  lat\_abst -1.784 1.530 -1.166 0.247652  Diagnostic tests:  df1 df2 statistic p-value  Weak instruments 1 67 9.269 0.00333 \*\*  Wu-Hausman 1 66 18.793 5.09e-05 \*\*\*  Sargan 0 NA NA NA  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 1.038 on 67 degrees of freedom  Multiple R-Squared: 0.06701, Adjusted R-squared: 0.03915  Wald test: 19.39 on 2 and 67 DF, p-value: 2.276e-07 |

Weak instruments: This is an F-test on the instruments in the first stage. The null hypothesis is essentially that we have weak instruments, so a rejection means our instruments are not weak, which is good.

R console : regression model with control and two instrument variable.

|  |
| --- |
| > iv3<-ivreg(logpgp95~avexpr+lat\_abst|lat\_abst+logem4+euro1900, data=ps5)  > summary(iv3, vcov = sandwich, diagnostics = TRUE)  Call:  ivreg(formula = logpgp95 ~ avexpr + lat\_abst | lat\_abst + logem4 +  euro1900, data = ps5)  Residuals:  Min 1Q Median 3Q Max  -2.516402 -0.658570 -0.002267 0.769096 1.886633  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 1.9755 1.0480 1.885 0.0638 .  avexpr 0.9735 0.1785 5.455 7.9e-07 \*\*\*  lat\_abst -1.6294 1.1228 -1.451 0.1514  Diagnostic tests:  df1 df2 statistic p-value  Weak instruments 2 65 10.162 0.000144 \*\*\*  Wu-Hausman 1 65 26.147 3.01e-06 \*\*\*  Sargan 1 NA 0.105 0.746008  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.9819 on 66 degrees of freedom  Multiple R-Squared: 0.1599, Adjusted R-squared: 0.1345  Wald test: 27.52 on 2 and 66 DF, p-value: 2.03e-09 |

Sargan: This is a test of instrument exogeneity using overidentifying restrictions, called the J-statistic in Stock and Watson. It can only be used if you have more instruments than endogenous regressors, as we do in iv3. Sargan test suggest that the null hypothesis is accepted, it means that our instruments are valid.

**2.2**

Regression discontinuity on the data of students from the schools in New York City during the year 2009-2013. 4053 observations from several schools which contains the grades and year.

R console : regression discontinuity

|  |
| --- |
| > RDestimate(averageclasssize~x|grade, data=ps5rd, bw=10)  Call:  Rdestimate(formula = averageclasssize ~ x | grade, data = ps5rd, bw = 10)  Coefficients:  LATE Half-BW Double-BW  -3.140 -2.441 -4.202 |

Using a bandwidth of 10, the estimated marginal average treatment effect is -3.140. The figure below illustrates the discontinuity:

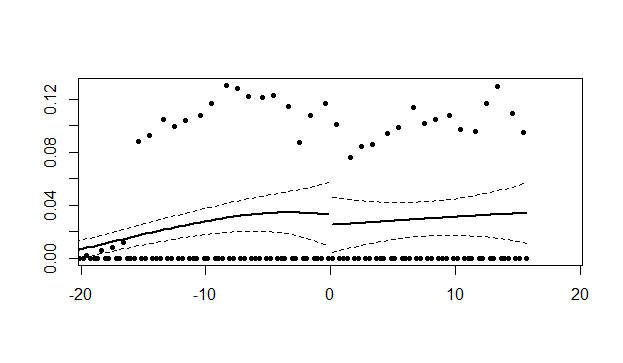


Figure : discontinuity in the average class size

The figure above shows that there is shift of intercept is much easier to see then the change in slope as it seems equal in both side of zero.

|  |
| --- |
| call:  RDestimate(formula = mathscore ~ x + averageclasssize | enroll +  pctblack + pctwhite + pcthisp + pctdisability + pctEL, data = ps5rd,  cutpoint = 0, bw = 10)  Type:  fuzzy  Estimates:  Bandwidth Observations Estimate Std. Error z value Pr(>|z|)  LATE 10 2534 -0.02350 0.017580 -1.3368 0.18128  Half-BW 5 1176 -0.00369 0.033877 -0.1089 0.91326  Double-BW 20 4027 -0.01888 0.009599 -1.9671 0.04917 \*  ---  Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  F-statistics:  F Num. DoF Denom. DoF p  LATE 409.4 9 2524 0.000e+00  Half-BW 193.1 9 1166 6.217e-224  Double-BW 659.8 9 4017 0.000e+00 |

Using a bandwidth of 10, the estimated marginal average effect of class size on student test score is -0.0235.