

Problem Set 1.2

Labour Economics, Winter Term 2025/26

Submit by Sunday, 16 November, 22:45h **on Moodle!**

Learning objectives

- Conduct instrumental variables (IV) regressions to address OVB.
- Conduct IV regressions to address measurement error.
- Interpret experimental data and compare results to OLS.

Tasks

Download the data `ps1_more_realistic_data.Rda` and open it in R Studio. Reality is often messier than what we saw in *Problem Set 1.1*. We commonly have variables measured with error, such as *ln_wage_ME* or *hours_ME*, and lack information on important factors (*motivation*). We will use the randomly assigned wage increases *wage_premium* to try to address these issues. As said before, these could be thought of as individuals being drafted into an income support program like the SSP in British Columbia / New Brunswick (Canada) that we discussed in lecture.

- a) Repeat part of the descriptive statistics (those you find informative) from questions a)–c) in *Problem Set 1.1*. What differences can you notice in the data?

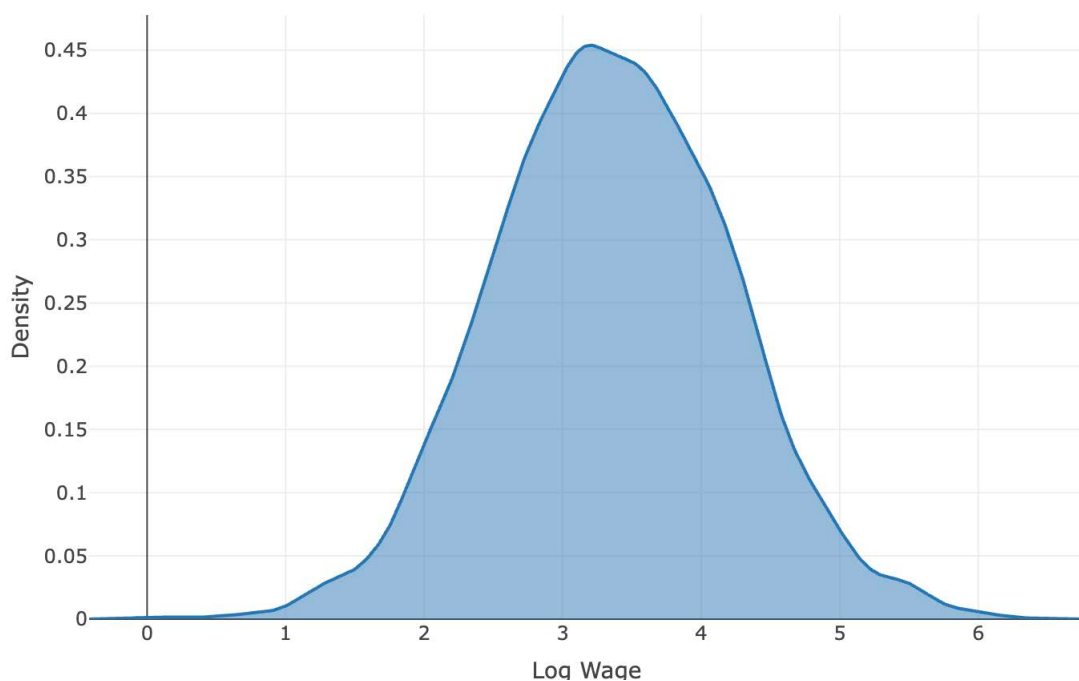
Solution

For better comparison, the first table shows the data from 1.1. The second table shows our new data, now with measurement error.

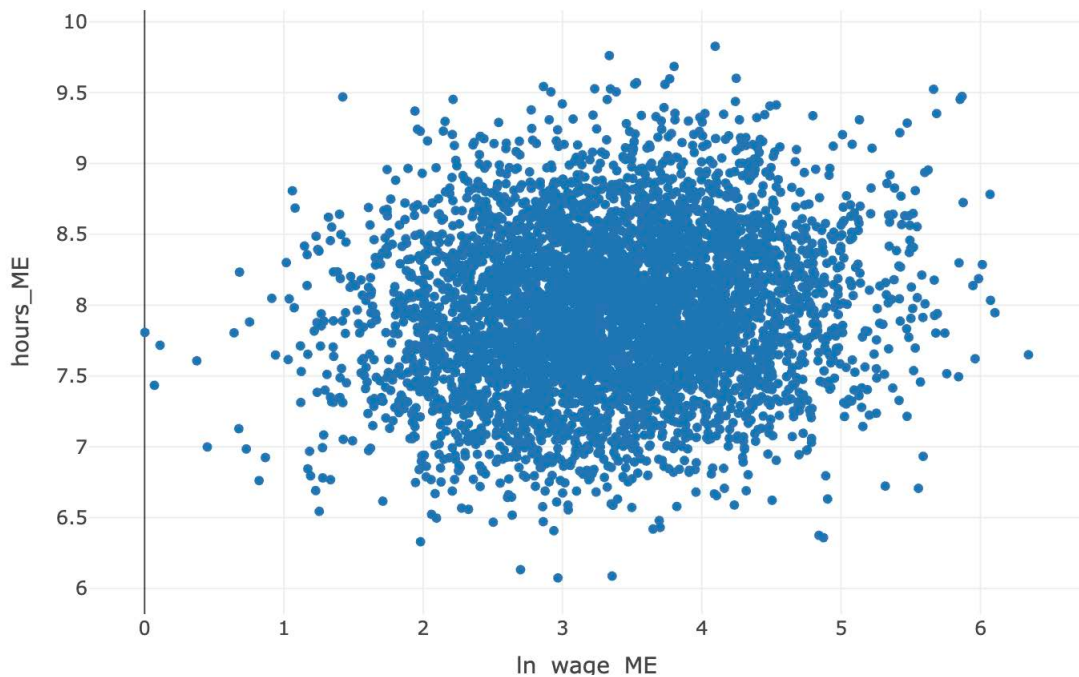
variable	n	min	max	median	iqr	mean	sd	se	ci
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
education	5000	10	24	12	4	12.8	2.42	0.034	0.067
motivation	5000	-4.56	5.14	-0.002	1.88	0.014	1.39	0.02	0.039
hours	5000	6.35	9.70	7.95	0.666	7.95	0.489	0.007	0.014
ln_wage	5000	2.37	4.12	3.35	0.317	3.36	0.23	0.003	0.006

variable	n	min	max	median	iqr	mean	sd	se	ci
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
education	5000	10	24	12	4	12.8	2.42	0.034	0.067
hours_ME	5000	6.07	9.83	7.94	0.75	7.95	0.549	0.008	0.015
ln_wage_ME	5000	0.002	6.34	3.35	1.17	3.36	0.855	0.012	0.024

- education: Exactly the same. Education is not effected by our measurement error, as it is easy to measure.
- hours_ME: When we compare our hours_ME to our hours from 1.1, we can see greater min and max as well as higher standard deviation (sd). The latter is expected with measurement error...
- ln_wage_ME: We can observe more extreme max and min in our data with measurement error. The mean is slightly different. The sd changed from 0.23 to 0.855.
- motivation is very hard to measure and which is why it is not in the dataset anymore.



- Our new density plot shows us that the data is significantly more dispersed than in 1.1 (notice the different scales on the x-axis). The highest point now has a density of only 0.45 instead of over 1.6 previously and is significantly more spread out (from -0.5 to 7).



- The scatter plot is much more scattered than in 1.1. The correlation can only be guessed, it is by far not as clear as before.

b) Run again the regression of *hours* on *ln_wages*. What is the difference now that you only have measurement-error versions of these variables?

Solution

Regression of *hours* on *ln_wages*:

regression using clean data

Call:
lm(formula = hours ~ ln_wage, data = df1)

Residuals:

Min	1Q	Median	3Q	Max
-1.33182	-0.27013	0.00022	0.26766	1.47676

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.63582	0.08071	45.05	<2e-16 ***
ln_wage	1.28478	0.02396	53.61	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.39 on 4998 degrees of freedom
Multiple R-squared: 0.3651, Adjusted R-squared: 0.365
F-statistic: 2875 on 1 and 4998 DF, p-value: < 2.2e-16

new regression with ME

Call:
lm(formula = hours_ME ~ ln_wage_ME, data = df2)

Residuals:

Min	1Q	Median	3Q	Max
-1.86052	-0.37018	-0.00967	0.36631	1.81559

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.651493	0.031170	245.475	<2e-16 ***
ln_wage_ME	0.088266	0.008999	9.808	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5443 on 4998 degrees of freedom
Multiple R-squared: 0.01888, Adjusted R-squared: 0.01869
F-statistic: 96.2 on 1 and 4998 DF, p-value: < 2.2e-16

The coefficient on *ln_wages* using the clean data was 1.28 now with measurement error the labour supply functions are very weak (coefficient of 0.088 or a 0.0088 hours = 31,7 seconds daily work increase per ten percent higher wage). Coefficient remains statistically significant though while R^2 is low.

- c) Before using the wage premium as a quasi-experiment, what check would you like to do?
Does the data pass this check?

Solution

One test we should do is to test for balancedness between the people who get the wage_premium and those who do not.

Balancedness checks are on pre-determined (before the treatment) or unchangeable features of the units analysed. Here the only such variable is education, since ln_wages and hours worked will be endogenous to the wage top-up.

wage_premium	variable	n	min	max	median	iqr	mean	sd	se	ci
<int>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0	education	3289	10	22	12	4	12.7	2.40	0.042	0.082
0	hours_ME	3289	6.07	9.83	7.91	0.741	7.92	0.548	0.01	0.019
0	ln_wage_ME	3289	0.002	6.34	3.23	1.15	3.24	0.84	0.015	0.029
1	education	1711	10	24	13	3	12.8	2.48	0.06	0.117
1	hours_ME	1711	6.36	9.60	7.99	0.764	8.00	0.548	0.013	0.026
1	ln_wage_ME	1711	0.642	6.10	3.58	1.14	3.59	0.837	0.02	0.04

Two Sample t-test

data: education by wage_premium

t = -0.96746, df = 4998, p-value = 0.3334

alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0

95 percent confidence interval:

-0.21159002 0.07175903

sample estimates:

mean in group 0 mean in group 1

12.74247 12.81239

We can see that both groups' mean of education is very similar (12.7 and 12.8). Accordingly, our formal significance test fails to reject the null hypothesis that they are the same (t-statistic: 0.9675, p-value: 0.3334). With 4998 degrees of freedom and our α of 0.05 our critical value is 1.96. $0.9675 \leq 1.96$ means we can't reject our H_0 . Alternatively, we could look at the 95% confidence interval, which is between -0.2 and 0.07. For a rejection of our H_0 , our confidence interval has to be different from 0. But here we can not say, that the true difference between Group 0 and 1 is not equal to 0).

This is in line with the assumption that the wage top up is quasi-random and therefore uncorrelated with workers' characteristics. Given successful randomization, this lack of correlation extends to unmeasured characteristics (here: the worker's motivation), too.

As an alternative to directly controlling for (unobserved) motivation, the successful quasi-experiment therefore solves the OVB problem as we see next.

- d) Do a balancedness check on *motivation* using the Data from PS1.1. Is this in line with your check from c)?

Solution

Two Sample t-test

```
data: motivation by wage_premium
t = -0.61427, df = 4998, p-value = 0.5391
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -0.10675720  0.05581737
sample estimates:
mean in group 0 mean in group 1
  0.00505130      0.03052121
```

We see the same as in c). The difference between Groups is very small and our t-test can't be rejected, so we seem to have balancedness of treatment and control group.

- e) Use the wage premium as an instrumental variable for log wages. What do you see; interpret also the first-stage regression.

Solution

Now with our IV *wage_premium*:

First Stage

```
Call:
lm(formula = ln_wage_ME ~ wage_premium, data = df2)

Residuals:
    Min       1Q   Median       3Q      Max
-3.2329 -0.5703 -0.0064  0.5727  3.1092

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.23515    0.01463   221.17  <2e-16 ***
wage_premium  0.35412    0.02501   14.16  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8389 on 4998 degrees of freedom
Multiple R-squared:  0.03858,    Adjusted R-squared:  0.03839
F-statistic: 200.6 on 1 and 4998 DF, p-value: < 2.2e-16
```

IV Regression

```
Call:
ivreg(formula = hours_ME ~ ln_wage_ME | wage_premium, data = df2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.946754 -0.374956 -0.008143  0.367165  1.977099

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.15818    0.15802   45.298  < 2e-16 ***
ln_wage_ME    0.23525    0.04702    5.003  5.85e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic tests:
            df1  df2 statistic p-value
Weak instruments    1 4998    200.56 < 2e-16 ***
Wu-Hausman         1 4997    10.72 0.00106 **
Sargan             0  NA         NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5587 on 4998 degrees of freedom
Multiple R-Squared: -0.03348,    Adjusted R-squared: -0.03369
Wald test: 25.03 on 1 and 4998 DF, p-value: 5.847e-07
```

The first stage F-statistic is 200.6, i.e., way above the rule-of-thumb value of 10. This shows the relevance of our instrument *wage_premium* for *ln_wage*, so the first condition for valid IV is satisfied. Coefficient of 0.35 log points indicates an approximately 40% higher wage on average for those workers who get the wage top-up. Given the standard error of 0.025 this sees within the range of sampling variation, i.e. not significantly different from the “truth” of 35% wage top-up.

In the second stage, the coefficient on `wage_premium` rises, from 0.088 to 0.235. The difference is also statistically significant, e.g., as shown in Wu-Hausman statistic which tests the difference between the OLS and the IV.

- f) From b) and e), what are the differences to question d) in *Problem Set 1.1*? Is this what you would expect (*hint*: consider also our lecture notes on IV)?

Solution

Clearly the IV delivers a different result from the OLS in b). The IV validity conditions are

1. Relevance, also called existence of a first stage. Verified in e).
2. Exogeneity, also called randomizer: we said that people got the wage top-up randomly and saw balancedness w.r.t. education (in actuality can't do the check in d) w.r.t. unobserved motivation).
3. Exclusion restriction, implying that the top-up does not directly affect hours worked but only via the wage. Can be debated but seems plausible (it is a wage top-up affecting the wage).

The IV is therefore arguably valid and solves two problems here: 1) measurement error as discussed in lecture 2b and 2) omitted variables bias (OVB) as discussed in lecture 2c. This suggests we have identified a causal effect. Wow!

OLS from b) seems problematic for exactly those two reasons: variables like `education` or `motivation` are not included, and there is measurement error in the wages... [Aside: R^2 comparison not helpful here, as discussed in lecture 2c, since R^2 hard to interpret in IV specifications]

Comparing to d) from problem set 1.1, which had a coefficient on `ln_wages` of 1.28, way higher. The problem there was an OVB of not including `motivation` in the regression, whereas here we do not need to do this as our instrument is randomly assigned and thus uncorrelated with `motivation`.

- g) Add *Education* as a control variable to the IV estimation. Do you obtain substantively different results and why (not)?

Solution

Adding education to our IV regression:

First Stage	IV Regression
Call: lm(formula = ln_wage_ME ~ wage_premium + education, data = df2)	Call: ivreg(formula = hours_ME ~ ln_wage_ME + education wage_premium + education, data = df2)
Residuals: Min 1Q Median 3Q Max -3.13711 -0.56294 -0.00371 0.55705 3.14651	Residuals: Min 1Q Median 3Q Max -1.820363 -0.320934 -0.007087 0.320817 1.917858
Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 2.594661 0.063385 40.94 <2e-16 *** wage_premium 0.350608 0.024745 14.17 <2e-16 *** education 0.050264 0.004843 10.38 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 5.723029 0.114740 49.878 < 2e-16 *** ln_wage_ME 0.211836 0.040317 5.254 1.55e-07 *** education 0.118571 0.003452 34.352 < 2e-16 *** --- Diagnostic tests: df1 df2 statistic p-value Weak instruments 1 4997 200.8 < 2e-16 *** Wu-Hausman 1 4996 22.1 2.66e-06 *** Sargan 0 NA NA NA --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.8301 on 4997 degrees of freedom Multiple R-squared: 0.05887, Adjusted R-squared: 0.05849 F-statistic: 156.3 on 2 and 4997 DF, p-value: < 2.2e-16	Residual standard error: 0.4742 on 4997 degrees of freedom Multiple R-Squared: 0.2556, Adjusted R-squared: 0.2553 Wald test: 1108 on 2 and 4997 DF, p-value: < 2.2e-16

This doesn't substantively change the relevant first and second-stage coefficient because wage top-up is random; thus not correlated with education and there is no OVB. All coefficients remain statistically significant and the first-stage sufficiently strong for the relevance condition.

- h) Now compare your IV results to the full regression f) from *Problem Set 1.1*. What is your preferred estimate of the elasticity of labour supply in this data?

Solution

Finally, compare this to the OLS in the full regression of Problem Set 1.1 f):

OLS	Effects
Call: lm(formula = hours ~ ln_wage + education + motivation, data = df1)	# print effects cat("% effect of ln_wage at the average of hours:", average_effect_ln_wage, "\n") effect of ln_wage at the average of hours: 0.02639629 cat("% effect of ln_wage at the average of hours:", average_effect_education, "\n") effect of ln_wage at the average of hours: 0.01256543 cat("% effect of ln_wage at the average of hours:", average_effect_motivation, "\n") effect of ln_wage at the average of hours: 0.03143381
Residuals: Min 1Q Median 3Q Max -0.36675 -0.06728 0.00121 0.06598 0.33396	
Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 5.9681690 0.0223144 267.46 <2e-16 *** ln_wage 0.2099211 0.0075306 27.88 <2e-16 *** education 0.0999287 0.0006756 147.91 <2e-16 *** motivation 0.2499828 0.0010803 231.39 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
Residual standard error: 0.09925 on 4996 degrees of freedom Multiple R-squared: 0.9589, Adjusted R-squared: 0.9589 F-statistic: 3.886e+04 on 3 and 4996 DF, p-value: < 2.2e-16	

The IV results in line with the ones we got in the previous problem set: coefficient of 0.212 in g) or 0.235 in e) very similar and not statistically different from 0.210 in PS1.1 f).

So either we have perfect measures of all relevant variables and we get the correct results from OLS. Or we have a valid instrument and can deal with issues like unobserved variables (here motivation) or measurement error by running IV estimations.

Standard errors in the IV are higher and statistical significance is lower in the IV, however, because we use less variation (only from wage_premium) and have more error in the regressor log wage. This is usually the cost of using IV, although the relevant estimate is still precise / highly significant.

The parameter estimate is reassuringly around $\frac{\partial hours}{\partial \log(wage)} = 0.20$ in both cases (this is how we actually simulated the data). But this is only a semi-elasticity. Dividing by average hours of 7.95, we get the labour supply elasticity at the mean of hours in the data:

$$\frac{\partial hours}{\partial (wage)} \frac{wage}{hours} = 0.026$$

When wages increase by 10 percent, hours worked per day increase by 0.27 percent. This elasticity seems low, even compared to what the literature finds for men (0.1 according to lecture 2b).

Notes: You can work in teams of 1–3 students. Please upload your code as well as a pdf-file with discussions on what you found in the data in response to the tasks above. It should be clear which lines of code and answers in the .pdf refer to which question.