Econometrics 1 - Problem Set 2 - Group 15

Felipe Montealegre[1], Paritosh Junare[2], and Ketki Balyan[3]

[1,2,3]Alma Mater Studiorum - Università di Bologna

October 2021

A Preliminary Tasks

1. In the table below we first report the average values (along with standard deviations) for each of the target variables, both over the entire sample and over the means of each of the states. Then we report the states that have the minimum and maximum average value for each of the target variables along with its values.

		Mean value (SD)		Minimum		Maximum	
		Overall	**Over-states Means**	**State**	**Value**	**State**	**Value**
Variables	**Exposure to Robots**	1.800013 (1.101769)	1.852853 (0.9572189)	NV	0.5719783	MI	5.246422
	Employment	0.8113613 (3.188102)	.3809222 (2.482054)	NC	-4.594934	ND	7.900403

2. No, the states with the highest and lowest variation in *expos_to_robots* are MI and NV respectively. However, the states with the largest and smallest variation in the outcome variable, i.e. *d_emppriv_1990_2011* are ND and NC respectively.

3. The average of the outcome variable, i.e., *d_emppriv_1990_2011* computed at the state level is the same value as the average of the variable computed on the overall sample. And the value is 0.8113613. The minimum and maximum value of the outcome variable computed at the state level is -4.594934 and 7.900403 respectively.

4. In the graph below, we present a scatterplot and a regression line of *d_emppriv_1990_2011* over *exposure_to_robots* with 95% confidence intervals.

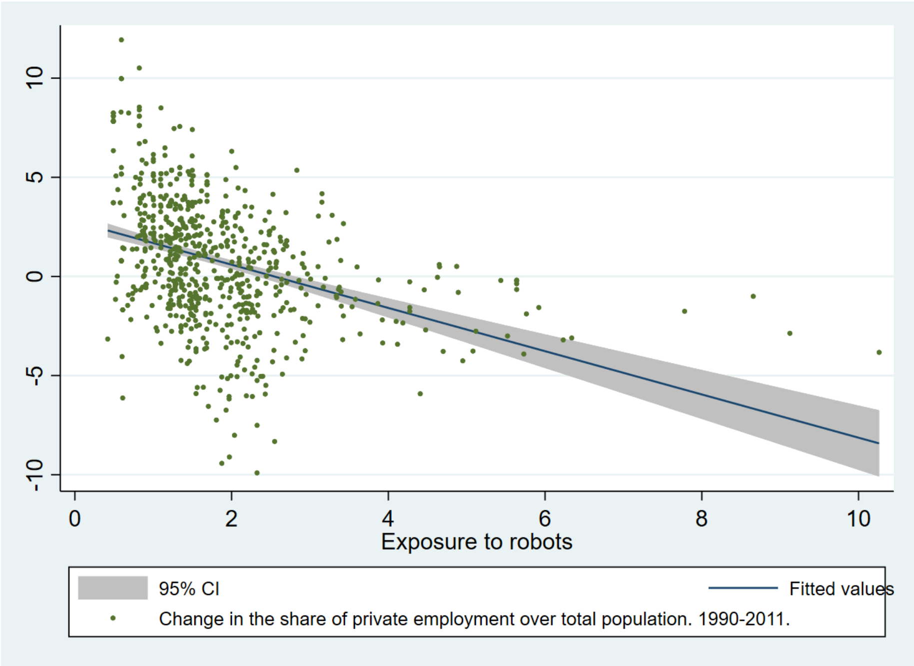


Figure 1: Scatterplot and regression line for employment over exposure to robots.

At an initial look, there appears to be a low yet negative correlation between the two variables. This is confirmed by their correlation value at -0.3769.

B Regression Analysis - Employment

1. The minimal assumptions required to make large sample inference are:

   (a) The variables $\{(X_1, Y_1), \ldots, , (X_i, Y_i), \ldots, (X_n, Y_n)\}$ are independent and identically distributed (i.i.d), that is, they are draws from a common distribution and they are independent from each other.

   (b) The variables (X, Y) satisfy the linear regression equation

   $$Y = X\beta + e \ . \tag{1}$$

   (c) The variables have finite second moments

   $$\mathbb{E}[Y^2] < \infty \ , \tag{2}$$

   $$\mathbb{E}||X||^2 < \infty \ . \tag{3}$$

   (d) The design matrix $\mathbf{Q}_{XX}$ is invertible, that is,

   $$\mathbf{Q}_{XX} = \mathbf{E}[XX'] > 0 \ . \tag{4}$$

   This assumption implies there is absence of perfect collinearity in regressors.

(e) The errors and regressors should be uncorrelated, i.e.,

$$\mathbf{E}[u_i X_i] = 0 \tag{5}$$

To make inference it is important to know something about the distribution of the errors, but by exploiting the Central Limit Theorem we can discard the assumption of normality of the errors because when we are studying the large-sample properties of the OLS.

2. Simple linear regression of employment on exposure to robots:

	(1)
VARIABLES	d_emppriv_1990_2011
expos_to_robots	-1.091***
	(0.110)
Constant	2.775***
	(0.234)
Observations	722
R-squared	0.142

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This simple regression model without control variables tells us that the effect of exposure to robots (measured as the number of robots per 1000 workers) on employment (measured as the change in the private employment to population ratio from 1990 to 2011) is predicted to be -1.09 percent points. This coefficient is statistically significant at the 1% level.

3. Regression with specified controls (ethnicity and education):

When we first run the regression, adding all the specified control variables from the group we find that STATA tells us that it has dropped two variables because of multicollinearity, one from each group variable. We decided to manually drop *ipums_white_1990* and *ipums_highschool_1990* and use them as reference values for ethnicity and education. We present the corrected model as

	(1)
VARIABLES	d_emppriv_1990_2011
expos_to_robots	-0.556***
	(0.0888)
ipums_logpop_1990	-0.447***
	(0.0894)
ipums_female_1990	0.634
	(15.73)
ipums_above65_1990	11.58**
	(5.439)
ipums_somecollege_1990	10.59***
	(3.289)
ipums_college_1990	23.29***
	(6.426)
ipums_masters_1990	-53.27***
	(11.69)
ipums_black_1990	-5.124***
	(1.047)
ipums_hispanic_1990	1.879***
	(0.687)
ipums_asian_1990	-26.92***
	(8.414)
Constant	2.293
	(7.817)
Observations	722
R-squared	0.385

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The *expos_to_robots* tells us that the community zones which use 1 additional robot per 1000 workers predict a decrease of 0.56 percentage points in the change in the private employment to population ratio over 1990 to 2011. *ipums_somecollege_1990* indicates the relative difference in the impact of some years in college vis-a-vis education till highschool. This implies that with other variables constant, on average, there is a 10.59 percentage point difference in the impact of some years in college and education till highschool on the change in employment levels. Similarly, *ipums_college_1990* indicates the relative difference in the impact of the share of population who completed college, against those with only highschool education, which is 23.29 percentage points. On the other hand, *ipums_black_1990* indicates the relative difference between the impact of share of black community vis-a-vis white community on the change in employment levels. A negative of value of 5.124 implies that for an increase in 1 percentage point in the share of black community, on average, the change in employment levels is 5.124 percentage points less than due to an increase in 1 percentage point in the share of white community.

All coefficients are statistically significant at the 1% level except for *ipums_female_1990* and *ipums_above65_1990*. The former one is not statistically significant while the latter one is statistically significant at the 5% level.

4. On adding the squared term of exposure to robots in the model, the r-squared marginally improves to 0.393. Also, with negative linear coefficient and positive quadratic coefficient to the variable exposure to robots, the model implies that though its impact is predicted to be negative to the change in employment, with increase in the value of exposure to robots, the overall negative marginal effect decreases. This may be due to better adaption of technology in the community zones with higher exposure to robots.

	(1)
VARIABLES	d_emppriv_1990_2011
expos_to_robots	-1.325***
	(0.246)
expos_2	0.106***
	(0.0272)
ipums_logpop_1990	-0.405***
	(0.0928)
ipums_female_1990	4.906
	(15.73)
ipums_above65_1990	10.57*
	(5.562)
ipums_somecollege_1990	8.749***
	(3.285)
ipums_college_1990	20.84***
	(6.386)
ipums_masters_1990	-50.08***
	(11.56)
ipums_black_1990	-5.611***
	(1.052)
ipums_hispanic_1990	1.223*
	(0.723)
ipums_asian_1990	-26.26***
	(8.474)
Constant	1.340
	(7.813)
Observations	722
R-squared	0.393

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

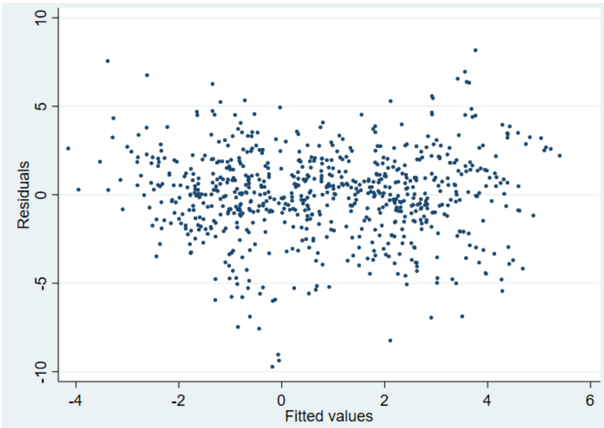
5. Since in the model we have specified the variable of exposure to robots as a quadratic function to capture its non-linear effects, the marginal effect of it is a function of the value it takes. We evaluate this function at the mean value of *expos_to_robots*. The p-value of the test is well below the standard value of 0.05, therefore, the null hypothesis of the marginal effect of exposure to robots is equal to zero is rejected.

```
. lincom _b[expos_to_robots] + (2 * _b[expos_2] * mean_etr)

 ( 1)  expos_to_robots + 3.600026*expos_2 = 0

------------------------------------------------------------------------------
d_emppr~2011 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.9451252   .1541927    -6.13   0.000    -1.247853   -.6423971
------------------------------------------------------------------------------
```

6. To check for heteroskedasticity, we check for visual patterns in the residual-fitted plot.



From the scatterplot, there is no distinct pattern visible between the residuals and fitted values, except for a wider spread of the residuals for extreme fitted values. Hence, we further test for heteroskedasticity using Breusch Pagan test for heteroskedasticity.

```
White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

   chi2(76) = 168.18
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

--------------------------------------------------
              Source |       chi2     df         p
---------------------+----------------------------
  Heteroskedasticity |     168.18     76    0.0000
            Skewness |      17.13     11    0.1042
            Kurtosis |       9.52      1    0.0020
---------------------+----------------------------
               Total |     194.83     88    0.0000
--------------------------------------------------
```

Since the p-value for the chi-squared statistic, under the null hypothesis, is well below the standard 0.05, we reject the null hypothesis of homoscedasticity in the model. Therefore, in our specification of the model we use robust standard errors to control for heteroskedasticity.

7. Huber-white robust standard errors help us to draw more accurate inference because, unlike in the classical OLS where the variance of the error term is assumed to be constant, robust standard errors are computed with the estimated values of error variances. This process is entirely empirical, without any restriction on the structure of heteroskedasticity. If there is hetereoskedasticity in the model, though the classical OLS estimators remain unbiased, they are no longer BLUE. The estimates no longer have the minimum variances. And their variances are not unbiased and hence, any inferences from their t-test or F-test are not accurate.

C Regression-Wage

1. For creating a dummy for $whites = 1$ and 0 otherwise we used the following code:

```
codebook race
gen white = 0 if race != 5
replace white = 1 if race == 5
tab race white
```

2. Following is the regression output of change in wage over exposure to robots and female, and white and education:

d_yrwage_ln_1990_2011	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
expos_to_robots	-.0452585	.0047369	-9.55	0.000	-.0545432	-.0359739
1.female	.1428625	.0145603	9.81	0.000	.1143233	.1714016
female#c.expos_to_robots						
1	.0078448	.006691	1.17	0.241	-.0052699	.0209595
education						
high school	-.0310826	.0138116	-2.25	0.024	-.0581543	-.0040109
some college	-.0251477	.0137477	-1.83	0.067	-.0520941	.0017986
college	.1015417	.015111	6.72	0.000	.0719231	.1311603
master	.1624391	.017479	9.29	0.000	.1281792	.1966991
1.white	-.0301436	.0187803	-1.61	0.108	-.0669543	.0066671
education#white						
high school#1	.0469742	.0261604	1.80	0.073	-.0043019	.0982502
some college#1	.104628	.0261267	4.00	0.000	.053418	.155838
college#1	.0593417	.0268691	2.21	0.027	.0066765	.1120068
master#1	.0211344	.0282687	0.75	0.455	-.0342741	.0765429
_cons	.0541556	.0140176	3.86	0.000	.0266802	.0816311

The relative difference in the marginal effect of exposure to robot between men and women, on average, is 0.07, however it is statistically insignificant since its p-value is equal to 0.241. On the other hand, in the categorical variable education, the groups of individuals with college degree and master degree have a positive wage premium over those groups of individuals with less than high school education, which is statistically significant too.

3. Chow test not included.

4. Following is the test output to check if marginal effect of exposure to robots differs between men and women is

```
. test 1.female#c.expos_to_robots

 ( 1)  1.female#c.expos_to_robots = 0

       F(  1, 23757) =    1.37
            Prob > F =    0.2410
```

Since the p-value of the test is well above the standard 0.05, the null hypothesis of the difference in marginal effect of exposure to robots between men and women is zero cannot be rejected.

5. The wage premium for the group of individuals with college degree over those group of individuals with less than high school education is, on average, 0.12 units.

```
. reg d_yrwage_ln_1990_2011 expos_to_robots female edu2 edu3 edu4 edu5

      Source |       SS           df       MS      Number of obs   =    23,770
-------------+----------------------------------   F(6, 23763)     =    147.25
       Model |  315.098903         6  52.5164838   Prob > F        =    0.0000
    Residual |   8475.1713    23,763  .356654097   R-squared       =    0.0358
-------------+----------------------------------   Adj R-squared   =    0.0356
       Total |   8790.2702    23,769  .369820783   Root MSE        =    .59721

------------------------------------------------------------------------------
d_yrwage_l~2011 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------+-------------------------------------------------------------
expos_to_robots |  -.0414809   .0033469   -12.39   0.000     -.048041   -.0349208
         female |   .1572163   .0077475    20.29   0.000     .1420307    .1724019
           edu2 |  -.0176851   .0117199    -1.51   0.131    -.0406569    .0052867
           edu3 |   .0027235   .0116779     0.23   0.816    -.0201659    .0256129
           edu4 |   .1202807   .0124924     9.63   0.000     .0957947    .1447667
           edu5 |   .1674346   .0135864    12.32   0.000     .1408045    .1940648
          _cons |   .0381767   .0112288     3.40   0.001     .0161675    .0601859
------------------------------------------------------------------------------
```

6. The marginal effect of exposure of robots for the group of individuals with masters degree is $-0.055 + 0.03631 = -0.01869$. As for the statistical significance of the difference with the marginal effect for the group of individuals with less than high school education, as observed from the regression output, the p-value is 0.002, well below the standard 0.05, hence the difference is statistically significant.

```
. //*===========================================================================
. //* C.6. Regression with interaction between robots and dummies for education.
. //*===========================================================================
.
. reg d_yrwage_ln_1990_2011 c.expos_to_robots##i.education female race, rob

Linear regression                               Number of obs     =     23,770
                                                F(11, 23758)      =      89.82
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0370
                                                Root MSE          =     .59693

------------------------------------------------------------------------------
                         |               Robust
  d_yrwage_ln_1990_2011  | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------------+----------------------------------------------------
         expos_to_robots |  -.0557862   .0083043    -6.72   0.000    -.0720631   -.0395093
                         |
               education |
             high school |  -.0100199   .0223958    -0.45   0.655    -.0539171    .0338772
            some college |  -.0223304   .0226929    -0.98   0.325    -.0668099    .0221491
                 college |   .0529709   .0248537     2.13   0.033     .0042561    .1016858
                  master |   .0995622   .0233687     4.26   0.000     .0537581    .1453663
                         |
education#c.expos_to_robots |
             high school |  -.0037766   .0103959    -0.36   0.716    -.0241532       .0166
            some college |   .0140973   .0102976     1.37   0.171    -.0060867    .0342813
                 college |   .0367525   .0115503     3.18   0.001     .0141132    .0593917
                  master |   .0363174   .0108219     3.36   0.001     .0151058    .0575291
                         |
                  female |   .1573782   .0077434    20.32   0.000     .1422006    .1725558
                    race |   .0040804   .0026976     1.51   0.130    -.0012071    .0093679
                   _cons |   .0505455   .0214288     2.36   0.018     .0085436    .0925473
------------------------------------------------------------------------------
```