The Impact of a Government Pay Reform in Mexico on the Public Sector Wage Gap

Erendira Leon (University of Westminster)

Barry Reilly (University of Sussex)

2022 UK Stata Conference

Outline

- Applied economics
- UQ regressions and D-i-D framework
- Validity: Difference in means
- Stata in applied economics
- Other econometric models
- Conclusions

Applied economics

- Analysis based on the **impacts of wage policies** established in the public sector on wages and public-private sector wage differentials
- **A Pay Reform** was introduced in 2018 to **regulate earnings of Public Servants up to 40%** and to narrow the wage gaps

Unconditional Quantile (UQ) regression

- Address **heterogeneity** that can emerge at different percentiles
- Analyse wages across the unconditional wage distribution

Difference-in-Differences (D-i-D) estimation

- Address **endogeneity of employment selection**
- The differences in time-invariant unobservable effects of the public and the private sector pre-treatment and post-treatment are potentially eliminated

Unconditional Quantile (UQ) regressions within Difference-in-Differences (D-i-D) framework

Based on Firpo et al. (2009) and (2018) in applying the **Re-centred Influence Function (RIF)** procedure

- Compare public and private sector wages before and after the policy
- Shed light on the public–private sector wage differentials across the unconditional wage distribution
- RIF procedure allows the average effects to be interpreted at different quantiles of unconditional hourly wage distribution as the dependent variables

  (e.g., the 5th, 95th percentiles or other intermediate quantiles)
- RIF centre the IF around the statistic of interest (e.g., the population mean "μ" E(Y)) and not zero (i.e., re-weighting the observations)

UQ regressions within D-i-D framework

- Estimates from a linear probability model and transformed into unconditional quantile effects using the reciprocal of the probability kernel density
- Then, RIF-quantiles within a D-i-D approach are estimated through linear regressions

$$\widehat{RIF}\,(w_i, \widehat{q_\tau})_i = \beta_{0\tau} + \beta_{1\tau}\,(POST18_i) + \beta_{2\tau}\,(POST18_i * TREAT_i) + \beta_{3\tau}\,TREAT_i + \beta_{4\tau}\,X_i + \delta_{k\tau} + e_{i\tau}$$

X_i are years of education, age and its square, urban residence status, marital status, head of household status and the economic sector. δ_k are municipality fixed effects

Validity: Difference in means

diff runs several difference in differences (D-i-D) treatment effect estimations of a given outcome variable.

diff is also suitable for estimating repeated cross-sections

diff *outcome_var* [*if*] [*in*] [*weight*] ,[*options*]

where the model requires...

period(*varname*) indicates the binary period variable (0: before; 1: after)

treated(*varname*) indicates the binary treatment variable (0: controls; 1:treated)

Validity: Difference in means

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 85867

	Before	After	
Control:	46093	25057	71150
Treated:	9889	4828	14717
	55982	29885	

Outcome var.	lg_inc	S. Err.	|t|	P>|t|
Before				
Control	3.494			
Treated	4.019			
Diff (T-C)	0.525	0.006	85.58	0.000***
After				
Control	3.522			
Treated	4.010			
Diff (T-C)	0.488	0.009	56.12	0.000***
Diff-in-Diff	-0.037	0.011	3.45	0.001***

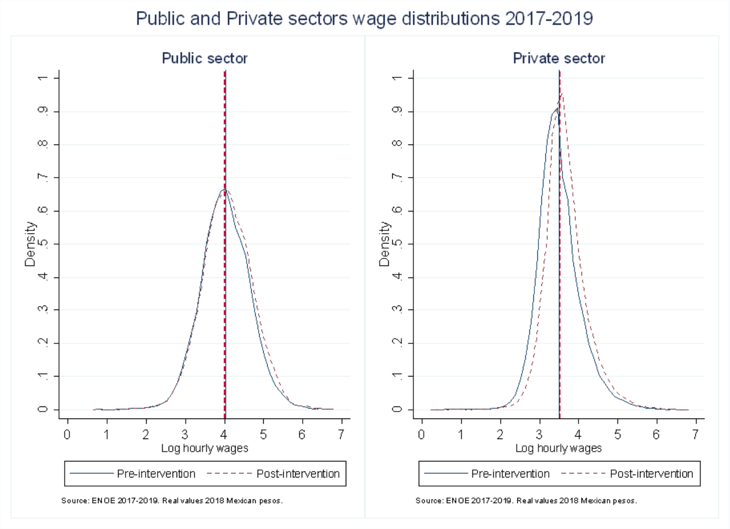
R-square: 0.11

* Means and Standard Errors are estimated by linear regression

**Inference: *** p<0.01; ** p<0.05; * p<0.1

The D-i-D estimate indicates the intervention may decrease the wages of the treated group, on average, 3.7% due to this reform

Validity


Public and Private sectors wage distributions 2017-2019
Public sector
Private sector
Density
Log hourly wages
Pre-intervention
Post-intervention
Source: ENOE 2017-2019. Real values 2018 Mexican pesos.

pctile – creates a new variable containing the percentiles of *exp*, where *exp* is typically another variable

pctile [*type*] *newvar* = *exp* [*if*] [*in*] [*weight*] [, *pctile_options*]

where *newvar* is the new variable containing percentages

pctile_options calculate percentiles corresponding to the specified percentages, e.g., 100

weight Stata allows four kinds of weights: fweights (frequency weights), pweights (sampling weights), aweights (analytic weights) or iweights (importance weights)

Stata in applied economics: RIF

kdensity – produces univariate kernel density estimates and graphs the result

kdensity *varname* [*if*] [*in*] [*weight*] [, *options*]

varname is the variable of interest

options;

at(*var_x*) estimate density using the values specified by *var_x*

generate(*newvar_x newvar_d*) store the estimation points in *newvar_x* and the density estimate in *newvar_d*

Stata in applied economics: RIF-quantiles within D-i-D framework

xtreg – fits regression models to panel data. In particular fixed-, between-, and random-effects and population-averaged linear models

Fixed-effects (FE) model

xtreg *depvar* [*indepvars*] [*if* [*in*] [*weight*], fe [*FE_options*]

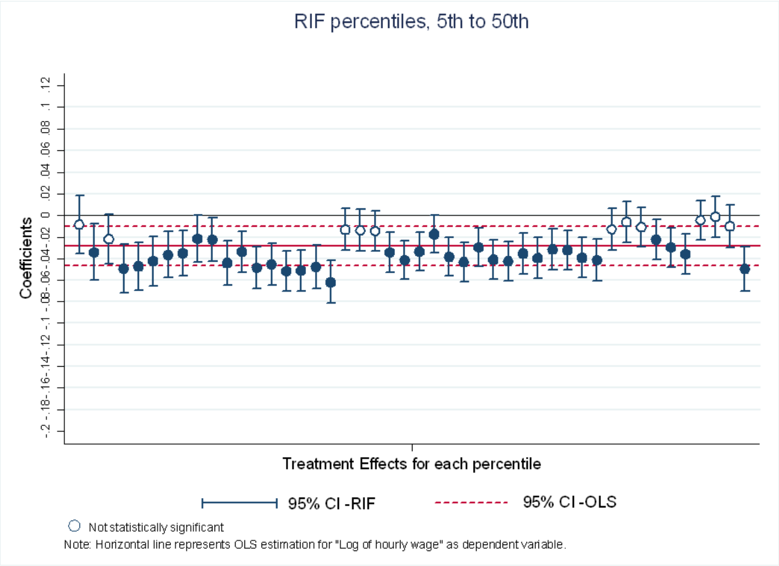
where *depvar* is the dependent variable, and *indepvars* are independent variables

weights are allowed for the fixed-effects model and for the population-averaged model

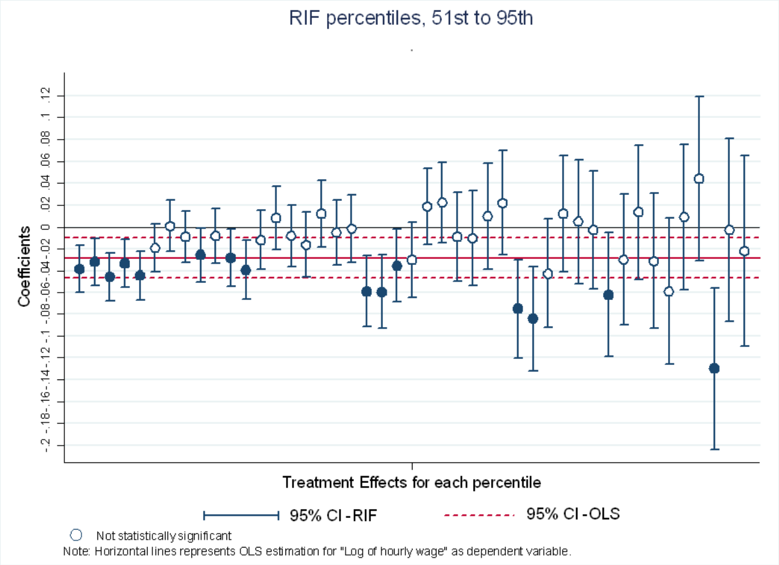
options for the type of standard error reported. E.g., *robust, cluster*, etc.

A panel variable must be specified with the use of **xtset**

RIF-quantiles within D-i-D framework: Results



RIF-quantiles within D-i-D framework: Results



RIF-quantiles within D-i-D framework: selected percentiles

The empirical evidence suggests the policy **reduced the public sector wage gap largely through a contraction of wages for the low-paid workers**

Higher-paid public sector earners do not appear to have incurred any pay penalties with this policy

	(1) Log hourly wage	(2) RIF 10	(3) RIF 25	(4) RIF 50	(5) RIF 75	(6) RIF 90
Treatment effects	-0.029***	-0.043***	-0.015	-0.050***	0.022	-0.060
	(0.009)	(0.012)	(0.009)	(0.010)	(0.019)	(0.034)
Obs.	85,867	85,867	85,867	85,867	85,867	85,867
Covariates	Yes	Yes	Yes	Yes	Yes	Yes

Notes: ** p<0.05, *** p<0.01
The sample is constructed from the 2017-2019 Mexican National Occupations and Employment Survey.
Municipality Fixed effects. Robust standard errors adjusted for 645 clusters at municipality level in parentheses.
Model (1) OLS standard estimation.

Other econometric models

RIF-quantiles within D-i-D framework can be applied also to:

- A measurement of inequality: Gini

Two-step model within a Heckman framework for dealing with sectoral employment attachment is more applicable for mean-based analysis and not UQ regressions

- Test orthogonality of the instruments to the variable of interest
- It can be calculated after **ivregress** or **ivreg2** by the command **ivendog**
- Head of the household status, the number of children, older people in the household, and other household members working in the public sector

RIF-Gini

rifvar() – is an egen extension that can be used to create RIFs for a large set of distributional statistics and in combination with other statistics

The flexibility and simplicity of this tool extends the analysis to the Gini inequality index, using linear regressions

egen *newvar* = **rifvar**(*varname*) [*if*] [*in*] [, *options*]

where *newvar* is the new variable created, and *varname* the variable of interest

Then

xtreg *depvar* [*indepvars*] [*if* [*in*] [*weight*], fe [*FE_options*]

Two-step Heckman procedure

- First step

dprobit – Rather than reporting the coefficients, dprobit reports the marginal effect, that is, the change in the probability for an infinitesimal change in each independent, continuous variable and, by default, reports the discrete change in the probability for dummy variables.

> **dprobit** [*depvar indepvars* [*if*] [*in*] [*weight*]] [, options]

where *depvar* is the dependent variable, and *indepvars* the independent variables

options for the type of standard error reported (e.g., *robust, cluster*, etc)

Two-step Heckman procedure

predict – calculates predictions, residuals, influence statistics, and the like after estimation. Exactly what **predict** can do is determined by the previous estimation command; command-specific options are documented with each estimation command

predict [type] *newvar* [*if*] [*in*] [, $single_options$]

where *newvar* contains the new variable with the predicted values

- **Second step**

Use the predicted values in the OLS regression with D-i-D

xtreg *depvar* [*indepvars*] [*if* [*in*] [*weight*], fe [*FE_options*]

Conclusions

- **RIF-quantile regressions** within a D-i-D framework can be implemented with Stata for analysing policy impacts
  - across the unconditional distribution
  - before and after the introduction of such policies
- Different in means test can be applied with Stata for **validating** the implementation of **D-i-D** methodology
  - diff
- **Stata commands**
  - pctile
  - kdensity
  - rifvar
  - xtreg

Thank you!

References

Firpo, S. P., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. Econometrica 77: 953-973.
https://doi.org/10.3982/ECTA6822

Firpo, S. P., N. M. Fortin, and T. Lemieux (2018). Decomposing wage distributions using recentered influence function regressions. Econometrics 6: 28. https://doi.org/10.3390/econometrics6020028.

Data

National Employment Survey (ENOE) from 2017 to 2019

- Report, *inter alia*, earnings, type of employment, schooling, etc
- Male formal sector employees aged between 15 to 65 years

Example: Stata commands

```
. foreach qt of numlist 5 10 15 20 25 30 35 40 45 50 {
  2. gen rif_`qt'=.
  3. }

. pctile eval2=lg_inc, nq(100)
. kdensity lg_inc, at(eval2) gen(eval_nw dens_nw) width(0.10) nograph

. foreach qt of numlist 5 10 15 20 25 30 35 40 45 50 {
  2. local qc = `qt'/100
  3. replace rif_`qt'=eval_nw[`qt']+`qc'/dens_nw[`qt'] if
lg_inc>=eval_nw[`qt']
  4. replace rif_`qt'=eval_nw[`qt']-(1-`qc')/dens_nw[`qt'] if
lg_inc<eval_nw[`qt']
  5. }


** Gini***

. egen wage_gini =rifvar(wage), gini

**********
```

Example: Stata commands

```
*****************************************

****** RIF quantiles with D-i-D

*****************************************

. xtset id_mun1
. estimates store fe_cov

. foreach x of numlist 10 25 50 75 90 {
  2. quietly xtreg rif_`x' post19 public post19_treat year_sch age age2
urban head_h ib2.sector ib2.marital, fe robust
  3. estimates store all_`x'
  4. }

*********
***GINI
*********
. xtreg wage_gini post19 public post19_treat year_sch age age2  urban
head_h ib2.sector ib2.marital, fe robust
  2. estimates store gini_all
```

Example: Stata output

```
Fixed-effects (within) regression               Number of obs     =     85,867
Group variable: id_mun1                         Number of groups  =        645

R-sq:                                           Obs per group:
     within  = 0.0528                                         min =          5
     between = 0.0158                                         avg =      133.1
     overall = 0.0437                                         max =      2,850

                                                F(15,644)         =      36.30
corr(u_i, Xb)  = -0.0932                        Prob > F          =     0.0000

                                    (Std. Err. adjusted for 645 clusters in id_mun1)
------------------------------------------------------------------------------------
                   |               Robust
            rif_10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
            post19 |   .0362564    .007631     4.75   0.000     .0212719     .051241
            public |   .2554758   .0173733    14.71   0.000     .2213605     .289591
      post19_treat |  -.0426318   .0116927    -3.65   0.000    -.0655922   -.0196714
          year_sch |   .0278765   .0020328    13.71   0.000     .0238848    .0318683
               age |   .0172619   .0017206    10.03   0.000     .0138833    .0206405
              age2 |  -.0002287   .0000212   -10.79   0.000    -.0002703    -.000187
            head_h |   .0539975    .006901     7.82   0.000     .0404463    .0675487
             urban |   .0233497   .0153079     1.53   0.128    -.0067097    .0534092
                   |
            sector |
      Construction |   .1138361   .0116268     9.79   0.000     .0910049    .1366672
          Commerce |  -.1747128   .0126122   -13.85   0.000    -.1994789   -.1499467
          Services |    -.15608   .0106168   -14.70   0.000    -.1769277   -.1352323
       Agriculture |  -.0914186   .0255851    -3.57   0.000    -.1416589   -.0411782
 Mining and energy |  -.0416519   .0173775    -2.40   0.017    -.0757753   -.0075285
                   |
           marital |
           Married |   .0176949   .0125828     1.41   0.160    -.0070135    .0424032
            Single |  -.0196115   .0134431    -1.46   0.145     -.046009    .0067861
                   |
             _cons |   2.315114   .0497659    46.52   0.000     2.217391    2.412837
-------------------+----------------------------------------------------------------
           sigma_u |  .30428324
           sigma_e |  .70731561
               rho |  .15616603   (fraction of variance due to u_i)
------------------------------------------------------------------------------------
```

coefplot

coefplot plots results from estimation commands, multiple models or matrices can be combined into one graph

The default behavior of coefplot is to draw markers for coefficients and horizontal spikes for confidence intervals

> **coefplot** subgraph [—— subgraph ...] [, globalopts]
> where subgraph is defined as
> (plot) [(plot) ...] [, subgropts]
> and plot is either _skip (to skip a plot) or
> model [model ...] [, plotopts]
> and model is either
> name [, modelopts]
> where name is the name of a stored model

globalopts are options that apply to the overall graph