# RIF-Regressions and RIF-Decompositions:

## Implementation and Interpretation of Distributional Effects

Fernando Rios-Avila

Levy Economics Institute

2022-05-30

# Introduction

## What do you mean RIF?

## and what do you mean **distributional** effects??

# Introduction

- Unconditional quantile regressions (UQR) via RIF (Recentered Influence functions) was introduced by Firpo, Fortin, and Lemieux (2009) as a computationally simple strategy to estimate **Unconditional Partial Effects (UPE)** on quantiles, caused by **small** changes in distributions of characteristics.

- These are effects that would be measured across the **whole** distribution, but not for individual observations.

  - This is in contrast with standard **Conditional quantile regressions**, which focus on effects conditional on all characteristics being known.

- Since then, RIF-regressions have been used to analyze other statistics. See FFL(2018), Firpo and Pinto (2016), Chung and Vankerm (2018), Cowell and Flachaire (2007), Essama-Nssah and Lambert (2012) and Heckley et al (2016).

- Despite its popularity, however, the correct interpretation of RIF regressions and decomposition remain a challenge.

# CQR vs UQR

## What is the difference between Conditional and Unconditional effects?

# LR vs CQR : Visual analysis

Simulated Data | What LR does: | What CQR does: | But What UQR does?

# LR vs CQR : Visual analysis

Simulated Data | What LR does: | What CQR does: | But What UQR does?

# LR vs CQR : Visual analysis

Simulated Data | What LR does: | What CQR does: | But What UQR does?

# LR vs CQR : Visual analysis

Simulated Data | What LR does: | What CQR does: | **But What UQR does?**

- With LR and CQR, we can use coefficient estimations to identify how individual changes, or group specific changes will affect **expected** outcomes.

$$\frac{\partial y_i}{\partial x_{i1}} \; or \; \frac{\partial E(y_i|X)}{\partial x_1} \; or \; \frac{\partial Q(y_i|X,\theta)}{\partial x_1}$$

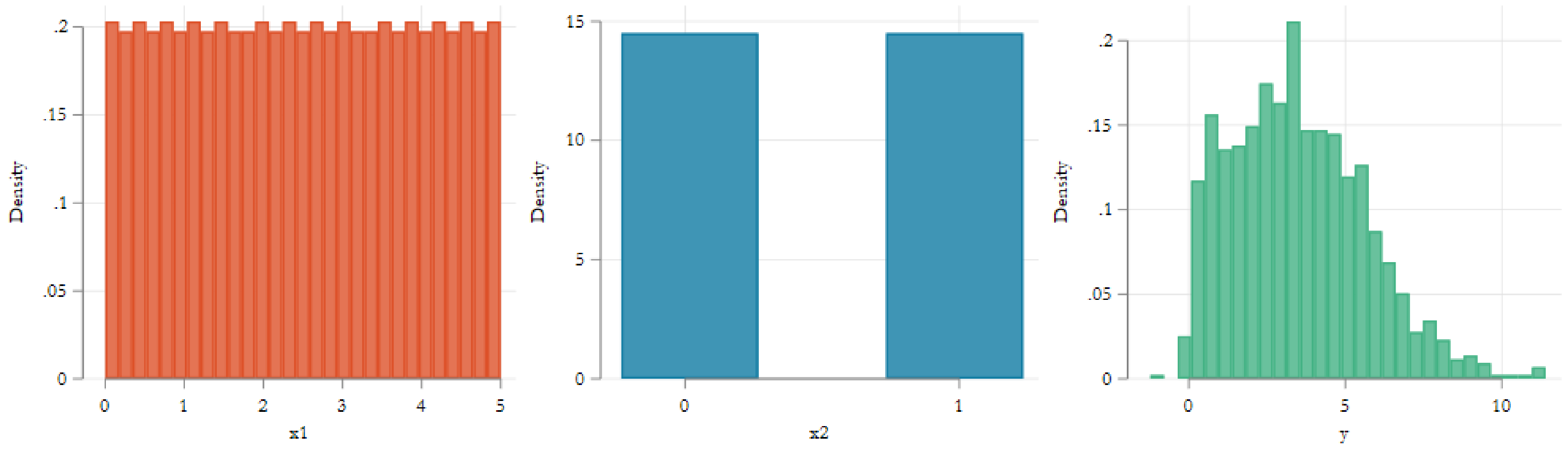
- But with **UQR** you can't measure the change on the distribution of individual values. (too small to be meaningful)

- Rather, we need to think in terms of distributional changes in $X's$, and how that will affect the change in the distribution of $y$ (but measured in quantiles)

# UQR: What it Does

UQR: From F(x1) and F(x2) to F(y) | UQR: D in x_1 | UQR: D in x_1 | UQR: Changes in x_2

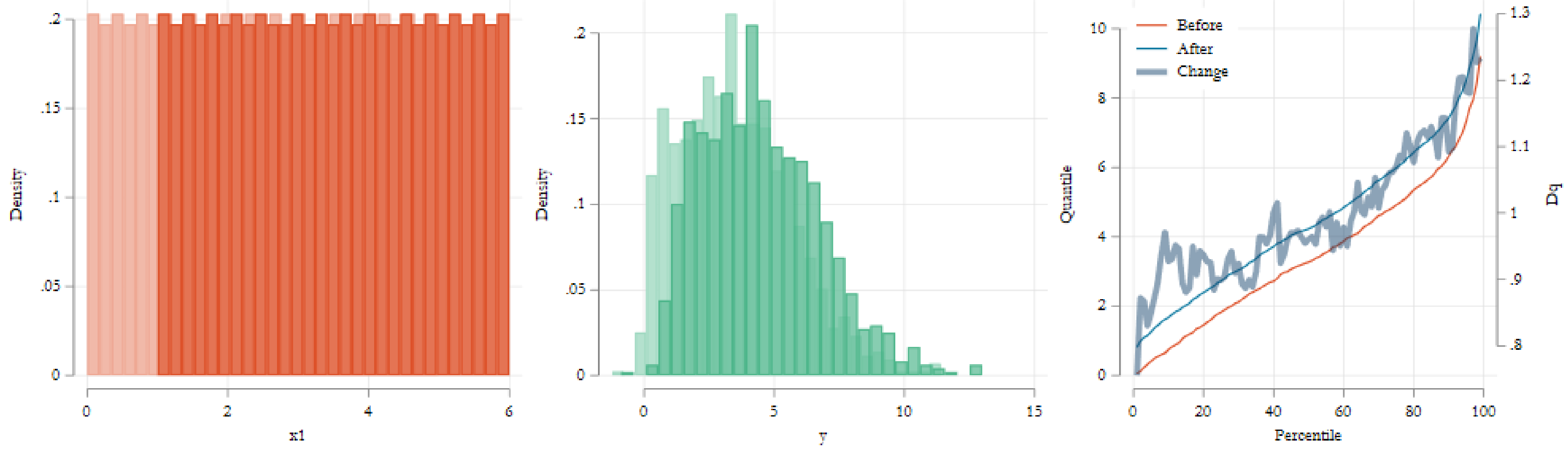
- In the current example, $F(Y)$ depends on the distribution of 3 other factors: $F(x_1)$, $F(x_2)$, $F(e)$. But, we assume $F(e)$ is independent, thus can be ignored.
- Thus, we can change the distribution of both $x_1$ or $x_2$!



# UQR: What it Does

UQR: From F(x1) and F(x2) to F(y) | **UQR: D in x_1** | UQR: D in x_1 | UQR: Changes in x_2

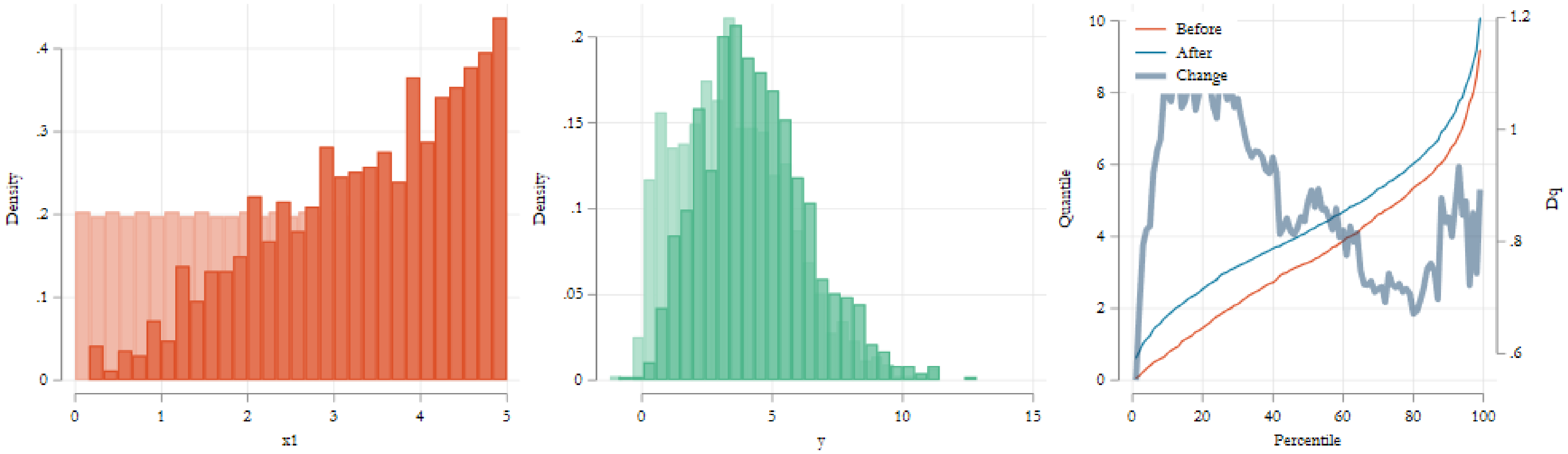
- The simplest kind of change could be one when the whole distribution shifts (change in means)



# UQR: What it Does

UQR: From F(x1) and F(x2) to F(y) | UQR: D in x_1 | **UQR: D in x_1** | UQR: Changes in x_2

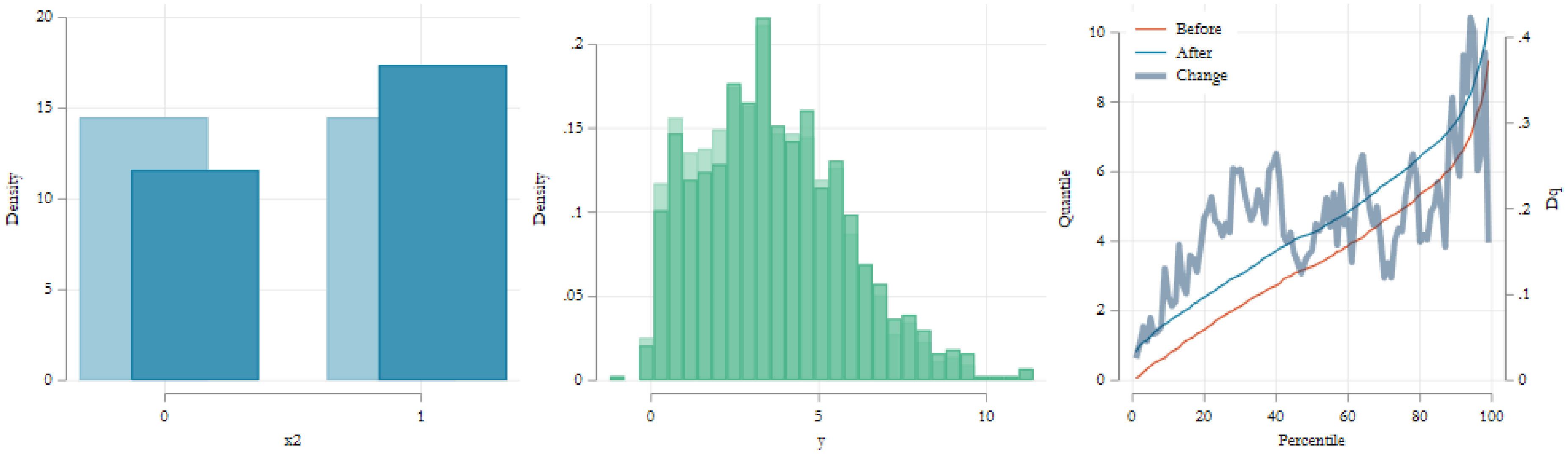
- But other kind of distributional changes could also be considered



# UQR: What it Does

UQR: From F(x1) and F(x2) to F(y) | UQR: D in x_1 | UQR: D in x_1 | **UQR: Changes in x_2**

- If $x_2$ is discrete, One may want to consider changes in proportions or shares of $1's$.



Why stop at quantiles ?

# Why stop at quantiles ?

Given the simulated data, we do not need to stop at **only** analyzing quantiles. We can, in fact, analyze changes in ANY statistic.

| | Original | After dx1 | After dx2 | Change dX1 | Change dX2 |
|---|---|---|---|---|---|
| Mean | 3.487 | 4.673 | 3.673 | 1.186 | 0.186 |
| Variance | 4.620 | 5.327 | 4.822 | 0.707 | 0.202 |
| SD | 2.150 | 2.308 | 2.196 | 0.159 | 0.046 |
| Skewness | 0.543 | 0.527 | 0.506 | -0.016 | -0.037 |
| Kurtosis | 3.031 | 2.998 | 2.940 | -0.033 | -0.091 |
| CV | 0.616 | 0.494 | 0.598 | -0.123 | -0.019 |

dX1: Change in the Mean of X1 in 1 unit

dX2: Change in the proportion of X2 in 10pp

## what about Quantile Treatment effects (QTE)?

# What about QTE?

Story Effect

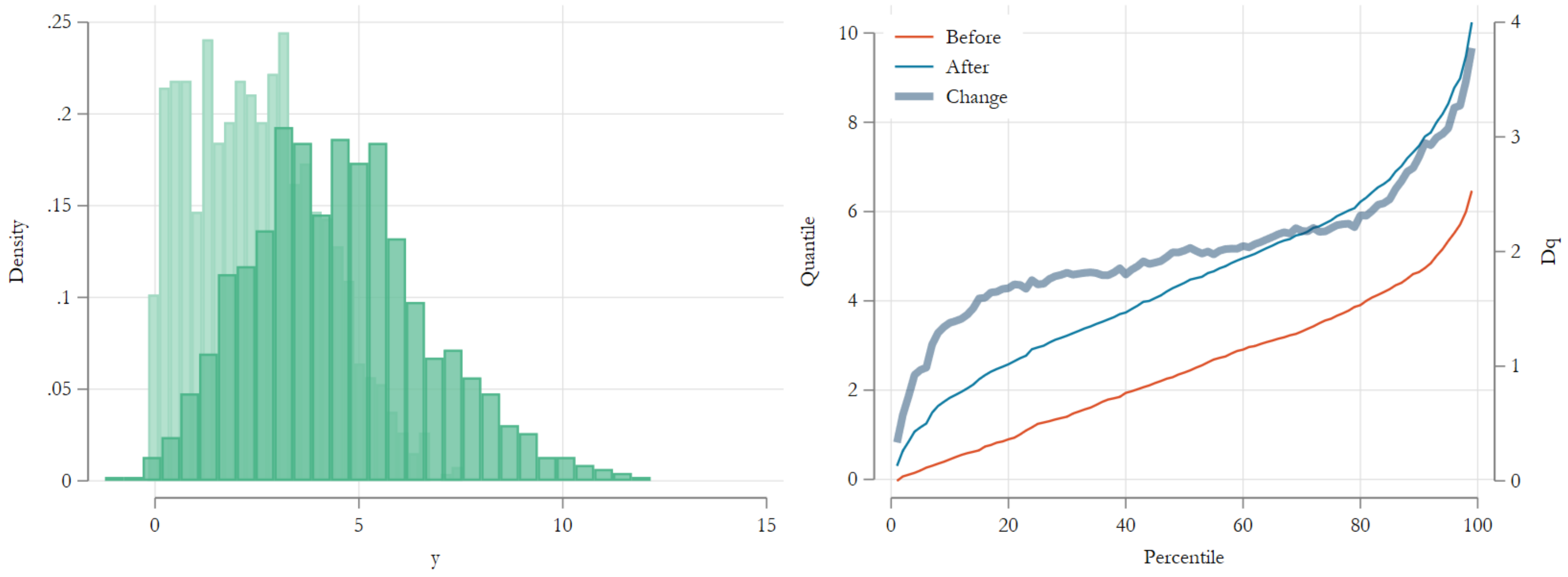
QTE can be understood as an extreme case of UQR, but also similar to CQR.

- CQR because you compare two distributions conditional on one [categorical] variable changin ( $x_2 = 0 \; vs \; x_2 = 1$ ), everything else constant (at some values).
- UQR because the distributions are unconditional on everything else, (conditional on distributions) making 2 changes to the distribution. (all $x_2$ to 0 and to 1)

With QTE you are making the thought experiment of comparing distributions when everyone is "treated" vs "untreated" (ATE). Although ATT or ATU are also feasible.

# What about QTE?

Story | Effect



# How are these effects estimated?

In the previous slides I have shown the intuition behind what it means to estimate Unconditional effect caused by changes in the distribution of $X's$.

The estimation of the effects or unconditional quantiles are *NOT* approximations. They are the true effects, because I know the data generating process exactly.

In real world applications we need to use approximations:

- Flexible multivariate modeling and Simulations/imputation (Machado and Mata, 2005).
  - Computationally intensive. Multiple Qreg or Flex Het Reg.
- Reweighting approaches (DiNardo, Fortin and Lemieux, 1996).
  - Two step approach. Logit $\rightarrow$ Comparing statistics of interest
- Local Approximations (Firpo, Fortin and Lemieux, 2009).
  - via RIF's!

# So why do we need RIF's?

- From the three strategies I described above, the first two can be very computationally intensive.
- In contrast, RIF regressions provide a faster approach to get **Local linear approximations** of changes in distributions.
- The approach, however, is not without limitations:
  - All changes are related to **small** location shifts in the distribution of explanatory variables (change in means)
  - Larger changes will create poor approximations of the effects
- One must keep this in mind when analyzing and interpreting estimation results

But what are RIFs (Recentered Influence functions) ?

# Building blocks: IF $\rightarrow$ RIF

# Building Blocks: Some Notation

To understand what (R)IF's are, we first need to establish some notation:

1) We are interested on analyzing **distributional** statistics $v$ , which summarize some aspect or property of the data of interest y .

2) To calculate any distributional statistic, we need to know $F_y$ or $f_y$ . (the distribution).

3) And if we assume that $y_i = g(x_i, e_i)$, we need $e_i$ to be fully independent from $x$.

4) $F_y$ can also be written as a weighted mean of all conditional distributions:

$$F_y = F(y) = \int F_{Y|x} f_x dx$$

$$F_{y|x} = F(y|x) = \int F(y|x, e) f_e \, de$$

# Building Blocks: What are Influence Functions (IF's)?

Influence functions are *directional* derivatives of how would $v(F_y)$ change when $F_y$ changes, at the margin.

$$IF(v(F_y), \partial F_y) = \frac{\partial v(F_y)}{\partial F_y}$$

However this definition of distribution change is too *broad*. Thus, for the formal definition of IF's, we need to redefine the definition of a change:

$$G_y^c = \varepsilon 1_{y \geq c} + (1-\varepsilon) F_y$$

Thus the IF is defined by:

$$IF(y, v(F_y)) = lim_{\varepsilon \to 0} \frac{v(G_y^c) - v(F_y)}{\varepsilon}$$

> what is the rate of change (slope) that would be caused by a change in the distribution of $Y$, that gives more weight to $y = c$?

# Building Blocks: Recentered Influence Functions

FFL (2009) proposes using an modification of IF, called recentered influence functions (RIF).

$$RIF(c, v(F_y)) = v(F_y) + IF(c, v(F_y))$$

Which has the nice properties

$$E\left[RIF(y_i, v(F_y))\right] = v(y)$$

$$N^{-1}Var\left[RIF(y_i, v(F_y))\right] = Var(v(F_y))$$

My interpretation: A **RIF** is the contribution of a single observation to the construction of the distributional $v$.

# Building Blocks: RIF and Regression

So how does a RIF regression work?

- First: we are interested in analyzing potential changes in $v(F_y)$
- Second: $v(F_y)$ can be approximated as the unconditional mean of RIF's:

$$v(F_y) = \int RIF(y, v(F_y)) f_y dy$$

- Third: the unconditional distribution $F_y$ can be written as a weighted mean of conditional distributions:

$$v(F_y) = \iint RIF(y, v(F_y)) F_{y|x} dy \; f_x dx = \int E\left[RIF(y, v(F_y))|X\right] f_x dx$$

# Building Blocks: RIF and Regression

Based on the last expression, there are two ways one can use to **simulate** changes on the statistic $v(F_y)$.

1) One either assumes changes in $f_x$ (long way (reweighting))

2) One models $E\left[RIF(y, v(F_y))|X\right]$ using linear regressions, And average "predicted" changes in the RIF.

$$RIF(y, v(F_y)) = \beta_0 + \beta_1 x + e$$

> Small changes in $X's$ will be reflected as changes in the RIF, which can be interpreted as potential changes in the distributional Statistic $v(F_y)$.

**Why Small?**

If $\Delta F_x \rightarrow 0$, then $F_y' \sim F_y$ and you do not need to Re-estimate the RIF.

$$RIF(y, v(F_y')) \sim RIF(y, v(F_y)) + \beta_1$$

For large changes, however, this may not hold.

# RIF-Regression: Interpretation

RIF regressions imply using a model specification as follows:

$$RIF(y, v(F_y)) = \beta_0 + \beta_1 x + e$$

- Standard approach, Obtain partial derivatives, and interpret:

$$\frac{\partial RIF_i}{\partial x_i} = \beta_1 \; or \; \frac{\partial E(RIF_i|X)}{\partial x} = \beta_1$$

  But this is a change in the RIF for observation $i$. Or all groups where $X = x$.

Effect of $v$ is too small to be meaningful. (What proportion of the sample experience the change?)

# RIF-Regression: Interpretation

- Better Approach: Aggregate and obtain a better interpretations:

$$v(F_y) = \int E\left[RIF(y, v(F_y))|X\right] f_x dx$$

$$\frac{\partial v(F_y)}{\partial \bar{x}} = \int \frac{\partial E(RIF|X)}{\partial x} f_x dx$$

  Or:

$$v(F_y) = \beta_0 + \beta_1 \bar{x} \rightarrow \frac{\partial v(F_y)}{\partial \bar{x}}$$

But the only "simulation" we can impose is a change in the distribution of $X$ that increases the mean in 1 unit. (FFL: Location shift)

unless:

$$RIF_i = \beta_0 + \beta_1 x + \beta_2 (x - \bar{x})^2 + e \rightarrow v(F_y) = \beta_0 + \beta_1 \bar{x} + \beta_2 \hat{\sigma}^2$$

# Example: rifhdreg

Panel Panel

```
ssc install rif, replace
ssc install qregplot, replace
use http://fmwww.bc.edu/RePEc/bocode/o/oaxaca.dta, clear

rifhdreg  lnwage educ exper tenure female [pw=wt], rif( q(10) )
Linear regression                                Number of obs     =      1,434
                                                 F(4, 1429)        =      18.53
                                                 Prob > F          =     0.0000
                                                 R-squared         =     0.1057
                                                 Root MSE          =     1.2185
------------------------------------------------------------------------------
             |               Robust
      lnwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]   Mean
-------------+----------------------------------------------------------------
        educ |   .1202519   .0176049     6.83   0.000     .0857177    .1547861   11.5
       exper |   .0265787   .0045188     5.88   0.000     .0177145     .035443   12.7
      tenure |   .0058231   .0039389     1.48   0.140    -.0019035    .0135497    7.8
      female |  -.0895745    .082585    -1.08   0.278    -.2515753    .0724262   45.8%
       _cons |   1.040692   .2704888     3.85   0.000     .5100938    1.571289
------------------------------------------------------------------------------
Distributional Statistic: q(10)
Sample Mean    RIF q(10) :  2.7635 -> 10th : 15.85
1pp increase women  -> -0.00089 log points ~ -0.089% wage decline At 10th
```

# Example: rifhdreg

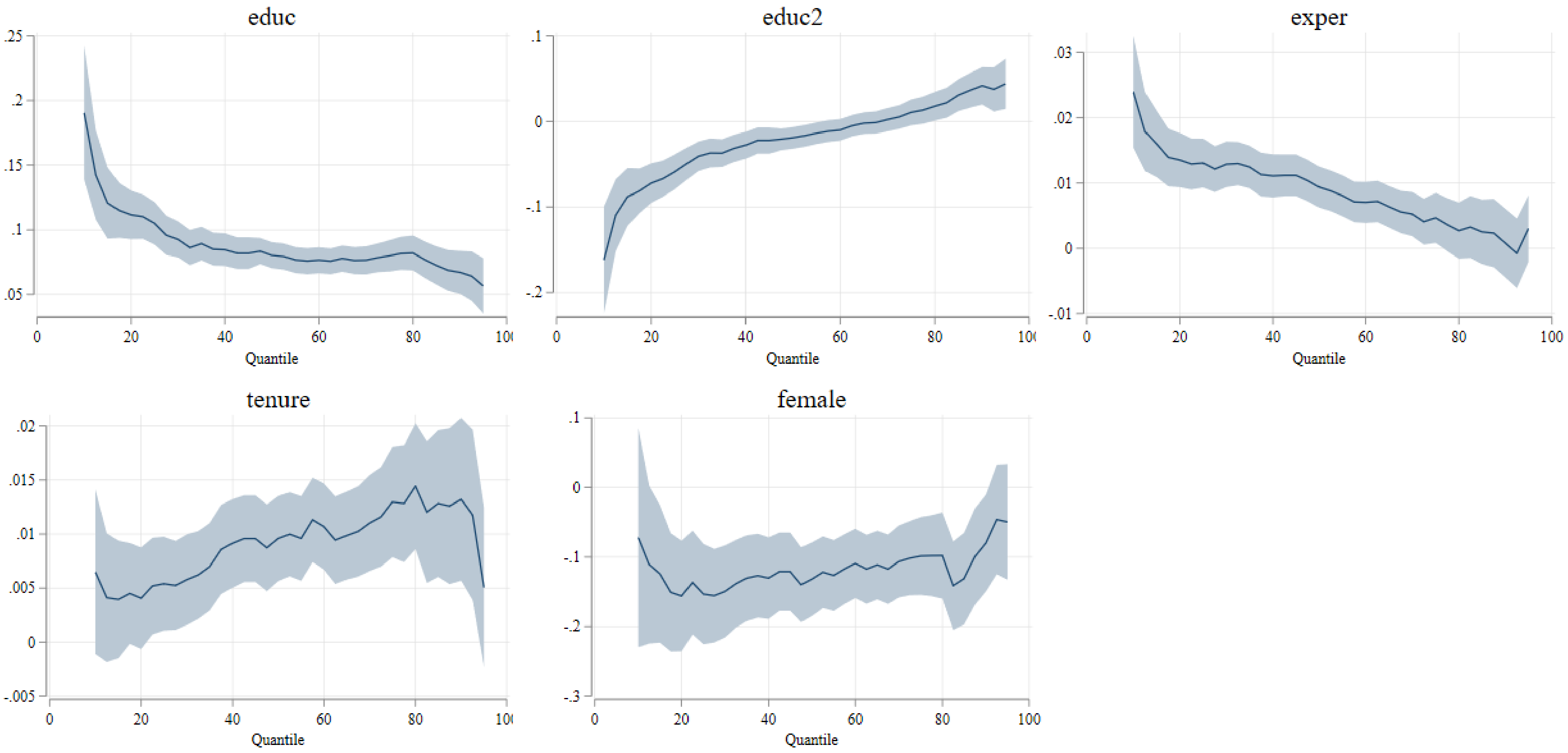
Panel Panel

```
sum educ if lnwage!=. [aw=wt]
gen educ2=(educ-r(mean))^2
sum educ2 if lnwage!=. [aw=wt]
replace educ2=educ2/r(mean)
rifhdreg  lnwage educ educ2 exper tenure female [pw=wt], rif( q(10) )
Linear regression                               Number of obs     =      1,434
                                                F(5, 1428)        =      17.87
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1458
                                                Root MSE          =     1.1913

------------------------------------------------------------------------------
             |               Robust
      lnwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]   Mean
-------------+----------------------------------------------------------------
        educ |   .1905978    .026438     7.21   0.000     .1387363    .2424594   11.5
       educ2 |  -.1621495   .0319977    -5.07   0.000     -.224917   -.099382    1    <- Std Variance
       exper |   .0239944   .0043647     5.50   0.000     .0154326    .0325562   12.7
      tenure |   .0065056   .0038795     1.68   0.094    -.0011046    .0141159    7.8
      female |  -.0717937   .0804414    -0.89   0.372    -.2295898    .0860024   45.8%
       _cons |   .4142694   .3241888     1.28   0.202    -.2216679    1.050207
------------------------------------------------------------------------------
Distributional Statistic: q(10) -> 10th : 15.85
10% increase in Educ variance -> -.016 Log points ~ 1.6% wage decline of the 10th per
```

# Example: rifhdreg

```
qregplot educ educ2 exper tenure female, hole(4) col(3) xsize(10) ysize(5) q(10(2.5)95)
```



# Example: rifhdreg

| | (1) gini | (2) Std | (3) IQRatio8020 |
|---|---|---|---|
| educ | -1.971*** | -0.0772*** | -0.0756** |
| | (0.404) | (0.0180) | (0.0237) |
| educ2 | 3.039*** | 0.100*** | 0.193*** |
| | (0.515) | (0.0227) | (0.0292) |
| exper | -0.406*** | -0.0123*** | -0.0239*** |
| | (0.0746) | (0.00226) | (0.00589) |
| tenure | 0.0560 | -0.000158 | 0.0205** |
| | (0.0871) | (0.00168) | (0.00696) |
| female | 3.097* | 0.0275 | 0.140 |
| | (1.438) | (0.0442) | (0.101) |
| _cons | 49.16*** | 1.521*** | 2.817*** |
| | (4.873) | (0.206) | (0.293) |
| N | 1434 | 1434 | 1434 |
| rifmean | 26.27 | 0.590 | 2.064 |

Standard errors in parentheses

p<0.05, ** p<0.01, *** p<0.001

# What about QTE?

# QTE via RIF

- As mentioned, Standard RIF regressions cannot be used to estimate "treatment" effects.
    - Changes in treatment Status is a large distributional change.
- However, one can use a modified version of RIF, conditioning only on 1 variable (treatment)

$$RIF(y|D) = \beta_0 + \beta_1 D + \beta_2 x + e$$

$$RIF(y|D=1) \times D + (1-D) \times RIF(y|D=0) = \beta_0 + \beta_1 D + \beta_2 x + e$$

- Thus we simply need to estimate the RIF Separately for treated and untreated groups, while controlling for other factors.
    - Other factors can be controlled linearly, with interactions, or IPW, or both.

# Example: QTE via RIF

code Plots Other

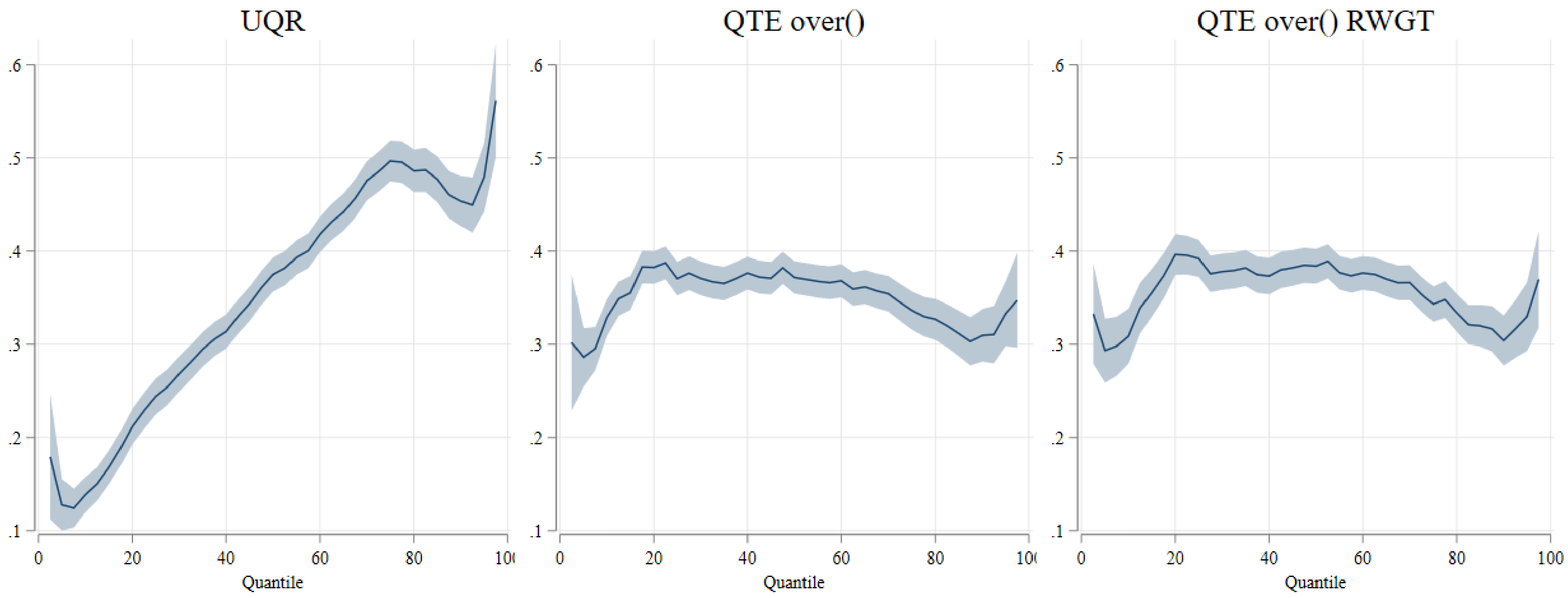
```
webuse nlswork, clear
  UQREG
rifhdreg ln_wage collgrad age i.race nev_mar union c_city ttl_exp, rif(q(10))
qregplot collgrad, q(2.5(2.5)97.5) name(m1, replace)

  QTE with Over()
rifhdreg ln_wage collgrad age i.race nev_mar union c_city ttl_exp, rif(q(10))  ///
   over(collgrad)
qregplot collgrad, q(2.5(2.5)97.5) name(m2, replace)

  QTE with Over() and RWGT
rifhdreg ln_wage collgrad age i.race nev_mar union c_city ttl_exp, rif(q(10)) ///
   over(collgrad) rwlogit( age i.race nev_mar union c_city ttl_exp)
qregplot collgrad, q(2.5(2.5)97.5) name(m3, replace)
```

# Example: QTE via RIF

code | Plots | Other



# Example: QTE via RIF

code Plots **Other**

| | Gini | TE-Gini | Std | TE-Std | IQRT | QTE-IQTR |
|---|---|---|---|---|---|---|
| collgrad | 5.517*** | -1.400* | 0.109*** | -0.009 | 0.574*** | -0.106*** |
| | (0.532) | (0.552) | (0.007) | (0.008) | (0.031) | (0.027) |
| age | 0.503*** | 0.380*** | 0.007*** | 0.006*** | 0.034*** | 0.022*** |
| | (0.048) | (0.049) | (0.001) | (0.001) | (0.003) | (0.002) |
| 1.race | | | | | | |
| 2.race | 1.125* | -0.064 | 0.017* | -0.004 | 0.244*** | 0.073** |
| | (0.488) | (0.506) | (0.007) | (0.007) | (0.028) | (0.024) |
| 3.race | 2.875 | 1.951 | 0.048 | 0.033 | 0.188 | 0.290** |
| | (1.990) | (2.063) | (0.028) | (0.028) | (0.116) | (0.100) |
| nev_mar | 1.756** | 1.296* | 0.039*** | 0.031*** | 0.103** | 0.104*** |
| | (0.557) | (0.578) | (0.008) | (0.008) | (0.032) | (0.028) |
| union | -0.767 | 0.482 | -0.015* | -0.002 | -0.001 | 0.046 |
| | (0.494) | (0.512) | (0.007) | (0.007) | (0.029) | (0.025) |
| c_city | -0.333 | -0.751 | -0.013* | -0.016* | -0.096*** | -0.058* |
| | (0.460) | (0.477) | (0.006) | (0.007) | (0.027) | (0.023) |
| ttl_exp | -0.074 | 0.161* | 0.000 | 0.003** | 0.012** | 0.023*** |
| | (0.062) | (0.064) | (0.001) | (0.001) | (0.004) | (0.003) |
| _cons | 9.830*** | 11.571*** | 0.213*** | 0.243*** | 0.857*** | 1.112*** |
| | (1.283) | (1.330) | (0.018) | (0.018) | (0.075) | (0.064) |
| **N** | 19215 | 19215 | 19215 | 19215 | 19215 | 19215 |
| rifmean | 26.508 | 24.563 | 0.468 | 0.437 | 2.171 | 2.000 |

# What about [K]OB Decompositions?

# RIF Decompositions

FFL(2018) proposes that it is possible to extend the use of RIF regressions to decomposition analysis, similar to what is commonly done with the Kitagawa-Oaxaca-Blinder decomposition.

They also suggest a refinement that introduces a reweighting step.

The logic is simple: As with RIF-regressions, simply swap the dependent variable with the corresponding RIF, and proceed as usual.

I would argue that KOB decompositions are in principle better at keeping the principles and assumptions of RIF-regressions:

- Analyze unconditional statistics, considering unconditional differences in distributions across two groups

However, Interpretation could be somewhat difficult.

# OB_RIF Decompositions

Step 1: RIF estimation

$$RIF_i(D=k) = \beta_{k0} + \beta_{k1}x_1 + \beta_{k2}x_2 + e_k \ \forall \ k = 0,1$$

Step 2: Express regressions in terms of unconditional means

$$F_{y|D=k} = F_y^k$$

$$E(RIF_i|D=k) = v(F_y^k) = \beta_{k0} + \beta_{k1}\bar{x}_1 + \beta_{k2}\bar{x}_2 = \bar{X}_k\beta_k = v( \ F^k_{y|x} \ , \ F^k_x)$$

This step explicitly says: group specific distributional statistics can be approximated as function of the Mean of the variables on control variables.

# OB_RIF Decompositions

Step 3: Decomposition

$$v(F_y^1) - v(F_y^0) = \bar{X}_1\beta_1 - \bar{X}_0\beta_0$$

$$v(F_y^1) - v(F_y^0) = (\bar{X}_1 - \bar{X}_0)\beta_0 + \bar{X}_1(\beta_1 - \beta_0)$$

$$v(F_y^1) - v(F_y^0) = \left[v(F^0_{y|x}, F^1_x) - v(F^0_{y|x},\ F^0_x)\right] + \left[v(F^1_{y|x}, F^1_x) - v(F^0_{y|x},\ F^1_x)\right]$$

Under certain assumptions, the last term can be use as a treatment effect on treated (k=1).

# OB_RIF Decompositions: Refinement

Using means may not be enough to capture differences in distributions across groups. Thus a better approach, Combine KOB with reweighted approach.

Step 1: logit/probit, and obtain IPW's

$$P(k=1|x) = G(X\beta)$$

Step 2: Estimate Auxiliary Regression (counterfactual) using IPW:

$$v\left(F^0_{y|x}, \tilde{F}^1_x\right) = \bar{X}^c_1\beta^c_0 \simeq \bar{X}_1\beta_0$$

$$v\left(F^1_{y|x}, \tilde{F}^0_x\right) = \bar{X}^c_0\beta^c_1 \simeq \bar{X}_0\beta_1$$

Step 3: Decomposition

$$\Delta X = \bar{X}^c_1\beta^c_0 - \bar{X}_0\beta_0 = (\bar{X}^c_1 - \bar{X}_0)\beta_0 + \bar{X}^c_1(\beta^c_0 - \beta_0)$$

$$\Delta\beta = \bar{X}_1\beta_1 - \bar{X}^c_1\beta^c_0 = \bar{X}_1(\beta_1 - \beta^c_0) + (\bar{X}_1 - \bar{X}^c_1)\beta^c_0$$

# OB_RIF Decompositions: Summary

In the RIF framework, we are attempting to determine what factors account for differences in the distribution between two groups: $\Delta v$.

This may be explained by:

- Differences in distribution of characteristics: differences in means.
- Differences in "returns" to those characteristics: Differences in $\beta's$ as proxy for Conditional distributions. $F(Y|X,D)$.

This can be done using `oaxaca_rif`

# Example: Oaxaca RIF

```
use http://fmwww.bc.edu/RePEc/bocode/o/oaxaca.dta, clear;gen wage=exp(wage);
oaxaca_rif wage educ exper tenure [pw=wt], rif( q(50) ) by(female)

RIF    : q(50)
Group 1: female = 0 x1*b1                       N of obs 1      = 751
Group c: x1*b2                                  N of obs C      =           .
Group 2: female = 1 x2*b2                       N of obs 2      = 683
------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   32.21719   .6021551    53.50   0.000     31.03698    33.39739
     group_2 |   26.31728   .5978081    44.02   0.000      25.1456    27.48897
  difference |   5.899902   .8485077     6.95   0.000     4.236857    7.562946
   explained |   2.394187   .4587227     5.22   0.000     1.495107    3.293267  (x1-x2)*b2
 unexplained |   3.505714   .7964557     4.40   0.000      1.94469    5.066739  ATT:x1*(b1-b2)
-------------+----------------------------------------------------------------
explained    |                                                                    M     F
        educ |    1.25169    .322888     3.88   0.000     .6188412    1.884539  (11.74-11.18)
       exper |   .6502986   .2382173     2.73   0.006     .1834013    1.117196  (13.74-11.46)
      tenure |   .4921986   .2726614     1.81   0.071    -.0422079    1.026605  ( 8.97- 6.39)
-------------+----------------------------------------------------------------
unexplained  |                                                                Diff in F_y|X
        educ |  -2.529993   3.086411    -0.82   0.412    -8.579247    3.519262
       exper |  -1.011601   1.330902    -0.76   0.447    -3.620122    1.596919
      tenure |   1.790919   1.120194     1.60   0.110    -.4046214    3.986459
       _cons |    5.25639   3.192297     1.65   0.100    -1.000397    11.51318  Diff b0
------------------------------------------------------------------------------
```

# Example: Oaxaca RIF

```
. oaxaca_rif wage educ exper tenure [pw=wt], rif( q(50) ) by(female) rwlogit(educ exper tenure)
Group 1: female = 0 x1*b1                         N of obs 1       = 751
Group c: X2~>rw~>X1 or x1*b2                      N of obs C       = 683
Group 2: female = 1 x2*b2                         N of obs 2       = 683

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Overall      |
     group_1 |   32.21719   .5686022    56.66   0.000     31.10275    33.33163   Men
     group_c |   28.87179   .5911752    48.84   0.000     27.71311    30.03047   Men Paid as Women
     group_2 |   26.31728   .5556553    47.36   0.000     25.22822    27.40635   Women
 tdifference |   5.899902   .7950228     7.42   0.000     4.341685    7.458118
 t_explained |   2.554504   .8113205     3.15   0.002     .9643449    4.144663
t_unexplained |  3.345398   .8202418     4.08   0.000     1.737753    4.953042   ATT
-------------+----------------------------------------------------------------
explained    |
       total |   2.554504   .8113205     3.15   0.002     .9643449    4.144663
 p_explained |    2.06466   .3891674     5.31   0.000     1.301906    2.827414
  specif_err |   .4898437    .770001     0.64   0.525     -1.01933    1.999018 -> Need Review Specification
-------------+----------------------------------------------------------------    Polynomial Interactions
unexplained  |
       total |   3.345398   .8202418     4.08   0.000     1.737753    4.953042
   rwg_error |   .2952259   .3468002     0.85   0.395      -.38449    .9749418 -> Similar to Balance Test
p_unexplained |  3.050172   .7424254     4.11   0.000     1.595045    4.505299
------------------------------------------------------------------------------
```

# Example: Other Stats

```
                     q10             q90             gini        iqratio8020
overall
group_1            18.43***        52.87***        23.78***        1.979***   Men
                  (0.867)         (1.177)         (0.826)        (0.0603)
group_2            13.90***        43.65***        28.20***        2.011***   Women
                  (0.875)         (1.192)         (1.279)        (0.0692)
difference         4.531***        9.221***       -4.422**        -0.0316
                  (1.232)         (1.676)         (1.523)        (0.0918)
explained          2.531***        2.816***       -1.552*         0.00544
                  (0.623)         (0.832)         (0.622)        (0.0368)
unexplained        2.000*          6.405***       -2.869          -0.0370
                  (1.014)         (1.917)         (1.537)        (0.0929)
----------------------------------------------------------------------------
explained
educ               0.934**         1.803**        -0.360           0.0116
                  (0.305)         (0.574)         (0.361)        (0.0171)
exper              1.307**         1.192*         -0.956*         -0.0424
                  (0.419)         (0.499)         (0.414)        (0.0249)
tenure             0.290          -0.179          -0.237           0.0362
                  (0.228)         (0.518)         (0.395)        (0.0305)
unexplained
educ               5.519           5.549          -0.533          -0.111
                  (5.816)         (10.61)         (9.030)         (0.462)
exper             -1.610          -9.895***       -0.862          -0.218
                  (1.927)         (3.000)         (2.351)         (0.161)
tenure             0.430           6.813**         1.282           0.0155
                  (0.966)         (2.508)         (1.637)         (0.125)
_cons             -2.339           3.938          -2.757           0.276
                  (7.570)         (10.18)         (9.700)         (0.519)
```

# Conclusions

- In this presentation, I provided a general review of RIF regressions, implementation and interpretation.
- RIF, by default estimates effects at the margin (UPE). But can be used to estimate distributional effects (over() and oaxaca_rif)
- IPW can be combined with RIF to estimate Distributional TE. But Standard errors need correction
- The implementation is straight forward with the commands `rifhdreg` and `oaxaca_rif`

# Thanks you!

If you have comments or questions, you can contact me at

email: friosa@gmail.com

twitter: friosavila

stataforum:FernandoRios

Site :playingwithStata

# References


Chung, Choe, and Philippe Van Kerm. 2018. "Foreign Workers and the Wage Distribution: What Does the Influence Function Reveal." Econometrics no. 6 (2):28.

Cowell, F. and E. Flachaire (2007). 'Income Distribution and Inequality Measurement: The problem of Extreme Values.' Journal of Econometrics, 141(2):1044-1072.

Deville, J.-C. 1999, ‘Variance estimation for complex statistics and estimators: linearization and residual techniques’, Survey Methodology 25, 193–204

Essama-Nssah, B. and Lambert, P. J. (2012), Influence functions for policy impact analysis, in J. A. Bishop and R. Salas, eds, ‘Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber’, Vol. 20 of Research on Economic Inequality, Emerald Group Publishing, chapter 6, pp. 135–159

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), ‘Unconditional quantile regressions’, Econometrica 77(3), 953–973.

Firpo, S., Fortin, N. M. and Lemieux, T. (2018), ‘Decomposing Wage Distributions using Recentered Influence Functions Regressions', Econometrics 6(41).

Firpo, S. P., and C. Pinto. (2016). "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures." Journal of Applied Econometrics no. 31 (3):457-486

Jann, Ben. 2008. "The Blinder-Oaxaca decomposition for linear regression models." Stata Journal no. 8 (4):453-479.