Unconditional Quantile Regressions (and RIF's)

When we care about everyone

Fernando Rios-Avila

Introduction

As we saw last class, conditional quantile regressions have only one purpose:

- Analyze relationships between conditional distributions.

This is a very useful tool!. As it allows you to move beyond Average relationships.

- How do people (who are not all average) would be affected by changes in Xs

There is a limitation, however. The effects you may estimate, will depend strongly on model specification.

- This is similar to OVB. Changing covariates could drastically change the conditional distributions and associated coefficients

What if, you are interested in distributional effects across the whole population! Not only a subsample?

E(q(y|X)) is not Q(y)

One common mistake one makes when analyzing QRegressions is to make interpretations as if the average effects on the qth conditional quantiles would be the same as the effect on the "overall" qth quantile.

Except for few cases (when Quantile regressions are not relevant), CQ effects do not translate **directly** into Changes into the unconditional quantile.

However, as a policy maker, this would be the most relevant estimand you may be interested in :

- How does improving education affect inequality?

- Would eliminating Unionization would increase wage inequality?

- Is there heterogeneity in consumption expenditure?

However, going from Conditional to unconditional statistics (not only Q) is not always straight forward.

Wait...What do we mean unconditional?

One of the questions I read a lot regarding UQR is what do we mean unconditional?

- This is perhaps a someone poor choice of words.

- Anytime we estimate **ANY** statistic, we condition on something.

  - We condition on all individual characteristics (including errors)

  - We condition on groups characteristics (CQREG and CEF)

  - or, We condition on all characteristics (distributions). We happen to call this, unconditional statistics.

- This, however, ***does*** make a big difference in interpretation.

From Condition on Individuals,

to conditioning on Distributions

$$y_i = b_0 + b_1 x_i + e_i + x_i e_i$$
$$\frac{dy_i}{dx_i} = b_1 + e_i$$
$$E(y_i | x_i = x) = b_0 + b_1 x$$
$$\frac{dE(y_i | x)}{dx} = b_1$$
$$E(E(y_i | x_i = x)) = E(y_i) = b_0 + b_1 E(x_i)$$
$$\frac{dE(y_i)}{dE(x_i)} = b_1$$

Same effects, but different interpretations (specially last one)

How are Unconditional effects Estimated?

Consider any distributional statistic v, which takes as arguments, all observations, density distributions $f()$, or cumulative distributions $F()$.

$$v = v(F_y) \ or \ v(f_y) \ or \ v(y_1, y_2, ..., y_n)$$

And to simplify notation, lets say this function is defined as follows:

$$v(f_y) = \int_{-\infty}^{\infty} h(y, \theta) f(y) dy$$

This simply considers distributional statistics v that can be estimated by simply integrating a transformation of $h(y, \theta)$ given a set of parameters θ.

But for now, lets consider only the Identify function $h(y, \theta) = y$

but...What about Controls??

Introducing controls

Assume there is a joint distribution of function $f(y, x)$, then

$$f(y, x) = f(y|x) f(x)$$
$$f(y) = \int f(y|x) f(x) dx$$

And all together:

$$v(f_y) = \int y \int f(y|x) f(x) dx \ dy$$
$$v(f_y) = \iint y f(y|x) dy \ f(x) dx$$
$$v(f_y) = \int E(y|X) f(x) dx$$

Or a bit more General

$$v(f_y) = \iint h(y,\theta)f(y|x)f(x)dxdy$$

So, the statistic v will change if:

- We change the function h or its parameters θ.

- Assume some shocks that change the conditional $f(y|x)$

- or the distribution of characteristics change!

Note:

$$f(y|x) \sim \beta \text{ and } f(x) \sim x$$

Again...How are Unconditional effects Estimated?

In an ideal scenario, you simple get the data under to regimes (before and after changes in x), and do the following:

$$\Delta v = v(f'_y) - v(f_y)$$

That is, just estimate the statistic in two scenarios (f' and f), and calculate the difference. (impossible!)

But there are (at least) three alternatives:

1. Using Reweighting approaches to "reshape" the data: $f(x)$ (non-parametric)
2. Identify $f(y|x)$ so one can simulate how ΔX affect y
3. Focus on the statistic v and indirectly identify the effects of interest. (RIF!)

Op1: Re-weighting

Consider the following. there is a policy such that you plan to improve education in a country. Every single person will have at least 7 years of education, and will have free access to two additional years of education if they want to.

In other words, characteristics change from $f(x) \rightarrow g(x)$. But you do not see this!

$$v(g_y) = \iint h(y,\theta)f(y|x)g(x)dxdy$$

but perhaps, we could see this:

$$\hat{v}(g_y) = \iint h(y, \theta) f(y|x) \hat{w}(x) f(x) dx dy$$

if we can come up with a set of weights $\hat{w}(x)$ such that $f(x)\hat{w}(x) = g(x)$

$$\hat{w}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}$$

Op1: Re-weighting

Simple, yet hard. Estimation of multivariate densities can be a difficult task. So assume the following

$$f(x) = h(x|s = 0); g(x) = h(x|s = 1)$$

This makes things "easier".

1. Bayes:
$$h(x|s = k) = \frac{h(x)p(s = k|x)}{p(s = k)}$$
$$\hat{w}(x) = \frac{h(x)p(s = 1|x)}{h(x)p(s = 0|x)} \frac{p(s = 0)}{p(s = 1)}$$
$$= \frac{p(s|x)}{1 - p(s|x)} \frac{1 - p(s)}{p(s)}$$

Easier to estimate conditional probabilities, (logit probit or other) than Densities

Example

```
. *** UQR
. *** Reweighting
. webuse dui, clear
(Fictional data on monthly drunk driving citations)

. ** Create Fake Sample
. gen id = _n

. expand 2
(500 observations created)
```

```
. bysort id:gen smp = _n ==2

. ** Now you have two of ever person. So lets do some Policy
. ** Fines increase lower fines more than higher ones, up to 12
. replace fines = 0.1*(12-fines)+fines if smp==1
(498 real changes made)

. ** Estimate logit
. logit smp c.fines##c.fines taxes i.csize college

Iteration 0:   log likelihood = -693.14718
Iteration 1:   log likelihood = -680.74735
Iteration 2:   log likelihood = -680.68931
Iteration 3:   log likelihood = -680.68931

Logistic regression                                  Number of obs =   1,000
                                                     LR chi2(6)    =   24.92
                                                     Prob > chi2   =  0.0004
Log likelihood = -680.68931                          Pseudo R2     =  0.0180


------------------------------------------------------------------------------
         smp | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       fines |   4.231032   1.762476     2.40   0.016     .7766421    7.685422
             |
c.fines#c.fines | -.1923691   .0877831    -2.19   0.028    -.3644209   -.0203173
             |
       taxes |     .00044   .1407545     0.00   0.998    -.2754338    .2763139
             |
       csize |
      Medium |   .0549913    .161242     0.34   0.733    -.2610373    .3710198
       Large |   .0399068   .1526637     0.26   0.794    -.2593086    .3391222
             |
     college |  -.0218305   .1488361    -0.15   0.883    -.3135439    .2698829
       _cons |  -22.99551   8.826624    -2.61   0.009    -40.29537   -5.695646
------------------------------------------------------------------------------

. predict pr_smp
(option pr assumed; Pr(smp))

. gen wgt = pr_smp / (1-pr_smp)
```

6

```
. replace wgt = 1 if smp==0
(500 real changes made)

.  xi:tabstat fines i.csize college  taxes [w=wgt],  by(smp)
i.csize             _Icsize_1-3        (naturally coded; _Icsize_1 omitted)
(analytic weights assumed)

Summary statistics: Mean
Group variable: smp

     smp |     fines  _Icsiz~2  _Icsiz~3   college     taxes
---------+------------------------------------------------
       0 |    9.8952       .29      .358      .248      .704
       1 |  10.24873  .2926133  .3584506  .2474627  .7045357
---------+------------------------------------------------
   Total |  10.07894  .2913582  .3582342  .2477208  .7042784
----------------------------------------------------------


two (kdensity citations if smp==0 ) ///
    (kdensity citations if smp==1 [w=wgt]) ///
    , legend(order(1 "Before Policy" 2 "After Policy"))

tabstat citations [w=wgt], by(smp) stats(p10 p25 p50 mean p75 p90  )
(analytic weights assumed)

Summary for variables: citations
Group variable: smp

     smp |       p10       p25       p50      Mean       p75       p90
---------+------------------------------------------------------------
       0 |      11.5        15        20    22.018        27      34.5
       1 |        11        15        19  20.69419        26        32
---------+------------------------------------------------------------
   Total |        11        15        20  21.32998        27        33
----------------------------------------------------------------------
```

Increasing fines, may reduce citations in about 1.3., but have almost no effect at the bottom of the distribution.

What about Standard errors? Bootstrap! (logit and estimation, probably clustering at individual level)

Easy to extend to other Statistics, but, can only provide results "within" support.

Op2: Model Conditional Distribution

Say that you are interested in the same Policy, but do not trust re-weighting. Instead you want to model the Outcome, using some parametric or nonparametric analysis

1. Define your model. Should be feasible enough to accommodate changes in the conditional distribution. (one "model" for each $X's$ combination?)
2. Use the model to make predictions of your outcome (quite a few times). and summarize all results.

Options for flexible mode?

- You can use Heteroskedastic OLS $y \sim N(x\beta, x\gamma)$ and predict from here

- You can use CQregressions to simulate the results.

One of this is similar to what we do in simulation analysis, and imputation. The other is similar to the work of Machado Mata (2005) and Melly(2005). Where you invert the whole distribution "globally"

Example #1: Hetregress

```
** Example for OPT2
webuse dui, clear
** Modeling OLS with heteroskedastic errors
    hetregress citations fines i.csize college taxes ,  het(fines i.csize college taxes )
```

```
-----------------------------------------------------------------------------
   citations | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------------
citations    |
       fines |   -6.18443    .3018298   -20.49   0.000    -6.776006   -5.592855
             |
       csize |
      Medium |   4.683941    .5028377     9.32   0.000     3.698397    5.669484
       Large |   9.655742    .5261904    18.35   0.000     8.624428    10.68706
             |
     college |   4.495635    .5283579     8.51   0.000     3.460072    5.531197
       taxes |  -3.640864    .4938209    -7.37   0.000    -4.608735   -2.672993
```

```
       _cons |    79.48011    3.118008     25.49   0.000     73.36892     85.59129
-------------+----------------------------------------------------------------
lnsigma2     |
       fines |   -.5261208     .082495     -6.38   0.000     -.687808    -.3644337
             |
       csize |
      Medium |     .331204    .1681709      1.97   0.049     .0015952     .6608129
       Large |    .5578834    .1662309      3.36   0.001     .2320768     .8836899
             |
     college |    .3186815    .1539424      2.07   0.038     .0169599     .6204032
       taxes |   -.3988692    .1437708     -2.77   0.006    -.6806548    -.1170836
       _cons |    8.257714    .8201063     10.07   0.000     6.650335     9.865093
------------------------------------------------------------------------------
LR test of lnsigma2=0: chi2(5) = 75.42                      Prob > chi2 = 0.0000
```

```
**  make Policy
clonevar fines_copy = fines
replace fines = 0.1*(12-fines)+fines

predict xb, xb
predict xbs, sigma

** Simulate results
capture program drop sim1
program sim1, eclass
    capture drop cit_hat
    gen cit_hat = rnormal(xb,xbs)
    qui:sum citations, d
    local lp10 = r(p10)
    local lp25 = r(p25)
    local lp50 = r(p50)
    local lpmn = r(mean)
    local lp75 = r(p75)
    local lp90 = r(p90)
    qui:sum cit_hat, d
    matrix b = r(p10)-`lp10',r(p25)-`lp25', r(p50)-`lp50' , r(mean) -`lpmn',r(p75)-`lp75',
    matrix colname b = p10 p25 p50 mean p75 p90
    ereturn post b
end
```

```
simulate, reps(1000): sim1
sum


------------+-------------------------------------------------------------
     _b_p10 |      1,000    -1.08147    .3913698   -2.31713    .1689796
     _b_p25 |      1,000   -.3262908    .3230118   -1.817808    .6465259
     _b_p50 |      1,000   -.2085465     .316455   -1.09237    .7785921
    _b_mean |      1,000   -1.675626    .2234377   -2.400322   -1.03909
     _b_p75 |      1,000   -1.541725    .4210822   -2.857586   -.2505198
------------+-------------------------------------------------------------
     _b_p90 |      1,000   -3.543298    .6079578   -5.464802   -1.682991
```

Effects larger than Reweigthing. Statistical inference here may be flawed. (first stage error
not carried over)

Example #2: Qregress

```
webuse dui, clear
gen id = _n
** Expand to 99 quantiles
expand 99
bysor id:gen q=_n
** make policy
gen fines_policy=0.1*(12-fines)+fines
gen fines_copy  =fines
** Estimate 99 quantiles (in theory one should do more..but choose at random)
ssc install qrprocess // Faster than qreg
** Save Cit hat (prediction)
** cit policy (with policy)
gen cit_hat=.
gen cit_pol=.

forvalues  i = 1 / 99 {
    if `i'==1   _dots 0 0
    _dots `i' 0
    qui {
        local i100=`i'/100
        capture drop aux
        qrprocess citations c.fines##c.fines  (i.csize college taxes) if q==1, q(`i100')
        ** predicts the values as if they were in q100
```

```
            predict aux
            replace cit_hat=aux if q==`i'
            drop aux
            replace fines = fines_policy
            predict aux
            replace cit_pol=aux if q==`i'
            replace fines = fines_copy
        }
}

  tabstat citations cit_hat cit_pol, stats(p10 p25 p50 mean p75 p90)

    Stats |   citati~s    cit_hat    cit_pol
  --------+------------------------------
      p10 |      11.5   10.70744   9.911633
      p25 |        15   15.42857   14.27302
      p50 |        20   21.15557   19.68303
     Mean |    22.018    22.0002   20.31824
      p75 |        27   27.65936   25.56173
      p90 |      34.5   34.03413   31.39192
  ----------------------------------------
```

Very demanding (computationally) and may only capture effects to the extend that we have good coverage of the distribution.

Standard Errors...Bootstrapping. Perhaps use random quantile assignment, and may have problems near boundaries.

Opt 1 and 2: Comments

The first two options allow you to estimate effects of changes in $f(x)$ on the unconditional distribution of y, and in consequence, the distributional statistics of interest.

They have limitations:

1. They both are limited to a single experiment. A different policy requires a change in the setup.
2. Reweighing is simple to apply, but has limitation on the type of policies. They all need to be within the support of X
3. Modeling the conditional distribution is a more direct approach, but more computationally intensive, specially for obtaining Stand errors.

Opt 3. Local Approximation: RIF regression

The third approach was first introduced by Firpo, Fortin and Lemieux 2009, as a computationally simple way to analyze how changes in $X's$ affect the unconditional quantiles of y.

This strategy was later extended to analyze the effects on a myriad of distributional statistics and rank dependent indices. as well as an approach to estimate distributional treatment effects. See Rios-Avila (2020).

In contrast with other approaches, it can be used to analyze multiple types of policies without re-estimating the model. However the identification and interpretation needs particular attention.

It also allows you to easily make Statistical inference. (except for quantiles...)

Opt 3. From ground up

Reconsider the Original question. How do you capture the effect of changes of distribution of x on the distribution of y.

$$\Delta v = v(G_y) - v(F_y)$$

Now, assume that G_y is just marginally different from F_y (different in a very particular way)

$$G_y(y_i) = (1 - \epsilon)F_y + \epsilon 1(y > y_i)$$

This function puts just a bit more weight on observation y_i. Think of it as "dropping" a new person in the pool.

If this is the case, the $\Delta v(y_i)$ Captures how would the Statistic v changes if the distribution puts just a bit extra weight on 1 observation. (this would be very small)

Opt 3. One more thing

Lets Rescale it:

$$IF(v, F_y, y_i) = lim_{\epsilon \to 0} \frac{v(G_y(y_i)) - v(F_y)}{\epsilon}$$

The **influence function** is a measure of direction of change, we should expect the statistic v will have as we change $F_y \to G_y$.

From here the RIF is just $RIF(v, F_y, y_i) = v + IF(v, F_y, y_i)$

Which has some properties:

$$\int IF(v, F_y, y_i) f_y dy = 0; \int RIF(v, F_y, y_i) f_y dy = v$$

$$v(F_y) \sim N\left(v(F_y), \frac{\sigma_{IF}^2}{N}\right); \int IF^2 f_y dy = \sigma_{IF}^2$$

Opt 3. RIF Regression

First:

$$v(F_y) = \iint RIF(v, F_Y, y_i) f(y|x) f(x) dy = \int E(RIF(.)|x) f(x)$$

From here is similar to Opt 3. Use **some** econometric model to estimate $E(RIF(.)|X)$, and use that to make predictions on how $v(F_y)$ would change, when there is a **distributional change** in X.

RIF-OLS: Unconditional effect!

$$RIF(v, F_y, y_i) = X\beta + e \rightarrow E(RIF) = v(F_y) = \bar{X}\beta \frac{dv(F_y)}{d\bar{X}} = \beta$$

Logic. When F_x changes, it will change the distribution of F_y, which will affect how the statistic v will change. But, we can only consider changes in means! (and Var)

Why it works, and why it may not

RIF regressions works by using a linear approximation of the statistic v with the changes in F_y which are caused by changes in F_x, proxied by changes in $\bar{X}$.

- Changes at the individual x_i are not interesting (in a population of 1million, what happens to person 99 may not be large enough to matter)

Depending on the model specification, however, we may only be able to identify changes in first and second moments of the distribution of x. (Mean and variance).

-

However, as any linear approximation to a non-linear function, the approximations are BAD when the changes in F_x are too large. The most relevant example...Dummies and treatment!

RIF-Reg and dummies

Dummies are a challenge. At individual or conditional level, we usually consider changes from 0 to 1 (off or on).

- For unconditional effects this is not correct (too large of a change) (No-one treated vs All treated). Thus you need to change the question...Not on and off changes, but Changes in proportion of treated!

  - Very important. a 1% increase in pop treated is different if current treatment is 10% or 90%.

- However, its possible to restructure RIF regressions to be partially conditional (Rios-Avila and Maroto 2023) (Combines CQREG with UQREG)

- Similar problems are experienced if the change in continuous variables is large!

  - Minor point. How do you construct RIFs? (analytically and Empirically)

Example

```
webuse dui, clear
**  Consider the policy change
gen change_fines= 0.1*(12-fines)
**  consider average change in fines.Since we are only considering this effect
sum change_fines

rifhdreg citations fines i.csize college taxes, rif(q(10))
est sto m1
rifhdreg citations fines i.csize college taxes, rif(q(50))
est sto m2
rifhdreg citations fines i.csize college taxes, rif(q(90))
est sto m3
** This are Rescaled to show true effect
rifhdreg citations fines i.csize college taxes, rif(q(10)) scale(.21048)
est sto m4
rifhdreg citations fines i.csize college taxes, rif(q(50)) scale(.21048)
est sto m5
rifhdreg citations fines i.csize collegetaxes, rif(q(90)) scale(.21048)
est sto m6

. esttab m1 m2 m3 m4 m5 m6, se mtitle(q10 q50 q90 r-q10 r-q50 r-q90) compress nogaps
```

```
----------------------------------------------------------------------------------------
                    (1)          (2)          (3)          (4)          (5)          (6)
                    q10          q50          q90        r-q10        r-q50        r-q90
----------------------------------------------------------------------------------------
fines           -4.476***    -6.700***    -9.887***    -0.942***    -1.410***    -2.081***
                (0.491)      (0.493)      (0.978)      (0.103)      (0.104)      (0.206)
1.csize              0            0            0            0            0            0
                  (.)          (.)          (.)          (.)          (.)          (.)
2.csize          4.603***     7.325***     6.370***     0.969***     1.542***     1.341***
                (0.963)      (0.966)      (1.917)      (0.203)      (0.203)      (0.404)
3.csize          6.504***    13.54***     12.97***     1.369***     2.851***     2.729***
                (0.914)      (0.917)      (1.820)      (0.192)      (0.193)      (0.383)
college          2.922**      5.948***     9.973***     0.615**      1.252***     2.099***
                (0.890)      (0.892)      (1.771)      (0.187)      (0.188)      (0.373)
taxes           -3.279***    -3.303***    -8.319***    -0.690***    -0.695***    -1.751***
                (0.842)      (0.844)      (1.676)      (0.177)      (0.178)      (0.353)
_cons           53.71***     81.04***    129.2***     11.30***     17.06***     27.20***
                (4.964)      (4.977)      (9.880)      (1.045)      (1.048)      (2.080)
----------------------------------------------------------------------------------------
N                  500          500          500          500          500          500
----------------------------------------------------------------------------------------
```

Consider the basic change. Fines increases in 1 unit, Cntys with taxes, increase 10%, etc

Or consider rescaled effects

How Do they Compare
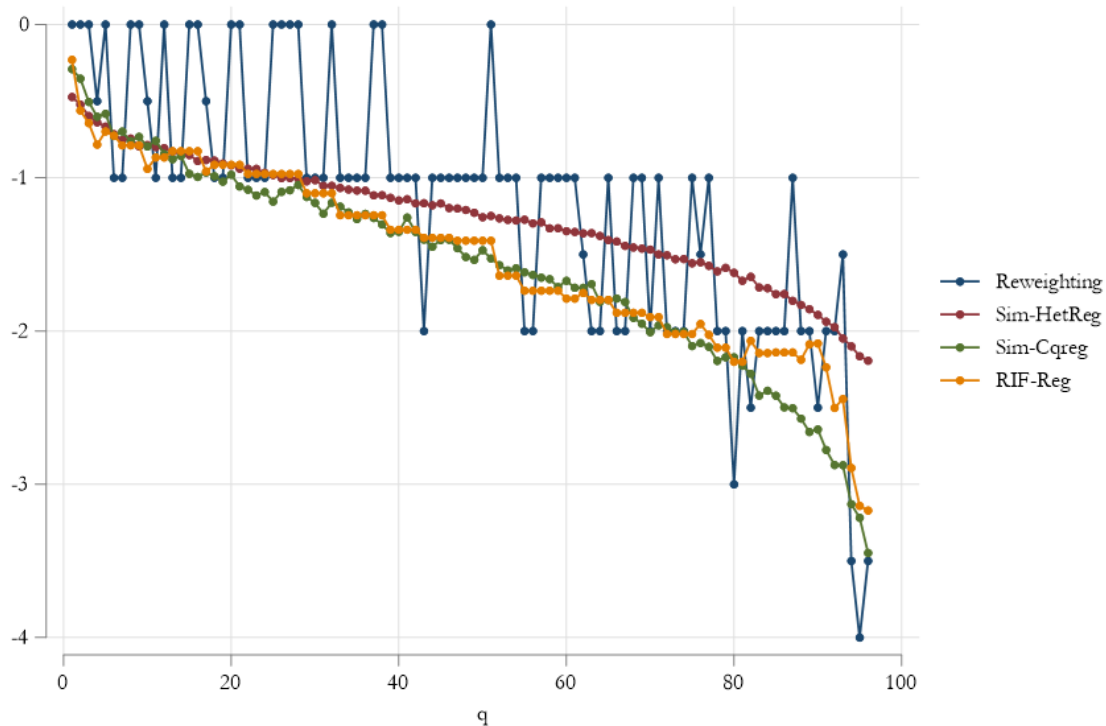
Other Considerations

RIF Regressions are useful, but again, one must use them with care.

- Only Small changes! Larger changes may be meaningless

Except for `Stata` (see rif and rifhdreg), the applications of RIF regressions outside Mean, Variance and Quantiles are non-existent. (paper?)

- For most Common Statistics, RIF's automatically provide correct Standard errors (which can be Robustized!). In fact, a simple **LR** can be considered as a special case of **RIF's**

$$RIF(mean, y_i, F_y) = y_i$$
$$RIF(variance, y_i, F_y) = (y_i - \bar{y})^2$$
$$RIF(Q, y_i, F_Y) = Q_y(\tau) + \frac{\tau - 1(y_i \leq Q_y(\tau))}{f_Y(y_i)}$$

Except for quantile related functions! (f_y also needs estimation, thus errors!)

- Accounting for "local" unconditional effects beyond means require Center Polynomials:

$$RIF(., y) = b_0 + b_1 x + b_2 (x - \bar{x})^2 + \varepsilon$$

- Quantile treatment effects (on and off) are possible using PC-RIF (When you condition the distribution on just 1 variable)

$$RIF(., F_{Y|D}, y) = b_0 + b_1 D + b_2 x + b_3 (x - \bar{x})^2 + \varepsilon$$

Final words on RIF

Because this implementation uses LR, you can add Multiple Fixed effects as well. (with limitations)

And you can skip LR all together, and model RIF using Other approaches! (which may be even better than OLS).

NEXT

Truly going nonlinear. When β is no longer linear in y (nor is the error)