R Example Quantile Regression and Testing with Quantreg

Fan Wang

2020-09-26

Contents

1 Quantile Regression Basics

Go to the **RMD**, **R**, **PDF**, or **HTML** version of this file. Go back to fan's REconTools Package, R Code Examples Repository (bookdown site), or Intro Stats with R Repository (bookdown site).

1.1 Estimate Mean and Quantile Coefficients using mtcars dataset

First, estimate the mean regression:

```
fit_mean <- lm(mpg ~ disp + hp + factor(am) + factor(vs), data = mtcars)
summary(fit_mean)
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + factor(am) + factor(vs), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7981 -1.9532  0.0111  1.5665  5.6321
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.832119   2.890418   8.591 3.32e-09 ***
## disp        -0.008304   0.010087  -0.823  0.41757
## hp          -0.037623   0.013846  -2.717  0.01135 *
## factor(am)1  4.419257   1.493243   2.960  0.00634 **
## factor(vs)1  2.052472   1.627096   1.261  0.21794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.812 on 27 degrees of freedom
## Multiple R-squared:  0.8104, Adjusted R-squared:  0.7823
## F-statistic: 28.85 on 4 and 27 DF,  p-value: 2.13e-09
```

Now estimate the quantile regressions at various quantiles, standard error obtained via bootstrap. Note that there is a gradient in the quantile hp coefficients as well as disp. disp sign reverses, also the coefficient on factor am is different by quantiles:

```
ls_fl_quantiles <- c(0.25, 0.50, 0.75)
fit_quantiles <- rq(mpg ~ disp + hp + factor(am),
               tau = ls_fl_quantiles,
               data = mtcars)
summary(fit_quantiles, se = "boot")
```

```
##
## Call: rq(formula = mpg ~ disp + hp + factor(am), tau = ls_fl_quantiles,
##     data = mtcars)
##
## tau: [1] 0.25
##
## Coefficients:
##             Value    Std. Error t value  Pr(>|t|)
## (Intercept) 25.34665  1.55372   16.31348  0.00000
## disp        -0.02441  0.00800   -3.05213  0.00494
## hp          -0.01672  0.01506   -1.10978  0.27654
## factor(am)1  1.39719  1.37941    1.01289  0.31979
##
## Call: rq(formula = mpg ~ disp + hp + factor(am), tau = ls_fl_quantiles,
##     data = mtcars)
##
## tau: [1] 0.5
##
## Coefficients:
##             Value    Std. Error t value  Pr(>|t|)
## (Intercept) 27.49722  1.89355   14.52154  0.00000
## disp        -0.02253  0.01558   -1.44637  0.15918
## hp          -0.02713  0.02327   -1.16575  0.25355
## factor(am)1  3.37328  1.96454    1.71708  0.09701
##
## Call: rq(formula = mpg ~ disp + hp + factor(am), tau = ls_fl_quantiles,
##     data = mtcars)
##
## tau: [1] 0.75
##
## Coefficients:
##             Value    Std. Error t value  Pr(>|t|)
## (Intercept) 28.06384  1.89727   14.79168  0.00000
## disp         0.00445  0.01490    0.29871  0.76736
## hp          -0.06662  0.01814   -3.67329  0.00100
## factor(am)1  7.91402  2.45424    3.22463  0.00320
```

1.2 Test Quantile Coefficients if Different

Use the rq.anova function frm the quantile regression packge to conduct WALD test. Remember WALD test says given unrestricted model's estimates, test where null is that the coefficients satisfy some linear restrictions.

To test, use the returned object from running rq with different numbers of quantiles, and set the option *joint* to true or false. When joint is true: "equality of slopes should be done as joint tests on all slope parameters", when joint is false: “separate tests on each of the slope parameters should be reported”. A slope parameter refers to one of the RHS variables.

Note that quantile tests are “parallel line” tests. Meaning that we should except to have different x-intercepts for each quantile, because they represents the levels of the conditional shocks distributions. However, if quantile coefficients for the slopes are all the same, then there are no quantile specific effects, mean effects would be sufficient.

see: - anova.rq() in quantreg package in R

1.2.1 Test Statistical Difference between 0.25 and 0.50

Given the quantile estimates above, the difference between 0.25 and 0.50 quantiles exists, but are they sufficiently large to be statistically different? What is the p-value? Reviewing the results below, they are not statistically different.

First, joint = TRUE. This is not testing if the coefficien on disp is the same as the coefficient on hp. This is testing jointly if the coefficients for different quantiles of disp, and different quantiles of hp are the same for each RHS variable.

```
ls_fl_quantiles <- c(0.25, 0.50)
fit_quantiles <- rq(mpg ~ disp + hp + factor(am),
               tau = ls_fl_quantiles,
               data = mtcars)
anova(fit_quantiles, test = "Wald", joint=TRUE)
```

```
## Quantile Regression Analysis of Deviance Table
##
## Model: mpg ~ disp + hp + factor(am)
## Joint Test of Equality of Slopes: tau in {  0.25 0.5  }
##
##   Df Resid Df F value Pr(>F)
## 1  3       61  0.7986 0.4994
```

Second, joint = False:

```
anova(fit_quantiles, test = "Wald", joint=FALSE)
```

```
## Quantile Regression Analysis of Deviance Table
##
## Model: mpg ~ disp + hp + factor(am)
## Tests of Equality of Distinct Slopes: tau in {  0.25 0.5  }
##
##             Df Resid Df F value Pr(>F)
## disp         1       63  0.0304 0.8621
## hp           1       63  0.5397 0.4653
## factor(am)1  1       63  1.0957 0.2992
```

1.2.2 Test Statistical Difference between 0.25, 0.50 and 0.75

The 1st quartile and median do not seem to be statistically different, now include the 3rd quartile. As seen earlier, the quartiles jointly show a gradient. Now, we can see that idisp, hp and am are separately have statistically different

First, joint = TRUE:

```
ls_fl_quantiles <- c(0.25, 0.50, 0.75)
fit_quantiles <- rq(mpg ~ disp + hp + factor(am),
               tau = ls_fl_quantiles,
```

```r
                data = mtcars)
anova(fit_quantiles, test = "Wald", joint=TRUE)
```

```
## Quantile Regression Analysis of Deviance Table
##
## Model: mpg ~ disp + hp + factor(am)
## Joint Test of Equality of Slopes: tau in {  0.25 0.5 0.75  }
##
##   Df Resid Df F value   Pr(>F)
## 1  6       90   3.957 0.001475 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Second, joint = False:

```r
anova(fit_quantiles, test = "Wald", joint=FALSE)
```

```
## Quantile Regression Analysis of Deviance Table
##
## Model: mpg ~ disp + hp + factor(am)
## Tests of Equality of Distinct Slopes: tau in {  0.25 0.5 0.75  }
##
##              Df Resid Df F value    Pr(>F)
## disp          2       94  9.2284 0.0002191 ***
## hp            2       94  6.5798 0.0021162 **
## factor(am)1  2       94  3.6669 0.0292803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```