Regression rank-scores tests in R

Jan Dienstbier

Jan Picek

contact: jan.picek@vslib.cz

Charles University, Prague
Technical University of Liberec

Czech Republic

UseR! 2006, Vienna

Quantile regression

Consider a linear model $Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + e_i$, where $e_i \sim F$ are $i.i.d.$

Definition: Regression quantiles

$$\hat{\boldsymbol{\beta}}(\tau) := \arg\min_{\boldsymbol{b} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - \mathbf{x}_i^\top b \right),$$

where ρ_τ denotes loss function

$$\rho_\tau(u) := u \cdot (\tau - I(u < 0)).$$

Simple example – quadratic regression

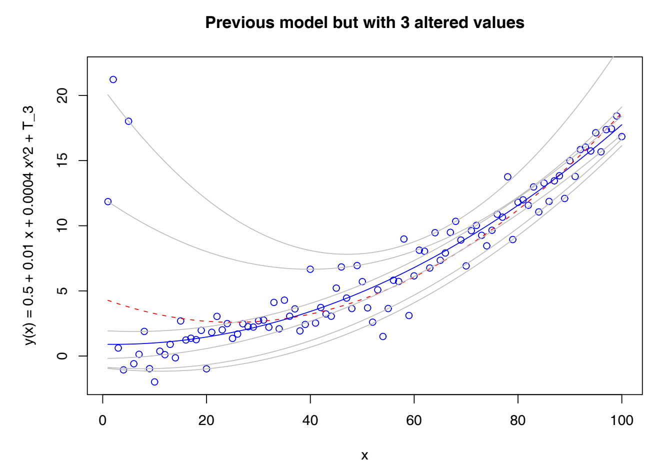
Regession quantiles are:

- direct generalization of “quantile principle” in a linear model
- robust as much as ordinary quantiles

Simple example – quadratic regression

Regession quantiles are:

- direct generalization of “quantile principle” in a linear model
- robust as much as ordinary quantiles



Regression ranks

Definition: Regression rank scores

$$\hat{a}(\tau) = \arg\max_{a \in \mathbb{R}^n} y^\top a$$

in conditions that

$$X^\top a = (1-\tau) X^\top 1_n, \ a \in [0,1]^n.$$

- solution of the dual linear programming problem
- behave similary as ordinary ranks ⇒ regression rank tests

Regression rank-scores tests - The Mechanism

1. calculate regression rank scores $\hat{a}(\tau)$ defined above
2. choose the proper score function φ – this determines the form of the test and should be done in respect to the used data
   - *usual selections are eg. logistic (Wilcoxon), normal (van der Waerden) or sign scores*
3. calculate scores $\hat{b}_{ni}$, $i = 1, \dots, n$

$$\hat{b}_{ni} = -\int_0^1 \varphi(u) d\hat{a}_{ni}(u), \quad i = 1, \dots, n$$

4. plug this to the statistic invariant to regression – e.g.

$$S_{n0}(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^{n} Y_i \hat{b}_{ni} = n^{-1} \mathbf{Y}^\top \hat{\mathbf{b}}_n$$

An example - Goodness-of-fit test

Example

In the linear model $Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma e_i$, where $e_i \sim F$ is continuous distribution it holds under $\mathbf{H}_0 : \ F(e) \equiv F_0(e/\sigma)$

$$T_n^* = n^{1/2} \left\{ \log \frac{S_{n0}(\mathbf{Y})}{\hat{\beta}_1(3/4) - \hat{\beta}_1(1/4)} \right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \xi^2(F_0)\right),$$

Common properties of such tests:

1. quite robust – they are not affected by a heavy tailed F
2. independent to regression parameters $(\boldsymbol{\beta}, \sigma)$ – they needn't to be estimated

. . . in R language it means . . .

1. useR! library `quantreg`, where are implemented basic methods of quantile regressions (regression ranks included)

. . . IN R LANGUAGE IT MEANS . . .

1. useR! library `quantreg`, where are implemented basic methods of quantile regressions (regression ranks included)
2. scores can be computed via `ranks`
   - e.g. `ranks(rq(data ~ covar, tau=-1, score="wilcoxon"))`

. . . IN R LANGUAGE IT MEANS . . .

1. useR! library `quantreg`, where are implemented basic methods of quantile regressions (regression ranks included)
2. scores can be computed via `ranks`
   - e.g. `ranks(rq(data ~ covar, tau=-1, score="wilcoxon"))`
3. score function can be specified as a parametr of `ranks`
   - e.g. `ranks(v, score="wilcoxon", tau=0.5)`

. . . IN R LANGUAGE IT MEANS . . .

1. useR! library `quantreg`, where are implemented basic methods of quantile regressions (regression ranks included)
2. scores can be computed via `ranks`
   - e.g. `ranks(rq(data ~ covar, tau=-1, score="wilcoxon"))`
3. score function can be specified as a parametr of `ranks`
   - e.g. `ranks(v, score="wilcoxon", tau=0.5)`
4. plug this to a suitable statistic and compare it with it's proper asymptotical p-value
   - eg. properly normalized T_n^* compare with `1 - pnorm(0.95)...`

. . . IN R LANGUAGE IT MEANS . . .

1. useR! library `quantreg`, where are implemented basic methods of quantile regressions (regression ranks included)
2. scores can be computed via `ranks`
   - e.g. `ranks(rq(data ~ covar, tau=-1, score="wilcoxon"))`
3. score function can be specified as a parametr of `ranks`
   - e.g. `ranks(v, score="wilcoxon", tau=0.5)`
4. plug this to a suitable statistic and compare it with it's proper asymptotical p-value
   - eg. properly normalized T_n^* compare with `1 - pnorm(0.95)...`

As an example we implemented described GOF test.

Autoregression

Regression rank scores concept can be generalized on AR time series

$$X_t = \theta_1 X_{t-1} + \ldots + \theta_p X_{t-p} + \epsilon_t, \quad t = 0, \pm 1, \pm 2 \ldots$$

Autorergression rank scores tests proposed in literature

- independence of two AR time series
- hypothesis $AR(p-1)$ against $AR(p)$

. . . other tests can be derived from quite general theory Hallin and Jurečková (1997).

Autoregression

Regression rank scores concept can be generalized on AR time series

$$X_t = \theta_1 X_{t-1} + \ldots + \theta_p X_{t-p} + \epsilon_t, \quad t = 0, \pm 1, \pm 2 \ldots$$

Autorergression rank scores tests proposed in literature

- independence of two AR time series
- hypothesis $AR(p-1)$ against $AR(p)$

. . . other tests can be derived from quite general theory Hallin and Jurečková (1997).

- we implemented these tests in `R`
- basic procedures taken from `quantreg`
- only minor changes needed

Proposals – possible improvements of implementation of regression rank scores in R

Althought there are excellent basic algoritms in `quantreg` the implementation of regression rank tests is still little bit a "stub".

We propose:

- eliminate gaps in the implementation of the linear submodel hypothesis
  - *basic procedure* `rrs.test.` *doesn't show p-values, user must know little bit more about the test to use it*
  - *implementation of the test in* `anova.rq` *isn't user-friendly, sometimes is hard to understand, what results user gets*
- more than three types of score functions – make it universally
- direct treatment of autoregression series with `quantreg`

Quantile regression ranks and R

- with `quantreg` package, there can be easily implemented various regression rank tests
- procedures are universal enough to extend these concept even on the AR series
- surprisingly large scale of hypothesis - eg. GOF, AR independence

Quantile regression ranks and R

- with quantreg package, there can be easily implemented various regression rank tests
- procedures are universal enough to extend these concept even on the AR series
- surprisingly large scale of hypothesis - eg. GOF, AR independence

Authors hope, that thanks to R it will be possible to evaluate, whether these tests interesting from the theoretical point of view can be used in the daily praxis.

Regression rank-scores tests in R

Jan Dienstbier
Jan Picek
contact: jan.picek@vslib.cz

Charles University, Prague
Technical University of Liberec

Czech Republic

UseR! 2006, Vienna