Hill's estimator based on two-step regression quantiles

J. Picek [1]

[1] Department of Applied Mathematics, Technical University of Liberec, Hálkova 6, CZ-461 17 Czech Republic

Keywords: Hill's estimator, regression quantile, two-step regression, R-estimator.

1 Abstract

Jurečková and Picek (2005) proposed a two-step regression quantiles in the linear regression model, ordering the residuals with respect to an initial R-estimate of the slope parameters. In this way they obtained a consistent estimator of $(\beta_0 + F^{-1}(\alpha), \beta_1, \ldots, \beta_p)'$, asymptotically equivalent to the regression α-quantile of Koenker and Bassett (1978).

Consider the linear regression model

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \tag{1}$$

with observations $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, i.i.d. errors $\mathbf{E} = (E_1, \ldots, E_n)'$ with an unknown distribution function F, and unknown parameter $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \ldots, \beta_p)'$. The $n \times p$ matrix $\mathbf{X} = \mathbf{X}_n$ is known and $\mathbf{1}_n = (1, \ldots, 1)' \in R^n$. The main point is to use a suitably chosen R-estimator $\widehat{\boldsymbol{\beta}}_{nR}$ (rank-estimator) as the initial estimate of $\boldsymbol{\beta}$, that itself is very close to the slope component of the α-regression quantile, and then order the residuals. Because the ranks are invariant to the shift, the R-estimator automatically estimates only the slope parameters in model (1). The possible R-estimator of $\boldsymbol{\beta}$ can be defined as follows:

$$\widehat{\boldsymbol{\beta}}_{nR} = \mathrm{argmin}_{\mathbf{b} \in R^p} \mathcal{D}_n(\mathbf{b}), \tag{2}$$

where

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i'\mathbf{b}) \varphi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})}{n+1} \right) \tag{3}$$

is Jaeckel's measure of rank dispersion (Jaeckel (1972)),

$$\varphi_\alpha(u) = \psi_\alpha(F_\alpha^{-1}(u)) = \alpha - I[u < \alpha], \ 0 < u < 1, \tag{4}$$

is the score function and $R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})$ is the rank of $Y_i - \mathbf{x}_i'\mathbf{b}$ among $(Y_1 - \mathbf{x}_1'\mathbf{b}, \ldots, Y_n - \mathbf{x}_n'\mathbf{b})$, $\mathbf{b} \in R^p, \ i = 1, \ldots, n$. $\widehat{\boldsymbol{\beta}}_{nR}$ estimates only the slope parameters, and there is no need to estimate the intercept for its computation. Having estimated $\boldsymbol{\beta}$ by R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$, consider the estimation problem of β_0. Its solution, denoted as $\hat{\beta}_{n0}$, is the $[n\alpha]$-th order statistic of the residuals $Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}, \ i = 1, \ldots, n$. Jurečková and Picek called the vector $\widehat{\boldsymbol{\beta}}_n^{(1)}(\alpha) = \left(\hat{\beta}_{n0}, \widehat{\boldsymbol{\beta}}_{nR}'\right)'$ the two-step α-regression quantile and showed that $\widehat{\boldsymbol{\beta}}_n^{(1)}(\alpha)$ very closely approximates the regression quantile $\widehat{\boldsymbol{\beta}}_n^{*}(\alpha)$. Similarly it is possible to construct a version of the autoregression quantile in the linear AR(p) model.

The $[n\alpha]$-order statistic $\hat{\beta}_{n0}$ of the residuals $Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}, \ i = 1, \ldots, n$ very closely approximates $E_{n:[n\alpha]} + \beta_0$, where $E_{n:1} \leq \ldots \leq E_{n:n}$ are the order statistics of errors $E_1, \ldots, E_n$. In this way we can also estimate the order statistics of unobservable errors $E_{n:1}, \ldots, E_{n:n}$. For example, we denote the estimate of $E_{n:n} + \beta_0$ as $\hat{E}_{n:n}$:

$$\hat{E}_{n:n} = \max\{Y_1 - \mathbf{x}_1'\widehat{\boldsymbol{\beta}}_{nR}, \ldots, Y_n - \mathbf{x}_n'\widehat{\boldsymbol{\beta}}_{nR}\}. \tag{5}$$

Estimates of extreme errors provide a tool for an inference on the tails of the distribution of the errors; for instance, the tail index of the distribution of errors can be estimated using the Hill's estimator based on estimated higher order quantiles:

$$T_H(k) = \frac{1}{k}\sum_{i=1}^{k} \log \hat{E}_{(n-i+1:n)} - \log \hat{E}_{(n-k:n)}, \quad k = 1, \ldots, n-1. \tag{6}$$

In this contribution we shall discuss properties of estimator (6). We also compare our estimator (6) with the approach of Resnick and Stărică (1997) in in the linear AR(p) model.

References

B.M. Hill (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3, 1163–1174.

L.A. Jaeckel (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, 43, 1449–1459.

J. Jurečková and J. Picek (2005). Quantiles and R-estimates in linear models, submitted.

R. Koenker and G. Bassett (1978). Regression quantiles. *Econometrica*, 46, 33–50.

S. Resnick and C. Stărică (1997). Asymptotic behavior of Hill's estimator for autoregressive data. *Comm. Statist. Stochastic Models*, 13, 703–721.