

EXTREMAL QUANTILE REGRESSION INFERENCE: AN R VINAIGRETTE

ROGER KOENKER

1. INTRODUCTION

Confidence intervals for extreme quantiles require a distinct theory from that describing the behavior of central quantiles. This theory has been pioneered for quantile regression by Victor Chernozhukov in his Phd thesis and a considerable body of subsequent work, as reviewed in Chernozhukov et al. (2018). In this note I would like to describe very briefly the implementation of some of these methods in the R package **quantreg**.

2. CONFIDENCE INTERVALS FOR EXTREME REGRESSION QUANTILES

In accordance with general R principles inference about various linear model coefficient estimates is available by subjecting a fitted object containing point estimates to a **summary** function. In the case of fitted objects produced by the function **rq** in the **quantreg** there are already quite a few options to choose among; the latest addition to these methods is the option **se = "extreme"**. It uses a relatively new bootstrapping option described in Chernozhukov et al. (2019) to do subsampling and produces percentile intervals for each of the estimated coefficients. There are several tuning parameter choices involved in this procedure: the size of the subsamples, denoted by **mofn**; a bandwidth, denoted **kex**; and the number of bootstrap replications, denoted by **R**. These parameters are described in more detail in the documentation of the function **summary.rq** and of course in the code itself.

To illustrate the use of this option we will consider the financial application in Chernozhukov et al. (2018), investigating value-at-risk estimates for Citigroup return from 2009 to 2015. The data is constructed as described in the supplemental material for the Handbook chapter and stored as the compressed image **D.Rda**. Then the code to reproduce a version of Figure 18.2 of their chapter is quite simple.

```
library(quantreg)
load("D.Rda")
p <- ncol(D)
mofn <- floor(50 + sqrt(nrow(D)))
taus <- (1:99)/100
fit <- rq(Y ~ ., tau = taus, data = D)
extreme <- summary(fit, se = "extreme", mofn = mofn, kex = 5+p)
pfit <- plot(extreme)
```

October 1, 2020. Thanks to Victor Chernozhukov and Ivan Fernandez-Val for making their code available and suggesting an implementation in **quantreg**. Thanks too to Shuowen Chen for helpful comments on an earlier version of the **quantreg** implementation.

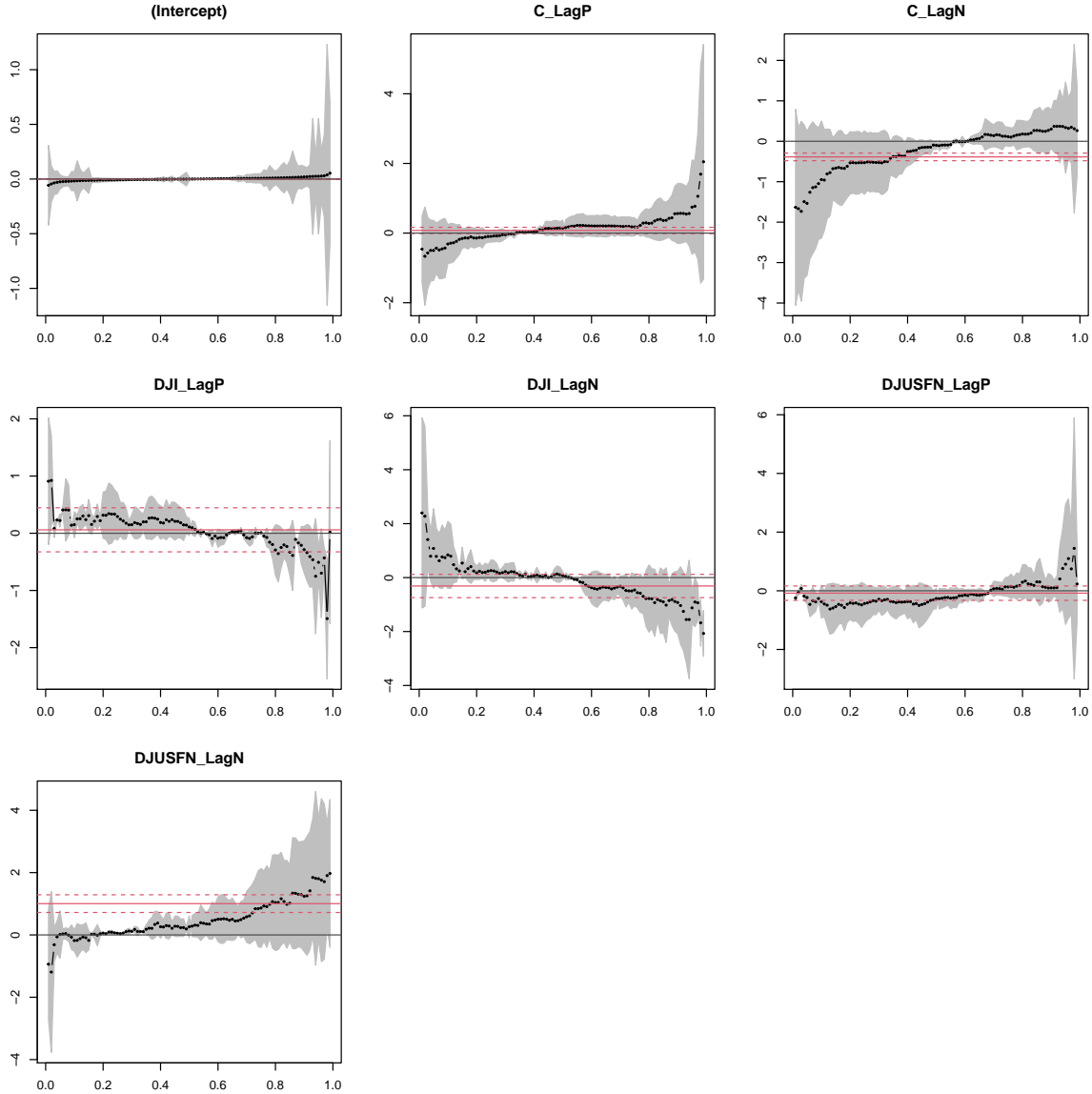


FIGURE 1. Extreme Value Quantile Inference for Citigroup Data: The point-wise bands illustrated here use the bootstrap method of Chernozhukov et al

```
central <- summary(fit, se = "ker")
plot(central, ylim = pfit$Ylim) # Use "extreme" plotting limits
```

One shouldn't draw any firm conclusions from a single realization like this, but I hope that the new syntax might encourage someone to make a more systematic comparison. It is slightly odd that the extremal method produces a few very short intervals for central quantiles, but of course it is not intended to be used for central quantiles. Chernozhukov et al. (2018) suggest the rule of thumb that $\tau n/p < 30$ as a threshold for the use of the extremal intervals.

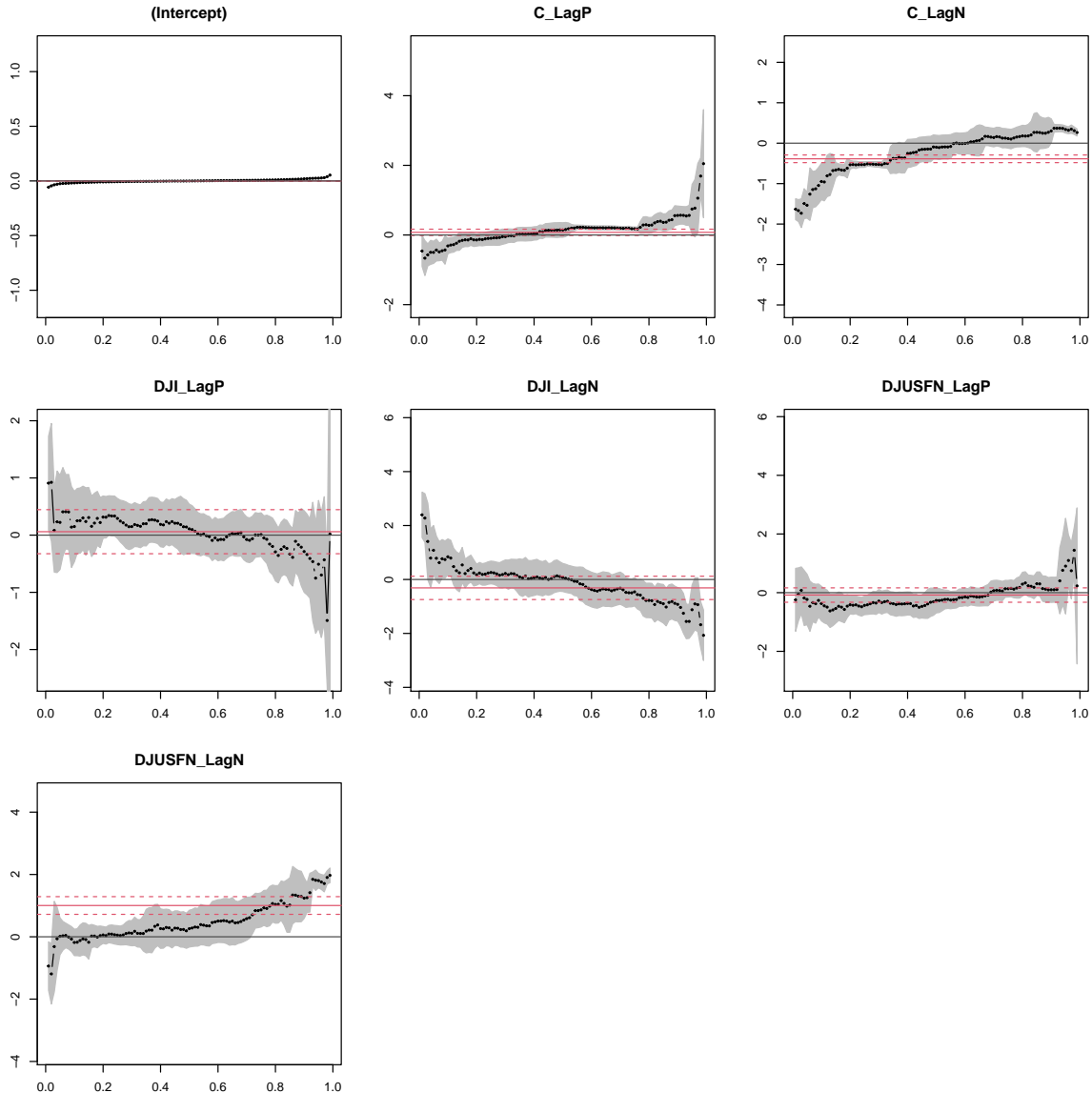


FIGURE 2. Conventional Central Quantile Inference for Citigroup Data: The pointwise bands illustrated here use the kernel method of Powell.

3. ESTIMATING PARETO TAIL EXPONENTS

The other new feature introduced for handling extremal quantile regression problems is a pair of functions for estimating the tail index of the response observations. These are implemented in a umbrella function `ParetoTest` with options for the Hill (1975) method and the Pickands (1975) method. To explore the performance of these two methods a very simple simulation experiment was undertaken. The setup is a very simple bivariate linear model with iid Student t_2 errors. The Pareto tail exponent is thus $1/2$, and we would hope and expect that both methods would produce estimates and confidence intervals that reflected this target

	Hill	Pickands
Estimate	0.583	0.333
BC-Estimate	0.507	0.375
Lower	0.399	0.068
Upper	0.604	0.685

TABLE 1. Simulation Results for the ParetoTest Function: A simple bivariate linear model with iid Student t errors with two degrees of freedom is considered. Mean estimates based on 500 replications for Hill and Pickands estimates are reported as well as bias-corrected estimates and percentile confidence bounds based on 200 bootstrap replications.

value. As can be seen in the summary table, the Hill estimator is quite well centered at the desired value, but the Pickands estimator is substantially downward biased, as is its bias corrected version and confidence interval. This is confirmed in an examination of histograms of the 500 replications of the two estimators illustrated in Figure 3.

```
# Toy Simulation of Hill vs Pickands for t2 model
require(quantreg)
set.seed(1729)
n = 2000
R = 500
A = array(0,c(R,6,2))
x = rnorm(n)
for(i in 1:R){
  y = x + rt(n,2)
  A[i,,1] = ParetoTest(y ~ x, 0.05, flavor = "Hill")$z
  A[i,,2] = ParetoTest(y ~ x, 0.05, flavor = "Pickands")$z
}
B = apply(A,2:3,mean)[1:4,]
dimnames(B) = list(c("Estimate", "BC-Estimate", "Lower", "Upper"),
                  c("Hill", "Pickands"))

par(mfrow = c(1,2))
load("sim1.Rda")
hist(A[,1,1],40, main = 'Hill Estimates', xlab = 'Estimate')
hist(A[,1,2],40, main = 'Pickands Estimates', xlab = 'Estimate')
```

4. CONCLUSION

Inference for extremal quantile regression is an important practical objective that has heretofore been neglected in the **quantreg** package. It is to be hoped that making the implementations described above available will encourage others to explore and improve upon these features.

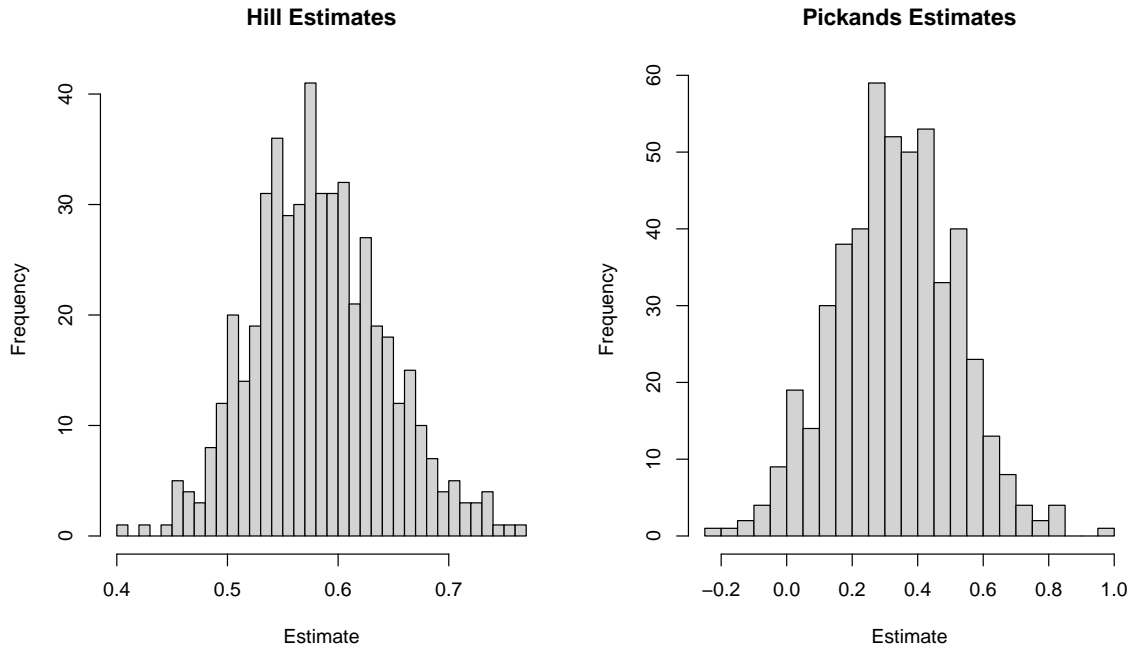


FIGURE 3. Histograms of 500 Hill and Pickands estimates of a Pareto Tail Exponent: The simulation setting has target tail exponent of $1/2$, so the Hill estimator is quite well centered, however the Pickands estimator is somewhat downwardly biased.

REFERENCES

- Chernozhukov, V., Fernandez-Val, I. and Kaji, T. (2018), Extremal quantile regression, *in* R. Koenker, V. Chernozhukov, X. He and L. Peng, eds, ‘Handbook of Quantile Regression’, CRC Press.
- Chernozhukov, V., Fernandez-Val, I. and Melly, B. (2019), ‘Fast algorithms for the quantile regression process’. available from: arXiv, 1909.05782.
- Hill, B. M. (1975), ‘A simple general approach to inference about the tail of a distribution’, *The Annals of Statistics* **3**, 1163–1174.
- Pickands, J. (1975), ‘Statistical inference using extreme order statistics’, *The Annals of Statistics* **3**, 119–131.