A Quantile Regression Memoir[1]

Gilbert W. Bassett Jr. and Roger Koenker

In the summers of 1972 and 1973 the two of us spent a lot of time playing tennis, in a successful effort to avoid working on our dissertations at the University of Michigan. Gib was working with Lester Taylor on theoretical aspects of l_1 regression, and Roger on hierarchical models for longitudinal data. Inevitably our anxiety about work intruded into the tennis conversation and there were frequent discussions of linear programming aspects of the l_1 regression problem. Gib had derived conditions under which the l_1 estimator was linear in the response vector, which explained some pathological simulation results of Taylor's. This might be called "breakdown" of the estimator due to influential design points now. More significantly, we frequently mentioned that the l_1 estimator seemed to be a regression analogue of the median since it was easily shown that essentially half the regression responses must lie above the fitted l_1 regression hyperplane and half must lie below as long as there was an intercept in the model. We also began to ask ourselves the question: if the l_1 estimator is a median regression estimator, mustn't there be other quantile regression estimators?

In the fall of 1973 we had both accepted positions at the University of Illinois, Gib at the University of Illinois at Chicago, and Roger at Urbana-Champaign, but we continued to discuss our research via the dedicated "WATS" line that connected the two campuses by telephone. At some point we asked ourselves: "suppose instead of weights 1 and -1 on positive and negative residuals as in median regression, or weights 0 and -1 as in a proposal of Aigner and Chu (1968) for an extremal estimator, we used weights, τ and $\tau - 1$ for τ in $(0, 1)$, couldn't we show that roughly, τn of the observations would lie below the fitted plane and $(1 - \tau)n$ above? The affirmative answer seemed to resolve our old question about how to define the rest of the regression quantiles, and we began an intensive effort to understand better how they behaved. Only later did we recall that we had both done an exercise on a one-sample version of this idea from Ferguson's decision theory text in a course given by Bruce Hill. And much, much later we discovered that the univariate germ of this idea in an influential paper of Edgeworth (1888).

Neither of us were asymptotically adept, but our background in economics and Gib's thesis work did provide a useful foundation on the linear programming aspects of the problem. Rather naively we began to attack the asymptotic theory via a combinatorial approach to the finite sample density. Our expression for the finite sample density didn't seem to be terribly practical since it required, when there were p parameters, summation of $\binom{n}{p}$ terms involving exact fits to all "elementary subsets" of p observations. However, eventually we were able to show limiting normality of the joint density of several regression quantiles for certain replicated designs. We presented a early version of this at the Winter Meetings the Econometric Society in San Francisco 1974. By a fortuitous circumstance the discussant on that occasion was Joe Gastwirth, who was very encouraging and suggested that we explore connections to the then rapidly expanding robustness literature. In January of 1975 we submitted our paper to *Econometrica* and in due course we we received a report stating, in essence, that we had failed to make a convincing case that $\tau \neq \frac{1}{2}$ was "interesting", but perhaps a revision could be considered. We were consequently pessimistic about our prospects at *Econometrica* so in June after some further revision we submitted the paper to the *Annals of Statistics.* The sole referee report, quoted in its entirety, was far briefer, but the message was the same:

> "I regret that I cannot see any point in this paper, and therefore cannot recommend its publication. It may be of interest to compute regression analyses to minimize the sum of absolute deviations between the observed and fitted responses, and there is a fair amount of literature on this topic. But why should one consider $\tau \neq \frac{1}{2}$?"

[1]Prepared for the forthcoming *Handbook of Quantile Regression* published by Chapman-Hall/CRC.

This report has continued to serve as a valuable reminder that however obvious the quantile regression idea may now appear to be, it was not always so apparent. Meanwhile, we had received some more positive feedback on the paper so we decided to prepare a revision for *Econometrica.* Steve Portnoy had joined the faculty at UIUC in the fall of 1975, and the next spring I decided to get his reaction to what we were doing. He was immediately enthuastic and this encouraged us further. The new manuscript contained an extended introduction in which we tried to motivate the idea of L-statistics for regression along the lines of the work in the late 1940's and early '50's by Mosteller and others. Bickel (1973) constituted a persuasive case for this idea, and we believed our approach had some advantages from an equivariance standpoint. In 1976 Roger moved to Bell Laboratories and was exposed over the next several years to a broad spectrum of current research in robust statistics, when our paper finally appeared in 1978, the introduction undoubtedly reflected some of this exposure. In retrospect the emphasis in our revised introduction on robust estimation of the conditional central tendency of the response was probably somewhat unfortunate since it tended to obscure the more important message concerning heterogeneity of the conditional quantile functions.

About this time Dave Ruppert and Ray Carroll began to look into the question of trimmed least squares estimation using our approach. In Ruppert and Carroll (1980) they showed, rather surprisingly, that trimming based on residuals from preliminary estimators such as least squares had much less satisfactory asymptotic behavior than the regression quantile methods we had proposed. Later Welsh (1987) showed that a modified version of trimming using Winsorized residuals *could* succeed in giving estimators with asymptotics like that of the trimmed mean location estimator. Ruppert and Carroll, using earlier work by Jana Jureckova, also provided a much more straightforward proof of the asymptotic normality of the regression quantiles than our density-based approach. In the one-parameter regression-through-the-origin model Laplace had already derived the asymptotic behavior of the l_1 (weighted median) estimator in the early part of the 19th century, as we learned eventually from Steve Stigler.

We continued to work on these ideas over the next several years, and gradually others became interested as well. Jana Jurečková was enthusiastic early on, and wrote several papers extending the results of Ruppert and Carroll on trimmed least squares, emphasizing their advantage in overcoming the lack of scale invariance of the Huber M-estimator, and even proposing higher breakdown versions to avoid difficulties with influential design points. Gutenbrunner and Jurečková (1992) provided a crucial link between quantile regression ideas and rank tests as exposited by Hájek and Šidák through the formal duality of the linear programming approach. Steve Portnoy also maintained a strong interest in these ideas, and when Roger returned to UIUC from Bell Labs in 1983, they began a close collaboration. Portnoy (1984) established tightness of the quantile regression process on $[\epsilon, 1-\epsilon]$, and this led to further work on more general L-statistics and adaptive estimation.

At Jana's suggestion, we were invited in 1983 to an Oberwolfach meeting on quantile processes and extreme value theory. We were, to put it mildly, not notably successful in conveying our enthusiasm about the potential value of regression quantiles to the distinguished participants of this meeting. Two indelible memories of this meeting remain: the Schumann Romance for concertina and piano played for the evening musicale by Henry Daniels and Richard Smith, and the comment by the conference organizer Willem van Zwet to us in the back of the lecture hall near the end of the sessions: "Erich Lehmann once told me that any good idea takes at least ten years to percolate to the surface of the field." Now, more than 40 years after the first glimmer of the idea, it is nice to see that it is still percolating.

Bibliography

D. Aigner and S.F. Chu. On estimating the industry production function. *American Economic Review*, 58:826–839, 1968.

P. J. Bickel. On some analogues to linear combinations of order statistics in the linear model. *The Annals of Statistics*, 1:597–616, 1973.

F.Y. Edgeworth. A mathematical theory of banking. *J. Royal Statistical Society.*, 51:113–127, 1888.

C. Gutenbrunner and J. Jurečková. Regression quantile and regression rank score process in the linear model and derived statistics. *Ann. Statist.*, 20:305–330, 1992.

S. Portnoy. Tightness of the sequence of empiric cdf processes defined from regression fractiles. In J. Franke, W. Hardle, and D. Martin, editors, *Robust and Nonlinear Time Series Analysis*. Springer-Verlag: New York, 1984.

D. Ruppert and R.J. Carroll. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, 75:828–838, 1980.

A. H. Welsh. The trimmed mean in the linear model. *The Annals of Statistics*, 15:20–36, 1987.