A CONVERSATION WITH ROGER KOENKER

XUMING HE

Author Manuscript

Summary: Roger William Koenker was born February 21, 1947. He graduated from Grinnell College in 1969, and obtained his PhD in economics from University of Michigan in 1974 under the direction of Saul Hymans. He was Assistant Professor of Economics at the University of Illinois at Urbana-Champaign (UIUC) from 1974 to 1976, and a member of Technical Staff at Bell Telephone Laboratories from 1976 to 1983, and returned to UIUC as Professor in 1983. He is currently William B. McKinley Professor of Economics and Professor of Statistics at UIUC. He is best known for his seminal work on quantile regression, which has emerged as a powerful regression analysis tool across many disciplines. He is Fellow of the American Statistical Association, Fellow of the Institute of Mathematical Statistics, Fellow of Econometric Society, and recipient of the 2010 Emanuel and Carol Parzen Prize for Statistical Innovation. The conversation covers part of Roger Koenker's career as an econometrician and statistician, starting from his college years.

Key Words: Econometrics; Least absolute deviation; Optimization; Quantile regression.

1. Early Days

You grew up in North Dakota, can you tell us a bit about your family background?

My father's parents were German emigrants, and he grew up on a farm in western North Dakota where farming was very near the margin of subsistence. My mother's father was Danish and ran a small bank also in the western part of North Dakota. Before the War my father was briefly a high school principal in a school where my mother was teaching music. After the War my father began teaching economics at the University of North Dakota in Grand Forks. It is a small university town, very pleasant, somewhat cool in the winter, but I liked it very much. When I was 12 my father took a one-year US AID assignment in Baghdad; this was very exotic and I enjoyed it tremendously. With that exception though, all my school years were spent in Grand Forks until I went to college.

What led you to Grinnell College?

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/insr.12183

This article is protected by copyright. All rights reserved.

My father must have very deftly hinted that Grinnell would be a good option. Grinnell is a small liberal arts college in Iowa, my father – who was an academic economist – had a very high opinion of Howard Bowen who happened to be president of Grinnell at the time I was applying. Bowen had had, in the early 1950's, a brief but very brilliant tenure as Dean of the College of Commerce at the University of Illinois. As Dean he hired two future Nobel Prize winners in economics, Franco Modigliani and Leo Hurwicz, two very influential senior women, Margaret Reid and Dorothy Brady, and a handfull of very promising junior faculty. I think that my father hoped that Bowen would have had a good influence on economics at Grinnell too. Unfortunately, Bowen's remarkable achievement at Illinois was short-lived; Bowen was removed by the Trustees in a one of the most notorious episodes of academic McCarthyism and all the faculty he hired resigned as well. So it is really quite ironic that I've spent most of my career at the University that failed so spectacularly to defend Bowen and his faculty.

Did your parents have a clear influence on how you pursue an academic career?

Yes, my father definitely encouraged me to consider an academic career and was very pleased when I decided to do economics. Although he didn't quite approve of my steady drift toward statistics. He had grown up on a subsistence farm in western North Dakota and had a variety of very difficult jobs during the great depression, so academic life came as quite a relief for him when he got to that point.

Why did you choose econometrics for your graduate study?

Like many other economics graduate students, I began with the idea that I would do macro economics, but I quickly realized that the theoretical and empirical underpinnings of macro were weak and I became increasingly interested in econometrics and empirical micro economics. I was initially attracted by the idea that most significant policy questions in economics required some form of empirical analysis, they couldn't be resolved by purely theoretical reasoning.

Can you tell us about one or two most influential teachers in your life?

My high school mathematics experience was rather uninspiring, but in my last year of high school we had quite a good course which definitely increased my interest and appreciation. As an undergraduate I was fortunate that in my last year at Grinnell Lynn Muchmore came from Wisconsin and taught an elementary econometrics course. There I estimated some very primitive Phillips's curve models and found the experience quite intriguing. In graduate school Lester Taylor was responsible for introducing Gib Bassett and me to ℓ_1 regression, which turned out to be quite important. And I would say that Bruce Hill's course in decision theory in the statistics department at Michigan was another major influence.

Author Manuscript

This article is protected by copyright. All rights reserved.

Do you recall some of the books you read in your youth, and whether they had an impact on your life?

I have always been a very assiduous reader mainly of literary fiction. The first book that I paid close attention to was probably Henrik van Loon's *Story of Mankind* which I read in Baghdad when I was 12. Later, Bertrand Russell was a favorite for a time. As a graduate student Edmond Malinvaud's *Statistical Methods of Econometrics* was a constant companion, and later Terry Rockafellar's *Convex Analysis* was always within easy reach. Still is, if the truth be told.

You published probably your first paper in Journal of Regional Science in 1972. Can you tell us about it?

It was a toy model of the Ann Arbor housing market, a rare city where the classical assumption of a monocentric spatial structure made some sense. The model had a simple differential equation that determined housing prices according to distance from the center and an estimated elasticity of substitution between land and capital for housing production.

Was quantile regression the topic of your PhD dissertation? If not, can you tell us how you got interested in quantile regression?

No, my dissertation was a rather mundane exercise in estimating systems of input demand equations from longitudinal data on trucking firms. But my friend and fellow graduate student Gib Bassett was writing about ℓ_1 regression at the suggestion of Lester Taylor. Not much was known about minimizing absolute errors in the early 1970s, but Taylor had done some forecasting comparisons that suggested that LAD methods, as they were called then, performed quite well. Gib and I had some background in linear programming from an earlier course given by Sid Winter, and Gib's thesis constituted a very thorough study of how to characterize ℓ_1 solutions based on Kuhn-Tucker optimality conditions. Gib and I completed our PhDs the same year, 1974. He took a position at the University of Illinois at Chicago, and I accepted a position at the University of Illinois at Urbana-Champaign, but immediately went on leave in the Fall to join my wife who had a fellowship to study in Moscow. When I returned to UIUC in January Gib and I resumed our discussions about what we were already calling "median regression." Fortunately, in this pre-email era our universities had a very cheap telephone connection that proved essential to our continued collaboration. I had an interest in estimating "production frontiers" – essentially the extremal quantile regression problem of estimating the maximal output producible with a given vector of inputs. It seemed dangerous though to focus entirely on the most extreme observations, so there was interest in estimating production models "near the frontier." I recall phoning Gib and asking: for median regression we know that roughly half the residuals must be positive and half negative (when there is an intercept in the model), what if we asymmetrically weighted positive and negative

Author Manuscript

This article is protected by copyright. All rights reserved.

FIGURE 1. At Moscow State University 1973

residuals? Couldn't we control the proportion of positive residuals this way? He instantly responded yes, and we became immediately obsessed with trying to understand better what these "regression quantiles" could do.

Koenker and Bassett 1978 is a seminal paper. What brought the two of you together? What kind of impact did Gib have on your research career?

Gib was absolutely essential, without his initial stimulus we never would have ventured down the quantile regression road, and his enthusiasm for the project was a essential element in maintaining my focus in the early days when other sources of encouragement were quite sparse. We have written quite a few papers together, and continue to do so, and it is always a pleasure to have the chance to collaborate with him.

Did anyone else attempt to formulate and analyze the conditional quantile problem prior to your work?

There is quite a long history, some of which we were aware of at the outset and some aspects we only learned about much later. Of course fitting linear models by minimizing sums of absolute residuals is quite an old idea. Before Gauss and

Author Manuscript

This article is protected by copyright. All rights reserved.

Legendre began writing about least squares, Boscovich and Laplace had advocated what Laplace called *la methode de situation*, estimating bivariate linear regression models by minimizing absolute errors while constraining the mean residual to be zero. F.Y. Edgeworth was apparently the first one to suggest removing the mean residual constraint, and proposed a very clever geometric strategy for computation in the bivariate case that anticipated the development of the simplex algorithm. Once simplex became familiar in the 1950s, there was rapid improvement in computational methods for median regression, but curiously there seemed to be little interest in its statistical behavior. Kernel methods prompted some work on nonparametric estimation of conditional quantiles, notably in the work of P.K. Bhattacharya and Charles Stone. Bob Hogg proposed a graphical method of estimating linear conditional quantile functions in a 1974 JASA paper that was closely connected to Wald's 1940 errors in variables estimator. And Peter Bickel had written a 1973 *Annals* paper called "On Some Analogues to Linear Combinations of Order Statistics for the Linear Model" – this was quite an inspiration for Gib and me since it expressed very precisely what we had hoped to accomplish with "regression quantiles." Fortunately, our naive conception had some important invariance advantages over the Bickel proposal that made it considerably easier to analyze.

2. Career Path

You started your academic career at the University of Illinois in 1974. What did Illinois have to attract you then?

Illinois made a very early offer in January, 1973 and I was feeling quite risk averse about the job market and decided that I should accept. This was a period of considerable growth in economics at Illinois and I was very fortunate to have excellent new colleagues in econometrics: Takamitsu Sawa who had come from Stanford and Dale Poirier who was coming from Wisconsin, in addition to George Judge and Tom Yancy who had been at Illinois for some years. In my second year Steve Portnoy arrived in the Statistics Department from Harvard. Steve was very enthusiastic about the quantile regression idea from our first conversations, and has continued to be an enormously positive influence on my research. Illinois was also an attractive place for my wife, who was completing her PhD in Russian history.

What led to your move to Bell Labs in 1976? Can you tell us a little bit about Bell Labs in the 70's and how it evolved over the years?

My wife finished her PhD early in 1976 and was offered a position at Temple University in Philadelphia for the Fall of 1976, so I began to look for a job on the east coast. Bell Labs had started a small research group in economics a couple years earlier and I'd been very fortunate to have met John Panzer who was one of their first hires. I had written a couple of papers on peakload pricing for electricity, and this was a topic

Author Manuscript

This article is protected by copyright. All rights reserved.

that was also relevant to telephone pricing since congestion and capacity constraints are common to both settings. Bell Labs in the period I was there 1976-83 was a fantastic research environment, especially for someone with my interests. Economics was physically and organizationally situated within the Math Center, and my immediate office neighbors, first at Holmdel and then at Murray Hill, were members of the Statistics Research Department. This meant there was a constant flow of interesting seminar visitors and other opportunities. There was some expectation that we would contribute something to the well-being of the parent company AT&T, but for the most part we were allowed to do our own thing. I continued to work on quantile regression ideas as well peakload pricing. When AT&T settled its major anti-trust case in 1983, there was a split of the company, and the regional telephone companies, the so-called "Baby Bells" were separated from the parent company AT&T. This separation induced an abrupt change in the research environment; most of the economists were told that they would be transferred to a new research entity administered by the regional companies. At that point many of us felt that there would be a much more directed consulting environment in the new setting and we chose to return to academia in 1983.

Did the job at Bell Labs change your research directions?

There was a very strong robustness focus among the statisticians at Bell Labs in the period I was there. Of course, robustness was quite a prominent topic throughout statistics at that time, but the influence of John Tukey, who visited regularly from Princeton, made this especially important at Bell Labs. The effect of this on my own work is most apparent in the revised introduction that I wrote for our 1978 *Econometrica* paper, which tried to hook "regression quantiles" to the robustness bandwagon. Thus, much of the emphasis was on analogues of L-statistics for the linear model with the unfortunate result that it may have appeared that we were mainly interested in yet another class of estimators for the central tendency of the data, while neglecting the more important heterogeneity motivation of the methods.

You took a Professor position at University of Illinois in 1983. What attracted you back to Illinois?

When the Bell Labs Economics Department began to break up there was a sense of panic, but I had friends that I'd kept in contact with at Illinois, and they encouraged me to return. A brilliant aspect of my return was that I'd managed to avoid all the unpleasantness of tenure and promotion reviews. When the History Department agreed to make an offer to my wife, this sealed the deal.

You have stayed in Illinois for many years. Did you ever consider moving?

Occasionally, there have been some whispers about a move, but this never progressed to the stage that it was very serious. I have been quite happy at Illinois, it has been a

Author Manuscript

This article is protected by copyright. All rights reserved.

good research environment for me. I've been very fortunate to have close connections with folks in the statistics department like you and Steve, and I've also been very lucky to have a steady flow of good graduate students to work with.

You published in both econometrics journals and statistics journals. How did you choose between the two?

Early on I felt it was important to publish in the econometrics literature, but later I often found a more receptive audience in the statistics journals. This is a very positive aspect of the statistics discipline: people seems to be generally quite open to new ideas coming from outside the discipline whether it is from biology, or machine learning, or the social sciences.

How did you view the relationship between econometrics and statistics? Would you feel comfortable to be in a statistics department?

Well, I suppose it would depend on whether they felt comfortable with me. (Laughs) Early on I was always quite intimidated by the prospect of giving talks to statisticians. One of my first talks about quantile regression was in the Statistics Department at UIUC and the audience included Joe Doob and Jack Wolfowitz in addition to Steve Portnoy and Walter Philipp, but everyone was quite friendly, something that can't always be said about seminars in economics departments.

3. Work on quantile regression

Quantile regression has emerged as an important alternative to the least squares regression. In your early work you motivated quantile regression as a robust alternative to the least squares regression. How would you characterize the robustness of quantile regression estimators? Compared to many other robust regression methods, is quantile regression a better alternative?

It was initially quite difficult to find a convincing motivation. When we first submitted the paper to *Econometrica* in 1975 the reaction was roughly: we understand why minimizing the sum of absolute residuals is interesting, but the paper fails to make a convincing case that the asymmetric solutions are interesting. Even though the editor suggested we might consider revising the paper to strengthen the motivation we were sufficiently discouraged by the reports that we decided to see whether the reception would be better at the *Annals of Statistics*. To our dismay, the reaction was even more dismissive: "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, $\tau \neq \frac{1}{2}$?"

Author Manuscript

This article is protected by copyright. All rights reserved.

This led to rather drastic reworking of the introductory motivation for the paper. Robustness was already a very well established research agenda, at least in statistics, and we were strongly influenced by this environment, me particularly since I was already at Bell Labs. There were close connections that could be established with L-estimators, or linear combinations of order statistics, including proposals for analogues of the trimmed mean for the linear regression model, so it was relatively easy to reorient the paper in this direction. Admittedly, from a longer term perspective it wasn't ideal since it made our objective appear to be just a few more estimators of conditional central tendency in what was already a very crowded field. In the short run, though, this had to be counted as a success since it satisfied the referees and the paper eventually appeared in *Econometrica* in 1978.

From a formal robustness perspective like that of Hampel, quantile regression estimators aren't robust at all. They have bounded influence in the response direction, but have unbounded influence in any of the design directions. Like other regression M-estimators one sufficiently outlying design point can cause breakdown of the procedure. There have been several proposals to improve the robustness of quantile regression methods with respect to influential x's; the proposal I like best is probably that of Rousseeuw and Hubert based on multivariate depth ideas.

It is probably not a good idea to think of quantile regression just as a robust estimator. Can you explain the main difference between a robust regression estimator such as Huber's M-estimator and a quantile regression estimator?

That's right, from my present vantage point I would stress simply that quantile regression estimators are intended to estimate conditional quantile functions, and efforts to combine them in some way to produce an estimator of central tendency may have the unfortunate consequence of obscuring what is most interesting about them, their heterogeneity.

You and Gib published a paper in JASA on the asymptotic theory of the LAD regression, also in 78. Can you talk about the difference between the JASA paper and Econometrica paper, both published in 78?

We felt that it would be useful to provide a more detailed argument for the asymptotic behavior of the LAD (median) estimator in the JASA paper, allowing us to be a bit more brief about some details in the *Econometrica* paper, asymptotics there were focused more on the joint distribution of several quantile regression estimators. It probably should be admitted that the technology of the proofs in those early papers was somewhat primitive. In effect, we were trying to follow the prescription of Cramér in his famous 1946 text: write down the finite sample joint density of several quantiles and then analyse its limiting behavior. The downside of this is that it involves a rather delicate local limit argument that we didn't handle very well, in addition to the awkward nature of considering all n choose p distinct "basic" solutions. Fortunately,

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

FIGURE 2. Roger Koenker and Gib Bassett in Neuchâtel in 1987

it didn't take very long before David Ruppert and Ray Carroll and others provided more straightforward empirical process arguments.

Your 78 paper in Econometrica is mostly about quantile regression at a specific quantile level. When did you start to think about the quantile regression process, and what is the significance of thinking that way?

We really wanted to think about joint distributions of the regression quantiles from the very beginning, but its true that the 1978 paper was stuck in the mindset of iid error linear models. This was convenient from the perspective of L-statistics analogues, but highly unrealistic from a broader data analytic perspective. By the time of our 1982 *Econometrica* paper on testing for heteroscedasticity Gib and I were much more focused looking for differences in the quantile regression estimates.

You also spent quite a bit of time working on better algorithms for quantile regression. Since quantile regression solves a linear program, why cannot we simply use a standard linear program package to do the computation? Has the computational technology changed much over the years?

Yes, I've always been quite obsessed by computational developments for quantile regression. Having learned the S language at Bell Labs, I've tried since then to maintain software for quantile regression, first in S, and now in R, that implements current developments on the research frontier. Partly, I find that this is a good discipline for my own research and certainly facilitates reproducibility, and partly it is an attempt to

This article is protected by copyright. All rights reserved.

encourage others to explore these methods. Computational methods for linear programming and therefore for quantile regression have changed quite dramatically over my lifetime. This can be seen in the options available in my R package, `quantreg`. Initially, there was just the modified simplex implementation based on the algorithm of Barrodale and Roberts. When interior point methods were developed in the 1980's Steve Portnoy and I wrote a paper that appeared in *Statistical Science* that described how together with some preprocessing these innovations made quantile regression methods computationally comparable in speed to least squares. The next step forward came with the recognition that developments in sparse linear algebra, especially for Cholesky decomposition, made large nonparametric additive models with thousands of parameters quite efficiently estimable. In large dense problems eventually interior point methods become impractical, and I've recently been exploring proximal operator methods that provide promising gradient descent type methods for these new challenges.

If someone chooses to use quantile regression in data analysis, what are the main challenges s/he has to overcome? Are there difficulties in computing or inference or interpretation of results?

I think that interpretation is always the most difficult aspect in any statistical analysis. Computation is now quite easy in most settings, and inference while it still poses numerous challenges has achieved what might be called a callow maturity. But conditional quantile functions are rather complicated beasts. As in ordinary regression settings causality is often a controversial aspect. But researchers are sometimes also a bit careless about explicitly recognizing the nature of the conditioning underlying the quantile regression paradigm. In economics this has led to a small literature on "unconditional quantile regression," this is essentially an effort to estimate a family of binary response models that taken together can be viewed as an estimate of the conditional distribution function of $Y|X$. Thus, instead of asking, what is the τth quantile of Y when $X = x$, we ask instead, what is the probability that Y exceeds some y when $X = x$.

You have collaborated a lot with Steve Portnoy in Illinois. What brought you two together in the first place? What is the most successful collaborative project you have had?

Steve came to UIUC in 1975, the year before I left for Bell Labs. I had heard that he was interested in robustness, so I went to see him shortly after I heard this, and he was very encouraging about the early ideas about quantile regression. We kept in touch while I was at Bell Labs; I was very intrigued by his seminal work on "large-p asymptotics." From an econometric viewpoint his work seemed to provide a much more realistic framework for analysing estimation and inference methods in model sequences than the conventional fixed-p setup. When I returned to Illinois in 1983,

Author Manuscript

This article is protected by copyright. All rights reserved.

we continued to talk about various projects. Our L-estimator papers were one of the first outgrowths of this, and our work with you and Jana Jurečková on tail behavior of regression estimators and the connection to breakdown came only a little later. Our paper with Pin Ng on total variation nonparametric smoothing methods for quantile regression has been a particular favorite of mine, in effect it proposes lasso shrinkage as a smoothing device *avant le lettre*.

You have published some fundamental work on quantile autoregression (e.g., your 2006 paper with Zhijie Xiao is a highly cited paper). What makes quantile modeling in autoregression interesting? How did Zhijie come to the field of quantile regression?

Zhijie came to Illinois in 1997 after finishing his PhD at Yale, we were extremely fortunate to attract him. We began to talk about various topics involving time series analysis and quantile regression. Autoregression was a natural problem, but there were quite a few immediate problems; not the least of which was that in linear autoregression models it is not at all obvious how to ensure monotonicity of the conditional quantile functions. After considerable preliminary exploration we convinced ourselves that these models were potentially useful at least as an initial approximation. We worked out some basic stationarity conditions and proposed some new ideas for inference. At the time there was considerable interest in testing for unit-root behavior in economic time-series, and one thing that we wanted to show was that QAR models offered some potential for "unit-root-like behavior" while still satisfying stationarity

FIGURE 3. Roger Koenker with Steve Portnoy, Jana Jurečková , and Gib Bassett in Neuchâtel in 1987

Author Manuscript

This article is protected by copyright. All rights reserved.

conditions and mean reversion. I still think that this is an interesting aspect of these models that deserves further investigation.

You supported the work of Ying Wei on growth chart and conditional growth chart construction where nonparametric quantile regression proves to be useful.

Yes, I regard our work on growth charts as one of my most successful empirical ventures. I had met a Finnish pediatrician, Anneli Pere, on a brief visit to Oxford. We had talked about a collaboration analysing a reference growth data on Finnish children using quantile regression methods. However, at the time I was unsure how to cope with the longitudinal nature of the data, twenty or so measurements on each child. So when you and Ying Wei expressed an interest in pursuing this, I was quite delighted. I always say that growth curves are the Ur-quantile regression experience since immediately after birth one is measured and slotted into some existing growth chart. Finding better ways to produce these charts to make them more useful to diagnosticians seems to be an important task. I hope that our *Statistics in Medicine* paper helped to some degree to show the way, both in terms of quantile regression methodology and by showing how longitudinal aspects could be incorporated into the analysis.

Yes, indeed. There has been some nice follow-up work in epidemiology. The wide application of quantile regression can be seen in Google Scholar. Just in the year 2014 alone, there are over 5000 entries in Google Scholar when a search on "quantile regression" is used, and most entries are applications of quantile regression. Did you anticipate such wide use of quantile regression some years ago? What made quantile regression so widely used these days?

No, [Laughs], of course I hoped that someone would find it useful eventually, but certainly its first ten years or so didn't bode very well for this. It is hard to account for the rapid growth in applications. Part of it, I suppose, is simply that researchers have more data and are looking for new ways to dig more deeply into their data. And in my experience one rarely finds that the classical iid error linear model assumptions are very plausible, once you start looking at a family of quantile regression fits. In economics there was a structural break at the moment that Gary Chamberlain gave his talk at the 1990 World Congress of the Econometric Society in Barcelona on union effects on wages. In ecology it was always common to make models for the largest sustainable population size as a function of environmental factors, so the idea caught on there somewhat earlier.

Can you name one most interesting application of quantile regression in science or economics?

It is hard to pick out one or two examples from the vast array of applications, especially when I'm quite unfamiliar with the basic science for most of them. Based

Author Manuscript

This article is protected by copyright. All rights reserved.

entirely on titles though it is hard to imagine improving on: "Cannibalism by female *Calanus finmarchicus* on naupliar stages." In my own work, the JASA paper with Olga Geling on mortality of medflies was instrumental in raising my awareness of the potential of quantile regression methods in survival analysis.

Although quantile regression research remains active in econometrics, it has gained popularity in statistics too. What do you see as the most important advance over the past 15 years in the area of quantile regression?

I suppose the easy answer would be, it's too soon to tell. But I think that Steve Portnoy's work on censored survival data, and the subsequent work of Peng and Huang on related methods has been very significant. An early advance that was hugely important was the link to rank statistics provided by the work of Jana Jurečková and Cornelius Gutenbrunner. By connecting the quantile regression dual problem with the classical Hájek rankscores they provided a beautiful new class of inference methods. Within econometrics, the work of Chesher and Chernozhukov and Hanson on causal models has been very influential. There has also been very exciting work on quantile regression for multi-dimensional response and I look forward to seeing how that develops.

We are now entering the era of big data. Is quantile regression a natural fit for the analysis of big data?

Well, at least the computation problem is convex, and solutions are therefore easily computable. There has been quite a lot of work in both genomics and economics on quite large problems. Only time will tell, I suppose, but as data sources become richer, I expect that interest in new sources of heterogeneity is likely to increase and I hope that quantile regression can play a constructive role in assessing this.

Some recent work by Yang, Meng and Mahoney demonstrated that algorithms implemented in MapReduce-like environments can solve quantile regression problems for terabyte-sized data. Computer scientists are now coming on board. What is your view on the future of computation when it comes to quantile regression?

Yes, this is very interesting, and I've been a bit slow to appreciate these developments. But in the last few months I've begun to explore some of these new ideas and I'm finding it very intriguing. The paper you mention is especially interesting since it was one of the first to offer methods that would enable researchers to compute quantile regression estimates for terabyte scale problems. On "thin" regression problems like their main empirical test problem that is a rather typical econometric wage equation with 5,000,000 observations and 11 covariates, their methods require only about 7 seconds, while my standard interior point algorithm takes about 45 seconds, and using the preprocessing approach suggested in the paper with Portnoy gets this down to only 9 seconds. The downside of the new methods is accuracy: They are quick but

Author Manuscript

This article is protected by copyright. All rights reserved.

somewhat dirty in the sense that at the reported speeds they are only accurate to about two decimal digits. I am still trying to come to grips with this tradeoff. I know that there are a number of prominent people in the statistical and computer science communities who are also very interested in this tradeoff; it seems quite important and should at some point be reconciled with our conventional view of asymptotic behavior. Of course when the parametric dimension of problems becomes much larger, especially when design matrices are quite dense the advantages of the new methods are much greater. This can be said of the recent work on ADMM (alternating direction method of multipliers) methods which seem very promising, but again make serious sacrifices in accuracy to achieve faster computational speed.

You have been maintaining the R package `quantreg`*, and it has been extremely valuable to researchers and data analysts. What is your plan for the package in the next ten years and beyond?*

For me R packages are essential to my whole research strategy. I greatly admire those who can prove beautiful theorems in abstract settings without any means of visible computational support, but I've always needed to see some evidence of practical performance before venturing into the thicket of theory. So R provides a good environment for gradually building methodology, and it is also good discipline for maintaining a archive of reproducible research results. My immediate plans for the `quantreg` package are quite modest. I've been experimenting a bit with the new "first order" algorithms and hope to add some functionality based on these ideas. As you are well aware there are very interesting new Bayesian ideas that I would like to be able to incorporate, but it may be preferable to let others push forward with these. I hope that I can find someone to take over this effort soon. This is always a question with open source software projects: is there a sustainable path into the future? I hope so, but a little "creative destruction" – to use a Schumpeterian phrase from economics, would probably be helpful too.

Like any good thing in life, quantile regression could be misused, especially in terms of interpretation and inference. Are you concerned about the possible misuse of quantile regression analysis?

Sure, it is scary sometimes to read some of the email inquiries that I receive, but this is certainly an inevitable consequence of any success of new methods. Fortunately, I don't have to police this sort of thing and there are now plenty of knowledgeable people who can help evaluate new applications.

One question is about the ad hoc nature in the choice of the quantile level. If one analyses the .75 quantile, but another looks at 0.8 quantile, would they get very different results? How would you advise users in this regard?

Author Manuscript

This article is protected by copyright. All rights reserved.

This is certainly a valid concern. Of course if nearby quantiles produce dramatically different results this is a clear indication that they are very imprecisely estimated. Sometimes it can be quite valuable, especially in the tails, to try to borrow strength from adjacent quantiles by some sort of smoothing tactic.

I am always a little uneasy about the notion of multivariate quantiles, but it is appealing to extend the notion of quantile to multivariate data and functional data. What is your take on it?

Yes, this has been quite an interesting research dynamic. There are quite a variety of proposals at this stage. Probal Chaudhuri's transformation-retransformation work provided a nice link to spatial median ideas. Marc Hallin and his colleagues as well as Ivan Mizera and his student Linglong Kong have pursued connections to Tukey half-space depth. Ying Wei's recursive conditioning approach seems very attractive to me, and is closely related to Andrew Chesher's work. There is also very appealing recent work by Victor Chernozhukov and his colleagues using Monge-Kantorovich mass transport ideas. I also like very much the recursive rank transformation approach that you and John Marden have developed and hope that you will pursue that. But given the inherent difficulties, it seems inevitable that there will continue to be a multitude of approaches appropriate for various applications.

For a graduate student in statistics who is interested in pursuing further research in quantile regression, what promising directions would you point him/her to?

I suppose time-series and longitudinal data are still interesting sources of problems. Functional data offers many challenges, Kengo Kato has provided some initial impetus and it would be great to see further developments there since the dominant Gaussian paradigm seems too restrictive in many circumstances. Survival analysis has received quite a lot of attention, but it is such an important topic that I would expect to continue to see many important new developments.

4. Other Aspects

It seems to me that optimization has been part of your passion. In recent years you also wrote about density estimation and empirical Bayes methods. Can you share with us how you got attracted into those areas, and whether and how optimization plays a role there?

I sometimes joke that in economics, if one can't formulate a model of optimizing behavior to describe a phenomenon then it doesn't really exist. My own journey to empirical Bayes methods is a rather long story, but illustrates how random walks can lead one to some unexpected, but very exciting places. I suppose the story begins with a paper that Ivan Mizera and I wrote about two dimensional nonparametric quantile regression with total variation smoothing spline penalties. This led us to

Author Manuscript

This article is protected by copyright. All rights reserved.

start thinking about whether we could do something similar with bivariate density estimation, and we wrote several conference papers about that sort of thing, using total variation penalization as a smoothness penalty for the log density. The problem with these conference papers was that people kept asking: How do you choose the smoothing parameter? And we didn't have any better ideas about this than anyone else, so we got tired of apologetically answering with some vague ideas about AIC/BIC methods. It finally occurred to us that we could avoid such questions entirely if we simply said that we wanted to impose a shape constraint, and since the form of the total variation penalty produced estimated densities whose logarithm was piecewise linear it was easy to see how to impose log-concavity. The shape constraint is sufficient to regularize the density estimation problem, no tuning parameter is necessary. Of course log-concavity is also a very nice property of densities and has a very extensive literature, in survival, quality control and throughout economics. On the other hand, there didn't seem to be anything available on the nonparametric estimation of log-concaves, so we thought we had found a nice quiet little research domain that we could mine for a while. This turned out to be an illusion, and we quickly learned that Lutz Dümbgen, Richard Samworth and others were also deeply engaged in the subject. Fortunately, our robustness ideology altered our trajectory once again. We were somewhat dissatisfied with the fact that log concave densities have to have sub-exponential tails, and we began to explore the possibility that similar convex optimization methods that we had been using for them could be used to get estimates of heavier tailed densities. Since the dual log-concave problem led to minimizing a form of Shannon entropy, it was natural to consider replacing Shannon by one of the family of Renyi entropies in the dual, and we focused on the requirement that $1/\sqrt{f}$ be concave. This class included all the Student t densities down to Cauchy, although in the transition we had to jettison the maximum likelihood criterion and replace it by a Hellinger objective. Our contribution was mainly to describe computational methods for these Renyi estimators, but fortunately Jon Wellner and Qiyang Han have recently provided a much more extensive theoretical underpinning for them.

My empirical Bayes work grew out of a very brief conversion with Larry Brown while on a seminar visit to Wharton to talk about the shape constrained density paper. Larry had recently written a paper with Eitan Greenshtein about a Gaussian compound decision problem in which they had used kernel methods to estimate a mixture density to produce a nonparametric Bayes rule, or Tweedie formula. Larry was dissatisfied with the kernel approach since it failed to impose a monotonicity requirement on the Bayes rule that was implied by the exponential family structure of the original Gaussian problem. He wondered whether shape constrained methods could be used to construct an alternative to the kernel estimate of the mixture density that would enforce this monotonicity. This turned out to be a surprisingly tractable homework problem, but as I looked at related literature it became obvious that I would need to compare performance of the new estimator with the Kiefer-Wolfowitz

This article is protected by copyright. All rights reserved.

Author Manuscript

nonparametric MLE for mixture models that had been proposed by Zhang and Jiang. Initial efforts to do this were rather frustrating due to the extremely lackadaisical computational behavior of the EM algorithm for the Kiefer-Wolfowitz estimation. Eventually it occurred to me that one could replace the EM algorithm with an interior point convex optimization approach that was substantially faster and more accurate. This opened the way for a variety of new applications, and has provided a very exciting new direction for my research. Again Ivan Mizera played a vital role in helping to develop the original algorithmic approach and Jiaying Gu has been an indispensable collaborator on many of the subsequent projects.

You have supervised over twenty PhD students at the University of Illinois, and many of them are now quite successful in their careers. What is the general approach you take in supervising students? Did the students come to you with their own vision and problems or you started them with a research project that you cared about or ?

I've been very fortunate with the PhD students I've worked with, mostly from economics and a few from statistics. Thesis topics seem to arise in quite a variety of ways: a few from term papers written in one of my courses, or a talk by a seminar visitor. Others are motivated by an empirical problem that came from another source, like the growth curve work of Ying Wei, or the sequential survival work by Yannis Bilias that was jointly supervised by Zhiliang Ying and I. I was especially fortunate recently to convince Jiaying Gu to begin working on empirical Bayes methods right at the beginning of my new obsession with them. Our collaboration perfectly illustrates a comment that I like very much by David Cox: An interviewer asked him: "The late Professor Dennis Lindley told me that 'One of the joys of life is teaching a really good graduate.' Would you be in agreement?" And Cox responded: "I would say that one of the joys of life is learning from a good graduate. The first duty of a doctoral student is clearly to educate their supervisor which my own doctoral students have done."

Your first PhD student, José Machado, became the Dean of Business School of Universidade Nova de Lisboa. Can you talk about one or two of your students who have influenced your research agenda over the years?

José was an exceptional talent and we continued to work together on a variety of topics after he finished his PhD which dealt with model selection for general regression M-estimators including the quantile regression case. We wrote a JASA paper on inference for the quantile regression process that was influential in stimulating my interest in the Khmaladze approach to testing. Pin Ng, who also finished around the same time as José, was also a big influence; he was involved in my first paper about total variation penalty methods, and also was deeply involved in the later attempt to exploit developments in sparse linear algebra for large quantile regression problems.

Author Manuscript

This article is protected by copyright. All rights reserved.

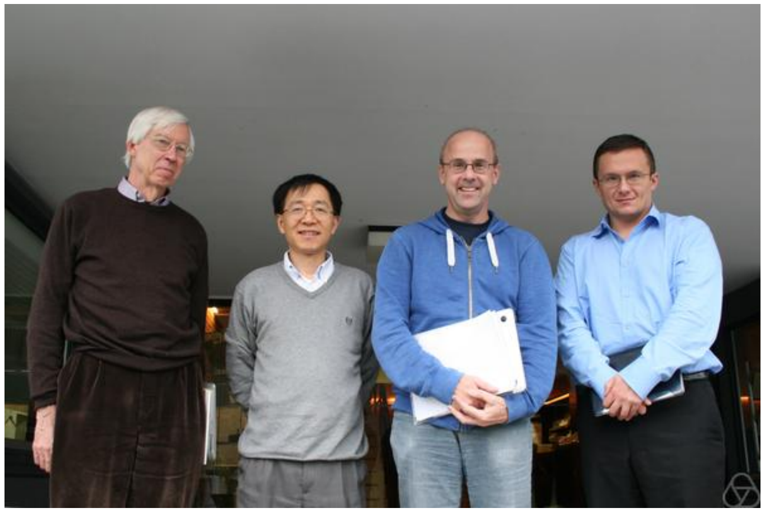

FIGURE 4. Roger Koenker with Xuming He, Holger Dette, and Victor Chernozhukov, co-organizers of the 2012 Oberwolfach workshop on Quantile Regression

In 2010, you received the Emanuel and Carol Parzen Prize for Statistical Innovation. Probably not accidentally, Emanuel has been a long-time advocate for quantile modeling. Can you tell us about the prize, and about Emanuel Parzen's work on quantile modeling?

This was an enormous surprise, but also hugely gratifying. Manny's work on quantile modeling was hugely influential in my own thinking about quantile regression and my transition away from the L-statistic view toward my current, more heterogeneity-centric view, to coin an oxymoron. I should also mention that Manny's much earlier work on reproducing kernel Hilbert spaces, and the wonderful work of Grace Wahba, was another source of inspiration. Certainly the idea of total variation smoothing penalties for quantile regression was directly inspired by the earlier success of RKHS methods for L_2 smoothing.

You have travelled to many parts of the world; England, Russia, China, Brazil, and many more. What is your favorite city/town? Are there any memorable travel stories to share?

Travel is always very illuminating. One of the many virtues of an academic career is the opportunity afforded by conference travel to meet exciting new people and visit new locales. Early in my career there were several conferences on L_1 regression in Neuchâtel where I met many influential statisticians interested in robustness. More recently I've visited Andrew Chesher's CEMMAP center at UCL quite regularly, which is always an exciting econometric environment. Two of my most inspiring travel experiences were my visit to Gabon to see our daughter who was a Peace

Author Manuscript

This article is protected by copyright. All rights reserved.

Corps volunteer, and a later trip to Mali where she had extended her Peace Corps stay for a third year.

You were a co-organizer of a Banff workshop in 2003 and an Oberwolfach workshop on quantile regression in 2012. How do you think about conference centers like Banff and Oberwolfach?

Oberwolfach is really a very special place. I have been very fortunate to participate in several meetings there, the first time in 1984 when the quantile regression project was still in its early adolescence. I was very intimidated to be speaking with very prominent statisticians and probabilists. I still remember fondly Willem van Zwet, who was an organizer of the meeting, being very kind and encouraging despite my obvious naivete.

There used to be regular conferences on L_1 statistical procedures in Neuchatel organized by Yadolah Dodge. Now there are annual international conferences on robust statistics and related fields. You have been to quite a few of them, and you may have noticed that many of the younger scholars in robustness are outside the United States. Do you see any new directions in robust statistics?

Big data seems to have pushed the robustness agenda aside in the last few years. There always seemed to be a tension in the robustness literature between statistical performance and algorithmic scalability, most visibly in the effort to design high breakdown procedures. This tension was really a precursor of the scalability concerns raised by our current obsession with big data. I sometimes get the feeling that some of the important progress that was made on robust methods has been lost in the transition. But there are some interesting echoes of robustness ideas in the big data scene. A prominent example is the recent work of Bin Yu, Peter Bickel, Noreddine El Karoui and Derek Bean on M-estimation in high dimensional regression, which provides some quite surprising results on the asymptotic un-optimality of the MLE. It is important to keep the flame of robustness burning, practical problems are rarely well served by methods that rely on standard Gaussian assumptions.

Statistics and econometrics are two closely related disciplines. Sometimes it is not easy to distinguish them, especially given that quantile regression has been a focus of both disciplines. Can you tell us the similarities and the differences between these two disciplines?

There is now much more interplay between econometrics and statistics than earlier. There are many examples of very slow diffusion of new ideas in the early history of the two disciplines. Econometrics has always provided challenging problems: errors in variables, causal modeling in equilibrium settings, unobserved heterogeneity in mixture models. Many econometricians like to stress the causal modeling aspect of their subject, and statisticians are sometimes uncomfortable with the assumptions

Author Manuscript

This article is protected by copyright. All rights reserved.

FIGURE 5. Roger and Diane Koenker in Mali in 2005

that underlie econometric methods like instrumental variables, but gradually with more dialogue between the two fields I think that there is a better understanding of the objectives and methods on both sides.

You have been Associate Editor for both statistics and econometrics journals. Any interesting stories you can share from your AE experience? If a statistician wishes to submit a paper to an econometrics journal, what does s/he need to keep in mind?

Serving as AE for *Econometrica* and *JASA* were very rewarding experiences. There is still some disciplinary chauvinism, but I think that this has gradually improved over my time in the profession. One now sees quite a lot of work by statisticians appearing in the econometrics journals, and vice-versa, but defensive citation behavior is always an important consideration in such ventures. My favorite AE story involves a paper that I submitted to JASA Applications & Case Studies several years ago. A few weeks later, I received a request to referee the paper, a request that I unfortunately I had to politely decline. This incident illustrates that if you stay in the academic research game long enough, all kinds of strange refereeing situations will arise.

I am sure that research, teaching and professional service always keep you busy. What are the hobbies that you consider almost as important as your academic life?

I regret that I'm not able to tell you that I'm composing string quartets or painting watercolors of rare birds on the side. Unfortunately, my hobbies are much more mundane. I'm a strong believer in the principle: Much depends on dinner. So cooking is very important. Music is also important but unfortunately only as a consumer, not

Author Manuscript

This article is protected by copyright. All rights reserved.

a producer, as we say in economics. Travel is also a long term obsession and I've been fortunate to visit quite a variety of places, beginning with a few months in Baghdad where my father was working for US AID in the late 50s. Most recently, I'm just back from two weeks in Georgia, including a visit to a remote town near Mt. Kazbegi, one of the highest mountains in Europe.

Diane is a prominent historian and I am sure she is also very busy as Chair of the History department at UIUC. Dual-career couples are becoming more common these days, but not without challenges. Do you have any wisdom to pass on to the younger generation who are facing challenges in dual-career development?

We have been very fortunate to have had two tenured positions at UIUC. Universities are increasingly aware that this is a serious concern for many couples, but it is still quite difficult particularly for couples in very disparate fields. It was clear as our children approached school age that our commuting arrangements while living in Princeton and working in Murray Hill and Philadelphia were going to be increasingly untenable. So moving to Urbana simplified many aspects of life. Of course simplicity is sometimes over-rated and we needed to do some things to complexify life too; it is important to find a balance. In our case our annual subscription to the Lyric Opera in Chicago helped considerably. Younger people need to keep in mind the immortal words of the Rolling Stones: "You can't always get what you want, but if you try sometimes, well, you just might find you get what you need."

Thank you, Roger, for sharing your thoughts and experiences.

Department of Statistics, University of Michigan,, Ann Arbor, MI 48109, USA. (email: xmhe@umich.edu)

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

Mali05.jpg

This article is protected by copyright. All rights reserved.

Author Manuscript

Moscow.jpg

This article is protected by copyright. All rights reserved.

Author Manuscript

Neuchatel87.jpg

This article is protected by copyright. All rights reserved.

Author Manuscript

Neuchatel92.jpg

This article is protected by copyright. All rights reserved.

Author Manuscript

Oberwolfach.jpg

This article is protected by copyright. All rights reserved.