Chapter 9: Quantile Regression

Overview:

In the previous chapters, we defined "best" by minimizing the sum of the squared residuals. An even function was required to ensure that the values in the target function were positive. In this chapter, we use the absolute value function.

While this may sound like a trivial move, it is not. Since the magnitude of the slope of the absolute value is constant, optimization is difficult. Furthermore, since it is undefined at its minimum, we can only use an iterative technique to estimate (approximate) our parameters of interest.

However, in building this structure, we are able to go beyond just estimating the median, we can estimate any quantile. This allows us to model the tails of the data-generating process.

Forsberg, Ole J. (10 DEC 2024). "Quantile Regression." In *Linear Models and Řurità Kràlovstvì*. Version 0.704442$\eta(\alpha)$.

Chapter Contents

9.1 Parameter Estimation 263
9.2 Quantile Regression 269
9.3 Conclusion 275
9.4 End-of-Chapter Materials 276

❧❧❧

least squares

In the previous sections we examined three types of least squares regressions — ordinary, weighted, and general. These three estimation methods have one thing in common: The estimates were obtained by minimizing the sum of squared residuals (properly weighted). We used the squaring function for two reasons. First, it is everywhere differentiable, especially at its minimum. Second, squaring the residuals ensures that you are adding non-negative values. All even functions attain the second goal. The class of functions that meet the first requirement is more restrictive.

The higher the even power, the more outliers affect the estimates; that is, the outliers will tend to have an increased effect on the estimator when the power is larger. One option to reduce the effect of these outliers is to use a different even function. The absolute value function has been used quite successfully in the past.

Unfortunately, the absolute value function is not everywhere differentiable. Even worse: it is not differentiable at its minimum — the point of interest. This means we cannot obtain a simple set of equations for our estimators. We can still, however, obtain estimators to an arbitrary degree of precision by using a set of equations that get us closer and closer to the true value of the estimate.

9.1: Parameter Estimation

Let us first think about how we *could* do this by hand:

In least squares, we just did calculus to get equations for the estimators. Here, since such solutions do not exist and since we need to use an iterative technique, I think working through a toy example may help understanding. And so, let us start with the data in the left two columns of Table 9.1.

L_1 Norm

Remember that we want to minimize the sum of the absolute values of the residuals.[1] Thus, the first step is to obtain residuals. This means we need to somehow obtain our first estimated regression line. Any will work as a starting point. So, let's say our first estimate of the line-of-best-fit is $\ell_1 : y = 3$, which is just the horizontal line at the median.

The next step is to calculate the residuals. This is the e_1 column in Table 9.1. The the target function is

$$Q_1 = \sum_{i=1}^{n} \mid e_i \mid \tag{9.1}$$

$$= \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{9.2}$$

$$= \sum_{i=1}^{n} \left| y_i - 3 \right| \tag{9.3}$$

$$= \left| y_1 - 3 \right| + \left| y_2 - 3 \right| + \left| y_3 - 3 \right| + \left| y_4 - 3 \right| \tag{9.4}$$

$$= \mid 1 - 3 \mid + \mid 3 - 3 \mid + \mid 3 - 3 \mid + \mid 9 - 3 \mid \tag{9.5}$$

$$= 2 + 0 + 0 + 6 \tag{9.6}$$

Thus, for line ℓ_1, the value of the target function is 8.

[1] In other words, we want to minimize the L_1 distance between the n-dimensional data vector and the p-dimensional parameter space. Recall Figure 3.2 where we illustrated this with least squares.

x	y	$\hat{y}_1$	e_1	$\hat{y}_2$	e_2	$\hat{y}_3$	e_3
0	1	3	-2.0	1.5	-0.5	0.0	1.0
1	3	3	0.0	2.5	0.5	2.0	1.0
2	3	3	0.0	3.5	-0.5	4.0	-1.0
3	9	3	6.0	4.5	3.5	6.0	3.0

Table 9.1: *Raw data and a few columns of the median regression estimation process. This is more heuristic than actual. The actual fitting method depends on the program used.*

The next step is to change the regression line. How? Well, that is the important question. Different methods may ultimately lead to slightly different answers. As this section only seeks to illustrate a method — and not even a good method — let's use logic to see what would be next. Note that the lower values have estimates that are too low, and the higher values have estimates that are too high. So, it makes sense to increase the slope. So, let us increase the slope to 1. *If* we force the line to pass through the dimension-wise median $(\widetilde{x},\ \widetilde{y}) = (1.5,\ 3.0)$, the linear equation will be $\ell_2 : y = 1.5 + 1x$. This produces the estimates and residuals in the next two columns of Table 9.1.

The value of the target function is

$$Q_2 = \sum_{i=1}^{n} \mid e_i \mid = \sum_{i=1}^{n} \Big| \, y_i - \hat{y}_i \, \Big| = \sum_{i=1}^{n} \Big| \, y_i - (1.5 + x_i) \, \Big| \tag{9.7}$$

Note that this value is 5. As this is lower than the previous value, we headed in the right direction; we are closer to the estimates because we have reduced the sum of the absolute errors.

We got closer. Note that the error for higher x-values is greater than for lower x-values. This suggests we should increase the slope yet again. So, let us select our third line as $\ell_3 : y = 0 + 2x$. Again, we are forcing the line to pass through the dimension-wise median.[2] The last pair of columns in Table 9.1 provide the predictions and residuals for this third line.

The value of the target function for this third line is $Q_3 = 6$. This value is not lower than Q_2. Thus, this line is a worse fit than line ℓ_2. The next line, ℓ_4, needs to take this into consideration.

This process would continue until the change in target function values is "small enough." Usually, we define "small enough" as being less than some tolerance, like $\tau = 0.000\,001$.

[2] Do we need to do this? No. There are algorithms that do not force this restriction. Again, the actual mathematics cannot reasonably be done by hand. I write this part so that you can get a feel for what the computer is doing.

9.1.1 The Big Question

The big question is how we get from one line-of-best-fit to the next, from ℓ_i to ℓ_{i+1}. Unfortunately, there is no "best method" to minimize the L_1 norm when there are more points than dimensions. It is even worse: We were able to find closed-form solutions to the unique estimators for the L_2 norm (squaring). That cannot be done when using the L_1 norm (absolute values). There are multiple appropriate algorithms. The estimators may not be unique. Those are just a *few* problems working with the L_1 norm.

For those interested, here are some methods:

- Barrodale and Roberts (1974),
- Koenecker and Bassett (1978),
- Koenker and d'Orey(1987, 1994),
- Li and Arce (2004), and
- Shu-guang and Jian-wen (1992).

Note: These algorithms make use of different paradigms, different ways of seeing the problems. That is what makes studying statistics fun and interesting. Looking at a problem differently may be the key to its solution.

In `R`, a function to perform median regression is `rq` in the package `quantreg`, which does not come with the default `R` installation. Its use is very similar to what we are used to. While the `rq` function allows you to select different optimization methods, the default is the Barrodale and Roberts (1974) method.

From my experience the optimization algorithm matters little for real data. If the data are all integers, there may be issues with non-unique solutions or non-convergent algorithms. The cause in these cases is the non-uniqueness of the median. **problem**

Example 1

Using median regression, what is the relationship between the violent crime rates in 2000 and 1990 in the `crime` data?

Solution: The following code estimates the median regression line for the relationship between the violent crime rates in 2000 and 1990 in the `crime` data:

```
library(quantreg)

dt = read.csv("http://rfs.kvasaheim.com/data/crime.csv")
attach(dt)

mod1 = rq(vcrime00 ~ vcrime90)
summary(mod1)
```

The following is the output:

```
Call: rq(formula = vcrime00  ~  vcrime90)

tau: [1] 0.50

Coefficients:
             coefficients lower bd  upper bd
(Intercept)  93.24525     72.83955 102.31731
vcrime90      0.57764      0.57518   0.62676
```

The output is the usual output. The value of `tau` is 0.50, because we are examining the regression line for the median, the 50th percentile.

The coefficients are the estimates for the intercept and slope. The lower and upper bounds are the 95% confidence interval for those parameters. There are no p-values, because the distribution of the estimators does not follow a nice test distribution. However, because we have a confidence interval, we have even more information than what a simple p-value would give. We are 95% confident that the relationship between the violent crime rate in 1990 and 2000 is between 0.575 and 0.627. Since this does not include the value 0, we can conclude that there is a significant relationship between the two variables. ♦

exercise

I leave it as an exercise for you to see that the OLS estimator for that effect is 0.581, with a 95% confidence interval from 0.518 to 0.643.

There is a difference between the two estimation methods. That difference is in how the method is affected by influential points like the District of Columbia. Median regression reduces the influence of DC, while ordinary least squares does not.

influential

The absolute value function increases linearly as the residual increases. The squaring function increases quadratically as the residual increases. Thus, ordinary least squares will work harder to avoid making the DC residual too big. Median regression will not weight it as heavily.

Example 2

Here is another example of using median regression. What is the relationship between the property crime rates in 1990 and 2000?

Solution: The code is quite similar to that above:

```
mod3 = rq(pcrime00 ~ pcrime90)
summary(mod3)
```

The following is the partial output:

```
Coefficients:
             coefficients lower bd   upper bd
(Intercept)  730.46936    349.56585 1093.31979
pcrime90       0.60584      0.50893    0.77457
```

Again, the relationship is positive. A point estimate for that relationship is $\tilde{\beta}_1 = 0.606$, with a 95% confidence interval from 0.509 to 0.775. I again leave it as an exercise for you to show that the OLS estimator is 0.582 with a 95% confidence interval from 0.458 to 0.707. ♦

Note: When the data are "well behaved" without influential points, there tends to be little difference in the estimators. Figure 9.1 illustrates this.

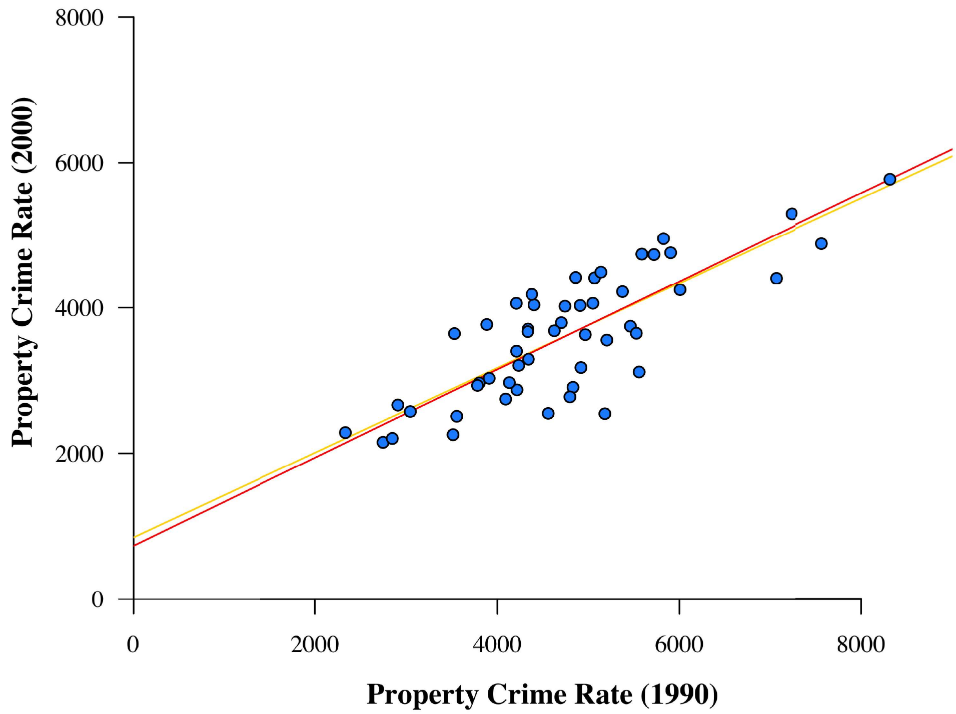

Figure 9.1: *A graphic comparing the estimated lines from ordinary least squares (gold) and median (red) regression.*

9.2: Quantile Regression

The previous section covered median regression. There, we motivated the method by focusing on minimizing the sum of the absolute value of the residuals. It turns out that this is equivalent to estimating the conditional median of the dependent variable (hence its name). In other words, the line of best fit is the line that best goes through the medians at each x-value.

conditional median

Compare this to how we motivated ordinary least squares in Chapter 2: by minimizing the sum of the squared errors. This is equivalent to estimating the conditional mean of the dependent variable.

conditional mean

In other words, OLS estimates $\mathbb{E}[Y \mid x]$ while median regression estimates $\text{Med}[Y \mid x]$, for want of better notation. (Perhaps $Q_2[Y \mid x]$ would be better notation?)

$P_{50}[Y \mid x]$**?**

There is absolutely no reason we need to focus *only* on the conditional median of the dependent variable (conditional on the independent variable). We may want to focus on other quantiles, like the 10th percentile. This happens a lot in sociology when studying poverty (10th percentile of income) or education (90th percentile of academic achievement).

The idea behind the fitting is the same (Koenker and Hallock 2001). The `R` function is also the same. The only difference is that you need to specify the quantile. To see this, let us see a couple familiar examples.

Example 3

What is the relationship between the violent crime rates in 2000 and 1990 in the `crime` data at the 10th percentile?

Solution: Here is the code the perform this estimation:

```
mod5 = rq(vcrime00 ~ vcrime90, tau=0.10)
summary(mod5)
```

The following is the output:

```
Call: rq(formula = vcrime00  ~  vcrime90, tau = 0.1)

tau: [1] 0.1

Coefficients:
             coefficients lower bd  upper bd
(Intercept)  40.29964    -14.80397 100.38757
vcrime90      0.55616      0.38948   0.60422
```

Thus, for those states near the 10th percentile, the relationship between the 1990 and 2000 violent crime rate is between 0.389 and 0.604, with a point estimate of 0.556. This is only a little different from the median results, which suggests those states that are less crime-ridden (at the 10th percentile) still followed the same "rule" with respect to violent crime rate changes between 1990 and 2000. ♦

Example 4

What is the relationship between the property crime rates in 2000 and 1990 in the `crime` data at the 90th percentile?

Solution: Here is the code to perform this estimation:

```
mod6 = rq(pcrime00 ~ pcrime90, tau=0.90)
summary(mod6)
```

The following is the output:

```
Call: rq(formula = pcrime00  ~  pcrime90, tau = 0.9)

tau: [1] 0.9

Coefficients:
             coefficients lower bd   upper bd
(Intercept) 1761.72465    327.54503 2436.15997
pcrime90       0.53326      0.40262    0.84489
```

Thus, for those near the 90th percentile, the relationship between the 1990 and 2000 property crime rate is between 0.403 and 0.845, with a point estimate of 0.533. This differs a little from the median results (Example 9.1.1), which suggests those states

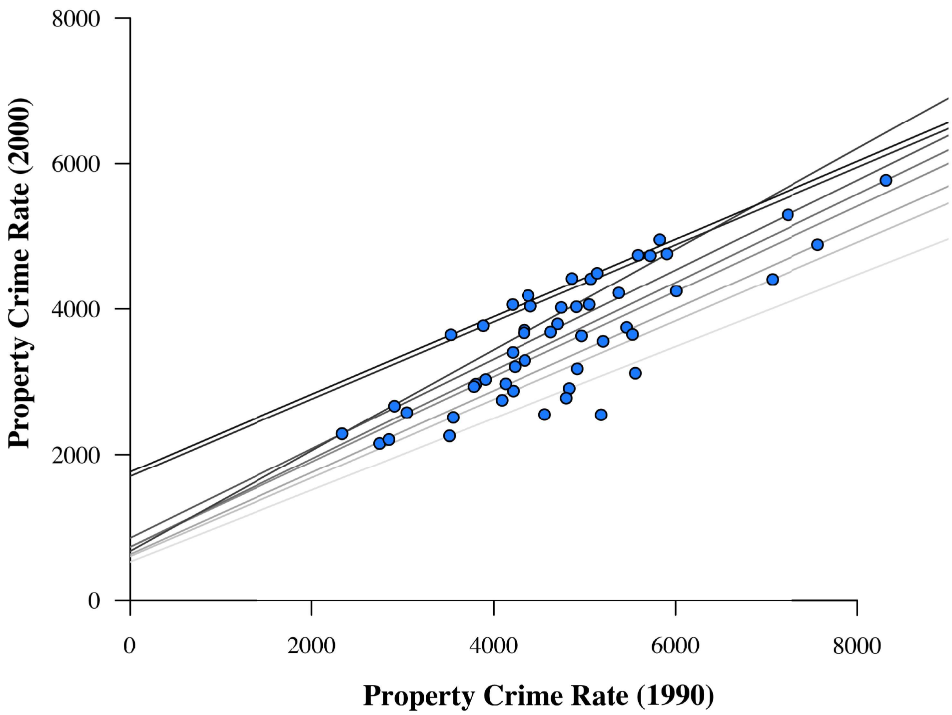

Figure 9.2: *Graphic illustrating the changing effect based on the quantile examined. The nine lines are regression lines for the deciles 0.10 through 0.90, with darker lines corresponding to higher quantiles.*

that are more (property) crime-ridden (at the 90th percentile) followed a similar "rule" with respect to violent crime rate changes between 1990 and 2000. Their rates dropped slightly more than did the typical (median) state. ♦

By the way, Figure 9.2 is a graphic of the deciles from 10 to 90% for the relationship between property crime rates in 1990 and 2000. Note that the effect does appear to change as one looks at middle-rate states. The highest levels, quantiles 80 and 90, are very similar in effect to the lower levels, quantile 10 and 20. However, those states near quantile 50 seem to have a greater slope. If we had only looked at the median, we would have only reported these steeper effects. This may have overstated the effect.

Example 5

What is the relationship between the state's wealth in 1990 and the property crime rate in 2000? Show the effects at the first, second, and third quartiles.

proxy

Solution: We will use the GSP per capita as a proxy measure of wealth in the state. I leave the coding to you. Here is the appropriate output for the median:

```
Call: rq(formula = pcrime00  ~  gspcap90)

tau: [1] 0.5

Coefficients:
            coefficients lower bd   upper bd
(Intercept) 3061.50674   2383.13082 4989.55600
gspcap90       0.02403     -0.08109    0.04996
```

Interpreting the table indicates that there is no significant evidence that there is a relationship between the average wealth in 1990 and the property crime rate in 2000 for the median state (the confidence interval contains 0).

For the first and third quartile, the conclusions will be the same. As both confidence interval contain both positive and negative numbers, we are unsure of the relationship between these two variables. ♦

I leave it as an exercise for you to show that ordinary least squares indicates a statistical significant relationship ($p-\text{value} = 0.0475$). It also provides a point estimate of that relationship of $b_1 = 0.03025$).

Figure 9.3 provides the results for all deciles. Note that the slopes also seem to vary according to the quantile examined. Thus, the effect of wealth on property crime rates seem to be a function of those property crime rates. The lowest quantiles suggest the steepest effect. However, performing the analysis shows that the relationship is not statistically significant at the $\alpha = 0.05$ level. In other words, we were unable to detect a relationship.

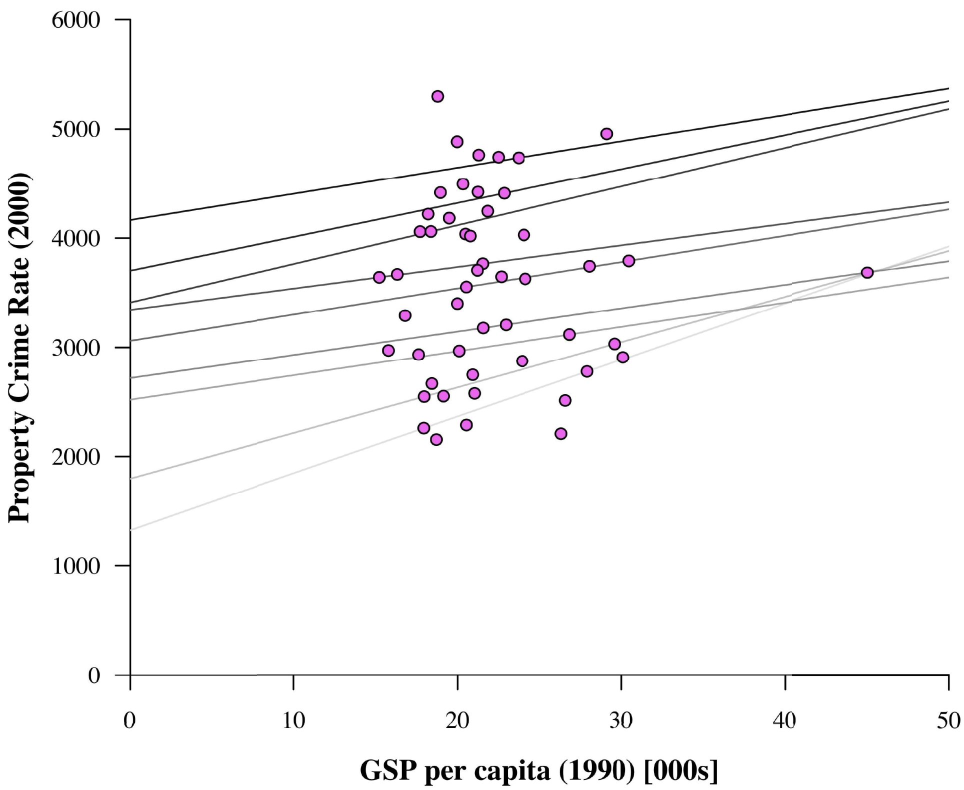

Figure 9.3: *Graphic illustrating the changing effect based on the quantile examined. The nine lines are regression lines for the quantiles* 0.10 *through* 0.90, *with darker lines corresponding to higher quantiles.*

9.2.1 The Ultimate Question

So, is there a relationship between average wealth in 1990 and the property crime rate in 2000? One thing we know is that if there is a relationship, then it is minor.

It is not surprising that median regression does not detect a relationship while ordinary least squares does. Median regression, like all statistics based on the median (ranks), has a lower power than ordinary least squares (these statistics require Normality).

Note: So, the answer to the ultimate question is "I'm not sure." This is unsatisfying. It is also reality. By using both OLS and median regression, we have a better understanding of the relationship between average wealth and property crime rates. That is the goal of statistics, not coming up with binary answers.

Douglas Adams

9.2.2 The Ultimate Answer

42

9.3: Conclusion

In this chapter, we covered quantile regression. We initially motivated the topic by modifying our definition of "best fit" to focus on the absolute value of the residuals in lieu of the square of the residuals. This led to an iterative process that allowed us to obtain estimates to any desired accuracy — at the cost of time and computing power.

This chapter then noted that median regression was just a specific instance of quantile regression, one in which the quantile was 0.500. This set the stage to introduce the results of quantile regression, in general. One may see quantile regression in research that focuses on better understanding the "wings" of the distribution instead of its middle.

Quantile regression uses **the entire data set**. It does not look at only the data corresponding to the q^{th} quantile. Such data may not actually exist. What states are at the 10th quantile of the property crime rates in 1990 and 2000? That's not enough data to obtain any meaningful estimates.

Quantile regression estimates the q^{th} quantile of the response variable given the value of the independent variable.

9.4: End-of-Chapter Materials

9.4.1 R Functions

In this chapter, we were introduced to several R functions that will be useful in quantile regression. These are listed here.

Packages:

quantreg This package contains many functions associated with quantile regression. This chapter just skimmed the surface of what can be done and what should be checked. As this package is not a part of the base R, you will need to install it before loading it with `library(quantreg)`.

Statistics:

rq(formula) This is the function that performs quantile regression. The formula is required. By default, the quantile examined is $\tau = 0.50$, but that can be changed by specifying the value of that τ.

summary(model) This is the familiar command that allows us to see the regression table produced by the regression method. Here, it provides the effect estimates (coefficients) and the central 95% confidence interval for that effect.

9.4.2 Exercises

1. In the setting of Example 9.1.1, perform ordinary least squares regression to calculate the effect estimate and its 95% confidence interval.

9.4.3 Applied Readings

- Lingxin Hao and Daniel Q. Naiman (2008). *Assessing Inequality*. Sage Publishing (Quantitative Applications in the Social Sciences; 166).

9.4.4 Theory Readings

- Ian Barrodale and F. D. K. Roberts (1974) "Solution of an overdetermined system of equations." *Communications of the ACM* 17(6), 319–320. doi: 10.1145/355616.361024
- Lingxin Hao and Daniel Q. Naiman (2007). *Quantile Regression*. Sage Publishing (Quantitative Applications in the Social Sciences; 149).
- Roger W. Koenker (2005). *Quantile Regression*. Cambridge University Press.
- Roger W. Koenker and Gilbert W. Bassett, Jr. (1978). "Regression quantiles," *Econometrica*, 46, 33–50.
- Roger W. Koenker and Kevin F. Hallock (2001). "Quantile Regression," *The Journal of Economic Perspectives*, 15: 143–156.
- Roger W. Koenker and Vasco D'Orey (1987). "Algorithm 229: Computing regression quantiles." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36, 383–393.
- Roger W. Koenker and Vasco D'Orey (1994). "Remark AS R92: A Remark on Algorithm AS 229: Computing Dual Regression Quantiles and Regression Rank Scores." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, 410–414.
- Yinbo Li and Gonzalo R. Arce (2004). "A Maximum Likelihood Approach to Least Absolute Deviation Regression." *EURASIP Journal on Applied Signal Processing*, 12: 1762–1769.
- Yang Shu-guang and Liao Jian-wen (1992). "Solution of an Overdetermined System of Linear Equations in L_2, L_∞, L_p Norm Using L.S. Techniques." *Journal of Computational Mathematics*, 10: 29–38.