

# QUANTILE REGRESSION LITERATURE REVIEW

## Motivations

- [3] “...the quantile regression results offer a much richer, more focused view of the applications than could be achieved by looking exclusively at conditional mean models. In particular, it provides a way to explore sources of heterogeneity in the response that are associated with the covariates. ” –*Quantile Regression, Koenker, page 25*  
  
Robustness: “The comparison of the influence functions of the mean and median graphically illustrates the fragility of the mean and the robustness of the median in withstanding the contamination of outlying observations. Much of what has already been said extends immediately to the quantiles generally, and from there to quantile regression. ” –*Quantile Regression, Koenker, page 43*
- [7] Quantile regression is gradually emerging as a comprehensive approach to the statistical analysis of linear and nonlinear response models. (...) Quantile regression supplements the exclusive focus of least squares based methods on the estimation of conditional mean functions with a general technique for estimating families of conditional quantile functions. This greatly expands the flexibility of both parametric and nonparametric regression methods.
- [6] Asymmetry as well as long tails (which means very extreme outcomes from a distribution have non-negligible probabilities), are common to not only in economics but also a number of other disciplines such as social sciences, medicine, public health, financial return, environment and engineering.  
  
Min and Kim (2004) claim that over a wide-class of non Gaussian errors, with asymmetric and long-tailed distributions, simple mean regression cannot satisfactorily capture the key properties of the data; even the conditional mean estimation can be misleading.
- [4] Taken together the ensemble of estimated conditional quantile functions offers a much more complete view of the effect of covariates on the location, scale and shape of the distribution of the response variable.  
  
Covariates may influence the conditional distribution of the response in myriad other ways: expanding its dispersion as in traditional models of heteroscedasticity, stretching one tail of the distribution, compressing the other tail, and even including multi modality. Explicit investigation of these effects via quantile regression can provide a more informative empirical analysis.
- [5] Quantile regression is a method that provides a more complete inferential picture than ordinary least-squares regression. It allows the full range of the data to be modeled and so can be beneficial when large or small response values are of particular interest. It also means that quantile regression can be viewed as a data-exploration technique.
- [1] Longitudinal data are very common in biological and medical research. Correlation may arise when data are measured repeatedly from the same subject. Various methods have been developed to analyze the longitudinal data, which can be only used to evaluate covariate effects on the mean of a response variable. To give a global assessment about covariate effects on the distribution of the response variable, a quantile regression model is an important alternative(Koenker and Bassett, 1978).

Each quantile regression characterizes a particular point of a distribution, and thus provides more complete description of the distribution. Furthermore, quantile regression is more robust against outliers and does not require specifying any error distribution.

Jung (1996) first extended the quantile regression to longitudinal data and developed a quasi-likelihood method for median regression. Koenker (2004) considered a linear quantile regression with subject specified fixed effects and made inferences from a penalized likelihood. Geraci and Bottai (2007) introduced random intercepts to account for the within- subject correlations and presented a likelihood-based approach by assuming the response variable following an asymmetric Laplace density. Liu and Bottai (2009) generalized the model by Geraci and Bottai (2007) to linear mixed-effects models with random regression coefficients and assumed random effects by a multivariate Laplace distribution.

- [2] Mixed effects models, also known as multilevel or hierarchical or random-effects models, represent highly popular and flexible models to analyze complex data. They model and estimate between-cluster variability by means of cluster-specific random effects. These, in turn, provide a modeling structure for estimating the intraclass correlation coefficient (ICC).

There are several reasons why an analyst might consider a random-effects approach for their data while wanting to go beyond a location-shift model. A quantile regression approach allows for departures from location-shift assumptions, typical of LMMs, in which covariates do not affect the shape of the distribution.

## Loss function (Wiki)

- Problems that seek to optimize a linear function subject to linear constraints are called *linear programs*. –*Quantile Regression, Koenker, page 173*
- In mathematical optimization, statistics, decision theory and machine learning, a **loss function** or **cost function** is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event. An optimization problem seeks to minimize a loss function. An **objective function** is either a loss function or its negative (sometimes called a **reward function** or a **utility function**), in which case it is to be maximized. –*Wikipedia*
  - In statistics, typically a loss function is used for *parameter estimation*, and the event in question is some function of the difference between estimated and true values for an instance of data.
  - *Formal definition*: More intuitively, we can think of  $X$  as our "data", perhaps  $X = (X_1, \dots, X_n)$ , where  $X_i \sim F_\theta$  i.i.d. The  $X$  is the set of things the decision rule will be making decisions on. There exists some number of possible ways  $F_\theta$  to model our data  $X$ , which our decision function can use to make decisions. For a finite number of models, we can thus think of  $\theta$  as the index to this family of probability models. For an infinite family of models, it is a set of parameters to the family of distributions.
  - From here, given a set  $A$  of possible actions, a *decision rule* is a function  $\delta : \mathcal{X} \rightarrow A$ .
  - A loss function is a real lower-bounded function  $L$  on  $\Theta \times A$  for some  $\theta \in \Theta$ . The value  $L(\theta, \delta(X))$  is the *cost of action*  $\delta(X)$  under parameter  $\theta$ .
  - **Expected loss**: The value of the loss function itself is a random quantity because it depends on the outcome of a random variable  $X$ . Both frequentist and Bayesian statistical theory involve making a decision based on the expected value of the loss function: however this quantity is defined differently under the two paradigms.
    - \* Frequentist: This is also referred to as the **risk function** of the decision rule  $\delta$  and the parameter  $\delta$ . Here the decision rule depends on the outcome of  $X$ .

$$R(\theta, \delta) = \mathbb{E}_\theta L(\theta, \delta(X)) = \int_{\mathcal{X}} L(\theta, \delta(x)) \, dP_\theta(x).$$

- \* Bayesian: the expectation is calculated using the posterior distribution  $\pi^*$  of the parameter  $\theta$ :

$$\rho(\pi^*, a) = \int_{\Theta} L(\theta, a) \, d\pi^*(\theta).$$

One then should choose the action  $a^*$  which minimizes the expected loss.

- **Decision rules:** A decision rule makes a choice using an optimality criterion. Some commonly used criteria are:

- \* *Minimax:* Choose the decision rule with the lowest worst loss – that is, minimize the worst-case (maximum possible) loss:

$$\arg \min_{\delta} \max_{\theta \in \Theta} R(\theta, \delta).$$

- \* *Invariance:* Choose the optimal decision rule which satisfies an invariance requirement.
- \* Choose the decision rule with the *lowest average loss* (i.e. minimize the expected value of the loss function):

$$\arg \min_{\delta} \mathbb{E}_{\theta \in \Theta} [R(\theta, \delta)] = \arg \min_{\delta} \int_{\Theta} R(\theta, \delta) p(\theta) \, d\theta.$$

- **Selecting a loss function:**

\*For most optimization algorithms, it is desirable to have a loss function that is globally continuous and differentiable.

\*Two very commonly used loss functions are the squared loss,  $L(a) = a^2$ , and the absolute loss,  $L(a) = |a|$ . However the absolute loss has the disadvantage that it is not differentiable at  $a = 0$ . The squared loss has the disadvantage that it has the tendency to be dominated by outliers—when summing over a set of  $a$ 's (as in  $\sum_{i=1}^n L(a_i)$ ), the final sum tends to be the result of a few particularly large  $a$ -values, rather than an expression of the average  $a$ -value.

\*Many common statistics, including  $t$ -tests, regression models, design of experiments, and much else, use least squares methods applied using linear regression theory, which is based on the quadratic loss function.

## References

- [1] Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.
- [2] Marco Geraci and Matteo Bottai. Linear quantile mixed models. *Statistics and Computing*, pages 1–19, 2013.
- [3] Roger Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.
- [4] Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.
- [5] Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649, 2012.
- [6] Craig Reed and Keming Yu. A partially collapsed gibbs sampler for bayesian quantile regression. 2009.
- [7] Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.