

# Linear Quantile Regression

The addition of [G02QFF](#) and [G02QGF](#) at Mark 23 adds linear quantile regression to supplement the wide variety of regression techniques already available in the NAG libraries.

Linear quantile regression is related to linear least-squares regression in that both are interested in studying the linear relationship between a response variable and one or more independent or explanatory variables. However, whereas least-squares regression is concerned with modelling the conditional mean of the response variable, quantile regression models the conditional  $\tau$ th quantile of the response variable, for some value of  $\tau \in (0, 1)$ . So, for example,  $\tau = 0.5$  would be the median.

Given the vector  $y$  it is a well known result that the sample mean,  $\hat{y}$ , solves the least squares problem

$$\underset{\mu \in \mathbb{R}}{\text{minimize}} \sum_{i=1}^n (y_i - \mu)^2.$$

This result leads to least-squares regression where, given a design matrix  $X$  and defining the conditional mean of  $y$  as  $\mu(X) = X\beta$ , an estimate of  $\beta$  is obtained from the solution to

$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \sum_{i=1}^n (y_i - x_i^T \beta)^2.$$

Quantile regression can be derived in a similar manner by specifying the  $\tau$ th conditional quantile as  $Q_y(\tau | X) = X\beta(\tau)$  and estimating  $\beta(\tau)$  as the solution to

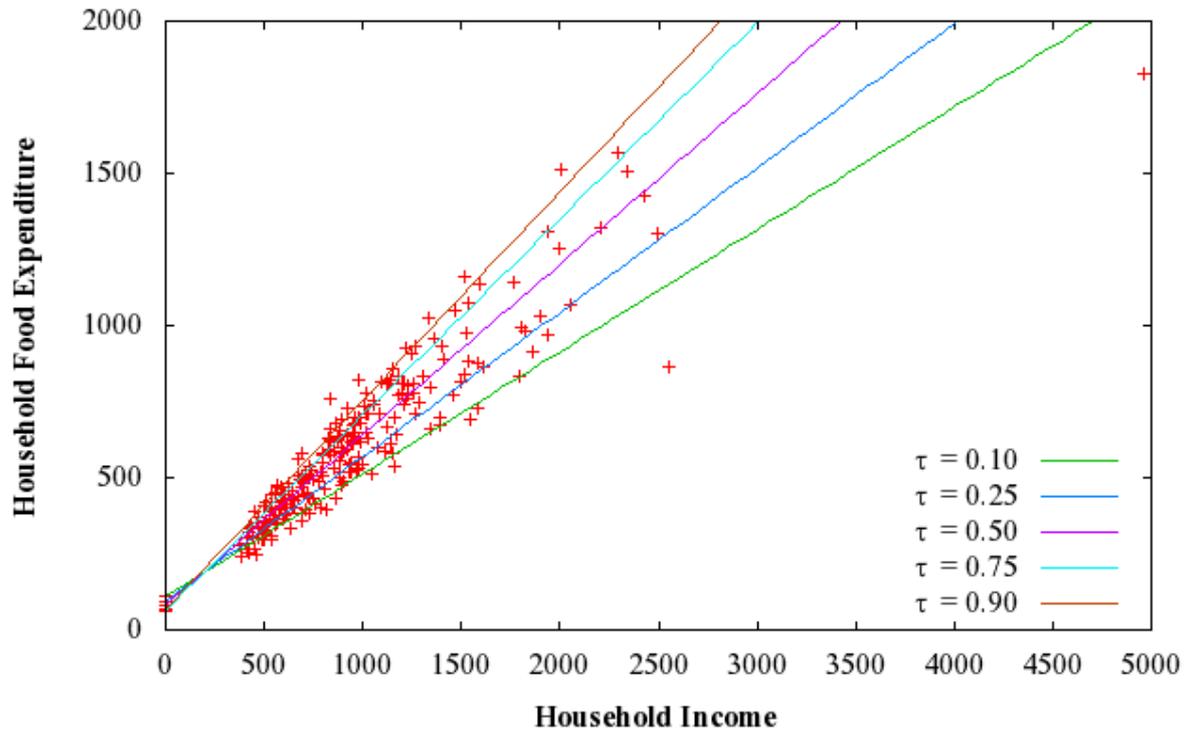
$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta),$$

where  $\rho_\tau(z)$  is a linear loss function defined as  $z(\tau - 1)$  if  $z < 0$  and  $z\tau$  otherwise.

As quantile regression allows multiple quantiles to be modelled it can allow for a more comprehensive analysis of the data to be carried out compared to least-squares regression where only the mean is considered. This potentially enables more insight into the data and any underlying relationships, in addition, it will tend to be less sensitive to large outlying observations. Additional information is given in the [G02 Chapter Introduction](#) and a comprehensive description of the theory, application and interpretation of quantile regression can be found in [Koenker \(2005\)](#).

The following simple example shows some results from using quantile regression to investigate the relationship between household food expenditure and income. The data is taken from Engels 1857 study of expenditure on food. It can be seen that the relationship is markedly different for those households with a high expenditure compared to those with a low expenditure

**Example Program**  
Quantile Regression - Simple Interface  
Engels 1857 Study of Household Expenditure on Food



Koenker R (2005) *Quantile Regression* Econometric Society Monographs, Cambridge University Press, New York