Appendix A: Mathematical Formulation of Quantile Regression

In classic Ordinary Least Squares (OLS) regression, we estimate the solution to the following equation for each regression coefficient \( \beta_i \).

\[
y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i
\]

The estimated solution \( \hat{\beta} \) minimizes the sum of squared residuals, and is estimated by

\[
\hat{\beta} = \arg\min_{b \in \mathbb{R}^p} S(b) = (\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i y_i
\]

or

\[
\hat{\beta} = (X^T X^{-1}) X^T y.
\]

In quantile regression, rather than estimating the linear conditional mean function \( E(Y|X = x) \), we will estimate the linear conditional quantile function, \( Q_Y(\tau|X = x) = x_i^T \beta(\tau) \). This can be estimated by solving

\[
\hat{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^p} \sum \rho_\tau(y_i - x_i^T \beta),
\]

where \( \tau \) is the quantile of interest (e.g. \( \tau = 0.5 \) corresponds to the median), and \( \rho_\tau(u) = u(\tau - I(u < 0)) \). The formulation of the quantile regression model for the \( \tau \)th quantile will be express like Equation (1), but the coefficients will correspond to the change in the outcome associated with the given covariate at that quantile of the conditional distributions.