线性分位 (Linear Quantile) 回归

在第 23 版中新建线性分位回归 **G02QFF** 与 **G02QGF** 函数，在现有 NAG 算法库中已经相当完整的回归部分再新建函数。

线性分位回归与线性最小二乘回归相同，两者都是针对依变数与一或多个自变量或实验变量间关系的研究。然而，最小二乘回归是由依变量的条件均值所建立的模型，分位回归是以依变量的第 τ 位数条件所构成的模型，其中 $\tau \in (0, 1)$。所以举例来说，$\tau = 0.5$ 表示是中位数。

假设向量 y，有一个已知的样本平均值，求最小二乘问题

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

此将会带出最小二乘回归，若定义矩阵 X 与 y 的条件平均值为 $\mu(X) = X\beta$，其中的 β 估计值将会由下式获得：

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

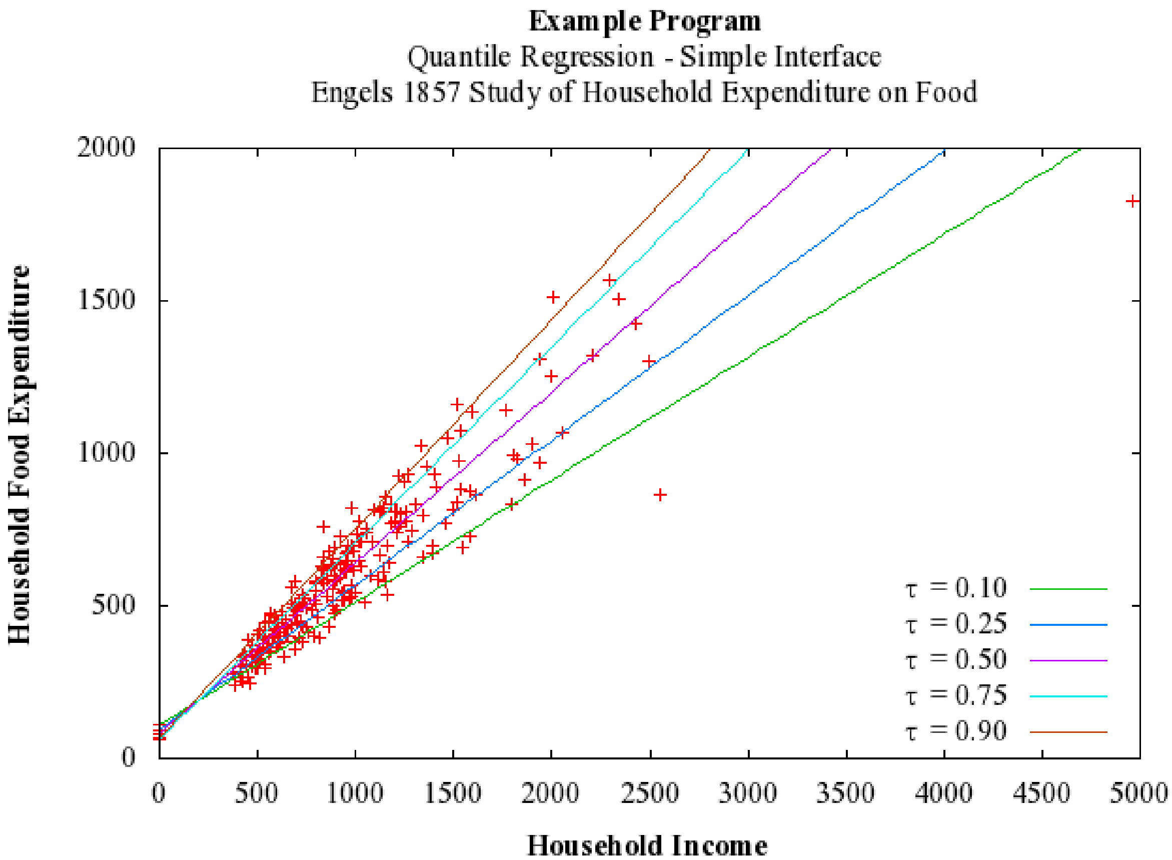
分位回归可以透过类似的方式指定 τth 条件分位数 $Q_y(\tau \mid X) = X\beta(\tau)$ 导出，其 $\beta(\tau)$ 求解如下：

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

其中 $\rho_\tau(z)$ 是线性损失 (loss) 函数，定义为：$z(\tau - 1)$ if $z < 0$，否则为 $z\tau$。

相较于最小二乘回归仅考虑平均值，分位回归允许多分位数建模，将能对资料做更全面的分析。潜在来说，这对资料与任何隐含的关系将能提供更多的观察，此外，其对大量离群值也较不敏感。其他更多资讯可以参考 **G02 Chapter 介绍**，关于分位回归的理论、应用及更详尽的解说可参考 **Koenker (2005)**。

以下的简单例子显示采用分位回归探讨家庭食品支出与收入之间的关系。资料取自 Engels 1857 食品消费研究。由此可以看出对家庭食品高消费与低消费间显著的不同。



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