行政院國家科學委員會專題研究計畫成果報告

Mallow 型式迴歸分位向量
Mallow Type Regression Quantile

計畫編號：NSC 87-2118-M-009-001
執行期限：86 年 8 月 1 日至 87 年 7 月 31 日
主持人：陳鄰安　　國立交通大學統計學研究所

一、　中文摘要

對於線性迴歸及聯立方程模型成份提出 Mallow 型式的方位向量。模擬分析顯示此一方位向量具有有效性。

二、Abstract

We present asymptotic distributions of the Mallow's type bounded-influence regression quantile for the linear regression model and also the simultaneous equations model. Monte Carlo simulation comparing means squared errors shows that the bounded-influence one is more efficient than the unbounded-influence one (Koenker and Bassett (1978)) when gross errors occur in the independent-variables-space.

三、Introduction

Consider the following linear regression model,

$$y_i = x_i'\beta + \varepsilon_i$$

where x_i is the observation of p independent variables including term of intercept, β is vector of regression parameters and ε_i's are i.i.d. disturbance variables with distribution function F. The conditional quantile of variable y is $x'\beta + F^{-1}(\alpha)$, 0< α <1 which can be expressed as $x'\beta(\alpha)$ with $\beta(\alpha) = \beta + \begin{pmatrix} F^{-1}(\alpha) \\ 0 \end{pmatrix}$ and where 0 is $(p-1)$-vector of zeros. As an extension of the sample quantile to the linear model, Koenker and Bassett (1978) introduced the regression quantile, as an estimator of $\beta(\alpha)$, as the solution for the following minimization problem

$$\min_{b \in R^p} \sum_{i=1}^{n} \rho_\alpha (y_i - x_i' b)$$

where $\rho_\alpha(u) = u\psi_\alpha(u)$, $\psi_\alpha(u) = \alpha - I(u < 0)$ with *I(A)* the indicator function of the event *A*. This has since been widely applied to construct robust estimators; see, for example, Ruppert, and Carroll (1980), Jurečková (1984), Koenker and Portnoy (1987) and Chen and Portnoy (1996). As obtained from Koenker and Bassett (1978) and Ruppert and Carroll (1980), the regression quantile $\hat{\beta}(\alpha)$ has asymptotic representation with the following influence function,

$$Q^{-1} f^{-1}\left(F^{-1}(\alpha)\right) x \psi_\alpha(\varepsilon)$$

where Q is a positive definite matrix which will be defined later and f and F^{-1} represent the p.d.f. and inverse distribution function of the error random variable. Note that the function ψ limits the effect of the residual and thus the influence function is bounded in the dependent-variable-space; however, it is not bounded in the independent-variable-space. Therefore, one can conjecture that in small samples the regression quantile will be able to handle outliers in the y space but not in the X space. For general discussion of influence analysis, see Cook and Wesley (1982).

In literature, considerarion has been given to the development of estimators of regression parameters β that limit the effects of the error variable and the independent variables. In light of the fact that bounded-influence type regression quantile has not been studied, our aim is to study the Mallows type regression quantile for the linear regression model and the simultaneous equations model.

四、Result and discussion

Let w_i, i=1, …, n, be real numbers. For 0< α <1, the Mallow's type bounded-inference regression quantile, denoted by $\hat{\beta}_{BI}(\alpha)$, is defined as the solution for the minimization problem

$$\min_{b\in R^p}\sum_{i=1}^{n} w_i \rho_\alpha\left(y_i - x_i'b\right).$$

Theorem 1.

(a)

$$n^{1/2}\left(\hat{\beta}_{BI}(\alpha)-\beta(\alpha)\right)= f^{-1}\left(F^{-1}(\alpha)\right)Q_w^{-1}n^{-1/2}\sum_{i=1}^{n} w_i x_i \Psi_\alpha\left(\varepsilon_i - F^{-1}(\alpha)\right)+o_p(1)$$

and

(b)

$$n^{1/2}\left(\hat{\beta}_{BI}(\alpha)-\beta(\alpha)\right)\to N\left(0,\alpha(1-\alpha)f^{-2}\left(F^{-1}(\alpha)\right)Q_w^{-1}Q_{ww}Q_w^{-1}\right)$$

where $\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n} x_i x_i' = Q$, $\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n} w_i x_i x_i' = Q_w$, and $\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n} w_i^2 x_i x_i' = Q_{ww}$, where Q, Q_w and Q_{ww} are $p\times p$ positive definite matrices.

Consider the simultaneous equations model

$$y = Y_1\beta_1 + Z_1\beta_2 + \lambda.$$

Let the reduced from of the simultaneous equations model be

$$Y = Z\Pi + V$$

where $Y=(y,Y_1)$, and $Z=(Z_1,Z_2)$ and rows of V are i.i.d. random vectors. The first stage is to estimate Π_2 by an initial estimator $\hat{\Pi}_2$ for the reduced model (3.2). Define $\hat{Y}_1 = Z\hat{\Pi}_2$. We have

$$y = D_n\beta + U$$

where $D_n = \left(Z\hat{\Pi}_2, Z_1\right)$, $\beta' = \left(\beta_1', \beta_2'\right)$, and $U = v_1 - Z\left(\hat{\Pi}_2 - \Pi_2\right)\beta_1$.

Let w_i, i = 1, …, n, be real numbers. For 0< α <1, we define the bounded-influence two stage regression quantile as an alternative estimator of $\beta(\alpha)$ as

$$\hat{\beta}_{BI}(\alpha) = \arg\min_{b\in R^{p_0+p_1-1}}\sum_{i=1}^{n} w_i \rho_\alpha\left(y_i - d_i'b\right)$$

where d_i' is i-th row of matrix D_n.

The following assumptions are needed $n^{-1}\sum_{i=1}^{n} z_i z_i' = Q + o(1)$, $n^{-1}\sum_{i=1}^{n} w_i z_i z_i' = Q_w + o(1)$, and

$n^{-1}\Sigma_{i=1}^{n} w_i^2 z_i z_i' = Q_{ww} + o(1)$ where Q, Q_w and Q_{ww} are all positive definite matrices.

Denote v_{ji} as (ij)-th element of matrix V where $i = 1, \ldots,$ n and $j = 1, \ldots, p_0$.

Theorem 2.

$$n^{1/2}\left(\hat{\beta}_{BI}(\alpha) - \beta(\alpha)\right) = f_1^{-1}\left(F_1^{-1}(\alpha)\right)\Sigma_w^{-1} n^{-1/2}$$
$$\sum_{i=1}^{n} w_i \tilde{d}_i (\alpha - I(v_{1i} < F_1^{-1}(\alpha)) -$$
$$\Sigma_w^{-1}\begin{bmatrix}\Pi_2 & I_{p1} \\ & 0_{p2\times p1}\end{bmatrix}' Q_w n^{1/2}\left(\hat{\Pi}_2 - \hat{\Pi}_2\right)\beta_1 + o_p(1))$$

where we denote by $\Sigma_w = \begin{bmatrix}\Pi_2 & I_{p_1} \\ & 0_{p_2\times p_1}\end{bmatrix}' Q_w \begin{bmatrix}\Pi_2 & I_{p_1} \\ & 0_{p_2\times p_1}\end{bmatrix}$.

五、Selected References

Amemyia, T. (1982), Two Stage Least Deviations Estimators, Econometrica, 50, 689-711.

Chen, L.-A. and Portnoy, S. (1996), Regression Quantiles and Trimmed Least Squares estimators for Structural Equations Models, Communications in Statistics - Theory and Methods, 25, 1005-1032.

Koenker, R. W. and Portnoy, S. (1987), L Estimation for Linear Model, Journal of American Statistical Association 82, 851-857.