Self-normalized subsampling method for non-standard time series regression models

Waseda University Fumiya Akashi
University of Georgia Shuyang Bai
Boston University Murad S. Taqqu

This talk considers parameter estimation problem of a linear regression model $y_t = \beta_0^\top x_t + e_t$ $(t = 1, \dots, n)$, where x_t is the p-dimensional stationary process, $\beta_0 \in \mathbb{R}^p$ is the unknown parameter and e_t is the unobserved error process. We assume that $\{(x_t, e_t) : t \in \mathbb{Z}\}$ is a jointly stationary process which may have long-range dependence and/or heavy-tails. The goal of this talk is to construct a confidence region for β_0. To achieve the goal, we approximate the distribution of the statistic $Q_{1,n}(\beta_0) = \Psi_{1,n}(\beta_0)^\top \Omega_{1,n}(\beta_0)^{-1} \Psi_{1,n}(\beta_0)$, where

$$\Psi_{i,j}(\beta) = \sum_{t=i}^{j} \psi(y_t, x_t, \beta),$$

$$\Omega_{i,j}(\beta) = \frac{1}{j-i+1} \sum_{k=i}^{j} \left\{ \Psi_{i,k}(\beta) - \frac{k-i+1}{j-i+1} \Psi_{i,j}(\beta) \right\}^\top \left\{ \Psi_{i,k}(\beta) - \frac{k-i+1}{j-i+1} \Psi_{i,j}(\beta) \right\}$$

and $\psi : \mathbb{R}^{1+2p} \to \mathbb{R}^p$ is a user-specified score function. We can choose any function ψ provided that a solution $\hat\beta$ of the equation $\mathbb{E}[\psi(y_1, x_1, \beta)] = 0_p$ converges to β_0 in probability as $n \to \infty$. For example, we can choose $(y - \beta^\top x)x$ (L_2-regression) or $\mathrm{sign}(y - \beta^\top x)x$ (LAD-regression) as $\psi(y, x, \beta)$. Now, let us define the empirical distribution function of subsamples $\{Q_{i,i+b-1}(\hat\beta) : i = 1, \dots, n-b+1\}$ as

$$\hat F_{n,b} := \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \mathbb{I}\{Q_{i,i+b-1}(\hat\beta) \le x\},$$

where b is the length of each subsample. Then, we have the following result.

Theorem 1 (Akashi, Bai and Taqqu (2017)). *Suppose that* $b = o(n)$. *Then, under some regularity conditions, we have* $\left|P(Q_{1,n}(\beta_0) \le x) - \hat F_{n,b}(x)\right| \xrightarrow{p} 0 \ (n \to \infty)$ *for* x *at any continuous point of* $F_Q(x) = \lim_{n\to\infty} P(Q_{1,n}(\beta_0) \le x)$.

We also check the finite sample performance of the proposed method via simulation study, where we consider long-memory and heavy-tailed error term. It is observed that the proposed method is robust against long-range dependence and heavy-tails of the model.

Reference

[1] Akashi, F., Bai, S. and Taqqu, M.S. (2017). Robust regression on stationary time series: a self-normalized resampling approach. *Submitted for publication*.

[2] Bai, S. and Taqqu, M.S. (2016). On the validity of resampling methods under long memory. *To appear in The Annals of Statistics*.