Nonparametric and Robust Methods

Zuo, Chen

School of Statistics, Renmin Univ. of China

E-Mail: John_Zuo@ruc.edu.cn

Content

1. Robust Analysis: Why Quantile is Robust?
2. Quantile Regression: Parametric or Nonparametric?
3. Local Polinomial Regression and Bandwidth Selection
4. Confidence Interval Inference: Twilight Zones...

Robustness Analysis

In statistical language, we treat the estimator $\hat{\theta}$ as a function of the underlying distribution F,i.e. $\hat{\theta}(F)$.

The contamination of the underlying distribution F is considered as a small amount of mass ϵ concentrated at point y.

The contaminated distribution is $F_\epsilon = \epsilon\delta_y + (1-\epsilon)F$

Influence Function

$$IF_{\hat{\theta}}(y, F) = lim_{\epsilon \to 0} \frac{\hat{\theta}(F_\epsilon) - \hat{\theta}(F)}{\epsilon}$$

Robust Analysis: Mean and Quantile

Mean Estimator: $IF_{\hat{\theta}}(y, F) = y - \hat{\theta}(F)$

Quantile Estimator:

$$IF_{\hat{\theta}}(y, F) = \frac{sgn(y - \theta(F))}{f(F^{-1}(\tau))}$$

For **General Moment Estimators** of mean, i.e.
$\hat{\theta} = argmin_{\theta}(\sum \psi(y_i - \theta))^2$,

$$IF_{\hat{\theta}}(y, F) = \frac{\psi(y - \theta(F))}{E\psi'(y - \theta(F))}$$

From Quantiles to Quantile Regression

We have shown that **Sample Quantiles** is more robust than common mean.Koenker(1978) extended the thinking to the regression analysis literature. Thus came the **Quantile Regression** literature.

Model Form:

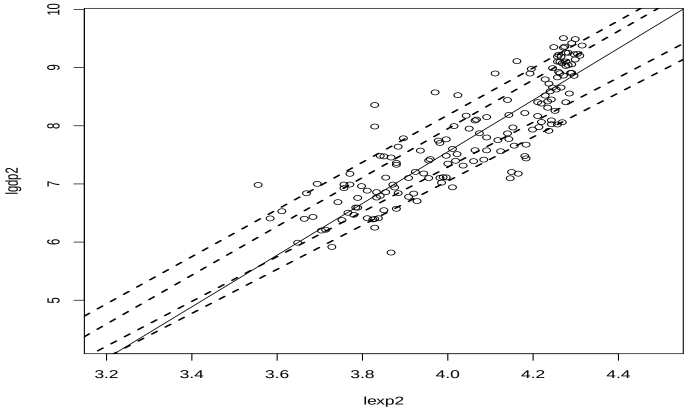
$$\hat{\beta}(\tau) = argmin \sum \rho_\tau(y_i - x_i^T \beta)$$

where the **check function** $\rho_\tau(u) = u(\tau - 1(u < 0))$

In R, the package **quantreg** provides estimation and testing for quantile regression model.

The problem: **Model Misspecification**.

Quantile Regression: Is Parametric Model Sufficient?



Nonparametric Smoothing

Model misspecification problem is inevitble because they are only convenient approximations but can not restore all the true trend functional forms, especially when such forms are complex.

Nonparametric estimation was first used for kernel estimation for density function.

$$\hat{f}(x) = \frac{1}{nh} \sum K(\frac{x_i - x}{h})$$

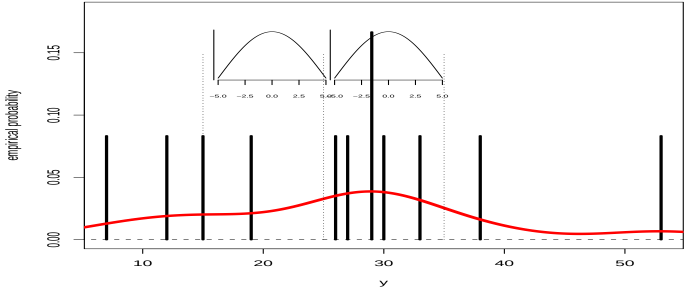
where $K(\cdot)$ is **kernel function** and h is the **bandwidth**

The **Nadaraya-Watson Nonparametric Estimation**:

$$m_n(x) = \frac{\sum y_i \cdot \frac{1}{h} K(\frac{x_i - x}{h})}{\sum \frac{1}{h} K((\frac{x_i - x}{h})}$$

Nonparametric Estimation as Moving Averages

The nonparametric regression is just a special type of moving average, where kernel function offers a certain set of weights. We can see the density estimation as a heauristic interpretation:



From Nonparametric Mean to Nonparametric Quantile

For quantile estimation, first express the estimator in terms of minimizer of conditional check function:

$$m_\tau(x) = argmin_{\theta \in \Theta} E\{\rho_p(Y - \theta) \mid X = x\}$$

Therefore, we can have the **local constant estimator**:

$$\hat{m}_\tau(x) = argmin_a \sum_i \rho_p(Y_i - a) K(\frac{X_i - x}{h})$$

This can also be extended to estimate to **local polynomial estimation**:

$$min_\beta \sum_i \rho_p(Y_i - \sum_{j=0}^{r} \beta_j (x_i - x)^j)) K(\frac{X_i - x}{h})$$

Statistical Properties and Bandwidth Selection

The bandwidth is the key parameter that should be treated seriously.

Generally, large h means small variance, but large bias.

The selection of regression order r is also important. For fixed j, when $r - j$ is odd, the bias of $\hat{\beta}_j$ is smaller. This is obtained by J. Fan,et al.(1994) when considering the asymptotic normality under general estimation functions.

The optimal bandwidth requires estimation of derivatives and is not very convenient.

Nonparametric Regression for Correlated Data

For dependent data, the aforementioned properties should be reconsidered.

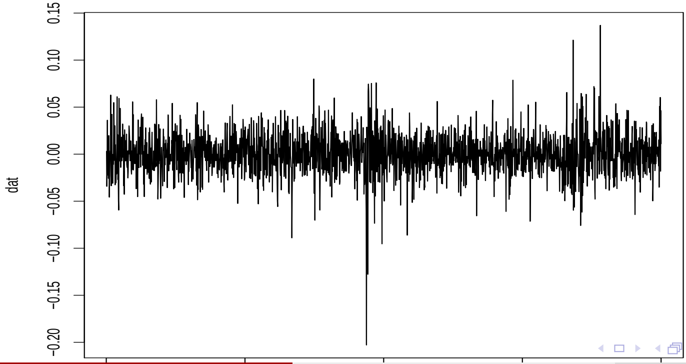
Intuitively, the **whitening by windowing principle** by Hart(1996) guarantees the consistency of estimators: *Given two points x_i and x_j in the design space, the random variable $\frac{1}{h}K(\frac{x_i - x}{h})$ and $\frac{1}{h}K(\frac{x_i - x}{h})$ is nearly uncorrelated as $h \to 0$*

But strict proof of consistency requires consideration of **mixing conditions**. The most common one is α-**mixing**:

$$\alpha(l) = max_{k \geq 1} sup_{A \subset \mathscr{F}_1^k, B \subset \mathscr{F}_{n+k}^{\infty}} \mid P(A)P(B) - P(A \cap B) \mid \to 0$$

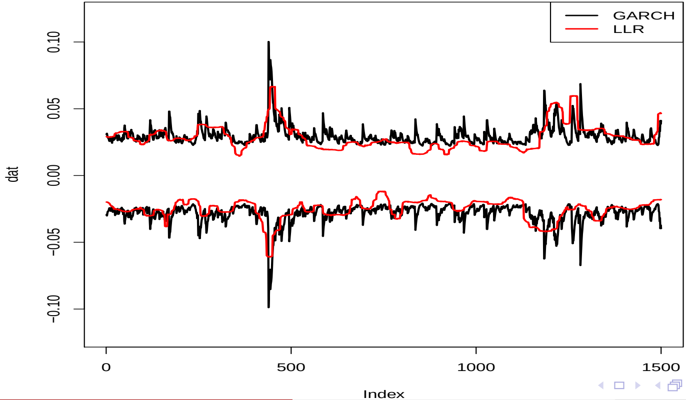
Nonparametric Quantile Regression: Example of Financial Data

Here we consider 1500 daily returns of Hewlett-Packard. The data is obtained from package **fEcofin**. The data shows classical features of nonlinear time series: *volatility clustering, assymetry, heavy tails*.



Nonparametric Quantile Regression V.S. GARCH Model

A GARCH model is fitted in comparison with a nonparametric local linear estimation. An important message is that the two tails show different volatile features. This information is missing in the GARCH model.



Bandwidth Selection: Adaptive Algorithm

1. Initialization: Set $s = 0$. For each X_i, we consider a small initial window $\Delta_0(X_i) = [X_i - h_0, X_i + h_0]$ for a small bandwidth h_0. Calculate $(\hat{a}_i^{(0)}(X_i), \hat{b}_i^{(0)}(X_i)) = argmin_{a,b} \sum_{X_j \in \Delta_0(X_i)} \rho_\tau(Y_j - a - b(X_j - X_i))K(\frac{X_j - X_i}{h_0})$ and estimate the variance $\hat{Var}(\hat{a}_i^{(0)}(X_i))$
2. Iteration: If $s > k_{max}$, with k_{max} the maximal iteration steps tolerable, then stop and set $m_\tau^*(X_i) = a_i^{(s-1)}(X_i)$.Else, denote $\Delta_s(X_i) = [X_i - h_s, X_i + h_s]$, with $h_s = h_{s-1} + h_0$ Calculate $\hat{a}_i^{(s)}(X_i)$ and update $\hat{Var}(\hat{a}_i^{(s)}(X_i))$ similar to above.

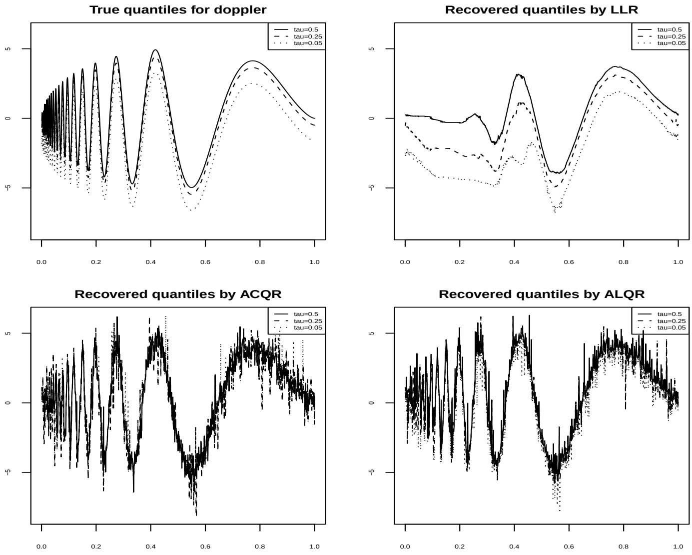
Adaptive Algorithm

1. Testing the homogeneity: If $\exists l < s$ s.t. $\mid \hat{a}_i^{(s)}(X_i) - \hat{a}_i^{(l)}(X_i) \mid > \eta\sqrt{\hat{Var}(\hat{a}_i^{(0)}(X_i))}$, stop and settle down the local linear estimator as $m_p^{\prime\prime *}(X_i) = a_i^{(s-1)}(X_i)$ Else set $s = s + 1$ and return to 2.

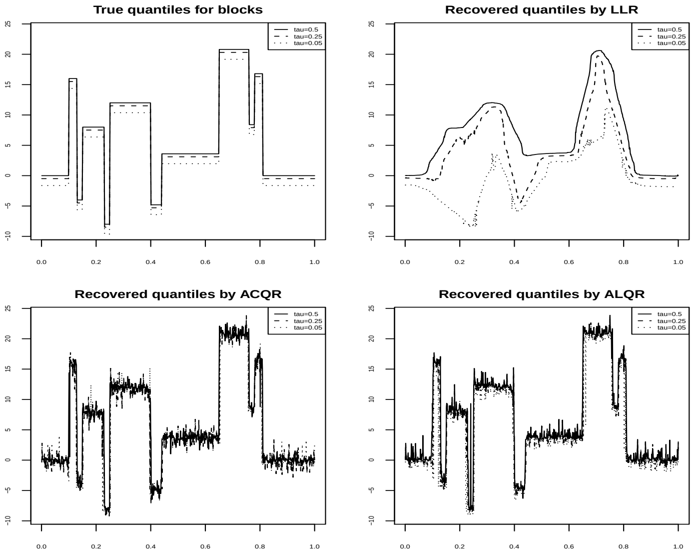
This algorithm was proposed by M. Tian(2009) and can overcome *curse of dimensionality* and notably, preserves *high frequency information*.

The consistency can also be proved under α-*mixing* conditions.

Adaptive Algorithm Simulation:Doppler



Adaptive Algorithm Simulation:Blocks



Confidence Interval Inference: An Open Problem

Sometimes, we are interested in the confidence interval inference of a perticular quantile, rather than the point estimation. Although we can rely on the asymptotia for large sample data, for small sample the variance estimator is very poor, and is still an open question.

One solution is to use the bootstrap method. Note the consistence of bootstrap requires that the resampled distribution F_n^* converges to the sample distribution F_n almost surely. Empirically, we can always assume it is true, but strict proof is difficult to obtain.

For one sample data, a better CI method has been proposed, thanks to **Kaplan-Meier** estimator.

One-sample Quantile CI

For one-sample data, using Kaplan-Meier estimator, the statistics becomes simple:

$$\frac{(\hat{S}^M(x) - \tau)^2}{var(S^M(x))} \sim \chi_1^2$$

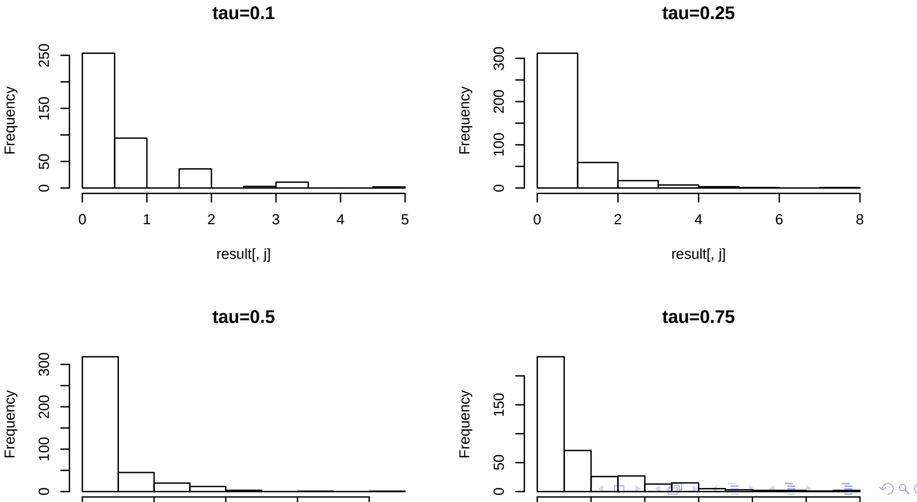
The efficiency is due to the **Greenwood Function** of the Kaplan-Meier estimator.

Intuitively, we may use

$$\frac{(\hat{x}_\tau - x_\tau)^2}{var(\hat{x}_\tau)} \sim \chi_1^2$$

In this case, the variance estimator is $\frac{1}{4\hat{f}(F^{-1}(\tau))}$, where the density function need be estimated using kernel. In small sample, this leads to very biased results.

The method adapts naturally to the censored data situation.



One-sample Estimation: Further Application

The method draws confidence intervals that are very robust, see the table below. indicating that it can possibly be adopted in related works, including ratio estimation, induced information cencoring problem, etc.

n=30			τ		
$1-\alpha$	0.1	0.25	0.5	0.75	0.9
95%	0.927	0.943	0.949	0.946	0.979
90%	0.860	0.885	0.882	0.897	0.947
85%	0.804	0.833	0.832	0.836	0.898
80%	0.710	0.782	0.771	0.788	0.817
75%	0.653	0.727	0.713	0.729	0.763

Conclusion

The presentation briefly discusses some robust analyis models, including nonparametric smoothing, quantile regression, and one-sample quantile estimation. In contradiction to a simple parametric model, these computationally intensive methods can avoid our arrogance in underestimating the real data structure. By applying such methods, we can let the data talk all by itself to reveal its own unique qualities.

For its application in R, there are already many built-in packages, including: **kde**, **ks**,(kernel smoothing) **quantreg**(quantile regression), **lowess**(Huber regression), **survival**(Kaplan-Meier estimator). Such packages offers quite fast algorithms and easy summarization. For some new algorithms shown in this result, the simulation often exceeds R's computation capacity. Therefore, I suggest mastering a low level language is still necessary for the research in this area.

Reference

- Fan, J. and Hu, T. and Truong, Y. K.(1994). Robust Non-parametric Function Estimation. *Scandinavian Journal of Statistics* **21** 433-446.
- Fan, J. and Yao, Q..(2003). Nonlinear Time Series: Nonparametric and Parametric Methods. *New York: Springer-Verlag*.
- Heiler, S.. A Survey on Nonparametric Time Series Analysis.
- Koenker, R.(2005). Quantile Regression. *Cambridge*
- Su, J. and Wei, L..(1993). Nonparametric Estimation for the Difference or Ratio of Median Failure Times. *Biometrics* **49** 603-607.
- Tian, M. (2009). Adaptive Quantile Regression. *Journal of the Royal Statistical Association*. **To appear**
- Zuo, C.(2009). Adaptive Estimation of Local Linear Quantiles. Junior Thesis, Renmin Univ. of China.