Estimating Conditional Value-at-Risk Using Kernel Conditional Quantile Estimation

Gulishana (Sarah) Adilijiang

University of Chicago

May 11, 2020

Contents

1.1 Conditional Value-at-Risk

Value-at-Risk (VaR)

By definition, the (1-p)-th VaR is the loss of a portfolio (measured in price or return) that is exceeded with a given probability p in a given time period (usually a day, a month or a year). Typically, $p = 5\%$ or 1%.

For example, if a daily 95% VaR of a stock return equals to 8%, then it means that there is a 5% chance that this stock will lose 8% or more on a given day.

Methods: historical simulation, Monte Carlo, variance-covariance approach.

Conditional Value-at-Risk (cVaR)

The cVaR is the estimation of VaR conditioning on a state process or the information variables, such as the prices (or returns) of portfolios.

Methods: parametric and nonparametric approaches.

1.2 Mathematical Models and Definitions

- **Model 1**: $Z_i = G(X_i, \varepsilon_i)$, where ε_i are i.i.d., $\varepsilon_i \perp X_j$ $(j \leq i-1)$, and (X_i, Z_i) is stationary.

  Example: the stochastic volatility (SV) model.

- Usually the sequence $\{X_t\}_{t\in\mathbb{Z}}$ has a dependence structure: $X_i = R(\eta_i, \eta_{i-1}, \cdots)$, where η_i are i.i.d., and R is a measurable function such as a linear or nonlinear function such that X_i is well defined.

- **Model 2**: $Z_i = G(Z_{i-1}, \varepsilon_i)$, where ε_i are i.i.d., and $\varepsilon_i \perp Z_{i-1}$. In this case, the sequence $\{Z_t\}_{t\in\mathbb{Z}}$ itself has a dependence structure. So we can set $X_i = Z_{i-1}$, then Model 2 will become a similar regression problem with Model 1.

  Examples: threshold autoregressive (TAR) model and autoregressive models with conditional heteroscedasticity (ARCH) model.

1.2 Mathematical Models and Definitions

- Distributions and densities:

  (a) cdf and pdf of X_t: $F(x) = \mathbb{P}(X_i \leq x),\ f(x) = F'(x)$

  (b) joint density $f(\xi, x) = f(\xi|x)f(x)$

  (c) conditional cdf and pdf of $Z_t|X_t$:
  $F(\xi|x) = \mathbb{P}(Z_i \leq \xi|X_i = x),\ f(\xi|x) = \frac{\partial F(\xi|x)}{\partial \xi}$

- Mathematical definition of cVaR:
  Let $\Omega^t = (..., X_{t-1}, X_t)$ be the information available up to time t.
  Then by the Markov property, the (1-p)-th cVaR, which is the p-th conditional quantile of $Z_t = X_{t+1}$ given $\Omega^t = (..., X_{t-1}, X_t)$, equals to:

$$\xi_p(x) = \inf_{\xi}\{\xi : F(\xi|x) = \mathbb{P}(Z_t \leq \xi|X_t = x) \geq p\} \quad (0 < p < 1)$$

Contents

2 Methodology and Theory

2.1 Kernel Conditional Quantile Estimation

- Kernels are widely used in nonparametric statistics for kernel density estimations and kernel regressions. A kernel meets the requirements:
  (a) $K(-u) = K(u)$ for all $u \in \mathbb{R}$
  (b) $\int uK(u)du = 0$ and $\int K(u)du = 1$
  (c) $\int u^2|K(u)|du < \infty$
- Most commonly used Gaussian kernel: $K(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$

2.1 Kernel Conditional Quantile Estimation

- Define $K_h(u) = K(u/h)$, where bandwidth $h > 0$ is a smoothing parameter.
- **Kernel Density Estimation**:

  Suppose $(X_1, X_2, \cdots, X_n)$ are the samples drawn from an unknown distribution with true density function $f(x)$. Then the kernel density estimate of $f(x)$ is:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x - X_i}{h}\right)$$

  where the bandwidth h satisfies that $h \to 0$ and $nh \to \infty$.

2.1 Kernel Conditional Quantile Estimation

- **General Kernel Regression**:

  Suppose $\{(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)\}$ are the samples drawn from an unknown joint distribution with true joint density function $f(x, y)$. Then the kernel density estimate of $f(x, y)$ is:

$$\hat{f}_{h_x,h_y}(x, y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_x}K_{h_x}(x - X_i)\frac{1}{h_y}K_{h_y}(y - Y_i)$$

  Thus the kernel regression estimate of $E(Y|X = x)$ is:

$$\begin{aligned}
\hat{E}(Y|X = x) &= \int y\hat{f}(y|x)dy = \int y\frac{\hat{f}(x, y)}{\hat{f}(x)}dy = \frac{\int y\hat{f}(x, y)dy}{\hat{f}(x)} \\
&= \frac{\int y\frac{1}{nh_xh_y}\sum_{i=1}^{n}K_{h_x}(x - X_i)K_{h_y}(y - Y_i)dy}{\frac{1}{nh_x}\sum_{i=1}^{n}K_{h_x}(x - X_i)} \\
&= \frac{\sum_{i=1}^{n}K_{h_x}(x - X_i)\int y\frac{1}{h_y}K_{h_y}(y - Y_i)dy}{\sum_{i=1}^{n}K_{h_x}(x - X_i)} \\
&= \frac{\sum_{i=1}^{n}K_{h_x}(x - X_i)\ Y_i}{\sum_{i=1}^{n}K_{h_x}(x - X_i)} \qquad \text{(Nadaraya-Watson estimator)}
\end{aligned}$$

- **Kernel Conditional Distribution Estimation**:
  Note that $F(\xi|x) = \mathbb{P}(Z \leq \xi|X = x) = E(\mathbf{1}_{Z\leq\xi}|X = x)$, so substituting the Y_i with $\mathbf{1}_{Z_i\leq\xi}$ in the kernel regression, we can get the Nadaraya-Watson estimator of the conditional distribution $F(\xi|x)$:

$$F_n(\xi|x) = \hat{E}(\mathbf{1}_{Z\leq\xi}|X = x) = \frac{\sum_{i=1}^{n} K_h(x - X_i)\ \mathbf{1}_{Z_i\leq\xi}}{\sum_{i=1}^{n} K_h(x - X_i)}$$
$$= \sum_{i=1}^{n} \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \mathbf{1}_{Z_i\leq\xi} = \sum_{i=1}^{n} w_i(x)\ \mathbf{1}_{Z_i\leq\xi}$$

  which is a weighted average of $\mathbf{1}_{Z_1\leq\xi}, \mathbf{1}_{Z_2\leq\xi}, \cdots, \mathbf{1}_{Z_n\leq\xi}$, and $w_i(x) = \frac{K_h(x-X_i)}{\sum_{i=1}^{n} K_h(x-X_i)}$ is the weight that satisfies $\sum_{i=1}^{n} w_i(x) = 1$.

2.1 Kernel Conditional Quantile Estimation

- **Kernel Conditional Quantile Estimation**:

  Our purpose is to get the kernel conditional quantile estimation of $\xi_p(x)$. We can estimate it by the sample conditional quantile:

$$\hat{\xi}_p(x) = \inf_{\xi}\{\xi : F_n(\xi|x) = \sum_{i=1}^{n} w_i(x)\, \mathbf{1}_{Z_i \leq \xi} \;\; \geq p\}$$

  By definition, $p = F(\xi_p(x)|x)$, and from equation of $F_n(\xi|x)$, we can get that $\hat{\xi}_p(x)$ satisfies:

$$\begin{aligned} |F_n(\hat{\xi}_p(x)|x) - F(\xi_p(x)|x)| &= |F_n(\hat{\xi}_p(x)|x) - p| \\ &= \frac{\sum_{i=1}^{n} K_h(x - X_i)\,(\mathbf{1}_{Z_i \leq \hat{\xi}_p(x)} - p)}{\sum_{i=1}^{n} K_h(x - X_i)} \\ &\leq \frac{\sup_u |K(u)|}{\sum_{i=1}^{n} K_h(x - X_i)} = \frac{K_0}{\sum_{i=1}^{n} K_h(x - X_i)} \end{aligned}$$

2.1 Kernel Conditional Quantile Estimation

- **Method to get the kernel conditional quantile estimate $\hat{\xi}_p(x)$:**

$$F_n(\hat{\xi}_p(x)|x) = \sum_{i=1}^{n} \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \mathbf{1}_{Z_i \leq \hat{\xi}_p(x)} = \sum_{i=1}^{n} w_i(x)\ \mathbf{1}_{Z_i \leq \hat{\xi}_p(x)} \ \rightarrow\ p$$

(1) Set up a grid values of $x \in [\min(X_i), \max(X_i)]$

(2) For each $X = x$, among all $\xi \in \{Z_1, Z_2, \cdots, Z_n\}$, find an optimal $\hat{\xi}_p(x)$ such that $F_n(\hat{\xi}_p(x)|x)$ is the closest to the given percentile p, i.e. $|F_n(\hat{\xi}_p(x)|x) - p|$ is the smallest.

Reason: suppose $Z^{(1)} \leq Z^{(2)} \leq \cdots \leq Z^{(n)}$ are the ordered observations, then for all the values of $\xi^* \in [Z^{(1)}, Z^{(2)})$, the vector $\mathbf{1}_{Z \leq \xi^*} = (1, 0, \cdots, 0)^T$ is the same for these ξ^* values, thus will get the same $F_n(\xi^*|x)$ value. And within these ξ^* values, the smallest value $Z^{(1)}$ would be chosen as a candidate due to the definition for $\hat{\xi}_p(x)$.

2.2 Theories and Properties of the Estimate

(*See Wu, et al. (2007) for detailed proof*)

- Two specific values of a kernel: $\kappa = \int K^2(u)du, \ \varrho = \frac{1}{2}\int u^2 K(u)du$
- **Bahadur Representation**:

  Asymptotic expansion for $\hat{\xi}_p(x) - \xi_p(x)$ is: $P_n(\xi) - Q_n(\xi)$, where $P_n(\xi) = \sum_{i=1}^{n} K_h(x - X_i)\left(\mathbf{1}_{Z_i \le \xi} - p\right)$, and $Q_n(\xi) = \mathbb{E}[P_n(\xi)]$.
- **Condition 1** (Short-Range Dependence):

  Let $f_\eta(x|\mathcal{F}_i)$ be the conditional density of X_{i+1} given $\mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i)$. Assume that $\sup_x f_\eta(x|\mathcal{F}_0) \le C_0$ almost surely for some $C_0 < \infty$, and that $\sum_{k=0}^{\infty} \omega_k < \infty$ where $\omega_k = \sup_{x\in\mathbb{R}} \|\mathcal{P}_0 f_\eta(x|\mathcal{F}_k)\|$ Here $\mathcal{P}_i$ is the projection operator: $\mathcal{P}_i \cdot = \mathbb{E}(\cdot|\mathcal{F}_i) - \mathbb{E}(\cdot|\mathcal{F}_{i-1})$, and $\|Z\| = [\mathbb{E}(|Z|^2)]^{1/2}$
- **Condition 2** (Regularity Conditions):
  1. (a) $f(x) > 0$ and $\sup_x f(x) < \infty$

     (b) The jth order derivatives $f^{(j)}(x), \ j = 1, \ldots, 4$, exist on $\mathbb{R}$ and are bounded.
  2. (a) $f(\xi, x) > 0$, where $F(\xi|x) = p$

     (b) All first order partial derivatives of $f(\cdot,\cdot)$ exist and are bounded.

2.2 Theories and Properties of the Estimate

(See Wu, et al. (2007) for detailed proof)

- **Theorem 1** (Short-Range-Dependent Processes):

  Let $r_n = (nh)^{-1/2} + h^2$, $\mu_n = r_n^2 + [(r_n \log n)/(nh)]^{1/2}$, $B(\xi, x) = F(\xi|x)f(x)$, $A(\xi, x) = \partial^2 B(\xi, x)/\partial x^2$, $C(\xi, x) = \varrho[pf''(x) - A(\xi, x)]/f(\xi, x)$. Assume that Conditions 1 and 2 hold and that $r_n(\log n)^2 \to 0$. Then

$$\hat{\xi}_p(x) - \xi_p(x) = \frac{Q_n(\xi_p(x)) - P_n(\xi_p(x))}{nhf(\xi_p(x), x)} + h^2 C(\xi_p(x), x) + O_{\mathbb{P}}(\mu_n)$$

- **Corollary 1** (Central Limit Theorem):

  Assume $nh^9 \to 0$ and $h \to 0$. Under conditions of Theorem 1, we have

$$\sqrt{nh}\ \left[\hat{\xi}_p(x) - \xi_p(x) - h^2 C(\xi_p(x), x)\right] \ \Rightarrow\ N\left(0,\ \frac{\kappa p(1-p)f(x)}{f^2(\xi_p(x), x)}\right)$$

  which can be applied to construct confidence intervals for $\xi_p(x)$.

- **Theorem 2** (Long-Range-Dependent Linear Processes):

  It shows that, for long-range-dependent linear processes, the Theorem 1 holds with $O_{\mathbb{P}}(\mu_n')$.

2.3 Bandwidth Selection

- By Corollary 1 (CLT), we can get the asymptotic MSE (mean squared error) of the kernel conditional quantile estimator $\hat{\xi}_p(x)$:

$$MSE = E\left(\hat{\xi}_p(x) - \xi_p(x)\right)^2 \approx h^4 C^2(\xi_p(x), x) + \frac{\kappa p(1-p) f(x)}{nhf^2(\xi_p(x), x)}$$

- By setting $\frac{\partial MSE}{\partial h} = 0$, we can get the optimal h_p that minimizes the MSE for a given percentile p:

$$h_p^5 = \frac{\kappa p(1-p) f(x)}{4nf^2(\xi_p(x), x)\ C^2(\xi_p(x), x)}$$

- Assume that the leading bias term of $C(\xi_p(x), x)$ is dominated by $-\xi_p''(x)$. Then since $\xi_{p_1}''(x)^2 \approx \xi_{p_2}''(x)^2$ and $f(\xi_p(x)|x) \approx \sigma_x^{-1}\phi(\Phi^{-1}(p))$, we can get the ratio between two bandwidths for different values of p:

$$\frac{h_{p_1}^5}{h_{p_2}^5} = \frac{p_1(1-p_1)\ \xi_{p_2}''(x)^2\ f^2(\xi_{p_2}(x)|x)}{p_2(1-p_2)\ \xi_{p_1}''(x)^2\ f^2(\xi_{p_1}(x)|x)} \approx \frac{p_1(1-p_1)\ \phi(\Phi^{-1}(p_2))^2}{p_2(1-p_2)\ \phi(\Phi^{-1}(p_1))^2}$$

2.3 Bandwidth Selection

- Setting $p_2 = 0.5$, we get:

$$h_p^5 = \frac{p(1-p)\ \phi(\Phi^{-1}(0.5))^2}{0.5^2\ \phi(\Phi^{-1}(p))^2} h_{0.5}^5 = \frac{2p(1-p)}{\pi\phi(\Phi^{-1}(p))^2} h_{0.5}^5$$

  Note that $h_p > h_{0.5}$ for all $p \neq 0.5$, which means estimating the conditional quantile for other $p's$ needs more smoothness than for the median.

- Usually median $\approx$ mean, so $h_{0.5}$ can be approximated by the opitmal bandwidth for the general kernel (mean) regression.

  In this article, we use the direct plug-in method to select the bandwidth, which is implemented by dpill() function in R package KernSmooth.

2.4 Bias Correction

- Biases:
  - Tail estimation bias of extreme quantiles (eg. 1% and 5%)
  - Bias that arises in kernel smoothing and finite sample fitting

- By Corollary 1 (CLT), we have

$$E\left[\hat{\xi}_p(x) - \xi_p(x)\right] \approx h^2 C(\xi_p(x), x)$$

So there is an unpleasant bias term $h^2 C(\xi_p(x), x)$.

- Introduce a jack-knife-type bias correction scheme:

$$\tilde{\xi}_p(x) = 2 \times \hat{\xi}_{p,h}(x) - \hat{\xi}_{p,\sqrt{2}h}(x)$$

Because $2 \times h^2 C(\xi_p(x), x) - (\sqrt{2}h)^2 C(\xi_p(x), x) = 0$.

2.5 Second Smoothing

- Recall that: we search for optimal $\hat{\xi}_p(x)$ within $\xi \in \{Z_1, Z_2, \cdots, Z_n\}$. Due to the finite sample sizes and the discrete values of the samples, we will get an piecewise-constant-function like estimation line. It need second smoothing.

- **Nonparametric Kernel Smoothing**:
  (1) **Kernel Regression Smoothing**: i.e. the Nadaraya-Watson kernel regression estimate:

$$\hat{g}(x) = \sum_{i=1}^{n} \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} Y_i = \sum_{i=1}^{n} w_i(x) Y_i$$

  (2) **Local Polynomial Smoothing**: using weighted least squares with weights $w_i(x) = K_h(x - X_i)$. For each x, find $\hat{\beta}_0(x)$ and $\hat{\beta}_1(x)$ that minimize the weighted residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} K_h(x - X_i)(Y_i - \beta_0 - \beta_1 X_i)^2 = \sum_{i=1}^{n} w_i(x)(Y_i - \beta_0 - \beta_1 X_i)^2$$

  and get the estimate $\hat{g}(x) = \hat{\beta}_0(x) + \hat{\beta}_1(x) \cdot x$ for x. So it needs to do this for each x, which takes longer time than kernel regression.

2.6 Algorithm Summary

- **Summary of the Algorithm for Estimating $\xi_p(x)$:**

  (1) **Kernel**: select a symmetric kernel function K (we use Gaussian kernel).

  (2) **Bandwidth Selection**: select optimal bandwidth $h_{0.5}$ by kernel (mean) regression estimation method.

  (3) **Bandwidth Calculation**: calculate the optimal bandwidth h_p for a given percentile p.

  (4) **Kernel Conditional Quantile Estimation** to get $\hat{\xi}_{p,h_p}(x)$:
  - Set up a grid values of $x \in [\min(X_i), \max(X_i)]$.
  - For each $X = x$, among all $\xi \in \{Z_1, Z_2, \cdots, Z_n\}$, find an optimal $\hat{\xi}_p(x)$ such that $|F_n(\hat{\xi}_p(x)|x) - p|$ is the smallest.

  (5) **Bias Correction**: Repeat the estimation one more time with bandwidth $\sqrt{2}h_p$ instead of h_p to get $\hat{\xi}_{p,\sqrt{2}h_p}(x)$, and get the jack-knife-type bias corrected $\tilde{\xi}_p(x)$

  (6) **Second Smoothing**: apply the kernel regression smoothing or the local polynomial smoothing method to get the final smoothed estimator $\xi_p^*(x)$

Contents

3 Simulation Study

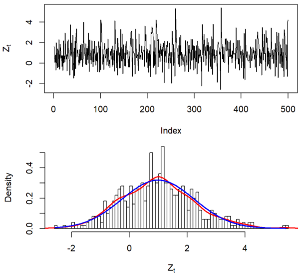
3.1 TAR Model

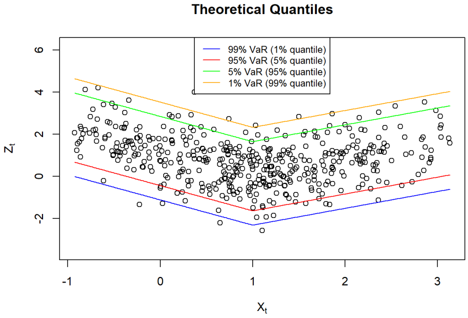
- The TAR (threshold autoregressive) model:

$$Z_t = \left(a \cdot \mathbf{1}_{Z_{t-1} \geq \alpha} + b \cdot \mathbf{1}_{Z_{t-1} < \alpha}\right) Z_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z}$$

  where $\varepsilon_t \sim N(0,1)$, $a = 0.8$, $b = 1.2$, and $\alpha = 1$, whose selections provide us a relatively sharp change of $E(Z_t | Z_{t-1})$ at the threshold value α.

- Set $X_t = Z_{t-1}$, then $Z_t | X_t = x \ \sim \ N(ax\mathbf{1}_{x \geq \alpha} + bx\mathbf{1}_{x < \alpha},\ 1)$
- Theoretical quantile: $\xi_p(x) = ax \cdot \mathbf{1}_{x \geq \alpha} + bx \cdot \mathbf{1}_{x < \alpha} + \Phi^{-1}(p)$

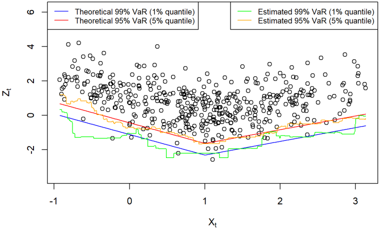




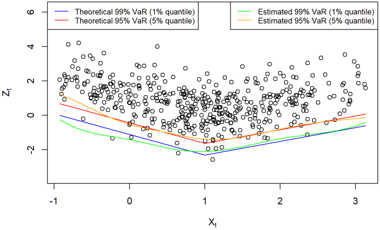
3.1 TAR Model

- Kernel Conditional Quantile Estimates:

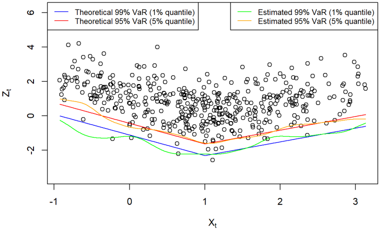
Original Quantile Estimation



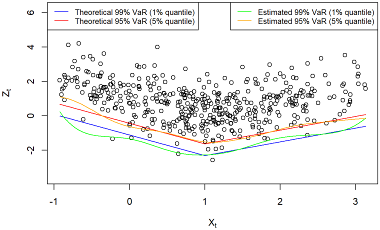
Smoothed Quantile Estimation (locpoly, degree=1)



Smoothed Quantile Estimation (ksmooth)



Smoothed Quantile Estimation (locpoly, degree=2)



3.1 TAR Model

- **Performance Evaluations**:
  - Discrete Mean Absolute Error (DMAE):

$$DMAE(p) = \frac{1}{N}\sum_{i=1}^{N}\left|\tilde{\xi}_p(x_i) - \xi_p(x_i)\right|$$

    where $N = 1000$ is the number of grid values of $x \in [\min(X_i), \max(X_i)]$.

  - Average DMAE(p):

$$aDMAE(p) = \frac{1}{S}\sum_{j=1}^{S} DMAE(p)_j$$

    where $S = 100$ is the number of simulations.

3.1 TAR Model

Table: aDMAE(p) for TAR Model

		ksmooth	locpoly(d=1)	locpoly(d=2)
n=250	p=0.01	0.3839	0.3367	0.3770
n=250	p=0.05	0.2839	0.2444	0.2775
n=500	p=0.01	0.3314	0.2825	0.3218
n=500	p=0.05	0.2172	0.1873	0.2114
n=1000	p=0.01	0.2517	0.2124	0.2439
n=1000	p=0.05	0.1652	0.1420	0.1611

Table: Range of Z_t and $\xi_p(x)$ of TAR Model

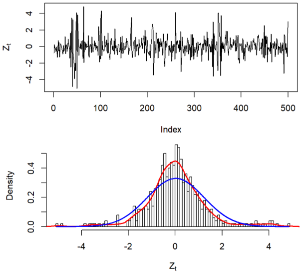
		range of $\{Z_t\}_{t=1}^n$	range of $\{\xi_p(x_i)\}_{i=1}^N$
n=250	p=0.01	[-3.2757, 5.5025]	[-2.3263, 0.4833]
n=250	p=0.05	[-3.2757, 5.5025]	[-1.6448, 1.1648]
n=500	p=0.01	[-3.5225, 5.6298]	[-2.3263, 0.3372]
n=500	p=0.05	[-3.5225, 5.6298]	[-1.6448, 1.0187]
n=1000	p=0.01	[-4.0900, 6.1246]	[-2.3263, 0.2689]
n=1000	p=0.05	[-4.0900, 6.1246]	[-1.6449, 0.9504]

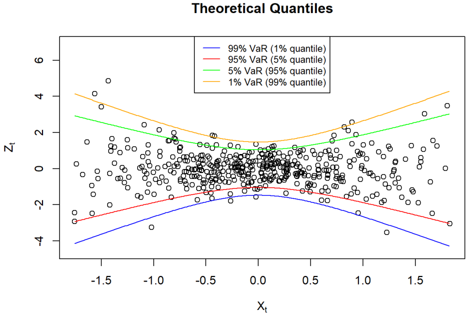
- The ARCH (autoregressive model with conditional heteroscedasticity) model:

$$Z_t = \varepsilon_t \sqrt{\omega + \alpha \; Z_{t-1}^2}, \qquad t \in \mathbb{Z}$$

where $\varepsilon_t \sim N(0,1)$, $\omega = 0.4$, and $\alpha = 0.9$, whose selections provide a distribution with heavier tails than a normal distribution.

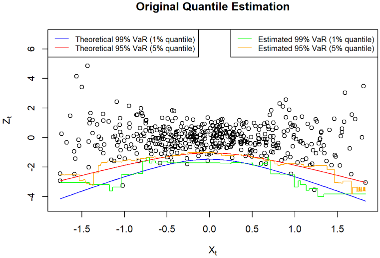
- Set $X_t = Z_{t-1}$, then $Z_t | X_t = x \;\sim\; N(0, \omega + \alpha x^2)$
- Theoretical quantile: $\xi_p(x) = \Phi^{-1}(p)\; \sqrt{\omega + \alpha x^2}$

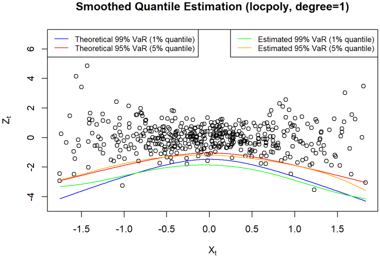


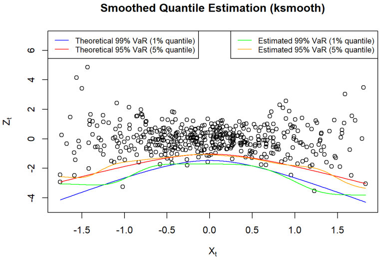


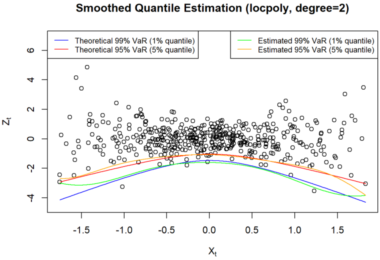
3.2 ARCH Model

- Kernel Conditional Quantile Estimates:









3.2 ARCH Model

Table: aDMAE(p) for ARCH Model

		ksmooth	locpoly(d=1)	locpoly(d=2)
n=250	p=0.01	0.6185	0.6202	0.6032
n=250	p=0.05	0.4270	0.4199	0.4188
n=500	p=0.01	0.4984	0.4964	0.4789
n=500	p=0.05	0.3388	0.3320	0.3344
n=1000	p=0.01	0.4007	0.4216	0.3849
n=1000	p=0.05	0.2415	0.2420	0.2359

Table: Range of Z_t and $\xi_p(x)$ of ARCH Model

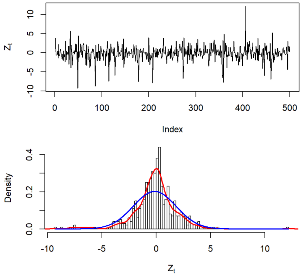
		range of $\{Z_t\}_{t=1}^n$	range of $\{\xi_p(x_i)\}_{i=1}^N$
n=250	p=0.01	[-7.0104, 6.6241]	[-8.6332, -1.4713]
n=250	p=0.05	[-7.0104, 6.6241]	[-6.1041, -1.0403]
n=500	p=0.01	[-8.2047, 7.5443]	[-9.3979, -1.4713]
n=500	p=0.05	[-8.2047, 7.5443]	[-6.6449, -1.0403]
n=1000	p=0.01	[-7.7339, 6.5270]	[-6.3188, -1.4713]
n=1000	p=0.05	[-7.7339, 6.5270]	[-4.4677, -1.0403]

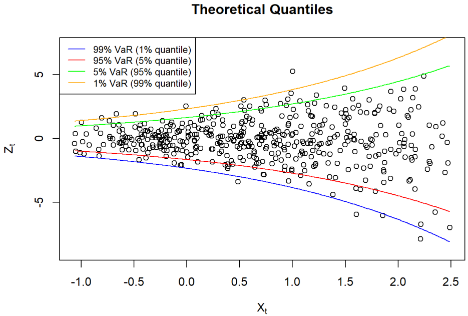
3.3 SV Model

- The SV (stochastic volatility) model:

$$Z_t = \varepsilon_t \exp(0.5X_t), \quad X_t = a + b\, X_{t-1} + c\, \eta_t, \qquad t \in \mathbb{Z}$$

  where $\varepsilon_t \sim N(0,1)$, $\eta_t \sim N(0,1)$, $a = 0.2$, $b = 0.6$, and $c = 0.9$, whose selections provide a distribution with heavier tails than a normal distribution.
- $Z_t | X_t = x \;\sim\; N(0, \exp(x))$
- Theoretical quantile: $\xi_p(x) = \Phi^{-1}(p)\, \exp(0.5x)$





3.3 SV Model

- Kernel Conditional Quantile Estimates:

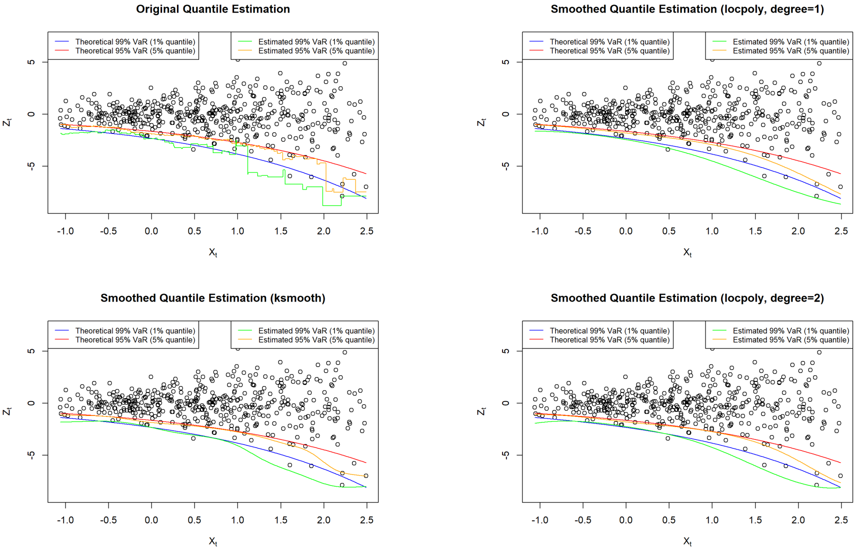


Table: aDMAE(p) for SV Model

		ksmooth	**locpoly(d=1)**	**locpoly(d=2)**
n=250	p=0.01	0.6567	0.5928	0.6314
n=250	p=0.05	0.4175	0.3833	0.4103
n=500	p=0.01	0.5360	0.4893	0.5136
n=500	p=0.05	0.3153	0.2835	0.3064
n=1000	p=0.01	0.3939	0.3751	0.3764
n=1000	p=0.05	0.2173	0.1980	0.2100

Table: Range of Z_t and $\xi_p(x)$ of SV Model

		range of $\{Z_t\}_{t=1}^n$	**range of** $\{\xi_p(x_i)\}_{i=1}^N$
n=250	p=0.01	[-11.5287, 10.0761]	[-9.0471, -0.8967]
n=250	p=0.05	[-11.5287, 10.0761]	[-6.3968, -0.6340]
n=500	p=0.01	[-9.6945, 11.2705]	[-9.4966, -1.0031]
n=500	p=0.05	[-9.6945, 11.2705]	[-6.7146, -0.7093]
n=1000	p=0.01	[-10.9458, 11.9877]	[-8.6332, -1.0262]
n=1000	p=0.05	[-10.9458, 11.9877]	[-6.1041, -0.7256]

3.4 Summary

- The estimation is reasonably well with relatively low $aDMAE(p)$.
- The performances of the estimates:
  - TAR Model: locpoly(d=1) $>$ locpoly(d=2) $>$ ksmooth
  - ARCH Model: locpoly(d=2) $>$ locpoly(d=1) $\approx$ ksmooth
  - SV Model: locpoly(d=1) $>$ locpoly(d=2) $>$ ksmooth
- In general, the local polynomial smoothing does better job than the kernel regression smoothing here, since kernel regression smoothing has relatively worse behaviour at the edges of the X range.
- Choice of degree of local polynomial smoothing:
  - Depends on the linearity and curvature of the true quantile line.
  - The choice can be made (1) by knowledge of the portfolio, (2) by the shape the scatterplot of (Z_i, X_i), (3) by measuring the heteroscedasticity of (Z_i, X_i).
  - If cannot decide, use the linear local polynomial.

Contents

4 Real Data Analysis

- Daily and Monthly Stock Data

4 Real Data Analysis

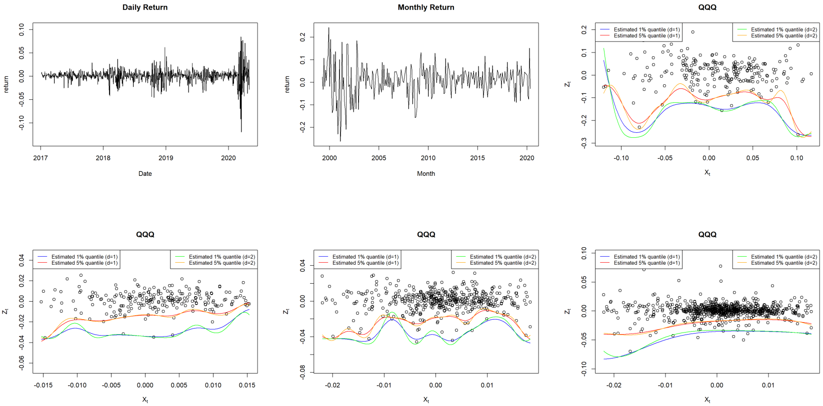
Table: Stock/Index

type	stock name	company/index	comment
index	SPY	S&P 500 Index	500 large-cap U.S. companies
index	QQQ	Invesco QQQ Trust	the Nasdaq 100 Index
tech	GOOG	Alphabet Inc.	
tech	AAPL	Apple Inc.	
commodity	MCD	McDonald's Corporation	
commodity	TGT	Target Corporation	

- Both the daily and monthly stock return data are used, which are downloaded from Yahoo! Finance.
- Daily data are from 2017-01-01 to 2020-05-01. Estimations of conditional quantiles for $p = 1\%$ and 5% have been implemented on the 1-year(2019), 2-year(2018-2019), over 3-year(2017 to now) data, respectively.
- Monthly data are from the first recorded date on Yahoo! Finance.
- Local polynomial smoothing is used for the second smoothing using both linear and quadratic regression.

4 Real Data Analysis

- QQQ

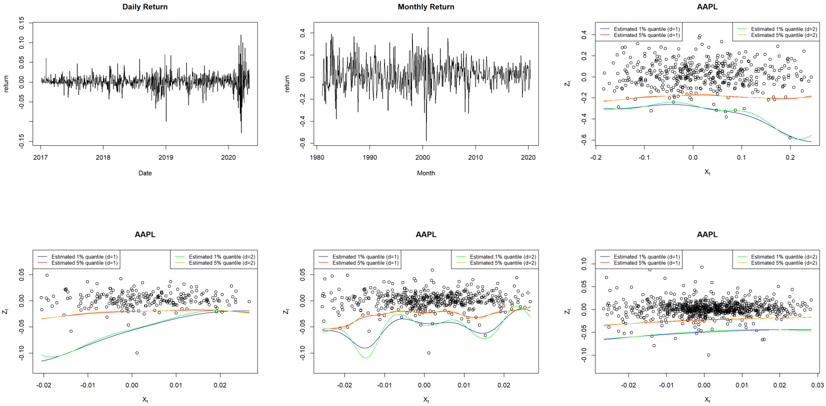


Topright figure is the monthly data estimation, which starts from Mar 1999 (n=256).

Bottom three figures are for daily data estimations using 1-year(2019), 2-year(2018-2019), over 3-year(2017 to now) data.

4 Real Data Analysis

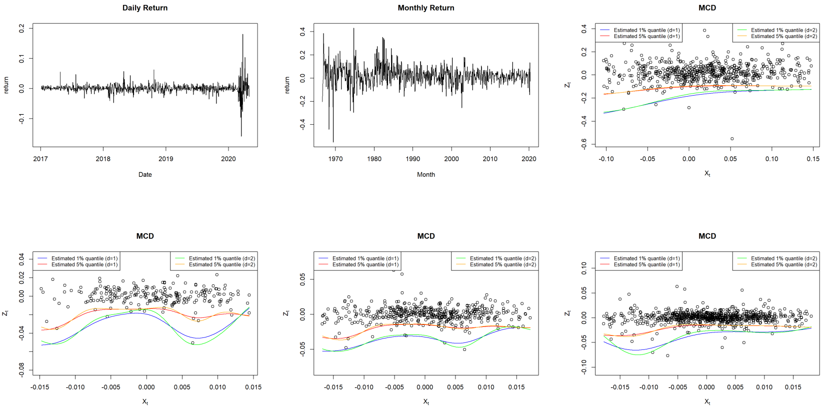
- AAPL



Topright figure is the monthly data estimation, which starts from Dec 1980 (n=475).

Bottom three figures are for daily data estimations using 1-year(2019), 2-year(2018-2019), over 3-year(2017 to now) data.

- MCD



Topright figure is the monthly data estimation, which starts from July 1966 (n=648).

Bottom three figures are for daily data estimations using 1-year(2019), 2-year(2018-2019), over 3-year(2017 to now) data.

5 Conclusions

- In this article, we mainly worked on the algorithm of kernel conditional quantile estimation, which is a good method for estimating conditional VaR. We studied the properties of the estimate and derived the optimal bandwidth for a given percentile. Then, we used bias correction and second smoothing to further improve the estimation.
- Later, we did simulation on three sets of weakly dependent data that are widely used in financial data analysis. The results show that our estimation algorithm does a reasonably well estimation of the extreme 1% and 5% quantiles. And local polynomial smoothing works generally better than the kernel regression smoothing as a second smoothing method.
- In the end, we performed real data analysis on the daily and monthly stock return data that are downloaded from Yahoo! Finance. We can see that without making any parametric model assumptions, the estimation of conditional VaR through our kernel conditional quantile estimation method works reasonably well for complex time series. This is a promising feature in situations where the underlying mechanisms are unknown.

Thank You

Questions