# Lec 18: VCM and Nonparametric Quantile regression

Weiping Zhang

December 16, 2020

# Functional Coefficient Models

- In the seminal work of Hastie and Tibshirani (1993), they propose a class of varying-coefficient models that admit the form

$$Y = a_0 + \sum_{j=1}^{d} a_j(U_j) X_j + \epsilon_j \tag{1}$$

where $\epsilon$ is a zero mean disturbance term, $X_j$, $j = 1, \ldots, d$, are regressors but whose coefficients depend on another random variable or vector $U_j$. Model (1) implies that the random variables $U_j$, $j = 1, \ldots, d$ change the effects of $X_j$ on $Y$ and hence we can call $U_j' s$ as modifying variables. The dependence of $a_j(\cdot)$ on $U_j$ implies a special kind of interaction between $U_j$ and $X_j$. In some cases, the variables $U_j$ are indistinguishable from the variable $X_j$, whereas in other cases $U_j$ might be a special variable such as "time".

- As Cai, Fan and Yao (2000) remark, the idea for the varying-coefficient models are not new, but the potential of this modeling techniques had not been fully explored until the seminal work of Cleveland et al. (1992), Chen and Tsay (1993), and Hastie and Tibshirani (1993), in which nonparametric techniques were proposed to estimate the unknown functional coefficients.
- **Example** (Hastie and Tibshirani, 1993) (a) If $a_j(U_j) = a_j$ (the constant function), then the corresponding term is linear in $X_j$. If all the terms in (1) are linear, then model (1) is the usual linear model.
- (b) If $X_j = c$ (say $c = 1$), then the $j$th term is simply $a_j(U_j)$, an unspecified function in $U_j$. If all terms have this form or are linear as in (a), then model (1) has the form of an *additive model*.

- **Example** (Hastie and Tibshirani, 1993) (c) A linear function $a_j(U_j) = a_j U_j$ leads to a product interaction of the form $a_j U_j X_j$.
- (d) Often $U_j$ will be a factor such as time or age that may change the effects of $X_1, \ldots, X_d$ on $Y$. For example, when $U_j = t_j$ (time trend), the model can be written as

$$Y_t = a_0(t) + \sum_{j=1}^{d} a_j(t) X_{j,t} + \epsilon_t, t = 1, 2, \ldots \tag{2}$$

  where we allow the intercept $a_0$ to change over time too.
- (e) Suppose that the modifying variable $U_j$ is same as $X_j$, then the model can be written as

$$Y = a_0 + \sum_{j=1}^{d} a_j(X_j) X_j + \epsilon$$

  in which the coefficient $a_j(X_j)$ will not signify the marginal effect of $X_j$ on $Y$.

- **Example** Chen and Tsay (1993) introduce a functional-coefficient autoregressive model, a special case of the functional coefficient model, which admits the form

$$X_t = a_1(X_{t-p})X_{t-1} + \cdots + a_d(X_{t-p})X_{t-d} + \sigma(X_{t-p})\epsilon_t, \quad (3)$$

where $\{\epsilon_t\}$ is a sequence of IID random variables with zero mean and unit variance, and $\epsilon_t$ is independent of $X_{t-1}, X_{t-2}, \ldots$. As in model (1), the coefficient functions $a_1(\cdot), \ldots, a_p(\cdot)$ are unknown. The model is a special case of the state-dependent model of Priestley (1981), and the variable $X_{t-p}$ is referred to as the model-dependent variable by Fan and Yao (2003). The model (4) can be denoted as $FAR(d, p)$.

- The state-dependent model is a natural extension of the threshold autoregressive model (TAR) in the nonlinear time series literature. It allows the coefficient functions to change gradually, rather than abruptly as in the TAR model. The FAR model also includes the generalized exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981) and Ozaki (1982):

$$X_t = \sum_{i=1}^{d} \left[\beta_{1i} + (\beta_{3i} X_{t-p}) exp(-\beta_{4i} X_{t-p}^2)\right] X_{t-i} + \epsilon_t, \quad (4)$$

where $\beta_{4i} \geq 0, i = 1, \ldots, d.$

- As Hastie and Tibshirani (1993) remark, model (1) is too general for most applications since no restrictions are imposed on the coefficient functions and unrestricted nonparametric estimation of these functions would be not possible except for some special cases. In this section we restrict ourselves to consider a special form:

$$Y = \sum_{j=1}^{d} a_j(U) X_j + \epsilon \tag{5}$$

where $U$ is a $p \times 1$ random vector.

# Estimation of Coefficient Functions

- Let $Z = (X_1, \ldots, X_d)'$ and $A(U) = (a_1(U), \ldots, a_d(U))'$. Then model (5) can be written as

$$Y = Z'A(U) + \epsilon \tag{6}$$

We wish to choose $A(\cdot)$ to minimize

$$E[Y - Z'A(U)]^2$$

A sufficient requirement for the solution is that it minimizes $E\{[Y - Z'A(U)]^2 | U = u\}$ for every $u$.

- Fortunately, the latter problem has a closed-form solution

$$A^*(u) = \{E[ZZ'|u]\}^{-1} E[ZY|u], \tag{7}$$

from the linear regression of $Y$ on $Z$ for each value $u$. Given data, we can estimate $A^*(u)$ by using a smoother to estimate each of the conditional expectations in (7). Equivalently, (7) suggests estimation of $A(u)$ by fitting a hyperplane to $Y$ as a function of $Z$ in the neighborhood of each $u$ value. This is an extension of the local linear fit of the previous section.

- We now estimate the unknown coefficient functions in model (5) by using a local linear regression technique. For any given $u$ and $\tilde{u}$ in a neighborhood of $u$, it follows from a first order Taylor expansion that

$$a_j(\tilde{u}) \approx a_j(u) + \nabla a_j(u)'(\tilde{u} - u) = a_j + b_j'(\tilde{u} - u),$$

where $a_j$ and $b_j$ are the local intercept and slope corresponding to $a_j(u)$ and $\nabla a_j(u)$.

- To estimate $a_j(u)$ and $\nabla a_j(u)$, we choose $\{a_j\}$ and $\{b_j\}$ to minimize

$$\sum_{i=1}^{n} \left[ Y_i - \sum_{j=1}^{d} \{a_j + b_j'(U_i - u)\} X_{ij} \right]^2 \mathcal{K}_h(U_i - u), \quad (8)$$

where $\mathcal{K}_h(u) = \prod_{s=1}^{p} h_s^{-1} K(u_s/h_s)$, $K$ is a univariate kernel function, and $h = (h_1, \ldots, h_p)$ is the bandwidth. Let $\{(\hat{a}_j, \hat{b}_j)\}$ be the local linear estimator. Then the local linear regression estimator for the functional coefficient is given by

$$\hat{a}_j(u) = \hat{a}_j, j = 1, \ldots, d. \quad (9)$$

- The local linear regression estimator for the functional coefficient can be easily obtained. To do so, let $e_{j,d(p+1)}$ be the $d(p+1) \times 1$ unit vector with 1 at the $j$th position and 0 elsewhere. Let $\tilde{X}$ denote an $n \times d(p+1)$ matrix with

$$\tilde{X}_i = (X_{i1}, \ldots, X_{id}, X_{i1}(U_i - u)', \ldots, X_{id}(U_i - u)')$$

as its $i$th row. Let $Y = (Y_1, \ldots, Y_n)'$. Set $W = diag\{\mathcal{K}_h(U_1 - u), \ldots, \mathcal{K}_h(U_n - u)\}$. Then the local regression problem (8) can be written as

$$(Y - \tilde{X}\theta)'W(Y - \tilde{X}\theta), \tag{10}$$

where $\theta = (a_1, \ldots, a_d, b_1', \ldots, b_d')$. So the local esitmator is simply

$$\hat{\theta} = (\tilde{X}'W\tilde{X})^{-1}\tilde{X}'WY, \tag{11}$$

which entails that

$$\hat{a}_j(u) = \hat{a}_j = e_{j,d(p+1)}'\hat{\theta}, j = 1, \ldots, d. \tag{12}$$

# Asymptotic Properties for the Local Linear Estimator

**Theorem**

*Under some regularity conditions, we have*

$$\sqrt{nh_1 \cdots h_p} H_1(\hat{\theta} - \theta) - S^{-1} b(h) \rightsquigarrow N(0, S^{-1} \Gamma S^{-1})$$

*where* $H_1 = diag\{1, \ldots, 1, h', \ldots, h'\}$ *is a* $d(p+1) \times d(p+1)$ *diagonal matrix with* $d$ *diagonal elements 1's and* $d$ *diagonal elements* $h's$. *In particular, for* $j = 1, \ldots, d$,

$$\sqrt{nh_1 \cdots h_d} \Big( \hat{a}_j - a_j(u) - \sum_{s=1}^{p} h_s^2 B_{j,s}(u) \Big) \rightsquigarrow N(0, \Sigma^*)$$

# Bandwidth Selection

- The above theorem implies that the leading term for the mean squared error (MSE) of $\hat{a}_0$ is

$$MSE(\hat{a}_0) = \Big[\sum_{s=1}^{p} h_s^2 B_{1s}(u)\Big]^2 + \frac{C_2}{nh_1 \cdots h_p}$$

By symmetry, all $h_s$ should have the same order. It is easy to obtain the optimal rate of bandwidth in terms of minimizing a weighted integrated version of $MSE(\hat{a}_0)$:

$$h_s \sim n^{-1/(4+p)}$$

Nevertheless, the exact formula for the optimal smoothing parameters is difficult to obtain except for the simplest cases (e.g.$p \leq 2$). Cai, Fan and Yao (2000) studied the case $p = 1$.

# Nonparametric Quantile Estimation

- Since the seminal work of Koenker and Bassett (1978) there has developed a large literature on (conditional) quantile estimation. There are a variety of approaches to estimating conditional quantiles. These can be divided into three categories according to whether a parametric assumption is made: fully parametric, semiparametric, and purely nonparametric.
- For a recent account for the parametric approach, see Kim and White (2003) and Komunjer (2003). The second approach includes Koenker and Zhao (1996), Engle and Manganelli (1999), and Lee (2003), whereas the third approach includes Chaudhuri (1991), Fan et al. (1994), Yu and Jones (1998), Cai (2002), and Hansen (2004a), among many others. Here we focus on the nonparametric estimation of conditional quantile functions

# The Local Linear Nonparametric Quantile Estimator

- Let $\rho_\tau(z) = z[\tau - 1(z \leq 0)]$ be the check function, it is well known that the $\tau-$th conditional quantile $q_\tau(x)$ of $Y_t$ given $X_t = x$ satisfies

$$q_\tau(x) = \arg\min_q E[\rho_\tau(Y_t - q(X_t)|X_t = x)], \qquad (13)$$

where we assume that the solution to the above minimization problem is unique (which is true if the conditional CDF $F(\cdot|x)$ of $Y$ given $X = x$ is strictly monotone) and $q$ belongs to a space of measurable functions defined on $\mathbb{R}^p$. In the parametric setup, it is frequently assumed that $q(x) = x'\beta$ where $\beta$ is a $p \times 1$ vector of parameters, and $x = (x_1, \ldots, x_p)' \in \mathbb{R}^p$.

- Denote by $\dot{q}_\tau(x) = (\partial q_\tau(x)/\partial x_1, \ldots, \partial q_\tau(x)/\partial x_p)'$ the first order derivative of $q_\tau(x)$ at $x = (x_1, \ldots, x_p)' \in \mathbb{R}^p$. The idea of the local linear fit is to approximate the unknown $\tau$-th quantile $q_\tau(\cdot)$ by a linear function

$$q_\tau(z) \approx q_\tau(x) + \dot{q}_\tau(x)'(z - x) = \beta_0 + \beta_1'(z - x)$$

for $z$ in a neighborhood of $x$. Locally, estimating $q_\tau(x)$ is equivalent to estimating $\beta_0$ and estimating $\dot{q}_\tau(x)$ is equivalent to estimating $\beta_1$. This motivates Yu and Jones (1998) to define a local linear quantile regression (LLQR) estimator of $q_\tau(x)$ and its derivative by $\hat{q}_\tau(x) = \hat{\beta}$ and $\hat{\dot{q}}(x) = \hat{\beta}_1$, respectively, where

$$\{\hat{\beta}_0, \hat{\beta}_1\} = \arg \min_{(\beta_0, \beta_1)} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \beta_0 - \beta_1'(X_i - x)) \mathcal{K}_h(X_i - x), \tag{14}$$

- Since the objective function is highly nonlinear in the parameter $(\beta_0, \beta_1)'$. There is no closed form solution to the above minimization problem. We need to resort to some numerical optimization routine to find the estimators. Fortunately, we show now that any routine for parametric quantile regression can be applied to our nonparametric framework. To see this, write

$$\begin{aligned}
&\rho_\tau(Y_i - \beta_0 - \beta_1'(X_i - x))\mathcal{K}_h(X_i - x) \\
&\quad = \mathcal{K}_h(X_i - x)\{Y_i - \beta_0 - \beta_1'(X_i - x)\} \\
&\qquad \cdot \{(\tau - 1(Y_i - \beta_0 - \beta_1'(X_i - x) \le 0)\} \\
&\quad = (\tilde{Y}_i - \beta'\tilde{X})[\tau - 1(\tilde{Y}_i - \beta'\tilde{X} \le 0)] \\
&\quad = \rho_\tau(\tilde{Y}_i - \beta'\tilde{X}_i)
\end{aligned}$$

where $\tilde{Y}_i = Y_i\mathcal{K}_h(X_i - x)$, $\beta' = (\beta_0, \beta_1')$, $\tilde{X}_i = (\mathcal{K}_h(X_i - x), (X_i - x)'\mathcal{K}_h(X_i - x))$.

- Consequently, for any given $x$, the local linear estimator for $(q_\tau(x), \dot{q}_\tau(x)')'$ can be obtained from the parametric quantile regression of $\tilde{Y}_i$ on $\tilde{X}_i$.
- Under suitable conditions, Lu et al. (2001) show that $\hat{q}_\tau(x)$ has the Bahadur representation

$$\begin{aligned} &\sqrt{nh^p}(\hat{q}_\tau(x) - q_\tau(x)) \\ &\qquad = \phi_\tau(x) \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} \psi(Y_i^*(x,\tau)) \mathcal{K}_h(X_i - x) + o_p(1) \end{aligned} \tag{15}$$

where $\psi_\tau(y) = \tau - 1(y \leq 0)$,
$Y_i^*(x,\tau) = Y_i - q_\tau(x) - \dot{q}_\tau(x)'(X_i - x)$,
$\phi_\tau(x) = (f_{Y|X}(q_\tau(x)|x) f_X(x))^{-1}$, $f_{Y|X}$ is the conditional density of $Y$ given $X = x$ and $f_X$ is the marginal density of $X$.

- The above result is the key to study the asymptotic properties of the quantile estimator $\hat{q}_\tau(x)$. One can obtain the Bahadur representation for the derivative estimator $\hat{\dot{q}}_\tau(x)$. It is worth mentioning that we can obtain results by resorting to the convexity lemma of Pollard (1991). See Fan et al. (1994) for details.

### Theorem

*Under some regularity conditions, we have*

$$\sqrt{nh^q}\Big(\hat{q}_\tau(x) - q_\tau(x) - \frac{1}{2}h^2 tr\Big[\ddot{q}_\tau(x)\int uu'K(u)du\Big]\Big)$$
$$\rightsquigarrow N\Big(0, \frac{\tau(1-\tau)\int K(u)^2du}{[f_{Y|X}(q_\tau(x)|x)]^2 f_X(x)}\Big).$$

*where $\ddot{q}_\tau(x)$ is the second derivative matrix of $q_\tau(x)$, whose $(i,j)th$ element is given by $\partial^2 q_\tau(x)/\partial x_i \partial x_j$.*

- The above result suggests that for an interior point $x$, the MSE of $\hat{q}_\tau(x)$ is given by

$$MSE(\hat{q}_\tau(x)) \backsimeq \left\{\frac{1}{2}h^2 tr\left[\ddot{q}_\tau(x)\int uu'\mathcal{K}(u)du\right]\right\}^2 + \frac{\tau(1-\tau)\int K(u)^2 du}{nh^p[f_{Y|X}(q_\tau(x)|x)]^2 f_X(x)}$$

  Consequently, the optimal rate of bandwidth in terms of minimizing the MSE is proportional to $n^{-1/(4+p)}$. When $x$ lies on the boundary of the support, the MSE formula looks similar. This reflects the two major advantages of local linear fitting and shows that these advantages apply to the local quantile regression too: (a) no dependence of the asymptotic bias on the density $f_X(x)$ and (b) automatic good behavior at boundaries.

# Bandwidth selection

- Unfortunately, to the best of our knowledge, there does not exist an automatic data-driven method for optimally selecting bandwidths when estimating a conditional quantile function in the sense that a weighted integrated MSE is minimized.
- Yu and Jones (1998) suggest choosing bandwidth by regressing the response $Y_i$ on the covariate $X_i$ and then modify the selected bandwidth by assuming normality. However, it is straightforward to use the principle of cross-validation to choose bandwidth. We leave the theoretical justification for future research.

## Two Other Nonparametric Quantile Estimators

- We now introduce briefly two other nonparametric quantile estimator in the literature. One is based upon the Nadaraya-Watson (NW) estimator for the conditional distributions; the other is a smoothed version of the local linear quantile estimator introduced previously.
- **Weighted Nadaraya-Watson (WNW) Estimator** Denote by $F_{Y|X}(y|x)$ the conditional distribution function $Y$ given $X = x$. Motivated by the good boundary properties of local polynomial estimators, Hall et al. (1999) suggests estimating $F_{Y|X}(y|x)$ by a weighted version of the well known NW estimator:

$$\hat{F}_{wnw}(y|x) = \frac{\sum_{t=1}^{n} p_t(x) K_{h_1}(X_t - x) 1(Y_t \leq y)}{\sum_{t=1}^{n} p_t(x) K_{h_1}(X_t - x)},$$

where $h_1 = h_1(n)$ the bandwidth, and one chooses the nonnegative weight functions $p_t(x), 1 \leq t \leq n$, such that

$$\sum_{t=1}^{n} p_t(x) = 1, \quad \sum_{t=1}^{n} p_t(x)(X_t - x) K_{h_1}(X_t - x) = 0$$

- More recently, Cai (2002) proposes to choose $\{p_t(x)\}$ based on the idea of empirical likelihood, i.e., to maximize $\sum_{t=1}^{n} log\{p_t(x)\}$ subject to the constraints specified above constraints. He proposes to invert $\hat{F}_{wnw}$ to get the conditional quantile estimator:

$$\hat{q}_\tau^{wnw}(x) = \inf\{y \in \mathbb{R} : \hat{F}_{wnw}(y|x) \geq \tau\}.$$

- **Smoothed Local Linear (SLL) Estimator** Let $l$ be a symmetric density function on $\mathbb{R}$ and $L$ be the corresponding distribution function. Yu and Jones (1998) propose a smoothed local linear estimator for conditional quantiles that is based on the observation $E[L((y - Y_t)/h_2)|X_t = x] \to F(y|x)$ as the bandwidth $h_2 \to 0$. To obtain the smoothed local linear (SLL) estimator for the conditional quantile function, one first obtains

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \arg\min_{\beta} \sum_{t=1}^{n} \{L((y - Y_t)/h_2) - \beta_0 - \beta_1'(X_t - x)\}^2 K_{h_2}(X_t - x),$$

where $\beta = (\beta_0, \beta_1')' \in \mathbb{R} \times \mathbb{R}^d$, $h_2 = h_2(n)$ is the bandwidth.

- Set $\hat{F}_{sll}(y|x) = \hat{\beta}_0$. Yu and Jones (1998) propose to invert $\hat{F}_{sll}$ to get the conditional quantile estimator:

$$\hat{q}_\tau^{sll}(x) = \inf\{y \in \mathbb{R} : \hat{F}_{sll}(y|x) \geq \tau\}$$

Note that $\hat{F}_{sll}(y|x)$ can range outside $[0, 1]$. In the special case where $d = 1$ it can be expressed as

$$\hat{F}_{sll}(y|x) = \sum_{t=1}^{n} w_t(x) L((y - Y_t)/h_2) / \sum_{s=1}^{n} w_s(x),$$

where $w_t(x) = K_{h_1}(X_t - x)(1 - \hat{\beta}_x(x - X_t))$, and

$$\hat{\beta}_x = \left(\sum_{t=1}^{n} K_{h_1}(X_t - x)(x - X_t)^2\right)^{-1} \sum_{t=1}^{n} K_{h_1}(X_t - x)(x - X_t).$$

To obtain an monotone estimator for $F(y|x)$ that lies between 0 and 1, Hansen (2004a) proposes to replace $w_t(x)$ by $w_t^*(x) = K_{h_1}(X_t - x)(1 - \hat{\beta}(x - X_t))1\{\hat{\beta}_x(x - X_t) \leq 1\}$.

# np package

- The np package implements recently developed kernel methods that seamlessly handle the mix of continuous, unordered, and ordered factor data types often found in applied settings.
- Functions relates to nonparametric regression and tests including

| Function | Description | Reference |
|---|---|---|
| `npcdens` | Nonparametric conditional density estimation | Hall *et al.* (2004) |
| `npcdensbw` | Nonparametric conditional density bandwidth selection | Hall *et al.* (2004) |
| `npcdist` | Nonparametric conditional distribution estimation | Li and Racine (2008) |
| `npcmstest` | Parametric model specification test | Hsiao *et al.* (2007) |
| `npconmode` | Nonparametric modal regression | |
| `npindex` | Semiparametric single index model | Ichimura (1993), Klein and Spady (1993) |
| `npindexbw` | Semiparametric single index model parameter and bandwidth selection | Ichimura (1993), Klein and Spady (1993) |
| `npksum` | Nonparametric kernel sums | |
| `npplot` | General purpose plotting of nonparametric objects | |

- 

| | | |
|---|---|---|
| `npplreg` | Semiparametric partially linear regression | Robinson (1988), Racine and Liu (2007) |
| `npplregbw` | Semiparametric partially linear regression bandwidth selection | Robinson (1988), Racine and Liu (2007) |
| `npqcmstest` | Parametric quantile regression model specification test | Zheng (1998), Racine (2006) |
| `npqreg` | Nonparametric quantile regression | Li and Racine (2008) |
| `npreg` | Nonparametric regression | Racine and Li (2004), Li and Racine (2004) |
| `npregbw` | Nonparametric regression bandwidth selection | Hurvich, Simonoff, and Tsai (1998), Racine and Li (2004), Li and Racine (2004) |
| `npscoef` | Semiparametric smooth coefficient regression | Li and Racine (2007b) |
| `npscoefbw` | Semiparametric smooth coefficient regression bandwidth selection | Li and Racine (2007b) |
| `npsigtest` | Nonparametric regression significance test | Racine (1997), Racine *et al.* (2006) |
| `npudens` | Nonparametric density estimation | Parzen (1962), Rosenblatt (1956), Li and Racine (2003) |
| `npudensbw` | Nonparametric density bandwidth selection | Parzen (1962), Rosenblatt (1956), Li and Racine (2003) |
| `npudist` | Nonparametric distribution estimation | Parzen (1962), Rosenblatt (1956), Li and Racine (2003) |
| – Utilities – | | |
| `gradients` | Extract gradients | |
| `se` | Extract standard errors | |
| `uocquantile` | Compute quantiles/modes for unordered, ordered, and numeric data | |

- Nonparametric quantile regression on the Italian GDP panel.