

Lec 17: Applications of Nonparametric Kernel regression

Weiping Zhang

2018.11.26

Functional Coefficient Models

- Estimation of Coefficient Functions

- Asymptotic Properties for the Local Linear Estimator

Nonparametric Quantile Estimation

- The Local Linear Nonparametric Quantile Estimator

- Two Other Nonparametric Quantile Estimators

Nonparametric Model Specification Tests

- Test for Correct Parametric Functional Form

- Nonparametric Test for Omitted Variables

Implementation in R

Functional Coefficient Models

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

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

where ϵ is a zero mean disturbance term, X_j , $j = 1, \dots, d$, are regressors but whose coefficients depend on another random variable or vector U_j . Model (1) implies that the random variables U_j , $j = 1, \dots, d$ 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, \dots, X_d on Y . For example, when $U_j = t_j$ (time trend), the model can be written as

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

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 ϵ_t is independent of X_{t-1}, X_{t-2}, \dots . As in model (1), the coefficient functions $a_1(\cdot), \dots, a_p(\cdot)$ 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 $\beta_{4i} \geq 0, i = 1, \dots, d$.

- 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:

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

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

Estimation of Coefficient Functions

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

$$Y = Z' A(U) + \epsilon \quad (6)$$

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

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

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 $h = (h_1, \dots, h_p)$ 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

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

- 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

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

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

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

where $\theta = (a_1, \dots, a_d, b'_1, \dots, b'_d)$. So the local estimator is simply

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

which entails that

$$\hat{a}_j(u) = \hat{a}_j = e'_{j,d(p+1)}\hat{\theta}, j = 1, \dots, d. \quad (12)$$

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 $H_1 = \text{diag}\{1, \dots, 1, h', \dots, h'\}$ 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 $j = 1, \dots, d$,

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

Bandwidth Selection

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

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

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 τ -th conditional quantile $q_\tau(x)$ of Y_t given $X_t = x$ satisfies

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

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 β is a $p \times 1$ vector of parameters, and $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$.

- Denote by $\dot{q}_\tau(x) = (\partial q_\tau(x)/\partial x_1, \dots, \partial q_\tau(x)/\partial x_p)'$ the first order derivative of $q_\tau(x)$ at $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$. The idea of the local linear fit is to approximate the unknown τ -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 β_0 and estimating $\dot{q}_\tau(x)$ is equivalent to estimating β_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 $\hat{\dot{q}}_\tau(x) = \hat{\beta}_1$, respectively, where

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

- 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

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

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

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

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

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

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

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

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.

- 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 $E[L((y - Y_t)/h_2)|X_t = x] \rightarrow F(y|x)$ as the bandwidth $h_2 \rightarrow 0$. 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 $\hat{F}_{sl}(y|x) = \hat{\beta}_0$. Yu and Jones (1998) propose to invert \hat{F}_{sl} to get the conditional quantile estimator:

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

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

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

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 $w_t^*(x) = K_{h_1}(X_t - x)(1 - \hat{\beta}(x - X_t))1\{\hat{\beta}_x(x - X_t) \leq 1\}$.

Nonparametric Model Specification Tests

- In practice, we must confront the fact that all models are potentially misspecified. Many popular parametric model specification tests require the specification of the set of parametric alternatives for which one will reject the null. If there exist some alternative models which the test cannot detect, then the test is said to be an “inconsistent test” since it lacks power in certain directions. One famous example is the Jarque-Bera test for normality. Since this test can only detect deviations from normality by the third and fourth moments, it does not have power in detecting distributions that differ from normal distributions only through higher order moments.
- A popular application of nonparametric methods turns out to be model specification test which is usually a consistent test in that it has power in detecting all kinds of deviations from the null at certain rates.

- To test whether a parametric model is correctly specified, we usually need to compare an estimate of the parametric model with its consistent nonparametric estimate. The null of interest is

$$H_0 : P[E(Y_i|X_i) = m(X_i; \theta)] = 1 \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p$$

where $m(x; \theta)$ is a known function, X_i is a $q \times 1$ vector of regressors, θ is a $p \times 1$ vector of unknown parameters. Θ is the parameter space on \mathbb{R}^p . The alternative hypothesis is

$$H_1 : P[E(Y_i|X_i) = m(X_i; \theta)] < 1 \quad \text{for all } \theta \in \Theta \subset \mathbb{R}^p$$

- Define $u_i = Y_i - m(X_i; \theta_0)$. The null is equivalent to

$$H_0 : E(u_i|X_i) = 0, a.s. \quad (16)$$

Noticing that

$$\begin{aligned} E[u_i E(u_i|X_i) f(X_i)] &= E\{E[u_i|X_i] f(X_i) |X_i]\} \\ &= E\{[E(u_i|X_i)]^2 f(X_i)\} = 0 \end{aligned}$$

under the null hypothesis and > 0 under the alternative hypothesis, where f is the pdf of X_i , we will construct a consistent model specification test based upon this observation under (16). Density weighting is used here simply to avoid a random denominator that would otherwise appear in the kernel estimator and make it hard to establish the asymptotic theory for the test.

- Clearly, the sample analogue of $E[u_i E(u_i|X_i) f(X_i)]$ is

$$\frac{1}{n} \sum_{i=1}^n u_i E(u_i|X_i) f(X_i),$$

which is infeasible since we don't observe u_i and don't know the functional form of f . To construct a feasible test statistic, we need to replace u_i and f by their consistent estimates. Let $\hat{u}_i = Y_i - m(X_i; \hat{\theta})$, where $\hat{\theta}$ is a \sqrt{n} -consistent estimator of θ based on the null model. Let

$\mathcal{K}_{h,ij} = \prod_{s=1}^q h_s^{-1} K((X_{is} - X_{js})/h_s)$. We can estimate $E(u_i|X_i) f(X_i) = E[u_i f(X_i)|X_i]$ by the leave-one-out kernel estimator

$$\frac{1}{n-1} \sum_{j=1, j \neq i}^n \hat{u}_j \mathcal{K}_{h,ij}$$

- Our test statistic is based upon

$$\begin{aligned} I_n &= \frac{1}{n} \sum_{i=1}^n \hat{u}_i \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \hat{u}_j \mathcal{K}_{h,ij} \right\} \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \hat{u}_i \hat{u}_j \mathcal{K}_{h,ij} \end{aligned}$$

The next theorem states the asymptotic null distribution of I_n .

Theorem

Under certain regularity conditions and H_0 ,

$$\frac{n(h_1 \cdots h_q)^{1/2} I_n}{\hat{\sigma}} \rightsquigarrow N(0, 1),$$

where $\hat{\sigma}^2 = 2n^{-2}h_1 \cdots h_q \sum_{i=1}^n \sum_{j \neq i}^n \hat{u}_i^2 \hat{u}_j^2 \mathcal{K}_{h,i,j}^2$ is a consistent estimator of $\sigma_e^2 = 2\kappa_{02}^q E[\sigma^4(X_i)f(X_i)]$ and $\sigma^2(x) = E(u_i^2 | X_i = x)$.

- Since the test is one-sided, we will reject the null when $T_n > z_\alpha$ at the significance level α . It is easy to show that the above test T_n is consistent. It has power in detecting any deviations from the null at the nonparametric rate $n^{-1/2}(h_1 \cdots h_q)^{-1/4}$.
- Li and Wang (1998) also proposed a **wild Bootstrap** test for finite sample. Fan and Linton (2003) further analyze the accuracy of the bootstrap test statistic.

- Härdle and Mammen (1993) also consider testing

$$H_0 : P[E(Y_i|X_i) = m(X_i; \theta)] = 1 \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p$$

where $m(x; \theta)$ is a known function, X_i is a $q \times 1$ vector of regressors, θ is a $p \times 1$ vector of unknown parameters. Θ is the parameter space on \mathbb{R}^p . The alternative hypothesis is the negation of H_0 . Let $m(x) = E(Y|X = x)$ and $\hat{m}(x)$ be its nonparametric kernel estimate. Härdle and Mammen (1993) propose a consistent test for parametric functional form based upon

$$I_n = \int [\hat{m}(x) - m(x; \hat{\theta})]^2 w(x) dx,$$

where $w(x)$ is a nonnegative weight function, and $\hat{\theta}$ is a \sqrt{n} -consistent estimator for θ .

- If one choose $w(x) = f^2(x)$ and use $\hat{f}(x) = n^{-1} \sum_{i=1}^n \mathcal{K}_h(X_i - x)$ to replace $f(x)$, then the above test statistic turns to be

$$\begin{aligned}
 I_n &\approx^* \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [Y_i - m(X_i, \hat{\theta})][Y_j - m(X_j, \hat{\theta})] \\
 &\quad \cdot \int \mathcal{K}_h(X_i - x) \mathcal{K}_h(X_j - x) dx \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{u}_i \hat{u}_j \bar{\mathcal{K}}_h(X_i, X_j)
 \end{aligned}$$

where $\hat{u}_i = Y_i - m(X_i, \hat{\theta})$,

$\bar{\mathcal{K}}_h(X_i, X_j) = \prod_{s=1}^q h_s^{-1} \bar{K}((X_{is} - X_{js})/h_s)$, and

$\bar{K}(v) = \int K(u)K(v - u)du$ is the convolution kernel for K .

- * The small terms are ignored in above derivation.

- The above test statistic has similar form to the test of Li and Wang (1998). By the latter result, we can prove the following theorem

Theorem

Under certain regularity conditions and H_0 ,

$$T_n = \frac{n(h_1 \cdots h_q)^{1/2}[I_n - c(n)]}{\hat{\sigma}} \rightsquigarrow N(0, 1),$$

where $\hat{\sigma}^2 = 2n^{-2}(h_1 \cdots h_q)^{-1} \sum_{i=1}^n \sum_{j \neq i}^n \hat{u}_i^2 \hat{u}_j^2 \bar{\mathcal{K}}_h^2(X_i, X_j)$ and $c(n) = (nh_1 \cdots h_q)^{-1} \bar{\kappa}^q(0) \sum_{i=1}^q \hat{u}_i^2$.

- If one chooses $w(x) = f(x)$, this can yields another version of the Härdle and Mammen's (1993) test.

Nonparametric Test for Omitted Variables

- Nonparametric tests for omitted variables have also been widely studied in the literature. Let $X \in \mathbb{R}^q$ be a $q \times 1$ vector of continuous random variables, and partition $X = (W, Z) \in \mathbb{R}^p \times \mathbb{R}^{q-p}$, where $1 \leq p < q$. The null hypothesis is that the conditional mean of Y does not depend on Z i.e.,

$$H_0 : P[E(Y|W, Z) = E(Y|W)] = 1.$$

The alternative is the negation of H_0 .

- Let $u_i = Y_i - E(Y_i|W_i)$, then $E(u_i|X_i) = 0$ under the null hypothesis. So we can construct a test statistic based on

$$E[u_i f_w(W_i) E\{u_i f_w(W_i)|X_i\} f(X_i)],$$

where f_w and f are the density functions of W_i and X_i , respectively. Let \hat{f}_{w_i} and \hat{Y}_i be the leave-one-out kernel estimators of $f_w(W_i)$ and $E(Y_i|W_i)$, respectively. That is

$$\hat{f}_{w_i} = \frac{1}{n-1} \sum_{j \neq i} \mathcal{K}_{h_w}(W_j - W_i),$$

$$\hat{Y}_i = \frac{1}{(n-1)\hat{f}_{w_i}} \sum_{j \neq i} Y_j \mathcal{K}_{h_w}(W_j - W_i),$$

where \mathcal{K} is the product kernel with bandwidth $h_w = (h_{w,1}, \dots, h_{w,p})$.

- A feasible test statistic is given by

$$I_n = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} (Y_i - \hat{Y}_i) \hat{f}_{w_i} (Y_j - \hat{Y}_j) \hat{f}_{w_j} \mathcal{K}_h(X_i - X_j),$$

where $\mathcal{K}_h(X_i - X_j) = \prod_{s=1}^q K_{h_s}(X_{is} - X_{js})$.

Theorem

Under certain regularity conditions and H_0 ,

$$\frac{n(h_1 \cdots h_q)^{1/2} I_n}{\hat{\sigma}} \rightsquigarrow N(0, 1),$$

where $\hat{\sigma}^2 = 2n^{-2} h_1 \cdots h_q \sum_{i=1}^n \sum_{j \neq i}^n \hat{u}_i^2 \hat{u}_j^2 \hat{f}_{w_i}^2 \hat{f}_{w_j}^2 \mathcal{K}_h^2(X_i, X_j)$ and $\hat{u}_i = Y_i - \hat{Y}_i$.

- Even though the above test statistic has asymptotic normal distribution under the null, simulations in Li (1999) and Lavergne and Vuong (2000) reveal that the asymptotic normal approximation does not work well for small to moderate samples. In practice, one can use the wild bootstrap principle to construct a wild bootstrap test statistic.

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



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

- Nonparametric quantile regression on the Italian GDP panel.

