

On Semiparametric Mode Regression Estimation

By ALI GANNOUN

Chaire de Statistique Appliquée and CEDRIC
CNAM

292, Rue Saint-Martin
75141 Paris, France

JEROME SARACCO

Institut de Mathématiques de Bordeaux, UMR CNRS 5251,
Université Bordeaux 1
351 cours de la libération
33405 TALENCE Cedex

and KEMING YU

Department of Mathematical Sciences
Brunel University, Uxbridge, UB8, 3PH, United Kingdom

Abstract

It has been found that, for a variety of probability distributions, there is a surprising linear relation between mode, mean and median. In this paper, the relation between mode, mean and median regression functions is assumed to follow a simple parametric model. We propose a semiparametric conditional mode (mode regression) estimation for an unknown (unimodal) conditional distribution function in the context of regression model, so that any m -step-ahead mean and median forecasts can then be substituted into the resultant model to deliver m -step-ahead mode prediction. In the semiparametric model, Least Squared Estimator (LSEs) for the model parameters and the simultaneous estimation of the unknown mean and median regression functions by the local linear kernel method are combined to infer about the parametric and nonparametric components of the proposed model. The asymptotic normality of these estimators is derived, and the asymptotic distribution of the parameter estimates is also given and is shown to follow usual parametric rates in spite of the presence of the nonparametric component in the model. These results are applied to obtain a data-based test for the dependence of mode regression over mean and median regression under a regression model.

Keywords. Asymptotic normality, hypothesis testing, local linear kernel estimate, mode, prediction, rate of convergence, semiparametric regression

1 Introduction

Mode, the most likely value of a distribution, has wide applications in astronomy, biology and finance. It is not uncommon in many fields to encounter data distributions that are skewed or contain outliers. In those cases, the arithmetic mean may not be an appropriate statistic to represent the center of location of the data. Alternative statistics with less bias are the median and the mode. The median is the value of the variable, in an ordered array, which has an equal number of data points on either side, whereas the mode is the value of the peak of the distribution. The mean or median of two densities may be identical, while the shapes of the two densities are quite different. Mode preserves some of the important features such as wiggles of underlying distribution function, whereas the mean or median tend to average out of the data. These important features of a probability distribution are usually very useful (Hedges and Shah, 2003). However, calculation of the mode is more difficult than the mean or median and this has limited its widespread application although kernel mode estimation has attracted much attention in literature. Under nonparametric kernel estimation, the mode estimator is defined as the maximum of the estimated conditional density. In fact, nonparametric estimation of mode had been discussed in decades (Grenander, 1956, Parzen, 1962, Eddy, 1980, Gasser and Hall, 1995, Bickel and Fan, 1996, Birgè, 1997, Berlinet *et al.*, 1998, Meyer, 2001). For deep discussion of the mode in regression, see, for example, Lee (1989, 1993), Shoung and Zhang (2001) and Ziegler (2002, 2003). However, kernel methods are very sensitive to smoothing parameter selection and not easy to implement in practice. Moreover, kernel methods and other existing methods for mode estimation do not provide prediction which is sometimes important. For example, in finance, the prediction of the single most likely future interest rate or the most likely share price or forecasting future inflation is of people's main interest. Our method in this paper is motivated by a closely insight of the relationship between mode, mean and median of a probability distribution. That is, the three simple averages *mean*, *median* and *mode* are identical for a symmetrical distributions and so, they satisfy the equation

$$\text{mode} = 3 \times \text{median} - 2 \times \text{mean}. \quad (1)$$

This relation also holds approximatively for a wide range of asymmetrical distributions (see the discussion and proof given by the Section 2.1 of Stuart and Ord, 1994 and the Appendix of Lee, 1994).

To extend the above relation to regression setting, let (X, Y) be a pair of random variables, where Y is the response variable and X its covariate. Suppose that the conditional distribution of Y given X is absolutely continuous. We denote by $f(\cdot | x)$ and $F(\cdot | x)$ the conditional density and distribution functions of Y giving $X = x$. The covariate X may be a vector. Suppose that (X, Y) is linked via a nonparametric regression model

$$Y = m(X) + \epsilon, \quad (2)$$

where m is an unknown mean regression function and ϵ is the model error. The mode function of Y given $X = x$, denoted by $\mu(x)$, is defined as the maximizer of the function $f(\cdot|x)$. Graphically, $\mu(x)$ displays the peak variation at different locations of X . Under model (2), let $m(x)$ and $q(x)$ be the conditional mean and median of Y given $X = x$ defined by

$$m(x) = E[Y|X = x],$$

$$q(x) = \operatorname{argmin}_{a \in \mathbb{R}} \{E[|Y - a||X = x]\}.$$

In this paper, we address the estimation of $\mu(x)$ via a linear parametric model

$$\mu(x) = \theta_0 + \theta_1 m(x) + \theta_2 q(x), \quad (3)$$

where $\theta = (\theta_0, \theta_1, \theta_2)$ is an unknown parameter vector.

In the context of financial returns, if we have historical time-series of mean and median estimates and records of mode, one-step ahead mean and median forecasts, $m_{t+1}(x)$ and $q_{t+1}(x)$ can then be substituted into the resultant model (3) to deliver one-step ahead mode prediction as in the following expression:

$$\mu_{t+1}(x) = \theta_0 + \theta_1 m_{t+1}(x) + \theta_2 q_{t+1}(x),$$

where θ_0 , θ_1 and θ_2 are the parameters estimated by the Least-Squared regression.

Combining model (3) with the nonparametric kernel estimation of $m(x)$ and $q(x)$, we develop a semiparametric mode function estimation.

Before proceeding the details of theoretical investigation, we carry out a simulation study to illustrate the performance of the parametric regression (3) in Section 2. The paper is then organized as follows. Section 3 discusses the estimation of regression parameters and functions involved in the models. Section 4 lists the main theoretical results of those estimators. The testing theory is derived in Section 5. Some proofs related to these conclusions are given in the Appendix.

2 Empirical study of semiparametric mode estimation under unique mode

Consider an unconditional regression, which is a specific case of model (2)

$$Y_i = \epsilon_i, \quad i = 1, \dots, n,$$

and let f be the underlying unknown probability density. The unique mode includes unimodal and multimodal with one largest peak.

We discuss the mode estimation of $f(\cdot)$. We compare the mode estimation by nonparametric kernel estimation of $f(\cdot)$ and the semiparametric model (3).

Let $f(\cdot)$ be the two-piece normal distribution with parameters μ , σ_1 and σ_2 , whose probability density function is for each $y \in \mathbb{R}$,

$$f(y) = \begin{cases} A \exp[-(y - \mu)^2/2\sigma_1^2], & y < \mu \\ A \exp[-(y - \mu)^2/2\sigma_2^2], & y \geq \mu \end{cases}$$

where $A = (\sqrt{2\pi}(\sigma_1 + \sigma_2)/2)^{-1}$. This density has been used as inflation forecasts by the Bank of England since 1997. It has mode μ and positive skewness when $\sigma_2 > \sigma_1$. Therefore, the distribution function of Y is given for each $y \in \mathbb{R}$ by

$$F(y) = \begin{cases} \frac{2\sigma_1}{\sigma_1 + \sigma_2} \Phi\left(\frac{y - \mu}{\sigma_1}\right), & y < \mu \\ \frac{2\sigma_2}{\sigma_1 + \sigma_2} \Phi\left(\frac{y - \mu}{\sigma_2}\right) + \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}, & y \geq \mu \end{cases}$$

where Φ is the standard normal distribution function $N(0, 1)$.

The probability of outcomes between L_1 and L_2 for the two-piece normal distribution is derived as

$$P(L_1 \leq Y \leq L_2) = \int_{L_1}^{L_2} f(y) dy =$$

- $\frac{\sigma}{\sigma_1 + \sigma_2} [\Phi(\frac{L_2 - \mu}{\sigma}) - \Phi(\frac{L_1 - \mu}{\sigma})]$ where $\sigma = \begin{cases} 2\sigma_1, & L_1 \leq L_2 \leq \mu \\ 2\sigma_2, & \mu \leq L_1 \leq L_2 \end{cases}$
- $\frac{2\sigma_2}{\sigma_1 + \sigma_2} \Phi(\frac{L_2 - \mu}{\sigma_2}) + \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} - \frac{2\sigma_1}{\sigma_1 + \sigma_2} \Phi(\frac{L_1 - \mu}{\sigma_1})$, where $L_1 \leq \mu \leq L_2$.

The exact relation (1) does not hold for this distribution, but we suppose the relation (3) does hold or approximately holds. When $\mu = 0$, $\sigma_1 = 1$ and $\sigma_2 = 2$, $f(\cdot)$ is skewed to right.

2.1 Kernel estimation of f

Suppose that Y_1, Y_2, \dots, Y_n , denotes a random sample from the unknown probability density function $f(\cdot)$. For each $y \in \mathbb{R}$, the kernel density estimator of $f(y)$ is defined as follows:

$$\hat{f}(y) = (nh)^{-1} \sum_{i=1}^n K\left(\frac{y - Y_i}{h}\right),$$

where K is a bounded nonnegative function satisfying $\int K(x) dx = 1$, and $h = h(n)$ is a sequence of positive numbers usually called the bandwidth.

Kernel smoothing provides a simple way of finding structures in data sets without knowing the probability density function. Theoretical and simulation analysis has shown that the choice of the kernel is not crucial for density estimation in the case of independent identically distributed (i.i.d.) random variables. The most important part in the kernel estimation method is to select the bandwidth. Some methods are listed below:

- Least Squares Cross-Validation (see Rudemo, 1982, and Bowman, 1984)
- Biased Cross-Validation (see Scott and Terrell, 1987)
- Plug-in Bandwidth Selection (see Sheather and Jones, 1991)
- Smoothed Cross-Validation (see Hall, Marron and Park, 1992)
- Root-n Bandwidth Selection (see Hall, Sheather, Jones and Marron, 1991)
- The Contrast Method (see Ahmad and Ran, 2004)

Most of these methods are based on minimizing the Mean Squared Error (MSE) or the Mean Integrated Squared Error (MISE).

2.2 Simulations

For convenience, we implement the following three different bandwidth rules for kernel density estimation. Note that, the first rule is the easiest one, but it assumes that the data are normally distributed. The other rules suppose that kernel function K is a second order kernel ($\int K(y)dy = 1$, $\int yK(y)dy = 0$ and $\int y^2K(y)dy \neq 0$). This is the case when K is the standard normal kernel. Useful comments and justification concerning rules 2 and 3 are in the nice paper of Grund and Hall (1995).

- Silverman's Rule of Thumb: $h = 1.06 S n^{-1/5}$, where S is the sample standard deviation of data (Silverman, 1986),
- oversmoothing bandwidth $h = c n^{-1/11}$, where c is a constant selected properly (Grund and Hall, 1995),
- the minimised L_p distance based bootstrap estimator on a grid of h -values (Grund and Hall, 1995).

We simulate 3 samples from f of size $n = 50, 100$ and 200 , respectively. For each sample we first obtain the kernel mode estimator. For each case of bandwidth selection, we then regress the kernel mode estimator over mean and median as in (3) to get the semiparametric mode estimation by Least Squares regression. To fit the parametric regression (3), we repeatedly estimate the kernel mode, mean and median via resampling. For example, we resample 50-times with sample size $n = 50$, then we obtain 50 observations of kernel mode, sample mean and sample median. To compare the Bias, Standard Deviation (SD) and Mean Squared Error (MSE) of kernel estimation and semiparametric estimation, we run the simulation 100 times.

Table 1. Mode estimation for two-piece normal distribution: Biases, Standard Deviations and MSEs

n	Kernel smoothing			Semiparametric model			
	h by	Bias	SD	MSE	Bias	SD	MSE
50	Rule-of-thumb	0.283 (0.021)	0.316 (0.014)	0.197 (0.007)	0.239 (0.096)	0.281 (0.020)	0.196 (0.005)
	Oversmoothing	-0.276 (0.088)	0.491 (0.025)	0.214 (0.016)	-0.208 (0.037)	0.321 (0.022)	0.193 (0.010)
	Bootstrap	0.207 (0.034)	0.298 (0.008)	0.198 (0.010)	0.195 (0.007)	0.171 (0.012)	0.116 (0.010)
100	Rule-of-thumb	0.179 (0.013)	0.143 (0.013)	0.117 (0.007)	0.123 (0.011)	0.108 (0.008)	0.154 (0.005)
	Oversmoothing	0.183 (0.012)	0.191 (0.008)	0.114 (0.006)	0.158 (0.031)	0.188 (0.002)	0.101 (0.010)
	Bootstrap	0.155 (0.006)	0.120 (0.014)	0.118 (0.009)	0.152 (0.021)	0.123 (0.012)	0.106 (0.005)
200	Rule-of-thumb	0.152 (0.016)	0.167 (0.013)	0.019 (0.008)	0.141 (0.012)	0.168 (0.014)	0.009 (0.007)
	Oversmoothing	0.157 (0.021)	0.111 (0.006)	0.021 (0.006)	0.142 (0.013)	0.105 (0.021)	0.019 (0.008)
	Bootstrap	0.112 (0.012)	0.122 (0.021)	0.009 (0.014)	0.112 (0.006)	0.115 (0.022)	0.007 (0.019)

Table 1 summarises the simulations. Clearly, the semiparametric method never does worse than the kernel estimation. In particular, when the rule-of-thumb gives small bandwidths which results in too many peaks of kernel density estimation where the highest peak does not correspond to the mode estimation, the semiparametric estimation is not affected. On the other hand, when the over-smoothing rule may result in flat top of kernel density estimation with too many candidates for the mode, the semiparametric estimation still overperforms. We have done extensive simulations based on a mixture of normal densities and found similar conclusions to the one reported here.

3 The formulation of the semiparametric model

In Section 1, we have introduced the idea behind the semiparametric mode estimation. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from a population (X, Y) . We consider the following regression model:

(S1) $Y_i = m(X_i) + \epsilon_i$, $i = 1, \dots, n$, where $m(\cdot)$ is an unknown regression function.

We assume that the median function $q(\cdot)$ and the mode function $\mu(\cdot)$ are unknown. Without loss of the generality, all of the three functions are defined on $[0, 1]$. The points X_i at which measurements Y_i are taken are assumed to be fixed or random, and in the latter case, the distribution density $g(\cdot)$ of X_i satisfies $0 < \inf_{[0,1]} g(\cdot) \leq \sup_{[0,1]} g(\cdot) < \infty$.

(S2) For the errors $\epsilon_i = \epsilon(X_i)$, it is assumed that they are independent with

$$E[\epsilon_i] = 0, \quad E[\epsilon_i^2] = \sigma^2(X_i)$$

where the conditional variance functions $\sigma^2(\cdot)$ are continuous on $[0, 1]$.

(S3) It is also assumed that there exists an $s > 4$ such that $E[|\epsilon_i|^s] < C < \infty$, where C is a positive constant.

Assumptions **(S1)** and **(S3)** are rather general, allowing for all kinds of error distributions, including multi-mode distributions.

The nonparametric part of our model consists of smoothness assumptions (given below) on regression mean $m(\cdot)$, mode $\mu(\cdot)$ and median $q(\cdot)$ functions.

(S4) Let $\mathcal{C}([0, 1])$ denotes the space of two times continuously differentiable functions on $[0, 1]$. Then,

$$m(\cdot), q(\cdot), \mu(\cdot) \in \mathcal{C}([0, 1]).$$

(S5) For fixed x , the parametric part of our semiparametric model relates the regression function $m(x)$, the median function $q(x)$ and the mode function $\mu(x)$ as follows:

$$\mu(x) = \theta_0 + \theta_1 m(x) + \theta_2 q(x).$$

(S6) For any constant vector $\beta = (\theta_0, \theta_1, \theta_2)^T$,

$$\int_0^1 g(u) (\theta_0 + \theta_1 m(u) + \theta_2 q(u))^2 du > 0,$$

where V^T denotes the transpose of the vector (or matrix) V .

Our first aim is to estimate the functions $m(\cdot)$, $q(\cdot)$ and $\mu(\cdot)$ and the parameter β . A particular interest is the question whether rates of convergence for the parameter estimates are the usual parametric ones in spite of the nonparametric smoothness assumptions **(S4)**, and the question whether the usual asymptotic normality holds for the smoothing functions in spite of the parametric structure **(S5)**. Theorems 4, 5 and 6 answer that, while Theorems 1, 2 and 3 guarantee the usual nonparametric rates for function estimates of $m(\cdot)$, $q(\cdot)$ and $\mu(\cdot)$. **(S6)** means that $\mu(x)$ defined under **(S5)** satisfies $E_X \mu(X) > 0$.

4 The estimation

4.1 Estimation of the nonparametric components

The nonparametric components of our semiparametric model **(S1)**–**(S6)** consist of the regression function $m(\cdot)$, median function $q(\cdot)$ and the mode function $\mu(\cdot)$. These functions, which are smooth according to **(S4)**, are studied and estimated by many authors over last decades. In the following, we will present their estimators based on kernel or linear kernel methods, and give some references to better understand them.

From now on we suppose that $X = x$. The classical kernel estimate $\hat{m}(x)$ of $m(x)$ is given by

$$m_n(x) = \frac{\sum_{i=1}^n W_i(x) Y_i}{\sum_{i=1}^n W_i(x)} \quad (4)$$

with

$$W_i(x) = K \left(\frac{X_i - x}{h} \right) [S_{n,2}(x) - (X_i - x) S_{n,1}(x)],$$

and

$$S_{n,l}(x) = \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) (X_i - x)^l, \quad l = 1, 2,$$

where the measurements (X_i, Y_i) come from model **(S1)**, and kernel $K(\cdot)$ and bandwidth h may satisfy some regularity conditions (see below). Equation (4) is the typical local linear regression function estimate whose properties are discussed in Fan and Gijbels (1992) and Wand and Jones (1996).

The local linear kernel estimate $q_n(x)$ of $q(x)$ is obtained as \hat{a} with

$$(\hat{a}, \hat{b}) = \arg \min_{a,b} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) |Y_i - a - b(X_i - x)|. \quad (5)$$

More details on this estimator can be found in Fan and Gijbels (1996), Yu and Jones (1998), Cheng and Peng (2002) and Yu and Lu (2004). Other estimators, based on conditional distribution, can be found in Samanta (1989) or Berline et al (2001).

Finally, the kernel estimator $\mu_n(x)$ of $\mu(x)$ is given by

$$\mu_n(x) = \arg \min_y f_n(y|x) \quad (6)$$

with $\hat{f}_n(y|x)$ denotes the kernel estimator of $f(y|x)$ such that

$$f_n(y|x) = \frac{f_n(x, y)}{g_n(x)},$$

where

$$f_n(x, y) = \frac{1}{nh^2} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right) K\left(\frac{Y_j - y}{h}\right) \quad \text{and} \quad g_n(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right),$$

see Samanta and Thavaneswarana (1990) and Berline *et al.* (1998) for more details.

These estimates would serve as the initial estimators in Section 4.3.

4.2 Parametric estimation

Given the set of observations $\{(X_i, Y_i), i = 1, 2, \dots, n\}$, define the $n \times 3$ matrix

$$\mathbb{X} = [t_{i,j}]_{1 \leq i \leq n, 1 \leq j \leq 3}$$

with $t_{i,1} \equiv 1$, $t_{i,2} = m(X_i)$, $t_{i,3} = q(X_i)$. Furthermore, let $Z = (\mu(X_1), \dots, \mu(X_n))^T$.

Then we could re-write the parametric relation **(S5)** as a parametric regression model

$$Z = \mathbb{X}\beta. \quad (7)$$

It is natural to estimate the parametric vector β by a weighted least squares method,

$$\hat{\beta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)^T = \arg \min_{\theta_0, \theta_1, \theta_2} \sum_{t=1}^n u(X_t) [\mu(X_t) - \theta_0 - \theta_1 m(X_t) - \theta_2 q(X_t)]^2, \quad (8)$$

where $u(\cdot)$ is a weight function, which is assumed to be Lipschitz continuous.

Let $\hat{\mathbb{X}} = (\hat{t}_{i,j})_{1 \leq i \leq n, 1 \leq j \leq 3}$, $\hat{Z} = (\hat{\mu}(X_1), \dots, \hat{\mu}(X_n))^T$, and $Q^{-1} = \text{diag}(u(X_1), \dots, u(X_n))$. The estimate for parameter β becomes the classical LSE

$$\hat{\beta} = \left(\hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}} \right)^{-1} \hat{\mathbb{X}}^T Q^{-1} \hat{Z}.$$

4.3 Simultaneous estimation of parametric and nonparametric components.

We advise here an iterative procedure for the simultaneous estimation of β , $m(\cdot)$, $q(\cdot)$ and $\mu(\cdot)$ by taking advantage of model **(S5)** and some additional assumptions. For simultaneous estimation of the parameters and functions, Gauss-Seidel type iterative procedure can be employed. Once $\hat{\beta}$ estimates have been obtained, they can be used to improve estimation of $m(\cdot)$, $q(\cdot)$ and $\mu(\cdot)$ in the following way.

Set the estimators of $m(\cdot)$, $q(\cdot)$ and $\mu(\cdot)$ defined by equations (4), (5) and (6) as the initial iteration values and denote them as $\hat{q}_{(0)}(\cdot) = q_n(\cdot)$, $\hat{m}_{(0)}(\cdot) = m_n(\cdot)$ and $\hat{\mu}_{(0)}(\cdot) = \mu_n(\cdot)$, respectively. Set $\hat{\beta}_{(0)} = \left(\hat{X}_{(0)}^T Q^{-1} \hat{X}_{(0)} \right)^{-1} \hat{X}_{(0)}^T Q^{-1} \hat{Z}_{(0)}$. Then iterate as follows until convergence:

$$\begin{cases} \hat{\beta}_{(i+1)} &= \left(\hat{X}_{(i)}^T Q^{-1} \hat{X}_{(i)} \right)^{-1} \hat{X}_{(i)}^T Q^{-1} \hat{Z}_{(i)}, \\ \hat{m}_{(i+1)}(\cdot) &= \frac{\hat{\mu}_{(i)}(\cdot) - \hat{\theta}_{0(i+1)} - \hat{\theta}_{2(i+1)} \hat{q}_{(i)}(\cdot)}{\hat{\theta}_{1(i+1)}}, \\ \hat{q}_{(i+1)}(\cdot) &= \frac{\hat{\mu}_{(i)}(\cdot) - \hat{\theta}_{0(i+1)} - \hat{\theta}_{1(i+1)} \hat{m}_{(i+1)}(\cdot)}{\hat{\theta}_{2(i+1)}}, \\ \hat{\mu}_{(i+1)}(\cdot) &= (1, \hat{m}_{(i+1)}(\cdot), \hat{q}_{(i+1)}(\cdot)) \hat{\beta}_{(i+1)}. \end{cases}$$

The iterations for $\hat{m}(\cdot)$ and $\hat{q}(\cdot)$ above were derived by the parametric relation (3). Experience told us that, instead of the equation (6), the relation (1) could be used to build an alternative simple initial value for estimating $\mu(\cdot)$ in the algorithm.

5 Asymptotic properties and rates of convergence

This section needs additional assumptions:

(C1) The kernel function $K(\cdot)$ is supposed to have compact support and to satisfy

$$\text{Supp}(K) = [-1, 1], \quad K(-u) = K(u), \quad \sup_u |K(u)| \leq c < \infty, \quad \int K(u) du = 1,$$

$$\int uK(u)du = 0 \quad \text{and} \quad I_2(K) \equiv \int u^2K(u)du \neq 0.$$

(C2) Basic requirements for the sequence of bandwidths are: $h \rightarrow 0$ and $nh \rightarrow \infty$, as $n \rightarrow \infty$.

Furthermore, for uniform convergence of $\hat{m}(\cdot)$, $\hat{q}(\cdot)$ and $\hat{\mu}(\cdot)$, we need the following condition:

(C3) For some $\zeta \in (0, 1]$, it holds that, for all sequences α_n ,

$$\alpha_n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty, \quad \text{and} \quad \int |K(v + \alpha_n) - K(v)|dv = O(\alpha_n^\zeta).$$

This condition is for instance satisfied if $K(\cdot)$ is Lipschitz continuous of order ζ on \mathbb{R} except at a finite number of points where $K(\cdot)$ could have a discontinuity.

For uniform considerations, we will need a further assumptions on the bandwidth sequence $h = h(n)$.

(C4) Let s be a constant as in **(S3)**, and let r be another constant such that

$$2 < r < s, \quad \text{with} \quad \lim_{n \rightarrow \infty} \inf \left(\frac{nh}{\log n} \right)^{1/2} n^{-2/r} > 0.$$

(C5) $nh^2(\log n)^2 \rightarrow \infty$, $nh \rightarrow \infty$, $\sqrt{nh^2} \rightarrow d^*$ for a constant d^* with $0 \leq d^* < \infty$.

Consider first uniform convergence of the nonparametric estimates $\hat{m}(\cdot)$, $\hat{q}(\cdot)$ and $\hat{\mu}(\cdot)$ defined by the equations (4), (5) and (6).

Theorem 1. *Under **(S1)–(S6)** and **(C1)–(C4)**,*

$$\sup_{x \in I} |\hat{m}(x) - m(x)| = O_P \left(\left[\frac{\log n}{nh} \right]^{1/2} + h^2 \right) \quad \text{where} \quad I \subset (0, 1). \quad (9)$$

Theorem 2. *Under **(S1)–(S6)** and **(C1)–(C4)**,*

$$\sup_{x \in I} |\hat{q}(x) - q(x)| = O_P \left(\left[\frac{\log n}{nh} \right]^{1/2} + h^2 \right) \quad \text{where} \quad I \subset (0, 1). \quad (10)$$

Theorem 3. Under (S1)–(S6) and (C1)–(C4),

$$\sup_{x \in I} |\hat{\mu}(x) - \mu(x)| = O_P \left(\left[\frac{\log n}{nh} \right]^{1/2} + h^2 \right) \quad \text{where } I \subset (0, 1). \quad (11)$$

Some comments.

- Lemmas 1 and 2 (given in the Appendix) lead to the following uniform rates; analogous result for $\hat{m}(\cdot)$ under nonparametric model is in Müller and Stadtmüller (1987) and Müller and Zhao (1995).
- These theorems provide consistency of the nonparametric parts, with the usual nonparametric rates of convergence. Under somewhat stricter moment conditions, the results (9), (10) and (11) can be modified to yield almost sure convergence results with the same rate.

Moreover, we have the asymptotic normality for the nonparametric estimators $\hat{m}(\cdot)$, $\hat{q}(\cdot)$ and $\hat{\mu}(\cdot)$. Let $\Phi(\cdot)$ be the standard normal distribution function, and from now on, for any function v , v'' denotes its second derivative.

Theorem 4. Under (S1)–(S6) and (C1)–(C4), with random design,

$$P \left(\frac{\hat{m}(x) - m(x) - \ell_m(x)h^2}{\sqrt{\tau_m^2(x)/(nh)}} \leq t \mid X_1, \dots, X_n \right) = \Phi(t) + o_P(1), \quad (12)$$

where $\ell_m(x) = (1/2)I_2(K)m''(x)$ and $\tau_m^2(x) = \frac{\int K^2(u)du}{g(x)}\sigma^2(x)$. The dominated convergence theorem implies that, unconditionally,

$$P \left(\frac{\hat{m}(x) - m(x) - \ell_m(x)h^2}{\sqrt{\tau_m^2(x)/(nh)}} \leq t \right) = \Phi(t) + o(1).$$

The conditional asymptotic normality also suggests that such a result holds for fixed design with design points $\{x_j = G^{-1}(j/n), j = 1, \dots, n\}$ and $G'(\cdot) = g(\cdot)$.

Theorem 5. Under (S1)–(S6) and (C1)–(C4), with random design,

$$P \left(\frac{\hat{q}(x) - q(x) - \ell_q(x)h^2}{\sqrt{\tau_q^2(x)/(nh)}} \leq t \mid X_1, \dots, X_n \right) = \Phi(t) + o_P(1), \quad (13)$$

where $\ell_q(x) = (1/2)I_2(K)q''(x)$ and $\tau_q^2(x) = \frac{\int K^2(u)du}{g(x)} \frac{1}{4f^2(q(x)|x)}$.

Theorem 6. Under (S1)–(S6) and (C1)–(C4), with random design,

$$P \left(\frac{\hat{\mu}(x) - \mu(x) - \ell_{\mu}(x)h^2}{\sqrt{\tau_{\mu}^2(x)/(nh)}} \leq t \mid X_1, \dots, X_n \right) = \Phi(t) + o_P(1), \quad (14)$$

where $\ell_{\mu}(x) = (1/2)I_2(K)\mu''(x)$ and

$$\tau_{\mu}^2(x) = \frac{\int K^2(u)du}{g(x)} \left(\frac{9}{4f^2(q(x)|x)} + 4\sigma^2(x) - \frac{3\sigma^2(x)}{f(q(x)|x)} \right).$$

Some comments.

- Theorem 6 shows that the median function has effect on the asymptotic variance of mode function estimation through the conditional distribution density function, and mean function has effect on the asymptotic variance of mode function estimation only if the regression model is really a heteroscedastic one, but both mean and median functions have no effect on the asymptotic bias.
- The unconditional results of theorems 5 and 6 are the same as that of Theorem 4.

Turning now to the parameter estimates $\hat{\beta}$, the following asymptotic normality result establishes parametric rates of convergence. It provides the basic tool to check if the peak of the model depends on the average qualities such as the mean and median in our semiparametric variance function model. The derivation which is given in the Appendix requires the uniform convergence results given by Theorems 1, 2 and 3.

Remark. Although all the previous results are known, we prove them in lighter and more compact way. Next theorems are quite new.

Recall that $u(\cdot)$ is a weight function introduced in equation (8). We have:

Theorem 7. Under (S1)–(S6) and (C1)–(C5),

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \rightarrow_d N(\Sigma_0^{-1}\eta, \Sigma),$$

where $\Sigma = \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$ and $\eta = (\eta_0, \eta_1, \eta_2)^T$, $\Sigma_0 = (\rho_{i,j})_{0 \leq i, j \leq 2}$ and the 3×3 diagonal matrix $\Sigma_1 = 3R(K)\text{diag}(\tau_0, \tau_1, \tau_2)$ with $R(K) = \int K^2(t)dt$; and with

$$\begin{aligned}
\eta_0 &= \frac{1}{2} d^* I_2(K) \left(\int_0^1 g(t)u(t)\mu''(t)dt - \theta_1 \int_0^1 g(t)u(t)m''(t)dt \right. \\
&\quad \left. - \theta_2 \int_0^1 g(t)u(t)q''(t)dt \right); \\
\eta_1 &= \frac{1}{2} d^* I_2(K) \left(\int_0^1 g(t)m(t)u(t)\mu''(t)dt - \theta_1 \int_0^1 g(t)m(t)u(t)m''(t)dt \right. \\
&\quad \left. - \theta_2 \int_0^1 g(t)m(t)u(t)q''(t)dt \right); \\
\eta_2 &= \frac{1}{2} d^* I_2(K) \left(\int_0^1 g(t)q(t)u(t)\mu''(t)dt - \theta_1 \int_0^1 g(t)q(t)u(t)m''(t)dt \right. \\
&\quad \left. - \theta_2 \int_0^1 g(t)q(t)u(t)q''(t)dt \right); \\
\rho_{0,0} &= \int_0^1 g(t)u(t)dt; \\
\rho_{0,1} &= \rho_{1,0} = \int_0^1 g(t)u(t)m(t)dt; \\
\rho_{0,2} &= \rho_{2,0} = \int_0^1 g(t)u(t)q(t)dt; \\
\rho_{1,1} &= \int_0^1 g(t)u(t)m^2(t)dt; \\
\rho_{1,2} &= \rho_{2,1} = \int_0^1 g(t)u(t)m(t)q(t)dt; \\
\rho_{2,2} &= \int_0^1 g(t)u(t)q^2(t)dt; \\
\tau_0 &= \int_0^1 u(t)^2 \frac{\sigma^2(t)}{f(q(t)|t)} dt; \\
\tau_1 &= \int_0^1 u(t)^2 m^2(t) \frac{\sigma^2(t)}{f(q(t)|t)} dt; \\
\tau_2 &= \int_0^1 u(t)^2 q(t)^2 \frac{\sigma^2(t)}{f(q(t)|t)} dt.
\end{aligned}$$

Remark. $d^* = 0$ in (C5) corresponds to further undersmoothing of nonparametric curve estimates for the purpose of estimating the parametric part β . Note also that the asymptotic bias and covariance matrices depend on the unknown parameter β .

6 Testing

The mode of some distributions is identical or approximately equal to mean or/and median. For example, for normal probability distribution $N(\mu, \sigma^2)$, $median = mean = mode = \mu$. This is generally true for any symmetric and unimodal distribution such as the Student's t -distribution. But for Binomial distri-

bution $B(m, \pi)$, the mode is approximately equal to mean but not to the median, as $\text{mean} = m\pi$, $\text{mode} = [(m+1)\pi]$, but median is found from $P(Y \leq \text{median}) = 1/2$ and $P(Y \geq \text{median}) = 1/2$. This is also true for Poisson distribution and hypergeometric distribution. However, the mode of most of distributions links to both mean and median as in (3). On the other hand, for the Fisher's Z distribution, which is defined as $Z = \frac{1}{2} \log Y$ with $Y \sim F_{\nu_1, \nu_2}$, Z has non-zero mean and median, the mode of Z is always zero for any degrees of freedom. This means that the mode of Fisher's Z distribution is independent of the mean and the median. Therefore, in (3) we are concerned with the following hypotheses:

$$H_0 : \theta_1 = 0, H_0 : \theta_2 = 0 \text{ and } H_0 : \theta_1 = \theta_2 = 0.$$

In this section, we develop a data-dependent test statistic whose distribution under the null hypotheses above is derived. All the three different tests mentioned above belong to the linear parametric testing, and we can write the linear testing as

$$H_0 : \Lambda\beta = 0,$$

where Λ is a linear operator. It follows under H_0 that,

$$\sqrt{n}\Lambda\hat{\beta} \rightarrow_d N(0, \Lambda\Sigma\Lambda^T),$$

where Σ is defined in Theorem 7.

To find a data-based test statistic, we still need to estimate Σ . To do that, we use sample-based sums to estimate the elements of these covariance matrix, which are some integrals. We can still prove that $\hat{\Sigma} \rightarrow \Sigma$.

Let Π be an $(m \times 3)$ -matrix of rank m , $m \leq 3$, and let ς_0 and ς_1 be m -vectors. To test

$$H_0 : \Pi\beta = \varsigma_0,$$

we consider the test statistic

$$T_n = n(\Pi\hat{\beta} - \varsigma_0)^T (\Pi\hat{\Sigma}\Pi^T)^{-1} (\Pi\hat{\beta} - \varsigma_0).$$

Theorem 8. *Under (S1) – (S6) and (C1) – (C5), it holds that, under the null hypothesis H_0 ,*

$$T_n \rightarrow_d \chi(m)^2,$$

where $\chi(m)^2$ is the central χ^2 distribution with m degrees of freedom.

Assume that the alternatives $H_1 : \Pi\beta = \varsigma_1$ satisfy

$$n(\varsigma_1 - \varsigma_0)^T (\Pi\Sigma\Pi^T)^{-1} (\varsigma_1 - \varsigma_0) \rightarrow \lambda^2,$$

for a fixed real constant λ . Then, under H_1 ,

$$T_n \rightarrow_d \chi^2(m, \lambda^2),$$

where $\chi^2(m, \lambda^2)$ is the noncentral χ^2 distribution with m degrees of freedom and noncentrality parameter λ^2 .

Consequence. The application to the construction of a level- α test is: reject H_0 if $T_n > \chi_\alpha^2(m)$, the 100(1- α) quantile of the corresponding χ^2 distribution.

Appendix: Some lemmas and proofs of Theorems

Proof of Theorem 1. Theorem 1 can be proved along the lines of Theorem 4.1 of Müller and Zhao (1995). In fact Lemma 6.1 of Müller and Zhao (1995) is true here and is still specified as Lemma 1. We may need to use the following facts throughout the proofs.

For any Lipschitz-continuous function u defined on $[0, 1]$,

$$\frac{1}{n} \sum_{i=1}^n u(X_i) = \int_0^1 g(t)u(t)dt + O\left(\frac{1}{n}\right). \quad (A1)$$

Analogously

$$\frac{1}{n} \sum_{i=1}^n u(X_i)K\left(\frac{X_i - x}{h}\right) = \int_0^1 g(t)u(t)K\left(\frac{x - t}{h}\right) dt + O\left(\frac{1}{n}\right). \quad (A2)$$

Lemma 1. Under (S1)–(S4), (C1)–(C4), for $x \in (0, 1)$,

$$\hat{m}(x) - m(x) = \left[\frac{1}{2}h^2 I_2(K)m^{(2)}(x) + O\left(\frac{1}{nh}\right) + o(h^2) \right] + \sum_{i=1}^n \frac{w_i(x)}{\sum_{j=1}^n w_j(x)} \epsilon_i, \quad (A3)$$

where $w_j(x)$ is the kernel weight function. The proof of this lemma was given by Section 4 of Müller (1988).

Proof of Theorem 2. We need the following Lemma 2 to establish Theorem 2.

Lemma 2. Under (S1)–(S4), (C1)–(C4), for $x \in (0, 1)$,

$$\hat{q}(x) - q(x) = \phi(x) \sum_{i=1}^n \psi(Y_i^*)K\left(\frac{X_i - x}{h}\right) + o(h^2), \quad (A4)$$

where $\psi(y) = 1/2 - \mathbb{I}(y < 0)$, $Y_i^* = Y_i - q(x)$ and $\phi(x) = (f(q(x)|x)g(x))^{-1}$.

This is the Bahadur equation (see Bahadur, 1966), and the proof was also given by Yu and Lu (2004).

Proof of Theorem 3. Theorem 3 is proved by $\hat{\mu}(\cdot) = \hat{\theta}_0 + \hat{\theta}_1 \hat{m}(\cdot) + \theta_2 \hat{q}(\cdot)$, and $\hat{\beta} \rightarrow \beta$.

Proof of Theorems 4 and 5. The proofs of Theorems 4 and 5 are standard by the equation (A3) with Lemma 1 and the equation (A4) with Lemma 2 respectively.

Proof of Theorem 6. Similarly, Theorem 6 can be established through Theorems 4 and 5 and $\hat{\mu} = \hat{\theta}_0 + \hat{\theta}_1 \hat{m} + \theta_2 \hat{q}$, $\hat{\beta} \rightarrow \beta$, and using further the Gram-Charlier expansion, see Kendall, Stuart and Ord (1987), Section 2.11, $\beta \approx (0, -2, 3)$ under large sample or $n \rightarrow \infty$. In particular, asymptotically, from

$$\begin{aligned} \text{Var}(\hat{\mu}(x)) &= \theta_1^2 \text{Var}(\hat{m}(x) - E(\hat{m}(x))) + \theta_2^2 \text{Var}(\hat{q}(x) - E(\hat{q}(x))) \\ &\quad - 2\theta_1\theta_2 \text{Cov}(\hat{m}(x) - E(\hat{m}(x)), \hat{q}(x) - E(\hat{q}(x))), \end{aligned}$$

we could obtain the asymptotic variance of Theorem 6.

Proof of Theorem 7. Write $S = \hat{Z} - Z$ and $V = \hat{\mathbb{X}} - \mathbb{X}$, then Theorems 1 to 6 indicate $S = [O_p(h^2)]$ and $V = [O_p(h^2)]$.

Lemma 3. As $n \rightarrow \infty$,

$$\frac{1}{n} \mathbb{X}^T Q^{-1} \mathbb{X} \rightarrow \Sigma_0 \quad \text{and} \quad \frac{1}{n} \hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}} \rightarrow_p \Sigma_0.$$

If we write $\hat{\beta} = (\hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}})^{-1} \hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}} \beta$, then

$$\sqrt{n}(\hat{\beta} - \beta) = (\hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}})^{-1} (\hat{\mathbb{X}}^T Q^{-1}) (\hat{Z} - \hat{\mathbb{X}} \beta) = (\hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}})^{-1} (\hat{\mathbb{X}}^T Q^{-1}) (S - V \beta).$$

Hence, $\sqrt{n}(\hat{\beta} - \beta) = (n^{-1} \hat{\mathbb{X}}^T Q^{-1} \hat{\mathbb{X}})^{-1} (n^{-1/2} \hat{\mathbb{X}}^T Q^{-1}) S - n^{-1/2} \hat{\mathbb{X}} Q^{-1} V \beta$.

Observe that $n^{-1/2} \mathbb{X}^T Q^{-1} S - n^{-1/2} \mathbb{X}^T Q^{-1} V \beta = (\xi_0, \xi_1, \xi_2)^T$, where

$$\begin{aligned} \xi_0 &= \sum_{i=1}^n u(X_i) (\hat{\mu}(X_i) - \mu(X_i)) - \theta_1 \sum_{i=1}^n u(X_i) (\hat{m}(X_i) - m(X_i)) \\ &\quad - \theta_2 \sum_{i=1}^n u(X_i) (\hat{q}(X_i) - q(X_i)), \\ \xi_1 &= \sum_{i=1}^n m(X_i) u(X_i) (\hat{\mu}(X_i) - \mu(X_i)) - \theta_1 \sum_{i=1}^n m(X_i) u(X_i) (\hat{m}(X_i) - m(X_i)) \\ &\quad - \theta_2 \sum_{i=1}^n m(X_i) u(X_i) (\hat{q}(X_i) - q(X_i)), \\ \xi_2 &= \sum_{i=1}^n q(X_i) u(X_i) (\hat{\mu}(X_i) - \mu(X_i)) - \theta_1 \sum_{i=1}^n q(X_i) u(X_i) (\hat{m}(X_i) - m(X_i)) \\ &\quad - \theta_2 \sum_{i=1}^n q(X_i) u(X_i) (\hat{q}(X_i) - q(X_i)). \end{aligned}$$

Also, note that the 6 different elements of the symmetric matrix $X^T Q^{-1} X$ are

given by

$$\sum_{i=1}^n u(X_i), \quad \sum_{i=1}^n u(X_i)m(X_i), \quad \sum_{i=1}^n u(X_i)q(X_i), \quad \sum_{i=1}^n u(X_i)m(X_i)^2, \\ \sum_{i=1}^n u(X_i)m(X_i)q(X_i) \quad \text{and} \quad \sum_{i=1}^n u(X_i)q(X_i)^2, \quad \text{respectively.}$$

Using Lemmas 1 and 2 and **(S5)**, we can replace all $\hat{m} - m$, $\hat{q} - q$ and $\hat{\mu} - \mu$ by a linear combination of normal variables in ξ_0 , ξ_1 and ξ_2 . Then compute the asymptotic bias and asymptotic variance of random vector $(\xi_0, \xi_1, \xi_2)^T$ to obtain the vector η and matrix Σ_1 respectively, so to establish Theorem 7 immediately. For example, the variance of ξ_0 could be simplified into

$$-4\theta_1\theta_2 \text{Cov} \left(\sum_{i=1}^n u(X_i)(\hat{m}(X_i) - m(X_i)); \sum_{i=1}^n u(X_i)(\hat{q}(X_i) - q(X_i)) \right),$$

which is asymptotically equal to $3R(K) \int_0^1 u(t)^2 \frac{\sigma^2(t)}{f(q(t)|t)} dt$.

Proof of Theorem 8. The proof of Theorem 8 is easy. In fact, Theorem 7 implies that

$$n^{1/2}(\Pi\hat{\Sigma}\Pi^T)^{-1/2}(\Pi\beta - \varsigma_0) \rightarrow N(0, I) \quad \text{under } H_0,$$

which implies part 1 of Theorem 8, whereas under H_1

$$n^{1/2}(\Pi\hat{\Sigma}\Pi^T)^{-1/2}(\Pi\beta - \varsigma_0) \rightarrow N(\rho, I)$$

where $\|\rho\| = \lambda$,

$$n^{1/2}(\Pi\hat{\Sigma}\Pi^T)^{-1/2}(\Pi\beta - \varsigma_0) = \\ n^{1/2}(\Pi\hat{\Sigma}\Pi^T)^{-1/2}(\Pi\beta - \varsigma_1) + n^{1/2}(\Pi\hat{\Sigma}\Pi^T)^{-1/2}(\Pi\beta - \varsigma_0) \rightarrow N(\rho, I).$$

References

Ahmad, I. A. and Ran, I. S., (2004) Selection of smoothing parameters via kernel contrasts, *J. of Nonparametric Statistics*, **16**, 841–877.

Bahadur, R.R. (1966) A note on quantiles in large samples. *Ann. Math. Statist.*, **37**, 577–580.

Berlinet, A., Gannoun, A., Matzner-Lober, E. (1998) Normalité asymptotique d'estimateurs convergents du mode conditionnel. *La Revue Canadienne de Statistique*, **26**, 365–380.

- Berlinet, A., Gannoun, A., Matzner-Lober, E. (2001) Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*. **35**, 139–169.
- Bowman, A. W., (1984) An alternative method of cross-validation for the smoothing density estimates. *Biometrika*, **71**, 353–360.
- Bickel, P. J. and Fan, J. (1996) Some problems on the estimation of unimodal densities. *Statistica Sinica*, **6**, 23–45.
- Birgè, L. (1997) Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* , **25**, 970–981.
- Cheng, M., Peng, L., (2002) Regression modeling for nonparametric estimation of distribution and quantile functions, *Statistica Sinica*, **12**, 1043–1060.
- Eddy, W. F. (1980) Optimal kernel estimators of the mode. *Ann. Statist.*, **8**, 870–882.
- Gasser, T., Hall, P., Presnell, B. (1998) Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc.*, **60**, 681–691.
- Fan, J. and Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *Ann. Statist.*, **20**, 2008–2036.
- Fan, J. and Gijbels, I. (1996) *Local polynomial modelling and its applications*. London : Chapman & Hall.
- Grenander, U. (1956) On the theory of mortality measurement, Part II. *Skand. Akt.*, **39**, 125–153.
- Grund, B. and Hall, P. (1995) On the minimisation of L^P error in mode estimation. *Ann. Statist.*, **23**, 2264–2284.
- Hall, P., Marron, J. S. and Park, B. U., (1992) Smoothed cross validation. *Probability Theory Related Fields*, **90**, 149–173.
- Hall, P., Sheather, S. J., Jones, M. C. and Marron, J. S., (1991), On optimal databased bandwidth selection in kernel density estimation. *Biometrika*, **78**, 263–269.
- Hedges, S.B. and Shah, P. (2003) Comparison of mode estimation methods and application in molecular clock analysis. *BMC Bioinformatics* , **4**, 1471–2105.
- Jones, M. C., Marron, J. S. and Sheather, S. J., (1996) A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, **91**, 401–407.
- Kendall ,M.G., Stuart,A. and Ord, J.K. (1987) *Kendall's advanced theory of statistics*. New York: Oxford University Press.

- Lee, M.J. (1989) Mode Regression. *J. Econometrics*, **42**, 337–349.
- Lee, M.J. (1993) Quadratic mode regression, *J. Econometrics*, **57**, 1–19.
- Lee, P. M. (1994) *Bayesian Statistics: an introduction*. Oxford University Press, New York.
- Meyer, M. C. (2001) An alternative unimodal density estimator with a consistent estimate of the mode. *Statistica Sinica*, **11**, 1159–1174.
- Müller, H.G. and Stadtmüller, U.(1987) Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, **15**, 610–625.
- Müller, H.G. (1988) Nonparametric Regression Analysis of Longitudinal Data. *Lecture Notes in Statistics*, **46**, Springer-Verlag, New York.
- Müller, H.G. and Zhao, P. (1995) On a semiparametric variance function model and a test for heteroscedasticity. *Ann. Statist.*, **23**, 946–967.
- Parzen, E. (1962) On the estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
- Rudemo, M., (1982) Empirical Choice of histograms and kernel density estimation. *Scand. Journal of Statistics*, **9**, 65–78.
- Samanta, M. (1989) Non-Parametric Estimation of Conditional Quantiles. *Statist. Prob. Letters*, **7**, 407–412.
- Samanta, M. (1989) Non-Parametric Estimation of Conditional Quantiles. *Statist. Prob. Letters*, **7**, 407–412.
- Samanta, M., Thavaneswarana, A. (1990) Non-Parametric Estimation of Conditional Mode. *Comm. Stat. Theo. and Method.*,**19**, 4515–4524.
- Scott, D. W. and Terrel, G. R., (1987) Biased and unbiased cross validation in density estimation. *J. Amer. Statist. Assoc.*, **82**, 1131–46.
- Sheather, S. J., Jones, M. C., (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc.*, **B 53**, 683–690.
- Shoung J. and Zhang C. (2001) Least squares estimators of the mode of a unimodal regression function. *Ann. Statist.*, **29**, 648–665.
- Silverman, B. W. (1986) *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Stuart, A. and Ord, J.K. (1994) *Kendall's advanced theory of statistics*. Vol. I, 6th edn. London.

Wand, M.P., and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

Yu, K., Lu, Z. (2004) Local linear additive quantile regression. *Scand. J. Statist.*, **31**, 333–346.

Yu, K., Jones, M.C. (1998) Local linear regression quantile estimation. *J. Amer. Statist. Assoc.*, **93**, 228–238.

Ziegler, K. (2002) On nonparametric kernel estimation of the mode of the regression function in the random design model. *J. Nonparametr. Statist.*, **14**, 749–774.

Ziegler, K. (2003) On the asymptotic normality of kernel regression estimators of the mode in the random design model. *J. Statist. Plann. Inf.*, **115**, 123–144.