One-Sided Cross-Validation in Local Exponential Regression

Caixia Zhao and Jeffrey D. Hart*

Key Words and Phrases: Prequential model choice; Local likelihood; Data-driven smoothing; Mean average squared error; Optimal window widths.

1 Introduction

Regression data for which the response has an exponential distribution arise in two fundamentally important problems: probability density estimation and spectral estimation. A little used means of obtaining a density estimate from independent and identically distributed observations is to smooth sample spacings, which are approximately exponentially distributed

Consider regression data (x_i, Y_i), $i = 1, ..., n$, where the x_i's are known constants and Y_i has the exponential distribution with mean $r_0(x_i)$, $i = 1, ..., n$. Suppose we wish to estimate r_0, but have no parametric model for it. Our approach will be to model r_0 locally by a function of the form

$$\exp(a + bx)$$

where the exponential function is used to ensure positivity of our estimate.

Given a candidate r for r_0, the log-likelihood function is

$$l(r) = -\sum_{i=1}^{n}[\log r(x_i) + Y_i/r(x_i)], \tag{1}$$

In the absence of a global parametric model, it makes sense to use local parametric models for r_0. Given a value of x, say x_0, let us assume that

$$r_0(x) \approx \exp(a + bx), \quad x \in (x_0 - h, x_0 + h),$$

*_Address for correspondence_: Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
Email: hart@stat.tamu.edu

for constants a and b (which depend on x_0) and some small positive number h. A local version of the log-likelihood function takes the form

$$l_{x_0}(a,b) = -\sum_{i=1}^{n} K\left(\frac{x_0 - x_i}{h}\right)[a + bx_i + Y_i \exp(-(a + bx_i))] \tag{2}$$

for some density K that is symmetric about 0 and unimodal. An estimate of $r_0(x_0)$ is found by choosing the values of a and b that maximize $l_{x_0}(a,b)$. Such an approach has been investigated by various authors, including Tibshirani and Hastie(1987), Staniswalis(1989), Chaudhuri and Dewanji(1995), Fan, Heckman and Wand(1995) and Aerts and Claeskens(1997).

An omnipresent problem with local methods is deciding how local to make the estimation. In the method just presented this amounts to deciding on an appropriate value for h, which we shall call the *window width*. The purpose of this report is to propose and analyze a data-driven method of selecting a window width in the exponential regression model. Many methods exist for selecting regression smoothing parameters, principal among them cross-validation and plug-in. Another, less well-known, method is that of prequential analysis (Dawid (1984), Modha and Masry(1998)). The prequential approach is based on the notion of selecting a model that yields good predictions of future observations. It is similar to cross-validation, but bases a prediction of Y_j only on the previous data $Y_1, ... Y_{j-1}$ as opposed to the data $Y_1, ..., Y_{j-1}, Y_{j+1}, ..., Y_n$. The results in this report are further support for the work of Hart and Yi(1998) that shows the prequential approach to be not simply an alternative to cross-validation, but rather a more efficient alternative.

Prequential analysis is applicable when the data can be ordered with respect to time or some other variable, such as a regressor. We shall describe the prequential method in our regression context and order the data with respect to the design variable x. Without loss of generality we assume that $x_1 < x_2 < ... < x_n$. Now, let $D(m, Y_i)$ be a measure of discrepancy between the number m and the response Y_i. An overall measure of how well the curve r fits the data is

$$\frac{1}{n}\sum_{i=1}^{n} D(r(x_i), Y_i).$$

For example,

$$D(r(x_i), Y_i) = (Y_i - r(x_i))^2,$$

and

$$D(r(x_i), Y_i) = -\log f(Y_i; r(x_i)),$$

where $f(\cdot;r)$ is a function that is known up to r. The former example is an all-purpose measure and the latter is likelihood based, being appropriate in our exponential regression model.

In this project, we investigate use of the prequential method for selecting the window width of local estimators of a regression curve when the responses have an exponential distribution.We shall make comparisons among several different methods: ordinary cross-validation, and two versions of the prequential method.

2 DESCRIPTION OF METHODOLOGY

2.1 Squared error methods

The observed data $(x_i, Y_i), i = 1, \ldots, n$ are independent and such that Y_i has an exponential distribution with mean $r(x_i), i = 1, \ldots, n$. We will estimate r using the local linear scheme described in Section 1. The estimate with window width h is denoted $\hat{r}_h(x)$.

The choice of window width h is crucial to the performance of estimate $\hat{r}_h(x)$. A straightforward method of window width selection is cross-validation(Stone 1974), the idea of which is to use a part of the data to construct an estimate of the regression model and then to predict the rest of the data with this estimate. The most often used form of cross-validation is the "least squares leave-one-out" cross-validation criterion:

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - \hat{r}_h^i(x_i)\right]^2, \tag{3}$$

where $\hat{r}_h^i(x)$ denotes an estimate of r computed without the ith data point.

Another version of cross-validation is to use only the "previous" data points $(x_1, Y_1), \ldots, (x_{(i-1)}, Y_{(i-1)})$ to construct an estimate $\tilde{r}_h^i(x_i)$ of $r(x_i)$. The corresponding criterion is

$$OSCV(h) = \frac{1}{n-m} \sum_{i=m+1}^{n} \left(Y_i - \tilde{r}_h^i(x_i)\right)^2, \tag{4}$$

where m is some small integer that is at least 1.

It can be shown that these two versions of cross-validation yield approximately unbiased estimators of MASE$(\hat{r}_h)+\frac{1}{n}\sum_{i=1}^{n} Var(Y_i)$ and MASE$(\tilde{r}_h)+\frac{1}{n-m}\sum_{i=m+1}^{n} Var(Y_i)$, respectively, where

$$MASE(\hat{r}_h) = E\left\{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{r}_h(x_i) - r(x_i)\right)^2\right\}, \tag{5}$$

and

$$MASE(\tilde{r}_h) = E\left\{\frac{1}{n-m}\sum_{i=m+1}^{n}\left(\tilde{r}_h(x_i) - r(x_i)\right)^2\right\}. \tag{6}$$

As $n \longrightarrow \infty$, the asymptotic minimizer, h_n of MASE for $\hat{r}_h$ is

$$h_n = C_K C_{model} n^{-1/5},$$

where C_K is a constant that depends only on the kernel K used in the local linear estimator(Fan 1992). The asymptotic minimizer, b_n, of MASE for $\tilde{r}_h$ is identical to h_n except that it has a different constant C_L in place of C_K. This implies that

$$\frac{h_n}{b_n} \longrightarrow \frac{C_K}{C_L},$$

which motivates the following definition of a window width for use in $\hat{r}_h$:

$$\hat{h}_{oscv} = \frac{C_K}{C_L}\hat{b} = M\hat{b}. \tag{7}$$

where $\hat{b}$ minimizes $OSCV(h)$.

2.2 Likelihood methods

We now describe CVLI and OSLI, which are criteria based on a log-likelihood function. Define

$$CVLI(h) = -\sum_{i=1}^{n}\log f(Y_i; \hat{r}_h^i(x_i)) \tag{8}$$

where $f(y;a) = \frac{1}{a}\exp(\frac{-y}{a})I_{(0,\infty)}(y)$ and $\hat{r}_h^i(x_i)$ denotes an estimate of r computed without the ith data point. Define also

$$OSLI(h) = -\sum_{i=m+1}^{n}\log f(Y_i; \tilde{r}_h^i(x_i)), \tag{9}$$

where $\tilde{r}_h^i(x_i)$ is an estimate using only the "previous" data points, and m is some small integer that is at least 1. The quantities CVLI(h) and OSLI(h) measure the discrepancy between the data and their estimates.

In Section 2.4, we will give a proof that these two likelihood criteria yield asymptotically unbiased estimators of MWASE($\hat{r}_h$)+C_1 and MWASE($\tilde{r}_h$)+C_2, respectively, where

$$MWASE(\hat{r}_h) = E\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{r}_h(x_i) - r(x_i))^2}{r^2(x_i)}\right\}, \tag{10}$$

$$MWASE(\tilde{r}_h) = E\left\{\frac{1}{n-m}\sum_{i=m+1}^{n}\frac{(\tilde{r}_h(x_i) - r(x_i))^2}{r^2(x_i)}\right\}, \tag{11}$$

C_1 and C_2 are constants independent of h. It will be shown that as $n \longrightarrow \infty$,

$$\frac{h_n}{b_n} \longrightarrow \frac{C_K}{C_L},$$

where h_n and b_n are the asymptotic minimizers of MWASE for $\hat{r}_h$ and $\tilde{r}_h$, respectively, C_K and C_L are constants that depend only on the kernel K.

This motivates the following definition of a window width for use in $\hat{r}_h$:

$$\hat{h}_{osli} = \frac{C_K}{C_L}\hat{b} = M\hat{b}. \tag{12}$$

where $\hat{b}$ is the minimizer of $OSLI(h)$. The constant M is the same as in Section 2.1 so long as the same kernel is used for the least squares and likelihood methods.

2.3 Computation of estimators

The estimate $\hat{r}_h(x)$ has the form $\exp(\hat{a}+\hat{b}x)$, where $\hat{a}$ and $\hat{b}$ maximize the local log-likelihood function

$$l_x(a,b) = -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)[a + bx_i + Y_i\exp(-(a+bx_i))]. \tag{13}$$

The Newton-Raphson method is used to find the maximizer, and the gradient and Hessian matrices are given by

$$G = \begin{bmatrix} -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)[1 - Y_i\exp(-(a+bx_i))] \\ -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)x_i[1 - Y_i\exp(-(a+bx_i))] \end{bmatrix},$$

$$H = \begin{bmatrix} -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)Y_i\exp(-(a+bx_i)) & -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)x_iY_i\exp(-(a+bx_i)) \\ -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)x_iY_i\exp(-(a+bx_i)) & -\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right){x_i}^2Y_i\exp(-(a+bx_i)) \end{bmatrix}.$$

To obtain initial estimates of a and b, we do a log-transformation of the data. The transformed data follow the model

$$\log Y_i - EZ_i = \log r(x_i) + Z_i - EZ_i, \quad i = 1,\ldots,n,$$

where Z_i is the logarithm of a r.v. having exp(1) distribution and $\mathrm{E}(Z_i) = -0.577216$. We compute an ordinary local linear estimate of $\log r(x)$ based on data $\log Y_1 + 0.577216, \ldots, \log Y_n + 0.577216$, and use the resulting intercept and slope estimates as our initial estimates of a and b, respectively.

The OSCV and OSLI methods use one-sided estimates $\tilde{r}_h(x)$. The estimates are computed in a similar way as follows:

- The estimate $\tilde{r}_h(x)$ maximizes the local log-likelihood function

$$l_x(a,b) = -\sum_{i \in S_x} K\left(\frac{x - x_i}{h}\right)[a + bx_i + Y_i \exp(-(a + bx_i))], \tag{14}$$

  where $S_x = \{i : x_i \le x\}$.

- The gradient and Hessian for Newton-Raphson are given by

$$G = \begin{bmatrix} -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right)[1 - Y_i \exp(-(a + bx_i))] \\ -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right) x_i[1 - Y_i \exp(-(a + bx_i))] \end{bmatrix},$$

$$H = \begin{bmatrix} -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right) Y_i \exp(-(a + bx_i)) & -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right) x_i Y_i \exp(-(a + bx_i)) \\ -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right) x_i Y_i \exp(-(a + bx_i)) & -\sum_{i \in S_x} K\left(\frac{x-x_i}{h}\right) {x_i}^2 Y_i \exp(-(a + bx_i)) \end{bmatrix}.$$

2.4 Connection of exponential likelihood to weighted square error

Consider the likelihood-based cross-validation criterion:

$$CVLI(h) = -\sum_{i=1}^{n} \log f(Y_i; \hat{r}_h^i(x_i)).$$

Expanding the summand in Taylor series about $r(x_i) = r_i$, we have

$$\begin{aligned} \log f(Y_i; \hat{r}_h^i(x_i)) = \log f(Y_i; r_i) \quad &+ \quad (\hat{r}_h^i(x_i) - r_i)\left[\frac{\partial}{\partial a} \log f(Y_i; a)\right]\Big|_{a=r_i} \\ &+ \quad \frac{1}{2}(\hat{r}_h^i(x_i) - r_i)^2 \left[\frac{\partial^2}{\partial a^2} \log f(Y_i; a)\right]\Big|_{a=\tilde{r}_i}, \end{aligned}$$

where $\tilde{r}_i$ is between r_i and $\hat{r}_h^i(x_i)$. Taking expectation,

$$\begin{aligned} E[\log f(Y_i; \hat{r}_h^i(x_i))] = E[\log f(Y_i; r_i)] \quad &+ \quad E[(\hat{r}_h^i(x_i) - r_i)] E\left\{\left[\frac{\partial}{\partial a} \log f(Y_i; a)\right]\Big|_{a=r_i}\right\} \\ &+ \quad E[\frac{1}{2}(\hat{r}_h^i(x_i) - r_i)^2] E\left\{\left[\frac{\partial^2}{\partial a^2} \log f(Y_i; a)\right]\Big|_{a=\tilde{r}_i}\right\}, \end{aligned}$$

where we use the fact that $\hat{r}_h^i(x_i)$ is independent of Y_i.

We now show that, in general, $E\{\frac{\partial}{\partial a}\log f(Y_i;a)|_{a=r_i}\} = 0$, since

$$\begin{aligned} E\{\frac{\partial}{\partial a}\log f(Y_i;a)|_{a=r_i}\} &= \int \frac{1}{f(y;r_i)}\cdot\frac{\partial f(y;a)}{\partial a}|_{a=r_i} f(y;r_i)dy \\ &= \int \frac{\partial f(y;a)}{\partial a}|_{a=r_i}dy \\ &= \frac{\partial}{\partial a}\int f(y;a)dy|_{a=r_i} \\ &= 0. \end{aligned}$$

The only assumption needed here is the interchange of derivative and integral, which is certainly warranted in the exponential case. Because $f(y;a) = \frac{1}{a}\exp(-\frac{y}{a})I_{(0,\infty)}(y)$, we have

$$E\log f(Y_i;r_i) = -\log r_i - 1$$

and

$$E\left[\frac{\partial^2}{\partial a^2}\log f(Y_i;a)|_{a=r_i}\right] = -\frac{1}{r_i^2}.$$

Under standard conditions ensuring the consistency of $\hat{r}_h^i(x_i)$, we thus have

$$E\left[-\sum_{i=1}^n \log f(Y_i;\hat{r}_h^i(x_i))\right] \doteq \sum_{i=1}^n \log r_i + n + \frac{1}{2}\sum_{i=1}^n \frac{E(\hat{r}_h^i(x_i)-r_i)^2}{r_i^2}.$$

Similarly,

$$E\left[-\sum_{i=m+1}^n \log f(Y_i;\tilde{r}_h^i(x_i))\right] \doteq \sum_{i=m+1}^n \log r_i + n - m + \frac{1}{2}\sum_{i=m+1}^n \frac{E(\tilde{r}_h^i(x_i)-r_i)^2}{r_i^2}.$$

For n large, $E(\hat{r}_h^i(x_i)-r_i)^2 \doteq E(\hat{r}_h(x_i)-r_i)^2$ and $E(\tilde{r}_h^i(x_i)-r_i)^2 \doteq E(\tilde{r}_h(x_i)-r_i)^2$ which completes the proof that the likelihood CV criteria provide asymptotically unbiased estimators of criteria that are equivalent to MWASE.

3 SIMULATION RESULTS

Here we present the results of a simulation study which compares four different methods to estimate the smoothing parameter under several regression settings. In each setting, we calculate $\hat{h}_{osli}$ (OSLI method), $\hat{h}_{oscv}$ (OSCV method), $\hat{h}_{li}$ (CVLI method) and $\hat{h}_{cv}$ (CV method).

We used the following function, where $0 \le x \le 1$:

$$r(x) = 1 + C2^8 x^4 (1-x)^4,$$

where C is some positive constant. Define the signal to noise ratio (SNR) by

$$SNR = \frac{\sup_{0\le x\le 1} r(x) - \inf_{0\le x\le 1} r(x)}{\int r(x)dx}.$$

The motivation for the denominator of SNR is that $r(x)$ is the standard deviation of an observation made at x, and hence the denominator measures average noise level. We chose C to yield the following SNR values: 0.3, 0.65, 1, 1.3, 1.65 and 2.

The sample sizes 50, 100, 150 are used. The design points were $x_i = (i - 0.5)/n$, $i = 1, \ldots, n$. We used the quartic kernel in all simulations, that is $K(u) = (15/16)(1 - u^2)^2 I_{(-1,1)}(u)$, and the constant M in (9) and (14) is 0.557297. Let $E_1, \ldots, E_n$ be a random sample from the exponential distribution with mean 1. The generated data are then

$$y_i = r(x_i)E_i, \quad i = 1, \ldots, n.$$

Define the weighted average squared error(WASE) for a linear estimator $\hat{r}_h$ by

$$WASE(h) = \frac{1}{n}\sum_{i=1}^{n} \frac{\left(\hat{r}_h(x_i) - r(x_i)\right)^2}{r(x_i)^2}.$$

The window width $\hat{h}_o$ which minimizes WASE(h) was approximated for each dataset, and WASE($\hat{h}$) was computed for each of the four data-driven window widths. We conducted 500 replications at each combination of function, C and n. Some results are summarized in Table 1. (See the Appendix for other results). Some typical data and fitted curves are given in Figures 1-3. Our results show that compared with the other methods, OSLI provides a smaller average WASE. A proxy for how close the WASE of an estimate is to the optimal WASE is the quantity $(\hat{h} - \hat{h}_o)^2$, where $\hat{h}$ and $\hat{h}_o$ denote data-driven and WASE-optimal window widths. Note in Table 1 that the average of $(\hat{h}_{osli} - \hat{h}_o)^2$ is smallest. The small variance of $\hat{h}_{osli}$ indicates that the OSLI method usually produces a more stable estimate(see Figure 4). The plots imply that the performance of the OSLI method is much better than the other methods, especially when the sample size is small and the SNR is small.

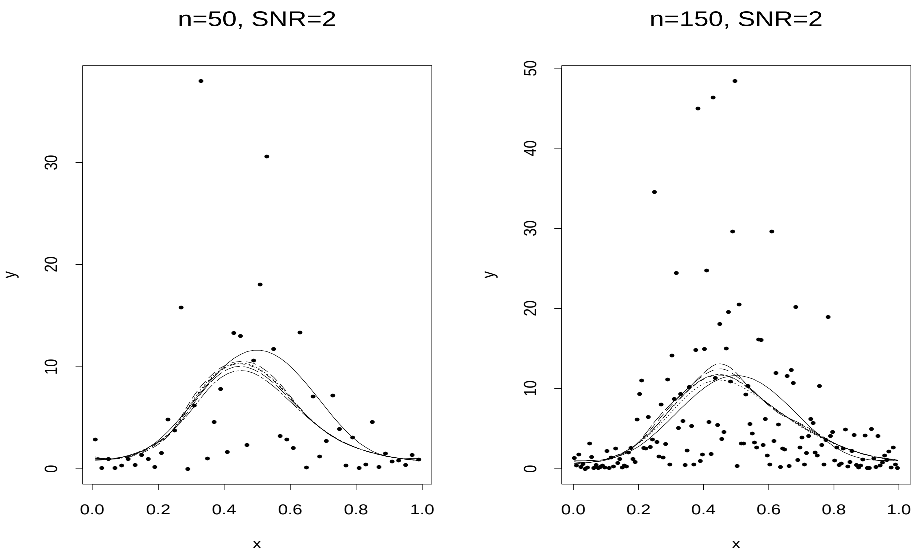

Figure 1: *Estimates of function* $1 + 10.638 \times 2^8 \times x^4(1-x)^4$, $SNR = 2$. *True curve(solid line), ASE(dotted line), OSLI(short dashed line), CVLI(long dashed line),OSCV(long-long dashed line), CV(long-short dashed line).*

TABLE 1: Simulation results for the function $1 + 10.638 \times 2^8 \times x^4(1-x)^4$**,**

		n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
		50	0.30836	0.29423	0.22616	0.23701	0.29783
	Mean($\hat{h}$)	100	0.25388	0.22100	0.19711	0.19887	0.23774
		150	0.22557	0.19479	0.18055	0.18521	0.21760
		50	0.09446	0.11189	0.10426	0.12944	0.08316
	Mean(WASE($\hat{h}$))	100	0.05740	0.06895	0.06084	0.08081	0.05103
$SNR = 2$		150	0.04208	0.04985	0.04396	0.05435	0.03730
		50	0.26477	1.59316	2.86368	1.11058	
	Mean($(\hat{h} - \hat{h_o})^2 \times 10^2$)	100	0.11586	0.63752	0.16099	0.67821	
		150	0.08038	0.39501	0.09927	0.44038	
		50	1.33644	2.75464	1.73374	2.57903	
	Variance ($\hat{h}$) $\times 10^2$	100	0.69564	1.32372	0.90085	1.50682	
		150	0.56636	0.97502	0.73211	1.08408	

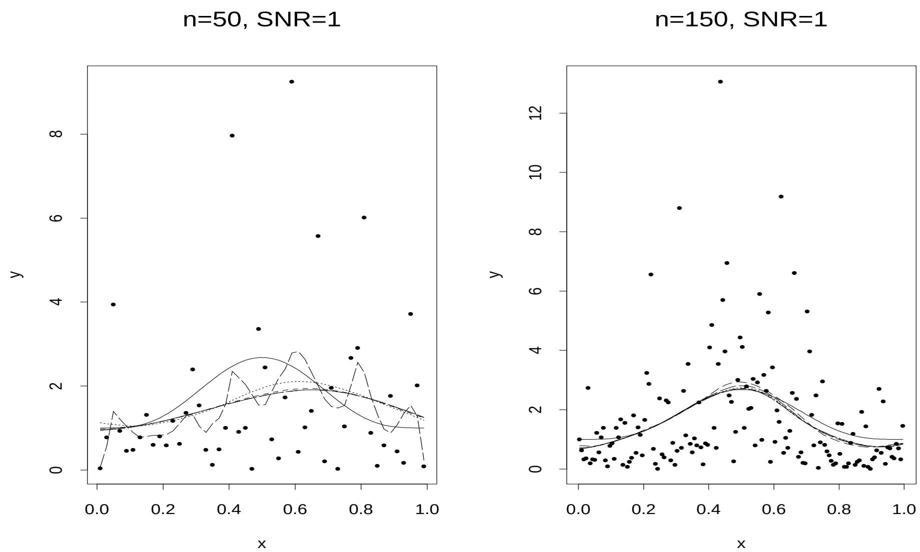

Figure 2: *Estimates of function* $1 + 1.68 \times 2^8 \times x^4(1-x)^4$, $SNR = 1$. *(See Figure 1 for legend)*

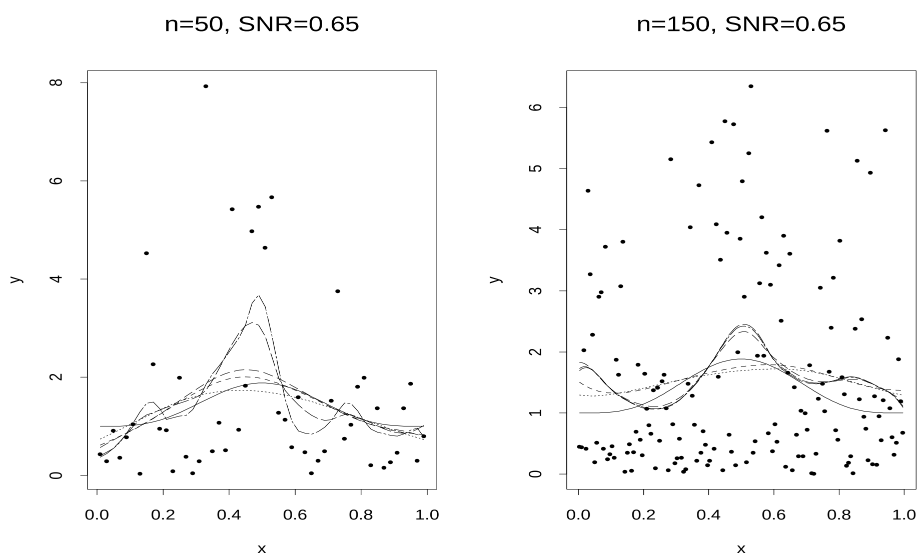

Figure 3: *Estimates of function* $1 + 0.883 \times 2^8 \times x^4(1-x)^4$, $SNR = 0.65$. *(See Figure 1 for legend)*

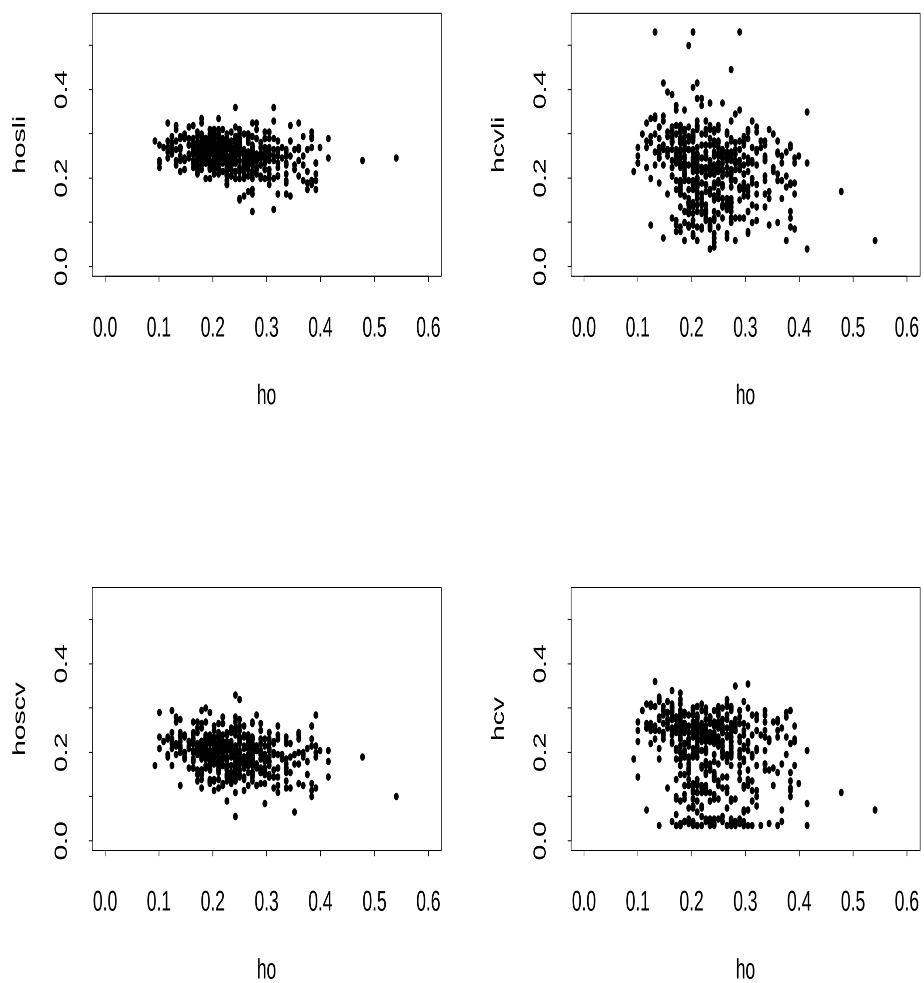

Figure 4: *Scatterplots of window widths from simulation for function* $1+10.638\times 2^8\times x^4(1-x)^4 (SNR=2)$*, Each plot is* $\hat{h}$ *versus* $\hat{h}_o$*.*

4 STUDY OF SPECIAL CASES

4.1 Density Quantile Estimation

Suppose F_X is the cumulative distribution function of a random variable X. The quantile function of X is the left continuous inverse of F_X:

$$Q_X(u) := \inf\{x : F_X(x) \geq u\}, \quad u \in (0,1). \tag{15}$$

If there exists a function $f_X : R \longrightarrow [0, \infty)$ such that

$$F_X(x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R},$$

then X is called absolutely continuous (with respect to Lebesgue measure), and f_X is called the probability density function. Let $x_F = \sup\{x : F_X(x) = 0\}$ and $x^F = \inf\{x : F_X(x) = 1\}$. If f_X is positive on (x_F, x^F), then F_X is one-to-one and consequently Q_X is a strictly increasing and differentiable function on (0, 1). In this case the function

$$q_X(u) = \frac{d}{du} Q_X(u), \quad u \in (0,1), \tag{16}$$

is called the quantile density function (qdf) of X. When the qdf exists,

$$q_X(u) f_X(Q_X(u)) = 1, \quad u \in (0,1). \tag{17}$$

The function $fQ_X(u) := f_X(Q_X(u)), u \in (0,1)$, is called by Parzen(1979) the *density − quantile* function. It is easy to see that the pdf is uniquely determined by the following curves:

$$\begin{cases} x(u) = Q_X(u) \\ y(u) = 1/q_X(u), \quad u \in (0,1). \end{cases} \tag{18}$$

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from some distribution with quantile density q, and let $Y_i = n(X_{(i)} - X_{(i-1)}), i = 2, \ldots, n$. As argued in Pyke(1965), $Y_1, \ldots, Y_n$ are approximately independent with Y_i approximately distributed exponential with mean $q(\frac{i}{n})$, $i = 2, \ldots, n$. So we can regress Y_i on $u_i (= i/n)$ to estimate $q(u)$, and thereby estimate $fQ_x(u)$ by using (17).

A simulation study was done with the normal, exponential and gamma density functions as test cases. The results show that the OSLI method estimates the density quantile function better, on average, than the other methods. We were also curious about comparing the methods with respect to the average squared error(ASE) criterion, where ASE of the estimate $\hat{r}_h$ is defined by

$$ASE(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_h(x_i) - r(x_i)\right)^2.$$

Some results are given in the Tables 2-4 and Tables 12-14.

In addition, we also tried the methodology on some real data, see Figures 1-4. The data are SPAD readings taken on different sets of cowpeas. A SPAD reading is the ratio of two measures of light transmission, and indicates the relative amount of chlorophyll present in a cowpea leaf. High SPAD readings are produced by green leaves, and low SPAD readings by yellow ones. The P1 and P2 data sets correspond to predominantly yellow and green cowpeas, respectively, while the F1 peas are a cross of the P1 and P2 types. The F2 cowpeas are a backcross of F1 peas with peas of type P1 and P2. The trimodal shape of the F2 estimate is interesting since it tends to support the hypothesis that a single gene controls cowpea color.

TABLE 2: Density quantile estimation(exponential, weighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.41379	0.38483	0.29115	0.31975	0.34627
	100	0.29412	0.26245	0.17165	0.21304	0.24687
Mean(WASE($\hat{h}$))						
	50	0.11287	0.12433	0.13402	0.14541	0.08459
	100	0.07751	0.08070	0.09154	0.11007	0.05898
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)						
	50	1.17847	2.85427	2.67470	4.40382	
	100	0.79966	1.51508	1.26801	2.82103	
Variance($\hat{h}$) $\times 10^2$						
	50	4.16468	5.15370	4.18559	6.03267	
	100	2.39648	2.70432	2.46543	3.81780	

TABLE 3: Density quantile estimation(normal, weighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.33804	0.33614	0.30337	0.33586	0.28240
	100	0.24388	0.22079	0.19356	0.21355	0.19546
Mean(WASE($\hat{h}$))						
	50	0.08761	0.10131	0.09865	0.11718	0.06770
	100	0.06690	0.07221	0.07648	0.11218	0.05405
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)						
	50	2.22351	3.59562	2.62312	4.95087	
	100	1.11947	1.65891	1.33570	3.10678	
Variance($\hat{h}$) $\times 10^2$						
	50	0.78433	2.18140	1.70714	3.98908	
	100	0.30694	0.96196	1.13278	2.84419	

TABLE 4: Density quantile estimation(gamma, weighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.36302	0.35240	0.29659	0.32626	0.29958
	100	0.25832	0.23914	0.17278	0.22977	0.21087
Mean(WASE($\hat{h}$))						
	50	0.09138	0.10449	0.10356	0.12003	0.06868
	100	0.06757	0.07264	0.08043	0.09763	0.05327
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)						
	50	2.96983	4.54373	3.21732	5.44116	
	100	1.31547	1.79328	1.35909	3.22895	
Variance($\hat{h}$) $\times 10^2$						
	50	1.04178	2.70393	2.32235	4.51948	
	100	0.41023	1.12705	1.07762	2.99921	

4.2 Spectral Estimation

The covariance sequence of a second-order stationary time series can always be expressed as

$$r_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t} dF(\omega)$$

for the *spectrum* F, a finite measure on $(-\pi, \pi]$. Under mild conditions, this measure has a density known as the spectral density f,

$$f(\omega) = \sum_{-\infty}^{\infty} r_t \exp(-i\omega t) = r_0[1 + 2\sum_{1}^{\infty} \rho_t \cos(\omega t)], \quad \omega \in (-\pi, \pi].$$

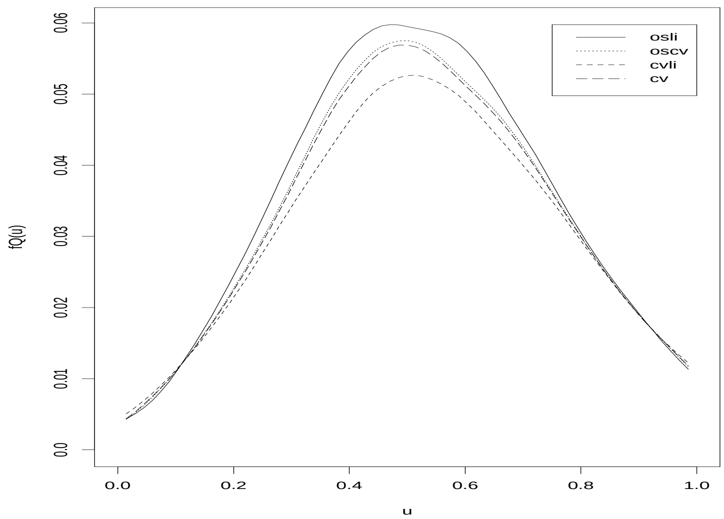

Figure 5: *Estimate of density quantile function for data F1.*

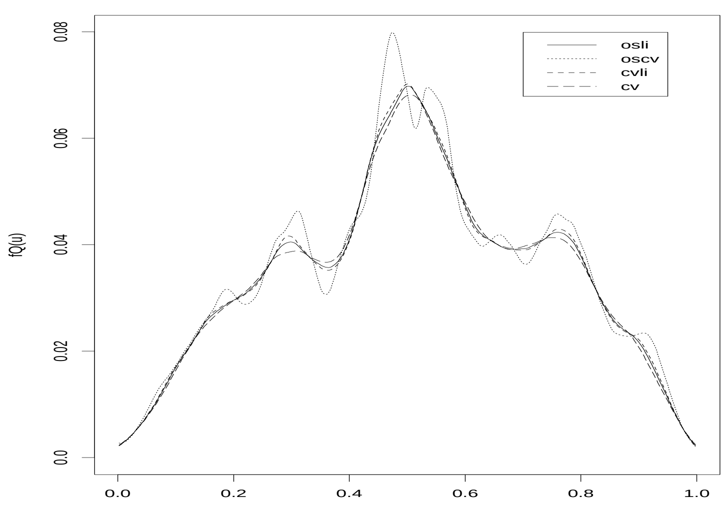

Figure 6: *Estimate of density quantile function for F2 data.*

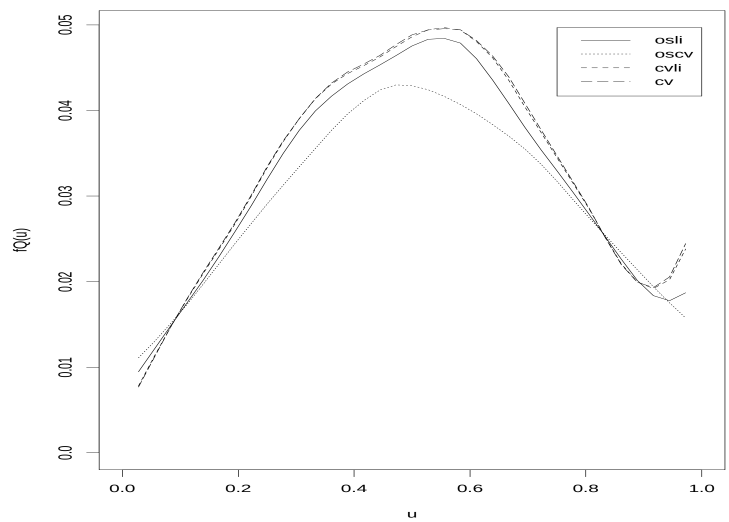

Figure 7: *Estimate of density quantile function for P1 data.*

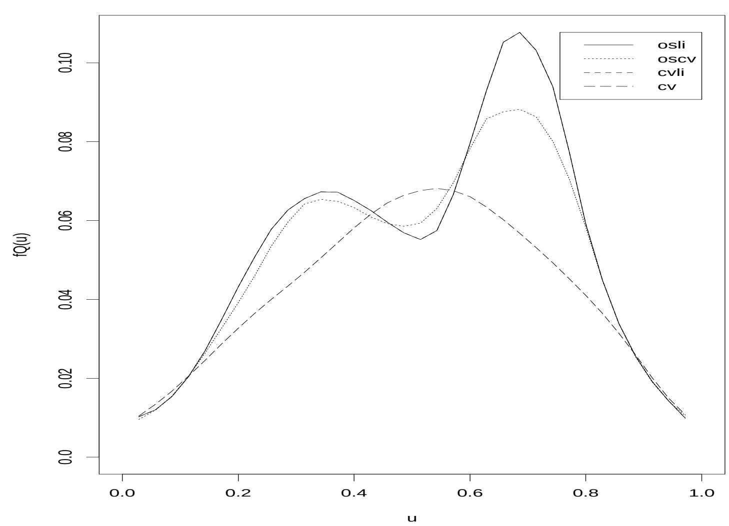

Figure 8: *Estimate of density quantile function for P2 data.*

The basic tool in estimating the spectral density is the periodogram. For a frequency ω we effectively compute the squared correlation between the series and sine/cosine waves of frequency ω by

$$I(\omega) = \left| \sum_{t=1}^{n} e^{-i\omega t} X_t \right|^2 / n = \frac{1}{n} \left[\left\{ \sum_{t=1}^{n} X_t \sin(\omega t) \right\}^2 + \left\{ \sum_{t=1}^{n} X_t \cos(\omega t) \right\}^2 \right]. \tag{19}$$

Asymptotic theory shows that $I(\omega)$ is approximately distributed $f(\omega)E$, where E has a standard exponential distribution, except at $\omega = 0$ and $\omega = \pi$, and that periodogram ordinates at frequencies $\omega_i = \frac{2\pi i}{n}, i = 1, \ldots, [\frac{n}{2}]$ are approximately independent. We can thus regress $I(\omega_i)$ on ω_i of the form $\omega_i = 2\pi i/n$, and estimate the spectral density using the regression techniques of Section 2.

In this simulation study, we use the AR(1) model

$$X_t = 0.35 X_{t-1} + \epsilon_t, \quad t = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, 0.04)$, and MA(1) model

$$X_t = \epsilon_t + 0.35\epsilon_{t-1}, \quad t = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, 0.04)$. The results show that compared with the other methods, OSLI provides a smaller WASE. The small variance of $\hat{h}_{osli}$ and the mean of $(\hat{h} - \hat{h}_o)^2$ indicate that OSLI method usually produces a better estimate than the other methods.

TABLE 5: Spectral density estimation (AR(1) model)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	1.71765	1.55942	1.42849	1.40537	1.66478
	100	1.68593	1.54837	1.41414	1.44613	1.62203
Mean(WASE($\hat{h}$))						
	50	0.06419	0.07606	0.08280	0.08435	0.05418
	100	0.03306	0.03966	0.04176	0.04080	0.02813
Mean($(\hat{h} - \hat{h}_o)^2 \times 10^2$)						
	50	0.22518	0.39160	0.56921	0.52576	
	100	0.24864	0.42713	0.52830	0.48976	
Variance($\hat{h}$) $\times 10^2$						
	50	0.08804	0.23882	0.37450	0.33786	
	100	0.10235	0.25156	0.35348	0.32280	

TABLE 6: Spectral density estimation (MA(1) model)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	2.57295	2.38350	1.89865	2.05340	2.33793
	100	2.44585	2.24855	1.90100	2.04340	2.18360
Mean(WASE($\hat{h}$))						
	50	0.05614	0.06431	0.07294	0.06732	0.04381
	100	0.03243	0.03641	0.03798	0.03628	0.02336
Mean($(\hat{h} - \hat{h}_o)^2 \times 10^2$)						
	50	1.26998	1.45133	1.97247	1.62678	
	100	1.47271	1.57920	1.94889	1.53943	
Variance($\hat{h}$) $\times 10^2$						
	50	0.46527	0.80203	1.14554	0.91287	
	100	0.54226	0.85709	1.09444	0.88649	

5 CONCLUSION

In this project, we proposed a method, OSLI, for selecting the window width of local generalized linear estimates in an exponential regression model. We also compared the performances of four different methods: OSLI, CVLI, OSCV and CV. The two versions of likelihood CV(CVLI and OSLI) provide estimates of an average weighted squared error criterion.

Under the exponential distribution setting, we investigate use of the prequential method for selecting the window width of local estimators of a regression curve. Two special cases: density quantile function and spectral estimation are studied. The OSLI gave the best results in all the simulations, based on MWASE.

The theoretical and simulation results in this project apply only to the case of fixed, evenly spaced x's. However, the method of OSLI is well defined regardless of the type of design. We also studied the performance of OSLI in real dataset consisting of chlorophyll measurements on cowpeas.

Generalizing OSLI to other model settings is of interest for further study.

6 APPENDIX

TABLE 7: Simulation results for the function $1 + 4.998 \times 2^8 \times x^4(1-x)^4$**,** $SNR = 1.65$

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)	50	0.34967	0.33445	0.25930	0.28437	0.334189
	100	0.28195	0.25066	0.22472	0.22179	0.278977
	150	0.25138	0.22237	0.20441	0.20292	0.240643
Mean(WASE($\hat{h}$))	50	0.09453	0.11422	0.10359	0.12324	0.081227
	100	0.05159	0.06448	0.05590	0.07487	0.044650
	150	0.03902	0.04589	0.04178	0.05117	0.034450
Mean($(\hat{h} - \hat{h}_o)^2 \times 10^2$)	50	0.48228	2.02845	0.43097	1.55462	
	100	0.18322	0.93469	0.22560	0.88203	
	150	0.11736	0.55782	0.17038	0.61060	
Variance($\hat{h}$) $\times 10^2$	50	1.87203	3.95823	2.37015	3.29513	
	100	1.19948	2.18121	1.50780	2.38636	
	150	0.80042	1.31245	0.96073	1.45096	

TABLE 8: Simulation results for the function $1 + 2.753 \times 2^8 \times x^4(1-x)^4$**,** $SNR = 1.3$

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)	50	0.39237	0.36827	0.29366	0.30831	0.38349
	100	0.31601	0.29224	0.25340	0.25335	0.32148
	150	0.28213	0.25859	0.23398	0.22594	0.28160
Mean(WASE($\hat{h}$))	50	0.07729	0.09876	0.08755	0.11362	0.06678
	100	0.04713	0.05848	0.05224	0.06614	0.04012
	150	0.03496	0.04212	0.03660	0.04755	0.03035
Mean($(\hat{h} - \hat{h}_o)^2 \times 10^2$)	50	0.69496	2.41963	0.72195	2.21447	
	100	0.34534	1.26826	0.36343	1.16117	
	150	0.18902	0.87047	0.23154	0.82029	
Variance($\hat{h}$) $\times 10^2$	50	2.51337	5.20807	3.43360	4.96081	
	100	1.63547	2.77200	2.09585	2.89505	
	150	1.05538	1.90350	1.31184	2.05874	

TABLE 9: Simulation results for the function $1 + 1.680 \times 2^8 \times x^4(1-x)^4$, $SNR = 1$

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)	50	0.44220	0.40737	0.34451	0.35773	0.42304
	100	0.36671	0.33205	0.29980	0.29873	0.35767
	150	0.31783	0.30615	0.26588	0.27076	0.32080
Mean(WASE($\hat{h}$))	50	0.07280	0.09098	0.08095	0.10689	0.06179
	100	0.04461	0.05481	0.04716	0.06316	0.03794
	150	0.03157	0.03761	0.03342	0.04132	0.02671
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)	50	0.85548	2.67827	1.17753	2.98451	
	100	0.60396	1.72553	0.60677	1.87451	
	150	0.37788	1.24885	0.41905	1.19915	
Variance($\hat{h}$) $\times 10^2$	50	3.22652	5.55836	3.96804	6.41647	
	100	2.09171	3.72725	2.35442	4.11106	
	150	1.54514	2.75761	1.73862	2.80569	

TABLE 10: Simulation results for the function $1 + 0.883 \times 2^8 \times x^4(1-x)^4$, $SNR = 0.65$

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)	50	0.48675	0.44771	0.40151	0.41141	0.48971
	100	0.44170	0.38281	0.37106	0.36568	0.42781
	150	0.40004	0.37305	0.33842	0.35201	0.38459
Mean(WASE($\hat{h}$))	50	0.06486	0.08242	0.07475	0.09730	0.05538
	100	0.03903	0.04956	0.04281	0.05556	0.03321
	150	0.02726	0.03354	0.02946	0.03638	0.02286
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)	50	0.68080	2.75850	1.58593	3.39153	
	100	0.77584	2.16326	1.22281	2.51618	
	150	0.86393	1.96760	0.99595	2.15599	
Variance($\hat{h}$) $\times 10^2$	50	2.99427	5.68897	4.83932	6.55072	
	100	2.60806	4.57087	3.44784	5.34095	
	150	2.12229	3.67340	2.30392	3.78463	

TABLE 11: Simulation results for the function $1 + 0.3416 \times 2^8 \times x^4(1-x)^4$, $SNR = 0.3$

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
	50	0.52303	0.79363	0.45145	0.68803	0.73635
Mean($\hat{h}$)	100	0.50890	0.72659	0.46056	0.68194	0.64703
	150	0.49804	0.71322	0.45654	0.63353	0.60650
	50	0.07172	0.08766	0.08476	0.10308	0.05395
Mean(WASE($\hat{h}$))	100	0.03957	0.05116	0.04394	0.05264	0.03043
	150	0.02792	0.03647	0.03161	0.03857	0.02252
	50	11.86332	19.28791	17.24091	21.58073	
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)	100	9.85701	21.62282	12.15919	20.01313	
	150	8.65789	21.58704	10.36656	18.68326	
	50	0.45594	9.67700	1.85099	11.73457	
Variance($\hat{h}$) $\times 10^2$	100	0.61863	11.60840	1.40051	10.94829	
	150	0.66172	11.08550	1.34868	11.10388	

TABLE 12: Density quantile estimation(Exponential, unweighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.41379	0.38483	0.29115	0.31975	0.22743
	100	0.29412	0.26245	0.17165	0.21304	0.11973
Mean(ASE($\hat{h}$))						
	50	23.2234	31.1463	39.9413	43.0162	12.2444
	100	48.5347	47.7488	58.3223	60.3079	22.3146
Mean($(\hat{h}-\hat{h}_o)^2 \times 10^2$)						
	50	7.35246	7.56885	5.55739	7.57692	
	100	4.96960	4.84095	2.32337	4.48676	
Variance($\hat{h}$) $\times 10^2$						
	50	1.11167	2.45928	2.60954	3.79240	
	100	0.79966	1.51508	1.26801	2.82103	

TABLE 13: Density quantile estimation(Normal, unweighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.33804	0.33614	0.30337	0.33586	0.24667
	100	0.24388	0.22109	0.19356	0.21583	0.13046
Mean(ASE($\hat{h}$))						
	50	6.86385	7.97369	8.21008	9.63062	4.93001
	100	11.91433	12.27353	15.41093	18.57002	7.94981
Mean($(\hat{h}-\hat{h_o})^2 \times 10^2$)						
	50	3.77008	5.58878	4.15529	7.39709	
	100	2.24776	2.59997	2.31455	4.33252	
Variance($\hat{h}$) $\times 10^2$						
	50	0.78433	2.18140	1.70714	3.98908	
	100	0.30694	0.95839	1.13278	2.83221	

TABLE 14: Density quantile estimation(Gamma, unweighted ASE)

	n	$\hat{h}_{osli}$	$\hat{h}_{li}$	$\hat{h}_{oscv}$	$\hat{h}_{cv}$	$\hat{h}_o$
Mean($\hat{h}$)						
	50	0.36302	0.35240	0.29659	0.32626	0.22889
	100	0.25832	0.23914	0.17278	0.22972	0.13160
Mean(ASE($\hat{h}$))						
	50	39.7017	45.4843	58.4892	55.8592	22.8069
	100	75.5480	77.1118	96.8821	99.1048	41.4630
Mean($(\hat{h}-\hat{h_o})^2 \times 10^2$)						
	50	5.55381	7.10047	5.16149	7.70358	
	100	3.20848	3.54081	2.14431	5.01529	
Variance($\hat{h}$) $\times 10^2$						
	50	1.04178	2.70393	2.32235	4.51948	
	100	0.41023	1.12705	1.07762	2.99921	

REFERENCES

Aerts, M. and Claeskens, G.(1997). Local polynomial estimation in multiparameter likelihood models. *J. Amer. Statist. Assoc.* **92** 1536-1545.

Chaudhuri, P. and Dewanji, A.(1995). On a likelihood-based approach in nonparametric smoothing and cross-validation. *Statist. Probab. Lett.* **22** 7-15.

Cheng, Cheng and Parzen, E.(1997). Unified estimators of smooth quantile density functions. *J. Statist. Plann. Inference* **59** 291-307.

Climov D., Hart J.D. and Simar, L. Automatic smoothing and estimating in single index poisson regression, submitted to *Journal of Nonparametric Statistics*.

Fan, Jianqing, Heckman, N. E, and Wand P.(1995). Local polynomial kernel regression for Generalized Linear Models and Quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90** 141-150.

Hart, J. D. and Yi, Seongbaek(1998). One-sided cross-validation. *J. Amer. Statist. Assoc.* **93** 620-631.

Hart, J.D.(1997). *Nonparametric Smoothing and lack-of-fit tests.* Springer-Verlag, New York.

Modha, D.S. and Masry, E(1998). Prequential and cross-validated regression estimation. *Machine Learning* **33** 5-39.

Pyke, R (1965). Spacing (with discussion). *J. Roy. Statist. Soc. Ser. B* **27** 395-449.

Staniswalis, J. G.(1994) Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89** 501-511.

Tibshirani, R. and Hastie, T.(1987) Local likelihood estimation. *J. Amer. Statist. Assoc.* **82** 559-567.

Venables, W.N. and Ripley, B.D.(1999). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.