KMITL Sci. Tech. J. Vol. 11 No. 2 Jul. - Dec. 2011

The Comparison of Bandwidth Selection Methods using Kernel Function

Autcha Araveeporn

Department of Applied Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

Abstract

This paper presents the bandwidth selection methods for local polynomial regression with Normal, Epanechnikov, and Uniform kernel function. The bandwidth selection methods are proposed by Histogram Bin Width method, Bandwidth for Kernel Density Estimation method, and Bandwidth for Local Linear Regression method to estimate the local polynomial regression estimator. Using Monte Carlo simulations, we compare the Mean Square Error (MSE) of the bandwidth selection methods. For simulation results, it can be seen that the MSE of Bandwidth for Kernel Density Estimation method provides the smallest in all situations. The bandwidth selection methods are applied to the Stock Exchange of Thailand (SET) index. The results show that the MSE of Kernel Density Estimation method with Normal kernel function is the smallest as the simulation study.

Keywords: Bandwidth, Epanechnikov kernel function, Local linear regression, Local polynomial regression

1. Introduction

A parametric regression model is known as the analyzing of the relationship between predictor variables and the response variable that required the assumption of the underlying regression such as linearity, stationary variance, and independence of explanatory variables. The restriction of parametric regression model is possible to produce misleading conclusions. This latter strategy leads to nonparametric regression model that relaxed the assumption of linearity.

Many nonparametric regression methods or called smoothing methods which produce a smoother exist. A smoother is a tool for summarizing the trend of a response variable as a function of one or more predictor variables. The single predictor case is called simple nonparametric model or scatterplot smoothing that can be used to enhance the visual appearance of the scatterplot of the response variable, to help our eyes pick out the trend in the plot [1].

The most popular nonparametric regression methods include kernel smoothing started by Nadaraya-Watson kernel estimator [2-3] and local polynomial regression. A vast literature on local polynomial regression are reported. Stone [4] examined the consistency properties of many nonparametric regression estimators in local polynomial. Cleveland [5] was a catalyst whose renewed interest in local polynomial, introducing LOWESS (locally weighted scatterplot smoothing) using a tricube kernel function with bandwidth based on neighbor distances. Müller [6] studied certain equivalences between local polynomial kernel estimators and kernel estimators.

*Corresponding author: E-mail: kaautcha@kmitl.ac.th

The local polynomial regression is a method based on a class of kernel smoothing that can be used to estimate the regression function at a particular point by locally fitting a *p*th degree polynomial to the data via weighted least squares. This class includes the Nadaraya-Watson kernel estimator since it corresponds to fitting degree zero polynomials (i.e., constants). The importance and simplicity is the local polynomial kernel estimators corresponding to *p = 1* or called the local linear regression.

The Nadaraya-Watson kernel estimator depends on the smoothing parameter or called bandwith that controls the trade-off between the goodness-of-fit and model complexity. Allen [7] and Stone [8] proposed the cross-validation method to select the smoothing parameter by the cross-validation criterion. Wahba [9] and Craven and Wahba [10] suggested replacing the ordinary cross-validation to generalized cross-validation for choosing the smoothing parameter.

The aim of this article is to compare the bandwidth selection methods of bandwidth selection for the Nadaraya-Watson kernel estimator under Normal, Epanechnikov, and Uniform kernel function. We consider the local polynomial regression in Section 2 and kernel function in Section 3. The bandwidth selection methods are illustrated in Section 4 and these methods with simulation data and real data were used in Sections 5-6. Finally, the conclusions are presented in Section 7.

2. The Local Polynomial Regression

Consider the simple nonparametric regression functions as

$$y_t = \mu(x_t) + \varepsilon_t \quad , t = 1, 2, \ldots, n \tag{1}$$

where $x_t, t = 1, 2, \ldots, n$ are known the predictor variable, $y_t, t = 1, 2, \ldots, n$ are known the response variable, $\mu(x_t)$ are the nonparametric regression function that we want to estimate, and $\varepsilon_t, t = 1, 2, \ldots, n$ denote the measurement errors.

The main concept of local polynomial regression estimator is to approximate parameter by a polynomial of some degree. The Taylor expansion can be approximated by a polynomial of degree *p* and assumed that $\mu(x_t)$, as

$$\mu(x_t) \approx \mu(x_t) + (x_t - x)\mu^{(1)}(x) + \frac{(x_t - x)^2 \mu^{(2)}(x)}{2!} + \ldots + \frac{(x_t - x)^p \mu^{(p)}(x)}{p!} \tag{2}$$

The local polynomial regression estimator is minimizer of;

$$(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p) = \arg\min \sum_{t=1}^{n} \left\{ y_t - \beta_0 - \beta_1(x_t - x) - \ldots - \beta_p(x_t - x)^p \right\}^2 K\left(\frac{x_t - x}{\lambda}\right) \tag{3}$$

where $K(.)$ is a kernel function, $\lambda > 0$ is called the bandwidth and $\underset{\sim}{\beta}$ denotes the vector of coefficient $(\beta_0, \beta_1, \ldots, \beta_p)^T$ evaluated at x.

The explicit expression in (3) can be made via matrix notation as

$$X = \begin{bmatrix} 1 & x_1 - x & \dots & (x_1 - x)^p \\ 1 & x_2 - x & \dots & (x_2 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & \dots & (x_n - x)^p \end{bmatrix} \quad , \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and

$$W = \begin{bmatrix} K\left(\frac{x_1 - x}{\lambda}\right) & 0 & \dots & 0 \\ 0 & K\left(\frac{x_2 - x}{\lambda}\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K\left(\frac{x_n - x}{\lambda}\right) \end{bmatrix}$$

We get the weight least squares estimator as

$$\hat{\beta}(x) = (X^T W X)^{-1} X^T W Y \qquad (4)$$

For *p=0*, the local constant estimator is Nadaraya-Watson kernel estimator written as

$$\hat{\beta}_0(x\,;\lambda) = \frac{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right) y_t}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right)} \qquad (5)$$

or $\hat{\mu}(x) = \sum_{t=1}^{n} w_t y_t$

where $w_t = \frac{K\left(\frac{x_t - x}{\lambda}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right)}$, λ is known the bandwidth, and $K(.)$ is a kernel function.

3. Kernel Function

The kernel function is a weighting function used in the local polynomial kernel estimator in nonparametric regression function. It gives the weight of the nearby data points in making the estimators. Given these characteristics, the specific choice of a kernel function is not critical that calculated weight by

$$K(x_t) = K\left(\frac{x_t - x}{\lambda}\right)$$

Three popular choices of kernel functions, illustrated in Figure 1 are the Gaussian or Normal kernel function, Epanechnikov kernel function, and Uniform kernel function.

The Normal kernel function is simply the standard normal density function

$$K(u) = (2\pi)^{-1/2} e^{-u^2/2} \quad , \quad u \in [-\infty, \infty]$$

The Epanechnikov kernel function is optimal in minimum variance sense [11]

$$K(u) = \frac{3}{2}\left(1 - u^2\right) \quad , \quad u \in [-1, 1]$$

The Uniform kernel function

$$K(u) = \frac{1}{2} \quad , \quad u \in [-1, 1]$$

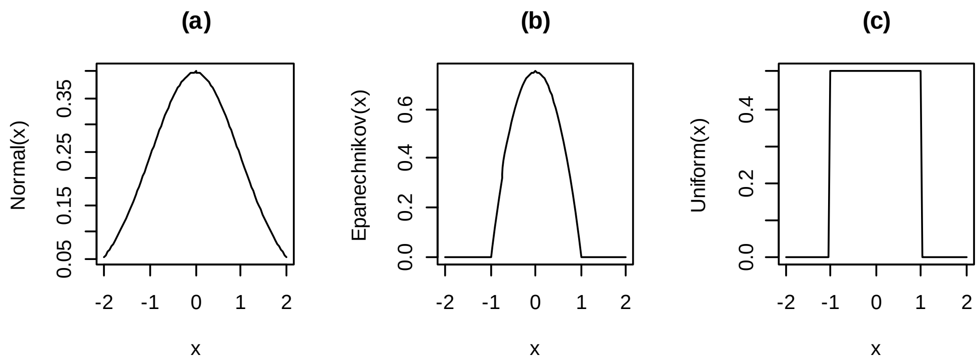

Figure 1 Gaussian or Normal kernel function (a), Epanechnikov kernel function (b), and Uniform kernel function (c).

4. Bandwidth Selection Methods

A bandwidth or a smoothing parameter (λ) is considered to be good if it produces a small prediction error, usually measured by Mean Square Error (MSE).

There are several methods to estimate bandwidth, so in this case we use the three good performance methods as follows:

4.1 A Histogram Bin Width [12]

This method uses direct plug-in methodology to select the bin width of a histogram denoted "dpih" that extended the bin width rules of Scott [13] to estimate bandwidth by

$$\hat{\lambda} = 3.49\,\hat{\sigma}\,n^{-1/3} \qquad (6)$$

where $\hat{\sigma}$ is an estimate of standard deviation, and n is a number observation.

Wand [12] studied the number stages l of functional estimation before a rough estimate is used is another variable that needs to be specified. This means that he actually has a family of plug-in rules indexed by l. Let $\hat{\lambda}_l$ denote the l-stage plug-in rule with initial estimates found using a normal scale rule. Examples of $\hat{\lambda}_l$ are :

The zero-stage rule of $\hat{\lambda}_0$

$$\hat{\lambda}_0 = \left(\frac{6}{-\hat{\omega}_2^{NS} n}\right)^{1/3} = \left(\frac{24\,\pi^{1/2}}{n}\right)^{1/3}\hat{\sigma} \approx 3.49\,\hat{\sigma}\,n^{-1/3} \qquad (7)$$

Note that $\hat{\lambda}_0$ is simply the normal scale bin width selection rule of Scott [13].

The one-stage rule of $\hat{\lambda}_1$

$$\hat{\lambda}_1 = \left(\frac{6}{-\hat{\omega}_2(g_{11})n}\right)^{1/3} \quad \text{where} \quad g_{11} = \left[\frac{-2L^{(2)}(0)}{\mu_2(L)\hat{\omega}_4^{NS} n}\right]^{1/5} \qquad (8)$$

The two-stage rule of $\hat{\lambda}_2$

$$\hat{\lambda}_2 = \left(\frac{6}{-\hat{\omega}_2(g_{21})n}\right)^{1/3} \quad \text{where} \quad g_{21} = \left[\frac{-2L^{(2)}(0)}{\mu_2(L)\hat{\omega}_4(g_{22})n}\right]^{1/5} \qquad (9)$$

$$,\ g_{22} = \left[\frac{-2L^{(4)}(0)}{\mu_4(L)\hat{\omega}_6^{NS} n}\right]^{1/7}, \text{ and } \mu_k(L) = \int u^k L(u)\,du$$

An estimate of ω_r is called a normal scale estimator, and denoted this by $\hat{\omega}_r^{NS}$. A useful result is

$$\hat{\omega}_r^{NS} = \frac{(-1)^{r/2}\,r!}{(2\hat{\sigma})^{r+1}(r/2)!\,\pi^{1/2}} \qquad (10)$$

4.2 A Bandwidth for Kernel Density Estimation [14]

This method uses plug-in methodology to select the bandwidth of a kernel density estimation denoted “dpik” that was studied from the concept of Park and Marron [15]. The Park and Marron’s λ selector based on the choosing λ to minimize a kernel-based estimate of Mean Integrated Squared Error (MISE) via the first two terms of its usual asymptotic expansion (AMISE) valids as $n \to \infty$ and $\lambda = \lambda(n) \to 0$:

$$AMISE(\lambda) = (n\lambda)^{-1} R(K) + \frac{1}{4}\lambda^4 \sigma_K^4 R(f'') \tag{11}$$

Here, the notation follows the convention $R(g) = \int g^2(x)\,dx,\quad \sigma_g^2 = \int x^2 g(x)\,dx$ for appropriate function g. Each objective function is thus of the form

$$\psi(\lambda) = (n\lambda)^{-1} R(K) + \frac{1}{4}\lambda^4 \sigma_K^4\, \hat{S}(\alpha) \tag{12}$$

where $\hat{S}(\alpha)$ is a kernel-based estimate of $R(f'')$, using some appropriate bandwidth α . Note that if α did not depend on λ , the minimization of ψ could be performed analytically to give

$$\tilde{\lambda} = \left[\frac{R(K)}{\sigma_K^4 \hat{S}(\alpha)}\right]^{1/5} n^{-1/5} \tag{13}$$

and equation (14) was improved from equation (13) by Jones and Sheather [14], i.e.

$$\tilde{\lambda} = \left[\frac{R(K)}{\sigma_K^4 \hat{S}(\alpha_2(\lambda))}\right]^{1/5} n^{-1/5} \tag{14}$$

where $\alpha_2(\lambda) = \hat{c}_1 \lambda^{5/7}$, noting that $\hat{c}_1$ estimates c_1 to use a scale model.

4.3 A Bandwidth for Local Linear Regression [16]

This method uses plug-in methodology to select the bandwidth of local linear Gaussian kernel regression denoted “dpill”. The direct plug-in approach, where unknown functions that appear in expressions for the asymptotically optimal bandwidths are replaced by kernel estimates, is used. The kernel is the standard normal density. Least squares quartic fits over blocks of data are used to obtain an initial estimate. Mallow’s C_p is used to select the number of blocks.

For the local linear kernel estimator, the MISE-optimal bandwidth is asymptotic to

$$\lambda_{AMISE} = C_1(K)\left[\frac{\sigma^2(b-a)}{\theta_{22} n}\right]^{1/5} \tag{15}$$

where $C_1(K) = [R(K)/\mu_2(K)^2]^{1/5}$, whereas the MSE-optimal bandwidth for estimation of

$$g_{AMSE} = C_2(K)\left[\frac{\sigma^2(b-a)}{|\theta_{24}| n}\right]^{1/7} \tag{16}$$

where $\sigma^2(b-a)$ is replaced by $\int_a^b v(x)dx$.

5. Simulation Study

The simulation study is to estimate the performance of bandwidth selection methods, dpih, dpik, and dpill, via kernel function of local polynomial regression method. Data are generated from an AutoRegressive (AR) in order 1 given by

$$y_t = \rho\, y_{t-1} + \varepsilon_t \quad , \quad t = 1, 2, \ldots, n$$

where $\varepsilon_t \sim$ N(0,1) ; the coefficient of AR(1) is ρ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1; sample sizes n = 25, 50, 100 and 250.

The R Statistical Software was used to simulate data at 500 replications until the results are stable in all cases. The efficiency of bandwidth selection method, dpih, dpik, and dpill is estimated using Mean Square Error (MSE) as follows:

$$MSE = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}$$

Table1 shows the estimated MSEs of kernel function, Normal kernel function, Epanechnikov kernel function, and Uniform kernel function, under bandwidth selection methods. The MSEs of Normal kernel function, Epanechnikov kernel function, and Uniform kernel function, of all sample sizes are also presented in Figures 2-4.

The estimated MSEs dpik are the smallest especially when n is small, but the MSEs of dpill are larger than the MSEs of the other methods except when ρ =1 at the sample sizes n = 25, 50, and 100. When n is large (n=250) , the MSEs of dpih and dpik are the same at Epanechnikov kernel function, and Uniform kernel function for ρ = 0.1-0.7 .

In this case, the estimated MSEs are too small that the horizontal axis of kernel function was divided into sub-intervals which covers the range of data.

Table 1 The estimated Mean Square Error (MSE) of bandwidth selection methods, dpih, dpik, and dpill, via Normal kernel function, Epanechnikov kernel function and Uniform kernel function.

n	ρ	Normal			Epanechnikov			Uniform		
		dpih	dpik	dpill	dpih	dpik	dpill	dpih	dpik	dpill
25	0.1	0.0007	0.0001	0.0012	0.0002	9.1e-35	0.0005	0.0007	1.0e-34	0.0011
	0.2	0.0006	0.0001	0.0011	0.0002	1.0e-34	0.0006	0.0005	1.0e-34	0.0009
	0.3	0.0008	0.0001	0.0012	0.0002	1.1e-34	0.0005	0.0005	1.1e-34	0.0008
	0.4	0.0007	0.0001	0.0011	0.0002	1.3e-08	0.0004	0.0005	1.4e-06	0.0007
	0.5	0.0008	0.0001	0.0012	0.0002	5.9e-07	0.0003	0.0004	1.4e-34	0.0006
	0.6	0.0009	0.0001	0.0011	0.0002	1.3e-06	0.0003	0.0004	4.5e-06	0.0006
	0.7	0.0011	0.0001	0.0012	0.0002	1.0e-06	0.0003	0.0004	1.2e-05	0.0006
	0.8	0.0012	0.0001	0.0011	0.0002	9.1e-06	0.0002	0.0004	1.2e-05	0.0004
	0.9	0.0024	0.0002	0.0014	0.0003	1.3e-05	0.0002	0.0007	5.9e-05	0.0004
	1	0.0049	0.0003	0.0017	0.0005	5.4e-05	0.0002	0.0012	0.0001	0.0005
50	0.1	0.0001	1.7e-05	0.0004	1.0e-05	4.8e-35	0.0003	6.4e-05	5.8e-35	0.0004
	0.2	0.0001	1.8e-05	0.0005	9.6e-06	5.0e-35	0.0002	7.9e-05	5.9e-35	0.0004
	0.3	0.0001	1.7e-05	0.0005	1.7e-05	4.7e-35	0.0002	6.4e-05	5.3e-35	0.0004
	0.4	0.0001	2.1e-05	0.0006	1.2e-05	5.8e-35	0.0002	7.3e-05	6.9e-35	0.0004
	0.5	0.0001	1.5e-05	0.0006	1.6e-05	6.7e-35	0.0001	8.1e-05	6.2e-35	0.0003
	0.6	0.0001	2.5e-05	0.0005	2.8e-05	7.6e-35	0.0001	8.3e-05	8.2e-35	0.0003
	0.7	0.0001	2.7e-05	0.0004	3.2e-05	1.2e-09	0.0001	8.6e-05	2.8e-07	0.0002
	0.8	0.0001	3.3e-05	0.0004	4.4e-05	2.1e-07	0.0001	9.4e-05	2.5e-06	0.0001
	0.9	0.0003	5.0e-05	0.0005	5.5e-05	3.5e-06	8.7e-05	0.0001	1.3e-05	0.0001
	1	0.0010	0.0001	0.0006	0.0001	1.3e-05	9.2e-05	0.0003	4.2e-05	0.0001
100	0.1	2.3e-05	1.6e-06	0.0002	2.3e-35	2.3e-35	0.0001	2.7e-35	2.7e-35	0.0002
	0.2	2.2e-05	1.5e-06	0.0002	2.6e-35	2.6e-35	0.0001	2.6e-35	2.6e-35	0.0002
	0.3	2.1e-05	1.5e-06	0.0002	4.4e-09	2.8e-35	0.0001	9.1e-09	2.5e-35	0.0002
	0.4	2.0e-05	1.9e-06	0.0002	2.8e-09	3.0e-35	0.0001	3.6e-07	3.2e-35	0.0001
	0.5	1.9e-05	2.4e-06	0.0002	1.9e-07	3.2e-35	7.4e-05	1.0e-06	3.4e-35	0.0001
	0.6	2.5e-05	3.7e-06	0.0001	9.4e-07	3.4e-35	4.7e-05	8.0e-06	4.2e-35	0.0001
	0.7	2.4e-05	4.6e-06	0.0001	2.7e-06	5.1e-35	3.4e-05	1.2e-05	5.0e-35	7.1e-05
	0.8	3.3e-05	6.5e-06	0.0001	5.7e-06	6.9 e-35	2.8e-05	2.0e-05	1.1e-07	6.2e-05
	0.9	5.4e-05	1.2e-05	0.0001	1.0e-05	1.8e-07	2.6e-05	2.5e-05	2.5e-06	5.7e-05
	1	0.0002	5.1e-05	0.0002	3.6e-05	5.3e-06	2.8e-05	6.8e-05	1.4e-05	5.7e-05
250	0.1	1.6e-06	3.5e-08	8.2e-05	9.4e-36	9.4e-36	4.4e-05	1.0e-35	1.0e-35	7.5e-05
	0.2	1.5e-06	3.3e-08	9.0e-05	8.9e-36	8.9e-36	5.5e-05	1.0e-35	1.0e-35	7.6e-05
	0.3	1.7e-06	5.0e-08	8.0e-05	1.0e-35	1.0e-35	5.2e-05	1.0e-35	1.0e-35	8.5e-05
	0.4	1.6e-06	6.0e-08	7.3e-05	1.1e-35	1.1e-35	4.5e-05	1.2e-35	1.2e-35	8.6e-05
	0.5	1.6e-06	9.4e-08	7.6e-05	1.1e-35	1.1e-35	3.1e-05	1.3e-35	1.3e-35	6.4e-05
	0.6	2.0e-06	1.7e-07	8.3e-05	1.6e-35	1.6e-35	3.3e-05	1.7e-35	1.7e-35	5.0e-05
	0.7	1.9e-06	3.1e-07	7.8e-05	1.8e-35	1.8e-35	2.3e-05	2.0e-35	2.0e-35	4.6e-05
	0.8	2.7e-06	7.5e-07	7.5e-05	6.6e-08	2.8 e-35	1.6e-05	8.1e-07	3.3e-35	4.0e-05
	0.9	4.6e-06	1.4e-06	7.6e-05	9.2e-07	3.1e-09	1.2e-05	3.3e-06	1.7e-07	3.0e-05
	1	3.5e-05	8.3e-06	9.6e-05	5.9e-06	1.4e-06	1.4e-05	1.1e-05	3.0e-06	2.4e-05

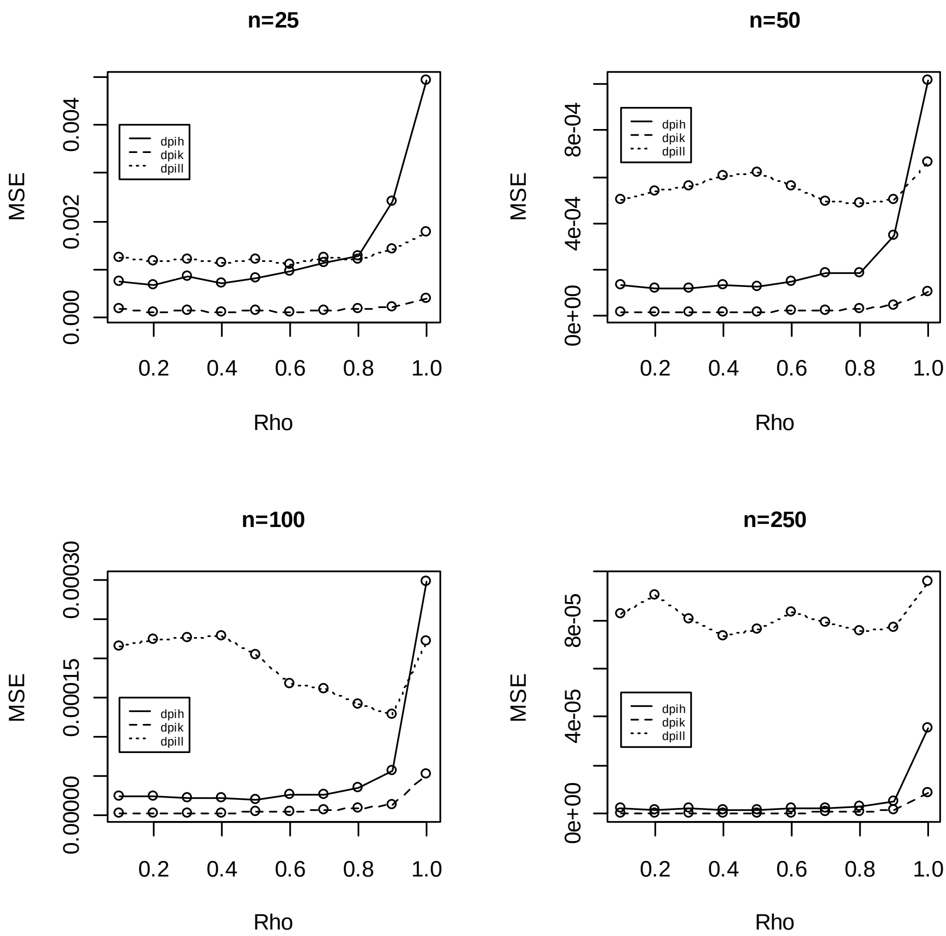

Figure 2 The estimated Mean Square Error (MSE) of bandwidth selection methods, dpih, dpik, and dpill, with Normal kernel function.

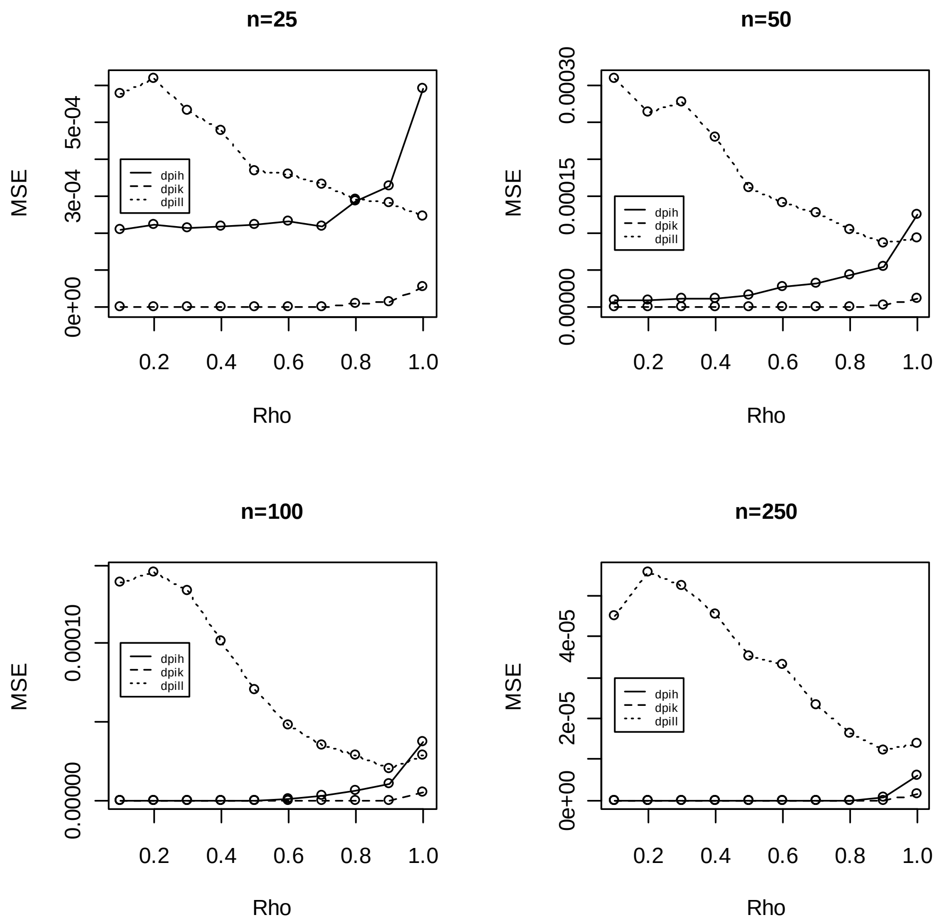

Figure 3 The estimated Mean Square Error (MSE) of bandwidth selection methods, dpih, dpik, and dpill, with Epanechnikov kernel function.

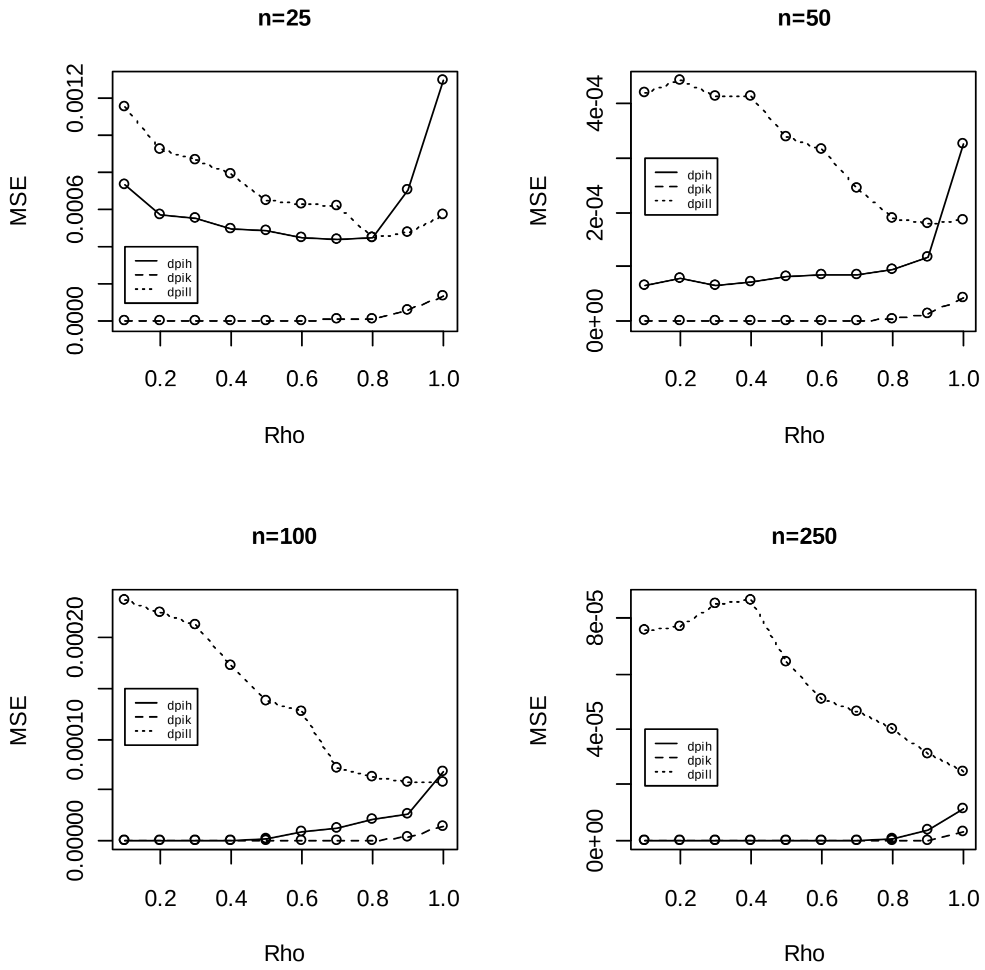

Figure 4 The estimated Mean Square Error (MSE) of bandwidth selection methods, dpih, dpik, and dpill, with Uniform kernel function.

6. Application in Real Data

In this section, we applied the bandwidth selection methods to an economic time series data. A real data is the monthly Stock Exchange of Thailand (SET) index that starting trading on April 30, 1975. These data were collected from 1976 to 2011 giving a total of 430 observations.

The MSEs obtained from bandwidth selection methods of kernel functions are shown in Table 2. In comparison for the kernel functions, we also proposed the observed values and fitted values of bandwidth selection methods in Figures 5-7.

Table 2 The estimated Mean Square Error (MSE) of bandwidth selection methods, dpih, dpik, and dpill, via Normal kernel function, Epanechnikov kernel function, and Uniform kernel function for SET index.

	Normal			**Epanechnikov**			**Uniform**		
	dpih	**dpik**	**dpill**	**dpih**	**dpik**	**dpill**	**dpih**	**dpik**	**dpill**
MSE	28.815	0.0002	263.48	10.115	0.0318	21.615	13.049	0.0451	62.758

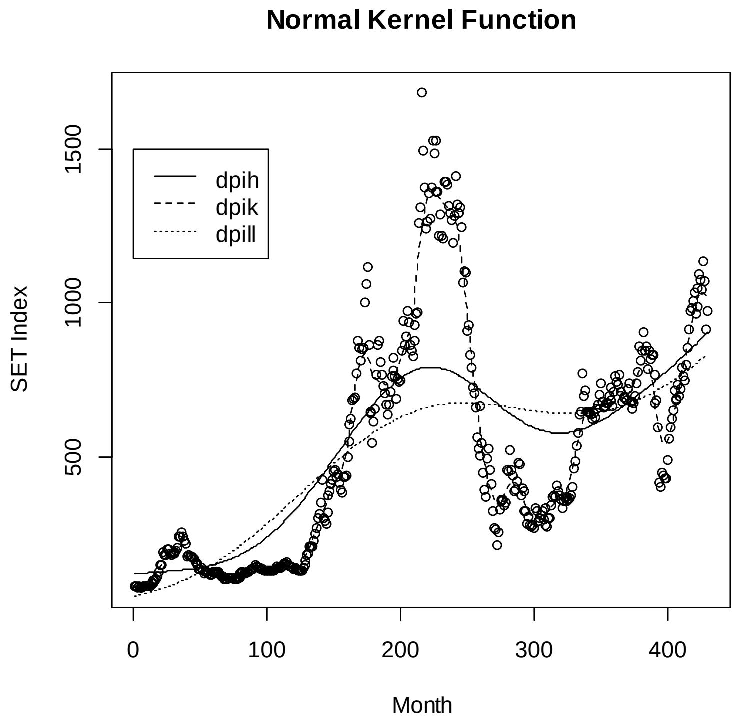

Figure 5 The scatterplot of SET index data, and local polynomial regression estimates of bandwidth selection methods, dpih, dpik, and dpill, with Normal kernel function.

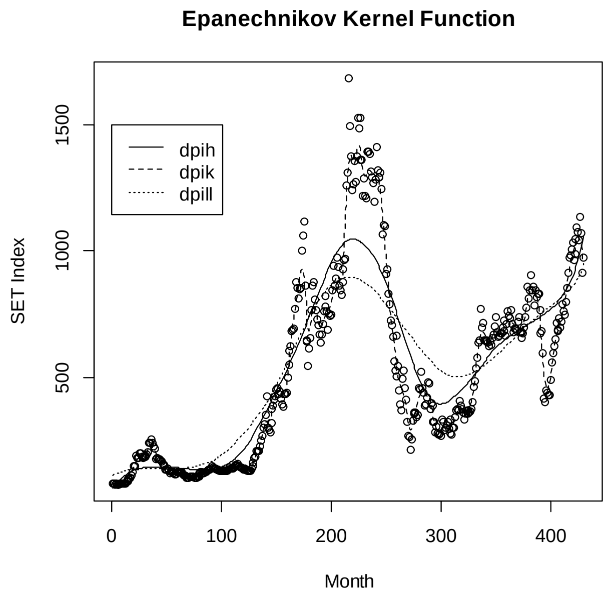

Figure 6 The scatterplot of SET index data, and local polynomial regression estimates of bandwidth selection methods, dpih, dpik, and dpill, with Epanechnikov kernel function.

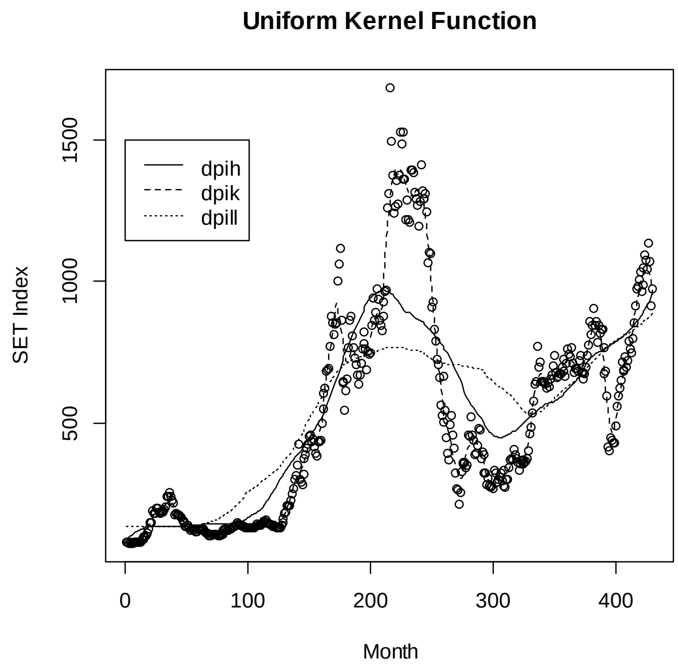

Figure 7 The scatterplot of SET index data, and local polynomial regression estimates of bandwidth selection methods, dpih, dpik, and dpill, with Uniform kernel function.

As presented in Table 2, the proposed bandwidth selection methods, dpik provides the MSEs less than those of the dpih, dpill in all kernel functions. It is easy to see the scatterplot of SET index data and different fitting of dpih, dpik, and dpill from Figures 5-7. The local polynomial regression estimates of dpik is closed to the observations of SET index.

7. Conclusions

We have compared a bandwidth selection methods of kernel function for the local polynomial estimator. Through a Monte Carlo simulation study, the selection of bandwidth for Kernel Density Estimation method (dpik) worked reasonably well for simulated data in all cases. One reason behind this method is choosing the bandwidth to minimize good quality estimates of the mean integrated squared error.

For application in real data, we are also interested in comparing the power of bandwidth selection method by considering Mean Square Error (MSE). The MSE of the selected bandwidth for Kernel Density Estimation method (dpik) with normal kernel function is shown the minimum. Therefore, the proposed bandwidth selection method is a good estimator based on the sample sizes and kernel function.

References

[1] Hastie, T.J. and Tibshirani, R., **1990**. *Generalized Additive Models*. Chapman and Hall, London.

[2] Nadaraya, E.A., **1964**. On estimating regression. *Theory of Probability and Its Application*, 9, 141-142.

[3] Watson, G.S., **1964**. Smooth regression analysis, *Sankhy Series A*, 26, 359-372.

[4] Stone, C.G., **1977**. Consistent nonparametric regression, *The Annals of Statistics*, 5, 595-620.

[5] Cleveland, W.S., **1979**. Robust locally weight regression and smoothing scatter plots, *Journal of American Statistics Association*, 74, 829-836.

[6] Müller, H.G., **1987**. Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of American Statistics Association*, 82, 231- 238.

[7] Allen, D.M., **1974**. The relationship between variable and data augmentation and a method of prediction, *Technometrics*, 16, 125-127.

[8] Stone, M., **1974**. Cross-validatory choice and assessment of statistical predictions (with discussion), *Journal of the Royal Statistical Society. Series B*, 36, 111-147.

[9] Wahba, G., **1977**. A survey of some smoothing problems and the method of generalized cross-validation for solving them. In: P.R. Krisnaiah, ed. *Application of Statistics*, North Holland, Amsterdam, pp. 507-523.

[10] Craven, P. and Wahba, G., **1979**. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized crossvalidation. *Numerische Mathematik*, 31, 377-403.

[11] Epanechnikov, V.A., **1969**. Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Application*, 14, 153-358.

[12] Wand, M.P., **1995**. Data-based choice of histogram winbidth. *The American Statistician*, 51, 59-64.

[13] Scott, D.W., **1979**. On optimal and data-based histogram, *Biometrika*, 66, 605- 610.

[14] Sheather, S.J. and Jones, M.C., **1991**. A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society, Series B*, 53, 683-690.

[15] Park, B.U. and Marron, J.S. **1990**. Comparison of data-driven bandwidth selectors, *Journal of American Statistics Association*, 85, 66 - 72.
[16] Ruppert, D., Sheather, S.J. and Wand, M.P., **1995**. An effective bandwidth selector for local least squares regression, *Journal of American Statistics Association*, 90, 1257-1270.