International Journal of
Mathematical Modelling & Computations
Vol. 04, No.02, 2014, 115- 123

The Comparison of Two Methods Nonparametric Approach on Small Area Estimation (Case: Approach with Kernel Methods and Local Polynomial Regression)

Dewi Juliah Ratnaningsih[a,*], Fia Fridayanti[b], and Andi Megawarni[a]

[a]*Department of Statisticss, Indonesia Open University, South Tangerang, Banten.*
[b]*Vocational Program, Indonesia University, Depok, Jakarta.*

Abstract.*Small Area estimation is a technique used to estimate parameters of subpopulations with small sample sizes. Small area estimation is needed in obtaining information on a small area, such as sub-districtor village. Generally, in some cases, small area estimation uses parametric modeling. But in fact, a lot of models have no linear relationship between the small area average and the covariate. This problemrequiresa non-parametric approach to solve, such asKernel approach and Local Polynomial Regression(LPR).The purpose of this study is comparing the results of smaller estimation using Kernel approach hand LPR. Data used in this study are generatedbysimulation resultsusing R language .Simulation data obtained by generating function m(x) are linear and quadratic pattern. The criteria used tocompare the results ofthe simulationareAbsoluteRelativeBias(ARB), MeanSquareError(MSE), GeneralizedCrossValidation(GCV), and risk factors. The simulation results showed: 1)Kernel gives smaller relative bias than LPR does on both linearandquadraticdata pattern. The relative bias obtained by Kernel tends to be more stable and consistent than the relative bias resulted by LPR, (2)the Kernel MSE is smaller than the LPR MSEeither on linear or quadratic pattern in any combination treatment, (3)the value of GCV and the risk factors in Kernel are smaller thanthese in LPRin any combination of the simulated data patterns, (4) on nonparametric data, for both linear data patternandquadratic data pattern, Kernel methods provide better estimation compared to LPR.*

Received: 2 February 2013; Revised: 8 August 2013; Accepted: 18 September 2013.

Keywords: Small Area Estimation, Nonparametric Methods, Kernel Methods, and Local Polynomial Regression.

Index to information contained in this paper

*Corresponding author. Email: dewijuliah@yahoo.com.

http://www.ijm2c.ir

1. Introduction

1. Introduction

Small Area Estimation is a statistical technique used to estimate parameters of asubpopulation with a small sample size[11]. This estimation is really necessary and very interested to obtain information on a small area such as on a city, county, district, rural or urban area.

Small area is defined as a subpopulation with a small samplesize. Small Area Estimation is a method that can deal with this problem. Small area estimation technique uses data from a large domain such as census data and data from National Social Economic Survey (SUSENAS) to estimate variables on a smaller domain which are observed Direct estimation on a small area can not produce an accurate estimation[11]. Direct-estimation in a small area will obtain a large variance if the samples are taken from survey data which are designed for a large scope such as a national scope.To handle this problem,indirect estimation will be used by adding covariate for estimating the parameters. The covariate can be taken from other area that has similarities, previous survey in the same area, or other variables that related to the predicted variables. Some indirect-estimation procedures were used to obtain such accuracy.

In general, small area estimation uses parametric modeling to connect the small area statistics with supported variables. Rao clearly describes various techniques of small area estimation that often used, which are synthetic approach, composite estimation, Estimated Best Linear Unbiased Predictor (EBLUP), Empirical Bayes and Hierarchy Bayes. All of these small area estimation use parametric procedures [11]. In fact, many models have nonlinear pattern of relationships between small area mean and covariates. Kismiantini found that adding or not adding covariates in the model yield a same estimation [4]. This is due to the nonlinear relationship between the direct estimation and the covariate. In such case, parametric modeling becomes less flexible. To overcome this case, nonparametric approach was developed in small area estimation.

According to some experts, some nonparametric approaches that can be used in small area estimation are the Kernel approach [5] and local polynomial regression [5,6]. Fridayanti stated unbiased and diversity of Kernel nonparametric model in nonlinear relationship patterns is better than the Fay-Herriot model which is a parametric model of small area estimation [2,3]. Local Polynomial regression, the other nonparametric methods, can be done to accommodate the nonlinear relationship in the statistical estimation in small area. In this research, the comparison between Kernel nonparametric methods and local polynomial regression will be studied in conducting small area estimation.

2. Basic Concept

Basic model of small area estimation can be divided into basic area level and basic unit level model [11]. Basic area level model is a model based on the availability of supported data that available only for a certain area level. Let $\mathbf{x_i} = (\mathbf{x_{1i}}, \ldots, \mathbf{x_{pi}})^{\mathbf{T}}$. $\mathbf{x_i}$ is a vector, i is the number of areas and p is the number of supported variables, and the parameters to be estimated are assumed to have relation with $\mathbf{x_i}$. The supported data were used to build the model:

$$\theta_i = x_i^T \beta + v_i \quad (1)$$

with $v_i \sim N(0, \sigma_v^2)$.

Parameter θ_i can be determined by assuming that the direct estimation model:

$$y_i = \theta_i + \epsilon_i \text{ is available} \quad (2)$$

with $e_i \sim N(0, \sigma_{ei}^2)$ and σ_{ei}^2 is known. By combining (1) and. (2), then a particular model from mixed linear model is ditermined as followed (3) :

$$y_i = x_i^T \beta + v_i + e_i \quad (3)$$

area-based model with one covariat, equation (1) and can be expressed as

$$\theta_i = \beta_0 + \beta_1 x_i + u_i \quad (4)$$

where ϵ_i and u_i are independently distributed as $N(0, D_i)$ and $N(0, \sigma_u^2)$[2].

Basic unit level model is a model in which the available supporting data and the response data are individually corresponded, ie $\mathbf{x_{ij}} = (\mathbf{x_{ij1}}, \ldots, \mathbf{x_{ijp}})^T$, so a nested regression model can be builtas followed $y_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta} + v_i + e_i$ with $v_i \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. However, the model is complex so that it requires relatively more complicated techniques to finish it[11].

In most small area estimation applications, mixed linear modelassumption is used. The estimation issensitive to this assumption. If there is no linearity between the small area average and the covariate, then taking power from other area using a linear model is not appropriate. Mukhopadhyay and Maiti used the following model [5]:

$$y_i = \theta_i + \epsilon_i \quad (7)$$

$$\theta_i = m(x_i) + u_i \quad (8)$$

where i = 1, 2, ..., m is the number of small areas. Function m (.) is a smooth function that defines the relationship between x and y. θ_i is a small area average that is not observed, y_i is a direct estimation of the small area average, u_i is a random error with freely and identically distributed, $E(u_i) = 0$ and $V(u_i) = \sigma_u^2$, ϵ_i is a free random sample , $E(\epsilon_i) = 0$ and $V(\epsilon_i) = D_i$, (assuming that D_i is known). Substituting equations (7) and (8) will yield the following equation

$$y = m(x_i) + u_i + e_i, \quad \text{i=1,2,…,n} \quad (9)$$

To determine the best model, an evaluation of the parameter estimate was conducted. One of the good parameter estimator characteristics is unbiased and small variance. The unbiased estimator is indicated by the absolute value of the relative bias or Absolute Relative Bias (ARB) and Mean Square Error (MSE). Formulations for calculating the ARB is as followed:

$$ARB(\hat{\theta}_i) = \frac{1}{R}\left|\frac{\sum_{j=1}^{R}(\hat{\theta}_{ij} - \theta_{ij})}{\theta_{ij}}\right| \quad (10)$$

Meanwhile, the formulation for calculating the MSE is:

$$MSE(\hat{\theta}_i) = \frac{1}{R}\sum_{j=1}^{R}(\hat{\theta}_{ij} - \theta_{ij})^2 \quad (11)$$

In addition, Eubank states that the size of the regression curve estimation performance can be determined from the MSE, the risk function (P), and the Generalized Cross-Validation (GCV) [1]. Each function will be outlined as followed:

1) Mean Square Error (MSE)

In small area estimation, the MSE value was calculated through the following stages [9]. First, MSE $\theta^* = \gamma_i y_i + (1-\gamma_i)m(x_i)$had to be calculated fromθ_i. And then calculate the squared average of the distance between $\theta^*$and$\tilde{\theta}_i$, in which the m (.) was replaced by $\hat{m}(.)$. Finally, calculate the squared average of the distance between $\tilde{\theta}_i$dan$\hat{\theta}_i$that depended on the$\sigma_u^2$by$\hat{\sigma}_u^2$. This calculation was biased. Therefore [9] and also [7] suggest to useBootstrap estimation to determine the value of the MSE. Bootstrap MSE estimator is given by:

$$mse^*\left(\hat{\theta}_i\right) = \frac{1}{J}\sum_{j=1}^{B}\left(\hat{\theta}_i^{*(j)} - \theta_i^{*(j)}\right)^2 \quad (12)$$

in which *J* is the number of bootstrap population, $\hat{\theta}_i^{*(j)}$is the i-th small area average estimator from the jth bootstrap population, and$\theta_i^{*(j)}$ is the 'true value' of the small area average estimator from the jth bootstrap population.

2) Risk Faktor (P) and GCV

The relationship between the MSE and the risk function (P) by Eubank is as followed [1]:

$$\hat{P} = MSE + 2\hat{\sigma}^2\frac{tr(H)}{n} \quad (13)$$

In which **H** is $\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathrm{X}$.

Meanwhile, the relationship between GCV and MSE is shown by the following formula [1].

$$GCV = \frac{MSE}{[n^{-1}tr(\boldsymbol{I}-\boldsymbol{H})]^2} \quad (14)$$

3. Methodology

The data used issimulation data that generatedby using R. The simulation data obtained from the 2 (two) possible functions m (x), the pattern of linear and quadratic relation. Both functions represent the explicit function pattern which is geometrically a nonparametric curve models (*www.me.mtu.edu*). The two models are: (1) $m(x) = 3x + 1$ and (2) $m(x) = x^2 + 2x + 1$.

The number of small areas (m) that weregenerated were m = 20, m = 30 and m = 40. The process can not be done in the case of m smaller than 20, for example if m = 10 there is an error in the calculation of the variance on the local polynomial regression method (LPR). Thus, for a small size data in this study is represented by m = 20. In performing the simulation, if the area estimation is small then the data size is also small.

The value of σ^2 used are 0.5; 1.0, and 1.5, while D chosen is equal to 1 to simplify the calculations.The width of the bandwidth used is $h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$ which is known as*Silverman's rule of thumb*. In this study, simulations performed on 36 possible combinations of data, 18 simulations on Kernel method and 18 simulations on LPR method. Steps of the simulations performed on these Kernel and LPR methods. Kernel Simulation approach has been done by Fridayanti[3]. However, the model and the bandwidth used is different. The simulation steps are as follows:

1. Generating$x$ dari $uniform(-1,1)$

2. For every i, generate $\theta_{i,j} = m(x_i) + u_{i,j}$ and $y_{i,j} = \theta_{i,j} + \epsilon_{i,j}$ in which $u_{i,j} \sim N(0, \sigma_u^2)$ and $\epsilon_{i,j} \sim N(0,1)$ are independent, j=1,2,...n and x_i is from step (1). By using $m(x) = 3x + 1$ and $m(x) = x^2 + 2x + 1$.
3. $\theta_i = \sum_{j=1}^{n} \frac{\theta_{i,j}}{n}$ and $y_i = \sum_{j=1}^{n} \frac{y_{i,j}}{n}$,for i = 20, 30, and 40
4. The bandwidth used is the same bandwidth with *Kernel'sh* $= \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$ because it is an optimum bandwidth.
5. Calculating $\hat{\sigma}_u^2 = \max\left\{0, \frac{1}{n-1}\sum_{i=1}^{n} W_{hi}(x)\{y_i - \hat{m}(x_i)\}^2 - 1\right\}$, in which $W_{hi}(x) = \frac{K_h(x-x_i)}{1/m \sum_i K_h(x-x_i)}$ and $K_h(x - x_i) = \frac{1}{h}K((x - x_i)/h)$, K is standard normal Kernel function in which$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$.
6. Calculating $\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)\hat{m}(x_i)$ in which $\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + 1}$

   The calculation of $mse(\hat{\theta}_i)$ was done by using *bootstrap* in the following way:
   a. Generating *bootstrap* sample $\theta_i^* \sim N(\hat{m}(x_k), \hat{\sigma}_u^2)$ and $y_i^* | \theta_i^* \sim N(\theta_i^*, 1)$, k=1,..,20 for n=20; k=1,2,…,30 for n=30;and k=1,2,...,40 for n=40.
   b. With x_i and y_i^* , $\hat{m}^*(x_k)$ is resulted from *k smooth* and calculate $\hat{\sigma}_u^{2*} = \max\left\{0, \frac{1}{n-1}\sum_{i=1}^{n} W_{hi}(x)\{y_i^* - \hat{m}^*(x_i)\}^2 - 1\right\}$ then substitute it to$\hat{\theta}_i^* = \hat{\gamma}_i^* y_i^* + (1 - \hat{\gamma}_i^*)\hat{m}^*(x_{ik})$ with $\hat{\gamma}_i^* = \frac{\hat{\sigma}_u^{2*}}{\hat{\sigma}_u^{2*} + 1}$
   c. Repeat (a) and (b) 1000 times in order to obtain:

$$\widehat{mse}_{bs}^*(\hat{\theta}_i) = \frac{1}{1000}\sum_{k=1}^{1000}\left(\hat{\theta}_{ik}^* - \theta_{ik}^*\right)^2$$

7. Calculate the value of the relative bias (ARB) in equation (22)using *bootstrap*, by repeating step (2) to step (6) 100 times, then foreach simulation (r = 1,..,100) calculate the ith small area relative bias :

$$RB(\hat{\theta}_i) = \frac{1}{100}\left|\frac{\sum_{r=1}^{100}\hat{\theta}_i^{(r)} - \theta_i^{(r)}}{\theta_i^{(r)}}\right|$$

8. Calculate the value of GCV in equation (26) and P(λ) in equation (25) by using MSE resulted from step (6).

Steps on LPR approach:

1. Generating $x$from$uniform(-1,1)$
2. For i, generate $\theta_{i,j} = m(x_i) + u_{i,j}$ and $y_{i,j} = \theta_{i,j} + \epsilon_{i,j}$ with $u_{i,j} \sim N(0, \sigma_u^2)and\epsilon_{i,j} \sim N(0,1)$are independent , j=1,2,...n and x_i are from step(1). By taking: $m(x) = 3x + 1$and$m(x) = x^2 + 2x + 1$
3. $\theta_i = \sum_{j=1}^{n} \frac{\theta_{i,j}}{n}$and $y_i = \sum_{j=1}^{n} \frac{y_{i,j}}{n}$, for i = 20, 30, and 40
4. The bandwidth used for calculating P_1 is the same bandwidth with *Kernel,* $h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$ because that bandwidth is an optimum bandwidth.
5. Do data entry by entering data $X_{p1}(x)$ into matrix form (equation 9). The matrix is adjusted for linear model (p=1), quadratic model (p=2), and cubic model (p=3).
6. Calculate matrix $W_{p1}(x)$ as equation(18) for n=20, 30, 40 respectively.
7. Determine thematrix P_1byentering data $[P_1]_{ij} = e_1^T[X_{p1}(x_i)^T W_{p1}(x_i) X_{p1}(x_i)]^{-1} X_{p1}(x_i)^T W_{p1}(x_i) e_{j_{p1}}(x)$ into each cell of the matrix , with vector e_i=1. The rank of the e_i is adjusted with the polynomial degree.

8. Repeat step(7) with different bandwidth, named P_2. Based on the simulation result, for producing a different bandwidth which is not very different from the optimum bandwith (in accordance with *Silverman's rule of tumb*) , using random number and variance of 0,07.
9. Determine another matrix, P_2 , P_1 matrix with different bandwidth. Based on the simulation result , by using random normal number with variance 0.07, we get different bandwidth, that is not far from optimum bandwidth (in accordance with *Silverman's rule of tumb*) .
   Determine:
   a. $r = y - P_1 y$.
   b. $\Delta_1 = diag\{P_1 P_1^T - 2P_1\}$and
   c. $\Delta_2 = \text{diag}\left\{\text{D} + \text{P}_1\text{DP}_1^{\text{T}} - 2\text{P}_1\text{D}\right\}$with $D_i = v\,(x_i)$
10. From step(6), $\widehat{m}(x) = e_i^T P_1 y$ can be determined by$\hat{v}_i = e_i^T \frac{P_2(r^2 - \Delta_2)}{1 + P_2\Delta_1}$
11. Calculate $\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)\widehat{m}(x_i)$ with $\hat{\gamma}_i = (\hat{v}_i + D_i)^{-1}\hat{v}_i$
    $mse(\hat{\theta}_i)$is calculated by using *bootstrap*, as followed:
    a. Generate a sampel of *bootstrap* $\theta^*_{ilokal} \sim N(\widehat{m}(x_{lokal}), \hat{\sigma}_u^2)$,and $y_i^* | \theta^*_{ilokal} \sim N(\theta_i^*, D_i)$, k=1,..,68.
       $\hat{v}_i = \hat{\sigma}_u^2 = e_i^T \frac{P_2(r^2 - \Delta_2)}{1 + P_2\Delta_1}$
    b. With $x_i$andy_i^*, analog with the step above, calculate $\widehat{m}(x_{lokal})$ and the variance estimator , than substitute the variance into

$\hat{\theta}_i^* = \hat{\gamma}_i^* y_i^* + (1 - \hat{\gamma}_i^*)\widehat{m}^*(x_{ilocal})$ with $\hat{\gamma}_i = (\hat{v}_i + D_i)^{-1}\hat{v}_i$

c. Repeat (a) and (b) 1000 times, so:

$$\widehat{mse}^*_{bs}(\hat{\theta}_i) = \frac{1}{1000}\sum_{k=1}^{1000}\left(\hat{\theta}^*_{ilokal} - \theta^*_{ilokal}\right)^2$$

12. Calculate the value of ARB in equation(22) using Bootstrap by repeating step(2) to step(6) 100 times, then for r = 1,..., 100 calculate the relative bias for ith small area

$$RB(\hat{\theta}_i) = \frac{1}{100}\left|\frac{\sum_{r=1}^{100}\hat{\theta}_i^{(r)} - \theta_i^{(r)}}{\theta_i^{(r)}}\right|$$

13. Calculate GCV in equation (14) and P(λ) in equation (13) by using MSE resulted from step (11).

The comparison of the two methods was evaluated through the relative bias value or Absolute Relative Bias (ARB) and the criteria for selecting the best model are MSE (Mean Square Error), GCV (Generalized Cross Validation)and risk factor P (λ). The three criteria are expected to have a minimum value in order to obtain optimum estimator.

4. Results and Discussion

The relative biasedvalue and the three optimal parameter values are determined based onthe simulation to the two methods of small area estimation, the Kernel and the Local polynomial Regression (LPR), on the linear and quadratic data pattern. To obtain a comprehensive result, simulation is done on three types of small data, ie, n = 20, n = 30 and n = 40 on three combinations of variances σ^2 (0.5; 1.0, and 1.5) and D = 1. For a sample size smaller than 20, a simulation using R program can not be done so that the value of these three parameters can be excluded. The choosing of value of variances = 0.5; 1.0, and 1.5 and the value of D = 1 were adopted from Fridayanti[4]. The statistical value to

compare the three kinds of the parameter size of the two methods which are simulated is a summary statistics and boxplot diagram.

a. Simulationon Linear Data Pattern.

As stated in research methods, data pattern examined in this study is linear and quadratic. Linear data pattern is represented by the linear model m (x) = 3 x +1. Relative bias value, MSE, GCV, and risk factor P(λ) are represented on Table 2. From Table 2, it can be seen that for linear data pattern, the relative bias value, for every sample size which is tested with different variance, generally Kernel approach hassmaller relative biasvalue thanLPRs. Likewise, the three parameter values in the both methods tend to have low value on the large sample (n = 40) for various combinations of variances. However, on the LPR methodwith a variance of 1.5 , the three lowest parameter valuesare on the smallest sample (n = 20).

Table 1.Simulation Results on Linear Data Pattern

D	σ^2	N	ARB		MSE		GCV		P(λ)	
			Kernel	LPR	Kernel	LPR	Kernel	LPR	Kernel	LPR
1	0.5	20	9.24	11,986.21	0.93	1,200.25	1.03	13,297.54	0.98	1,200.30
		30	9.24	8,881.29	0.93	888.33	0.99	950.65	0.96	888.37
		40	9.23	5,741.90	0.92	580.00	0.97	610.13	0.95	580.03
	1	20	9.32	55,938.62	0.93	5,650.43	1.04	6,261.58	1.03	5,650.53
		30	9.31	35,974.79	0.93	3,603.03	1.00	3,856.91	1.00	3,603.10
		40	9.17	27,072.76	0.93	2,710.26	0.97	2,851.03	0.98	2,710.31
	1.5	20	10.31	16,974.50	0.94	1,716.33	1.04	1,805.73	1.09	1,716.40
		30	9.32	42,674.22	0.93	4,273.10	1.00	4,734.74	1.03	4,273.25
		40	9.18	19,672.38	0.93	1,989.33	0.98	2,129.85	1.01	1,989.43

From Table 1, it can also be seen that for linear data pattern, generally the three parameter values in the both methods tend to have low value on the large sample (n = 40) for various combinations of variances. However, the LPR method,with a variance of 1.5 , the three lowest parameter value are on the smallest sample (n = 20).

From Figure 1, based on the obtained estimated theta value, drawn through a boxplot with n=40 and variance=0,5, Kernel Method yields a better estimation. The taken sample of n=40 and variance of 0,5 with the case consideration, Kernel Method yields relatively smaller parameter value compared to any other alternative combinations.

Figure 1 shows that the estimated theta resulted by Kernel Method is more compact and symetric compared to LPRs. Although the median value resulted byLPR is relatively lower than the Kernel's, butoverall the value of the parameters generated by Kernel approach is smaller, stable and consistent. So, it can be concluded that using mean as a measure of central data, Kernel method is better than the LPR method in the case of linear data pattern. Kernel approach yieldsan optimum estimation than LPR method in the case of linear data pattern simulation[1].

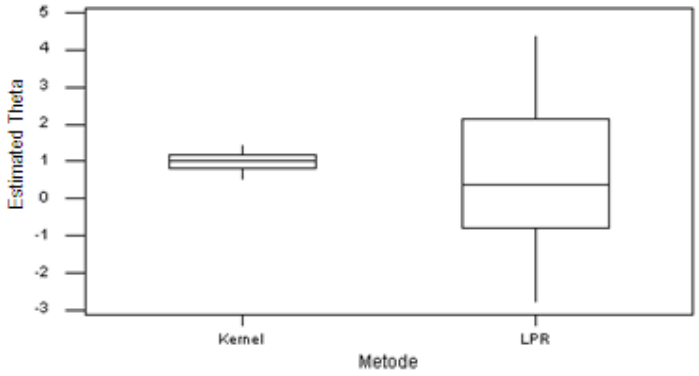

Figure 1. Boxplot of the estimatedtheta by Kernel and LPR on Linear Data Pattern with n=40 andσ^2=0.5

Summary statisticsof the estimated theta and the optimum smoothing parameter value is presented on Table 2. Tabel 2 shows that generally Kernel Method yield a bettersummary statisticsthan LPRdoesthough by using median MSE, GCV, and P(λ) which are resulted are smaller than the Kernel Method does.

Table 2. Summary statistics in Linear Data Pattern on Kernel Method and Local Polynomial Regression

Statistics	Kernel				Local Polynomial Regression (LPR)			
	Estimated Theta	MSE	GCV	P(λ)	Estimated Theta	MSE	GCV	P(λ)
Mean	1.010	0.923	0.970	0.948	0.482	580.002	610.127	580.027
Median	1.021	0.918	0.966	0.943	0.357	0.451	0.474	0.476
Maximum	1.431	0.958	1.008	0.983	4.375	1846.596	1942.794	1846.621
Minimum	0.552	0.904	0.951	0.929	-2.762	0.176	0.186	0.201
Quantil 1	0.848	0.909	0.956	0.934	-0.735	0.258	0.271	0.283
Quantil 3	1.175	0.931	0.980	0.956	1.626	0.705	0.741	0.730

b. Simulationon Quadratic Data Pattern

Other case on non-parametric model is data in quadratic pattern. In this simulation the quadratic model is $m(x) = x^2 + 2x + 1$. The results of the simulationof the relative bias value and the three parameters on various combinations of quadratic data pattern are presented in Table 3. Table 3 shows that on the data with quadratic pattern, generally Kernel gives smaller relative bias and parameter than LPR does. The parameter values given by Kernel tend to be more stable and provide consistent values for any given sample size. These can be seen from the values of MSE, GCV, and P(λ). They are always smaller for n=40 compared tothe values for other two sample sizes in each combination treatment. By using Kernel Methods, with larger sample size, the parameter values in quadratic data pattern are smaller than the parameter values in linear data pattern.

Meanwhile, by using LPR method, parameter values resulted in various combinations treatmentare huge, even different for each given sample size and variance. In general, it appears that the data in quadratic pattern, Kernel methods still produce a more optimum parameter values compared with LPR. Thus, it can be said that Kernel approach on the data quadratic pattern is better than LPR approach. This is indicated by the three parameters discussed in advance.

Table 3. Data Simulation Results on Quadratic Model

D	σ^2	n	ARB		MSE		GCV		P(λ)	
			Kernel	LPR	Kernel	LPR	Kernel	LPR	Kernel	LPR
1	0.5	20	7.77	23,695.26	0.88	6,053.76	0.98	1,778.10	0.93	1,605.81
		30	7.77	22,016.84	0.88	5,614.54	0.94	6,008.42	0.91	5,614.57
		40	7.68	37,470.03	0.87	9,628.76	0.92	8,284.69	0.90	8,628.78
	1	20	7.93	32,854.71	0.90	8,424.47	0.99	8,334.59	1.00	8,424.57
		30	7.79	38,149.99	0.89	9,746.71	0.95	9,490.67	0.95	1,007.61
		40	7.72	35,583.46	0.88	9,073.54	0.93	9,780.66	0.93	4,704.76
	1.5	20	7.95	27,068.94	0.91	6,915.68	1.05	6,631.23	1.06	6,915.83
		30	7.86	6,862.06	0.90	1,740.39	0.96	8,628.59	1.00	1,740.49
		40	7.84	28,563.46	0.89	73,22.25	0.94	7,031.69	0.97	7,322.32

In the case of quadratic data pattern, both variance and sample size used same with in the case of linear data pattern.It can also be seen from Figure 2that the boxplot in LPR methods has two outliers. Meanwhile, the obtained boxplot in Kernel methods tend to be stable and almost equal to those in the linear data pattern, but it has a smaller median value.

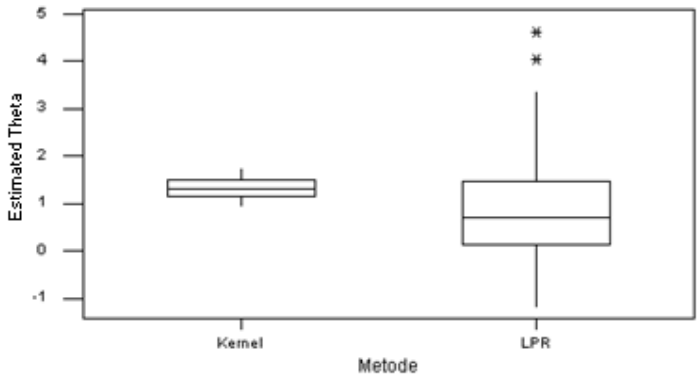

Figure 2.Boxplot of the estimated theta by Kernel and Local Polynomial Regression on quadratic pattern with n=40 and σ^2=0.5

Summary statistics for quadratic data pattern is presented in Table 4.

Table 4. Summary Statistics forQuadratic Data Pattern by Kernel Method and Local Polynomial Regression

Statistical value	Kernel				Local Polynomial Regression (LPR)			
	Estimated Theta	MSE	GCV	P(λ)	Estimated Theta	MSE	GCV	P(λ)
Mean	1.348	0.872	0.918	0.897	0.995	1,295.553	1,365.302	1,292.578
Median	1.316	0.865	0.910	0.890	0.712	14.879	15.652	14.904
Maximum	1.731	0.939	0.988	0.964	4.628	9,728.582	20,743.058	9,728.607
Minimum	1.025	0.853	0.898	0.878	-1.156	0.527	0.554	0.552
Quantil 1	1.182	0.856	0.900	0.881	0.368	3.041	3.199	3.066
Quantil 3	1.484	0.877	0.923	0.902	1.386	1,789.004	1,881.924	1,789.029

5. Conclusions and Suggestions

Conclusions:

1. In the linear and quadratic data pattern, the relative bias values resulted by Kernel method are smaller than the relative bias values resulted by the Local polynomial Regression (LPR). The relative bias values resultedby Kernel method tend to be more stable and consistent than the relative bias values resulted by LPR methods.
2. The Mean Square Errors (MSE)produced by Kernel methods on every combination treatment are smaller than the MSEs obtained by LPR in both linear and quadratic data pattern.
3. Likewise, the values of GCV and the risk factors P(λ) in every combination of simulated data patterns, Kernel methods provide smaller values than the LPR does.

Thus, it can be said that Kernel methods is better than LPR method in dealing with both linear and kuadratic data with non-parametric characteristic.

Suggestions:

1. The optimum bandwidth usage greatly affect thevalues obtained.,The bandwidth used in this study is an optimum bandwidth for Kernel methods.
2. It is need to studyin further about optimum bandwidth determination for LPR in small area estimation on non-parametric data pattern.

References

[1] Eubank, R.L. Spline Smoothing and Nonparametric Regression. Amazon: Marcel Dekker Inc (1988).

[2] Fay, R.E, Herriot, R.A.Estimates income for small places : an application of James-Stein procedures to census data. *Journal of the American Statistical Association*. 74:269-277 (1979).

[3] Fridayanti., *Pendekatan Nonparametrik dalam Pendugaan Statisti Area Kecil.* Tesistidakdipublikasikan. Bogor: Program Pascasarjana, InstitutPertanian Bogor(2009).

[4] Kismiantini., *Pendugaan Statistik Area Kecil Berbasis Model Poisson Gamma*. Tesistidakdipublikasikan. Bogor: Program Pascasarjana, InstitutPertanian Bogor (2007).

[5] Mukhopadhyay P, Maiti T., *Two Stage Non-Parametric Approach for Small Area Estimation*. Proceedings of ASA Section on Survey Research Methods: **4058-4065**(2004).

[6] Mukhopadhyay P,Maiti T., *Local Polynomial Regression for Small Area Estimation*. Proceedings of ASA Section on Survey Research Methods: **3447-3452** (2006).

[7] Opsomer et al., *Nonparametric Small Area Estimation Using Penalized Spline Regression*. Proceedings of ASA Section on Survey Research Methods:**1-8**(2004).

[8] Prasad, N.G.N, dan Rao, J.N.K., The estimation of the mean squared error of the small area estimators. Journal of the American Statistical Association, **85, 163-171** (1990).

[9] Pawitan, Y., *In All Likelihood*. Oxford: Clarendon Press(2001).

[10] Rao, J.N.K., Some recent advances in model-based small area estimation.*Survey Methodology* **25:175-186**(1999).

[11] Rao JNK., *Small Area Estimation*. New York: John Wiley and Sons(2003).
[12] Silverman, B.S., *Density Estimation forStatistical Analysis*. New York: Chapman and Hall(1986).
[13] Takezawa, K., *Introduction Nonparametric Regression*. New York: John Wiley & Sons, Inc(2006).