Journal of Mathematics and Statistics 8 (3): 342-347, 2012
ISSN 1549-3644

Designing of Child Growth Chart Based on Multi-Response Local Polynomial Modeling

[1]Nur Chamidah, [2]I. Nyoman Budiantara, [2]Sony Sunaryo and [2]Ismaini Zain
[1]Department of Mathematics, Faculty of Sciences and Technology,
Airlangga University, Surabaya, Indonesia
[2]Department of Statistics, Faculty of Mathematics and Natural Sciences,
Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

Abstract: Problem statements: Anthropometry measures used to measure physical children growth are not onlyweight but also height and head circumference. In this study we develop the estimation of multi-response localpolynomial regression and apply it to design growth chart for children up to five years old based on three responsevariables i.e., weight, height and head circumference. **Approach:** Based on local polynomial estimator, wedescribe the estimation of multi-response nonparametric regression model by using weighted least squared. Themodel is applied to design health card of children up to five years old by using children data in Surabaya, Indonesia. Generalized Cross Validation (GCV) method is used to determine the order of local polynomial fit and the bandwidthfor each response variable. **Results:** We formulate the multi-response local polynomial modeling and give a design of health card of children up to five years old in Surabayacity, Indonesia. **Conclusion:** The child growth chart based on multi-response local polynomial modeling showsincreasing of children nutrition in Surabaya 2010.Because of the strong correlations among all three response variables,the simultaneosly approach for model estimationis better than partly single response approach. The result of simultaneosly model estimation based on multi-response local polynomial modeling satisfies goodness of fit criterion i.e., mean squared error value tend to zero and determination coefficient value tend to one.

Key words: Local polynomial, multi-response, Generalized Cross Validation (GCV), growth chart

INTRODUCTION

There are many cases that involve the regressionmodel has more than one response variables thatcorrelate each others. In that case, the multi-responsenonparametric regression model provides powerfultools to model the functions which draw association ofthese variables. Local polynomial estimation is widelyused for estimating regression function because it issimple and easy to understand. Localpolynomial estimator is obtained by locally fitting a dth degree polynomial to data by weighted least squares. The local polynomial estimator depends on twoparameters, which must be specified i.e., the order oflocal polynomial fit (d) and the smoothing parameternamed bandwidth (h). These two parameters have asimilar effect, in that a higher order fit or smallerbandwidth reduces bias but increases variance; while alower order or larger bandwidth increases biasbut reduces variance.

Many authors studied multi-responsenonparametric regression model. Wang *et al*. (2000) proposed spline smoothing for estimatingnonparametric functions from bivariate data with thesame variances of errors for each same response (i.e., $\sigma_j^2, j = 1,2$) and applied it to hormone data. Welsh and Yee (2006) considered bi-response local linearregression and applied to blood pressure data which hasresponse variables i.e., systolic and diastolic andpredictor variable i.e., Body Mass Index (BMI). Lestari *et al*. (2010) studied the estimating of multi-responsenonparametric regression based on spline estimator. In case of heteroscedasticity, Chamidah (2012) studied estimation of biresponses local polynomialregression model and applied the model to estimategrowth curve of children up to 5 years of age basedon their weight and height.

According to pediatrician Roumeliotis (2012) the growthof childrenduring the first 18 months grows rapidly andthen it decreases parallel with increasing of age. It means locally model approach moreappropriate

Corresponding Author: Nur Chamidah, Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia

to this data. Also, Roumeliotis (2012) stated anthropometrymeasures which is used to measure physical child growth are weight,height and head circumference of children. It meansthat physical child growth is more realistic if it ismodeled by multi-response nonparametricregression approach.

In this study, we discuss the estimation ofmulti-response local polynomial modeling whichhas bandwidth and polynomial degree of eachresponse must not be equal. For determining thesesmoothing parameters, we use GCV method andthen apply the model to children growth data inSurabaya, Indonesia 2010. It is necessary fordesigning health card that in Indonesia is called as KartuMenuju Sehat (KMS) based on children condition inIndonesia. Currently, the KMS is used formonitoring health and growth children in Indonesiabased on National Center Health Statistics (NCHS)chart, USA. The chart may not appropriate to thecondition of Indonesian children.

MATERIALS AND METHODS

Given multi-response nonparametric regressionas follows Eq. 1:

$$\underset{\sim}{y}_i = \underset{\sim}{f}(t_i) + \underset{\sim}{\varepsilon}_i \quad , \ i = 1,2,...,n \qquad (1)$$

where, $\underset{\sim}{f}(t_i) = (f_1(t_i), f_2(t_i), ..., f_r(t_i))^T$ is a vector of theunknown smooth function, $\underset{\sim}{y}_i = (y_i^{(1)}, y_i^{(2)}, ..., y_i^{(r)})^T$ and $\underset{\sim}{\varepsilon}_i = (\varepsilon_i^{(1)}, \varepsilon_i^{(2)}, ..., \varepsilon_i^{(r)})^T$, Assume $\underset{\sim}{\varepsilon}_i \sim \text{i.i.d.}(\underset{\sim}{0}, \Sigma)$ where:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1r}\sigma_1\sigma_r \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{12}\sigma_2\sigma_r \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{r1}\sigma_r\sigma_1 & \rho_{r2}\sigma_r\sigma_2 & \cdots & \sigma_r^2 \end{bmatrix}$$

We can write the multi-response nonparametricregression model (1) into vector expression as follows Eq. 2:

$$\underset{\sim}{Y} = \underset{\sim}{f} + \underset{\sim}{\varepsilon} \qquad (2)$$

where:

$$\underset{\sim}{Y} = \left(\underset{\sim}{y}^{(1)}, \underset{\sim}{y}^{(2)}, ..., \underset{\sim}{y}^{(r)}\right)^T, \underset{\sim}{f} = \left(\underset{\sim}{f}^{(1)}, \underset{\sim}{f}^{(2)}, ..., \underset{\sim}{f}^{(r)}\right)^T, \underset{\sim}{\varepsilon} = \left(\underset{\sim}{\varepsilon}^{(1)}, \underset{\sim}{\varepsilon}^{(2)}, ..., \underset{\sim}{\varepsilon}^{(r)}\right)^T$$

By expression (2), of course, we can take variance oferrors vector $\underset{\sim}{\varepsilon}$ easily:

$$\text{var}(\underset{\sim}{\varepsilon}) = E[\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon})][\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon})]^T = E[\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}^T]$$

$$= E\begin{bmatrix} \varepsilon_1^{(1)}\varepsilon_1^{(1)} & \cdots & \varepsilon_1^{(1)}\varepsilon_n^{(1)} & \varepsilon_1^{(1)}\varepsilon_1^{(2)} & \cdots & \varepsilon_1^{(1)}\varepsilon_n^{(2)} & \cdots & \varepsilon_1^{(1)}\varepsilon_1^{(r)} & \cdots & \varepsilon_1^{(1)}\varepsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \ddots \\ \varepsilon_n^{(1)}\varepsilon_1^{(1)} & \cdots & \varepsilon_n^{(1)}\varepsilon_n^{(1)} & \varepsilon_n^{(1)}\varepsilon_1^{(2)} & \cdots & \varepsilon_n^{(1)}\varepsilon_n^{(2)} & \cdots & \varepsilon_n^{(1)}\varepsilon_1^{(r)} & \cdots & \varepsilon_n^{(1)}\varepsilon_n^{(r)} \\ \varepsilon_1^{(2)}\varepsilon_1^{(1)} & \cdots & \varepsilon_1^{(2)}\varepsilon_n^{(1)} & \varepsilon_1^{(2)}\varepsilon_1^{(2)} & \cdots & \varepsilon_1^{(2)}\varepsilon_n^{(2)} & \cdots & \varepsilon_1^{(2)}\varepsilon_1^{(r)} & \cdots & \varepsilon_1^{(2)}\varepsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \varepsilon_n^{(2)}\varepsilon_1^{(1)} & \cdots & \varepsilon_n^{(2)}\varepsilon_n^{(1)} & \varepsilon_n^{(2)}\varepsilon_1^{(2)} & \cdots & \varepsilon_n^{(2)}\varepsilon_n^{(2)} & \cdots & \varepsilon_n^{(2)}\varepsilon_1^{(r)} & \cdots & \varepsilon_n^{(2)}\varepsilon_n^{(r)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \varepsilon_1^{(r)}\varepsilon_1^{(1)} & \cdots & \varepsilon_1^{(r)}\varepsilon_n^{(1)} & \varepsilon_1^{(r)}\varepsilon_1^{(2)} & \cdots & \varepsilon_1^{(r)}\varepsilon_n^{(2)} & \cdots & \varepsilon_1^{(r)}\varepsilon_1^{(r)} & \cdots & \varepsilon_1^{(r)}\varepsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \varepsilon_n^{(r)}\varepsilon_1^{(1)} & \cdots & \varepsilon_n^{(r)}\varepsilon_n^{(1)} & \varepsilon_n^{(r)}\varepsilon_1^{(2)} & \cdots & \varepsilon_n^{(r)}\varepsilon_n^{(2)} & \cdots & \varepsilon_n^{(r)}\varepsilon_1^{(r)} & \cdots & \varepsilon_n^{(r)}\varepsilon_n^{(r)} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \cdots & 0 & \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & \cdots & \rho_{1r}\sigma_1\sigma_r & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_1^2 & 0 & \cdots & \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & \cdots & \rho_{1r}\sigma_1\sigma_r \\ \rho_{12}\sigma_1\sigma_2 & \cdots & 0 & \sigma_2^2 & \cdots & 0 & \cdots & \rho_{2r}\sigma_2\sigma_r & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_{12}\sigma_1\sigma_2 & 0 & \cdots & \sigma_2^2 & \cdots & 0 & \cdots & \rho_{2r}\sigma_2\sigma_r \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \rho_{r1}\sigma_r\sigma_1 & \cdots & 0 & \rho_{r2}\sigma_r\sigma_2 & \cdots & 0 & \cdots & \sigma_r^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_{r1}\sigma_r\sigma_1 & 0 & \cdots & \rho_{r2}\sigma_r\sigma_2 & \cdots & 0 & \cdots & \sigma_r^2 \end{bmatrix} = V$$

Matrix V is used to obtain weighted matrix forestimating.

In the local polynomial regression, there are twoparameters, i.e., the bandwidth (h) and the order of localpolynomial fit (d) that control the smoothness of thefit and also affect the bias-variance trade-off. Weuse generalized cross-validation method forchoosing optimal h and d of each response.

Estimation of the function $\hat{f}_j(t_i)$ for each responseis:

$$\hat{f}_j(t_i) = A(h_j)\underset{\sim}{y}^{(j)}, \ j = 1,2,...,r$$

where,A (h_j)represents matrix as follows:

$$A(h_j) = e_1^T \left[X_{ti}^{(dj)T} K_{hj}(t_i) X_{ti}^{(d_j)} \right]^{-1} X_{t_t}^{(d_j)T} K_{hj}(t_i)$$

To obtain optimal $h_j$and $d_j$based on GCVmethod given by Wu and Zhang (2006), we minimize:

$$GCV(h_j) = \frac{n^{-1}\sum_{i=1}^{n}(y_i^{(j)} - \hat{y}_i^{(j)})^2}{(n^{-1}tr[I - A(h_j)])^2}$$

RESULTS

From model (1), we estimate every element of $\underset{\sim}{f}(t) = (f_1(t), ..., f_r(t))^T$ by local polynomial fitting which uses Taylor expansion approach.By Taylor expansion, for t in a neighborhood of t_0,we have Eq. 3:

$$\left.\begin{array}{l} f_1(t) \approx f_1(t_0) + (t-t_0)f_1(t_0) + \\ (t-t_0)^2 \frac{f_1^{(2)}(t_0)}{2!} + \ldots + (t-t_0)^{d1} \frac{f_1^{(d_1)}t_0}{d_1!} \\ f_2(t) \approx f_2(t_0) + (t-t_0)f_2(t_0) + (t-t_0)^2 \\ \frac{f_2^{(2)}(t_0)}{2!} + \ldots + (t-t_0)^{d_2} \frac{f_2^{(d_2)}(t_0)}{d_2!} \\ \vdots \\ f_r(t) \approx f_r(t_0) + (t-t_0)f_r(t_0) + \\ (t-t_0)^2 \frac{f_r^{(2)}(t_0)}{2!} + \ldots + (t-t_0)^{d_r} \frac{f_r^{(d_r)}(t_0)}{d_r!} \end{array}\right\} \quad (3)$$

In terms of statistical modeling locally around t_0,we model (3) as follows Eq. 4:

$$\begin{array}{l} f_j(t) \approx \beta_0^{(j)}(t_0) + \beta_1^{(j)}(t_0)(t-t_0) + \beta_2^{(j)}(t_0)(t-t_0)^2 \\ +\ldots+\beta_{d_1}^{(j)}(t_0)(t-t_0)^{d_1}, j=1,\ldots,r \end{array} \quad (4)$$

Where:

$$\beta_I^{(j)}(t_0) = \frac{f_j^{(I)}(t_0)}{I!}, I = 0,1,2,\ldots d_j$$

To estimate model (4) from the sample data $\{t_i, \underset{\sim}{y}_i\}_{i=1}^n$ we use local polynomial estimator definedas follows:

$$\min_{\underset{\sim}{\beta}(t_0)} Q(t_0) = (\underset{\sim}{Y} - X_{t0}\underset{\sim}{\beta}(t_0))^T V^{-1} K_h(t_0)(\underset{\sim}{Y} - X_{t0}\underset{\sim}{\beta}(t_0)) \quad (5)$$

where K(.) is a Kernel function, h > 0:

$$K_h(t_0) = \text{diag}\left\{\underset{\sim}{K}_{h_1}(t_1-t_0), \underset{\sim}{K}_{h_2}(t_i-t_0), \ldots, \underset{\sim}{K}_{h_r}(t_i-t_0)\right\}$$

$$\underset{\sim}{K}_{h_j}(t_i-t_0) = \left(K_{hj}(t_1-t_0), K_{h_j}(t_2-t_0), \ldots, K_{h_j}(t_n-t_0)\right)^T$$

$$\underset{\sim}{\beta}(t_0) = \begin{bmatrix} \underset{\sim}{\beta}_{d_1}^{(1)}(t_0) \\ \underset{\sim}{\beta}_{d_2}^{(2)}(t_0) \\ \vdots \\ \underset{\sim}{\beta}_{d_r}^{(r)}(t_0) \end{bmatrix}, \underset{\sim}{\beta}_{dj}^{(j)}(t_0) = \left(\beta_0^{(j)}(t_0), \beta_1^{(j)}(t_0), \ldots, \beta_{d_j}^{(j)}(t_0)\right)^T$$

$$X_{t_0} = \begin{bmatrix} X_{t_0}^{(d_1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{t_0}^{(d_r)} \end{bmatrix}$$

$$X_{t_0}^{d_j} = \begin{bmatrix} 1 & (t_1-t_0) & \cdots & (t_1-t_0)^{d_j} \\ 1 & (t_2-t_0) & \cdots & (t_2-t_0)^{d_j} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & (t_n-t_0) & \cdots & (t_1-t_0)^{d_j} \end{bmatrix}$$

and V^{-1} is invert of variance-covariance matrix of errorswhich is estimated from sample data. The solution of Eq. 5 is:

$$\hat{\underset{\sim}{\beta}}(t_0) = \left(X_{t_0}^T V^{-1} K_h(t_0) X_{t_0}\right)^{-1} X_{t_0}^T V^{-1} K_h(t_0) \underset{\sim}{Y} \quad (6)$$

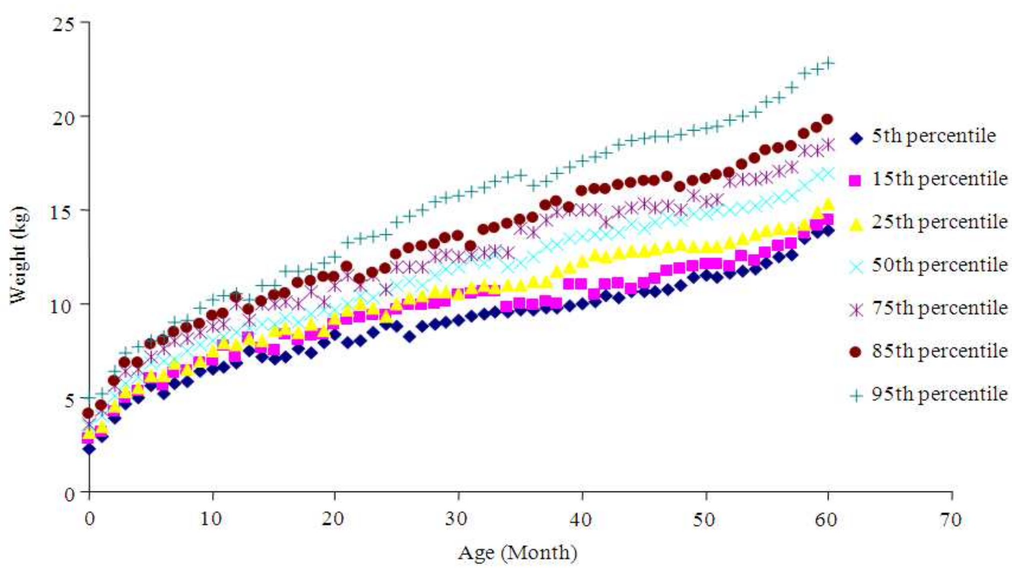


Fig. 1: Percentiles of weight children plot

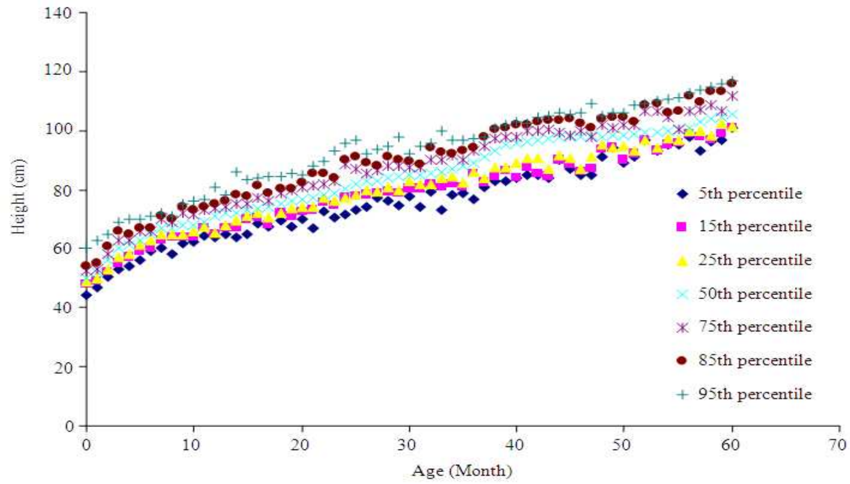


Fig. 2: Percentiles of height children plot

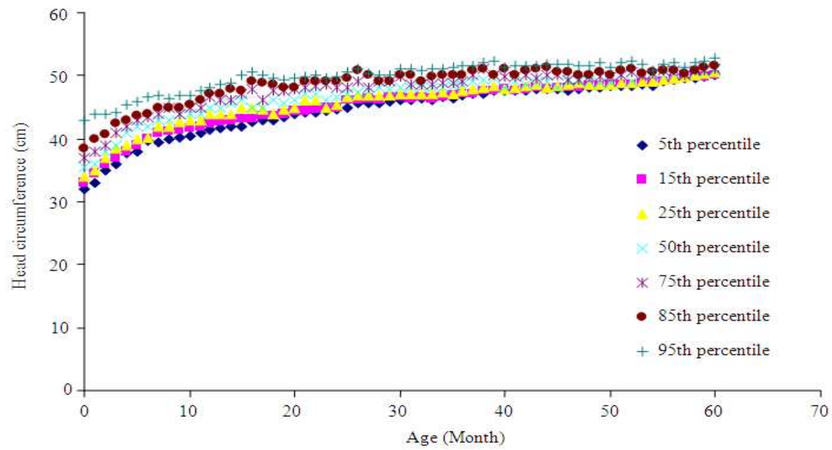


Fig. 3: Percentiles of head circumference childrenplot

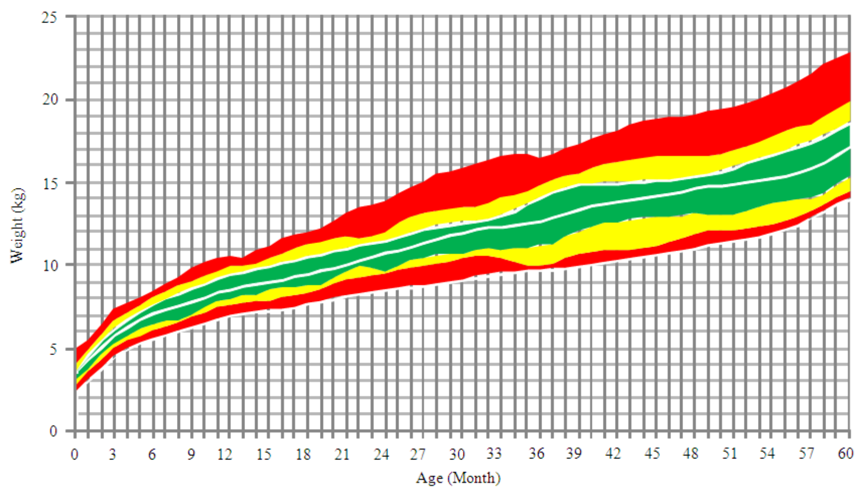
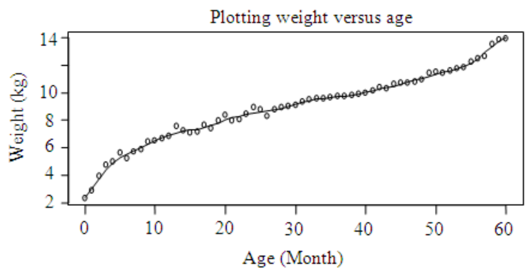
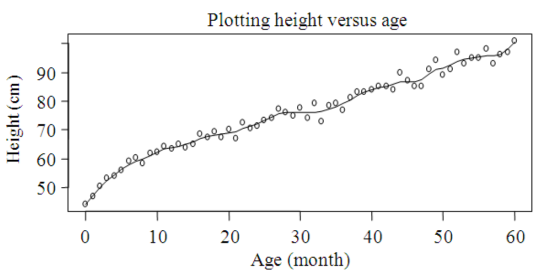


Fig. 4: Design of KMS for children in Surabaya



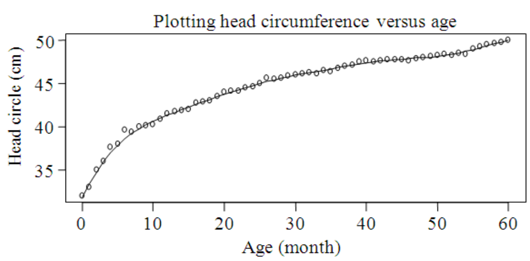
Point: Observation point dan solid line: fitted curve

(a)



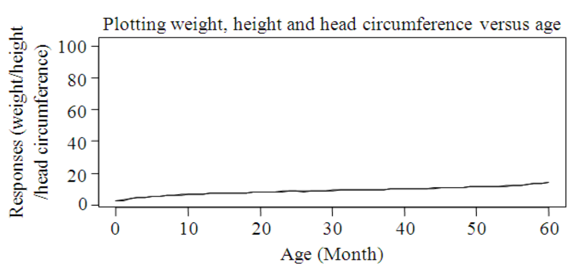
Point: Observation point dan dash line: Fitted curve

(b)



Point: Observation point dan dot dash line: Fitted curve

(c)



Solid line: Plotting weight/age, dot dash line: Plotting head circumference/age dan dash line: Plotting height/age

(d)

Fig. 5: Plots the 50th percentiles estimation of weight, height and head circumferenceversus age

So, from Eq. 6 we get the local polynomialestimator of $\underset{\sim}{f}(t_i)$ Eq. 7:

$$\hat{\underset{\sim}{f}}(t_i) = e^T (X_{t_i}^T V^{-1} K_h(t_i) X_{t_i})^{-1} X_{t_i}^T V^{-1} K_h(t_i) \underset{\sim}{Y} \quad (7)$$

Table 1: Optimal bandwidth, polynomial order and GCV value

Percentil	Response	Optimal order	Optimal bandwidth	GCV value
5	(y1)	3	2.64	0.049270
	(y2)	1	0.13	0.042860
	(y3)	2	4.70	4.279840
15	(y1)	2	1.77	0.025560
	(y2)	2	2.46	0.072490
	(y3)	3	8.35	2.643760
25	(y1)	1	0.63	0.114120
	(y2)	1	1.12	0.045420
	(y3)	3	7.83	2.214360
50	(y1)	3	2.72	0.212460
	(y2)	3	2.02	0.028280
	(y3)	3	1.06	0.152698
75	(y1)	2	2.83	0.252890
	(y2)	2	2.96	0.113100
	(y3)	2	3.57	2.535190
85	(y1)	1	0.87	0.238570
	(y2)	2	2.83	0.068760
	(y3)	2	3.70	2.972110
95	(y1)	2	0.67	0.103140
	(y2)	3	1.32	0.034180
	(y3)	1	2.63	2.770170

where $e^T = \left(e_1, 0, \ldots, e_{d_1+2}, 0, \ldots, e_{d_1+d_2+3}, 0, \ldots, e_{d_1+d_2+\ldots+d_{r-1}+r}\right)^T$ is vector whichhas elements $e_1 = 1, e_{d_1+2} = 1, e_{d_1+d_2+3} = 1, \ldots, e_{d_1+d_2+\ldots+d_{r-1}+r} = 1$ and 0 elsewhere.

The data used for applying the modelcontainsof 1700 children obtained from community healthcenter in Surabaya 2010. The data that describes childrengrowth in Surabaya consists of 3 response variables.These are y_1: weight (kg), y_2: height (cm) and y_3:head circumference. (cm). While a predictor variable isage (month). In every month of children age i.e., from 0 until 60, we determine 5th percentile, 15th percentile, 25th percentile, 50th percentile, 75th percentile, 85th percentile and 95th percentile.

Plotting of percentiles of weight, height and headcircumference versus age are shown in Fig. 1-3, respectively.

Based on GCV method, we create R-code forchoosing bandwidth and optimal order of polynomial for each response. These results are given in Table 1.

These results given in Table 1 are used for designing child growth chart in KMS based on the three responses local polynomial estimation. The chart is shown in Fig. 4 as follows. Based on Table 1, the results of the 50th percentiles estimation of weight, height and head circumference versus age give the mean squared error value 0.0514 and coefficient determination 99%. Plots the estimation of 50th percentiles of weight, height and head circumference versus age are shown in Fig. 5.

DISCUSSION

In Fig. 4, green area indicates good health, lower yellow area indicates warning for underweight and upper yellow area indicates warning for overweight. Lower red area indicates underweight and upper red area indicates overweight. Based on correlation Pearson formula, we get correlation coefficient between weight and height of children 0.996; correlation coefficient between weight and head circumference of children 0.953; and correlation between height and circumferenceof children 0.946. It means that there are strong correlations among all three response variables. Design of the KMS of child growth in Surabaya 2010 as given in Fig. 4 is quite higher than that currently used to control children health in Surabaya. The simultaneously estimationgives mean squared errorvalue tend to zeroand determinationcoefficient value tend to one. These facts mean that the simultaneously model has satisfied goodness of fit criterion.

CONCLUSION

Designing of child growth chart based on multi-responselocal polynomial modeling shows increasingof children nutrition in Surabaya 2010. Because of the strongcorrelationsamong all three response variables, the simultaneosly three responses model estimationis better than partly single response model estimation.The result of simultaneosly model estimation based on multi-response local polynomial modeling satisfies goodness of fit criterion.

ACKNOWLEDGEMENT

This Research was funded by the DirectorateGeneral of Higher Education of Indonesia throughDoctoral Research Grant 2010 with contractnumber: 502/SP2H/PP/DP2M/VI/2010.

REFERENCES

Chamidah, N., 2012. Biresponse local polynomial regression of baby growth curve in case of heteroscedasticity. Airlangga University.

Lestari, B., I.N. Budiantara, S. Sunaryo and M. Mashuri, 2010. Spline estimator in multiresponse nonparametric regression model with unequal correlation of errors. J. Math. Stat., 6: 327-332. DOI: 10.3844/jmssp.2010.327.332

Roumeliotis, P., 2012. Children's health and wellness. Growth and Development.

Wang, Y.,W. Guo and M.B. Brown, 2000. Spline smoothing for bivariate data with applications to association between hormones. Stat. Sinica, 10: 377-397.

Welsh, A.H. and T.W. Yee, 2006. Local regression for vector responses. J. Stat. Plann. Inform., 136: 3007-3031. DOI: 10.1016/j.jspi.2004.01.024

Wu, H. and J.T. Zhang, 2006. Nonparametric Regression Methods for Longitudinal Data Analysis. 1st Edn., John Wiley and Sons, Inc., Hoboken, New Jersey, ISBN-10: 0471483508, pp: 369.