

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

<sup>1</sup>Nur Chamidah, <sup>2</sup>I. Nyoman Budiantara, <sup>2</sup>Sony Sunaryo and <sup>2</sup>Ismaini Zain

<sup>1</sup>Department of Mathematics, Faculty of Sciences and Technology,  
Airlangga University, Surabaya, Indonesia

<sup>2</sup>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 only weight but also height and head circumference. In this study we develop the estimation of multi-response local polynomial regression and apply it to design growth chart for children up to five years old based on three response variables i.e., weight, height and head circumference. **Approach:** Based on local polynomial estimator, we describe the estimation of multi-response nonparametric regression model by using weighted least squares. The model 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 bandwidth for 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 Surabaya city, Indonesia. **Conclusion:** The child growth chart based on multi-response local polynomial modeling shows increasing of children nutrition in Surabaya 2010. Because of the strong correlations among all three response variables, the simultaneously approach for model estimation is better than partly single response approach. The result of simultaneously 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 regression model has more than one response variables that correlate each others. In that case, the multi-response nonparametric regression model provides powerful tools to model the functions which draw association of these variables. Local polynomial estimation is widely used for estimating regression function because it is simple and easy to understand. Local polynomial estimator is obtained by locally fitting a  $d$ th degree polynomial to data by weighted least squares. The local polynomial estimator depends on two parameters, which must be specified i.e., the order of local polynomial fit ( $d$ ) and the smoothing parameter named bandwidth ( $h$ ). These two parameters have a similar effect, in that a higher order fit or smaller bandwidth reduces bias but increases variance; while a lower order or larger bandwidth increases bias but reduces variance.

Many authors studied multi-response nonparametric regression model. Wang *et al.* (2000) proposed spline smoothing for estimating nonparametric functions from bivariate data with the same 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 linear regression and applied to blood pressure data which has response variables i.e., systolic and diastolic and predictor variable i.e., Body Mass Index (BMI). Lestari *et al.* (2010) studied the estimating of multi-response nonparametric regression based on spline estimator. In case of heteroscedasticity, Chamidah (2012) studied estimation of bi-responses local polynomial regression model and applied the model to estimate growth curve of children up to 5 years of age based on their weight and height.

According to pediatrician Roumeliotis (2012) the growth of children during the first 18 months grows rapidly and then it decreases parallel with increasing of age. It means locally model approach more appropriate

---

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

to this data. Also, Roumeliotis (2012) stated anthropometry measures which is used to measure physical child growth are weight, height and head circumference of children. It means that physical child growth is more realistic if it is modeled by multi-response nonparametric regression approach.

In this study, we discuss the estimation of multi-response local polynomial modeling which has bandwidth and polynomial degree of each response must not be equal. For determining these smoothing parameters, we use GCV method and then apply the model to children growth data in Surabaya, Indonesia 2010. It is necessary for designing health card that in Indonesia is called as Kartu Menuju Sehat (KMS) based on children condition in Indonesia. Currently, the KMS is used for monitoring health and growth children in Indonesia based on National Center Health Statistics (NCHS) chart, USA. The chart may not appropriate to the condition of Indonesian children.

**MATERIALS AND METHODS**

Given multi-response nonparametric regression as follows Eq. 1:

$$y_i = f_i(t_i) + \epsilon_i, \quad i = 1, 2, \dots, n \tag{1}$$

where,  $f_i(t_i) = (f_1(t_i), f_2(t_i), \dots, f_r(t_i))^T$  is a vector of the unknown smooth function,  $y_i = (y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(r)})^T$  and  $\epsilon_i = (\epsilon_i^{(1)}, \epsilon_i^{(2)}, \dots, \epsilon_i^{(r)})^T$ , Assume  $\epsilon_i \sim \text{i.i.d.}(0, \Sigma)$  where:

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

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

$$Y = f + \epsilon \tag{2}$$

where:

$$Y = (y^{(1)}, y^{(2)}, \dots, y^{(r)})^T, f = (f^{(1)}, f^{(2)}, \dots, f^{(r)})^T, \epsilon = (\epsilon^{(1)}, \epsilon^{(2)}, \dots, \epsilon^{(r)})^T$$

By expression (2), of course, we can take variance of errors vector  $\epsilon$  easily:

$$\text{var}(\epsilon) = E[\epsilon - E(\epsilon)][\epsilon - E(\epsilon)]^T = E[\epsilon\epsilon^T]$$

$$= E \begin{bmatrix} \epsilon_1^{(1)}\epsilon_1^{(1)} & \dots & \epsilon_1^{(1)}\epsilon_n^{(1)} & \epsilon_1^{(1)}\epsilon_1^{(2)} & \dots & \epsilon_1^{(1)}\epsilon_n^{(2)} & \dots & \epsilon_1^{(1)}\epsilon_1^{(r)} & \dots & \epsilon_1^{(1)}\epsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \epsilon_n^{(1)}\epsilon_1^{(1)} & \dots & \epsilon_n^{(1)}\epsilon_n^{(1)} & \epsilon_n^{(1)}\epsilon_1^{(2)} & \dots & \epsilon_n^{(1)}\epsilon_n^{(2)} & \dots & \epsilon_n^{(1)}\epsilon_1^{(r)} & \dots & \epsilon_n^{(1)}\epsilon_n^{(r)} \\ \epsilon_1^{(2)}\epsilon_1^{(1)} & \dots & \epsilon_1^{(2)}\epsilon_n^{(1)} & \epsilon_1^{(2)}\epsilon_1^{(2)} & \dots & \epsilon_1^{(2)}\epsilon_n^{(2)} & \dots & \epsilon_1^{(2)}\epsilon_1^{(r)} & \dots & \epsilon_1^{(2)}\epsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \epsilon_n^{(2)}\epsilon_1^{(1)} & \dots & \epsilon_n^{(2)}\epsilon_n^{(1)} & \epsilon_n^{(2)}\epsilon_1^{(2)} & \dots & \epsilon_n^{(2)}\epsilon_n^{(2)} & \dots & \epsilon_n^{(2)}\epsilon_1^{(r)} & \dots & \epsilon_n^{(2)}\epsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \epsilon_1^{(r)}\epsilon_1^{(1)} & \dots & \epsilon_1^{(r)}\epsilon_n^{(1)} & \epsilon_1^{(r)}\epsilon_1^{(2)} & \dots & \epsilon_1^{(r)}\epsilon_n^{(2)} & \dots & \epsilon_1^{(r)}\epsilon_1^{(r)} & \dots & \epsilon_1^{(r)}\epsilon_n^{(r)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \epsilon_n^{(r)}\epsilon_1^{(1)} & \dots & \epsilon_n^{(r)}\epsilon_n^{(1)} & \epsilon_n^{(r)}\epsilon_1^{(2)} & \dots & \epsilon_n^{(r)}\epsilon_n^{(2)} & \dots & \epsilon_n^{(r)}\epsilon_1^{(r)} & \dots & \epsilon_n^{(r)}\epsilon_n^{(r)} \end{bmatrix}$$

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

Matrix V is used to obtain weighted matrix for estimating.

In the local polynomial regression, there are two parameters, i.e., the bandwidth (h) and the order of local polynomial fit (d) that control the smoothness of the fit and also affect the bias-variance trade-off. We use generalized cross-validation method for choosing optimal h and d of each response.

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

$$\hat{f}_j(t_i) = A(h_j)y_i^{(j)}, \quad j = 1, 2, \dots, r$$

where, A(h<sub>j</sub>) represents matrix as follows:

$$A(h_j) = e_i^T \left[ X_u^{(d_j)T} K_{h_j}(t_i) X_u^{(d_j)} \right]^{-1} X_{t_i}^{(d_j)T} K_{h_j}(t_i)$$

To obtain optimal h<sub>j</sub> and d<sub>j</sub> based on GCV method given by Wu and Zhang (2006), we minimize:

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

**RESULTS**

From model (1), we estimate every element of  $f(t) = (f_1(t), \dots, f_r(t))^T$  by local polynomial fitting which

uses Taylor expansion approach. By Taylor expansion, for  $t$  in a neighborhood of  $t_0$ , we have Eq. 3:

$$\left. \begin{aligned} f_1(t) &\approx f_1(t_0) + (t-t_0)f_1'(t_0) + \frac{(t-t_0)^2}{2!}f_1^{(2)}(t_0) + \dots + \frac{(t-t_0)^{d_1}}{d_1!}f_1^{(d_1)}(t_0) \\ f_2(t) &\approx f_2(t_0) + (t-t_0)f_2'(t_0) + \frac{(t-t_0)^2}{2!}f_2^{(2)}(t_0) + \dots + \frac{(t-t_0)^{d_2}}{d_2!}f_2^{(d_2)}(t_0) \\ &\vdots \\ f_r(t) &\approx f_r(t_0) + (t-t_0)f_r'(t_0) + \frac{(t-t_0)^2}{2!}f_r^{(2)}(t_0) + \dots + \frac{(t-t_0)^{d_r}}{d_r!}f_r^{(d_r)}(t_0) \end{aligned} \right\} \quad (3)$$

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

$$\begin{aligned} f_j(t) &\approx \beta_0^{(j)}(t_0) + \beta_1^{(j)}(t_0)(t-t_0) + \beta_2^{(j)}(t_0)(t-t_0)^2 \\ &+ \dots + \beta_{d_j}^{(j)}(t_0)(t-t_0)^{d_j}, j=1, \dots, r \end{aligned} \quad (4)$$

Where:

$$\beta_r^{(j)}(t_0) = \frac{f_j^{(r)}(t_0)}{r!}, r=0,1,2,\dots,d_j$$

To estimate model (4) from the sample data  $\{t_i, y_i\}_{i=1}^n$  we use local polynomial estimator defined as follows:

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

where  $K(\cdot)$  is a Kernel function,  $h > 0$ :

$$K_h(t_0) = \text{diag}\{K_{h_1}(t_1-t_0), K_{h_2}(t_2-t_0), \dots, K_{h_r}(t_r-t_0)\}$$

$$K_{h_j}(t_i-t_0) = (K_{h_j}(t_1-t_0), K_{h_j}(t_2-t_0), \dots, K_{h_j}(t_n-t_0))^T$$

$$\beta(t_0) = \begin{bmatrix} \beta_{d_1}^{(1)}(t_0) \\ \beta_{d_2}^{(2)}(t_0) \\ \vdots \\ \beta_{d_r}^{(r)}(t_0) \end{bmatrix}, \beta_{d_j}^{(j)}(t_0) = (\beta_0^{(j)}(t_0), \beta_1^{(j)}(t_0), \dots, \beta_{d_j}^{(j)}(t_0))^T$$

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

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

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

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

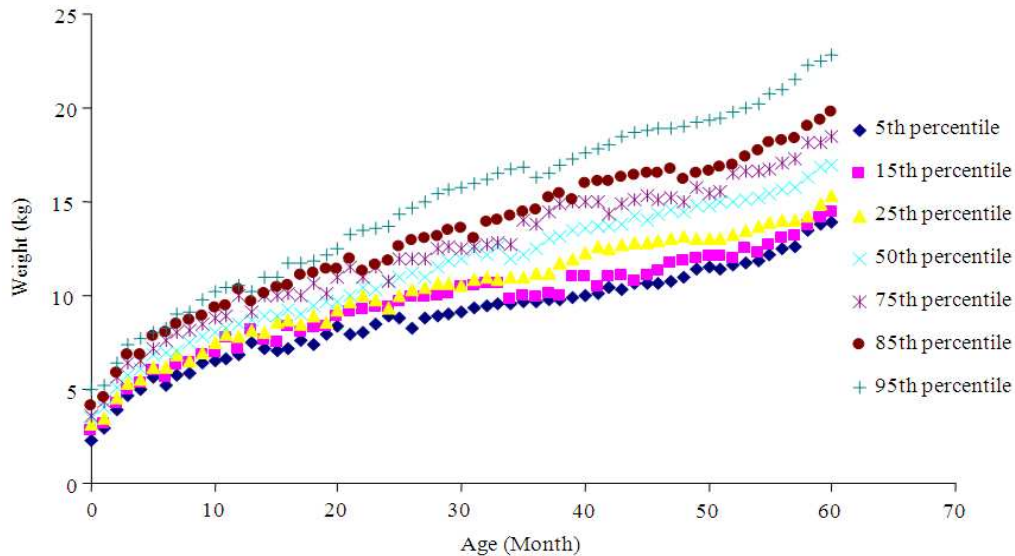


Fig. 1: Percentiles of weight children plot

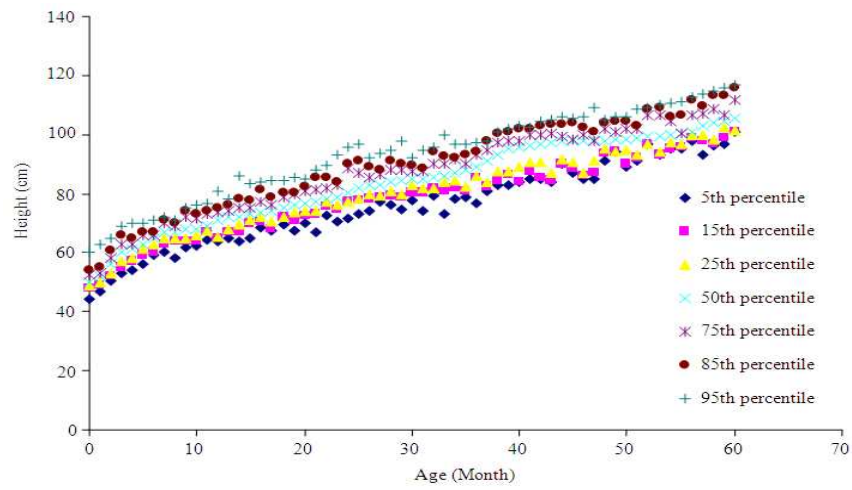


Fig. 2: Percentiles of height children plot

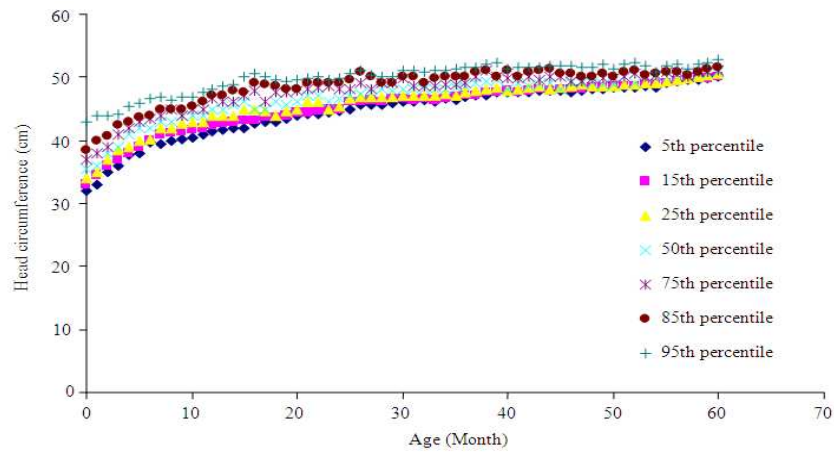


Fig. 3: Percentiles of head circumference childrenplot

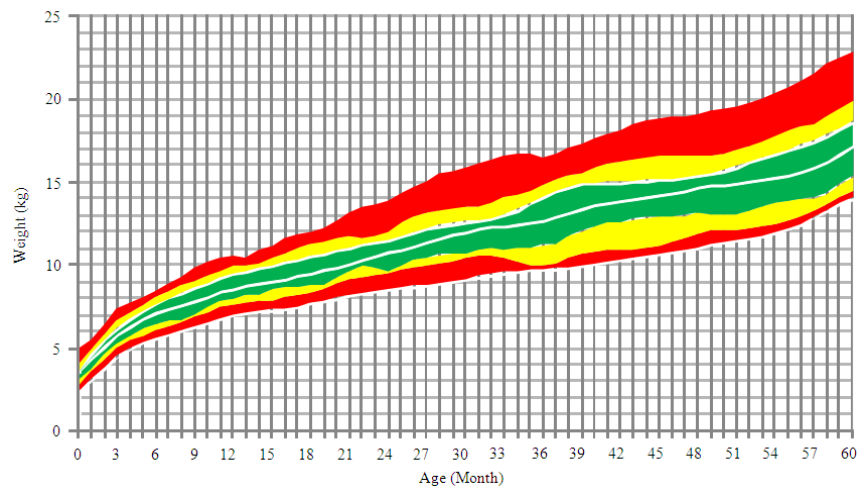


Fig. 4: Design of KMS for children in Surabaya

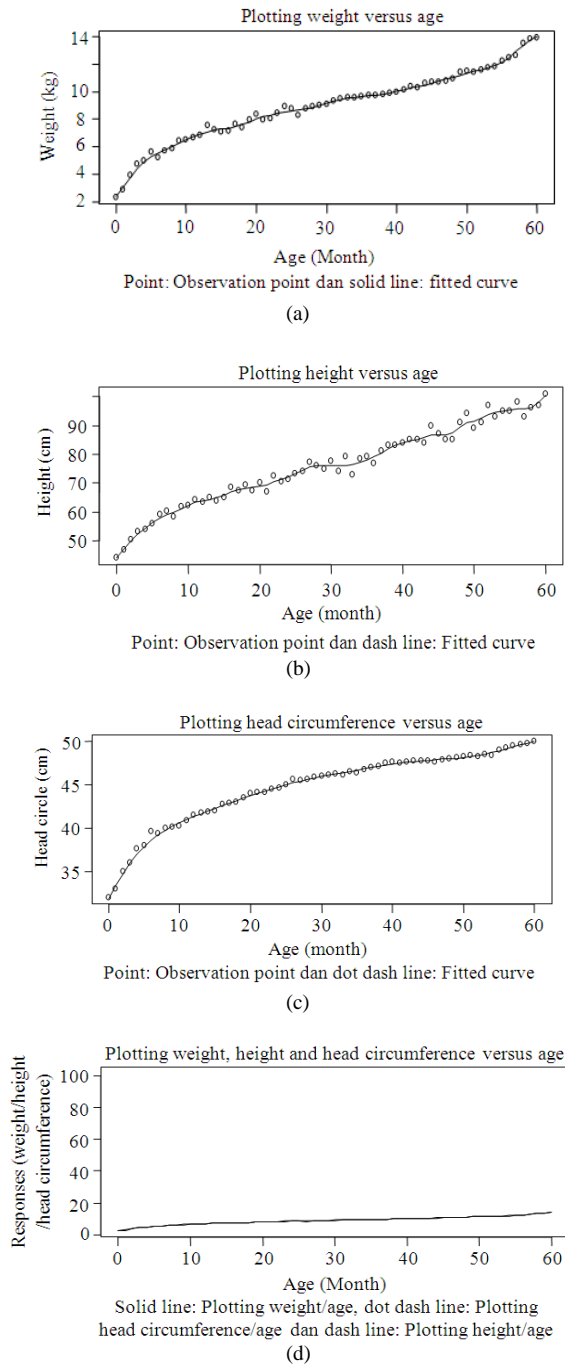


Fig. 5: Plots the 50<sup>th</sup>percentiles estimation of weight, height and head circumference versus age

So, from Eq. 6 we get the local polynomial estimator of  $f(t_i)$  Eq. 7:

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

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  $e^T = (e_1, 0, \dots, e_{d_1+2}, 0, \dots, e_{d_1+d_2+3}, 0, \dots, e_{d_1+d_2+\dots+d_{l-1}+r})^T$  is vector which has elements  $e_1 = 1, e_{d_1+2} = 1, e_{d_1+d_2+3} = 1, \dots, e_{d_1+d_2+\dots+d_{l-1}+r} = 1$  and 0 elsewhere.

The data used for applying the model contains of 1700 children obtained from community health center in Surabaya 2010. The data that describes child growth 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 is age (month). In every month of children age i.e., from 0 until 60, we determine 5<sup>th</sup> percentile, 15<sup>th</sup> percentile, 25<sup>th</sup> percentile, 50<sup>th</sup> percentile, 75<sup>th</sup> percentile, 85<sup>th</sup> percentile and 95<sup>th</sup> percentile.

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

Based on GCV method, we create R-code for choosing 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 50<sup>th</sup> 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 50<sup>th</sup> 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 circumference of 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 estimation gives mean squared error value tend to zero and determination coefficient 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-response local polynomial modeling shows increasing of children nutrition in Surabaya 2010. Because of the strong correlations among all three response variables, the simultaneously three responses model estimation is better than partly single response model estimation. The result of simultaneously model estimation based on multi-response local polynomial modeling satisfies goodness of fit criterion.

## ACKNOWLEDGEMENT

This Research was funded by the Directorate General of Higher Education of Indonesia through Doctoral Research Grant 2010 with contract number: 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.