MODEL REGRESI NONPARAMETRIK ADITIF DENGAN FUNGSI *LINK*

Siti Nasrifah, I. Nyoman Budiantara, Kartika Fitriasari

Abstract. If given a couple of data $\{(x_i, y_i)\}_{i=1}^{n}$ and the relationship between variable x and y is revealed as nonparametric regression model with $\{u_i\}_{i=1}^{n}$ is an error which is random variable with zero mean and variance σ^2 and $m(x_i)$ is unknown function. If $m(x)$ is an additive function with $x \in \mathbf{R}^d$, d ≥ 2. By using link function *F*, this research is focused on the way to get the estimator in additive component with link function by using two-stage methode and completed with simulation using logit link function. The result shows that two-stage estimator is eisier be used to model nonparametric regression by link function rather than Nadaraya-Watson and local linear. From the simulation, it's obtained that of the mean and R^2 value of the estimator is similar.

Keywords: Local linear estimator, Two-stage estimator, Nadaraya-Watson estimator, Nonparametric regression, Additive model.

1. PENDAHULUAN

Andaikan suatu sampel berukuran *n* yaitu $\{(x_i, y_i)\}_{i=1}^{n}$, *x* adalah variabel prediktor dan *y* adalah variabel respon. Diasumsikan variabel-variabel tersebut memenuhi hubungan yang dapat dinyatakan dengan model regresi sebagai berikut.

$$y_i = m(x_i) + u_i,\ i = 1, 2, \ldots, n, \qquad (1.1)$$

dengan $\{u_i\}_{i=1}^{n}$ adalah *error* yang merupakan va-riabel random dengan *mean* nol dan *varian*s σ^2 dan $m(x_i)$ adalah nilai-nilai dari fungsi *m* dititik $x_1, \ldots, x_n$. Fungsi *m* disebut kurva regresi.

Nonparametrik digunakan jika bentuk dari fungsi $m(x_i)$ tidak diketahui atau sedikit sekali informasi mengenai fungsi tersebut. Model regresi nonparametrik hanya mengasumsikan bahwa *m smooth* dalam pengertian kontinu dan diferensiabel [1].

Dalam banyak kasus, kita sering dihadapkan dengan adanya banyak variabel prediktor atau multivariabel, sehingga dapat digunakan dua model pendekatan untuk fungsi $m(x_i)$, yaitu model nonaditif dan aditif. Model nonaditif tidak banyak digunakan karena model ini rumit, sehingga untuk estimasi menjadi sulit serta interpretasi hasil juga tidak mudah. Maka untuk mempermudah analisis dari fungsi $m(x_i)$ digunakan model aditif.

Apabila fungsi $m(x_i)$ bersifat aditif dengan $x \in \mathrm{R}^d$, d ≥ 2, maka persamaan (1.1) menjadi

$$\mathrm{y} = F[\, m_1(x_1) + \ldots + m_d(x_d)] + \mathrm{u}, \quad (1.2)$$

dengan

y : vektor berukuran $n \times 1$

x_j : komponen ke-*j* (*j* = 1, …, d) pada vektor random $x \in R^d$, $d \geq 2$

$m_1, \ldots, m_d$: fungsi yang tidak diketahui

F : fungsi yang diketahui

u : vektor ukuran $n \times 1$.

2. ESTIMATOR *TWO-STAGE*

2.1 Estimator *first-stage*

Didefinisikan $m(x) = m_1(x_1) + \ldots + m_d(x_d)$, dimana x_j adalah komponen ke-j dari x. Misalkan x merupakan variabel random kontinu pada interval $[-1,1]$ dan diberikan $\{P_\kappa : \kappa = 1, 2, \ldots\}$ merupakan basis yang memenuhi syarat di bawah ini:

1. $\int_{-1}^{1} p_\kappa(v)\,dv = 0$,

2. $\int_{-1}^{1} p_j(v)p_\kappa(v)dv = \begin{cases} 1 & \text{jika } j=\kappa \\ 0 & \text{yang lain} \end{cases}$,

3. $m_j(x_j) = \sum_{\kappa=1}^{\infty} \theta_{j\kappa} p_\kappa(x_j)$,

untuk $j = 1, \ldots, d$, dan $x_j \in [0,1]$

dimana untuk sembarang bilangan bulat $\kappa > 0$ didefinisikan

$$P_\kappa(x) = \left[p_1(x_1), \ldots, p_\kappa(x_1) \ldots p_1(x_d), \ldots, p_\kappa(x_d) \right]$$

dan misalkan $\theta_{j\kappa}$ merupakan koefisien dari ekspansi deret yang didefinisikan sebagai $\theta_\kappa = [\theta_{11}, \ldots, \theta_{1\kappa} \quad \cdots \quad \theta_{d1}, \ldots, \theta_{d\kappa}]'$ maka untuk $\theta_\kappa \in R^{\kappa d+1}$. Sehingga $P_\kappa(x)'\theta_\kappa$ merupakan pendekatan deret pada $m(x)$.

Misalkan diberikan data $\{(x_i, y_i) : i = 1, \ldots, n)\}$ merupakan sampel random dari (x, y), maka estimator $\hat{\theta}_{n\kappa}$ merupakan penyelesaian dari meminimumkan :

$$\min_{\theta \in \Theta_\kappa} S_{n\kappa}(\theta) = n^{-1} \sum_{i=1}^{n} \{ y_i - F[P_\kappa(x_i)'\theta] \}^2 \qquad (2.1)$$

dengan meminimumkan persamaan (2.1), diperoleh

$$\hat{\theta} = (P_\kappa(x)P_\kappa(x)')^{-1} P_\kappa(x) G(y)$$

dengan $G = F^{-1}$

Jadi, estimator *first stage* dari $m_j(x_j)$ adalah

$$\tilde{m}_1(x_1) = p_1(x_1)\hat{\theta}_{11} + \ldots + p_\kappa(x_1)\hat{\theta}_{1\kappa}$$

$$\tilde{m}_2(x_2) = p_1(x_2)\hat{\theta}_{21} + \ldots + p_\kappa(x_2)\hat{\theta}_{2\kappa}$$

$$\vdots$$

$$\tilde{m}_d(x_d) = p_1(x_d)\hat{\theta}_{21} + \ldots + p_\kappa(x_d)\hat{\theta}_{d\kappa}$$

2.2 Estimator *second-stage*

Untuk memperoleh estimator *second-stage* dari $m_1(x_1)$, $m_2(x_2)$, …, dan $m_d(x_d)$, terlebih dahulu dicari estimator dari $m_1(x_1)$.

Didefinisikan:

$\tilde{x}_{ij} = (x_{12}, \ldots, x_{i2})$ dan

$\tilde{m}_{-1}(\tilde{x}_{ij}) = \tilde{m}_2(x_{i2}) + \tilde{m}_3(x_{i3}) + \ldots + \tilde{m}_d(x_{id})$,

dimana $\tilde{x}_{ij}$ adalah observasi ke-i dari komponen ke-j pada variabel x dan $\tilde{m}_{-j}(\tilde{x}_{ij})$ adalah estimator deret dari $m_j(x_j)$ tidak termasuk estimator yang akan dicari maka estimator linear lokal dari $m_1(x_1)$ dipeoleh dari meminimumkan:

$$S_{n1}(x_1, b_0, b_1) = \sum_{i=1}^{n} \{ y_i - F[b_0 + b_1(x_{i1} - x_1) + \tilde{m}_{-1}(\tilde{x}_i)] \}^2 K_h(x_1 - x_{i1}) \qquad (2.2)$$

dengan $b_0 = \tilde{m}_1(x_1)$ dan $b_1 = 0$. Untuk mendapatkan derivatif pertama dari (2.2) adalah:

$$S'_{nj1}(x_1, \tilde{m}) = \frac{\partial S_{n1}(x_1, b_0, b_1)}{\partial b_j}, \quad j = 0,1$$

yaitu

$$S'_{n01}(x_1, \tilde{m}) = -2 \sum_{i=1}^{n} \{ y_i - F[b_0 + b_1(x_{i1} - x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] \} \times R$$

$$= -2 \sum_{i=1}^{n} \{ y_i - F[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] \} F' [\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] K_h(x_{i1} - x_1)$$

dengan

$$R = F'[b_0 + b_1(x_{i1} - x_1) + m_{-1}(\tilde{x}_{ij})] K_h(x_{i1} - x_1)$$

$$S'_{n11}(x_1, \tilde{m}) = -2 \sum_{i=1}^{n} \{ y_i - F[b_0 + b_1(x_{i1} - x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] \} \times Q$$

$$= -2 \sum_{i=1}^{n} \{ y_i - F[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] \} \times Q$$

dengan

$$Q = F'[b_0 + b_1(x_{i1} - x_1) + m_{-1}(\tilde{x}_{ij})] (x_{i1} - x_1) K_h(x_{i1} - x_1)$$

$$= F'[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})] (x_{i1} - x_1) K_h(x_{i1} - x_1)$$

Selanjutnya untuk derifatif kedua dari (2.2) adalah $S''_{nj1}(x_1, \tilde{m}) = \dfrac{\partial^2 S_{n1}(x_1, b_0, b_1, b_2)}{\partial b_j^2}$,

$j = 0, 1, 2$, dengan $b_0 = \tilde{m}_1(x_1)$, $b_1 = b_2 = 0$

$$S''_{n01}(x_{1j},\tilde{m}) = 2\sum_{i=1}^{n} F'[b_0 + b_1(x_{i1} - x_1) +$$
$$b_2(x_{i1} - x_1)^2 + \tilde{m}_{-1}(\tilde{x}_{ij})]^2 K_h(x_1 - x_{i1}) -$$
$$2\sum_{i=1}^{n}\{y_i - F[b_0 + b_1(x_{i1} - x_1) + b_2(x_{i1} - x_1)^2 +$$
$$\tilde{m}_{-1}(\tilde{x}_{ij})]\} \times D$$
$$= 2\sum_{i=1}^{n} F'[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})]^2 K_h(x_1 - x_{i1}) +$$
$$-2\sum_{i=1}^{n}\{Y_i - F[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_{ij})]\}$$
$$F''[\tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{x}_{ij})] K_h(x_1 - x_{i1})$$

dengan

$$D = F''[b_0 + b_1(x_{i1} - x_1) + b_2(x_{i1} - x_1)^2 +$$
$$\tilde{m}_{-1}(\tilde{x}_{ij})] K_h(x_1 - x_{i1}).$$

$$S''_{n11}(x_1,\tilde{m}) = 2\sum_{i=1}^{n} F'[b_0 + b_1(x_{i1} - x_1) +$$
$$b_2(x_{i1} - x_1)^2 \tilde{m}_{-1}(\tilde{x}_i)]^2 \times (x_{i1} - x_1) K_h$$
$$(x_1 - x_{i1}) - 2\sum_{i=1}^{n}\{y_i - F[b_0 + b_1(x_{i1} - x_1) +$$
$$b_2(x_{i1} - x_1)^2 + \tilde{m}_{-1}(\tilde{x}_i)]\} \times B$$
$$= 2\sum_{i=1}^{n} F'[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_i)]^2 (x_{i1} - x_1)$$
$$K_h(x_1 - x_{i1}) - 2\sum_{i=1}^{n}\{Y_i - F[\tilde{m}_1(x_1) +$$
$$\tilde{m}_{-1}(\tilde{x}_i)]\} F''[\tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{x}_i)]$$
$$(x_{i1} - x_1) K_h(x_1 - x_{i1}).$$

dengan

$$B = F''[b_0 + b_1(x_{i1} - x_1) + b_2(x_{i1} - x_1)^2$$
$$+ \tilde{m}_{-1}(\tilde{x}_i)](x_{i1} - x_1) K_h(x_1 - x_{i1})$$

$$S''_{n21}(x_1,\tilde{m}) = 2\sum_{i=1}^{n} F'[b_0 + b_1(x_{i1} - x_1) +$$
$$b_2(x_{i1} - x_1)^2 + \tilde{m}_{-1}(\tilde{x}_i)]^2 (x_{i1} - x_1)^2$$
$$K_h(x_1 - x_{i1}) - 2\sum_{i=1}^{n}\{y_i - F[b_0 + b_1(x_{i1} - x_1) +$$
$$b_2(x_{i1} - x_1)^2 + \tilde{m}_{-1}(\tilde{x}_i)]\} \times C$$
$$= 2\sum_{i=1}^{n} F'[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_i)]^2 (x_{i1} - x_1)^2$$
$$K_h(x_1 - x_{i1}) - 2\sum_{i=1}^{n}\{Y_i - F[\tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{x}_i)]\}$$
$$F''[\tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{x}_i)](x_{i1} - x_1)^2 K_h(x_1 - x_{i1})$$

dengan

$$C = F''[b_0 + b_1(x_{i1} - x_1) + b_2(x_{i1} - x_1)^2 +$$
$$\tilde{m}_{-1}(\tilde{x}_i)]\ (x_{i1} - x_1)^2 K_h(x_1 - x_{i1})$$

Dengan menggunakan metode Newton-Raphson, maka diperoleh estimator linear lokal dari $m_1(x_1)$ adalah:

$$\hat{m}_{1,LL}(x_1) = \tilde{m}_1(x_1) - e'\begin{pmatrix} S''_{n01}(x_1,\tilde{m}) & S''_{n11}(x_1,\tilde{m}) \\ S''_{n11}(x_1,\tilde{m}) & S''_{n21}(x_1,\tilde{m}) \end{pmatrix}^{-1} \begin{pmatrix} S'_{n01}(x_1,\tilde{m}) \\ S'_{n11}(x_1,\tilde{m}) \end{pmatrix}$$

dengan $e = (1,0)$, sehingga estimator linear lokal dari $m_1(x_1)$ dapat ditulis:

$$\hat{m}_{1,LL}(x_1) = \tilde{m}_1(x_1) - \frac{S''_{n21}(x_1,\tilde{m})S'_{n01}(x_1,\tilde{m}) - S''_{n11}(x_1,\tilde{m})S'_{n11}(x_1,\tilde{m})}{S''_{n01}(x_1,\tilde{m})S''_{n21}(x_1,\tilde{m}) - S''_{n11}(x_1,\tilde{m})^2} \tag{2.3}$$

Sedangkan untuk estimator konstant lokal dari $m_1(x_1)$ adalah

$$\hat{m}_{1,LC}(x_1) = \tilde{m}_1(x_1) - \frac{S'_{n01}(x_1,\tilde{m})}{S''_{n01}(x_1,\tilde{m})}. \tag{2.4}$$

Untuk mendapatkan estimator $m_2, \ldots, m_d$ caranya sama seperti mendapatkan $m_1(x_1)$, estimator linear lokal dapat ditulis seba-gai berikut.

$$\hat{m}_{2,LL}(x_1) = \tilde{m}_2(x_2) - \frac{S''_{n21}(x_2,\tilde{m})S'_{n01}(x_2,\tilde{m}) - S''_{n11}(x_2,\tilde{m})S'_{n11}(x_2,\tilde{m})}{S''_{n01}(x_2,\tilde{m})S''_{n21}(x_2,\tilde{m}) - S''_{n11}(x_2,\tilde{m})^2}$$

$$\hat{m}_{3,LL}(x_1) = \tilde{m}_3(x_3) - \frac{S''_{n21}(x_3,\tilde{m})S'_{n01}(x_3,\tilde{m}) - S''_{n11}(x_3,\tilde{m})S'_{n11}(x_3,\tilde{m})}{S''_{n01}(x_3,\tilde{m})S''_{n21}(x_3,\tilde{m}) - S''_{n11}(x_3,\tilde{m})^2}$$

$$\hat{m}_{4,LL}(x_1) = \tilde{m}_4(x_4) - \frac{S''_{n21}(x_4,\tilde{m})S'_{n01}(x_4,\tilde{m}) - S''_{n11}(x_4,\tilde{m})S'_{n11}(x_4,\tilde{m})}{S''_{n01}(x_4,\tilde{m})S''_{n21}(x_4,\tilde{m}) - S''_{n11}(x_4,\tilde{m})^2}$$

$$\vdots$$

$$\hat{m}_{\mathrm{d},LL}(x_1) = \tilde{m}_{\mathrm{d}}(x_{\mathrm{d}}) - \frac{S''_{n21}(x_{\mathrm{d}},\tilde{m})S'_{n01}(x_{\mathrm{d}},\tilde{m}) - S''_{n11}(x_{\mathrm{d}},\tilde{m})S'_{n11}(x_{\mathrm{d}},\tilde{m})}{S''_{n01}(x_{\mathrm{d}},\tilde{m})S''_{n21}(x_{\mathrm{d}},\tilde{m}) - S''_{n11}(x_{\mathrm{d}},\tilde{m})^2}$$

Cara yang sama pula untuk mendapatkan estimator konstant lokal, sehingga diperoleh sebagai berikut.

$$\hat{m}_{2,\,LC}(x_1) = \tilde{m}_2(x_2) - \frac{S'_{n01}(x_2,\tilde{m})}{S''_{n01}(x_2,\tilde{m})},$$

$$\hat{m}_{3,\,LC}(x_3) = \tilde{m}_3(x_3) - \frac{S'_{n01}(x_3,\tilde{m})}{S''_{n01}(x_3,\tilde{m})},$$

$$\vdots$$

$$\hat{m}_{d,\,LC}(x_1) = \tilde{m}_d(x_d) - \frac{S'_{n01}(x_d,\tilde{m})}{S''_{n01}(x_d,\tilde{m})}.$$

3. Estimator *Nadaraya-Watson* dan Linear lokal

Diberikan model $y = m(x) + u$. Fungsi $m(x)$ didekati dengan persamaan

$$m(x) \approx b_0 + b_1(x_i - x) + b_2(x_i - x)^2 + \cdots + b_p(x_i - x)^p.$$

Dengan metode *least square* meminimumkan

$$\min \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - m(x))^2 K_h(x - x_i)$$

$$= \sum_{i=1}^{n} \{ y_i - (b_0 + b_1(x_i - x) + b_2(x_i - x)^2 + \cdots + b_p(x_i - x)^p)\}^2 K_h(x - x_i). \quad (2.5)$$

Dimana p merupakan derajat dari estimasi fungsi regresi. Berarti p = 0, 1, ... serta h adalah *bandwidth*. Jika p = 0, maka $b_2 = b_3 = \cdots = b_p = 0$, sehingga untuk meminimumkan persamaan (2.5) adalah

$$-2\sum_{i=1}^{n}\{y_i - b_0\}K_h(x - x_i) = 0.$$

$$\sum_{i=1}^{n} K_h(x - x_i)y_i = \sum_{i=1}^{n} K_h(x - x_i)b_0.$$

$$b_0 = \frac{\sum_{i=1}^{n} K_h(x - x)y_i}{\sum_{i=1}^{n} K_h(x - x)} \text{ yang artinya}$$

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K_h(x - x)y_i}{\sum_{i=1}^{n} K_h(x - x)}.$$

Dengan demikian jika p = 0 maka estimator tersebut disebut kostant lokal disebut juga estimator *Nadaraya-Watson*. Misalkan

$$S_{h,p}(x) = \sum_{i=1}^{n} (x_i - x)^p K_h(x - x_i), \text{ dan}$$

$$T_{h,p}(x) = \sum_{i=1}^{n} (x_i - x)^p K_h(x - x_i) y_i$$

Jadi estimator konstant lokal dapat ditulis sebagai

$$\hat{m}_{h,0}(x) = \frac{T_{h,0}(x)}{S_{h,0}(x)}, \quad (2.6)$$

dan jika p = 1, maka estimator tersebut disebut linear lokal, dinyatakan dalam persamaan :

$$\hat{m}_{h,0}(x) = e' \begin{pmatrix} S_{h,0}(x) & S_{h,1}(x) \\ S_{h,1}(x) & S_{h,2}(x) \end{pmatrix}^{-1} \begin{pmatrix} T_{h,0}(x) \\ T_{h,1}(x) \end{pmatrix}$$

$$, e = (1,0).$$

Jadi estimator linear lokal adalah:

$$\hat{m}_{h,0}(x) = \frac{T_{h,0}(x)S_{h,2}(x) - T_{h,1}(x)S_{h,1}(x)}{S_{h,0}(x)S_{h,2}(x) - S^2{}_{h,0}(x)}. \quad (2.7)$$

Kelebihan estimator *two-stage* daripada *Nadaraya-Watson* adalah jika dilihat dari bentuk persamaan estimator yang telah diperoleh pada persamaan (2.3) dan persamaan (2.4), maka metode tersebut dapat digunakan untuk mengestimasi suatu model yang menggunakan sembarang fungsi *link* dengan langsung mensubstitusikan fungsi tersebut. Sedangkan untuk mendapatkan estimator yang menggunakan metode *Nadaraya-Watson*, fungsi *link* harus terlebih dahulu ditentukan, kemudian dapat dicari estimatornya.

4. Simulasi dengan Fungsi *Link* Logit

Diberikan model regrasi

$$y = F[m_1(x_1) + \ldots + m_d(x_d)] + u.$$

Simulasi ini dilakukan dengan menggunakan model berbentuk

$$m(x) = m_1(x_1) + m_2(x_2)$$

dengan $m_1(x_1) = \sin(\pi x_1)$ dan

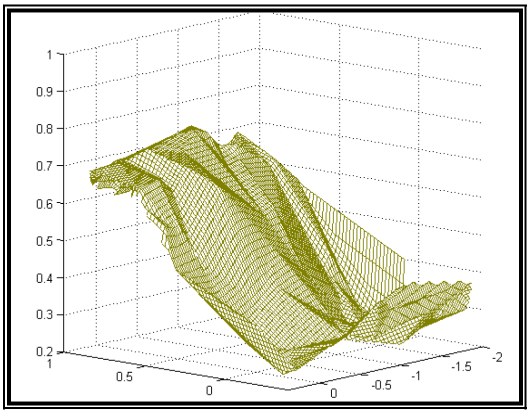
$$m_2(x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(3x_2)^2}{2}} \text{ dan } F(v) = \frac{e^v}{1 + e^v}.$$

Sehingga model menjadi

$$y = \frac{e^{\sin(\pi x_1) + \frac{1}{\sqrt{2\pi}} e^{-\frac{9x_2^2}{2}}}}{1 + e^{\sin(\pi x_1) + \frac{1}{\sqrt{2\pi}} e^{-\frac{9x_2^2}{2}}}} + u.$$

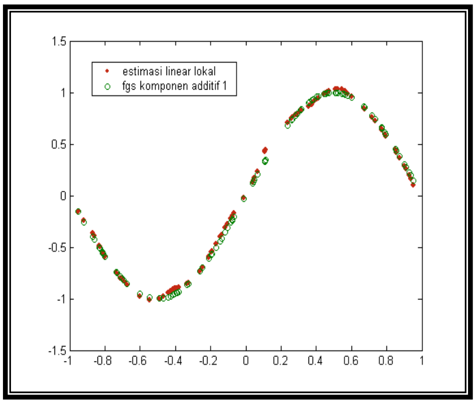
Simulasi ini dicobakan n = 50, 100, dan 200 dengan varians masing-masing $\sigma^2 = 0{,}1$, $\sigma^2 = 0{,}2$ dan $\sigma^2 = 0{,}4$ serta replikasi sebanyak 20. Dengan berbagai banyak sampel dan varians, digunakan kernel *gaussian* sebagai pembobot untuk estimator, selanjutnya dicari *banwidth* optimum berdasarkan GCV yang paling minimum. Dari hasil estimasi, akan dibandingkan antara *two-stage, Nadaraya-Watson*, dan linear lokal berdasarkan R^2 dan MSE.

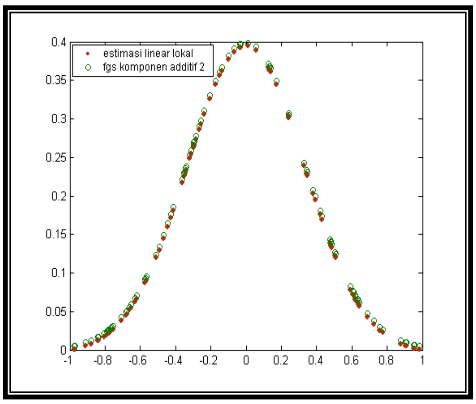
Dalam simulasi untuk n = 100 dan σ^2= 0,2 pada replikasi ke-7 didapatkan *bandwidth* optimum sebesar 0.0590 untuk x_1 dan 0.0370 untuk x_2. Berikut adalah plot antara y dengan x_1 dan x_2 .



Gambar 1. Plot antara y dengan x_1 dan x_2

Model Hasil estimasi adalah : $\hat{y} = \dfrac{e^{\hat{m}_1(x_1) + \hat{m}_2(x_2)}}{1 + e^{\hat{m}_1(x_1) + \hat{m}_2(x_2)}}$



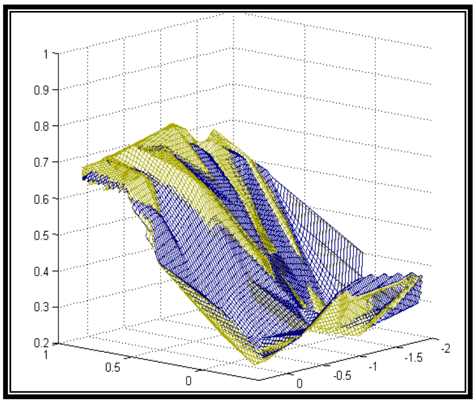


Gambar 2. Plot antara $m_1(x_1)$ dan $\hat{m}_1(x_1)$ dengan x_1 untuk n = 100 danσ^2=0.2

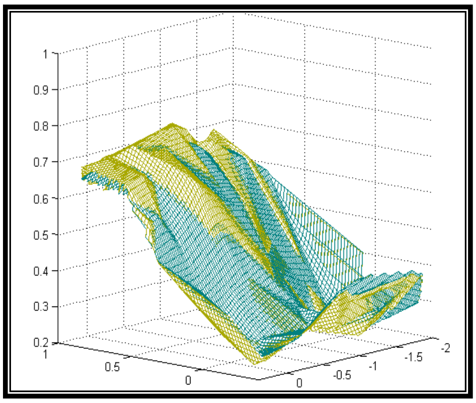
Gambar 3. Plot antara $m_2(x_2)$ dan $\hat{m}_2(x_2)$ dengan x_2 untuk n = 100 dan σ^2=0.2

Tabel 1. Hasil Simulasi *two- stage* mean dari y, $\hat{y}_{LL}$, $\hat{y}_{LC}$ n = 100 ,σ^2 = 0.1,0.2, 0,4

rep	σ^2 = 0.1			σ^2 = 0.2			σ^2 = 0.4		
	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$
1	0.5755	0.5754	0.5762	0.5759	0.5755	0.5756	0.5755	0.5758	0.5756
2	0.5445	0.5443	0.5444	0.5464	0.5466	0.5464	0.5418	0.5416	0.5418
3	0.5496	0.5492	0.5494	0.5451	0.5450	0.5450	0.5407	0.5415	0.5423
4	0.5226	0.5226	0.5221	0.5220	0.5220	0.5224	0.5244	0.5246	0.5245
5	0.5623	0.5624	0.5623	0.5610	0.5614	0.5616	0.5545	0.5542	0.5548
6	0.5347	0.5342	0.5345	0.5349	0.5345	0.5347	0.5388	0.5377	0.5392
7	0.5435	0.5441	0.5436	0.5425	0.5426	0.5425	0.5496	0.5486	0.5487
8	0.5457	0.5456	0.5455	0.5435	0.5440	0.5442	0.5460	0.5460	0.5460
9	0.5430	0.5434	0.5431	0.5471	0.5473	0.5472	0.5437	0.5448	0.5443
10	0.5163	0.5168	0.5166	0.5135	0.5135	0.5136	0.5202	0.5202	0.5204
11	0.5459	0.5460	0.5460	0.5440	0.5442	0.5447	0.5419	0.5430	0.5425
12	0.5271	0.5265	0.5272	0.5321	0.5321	0.5335	0.5314	0.5312	0.5311
13	0.5229	0.5229	0.5232	0.5239	0.5235	0.5239	0.5169	0.5165	0.5166
14	0.5329	0.5326	0.5329	0.5310	0.5312	0.5311	0.5332	0.5329	0.5331
15	0.5205	0.5207	0.5206	0.5232	0.5235	0.5240	0.5210	0.5212	0.5212
16	0.5135	0.5136	0.5137	0.5137	0.5138	0.5136	0.5102	0.5104	0.5104
17	0.5282	0.5284	0.5285	0.5260	0.5258	0.5259	0.5425	0.5423	0.5426
18	0.5466	0.5465	0.5466	0.5475	0.5475	0.5475	0.5399	0.5399	0.5399
19	0.5417	0.5426	0.5416	0.5383	0.5384	0.5382	0.5373	0.5362	0.5371
20	0.5301	0.5303	0.5302	0.5286	0.5287	0.5291	0.5283	0.5280	0.5284



Gambar 4. Plot fungsi y (--) dan $\hat{y}_{LL}$ (--) dengan x_1 dan x_2 *two-stage*



Gambar 5. Plot fungsi y (--) dan $\hat{y}_{LC}$ (--) dengan x_1 dan x_2 *two-stage*

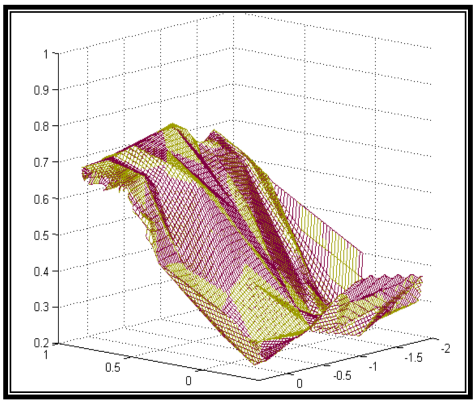
Untuk mengetahui apakah mean hasil estimasi dari *two-stage*, *Nadaraya-Watson*, linear lokal cenderung sama, dilakukan analisis dengan *Compare mean test* yaitu dengan *Independent-Sample T-Test*, pengujian hipotesisnya adalah:

$$H_0 : \mu_1 = \mu_2$$

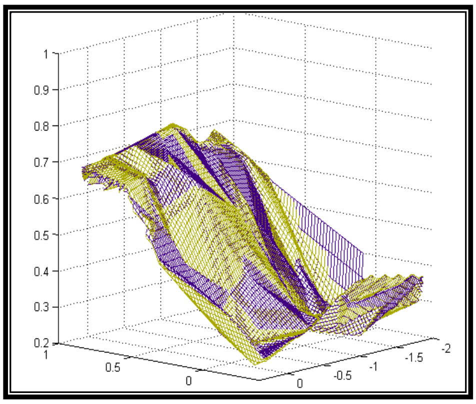
$$H_1 : \mu_1 \neq \mu_2$$

Hasil analisis untuk n = 100, σ^2 = 0.2 dan replikasi ke-7, Linear lokal memberikan hasil *P-Value* = 0.999 > 0.05, sehingga dapat disimpulkan bahwa pada tingkat kepercayaan 95%, mean antara linear lokal *two-stage* dengan linear lokal adalah cenderung sama. Hal yang sama dihasilkan oleh konstant lokal, yaitu *P-Value* = 0.998 > 0.05, dapat disimpulkan pula bahwa pada tingkat kepercayaan 95%, mean antara konstant lokal *two-stage* dengan *Nadaraya-Watson* juga cenderung sama.

Hasil simulasi yang ditampilkan pada Tabel 3 menunjukkan R^2 yang didapatkan dengan konstant lokal dan linear lokal pada metode *two-stage*, *Nadaraya-Watson* dan linear lokal adalah cenderung sama. Uji analisisnya untuk linear lokal dengan menghasilkan *P-value* = 0.115>0.05, sehingga dapat disimpulkan bahwa R^2 linear lokal *two-stage* dan linear lokal cenderung sama. Untuk konstant lokal, menghasilkan *P-value* = 0.063 >0.05, sehingga dapat disimpulkan bahwa R^2 konstant lokal *two-stage* dan konstant lokal *Nadaraya-Watson* cenderung sama. Walaupun terdapat sedikit perbedaan, namun tidak signifikan.



Gambar 6. Plot fungsi y (--) dan $\hat{y}_{LC}$ (--) dengan x_1 dan x_2 pada linear lokal



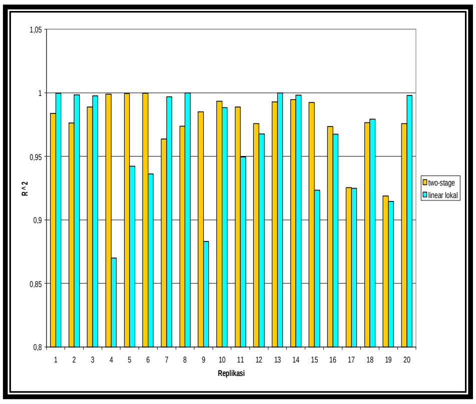
Gambar 7. Plot fungsi y (--) dan $\hat{y}_{LC}$ (--) dengan x_1 dan x_2 metode *NW*

Tabel 2. Hasil Simulasi *NW* mean dari y, $\hat{y}_{LL}$, $\hat{y}_{LC}$ n = 100 ,σ^2 = 0.1,0.2, 0,4

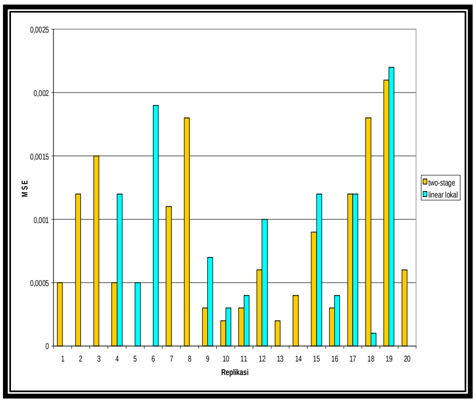
rep	**σ^2 = 0.1**			**σ^2 = 0.2**			**σ^2 = 0.4**		
	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$	y	$\hat{y}_{LL}$	$\hat{y}_{LC}$
1	0.5755	0.5754	0.5755	0.5759	0.5758	0.5759	0.5755	0.5755	0.5755
2	0.5445	0.5445	0.5444	0.5464	0.5466	0.5462	0.5418	0.5417	0.5419
3	0.5496	0.5496	0.5496	0.5451	0.5452	0.5450	0.5407	0.5403	0.5410
4	0.5226	0.5244	0.5222	0.5220	0.5239	0.5216	0.5244	0.5259	0.5242
5	0.5623	0.5618	0.5623	0.5610	0.5607	0.5610	0.5545	0.5543	0.5544
6	0.5347	0.5328	0.5337	0.5349	0.5331	0.5340	0.5388	0.5367	0.5381
7	0.5435	0.5435	0.5435	0.5425	0.5426	0.5425	0.5496	0.5492	0.5497
8	0.5457	0.5459	0.5459	0.5435	0.5435	0.5435	0.5460	0.5460	0.5459
9	0.5430	0.5423	0.5431	0.5471	0.5465	0.5470	0.5437	0.5435	0.5436
10	0.5163	0.5167	0.5160	0.5135	0.5138	0.5133	0.5202	0.5205	0.5202
11	0.5459	0.5472	0.5470	0.5440	0.5453	0.5452	0.5419	0.5429	0.5428
12	0.5271	0.5269	0.5270	0.5321	0.5357	0.5334	0.5314	0.5313	0.5314
13	0.5229	0.5229	0.5229	0.5239	0.5238	0.5239	0.5169	0.5170	0.5170
14	0.5329	0.5329	0.5329	0.5310	0.5311	0.5310	0.5332	0.5336	0.5332
15	0.5205	0.5219	0.5200	0.5232	0.5245	0.5227	0.5210	0.5224	0.5204
16	0.5135	0.5145	0.5133	0.5137	0.5147	0.5135	0.5102	0.5098	0.5098
17	0.5282	0.5298	0.5275	0.5260	0.5274	0.5248	0.5425	0.5429	0.5413
18	0.5466	0.5466	0.5464	0.5475	0.5476	0.5472	0.5399	0.5392	0.5399
19	0.5417	0.5420	0.5416	0.5383	0.5386	0.5384	0.5373	0.5381	0.5375
20	0.5301	0.5301	0.5301	0.5286	0.5285	0.5285	0.5283	0.5282	0.5282

Tabel 3. Hasil Simulasi MSE dan R^2 untuk n=100 dan $\sigma^2 = 0{,}2$

REp	**Metode *two- stage***				**Linear Lokal**		**Lokal Konstan (NW)**	
	Linear Lokal		**Lokal Konstan**					
	MSE	**R^2**	**MSE**	**R^2**	**MSE**	**R^2**	**MSE**	**R^2**
1	0.0005	0.9838	0.0002	0.9920	0.0000	0.9996	0.0001	0.9963
2	0.0012	0.9763	0.0006	0.9830	0.0000	0.9985	0.0001	0.9954
3	0.0015	0.9888	0.0003	0.9933	0.0000	0.9977	0.0000	0.9976
4	0.0005	0.9989	0.0004	0.9998	0.0012	0.8699	0.0014	0.8279
5	0.0000	0.9995	0.0000	0.9999	0.0005	0.9424	0.0007	0.9185
6	0.0000	0.9996	0.0000	0.9999	0.0019	0.9363	0.0029	0.9496
7	0.0011	0.9637	0.0007	0.9659	0.0000	0.9971	0.0001	0.9900
8	0.0018	0.9738	0.0010	0.9863	0.0000	1.0000	0.0000	0.9996
9	0.0003	0.9850	0.0006	0.9864	0.0007	0.8832	0.0009	0.8571
10	0.0002	0.9935	0.0001	0.9983	0.0003	0.9883	0.0002	0.9873
11	0.0003	0.9890	0.0002	0.9941	0.0004	0.9497	0.0005	0.9298
12	0.0006	0.9757	0.0005	0.9822	0.0010	0.9677	0.0012	0.9672
13	0.0002	0.9930	0.0001	0.9995	0.0000	1.0000	0.0002	0.9985
14	0.0004	0.9946	0.0001	0.9970	0.0000	0.9983	0.0001	0.9952
15	0.0009	0.9923	0.0000	0.9999	0.0012	0.9234	0.0005	0.9044
16	0.0003	0.9736	0.0002	0.9840	0.0004	0.9674	0.0003	0.9653
17	0.0012	0.9254	0.0010	0.9445	0.0012	0.9250	0.0020	0.8582
18	0.0018	0.9766	0.0004	0.9861	0.0001	0.9792	0.0005	0.9476
19	0.0021	0.9190	0.0008	0.9895	0.0022	0.9147	0.0006	0.9046
20	0.0006	0.9759	0.0001	0.9909	0.0000	0.9980	0.0001	0.9885



Gambar 8. Histogram R^2 untuk linear lokal



Gambar 9. Histogram MSE untuk linear lokal

5. PENUTUP

a. Kesimpulan

Berdasarkan hasil analisis, diperoleh kesimpulan sebagai berikut:

1. Estimasi yang dihasilkan dari estimasi *two-stage*, *Nadaraya Watson* dan linear lokal adalah cenderung sama.
2. Pada metode *two*-stage, dapat diperoleh estimator tiap komponen fungsi aditif sedangkan pada metode *Nadaraya Watson* tidak, hal itu karena pada metode *Nadaraya Watson* langsung menggunakan estimasi *full-dimensional*.

b. Saran

Dalam tulisan ini membahas tentang model aditif nonparametrik dengan fungsi *link* yang difokuskan dalam mengkaji metode untuk memperoleh estimator. Saran yang diberikan untuk penelitian selanjutnya adalah:

1. Perlu dicobakan untuk fungsi *link* lain untuk melihat kecenderungan bahwa metode *two stage* dapat

digunakan untuk mendapatkan estimator suatu model aditif dengan menggunakan fungsi link manapun

2. Perlu dicari distribusi asimtotik dari tiap *komponen* aditif dari model aditif yang menggunakan fungsi *link* manapun.

6. DAFTAR PUSTAKA

[1] Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, Inc., New York.

[2] Fan,J., Hardle,W., dan Mammen, E. (1998), Direct Estimation Of Low-Dimensional Components in Additive Models, *Annals of Statistics*, **26**: 943-971.

[3] Green, P.J. (1994), *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London.

[4] Greybill, F. A., Mood, A.M., dan Boes, D. C. (1974), *Applied Regression Analisys*, 2nd ed., John Willey & Sons, Chapman & Hall., New York.

[5] Hastie, T. J. dan Tibshirani, R.J. (1990), *Generalized Additive Models*, Chapman & Hall, London.

[6] Horowitz, J. dan Mammen, E. (2003), *Nonparametric Esimation of an Additive Model with a Link Function*, Cemmap Working Paper CWP19/02.

[7] Kreyszig, E. (1988), *Introductory Numerical Analisys*, John Willey & Sons, Chapman & Hall., New York.

[8] Linton, O. B. dan Hardle, W. (1996), Estimating Additive Regression with Known Links. *Biometrika*, **83**: 529-540.

[9] Linton, O. B. dan Nielsen, J.P. (1995), *A Kernel Method of estimating structured nonparametric regression based on Marginal Integration. Biometrika*, **82**: 93-100.

[10] McCullagh, P. (1983), *Generalized Linear Models,* Chapman & Hall, London.

[11] Newey, W. K. (1997), *Convergence rates and Asymtotic Normality for series estimators*, Journal Of Econometric, **79:** 147-168.

[12] Opsomer, J. D. (2000), *Asymtotic Properties of Backfitting Estimators,* Journal of Multivariate Analysis, **73:** 166-179.

[13] Opsomer, J. D. dan Ruppert, D. (1997), *Fitting a Bivariate Additive Model by Local Polynomial Regression*, Annals of Statistics, **25**: 186-211.

[14] Stone, C. J. (1985), *Additive Regression and Other nonparametric models*, Annals of Statistics, **13**: 689-705.

[15] Stone, C. J. (1986), *The Dimensionality Reduction Principle for Generalized Additive Models*, Annals of Statistics, **14**: 590-606.

[16] Wahba, G. (1990), Spline Model for *Observation Data*, SYAM, Pensylvania.