Pemodelan Regresi Semiparametrik menggunakan Estimator Campuran Spline Truncated dan Kernel

Seminar Tesis

Hesikumalasari
1314201205
Dosen Pembimbing : Prof Dr. I. Nyoman Budiantara, M.Si
Co. Pembimbing : Dr. Vita Ratnasari, M.Si

Pemodelan Regresi Semiparametrik menggunakan Estimator Campuran Spline Truncated dan Kernel

BAB I. Pendahuluan

BAB II. Tinjauan Pustaka

BAB III. Metodologi Penelitian

BAB IV. Hasil dan Pembahasan

BAB IV. Saran dan Kesimpulan

Latar Belakang

Rumusan Masalah

Tujuan Penelitian

Manfaat Penelitian

Batasan Masalah

Regresi Parametrik dan Nonparametrik

Regresi Semiparametrik

Regresi Parametrik Linear

Estimator Spline

Estimator Kernel

GCV Dan Koefisien Determinasi

Produksi Padi

Sumber Data

Variabel Penelitian dan Struktur Data

Metodologi Penelitian

Model Regresi Semiparametrik Campuran

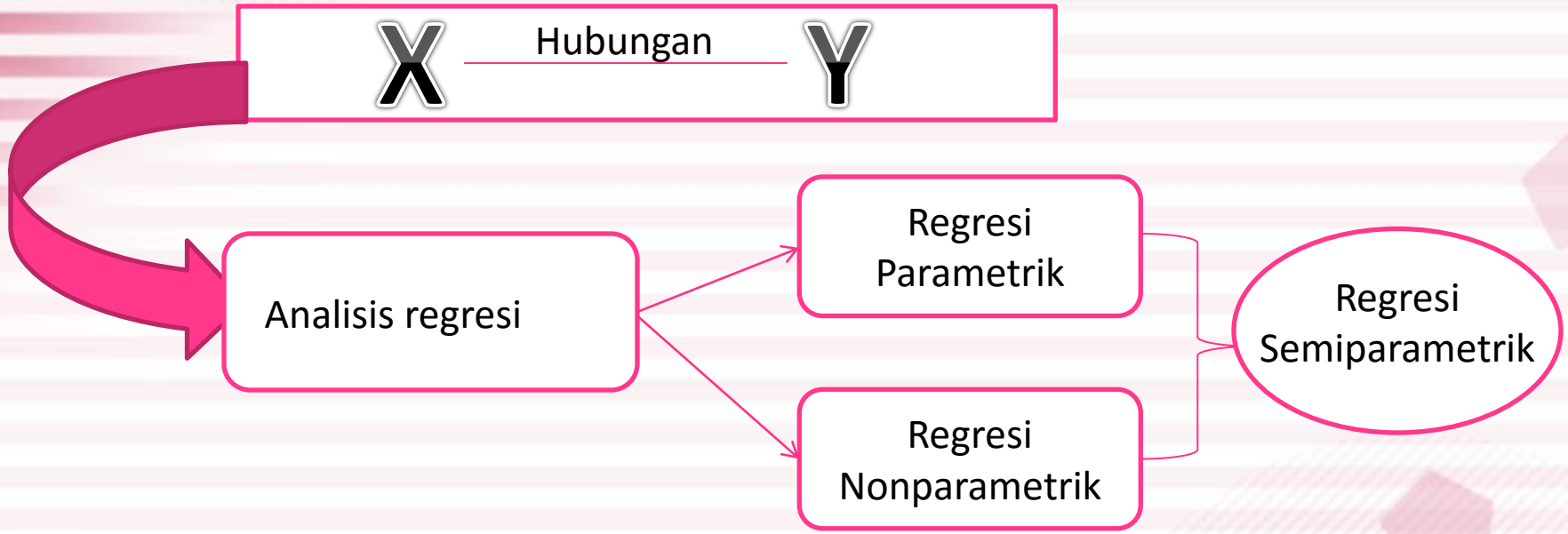
Estimasi Kurva Regresi Semiparametrik Campuran

Aplikasi pada Data Produksi Padi di Jawa Tengah

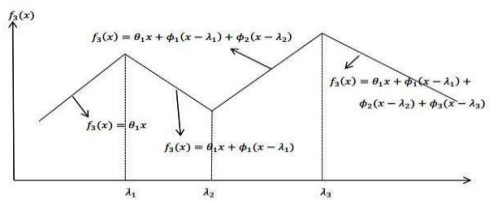
Kesimpulan

Saran

Latar Belakang



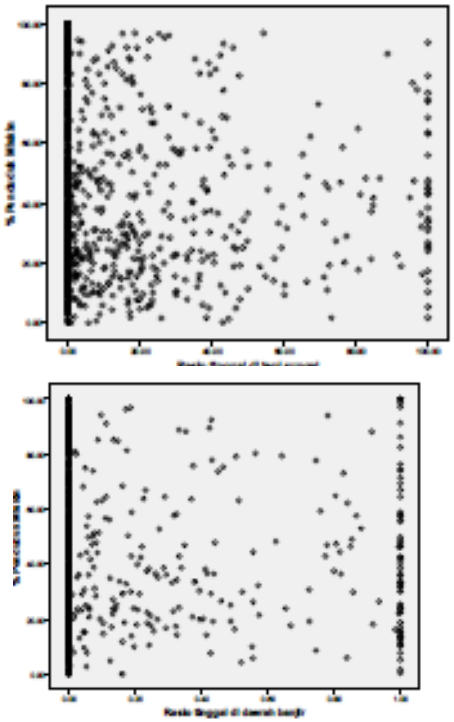
Latar Belakang



Semiparametrik

Estimator Campuran Spline Truncated dan Kernel

Produksi Padi di Jawa Tengah Tahun 2008



Rumusan Masalah

1

Bagaimana bentuk estimasi kurva regresi semiparametrik campuran *spline truncated* dan kernel?

2

Bagaimana menerapkan estimasi campuran *spline truncated* dan kernel untuk memodelkan data produksi padi di Jawa Tengah?

Tujuan dan Manfaat Penelitian

Tujuan Penelitian

1. Mengkaji bentuk estimasi kurva regresi semiparametrik campuran *spline truncated* dan kernel.
2. Mengaplikasikan/menerapkan estimasi campuran *spline truncated* dan kernel untuk memodelkan data produksi padi di Jawa Tengah.

Manfaat Penelitian

1. Memberikan wawasan baru mengenai pemodelan, khususnya kurva regresi semiparametrik campuran *spline truncated* dan kernel.
2. Mengetahui penggunaan hasil estimasi yang diperoleh untuk memodelkan data produksi padi di Jawa Tengah.

Batasan Masalah

- Data yang digunakan adalah data produksi padi di Jawa Tengah Tahun 2008.
- Fungsi *spline* yang digunakan dalam estimator campuran adalah *spline truncated* linear.
- Estimator kernel yang digunakan dalam estimator campuran adalah estimator kernel Nadaraya-Watson.
- Pemilihan titik knot dan parameter *bandwidth* optimal menggunakan metode GCV.
- Dalam aplikasi model digunakan *spline truncated* linear dengan satu, dua dan tiga knot.

Kajian Pustaka

Analisis Regresi (Parametrik dan Nonparametrik)

Analisis regresi merupakan sebuah metode Statistika yang memberikan penjelasan tentang pola hubungan (model) antara dua variabel atau lebih.

Bentuk pola hubungan kedua variabel tersebut dapat diidentifikasi berdasarkan suatu *scatter plot* atau pengalaman masa lalu yang memuat informasi tentang kurva regresi.

Misalkan dberikan data $(x_i, y_i),\ i = 1, 2, \ldots, n$ maka model regresiya dapat ditulis :

$$y_i = f(x_i) + \varepsilon_i$$

Kurva Regresi

- Diketahui → Regresi Parametrik
  - Linear
  - Kuadratik
  - Kubik
  - Eksponensial
- Tidak Diketahui → Regresi Nonparametrik
  - Spline
  - Kernel
  - Deret fourier
  - wavelet,

Kajian Pustaka

Regresi Semiparametrik

Diberikan pasangan data (x_i, t_i, y_i), $i = 1, 2, \ldots, n$ maka model semiparametrik dirumuskan :

$$y_i = \mu(x_i, t_i) + \varepsilon_i \qquad \mu(x_i, t_i) = f(x_i) + g(t_i)$$

$$y_i = f(x_i) + g(t_i) + \varepsilon_i$$

Komponen Parametrik

Komponen Nonparametrik

Kajian Pustaka

Regresi Parametrik Linear

$$y_i = f(x_i) + \varepsilon_i,$$

linear $f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$

kuadratik $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

Kubik $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

Eksponensial $f(x) = e^{\beta_0 + \beta_1 x}$

model regresi linear berganda dengan *p* variabel prediktor

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},\ i = 1, 2, \ldots, n$$

dimana y_i adalah variabel respon, β_0 adalah konstanta, $\beta_1, \beta_2, \cdots, \beta_p$ adalah koefisien variabel prediktor, dan ε_i adalah *error* dimana $\varepsilon_i \sim N(0, \sigma^2)$.

Estimator Spline

Secara umum, fungsi *spline* derajat *q* adalah sembarang fungsi yang dapat disajikan dalam bentuk :

$$g(t_i) = \theta_0 + \theta_1 t_i^1 + \ldots + \theta_q t_i^q + \lambda_1 (t_i - k_1)_+^q + \lambda_2 (t_i - k_2)_+^q + \ldots + \lambda_m (t_i - k_m)_+^q$$

$$= \sum_{j=0}^{q} \theta_j t_i^j + \sum_{k=1}^{m} \lambda_k (t_i - k_k)_+^q$$

dimana : $(t_i - k_k)_+^q = \begin{cases} (t_i - k_k)^q ; t_i \geq k_k \\ 0 ; t_i < k_k \end{cases}$

θ_j adalah konstanta real, k_k adalah Knot ke-*k*, $k = 1,2, \ldots, m$.

Kajian Pustaka

Estimator Kernel

Untuk mengestimasi fungsi regresi $h(x_i)$ pada model regresi nonparametrik, Nadaraya dan Watson pada tahun 1964 mendefinisikan estimator regresi kernel :

$$\hat{h}_\alpha(z) = n^{-1}\sum_{j=1}^{n} W_{\alpha i}(z) y_i; \qquad W_{\alpha i}(z) = \frac{K_\alpha(z - z_i)}{n^{-1}\sum_{j=1}^{n} K_\alpha(z - z_j)}; \quad K_\alpha(z - z_i) = \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right).$$

Dimana $\hat{h}_\alpha(z)$ adalah fungsi taksiran regresi kernel, z adalah variabel prediktor yang nilainya tidak teramati namun akan digunakan untuk menaksir, z_i adalah variabel prediktor pada data ke-i, Y_i adalah variabel respon data ke-i, K merupakan fungsi kernel, n adalah banyak pengamatan dan α adalah lebar *bandwidth*. Estimator kernel tergantung pada fungsi kernel K dan parameter *bandwith* α.

Beberapa jenis fungsi kernel Uniform, Cosinus, Kuartik, Gaussian, dsb.

$$\text{Gaussian}: \; K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \; -\infty < x < \infty$$

GCV

$$GCV(k) = \frac{MSE(k)}{\left(n^{-1}tr\left(I - H(k)\right)\right)^2}; \quad MSE(k) = n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

Koefisien Determinasi

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Kajian Pustaka

yang mempengaruhi Produksi Padi

- Luas Panen
- Penggunaan Pupuk
- Penggunaan Pestisida
- Penggunaan Benih
- Penggunaan Tenaga Kerja

Struktur Data Penelitian

Struktur Data Penelitian

Data	X_1	X_2	X_3	X_4	X_5	Y
1	$X_{1(1)}$	$X_{2(1)}$	$X_{3(1)}$	$X_{4(1)}$	$X_{5(1)}$	$Y_{1(1)}$
2	$X_{1(2)}$	$X_{2(2)}$	$X_{3(2)}$	$X_{4(2)}$	$X_{5(2)}$	$Y_{1(2)}$
3	$X_{1(3)}$	$X_{2(3)}$	$X_{3(3)}$	$X_{4(3)}$	$X_{5(3)}$	$Y_{1(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
691	$X_{1(691)}$	$X_{2(691)}$	$X_{3(691)}$	$X_{4(691)}$	$X_{5(691)}$	$Y_{1(691)}$

Metode Penelitian

Mengestimasi bentuk kurva regresi semiparametrik campuran *spline truncated* dan kernel

f. Menyelesaikan optimasi :

$$Q(\tilde{\beta},\tilde{\theta}) = \underset{\tilde{\beta},\tilde{\theta}}{Min}\left\|\tilde{y} - X\tilde{\beta} - G(k)\tilde{\theta} - D(\alpha)\tilde{y}\right\|^2$$

Dengan menggunakan derivatif parsial.

g. Menyamakan derivatif partial tersebut dengan nol :

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial \tilde{\beta}} = 0$$

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial \tilde{\theta}} = 0$$

h. Mendapatkan estimasi campuran

$$\hat{\tilde{\mu}}(x,t,z) = \hat{\tilde{f}}(x) + \hat{\tilde{g}}(t) + \hat{\tilde{h}}(z)$$

Metode Penelitian

Mengaplikasikan estimasi campuran yang diperoleh untuk memodelkan produksi padi di Jawa Tengah

a. Melakukan analisis deskriptif data untuk mengetahui gambaran umum tentang data produksi padi di Jawa Tengah, dengan variabel-variabel prediktornya.
b. Membuat plot data antara variabel respon dengan masing-masing variabel prediktor.
c. Menentukan variabel prediktor yang merupakan komponen parametrik, komponen nonparametrik yang didekati dengan fungsi *spline truncated* dan yang didekati dengan fungsi kernel.
d. Memodelkan produksi padi di Jawa Tengah dengan menggunakan pendekatan semiparametrik campuran *spline truncated* dan kernel.
e. Memilih titik knot dan *bandwith* optimal menggunakan metode GCV.
f. Mendapatkan estimasi kurva regresi semiparametrik campuran *spline truncated* dan kernel.
g. Menghitung R^2
h. Menentukan kesimpulan yang sesuai dengan tujuan penelitian.

Metode Penelitian

Model regresi semiparametrik :

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

↓

Dekomposisi kurva regresi μ :

$$\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$$

↓

Menghampiri $f(x_i)$ menggunakan fungsi linear, $g(t_i)$ menggunakan *spline truncated*, dan $h(z_i)$ menggunakan kernel

↓

Mencari bentuk estimasi kurva regresi $\mu(x_i, t_i, z_i)$ menggunakan metode OLS :

$$Q(f, g, h) = \underset{f,g,h}{Min}\left\{\left\|\tilde{y} - \tilde{f}(x) - \tilde{g}(t) - \tilde{h}(z)\right\|^2\right\}$$

↓

$$\tilde{f}(x) = X\tilde{\beta}, \quad \tilde{g}(t) = G(k)\tilde{\theta} \text{ dan } \tilde{h}(z) = D(\alpha)\tilde{y}$$

↓

$$Q(\tilde{\beta}, \tilde{\theta}) = \underset{\tilde{\beta},\tilde{\theta}}{Min}\left\|\tilde{y} - X\tilde{\beta} - G(k)\tilde{\theta} - D(\alpha)\tilde{y}\right\|^2$$

↓

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0, \quad \frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0$$

↓

Mendapatkan estimasi campuran : $\hat{\tilde{\mu}}(x, t, z) = \hat{\tilde{f}}(x) + \hat{\tilde{g}}(t) + \hat{\tilde{h}}(z)$

Diagram alir Langkah Analisis untuk Tujuan Pertama

Metode Penelitian

Mengaplikasikan estimasi campuran yang diperoleh untuk memodelkan produksi padi di Jawa Tengah

Diagram alir Langkah Analisis untuk Tujuan Kedua

Analisi Deskriptif Data

↓

Membuat Plot data masing-masing variabel prediktor dengan variabel respon.

↓

Menentukan variabel prediktor yang merupakan komponen parametrik linear dan komponen nonparametrik (*spline truncated*, dan kernel)

↓

Memodelkan produksi padi

Memilih titik knot dan bandwith optimal menggunakan metode GCV

↓

Mendapatkan estimasi kurva regresi

↓

Menghitung nilai R^2

↓

Kesimpulan

Model Regresi Semiparametrik Campuran

Diberikan data berpasangan (x_i, t_i, z_i, y_i), $i = 1, 2, \ldots, n$ yang diasumsikan mengikuti model regresi semiparametrik :

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i \quad ; \quad \mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$$

$$y_i = f(x_i) + g(t_i) + h(z_i) + \varepsilon_i$$

komponen nonparametrik kernel :

$$\hat{h}_\alpha(z_i) = n^{-1} \sum_{i=1}^{n} W_{\alpha i}(z_i) y_i,$$

$$\begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} = \begin{bmatrix} n^{-1}W_{\alpha 1}(z_1) & n^{-1}W_{\alpha 2}(z_1) & \cdots & n^{-1}W_{\alpha n}(z_1) \\ n^{-1}W_{\alpha 1}(z_2) & n^{-1}W_{\alpha 2}(z_2) & \cdots & n^{-1}W_{\alpha n}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{\alpha 1}(z_n) & n^{-1}W_{\alpha 2}(z_n) & \cdots & n^{-1}W_{\alpha n}(z_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{\tilde{h}}(z) = D(\alpha)\tilde{y},$$

Model regresi semiparametrik dapat ditulis menjadi :

$$\tilde{y} = f(x) + g(t) + \hat{h}(z) + \tilde{\varepsilon}$$

$$= X\tilde{\beta} + G(k)\tilde{\theta} + D(\alpha)\tilde{y} + \tilde{\varepsilon}$$

Model Regresi Semiparametrik Campuran Multivariabel

Diberikan data berpasangan $(x_i, t_{1i}, t_{2i}, \ldots, t_{pi}, z_{1i}, z_{2i}, \ldots, z_{qi}, y_i)$ yang diasumsikan mengikuti model regresi semiparametrik :

$$y_i = \mu(x_i, t_{1i}, t_{2i}, \ldots, t_{pi}, z_{1i}, z_{2i}, \ldots, z_{qi}) + \varepsilon_i$$

$$\mu(x_i, t_{1i}, t_{2i}, \ldots, t_{ri}, z_{1i}, z_{2i}, \ldots, z_{si}) = f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + \ldots + g_p(t_{pi}) + h_1(z_{1i}) + h_2(z_{2i}) + \ldots + h_q(z_{qi}).$$

$$\begin{aligned} y_i &= f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + \ldots + g_p(t_{pi}) + h_1(z_{1i}) + h_2(z_{2i}) + \ldots + h_q(z_{qi}) + \varepsilon_i \\ &= f(x_i) + \sum_{r=1}^{p} g_r(t_{ri}) + \sum_{s=1}^{q} h_s(z_{si}) \end{aligned}$$

Model Regresi Semiparametrik Campuran Multivariabel

Model regresi semiparametrik multivariabel dapat ditulis :

$$y_i = f(x_i) + \sum_{r=1}^{p} g_r(t_{ri}) + \sum_{s=1}^{q} h_s(z_{si}) + \tilde{\varepsilon}$$

$$\tilde{y} = X\tilde{\beta} + G(\tilde{k})\tilde{\theta} + D(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon}$$

Estimasi Kurva Regresi Semiparametrik Campuran

Model regresi semiparametrik campuran :

$$\tilde{y} = X\tilde{\beta} + G(\tilde{k})\tilde{\phi} + D(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon}$$

Error :

$$\begin{aligned}\tilde{\varepsilon} &= \tilde{y} - X\tilde{\beta} - G(\tilde{k})\tilde{\theta} - D(\tilde{\alpha})\tilde{y} \\ &= \tilde{y} - D(\tilde{\alpha})\tilde{y} - X\tilde{\beta} - G(\tilde{k})\tilde{\theta} \\ &= (I - D(\tilde{\alpha}))\tilde{y} - X\tilde{\beta} - G(\tilde{k})\tilde{\theta}\end{aligned}$$

OLS :

$$\begin{aligned}\sum_{i=1}^{n}\varepsilon_i^2 &= \tilde{\varepsilon}^T\tilde{\varepsilon} \\ &= \left[\left(I - D(\tilde{\alpha})\right)\tilde{y} - X\tilde{\beta} - G(\tilde{k})\tilde{\theta}\right]^T\left[\left(I - D(\tilde{\alpha})\right)\tilde{y} - X\tilde{\beta} - G(\tilde{k})\tilde{\theta}\right] \\ &= \tilde{y}^T\left(I - D(\tilde{\alpha})\right)^T\left(I - D(\tilde{\alpha})\right)\tilde{y} - 2\tilde{\beta}^T X^T\left(I - D(\tilde{\alpha})\right)\tilde{y} + \\ &\quad - 2\tilde{\theta}^T G(\tilde{k})^T\left(I - D(\tilde{\alpha})\right)\tilde{y} + \tilde{\beta}^T X^T X\tilde{\beta} + 2\tilde{\beta}^T X^T G(\tilde{k})\tilde{\theta} + \\ &\quad \tilde{\theta}^T G(\tilde{k})^T G(\tilde{k})\tilde{\theta} \\ &= Q(\tilde{\beta},\tilde{\theta})\end{aligned}$$

Diturunkan, kemudian disamakan dengan nol :

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\beta}} = 0 - 2X^T\left(I - D(\tilde{\alpha})\right)\tilde{y} + 2X^T X\hat{\beta} + 2\tilde{\beta}^T X^T G(\tilde{k})\tilde{\theta}$$

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\beta}} = 0,$$

$$\begin{aligned}\hat{\tilde{\beta}} &= \left(X^T X\right)^{-1}\left(X^T\left(I - D(\tilde{\alpha})\right)\tilde{y} - X^T G(\tilde{k})\tilde{\theta}\right) \\ &= \left(X^T X\right)^{-1} X^T\left(I - D(\tilde{\alpha})\right)\tilde{y} - \left(X^T X\right)^{-1} X^T G(\tilde{k})\tilde{\theta}\end{aligned}$$

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\theta}} = 0 - 2G(\tilde{k})^T\left(I - D(\tilde{\alpha})\right)\tilde{y} + 2G(\tilde{k})^T X\tilde{\beta} + 2G(\tilde{k})^T G(\tilde{k})\tilde{\theta}$$

$$\frac{\partial Q(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\theta}} = 0$$

$$\begin{aligned}\hat{\tilde{\theta}} &= \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1}\left(G(\tilde{k})^T\left(I - D(\tilde{\alpha})\right)\tilde{y} - G(\tilde{k})^T X\tilde{\beta}\right) \\ &= \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T\left(I - D(\tilde{\alpha})\right)\tilde{y} - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\tilde{\beta}\end{aligned}$$

Estimasi Kurva Regresi Semiparametrik Campuran

Estimator $\hat{\tilde{\beta}}$

$$\hat{\tilde{\beta}} = \left[I - \left(X^T X\right)^{-1} X^T G(\tilde{k}) \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \right]^{-1} \left[\left(X^T X\right)^{-1} X^T (I - D(\tilde{\alpha})) + \right.$$
$$\left. - \left(X^T X\right)^{-1} X^T G(\tilde{k}) \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) \right] \tilde{y}$$
$$= A(\tilde{k}, \tilde{\alpha}) \tilde{y},$$

dimana :

$$A(\tilde{k}, \tilde{\alpha}) = \left[I - \left(X^T X\right)^{-1} X^T G(\tilde{k}) \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \right]^{-1} \left[\left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right) + \right.$$
$$\left. - \left(X^T X\right)^{-1} X^T G(\tilde{k}) \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) \right]$$

Estimator $\hat{\tilde{\theta}}$

$$\hat{\tilde{\theta}} = \left[I - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \left(X^T X\right)^{-1} X^T G(\tilde{k}) \right]^{-1} \left[\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) + \right.$$
$$\left. - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right) \right] \tilde{y}$$
$$= B(\tilde{k}, \tilde{\alpha}) \tilde{y}$$

dimana :

$$B(\tilde{k}, \tilde{\alpha}) = \left[I - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \left(X^T X\right)^{-1} X^T G(\tilde{k}) \right]^{-1} \left[\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) + \right.$$
$$\left. - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X \left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right) \right]$$

Estimasi Kurva Regresi Semiparametrik Campuran

Estimator Kurva parametrik

$$\hat{\tilde{f}}(x) = X\hat{\tilde{\beta}}$$

$$= X\left[I - \left(X^T X\right)^{-1} X^T G(\tilde{k})\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\right]^{-1}\left[\left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right) + \right.$$

$$\left. -\left(X^T X\right)^{-1} X^T G(\tilde{k})\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right)\right]\tilde{y}$$

$$= C(\tilde{k},\tilde{\alpha})\tilde{y}$$

$$C(\tilde{k},\tilde{\alpha}) = X\left[I - \left(X^T X\right)^{-1} X^T G(\tilde{k})\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\right]^{-1}\left[\left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right) + \right.$$

$$\left. -\left(X^T X\right)^{-1} X^T G(\tilde{k})\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right)\right]$$

Estimator Kurva spline

$$\tilde{g}(t) = G(\tilde{k})\hat{\tilde{\theta}}$$

$$= G(\tilde{k})\left[I - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\left(X^T X\right)^{-1} X^T G(\tilde{k})\right]^{-1}\left[\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) + \right.$$

$$\left. -\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right)\right]\tilde{y}$$

$$= K(\tilde{k},\tilde{\alpha})\tilde{y}$$

$$K(\tilde{k},\tilde{\alpha}) = G(\tilde{k})\left[I - \left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\left(X^T X\right)^{-1} X^T G(\tilde{k})\right]^{-1}\left[\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T \left(I - D(\tilde{\alpha})\right) + \right.$$

$$\left. -\left(G(\tilde{k})^T G(\tilde{k})\right)^{-1} G(\tilde{k})^T X\left(X^T X\right)^{-1} X^T \left(I - D(\tilde{\alpha})\right)\right].$$

Estimasi Kurva Regresi Semiparametrik Campuran

Estimator Kurva semiparametrik campuran

$$\hat{\mu}(x,t,z) = \hat{\tilde{f}}(x) + \hat{\tilde{g}}(t) + \hat{\tilde{h}}(z)$$

$$= C(\tilde{k},\tilde{\alpha})\tilde{y} + K(\tilde{k},\tilde{\alpha})\tilde{y} + D(\alpha)\tilde{y}$$

$$= (C(\tilde{k},\tilde{\alpha}) + K(\tilde{k},\tilde{\alpha}) + D(\alpha))\tilde{y}$$

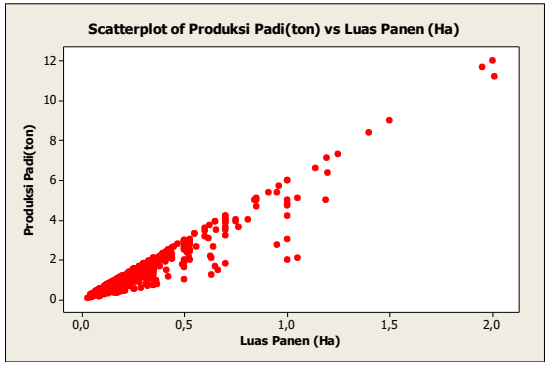
$$= M(\tilde{k},\tilde{\alpha})\tilde{y},$$

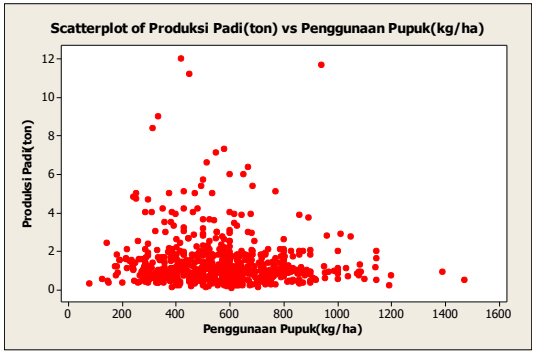
dimana :

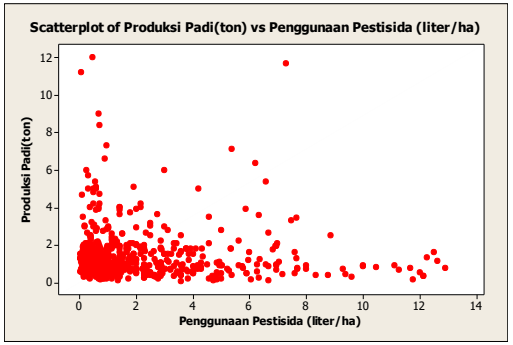
$$M(\tilde{k},\tilde{\alpha}) = C(\tilde{k},\tilde{\alpha}) + K(\tilde{k},\tilde{\alpha}) + D(\alpha)$$

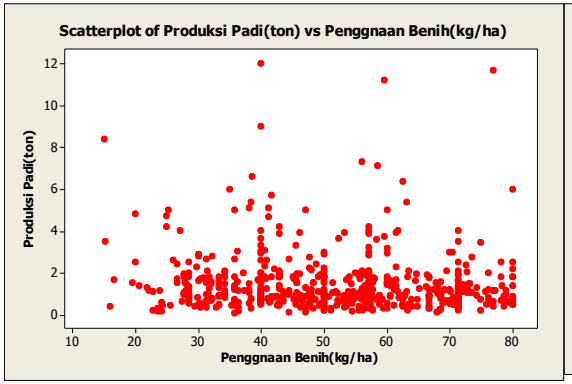
Aplikasi Pada Data Produksi Padi di Jawa Tengah

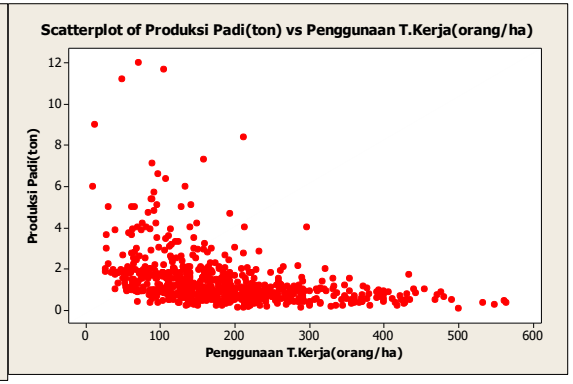
Scatter plot











Aplikasi Pada Data Produksi Padi di Jawa Tengah

Tabel 4.2 GCV dari model-Model

No	Jumlah Variabel	Variabel		GCV
		Spline	Kemel	
1	1 *spline* dan 3 kemel	x_2	x_3, x_4, x_5	0,00031328792
		x_3	x_2, x_4, x_5	0,00031327406
		x_4	x_2, x_3, x_5	0,00038182276
		x_5	x_2, x_3, x_4	0,00025154762
2	2 *spline* dan 2 kemel	x_2, x_3	x_4, x_5	0,00047521137
		x_2, x_4	x_3, x_5	0,00048107085
		x_2, x_5	x_3, x_4	0,00022037370
		x_3, x_4	x_2, x_5	0,0002203538
		x_3, x_5	x_2, x_4	0,00021695737
		x_4, x_5	x_2, x_3	0,00022221830
3	3 *spline* dan 1 kemel	x_3, x_4, x_5	x_2	0,0020932217
		x_2, x_4, x_5	x_3	0,00021272675*
		x_2, x_3, x_5	x_4	0,00021348236
		x_2, x_3, x_4	x_5	0,0019302908

Aplikasi Pada Data Produksi Padi di Jawa Tengah

Tabel 4.5 Komponen Parametrik dan Nonparametrik

No	Variabel Prediktor	Kurva Regresi	Pendekatan yang digunakan	Simbol Variabel
1	Luas Panen	Parametrik	Parametrik Linear	x
2	Penggunaan pupuk	Nonparametrik	Fungsi Spline Truncated	t_1
3	Penggunaan benih		Fungsi Spline Truncated	t_2
4	Penggunaan Tenaga Kerja		Fungsi Spline Truncated	t_3
5	Penggunaan pestisida		Fungsi Kernel	z_1

Aplikasi Pada Data Produksi Padi di Jawa Tengah

model regresi semiparametrik campuran *spline truncated* dan kernel pada data produksi padi dengan *k* titik knot dapat ditulis :

$$y_i = \mu(x_i, t_{1i}, t_{2i}, t_{3i}, z_{1i}) + \varepsilon_i$$
$$= f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + g_3(t_{3i}) + h_1(z_{1i}) + \varepsilon_i$$

dapat ditulis :

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - k_{11})_+ + \lambda_{12}(t_{1i} - k_{12})_+ + \ldots + \lambda_{1k}(t_{1i} - k_{1k}) +$$
$$\theta_{21} t_{2i} + \lambda_{21}(t_{2i} - k_{21})_+ + \lambda_{22}(t_{2i} - k_{22})_+ + \ldots + \lambda_{2k}(t_{2i} - k_{2k})_+ +$$
$$\theta_{31} t_{3i} + \lambda_{31}(t_{3i} - k_{31})_+ + \lambda_{32}(t_{3i} - k_{32})_+ + \ldots + \lambda_{3k}(t_{3i} - k_{3k})_+ +$$
$$\sum_{i=1}^{n} \frac{\frac{1}{\alpha_1} K\left(\frac{z_1 - z_{1i}}{\alpha_1}\right)}{\sum_{i=1}^{n} \frac{1}{\alpha_1} K\left(\frac{z_1 - z_{1i}}{\alpha_1}\right)} y_i + \varepsilon_i$$

Aplikasi Pada Data Produksi Padi di Jawa Tengah

Model dengan Komponen Spline Truncated 3 Titik Knot :

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - k_{11})_+ + \lambda_{12}(t_{1i} - k_{12})_+ + \lambda_{13}(t_{1i} - k_{13})_+ +$$
$$\theta_{21} t_{2i} + \lambda_{21}(t_{2i} - k_{21})_+ + \lambda_{22}(t_{2i} - k_{22})_+ + \lambda_{23}(t_{2i} - k_{23})_+ + \theta_{31} t_{3i} +$$
$$\lambda_{31}(t_{3i} - k_{31})_+ + \lambda_{32}(t_{3i} - k_{32})_+ + \lambda_{33}(t_{3i} - k_{33})_+ + \sum_{i=1}^{n} \frac{\frac{1}{\alpha_1} K\left(\frac{z_1 - z_{1i}}{\alpha_1}\right)}{\sum_{i=1}^{n} \frac{1}{\alpha_1} K\left(\frac{z_1 - z_{1i}}{\alpha_1}\right)} y_i + \varepsilon_i$$

Berdasarkan hasil estimasi yang diperoleh dengan GCV terkecil 0,00021036855, maka model regresi semiparametrik campuran spline truncated dengan 3 titik knot dan kernel dapat ditulis :

$$\hat{y}_i = 1{,}0369 + 5{,}2964 x_i + 0{,}0006 t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ - 0{,}0012(t_{1i} - 330{,}3146)_+ +$$
$$-0{,}0002(t_{1i} - 710{,}4064)_+ - 0{,}1097 t_{2i} + 0{,}1295(t_{2i} - 20{,}9091)_+ - 0{,}0148(t_{2i} - 26{,}8182)_+ +$$
$$-0{,}0067(t_{2i} - 44{,}5455)_+ - 0{,}0153 t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ - 0{,}0038(t_{3i} - 110{,}9091)_+ +$$
$$-0{,}0005(t_{3i} - 262{,}2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)} y_i$$

Aplikasi Pada Data Produksi Padi di Jawa Tengah

Tabel 4.12 Perbandingan GCV untuk Model dengan Komponen Spline Truncated 1,2 dan 3 Knot

No.	Model dengan Komponen Spline	GCV
1	1 Titik Knot	0,00021272675
2	2 Titik Knot	0,00021113277
3	3 Titik Knot	0,00021036855*

Interpretasi Model

Model untuk variabel luas panen

$$\hat{y}_i = 5,2964x_i + c_1$$

dimana :

$$c_1 = 1,0369 + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ - 0,0012(t_{1i} - 330,3146)_+ +$$
$$-0,0002(t_{1i} - 710,4064)_+ - 0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ +$$
$$-0,0148(t_{2i} - 26,8182)_+ - 0,0067(t_{2i} - 44,5455)_+ - 0,0153t_{3i} +$$
$$0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ - 0,0005(t_{3i} - 262,2727)_+ +$$
$$\sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)}{\sum_{i=1}^{n} \frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)} y_i$$

Sehingga dengan mengasumsikan data selain luas panen konstan disimpulkan bahwa ketika luas panen naik sebesar 1 hektar maka produksi padi akan naik sebesar 5,2964 ton

Interpretasi Model

Model untuk variabel penggunaan pupuk

$$\hat{y}_i = 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ - 0,0012(t_{1i} - 330,3146)_+ - 0,0002(t_{1i} - 710,4064)_+ + c_2$$

dimana :

$$\begin{aligned} c_2 = {} & 1,0369 + 5,2964x - 0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ + \\ & - 0,0148(t_{2i} - 26,8182)_+ - 0,0067(t_{2i} - 44,5455)_+ + \\ & - 0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\ & - 0,0005(t_{3i} - 262,2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)}{\sum_{i=1}^{n} \frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)} y_i \end{aligned}$$

Interpretasi Model

$$\hat{y} = \begin{cases} 0,0006t_{1i} + c_2, & t_{1i} \le 203,6173 \\ -0,18326 + 0,0015t_{1i} + c_2, & 203,6173 < t_{1i} \le 330,3146 \\ 0,213122 + 0,0003t_{1i} + c_2, & 330,3146 < t_{1i} \le 710,4064 \\ 0,355203 + 0,0001t_{1i} + c_2, & t_{1i} > 710,4064 \end{cases}$$

- Untuk penggunaan pupuk kurang dari 203,6173 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0006 ton.
- Untuk penggunaan pupuk berada di rentang nilai 203,6173 sampai 330,3146 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0015 ton.
- Untuk penggunaan pupuk berada di rentang nilai 330,3146 sampai 710,4064 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0001 ton.
- Untuk penggunaan pupuk lebih dari dari 710,4064 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0001 ton

Model untuk variabel penggunaan benih

$$\hat{y}_i = -0{,}1097t_{2i} + 0{,}1295(t_{2i} - 20{,}9091)_+ - 0{,}0148(t_{2i} - 26{,}8182)_+ - 0{,}0067(t_{2i} - 44{,}5455)_+ + c_3$$

dimana :

$$\begin{aligned} c_3 = {} & 1{,}0369 + 5{,}2964x_i + 0{,}0006t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ + \\ & -0{,}0012(t_{1i} - 330{,}3146)_+ - 0{,}0002(t_{1i} - 710{,}4064)_+ + \\ & -0{,}0153t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ - 0{,}0038(t_{3i} - 110{,}9091)_+ + \\ & -0{,}0005(t_{3i} - 262{,}2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)} y_i \end{aligned}$$

Interpretasi Model

$$\hat{y} = \begin{cases} -0,1097t_{2i} + c_3, & t_{2i} \le 20,9091 \\ -2,70773 + 0,01985t_{2i} + c_3, & 20,9091 < t_{2i} \le 26,8182 \\ -2,31082 + 0,005t_{2i} + c_3, & 26,8182 < t_{2i} \le 44,5455 \\ -2,01236 - 0,0017t_{2i} + c_3, & t_{2i} > 44,5455 \end{cases}$$

- Untuk penggunaan benih kurang dari 20,9091 kg, dilakukan penambahan benih sebesar 1 kg maka produksi padi turun sebesar 0,1097 ton.
- Untuk penggunaan benih berada di rentang nilai 20,9091 sampai 26,8182 kg , ketika dilakukan penambahan benih sebesar 1 kg maka produksi padi naik sebesar 0,01985 ton.
- Untuk penggunaan benih berada di rentang nilai 26,8182 sampai 44,5455 kg, ketika penggunaan benih naik sebesar 1 kg maka produksi padi naik sebesar 0,005 ton.
- Untuk penggunaan benih lebih dari dari 44,5455 kg, ketika penggunaan benih naik sebesar 1 kg maka produksi padi turun sebesar 0,001 ton.

Interpretasi Model

Model untuk variabel penggunaan tenaga kerja

$$\hat{y}_i = -0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ - 0,0005(t_{3i} - 262,2727)_+$$

dimana :

$$\begin{aligned} c_4 = {} & 1,0369 + 5,2964x_i + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ + \\ & -0,0012(t_{1i} - 330,3146)_+ - 0,0002(t_{1i} - 710,4064)_+ + \\ & -0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\ & -0,0005(t_{3i} - 262,2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)}{\sum_{i=1}^{n} \frac{1}{3,0135} K\left(\frac{z_1 - z_{1i}}{3,0135}\right)} y_i \end{aligned}$$

Interpretasi Model

$$\hat{y} = \begin{cases} -0,0153t_{3i} + c_4, & t_{2i} \le 60,4545 \\ -1,15468 + 0,0038t_{3i} + c_4, & 60,4545 < t_{2i} \le 110,9091 \\ -0,73323 + 0,0001t_{3i} + c_4, & 110,9091 < t_{2i} \le 262,2727 \\ -0,60209 - 0,0005t_{3i} + c_4, & t_{2i} > 262,2727 \end{cases}$$

- Untuk penggunaan tenaga kerja kurang dari 60 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0153 ton.
- Untuk penggunaan tenaga kerja berada di rentang nilai 60 sampai 110 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi naik sebesar 0,0038 ton.
- Untuk penggunaan tenaga kerja berada di rentang nilai 110 sampai 262 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi naik sebesar 0,0001 ton.
- Untuk penggunaan tenaga kerja lebih dari 262 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0005 ton.

Kesimpulan

Model regresi semiparametrik campuran dengan komponen nonparametrik :

$$\hat{\tilde{y}} = \hat{\tilde{\mu}}(x,t,z)$$
$$= X\hat{\tilde{\beta}} + G(\tilde{k})\hat{\tilde{\theta}} + D(\tilde{\alpha})\tilde{y}$$

$$\hat{\tilde{\beta}} = A(\tilde{k},\tilde{\alpha})\tilde{y}$$
$$\hat{\tilde{\theta}} = B(\tilde{k},\tilde{\alpha})\tilde{y}$$
$$\hat{\tilde{f}}(x) = C(\tilde{k},\tilde{\alpha})\tilde{y}$$
$$\hat{\tilde{g}}(t) = K(\tilde{k},\tilde{\alpha})\tilde{y}$$
$$\hat{\tilde{h}}(z) = D(\tilde{\alpha})\tilde{y}$$

Estimator kurva Campuran :

$$\hat{\tilde{\mu}}(x,t,z) = M(\tilde{k},\tilde{\alpha})\tilde{y}$$

dengan : $M(\tilde{k},\tilde{\alpha}) = C(\tilde{k},\tilde{\alpha}) + K(\tilde{k},\tilde{\alpha}) + D(\tilde{\alpha}).$

Kesimpulan

Model regresi semiparametrik campuran *spline truncated* dan kernel diterapkan pada data produksi padi di Jawa Tengah dengan variabel responnya adalah produksi padi (y), sementara itu variabel prediktornya adalah luas panen (x) yang dihampiri dengan parametrik linear, penggunaan pupuk (t_1), penggunaan benih (t_2) dan penggunaan tenaga kerja (t_3) dihampiri dengan fungsi *spline truncated* linear, serta penggunaan pestisida (z) dihampiri dengan fungsi kernel.

Model terbaik diperoleh dari model yang memiliki nilai GCV paling yaitu model dengan komponen *spline truncated* 3 titik knot, dengan R^2 sebesar 91,91%.

Estimasi yang diperoleh:

$$\begin{aligned}\hat{y}_i = {} & 1{,}0369 + 5{,}2964x_i + 0{,}0006t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ - 0{,}0012(t_{1i} - 330{,}3146)_+ + \\ & -0{,}0002(t_{1i} - 710{,}4064)_+ - 0{,}1097t_{2i} + 0{,}1295(t_{2i} - 20{,}9091)_+ - 0{,}0148(t_{2i} - 26{,}8182)_+ + \\ & -0{,}0067(t_{2i} - 44{,}5455)_+ - 0{,}0153t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ - 0{,}0038(t_{3i} - 110{,}9091)_+ + \\ & -0{,}0005(t_{3i} - 262{,}2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z_1 - z_{1i}}{3{,}0135}\right)} y_i\end{aligned}$$

Saran

Penelitian ini menggunakan campuran dua kurva regresi nonparametrik, sehingga untuk penelitian berikutnya dapat menggunakan campuran lebih dari dua komponen nonparametrik yang sesuai dengan pola data. Selain itu, dapat pula dikembangkan dengan menggunakan fungsi *spline truncated* kuadratik maupun kubik dan untuk kernel dapat menggunakan kernel selain Gaussian misalnya kernel *uniform, triangle* dan lain sebagainya.

Daftar Pustaka

Ratnasari, V. Budiantara, I. N., Zain, I. Ratna, M. & Mariati, N. P .A . M. 2015. Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regresssion Model (Application: Data of Poverty in Papua, Indonesia). *International Journal of Basic & Applied Sciences IJBAS-IJENS* Vol : 15 N0:04, 9-14.

Samsodin, M. 2011. Regresi *Spline Polinomial Truncated* Multirespon untuk Pemodelan Indikator Kemiskinan di Provinsi Jawa Timur. *Theses*: Institut Teknologi Sepuluh November Surabaya.

Srinadi. 2002. Estimator Spline pada Model Semiparaetrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.

Susianto, Y. 2008. Model Regresi Semiparametrik kernel untuk Menduga Produksi Padi Sawah di Jawa tengah. *Theses*: Institut Teknologi Sepuluh November Surabaya.

Triyanto, J. 2006. Analisis Produksi Padi di Jawa Tengah. *Theses*: Universitas Diponegoro.

Wahyuningdyawati, F.,Kasijadi., & Heriyanto . 2003. Tingkat Adopsi Teknologi Usaha Tani padi Lahan Sawah di Jawa Timur, Jurnal Pengkajian dan Pengembangan Teknologi Pertanian,Vol. 6, No. 1, hlm 40-49.

Wand, M. P., & Jones, M. C. 1995. *Kernel Smoothing.* London: Chapman & Hall.

Widiantini, D. 2013. Estimasi Model Regresi Nonparametrik Multivariat berdasarkan Estimator Polinomial Lokal Orde Dua, Jurnal Matematika,Vol. 1, No. 1, hlm 29-39.

Zulfikar. 2005. Komponen Bias, Variansi dan MSE Estimator Kernel dalam Regresi Nonparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.

Zulkarnain, I. 2004. Analisis Produksi dan Keuntungan pada Usaha Tani Padi Sawah Jawa Tengan. *Theses*: Universitas Diponegoro.

Zuriani. 2013. Analisis Produksi Padi di Kabupaten Aceh Utara. *Theses*: Universitas Sumatera Utara.