PEMODELAN REGRESI SEMIPARAMETRIK MENGGUNAKAN ESTIMATOR CAMPURAN *SPLINE* TRUNCATED DAN KERNEL

Hesikumalasari(1), I Nyoman Budiantara(2), Vita Ratnasari(3)
(1)(2)(3)Jurusan Statistika, FMIPA, ITS, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail : hesikumala@gmail.com

Abstrak--Variabel respon pada analisis regresi dapat memiliki hubungan linear dengan salah satu variabel prediktor, namun dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Pola data seperti ini dapat dimodelkan dengan model regresi semiparametrik. Diberikan data berpasangan (x_i, t_i, z_i, y_i) dan diasumsikan hubungan antar variabel prediktor dengan variabel respon mengikuti model regresi semiparametrik $y_i = \mu(x_i, t_i, z_i) + \varepsilon_i$. Kurva $\mu(x_i, t_i, z_i)$ dapat ditulis dalam bentuk $\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$. Kurva regresi $f(x_i)$ dihampiri dengan parametrik linear, $g(t_i)$ dihampiri dengan fungsi *spline truncated* dan $h(z_i)$ dihampiri dengan fungsi kernel. Tujuan dari penelitian ini adalah untuk memperoleh bentuk estimator campuran dalam regresi semiparametrik dengan menggunakan estimator campuran *spline truncated* dan kernel. Berdasarkan hasil kajian diperoleh estimator kurva parametrik linear adalah $\hat{\tilde{f}}(x) = \mathbf{C}(\tilde{k}, \tilde{\alpha})\tilde{y}$, estimator kurva *spline truncated* adalah $\hat{\tilde{g}}(t) = \mathbf{K}(\tilde{k}, \tilde{\alpha})\tilde{y}$ dan estimator kurva kernel adalah $\hat{\tilde{h}}(z) = \mathbf{D}(\tilde{\alpha}))\tilde{y}$. Selanjutnya diperoleh estimator kurva regresi semiparametrik campuran *spline* truncated dan kernel $\hat{\mu}(x_i, t_i, z_i) = \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y}$, dengan $\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\alpha)$. Estimator campuran ini tergantung pada titik knot dan parameter *bandwith*. Estimator terbaik diperoleh dari model dengan nilai *Generalized Cross Validation* minimum. Model yang diperoleh kemudian diterapkan pada data produksi padi di Jawa Tengah, dimana variabel responnya adalah produksi padi (y). Sementara itu, variabel prediktornya adalah luas panen (x) yang didekati dengan parametrik linear, penggunaan pupuk (t_1), penggunaan benih (t_2) dan penggunaa tenaga kerja (t_3) yang mengikuti kurva regresi *spline truncated*, serta penggunaan pestisida yang didekati dengan fungsi kernel (z). Model yang diperoleh kemudian diterapkan pada data produksi padi di Jawa Tengah. Model yang diperoleh adalah model estimator campuran yang menghasilkan R^2 sebesar 91,91%.

***Keywords**: Regresi Semiparametrik, Spline Truncated, Kernel, Estimator Campuran.*

I. PENDAHULUAN

Analisis regresi merupakan salah satu metode Statistika yang digunakan untuk mengestimasi pola hubungan antara variabel prediktor (*X*) dengan variabel respon (*Y*). Tujuan utama dalam analisis regresi adalah mencari bentuk estimasi kurva regresi. Pendekatan yang biasa digunakan untuk mengestimasi kurva regresi yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik. Regresi parametrik dilakukan apabila bentuk kurva regresinya diketahui. Sedangkan regresi nonparametrik dilakukan jika bentuk kurva regresinya tidak diketahui.

Dalam analisis regresi terdapat beberapa kasus dimana fungsi regresinya memuat dua komponen yaitu komponen parametrik dan komponen nonparametrik. Model regresi yang memuat dua komponen tersebut dinamakan model regresi semiparametrik atau biasa disebut model linear parsial. Penelitian sebelumnya mengenai regresi semiparametrik telah dilakukan oleh [20], [3], [2], dan lainnya.

Pendekatan regresi parametrik untuk mengestimasi kurva regresi diasumsikan bentuk kurvanya diketahui seperti linier, kuadratik, kubik, eksponensial dan lain sebagainya. Sementara dalam permasalahan pola hubungan yang bentuk kurvanya tidak diketahui atau tidak mempunyai bentuk pola hubungan tertentu digunakan pendekatan regresi nonparametrik [11]. Pendekatan regresi nonparametrik tidak mensyaratkan variabel yang diteliti harus mempunyai kurva regresi dengan pola tertentu atau bentuk kurvanya diketahui. Oleh karenanya, dengan menggunakan pendekatan ini, akan lebih fleksibel. Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain estimator histogram, *spline*, kernel, polinomial lokal, *deret fourier*, dan *wavelet* [10].

Spline merupakan salah satu teknik estimasi regresi nonparametrik yang pertama kali dikembangkan oleh Whittaker pada tahun 1923[11]. *Spline* merupakan salah satu model yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik [9]. Sementara itu Estimator kernel diperkenalkan pertama kali oleh Rosenblatt (1956) dan Parzen (1962) yang disebut estimator densitas kernel Rosenblatt-Parzen. Kelebihan estimator kernel adalah fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat [5]. Penelitian sebelumnya mengenai regresi nonparametrik *spline* maupun kernel telah dilakukan oleh [24], [23], [10], [25],

[12], [15], [1], [19], [14], [27], [16], [13], [8], [7], [18], [22] dan lainnya.

Model-model regresi nonparametrik maupun semiparametrik yang dikembangkan oleh para peneliti selama ini pada umumnya menggunakan jenis metode estimasi yang sama untuk sebagian bahkan semua prediktornya. Hal ini disebabkan oleh anggapan bahwa pola dari masing-masing prediktor dianggap mempunyai pola yang sama sehingga peneliti menggunakan hanya satu bentuk estimator model untuk semua variabel prediktor. Sementara itu, pada kenyataannya sering dijumpai kasus-kasus dengan pola data yang berbeda dari masing-masing variabel prediktor. Oleh karena itu, untuk mengatasi masalah tersebut beberapa peneliti telah mengembangkan estimator kurva regresi campuran nonparametrik dimana masing-masing pola data dalam model regresi nonparametrik dihampiri dengan estimator kurva yang sesuai dengan pola data. Penelitian sebelumnya yang menggunakan estimator campuran pernah dilakukan oleh [21] yaitu melibatkan estimator deret fourier dan *spline truncated*. Sedangkan penelitian yang dilakukan [6], [17] dan [18] telah melibatkan kernel dan *spline truncated*. Penelitian-penelitian tersebut melibatkan estimator campuran pada regresi nonparametrik. Sementara itu, belum ada peneliti yang melibatkan estimator campuran pada regresi semiparametrik. Oleh karena itu, pada penelitian ini akan digunakan model campuran yaitu *spline truncated* dan kernel pada regresi semiparametrik.

Tanaman padi adalah salah satu komoditas pangan utama karena merupakan tanaman penghasil beras yang menjadi makanan pokok bagi mayoritas penduduk di Indonesia. Kebutuhan padi di Indonesia terus meningkat dari tahun ke tahun dikarenakan jumlah penduduk yang terus meningkat. Peningkatan jumlah penduduk secara otomatis akan menambah konsumsi beras dalam negeri. Oleh karena itu produksi padi perlu ditingkatkan agar dapat memenuhi kebutuhan tersebut. Pusat produksi tanaman padi di Indonesia terdapat di Pulau Jawa, dengan produksi padi terbanyak ini dihasilkan oleh Jawa Timur dan Jawa Barat, sementara Jawa Tengah masih berada dibawah angka produksi dua provinsi tersebut. Bahkan dalam kurun waktu antara tahun 2010-2014. Penelitian sebelumnya tentang produksi padi pernah dilakukan oleh [22] dan [24], dimana pada penelitian-penelitian tersebut diketahui bahwa luas panen berhubungan linear terhadap produksi padi dan variabel lainnya berhubungan nonparametrik terhadap produksi padi. [22] mendekati kurva regresi untuk variabel nonparametriknya menggunakan kernel dan [24] menggunakan deret fourier. Sementara itu, dengan mempertimbangkan perbedaan pola data, maka bisa digunakan/didekati dengan estimator yang berbeda yang sesuai dengan pola data. Sehingga, berdasarkan perumusan masalah di atas, maka tujuan berikutnya dari penelitian ini adalah untuk menganalisis faktor-faktor yang berpengaruh terhadap produksi padi di Provinsi Jawa Tengah menggunakan regresi semiparametrik campuran *spline truncated* dan kernel.

II. KAJIAN PUSTAKA

A. Regresi Parametrik Linear

Diberikan data berpasangan (x_i, y_i), $i = 1, 2, ..., n$, dimana pola hubungannya dapat dinyatakan dalam model regresi $y_i = f(x_i) + \varepsilon_i$. Kurva regresi $f(x_i)$ dihampiri dengan kurva regresi parametrik linear dimana :

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

Sehingga model regresi linear berganda dengan *p* variabel prediktor dapat disajikan dalam bentuk :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (2)$$

Atau dapat ditulis dalam notasi matriks :

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \tilde{\varepsilon}$$

Estimator parameter model diperoleh berdasarkan berbagai metode yang telah dikenal dalam Statistika, yaitu *Least Square* atau *Maximum Likelihood Estimator*, jika digunakan metode *least square* didapat estimator parameter :

$$\hat{\tilde{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{y}$$

B. Estimator *Spline*

Diberikan data berpasangan (t_i, y_i), $i = 1, 2, ..., n$, dimana pola hubungannya dapat dinyatakan dalam model regresi $y_i = g(t_i) + \varepsilon_i$. Kurva regresi $g(t_i)$ dihampiri dengan fungsi *spline truncated*. Secara umum, fungsi *spline* derajat *q* adalah sembarang fungsi yang dapat disajikan dalam bentuk [9] :

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

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

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

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

Sehingga model regresinya menjadi :

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

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

atau dapat ditulis dalam notasi matriks :

$$\tilde{y} = \mathbf{G}(k)\tilde{\theta} + \tilde{\varepsilon}$$

Jika digunakan metode *least square* diperoleh estimator parameter :

$$\hat{\tilde{\theta}} = \left[\mathbf{G}(k)'\mathbf{G}(k)\right]^{-1}\mathbf{G}(k)'\tilde{y}$$

C. Estimator Kernel

Diberikan data berpasangan (z_i, y_i), $i = 1, 2, ..., n$, dimana pola hubungannya dapat dinyatakan dalam model

regresi $y_i = h(z_i) + \varepsilon_i$. Kurva regresi $h(z_i)$ dihampiri dengan fungsi kernel. Untuk mengestimasi fungsi regresi $h(z_i)$ pada model regresi nonparametrik, Nadaraya dan Watson pada tahun 1964 mendefinisikan estimator regresi kernel sehingga disebut estimator Nadaraya-Watson [11].

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

dengan

$$W_{\alpha i}(z) = \frac{\frac{1}{\alpha}K\left(\frac{z - z_i}{\alpha}\right)}{n^{-1}\sum_{j=1}^{n}\frac{1}{\alpha}K\left(\frac{z - z_i}{\alpha}\right)},$$

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 yang bisa berupa kernel gaussian, uniform, kuadratik, dan lain-lain. Sementara itu, n adalah banyak pengamatan dan α adalah lebar *bandwidth*. Estimator kernel tergantung pada fungsi kernel K dan parameter *bandwith* α.

Estimator kernel dapat ditulis dalam notasi matriks :

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

D. Produksi Padi

Untuk mencapai kemandirian dan kedaulatan pangan, ketersediaan komoditas tanaman pangan untuk konsumsi dalam negeri harus terjamin. Komoditas yang nilainya paling strategis adalah padi. Padi merupakan tanaman penghasil beras yang merupakan makanan pokok sebagian besar penduduk indonesia. Karena demikian strategisnya padi termasuk dalam salah satu sasaran utama pemerintah dalam target swasembada sebagaimana tertuang dalam RPJM 2015-2019. Sebagai komoditas terpenting di dalam pembangunan pertanian maka berbagai usaha dilakukan untuk meningkatkan produksi padi.

Pertanian mempunyai kontribusi penting terhadap perekonomian yaitu kontribusi produk dalam sumbangannya terhadap Produk Domestik Regional Bruto (PDRB) dan kontribusi pasar dibandingkan dengan usaha-usaha lainnya. Menurut publikasi Pusdatin Kementerian Pertanian Republik Indonesia, sekitar 70 persen dari total produk domestik bruto Indonesia yang disumbangkan oleh sektor pertanian berasal dari pertanian padi. Pada tahun 2010, pertanian padi berkontribusi terhadap tujuh persen dari total produk domestik bruto Indonesia secara keseluruhan. Peran penting lainnya adalah dalam penyediaan kebutuhan pangan penduduk, apalagi dengan semakin meningkatnya jumlah penduduk yang berarti bahwa kebutuhan akan pangan juga semakin meningkat. Oleh karena itu peningkatan produksi padi perlu dilakukan dan perlu diperhatikan secara intensif pula apa saja yang mempengaruhi faktor produksinya. Produksi padi sangat ditentukan oleh penggunaan faktor-faktor produksi seperti luas panen, penggunaan pupuk, benih, pestisida dan tenaga kerja.

III. METODOLOGI PENELITIAN

A. Data dan Variabel

Data yang akan digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi BPS tahun 2008. Data yang diambil merupakan data yang berkaitan dengan produksi padi di Jawa Tengah, dengan unit penelitian adalah rumah tangga yang sumber pendapatan utamanya berasal dari tanaman padi (rumah tangga usaha tani padi) sejumlah 691. Variabel responnya adalah produksi padi. Sementara itu, variabel prediktornya adalah luas panen, penggunaan pupuk, penggunaan pestisida, penggunaan benih dan penggunaan tenaga kerja.

B. Tahapan Penelitian

Tahapan penelitian yang dilakukan adalah menentukan model regresi semiparametrik campuran *spline* trunated linear dan kernel, selanjutnya mengkaji estimasi dari model campuran tersebut dan terakhir mengaplikasikan model pada data produksi padi di Jawa Tengah.

IV. HASIL DAN PEMBAHASAN

A. Model Regresi Semiparametrik Campuran *Spline Truncated* dan Kernel

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

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i$$

kurva regresi μ diasumsikan bersifat additif sehingga dapat ditulis menjadi :

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

Bentuk pola hubungan antara variabel respon y_i dengan variabel prediktor x_i diasumsikan mengikuti pola linear. Sementara itu, pola hubungan antara variabel respon y_i dengan variabel prediktor t_i dan z_i diasumsikan tidak diketahui, dimana bentuk pola hubungan variabel respon y_i dengan variabel prediktor t_i diasumsikan berubah-ubah pada sub-sub interval tertentu. Sementara itu, bentuk pola hubungan variabel respon y_i dengan variabel prediktor z_i diasumsikan tidak memiliki pola tertentu. Kurva $g(t_i)$ didekati dengan fungsi *spline truncated*. Sementara itu, kurva $h(z_i)$ dapat dihampiri dengan fungsi kernel. Dengan demikian, kurva $\mu(x_i, t_i, z_i)$ secara keseluruhan bisa didekati dengan kurva regresi semiparametrik campuran *spline truncated* linear dan kernel.

Komponen kurva regresi parametrik linear didefinisikan oleh :

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

Dari persamaan fungsi linear sederhana di atas, berlaku untuk $i = 1$ sampai n, sehingga :

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \ldots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \ldots + \beta_p x_{np} \end{bmatrix}$$

Selanjutnya dapat ditulis menjadi :

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

atau dapat ditulis dalam notasi matriks :

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

Sementara itu, komponen kurva regresi nonparametrik *spline truncated* linear didefinisikan oleh :

$$g(t_i) = \theta_1 t_i + \sum_{k=1}^{m} \lambda_k (t_i - k_k)_+$$

$$= \theta_1 t_i + \lambda_1 (t_i - k_1)_+ + \ldots + \lambda_m (t_i - k_m)_+$$

Dengan

$$(t_i - k_k)_+ = \begin{cases} (t_i - k_k)_+ & , t_i \geq k_k \\ 0 & , t_i < k_k \end{cases}$$

Dimana $\theta_1, \lambda_1, \ldots, \lambda_m$ merupakan parameter-parameter yang tidak diketahui, $k_1, k_2, \ldots, k_k$ merupakan titik knot dimana $k_1 < k_2 < \ldots < k_k$.

Dari persamaan fungsi *spline truncated* di atas, berlaku untuk $i = 1$ sampai n, sehingga

$$\begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} = \begin{bmatrix} \theta_1 t_1 + \lambda_1 (t_1 - k_1)_+ + \ldots + \lambda_m (t_1 - k_m)_+ \\ \theta_1 t_2 + \lambda_1 (t_2 - k_1)_+ + \ldots + \lambda_m (t_2 - k_m)_+ \\ \vdots \\ \theta_1 t_n + \lambda_1 (t_n - k_1)_+ + \ldots + \lambda_m (t_n - k_m)_+ \end{bmatrix}$$

Selanjutnya dapat ditulis :

$$\begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} = \begin{bmatrix} t_1 & (t_1 - k_1)_+ & \cdots & (t_1 - k_m)_+ \\ t_2 & (t_2 - k_1)_+ & \cdots & (t_2 - k_m)_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_n & (t_n - k_1)_+ & \cdots & (t_n - k_m)_+ \end{bmatrix} \begin{bmatrix} \theta_1 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

Sehingga dapat ditulis dalam notasi matriks :

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

Sementara itu, komponen kurva regresi nonparametrik kernel didefinisikan oleh :

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

Dimana :

$$W_{\alpha i}(z_i) = \frac{K_\alpha(z_i - z_j)}{n^{-1} \sum_{j=1}^{n} K_\alpha(z_i - z_j)}$$

Dan

$$K_\alpha(z_i - z_j) = \frac{1}{\alpha} K\left(\frac{z_i - z_j}{\alpha}\right)$$

Dari persamaan fungsi kernel di atas, berlaku $i = 1$ sampai *n*, sehingga :

$$\begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} = \begin{bmatrix} n^{-1} \sum_{i=1}^{n} W_{\alpha i}(z_1) y_i \\ n^{-1} \sum_{i=1}^{n} W_{\alpha i}(z_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^{n} W_{\alpha i}(z_n) y_i \end{bmatrix}$$

$$= \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) y_1 + \cdots + n^{-1} W_{\alpha n}(z_1) y_n \\ n^{-1} W_{\alpha 1}(z_2) y_1 + \cdots + n^{-1} W_{\alpha n}(z_2) y_n \\ \vdots \\ n^{-1} W_{\alpha 1}(z_t \quad (4.5) \quad n^{-1} W_{\alpha n}(z_n) y_n \end{bmatrix}$$

Selanjutnya dapat ditul :

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

Atau dapat ditulis dalam notasi matriks :

$$\hat{\tilde{h}}(z_i) = \mathbf{D}(\alpha)\tilde{y}$$

Sehingga model regresi semiparametrik campuran *spline truncated* dan kernel adalah:

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

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

Selanjutnya model regresi semiparametrik campuran *spline truncated* dan kernel multivariabel diperoleh :

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

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

B. Estimasi Kurva Regresi Semiparametrik Campuran

Estimator semiparametrik campuran *spline truncated* dan kernel dapat diperoleh dengan menggunakan metode *ordinary least square* (OLS). Dari model kemudian diperoleh :

$$\tilde{\varepsilon} = \tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} - \Delta(\tilde{\alpha})\tilde{y}$$

$$= (I - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta}$$

Sehingga menggunakan optimasi *ordinary least square*, maka :

$$\sum_{i=1}^{n} \varepsilon_i^2 = \tilde{\varepsilon}^T \tilde{\varepsilon}$$

Kemudian diperoleh :

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

Selanjutnya melakukan derivatif parsial terhadap $\tilde{\beta}$ dan $\tilde{\theta}$, Untuk $\tilde{\beta}$:

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

Sementara itu untuk $\tilde{\theta}$:

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

Selanjutnya disamakan dengan nol, untuk parameter $\tilde{\beta}$ diperoleh :

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

Sementara itu untuk parameter $\tilde{\theta}$ diperoleh :

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

Kemudian mensubstitusikan $\tilde{\theta}$ ke dalam $\hat{\tilde{\beta}}$, sehingga diperoleh **:**

$$\hat{\tilde{\beta}} = \mathbf{A}(\tilde{k}, \tilde{\alpha}) \tilde{y},$$

dimana :

$$\mathbf{A}(\tilde{k}, \tilde{\alpha}) = P_1 Q_1$$

$$P_1 = \left[I - \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k})\right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X}\right]^{-1}$$

$$\begin{aligned}Q_1 &= \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \left(I - \mathbf{D}(\tilde{\alpha})\right) + \\ &\quad - \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k})\right)^{-1} \mathbf{G}(\tilde{k})^T \left(I - \mathbf{D}(\tilde{\alpha})\right)\end{aligned}$$

Selanjutnya, substitusikan $\tilde{\beta}$ ke dalam $\hat{\tilde{\theta}}$, diperoleh :

$$\hat{\tilde{\theta}} = \mathbf{B}(\tilde{k}, \tilde{\alpha}) \tilde{y},$$

dimana :

$$\mathbf{B}(\tilde{k}, \tilde{\alpha}) = P_1 Q_1$$

$$P_1 = \left[I - \left(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k})\right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k})\right]^{-1}$$

$$\begin{aligned}Q_1 &= \Big[\left(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k})\right)^{-1} \mathbf{G}(\tilde{k})^T \left(I - \mathbf{D}(\tilde{\alpha})\right) - \\ &\quad \left(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k})\right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \left(I - \mathbf{D}(\tilde{\alpha})\right)\Big]\end{aligned}$$

Sehingga estimator kurva parametrik linear adalah :

$$\begin{aligned}\hat{\tilde{f}}(x) &= \mathbf{X} \hat{\tilde{\beta}} \\ &= \mathbf{C}(\tilde{k}, \tilde{\alpha}) \tilde{y},\end{aligned}$$

dimana :

$$\mathbf{C}(\tilde{k}, \tilde{\alpha}) = X P_1 Q_1$$

Selanjutnya, estimator kurva regresi *spline truncated* linear diperoleh :

$$\begin{aligned}\tilde{g}(t) &= G(\tilde{k}) \hat{\tilde{\theta}} \\ &= \mathbf{K}(\tilde{k}, \tilde{\alpha}) \tilde{y},\end{aligned}$$

dimana :

$$\mathbf{K}(\tilde{k}, \tilde{\alpha}) = \mathbf{G}(\tilde{k}) P_2 Q_2$$

Sehingga estimator campuran *spline truncated* dan kernel dalam regresi semiparametrik diperoleh :

$$\begin{aligned}\hat{\mu}(x_i, t_i, z_i) &= \hat{\tilde{f}}(x) + \hat{\tilde{g}}(t) + \hat{\tilde{h}}(z) \\ &= \mathbf{C}(\tilde{k}, \tilde{\alpha}) \tilde{y} + \mathbf{K}(\tilde{k}, \tilde{\alpha}) \tilde{y} + \mathbf{D}(\tilde{\alpha}) \tilde{y} \\ &= (\mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha})) \tilde{y} \\ &= \mathbf{M}(\tilde{k}, \tilde{\alpha}) \tilde{y}\end{aligned}$$

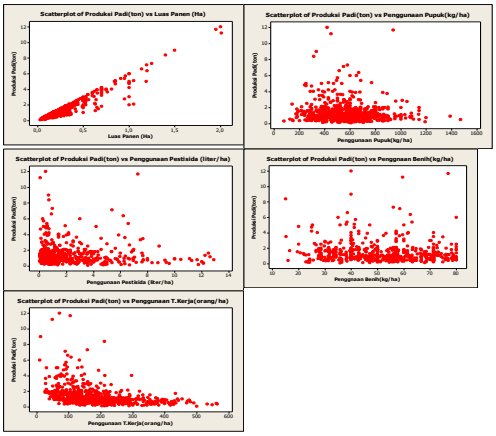
Dimana :

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

C. Aplikasi pada Data Produksi Padi di Jawa Tengah

Bentuk pola hubungan antara variabel respon dengan masing-masing variabel prediktor dapat dilihat dari grafik *scatter plot*. Gambar I menunjukkan bahwa variabel luas panen mengikuti bentuk pola data linear. Sementara itu bentuk pola hubungan variabel selain dari variebel luas panen tidak diketahui, sehingga secara teoritis keempatnya dapat dihampiri dengan nonparametrik. Selanjutnya bentuk pola hubungan antara

variabel respon produksi dengan masing-masing variabel prediktor penggunaan pupuk, penggunaan benih dan penggunaan tenaga kerja berubah-ubah pada sub-sub interval tertentu, sehingga dapat didekati dengan fungsi *spline truncated*. Sementara itu, bentuk pola hubungan antara variabel respon produksi padi dengan penggunaan pestisida tidak mengikuti pola tertentu sehingga dapat didekati dengan fungsi kernel.



Gambar 1. *Scatter Plot* Variabel Respon dengan Variabel Prediktor

Selain melihat bentuk pola hubungan data melalui *scatter plot*, untuk menentukkan variabel mana yang didekati dengan *spline truncated* atau kernel untuk variabel-variabel nonparametrik, juga dilakukan dengan memeriksa nilai GCV dari masing-masing variabel prediktor. Penjelasan untuk hal ini dapat dilihat pada Tabel 1 berikut :

Tabel 1. GCV dari Model Campuran

No	Jumlah Variabel	Variabel		GCV
		Spline	Kernel	
1	1 *spline* dan 3 kernel	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 kernel	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 kernel	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

Berdasarkan Tabel 1 dari semua kemungkinan diperoleh GCV minimum 0,00021695737 jika komponen *spline*nya adalah x_2, x_4, x_5 dan komponen kernelnya adalah x_3. Komponen parametrik selanjutnya akan disimbolkan dengan *x*, nonparametrik *spline* disimbolkan dengan *t*, dan nonparametrik kernel disimbolkan dengan *z*. Dengan demikian keterangan variabel prediktornya seperti yang terlihat pada Tabel 2.

Tabel 2. Komponen Parametrik dan Nonparametrik

No	Variabel Prediktor	Pendekatan yang digunakan	Simbol Variabel
1	Luas Panen	Parametrik Linear	x
2	Penggunaan pupuk	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

Sehingga model regresi semiparametrik campuran *spline truncated* dan kernel pada data produksi padi tersebut dapat ditulis sebagai berikut :

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

Kemudian dapat ditulis menjadi :

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

Pemilihan model regresi semiparametrik campuran *spline truncated* dan kernel terbaik diperoleh dari penentuan titik knot optimal dan *bandwidth* optimal yang didapatkan dari membandingkan nilai-nilai GCV yang diperoleh. Pada penelitian ini dicobakan untuk 1 titik knot, 2 titik knot dan 3 titik knot, dengan model terbaik yaitu yang mempunyai nilai GCV terkecil di antara nilai-nilai yang diperoleh tersebut. Berikut ini perbandingan nilai GCV dari model-model dengan 1,2 dan 3 titik knot

Tabel 2. Perbandingan GCV

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

Berdasarkan Tabel diketahui bahwa nilai GCV minimum dihasilkan oleh model dengan 3 titik knot, sehingga model regresi semiparametrik campuran *spline* truncated dan kernel diperoleh :

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

Dengan nilai R^2 sebesar 91,91%, ini berarti variabel yang digunakan dapat menjelaskan model sebesar 91,91 %. Dari model kita dapat menginterpretasikan sebagai berikut :

1) Model untuk variabel luas panen
   Dengan mengasumsikan data selain luas panen konstan, maka diperoleh model umumnya :
   $$\hat{y}_i = 5{,}2964x_i + c_1$$
   dimana
   $$c_1 = 1{,}0369 + 0{,}0006t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ +$$
   $$-0{,}0012(t_{1i} - 330{,}3146)_+ - 0{,}0002(t_{1i} - 710{,}4064)_+ +$$
   $$-0{,}1097t_{2i} + 0{,}1295(t_{2i} - 20{,}9091)_+ +$$
   $$-0{,}0148(t_{2i} - 26{,}8182)_+ - 0{,}0067(t_{2i} - 44{,}5455)_+ +$$
   $$-0{,}0153t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ +$$
   $$-0{,}0038(t_{3i} - 110{,}9091)_+ - 0{,}0005(t_{3i} - 262{,}2727)_+ +$$
   $$\sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)} y_i$$
   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.
2) Model untuk variabel penggunaan pupuk
   Dengan mengasumsikan data selain penggunaan pupuk konstan, maka diperoleh model umumnya:
   $$\hat{y}_i = 0{,}0006t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ +$$
   $$-0{,}0012(t_{1i} - 330{,}3146)_+ +$$
   $$-0{,}0002(t_{1i} - 710{,}4064)_+ + c_2$$
   dimana :
   $$c_2 = 1{,}0369 + 5{,}2964x - 0{,}1097t_{2i} +$$
   $$0{,}1295(t_{2i} - 20{,}9091)_+ - 0{,}0148(t_{2i} - 26{,}8182)_+ +$$
   $$-0{,}0067(t_{2i} - 44{,}5455)_+ - 0{,}0153t_{3i}$$
   $$+0{,}0191(t_{3i} - 60{,}4545)_+ - 0{,}0038(t_{3i} - 110{,}9091)_+ +$$
   $$-0{,}0005(t_{3i} - 262{,}2727)_+ +$$
   $$\sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)} y_i$$
   Untuk interpretasi model ini digunakan fungsi seperti berikut :
   $$\hat{y} = \begin{cases} 0{,}0006t_{1i} + c_2, & t_{1i} \le 203{,}6173 \\ -0{,}18326 + 0{,}0015t_{1i} + c_2, & 203{,}6173 < t_{1i} \le 330{,}3146 \\ 0{,}213122 + 0{,}0003t_{1i} + c_2, & 330{,}3146 < t_{1i} \le 710{,}4064 \\ 0{,}355203 + 0{,}0001t_{1i} + c_2, & t_{1i} > 710{,}4064 \end{cases}$$
   a. Untuk penggunaan pupuk kurang dari 203,6173 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0006 ton.
   b. 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.
   c. 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.
   d. Untuk penggunaan pupuk lebih dari dari 710,4064 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0001 ton.
3) Model untuk variabel penggunaan benih
   Dengan mengasumsikan data selain penggunaan benih konstan, maka diperoleh model umumnya :
   $$\hat{y}_i = -0{,}1097t_{2i} + 0{,}1295(t_{2i} - 20{,}9091)_+ +$$
   $$-0{,}0148(t_{2i} - 26{,}8182)_+ +$$
   $$-0{,}0067(t_{2i} - 44{,}5455)_+ + c_3$$
   dimana :

$$c_3 = 1{,}0369 + 5{,}2964x_i + 0{,}0006t_{1i} +$$

$$0{,}0009(t_{1i} - 203{,}6173)_+ - 0{,}0012(t_{1i} - 330{,}3146)_+ +$$

$$-0{,}0002(t_{1i} - 710{,}4064)_+ - 0{,}0153t_{3i} +$$

$$0{,}0191(t_{3i} - 60{,}4545)_+ - 0{,}0038(t_{3i} - 110{,}9091)_+ +$$

$$-0{,}0005(t_{3i} - 262{,}2727)_+ +$$

$$\sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)} y_i$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

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

a. Untuk penggunaan benih kurang dari 20,9091 kg, dilakukan penambahan benih sebesar 1 kg maka produksi padi turun sebesar 0,1097 ton.
b. 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.
c. 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.
d. Untuk penggunaan benih lebih dari dari 44,5455 kg, ketika penggunaan benih naik sebesar 1 kg maka produksi padi turun sebesar 0,001 ton.

4) Model untuk variabel penggunaan tenaga kerja
Dengan mengasumsikan data selain penggunaan tenaga kerja konstan, maka diperoleh model umumnya :

$$\hat{y}_i = -0{,}0153t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ +$$

$$-0{,}0038(t_{3i} - 110{,}9091)_+ +$$

$$-0{,}0005(t_{3i} - 262{,}2727)_+ + c_4$$

dimana :

$$c_4 = 1{,}0369 + 5{,}2964x_i + 0{,}0006t_{1i} + 0{,}0009(t_{1i} - 203{,}6173)_+ +$$

$$-0{,}0012(t_{1i} - 330{,}3146)_+ - 0{,}0002(t_{1i} - 710{,}4064)_+ +$$

$$-0{,}0153t_{3i} + 0{,}0191(t_{3i} - 60{,}4545)_+ +$$

$$-0{,}0038(t_{3i} - 110{,}9091)_+ - 0{,}0005(t_{3i} - 262{,}2727)_+ +$$

$$\sum_{i=1}^{691} \frac{\frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)}{\sum_{i=1}^{n} \frac{1}{3{,}0135} K\left(\frac{z - z_i}{3{,}0135}\right)} y_i$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

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

a. Untuk penggunaan tenaga kerja kurang dari 60 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0153 ton.
b. 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.
c. 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.
d. Untuk penggunaan tenaga kerja lebih dari 262 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0005 ton.

V. KESIMPULAN DAN SARAN

Kurva regresi semiparametrik campuran *spline truncated* dan kernel merupakan model yang terdiri dari tiga komponen utama, yaitu terdapat komponen parametrik linear, komponen *spline truncated* dan komponen kernel. Kurva ini diharapkan mampu mendekati pola data dengan baik karena masing-masing data dihampiri dengan kurva yang sesuai. *Spline* yang digunakan dalam penelitian ini adalah *spline truncated* linear, dengan knot dibatasi hanya 1, 2 dan 3 titik knot. Sementara itu, estimator dan fungsi kernel yang digunakan adalah estimator kernel Nadaraya-Watson dan kernel gaussian. Sebagai saran untuk penelitian selanjutnya dapat menggukan fungsi *spline* yang berbeda seperti kuadratik dan kubik, serta bisa menggunakan fungsi kernel yang lain seperti triangel, uniform, dan sebagainya. Selain itu, bisa juga dilakukan kajian model dengan campuran selain *spline* dan kernel,

serta kajian model dengan lebih dari dua kurva nonparametrik yang berbeda.

n DAFTAR PUSTAKA

[1]Adyana, I. G. 2011. Estimator *Spline* dalam Regresi Nonparametrik Multirespon. *Theses* : Institut Teknologi Sepuluh November Surabaya.

[2]Amelia, D. 2014. Penerapan Regresi Semiparametrik *Spline* pada Pemodelan Nilai UN Siswa SMKN 1 Nguling Pasuruan. *Theses*: Institut Teknologi Sepuluh November Surabaya.

[3]Asmin, S. 2009. Estimator *Spline* pada Model Semiparaetrik. *Theses*: Institut Teknologi sepuluh November Surabaya.

[4]Asriani, L.J. 2013. Model Regresi Semiparametrik Deret Fourier. *Theses*: Institut Teknologi Sepuluh November Surabaya.

[5]Budiantara, I N. dan Mulianah. 2007, Pemilihan *Bandwidth* Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, Vol. 10, No. 2, Hal. 159-166.

[6]Budiantara, I.N., Ratnasari, V., Ratna, M., and Zain, I. 2015. The Combination of *Spline* and Kernel estimator for Nonparametrik Regression and Its Properties, *Applied Mathematical Science*, 9, No 122, 6083-6094.

[7]Chamidah, N. 2008. Inferensi Kurva Regresi Nonparametrik Berdasarkan Estimator Polinomial Lokal dengan *Error* Lognormal, Jurnal Penelitian Eksakta, Vol. 7, No. 1, hlm 61-69.

[8]Eichner, G., & Stute, W. 2012. Kernel Adjusted Nonparametric regression. *Journal of Statistical Planning and Inference*, No. 142, hal 2537-2544.

[9]Eubank, R.L. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Deker : New York.

[10]Eubank, R. L. 1999. *Spline Smoothing and Nonparametric Regression*. Marcel Deker: New York.

[11]Hardle, W. 1990. *Applied Nonparametric Regression*. Cambrige University Press: New York.

[12]Hazelton, M. L., & Turlach, B. A. 2011. Semiparametric Regression with Shape-constrained Penalized Splines. *Computational Statistics and Data Analysis*, No. 55, hal 2871-2879.

[13]Kulasekera, Q. Z., & Gallagher, C. 2011. Local Adaptive Smoothing in Kernel Regression Estimation. *Statistics and Probability Letters*, No. 88, hal 540-547.

[14]Muller, H. G. 1984. Optimal Designs for Nonparametric Kernel Regression. *Statistics and Probability Letters*, No. 2, hal 285-290.

[15]Nafi, M. 2010. Estimasi Interval *Spline* dalam regresi Nonparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.

[16]Okumura, H., & Naito, K. 2006. Non-parametric Kernel Regression for Multinominal Data. *Journal of Multivariate Analysis*, No. 97, hal 2009-2022.

[17]Purnomo, A. A. 2016. Estimator Campuran Kernel dan Regresi Spline Truncated Linear Multivariabel dalam Regresi Nonparametrik. *Theses* : Institut Teknologi Sepuluh November Surabaya.

[18]Rory. 2016. Regresi Campuran Nonparametrik Spline Linear Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua. *Theses* : Institut Teknologi Sepuluh November Surabaya.

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

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

[21]Sudiarsa, I. W., Budiantara, I. N., Suhartono, Purnami, S. W. 2015. *Combined Estimator Fourier Series and Spline Truncated in Multivariabel Nonparametrik Regression*, *Applied Mathematical Science*, Vol.9, 2015, no. 100, 4997-5010, HIKARI Ltd.

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

[23]Wahba G. 1990. *Spline Models for Observational Data*, SIAM Pensylvania.

[24]Wang, R. 1985. The Dimention and Basis of Space of Multivariate Splines. *Journal of Computational and Applied Mathematics*, No. 12&13, hal 163-177.

[25]Weinert, H. L. 2009. A Fast Compact Algorithm for Cubic Spline Smoothing. *Computational Statistics and Data Analysis*, No. 53, hal 932-s940.

[27]William, R. 1989. On Nonparaetric Regression with Higher-Order Kernels. *Journal of Statistical Planning and Inference*, No. 23, hal 141-151.

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