
**Pemodelan Regresi Semiparametrik *B Spline*
(Studi Kasus: Pengaruh Harga Emas dan Minyak Mentah Dunia
Terhadap Indeks Harga Saham Gabungan)**

M. Pratama Aryansah¹, Suparti^{2*}

^{1,2}Program Studi Statistika, Universitas Diponegoro, Tembalang, Indonesia

**Corresponding author:* suparti@live.undip.ac.id

Abstract. *The increases of the world gold and crude oil prices have a big role as a main factor that effect composite stock price index, the effect can make investors to buy stock from Bursa Efek Indonesia. Regression semiparametric used in this research for a purpose to get combined parametric and nonparametric with B Spline approach. B Spline is a development of spline to overcome weaknesses in making singular matrix at a high order spline with many knot points and close together. Variable parametric component is composite stock price index with crude oil price, and variable nonparametric component is composite stock price index with gold price that got obtained from January 2015 until December 2022. The result from this research is best regression semiparametric B-Spline modelling can be obtained using some combination of order and knot points. The optimal point is obtained on 2nd order using 4 knot point (1.135;1.319,15;1.320,75;1.323,25) with a minimum GCV value is 100.227,8. The best measure of goodness with a coefficient of determination value (R-Square) obtained a value 78,8%, because the value is more than 67% make it as a strong model. MAPE value is 3,37% that has a value less than 10 %, make this model have a perfect forecasting ability.*

Keywords: *Gold; Crude Oil; Composite Stock Price Index; Semiparametric B Spline; GCV*

1. PENDAHULUAN

Investasi menjadi suatu keharusan karena merupakan salah satu penggerak perekonomian yang mempercepat perkembangan ekonomi suatu negara. Keterkaitan antara investasi, pasar modal dengan perkembangan ekonomi adalah pasar modal dijadikan tempat mengumpulkan modal dari investor yang kemudian disalurkan untuk perkembangan negara. Indeks Harga Saham Gabungan menjadi alat ukur untuk memeriksa perkembangan kinerja pasar modal di Indonesia [1].

Terdapat beberapa faktor yang dapat mempengaruhi indeks harga saham gabungan, diantaranya adalah perubahan harga minyak mentah dunia dan harga emas dunia [2]. Besarnya pengaruh harga emas dan minyak mentah dunia terhadap indeks harga saham gabungan dapat ditemukan menggunakan pendekatan analisis regresi. Penentuan metode regresi ini dilakukan secara visual, dimana IHSG memiliki bentuk pola data yang diketahui bentuk fungsinya adalah linier pada harga minyak mentah, tetapi pada harga emas memiliki bentuk pola data yang naik turun sehingga tidak diketahui bentuk fungsi persamaannya yang menandakan jika analisis

regresi dapat dilakukan dengan gabungan parametrik dan nonparametrik. Pemodelan regresi yang menggunakan dua pendekatan ini dinamakan regresi semiparametrik [3].

Regresi *spline* merupakan pendekatan nonparametrik yang mampu dalam menangani pola data yang berfluktuasi. Regresi *spline* memiliki kelemahan, ketika orde *spline* tinggi, terdapat banyak titik knot dan berdekatan sehingga menghasilkan matriks singular, oleh karena itu diperlukan pendekatan yang mampu menanggulangi kelemahan *spline*, yaitu regresi *B Spline* [4]. Metode regresi semiparametrik *B Spline* dapat digunakan dalam melakukan pemodelan regresi 2 variabel prediktor terhadap 1 variabel respon studi kasus harga emas dan minyak mentah dunia terhadap indeks harga saham gabungan.

2. METODE PENELITIAN

2.1 Tinjauan Pustaka

Indeks Harga Saham Gabungan (IHSG) merupakan salah satu indikator yang menunjukkan pergerakan harga saham yang biasanya digunakan oleh Bursa Efek Indonesia. Sekarang ini (Februari 2023) jumlah emiten yang terdaftar di Bursa Efek Indonesia mencapai 852 emiten. Emas merupakan standar keuangan yang digunakan sebagai alat tukar yang bertahan secara lama (abadi) dan diterima oleh semua negara di dunia [5]. Patokan harga emas yang sampai sekarang digunakan oleh seluruh dunia adalah standar pasar emas London, atau dinamakan London Gold Fixing (LGF). Minyak mentah yang telah diolah akan dijadikan sebagai energi lalu digunakan untuk kegiatan produksi, semakin besar kebutuhan minyak mentah maka semakin besar pula harga saham pada sektor pertambangan mengalami kenaikan. Patokan harga minyak mentah dunia yang umum digunakan saat ini adalah West Texas Intermediate (WTI).

2.2 Regresi Parametrik

Analisis regresi merupakan teknik statistika yang umumnya digunakan untuk meneliti dan memodelkan hubungan antar variabel [6]. Variabel yang digunakan pada analisis regresi terdiri dari variabel dependen (y) dan independen (x). Regresi parametrik adalah pendekatan regresi untuk pola data yang diketahui bentuk fungsi persamaannya adalah linier, eksponensial, dan lain – lain. Asumsi harus dipenuhi, jika tidak terpenuhi maka pendekatan parametrik akan mendapatkan hasil yang menyesatkan sehingga perlu diatasi dengan pendekatan nonparametrik [7]. Model regresi linier sederhana dari data x_i dan y_i yang diberikan n pengamatan dapat ditulis seperti pada Persamaan 1[6].

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i=1,2, \dots, n \tag{1}$$

Persamaan 1 dapat dirumuskan menjadi matriks yang dapat dilihat pada Persamaan 1.

$$Y = X\beta + \varepsilon \tag{2}$$

dengan

$$Y = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}_{n \times 1}; X = \begin{bmatrix} 1 & x_1 & 1 & x_2 & \dots & 1 & x_n \end{bmatrix}_{n \times 1}; \beta = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix}_{n \times 1}$$

Estimasi parameter dapat diperoleh menggunakan *Ordinary Least Square* (OLS) dengan meminimumkan *Sum Square Error* (SSE), yaitu jumlah dari kuadrat *error* antara nilai aktual dengan nilai estimasi. *Error* dapat diperoleh dari Persamaan 2.

$$\varepsilon = Y - X\beta \tag{3}$$

Error yang telah diperoleh pada Persamaan 3, dapat diperoleh fungsi SSE yang terlihat pada Persamaan 4.

$$\begin{aligned} SSE &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - \beta^T X^T Y - (\beta^T X^T Y)^T + \beta^T X^T X \beta \\ &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta \end{aligned} \tag{4}$$

Persamaan 5 merupakan persamaan 4 yang diturunkan terhadap β , kemudian disamakan.

$$\begin{aligned} -2X^T Y + 2X^T X \hat{\beta} &= 0 \\ X^T X \hat{\beta} &= X^T Y \\ \hat{\beta} &= (X^T X)^{-1} X^T Y \end{aligned} \tag{5}$$

Estimator persamaan model linier pada Persamaan 2 dapat dilihat pada Persamaan 6.

$$\hat{Y} = X\hat{\beta} \tag{6}$$

2.3 Regresi Nonparametrik B-Spline

Regresi nonparametrik memiliki fungsi yang tidak diketahui dan umumnya diasumsikan termuat dalam suatu ruang fungsi dengan dimensi yang tak hingga dan memiliki fungsi f yang bersifat mulus [8]. Model regresi nonparametrik dengan variabel respon (y_i) dan prediktor (z_i) dapat ditulis sebagai Persamaan 7.

$$y_i = f(z_i) + \varepsilon_i \tag{7}$$

dengan $f(z_i)$ merupakan fungsi regresi nonparametrik yang tidak diketahui. Jika $f(z_i)$ didekati dengan fungsi *B Spline* berorde m dengan p titik knot yang dapat dilihat pada Persamaan 8.

$$y_i = \sum_{j=1}^{m+p} B_j S_{-(m-j),m}(z_i) + \varepsilon_i \tag{8}$$

dengan:

- m : Orde
- p : Titik Knot
- B_j : Koefisien ke - j pada regresi nonparametrik
- S : Basis fungsi *B Spline*

Knot tambahan berjumlah $2m$ diperlukan untuk membangun basis fungsi *B Spline* sebagai berikut:

$K_{-(m-1)}, \dots, K_{-1}, K_0, K_{p+1}, \dots, K_{p+m}$, dengan

$$K_{-(m-1)} = \dots = K_0 = b_0 \text{ dan } K_{p+1} = \dots = K_{p+m} = b_1$$

dengan b_0 adalah nilai minimum data z_i pada *in sample* dan b_1 adalah nilai maksimum data y_i pada *in sample*.

Basis fungsi *B-Spline* menggunakan titik–titik knot K_q di orde m dimana $q = -(m-1), \dots, p$ secara rekursif dapat didefinisikan menjadi Persamaan 9 [9].

$$S_{q,m}(z_i) = \frac{z_i - K_q}{K_{q+m-1} - K_q} S_{q,m-1}(z_i) + \frac{K_{q+m} - z_i}{K_{q+m} - K_{q+1}} S_{q+1,m-1}(z_i) \tag{9}$$

Persamaan 8 jika dijadikan matriks dapat dirumuskan menjadi Persamaan 10.

$$Y = SB + \varepsilon \tag{10}$$

dengan

$$Y = [y_1 \ y_2 \ \dots \ y_n]_{n \times 1}; S = [S_{1-m,m}(z_1) \ S_{2-m,m}(z_1) \ \dots \ S_{p,m}(z_1) \ S_{1-m,m}(z_2) \ S_{2-m,m}(z_2) \ \dots \ S_{p,m}(z_2) \ \dots \ \dots \ \dots \ \dots]_{n \times (m+p)}$$

$$; B = [B_1 \ B_2 \ \dots \ B_{(m+p)}]_{(m+p) \times 1}; \varepsilon = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]_{n \times 1}$$

Fungsi *B Spline* berorde m dengan titik knot adalah $\lambda = \{K_1, \dots, K_p\}$ pada Persamaan 8 dapat disajikan dalam bentuk Persamaan 11 [6].

$$y_i = \sum_{j=1}^{m+p} B_{\lambda_j} S_{-(m-j),m}(z_i) + \varepsilon_i \tag{11}$$

Persamaan 11 dapat ditulis menjadi matriks dengan perumusan sebagai Persamaan 12.

$$Y = S_\lambda B_\lambda + \varepsilon \tag{12}$$

Estimasi parameter dapat diperoleh menggunakan OLS dengan meminimumkan SSE, dimana *error* diperoleh dari Persamaan 10 seperti dapat dilihat pada Persamaan 13.

$$\varepsilon = Y - S_\lambda B_\lambda \tag{13}$$

Error yang telah diperoleh pada Persamaan 13, maka dapat diperoleh fungsi SSE yang terlihat pada Persamaan 14.

$$SSE = \sum_{i=1}^n \varepsilon_i^2$$

$$= \varepsilon^T \varepsilon$$

$$= (Y - S_\lambda B_\lambda)^T (Y - S_\lambda B_\lambda) \tag{14}$$

$$SSE = Y^T Y - B_\lambda^T S_\lambda^T Y - (B_\lambda^T S_\lambda^T Y)^T + B_\lambda^T S_\lambda^T S_\lambda B_\lambda$$

$$= Y^T Y - 2B_\lambda^T S_\lambda^T Y + B_\lambda^T S_\lambda^T S_\lambda B_\lambda$$

Persamaan 14 diturunkan terhadap B_λ , kemudian disamakan dengan 0 menjadi Persamaan 15.

$$- 2S_\lambda^T Y + 2S_\lambda^T Y \hat{B}_\lambda = 0 \tag{15}$$

$$S_\lambda^T S_\lambda \hat{B}_\lambda = S_\lambda^T Y$$

$$\hat{B}_\lambda = (S_\lambda^T S_\lambda)^{-1} S_\lambda^T Y$$

Estimator persamaan model *B Spline* pada Persamaan 12 dapat dilihat pada Persamaan 16.

$$\hat{Y} = S_\lambda \hat{B} \tag{16}$$

2.4 Regresi Semiparametrik *B-Spline*

Regresi semiparametrik adalah metode statistika untuk mendapatkan pemodelan antara variabel respon dan prediktor dengan sebagian diketahui bentuk fungsi dan sebagian lagi tidak diketahui bentuk fungsinya dari pola data [10]. Bentuk umum model persamaan regresi semiparametrik dari data x_i, z_i , dan y_i adalah sebagai Persamaan 17.

$$y_i = \beta_0 + \beta_1 x_i + f(z_i) + \varepsilon_i \tag{17}$$

Model persamaan regresi semiparametrik apabila didekati menggunakan pendekatan *B Spline*, maka model Persamaan 17 akan menjadi Persamaan 18.

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{m+p} B_{\lambda j} S_{-(m-j),m}(z_i) + \varepsilon_i \tag{18}$$

Persamaan 18 dapat dijadikan matriks sebagai Persamaan 19.

$$Y = X\beta + S_{\lambda} B_{\lambda} + \varepsilon \tag{19}$$

Selanjutnya ditentukan $C = [X S_{\lambda}]$, $\gamma = [\beta B_{\lambda}]$ sehingga Persamaan 19 dapat dirumuskan menjadi Persamaan 20.

$$Y = C\gamma + \varepsilon \tag{20}$$

dengan,

$$Y = [y_1 y_2 \dots y_n]_{nx1}; S_{\lambda} = [S_{1-m,m}(z_1) S_{2-m,m}(z_1) \dots S_{p,m}(z_1) S_{1-m,m}(z_2) S_{2-m,m}(z_2) \dots S_{p,m}(z_2) \dots \dots \dots]_{(m+p) \times n}$$

$$B_{\lambda} = [B_{\lambda 1} B_{\lambda 2} \dots B_{\lambda(m+p)}]_{(m+p) \times 1}^T;$$

$$X = [1 x_1 \ 1 x_2 \ \dots \ 1 x_n]_{(nx1)+1}; \beta = [\beta_0 \ \beta_1]; \varepsilon = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]_{nx1}$$

Estimasi parameter dapat diperoleh menggunakan OLS dengan meminimumkan SSE dengan *error* dapat diperoleh dari Persamaan 20 sebagai Persamaan 21.

$$\varepsilon = Y - C\gamma \tag{21}$$

Error yang telah diperoleh pada Persamaan 21, dapat diperoleh fungsi SSE pada Persamaan 22.

$$\begin{aligned} SSE &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \varepsilon^T \varepsilon \\ &= (Y - C\gamma)^T (Y - C\gamma) \\ &= Y^T Y - \gamma^T C^T Y - (Y^T C^T Y)^T + \gamma^T C^T C \gamma \\ &= Y^T Y - 2\gamma^T C^T Y + \gamma^T C^T C \gamma \end{aligned} \tag{22}$$

Persamaan 4 diturunkan terhadap γ , kemudian disamakan dengan 0 menjadi Persamaan 23.

$$\begin{aligned} -2C^T Y + 2C^T \hat{C}\gamma &= 0 \\ C^T \hat{C}\gamma &= C^T Y \\ \hat{\gamma} &= (C^T C)^{-1} C^T Y \end{aligned} \tag{23}$$

Estimator persamaan model semiparametrik *B Spline* pada Persamaan 20 dapat dilihat pada Persamaan 24.

$$\hat{Y} = C\hat{\gamma}$$

$$\hat{Y} = C(C^T C)^{-1} C^T Y \tag{24}$$

$$\hat{Y} = H_{(\lambda=K_1, K_2, \dots, K_p)} Y$$

dengan $H_{(\lambda=K_1, K_2, \dots, K_p)}$ adalah matriks simetris dan definit positif.

Model terbaik semiparametrik *B Spline* merupakan titik knot optimal pada nilai GCV minimum, maka perlu dilakukan perbandingan nilai GCV di setiap titik knot dan orde. Secara teoritis, GCV mempunyai sifat optimal asimptotik yang tidak terdapat pada metode lain [11]. Kriteria GCV dirumuskan sebagai Persamaan 25.

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(\frac{1}{n} \text{trace}[I-H_\lambda]\right)^2} \tag{25}$$

dengan:

$$MSE(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

I : Matriks identitas

H_λ : Matriks Hat

2.5 Ukuran Kebaikan dan Evaluasi Kinerja Model

Koefisien determinasi adalah suatu ukuran yang memperlihatkan persentase keberagaman pada variabel respon dapat dijelaskan oleh variabel prediktor. Perumusan untuk menghitung koefisien determinasi [12] dapat dilihat pada Persamaan 26.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{26}$$

Nilai R^2 memiliki interval $0 \leq R^2 \leq 1$, dengan R^2 yang mendekati 1 sebagai nilai terbaik adalah variabel respon dapat dijelaskan dari variabel prediktor. Nilai $R^2 > 0,67$ mengindikasikan model tergolong kuat. *Mean Absolute Percentage Error* (MAPE) merupakan ukuran kesalahan dari suatu peramalan dengan memperlihatkan rata-rata persentase nilai absolut *error* antara nilai estimasi dan nilai aktual. Perumusan MAPE didefinisikan pada Persamaan 27.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{27}$$

Nilai MAPE dengan hasil kurang dari 10%, maka pemodelan akan mendapatkan peramalan yang sangat baik [13].

2.6 Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder bulanan yang diperoleh dari situs investing.com yang terhitung dari bulan Januari 2015 hingga Desember 2022. Data terdiri dari data *in sample* dari bulan Januari 2015 hingga Desember 2021 dan data *out sample* dari bulan Januari 2022 hingga Desember 2022. Variabel penelitian yang digunakan telah diuraikan dan dapat dilihat pada Tabel 1. *Software* yang digunakan pada penelitian ini ialah aplikasi R 4.1.3 serta Rstudio menggunakan *package shiny* yang dikemas dalam antarmuka GUI.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Uraian	Satuan
Y	Respon	Indeks Harga Saham Gabungan (<i>Closing Price</i>)	Rupiah
X	Prediktor 1	Harga Minyak Mentah Dunia (<i>Closing Price</i>)	USD
Z	Prediktor 2	Harga Emas Dunia (<i>Closing Price</i>)	USD

Tahapan analisis dalam pemodelan regresi semiparametrik *B Spline* terdiri dari beberapa tahapan sebagai berikut:

1. Mempersiapkan *software* GUI R.
2. Menetapkan data *in sample* dan *out sample* yang akan digunakan.
3. Melakukan input data *in sample* dan *out sample* sesuai dengan variabel yang ditentukan.
4. Membuat analisis deskriptif menggunakan data *in sample*.
5. Membuat *scatterplot* variabel dependen (respon) terhadap setiap variabel independent (prediktor).
6. Menentukan variabel komponen parametrik serta nonparametrik sesuai dengan hasil *scatterplot*.
7. Menentukan orde (2,3,4) dan banyaknya knot yang dibatasi maksimal hingga 4 knot untuk diuji.
8. Menghitung estimasi parameter model semiparametrik *B Spline*
9. Menghitung nilai GCV.
10. Menentukan model semiparametrik *B Spline* terbaik berdasarkan nilai GCV minimum.
11. Menghitung kriteria kebaikan model terbaik dengan nilai R^2 data *in sample*.
12. Menghitung evaluasi kinerja model terbaik menggunakan nilai MAPE di data *out sample*.

3. HASIL DAN PEMBAHASAN

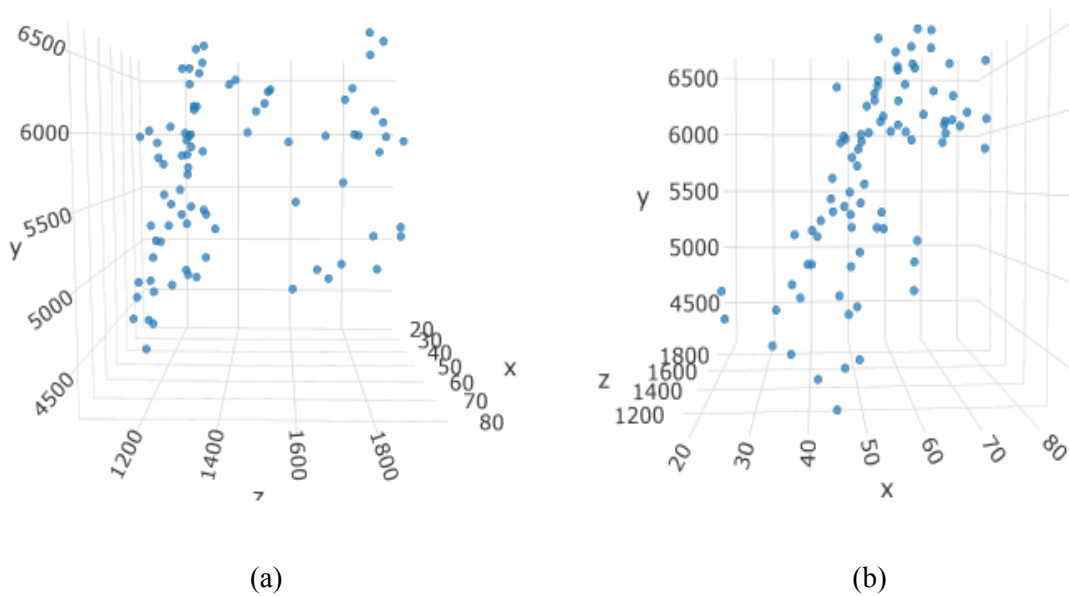
Regresi semiparametrik pada umumnya menggunakan variabel yang terdiri dari 1 variabel respon, 1 variabel prediktor parametrik dan 1 variabel prediktor nonparametrik. *Scatterplot* 3D dalam penentuan komponen dapat diperoleh seperti pada Gambar 1. Gambar 1 sebelah kiri merupakan tampilan plot variabel IHSG dengan Harga Emas sebagai nonparametrik karena memiliki bentuk pola data yang naik turun dan tersebar acak, sedangkan Gambar 1 sebelah kanan merupakan tampilan plot variabel IHSG dengan Harga Minyak Dunia sebagai parametrik karena memiliki pola data linier. Tabel statistik deskriptif pada data *in sample* dapat dilihat pada Tabel 2.

Model semiparametrik *B Spline* terbaik diperoleh dengan pengujian menggunakan beberapa kombinasi dari setiap orde dan titik knot hingga mendapatkan nilai GCV yang dapat dilihat pada Tabel 3, Tabel 4, dan Tabel 5.

Titik optimal GCV untuk model terbaik berada di orde ($m = 2$), jumlah titik knot ($p = 4$) yaitu $K_1 = 1.135$; $K_2 = 1.319,15$; $K_3 = 1.320,75$; $K_4 = 1.323,25$, dengan nilai GCV minimum adalah 100.227,8. Basis fungsi *B Spline* dibangun menggunakan knot tambahan $2m$ menggunakan nilai minimal harga emas *in sample* ($b_0 = 1.061,9$), nilai maksimal harga emas *out sample* ($b_1 = 1.964,9$) sebagai berikut:

$$K_{-1} = K_0 = 1.061,9; K_1 = 1.135; K_2 = 1.319,15; K_3 = 1.320,75;$$

$$K_4 = 1.323,25; K_5 = K_6 = 1.964,9$$



Gambar 1. (a) Scatterplot 3D IHSG (Y) dengan Harga Emas (Z), (b) Scatterplot 3D IHSG (Y) dengan Minyak Mentah Dunia (X)

Tabel 2. Tabel Statistik Deskriptif *In Sample*

Variabel	Jumlah	Minimum	Median	Mean	Maksimum
IHSG	84	4.224	5.836	5.646	6.606
Minyak Mentah	84	18,84	52,51	53,44	83,57
Emas	84	1.062	1.307	1.415	1.965

Tabel 3. Perbandingan GCV Orde 2

Jumlah Knot	Letak Titik Knot	GCV
1	1.345,05	110.658,9
2	1.135 ; 1.295,4	107.280,1
3	1.319,15 ; 1.320,75 ; 1.323,25	104.685,3
4	1.135 ; 1.319,15 ; 1.320,75 ; 1.323,25	100.227,8

Tabel 4. Perbandingan GCV Orde 3

Jumlah Knot	Letak Titik Knot	GCV
1	1.584,2	114.809,7
2	1.159,1 ; 1.528,4	105.946,1
3	1.061,9 ; 1.159,1 ; 1.528,4	105.946,1
4	1.319,15 ; 1.320,75 ; 1.322,5 ; 1.323,25	105.428,9

Tabel 5. Perbandingan GCV Orde 4

Jumlah Knot	Letak Titik Knot	GCV
1	1.202,45	107.394,7
2	1.061,9 ; 1.202,45	107.394,7
3	1.061,9 ; 1.135 ; 1.142,35	108.126,9
4	1.062,25 ; 1.098,4 ; 1.111,80 ; 1.114	108.896,8

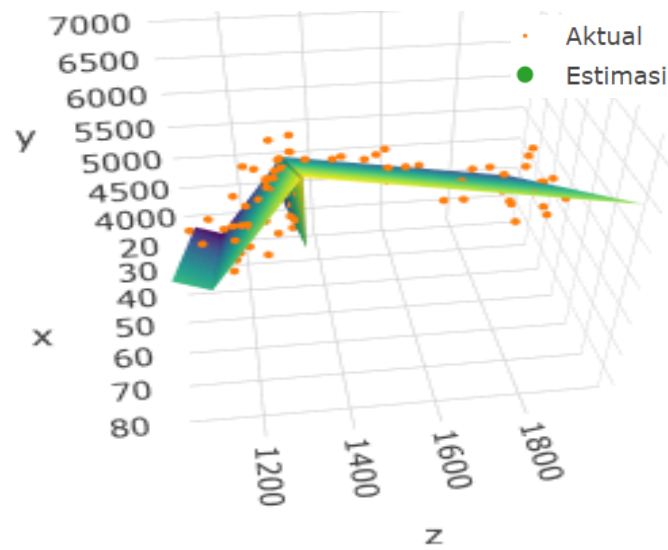
Titik knot tersebut digunakan pada basis fungsi *B Spline* $S_{q,2}(z_i)$ untuk $q = -1,0,1,2,3,4$ sebagai berikut:

$$S_{-1,2}(z_i) = \left\{ \frac{1.135-z_i}{73,1}; 1.061,9 \leq z_i \leq 1.135 \right\}; \quad S_{0,2}(z_i) = \left\{ \frac{z_i-1.061,9}{73,1}, 1.061,9 \leq z_i \leq 1.135 \right\}; \quad S_{1,2}(z_i) = \left\{ \frac{z_i-1.135}{184,15}; 1.135 \leq z_i \leq 1.319,15 \right\}; \quad S_{2,2}(z_i) = \left\{ \frac{z_i-1.319,15}{1,6}; 1.319,15 \leq z_i \leq 1.320,75 \right\}; \quad S_{3,2}(z_i) = \left\{ \frac{z_i-1.320,75}{2,5}; 1.320,75 \leq z_i \leq 1.322,5 \right\}; \quad S_{4,2}(z_i) = \left\{ \frac{z_i-1.323,25}{641,65}; 1.323,25 \leq z_i \leq 1.964,9 \right\}$$

Model regresi semiparametrik *B Spline* terbaik pada orde 2 dengan letak titik knot di 1135; 1319,15; 1320,75; 1323,25 memperoleh model estimasi parameter regresi semiparametrik *B Spline* yaitu:

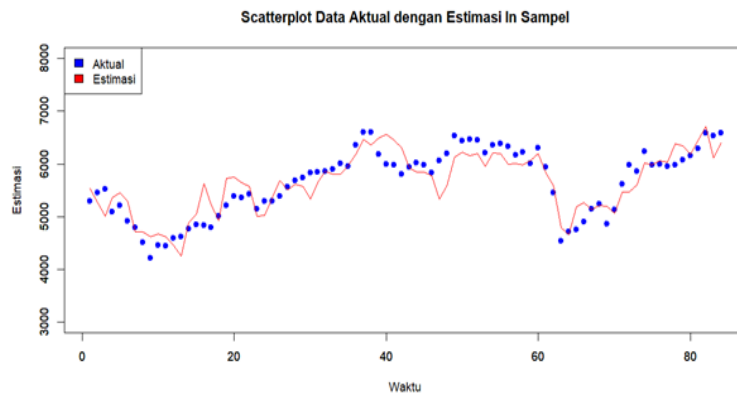
$$\hat{y}_i = 3.188.669 + 32,474(x_i) + 75,388S_{-1,2}(z_i) - 77,094S_{0,2}(z_i) + 1. \tag{27}$$

Persamaan 27 dapat memperoleh nilai estimasi y_i yang kemudian dimasukkan nilainya ke dalam *scatterplot* 3D untuk membandingkan nilai estimasi dan aktual yang dapat dilihat pada Gambar 2, sedangkan pada Gambar 3 merupakan perbandingan nilai data aktual y *in sample* dengan nilai estimasi \hat{y} berdasarkan waktu.

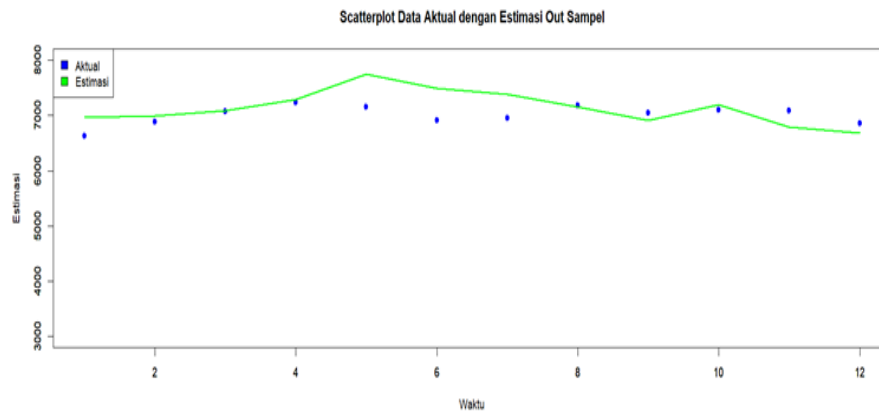


Gambar 2. *Scatterplot* 3D Estimasi IHSG dengan Harga Minyak Mentah dan Emas

Model yang dianggap terbaik akan dilakukan perhitungan ukuran kebaikan model menggunakan koefisien determinasi, Model regresi semiparametrik *B Spline* terbaik di orde 2 menggunakan letak titik knot di 1.135; 1.319,15; 1.320,75; 1.323,25 mendapatkan nilai koefisien determinasi sebesar 0,788 atau 78,8%, karena memiliki nilai lebih besar dari 67% maka dapat disimpulkan jika model yang diperoleh merupakan model yang kuat. Model terbaik kemudian diukur kinerja modelnya dalam melakukan peramalan menggunakan data *out sample* dan diperoleh nilai MAPE sebesar 3,37% sehingga memiliki peramalan yang sangat baik. Gambar 4 menunjukkan *scatterplot* pada data *out sample*.



Gambar 3. *Scatterplot* Data Aktual Estimasi *In Sample* Berdasarkan Waktu



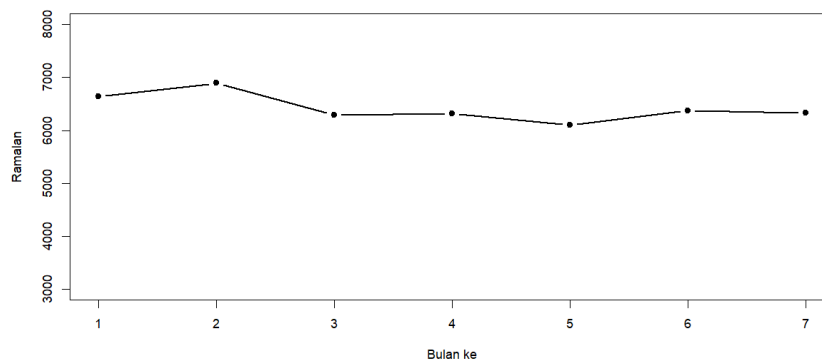
Gambar 4. Scatterplot Data Aktual dengan Nilai Estimasi Out Sample

4. KESIMPULAN

Pemodelan regresi semiparametrik *B Spline* terbaik menggunakan orde 2 dengan 4 titik knot, yaitu 1.135 ; 1.319,15 ; 1.320,75 ; 1.323,25. Pemodelan tersebut diperoleh nilai GCV minimum sebesar 100.227,8 ; nilai R-Square sebesar 0,788 atau 78,8% ; dan nilai MAPE sebesar 3,366%. Model semiparametrik *B Spline* merupakan model kuat karena $R^2 > 0,67$ dengan kinerja model sangat bagus karena $MAPE < 10\%$. Graphical User Interface yang tersusun dapat mempermudah dalam melakukan pemodelan. Model regresi semiparametrik *B Spline* terbaik menggunakan orde 2 dengan 4 titik knot adalah sebagai berikut:

$$\hat{y}_i = 3.188,669 + 32,474(x_i) + 75,388S_{-1,2}(z_i) - 77,094S_{0,2}(z_i) + 1.182,115S_{1,2}(z_i) + 174,101S_{2,2}(z_i)$$

Model tersebut dapat digunakan untuk melakukan peramalan beberapa bulan ke depan, sehingga indeks harga saham gabungan dapat diketahui perubahan harganya berdasarkan pengaruh dari harga emas dan minyak mentah dunia.



Gambar 5. Scatterplot Peramalan 7 Bulan Ke Depan

Penelitian selanjutnya dapat dicoba pemilihan *bandwidth* optimal dengan optimasi lain seperti MSE, CV, UBR.

DAFTAR PUSTAKA

- [1] A. Basit, “Pengaruh Harga Emas dan Minyak Dunia terhadap Indeks Harga Saham Gabungan (IHSG) Periode 2016-2019,” *J. Manajemen Bisnis Islam.*, vol. 1, no. 2, pp. 95-110, 2020.

- [2] O. Blanchard, *Macroeconomics* 7th edn. Pearson Prentice Hall, 2017.
- [3] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric Regression*. Cambridge University Press, 2003.
- [4] S. Wulandary and D. I. Purnama, "Perbandingan Regresi Nonparametrik Kernel NWE dan B Splines Pada Pemodelan Rata-Rata Lama Sekolah dan Pengeluaran Perkapita Di Indonesia," *J. Journal of Probability and Statistics.*, vol. 1, no. 2, pp. 89-97, 2020.
- [5] H. Mariani, *Emas: Kandungan dan Penggunaan*. Elex Media Komputindo, 2010.
- [6] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis* 5th edn. John Wiley & Sons, 2013.
- [7] Suparti, R. Santoso, A. Prahutama, and A. R. Devi, *Regresi Nonparametrik*. Wade Group, 2018.
- [8] R. L. Eubank, *Nonparametric Regression and Spline Smoothing* 2nd edn. Marcel Dekker, 1999.
- [9] D. Eberly. "B Spline Interpolation on Lattices." <https://www.geometrictools.com> (accessed Feb. 12, 2023).
- [10] A. P. Sugiantari and I. N. Budiantara, "Analisis Faktor – Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline," *J. Sains dan Seni POMITS*, vol. 2, no. 1, pp. 2337-3520, 2013.
- [11] G. Wahba, *Spline Models for Observational Data*. SIAM, 1990.
- [12] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient of Determination R-squared is More Informative than SMAPE, MAE, MAPE, MSE, and RMSE in Regression Analysis Evaluation." *PeerJ Compt. Sci.*, 2021.
- [13] P. C. Chang, Y. W. Wang, and C. H. Liu, "The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting Expert Systems with Applications," *Expert Systems with Applications*, vol. 32, pp. 86-96, 2007.