Indonesian Journal of Applied Statistics

E-ISSN 2621–086X

Vol. 6, No. 1, pp. 22-32, May 2023
https://jurnal.uns.ac.id/ijas/article/view/79385
https://doi.org/10.13057/ijas.v6i1.79385

Analisis Data *Time Series* Menggunakan Model Kernel: Pemodelan Data Harga Saham MDKA

Suparti*, Rukun Santoso
Departemen Statistika, Universitas Diponegoro

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

Abstract. *Classic time series data analysis techniques, such as autoregressive, model stationary data in which the values of prior observations influence the current observations through a process known as linear regression. There are several requirements for error assumptions in autoregressive, including independence, normal distribution with a zero mean and constant variance. It is frequently discovered that these assumptions are challenging to verify when modelling real data. Kernel time series regression is an alternative model that does not require error assumptions. Non-stationary time series data can be effectively modelled using the kernel time series method. Time series data that isn't yet stationary is made stationary first, then the data is modified by forming the current stationary time series data as the response variable and the previous period data as the predictor variable. Next, regression kernel modelling is carried out while applying kernel weight function and determining the optimal bandwidth. For development of science, the optimal bandwidth can be achieved by minimizing the MSE, CV, GCV, or UBR values. It is possible to use* R^2 *or MAPE as the kernel time series regression model's goodness metric. A strong model is generated while modelling MDKA stock price data using kernel regression utilizing the Gaussian kernel function and optimal bandwidth selection using GCV since* R^2 *is 0.9828372 more than 0.67 and MAPE is 1.985681% under 10%.*

Keywords: *3 time series; kernel regression; GCV; MDKA stock price.*

1. PENDAHULUAN

Analisis regresi merupakan suatu teknik statistik untuk menentukan hubungan antara variabel prediktor dan variabel respon. Hubungan tersebut dinyatakan dalam suatu model matematika. Dalam pemodelan data ada 2 pendekatan yaitu pendekatan parametrik dan nonparametrik. Pendekatan parametrik dilakukan jika *scatter plot* datanya membentuk pola yang sudah diketahui bentuk fungsinya seperti garis lurus, kuadratik, dan sebagainya. Pendekatan nonparametrik dilakukan jika scatterplot datanya tidak membetuk pola yang sudah diketahui bentuk fungsinya, misalnya pola datanya menyebar secara acak, bergerombol di suatu titik tertentu atau mempunyai pola fluktuasi naik dan turun di beberapa bagian, dan sebagainya. Beberapa metode dalam regresi nonparametrik diantaranya model kernel, spline, polinomial lokal, deret Fourier, dan wavelet [1]. Setiap model mempunyai karakteristik pola data masing-masing, seperti jika pola data cenderung menyebar secara acak, maka data tersebut lebih cocok dimodelkan menggunakan pendekatan kernel dan jika pola data bergerombol lebih cocok mengginakan regresi polinomial lokal [2,3].

Submitted: 07 October 2023; ***Accepted:*** 28 November 2023

Pada pemodelan regresi parametrik ada asumsi yang mengikat yaitu *error* model harus memenuhi asumsi independensi, memiliki distribusi normal dengan *mean* nol dan varians konstan. Jika asumsi tersebut tidak dipenuhi maka pengambilan keputusan dalam uji signifikansi model akan menyesatkan [1]. Dalam regresi nonparametrik asumsi *error* model tersebut tidak diharuskan dan pengambilan keputusan kebaikan model dapat menggunakan ukuran kuantitatif seperti R^2 dan nilai *Mean Absolute Percentage Error* (MAPE) [4,5].

Berbeda dengan analisis regresi yang digunakan untuk memodelkan 2 variabel atau lebih, analisis *time series* dipergunakan dalam pemodelan data dengan satu variabel respon yang diamati dalam urutan waktu. Metode yang popular untuk menganalisis data *time series* adalah model ARIMA Box-Jenkins [6]. Dalam model ARIMA, data harus stasioner dan jika belum stasioner, data harus distasionerkan terlebih dahulu. Model ARIMA Box-Jenkins terdiri model *Autoregressive* (AR), *Moving* Average (MA), *Autoregressive Moving* Average (ARMA) dan *Autoregressive Integrated Moving* Average (ARIMA). Model ARIMA merupakan model parametrik dan *error* modelnya terikat dengan asumsi-asumsi independensi, berdistribusi normal dengan *mean* nol dan varians konstan. Jika asumsi tersebut dilanggar maka pengambilan keputusan akan menyesatkan. Dalam pemodelan data riil, asumsi *error* model ARIMA tersebut terkadang sulit untuk dipenuhi. Sebagai alternatifnya, analisis *time series* dapat dilakukan dengan model regresi nonparametrik dengan tetap memperhatikan stasioneritas data harus dipenuhi. Hardle [7] memberikan solusi bahwa dari data *time series* yang stasioner dapat dibawa ke dalam permasalahan regresi dengan mendefinisikan data sekarang sebagai variabel respon dan data satu periode sebelumnya sebagai variabel prediktor. Fitri, et. al. [8] telah membandingkan model data *time series* pada harga Jakarta Islamic Index (JII) menggunakan ARIMA dan regresi kernel. Semua model ARIMA yang terbentuk tidak memenuhi asumsi normalitas *error* dan model regresi kernel memenuhi kriteria model yang bagus karena R^2 lebih dari 0,67 dan MAPE kurang dari 10%. Kekurangan penelitian ini adalah data *time series*-nya dibiarkan tidak stasioner sehingga konsep dasar pada analisis data *time series* terabaikan.

Permasalahannya adalah ketika data *time series*-nya tidak stasioner, diperlukan prosedur pembentukan model regresi nonparametrik dan mencari model yang sesuai dengan karakteristik datanya. Tujuan penelitian ini adalah memodelkan data *time series* ke dalam model alternatif yang bebas asumsi model yaitu model regresi nonparametrik dengan tetap memperhatikan konsep penting stasioneritas pada data *time series*. Model nonparametrik yang dipilih adalah model kernel karena pada data stasioner akan cenderung berpola acak sehingga model kernel merupakan salah satu model yang sesuai dengan karakteristik data tersebut. Pemodelan data dipilih data harga penutupan saham harian PT. Merdeka Copper Gold Tbk. yang diberi kode MDKA. Data selama 5 tahun terakhir sebelum tahun 2022 mengalami fluktuasi cenderung naik dan akan memberikan keuntungan bagi investor dan setelah dilakukan penelitan pendahuluan dengan pemodelan ARIMA Box-Jenkins ada asumsi model yang tidak terpenuhi, sehingga dilakukan pemodelan alternatif menggunakan regresi *time series* kernel.

2. LANDASAN TEORI

2.1. Model *Autoregressive*

Model *Autoregressive* (AR) merupakan model *time series* klasik yang populer dan bagian dari Model ARIMA. Model *AR* orde $p \geq 1$ didefinisikan seperti pada Persamaan (1).

$$\mathrm{Z_t} = \rho_1 \mathrm{Z_{t\text{-}1}} + \cdots + \rho_p \mathrm{Z_{t\text{-}p}} + \varepsilon_t, \text{ dengan } \{\varepsilon_t\} \sim NID(0, \sigma^2). \tag{1}$$

Persamaan (1) merepresentasikan kondisi saat ini ($\mathrm{Z_t}$) melalui *p* nilai sebelumnya yaitu Z_{t-1}, Z_{t-2}, …, Z_{t-p} dalam bentuk regresi linier [9]. Secara khusus, pada model AR (1) menyatakan nilai pengamatan saat ini ($\mathrm{Z_t}$) dipengaruhi oleh satu periode nilai pengamatan sebelumnya (Z_{t-1}) dalam bentuk regresi linier.

2.2. Stasioneritas

Salah satu syarat yang penting dalam analisis data *time series* adalah sifat stasioneritas. Suatu data *time series* dikatakan stasioner apabila *mean* dan variansnya tetap sepanjang waktu [6]. Proses stasioner dapat diasumsikan seperti pada Persamaan (2) dan (3).

$$E(Z_t) \quad = \mu \tag{2}$$

$$Var\ (Z_t) = E(Z^t - \mu)^2 = \gamma_o \tag{3}$$

dengan μ untuk semua lag *k* adalah konstan. Uji akar unit *Dickey-Fuller* (DF) merupakan sakah satu metode untuk menguji stasioneritas data dalam *mean.* Uji akar unit mengaplikasikan model regresi yang dituliskan pada Persamaan (4).

$$\begin{aligned} \mathit{\Delta} Z_t &= (\rho - 1)Z_{t-1} + \varepsilon_t \\ \mathit{\Delta} Z_t &= \delta Z_{t-1} + \varepsilon_t \end{aligned} \tag{4}$$

Z_t dikatakan tidak stasioner (memiliki akar unit) apabila δ = 0 atau ρ = 1. Hipotesis untuk uji akar unit *Dickey-Fuller* (DF) adalah:

H_0: δ = 0 (Data tidak stasioner dalam *mean*)

H_1: δ < 0 (Data stasioner dalam *mean*)

Statistik uji untuk uji DF dinyatakan oleh Persamaan (5).

$$\tau_{\text{hitung}} = \frac{\hat{\delta} - \delta}{\mathrm{SE}(\hat{\delta})} \tag{5}$$

dengan $\hat{\delta}$: penduga dari δ, dan SE($\hat{\delta}$): *standard error* dari δ.

Kriteria uji: Ho ditolak jika nilai $\tau_{\text{hitung}} < \tau_{(\alpha;\text{n-1})}$ atau $p -$ value $< \alpha$.

Data *time series* yang tidak stasioner dalam *mean* dapat diatasi dengan melakukan diferensiasi (*differencing*). Proses *differencing* orde ke-*d* dapat dinyatakan oleh Persamaan (6).

$$\mathit{\Delta}^d Z_t = (1 - B)^d Z_t \tag{6}$$

dengan $\Delta^d Z_t$ adalah hasil *differencing* pengamatan ke–*t* orde ke-*d*, *d* adalah orde *differencing*, dan B adalah operator *backshift.*

Pengujian stasioneritas dalam varians dapat dilakukan dengan melihat nilai λ pada transformasi Box-Cox. Nilai λ yang tidak sama dengan satu menandakan bahwa data tidak stasioner dalam varians, dan data harus ditransformasi Box-Cox. Transformasi Box-Cox yang dinyatakan oleh Persamaan (7).

$$Z_t^* = \begin{cases} \frac{Z_t^\lambda - 1}{\lambda}, & \lambda \neq 1 \\ ln\, Z_t\ , & \lambda = 1 \end{cases} \tag{7}$$

2.3. Analisis Regresi Nonparametrik

Analisis regresi nonparametrik digunakan untuk menjelaskan hubungan antara variabel prediktor dengan variabel respon apabila bentuk fungsi atau pola dari data tersebut tidak diketahui. Berbeda dengan analisis regresi parametrik yang memiliki beberapa asumsi yang harus dipenuhi, analisis regresi nonparametrik tidak mengharuskan adanya asumsi pada residual model sehingga metode ini sangat cocok digunakan untuk menganalisis data yang sangat fluktuatif seperti harga saham yang berjalan dinamis terutama karena fleksibilitas dari metode ini. Model regresi nonparametrik secara umum dituliskan pada Persamaan (8).

$$Y_t = m(X_t) + \varepsilon_t \quad (8)$$

dengan Y_t merupakan pengamatan ke-t dari variabel dependen, fungsi m merupakan fungsi regresi yang polanya tidak diketahui bentuk kurvanya, ε_t merupakan pengamatan ke-*t* dari variabel independen, serta merupakan residual pengukuran random. Fungsi regresi diasumsikan mulus sehingga lebih fleksibel dalam menduga fungsi regresinya.

2.4. Regresi Kernel

Regresi kernel adalah suatu regresi terboboti dengan menggunakan bobot fungsi kernel. Fungsi kernel K adalah fungsi yang bernilai riil non negatif, simetris, terbatas dan integralnya sama dengan 1. Misalkan diberikan data $\{(X_t, Y_t)\}$, $t = 1,2...,n$ dan akan dilakukan pendugaan fungsi regresi $m(x)$ pada Persamaan (8). Penduga Nadaraya-Watson dari $m(x)$ pada Persamaan (8) didefinisikan pada Persamaan (9).

$$m(x) = E(Y|X = x) \quad (9)$$

Variabel Y dan X bersifat random, sehingga estimator m(x) dari Nadaraya-Watson adalah

$$\hat{m}(x) = \frac{\sum_{t=1}^{n} K\left(\frac{x-X_t}{h}\right) Y_t}{\sum_{t=1}^{n} K\left(\frac{x-X_t}{h}\right)} \quad (10)$$

dengan:

$$W_t(x) = \frac{K\left(\frac{x-X_t}{h}\right)}{\sum_{t=1}^{n} K\left(\frac{x-X_t}{h}\right)} \quad (11)$$

Simbol $\{W_t(x)\}$ pada Persamaan (11) menyatakan bobot-bobot yang bernilai positif dan menpunyai sifat $\sum_{t=1}^{n} W_t(x) = 1$. Persamaan (10) dapat ditulis seperti Persamaan (12).

$$\hat{m}(x) = \sum_{t=1}^{n} W_t(x) Y_t \quad (12)$$

Dengan mengganti indeks t dengan j diperoleh Persamaan (13).

$$\hat{m}(x) = \sum_{j=1}^{n} W_j(x) Y_j \quad (13)$$

Ketika nilai $x = X_t$, maka $\hat{m}(x)$ disebut $\hat{Y}_t$ dan diperoleh Persamaan (14).

$$\hat{Y}_t \quad = \sum_{j=1}^{n} W_j\,(X_t) Y_j \quad (14)$$

Dengan menotasikan $W_j(X_t) = H_{\mathrm{tj}}$ maka $\hat{Y}_t = \sum_{j=1}^{n} H_{\mathrm{tj}}\, Y_j$ dengan $H_{\mathrm{tj}} = \frac{K\left(\frac{X_t - X_j}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_t - X_j}{h}\right)}$ disebut matriks Hat.

Performa dari $\hat{Y}_t$ dipengaruhi oleh fungsi bobot kernel K yang dipilih dan *bandwidth h* yang ditentukan. Namun pengaruh pemilihan fungsi bobot kernelnya tidak bersifat signifikan, tetapi penentuan nilai *bandwidth* merupakan faktor yang paling dominan. Salah satu ukuran untuk memilih nilai *bandwidth* optimal adalah dengan meminimalkan nilai *Generalized Cross Validation* (GCV) yang disajikan pada Persamaan (15) [1].

$$\mathrm{GCV}(h) = \frac{n^{-1} \sum_{t=1}^{n} [Y_t - \hat{m}_h(X_t)]^2}{\left[1 - \frac{\sum_{t=1}^{n} H_{\mathrm{tt}}}{n}\right]^2} = \frac{n^2 \mathrm{MSE}}{\left[n - \sum_{t=1}^{n} H_{\mathrm{tt}}\right]^2} \quad (15)$$

2.5. Analisis Data *Time Series* Nonparametrik

Pada pemodelan regresi, pada dasarnya data $\{(X_t, Y_t)\}$, $t = 1, 2, ..., n$ merupakan pengamatan saling independen. Praktiknya banyak ditemukan bahwa asumsi independensi data tersebut tidak dipenuhi misalnya data *time series* dengan respon obyek sekarang tergantung dari respon sebelumnya. Hardle [7] menyusun konsep dasar pemodelan matematika yang mendasari pemodelan ini, yaitu:

1. Model (S): Suatu barisan stasioner $\{(X_t, Y_t)\}$, $t = 1, 2, ..., n$ (antar pengamatan boleh dependen stokastik) telah diamati dan akan diduga $m(x) = E(Y|X = x)$.
2. Model (T): Suatu data *time series* telah diamati dan akan diprediksi Z_{n+1} dengan $m(x) = E(Z_{n+1}|Z_n = x)$

Permasalahan model (T) dapat dibawa ke permasalahan dalam model (S) dengan mendefinisikan dalam bentuk *time series* $\{Z_t, t = 1, 2, ..., n\}$, nilai Z_{t-1} sebagai variabel prediktor X_t dan Z_t sebagai variabel respon Y_t,. Selanjutnya, masalah prediksi Z_{n+1} dari $\{Z_t\}$ dapat dipandang sebagai masalah pemodelan regresi untuk $\{(X_t, Y_t)\} = \{(Z_{t-1}, Z_t), t = 1, 2, ..., n\}$. Jadi, masalah prediksi $\{Z_t\}$ ekivalen dengan menduga $m(x) = E(Y|X = x)$untuk data *time series* dua dimensi $\{(X_t, Y_t), t = 1, 2, ..., n\}$. Selanjutnya $m(x)$ dapat diduga dengan regresi kernel Nadaraya-Watson.

2.6. Ukuran Kebaikan dan Kinerja Model

Pada analisis pemodelan data, sebelum data dibagi menjadi 2 bagian yaitu data *in sample* yang digunakan untuk membetuk model dan data *out* sample yang digunakan untuk evaluasi kinerja model *in sample* yang sudah terbentuk. Ukuran kebaikan model *in sample* dapat diukur dari koefisien determinasinya. Koefisien determinasi adalah ukuran yang menyatakan seberapa besar variabel prediktor dapat menjelaskan variabel responnya. Koefisien determinasi dirumuskan dalam Persamaan (16).

$$R^2 = \frac{\sum_{t=1}^{n}(\hat{Z}_t - \bar{Z})^2}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2 + \sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2} \tag{16}$$

Nilai koefisien determinasi yang semakin besar mendekati 1 maka model yang dihasilkan semakin baik untuk digunakan dan sebaliknya. Nilai R^2 yang sudah melebihi 0,67, modelnya sudah tergolong model yang kuat [10].

Ukuran kinerja model dapat diukur dari nilai MAPE *out sample*. Nilai MAPE yang lebih kecil dari 10% mengindikasikan kinerja modelnya sangat bagus [11]. Nilai MAPE dirumuskan pada Persamaan (17).

$$\text{MAPE} = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|\right) \text{x } 100\% \tag{17}$$

dengan Z_t adalah data aktual pada periode ke-t, $\hat{Z}_t$adalah hasil prediksi pada periode ke-t, dan n adalah banyak pengamatan.

3. METODE PENELITIAN

3.1. Data Penelitian dan Pengolahan.

Data penelitian yang digunakan adalah data sekunder yang diambil dari situs web Yahoo Finance [12], dengan variabel penelitian adalah harga penutupan saham harian PT. Merdeka Copper Gold Tbk. yang diberi kode MDKA dinotasikan dengan Z_t. Data tersebut merupakan data perdagangan aktif mulai dari tanggal 4 Januari 2021 hingga 22 September 2022 ada sebanyak 422 data. Pengolahan data dilakukan dengan software R Studio dengan antarmuka GUI.

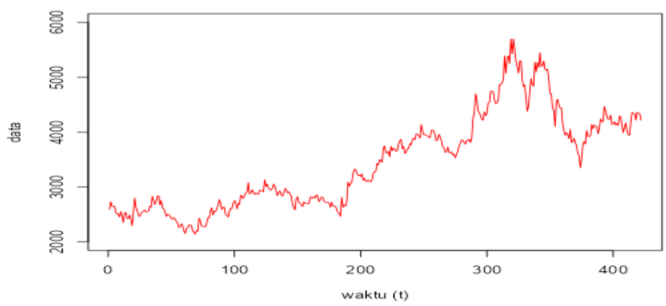
3.2. Prosedur Analisis Data.

Prosedur dalam pemodelan data *time series* kernel adalah sebagai berikut:

1. Data *time series* $\{Z_t$, t = 1, 2, ..., n$\}$ diuji stasioneritasnya baik dalam *mean* maupun varians. Pegujian pertama stasioneritas dalam varians, jika data belum stasioner dilakukan transformasi Box-Cox menjadi Z^*_t. Selajutnya data yang telah stasioner dalam varians diuji stasioneritasnya dalam *mean*. Data yang belum memenuhi stasioneritas dalam *mean* dilakukan differencing pertama (ΔZ^*_t). Data yang didifferencing akan berkurang 1 pengamatan dan datanya menjadi $\{\Delta Z^*_t$, t = 2, 3, …, n$\}$.
2. Data stationer dimodifikasi dengan nilai ΔZ^*_{t-1} sebagai X_t dan ΔZ^*_t sebagai Y_t. Selanjutnya masalah prediksi ΔZ^*_{n+1} dari $\{\Delta Z^*_t\}$ dapat dipandang sebagai masalah pemodelan regresi untuk $\{(X_t, Y_t)\} = \{(\Delta Z^*_{t-1}, \Delta Z^*_t)$, t = 3, 4, ..., n$\}$.
3. Data modifikasi dibagi menjadi 2 bagian yaitu data *in sample* dan *out sample* dengan proporsi 90%:10%. Data *in sampel* dilakukan pemodelan dengan regresi kernel dengan fungsi kernel Gaussian dan model terbaik diperoleh pada *bandwidth* optimal yang meminimumkan nilai GCV. Pencarian *bandwidth* optimal dilakukan dengan *trial* dan *error* pada selang interval tertentu.
4. *Bandwidth* optimal digunakan untuk menghitung dugaan data $\{\Delta Z^*_t)$, t = 3, 4, ..., n$\}$ dan hasil dugaan dikembalikan ke bentuk data semula seperti sebelum distasionerkan.
5. Model pada data *in sample* diukur kebaikan modelnya dengan menghiung koefisien determinasi dan model yang terbentuk dievaluasi kinerja modelnya menggunakan nilai MAPE data *out sample*.
6. Langkah terakhir adalah melakukan prediksi data ke depan.

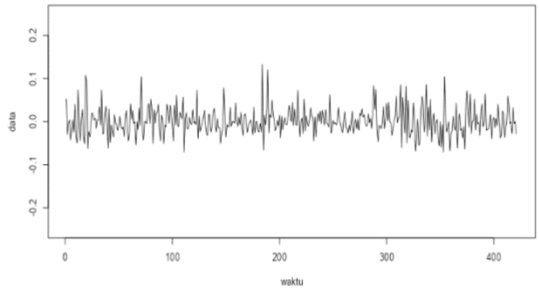
4. Hasil dan Pembahasan

Data harga penutupan pada perdagangan harian saham MDKA tanggal 4 Januari 2021 hingga 22 September 2022 sebanyak 422 data. Plot data *time series* harga saham tersebut disajikan dalam Gambar 1.



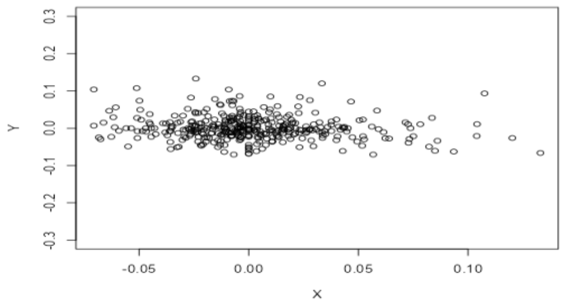
Gambar 1. Grafik Data Harga Penutuan Harian Saham MDKA

Gambar 1 secara visual terlihat data belum stationer dalam *mean* dan varians. Oleh karena itu data harus ditarnsformasi. Hasil dari transformasi Box-Cox diperoleh nilai λ sebesar -0,01 (tidak sama dengan 1) sehingga data tidak stasioner dalam varians dan dilakukan transformasi Box-Cox (Persamaan 6) dengan nilai λ sebesar (-0,01). Setelah dilakukan uji stasioneritas ulang diperoleh nilai λ pada transformasi Box-Cox sebesar 1, sehingga data setelah ditransformasi sudah stasioner dalam varians. Langkah berikutnya adalah pengujian stasioneritas dalam *mean* menggunakan uji akar unit *Dickey Fuller*. Berdasarkan hasil pengujian *Dickey Fuller* dengan α=5% dapat disimpulkan bahwa data tidak stasioner dalam *mean* karena p-value = 0,4571 > 5%, kemudian data dilakukan differencing 1 kali. Selanjutnya diuji *Dickey Fuller* kembali menghasilkan data yang sudah stasioner dalam *mean* karena p-value = 0,01 < 5%. Data yang dideferencing 1 kali berkurang 1 pengamatan dan datanya menjadi $\{\Delta Z^*_t$, t = 2, 3, ..., 422}. Plot data setelah distasionerkan dalam *mean* dan varians disajikan dalam Gambar 2. Data yang telah stasioner selanjutnya dimodifiksi menjadi untuk $\{(X_t, Y_t)\} = \{(\Delta Z^*_{t-1}, \Delta Z^*_t)$, t = 3, 4, ..., 422} dan dibagi menjadi data *in sample* dan *out sample* dengan perbandingan 90% dan 10%.



Gambar 2. Plot Data Stasioner Keseluruhan

Data modifikasi *in sampel* $\{(X_t, Y_t)$, t= 3, 4, …, 380} sebanyak 378 pengamatan dan data *out sample* $\{(X_t, Y_t)$, t= 381, 382, …, 422} sebanyak 42 pengamatan. Data modifikasi *in sample* disajikan pada Gambar 3.

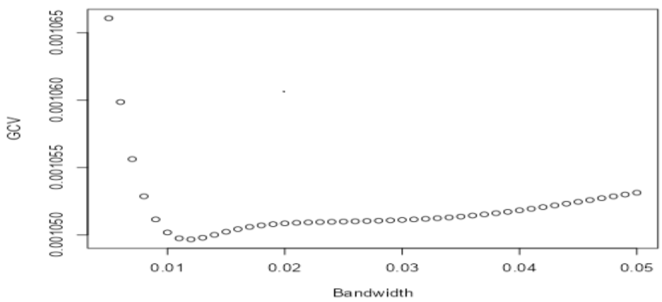


Gambar 3. *Scatter* Plot Data Modifikasi *In Sample*

Gambar 3 terlihat data modifikasi mempunyai pola yang cenderung acak. Data *in sample* pada Gambar 3, selanjutnya dimodelkan menggunakan regresi kernel dengan mengambil fungsi kernel Gaussian yang rumus pada Persamaan (18).

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{untuk} -\infty < x < \infty, \quad (18)$$

dengan fungsi kernel diskala $K_h(x) = \frac{K(\frac{x}{h})}{h}$.



Gambar 4. Plot Antara *Bandwidth* dan GCV

Proses pemodelan regresi kernel dengan menentukan nilai *bandwidth* h dengan *trial* dan *error* pada interval tertentu sehingga didapatkan nilai GCV terkecil. Setelah melalui proses pencarian h optimal, pada penelitian ini ditampilkan gambar fungsi GCV pada nilai 0,005 < h < 0,05 dengan selisih penambahan 0,001. Plot nilai h dan GCV disajikan pada Gambar 4 dan sepuluh nilai GCV terkecil ditampilkan pada Tabel 1.

Tabel 1. Plot Antara *Bandwidth* dan GCV

No	h	GCV
1.	0,012	0,001049671
2.	0,011	0,001049749
3.	0,013	0,001049795
4.	0,014	0,001050007
5.	0,010	0,001050182
6.	0,015	0,001050233
7.	0,016	0,001050433
8.	0,017	0,001050593
9.	0,018	0,001050712
10.	0,019	0,001050797

Pada Tabel 1 diperoleh nilai *bandwidth* optimal dicapai pada nilai h = 0,012. Gambar model terbaik dari data stasioner dengan nilai h = 0,012 disajikan pada Gambar 5. Model regresi kernel pada Gambar 5 disajikan pada Persamaan (19) dengan fungsi kernel K(x) pada Persamaan (18).

$$\hat{m}(x) = \frac{\sum_{t=3}^{380} K\left(\frac{x-X_t}{0,012}\right) Y_t}{\sum_{t=3}^{380} K\left(\frac{x-X_t}{0,012}\right)}, \text{ dengan } X_t = \Delta Z^*_{t-1},\ Y_t = \Delta Z^*_t \tag{19}$$

Prediksi pada nilai pengamatan keseluruhan $x=X_i$ i= 3, 4, …, 422 disajikan pada Persamaan (20)

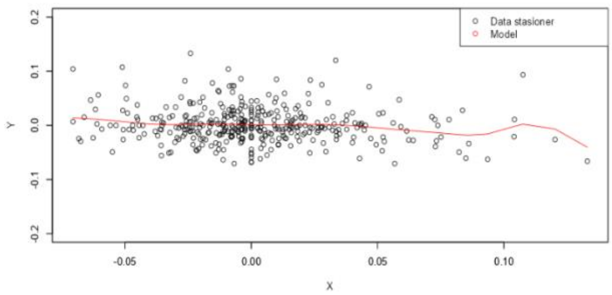
$$\hat{Y}_i = \hat{m}(X_i) = \frac{\sum_{t=3}^{380} K\left(\frac{X_i-X_t}{0,012}\right) Y_t}{\sum_{t=3}^{380} K\left(\frac{X_i-X_t}{0,012}\right)} \tag{20}$$

Dari $Y_t = \Delta Z^*_t = Z^*_t - Z^*_{t-1}$, maka $\hat{Y}_i = \hat{Z}_i - Z^*_{i-1}$ sehingga model Persamaan (20) menjadi Persamaan (21) dengan $X_t = Z^*_{t-1} - Z^*_{t-2}$, $Y_t = Z^*_t - Z^*_{t-1}$, $Zt^* = \frac{Z_t^{\lambda} - 1}{\lambda}$, t = 3,4,…,380 dan

$X_i = Z^*_{i-1} - Z^*_{i-2}$, $Zi^* = \frac{Zi^{\lambda} - 1}{\lambda}$, untuk i = 3, 4, 5, …, $\lambda = -0,01$.

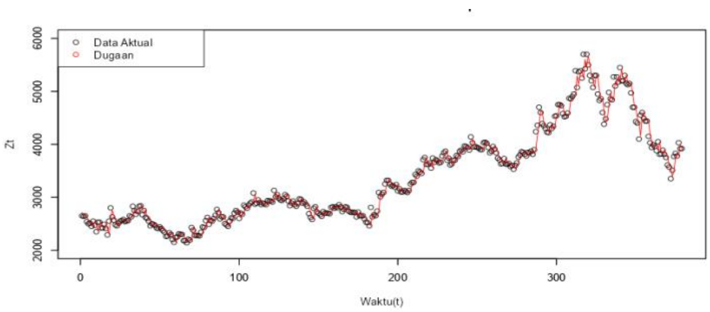
$$\hat{Z}_i^* = Z^*_{i-1} + \frac{\sum_{t=3}^{380} K\left(\frac{X_i-X_t}{0,012}\right) Y_t}{\sum_{t=3}^{380} K\left(\frac{X_i-X_t}{0,012}\right)} \tag{21}$$

Selanjutnya untuk mengembalikan prediksi $\hat{Z}_i^*$ ke data *time series* semula $\hat{Z}_i$ dengan melakukan transformasi invers dari Box-Cox yaitu $\hat{Z}_i = (\lambda \hat{Z}_i^* + 1)^{1/\lambda}$,dengan$\lambda = -0,01$, i = 3, 4, 5, …



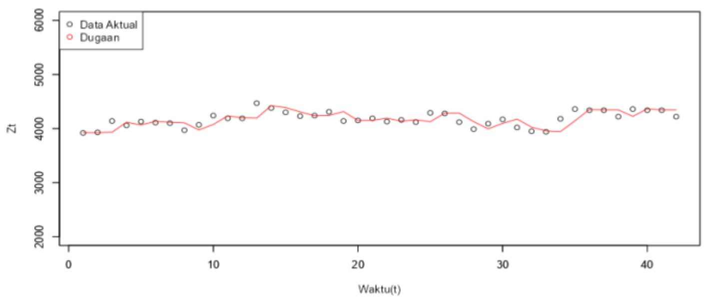
Gambar 5. Grafik Data Modifikasi *In Sample* dan Model Dugaan

Hasil penduga model pada data stasioner *in sample* (Gambar 5) yang telah dikembalikan ke data aslinya ditampilkan pada Gambar 6 yang secara visual terlihat data aktual dan prediksinya sangat dekat sekali. Model terbaik memiliki nilai R^2 sebesar 0,9828372 artinya bahwa variabel prediktor mampu menjelaskan variabel respon sebesar 98,28372 %. Model kernel yang terbentuk tergolong model yang kuat karena mempunyai R^2 lebih dari 0,67.



Gambar 6. Grafik Data *in sample* Asli dan Dugaannya

Model terbaik selanjutkan digunakan untuk prediksi data o*ut sample* yang hasilnya disajikan pada Gambar 7 dengan nilai MAPE *out sample* sebesar 1,985681 % yang menunjukkan bahwa kinerja model kernel sangat bagus karena nilai MAPE < 10%.



Gambar 7. Grafik Data *Out Sample* Asli dan Dugaannya

Dari model terbaik dapat digunakan untuk prediksi beberapa hari ke depan asalkan 2 periode sebelumnya diketahui nilainya. Misalnya untuk memprediksi harga saham harian penutupan MDKA pada minggu pertama bulan Juni 2023 (5 - 9 Juni 2023), maka data 2 periode sebelum tanggal 5 Juni harus sudah ada yaitu tanggal 30-31 Mei 2023 mempunyai harga penutupan 2960 dan 3000. Prediksi saham penutupan untuk 5 hari ke depan menggunakan model Persamaan (21) disajikan Tabel 2.

Tabel 2. Prediksi Harga Saham Tanggal 5-9 Juni 2023

Tanggal	Harga (Rp.)
05 Juni 2023	3105
06 Juni 2023	3108
07 Juni 2023	3112
08 Juni 2023	3115
09 Juni 2023	3119

5. KESIMPULAN

Pemodelan data *time series* selain menggunakan model klasik *Autoregressive* (AR) dapat juga menggunakan model regresi nonparametrik kernel. Dalam analisis AR, datanya harus memenuhi asumsi yang ketat agar penarikan kesimpulannya tidak menyesatkan. Dalam analisis regresi kernel, asumsi *error* model tidak diperlukan. Pengambilan keputusan untuk kebaikan model kernel berdasarkan nilai R^2 dan MAPE. Penentuan *bandwidth* optimal secara *trial* dan *error* pada interval tertentu, oleh karenanya pemrograman dengan *softwre* R Studio dibuat dalam tampilan GUI agar efisien waktu. Pada pemodelan harga harian saham penutupan MDKA menghasilkan model regresi nonparametrik kernel yang sangat bagus dengan ukuran R^2 lebih dari 0,67 dan MAPE *out sample* kurang dari 10%. Penelitian selanjutnya dapat dilakukan dengan model nonparametrik polinomial lokal karena data stasioner juga bersifat menggerombol di suatu titik-titik tertentu. Untuk pengembangan ilmu pengetahuan, pemilihan *bandwidth* optimal dapat dicoba dengan metode lainnya seperti *Mean Squared Error* (MSE), *Cross Validation* (CV), *Akaike Information Criterion* (AIC), *Unbiased Estimator* (UBR).

DAFTAR PUSTAKA

[1] Suparti, R. Santoso, A. Prahutama, dan A.R. Devi, Regresi Nonparametrik. Ponorogo: Wade Group. 2018.

[2] Putri, A.P., Santoso, R., dan Sugito, “Analisis Regresi Nonparametrik Kernel Menggunakan Metode Jackknife Sampel Terhapus-1 Dan Sampel Terhapus-2,” Jurnal Gaussian, vol. 6, no. 1, pp. 1-10, 2017.

[3] Suparti. dan A. Prahutama, “Pemodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal Pada Beban Listrik di Kota Semarang,” Media Statistika, vol. 9, no. 2, pp. 85–93, 2016.

[4] Suparti et. al., “Biresponses Kernel Nonparametric Regression: Inflation and Economic Growth,” International Journal of Criminology and Sociology, vol. 10, pp. 465-471. 2021

[5] 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.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis Forecasting and Control Fifth Edition. John Wiley & Sons. 2016.

[7] Hardle, W., Applied Nonparametric Regression. New York: Cambridge University Press. 1990.

[8] R. Fitri, S. Suparti, and P. Kartikasari, "Peramalan Indeks Jakarta Islamic Index (JII) dengan Pendekatan Regresi Parametrik Linier Sederhana dan Regresi Nonparametrik Kernel Dilengkapi GUI R-Shiny," *Jurnal Gaussian*, vol. 12, no. 2, pp. 221-230. 2023. https://doi.org/10.14710/j.gauss.12.2.221-230

[9] J. Fan and Q. Yao, Nonlinear Time Series : Nonparametric and Parametric Methods. New York: Springer. 2003.

[10] W.W. Chin, The Partial Least Square Aproach to Structural Equation Modeling. London: Lawrence Erlbaum Associates. 1998.

[11] 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 Application. vol. 32, pp. 88-89. 2007.

[12] Yahoo Finance: PT Merdeka Copper Gold Tbk (MDKA.JK) Stock Historical Prices & Data: https://finance.yahoo.com/quote/MDKA.JK/history?p=MDKA.JK. 2023.