

MODEL ARIMA SEMIPARAMETRIK

Rosi Pratiwi, Yundari, Helmi

INTISARI

Pada pemodelan stokastik asumsi error diperlukan untuk memvalidasi suatu model. Error yang baik adalah error yang kecil, berdistribusi normal dan memenuhi sifat keacakan. Metode Autoregressive Integrated Moving Average (ARIMA) merupakan salah satu model stokastik yang memperhatikan asumsi error tersebut. Pada penelitian ini, error model dimodifikasi agar diperoleh error yang lebih kecil namun tetap memenuhi asumsi error. Model yang diusulkan adalah model ARIMA Semiparametrik. Model ARIMA semiparametrik merupakan teknik pemodelan yang menggabungkan model ARIMA parametrik (konvensional) dengan model nonparametrik. Tujuan dari penelitian ini adalah untuk membandingkan hasil pemodelan ARIMA semiparametrik dengan model ARIMA parametrik pada data Indeks Harga Saham Gabungan (IHSG). Langkah-langkahnya diawali dengan metode ARIMA secara parametrik, selanjutnya dilakukan pemodelan nonparametrik pada error yang dihasilkan oleh metode ARIMA parametrik. Hasil penelitian menunjukkan bahwa analisis data deret waktu menggunakan model ARIMA semiparametrik ini menghasilkan model estimasi yang lebih baik dari model ARIMA parametrik. Hal ini dapat dilihat dari nilai RMSE yang diperoleh yaitu senilai 59,224.

Kata Kunci : estimasi, deret waktu, regresi kernel, ARIMA

PENDAHULUAN

Model *Autoregressive Integrated Moving Average* (ARIMA) semiparametrik merupakan teknik pemodelan yang menggabungkan model ARIMA secara parametrik dan nonparametrik. Beberapa peneliti yang mengkaji model semiparametrik diantaranya adalah [1] Model Regresi Semiparametrik Deret Fourier, [2] Regresi Semiparametrik Berdasarkan Deret Fourier untuk Data Longitudinal dengan Optimisasi *Weighted Least Square* (WLS), [3] Analisis Runtun Waktu Menggunakan Regresi Semiparametrik pada Produksi Minyak Kelapa Sawit. Merujuk pada penelitian [4] Model Semiparametrik untuk Data Panel Heterogen dengan *Fixed Effect*, maka digunakan pendekatan ARIMA semiparametrik yang dikenakan pada *error* model ARIMA parametrik. Umumnya, asumsi *error* pada pemodelan ARIMA adalah memenuhi asumsi kenormalan, keacakan (tidak berkorelasi) dan kehomogenan. Selain memenuhi ketiga asumsi tersebut, hasil *error* yang kecil juga sangat diperhatikan. Salah satu cara untuk menghasilkan *error* yang kecil tersebut adalah dengan memodifikasi *error* model yang telah diperoleh dengan suatu analisis pemulusan *error*. Pada penelitian ini dilakukan pemulusan *error* menggunakan metode nonparametrik kernel. Kernel yang digunakan adalah kernel *Gaussian* karena fungsi kernel *Gaussian* berlaku untuk semua domain pada semua bilangan real.

Beberapa penelitian tentang analisis semiparametrik khususnya pada model deret waktu banyak diterapkan pada data perekonomian/harga saham. Indeks Harga Saham Gabungan (IHSG) adalah salah satu acuan sering dimanfaatkan sebagai penanda arah pasar saham di Indonesia apakah sedang mengalami keuntungan atau kerugian karna dalam perhitungannya menggunakan hampir seluruh saham yang ada di pasar saham. Pergerakan angka IHSG dapat mencerminkan kenaikan dan penurunan harga-harga saham yang ada walau pada beberapa waktu terdapat saham yang kenaikan dan penurunannya melebihi bahkan berlawanan dengan nilai IHSG saat itu, hal ini pastinya sangat membantu bagi para *investor*.

Salah satu cara agar para *investor* mendapatkan keuntungan dari perdagangan saham secara maksimal adalah dengan melakukan prediksi yang akurat terhadap pergerakan harga saham khususnya pada IHSG di dalam pasar saham. Terdapat banyak metode yang digunakan dalam memprediksikan IHSG, salah satunya yaitu dengan menggunakan model ARIMA semiparametrik yang dibahas lebih lanjut dalam penelitian ini.

Model Autoregressive Integrated Moving Average (ARIMA)

Model *Autoregressive* (AR) merupakan bentuk regresi yang menghubungkan nilai pengamatan suatu waktu dengan nilai-nilai pengamatan sebelumnya pada selang waktu tertentu. Model AR(p) dapat dinyatakan sebagai berikut [5]:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

dengan Y_t adalah data deret waktu stasioner pada waktu t ; $\phi_1, \phi_2, \dots, \phi_p$ merupakan parameter model AR yang berordo $1, 2, 3, \dots, p$, dan e_t adalah nilai *error* pada waktu ke- t .

Selanjutnya model *Moving Average* (MA) yang merupakan bentuk regresi yang digunakan untuk memperkirakan nilai pengamatan waktu tertentu dengan menggunakan nilai *error*-nya. Model MA(q) dapat dinyatakan sebagai berikut [5]

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

dengan $\theta_1, \theta_2, \dots, \theta_q$ adalah koefisien model MA dengan ordo $1, 2, 3, \dots, q$, dan e_t adalah nilai *error* pada waktu ke- t .

Model ARMA merupakan gabungan dari model AR dan MA. Model ini dinyatakan dengan notasi ARMA(p, q). Bentuk model ARMA(p, q) yaitu sebagai berikut [5]:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

dengan $\phi_1, \phi_2, \dots, \phi_p$ merupakan parameter model AR yang berordo $1, 2, 3, \dots, p$, sedangkan $\theta_1, \theta_2, \dots, \theta_q$ adalah parameter model MA dengan ordo $1, 2, 3, \dots, q$, dan e_t adalah nilai *error* pada waktu ke- t .

Model ARMA dilakukan proses *differencing* untuk menghilangkan ketidastasionerannya maka model itu akan menjadi model ARIMA. Bentuk umum model ARIMA adalah sebagai berikut [5]

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

dengan $W_t = Y_t - Y_{t-1}$

Adapun tahap-tahap dalam proses permodelan dengan metode ARIMA adalah identifikasi model, estimasi, dan validasi. Tahapan identifikasi model merupakan tahap untuk menentukan beberapa model yang mungkin untuk pemodelan. Penentuan orde dari model ARIMA yaitu p, d, q diperoleh dari plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Sedangkan banyaknya proses *differencing* digunakan untuk menentukan nilai d .

Penentuan nilai lag residual (q) dan nilai lag independen (p) yang digunakan dalam model, yaitu dengan melihat plot fungsi ACF dan plot fungsi PACF pada Tabel 1:

Tabel 1. Kriteria ACF dan PACF

Proses	ACF	PACF
AR (p)	Turun cepat secara eksponensial atau sinusoidal	Terputus setelah lag p
MA (q)	Terputus setelah lag q	Turun cepat secara eksponensial atau sinusoidal
ARMA (p, q)	Turun cepat Secara eksponensial atau sinusoidal	Turun cepat secara eksponensial atau sinusoidal

Setelah menentukan nilai dari p, d , dan q pada tahap identifikasi, maka selanjutnya dilakukan estimasi pada parameter-parameter model, salah satu metode yang dapat dipakai untuk mengestimasi nilai-nilai parameter yaitu *Maksimum Likelihood Estimation* (MLE). MLE diperoleh dengan menurunkan fungsi *Likelihood* (L) terhadap parameternya dan dinyatakan dengan nilai sama dengan nol sebagai berikut [7].

$$\frac{1}{f(x_1, \theta)} \frac{\partial f(x_1, \theta)}{\partial \theta} + \dots + \frac{1}{f(x_n, \theta)} \frac{\partial f(x_n, \theta)}{\partial \theta} = 0 \quad (1)$$

Berdasarkan Persamaan (1), θ dalam bentuk X_k merupakan estimator *maksimum likelihood*. Adapun metode MLE ini dapat dijelaskan dalam definisi berikut:

Definisi 1 [7] Misalnya X_1, X_2, \dots, X_n adalah peubah acak dengan fungsi distribusi $F(x_1, x_2, \dots, x_n | \theta)$ dengan $\theta \in \Theta$ merupakan estimator yang tidak diketahui nilainya. Fungsi *Likelihoodnya* adalah

$$L(\theta) = \begin{cases} f(x_1, x_2, \dots, x_n | \theta) & \text{jika } F \text{ mempunyai fungsi padat } f \\ p(x_1, x_2, \dots, x_n | \theta) & \text{jika } F \text{ mempunyai fungsi padat } p \end{cases}$$

Setiap $\hat{\theta} = \hat{\theta}_n(X_1, X_2, \dots, X_n) \in \Theta$ sehingga, $L(\hat{\theta}) = \sup\{L(\theta) : \theta \in \Theta\}$ disebut *maximum likelihood estimation*.

Tahap terakhir dilakukan uji signifikansi dan uji residual untuk meyakinkan kembali model yang didapatkan sudah benar-benar tepat atau belum. Pada tahap uji diagnostik dilakukan untuk melihat sifat keacakan *error* melalui plot ACF dan PACF *error*. Selanjutnya untuk asumsi kenormalan dilihat melalui histogram *errornya*

Estimasi Kernel Nadaraya-Watson

Fungsi kernel merupakan fungsi pemulus yang digunakan sebagai estimator untuk mengetahui distribusi suatu data. Pada pendekatan kernel terdapat sebuah parameter yang menentukan tingkat kemulusan kurva. Parameter tersebut dinamakan *bandwith* (h). Semakin kecil nilai h menyebabkan bentuk kurva semakin kasar dan sebaliknya, semakin besar nilai h menyebabkan kurva semakin mulus.

Penduga kernel k dengan parameter pemulus h dirumuskan sebagai berikut:

$$h(x) = \frac{1}{h} k\left(\frac{x}{h}\right) \text{ untuk } -\infty < x < \infty \text{ dan } h > 0$$

Sifat-sifat dari fungsi kernel yaitu sebagai berikut:

1. $k(x) \geq 0$
2. $\int_{-\infty}^{\infty} k(x) dx = 1$
3. $k(-x) = k(x)$
4. $\int_{-\infty}^{\infty} x k(x) dx = 0$
5. $\int_{-\infty}^{\infty} x^2 k(x) dx < \infty$

Pemilihan *bandwith* pada regresi kernel jauh lebih penting dibandingkan dengan pemilihan fungsi kernel. Hal ini disebabkan penggunaan fungsi kernel yang berbeda dengan nilai *bandwith* optimum menghasilkan kurva regresi yang hampir sama. Pemilihan *bandwith* yang terlalu kecil dapat menyebabkan fungsi estimasi menjadi sangat kasar sehingga hubungan variansinya tinggi dan memiliki potensi bias yang rendah, sebaliknya jika *bandwith* yang dipilih terlalu besar dapat menyebabkan fungsi yang diestimasi menjadi sangat mulus sehingga hubungan variansinya rendah dan memiliki potensi bias yang besar [8].

Salah satu metode yang digunakan untuk pemilihan *bandwith* optimal yaitu dengan metode *Cross Validation* (CV) yang dirumuskan sebagai berikut :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}(X_{\neq i}))^2$$

Beberapa jenis fungsi kernel yaitu:

1. Kernel *Uniform*

$$k(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

2. Kernel *Triangle*

$$k(x) = \begin{cases} (1 - |x|), & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

3. Kernel *Epanechnikov*

$$k(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

4. Kernel *Kuartik*

$$k(x) = \begin{cases} \frac{15}{16}(1 - x^2)^3, & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

5. Kernel *Triweigh*

$$k(x) = \begin{cases} \frac{35}{32}(1 - x^2)^3, & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

6. Kernel *Gaussian*

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty$$

7. Kernel *Cosinus*

$$k(x) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi x}{2}\right), & |x| \leq 1 \\ 0, & \text{lainnya} \end{cases}$$

Jika terdapat sebuah data pengamatan $x_1, x_2, x_3, \dots, x_n$, dari suatu distribusi dengan densitas yang tidak diketahui, maka estimator densitas kernel K dengan h merupakan *bandwith* dapat dirumuskan sebagai berikut:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right)$$

pemilihan *bandwith* yang terlalu kecil dalam regresi kernel dapat menyebabkan fungsi estimasi menjadi sangat kasar.

Terdapat sebuah fungsi regresi $m(x)$:

$$Y_i = m(X_i) + e_i \quad (2)$$

dengan $m(X_i)$ merupakan fungsi regresi yang tidak diketahui dan e_i adalah *error* dan $m(X_i) = E(Y|X = x)$ dengan syarat $X = x$ diketahui. Variabel dependen dan independen diasumsikan sebagai variabel random, sehingga $m(X_i)$ dapat dijabarkan menjadi

$$m(X_i) = \int_{-\infty}^{\infty} Y_i f(Y_i|X_i) dy \quad (3)$$

$$= \int_{-\infty}^{\infty} Y_i \frac{f(X_i, Y_i)}{f_X(X_i)} dy \quad (4)$$

Fungsi $f(X_i, Y_i)$ dan $f_X(X_i)$ adalah fungsi estimator densitas kernel yang persamaannya dapat ditulis sebagai berikut:

$$\hat{f}(X_i, Y_i) = \frac{1}{nh_x h_y} \sum_{j=1}^n k_x\left(\frac{X_i - X_j}{h_x}\right) k_y\left(\frac{Y_i - Y_j}{h_y}\right) \quad (5)$$

dengan K_x dan K_y adalah fungsi kernel, sedangkan h_x dan h_y adalah *bandwith*. Persamaan (4) dan (5) disubstitusikan ke dalam Persamaan (3) maka diperoleh $\hat{m}(X_i)$ sebagai berikut

$$\hat{m}(X_i) = \int_{-\infty}^{\infty} Y_i \frac{\hat{f}(X_i, Y_i)}{\hat{f}_X(X_i)} dy$$

$$\hat{m}(X_i) = \int_{-\infty}^{\infty} Y_i \frac{\frac{1}{nh_X h_Y} \sum_{j=1}^n k_X\left(\frac{X_i - X_j}{h_X}\right) k_Y\left(\frac{Y_i - Y_j}{h_Y}\right)}{\frac{1}{nh_X} \sum_{j=1}^n k_X\left(\frac{X_i - X_j}{h_X}\right)} dy \quad (6)$$

dengan $\eta = \frac{Y_i - Y_j}{h_Y}$ maka diperoleh persamaan sebagai berikut :

$$\frac{1}{h_Y} \int_{-\infty}^{\infty} Y_i k_Y\left(\frac{Y_i - Y_j}{h_Y}\right) dy = \int_{-\infty}^{\infty} (h_Y \eta + Y_j) k_Y(\eta) d\eta = Y_j \quad (7)$$

estimator *Nadaraya-Watson* (N-W) diperoleh dengan menstutitusikan $\sum_{j=1}^n k\left(\frac{X_i - X_j}{h}\right)$ pada Persamaan (7) ke dalam Persamaan (6) dan mengasumsikan $h_X = h$ sebagai berikut:

$$\hat{m}(X_i) = \frac{\sum_{j=1}^n k\left(\frac{X_i - X_j}{h}\right) Y_j}{\sum_{j=1}^n k\left(\frac{X_i - X_j}{h}\right)} \quad (8)$$

Bentuk umum fungsi kernel *Gaussian* dapat ditulis sebagai berikut:

$$k(X) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{X_i - X_j}{h}\right)^2\right) \quad (9)$$

dengan mensubtitusikan $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{X_i - X_j}{h}\right)^2\right)$ pada Persamaan (9) sebagai $k(X)$ ke dalam Persamaan (8) maka didapatkan Persamaan (10) sebagai berikut [9]:

$$\hat{m}(X_i) = \frac{\sum_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(\frac{(e_i - e_j)^2}{h}\right) e_{i+1}}{\sum_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(\frac{(e_i - e_j)^2}{h}\right)} \quad (10)$$

ARIMA Semiparametrik

ARIMA semiparametrik adalah model pendekatan yang menggabungkan antara pendekatan parametrik dengan pendekatan nonparametrik. Metode semiparametrik dalam penelitian ini diawali dengan metode ARIMA yang dikenakan pada data *close price* IHSG kemudian dilanjutkan dengan metode nonparametrik kernel *Gaussian* yang dikenakan pada *error* model yang didapat dari model ARIMA sebelumnya.

Langkah pertama dari metode ARIMA semiparametrik yaitu mengkonstruksi pasangan data (t, e_t) , dengan e_t adalah *error* model ARIMA pada waktu t . Selanjutnya pasangan data (t, e_t) dihaluskan menggunakan kernel *Gaussian* sebagai berikut:

$$k(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

Setelah pensubstitusian $x_i = e_t$, diperoleh e_t^* . Selanjutnya di e_t^* ditambahkan ke data asli y_t , sehingga diperoleh \hat{y}_t^* yang merupakan hasil estimasi data ARIMA semiparametrik yang persamaannya ditulis sebagai berikut:

$$Y_t = \phi Y_{t-1} + m(e_t),$$

suku pertama merupakan model ARIMA yang parameternya dicari menggunakan MLE. Sedangkan fungsi $m(x_i)$ merupakan *error* dari model ARIMA yang diestimasi menggunakan kernel. Pada penelitian ini fungsi kernel dikenakan pada *error* model ARIMA sehingga diperoleh model ARIMA semiparametrik sebagai berikut :

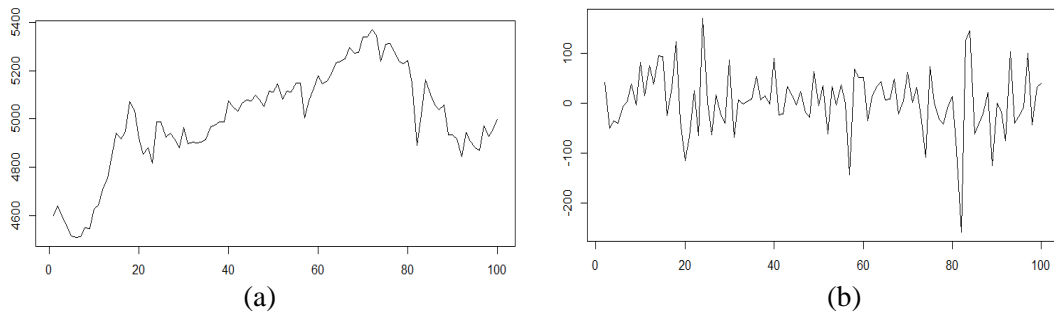
$$\widehat{W}_t = \sum_k^p \hat{\phi} \widehat{w}_{t-1} + \sum_k^p \hat{\phi} \hat{e}_{t-1} + \hat{m}(e_i) \quad (11)$$

dengan $W_t = Y_t - Y_{t-1}$

STUDI KASUS

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data harian Indeks Harga Saham Gabungan (IHSG) dari tanggal 8 Mei 2020 sampai dengan 7 Oktober 2020 sebanyak 100

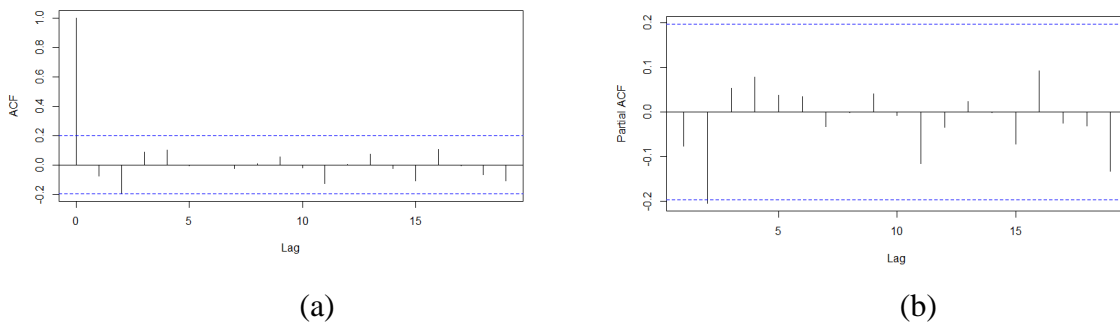
observasi. Plot dari pergerakan data harian *close price* IHSG serta hasil *differencing* data IHSG disajikan pada Gambar 1 berikut:



Gambar 1. (a) Data *Close Price* IHSG (b) Hasil *Differencing* Data *Close Price* IHSG

Berdasarkan Gambar 1.(a) penyebaran data pada plot data yang disajikan menunjukkan bahwa secara visual data stasioner dalam variansi namun tidak stasioner dalam rata-rata. Proses *differencing* pada data dilakukan untuk membuat data tersebut menjadi stasioner dalam rata-ratanya disajikan pada Gambar 1.(b). Gambar 1.(b) menunjukkan bahwa data bergerak stabil dan konvergen di sekitar nilai rata-ratanya sehingga dapat disimpulkan bahwa data tersebut sudah stasioner pada rata-ratanya setelah dilakukan proses *differencing* satu kali.

Tahapan selanjutnya adalah identifikasi model. Pada tahap ini untuk mengidentifikasi model-model yang cocok untuk estimasi adalah dengan plot ACF dan PACF dari *correlogram*. Plot dari ACF dan PACF setelah dilakukan *differencing* satu kali dan disajikan pada Gambar 2 berikut:



Gambar 2. (a) Plot ACF (b) Plot PACF

Berdasarkan Gambar 2 dapat dilihat bahwa lag ke-2 berpotongan dengan garis signifikansi pada plot ACF, sedangkan pada plot PACF garis signifikansi pada plot juga berpotongan dengan lag ke-2 sehingga dapat ditentukan bahwa ada dua kombinasi model ARIMA yang mungkin bagi data *close price* IHSG, yaitu model ARIMA (2,1,2) dan model ARIMA (2,1,0). Hasil estimasi parameter-parameter model ARIMA (2,1,2) dan ARIMA (2,1,0) dengan bantuan program *R Studio* disajikan pada Tabel 2 berikut:

Tabel 2. Estimasi Parameter ARIMA (2,1,0) dan (2,1,2)

Parameter	ARIMA (2,1,0)	ARIMA (2,1,2)
ϕ_1	-0,092	-0,118
ϕ_2	-0,205	-0,433

θ_1	-	-0,025
θ_2	-	-0,236
Konstanta	4,058	4,058

Berdasarkan hasil estimasi parameter ARIMA (2,1,0) pada Tabel 2 maka didapat persamaan untuk model ARIMA (2,1,0) sebagai berikut :

$$\begin{aligned}
 W_t &= 4,059 - 0,092W_{t-1} - 0,205W_{t-2} \\
 Y_t - Y_{t-1} &= 4,059 - 0,092(Y_{t-1} - Y_{t-2}) - 0,205(Y_{t-2} - Y_{t-3}) \\
 Y_t &= 4,059 + Y_{t-1} - 0,092Y_{t-1} + 0,092Y_{t-2} - 0,205Y_{t-2} - 0,205Y_{t-3} \\
 Y_t &= 4,059 + 0,008Y_{t-1} - 0,113Y_{t-2} - 0,205Y_{t-3}
 \end{aligned}
 \tag{12}$$

sedangkan untuk persamaan model ARIMA (2,1,2) dapat dituliskan sebagai berikut :

$$\begin{aligned}
 W_t &= 4,058 - 0,118W_{t-1} - 0,433W_{t-2} - 0,025e_{t-1} - 0,236e_{t-2} \\
 Y_t - Y_{t-1} &= 4,058 - 0,118(Y_{t-1} - Y_{t-2}) - 0,433(Y_{t-2} - Y_{t-3}) - 0,025e_{t-1} - 0,236e_{t-2} \\
 Y_t &= 4,058 + Y_{t-1} - 0,118Y_{t-1} + 0,118Y_{t-2} - 0,433Y_{t-2} - 0,433Y_{t-3} - 0,025e_{t-1} \\
 &\quad - 0,236e_{t-2} \\
 Y_t &= 4,058 + 0,882Y_{t-1} - 0,215Y_{t-2} - 0,433Y_{t-3} - 0,025e_{t-1} - 0,236e_{t-2}
 \end{aligned}
 \tag{13}$$

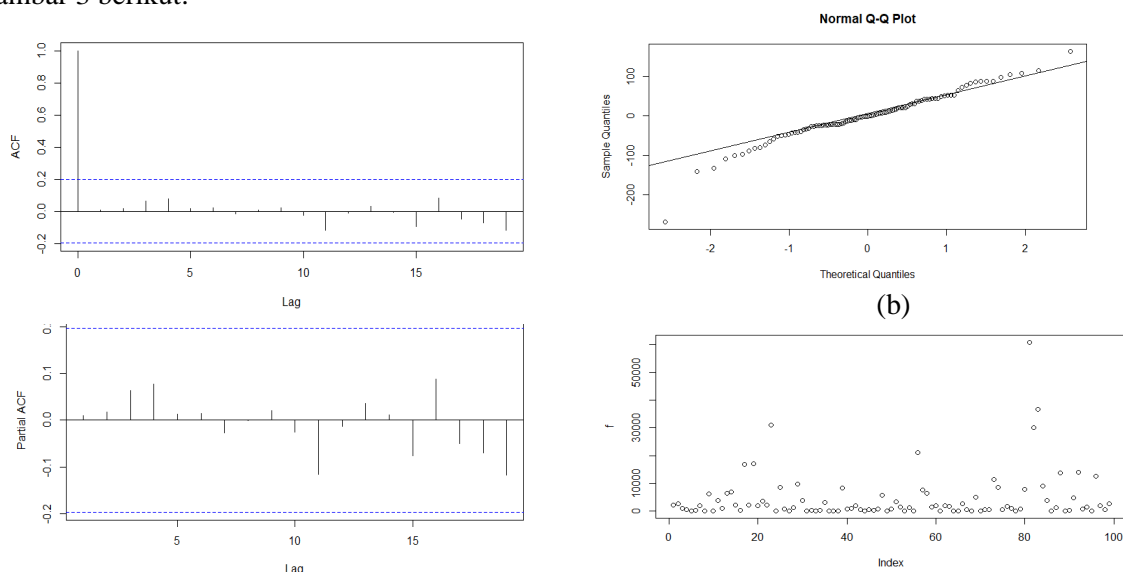
Model yang paling baik adalah model yang memiliki nilai RMSE terkecil. Nilai RMSE dari model ARIMA (2,1,0) dan ARIMA (2,1,2) disajikan pada Tabel 3 berikut:

Tabel 3. Nilai RMSE pada Model ARIMA

Model	RMSE
ARIMA (2,1,0)	61,629
ARIMA (2,1,2)	62,119

Berdasarkan Tabel 3 dapat dilihat bahwa model ARIMA (2,1,0) memiliki nilai RMSE yang lebih kecil dibandingkan dengan model ARIMA (2,1,2) sehingga dapat disimpulkan bahwa model ARIMA (2,1,0) merupakan model yang lebih baik dari model ARIMA (2,1,2)

Tahapan terakhir pada ARIMA parametrik adalah uji diagnostik. Pada tahap ini diagnostik dilakukan pada model ARIMA (2,1,0) dengan melihat sifat keacakan *error* melalui plot ACF dan PACF pada *error*. Selanjutnya untuk asumsi kenormalan pada model dapat dilihat melalui Q-Q plot *error*-nya. *Correlogram* ACF, PACF dan Q-Q plot dari *error* model ARIMA (2,1,0) dapat dilihat pada Gambar 3 berikut:

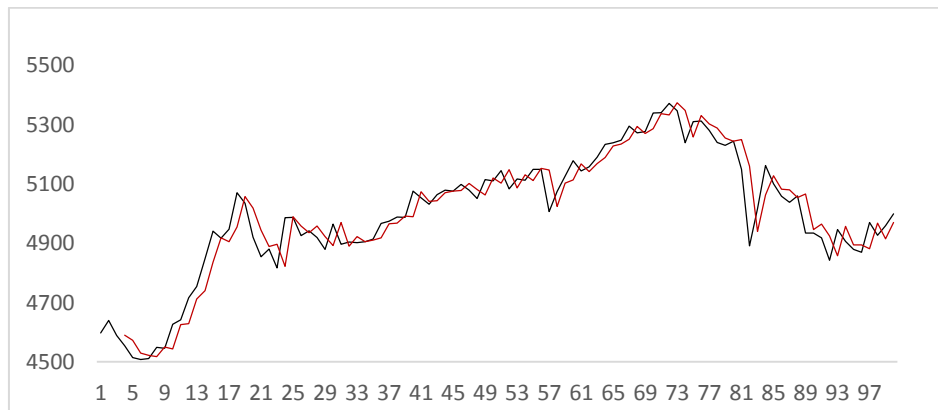


(a) (c)

Gambar 3. (a) Plot ACF & PACF *Error* ARIMA (2,1,0) (b) Q-Q Plot *Error* ARIMA (2,1,0) (c) Plot Kuadrat *Error* ARIMA (2,1,0)

Berdasarkan Gambar 3.(a) dapat dilihat bahwa tidak terdapat satu lag pun yang keluar batas pada plot ACF dan PACF *error* sehingga dapat dikatakan bahwa *error* model bersifat *white noise*. Pada Gambar 3.(b) dapat dilihat bahwa sebaran data *error* pada plot sebagian besar mendekati garis diagonal menandakan bahwa *error* pada model berdistribusi normal. Pada Gambar 3.(c) plot kuadrat *error* ARIMA tidak memiliki pola yang sistematis seperti pola menyempit, melebar, melengkung dan sebagainya. Karena plot kuadrat dari *error* model ARIMA memiliki pola yang acak dan tidak memiliki pola yang sistematis, maka dapat disimpulkan bahwa model yang didapat merupakan model yang baik untuk digunakan dalam prediksi IHSG.

Model yang dipilih adalah model pada Persamaan (12) karena telah memenuhi kriteria-kriteria model yang baik, maka selanjutnya model yang terpilih digunakan dalam pengestimasi nilai estimasi *close price* Indeks Harga Saham Gabungan (IHSG). Plot dari data hasil estimasi menggunakan model ARIMA (2,1,0) disajikan pada Gambar 4 berikut:



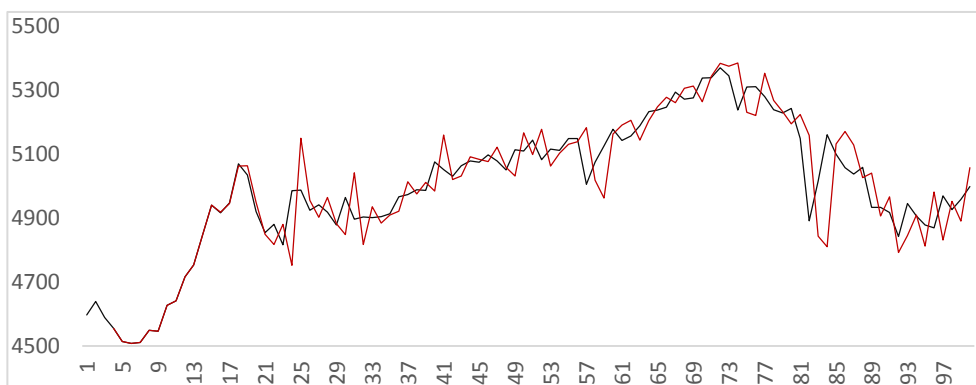
Gambar 4. Plot Data Hasil Estimasi ARIMA

Berdasarkan Gambar 4 garis berwarna merah menunjukkan data estimasi menggunakan ARIMA (2,1,0) dan garis berwarna hitam merupakan data asli nilai *close price* IHSG. Hasil estimasi yang diperoleh ini merupakan hasil dari pemodelan ARIMA parametrik. Selanjutnya dengan *error* yang diperoleh dilanjutkan pada tahapan nonparametrik.

Pada tahapan nonparametrik, *error* model yang telah didapat dari tahap sebelumnya dimodifikasi dengan menggunakan pendekatan kernel *Gaussian* pada *error* yang dihasilkan oleh model ARIMA (2,1,0). Langkah pertama yaitu mengkonstruksi pasangan data (t, e_t) , dengan e_t adalah *error* model ARIMA (2,1,0) pada waktu t . Selanjutnya pasangan data (t, e_t) dihaluskan menggunakan kernel *Gaussian* sebagai berikut:

$$k(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$$

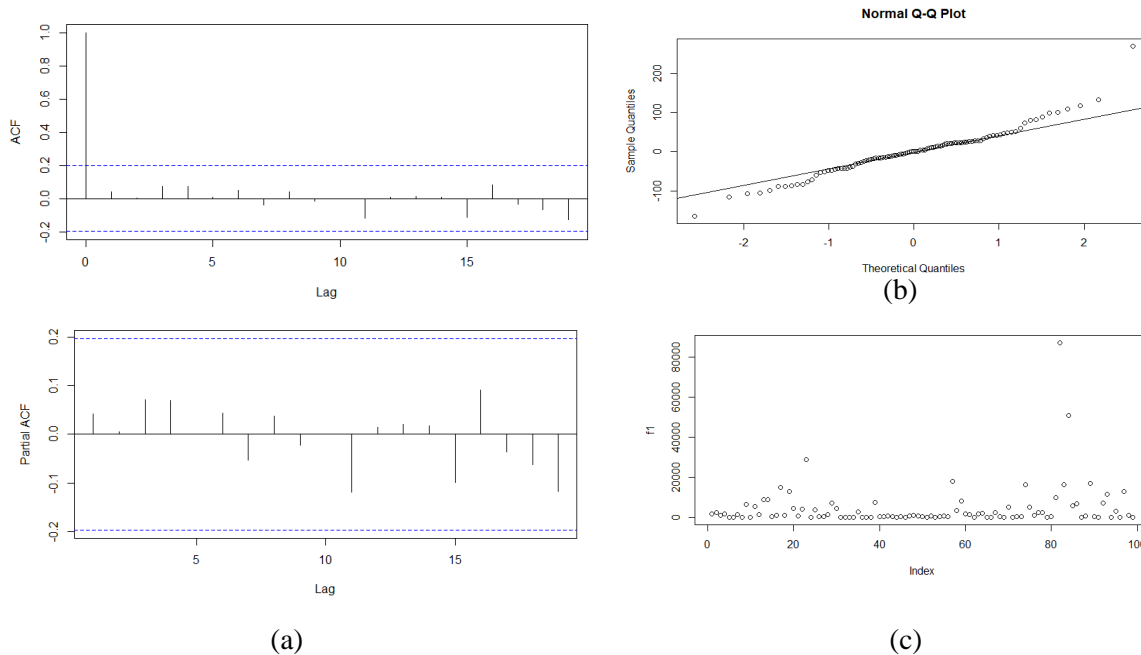
Setelah pensubstitusian $x_i = e_t$, diperoleh e_t^* . Selanjutnya di e_t^* ditambahkan ke data asli y_t , sehingga diperoleh \hat{y}_t^* yang merupakan hasil estimasi data ARIMA semiparametrik. Plot data hasil estimasi dari proses tahap semiparametrik disajikan pada Gambar 5 berikut :



Gambar 5. Plot Data Hasil Estimasi Semiparametrik

Pada Gambar 5 garis berwarna merah menunjukkan data hasil estimasi menggunakan pendekatan semiparametrik dan garis berwarna hitam merupakan data asli nilai *close price* IHS

Selanjutnya uji diagnostik dilakukan dengan melihat sifat keacakan *error* melalui plot ACF dan PACF pada *error* model ARIMA Semiparametrik, serta melihat asumsi kenormalan pada model melalui Q-Q plot *error*-nya. *Correlogram* ACF, PACF dan Q-Q plot dari *error* model ARIMA (2,1,0) dapat dilihat pada Gambar 6 berikut:



Gambar 6. (a) Plot ACF & PACF *Error* ARIMA Semiparametrik (b) Q-Q Plot *Error* ARIMA Semiparametrik (c) Plot Kuadrat *Error* ARIMA Semiparametrik

Berdasarkan Gambar 6.(a) dapat dilihat bahwa tidak terdapat satu lag pun yang keluar batas pada plot ACF dan PACF *error* sehingga dapat dikatakan bahwa *error* model bersifat *white noise*. Pada Gambar 6.(b) dapat dilihat bahwa sebaran data *error* pada plot sebagian besar mendekati garis diagonal menandakan bahwa *error* pada model berdistribusi normal. Pada Gambar 6.(c) plot kuadrat *error* ARIMA memiliki pola yang tidak beraturan sehingga dikatakan bahwa tidak terdapat heterokedastisitas pada data dan dapat disimpulkan bahwa model yang didapat merupakan model yang baik untuk digunakan dalam prediksi IHS.

Setelah didapatkan hasil estimasi dari metode ARIMA Parametrik dan ARIMA Semiparametrik, maka dapat dilakukan perbandingan keakuratan hasil estimasi data dari model ARIMA parametrik dan ARIMA semiparametrik dengan melihat nilai RMSE dari kedua model tersebut. Adapun perhitungan

nilai RMSE masing-masing model dilakukan dengan bantuan *software R Studio*. Nilai RMSE ARIMA parametrik dan ARIMA semiparametrik disajikan pada Tabel 4 berikut:

Tabel 4. Perbandingan *Error Model* ARIMA Parametrik dan ARIMA Semiparametrik

Model	RMSE
ARIMA Parametrik	61,629
ARIMA Semiparametrik	59,224

Berdasarkan Tabel 4 terlihat bahwa model ARIMA Semiparametrik menghasilkan nilai RMSE lebih kecil dibanding metode ARIMA parametrik, sehingga dapat disimpulkan bahwa model ARIMA semiparametrik memberikan nilai estimasi yang lebih akurat dari pada model ARIMA parametrik.

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan untuk menentukan estimasi model ARIMA parametrik dan ARIMA semiparametrik pada data runtun waktu Indeks Harga Saham Gabungan (IHSG), diperoleh bahwa model ARIMA semiparametrik memberikan hasil estimasi yang lebih akurat dibanding model ARIMA. Hal ini terlihat dari nilai RMSE pada model ARIMA semiparametrik yang bernilai 59,224 lebih kecil dari pada ARIMA parametrik yang bernilai 61,629.

DAFTAR PUSTAKA

- [1] Asrini LJ dan Budiantara IN. Fourier Series Semiparametric Regression Models. *ARPN Journal of Engineering and Applied Sciences*. 2014; 9:1501-1506.
- [2] Kuzairi, Miswanto, and Mardianto MFF. Semiparametric Regression Based on Fourier Series for Longitudinal Data with Weighted Least Square(WLS) Optimization. Di dalam: Kuzairi. *The 4th International Conference on Combinatorics. Graph Theory and Network Topology 2020*. East Java 22-23 August 2020. IOP Publishing. Jember.
- [3] Yundari, Pasaribu U, and Mukhaiyar U. Time Series Analysis Using Semiparametric Regression on Oil Palm Production. Di dalam: Mochamad Apri. *Symposium on Biomathematics 2015*. Bandung 4-6 November 2015. AIP Publisher. Mellville.
- [4] Boneva L, Linton O, Vogt MA. Semiparametric Model for Heterogeneous Panel Data with Fixed Effects. *Journal of Econometric*. 2015; 188:327-345.
- [5] Wei WWS. *Time Series Analysis Univariate and Multivariate Methods*. California: Pearson Education Inc; 2006.
- [6] Darsyah MY, Nur MS. Model Terbaik ARIMA dan Winter pada Peramalan Data Saham Bank. *Jurnal Statistika*. 2016; 4:31-32.
- [7] Dudewich dan Mishra. *Statistik Matematika Modern*. Bandung: ITB; 1995.
- [8] Saputra JA. Pemilihan Badwith pada Estimator Nadaraya-Watson dengan Tipe Kernel Gaussian Pada Data Runtun Waktu. *Jurnal Matematika*. 2016; 2:2-7
- [9] Suparti, Prahutama A. Permodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal Pada Beban Listrik di Kota Semarang. *Media Statistika*. 2016; 2:85-93.

ROSI PRATIWI : Jurusan Matematika FMIPA UNTAN, Pontianak
rosipratiwi@student.untan.ac.id

YUNDARI : Jurusan Matematika FMIPA UNTAN, Pontianak
yundari@math.untan.ac.id

HELMI : Jurusan Matematika FMIPA UNTAN, Pontianak
helmi@math.untan.ac.id

