

BAB I PENDAHULUAN

1.1 Latar Belakang

Bank Central Asia merupakan bank swasta terbesar di Indonesia yang memiliki peran penting dalam menjaga stabilitas sistem keuangan nasional. Saham Bank Central Asia tercatat di Bursa Efek Indonesia (BEI) dengan kode emiten BBKA. Berdasarkan laporan keuangan yang dirilis oleh PT Bank Central Asia Tbk. (2025), total aset BBKA per Januari 2025 tercatat sebesar Rp1.430 triliun. Selain itu, BBKA juga secara konsisten masuk dalam daftar 50 perusahaan dengan kapitalisasi pasar terbesar di Bursa Efek Indonesia (IDX, 2024). Stabilitas operasional, kinerja keuangan yang kuat, dan reputasi yang baik membuat saham BBKA menjadi pilihan utama bagi banyak investor. Kondisi tersebut mendorong perlunya analisis peramalan terhadap pergerakan harga saham BBKA, agar dapat dijadikan dasar dalam pengambilan keputusan investasi yang lebih bijak (Alfian et al., 2024).

Harga saham BBKA memiliki pola pergerakan yang tidak selalu stabil. Perubahan harga tersebut dapat terjadi secara *smooth* maupun secara mendadak (*abrupt*). Oleh karena itu, pola pergerakan data tidak sepenuhnya bersifat linier, melainkan memiliki karakteristik nonlinier yang dipengaruhi oleh berbagai faktor, seperti suku bunga, inflasi, kondisi global, serta sentimen pasar yang terus berubah. Perubahan ini sering kali membentuk beberapa *regime*, yaitu segmen-segmen dalam data yang masing-masing memiliki karakteristik statistik yang relatif konsisten dalam suatu rentang waktu. Ketika data berpindah dari satu *regime* ke *regime* lainnya, struktur hubungan antar variabel pun ikut berubah. Fenomena ini dikenal sebagai *regime switching*, yang menunjukkan adanya perubahan pola hubungan dalam data deret waktu akibat pergeseran kondisi (Guo et al., 2022). Kegagalan model dalam menangkap perubahan pola data dapat menurunkan akurasi peramalan, sehingga diperlukan pendekatan yang adaptif agar mampu merepresentasikan kondisi pasar secara lebih tepat.

Pendekatan yang digunakan dalam peramalan harga saham adalah analisis *time series* (Ridla et al., 2023). Salah satu model *time series* yang umum digunakan adalah *Autoregressive* (AR), yang memprediksi nilai saat ini berdasarkan kombinasi linier dari satu atau lebih nilai masa lalu (Yunita, 2019). Akan tetapi, model ini kurang tepat digunakan ketika data menunjukkan pola nonlinier. Oleh karena itu, model *Threshold Autoregressive* (TAR) diperkenalkan untuk menangani data yang berpola nonlinier dengan membagi data ke dalam beberapa *regime* berdasarkan nilai ambang batas (*threshold*) tertentu. Setiap *regime* pada model ini selanjutnya memiliki pola yang berbeda (Giordano et al., 2023). Model TAR kemudian menjadi *Self-Exciting Threshold Autoregressive* (SETAR), yang memiliki batas berdasarkan nilai *lag* dari variabel itu sendiri. Kedua model ini mampu mengidentifikasi transisi pola data yang bersifat mendadak dan mampu dalam menangani perubahan pola data yang terjadi secara mendadak (Chukwu et al., 2022).



Model *Smooth Transition Autoregressive* (STAR) diperkenalkan oleh Chan & Tong (1986) sebagai pengembangan dari model sebelumnya. Berbeda dengan TAR dan SETAR, STAR menggunakan fungsi transisi *smooth* seperti fungsi logistik atau eksponensial untuk menggambarkan perpindahan antar-*regime*. Perpindahan ini dikendalikan oleh parameter *threshold* dan *smoothness parameter*. Melalui mekanisme tersebut, parameter regresi dalam STAR dapat beradaptasi secara lebih fleksibel terhadap perubahan struktur data. Kemampuan tersebut membuat STAR lebih sesuai diterapkan pada data *time series* yang mengandung komponen nonlinier dengan pola pergerakan yang berubah secara *smooth*. Selain itu, pendekatan ini dianggap lebih representatif dalam menggambarkan dinamika sektor ekonomi dan keuangan, karena perubahan struktur data umumnya bersifat *smooth* akibat akumulasi berbagai faktor pasar (Lei et al., 2021). Karakteristik tersebut membuat STAR lebih unggul dibandingkan model *threshold* lainnya dan layak untuk digunakan dalam analisis pergerakan harga saham.

Sejumlah penelitian telah membuktikan efektivitas model STAR, seperti Blasques et al. (2020) yang membandingkan performa STAR dan AR pada tiga deret waktu makroekonomi di Amerika Serikat, dan hasilnya menunjukkan bahwa STAR memberikan peramalan yang lebih akurat berdasarkan *Root Mean Square Error* (RMSE), *log-likelihood*, dan *Akaike Information Criterion* (AIC). Chandra et al. (2021) juga membuktikan bahwa model STAR juga efektif digunakan dalam peramalan produksi kelapa sawit, dengan MAPE sebesar 17,07%. Selain itu, Aidoo et al. (2022) menunjukkan bahwa model STAR menghasilkan peramalan yang lebih baik dibandingkan AR dalam memproyeksikan penyebaran COVID-19 di Afrika, dengan nilai RMSE dan *Mean Absolute Error* (MAE) yang lebih rendah.

Karakteristik data harga saham BBKA yang bersifat nonlinier serta menunjukkan pola perubahan data yang bersifat *smooth* menunjukkan bahwa model STAR merupakan model yang tepat digunakan pada penelitian ini. Jenis estimasi parameter yang digunakan adalah *Nonlinear Least Squares*, karena estimasi ini mampu memberikan hasil peramalan yang sesuai dengan karakteristik model nonlinier. Berdasarkan latar belakang tersebut, peneliti akan melakukan penelitian dengan judul “Penerapan *Smooth Transition Autoregressive* Menggunakan Estimasi *Nonlinear Least Squares* (Studi Kasus: Peramalan Harga Saham Bank Central Asia)”.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data harga penutupan (*close*) saham harian Bank Central Asia pada tahun 2014-2024.



si yang digunakan adalah fungsi logistik.

re pada model STAR yang digunakan adalah dua, karena dua
 1 mewakili perubahan utama dalam pola data tanpa
 1 kompleksitas model.

4. Pemilihan *lag* optimal dibatasi hingga maksimal 3, sementara *delay* optimal dievaluasi hingga maksimal 5 untuk menjaga kestabilan model dan memperoleh estimasi yang lebih efisien.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan model STAR dengan estimasi *Nonlinear Least Squares* untuk peramalan harga saham Bank Central Asia.
2. Mendapatkan hasil peramalan harga saham Bank Central Asia tahun 2024 menggunakan model STAR.

Berdasarkan tujuan penelitian di atas maka setelah melakukan penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi mahasiswa, penelitian ini dapat menambah wawasan baik secara teori maupun praktik terkait penerapan metode statistik dalam peramalan harga saham yang memiliki pola data nonlinier.
2. Bagi investor, hasil penelitian ini dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan investasi, khususnya dengan memilih model yang lebih sesuai untuk melakukan peramalan harga saham BBKA.
3. Bagi peneliti selanjutnya, penelitian ini dapat menjadi acuan awal dalam mengembangkan model peramalan harga saham yang lebih akurat dan sesuai dengan karakteristik data yang bersifat nonlinier.

1.4 Teori

1.4.1 Time Series

Time series merupakan sekumpulan observasi yang dicatat secara berurutan berdasarkan interval waktu tertentu, seperti harian, mingguan, bulanan, maupun tahunan. Analisis *time series* digunakan secara luas di berbagai bidang seperti ekonomi, keuangan, dan meteorologi untuk memahami pola historis serta melakukan peramalan terhadap nilai masa depan (Danandeh Mehr et al., 2022).

Karakteristik utama dalam data *time series* meliputi tren (*trend*), musiman (*seasonality*), siklus (*cyclical*), dan variasi acak (*random variation*). Tren menggambarkan arah pergerakan data dalam jangka panjang, sedangkan musiman mencerminkan pola berulang yang terjadi secara periodik. Siklus merupakan fluktuasi jangka panjang yang dipengaruhi oleh kondisi ekonomi makro, sementara variasi acak mencerminkan komponen yang tidak terprediksi atau bersifat *noise* (Bhowmik, 2020).

Data *time series* dapat menunjukkan hubungan linier maupun nonlinier antar

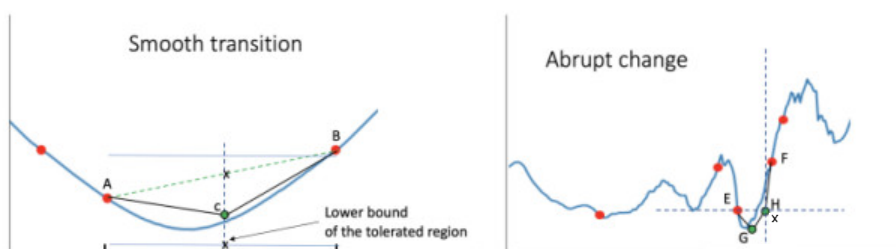


linier, pola hubungan antar variabel bersifat tetap sepanjang digunakan apabila data menunjukkan pola linier adalah *Moving Average* (MA), *Autoregressive Moving Average* (ARMA), *Integrated Moving Average* (ARIMA). Model ini banyak remodelkan data linier karena kesederhanaannya dalam ungkan yang stabil. Sebaliknya, data nonlinier menunjukkan pola

perubahan yang berubah-ubah bergantung pada kondisi tertentu dalam rentang waktu, sehingga hubungan antar variabel menjadi tidak konstan. Model yang digunakan pada data yang memiliki pola ini adalah *Threshold Autoregressive* (TAR), *Self-Exciting Threshold Autoregressive* (SETAR), dan *Smooth Transition Autoregressive* (STAR).

Fenomena perubahan dalam data *time series* nonlinier sering kali melibatkan konsep *regime*, yaitu segmen waktu dengan karakteristik statistik yang relatif stabil, seperti rata-rata (*mean*), variansi (*variance*), atau struktur autokorelasi. Setiap *regime* merepresentasikan dinamika data yang bersifat konsisten dalam suatu interval tertentu (Tsay, 2005). Pergeseran antar *regime* (*regime switching*) mencerminkan perubahan dalam hubungan antar variabel yang dapat terjadi secara alami di berbagai bidang seperti ekonomi, keuangan, iklim, kesehatan, dan rekayasa.

Pergeseran antar *regime* pada data nonlinier dapat terjadi melalui dua jenis transisi, yaitu *smooth* dan mendadak (*abrupt*). Transisi *smooth* menunjukkan perubahan antar kondisi yang berlangsung secara bertahap tanpa melewati batas (*threshold*) yang menandai adanya perubahan ekstrem, sedangkan transisi mendadak menggambarkan perubahan yang terjadi secara tiba-tiba ketika nilai data melampaui ambang batas tersebut. Gambar 1 menunjukkan perbedaan visualisasi pola perubahan yang terjadi secara *smooth* dan mendadak pada data *time series* (Bellucci et al., 2019).



Gambar 1. Perbedaan Visualisasi Transisi *Smooth* dan Mendadak

Gambar 1 menunjukkan bahwa pada transisi *smooth*, perubahan dari titik A menuju titik B melalui titik C berlangsung secara bertahap dalam rentang batas (*threshold*) yang telah ditentukan. Transisi antar kondisi tetap *smooth* karena perubahan nilai terjadi secara perlahan dan konsisten tanpa menimbulkan lonjakan ekstrem. Transisi ini mencerminkan karakteristik dinamika pola yang nonlinier dengan transisi *smooth*.

Sebaliknya, pada transisi mendadak, perubahan dari titik E menuju titik F terjadi secara tiba-tiba setelah nilai data melewati ambang batas. Perubahan ini menunjukkan adanya lonjakan ekstrem dalam peralihan kondisi yang tidak *smooth* dengan perubahan. Transisi ini menggambarkan pergantian *regime* secara cepat ketika nilai data melewati ambang batas yang ditetapkan.



1.4.2 Uji Augmented Dickey Fuller

Pembentukan model AR memerlukan data yang telah bersifat stasioner agar estimasi parameter menjadi valid dan hasil peramalan lebih akurat. Suatu *time series* dikatakan stasioner apabila karakteristik statistik deskriptifnya, seperti rata-rata dan varians, tetap konstan sepanjang waktu. Salah satu metode yang sering digunakan untuk menguji stasioneritas ialah *Augmented Dickey-Fuller* (ADF).

Uji ADF merupakan pengembangan dari uji *Dickey-Fuller* standar yang memasukkan *lag* dari variabel dependen untuk mengatasi autokorelasi pada *residual* (Ajewole KP et al., 2020). Bentuk umum uji ADF dijabarkan pada Persamaan (1):

$$\Delta Y_t = \delta Y_{t-1} + \sum_{j=1}^{q-1} \alpha_j \Delta Y_{t-j} + \varepsilon_t \quad (1)$$

dengan,

ΔY_t : *first difference* dari nilai *time series*

Y_t : nilai *time series* pada waktu ke- t , $t = 1, 2, 3, \dots, T$

δ : koefisien unit *root* yang diuji

α_j : koefisien *lag difference* ke- j

q : jumlah *lag* yang digunakan dalam uji ADF

ε_t : nilai *residual* pada waktu ke- t

Hipotesis uji:

H_0 : data memiliki unit *root* (tidak stasioner)

H_1 : data tidak memiliki unit *root* (stasioner)

Keputusan ditentukan dengan membandingkan nilai statistik uji ADF terhadap nilai kritis pada taraf signifikansi ($\alpha = 5\%$). Nilai kritis ini diperoleh dari distribusi khusus statistik ADF yang bergantung pada ukuran sampel dan spesifikasi model regresi.

Jika nilai statistik uji lebih kecil daripada nilai kritis (ADF < nilai kritis), maka H_0 ditolak, yang menunjukkan bahwa data bersifat stasioner. Sebaliknya, jika nilai statistik uji lebih besar atau tidak melewati batas kritis, maka tidak terdapat cukup bukti untuk menolak H_0 , sehingga data dianggap mengandung unit *root* atau tidak stasioner.

1.4.3 Autocorrelation Function

Fungsi *Autocorrelation Function* (ACF) digunakan untuk mengukur kekuatan hubungan linier antara nilai observasi dalam deret waktu dengan nilai-nilai sebelumnya yang terpisah oleh sejumlah *lag* tertentu. ACF sangat berguna untuk mengidentifikasi pola ketergantungan waktu, khususnya dalam menentukan orde model yang tepat (Yekubu & Saputra, 2022).



si pada *lag* ke- k , yang dinotasikan sebagai ρ_k secara teoritis in Persamaan (2):

$$\frac{\text{cov}(Y_t, Y_{t-k})}{\sqrt{\text{var}(Y_t)} \cdot \sqrt{\text{var}(Y_{t-k})}} = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sqrt{\text{var}(Y_t)} \cdot \sqrt{\text{var}(Y_{t-k})}} \quad (2)$$

Data *time series* diasumsikan stasioner, oleh karena itu $var(Y_t) = var(Y_{t-k}) = \gamma_0$, maka persamaan (2) dapat disederhanakan menjadi $\rho_k = \frac{\gamma_k}{\gamma_0}$, dengan γ_k adalah fungsi autokovariansi pada *lag* ke- k dan γ_0 adalah *varians* dari Y_t .

Estimasi ρ_k dari data sampel diperoleh melalui Persamaan (3):

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (3)$$

dengan $\bar{Y}$ menyatakan nilai rata-rata sampel dari data *time series*. Hipotesis pengujian terhadap ACF dinyatakan sebagai berikut:

H_0 : tidak terdapat autokorelasi antara Y_t dengan Y_{t-k} ($\rho_k = 0$)

H_1 : terdapat autokorelasi antara Y_t dengan Y_{t-k} ($\rho_k \neq 0$)

Keputusan pengujian didasarkan pada nilai estimasi $\hat{\rho}_k$. Penolakan terhadap H_0 dilakukan jika nilai $|\hat{\rho}_k|$ berada di luar batas interval kepercayaan, yaitu:

$$|\hat{\rho}_k| > \frac{z_{1-\alpha/2}}{\sqrt{T}}$$

dengan $z_{1-\alpha/2}$ merupakan nilai kritis dari distribusi normal standar pada taraf signifikansi α yang pada penelitian ini digunakan $\alpha = 5\%$. Apabila nilai $\hat{\rho}_k$ melebihi batas interval kepercayaan, maka *lag* ke- k dianggap signifikan dan menunjukkan adanya pola ketergantungan waktu dalam data.

1.4.4 Partial Autocorrelation Function

Fungsi *Partial Autocorrelation Function* (PACF) digunakan untuk mengukur kekuatan hubungan antara suatu variabel dan nilai masa lalunya dengan mengendalikan efek dari *lag* yang berada di antaranya. Informasi yang diberikan oleh PACF melengkapi struktur ketergantungan dalam *time series*, terutama pada proses identifikasi orde model AR yang optimal (Yakubu & Saputra, 2022).

Hipotesis pengujian PACF dirumuskan sebagai berikut.

H_0 : tidak terdapat korelasi parsial antara Y_t dengan Y_{t-k} ($\phi_{kk} = 0$)

H_1 : terdapat korelasi parsial antara Y_t dengan Y_{t-k} ($\phi_{kk} \neq 0$)

Nilai PACF pada *lag* ke- k dihitung menggunakan Persamaan (4) berikut.

$$\phi_{kk} = \rho_k - \sum_{j=1}^{k-1} \phi_j \rho_{k-j} \quad (4)$$

Pengujian signifikansi terhadap nilai ϕ_{kk} dilakukan berdasarkan Persamaan (5) sebagai berikut.

$$t = \frac{\phi_{kk}}{SE(\phi_{kk})} \quad (5)$$

dengan ϕ_{kk} fungsi PACF pada *lag* ke- k , dan $SE(\phi_{kk}) = \frac{1}{\sqrt{T}}$ merupakan simpangan



k .

didasarkan pada nilai estimasi ϕ_{kk} . Penolakan terhadap H_0 dilakukan jika nilai $|\phi_{kk}|$ berada di luar batas interval kepercayaan, yaitu:

$$|\phi_{kk}| > t_{\alpha/2, T-1}$$

di mana $t_{\alpha/2, T-1}$ adalah nilai kritis distribusi *t-student* pada taraf signifikansi α dan T adalah jumlah observasi. Jika kondisi tersebut terpenuhi, maka korelasi parsial pada *lag*

ke- k dinyatakan signifikan secara statistik dan berkontribusi terhadap penentuan struktur model *time series*.

1.4.5 Model Autoregressive

Model *Autoregressive* (AR) merupakan metode linier dalam analisis *time series* yang memanfaatkan hubungan antara nilai saat ini dan nilai-nilai pada periode sebelumnya (Kaur et al., 2023). Model ini mengasumsikan bahwa nilai saat ini memiliki ketergantungan pada beberapa nilai masa lalu melalui kombinasi bobot yang direpresentasikan oleh parameter tertentu. Bentuk umum model AR(p) dapat dituliskan menggunakan Persamaan (6) berikut.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (6)$$

dengan $\phi_1, \phi_2, \dots, \phi_p$ adalah parameter *autoregressive*.

1.4.6 Ordinary Least Squares

Estimasi parameter pada model AR dilakukan menggunakan *ordinary least squares*, yaitu dengan meminimalkan jumlah kuadrat *residual* dari hasil peramalan model (Ding, 2023). Berdasarkan bentuk model AR dalam Persamaan (6), proses estimasi dilakukan melalui turunan parsial dari fungsi jumlah kuadrat *residual* terhadap setiap parameter $\phi_1, \phi_2, \dots, \phi_p$, untuk memperoleh bentuk estimasinya. Turunan parsial dari fungsi jumlah kuadrat *residual* S terhadap setiap parameter $\phi_1, \phi_2, \dots, \phi_p$ menghasilkan sistem persamaan normal berikut:

$$\left. \frac{\partial S}{\partial \phi_1} \right|_{\phi_1 = \hat{\phi}_1} = -2 \sum_{t=p+1}^T Y_{t-1} (Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}) = 0$$

$$\left. \frac{\partial S}{\partial \phi_2} \right|_{\phi_2 = \hat{\phi}_2} = -2 \sum_{t=p+1}^T Y_{t-2} (Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}) = 0$$

$$\left. \frac{\partial S}{\partial \phi_p} \right|_{\phi_p = \hat{\phi}_p} = -2 \sum_{t=p+1}^T Y_{t-p} (Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}) = 0$$

Hasil turunan tersebut membentuk sistem persamaan normal yang menyatakan hubungan linier antar parameter dengan data pengamatan. Sistem ini dapat dituliskan dalam bentuk eksplisit sebagai berikut:

$$\hat{\phi}_1 \sum_{t=p+1}^T Y_{t-1}^2 + \hat{\phi}_2 \sum_{t=p+1}^T Y_{t-1} Y_{t-2} + \dots + \hat{\phi}_p \sum_{t=p+1}^T Y_{t-1} Y_{t-p} = \sum_{t=p+1}^T Y_{t-1} Y_t$$



$$\hat{\phi}_1 \sum_{t=p+1}^T Y_{t-2}^2 + \dots + \hat{\phi}_p \sum_{t=p+1}^T Y_{t-2} Y_{t-p} = \sum_{t=p+1}^T Y_{t-2} Y_t$$

⋮

$$\hat{\phi}_1 \sum_{t=p+1}^T Y_{t-p} Y_{t-2} + \dots + \hat{\phi}_p \sum_{t=p+1}^T Y_{t-p}^2 = \sum_{t=p+1}^T Y_{t-p} Y_t$$

Sistem tersebut dapat direpresentasikan dalam bentuk notasi matriks pada Persamaan (7) berikut:

$$X'X\hat{\phi} = X'Y \quad (7)$$

dengan,

$$X = \begin{bmatrix} Y_p & Y_{p-1} & \cdots & Y_1 \\ Y_{p+1} & Y_p & \cdots & Y_2 \\ \vdots & \vdots & \ddots & \vdots \\ Y_{T-1} & Y_{T-2} & \cdots & Y_{T-p} \end{bmatrix}, Y = \begin{bmatrix} Y_{p+1} \\ Y_{p+2} \\ \vdots \\ Y_T \end{bmatrix}, \text{ dan } \hat{\phi} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix}$$

Berdasarkan Persamaan (7), nilai estimasi parameter $\hat{\phi}$ dirumuskan melalui Persamaan (8) berikut.

$$\hat{\phi} = (X'X)^{-1}X'Y \quad (8)$$

1.4.7 Ljung-Box Test

Uji *Ljung-Box* merupakan metode statistik yang digunakan untuk mendeteksi keberadaan autokorelasi dalam *residual* suatu model *time series* pada beberapa *lag* sekaligus. Tujuan dari uji ini adalah mengevaluasi keberadaan pola yang tersisa dalam *residual* dan menentukan kesesuaiannya dengan karakteristik *white noise*, yaitu tidak memiliki struktur yang dapat dimodelkan lebih lanjut (Danioko et al., 2022). Statistik uji *Ljung-Box* dihitung menggunakan Persamaan (9) berikut.

$$Q = T(T+2) \sum_{k=1}^h \frac{\hat{p}_k^2}{T-k} \quad (9)$$

dengan,

Q : statistik uji *Ljung-Box*

$\hat{p}_k^2$: estimasi koefisien autokorelasi pada *lag* ke- k

h : jumlah maksimum *lag* yang diuji dalam uji *Ljung-Box*

Hipotesis pengujian disusun sebagai berikut

H_0 : *residual* tidak memiliki autokorelasi (*white noise*)

H_1 : *residual* memiliki autokorelasi (terdapat pola dalam *residual*)

Statistik uji Q mengikuti distribusi *Chi-Square* dengan derajat bebas h . Pengambilan keputusan dinyatakan sebagai berikut:

$$Q_{hitung} > \chi_{(1-\alpha, h-p)}^2$$

Kriteria pengujian:

Tolak H_0 jika nilai Q lebih besar dari nilai kritis $\chi_{(1-\alpha, h-p)}^2$, atau jika *p-value* lebih kecil dari tingkat signifikansi α yaitu 5%.



akan bahwa suatu fungsi yang memiliki turunan kontinu hingga diaproksimasi (didekati) oleh polinomial berdasarkan nilai fungsi di suatu titik pusat. Aproksimasi dalam konteks ini merujuk pada aproksimasi fungsi asli dengan polinomial yang nilainya mendekati fungsi asli di sekitar titik tertentu di sekitar titik pusat (Stewart, 2016).

Secara matematis, deret Taylor untuk fungsi $f(x)$ yang terdiferensiasi secara kontinu hingga orde ke- n di sekitar titik pusat a dinyatakan pada Persamaan (10):

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^n(a)}{n!}(x-a)^n + R_n(x) \quad (10)$$

dengan $f'(a), f''(a), \dots, f^n(a)$ adalah turunan pertama hingga ke- n dari $f(x)$ yang dievaluasi pada titik pusat a , $n!$ adalah faktorial dari bilangan bulat positif n , dan $(x-a)$ adalah selisih antara variabel bebas dengan titik pusat. Suku $R_n(x)$ menunjukkan sisa aproksimasi yang menggambarkan tingkat kesalahan yang menggambarkan tingkat kesalahan akibat penghentian deret pada orde ke- n .

Tujuan utama dari penggunaan deret Taylor adalah untuk menyederhanakan bentuk fungsi yang kompleks sehingga lebih mudah dianalisis, dihitung, atau digunakan dalam pendekatan numerik. Dalam bidang statistika, deret Taylor digunakan untuk mendekati hubungan nonlinier antar variabel melalui bentuk polinomial, sehingga pola hubungan yang sulit dianalisis secara langsung dapat direpresentasikan secara lebih sederhana. Deret Taylor juga banyak diterapkan dalam disiplin ilmu lain seperti fisika, teknik, ekonomi, serta bidang lainnya yang memerlukan formulasi matematis untuk menyelesaikan permasalahan komputasi dan analisis data berskala besar.

1.4.9 Uji Nonlinieritas

Uji nonlinieritas merupakan langkah penting dalam mengidentifikasi data *time series* mengikuti pola nonlinier atau cukup direpresentasikan dengan model linier seperti AR. Jika data menunjukkan karakteristik nonlinier, maka model STAR dapat digunakan untuk menangkap pola transisi *smooth* dalam data.

Variabel dependen yang digunakan dalam pengujian nonlinieritas ini dilambangkan dengan z_t , yaitu transformasi dari data Y_t . Notasi ini diterapkan sebagai pembeda untuk memudahkan identifikasi antara tahapan pengujian nonlinieritas dan tahapan pemodelan utama, sehingga

$$z_t = Y_t.$$

Variabel transisi dinotasikan dengan s_{t-d} , yang didefinisikan sebagai:

$$s_{t-d} = Y_{t-d} - c,$$

dengan c adalah nilai *threshold* yang menentukan titik perpindahan antar *regime* dan d adalah *delay* yang menunjukkan *lag* observasi dalam mendeteksi transisi antar *regime*.

Salah satu metode yang umum digunakan untuk menguji nonlinieritas adalah uji *Teräsvirta* (Hepkorucu, 2022). Hipotesis nol dalam pengujian ini menyatakan bahwa model linier cukup dalam menjelaskan pola data, sedangkan hipotesis alternatif menunjukkan adanya pola nonlinier pada data. Secara matematis, dapat dinyatakan sebagai:

$$H_0: \beta_j = 0, j \in \{1, 2, 3\}$$

Sedangkan dalam pengujian ini adalah ketidakmampuan parameter fungsi transisi $G(Y_{t-d}; \gamma, c)$ secara langsung ketika H_0 diterima. Pada kondisi model linier, parameter γ dan c tidak



memiliki nilai estimasi yang valid, sehingga interpretasi terhadap proses transisi antar regime tidak dapat dilakukan.

Pendekatan aproksimasi deret Taylor terhadap fungsi transisi digunakan sebagai solusi dalam menghadapi permasalahan tersebut. Dijk et al. (2002) menjelaskan bahwa fungsi transisi $G(Y_{t-d}; \gamma, c)$ direpresentasikan menggunakan perluasan deret Taylor hingga orde tiga agar mengatasi persoalan parameter yang tidak teridentifikasi dalam pengujian. Bentuk deret yang digunakan merupakan deret Maclaurin, yaitu bentuk khusus dari deret Taylor yang dikembangkan di sekitar nol. Aproksimasi ini digunakan karena fungsi transisi dalam model STAR didekati di sekitar $\gamma = 0$. Transformasi ini menghasilkan bentuk polinomial $\beta_1 s_{t-d} + \beta_2 s_{t-d}^2 + \beta_3 s_{t-d}^3$ dengan koefisien regresi yang diestimasi melalui metode *Ordinary Least Squares* (OLS) dalam *auxiliary regression*.

Perluasan deret Taylor ini menjadi dasar dalam membangun *auxiliary regression* untuk menguji keberadaan nonlinieritas menggunakan statistik *Lagrange Multiplier* (LM). Statistik LM berfungsi mengevaluasi signifikansi keberadaan efek nonlinier dengan membandingkan model linier awal dan model yang memuat komponen nonlinier. Pengujian dilakukan melalui beberapa tahap. Tahap pertama dimulai dengan membangun model linier untuk variabel dependen z_t , yang dirumuskan sebagai

$$z_t = \beta'_0 X_t + \varepsilon_t$$

dengan,

$$\beta_0 = \begin{bmatrix} \beta_{0,0} \\ \beta_{0,1} \\ \beta_{0,2} \\ \vdots \\ \beta_{0,p} \end{bmatrix}, X_t = \begin{bmatrix} 1 \\ z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{bmatrix}$$

Apabila dijabarkan, maka model linier z_t menjadi:

$$z_t = \beta_{0,0} + \beta_{0,1} z_{t-1} + \beta_{0,2} z_{t-2} + \dots + \beta_{0,p} z_{t-p} + \varepsilon_t^{(0)}$$

Setelah model linier ini diestimasi, *residual* ($\hat{\varepsilon}_t^{(0)}$) dihitung dan jumlah kuadrat *residual* diperoleh berdasarkan persamaan berikut:

$$SSR_0 = \sum_{t=1}^T (\hat{\varepsilon}_t^{(0)})^2$$

Tahap berikutnya adalah menyusun model regresi bantu (*auxiliary regression*) dengan menambahkan komponen nonlinier berupa perluasan deret Taylor terhadap variabel transisi s_{t-d} hingga orde tiga yang dijabarkan pada Persamaan (11)



$$z_t = \beta'_0 X_t + \beta_1 s_{t-d} + \beta_2 s_{t-d}^2 + \beta_3 s_{t-d}^3 + \varepsilon_t^{(1)} \quad (11)$$

lel regresi bantu ini diestimasi, *residual* $\hat{\varepsilon}_t^{(1)}$ dihitung dan drat *residual* dirumuskan sebagai berikut:

$$SSR_1 = \sum_{t=1}^T (\hat{\varepsilon}_t^{(1)})^2$$

Berdasarkan perbedaan jumlah kuadrat *residual* dari kedua model tersebut, uji LM dilakukan dengan menggunakan statistik uji F menggunakan Persamaan (12) berikut:

$$F = \frac{\frac{SSR_0 - SSR_1}{w}}{\frac{SSR_1}{(T - p - 1 - w)}} \sim F_{w, T-p-1-w} \quad (12)$$

dengan w merupakan jumlah parameter nonlinier yang digunakan pada regresi bantu. Keputusan pengujian diambil berdasarkan nilai F . Jika nilai $F \leq F_{w, T-p-1-w}$ atau hasil dari $p - value \geq 0.05$, maka H_0 diterima dan data dianggap linier. Sebaliknya, jika $F > F_{w, T-p-1-w}$ atau $p - value < 0.05$, maka H_0 ditolak dan disimpulkan bahwa data menunjukkan karakteristik nonlinier, sehingga model STAR layak digunakan.

1.4.10 Pemilihan Fungsi Transisi Model *Smooth Transition Autoregressive*

Pemilihan fungsi transisi menjadi tahapan penting dalam pengembangan model STAR setelah pola nonlinier teridentifikasi pada data deret waktu. Fungsi transisi mengatur perpindahan antar *regime* secara *smooth*, sehingga model mampu menggambarkan perubahan kondisi secara realistis (Terasvirta, 1994).

Fungsi transisi direpresentasikan dalam bentuk perluasan deret Taylor hingga orde tiga, yang kemudian diimplementasikan dalam model regresi bantu (*auxiliary regression*) dengan struktur sebagai berikut:

$$z_t = \beta'_0 X_t + \beta'_1 X_t s_{t-d} + \beta'_2 X_t s_{t-d}^2 + \beta'_3 X_t s_{t-d}^3 + \varepsilon_t$$

Setiap vektor parameter β_j pada persamaan tersebut memiliki dimensi $(p+1) \times 1$ dan mengandung koefisien yang mewakili kontribusi *lag-lag* variabel dependen terhadap komponen nonlinier pangkat ke- j dari variabel transisi. Dengan demikian, β_j dapat diuraikan sebagai

$$\beta_j = \begin{bmatrix} \beta_{j,0} \\ \beta_{j,1} \\ \beta_{j,2} \\ \vdots \\ \beta_{j,p} \end{bmatrix}$$

Penyertaan X_t pada setiap komponen nonlinear bertujuan untuk menangkap pengaruh *lag* variabel dependen terhadap efek nonlinear dari variabel transisi, sehingga model tidak hanya menguji variabel transisi s_{t-d}^j secara terpisah, tetapi juga memperhitungkan interaksi dengan kondisi data masa lalu.

Prosedur pemilihan fungsi transisi dilakukan melalui pengujian bertahap terhadap parameter β_j dengan hipotesis:

1. $H_{0,3}: \beta_3 = 0$
2. $H_{0,2}: \beta_2 = 0$ dengan $\beta_3 = 0$
3. $H_{0,1}: \beta_1 = 0$ dengan $\beta_2 = \beta_3 = 0$



pengujian tersebut, interpretasi model adalah sebagai berikut:

1. Jika $\beta_3 = 0$, maka digunakan fungsi transisi logistik.

2. Jika $\beta_2 \neq 0$, maka digunakan fungsi transisi eksponensial.

3. Jika $\beta_2 = 0$, tetapi $\beta_1 \neq 0$, maka digunakan fungsi transisi logistik.

1.4.11 Smooth Transition Autoregressive

Model *Smooth Transition Autoregressive* (STAR) merupakan pengembangan dari model AR yang mendukung adanya transisi antara dua atau lebih *regime* dalam data *time series*. Model ini digunakan untuk menangkap pola nonlinier yang tidak dapat dijelaskan oleh model linier (Ubilava, 2022).

Secara umum, bentuk dasar dari model STAR dengan dua *regime* yang memiliki orde *lag* maksimum sebesar *m* (yang diperoleh dari *lag p* dari model AR) dan variabel transisi Y_{t-d} , dengan *d* menunjukkan *delay*, dapat dituliskan pada Persamaan (13).

$$Y_t = \phi'_1 X_t (1 - G(Y_{t-d}; \gamma, c)) + \phi'_2 X_t G(Y_{t-d}; \gamma, c) + \varepsilon_t \quad (13)$$

dengan,

ϕ_1 : vektor parameter pada *regime* pertama, berbentuk

$$\phi_1 = \begin{bmatrix} \phi_{1,0} \\ \phi_{1,1} \\ \phi_{1,2} \\ \vdots \\ \phi_{1,m} \end{bmatrix}$$

ϕ_2 : vektor parameter pada *regime* kedua, berbentuk

$$\phi_2 = \begin{bmatrix} \phi_{2,0} \\ \phi_{2,1} \\ \phi_{2,2} \\ \vdots \\ \phi_{2,m} \end{bmatrix}$$

X_t : vektor regressor, yang mencakup

$$X_t = \begin{bmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-m} \end{bmatrix}$$

$G(Y_{t-d}; \gamma, c)$: fungsi transisi dengan nilai antara 0 hingga 1

c : parameter lokasi (*threshold*) yang menentukan titik transisi antar *regime*

γ : parameter *smoothness* yang mengontrol transisi antar *regime*

Model pada Persamaan (13) dapat dituliskan kembali secara eksplisit sebagai berikut:

$$Y_t = \left(\phi_{1,0} + \sum_{j=1}^m \phi_{1,j} Y_{t-j} \right) (1 - G(Y_{t-d}; \gamma, c)) + \left(\phi_{2,0} + \sum_{j=1}^m \phi_{2,j} Y_{t-j} \right) G(Y_{t-d}; \gamma, c) + \varepsilon_t$$



ini, komponen X_t menjadi komponen individual dalam bentuk dependen, dengan $\phi_{1,0}, \phi_{1,1}, \dots, \phi_{1,m}$ sebagai parameter *regime* pertama dan $\phi_{2,0}, \phi_{2,1}, \dots, \phi_{2,m}$ sebagai parameter *regime* kedua. Fungsi transisi $G(Y_{t-d}; \gamma, c)$ berfungsi mengatur *regime* berdasarkan nilai dari variabel transisi.

isi dalam model STAR bertanggung jawab atas pergerakan (Pratama et al., 2022). Terdapat dua jenis fungsi transisi utama

yang digunakan dalam model ini, yaitu *Logistic Smooth Transition Autoregressive* (LSTAR) yang dapat dilihat pada Persamaan (14) berikut:

$$G(Y_{t-d}; \gamma, c) = \frac{1}{1 + \exp(-\gamma(Y_{t-d} - c))}, \gamma > 0 \quad (14)$$

Fungsi ini digunakan ketika perubahan antar *regime* bersifat asimetris dan terjadi secara bertahap. Sedangkan *Exponential Smooth Transition Autoregressive* (ESTAR) yang dapat dilihat pada Persamaan (15) berikut:

$$G(Y_{t-d}; \gamma, c) = 1 - \exp(-\gamma(Y_{t-d} - c)^2), \gamma > 0 \quad (15)$$

Fungsi ini digunakan ketika perubahan antar *regime* bersifat simetris di sekitar nilai ambang c .

1.4.12 Nonlinear Least Squares

Nonlinear Least Squares (NLS) digunakan untuk mengestimasi parameter dari model regresi nonlinier, yaitu model dengan hubungan tidak linier antara variabel dependen dan parameternya. Secara umum, model nonlinier dinyatakan pada Persamaan (16):

$$Y_t = F(X_t; \theta) + \varepsilon_t \quad (16)$$

dengan θ adalah vektor parameter yang akan diestimasi dan $F(X_t; \theta)$ adalah fungsi prediksi nonlinier bergantung pada vektor parameter θ . Tujuan utama dari metode NLS adalah memperoleh estimasi parameter $\hat{\theta}$ yang diperoleh dengan cara meminimalkan jumlah kuadrat *residual* antara nilai aktual dan nilai peramalan sesuai dengan persamaan berikut:

$$\hat{\theta} = \arg \min_{\theta} \sum_{t=1}^T (Y_t - F(X_t; \theta))^2$$

Pada konteks model STAR, bentuk fungsi $F(X_t; \theta)$ merepresentasikan struktur model dengan transisi *smooth* dan dirumuskan sebagai berikut:

$$F(X_t; \theta) = \left(\phi_{1,0} + \sum_{j=1}^m \phi_{1,j} Y_{t-j} \right) (1 - G(Y_{t-d}; \gamma, c)) + \left(\phi_{2,0} + \sum_{j=1}^m \phi_{2,j} Y_{t-j} \right) G(Y_{t-d}; \gamma, c)$$

Prosedur estimasi dilakukan secara iteratif karena struktur nonlinier dari fungsi tersebut tidak memberikan penyelesaian eksplisit seperti pada regresi linier. Salah satu metode numerik yang digunakan untuk menyelesaikan persoalan ini adalah metode *Gauss-Newton* (Cartis et al., 2022). Metode ini bertumpu pada prinsip bahwa fungsi nonlinier dapat direpresentasikan secara lokal dalam bentuk linier menggunakan pengembangan deret *Taylor* orde satu (*linear approximation*) di sekitar estimasi parameter awal, yang dinotasikan sebagai $g^{(0)}$.

Secara umum iterasi *Gauss-Newton* dinyatakan pada Persamaan (17) berikut:

$$\theta^{(i+1)} = \theta^{(i)} - [D^{(i)'} D^{(i)}]^{-1} D^{(i)'} [Y - F^{(i)}] \quad (17)$$

matriks *Jacobian* dari fungsi $F(X_t; \theta)$ terhadap parameter θ , a iterasi ke- i , Y adalah vektor nilai aktual Y_t , dan $F^{(i)}$ adalah fungsi model.

tama dalam algoritma *Gauss-Newton* adalah menentukan parameter $g^{(0)}$, dengan:



$$\mathbf{g}^{(0)} = \begin{bmatrix} g_0^{(0)} \\ g_1^{(0)} \\ \vdots \\ g_M^{(0)} \end{bmatrix}$$

Vektor $\mathbf{g}^{(0)}$ berukuran $M \times 1$, dengan M merupakan jumlah parameter bebas dalam model. Kemudian, menyusun bentuk linier dari fungsi $F(\mathbf{X}_t; \boldsymbol{\theta})$ didekati dengan bentuk linier menggunakan ekspansi deret Taylor orde satu di sekitar titik $\boldsymbol{\theta} = \mathbf{g}^{(0)}$, sebagaimana ditunjukkan pada Persamaan (18) berikut.

$$F(\mathbf{X}_t; \boldsymbol{\theta}) \approx F(\mathbf{X}_t; \mathbf{g}^{(0)}) + \left[\frac{\partial F(\mathbf{X}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\mathbf{g}^{(0)}} (\boldsymbol{\theta} - \mathbf{g}^{(0)}) \quad (18)$$

Notasi disederhanakan dengan $F(\mathbf{X}_t; \mathbf{g}^{(0)})$ menjadi $F_t^{(0)}$, $\boldsymbol{\theta} - \mathbf{g}^{(0)}$ menjadi $\boldsymbol{\beta}^{(0)}$, dan $\left[\frac{\partial F(\mathbf{X}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\mathbf{g}^{(0)}}$ menjadi $\mathbf{D}_t^{(0)}$. Maka, persamaan (18) dapat ditulis menjadi:

$$F(\mathbf{X}_t; \boldsymbol{\theta}) \approx F_t^{(0)} + \mathbf{D}_t^{(0)} \boldsymbol{\beta}^{(0)}$$

Bentuk ini kemudian disubstitusikan ke dalam model regresi nonlinier pada Persamaan (16), sehingga persamaan regresi menjadi:

$$Y_t \approx F_t^{(0)} + \mathbf{D}_t^{(0)} \boldsymbol{\beta}^{(0)} + \varepsilon_t$$

Selanjutnya, kedua ruas persamaan dikurangi dengan $F_t^{(0)}$, sehingga menghasilkan persamaan berikut:

$$Y_t - F_t^{(0)} \approx \mathbf{D}_t^{(0)} \boldsymbol{\beta}^{(0)} + \varepsilon_t$$

Residual awal didefinisikan sebagai:

$$Y_t^{(0)} = Y_t - F_t^{(0)}$$

Dengan demikian, model regresi linier pada iterasi pertama dapat diformulasikan dalam bentuk:

$$Y_t^{(0)} = \mathbf{D}_t^{(0)} \boldsymbol{\beta}^{(0)} + \varepsilon_t$$

Model ini kemudian dikembangkan untuk seluruh observasi dalam bentuk Persamaan (19):

$$\mathbf{Y}^{(0)} = \mathbf{D}^{(0)} \boldsymbol{\beta}^{(0)} + \boldsymbol{\varepsilon} \quad (19)$$

dengan,

$$\mathbf{Y}^{(0)} = \begin{bmatrix} Y_1 - F_1^{(0)} \\ Y_2 - F_2^{(0)} \\ \vdots \\ Y_T - F_T^{(0)} \end{bmatrix}$$



$$\begin{bmatrix} \frac{\partial F_1^{(0)}}{\partial \phi_{2,0}} & \dots & \frac{\partial F_1^{(0)}}{\partial \phi_{2,m}} & \frac{\partial F_1^{(0)}}{\partial \gamma} & \frac{\partial F_1^{(0)}}{\partial c} \\ \frac{\partial F_2^{(0)}}{\partial \phi_{2,0}} & \dots & \frac{\partial F_2^{(0)}}{\partial \phi_{2,m}} & \frac{\partial F_2^{(0)}}{\partial \gamma} & \frac{\partial F_2^{(0)}}{\partial c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_T^{(0)}}{\partial \phi_{2,0}} & \dots & \frac{\partial F_T^{(0)}}{\partial \phi_{2,m}} & \frac{\partial F_T^{(0)}}{\partial \gamma} & \frac{\partial F_T^{(0)}}{\partial c} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1^{(0)}}{\partial \boldsymbol{\theta}} \\ \frac{\partial F_2^{(0)}}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial F_T^{(0)}}{\partial \boldsymbol{\theta}} \end{bmatrix}$$

$$\boldsymbol{\beta}^{(0)} = \begin{bmatrix} \beta_0^{(0)} \\ \beta_1^{(0)} \\ \vdots \\ \beta_M^{(0)} \end{bmatrix}$$

Vektor $\boldsymbol{\beta}^{(0)}$ kemudian ditaksir dari model regresi linier dengan notasi $\mathbf{b}^{(0)}$ menggunakan metode kuadrat terkecil sebagai berikut:

$$\mathbf{b}^{(0)} = [\mathbf{D}^{(0)'} \mathbf{D}^{(0)}]^{-1} \mathbf{D}^{(0)'} \mathbf{Y}^{(0)}$$

Kemudian, parameter diperbarui dengan menambahkan hasil estimasi tersebut pada estimasi awal, sehingga dijabarkan Persamaan (20).

$$\mathbf{g}^{(1)} = \mathbf{g}^{(0)} + \mathbf{b}^{(0)} \quad (20)$$

Proses ini diulang secara iteratif, dan pada setiap iterasi ke- i , parameter diperbarui sesuai rumus:

$$\mathbf{g}^{(i+1)} = \mathbf{g}^{(i)} + [\mathbf{D}^{(i)'} \mathbf{D}^{(i)}]^{-1} \mathbf{D}^{(i)'} \mathbf{Y}^{(i)}$$

Iterasi dihentikan ketika norma perbedaan antara dua estimasi berturut-turut $\|\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)}\|$ cukup kecil. Nilai yang mendekati nol menunjukkan bahwa estimasi parameter telah mencapai titik konvergensi. Estimasi terakhir dari $\boldsymbol{\theta}$ pada iterasi akhir dianggap sebagai hasil optimal dari proses NLS untuk model STAR.

1.4.13 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah salah satu metrik yang digunakan untuk mengevaluasi kualitas peramalan dengan mengukur selisih antara nilai aktual dan nilai peramalan dalam bentuk persentase absolut (Chicco et al., 2021). Metrik ini memberikan informasi mengenai akurasi relatif model terhadap skala data. Nilai MAPE yang lebih kecil menunjukkan kualitas peramalan yang lebih baik. Berikut pada Persamaan (21) dijabarkan rumus dari MAPE.

$$MAPE = \frac{100}{T} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (21)$$

1.4.14 Akaike Information Criterion

Akaike Information Criterion (AIC) adalah salah satu kriteria yang digunakan untuk mengevaluasi dan membandingkan kualitas model statistik dengan mempertimbangkan keseimbangan antara kesesuaian model terhadap data (*goodness of fit*) dan kompleksitas model yang diukur melalui jumlah parameter. Secara umum, AIC didefinisikan sebagai berikut:

$$AIC = -2\ln\hat{L} + 2M$$

maksimum fungsi *likelihood* model.

Kriteria AIC dalam model *time series*, terutama model nonlinier *Self-Exciting Threshold Autoregressive* (STAR), sering mengalami kendala akibat tidak tersedia secara langsung pada aplikasi pengolahan data secara eksplisit. Solusi dari kendala ini adalah dengan melakukan AIC dengan mempertimbangkan asumsi distribusi *residual*.



Residual dalam model diasumsikan mengikuti distribusi normal identik dan independen (i.i.d), yang secara matematis dinyatakan sebagai:

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Berdasarkan asumsi tersebut, fungsi *likelihood* untuk T observasi dapat dinyatakan sebagai hasil perkalian dari distribusi normal *residual* di setiap waktu t . Bentuk fungsi *likelihood* tersebut dinyatakan sebagai:

$$L = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right)$$

Fungsi *likelihood* ini kemudian ditransformasi ke dalam bentuk logaritmik guna menyederhanakan perhitungan. Penerapan logaritma natural terhadap fungsi *likelihood* menghasilkan bentuk *log-likelihood* sebagai berikut:

$$\ln L = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2$$

Tahap selanjutnya adalah menggantikan varians *residual* σ^2 dengan estimasi varians *residual* berbasis data sampel, yang dihitung melalui rata-rata kuadrat *residual*:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2$$

Sehingga diperoleh:

$$\sum_{t=1}^T \varepsilon_t^2 = T \hat{\sigma}_\varepsilon^2.$$

Substitusi nilai estimasi varians *residual* ke dalam persamaan *log-likelihood* menghasilkan *log-likelihood* maksimum berikut:

$$\ln \hat{L} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \hat{\sigma}_\varepsilon^2 - \frac{T}{2}$$

karena:

$$\frac{1}{2\hat{\sigma}_\varepsilon^2} T \hat{\sigma}_\varepsilon^2 = \frac{T}{2}.$$

Substitusi *log-likelihood* maksimum ke dalam formulasi AIC:

$$\begin{aligned} AIC &= -2 \ln \hat{L} + 2M \\ &= -2 \left(-\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\hat{\sigma}_\varepsilon^2) - \frac{T}{2} \right) + 2M \\ &= T \ln(2\pi) + T \ln(\hat{\sigma}_\varepsilon^2) + T + 2M \end{aligned}$$

Konstanta $T \ln(2\pi) + T$ dapat diabaikan dalam perbandingan model karena semua model yang dibandingkan. Oleh karena itu, rumus AIC dalam penelitian ini dijabarkan pada Persamaan (22).

$$AIC = T \ln \hat{\sigma}_\varepsilon^2 + 2M \quad (22)$$

yang dipengaruhi oleh besar kecilnya varians *residual*. Varians yang kecil akan menghasilkan AIC bernilai negatif, sedangkan varians yang besar akan menghasilkan AIC bernilai positif (Burnham & Anderson, 2002).



1.4.15 Korelasi

Korelasi merupakan salah satu metrik statistik untuk mengevaluasi kinerja model peramalan dalam mencerminkan pola pergerakan data aktual. Metrik ini mengidentifikasi kebaikan model dalam menghasilkan nilai peramalan yang selaras dengan arah perubahan nilai sesungguhnya. Salah satu metode yang umum diterapkan adalah korelasi *Pearson*, dengan rumus dituliskan pada Persamaan (23):

$$\rho = \frac{\sum_{t=1}^T (Y_t - \bar{Y})(\hat{Y}_t - \bar{\hat{Y}})}{\sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2} \sqrt{\sum_{t=1}^T (\hat{Y}_t - \bar{\hat{Y}})^2}} \quad (23)$$

dengan $\hat{Y}_t$ adalah nilai hasil peramalan dari model dan $\bar{\hat{Y}}$ adalah rata-rata nilai peramalan. Koefisien korelasi ρ berada dalam rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa hasil peramalan sesuai dengan arah dan perubahan dari nilai aktual. Semakin tinggi nilai korelasi, semakin baik model dalam mengikuti pola data sesungguhnya (Sari et al., 2023).

1.4.16 Root Mean Square Error

Root Mean Square Error (RMSE) adalah salah satu metrik yang digunakan untuk mengukur seberapa besar rata-rata residual antara nilai aktual dan nilai peramalan. Penggunaan RMSE sangat umum digunakan dalam analisis, baik di bidang statistika maupun *machine learning* dan dianggap sebagai salah satu metrik kesalahan yang efektif untuk prediksi numerik secara umum. Nilai RMSE yang lebih kecil menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dalam peramalan. RMSE dapat dihitung berdasarkan persamaan berikut:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2}$$

RMSE lebih tepat digunakan untuk membandingkan kesalahan prediksi antar model atau antar konfigurasi model pada variabel yang sama. Hal ini disebabkan karena RMSE bersifat bergantung pada skala data, sehingga besar kecilnya nilai RMSE sangat dipengaruhi oleh satuan dari variabel yang diprediksi. Jika digunakan pada variabel yang berbeda, perbandingan nilai RMSE menjadi tidak valid karena masing-masing variabel memiliki skala pengukuran yang berbeda. Oleh sebab itu, dalam membandingkan model pada variabel yang berbeda, disarankan menggunakan metrik yang tidak bergantung pada skala, seperti MAPE, yang mengukur kesalahan dalam bentuk persentase (Elsaraiti & Merabet, 2021).

1.4.17 Deteksi Data Ekstrem



akan nilai data observasi yang nilainya menyimpang dari nilai rata-rata ini berada di luar rentang yang diharapkan berdasarkan standar deviasi dari distribusi data, dan dapat diidentifikasi sebagai *outlier*). Keberadaan data ekstrem dapat memengaruhi struktur data sehingga menghasilkan estimasi parameter yang tidak akurat. Karakteristik populasi sebenarnya. Salah satu pendekatan yang digunakan dalam mendeteksi data ekstrem adalah metode ambang

batas berbasis simpangan baku (*standard deviation thresholding*), yang mengasumsikan bahwa data berdistribusi normal. Nilai perubahan antar waktu dihitung sebagai selisih antara dua periode berturut-turut melalui Persamaan (24):

$$\Delta_t = Y_t - Y_{t-1} \quad (24)$$

Ambang batas deteksi ditentukan menggunakan rata-rata perubahan ($\mu\Delta$) dan simpangan baku ($\sigma\Delta$) sebagai berikut:

$$\mu\Delta \pm k \sigma\Delta$$

Nilai ambang $k = 3$ digunakan dengan mengacu pada *Three Sigma Rule*, yang menyatakan bahwa dalam distribusi normal, sekitar 99,7% data berada dalam rentang tiga simpangan baku dari rata-rata. Oleh karena itu, observasi yang terletak di luar rentang ini diklasifikasikan sebagai data ekstrem karena hanya mewakili sekitar 0,3% dari distribusi (Pukelsheim, 1994). Batas deteksi data ekstrem dirumuskan melalui Persamaan (25) berikut:

$$|\Delta_t| > \mu\Delta \pm 3\sigma\Delta \quad (25)$$

1.4.18 Saham Bank Central Asia

Saham merupakan salah satu instrumen investasi yang memberikan hak kepemilikan terhadap suatu perusahaan. Investor yang membeli saham memperoleh kepemilikan atas sebagian dari perusahaan tersebut, sehingga memiliki hak atas pembagian keuntungan perusahaan, baik dalam bentuk dividen maupun melalui potensi *capital gain* dari kenaikan harga saham (Suryaman & Hindriari, 2021). Di antara berbagai saham yang diperdagangkan di Bursa Efek Indonesia (BEI), PT Bank Central Asia Tbk (BBCA) merupakan salah satu saham unggulan dengan kapitalisasi pasar terbesar. Keunggulan fundamental serta reputasi BBCA yang kuat menjadikannya sebagai salah satu saham *blue chip* yang diminati oleh masyarakat (Nugroho et al., 2021).

Salah satu alasan utama tingginya minat investor terhadap saham BBCA tidak hanya terletak pada fundamental keuangan yang kuat, tetapi juga pada karakteristik pergerakan harganya yang relatif stabil di pasar modal. Berbeda dengan saham perbankan lainnya yang cenderung lebih volatil, saham BBCA menunjukkan kecenderungan untuk mengalami kenaikan harga secara bertahap dan jarang menunjukkan lonjakan yang ekstrem dalam jangka pendek (Sadikin & Agustina, 2023). Stabilitas ini mencerminkan kepercayaan pasar terhadap kinerja dan prospek jangka panjang BBCA. Oleh karena itu, banyak investor memilih saham BBCA sebagai bagian dari portofolio investasi jangka panjang (Wibiksana, 2023).

Pada analisis saham, harga penutupan (*close price*) merupakan salah satu indikator utama karena mencerminkan harga terakhir saham diperdagangkan dalam satu periode tertentu. Harga penutupan sering digunakan sebagai dasar dalam



am karena lebih stabil dibandingkan harga *intraday* dan lebih di pasar terhadap informasi yang tersedia (Julian & Pribadi, itu, dalam penelitian ini, data yang digunakan adalah harga ce) sebagai dasar dalam membentuk model *time series*.

1.4.19 Return

Return saham adalah indikator utama dalam analisis keuangan yang menggambarkan tingkat keuntungan atau kerugian dari suatu investasi saham dalam periode tertentu. Dalam analisis keuangan berbasis *time series*, *return* lebih sering digunakan dibandingkan harga saham itu sendiri karena *return* lebih mencerminkan pola fluktuasi yang terjadi di pasar dan cenderung lebih stasioner (Amri & Ramdani, 2020). Terdapat dua metode utama dalam perhitungan *return*, yaitu *simple return* dan *log return*. *Simple return* dihitung sebagai perubahan harga saham antara dua periode berturut-turut, yang secara matematis dapat dinyatakan sebagai berikut

$$R_t^s = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

Sedangkan *log return* dihitung sebagai perubahan logaritmik dari harga saham antar periode, yang dirumuskan pada Persamaan (26) berikut

$$R_t = \ln\left(\frac{Y_t}{Y_{t-1}}\right) \quad (26)$$

Dalam analisis deret waktu keuangan, *log return* lebih banyak digunakan karena memiliki sifat yang lebih stabil dan cenderung memenuhi asumsi stasioneritas, sehingga penerapan model statistik diharapkan menjadi akurat dan valid. Oleh karena itu, dalam penelitian ini perhitungan *return* menggunakan metode *log return*. Dalam praktiknya, perhitungan *return* menyebabkan jumlah observasi menjadi berkurang satu, karena nilai *return* pada waktu t dihitung berdasarkan selisih terhadap nilai pada waktu $t - 1$. Dengan demikian, apabila jumlah data harga saham adalah T , maka jumlah data *return* yang dihasilkan adalah $T - 1$.



BAB II METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa harga saham penutupan (*Close*) BBKA yang diperoleh dari situs finance.yahoo.com. Data mencakup periode 2 Januari 2014 hingga 29 Desember 2024 dan ditransformasi menjadi data *return* untuk keperluan analisis.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah data *return* harian saham BBKA, yang dihitung dari *close price* menggunakan metode *log return*. Data harga saham dibagi menjadi dua bagian, yaitu data *training* sebanyak 2.482 observasi (periode 2 Januari 2014 hingga 29 Desember 2023) dan data *testing* sebanyak 237 observasi (periode 2 Januari 2024 hingga 30 Desember 2024). Perhitungan *return* menggunakan selisih logaritma dari harga penutupan pada dua waktu berturut-turut, sehingga jumlah observasi *return* pada data *training* berkurang satu menjadi 2.481. Pembagian ini bertujuan untuk membangun model peramalan dan mengevaluasi performanya terhadap data *testing*.

2.3 Metode Analisis

Langkah-langkah analisis ini dirancang untuk memperoleh hasil yang sesuai dengan tujuan penelitian. Berikut adalah tahapan analisis yang akan dilakukan:

1. Melakukan analisis deskriptif terhadap data harga penutupan saham
 - a. Melakukan analisis deskriptif menggunakan grafik runtun waktu pada data harga saham untuk mengamati pola tren, musiman, serta potensi data ekstrem. Deteksi data ekstrem dilakukan menggunakan aturan *three sigma rule*.
 - b. Menghitung nilai statistik deskriptif, seperti rata-rata, simpangan baku, nilai minimum, dan maksimum untuk memahami karakteristik penyebaran data.
2. Melakukan uji stasioneritas pada data *return* harga saham
 - a. Membuat grafik runtun waktu untuk data *return* guna mengamati pola data dan indikasi stasioneritas.
 - b. Menggunakan uji *Augmented Dickey-Fuller* (ADF) untuk mendeteksi keberadaan unit *root* dalam data.
 - c. Menerapkan proses *differencing* apabila data belum stasioner hingga memenuhi syarat kestasioneran secara statistik.
3. Mengestimasi model *Autoregressive* (AR) pada data *return* yang sudah



alisis grafik *Autocorrelation Function* (ACF) dan *Partial Correlation Function* (PACF) untuk mengidentifikasi struktur korelasi dalam data serta menentukan orde potensial dari model AR. Beberapa model AR dengan kombinasi *lag* yang berbeda akan dihasilkan dari hasil analisis ACF dan PACF.

- c. Mengestimasi parameter dari masing-masing model AR.
- d. Melakukan uji signifikansi terhadap parameter model untuk memastikan bahwa hanya parameter yang signifikan secara statistik yang dipertahankan dalam model.
- e. Membandingkan seluruh model yang memiliki parameter signifikan menggunakan nilai *Akaike Information Criterion* (AIC), kemudian memilih model dengan nilai AIC terkecil sebagai model AR terbaik.
- f. Menentukan jumlah *lag p* dari model terbaik sebagai nilai *lag m* ($lag\ p = lag\ m$) yang akan digunakan dalam proses pemodelan lanjutan dengan model *Smooth Transition Autoregressive* (STAR).
4. Melakukan uji nonlinieritas pada data *residual* model AR
 - a. Menerapkan uji *Lagrange Multiplier* (LM) untuk mengidentifikasi keberadaan unsur nonlinier dalam struktur data.
 - b. Melanjutkan ke tahap pemodelan STAR apabila hasil uji menunjukkan adanya karakteristik nonlinier yang signifikan.
5. Membangun model STAR pada data *return*
 - a. Menetapkan nilai *delay* optimal untuk menggambarkan waktu peralihan antar *regime* secara representatif.
 - b. Menentukan fungsi transisi yang sesuai, yaitu antara fungsi logistik (LSTAR) atau eksponensial (ESTAR), berdasarkan hasil regresi bantu dan pengujian parameter.
 - c. Mengestimasi parameter model STAR dengan menggunakan metode *Nonlinear Least Squares* (NLS) yang sesuai untuk model nonlinier.
 - d. Menggunakan model STAR yang telah terestimasi untuk melakukan peramalan data *return* pada periode berikutnya.
6. Melakukan evaluasi terhadap performa model STAR
 - a. Menghitung nilai *Akaike Information Criterion* pada data *training* untuk menilai efisiensi model dalam menyeimbangkan akurasi dan kompleksitas.
 - b. Menghitung nilai *Root Mean Square Error* pada data *training* untuk mengukur besarnya kesalahan prediksi secara absolut terhadap data aktual.
 - c. Menghitung nilai *Mean Absolute Percentage Error* pada data *testing* untuk menilai tingkat kesalahan relatif antara hasil peramalan dan nilai aktual.
 - d. Menghitung nilai korelasi *Pearson* pada data *testing* untuk mengevaluasi pola pergerakan antara data peramalan dan data aktual.

