ESTIMASI INDEKS HARGA SAHAM GABUNGAN BERDASARKAN MODEL REGRESI KUANTIL *PENALIZED SPLINE*

AZALIA FILADELFIA PAGALO
H051201049

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR

2024

ESTIMASI INDEKS HARGA SAHAM GABUNGAN BERDASARKAN MODEL REGRESI KUANTIL *PENALIZED SPLINE*

AZALIA FILADELFIA PAGALO
H051201049

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

ESTIMASI INDEKS HARGA SAHAM GABUNGAN BERDASARKAN MODEL REGRESI KUANTIL *PENALIZED SPLINE*

AZALIA FILADELFIA PAGALO
H051201049

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

SKRIPSI

ESTIMASI INDEKS HARGA SAHAM GABUNGAN BERDASARKAN MODEL REGRESI KUANTIL *PENALIZED SPLINE*

AZALIA FILADELFIA PAGALO
H051201049

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada tanggal 18 Agustus 2024 dan dinyatakan telah memenuhi syarat kelulusan pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing Tugas Akhir,

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

Mengetahui:
Ketua Program Studi,

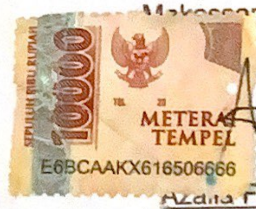
Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Estimasi Indeks Harga Saham Gabungan Berdasarkan Model Regresi Kuantil *Penalized Spline*" adalah benar karya saya dengan arahan dari pembimbing Dr. Anna Islamiyati, S.Si., M.Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbirkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 21 Agustus 2024



Azalia Filadelfia Pagalo
NIM H051201049

UCAPAN TERIMA KASIH

Segala puji dan syukur kepada Tuhan Yang Maha Esa yang oleh kasih karunia, dan penyertaan-Nya di setiap saat, sehingga penulis dapat menyelesaikan skripsi dengan judul **"Estimasi Indeks Harga Saham Gabungan Berdasarkan Model Regresi Kuantil *Penalized Spline*"**. Segala kemuliaan hanya bagi-Nya yang telah menuntun dan menguatkan saya dalam setiap langkah yang saya tempuh.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Pembimbing Utama yang dengan sabar dan tulus meluangkan waktu di tengah berbagai kesibukannya dalam membimbing, memberi masukan, dan motivasi dalam penulisan skripsi ini. Terima kasih kepada **Ibu Dra. Nasrah Sirajang, M.Si.** selaku Penasehat Akademik sekaligus Penguji yang senantiasa membimbing dan mengarahkan penulis selama berada di bangku perkuliahan. Terima kasih kepada **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku penguji yang telah memberikan pengetahuan, saran serta kritikan yang membangun dalam penyempurnaan penulisan skripsi ini. Terima kasih kepada **Pimpinan Universitas Hasanuddin**, **Departemen Statistika**, **Jajaran Dosen**, dan **Staf Departemen Statistika** yang telah membekali ilmu dan memberikan fasilitas yang melancarkan penulis dalam menempuh masa studi.

Dengan rasa hormat, penulis mengucapkan terima kasih kepada kedua orang tua, Ibunda **Johana** dan Ayahanda **Reki Pagalo** (almarhum) atas pendidikan, nasehat, dukungan materi, cinta dan kasih sayang, serta doa yang tak pernah berhenti mengiringi dalam setiap langkah penulis. Terima kasih juga kepada adik tersayang **Kores dan Jere**, atas segala bantuan, motivasi, dan semangat tiada henti yang telah diberikan kepada penulis. Tak lupa, penulis juga berterima kasih kepada keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini.

Terima kasih kepada teman-teman di **STATISTIKA 2020** atas kebersamaan selama menjalani pendidikan di Departemen Statistika. Juga, terima kasih kepada keluarga besar **Himastat FMIPA Unhas**, khususnya **POIS20N** atas ilmu dan pengalaman yang tak terlupakan. Penulis bangga menjadi bagian dari keluarga ini. Terima kasih kepada kakak-kakak **INTEGRAL 2018**, khususnya Kak **Fadhil** dan Kak **Juni** atas seluruh pengalaman, pembelajaran, dan motivasi yang diberikan selama berorganisasi. Kepada sahabat **HOPELESS ROMANTIC**, Uji, Bels, Ririn, Nola, dan Adel terima kasih karena telah menjadi tempat *recharged energy* bagi penulis. Dan kepada sahabat-sahabat perkuliahan penulis, **Rani**, **Uci**, **Ara**, **Naje**, **Liza**, **Ayu, Linda, Shafa, Nahda**, **Irma, Febi**, **Ayuni, Heri, Sabila**, **Nahdi**, **Isti, Dea, Gav, Indah, Halima, Razy**, **Kur**, **Ngkal, Faldi**, **Fadhil**, **Farhan**, **Hakam**, **Izzul**, **Azhar, Rivaldi**, **Iskar, Fadlan**, **Mukhlis**, **Fahmi**, **Ryan**, **Ridwan**, **Much**, **Vyn, Reza**, dan yang tak sempat saya sebutkan namanya, terima kasih atas segala momen kebersamaan, pembelajaran, dan diskusi yang terus membangun penulis.

Penulis menyadari bahwa masih banyak terdapat kekurangan dalam penyusunan skripsi ini. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Penulis,

Azalia Filadelfia Pagalo

ABSTRAK

AZALIA FILADELFIA PAGALO. **Estimasi Indeks Harga Saham Gabungan Berdasarkan Model Regresi Kuantil *Penalized Spline*** (dibimbing oleh Anna Islamiyati).

Latar Belakang. Indeks Harga Saham Gabungan (IHSG) adalah indikator yang mencerminkan kinerja keseluruhan saham yang terdaftar di Bursa Efek Indonesia. Pergerakan IHSG dipengaruhi oleh sentimen pasar global yang salah satunya adalah *Hang Seng* Indeks (HSI). Pola fluktuasi HSI yang seringkali diikuti oleh pergerakan serupa pada IHSG menyebabkan data dalam penelitian ini dianalisis menggunakan regresi kuantil *penalized spline*. Metode ini efektif untuk mengatasi data yang mengandung pencilan serta memiliki pola perubahan yang tajam dan kompleks serta dapat memodelkan hubungan variabel dependen dan independen di berbagai kuantil. **Tujuan**. Penelitian ini bertujuan untuk mengestimasi fungsi regresi kuantil menggunakan estimator *penalized spline* dan menerapkannya untuk memodelkan hubungan antara Indeks Harga Saham Gabungan dan *Hang Seng* Indeks pada kuantil $\tau = 0{,}25;\ 0{,}50;$ dan $0{,}75$. Model dengan nilai *Generalized Cross Validation* (GCV) minimum dipilih sebagai model optimal. **Metode**. Penelitian ini terdiri dari dua tahap, estimasi fungsi regresi kuantil dengan estimator *penalized spline* dan penerapan model regresi kuantil *penalized spline* pada data Indeks Harga Saham Gabungan dan *Hang Seng* Indeks. **Hasil**. Berdasarkan nilai GCV minimum, yaitu 270512.6506 diperoleh model optimal adalah model regresi kuantil *penalized spline* pada kuantil $\tau = 0{,}50$ dengan 3 titik knot.

Kata Kunci: Indeks Harga Saham Gabungan (IHSG) , Regresi Kuantil, *Spline*, Pencilan, *Generelized Cross Validation*

ABSTRACT

AZALIA FILADELFIA PAGALO. **Estimation of Composite Stock Price Index Based on Penalized Spline Quantile Regression Model** (supervised by Anna Islamiyati).

Background. The Jakarta Composite Index (JCI) is an indicator that reflects the overall performance of stocks listed on the Indonesia Stock Exchange. The movement of the JCI is influenced by global market sentiment, one of which is the Hang Seng Index (HSI). The pattern of HSI fluctuations that is often followed by similar movements in the JCI causes the data in this study to be analyzed using penalized spline quantile regression. This method is effective for dealing with data that contains outliers and has sharp and complex patterns of change and can model the relationship between dependent and independent variables in various quantiles. **Objective**. This study aims to estimate the quantile regression function using the penalized spline estimator and apply it to model the relationship between the Composite Stock Price Index and the Hang Seng Index at quantiles $\tau = 0{,}25;\ 0{,}50;$ and $0{,}75$. The model with the minimum Generalized Cross Validation (GCV) value is selected as the optimal model. **Methods**. This study consists of two stages, estimation of quantile regression function with penalized spline estimator and application of penalized spline quantile regression model on data of Composite Stock Price Index and Hang Seng Index. **Results**. Based on the minimum GCV value, which is 270512.6506, the optimal model is the penalized spline quantile regression model at quantile $\tau = 0{,}50$ with 3 knot points.

Keywords: Jakarta Composite Index (JCI), Quantile Regression, *Spline*, Outliers, *Generelized Cross Validation*

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Regresi	Analisis statistik yang digunakan untuk mengidentifikasi hubungan antara variabel respon berdasarkan satu atau lebih variabel prediktor
Kuantil	Nilai yang membagi data menjadi bagian-bagian yang sama
Estimator	Metode yang digunakan untuk memperkirakan nilai parameter populasi berdasarkan data sampel
Spline	Fungsi matematika yang digunakan untuk membuat model data yang fleksibel dan dapat menangani hubungan yang kompleks antara variabel
Penalized	Fungsi yang digunakan untuk mengontrol kompleksitas, mencegah overfitting, dan mendorong model yang lebih sederhana dan efektif
Estimasi	Proses memperkirakan nilai dari parameter populasi berdasarkan data sampel
Parameter	Nilai yang menggambarkan sifat atau karakteristik suatu populasi
Pencilan	Nilai yang sangat berbeda dari sebagian besar data lainnya dalam satu set data
Distribusi	Pola penyebaran data atau probabilitas kejadian dalam ruang sampel
Parametrik	Metode yang mengasumsikan bahwa data mengikuti distribusi tertentu
Nonparametrik	Metode yang tidak mengasumsikan bentuk distribusi data
Lamda	Parameter penghalus yang digunakan untuk mengontrol tingkat regularisasi.
Heterogenitas	Variasi atau perbedaan yang ada dalam suatu kumpulan data atau populasi
Least Absolute Deviation	Metode estimasi parameter model regresi yang meminimumkan jumlah absolut deviasi
Tersegmen	Pembagian fungsi menjadi segmen-segmen yang masing-masing diwakili oleh polinomial berbeda
Titik knot	Titik-titik tertentu pada domain variabel independen di mana bentuk fungsi *spline* berubah

DAFTAR LAMBANG DAN SINGKATAN

Lambang/singkatan	Arti dan Penjelasan
y_i	Variabel respon pada pengamatan ke-i
x_i	Variabel prediktor pada pengamatan ke-i
ε_i	*Error* ke-i
x_i	Pengamatan ke-i pada variabel prediktor
$y_i(\tau)$	Variabel respon ke-i pada kuantil ke-τ
$\beta(\tau)$	Penduga parameter pada kuantil ke-τ
$\varepsilon_i(\tau)$	*Error ke-*i pada kuantil ke-τ
$\boldsymbol{y}$	Vektor kolom berukuran $n \times 1$ dari variabel respon
$\boldsymbol{X}$	Matriks berukuran $n \times (k+1)$ dari variabel prediktor
$\boldsymbol{\beta}$	Vektor kolom berukuran parameter regresi
$\boldsymbol{\varepsilon}$	Vektor kolom berukuran $n \times 1$ dari *error*
$\rho_\tau(\varepsilon)$	*Loss function* ke-τ dari ε
$Q_\tau(y\|x)$	Fungsi kuantil ke-τ dari variabel y dengan syarat x
$\hat{\beta}(\tau)$	Penduga dar parameter regresi pada kuantil ke-τ
c_j	Nilai koefisien fungsi tujuan
c_b	Nilai koefisien dari variabel yang menjadi basis
v_b	Nama-nama dari variabel yang menjadi basis
w_b	Nilai ruas kanan dari kendala
z_j	Nilai koefisien fungsi objektif
a_{ij}	Nilai koefisien kolom kunci
q	Orde polinomial *spline truncated*
K_h	Nilai titik knot ke-h
$f(x_i)$	Fungsi regresi dari pengamatan ke-i
β_{lj}	Parameter polinomial pada orde ke-l dan prediktor ke-j
$\beta_{(q+h)}$	Parameter *truncated* pada titik knot ke-$(q+h)$
K_h	Nilai titik knot ke-h
$\boldsymbol{X}[K]$	Matriks berukuran $n \times (1+q+qr)$ dari model yang bergantung pada titik knot
$\boldsymbol{\tau}$	Merepresentasikan tingkat atau persentil dari kuantil yang diinginkan.
$\boldsymbol{\lambda}$	Regularisasi dalam model, mengontrol seberapa kuat penalti diterapkan untuk kompleksitas model
OLS	*Ordinary Least Square*
MSE	*Mean Square Error*
GCV	*Generelized Cross Validation*
LAD	*Least Absolute Deviation*

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

Lampiran **Halaman**

BAB I
PENDAHULUAN

1.1. Latar Belakang

Indeks Harga Saham Gabungan (IHSG) adalah indeks gabungan yang berasal dari seluruh jenis saham yang tercatat di Indonesia. IHSG sering digunakan oleh para investor untuk melihat perkembangan pasar saham di Indonesia karena merupakan indikator pergerakan saham yang tercatat di Bursa Efek Indonesia (BEI). Pergerakan IHSG ini akan mempengaruhi para investor untuk membeli, menahan, atau menjual sahamnya. Hang Seng Indeks (HSI) sebagai salah satu indikator utama di pasar Asia seringkali memiliki pengaruh signifikan terhadap pergerakan Indeks Harga Saham Gabungan (IHSG) terutama dalam konteks dinamika pasar global. Fluktuasi HSI yang seringkali diikuti oleh pergerakan serupa pada IHSG menunjukkan adanya korelasi yang kuat antara kedua indeks ini. Menurut penelitian Sutrisno dan Wahyudi (2019) hubungan ini disebabkan oleh keterkaitan ekonomi dan perdagangan antara negara-negara di Asia serta arus modal yang melibatkan pasar keuangan di Hong Kong dan Indonesia. Investor perlu memperhatikan pergerakan saham dan hubungan antara IHSG dan HSI karena kedua indeks ini mencerminkan dinamika ekonomi yang berbeda, yang dapat memberikan wawasan penting untuk diversifikasi dan strategi investasi. Dengan memantau hubungan antara kedua indeks ini, investor dapat memahami arus modal internasional, mengantisipasi potensi risiko dan peluang di pasar global, serta membuat keputusan investasi yang lebih terinformasi dan seimbang. Diversifikasi berdasarkan analisis pergerakan dua indeks ini juga membantu mengurangi risiko dan meningkatkan potensi pengembalian portofolio.

Oleh karena itu, analisis terhadap hubungan antara HSI dan IHSG menjadi sangat penting bagi investor dan pelaku pasar modal untuk mengantisipasi pergerakan pasar dan membuat keputusan investasi yang lebih baik. Melalui analisis teknis investor dapat memprediksi harga saham di masa depan dengan memeriksa pola dan tren pada grafik harga historis dan volume perdagangan serta menggunakan harga historis dan volume perdagangan sebagai landasan utama (Permathasari dkk., 2018).

Analisis regresi merupakan salah satu teknik analisis data dalam ilmu statistika untuk melihat pola hubungan dan pengaruh antara variabel prediktor terhadap variabel respon dengan mengestimasi kurva regresinya. Pada analisis regresi linier dapat ditemukan beberapa estimasi parameter model salah satunya adalah metode *Ordinary Least Squares* (OLS). Metode OLS merupakan metode yang umum digunakan dalam regresi dan memiliki prinsip meminimumkan jumlah kuadrat *error*. Namun, metode ini sangat sensitif terhadap *outlier* karena dapat menyebabkan hasil estimasi parameter yang bias. Pendekatan menggunakan metode OLS tidak dapat mewakili keseluruhan data dari distribusi karena didasarkan pada distribusi *mean* yang hanya menunjukkan ukuran pemusatan dari suatu distribusi. Pada analisis regresi telah dikembangkan regresi *robust* untuk menangani data yang mengandung *outlier*. Namun, metode ini terfokus untuk mengurangi pengaruh pencilan terhadap estimasi parameter dengan tetap memodelkan data hanya pada rata-rata distribusi sehingga kurang mampu menggambarkan variansi dalam berbagai bagian distribusi. Koenker (2005) memperkenalkan regresi kuantil, yaitu metode analisis regresi yang dapat menangani pencilan dan dapat memodelkan hubungan variabel dependen dan independen di berbagai kuantil.

Regresi kuantil adalah metode estimasi parameter yang tidak mudah terpengaruh oleh adanya pencilan, sehingga pencilan tidak mengganggu kestabilan data yang diperoleh (Zain, 2014). Pendugaan fungsi kuantil dari sebaran bersyarat respon dilakukan pada berbagai nilai kuantil yang diinginkan dan menggambarkan titik tertentu dari distribusi bersyarat. Regresi kuantil memiliki beberapa keunggulan dibandingkan metode OLS, salah satunya adalah fleksibilitas dalam memodelkan data dengan distribusi bersyarat heterogen (Puteri dkk., 2020).

Kajian penelitian dengan menggunakan regresi kuantil dapat digunakan pada kasus regresi parametrik dengan asumsi bahwa bentuk fungsional yang digunakan untuk model regresi telah diketahui sebelumnya, seperti bentuk linier, kuadratik, atau kubik. Namun, tidak semua data dapat dianalisis dengan pendekatan parametrik karena keterbatasan informasi mengenai bentuk dan pola hubungan antara variabel respon dan variabel prediktor. Regresi semi parametrik dapat mengurangi ketergantungan pada asumsi bentuk fungsi tertentu, namun regresi ini tetap memerlukan asumsi parametrik dalam mengestimasi model. Oleh karena itu, digunakan regresi nonparametrik kuantil untuk mengestimasi fungsi regresi ketika asumsi tentang bentuk kurva regresi tidak diketahui dan hanya diasumsikan mulus (*smooth*) dengan melibatkan nilai-nilai kuantil (Aprilia dkk., 2020).

Terdapat beberapa estimator pada regresi nonparametrik yang telah dikembangkan antara lain, *spline*, *kernel, wavelet,* deret *fourier*, dan polinomial lokal (Eubank, 1999). Estimator yang sering digunakan adalah estimator *spline*. *Spline* merupakan suatu potongan-potongan polinom yang memiliki sifat tersegmen dan kontinu serta memiliki orde tertentu yang saling bersambung pada titik-titik knot. Titik knot adalah titik perpaduan bersama yang terjadi karena terdapat perubahan perilaku pola pada interval yang berlainan (Aprilia dkk., 2020). Estimator *spline* terdiri dari *spline truncated, spline smoothing,* dan *spline penalized.*

Penelitian mengenai regresi kuantil nonparametrik telah banyak dilakukan diantaranya, Aprilia dkk.,(2020) yang mengestimasi parameter regresi kuantil menggunakan *spline* kuadratik. Matdoan & Balami, (2019) mengestimasi parameter regresi kuantil menggunakan fungsi *spline truncated*. Chen dkk., (2019) mengestimasi kuantil *small area* melalui regresi *spline*. Penelitian-penelitian tersebut belum menambahkan matriks penalti dalam proses analisisnya. Penerapan penalti pada *spline* bertujuan untuk mengendalikan dan menyesuaikan kompleksitas model dan kehalusan kurva serta menghindari *overfitting*.

Penelitian mengenai hubungan antara IHSG dan HSI telah banyak dilakukan sebelumnya dengan berbagai metode diantaranya Wilona, (2012) menggunakan metode korelasi *pearson* yang menunjukkan adanya korelasi positif signifikan, I Gusti, (2012) menggunakan metode analisis regresi sederhana dalam penelitiannya yang menunjukkan adanya pengaruh signifikan. Penelitian tersebut telah memberikan wawasan yang signifikan mengenai hubungan IHSG dan HSI. Namun, metode yang digunakan memiliki keterbatasan dalam menangkap hubungan yang lebih kompleks dan nonlinear antara kedua indeks. Oleh karena digunakan regresi kuantil *penalized spline* yang mampu menganalisis hubungan nonlinier antara kedua indeks dan memodelkan serta memberi gambaran perubahan HSI dapat mempengaruhi IHSG di berbagai kuantil dari distribusi hasil. Oleh sebab itu, dalam penelitian ini model regresi kuantil *penalized spline* kemudian akan diaplikasikan pada data Indeks Harga Saham Gabungan dan *Hang Seng* Indeks.

1.2. Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, batasan masalah pada penelitian ini adalah:

1. Data penelitian yang digunakan adalah kasus adalah data Indeks Harga Saham Gabungan (IHSG) dan *Hang Seng* Indeks (HSI) yang diperoleh dari publikasi Indonesia Stock Excange melalui situs www.idx.co.id dan www.investing.com dengan kurun waktu Januari 2020-Desember 2023.
2. Nilai kuantil yang digunakan terdiri dari τ = 0,25; 0,50; dan 0,75.
3. Parameter penghalus yang digunakan yaitu interval dari 0,5 hingga 1

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter Regresi Kuantil *Penalized Spline* yang bersesuaian dengan data Indeks Harga Saham Gabungan.
2. Memperoleh hubungan antara Indeks Harga Saham Gabungan dan *Hang Seng* Indeks berdasarkan model Regresi Kuantil *Penalized Spline*.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan informasi tentang estimasi parameter Regresi Kuantil *Penalized Spline* yang bersesuaian dengan data Indeks Harga Saham Gabungan.
2. Hasil analisis dapat menjadi referensi bagi regulator pasar modal dalam merumuskan kebijakan yang mendukung stabilitas dan pertumbuhan pasar saham Indonesia.
3. Sebagai inspirasi bagi pembaca untuk mengembangkan analisis regresi nonparametrik dengan metode lain dan/atau kasus lain yang dapat ditemukan dalam kehidupan sehari-hari.

1.5. Landasan Teori

1.5.1. Regresi Nonparametrik

Regresi nonparametrik merupakan analisis regresi yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor yang tidak diketahui fungsinya, hanya diasumsikan *smooth* (mulus). Kurva regresi hanya diasumsikan *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu. Regresi nonparametrik merupakan regresi yang sangat fleksibel dalam memodelkan pola data (Eubank, 1999).

Jika diberikan pasangan data (x_i, y_i) dengan $i = 1,2,3, \ldots, n$, dan pola hubungan antar variabel respon dan variabel prediktor tidak diketahui bentuknya, maka dapat digunakan pendekatan nonparametrik. Model regresi nonparamterik secara umum adalah sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i, i = 1,2,\dots,n \tag{1}$$

dengan y_i variabel respon pada pengamatan ke- i, $f(x_i)$ merupakan persamaan kurva regresi dengan x_i sebagai variabel prediktor pada pengamatan ke-i sedangkan ε_i merupakan *error* yang berdistribusi normal, independen dengan *mean* 0 dan variansi σ^2 (Eubank, 1999).

1.5.2. Regresi Kuantil

Regresi kuantil pertama kali diperkenalkan oleh Koenker dan Basset pada tahun 1978. Regresi kuantil adalah teknik statistika yang digunakan untuk menduga hubungan antara peubah respon dengan peubah prediktor pada fungsi kuantil bersyarat tertentu. Regresi kuantil meminimumkan jumlah galat mutlak terboboti dan menduga model dengan menggunakan fungsi kuantil bersyarat pada suatu sebaran data. Metode regresi kuantil tidak membutuhkan asumsi parametrik dan regresi kuantil sangat bermanfaat untuk menganalisis bagian tertentu dari suatu sebaran bersyarat (Buhai, 2004). Keuntungan dari regresi kuantil yaitu efesien jika sisaan tidak menyebar normal dan kekar terhadap adanya pencilan. Regresi kuantil dapat mengukur efek peubah prediktor tidak hanya di pusat sebaran data tetapi juga pada bagian atas dan bawah ekor sebaran. Metode ini sangat berguna dalam penerapan, khususnya bila nilai ekstrem merupakan permasalahan penting (Retno dkk., 2011).

Misalkan diberikan data $\{x_{i1}, x_{i2}, \dots, x_{ik}, y_i\}$, dengan $i = 1,2,\dots,n$ dan $j = 1,2,\dots,k$ adalah himpunan berpasangan variabel acak yang berdistribusi secara independen dan identik dengan kuantil $\tau \epsilon (0,1)$. Data tersebut memiliki fungsi distribusi peluang bersyarat $F(y|x_i) = P(Y \le y|x_i)$ dan fungsi invers $F^{-1}(\tau) = \inf\{y: F(y) \ge \tau\}$ yang merupakan kuantil ke-τ didefinisikan sebagai $Q_\tau(y) = \inf\{y: F(y) \ge \tau\} = F^{-1}(\tau)$ yang merupakan kuantil ke-τ dari variabel respon y (Davino, Furno, 2017). Persamaan umum regresi kuantil linier khusus untuk kuantil bersyarat dari variabel respon y_i yaitu:

$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau) \tag{2}$$

dengan $i = 1,2,\dots,n$, $j = 1,2,\dots,k$.

$y_i(\tau)$: variabel respon pada pengamatan ke- i

β_0 : konstanta parameter

τ : nilai kuantil (0,1)

$\beta_{(\tau)}$: penduga parameter pada kuantil ke-τ

x_{ij} : pengamatan ke-i pada variabel prediktor ke-j

ε_i : *error* ke-i pada kuantil ke-τ

Jika model regresi kuantil disajikan dalam bentuk matriks, Persamaan (2) dapat ditulis seperti pada Persamaan berikut:

$$\begin{bmatrix} y_1(\tau) \\ y_2(\tau) \\ \vdots \\ y_n(\tau) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_k(\tau) \end{bmatrix} \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \tag{3}$$

Selanjutnya Persamaan (3) dapat ditulis dalam bentuk model linier seperti Persamaan berikut:

$$\boldsymbol{y}(\tau) = \boldsymbol{X\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \tag{4}$$

dengan:

$\boldsymbol{y}(\tau)$: vektor kolom berukuran $n \times 1$ dari variabel respon y

$\boldsymbol{X}$: matriks berukuran $n \times (k+1)$ dengan baris n merupakan observasi pada kolom k variabel prediktor ke-j dengan $j = 1,2, \dots k$

$\boldsymbol{\beta}$: vektor kolom berukuran $(k + 1) \times 1$ dari parameter β_i dengan $j = 1,2, \dots k$

$\boldsymbol{\varepsilon}$: vektor kolom berukuran $n \times 1$ dari *error* ε_i

Jika fungsi bersyarat dari kuantil ke−τ dengan variabel independen x tertentu, maka fungsi bersyarat tersebut didefiniskan dalam Persamaan berikut:

$$Q_\tau(y_i|x_{i1}, x_{i2}, \dots, x_{ik}) = Q_\tau(y|X) = X_i^T \beta(\tau) \tag{5}$$

Koefisien β pada Persamaan (5) dapat diestimasi dengan meminimumkan fungsi objektif kuantil seperti pada Persamaan berikut:

$$\underset{\beta \ \in \ \mathbb{R}^p}{min} \sum\nolimits_{i=1}^{n} \rho_\tau(y_i - X_i^T \beta) \tag{6}$$

dengan yi = $\{y_1, y_2, \dots, y_n\}$ merupakan sampel acak dengan variabel dependen Y dan $x_i \in \mathbb{R}^p$ merupakan vektor kovariat, sedangkan ρ_τ merupakan fungsi *loss* (Balami, 2017).

1.5.3. Estimasi Parameter Regresi Kuantil

Estimasi parameter dalam regresi OLS, hanya dapat digunakan untuk memberi solusi pada masalah *mean* sehingga Koenker & Bassett (1978) mengembangkan metode alternatif yaitu regresi kuantil. Regresi dengan metode OLS diestimasi dengan meminimumkan jumlah absolut *error* yang lebih dikenal dengan *Least Absolut Deviation* (LAD).

Pada regresi kuantil, *error* diberi bobot yang berbeda. Bobot yang digunakan yaitu τ untuk nilai *error* yang lebih besar atau sama dengan nol, dan $1 - \tau$ untuk *error* yang kurang dari nol. Perkalian antara *error* dengan bobot yang diberikan disebut fungsi *loss* ρ_τ, yang dinyatakan pada Persamaan berikut:

$$\rho_\tau(\varepsilon) = \sum\nolimits_{i=1, \varepsilon_i \geq 0}^{n} \tau|\varepsilon_i| + \sum\nolimits_{i=1, \varepsilon_i < 0}^{n} (1 - \tau)|\varepsilon_i| \tag{7}$$

Dengan demikian, dalam regresi kuantil terdapat fungsi kuantil ke-τ dari variabel y dengan syarat x yang mempertimbangkan penduga $\beta(\tau)$, sehingga diperoleh solusi untuk permasalahan tersebut yang dinyatakan pada Persamaan berikut:

$$\underset{\beta \in \mathbb{R}^p}{min} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \underset{\beta \in \mathbb{R}^p}{min} \sum_{i=1}^{n} \rho_\tau(y_i - Q_\tau(y|x)) \quad (8)$$

dengan:

y_i : variabel respon ke- i

$\rho_\tau(\varepsilon)$: fungsi loss

τ : indeks kuantil dengan $\tau \in (0,1)$

$Q_\tau(y|x)$: fungsi kuantil ke- τ dari variabel y dengan syarat x

Fungsi untuk kuantil bersyarat $Q_\tau(y|x)$ didefinisikan sebagai berikut:

$$Q_\tau(y|x) = \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau) \quad (9)$$

Pada regresi kuantil, parameter $\hat{\beta}(\tau)$ diperoleh dengan meminimumkan fungsi *loss* dari Persamaan (8) seperti pada Persamaan berikut:

$$\hat{\beta}(\tau) = \underset{\beta}{min} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \underset{\beta}{min} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)) \quad (10)$$

dengan $\rho_\tau(\varepsilon)$ pada Persamaan (10) didefinisikan

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon, & \varepsilon \geq 0 \\ (1-\tau)\varepsilon, & \varepsilon < 0 \end{cases}$$

Selanjutnya, akan dibuktikan bahwa fungsi *loss* berbentuk asimetris dengan penjelasan sebagai berikut:

$$\rho_\tau(\varepsilon) = \left[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\right] = [\tau - I(\varepsilon < 0)]\varepsilon \quad (11)$$

Dengan

$$I(\varepsilon \geq 0) = \begin{cases} 1, \varepsilon \geq 0 \\ 0, \varepsilon < 0 \end{cases}$$

ε : residual dari estimasi parameter

$I(\varepsilon)$: fungsi indikator yang terdefinisikan

Sehingga dapat dibuktikan:

$$\rho_\tau(\varepsilon) = \begin{cases} \tau_\varepsilon, \varepsilon \geq 0 \\ (1-\tau)_\varepsilon < 0 \end{cases}$$

1. Untuk $\varepsilon \geq 0$

   $\rho_\tau = \left[\tau I(\varepsilon \geq 0) + (1-\tau) I\big((\varepsilon < 0)\big)\right]|\varepsilon|$

$= [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]\varepsilon$
$= [\tau I + (1-\tau)I(\varepsilon < 0)]\varepsilon$
$= [\tau + I(\varepsilon < 0) - \tau I(\varepsilon < 0)]\varepsilon$
$= [\tau + (1 - I(\varepsilon \geq 0)) - \tau(1 - I(\varepsilon \geq 0))]\varepsilon$
$= [\tau + (1-1) - \tau(1-1)]\varepsilon$
$= \tau\varepsilon$

atau

$\rho_\tau = [\tau - I(\varepsilon < 0)]\varepsilon$
$= \left[\tau - \left(1 - I(\varepsilon \geq 0)\right)\right]\varepsilon$
$= [\tau - (1-1)]\varepsilon$
$= \tau\varepsilon$

2. Untuk $\varepsilon < 0$

$\rho_\tau = [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]|\varepsilon|$
$= \left[\tau I(\varepsilon \geq 0) + (1-\tau)I\left((\varepsilon < 0)\right)\right](-\varepsilon)$
$= [\tau 0 + (1-\tau)I(\varepsilon < 0)](-\varepsilon)$
$= [(\tau - 1)I(\varepsilon < 0)]\varepsilon$
$= \left[(\tau-1)\left(1 - I(\varepsilon < 0)\right)\right]\varepsilon$
$= [(\tau-1)(1-0)]\varepsilon$
$= (\tau - 1)\varepsilon$

atau

$\rho_\tau = [\tau - I(\varepsilon < 0)]\varepsilon$
$= \left[\tau - \left(1 - I(\varepsilon \geq 0)\right)\right]\varepsilon$
$= [\tau - (1-0)]\varepsilon$
$= (\tau - 1)\varepsilon$

Sehingga menjadi

$$\rho_\tau = [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon, \forall_\varepsilon$$

Apabila y merupakan fungsi x yang diketahui dan memiliki fungsi probabilitas $F_{y|x}(y)$, maka kuantil ke-τ dari fungsi tersebut dapat dituliskan sebagai berikut.

$$\min_\beta \tau \int_{i=1;\varepsilon_i \geq 0}^{n} |\varepsilon_i| dF_y(y) + (1-\tau) \int_{i=1;\varepsilon_i < 0}^{n} |\varepsilon_i| dF_y(y) \qquad (12)$$

dengan mempertimbangkan $\hat{\beta}(\tau)$, sehingga diperoleh solusi untuk permasalahan tersebut yang dinyatakan pada Persamaan berikut:

$$\hat{\beta}(\tau) = \min_\beta \left\{ \tau \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)| + (1-\tau) \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)| \right\} \qquad (13)$$

Solusi dari Persamaan (10) tidak dapat diperoleh secara analitik, tetapi secara numerik. Salah satu metode numerik yang digunakan adalah algoritma simpleks. Algoritma simpleks merupakan metode yang dikembangkan oleh Barrodale dan Robert pada tahun 1974. Algoritma simpleks memberikan solusi permasalahan program linier yang melibatkan beberapa variabel keputusan dengan bantuan komputasi (Davino, Furno, 2017).

Adapun beberapa istilah yang terdapat dalam algoritma simpleks adalah sebagai berikut.

1. Variabel *Slack*

   Variabel *slack* berfungsi untuk menampung sisa kapasitas pada kendala yang berupa pembatas.
2. Variabel *Surplus*

   Variabel *surplus* berfungsi untuk menampung kelebihan nilai ruas kiri pada kendala yang berupa syarat. Pada kasus regresi kuantil dengan menggunakan metode LAD, variabel *surplus* adalah deviasi bawah yang diboboti dengan $(1-\tau)$.
3. Variabel *Artificial*

   Variabel *artificial* adalah variabel positif yang berfungsi untuk memulai penyelesaian dan harus dijadikan nol pada solusi akhir. Variabel ini digunakan untuk setiap persamaan yang tidak memiliki variabel basis. Pada kasus regresi kuantil dengan menggunakan metode LAD, variabel *artificial* adalah deviasi atas yang diboboti dengan τ.
4. Variabel Basis dan Nonbasis

   Variabel basis dan nonbasis merupakan dua terminologi penting yang akan selalu digunakan di dalam algoritma simpleks. Variabel basis adalah variabel yang bernilai positif dan variabel nonbasis adalah variabel yang bernilai nol.

Algoritma simpleks memerlukan sebuah tabel simpleks atau yang biasa dikenal dengan tabulasi simpleks seperti Tabel 1.

Tabel 1. Tabel metode simpleks untuk kasus regresi kuantil

c_j			0	0	...	0	τ	...	τ	$(1-\tau)$		...	$(1-\tau)$
c_b	v_b	w_b	x_1	x_2	...	x_n	d_{11}	...	d_{1n}	d_{1n}		...	d_{1n}
d_{11}^{+}	x_1	b_1	a_{ij}										
d_{21}^{+}	x_2	b_2											
$\vdots$	$\vdots$	$\vdots$											
d_{n1}^{+}	x_n	b_n											
z_j													
$c_j - z_j$													

Pengisian Tabel 2.1 diuraikan sebagai berikut:

a. Baris c_j diisi dengan koefisien fungsi tujuan.
b. Kolom c_b diisi dengan koefisien variabel yang menjadi basis.
c. Kolom v_b diisi dengan nama-nama variabel yang menjadi basis (variabel yang menyusun matriks identitas).
d. Kolom w_b diisi dengan nilai ruas kanan dari kendala.
e. Baris z_j diisi dengan rumus $z_j = \sum d_i a_{ij}, j = 1,2, \dots, n$.

Berikut ini akan diberikan proses algoritma simpleks, yaitu:

1. Mengubah terlebih dahulu masalah optimasi linear ke bentuk standar.
2. Menentukan kolom kunci, yaitu untuk masalah maksimum memilih $c_j - z_j$ yang terbesar, sedangkan untuk masalah minimum memilih $c_j - z_j$ yang terkecil.
3. Menentukan baris kunci, yaitu dari nilai rasio antara nilai ruas kiri (b_i) dengan koefisien kolom kunci (a_{ij}), pilih yang terkecil (untuk masalah minimum atau maksimum). Rasio = $\frac{b_i}{a_{ij}}$, dengan rasio > 0.
4. Menentukan pivot dari perpotongan antara kolom kunci dan baris kunci yang dinamakan elemen kunci atau elemen penentu iterasi algoritma simpleks dan akan diubah nilainya menjadi 1.
5. Selanjutnya, melakukan operasi baris dasar (OBD) berdasarkan pivot untuk baris lainnya, termasuk baris $c_j - z_j$ dengan nilai elemen-elemen yang termasuk di dalam kolom kunci dijadikan nol (selain elemen yang dijadikan pivot).
6. Proses iterasi untuk masalah maksimum berhenti jika nilai pada baris $c_j - z_j \leq 0$, berarti solusi sudah optimal. Apabila masih ada $c_j - z_j > 0$ (positif), maka iterasi algoritma simpleks masih berlanjut. Untuk masalah minimum berhenti jika semua nilai pada baris $c_j - z_j \geq 0$. Apabila masih ada $c_j - z_j < 0$, maka iterasi algoritma simpleks masih berlanjut (Khairunnisa, 2015).

1.5.4. Spline

Spline merupakan salah satu model regresi nonparametrik yang terdiri dari potongan-potongan polinomial yang memiliki sifat tersegmen yang memiliki sifat fleksibilitas sehingga mampu mengatasi pola data yang menunjukkan nilai naik atau turun dengan bantuan knot sehingga kurva yang dihasilkan relatif mulus (Asriyanti dkk., 2022). Titik knot merupakan titik perpaduan bersama yang memperlihatkan terjadinya perubahan perilaku dari fungsi *spline* pada interval-interval yang berbeda (Eubank, 1999). Secara umum, fungsi *spline* berorde q dengan titik knot $K_1, K_2, \dots, Kr$ dapat ditulis sebagai berikut:

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h} (x_i - K_h)_+^q \tag{14}$$

dengan fungsi *truncated*

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)^q, & jika\ x_i \geq K_h \\ 0 & jika\ x_i < K_h \end{cases}$$

(Islamiyati, 2017)

dengan:

$f(x_i)$: fungsi regresi

x_i : variabel prediktor

$\beta_1, \beta_2, \dots, \beta_{q+h}$: parameter regresi

l : orde, $(l = 1,2, \dots, q)$

K_h : titik knot ke-h, $(h = 1,2, \dots, r)$

$(x_i - K_h)_+^q$: fungsi polinomial *penalized*

Dari bentuk matematis *spline* pada persamaan (14) dapat dikatakan bahwa *spline* merupakan model polinomial yang tersegmen (*piecewise* polinomial*)* walaupun *spline* bersifat kontinu pada knot-knotnya. *Spline* orde ke-*q* dapat juga diartikan sebagai model polinomial orde ke-*q* pada setiap segmennya sehingga *spline* merupaka suatu fungsi potongan polinomial yanag setiap fungsinya tergabung dengan titik knot yang menjamin sifat kontinuitas (Eubank, 1999)

1.5.5. Regresi Kuantil *Penalized Spline*

Regresi kuantil *penalized* adalah teknik statistik yang digunakan untuk memperkirakan kuantil dari distribusi variabel dependen dengan menambahkan penalti pada model regresi kuantil. Penaksir *Penalized spline* melibatkan titik simpul dan pemulusan parameter secara bersamaan dalam mengendalikan kelancaran kurva. Penaksir *penalized spline* diperoleh dengan meminimukan fungsi *Penalized Least Square* (PLS) yang merupakan kriteria estimasi dengan menggabungkan fungsi *goodness of fit* dengan *penalty*. Fungsi *goodness of fit* yang dimaksud adalah *Least Square* $\sum_{i=1}^{n}(y_i - f(x_i))$ dan fungsi *penalty* adalah fungsi yang mengontrol kemulusan kurva yaitu $\sum_{l=1}^{k}\beta_{p+l}^2$ maka diperoleh model regresi kuantil dengan *penalized spline* sebagai berikut:

$$\begin{aligned} y_i(\tau) &= f(x_1) + f(x_2) + \cdots + f(x_i) + \varepsilon_i(\tau) \\ &= \sum_{i=1}^{k} f(x_i) + \varepsilon_i(\tau) \end{aligned} \tag{15}$$

Berdasarkan persamaan (15) fungsi f dihampiri dengan fungsi regresi kuantil dengan *penalized spline* dengan orde q dan jumlah knot sebanyak r sebagai berikut:

$$f(x_i) = \sum_{l=0}^{q} \beta_l(\tau) x_i^l + \sum_{h=1}^{r} \beta_{(q+h)}(\tau)(x_i - K_h)_+^q \tag{16}$$

dengan

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)^q, jika\ x_i > K_h \\ 0, \qquad jika\ x_i < K_h \end{cases}$$

$y_i(\tau)$: respon kuantil ke-τ pada pengamatan ke-i

$f(x_i)$: fungsi regresi kuantil dengan *spline* dari pengamatan ke-i pada prediktor ke-j

x_i : pengamatan ke-i

$\beta_l(\tau)$	: parameter polinomial kuantil ke-τ pada orde ke l
$\beta_{(q+h)j}(\tau)$	: parameter *penalized* kuantil ke-τ pada titik knot ke-$(q + h)$
K_h	: nilai titik knot ke-h
r	: banyaknya titik knot
q	: orde polinomial *penalized spline*
k	: banyaknya prediktor
ε_i	: *error* kuantil ke-τ pada persamaan ke-i

Uraian fungsi 16 dinyatakan sebagai berikut:

$$\sum_{i=1}^{k} f(x_i) = f(x_1) + f(x_2) + \cdots + f(x_k) \tag{17}$$

Dengan menguraikan fungsi f dan memisahkan antara parameter dan variabel, model regresi kuantil *penalized spline* dapat dinyatakan ke dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y}(\tau) = \boldsymbol{X}[K]\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \tag{18}$$

Dengan $\boldsymbol{y}(\tau) = [y_1(\tau), y_2(\tau), \ldots, y_n(\tau)]^T$ merupakan vektor kolom berukuran $n \times 1$ dari variabel respon y pada kuantil ke-τ; $\boldsymbol{X}[K] = [1 \quad X_1 \quad \cdots \quad X_k]$ merupakan matriks $\boldsymbol{X}$ dalam bentuk *spline* dengan orde q dan r knot berukuran $n \times (1 + q + qr)$ dengan n observasi pada k variabel x; $\boldsymbol{\beta}(\tau)$ merupakan vector kolom berukuran $(1 + q + qr) \times 1$ dari parameter $\boldsymbol{\beta}$ pada kuantil ke-τ; dan $\boldsymbol{\varepsilon}(\tau)$ merupakan vektor kolom berukuran $n \times 1$ dari *error* (ε) pada kuantil ke-τ.

Penalized spline merupakan salah satu bentuk penaksir dari regresi nonparametrik yang hasil estimasinya diperoleh dengan meminimumkan *penalized least square* (PLS), yaitu kriteria estimasi model yang menggabungkan antara *goodness of fit* dengan fungsi penalti (Islamiyati, 2019). Berikut merupakan bentuk dari fungsi PLS:

$$PLS = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 + \lambda \sum_{k}^{r} \beta_{q+h}^2 \tag{19}$$

dengan λ merupakan parameter penghalus prediktor (Ruppert dkk., 2003). Sebelum meminimumkan fungsi PLS, Persamaan (19) akan diubah ke dalam bentuk matriks sebagai berikut:

$$PLS = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X_j}(K)\boldsymbol{\beta_j}\right)^T \left(\boldsymbol{Y} - \boldsymbol{X_{ji}}(K)\boldsymbol{\beta_j}\right) + \boldsymbol{\lambda_j \beta_j^T D_j \beta_j}$$

$$PLS = \frac{1}{n}\left(\boldsymbol{Y^T Y} - \boldsymbol{2\beta_j^T X_{ji}^T}(K)\boldsymbol{Y} + \boldsymbol{\beta_j^T X_{ji}^T}(K)\boldsymbol{X_{ji}}(K)\boldsymbol{\beta_j}\right) + \boldsymbol{\lambda_j \beta_j^T D_j \beta_j} \tag{20}$$

dengan D_j didefinisikan sebagai matriks diagonal sebagai berikut:

$$\boldsymbol{D_j} = \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \mathrm{I} \end{bmatrix}$$

dengan I merupakan matriks identitas untuk $\boldsymbol{\beta}_{(q+1)}, \boldsymbol{\beta}_{(q+2)}, \dots, \boldsymbol{\beta}_{(q+k)}$.

Selanjutnya menghitung nilai diferensiasi dari fungsi PLS pada Persamaan (20) terhadap $\boldsymbol{\beta_j}$ dan hasilnya disamakan dengan 0, sehingga diperoleh $\boldsymbol{\widehat{\beta}_j}$ sebagai berikut:

$$\frac{\partial PLS}{\partial \beta_j} = 0$$

$$\frac{1}{n}(0 - 2\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{Y} + 2\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{X}_{\boldsymbol{ji}}(K)\boldsymbol{\beta_j} + 2\boldsymbol{\lambda_j}\boldsymbol{D_j}\boldsymbol{\beta_j}) = 0$$

$$\boldsymbol{\widehat{\beta}_j} = \left(\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{X}_{\boldsymbol{ji}}(K) + n\boldsymbol{\lambda_j}\boldsymbol{D_j}\right)^{-1}\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{Y} \tag{21}$$

Selanjutnya mensubstitusi nilai $\hat{\beta}_j$ pada model sebagai berikut:

$$\boldsymbol{\hat{y}} = \boldsymbol{X_j}(K)\left(\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{X}_{\boldsymbol{ji}}(K) + n\boldsymbol{\lambda_j}\boldsymbol{D_j}\right)^{-1}\boldsymbol{X}_{\boldsymbol{ji}}^{\boldsymbol{T}}\boldsymbol{Y} \tag{22}$$

atau dapat juga ditulis dalam persamaan berikut:

$$\boldsymbol{\hat{y}} = \boldsymbol{H}(\boldsymbol{\lambda_j})\boldsymbol{Y} \tag{23}$$

dengan, $\boldsymbol{H}(\boldsymbol{\lambda_j}) = \boldsymbol{X}_{\boldsymbol{ji}}(K)\left(\boldsymbol{X}_{\boldsymbol{ji}}^T(K)\boldsymbol{X}_{\boldsymbol{ji}}(K) + n\boldsymbol{\lambda_j}\boldsymbol{D_j}\right)^{-1}\boldsymbol{X}_{\boldsymbol{ji}}^T$ (Ruppert, 2002).

1.5.6. Generalized Cross Validation

Salah satu tahapan penting dalam pendekatan *spline* adalah pemilihan titik knot yang optimal. Titik knot adalah titik perpaduan bersama, di mana terjadi perubahan dalam perilaku fungsi di interval yang berbeda. Pemilihan titik knot optimal diperlukan untuk menentukan model *spline* yang optimal. Salah satu metode pemilihan titik knot optimal adalah dengan menggunakan *Generalized Cross Validation* (GCV) minimum. Titik knot optimal untuk model *spline* terbaik diperoleh dari nilai GCV yang terkecil. Metode GCV dapat dituliskan sebagai berikut.

$$GCV(K) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(1 - n^{-1}tr[H(\lambda)])^2} \tag{24}$$

Dengan $K = [K_{11}, K_{12}, \dots, K_{1h}, \dots, K_{p1}, \dots, K_{pr}]$ adalah titik knot, $MSE(K) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

1.5.7. Indeks Harga Saham Gabungan dan Hang Seng Indeks

Indeks Harga Saham Gabungan (IHSG) atau sering disebut *Jakarta Composite Index* adalah suatu rangkaian informasi Historis yang mencerminkan pergerakan harga saham gabungan di Bursa Efek Indonesia. Indeks ini memberikan informasi tentang perubahan harga saham-saham yang terdaftar di Bursa Efek Indonesia dalam periode tertentu, hingga tanggal yang spesifik. IHSG berfungsi sebagai pengukuran kinerja atau performa saham-saham yang termasuk dalam indeks tersebut. Melalui IHSG, investor dan pelaku pasar dapat melihat tren dan pergerakan pasar saham secara keseluruhan. Perubahan nilai IHSG dapat memberikan indikasi tentang kondisi pasar saham, baik dalam hal kenaikan, penurunan, atau stabilnya harga saham di Indonesia (Tricahyadinata, 2016).

Hang Seng Indeks (HSI) merupakan indeks harga saham gabungan negara Hong Kong. Indeks ini digunakan untuk merekam dan memantau perubahan harian perusahaan terbesar di pasar saham Hong Kong dan sebagai indikator utama kinerja pasar secara keseluruhan di Hong Kong. *Hang Seng* Indeks berisi 40 perusahaan terbesar yang diperdagangkan di Hong Kong *Exchange*. Keseluruhan dari nilai saham-saham ini merupakan 65 persen dari nilai kapitalisasi seluruh nilai saham yang tercatat pada Hong Kong Exchange. Karena itu naik atau turunnya *Hang Seng* Indeks merupakan refleksi *performance* dari keseluruhan saham-saham yang diperdagangkan (Sari, 2012).

Indeks Harga Saham Gabungan (IHSG) dan Hang Seng Index (HSI) merupakan dua indeks pasar saham yang paling dikenal di Asia. IHSG adalah indeks saham yang mencakup seluruh perusahaan yang terdaftar di Bursa Efek Indonesia, yang meliputi berbagai sektor seperti perbankan, properti, telekomunikasi, dan infrastruktur. Contoh perusahaan besar dalam IHSG antara lain Bank Central Asia (BBCA), Telkom Indonesia (TLKM), dan Astra International (ASII). Sementara itu, HSI adalah indeks saham yang terdiri dari 50 perusahaan terdaftar di Bursa Efek Hong Kong, mewakili sekitar 58% kapitalisasi pasar. Jenis saham dalam HSI meliputi sektor keuangan, properti, teknologi, dan energi, dengan contoh perusahaan seperti HSBC Holdings (0005.HK), Tencent Holdings (0700.HK), dan China Mobile (0941.HK). Kedua indeks ini mencerminkan kondisi ekonomi negara masing-masing dan menjadi acuan bagi investor untuk menilai kinerja pasar saham lokal dan regional.

HSI dipilih karna indeks ini adalah sebuah indeks yang besar milik Hong Kong. Pergerakan *Hang Seng* sebagai salah satu indeks utama di Asia dan dunia seringkali mencerminkan sentimen pasar global. Peristiwa-peristiwa ekonomi dan politik yang memengaruhi HSI juga bisa mempengaruhi IHSG dalam hal sentimen investor global yang didalamnya termasuk investor di Indonesia. Sebagai salah satu negara tujuan ekspor Indonesia yang mencapai 1,48 miliar dollar AS per tahun 2015, pertumbuhan ekonomi Hong Kong dapat mendorong pertumbuhan ekonomi Indonesia melalui kegiatan ekspor maupun aliran modal masuk baik investasi langsung maupun melalui pasar modal. Dengan demikian, naik turunnya *Hang Seng* Indeks akan berpengaruh pada pergerakan IHSG di BEI (Syarofi, 2014). Sebuah hasil riset yang ditulis oleh Mauliano (2009) yang menyatakan bahwa *Hang Seng* Indeks mempengaruhi pergerakan IHSG. Penelitian tersebut juga didukung oleh Mutakif & Nurwulandari (2012) serta Wicaksono & Yasa (2017) yang menyatakan HSI memiliki pengaruh signifikan dan positif terhadap IHSG.