PENDUGAAN *ALMON DISTRIBUTED LAG MODELS* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA PROVINSI LAMPUNG 2005-2024

(SKRIPSI)

Oleh

Rafika Aulia Arpan

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

ABSTRACT

ESTIMATION OF ALMON DISTRIBUTED LAG MODELS ON HUMAN DEVELOPMENT INDEX INDICATORS IN LAMPUNG PROVINCE 2005-2024

By

Rafika Aulia Arpan

Almon distributed lag models are used to analyze the influence of independent variables on a dependent variable by considering lag effects in both the short and long term. This approach estimates the weights of past values of the independent variables through polynomial interpolation, allowing for a complex relationship pattern within the distributed lag model. This study aims to analyze the influence of human development index indicators on the human development index in Lampung Province during the period 2005–2024. Based on the estimation results, only the variables of expected years of schooling and per capita expenditure have a significant effect on the human development index. Therefore, re-estimation of the model is necessary. After re-estimation, the best model is obtained with a lag length of 5 and a polynomial degree of 2, as indicated by the lowest AIC and BIC values of -27.59519 and -22.48273, respectively. These two variables continue to influence the HDI in both the short and long term. They show a positive impact in the initial periods, experience a decline in the medium term, but return to having a positive effect in the long run.

Keywords: Lag distribution, Almon distributed lag models, human development index

ABSTRAK

PENDUGAAN *ALMON DISTRIBUTED LAG MODELS* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA PROVINSI LAMPUNG 2005-2024

Oleh

Rafika Aulia Arpan

Almon distributed lag models digunakan untuk menganalisis pengaruh variabel bebas terhadap variabel terikat dengan mempertimbangkan efek *lag* dalam jangka pendek dan panjang. Pendekatan ini memperkirakan bobot pengaruh dari nilai masa lalu variabel bebas melalui interpolasi polinomial, sehingga dapat menggambarkan pola hubungan yang kompleks pada model distribusi *lag*. Penelitian ini bertujuan untuk menganalisis pengaruh indikator-indikator indeks pembangunan manusia terhadap indeks pembangunan manusia di Provinsi Lampung periode 2005-2024. Berdasarkan hasil estimasi, hanya variabel harapan lama sekolah dan pengeluaran per kapita yang memberikan pengaruh signifikan terhadap indeks pembangunan manusia. Oleh sebab itu perlu melakukan re-estimasi model, setelah dilakukan re-estimasi model hasil menunjukkan bahwa model terbaik diperoleh pada panjang lag 5 dan derajat polinomial 2 karena model memiliki nilai AIC dan BIC terkecil sebesar -27,59519 -22,48273. Kedua variabel tersebut tetap memberikan pengaruh dalam jangka pendek maupun jangka panjang. Keduanya menunjukkan pengaruh positif pada periode awal, mengalami penurunan pengaruh di jangka menengah, namun kembali berdampak positif dalam jangka panjang.

Kata Kunci: Distribusi *lag, Almon distributed lag models,* indeks pembangunan manusia

PENDUGAAN *ALMON DISTRIBUTED LAG MODELS* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA PROVINSI LAMPUNG 2005-2024

Oleh

Rafika Aulia Arpan

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

Judul Skripsi : **Pendugaan *Almon Distributed Lag Models* pada Indikator Indeks Pembangunan Manusia Provinsi Lampung Tahun 2005-2024**

Nama Mahasiswa : **Rafika Aulia Arpan**

Nomor Pokok Mahasiswa : **2117031004**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Prof. Drs. Mustofa Usman, MA., Ph.D.
NIP. 195701011984031020

Siti Laelatul Chasanah, M.Si.
NIP. 199306012019032021

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Prof. Drs. Mustofa Usman, MA., Ph.D.**	....
Sekretaris	: **Siti Laelatul Chasanah, M.Si.**	....
Penguji Bukan Pembimbing	: **Dr. Khoirin Nisa, S.Si., M.Si.**	....

2. Dekan Fakultas Matematika dan Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : **15 Mei 2025**

PERNYATAAN SKRIPSI MAHASISWA

Saya bertanda tangan di bawah ini:

Nama	:	**Rafika Aulia Arpan**
Nomor Pokok Mahasiswa	:	**2117031004**
Jurusan	:	**Matematika**
Judul	:	**Pendugaan *Almon Distributed Lag Models* pada Indikator Indeks Pembangunan Manusia Provinsi Lampung Tahun 2005-2024**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 15 Mei 2025
Penulis,

METERAI TEMPEL
1AAMX289965148

Rafika Aulia Arpan

RIWAYAT HIDUP

Penulis bernama Rafika Aulia Arpan, anak pertama dari tiga bersaudara, yang lahir di Kediri, Kabupaten Pringsewu, pada tanggal 31 Januari 2003. Penulis merupakan putri dari pasangan Bapak Eka N. Arpan dan Ibu Fitri Margiyati.

Penulis memulai pendidikan di TK IKI PTPN 7 Bandar Lampung dan lulus pada tahun 2009. Pendidikan dasar dilanjutkan di SD Al-Kautsar Bandar Lampung, yang diselesaikan pada tahun 2015. Setelah itu, penulis melanjutkan ke jenjang pendidikan menengah pertama di SMP Al-Kautsar Bandar Lampung dan lulus pada tahun 2018. Pendidikan menengah atas ditempuh di SMA Al-Kautsar Bandar Lampung, dan berhasil diselesaikan pada tahun 2021.

Penulis terdaftar sebagai mahasiswa Program Studi S1 Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam melalui jalur SNMPTN pada tahun 2021. Selama menjalani perkuliahan, penulis cukup aktif mengikuti organisasi kampus, salah satunya dengan menjadi anggota Minat dan Bakat Himpunan Mahasiswa Matematika (HIMATIKA) pada tahun 2022. Penulis juga cukup aktif dalam kepanitiaan acara di lingkungan kampus.

Tahun 2023 tepatnya di bulan Desember, penulis melaksanakan Kerja Praktik (KP) di Dinas Komunikasi, Informasi dan Statistik Provinsi Lampung. Kemudian pada bulan Juni 2024, penulis melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Braja Harjosari, Kecamatan Braja Selebah, Kabupaten Lampung Timur. Selanjutnya, di akhir masa studinya, penulis meluangkan waktu untuk mengajar les privat matematika, sebagai wujud kontribusi dan berbagi ilmu kepada sesama.

KATA INSPIRASI

“Sesungguhnya sesudah kesulitan itu ada kemudahan, maka apabila kamu telah selesai dari sesuatu urusan, kerjakanlah dengan sungguh-sungguh urusan yang lain”
(QS. Al-Insyirah: 6-7)

*“*Pada akhirnya, ini semua hanyalah permulaan*”*
(Nadin Amizah – Beranjak Dewasa)

*“*Karena pelaut hebat tak pernah lahir di laut yang tenang*”*
(HIVI – Jatuh, Bangkit Kembali)

*“*Kapankah terakhir kali kamu menyadari bersyukur atas anugerah-Nya?”
(Quinn– Dengar Hatimu)

PERSEMBAHAN

Segala puji bagi Allah SWT, Tuhan yang Maha Esa, yang dengan rahmat, hidayah, dan kasih-Nya, penulis dapat menyelesaikan skripsi ini. Tidak ada kata yang cukup untuk mengungkapkan rasa syukur saya atas segala nikmat-Nya, baik ilmu, kesehatan, dan kesempatan yang diberikan. Penulis persembahkan skripsi ini yang dibuat dengan ketulusan hati kepada:

Ayah dan Bunda tersayang.
Terima kasih atas dukungan, doa, dan kasih sayang sepanjang perjalanan hidup penulis. Ayah, yang menjadi sumber keteguhan dan kebijaksanaan, dan Bunda, yang tak pernah lelah memberi perhatian dan kekuatan kepada penulis. Semoga Ayah dan Bunda selalu diberikan kesehatan, kebahagiaan dan keberkahan.

Adik-Adik dan Teti.
Terima kasih atas semangat, kebersamaan, dan dukungan yang selalu kalian berikan, baik dalam suka maupun duka.

Dosen pembimbing dan pembahas.
Terima kasih kepada dosen pembimbing dan dosen pembahas yang telah memberikan ilmu, bimbingan, serta dukungan yang sangat membangun sehingga penulis dapat menyelesaikan tugas akhir ini tepat waktu.

Diri saya sendiri.
Terima kasih telah berjuang, menghadapi tantangan, dan tetap gigih menyelesaikan tugas ini meskipun banyak hambatan. Semoga perjalanan ini menjadi pembelajaran yang berharga.

Almamater tercinta Universitas Lampung.

SANWACANA

Puji syukur kehadirat Allah SWT, atas segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Pendugaan *Almon Distributed Lag Models* pada Indikator Indeks Pembangunan Manusia Provinsi Lampung Tahun 2005-2024”.

Dalam menyelesaikan skripsi ini, penulis mendapat dukungan, bimbingan, motivasi, saran dan bantuan dari beberapa pihak sehingga skripsi ini dapat terselesaikan dengan baik dan tepat waktu. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Drs. Mustofa Usman, MA., Ph.D., selaku Pembimbing I atas bimbingan, kritik dan saran kepada penulis dalam menyelesaikan skripsi ini.
2. Ibu Siti Laelatul Chasanah, M.Si., selaku Pembimbing II yang telah memberikan waktunya untuk memberi bimbingan serta saran selama proses penyusunan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen pembahas yang telah bersedia memberikan kritik dan saran serta evaluasi kepada penulis sehingga skripsi ini dapat lebih baik lagi.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku dosen Pembimbing Akademik dan Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah membimbing penulis sampai akhir perkuliahan.
5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staf Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Orangtua tercinta, Ayah dan Bunda yang telah memberikan seluruh kasih sayang, dukungan, kekuatan dan doa kepada penulis.
8. Raifa, Rabiya dan Teti yang selalu memberikan motivasi, semangat dan doa.
9. Bagus Satriyo yang selalu hadir menemani dalam keadaan suka maupun duka dan senantiasa mendengarkan keluh kesah penulis. Terima kasih atas segala dukungan, kebahagiaan, ketulusan, dan momen indah yang selalu diberikan dari awal sampai akhir perjalanan kuliah ini.
10. Putie dan Miko, kedua kucingku yang telah menjadi sumber kebahagiaan kecil di Tengah kesibukan proses penyusunan skripsi ini.
11. Teman-teman tersayang, Istri Konglomerat, Saburakado, Unik, Anida, Mikat'22 terima kasih atas kebersamaan, dukungan serta saling menguatkan dan berbagi pikiran satu sama lain selama proses penyusunan skripsi ini.
12. HIVI, Maliq *D'Essentials*, Nadine Amizah, dan Diskoria, terima kasih atas lagu-lagu indah yang selalu menemani dan memberikan semangat selama proses penulisan skripsi ini.
13. Teman-teman satu pembimbing skripsi, Afika, Amanda, Ranara, Salma dan Zahra yang selalu memotivasi dan saling memberikan dukungan satu sama lain selama penyusunan skripsi.
14. Semua pihak yang membatu penulis dalam menyelesaikan skripsi ini.

Bandar Lampung, 15 Mei 2025
Penulis,

Rafika Aulia Arpan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan metode analisis statistik yang digunakan untuk menjelaskan keterkaitan antara satu atau lebih variabel independen dengan variabel dependen yang dapat dinyatakan ke bentuk persamaan matematika. Analisis regresi juga digunakan untuk mengetahui bagaimana perubahan nilai variabel dependen apabila nilai variabel independennya naik atau turun (Alamsyah, dkk., 2022). Analisis regresi dengan serangkaian pengamatan terhadap suatu peristiwa diambil dari waktu ke waktu disebut sebagai analisis regresi deret waktu. Analisis regresi deret waktu dilakukan untuk melihat kondisi pada masa mendatang (Gujarati & Porter, 2009). Model regresi yang menggunakan data deret waktu tidak hanya menggunakan variabel independen terhadap variabel dependen dalam waktu periode pengamatan yang sama, namun dapat menggunakan periode waktu sebelumnya (Aqibah, dkk., 2020).

Dalam ilmu ekonomi, pada umumnya suatu penyebab akan menimbulkan pengaruh setelah suatu selang waktu tertentu. Selang waktu antara sebab dan pengaruh ini disebut *lag*. Oleh sebab itu, perumusan realistis dari hubungan-hubungan ekonomi membutuhkan nilai-nilai *lag* dari variabel independen. Banyak persoalan yang menunjukkan pentingnya pemakaian nilai-nilai *lag* dari variabel independen dalam merumuskan model-model ekonomi untuk memahami sifat nyata dari masalah ekonomi tersebut. Model yang membahas perubahan variabel dependen yang dipengaruhi nilai-nilai *lag* dari variabel independennya ini disebut model distribusi *lag* (Sumodiningrat, 1995).

Parameter distribusi *lag* dapat ditaksir dengan efisien menggunakan metode *Ordinary Least Square* (OLS) dalam kondisi tertentu, asalkan beberapa asumsi terpenuhi. Asumsi tersebut seperti adanya *lag* maksimum yang tetap, *error* yang terdistribusi secara normal, dan tidak adanya korelasi yang kuat antar variabel independen dari periode yang berbeda. Namun, pada model distribusi *lag*, jika jumlah *lag* besar dan ukuran sampel kecil tidak mungkin untuk menaksir pamareter-parameternya. Sedangkan, setiap menambah variabel *lag* maka sebuah pengamatan bisa hilang yang berakibat berkurangnya derajat bebas. Hal ini dapat mengakibatkan masalah multikolinearitas dan dalam tipe-tipe model semacam itu yang membuat penaksiran dengan metode OLS menyimpang dan tidak efisien. Oleh karena itu, untuk menaksir model-model distribusi *lag,* metode Almon merupakan salah satu metode pendekatan yang dapat digunakan untuk memperkirakan parameter model distribusi *lag* untuk mengurangi efek multikolinearitas (Lukman & Kibria, 2021).

Metode pendekatan Almon mengasumsikan bahwa pengaruh variabel independen terhadap variabel dependen yang nilainya bisa naik bisa turun mengikuti pola dari skema *lag* polinomial sehingga metode ini jauh lebih fleksibel. Oleh sebab itu, pendekatan Almon banyak digunakan untuk menduga model regresi distribusi *lag* karena memungkinkan pendugaan langsung faktor penimbang untuk pengamatan masa lalu yang berkaitan dengan variabel independen melalui interpolasi polinomial (Virgantari & Rahayu, 2021).

Penelitian pemodelan dinamis distribusi *lag* dengan menggunakan metode Kyock dan metode Almon, sebelumnya telah dilakukan oleh Sitorus, dkk. (2023) dengan membandingkan metode Kyock dan metode Almon pada kasus pengaruh kurs rupiah Indonesia terhadap nilai ekspor garmen PT. Shinwon ke mancanegara, penelitian menunjukkan bahwa metode Almon terbaik dibandingkan metode Kyock. Selain itu, penelitian dilakukan juga oleh Virgantari dan Rahayu (2021) dengan mengaplikasikan metode Almon untuk melihat pengaruh rasio biaya operasional dan pendapatan operasional (BOPO) terhadap rasio *return of asset* (ROA) di sebuah bank pemerintah berdasarkan data triwulan yang menunjukkan

BOPO memiliki pengaruh signifikan terhadap ROA dan hasil asumsi tidak terdapat autokorelasi dalam model.

Dalam masalah ekonomi, ekonometrika berperan dalam menemukan solusinya untuk mengukur hubungan antarvariabel berdasarkan data dan menginterpretasikan menggunakan hasil yang akurat. Analisis ekonometrika sangat relevan untuk mengamati perubahan dan perkembangan dari waktu ke waktu, karena memungkinkan pengukuran terhadap data deret waktu (Kusumawardhani,dkk., 2021). Salah satu indikator yang dapat digunakan dalam analisis ekonometrika adalah indeks pembangunan manusia (IPM) yang dapat memberikan informasi tentang perubahan dan faktor-faktor pembangunan manusia dari waktu ke waktu.

Berdasarkan data dari Badan Pusat Statistik Provinsi Lampung (2025), IPM di Provinsi Lampung menunjukkan tren peningkatan setiap tahunnya dari 2020 sampai 2024. Di tingkat kabupaten, Kabupaten Lampung Tengah mencatatkan IPM tertinggi, yaitu 74,16. Sementara untuk wilayah perkotaan, Kota Bandar Lampung dan Kota Metro memiliki nilai IPM yang hampir sama, dengan Bandar Lampung sebesar 80,46 dan Metro lebih sedikit rendah, yaitu 80,41. Pada tahun 2024 di Pulau Sumatera tercatat nilai IPM Provinsi dengan nilai IPM tertinggi adalah Kepulauan Riau, yang mencapai sekitar 76,00 diikuti oleh Sumatera Utara dan Sumatera Barat. Di sisi lain, Provinsi Lampung menunjukkan nilai IPM yang paling rendah dibandingkan provinsi lainnya, yaitu sedikit di atas 71,00.

Berdasarkan informasi tersebut, perkembangan IPM di Provinsi Lampung menunjukkan tren positif, terlihat dari peningkatan persentase IPM yang konsisten antara tahun 2020 hingga 2024. Meskipun demikian, angka IPM tersebut masih berada di bawah capaian provinsi lain di Pulau Sumatera. Ini menunjukkan bahwa peningkatan yang terjadi belum cukup signifikan untuk memperbaiki peringkat IPM Provinsi Lampung, sehingga kondisi pembangunan manusia di provinsi ini masih tertinggal dibandingkan dengan provinsi-provinsi lain di pulau tersebut.

Dengan lokasinya yang strategis dekat ibu kota DKI Jakarta, IPM Provinsi Lampung seharusnya tidak lebih rendah daripada provinsi lain di Pulau Sumatera melainkan menjadi provinsi yang lebih unggul di pulau ini, karena juga berfungsi sebagai gerbang utama menuju Pulau Sumatera. Oleh karena itu, pemerintah perlu berperan lebih aktif dalam meningkatkan kesejahteraan rakyat dengan lebih merata mendistribusikan sumber daya pendidikan, ekonomi dan kesehatan di Provinsi Lampung. Rendahnya investasi publik dalam sektor pendidikan, ekonomi dan kesehatan, dapat menyebabkan rendahnya produktivitas, dimana sektor pendidikan dapat tercermin dari harapan lama sekolah dan rata-rata lama sekolah, sektor ekonomi untuk standar hidup yang layak dapat dilihat melalui indikator pengeluaran per kapita, sementara itu sektor kesehatan dapat tercermin dari umur harapan hidup (Ramadanisa & Triwahyuningtyas, 2022).

Beberapa penelitian sebelumnya, seperti yang dilakukan oleh Syafira, dkk (2024), menunjukkan pengujian secara parsial bahwa harapan lama sekolah, rata-rata lama sekolah, pengeluaran pengeluaran per kapita serta umur harapan hidup memiliki pengaruh yang signifikan terhadap IPM dan memiliki pengaruh yang positif, yang artinya indikator tersebut mengalami peningkatan, maka IPM akan meningkat.. Namun penelitian oleh Nasyri, dkk (2024), menyatakan hasil uji secara parsial bahwa rata-rata lama sekolah tidak berpengaruh secara signifikan terhadap IPM dan penelitian lain oleh Prahasta, dkk (2023), menunjukkan hasil pengujian secara parsial bahwa umur harapan hidup tidak berpengaruh secara signifikan terhadap IPM.

Berdasarkan hal tersebut, Provinsi Lampung memiliki tantangan besar dalam meningkatkan kualitas pembangunan manusia melalui indikator pendidikan, perekonomian dan kesehatan. Pada penelitian ini berfokus pada analisis ekonometrika yaitu mengenai pemodelan dinamis distribusi *lag* dengan kasus pengaruh indikator-indikator penyusun IPM terhadap IPM di Provinsi Lampung menggunakan metode pendekatan Almon. Melalui pendekatan ini diharapkan mampu memberikan gambaran yang menyeluruh mengenai pengaruh pembangunan terhadap perubahan IPM di Provinsi Lampung, dengan memperhitungkan jangka pendek maupun jangka panjang.

1.2 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian ini adalah:

1. Menentukan model estimasi distribusi *lag* Almon terbaik dengan membandingkan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) pada panjang *lag* yang berbeda.
2. Menentukan model dinamis distribusi *lag* dengan mencari koefisien lag δ_i menggunakan koefisien model estimasi distribusi *lag* almon terbaik yang telah dipilih.
3. Menganalisis pengaruh indikator-indikator penyusun IPM terhadap IPM Provinsi Lampung pada tahun 2005-2024.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan pemahaman tentang model dinamis distribusi *lag* hasil estimasi dengan metode transformasi Almon.
2. Dapat digunakan sebagai bahan pertimbangan refrensi bagi penelitian selanjutnya yang berkaitan dengan analisis distribusi *lag*.
3. Memberikan kontribusi dalam penyusunan kebijakan pemerintah dalam pembangunan yang berkelanjutan untuk dapat meningkatkan kesejahteraan masyarakat di Provinsi Lampung secara efektif dan efisien.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Setiawan & Kusrini (2010) mengatakan analisis regresi merupakan metode statistik yang digunakan untuk menggambarkan dan mempelajari hubungan matematis antara satu variabel dependen dengan satu atau lebih variabel independen. Variabel dependen merupakan variabel yang dipengaruhi oleh variabel lain. Sedangkan, variabel independen merupakan variabel yang mempengaruhi variabel dependen. Tujuannya untuk memprediksi nilai rata-rata dari variabel dependen berdasarkan nilai-nilai variabel dependen yang sudah diketahui. Dalam analisis regresi terbagi menjadi dua yaitu analisis regresi linear sederhana dan analisis regresi linear berganda.

Dalam analisis regresi linear, model dapat direpresentasikan dalam bentuk matriks sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

dengan,

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \ddots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dimana :

$\boldsymbol{Y}$ = vektor yang berisi nilai-nilai variabel dependen

$\boldsymbol{X}$ = matriks yang berisi nilai-nilai variabel independent

$\boldsymbol{\beta}$ = vektor koefisien regresi atau vektor parameter yang ingin diprediksi

$\boldsymbol{\varepsilon}$ = vektor galat atau *error*, dimana $\varepsilon_i \sim N(0, \sigma^2)$

Menurut Basuki & Prawoto (2019), analisis regresi memiliki tiga fungsi utama. Pertama, digunakan untuk mendeskripsikan fenomena dalam data, di mana hubungan numerik antara variabel dapat dipahami secara rinci. Kedua, regresi dapat membantu dalam pengendalian situasi atau peristiwa yang sedang diteliti, karena model regresi memungkinkan peneliti memprediksi perubahan pada variabel dependen berdasarkan perubahan pada variabel independen. Ketiga, analisis regresi digunakan sebagai alat prediksi, dengan model yang terbentuk mampu meramalkan nilai variabel dependen di masa depan berdasarkan informasi variabel independen yang diketahui.

2.1.1 Analisis Regresi Linear Sederhana

Analisis regresi sederhana merupakan analisis regresi yang menggunakan dua variabel, dimana pada suatu model hanya terdapat satu variabel independen dan variabel dependen yang dinyatakan sebagai fungsi linear. Analisis regresi linear sederhana diterapkan dengan beberapa pertimbangan seperti, analisis ini praktis digunakan, analisis ini menyajikan ide dasar bagi analisis regresi dengan sesederhana mungkin, dan analisis ini mudah dijelaskan dengan hanya diagram dua dimensi (Firdaus, 2019).

Hubungan antar variabel di dalam analisis regresi sederhana bersifat linear, yaitu perubahan pada variabel independen akan diikuti oleh perubahan variabel dependen secara konstan. Sedangkan, pada hubungan non linear, perubahan variabel independen tidak diikuti variabel dependen secara seimbang (Muhartini, dkk., 2021).

Persamaan atau hubungan dalam ilmu ekonomi biasanya mempunyai spesifikasi hubungan yang jelas atau bersifat pasti. Namun, pada kenyataannya, hubungan yang sepenuhnya pasti jarang, bahkan hampir tidak pernah ada dalam ekonomi. Faktor-faktor ekonomi sangat dinamis dan dipengaruhi oleh banyak variabel yang sering kali berada di luar kendali, sehingga hubungan antarvariabel ekonomi tidak bisa selalu dijelaskan secara mutlak, maka faktor-faktor acak atau stokastik harus ada dalam hubungan ekonomi. Ini berarti ada unsur variabel lain yang dapat mempengaruhi hasil dan membuat hubungan antara variabel ekonomi tidak selalu bisa diprediksi secara tepat. Dalam regresi linear sederhana bentuk hubungan stokastik antara dua variabel independen dan variabel dependen dapat ditulis sebagai berikut (Sumodiningrat, 1995):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{2.2}$$

dengan:

Y_i = variabel dependen pada objek ke$-i$

X_i = variabel independen

β_0 = konstanta atau intersep

β_1 = koefisien regresi atau nilai parameter variabel independen

ε_i = galat atau *error* pada objek ke$-i$, dimana $\varepsilon_i \sim N(0, \sigma^2)$

i = $1, 2, 3, \dots, n$

Persamaan berikut dapat digunakan untuk menghitung intersep dan koefisien variabel dalam model regresi linear sederhana (Almumtazah, dkk., 2021):

$$\beta_0 = \frac{(\sum X^2)(\sum Y) - (\sum XY)(\sum X)}{n(\sum X^2)(\sum X)^2} \tag{2.3}$$

$$\beta_1 = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \tag{2.4}$$

dengan:

ΣX = jumlah seluruh nilai X

ΣX = jumlah seluruh nilai Y

ΣXY = jumlah dari hasil kali X dan Y

ΣX^2 = jumlah dari nilai X kuadrat

n = jumlah total pengamatan

Dalam regresi, intersep membantu kita mengetahui nilai dasar prediksi saat variabel independen tidak berkontribusi. Dan nilai koefisien β menunjukkan seberapa besar perubahan yang terjadi pada variabel dependen untuk setiap satu unit perubahan pada variabel independen. Jika β positif, maka ada hubungan positif antara variabel independen dan variabel dependen, dan jika β negatif artinya ada hubungan yang negatif.

2.1.2 Analisis Regresi Linear Berganda

Analisis regresi linear berganda merupakan pengembangan dari analisis regresi sederhana. Pada regresi linear berganda, terdapat lebih dari satu variabel independen. Analisis ini digunakan untuk melihat pengaruh beberapa variabel independen $(X_1, X_2, \dots, X_k)$ terhadap variabel dependen, dengan mempertimbangkan nilai yang dimiliki oleh setiap variabel independen tersebut (Wisudaningsi, dkk., 2019). Perbedaan utama antara regresi linear berganda dan regresi linear sederhana adalah jumlah variabel independen yang digunakan. Dalam regresi linear sederhana,hanya ada satu variabel independen yang digunakan untuk memprediksi variabel dependen. Sedangkan, dalam regresi linear berganda, lebih dari satu variabel independen digunakan untuk memprediksi variabel dependen.

Nilai-nilai $X_1, X_2, \dots, X_k$ ditaksir menggunakan metode OLS dengan menggunakan sampel sebanyak n. Dengan bertambahnya variabel independen, maka bentuk umum model persamaan dari regresi linear berganda dapat dituliskan sebagai berikut (Fitriyah, dkk., 2021):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{2.5}$$

dengan:

Y_i = variabel dependen pada objek ke$-i$

X_{ki} = variabel independent pada objek ke$-i$

β_0 = konstanta atau intersep

β_k = koefisien regresi nilai parameter variabel independent ke-k

ε_i = galat atau *error* pada objek ke$-i$, dimana $\varepsilon_i \sim N(0, \sigma^2)$

i = 1, 2, 3,…, n

2.2 Metode *Ordinary Least Square* (OLS)

Metode OLS adalah cara yang digunakan untuk menemukan hubungan linear antara dua variabel. Caranya dengan menentukan garis lurus (garis tren) yang meminimalkan selisih antara data asli dan titik-titik pada garis tersebut. Hasil dari metode ini adalah persamaan garis yang menunjukkan kecenderungan positif atau negatif. Dari persamaan ini, kita bisa memprediksi nilai di masa depan. Metode OLS sering digunakan dalam peramalan, terutama untuk memprediksi data masa depan berdasarkan data sebelumnya (Sihotang, dkk., 2019).

Gujarati & Porter (2009) mengatakan bahwa dalam metode OLS, penduga yang disebut sebagai penduga tak bias linear terbaik (*Best Linear Unbiased Estimator*) jika memenuhi tiga sifat penting yaitu, linearitas, tak bias, dan variansi minimum. Saat asumsi metode OLS telah terpenuhi kita perlu membahas sifat penduga tak bias linear terbaik dari parameter-parameternya.

Menurut Sumodiningrat (1995), terdapat beberapa asumsi yang digunakan untuk menguji sifat-sifat penduga parameter:

1. Nilai harapan (ekspetasi) dari *error* adalah nol atau $E[\boldsymbol{\varepsilon}] = 0$.
2. Nilai harapan dari hasil kali *error* dengan transposnya sama dengan varians *error* yang konstan atau $E[\boldsymbol{\varepsilon\varepsilon^T}] = \sigma^2\boldsymbol{I}$.

Penduga metode OLS diperoleh dengan cara meminimalkan jumlah kuadrat residual yaitu, selisih antara nilai pengamatan sebenarnya dan nilai prediksi dalam model regresi linier. Dengan kata lain, metode OLS mencari estimasi parameter yang menghasilkan nilai prediksi paling mendekati data aktual, sehingga total kuadrat dari residual dapat diperkecil sebanyak mungkin (Sumodiningrat,1995). Dimulai dari model regresi linear yaitu:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.6}$$

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta}$$

Jumlah kuadrat residual dinyatakan sebagai:

$$\begin{aligned} S(\widehat{\boldsymbol{\beta}}) &= \sum \boldsymbol{\varepsilon_i^2} \\ &= \boldsymbol{\varepsilon^T\varepsilon} \\ &= (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \left(\boldsymbol{Y^T} - \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^T}\right)(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \boldsymbol{Y^TY} - \boldsymbol{Y^TX}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^TY} + \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^TX}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{Y^TY} - 2\widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^TY} + \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^TX}\widehat{\boldsymbol{\beta}} \end{aligned} \tag{2.7}$$

Turunkan $S(\widehat{\boldsymbol{\beta}})$ terhadap $\widehat{\boldsymbol{\beta}}$ secara parsial dan buat turunan tersebut sama dengan nol, sehingga diperoleh:

$$\frac{\partial S(\widehat{\beta})}{\partial \widehat{\beta}} = 0$$

$$\frac{\partial(Y^TY - 2\widehat{\beta}^TX^TY + \beta^TX^TX\widehat{\beta})}{\partial \widehat{\beta}} = 0$$

$$\frac{\partial(Y^TY)}{\partial \widehat{\beta}} - \frac{\partial(2\widehat{\beta}^TX^TY)}{\partial \widehat{\beta}} + \frac{\partial(\widehat{\beta}^TX^TX\widehat{\beta})}{\partial \widehat{\beta}} = 0$$

$$0 - \boldsymbol{2X^TY} + \boldsymbol{2X^TX\widehat{\beta}} = 0$$

$$\boldsymbol{2X^TX\widehat{\beta}} = \boldsymbol{2X^TY}$$

$$\boldsymbol{X^TX\widehat{\beta}} = \boldsymbol{X^TY}$$

$$(\boldsymbol{X^TX})^{-1}\boldsymbol{X^TX\widehat{\beta}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TY}$$

$$\boldsymbol{I\widehat{\beta}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TY}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TY} \qquad (2.8)$$

Setelah mendapatkan penduga OLS, kita dapat menurunkan sifat-sifat penting dari penduga OLS, yaitu:

1. Linear

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TY}\\ &= (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})\\ &= (\boldsymbol{X^TX})^{-1}\boldsymbol{X^TX\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon}\\ &= \boldsymbol{I\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon}\\ &= \boldsymbol{\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon} \qquad (2.9)\end{aligned}$$

   Diperoleh $\hat{\beta}$ merupakan fungsi linear dari β dan ε.

2. Tak Bias

   Sifat tak bias berarti rata-rata penduga atau nilai harapan sama dengan nilai sebenarnya dari parameter yaitu $\mathrm{E}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

$$\begin{aligned}\mathrm{E}[\widehat{\boldsymbol{\beta}}] &= E[\boldsymbol{\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon}]\\ &= E[\boldsymbol{\beta}] + E[(\boldsymbol{X^TX})^{-1}\boldsymbol{X^T\varepsilon}]\\ &= \boldsymbol{\beta} + (\boldsymbol{X^TX})^{-1}\boldsymbol{X^T}E[\boldsymbol{\varepsilon}] \qquad (2.10)\end{aligned}$$

   Karena $E[\boldsymbol{\varepsilon}] = 0$, maka $\mathrm{E}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$

   Diperoleh $\hat{\beta}$ merupakan penduga tak bias.

3. Variansi Minimum

Penduga yang memiliki variansi paling kecil dibandingkan dengan penduga lainnya yang memiliki parameter sama disebut penduga variansi minimum. Misalnya jika ada dua penduga $\hat{\beta}_1$ dan $\hat{\beta}_2$, dan variansi $\hat{\beta}_1$lebih kecil dari $\hat{\beta}_2$, maka $\hat{\beta}_1$ merupakan penduga variansi minimum. Variansi $\hat{\beta}_1$ dapat dihutung sebagai berikut:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathbf{1}} &= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon} \\
&= \boldsymbol{I}\boldsymbol{\beta} + (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon} \\
\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta} &= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon} \qquad (2.11)
\end{aligned}$$

Selanjutnya, subtitusi hasil $\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta}$ kedalam persamaan $Var(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) = E\left[(\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta})^{\mathbf{2}}\right]$ sebagai berikut:

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) &= E\left[(\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta})^{\mathbf{2}}\right] \\
&= E\left[\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta})^{\boldsymbol{T}}\right] \\
&= E[\{(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon}\}\{(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\varepsilon}\}^{\boldsymbol{T}}\ ] \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\boldsymbol{T}}]\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}} \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\sigma}^{\mathbf{2}}\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}} \\
&= \boldsymbol{\sigma}^{\mathbf{2}}\boldsymbol{I}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}} \\
&= \boldsymbol{\sigma}^{\mathbf{2}}\boldsymbol{I}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}} \qquad (2.12)
\end{aligned}$$

Ditunjukkan $Var(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) \leq Var(\widehat{\boldsymbol{\beta}}_{\mathbf{2}})$, misalkan $\widehat{\boldsymbol{\beta}}_{\mathbf{2}} = [(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{B}]\boldsymbol{Y}$ dengan,

$\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$: penduga alternatif yang linear dan tak bias bagi $\boldsymbol{\beta}$

$\boldsymbol{B}$: matriks konstanta yang diketahui

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathbf{2}} &= [(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{B}]\boldsymbol{Y} \\
&= [(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}} + \boldsymbol{B}](\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) + \boldsymbol{B}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \qquad (2.13)
\end{aligned}$$

$$
\begin{aligned}
E[\widehat{\boldsymbol{\beta}}_2] &= E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})+\boldsymbol{B}(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})] \\
&= E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X\beta}+(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{BX\beta}+\boldsymbol{B\varepsilon}] \\
&= E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X\beta}]+E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}]+E[\boldsymbol{BX\beta}]+E[\boldsymbol{B\varepsilon}] \\
&= \boldsymbol{\beta}+\boldsymbol{0}+\boldsymbol{BX\beta}+\boldsymbol{0} \\
&= \boldsymbol{\beta}+\boldsymbol{BX\beta} \qquad (2.14)
\end{aligned}
$$

Misal $\widehat{\boldsymbol{\beta}}_2$ adalah penduga tak bias untuk $\boldsymbol{\beta}$ artinya $E[\widehat{\boldsymbol{\beta}}_2]=\boldsymbol{\beta}$ atau dapat disebut $\boldsymbol{BX\beta}$ adalah matriks 0. Variansi dari penduga alternatif ini dapat dihitung dengan:

$$
\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_2) &= E\left[(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta})^2\right] \\
&= E[(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta})^T] \\
&= E[\{[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}]\boldsymbol{Y}-\boldsymbol{\beta}\}\{[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}]\boldsymbol{Y}-\boldsymbol{\beta}\}^T] \\
&= E[\{[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}](\boldsymbol{X\beta}+\boldsymbol{\varepsilon})-\boldsymbol{\beta}\}\{[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}](\boldsymbol{X\beta}+ \\
&\quad \boldsymbol{\varepsilon})-\boldsymbol{\beta}\}^T] \\
&= E[\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X\beta}+(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{BX\beta}+\boldsymbol{B\varepsilon}- \\
&\quad \boldsymbol{\beta}\}\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X\beta}+(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{BX\beta}+\boldsymbol{B\varepsilon}-\boldsymbol{\beta}\}^T] \\
&= E[\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{B\varepsilon}\}\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{B\varepsilon}\}^T] \quad \text{karena } \boldsymbol{BX}=\boldsymbol{0} \\
&= E[\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}+\boldsymbol{B\varepsilon}\}\{\boldsymbol{\varepsilon}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}+\boldsymbol{\varepsilon}^T\boldsymbol{B}^T\}^T] \\
&= E[\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}\}\boldsymbol{\varepsilon\varepsilon}^T\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}+\boldsymbol{B}^T\}] \\
&= \{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}\}E[\boldsymbol{\varepsilon\varepsilon}^T]\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}+\boldsymbol{B}^T\} \\
&= \boldsymbol{\sigma}^2\boldsymbol{I}\{\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}\}\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{B}^T\}\} \\
&= \boldsymbol{\sigma}^2\boldsymbol{I}\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}+\boldsymbol{BX}(\boldsymbol{X}^T\boldsymbol{X})^{-1}+(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{B}^T+ \\
&\quad \boldsymbol{BB}^T\} \\
&= \boldsymbol{\sigma}^2\boldsymbol{I}\{(\boldsymbol{X}^T\boldsymbol{X})^{-1}+\boldsymbol{BB}^T\} \\
&= \boldsymbol{\sigma}^2\boldsymbol{I}(\boldsymbol{X}^T\boldsymbol{X})^{-1}+\boldsymbol{\sigma}^2\boldsymbol{IBB}^T \qquad (2.15)
\end{aligned}
$$

Diperoleh hasil $Var(\widehat{\boldsymbol{\beta}}_1)=\boldsymbol{\sigma}^2\boldsymbol{I}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\leq\boldsymbol{\sigma}^2\boldsymbol{I}(\boldsymbol{X}^T\boldsymbol{X})^{-1}+\boldsymbol{\sigma}^2\boldsymbol{IBB}^T=Var(\widehat{\boldsymbol{\beta}}_2)$, sehingga $Var(\widehat{\boldsymbol{\beta}}_1)\leq Var(\widehat{\boldsymbol{\beta}}_2)$.

Terbukti $Var(\widehat{\boldsymbol{\beta}}_1)$ merupakan variansi minimum.

2.3 Uji Stasioneritas

Menurut Box, dkk. (2015), stasioneritas adalah kondisi di mana statistik dasar dari suatu deret waktu, seperti rata-rata, varians, dan kovarians, tetap stabil atau tidak berubah seiring waktu. Stasioneritas dalam rata-rata berarti rata-rata data tetap konstan dan tidak berubah seiring waktu. Jika rata-rata ini terus berubah, data dianggap tidak stasioner. Sementara itu, stasioneritas dalam varians berarti perubahan data tetap konsisten, tidak semakin besar atau kecil seiring waktu. Jika variansnya berubah, maka data tersebut tidak stasioner. Dengan kata lain, jika rata-rata atau varians data tidak stabil, maka data dianggap tidak stasioner.

Dalam analisis deret waktu, penting untuk memastikan data yang digunakan stasioner karena banyak metode analisis yang membutuhkan data stasioner agar hasil estimasi koefisien regresi menjadi valid dan efisien. Jika data yang digunakan mengandung variabel yang tidak stasioner, misalnya yang memiliki unit *root*, dan tetap dimasukkan dalam model, maka estimasi yang dihasilkan bisa menjadi bias dan tidak akurat.

Bias ini akan mempengaruhi hasil analisis dan membuat tes-tes yang biasanya digunakan untuk menentukan hubungan sebab-akibat antar variabel menjadi tidak valid. Dengan kata lain, menggunakan data non-stasioner dalam estimasi bisa menghasilkan kesimpulan atau prediksi yang salah, karena koefisien yang diperoleh tidak akurat dan tidak efisien (Sujatmiko, dkk., 2021).

Salah satu metode yang digunakan untuk menguji stasioneritas pada data deret waktu adalah Uji KPSS (*Kwiatkowski Phillips Schmidt Shin*), yang berfungsi untuk mengetahui apakah data stasioner di sekitar tren yang tetap atau tidak. Uji KPSS ini memeriksa dua jenis stasioneritas utama, yaitu stasioneritas rata-rata, yang memastikan rata-rata data tetap konstan sepanjang waktu, dan stasioneritas varians, yang memastikan varians data tetap stabil tanpa perubahan besar yang tidak diinginkan.

Statistik uji untuk uji KPSS dihitung dengan rumus (Marsani & Shabri, 2019):

$$LM = \sum_{t=1}^{T} \frac{\varepsilon_t^2}{\hat{\sigma}_\varepsilon^2} \tag{2.16}$$

dengan:

LM = *Lagrange multiplier*

ε_t^2 = kuadrat residu pada waktu-t

$\hat{\sigma}_\varepsilon^2$ = estimasi varians dari *error*

T = jumlah observasi dalam deret waktu

Hipotesis pengujian uji stasioneritas menggunakan uji KPSS , yaitu:

1. H_0: Data stasioner.
2. H_1: Data tidak stasioner.

Dasar pengambilan keputusan uji KPSS umumnya seperti:

1. Jika nilai $P\ value \geq 0{,}05$, menyatakan bahwa data stasioner.
2. Jika nilai $P\ value \leq 0{,}05$, menyatakan bahwa data tidak stasioner.

2.4 Pembeda (*Differencing*)

Gujarati & Porter (2009) mengatakan untuk menghindari masalah regresi semu yang bisa muncul ketika kita melakukan regresi antara deret waktu yang tidak stasioner, kita perlu mengubah deret waktu yang tidak stasioner menjadi stasioner dengan melakukan transformasi pada data. Salah satu cara untuk membuat deret waktu stasioner adalah dengan melakukan *differencing*. Metode *differencing* ini dilakukan dengan cara membuat data baru yang diperoleh dari selisih antara nilai pengamatan pada waktu tertentu (t) dengan nilai pengamatan pada waktu sebelumnya $(t-1)$. Teknik ini akan membantu menghilangkan tren dalam data, sehingga data menjadi lebih stabil dan siap untuk dianalisis lebih lanjut.

Proses ini dilakukan dengan *differencing* pertama $(d = 1)$ yang secara sistematis dapat ditulis sebagai:

$$d_t' = X_t - X_{t-1} \tag{2.17}$$

dengan:

d_t' = nilai hasil *differencing* pertama pada waktu t

X_t = nilai data pada waktu t

X_{t-1} = nilai data pada waktu $t-1$

Jika setelah dilakukan *differencing* pertama, data masih belum stasioner, maka langkah selanjutnya adalah melakukan *differencing* kedua ($d = 2$). *Differencing* kedua ini dilakukan dengan cara yang sama seperti *differencing* pertama, namun pada data yang sudah mengalami *differencing* pertama. Jadi, kita mengurangi nilai hasil *differencing* pertama pada waktu t dengan nilai hasil *differencing* pertama pada waktu $t-1$. Secara sistematis dapat ditulis sebagai:

$$d_t'' = d_t' - d_{t-1}' \tag{2.18}$$

dengan:

d_t'' = nilai hasil *differencing* kedua pada waktu t

d_t' = nilai hasil *differencing* pertama pada waktu t

d_{t-1}' = nilai hasil *differencing* pertama pada waktu $t-1$

2.5 *Akaike Information Criterion* (AIC)

Salah satu ukuran yang digunakan untuk membandingkan kualitas relatif dari beberapa model statistik yaitu AIC yang dikembangkan oleh Akaike. AIC adalah metode yang digunakan untuk memilih model terbaik dengan mempertimbangkan keseimbangan antara kecocokan model terhadap data dan kompleksitas model itu sendiri. AIC dihitung berdasarkan nilai maksimum dari fungsi likelihood, yang menunjukkan seberapa baik model dapat menjelaskan data, serta penalti yang diberikan terhadap jumlah parameter dalam model. Semakin banyak parameter yang ditambahkan, model mungkin semakin cocok dengan data (*goodness of fit* meningkat), tetapi ada risiko overfitting, yaitu model menjadi terlalu kompleks dan kurang dapat digeneralisasi untuk data baru.

AIC menambahkan penalti pada model yang memiliki terlalu banyak parameter agar tidak hanya berfokus pada kecocokan terhadap data, tetapi juga memperhitungkan kesederhanaan model. Dalam penerapannya, model dengan nilai AIC paling kecil dianggap sebagai yang terbaik karena dapat menemukan keseimbangan yang optimal antara akurasi dan kompleksitas model (Pham, 2019). Rumus statistik untuk AIC, yaitu sebagai berikut:

$$AIC = 2k + n\ln\left(\frac{\sum \varepsilon_i^2}{n}\right) \tag{2.19}$$

dengan:

AIC = statistik *Akaike Information Criterion*

k = jumlah parameter dalam model

n = jumlah observasi

ln = logaritma natural

ε_i^2 = jumlah kuadrat residual

2.6 *Bayesian Information Criterion* (BIC)

Menurut Pham (2019), BIC juga merupakan salah satu ukuran yang digunakan untuk membandingkan kualitas relatif dari beberapa model statistic. BIC merupakan metode yang diperkenalkan oleh Schwarz yang digunakan untuk memilih model terbaik dengan mempertimbangkan keseimbangan antara kecocokan model terhadap data dan kompleksitas model dengan bergantung pada jumlah sampel. Tujuan utama dari BIC adalah menghindari pemilihan model yang terlalu kompleks dengan memberikan penalti terhadap jumlah parameter yang digunakan. Semakin banyak parameter dalam model, semakin besar nilai BIC, sehingga model dengan struktur yang lebih sederhana cenderung lebih diprioritaskan.

Perbedaan utama antara AIC dan BIC terletak pada cara mereka memberikan penalti terhadap jumlah parameter dalam model. AIC menggunakan penalti tetap sebesar $2k$, di mana k adalah jumlah parameter dalam model. Sementara itu, BIC memiliki penalti yang bergantung pada ukuran sampel, yaitu $kln(n)$. Dalam praktiknya, AIC lebih cenderung memilih model yang lebih kompleks (dengan lebih banyak parameter), terutama jika sampel kecil, karena penalti yang diberikan lebih ringan.

Sebaliknya, BIC lebih cenderung memilih model yang lebih sederhana, karena penalti terhadap parameter tambahan semakin besar saat jumlah data meningkat. Dalam seleksi model dengan BIC, model dengan BIC terkecil dipilih sebagai yang terbaik karena menyeimbangkan kecocokan data dan kompleksitas model. BIC dapat dihitung menggunakan rumus:

$$BIC = k\ln(n) + n\ln\left(\frac{\sum \varepsilon_i^2}{n}\right) \tag{2.20}$$

dengan:

BIC = statistik *Bayesian Information Criterion*

k = jumlah parameter dalam model

n = jumlah observasi

ln = logaritma natural

ε_i^2 = jumlah kuadrat residual

2.7 Uji Asumsi Klasik

Uji asumsi klasik adalah syarat statistik yang perlu dipenuhi dalam analisis regresi linear menggunakan OLS. Pengujian ini dilakukan untuk memastikan bahwa persamaan regresi yang digunakan adalah tepat dan akurat. Dengan melakukan pengujian asumsi klasik, kita dapat memastikan bahwa model regresi yang dihasilkan memiliki estimasi yang akurat, bebas dari bias, dan konsistensi yang optimal (Sholihah, dkk*.,* 2023).

Penduga OLS pada model regresi membutuhkan beberapa asumsi penting untuk menghasilkan estimasi yang akurat. Beberapa asumsi tersebut seperti, rata-rata *error* harus bernilai nol, distribusi *error* harus normal, tidak ada autokorelasi, tidak terjadi heteroskedastisitas, tidak ada multikolinearitas, serta variabel independen tidak bersifat acak (Gunawan, dkk., 2024).

2.7.1 Uji Multikolinearitas

Asumsi mengenai tidak adanya multikolinearitas menyatakan bahwa tidak boleh ada hubungan yang kuat atau korelasi antara dua atau lebih variabel independen dalam model regresi. Artinya, model regresi yang baik harus menunjukkan bahwa setiap variabel independen berdiri sendiri dalam menjelaskan variabilitas variabel dependen atau dengan kata lain $cov(x_i, x) = 0, i \neq j$. Jika model regresi menunjukkan adanya korelasi yang tinggi atau sempurna di antara variabel independen, maka dapat disimpulkan bahwa model tersebut mengandung masalah multikolinearitas. (Gunawan, dkk., 2024).

Salah satu metode umum untuk mendeteksi multikolinearitas dalam analisis regresi adalah dengan memeriksa nilai *Variance Inflation Factor* (VIF). Nilai VIF menunjukkan seberapa besar ketidakstabilan atau ketidaktepatan dalam menghitung koefisien regresi yang disebabkan oleh adanya hubungan antar variabel independen. Untuk setiap variabel independen dalam model.

VIF dihitung sebagai (Gujarati & Porter, 2009):

$$VIF_J = \frac{1}{1-R_j^2} \tag{2.21}$$

dengan:

VIF_J = VIF untuk variabel independent X_j dalam model regresi

R_j^2 = koefisien determinasi dari regresi yang dilakukan dengan X_j

Ambang batas umum untuk mendeteksi multikolinearitas menggunakan VIF adalah 10. Nilai VIF ≥ 10 mengindikasikan multikolinearitas kuat yang dapat memengaruhi stabilitas koefisien regresi, sedangkan VIF < 10 menunjukkan tidak adanya multikolinearitas serius, sehingga setiap variabel independen dianggap memberikan kontribusi unik terhadap model (Sholihah, dkk., 2023).

Menurut O'Brien (2007), penggunaan batas VIF seperti 4, 5, atau 10 sebagai indikator multikolinearitas sering dilakukan tanpa mempertimbangkan konteks analisis secara menyeluruh. VIF yang tinggi tidak otomatis membatalkan validitas regresi, selama koefisien signifikan dan asumsi model terpenuhi. Oleh karena itu, VIF sebaiknya tidak dijadikan satu-satunya acuan dalam menghapus variabel, melainkan dipertimbangkan bersama nilai signifikansi koefisien, R^2, dan varians variabel independen.

2.7.2 Uji Normalitas

Distribusi *error* mengikuti normal rata-rata 0 dan variansi σ^2, yang artinya *error* terdistribusi secara normal di sekitar nol dengan penyebaran yang ditentukan oleh variansi σ^2. Uji normalitas merupakan uji yang bertujuan untuk memeriksa apakah *error* dalam model regresi menyebar mengikuti distribusi normal. Ini penting karena banyak metode statistik, termasuk regresi, mengandalkan asumsi bahwa *error* berdistribusi normal untuk menghasilkan hasil yang lebih akurat dan dapat dipercaya (Gunawan, dkk., 2024).

Terdapat beberapa metode yang dapat digunakan untuk uji normalitas dan dalam penelitian ini digunakan metode Uji *Jarque-Bera* (JB *test*). JB *test* merupakan salah satu uji normalitas dalam kategori *goodness of fit test* yang mengevaluasi apakah *skewness* (kemencengan) dan *kurtosis* (keruncingan) dari sampel sama dengan karakteristik distribusi normal. Dalam uji ini, nilai statistik JB dibandingkan dengan nilai kritis *chi-square* (X^2) dari tabel. Jika nilai JB kurang dari atau sama dengan nilai X^2 tabel, residual dianggap berdistribusi normal (Sholihah, dkk., 2023). Taraf signifikansi yang umum digunakan adalah 0,05.

Untuk menghitung nilai statistik JB, kita dapat menggunakan rumus berikut (Jarque & Bera, 1980):

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3^2)}{24}\right) \tag{2.22}$$

dengan:

JB = statistik *jarque-bera*

S = koefisien *skewness*, yang mengukur asimetri distribusi data

K = koefisien *kurtosis*, yang mengukur puncak distribusi data

Hipotesis pengujian uji normalitas menggunakan JB *test,* yaitu:

1. H_0: Data berdistribusi normal.
2. H_1: Data tidak berdistribusi normal.

Dasar pengambilan keputusan JB *test* umumnya seperti:

1. Jika nilai $P\ value \geq 0{,}05$, menyatakan bahwa data berdistribusi normal.
2. Jika nilai $P\ value \leq 0{,}05$, menyatakan bahwa data tidak berdistribusi normal.

Menurut Kusumawardhani, dkk. (2021), pengujian apakah residual dalam model regresi mengikuti distribusi normal dapat dilakukan dengan menggunakan *Q-Q* plot *(Quantile-Quantile plot*), yang juga dikenal sebagai plot probabilitas normal. *Q-Q* plot adalah sebuah grafik dua dimensi yang membandingkan nilai kuantil residual (*error* prediksi dalam model) dengan nilai kuantil dari distribusi normal teoretis. Pada grafik ini, sumbu variabel independen merepresentasikan kuantil yang diharapkan dari distribusi normal, sementara sumbu variabel dependen menunjukkan kuantil yang diperoleh dari residual yang dihasilkan model. Tujuan utama dari *Q-Q* plot adalah untuk menguji apakah residual berdistribusi normal.

Jika titik-titik pada *Q-Q* plot terletak sepanjang garis lurus, maka residual dapat dikatakan mengikuti distribusi normal. Hal ini menunjukkan bahwa *error* model tersebar secara acak dan tidak ada pola yang jelas. Sebaliknya, jika titik-titik dalam *Q-Q* plot menyimpang jauh dari garis lurus, maka ini mengindikasikan bahwa residual tidak berdistribusi normal, yang bisa menjadi indikasi adanya masalah dalam model yang digunakan.

2.7.3 Uji Non Autokorelasi

Asumsi non autokorelasi berarti bahwa *error* dalam model pada satu waktu tidak memiliki keterkaitan dengan *error* pada waktu lainnya atau berarti $cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$. Jika asumsi ini terpenuhi, maka nilai *error* dari suatu waktu tidak memiliki pola tertentu atau ketergantungan dengan *error* di waktu lainnya, yang penting untuk memastikan bahwa model regresi tidak memiliki bias akibat pola berulang dalam *error* (Gunawan, dkk., 2024).

Autokorelasi dapat diuji menggunakan beberapa metode, salah satunya adalah uji *Breusch-Godfrey*. Uji ini merupakan teknik untuk mendeteksi masalah autokorelasi yang dikembangkan oleh *Breusch* dan *Godfrey*. Dalam uji ini, terdapat asumsi bahwa faktor *error* mengikuti pola yang diturunkan dari skema *autoregressive order*. Setelah melaksanakan uji non autokorelasi dengan metode ini, hasilnya akan dianalisis dengan membandingkan nilai X^2_{hitung} dengan X^2_{tabel}. Jika nilai X^2_{hitung} lebih kecil dari X^2_{tabel}, maka disimpulkan bahwa model regresi tidak mengalami masalah autokorelasi (Sholihah, dkk., 2023).

Rumus statistik uji *Breusch-Godfrey* adalah (Breusch, 1978):

$$BG = (n - p)R^2 \tag{2.23}$$

dengan:

BG = statistik *Breusch-Godfrey*

n = jumlah total pengamatan

p = jumlah lag dari residual yang digunakan pada regresi tambahan

R^2 = koefisien determinasi dari regresi tambahan yang melibatkan residual dan lag residual

Hipotesis pengujian uji autokorelasi menggunakan uji *Breusch-Godfrey,* yaitu:

1. H_0: Tidak ada autokorelasi pada residual model.
2. H_1: Terdapat autokorelasi pada residual model.

Dasar pengambilan keputusan uji ini, yaitu:

1. Jika nilai $P\ value \geq 0{,}05$, maka artinya tidak terdapat autokorelasi.
2. Jika nilai $P\ value \leq 0{,}05$, maka artinya tidak terdapat autokorelasi.

2.7.4 Uji Non Heteroskedastisitas

Asumsi ini menyatakan bahwa variansi (penyebaran) dari *error* harus sama atau konstan untuk semua nilai variabel independen. Dengan kata lain, data tidak boleh menunjukkan pola penyebaran yang berubah-ubah atau berarti $var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$. Jika variansi *error* tidak konstan dan berubah, hal ini dikenal sebagai heterokedastisitas, yang merupakan pelanggaran terhadap asumsi OLS (Gunawan, dkk., 2024).

Dalam penelitian ini, uji yang digunakan untuk mendeteksi adanya heteroskedastisitas adalah uji *Breusch-Pagan.* Proses uji ini melibatkan regresi variabel independen terhadap nilai ρ yang diperoleh dengan membandingkan rata-rata *error* dan varian dari *error* atau $\frac{\mu^2}{\sigma^2}$, nilai σ^2 didapat dari $\sum \frac{\varepsilon_i^2}{n}$, dimana n adalah jumlah pengamatan (Sholihah, dkk., 2023).

Rumus statistik uji *Breusch-Pagan* adalah (Breusch & Pagan, 1979):

$$BP = n \times R^2 \tag{2.24}$$

dengan:

BG = statistik *Breusch-Pagan*

n = jumlah total pengamatan

R^2 = koefisien determinasi dari regresi antara residual kuadrat dan variabel independen

Hipotesis pengujian uji heteroskedastisitas menggunakan uji *Breusch-Pagan,* yaitu:

1. H_0: Tidak terjadi heteroskedastisitas pada residual model.
2. H_1: Terjadi heteroskedastisitas pada residual model.

Pengambilan keputusan uji ini, yaitu:

1. Jika nilai $P\ value \geq 0{,}05$, dapat dikatakan bawha tidak terjadi heteroskedastisitas.
2. Jika nilai $P\ value \leq 0{,}05$, dapat dikatakan terjadi heteroskedastisitas.

2.8 Kesesuaian Model Regresi

Sebelum melakukan analisis pada parameter model regresi, penting untuk memastikan bahwa model tersebut sudah cocok dengan data yang ada. Dengan memastikan model regresi sesuai, kita bisa lebih akurat dalam menilai pengaruh signifikan terhadap variabel dependen (Damanik, 2019). Terdapat beberapa pengukuran yang bisa digunakan untuk menilai apakah model yang dihasilkan sudah sesuai dengan data. Beberapa di antaranya adalah koefisien determinasi, uji simultan, dan uji parsial.

2.8.1 Koefisien Determinasi (R^2)

Menurut Damanik (2019), R^2 pada dasarnya mengukur sejauh mana variabel independen mampu menjelaskan variasi pada variabel dependen. Nilai R^2 berkisar antara 0 hingga 1. Jika nilai R^2 rendah, itu berarti variabel-variabel independen hanya sedikit menjelaskan variansi pada variabel dependen. Sebaliknya, jika nilai R^2 mendekati 1, berarti variabel-variabel independen hampir sepenuhnya menjelaskan variansi yang terjadi pada variabel dependen.

Menurut Setiawan & Kusrini (2010), R^2 memiliki beberapa sifat penting yang menjelaskan kegunaannya dalam analisis regresi, yaitu:

1. Nilai R^2 selalu positif karena merupakan perbandingan dari jumlah kuadrat.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.25}$$

dengan:

R^2 = koefisien determinasi

$\hat{y}_i$ = nilai prediksi variabel dependen pada objek ke$-i$

y_i = nilai pengamatan variabel depeden pada objek ke$-i$

$\bar{y}$ = rata-rata semua nilai pengamatan variabel dependen

n = $1, 2, 3, \dots, n$

2. Nilai $0 \leq R^2 \leq 1$

   $R^2 = 0$, artinya tidak ada pengaruh hubungan antara variabel independen dan variabel dependen, atau model regresi yang dibuat kurang tepat untuk meramalkan variabel dependen. $R^2 = 1$, artinya model regresi yang dibuat dapat meramalkan variabel dependen secara sempurna. Tidak ada standar yang pasti mengenai nilai R^2 yang dianggap baik atau sesuai. Ini karena nilai R^2 sangat dipengaruhi oleh jumlah variabel independen yang dimasukkan ke dalam model dan juga jumlah observasi yang digunakan.

2.8.2 Uji Parsial (Uji t)

Dalam analisis regresi uji t digunakan untuk mengevaluasi apakah setiap variabel independen secara individual memiliki pengaruh yang signifikan terhadap variabel dependen. Dengan kata lain, untuk mengetahui apakah setiap variabel independen dalam model benar-benar berkontribusi terhadap perubahan yang terjadi pada variabel tidak dependen (Romadhoni, dkk., 2022).

Jika hasil uji menunjukkan bahwa variabel independen berpengaruh signifikan, itu berarti variabel tersebut penting dalam model dan mempengaruhi variabel yang kita coba prediksi. Sebaliknya, jika tidak signifikan, maka variabel independen tersebut mungkin tidak memiliki pengaruh yang berarti pada variabel dependen dan bisa dipertimbangkan untuk dikeluarkan dari model. Menurut Setiawan & Kusrini (2010), hipotesis dari pengujian secara parsial yaitu:

$H_0 = \beta_i = 0$

$H_1 = \beta_i \neq 0 = 1, 2, 3, \ldots, k$

Menggunakan uji t, statistik pengujian yang digunakan, yaitu:

$$t_{hitung} = \frac{\hat{\beta}_i}{SE\hat{\beta}_i} \tag{2.26}$$

dengan:

$\hat{\beta}_i$ = estimasi koefisien regresi variabel independent pada objek ke$-i$

$SE\hat{\beta}_i$ = standar *error* estimasi koefisien regresi $\hat{\beta}_i$

Dasar pengambilan keputusan uji ini yaitu, jika nilai $P\ value < 0{,}05$, maka H_0 akan ditolak. Ini menunjukkan bahwa variabel independen ke$-i$ memiliki pengaruh signifikan terhadap variabel dependen. Dan apabila jika nilai $P\ value > 0{,}05$, maka H_0 tidak akan ditolak, yang berarti variabel independen ke$-i$ tidak memiliki pengaruh signifikan terhadap variabel dependen.

2.8.3 Uji Simultan (Uji *F*)

Dalam regresi, kita sering kali memiliki lebih dari satu variabel independen. Kita juga perlu tahu apakah gabungan semua variabel independen secara bersama-sama signifikan dalam menjelaskan variasi pada variabel dependen, di sinilah uji F digunakan (Romadhoni, dkk., 2022).

Jika nilai F signifikan ini menunjukkan bahwa semua variabel independen secara bersama-sama memberikan kontribusi signifikan dalam mempengaruhi variabel dependen, sehingga model yang melibatkan variabel-variabel tersebut layak digunakan untuk prediksi atau penjelasan variabel dependen. Namun, jika nilai F tidak signifikan, berarti variabel-variabel independen dalam model mungkin tidak memberikan informasi yang cukup berarti untuk memprediksi variabel dependen, sehingga model tersebut mungkin perlu diperbaiki dengan menambah atau menghapus variabel. Menurut Setiawan & Kusrini (2010), hipotesis dari pengujian secara simultan,yaitu:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1: minimal ada satu $\beta_j \neq 0,\ j = 1, 2, 3, \ldots, k$

Statistik pengujian yang digunakan untuk uji F, yaitu:

$$F_{hitung} = \frac{\frac{R^2}{k}}{\frac{(1-R^2)}{(n-k-1)}} \tag{2.27}$$

dengan:

R^2 = koefisien determinasi

k = jumlah variabel independen

n = jumlah total pengamatan

Jika Keputusan $P\ value < 0{,}05$, , maka H_0 akan ditolak. Ini berarti setidaknya ada satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Sebaliknya, jika keputusan $P\ value > 0{,}05$, maka H_0 tidak akan ditolak, yang artinya tidak ada variabel independen yang berpengaruh signifikan terhadap variabel dependen.

2.9 Model Dinamis

Menurut Setiawan & Kusrini (2010), model dinamis adalah model yang menunjukkan bagaimana nilai-nilai dari masa lalu dapat mempengaruhi variabel yang akan kita prediksi, yaitu variabel dependen. Model ini memiliki kemampuan mengubah teori statis menjadi teori yang lebih realistis dengan memperhitungkan pengaruh waktu secara jelas. Dalam model ini, variabel independen mempengaruhi variabel dependen dengan mempertimbangkan nilai-nilai dari periode waktu sebelumnya (*lag*), sehingga kita bisa melihat bagaimana kejadian di masa lalu berdampak pada kondisi saat ini.

2.9.1 Model Distribusi *Lag* (DL)

Model distribusi *lag* (DL) adalah model yang digunakan dalam analisis data deret waktu untuk melihat pengaruh variabel independen tidak hanya pada waktu sekarang, tetapi juga pada waktu-waktu sebelumnya (masa lalu) melalui konsep *lag* atau keterlambatan. Dengan kata lain, model ini mempertimbangkan bahwa variabel independen dari beberapa periode sebelumnya juga bisa memengaruhi variabel dependen, dan pengaruh ini dapat berlangsung untuk beberapa periode ke belakang (Dewi, dkk., 2023). Dalam persamaan umum, model ini dinyatakan sebagai:

$$Y_t = \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_k X_{t-k} + \varepsilon_t \qquad (2.28)$$

dengan:

Y_t = variabel dependen pada waktu ke$-t$ (periode waktu saat ini)

X_t = variabel independent pada waktu ke$-t$ (periode waktu saat ini)

X_{t-k} = variabel independent pada waktu ke$-\ t-k$ (periode waktu sebelumnya)

β_0 = konstanta atau intersep

δ_k = koefisien regresi nilai parameter variabel independent ke$-k$

ε_t = galat atau *error* pada waktu ke$-t$, dimana $\varepsilon_i \sim N(0, \sigma^2)$

k = $1, 2, 3, \dots, n$

Model distribusi lag diklasifikasikan menjadi dua jenis, yaitu:

1. Model *Lag* Tak Terhingga (*Infinite Lag*)

   Pada model ini, panjang *lag* tidak diketahui atau diasumsikan sangat panjang, artinya pengaruh variabel bebas di masa lalu dapat bertahan untuk waktu yang tidak terbatas. Bentuk umumnya, sebagai berikut:

   $$Y_t = \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \varepsilon_t \tag{2.29}$$

   Pada kasus ini, kita tidak mengetahui secara pasti seberapa jauh ke belakang nilai dari variabel independen pada periode-periode sebelumnya berpengaruh terhadap variabel dependen saat ini.

2. Model *Lag* Terhingga (*Finite Lag*)

   Pada model ini, panjang *lag* diketahui secara pasti, yang berarti kita tahu berapa periode waktu sebelumnya yang mempengaruhi variabel dependen. Misalnya, jika kita menggunakan *lag* hingga k periode, maka model hanya memasukkan pengaruh variabel independen dari waktu ke$-t$ (saat ini), serta k periode sebelumnya seperti, X_{t-1}, X_{t-2}, …, X_{t-k}. Jadi, panjang *lag* ini dibatasi sampai periode ke$-k$. Bentuk umumnya, sebagai berikut:

   $$Y_t = \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_k X_{t-k} + \varepsilon_t \tag{2.30}$$

   Model ini disebut "*lag* terhingga" karena panjang *lag* diketahui dan dibatasi pada jumlah periode tertentu.

2.10 Pendekatan Metode Almon Terhadap Distribusi *Lag*

Menurut Sumodiningrat (1995), dalam analisis distribusi *lag*, terdapat dua pendekatan utama untuk mengestimasi model, salah satunya yaitu metode pendekatan Almon. Pendekatan Almon memiliki fleksibilitas tinggi dalam menentukan pola atau bentuk skema *lag*. Almon tidak membuat asumsi sebelumnya tentang bagaimana pengaruh variabel masa lalu terhadap variabel dependen berkurang seiring waktu. Pendekatan ini memungkinkan pola pengaruh dari variabel *lag* lebih bervariasi dan tidak harus menurun secara geometris. Metode Almon memiliki kemampuan untuk mengatasi masalah dengan model yang menggunakan banyak variabel *lag* tanpa mengurangi jumlah pengamatan. Dalam model dengan banyak *lag*, biasanya semakin banyak periode *lag* yang ditambahkan, semakin sedikit pengamatan yang tersedia untuk analisis. Namun, metode Almon mengurangi jumlah parameter yang perlu diestimasi, sehingga tidak mengorbankan pengamatan sebanyak metode lainnya.

Almon menggunakan polinomial untuk menggambarkan hubungan antara koefisien *lag* δ_i dan periode *lag*. Ini berarti bahwa metode ini mengasumsikan pengaruh variabel *lag* bisa diwakili oleh suatu fungsi polinomial. Polinomial ini membantu menggambarkan bentuk hubungan *lag* secara lebih fleksibel , artinya koefisien *lag* δ_i bisa naik dan bisa turun, karena bentuk kurva yang dihasilkan bisa lebih kompleks dan sesuai dengan pola yang ada dalam data (Gujarati & Porter, 2009).

Model yang digunakan dalam metode almon merupakan model dengan bentuk:

$$Y_t = \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_k X_{t-k} + \varepsilon_t \qquad (2.31)$$

Atau dapat ditulis,

$$Y_t = \beta_0 + \sum_{i=0}^{k} \delta_i X_{t-i} + \varepsilon_t$$

Berdasarkan pada teorema *Weir-Strass's*, Almon berpendapat bahwa koefisien *lag* δ_i dapat didekati menggunakan sebuah polinomial yang tergantung pada panjang beda waktu (*lag*). Panjang beda waktu ini dinyatakan dengan i, dan polinomial tersebut dapat memiliki derajat 0, 1, 2, dan seterusnya. Secara umum, polinomial untuk δ_i dinyatakan sebagai berikut:

$$\delta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_m i^m \qquad (2.32)$$

dengan:

δ_i = koefisien lag untuk period ke-i

α_0 = konstanta atau intersep dari polinomial

$\alpha_2 i^2$ = komponen linear dari polinomial

$\alpha_2 i^2$ = komponen kuadratik dari polinomial

$\alpha_m i^m$ = komponen berderajat m

Sebagai contoh, jika δ_i dituliskan dalam bentuk polinomial derajat kedua, maka akan dihasilkan persamaan baru, yaitu:

$$\begin{aligned} Y_t &= \beta_0 + \sum_{i=0}^{k}(\alpha_0 + \alpha_1 i + \alpha_2 i^2)X_{t-k} + \varepsilon_t \\ &= \beta_0 + \alpha_0 \sum_{i=0}^{k} X_{t-k} + \alpha_1 \sum_{i=0}^{k} iX_{t-k} + \alpha_2 \sum_{i=0}^{k} i^2 X_{t-k} + \varepsilon_t \end{aligned} \tag{2.33}$$

dengan,

$$Z_{0t} = \sum_{i=0}^{k} X_{t-k},\ Z_{1t} = \sum_{i=0}^{k} iX_{t-k},\ Z_{2t} = \sum_{i=0}^{k} i^2 X_{t-k} \tag{2.34}$$

persamaan menjadi,

$$Y_t = \beta_0 + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \varepsilon_t \tag{2.35}$$

Koefisien pada persamaan (2.35) dapat diestimasi menggunakan metode kuadrat terkecil. Jika dinyatakan dalam bentuk persamaan regresi perkiraan, maka akan ditulis sebagai berikut:

$$\widehat{Y_t} = \hat{\beta}_0 + \hat{\alpha}_0 Z_{0t} + \hat{\alpha}_1 Z_{1t} + \hat{\alpha}_2 Z_{2t} + \varepsilon_t \tag{2.36}$$

Estimasi untuk $\hat{\beta}_0$ dan $\hat{\alpha}_i$ akan memiliki sifat-sifat yang diinginkan selama residual memenuhi asumsi-asumsi dari model regresi linear klasik. Berdasarkan persamaan (2.36), koefisien *lag* $\hat{\delta}_i$ dapat dihitung dengan menggunakan persamaan (2.32) sebagai berikut:

$$\begin{aligned} \hat{\delta}_0 &= \hat{\alpha}_0 \\ \hat{\delta}_1 &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 \\ \hat{\delta}_2 &= \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2 \\ \hat{\delta}_3 &= \hat{\alpha}_0 + 3\hat{\alpha}_1 + 9\hat{\alpha}_2 \\ &\vdots \\ \hat{\delta}_k &= \hat{\alpha}_0 + k\hat{\alpha}_1 + k^2\hat{\alpha}_2 \end{aligned} \tag{2.37}$$

Metode Almon hanya bisa diterapkan jika dua hal diketahui sebelumnya. Pertama, harus diketahui panjang *lag* dalam model aslinya, yaitu nilai dari *k*. Kedua, pola umum dari koefisien δ juga harus dipahami, sehingga kita bisa menentukan derajat polinomial yang sesuai dalam model tersebut. Setelah menentukan panjang *lag* dan derajat polinomial kita dapat menghitung nilai Z. Artinya, sebelum menerapkan metode Almon, kita perlu mengetahui seberapa lama efek dari variabel independen mempengaruhi variabel dependen (panjang *lag*). Selain itu, pola perubahan koefisien δ juga harus diperkirakan agar kita bisa memilih derajat polinomial yang tepat untuk mewakili distribusi *lag*.

2.11 Indeks Pembangunan Manusia (IPM)

Menurut Badan Pusat Statistik (2024), pembangunan manusia merupakan konsep yang lebih luas daripada pembangunan ekonomi konvensional. Pembangunan manusia menekankan pentingnya partisipasi masyarakat dalam proses pembangunan, di mana setiap individu harus memiliki kesempatan yang sama untuk meraih kesejahteraan. Pembangunan nasional yang sukses tidak dapat dicapai tanpa pembangunan manusia yang kuat, dan sebaliknya, proses pembangunan manusia akan terhambat tanpa adanya pembangunan nasional yang berkelanjutan. Oleh karena itu, kualitas manusia menjadi kunci penting dalam mencapai pembangunan yang berkelanjutan.

Nilai capaian pembangunan manusia di suatu negara atau wilayah diukur menggunakan IPM. IPM merupakan salah satu indikator yang digunakan untuk mengevaluasi kemajuan dalam kualitas hidup masyarakat. IPM memberikan gambaran yang lebih menyeluruh tentang kesejahteraan penduduk dibandingkan hanya menggunakan indikator ekonomi seperti PDB (Produk Domestik Bruto) per kapita. Hal ini karena IPM tidak hanya memperhatikan kekayaan ekonomi, tetapi juga memperhitungkan kualitas hidup masyarakat dalam hal kesehatan dan pendidikan.

IPM mencakup tiga aspek utama yang dianggap krusial dalam menentukan kualitas hidup yang terus berkembang untuk mendukung produktivitas individu, yaitu harapan hidup yang panjang dan kesehatan yang baik, tingkat pengetahuan, serta kemampuan untuk memenuhi kebutuhan hidup yang layak. Ketiga aspek tersebut diberikan bobot yang setara dalam bentuk indeks, kemudian nilai dari masing-masing aspek digabungkan untuk menghasilkan angka IPM secara keseluruhan.

Berdasarkan data dari Badan Pusat Statistik Provinsi Lampung (2025), IPM di Provinsi Lampung menunjukkan tren peningkatan, yaitu pada tahun 2020 memiliki presentase sebesar 69,69, pada tahun 2021 memiliki presentase sebesar 69,90, pada tahun 2022 memiliki presentase sebesar 70,45, pada tahun 2023 memiliki presentase sebesar 71,15 dan pada tahun 2024 memiliki presentase sebesar 71,81.

IPM dihitung dengan mengombinasikan tiga indikator utama, yaitu:

$$IPM = \sqrt[3]{I_{kesehatan} + I_{pendidikan} + I_{pengeluaran}} \tag{2.38}$$

dengan:

IPM = indeks pembangunan manusia

$I_{kesehatan}$ = indeks kesehatan

$I_{pendidikan}$ = indeks pendidikan

$I_{pengeluaran}$ = indeks pengeluaran

2.12 Harapan Lama Sekolah (HLS)

HLS merupakan indikator yang digunakan untuk menggambarkan sejauh mana masyarakat memiliki kesempatan untuk mengakses pendidikan formal. HLS menggambarkan durasi pendidikan yang diharapkan dapat ditempuh oleh individu tepatnya oleh anak yang berumur 7 tahun pada suatu wilayah atau negara untuk dapat melanjutkan pendidikan formal mereka di masa depan, dan ini berkaitan erat dengan kebijakan pendidikan serta aksesibilitas sistem pendidikan yang ada. Indikator ini tidak hanya mengukur rata-rata tingkat pencapaian pendidikan yang sudah tercapai, tetapi juga memperkirakan kesempatan yang dimiliki oleh generasi mendatang untuk mendapatkan pendidikan yang lebih baik dan lebih lama.

HLS dihitung dengan cara berikut (Badan Pusat Statistik, 2024):

$$HLS_a^t = \sum_{i=a}^{t} \frac{E_i^t}{P_i^t} \tag{2.39}$$

dengan:

HLS_a^t = harapan lama sekolah pada usia a dan pada tahun t

E_i^t = partisipasi sekolah penduduk pada usia i dan pada tahun t

P_i^t = populasi penduduk pada usia i yang bersekolah pada tahun t

i = usia $(a, a+1, \dots, n)$

2.13 Rata-Rata Lama Sekolah (RLS)

Menurut Badan Pusat Statistik (2024), RLS menjadi salah satu indikator dalam penyusunan IPM, khususnya dalam dimensi pendidikan, bersama dengan indikator HLS. RLS adalah indikator yang digunakan untuk menggambarkan tingkat pendidikan rata-rata yang telah ditempuh oleh penduduk usia 25 tahun ke atas dalam satu wilayah tertentu. Indikator ini memberikan gambaran mengenai capaian pendidikan formal yang telah dijalani oleh masyarakat dan mencerminkan hasil nyata dari sistem pendidikan suatu daerah di masa lalu.

Indikator ini penting karena kelompok usia 25 tahun ke atas dianggap telah menyelesaikan sebagian besar masa pendidikan formalnya, sehingga RLS mampu merepresentasikan kondisi nyata dari pendidikan masyarakat. Berbeda dengan indikator HLS yang bersifat proyeksi dan mencerminkan harapan terhadap lama pendidikan anak-anak usia sekolah, RLS lebih menekankan pada capaian pendidikan yang benar-benar telah ditempuh oleh penduduk. Oleh karena itu, RLS sering disebut sebagai indikator pengalaman pendidikan, sedangkan HLS sebagai indikator potensi pendidikan.Berikut adalah rumus untuk menghitung RLS:

$$RLS = \frac{\sum_{i=1}^{n}(P_i \times T_i)}{\sum_{i=1}^{n} P_i} \tag{2.40}$$

dengan:

RLS = rata-rata lama sekolah

P_i = jumlah penduduk usia 25 tahun ke atas pada tingkat pendidikan ke-i

T_i = tahun sekolah yang setara untuk tingkat pendidikan ke-i

n = jumlah kategori tingkat pendidikan

2.14 Pengeluaran Per Kapita (PPP)

Badan Pusat Statistik (2024) mengatakan aspek standar hidup yang layak dapat mengukur kualitas kehidupan masyarakat, aspek ini menggunakan pengeluaran per kapita karena dapat mencerminkan tingkat pendapatan dan kesejahteraan masyarakat secara lebih merata di berbagai wilayah dan waktu. Pengeluaran per kapita adalah jumlah total pengeluaran yang dilakukan oleh seluruh anggota rumah tangga dalam suatu periode tertentu, yang kemudian dibagi dengan jumlah penduduk (baik mengonsumsi maupun tidak).

Untuk menghitung rata-rata pengeluaran per kapita yang disesuaikan dengan paritas daya beli menggunakan rumus yang dikembangkan. oleh Rao, yaitu:

$$PPP_j = \prod_{i=1}^{m} \frac{p_{ij}}{p_{ik}} \tag{2.41}$$

dengan:

PPP_j = paritas daya beli di wilayah j

p_{ij} = harga komoditas i di wilayah j

p_{ik} = harga komoditas i di wilayah acuan k

m = jumlah komoditas yang digunakan dalam perhitungan (96 komoditas).

Setelah mendapatkan nilai paritas daya beli untuk setiap wilayah, langkah selanjutnya adalah mencari pengeluaran per kapita yang disesuaikan dengan paritas daya beli tersebut, yaitu:

$$PPK' = \frac{PPK}{PPP_j} \tag{2.42}$$

dengan:

PPK' = pengeluaran perkapita disesuaikan

PPK = pengeluaran per kapita harga konstan

PPP_j = paritas daya beli di wilayah j

2.15 Umur Harapan Hidup (UHH)

Berdasarkan Badan Pusat Statistik (2024), UHH atau sering disimbolkan dengan e_0 adalah salah satu indikator utama yang digunakan untuk merepresentasikan dimensi umur panjang dan hidup sehat dalam pengukuran IPM. UHH menunjukkan perkiraan rata-rata lamanya waktu hidup (dalam tahun) yang akan dijalani oleh seseorang sejak lahir, dengan asumsi bahwa pola angka kematian yang berlaku saat ini tetap konstan sepanjang hidupnya.

Indikator ini memberikan gambaran tentang derajat kesehatan masyarakat di suatu wilayah. Semakin tinggi nilai UHH, maka semakin baik pula kondisi kesehatan, gizi, layanan kesehatan, dan sanitasi yang diterima oleh penduduk di daerah tersebut. Sebaliknya, nilai UHH yang rendah mengindikasikan adanya masalah serius dalam sistem kesehatan, lingkungan, atau tingkat kesejahteraan masyarakat.

Secara teknis, perhitungan UHH melibatkan penyusunan tabel kehidupan (*life table*) berdasarkan data kematian dan penduduk. Namun secara ringkas, UHH dihitung dengan rumus berikut:

$$e_0 = \frac{T_0}{l_0} \tag{2.43}$$

dengan:

e_0 = umur harapan hidup

T_0 = total tahun hidup seluruh kohor sejak lahir

l_0 = jumlah penduduk hidup pada usia 0 (biasanya 100.000)

Karena metode ini cukup teknis, BPS biasanya langsung menyajikan e_0 hasil olahan dari tabel kehidupan, dan bukan dari penghitungan manual oleh pengguna umum.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Kegiatan penelitian ini dilakukan pada semester ganjil tahun ajaran 2024/2025 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Dalam penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (2024) pada laman https://lampung.bps.go.id/id. Data ini merupakan data deret waktu yang mencakup informasi tahunan di Provinsi Lampung mengenai angka harapan lama sekolah (tahun) sebagai variabel $X1$, rata-rata lama sekolah (tahun) sebagai variabel $X2$, pengeluaran per kapita (ribu rupiah/orang/tahun) sebagai variabel $X3$, angka harapan hidup (tahun) sebagai variabel $X4$ serta indeks pembangunan manusia (IPM) sebagai variabel Y. Rentang waktu pengambilan data dimulai dari tahun 2005 hingga tahun 2024.

3.3 Metode Penelitian

Kegiatan penelitian ini dilaksanakan melalui pendekatan studi pustaka, yaitu dengan mengumpulkan dan menganalisis informasi dari berbagai sumber tertulis yang relevan, seperti buku teks pendukung dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk menyelesaikan penelitian ini , penulis menggunakan perangkat lunak *Rstudio* yang memudahkan dalam melakukan perhitungan dan memastikan hasil yang diperoleh akurat. Berikut adalah langkah-langkah yang dilakukan dalam penelitian ini:

1. Melakukan uji statistik deskriptif untuk melihat gambaran umum dari data yang diteliti.
2. Melakukan uji stasioneritas data pengamatan menggunakan uji *Kwiatkowski Phillips Schmidt Shin* (KPSS). Jika uji stasioner terpenuhi maka penelitian dapat dilanjutkan, sedangkan uji stasioner tidak terpenuhi akan dilakukan *differencing* data.
3. Menentukan panjang *lag* maksimum (*k)* dan derajat polinomial (*m*). Dengan membandingkan nilai AIC dan BIC pada model *lag* yang berbeda.
4. Menghitung nilai-nilai Z_{mt}, dimana:
$$Z_{0t} = \sum_{i=0}^{k} X_{t-i},\ Z_{1t} = \sum_{i=0}^{k} iX_{t-i},\ Z_{2t} = \sum_{i=0}^{k} i^2 X_{t-i} \tag{3.1}$$
5. Mencari model estimasi dengan metode Almon melalui regresi linear, yang dinyatakan sebagai:
$$\widehat{Y_t} = \hat{\alpha} + \hat{\alpha}_0 Z_{0t} + \hat{\alpha}_1 Z_{1t} + \hat{\alpha}_2 Z_{2t} + \cdots + \hat{\alpha}_m Z_{mt} \tag{3.2}$$
6. Melakukan pengujian asumsi klasik pada model transformasi almon, menggunakan uji normalitas dengan melihat nilai hasil uji *Jarque Berra* (JB) dan *QQ* plot residual, uji non-autokorelasi dengan nilai hasil uji *Breusch-Godfrey*, uji homoskedastisitas dengan nilai hasil uji *Breusch-Pagan* dan uji multikolinearitas menggunakan nilai hasil *Variance Inflation Factor* (VIF). Jika asumsi klasik terpenuhi akan dilakukan uji signifikansi parameter secara parsial dan simultan. Jika tidak terpenuhi akan dilakukan transformasi data.
7. Menguji signifikansi parameter secara simultan (uji F) dan parsial (uji t).
8. Mencari nilai $\hat{\delta}_i$ menggunakan :
$$\hat{\delta}_0 = \hat{\alpha}_0,\ \ \hat{\delta}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2, \dots, \hat{\delta}_k = \hat{\alpha}_0 + k\hat{\alpha}_1 + k^2\hat{\alpha}_2 \tag{3.3}$$
9. Menentukan model dinamis distribusi *lag* dengan menggunakan transformasi Almon.
10. Menginterpretasikan hasil model yang dihasilkan.

V. KESIMPULAN

Berdasarkan hasil analisis *Almon distributed lag models* pada indikator IPM, diperoleh kesimpulan sebagai berikut:

1. Model hasil modifikasi dengan panjang *lag* 5 dan derajat polinomial 2 memiliki nilai AIC dan BIC terkecil, yaitu sebesar -27,59519 dan -22,48273. Hal ini menunjukkan bahwa model regresi dapat dianggap lebih baik dalam hal keseimbangan antara *goodness of fit* (seberapa baik model menyesuaikan data) dan kompleksitas model (jumlah parameter yang digunakan). sehingga dipilih sebagai model terbaik untuk mengestimasi distribusi *lag* metode Almon. Bentuk model estimasi distribusi *lag* Almon dengan panjang *lag* 5 dan derajat polinomial 2 dalam bentuk sebagai berikut:

$$Y_t = 0{,}18160 + 0{,}76280Z_{ot} - 0{,}53950Z_{1t} + 0{,}08276Z_{2t}$$
$$+0{,}00077z_{0t} - 0{,}00054Z_{1t} + 0{,}00008Z_{2t} + \varepsilon_t$$

2. Model dinamis *Almon distributed lag* yang diperoleh dalam penelitian ini adalah sebagai berikut:

$$Y_t = 0{,}18160 + (0{,}76280X1_t + 0{,}30606X1_{t-1} + 0{,}01484X1_{t-2}$$
$$-0{,}11086X1_{t-3} - 0{,}07104X1_{t-4} + 0{,}13430X1_{t-5}) +$$
$$(0{,}00077X3_t + 0{,}00031X3_{t-1} + 0{,}00001X3_{t-2} -$$
$$0{,}00013X3_{t-3} - 0{,}00011X3_{t-4} + 0{,}00007X3) + \varepsilon_t$$

3. Dari empat indikator IPM, yaitu harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita, dan umur harapan hidup, hanya harapan lama sekolah dan pengeluaran per kapita yang berpengaruh terhadap IPM di Provinsi Lampung, baik jangka pendek maupun panjang. Keduanya berdampak positif di awal, namun menurun di jangka menengah. Tetapi, dalam jangka panjang, keduanya kembali memberikan kontribusi positif terhadap peningkatan IPM.

DAFTAR PUSTAKA

Alamsyah, Iqbal, F., Rut, E., Salwa, A., & Darnah, A. N. 2022. *Analisis Regresi Data Panel untuk Mengetahui Faktor yang Memengaruhi Jumlah Penduduk Miskin di Kalimantan Timur*, hlm. 254-266. In Prosiding Seminar Nasional Matematika dan Statistika, Kalimantan Timur.

Almumtazah, N., Azizah, N., Putri, Y. L., & Novitasari, D. C. 2021. Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana. *Jurnal Ilmiah Matematika dan Terapan.* **1**(1): 31-40.

Aqibah, M., Putu, S. N., & Sumarjaya, W. 2020. Model Dinamis Autoregressive Distributed Lag (Studi Kasus: Pengaruh Kurs Dolar Amerika dan Inflasi Terhadap Harga Saham Tahun 2014-2018). *E-Journal Matematika.* **9**(4): 240-250.

Badan Pusat Statistik Lampung. 2024. *Harapan Lama Sekolah.* https://lampung.bps.go.id/id. Diakses pada 10 Januari 2025.

Badan Pusat Statistik Lampung. 2024. *Indeks Pembangunan Manusia.* https://lampung.bps.go.id/id. Diakses pada 10 Januari 2025.

Badan Pusat Statistik Lampung. 2024. *Pengeluaran Per Kapita.* https://lampung.bps.go.id/id. Diakses pada 10 Januari 2025.

Badan Pusat Statistik. 2024. *Indeks Pembangunan Manusia 2023.* Badan Pusat Statistik, Jakarta

Badan Pusat Statistik Lampung. 2025. *Provinsi Lampung Dalam Angka 2024.* BPS Provinsi Lampung, Bandar Lampung.

Basuki, A. T., & Prawoto, N. 2019. *Analisis Regresi: Dalam Penelitian Ekonomi dan Bisnis.* Rajawali Pers, Jakarta.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. 2015. *Time Series Analysis: Forecasting and Control*. 5th edition. NJ: Jhon Wiley & Sons, Hoboken.

Breusch, T. S. 1978. Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers.* **17**(31): 334-355.

Breusch, T. S., & Pagan, A. R. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*. **47**(5): 1287–1294.

Dewi, A. P., Martha, S., & Perdana, H. 2023. Model Autoregressive Distributed Lag dengan Metode Kyock. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapan.* **1**(4): 89-96.

Damanik, B. E. 2019. Pengaruh Fasilitas dan Lingkungan Belajar terhadap Motivasi Belajar. *Publikasi Pendidikan.* **9**(1*)*: 46-52.

Firdaus, M. 2019. *Ekonometrika Suatu Pendekatan Aplikatif.* Bumi Aksara, Jakarta.

Fitriyah, Z., Irsalina, S., & Widodo, E. 2021. Analisis Faktor yang Berpengaruh terhadap IPM Menggunakan Regresi Linear Berganda. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika.* **2**(3): 282-291.

Gujarati, D. N., & Porter, D. C. 2009. *Basic Econometrics.* 5th edition. McGraw-Hill, New York.

Gunawan, R. A., Zulkarmain, D. P., & Arianto, S. T. 2024. Perbandingan Metode Ordinary Least Square (OLS) dan Metode Partial Least Square (PLS) untuk Mengatasi Multikolinearitas. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial.* **1**(6): 97-103.

Jarque, C. M., & Bera, A. K. 1980. Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals. *Economics Letters*. **6**(3): 255–259.

Kusumawardhani, R., Rizqiena, Z. D., & Astuti, S. P. 2021. *Ekonometrika Suatu Pengantar.* Gerbang Media, Yogyakarta.

Lukman, A. F., & Kibria, G. B. 2021. Almon-KL Estimator for The Distributed Lag Model. *Arab Journal of Basic and Applied Sciences.* **28**(1): 406-412.

Marsani, M. F., & Shabri, A. 2019. Random Walk Behaviour of Malaysia Share Return in Different Economic Circumstance. *MATEMATIKA: MJIAM*. **35**(3): 297–308.

Muhartini, A. A., Sahroni, O., Rahmawati, S. D., Febrianti, T., & Mahmuda, I. 2021. Analisis Peramalan Jumlah Penerimaan Mahasisma Baru dengan Metode Regresi Linear Sederhana. *Jurnal Bayesian: Jurnal Ilmiah dan Ekonometrika.* **1**(1): 17-23.

Nasyri, I. A., Harsono, I., Yuniarti, T., Sutanto, H., & Suprapti, I. A. P. 2024. Pengaruh Komponen Indeks Pembangunan Manusia Terhadap Pertumbuhan Ekonomi Provinsi Nusa Tenggara Barat Tahun 2018-2022. *MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*. **1**(2): 96-109.

O'brien, R. M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*. **41**(5): 673-690.

Pham, H. 2019. A New Criterion for Model Selection. *Mathematics.* **7**(10): 1020.

Prahasta, K. D., Isnaniah, U. N., Nurlaily, D., Nurhayati, F., & Silfiani, M. 2023. Analisis Pengaruh Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah dan Tingkat Pengangguran Terbuka terhadap Indeks Pembangunan Manusia berdasarkan Kabupaten/Kota di Pulau Kalimantan Tahun 2022. *Equiva Journal*. **1**(2):33-40 .

Ramadanisa, N., & Triwahyuningtyas, N. 2022. Analisis Faktor Yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Lampung. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*. **1**(7): 1049-1062.

Romadhoni, R., Nasution, R. Y., & Anam, K. 2022. Analisis Faktor Hasil Produksi Kelapa Sawit Menggunakan Regresi Linier Berganda Studi Kasus: Koperasi Unit Desa (KUD) Setia Kawan Desa Koto Damai. *Formosa Journal of Science and Technology.* **1**(4): 217-234.

Setiawan, & Kusrini, D. E. 2010. *Ekonometrika.* Gadjah Mada Press, Yogyakarta.

Sholihah, S. M., Aditiya, N. Y., & Maghfiroh, E. S. 2023. Konsep Uji Asumsi Klasik pada Regresi Linier Berganda. *Jurnal Riset Akutansi Soedirman.* **2**(2): 102-110.

Sihotang, M. A., Amelia, A., & Saumi, F. 2019. Penerapan Metode Kuadrat Terkecil dalam Menentukan Saldo JHT (Jaminan Hari Tua) dalam Waktu 3 Tahun ke Depan. *Jurnal Gamma-Pi*. **1**(1): 25–28.

Sitorus, D. P. N., Widiarti, W., Sutrisno, A., & Usman, M. 2023. Pemodelan Dinamis Distributed Lag dengan Menggunakan Metode Koyck dan Metode Almon. *Jurnal Siger Matematika*. **17**(1): 17–22.

Sujatmiko, F., Bawunuris, R., & Gunawati, O. 2021. Eksistensi Midle Income Trap Sebuah Kajian Empiris Tentang Fenomena Perlambatan Ekonomi di Indonesia. *Inspire Journal: Economics and Development Analysis*. **1**(1): 13-30.

Sumodiningrat, D. G. 1995. *Ekonometrika Pengantar.* Gadjah Mada University Press, Yogyakarta.

Syafira, R., Khoirudin, R., & A'yun, I. Q. 2024. Pengaruh Dana Otonomi Khusus, Pengeluaran Per Kapita, Umur Harapan Hidup Saat Lahir, Harapan Lama Sekolah, dan Rata-Rata Lama Sekolah terhadap Indeks Pembangunan Manusia di Provinsi Papua Tahun 2014-2022. *Jurnal Simki Economic*. **7**(1): 96-105.

Virgantari, F., & Rahayu, W. 2021. Pendugaan Parameter Model Distribusi Lag Pola Polinomial Menggunakan Metode Almon. *BAREKENG: Jurnal Ilmu Matematika dan Terapan.* **15**(4): 761-772.

Wisudaningsi, B. A., Arofah, I., & Belang, K. A. 2019. Pengaruh Kualitas Pelayanan dan Kualitas Produk Terhadap Kepuasan Konsumen dengan Menggunakan Metode Analisis Regresi Linear Berganda. *Statmat: Jurnal Statistika dan Matematika.* **1**(1): 103-116.