MODEL REGRESI DATA PANEL DENGAN *STANDARD ERROR* BERDASARKAN HASIL ESTIMASI MATRIKS VARIANSI *CLUSTER ROBUST* PADA KASUS TINGKAT PENGANGGURAN TERBUKA DI SULAWESI SELATAN

SRI RAHAYU YUSRI
H051201028

M STUDI STATISTIKA DEPARTEMEN STATISTIKA
S MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

MODEL REGRESI DATA PANEL DENGAN *STANDARD ERROR* BERDASARKAN HASIL ESTIMASI MATRIKS VARIANSI *CLUSTER ROBUST* PADA KASUS TINGKAT PENGANGGURAN TERBUKA DI SULAWESI SELATAN

SRI RAHAYU YUSRI
H051201028

Skripsi

sebagai salah satu syarat memperoleh gelar Sarjana Statistika

Program Studi Statistika

pada

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

SKRIPSI

MODEL REGRESI DATA PANEL DENGAN *STANDARD ERROR* BERDASARKAN HASIL ESTIMASI MATRIKS VARIANSI *CLUSTER ROBUST* PADA KASUS TINGKAT PENGANGGURAN TERBUKA DI SULAWESI SELATAN

yang disusun dan diajukan oleh

SRI RAHAYU YUSRI
H051201028

Skripsi,

Telah dipertahankan di depan Panitian Ujian Sarjana Statistika pada 20 Juni 2024 dan dinyatakan telah memenuhi syarat kelulusan

pada

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

Mengesahkan:	Mengetahui:
Pembimbing tugas akhir,	Ketua Program Studi
	KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN DEPARTEMEN STATISTIKA FAKULTAS MIPA
.,	Dr. Anna Islamiyati, S.Si., M.Si.
810 1 001	NIP. 19770808 200501 2 002

PERNYATAAN KEASLIAN SKRIPSI
DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Model Regresi Data Panel dengan *Standard Error* Berdasarkan Hasil Estimasi Matriks Variansi *Cluster Robust* pada Kasus Tingkat Pengangguran Terbuka di Sulawesi Selatan" adalah benar karya saya dengan arahan dari pembimbing skripsi saya (Drs. Raupong, M.Si.). 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 diterbitkan 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, 20 Juni 2024

METERAI TEMPEL 10000 8FFE0ALX249906836

Sri Rahayu Yusri
NIM H051201028

UCAPAN TERIMA KASIH

Assalamu'alaikum Warahmatullahi Wabaraktuh

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah SAW beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, skripsi ini dapat terampungkan atas bimbingan, diskusi dan arahan pembimbing dari bapak **Drs. Raupong, M.Si.** Penulis mengucapakan banyak terima kasih kepada beliau. Ucapan terima kasih juga penulis sampaikan kepada ibu **Anisa, S.Si., M.Si.,** dan ibu **Sitti Sahriman, S.Si., M.Si.,** selaku tim dosen penguji yang telah meluangkan waktu dalam memberikan saran dan kritikan yang membangun dalam penyempurnaan tugas akhir ini. Serta, ucapan terima kasih kepada pimpinan Hasanuddin beserta jajarannya, dosen pengajar, serta staf lingkup Departemen Statistika dan Fakultas MIPA Universitas Hasanuddin.

Penulis menyampaikan ucapan terima kasih yang mendalam dan penghargaan yang setinggi-tingginya, terutama kepada Ayahanda **Yusri Yusuf** dan Ibunda **Paida Umar** tercinta. Terima kasih atas kasih sayang, pengorbanan, dan dukungan tanpa batas yang telah menjadi pelita di kala ragu, semangat di kala lelah, dan kekuatan di kala terjatuh. Semoga Allah *Subhanahu Wa Ta'ala* membalas segala kebaikan dan pengorbanan Bapak dan Mama dengan limpahan pahala dan kebahagiaan. Amin. Serta keluarga besar penulis, terima kasih atas doa mulia dan dukungannya selama ini.

Ucapan terima kasih dengan penuh keikhlasan kepada sahabat tercinta penulis sejak kecil, **Annisa Indah Nur** dan **Happy Helmalia Kartika Armys** yang telah menjadi sahabat dan saudara terbaik. Ucapan yang serupa juga kepada **Liza, Nahda, Uci, Ara, Aslisha, Irma, Linda, Azal,** dan **Najla** atas kebersamaan serta kebaikannya menjadi sosok guru bagi penulis. Kepada Ciwi-Ciwi Among terkhusus kepada **Aliah, Nadia, Nahla, Laurine, Rahmi, Krisna, Afika,** dan **Aulia** atas suka, duka, dan kenangan indah yang dilalui bersama. Kepada teman-teman **Statistika 2020** yang senantiasa berproses bersama selama perkuliahan dan penyelesaian tugas akhir ini. Kepada sahabat seperjuangan dalam menyusun skripsi, **Ryan** terima kasih atas motivasi dan semangatnya. Serta, **BSI Maslahat** atas bantuan finansial dan kesempatan pengembangan diri yang telah diberikan. Semoga niat baik dan ketulusan BSI Maslahat dapat memberikan manfaat yang lebih luas lagi.

Akhirnya, penulis juga mengucapkan banyak terima kasih kepada seluruh pihak yang tidak dapat disebutkan satu persatu oleh penulis.

Warahmatullahi Wabarakatuh

Penulis,

Sri Rahayu Yusri

ABSTRAK

SRI RAHAYU YUSRI. **Model Regresi Data Panel dengan *Standard Error* Berdasarkan Hasil Estimasi Matriks Variansi *Cluster Robust* pada Kasus Tingkat Pengangguran Terbuka di Sulawesi Selatan** (dibimbing oleh Raupong).

Latar Belakang. Regresi data panel merupakan teknik untuk memodelkan pengaruh variabel bebas terhadap variabel terikat pada pengamatan dari individu yang sama dalam selang waktu tertentu. Pada umumnya, untuk mengestimasi koefisien regresi digunakan metode *Ordinary Least Square* (OLS) yang memiliki asumsi yang harus dipenuhi untuk memperoleh estimasi parameter yang bersifat *best, linear, dan unbiased*. Namun, dalam penerapan regresi data panel sering terjadi pelanggaran asumsi heteroskedastisitas dan autokorelasi. Hal ini akan menimbulkan bias pada *standard error* parameter. Hal ini dapat diatasi dengan menghitungkan Matriks Variansi *Cluster Robust.* **Tujuan**. Penelitian ini bertujuan untuk menerapkan *Cluster Robust Standard Error* dalam mengatasi *standard error* yang bias ketika terjadi heteroskedastisitas dan autokorelasi pada data panel. **Metode**. Analisis regresi data panel pada data kasus tingkat pengangguran terbuka di Provinsi Sulawesi Selatan dari tahun 2020 sampai 2022. **Hasil**. Hasil analisis menunjukkan perhitungan Matriks Variansi *Cluster Robust* menghasilkan nilai *standard error* yang lebih kecil 0.00018; 0.0312; dan 0.1682. Hal ini berimplikasi pada pengujian parsial yang sebelumnya hanya terdapat dua variabel signifikan secera parsial, setelah memperhitungkan Matriks Variansi *Cluster Robust* diperoleh tiga variabel signifikan secera parsial. Oleh karena itu, Matriks Variansi *Cluster Robust* direkomendasikan untuk menangani heteroskedastisitas dan autokorelasi dalam analisis regresi data panel karena menghasilkan nilai standard error yang kecil dalam analisis regresi data panel.

Kata Kunci: Matriks Variansi *Cluster Robust*, Regresi data panel, *Standard error*, Tingkat pengangguran terbuka.

ABSTRACT

SRI RAHAYU YUSRI. ***Panel Data Regression Model with Standard Error Based on the Estimation Result of Cluster Robust Variance Matrix in the Case of Open Unemployment Rate in South Sulawesi*** *(supervised by Raupong).*

Background. *Panel data regression is a technique to model the effect of independent variables on the dependent variable on observations of the same individual in a certain time interval. In general, to estimate regression coefficients, the Ordinary Least Square (OLS) method is used, which has assumptions that must be met to obtain parameter estimates that are best, linear, and unbiased. However, in the application of panel data regression, heteroscedasticity and autocorrelation assumptions are often violated. This will cause bias in the standard error of the parameters. This can be overcome by calculating the Cluster Robust Variance Matrix.* ***Objective.*** *This study aims to apply Cluster Robust Standard Error in overcoming biased standard errors when heteroscedasticity and autocorrelation occur in panel data.* ***Methods.*** *Panel data regression analysis on open unemployment rate case data in South Sulawesi Province from 2020 to 2022.* ***Results.*** *The analysis results show that the Cluster Robust Variance Matrix produces smaller standard error values of 0.00018; 0.0312; and 0.1682. This has implications for partial testing where previously there were only two partially significant variables, after taking into account the Cluster Robust Variance Matrix, three partially significant variables were obtained. Therefore, the Cluster Robust Variance Matrix is recommended for handling heteroscedasticity and autocorrelation in panel data regression analysis because it produces a small standard error value in panel data regression analysis.*

Keywords: *Cluster Robust Variance Matrix, Panel data regression, Standard error, Open unemployment rate.*

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Cluster	Pengumpulan data pada suatu pengamatan
Distribusi	Fungsi matematika yang menggambarkan probabilitas atau Frekuensi relatif dari kemungkinan nilai-nilai yang dapat diambil oleh suatu variabel acak.
Kurtosis	Ukuran tingkat kepuncakan dari suatu distribusi data.
Parameter	Nilai konstan yang mendeskripsikan karakteristik populasi.
Plot	Grafik atau diagram yang digunakan untuk menggambarkan atau memvisualisasikan data.
Skewness	Ukuran kemencengan dari suatu distribusi data.
Standard error	Standar deviasi dari distribusi sampling suatu statistik
Variabel *dummy*	Variabel numerik yang digunakan dalam analisis regresi untuk mewakili subkelompok

DAFTAR LAMBANG/SINGKATAN

Lambang/ Singkatan	Arti dan Penjelasan
β_0	Konstanta
$\hat{\beta}$	Estimator koefisien regresi
D_N	Matriks variabel *dummy* individu berukuran $NT \times N$
u_i	Galat pada pengamatan ke-i
u_{it}	komponen galat regresi dari individu ke-i dan periode waktu ke-t.
e	Komponen residual bersifat umum
v_i	Komponen spesifik dari individu
w_t	Komponen spesifik dari waktu
M_D	Matriks *idempotent*
$M_D X$	Matriks rata-rata variabel bebas
$M_D Y$	Matriks rata-rata variabel terikat
R^2_{MEU}	R-*square* Model Efek Umum
R^2_{MET}	R-*square* Model Efek Tetap
W	Statistik Hausman
χ^2	Distribusi *chi-squares*
μ	(mu)/Parameter lokasi (*location*).
σ	(sigma)/Parameter skala (*scale*).
λ	(lambda)/ Parameter-parameter yang terkait dengan *moment* (rataan, standar deviasi, *skewness*, dan kurtosis).
$\hat{\rho}_{ij}$	Estimasi sederhana korelasi galat *pair-wise*
α	Tingakat signifikansi.
N	Jumlah pengamatan individu
T	Jumlah pengamatan waktu
BLUE	*Best, Linear, and Unbiased Estimator*
CRSE	*Cluster Robust Standard Error*
MVCR	Matriks Variansi *Cluster Robust*
MEU	Model Efek Umum
MET	Model Efek Tetap
MEA	Model Efek Acak
LSDV	*Least Square Dummy Variable*
VIF	*Variance Inflation Factors*
BPS	Badan Pusat Statistik

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

Lampiran **Halaman**

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan teknik analisis statistika untuk mengetahui bentuk hubungan kausal antara variabel terikat dengan satu atau lebih variabel bebas. Salah satu jenis data menurut waktu pengambilannya yaitu data panel yang merupakan gabungan data lintas bagian (*cross-sectional*) dan data deret waktu (*time series*). Analisis regresi dengan menggunakan data panel disebut regresi data panel. Regresi data panel merupakan teknik untuk memodelkan pengaruh variabel bebas terhadap variabel terikat pada data panel (Sembodo, 2013). Hsiao (2022) memaparkan beberapa keunggulan penggunaan regresi data panel, yaitu: inferensi parameter model yang lebih akurat; menghasilkan estimasi yang lebih akurat untuk hasil individu; jumlah pengamatan yang banyak berimplikasi pada data yang lebih informatif, lebih bervariasi, kolinearitas antar variabel yang semakin berkurang, dan peningkatan derajat kebebasan sehingga dapat memperoleh hasil estimasi yang lebih efisien.

Pada umumnya, untuk mengestimasi koefisien regresi digunakan metode *Ordinary Least Square* (OLS) dengan meminimumkan jumlah kuadrat galat. Penggunaan metode OLS memiliki asumsi yang harus dipenuhi untuk mendapatkan estimasi parameter pada model yang bersifat *Best Linear Unbiased Estimator* (BLUE). Namun, dalam penerapan regresi data panel sering terjadi pelanggaran asumsi heteroskedastisitas dan autokorelasi. Heteroskedastisitas adalah kondisi galat dari model yang tidak memiliki variansi yang konstan. Sedangkan, autokorelasi adalah ketergantungan nilai galat antarpengamatan, baik antarwaktu maupun antarindividu (Nachrowi, 2006).

Wooldridge (2010) menjelaskan bahwa keberadaan heteroskedastisitas dan autokorelasi dalam model regresi tidak mempengaruhi sifat tak bias dan konsistensi dari parameter regresi pada metode OLS. Namun demikian, *standard error* dari parameter yang diperoleh menjadi tidak akurat serta variansi yang tidak konstan. Hal tersebut membuat uji simultan dan uji parsial menjadi tidak valid, sehingga dapat mengakibatkan kesimpulan yang ditarik tidak menggambarkan keadaan yang sebenarnya (Kutner dkk., 2005). *Standard error* adalah komponen penting dalam melakukan estimasi suatu nilai parameter (Arieska dan Pusponegoro, 2016).

Standard error yang bias dapat dikoreksi dengan menggunakan *Robust Standard Error*, salah satunya adalah metode Newey West. Nurlaila dkk. (2017) telah melakukan penelitian tentang mengatasi heteroskedastisitas dan autokorelasi dalam ...sil penelitian ini menunjukkan bahwa metode Newey West ...*tandard error* ketika terjadi heteroskedastisitas dan autokorelasi ...si. Namun, penelitian lain dilakukan oleh Petersen (2008) ...metode Newey West yang dimodifikasi untuk data panel belum ...as pada data pengaruh individu.

Pendekatan alternatif untuk mengoreksi *standard error* yang bias pada data panel adalah *Cluster Robust Standard Error* (CRSE). Penelitian yang dilakukan Huang dan Li (2022) menunjukkan bahwa CRSE efektif untuk memperkirakan *standard error* pada data *Cluster*. Lebih lanjut, Metanda dan Oktora (2022) mengkaji metode CRSE untuk mengoreksi *standard error* regresi data panel. Hasil penelitian ini menjelaskan bahwa penggunaan CRSE mengakomodasi kemungkinan heteroskedastisitas dan autokorelasi yang terdapat dalam analisis regresi data panel. Berdasarkan hal-hal tersebut maka dapat disimpulkan bahwa metode ini mampu mengatasi heteroskedastisitas dan autokorelasi serta dapat mengoreksi *standard error* pada regresi data panel.

CRSE merupakan jenis khusus dari *Robust Standard Error.* Struktur data ini melibatkan pengamatan unit yang sama, seperti individu atau negara bagian, dari waktu ke waktu. Pengamatan untuk unit yang sama cenderung berkorelasi yang menyebabkan secara alami membentuk *cluster*. Dalam analisis data panel, terdapat dua pendekatan utama untuk melakukan *cluster*, yaitu *cluster* satu arah dan *cluster* dua arah (Millo, 2017). *Cluster* satu arah berfokus pada satu indeks, baik indeks individu ataupun indeks waktu, tergantung karakteristik data, sedangkan *cluster* dua arah menggunakan kedua indeks (individu dan waktu) secara simultan untuk mengelompokkan data. CRSE dapat diperoleh dengan menghitung Matriks Variansi *Cluster Robust* (MVCR) yang umumnya direkomendasikan ketika menganalisis regresi data panel (Cameron dan Miller, 2015).

Penelitian sebelumnya berfokus pada angka partisipasi murni sekolah menunjukkan bahwa CRSE mengakomodasi terjadinya heteroskedastisitas dan autokorelasi serta dapat mengoreksi galat baku pada regresi data panel. Sehingga, pada penelitian ini akan diimplemetasikan metode tersebut pada tingkat pengangguran terbuka. Menurut Badan Pusat Statistik (BPS), tingkat pengangguran terbuka di Sulawesi Selatan pada tahun 2020 mencapai 6.31% dan pada tahun 2022 mengalami penurunan hingga mencapai 4.51% yang menunjukkan kasus tingkat pengangguran terbuka di Sulawesi Selatan cenderung mengalami penurunan.

Beberapa faktor yang mempengaruhi tingkat pengangguran terbuka adalah kepadatan penduduk (Nurlaily dkk., 2022), angka partisipasi kasar SMA/SMK, produk domestik regional bruto (Prasanti dkk., 2015) dan indeks pembangunan manusia (Rizki dkk., 2022). Berdasarkan uraian di atas, penelitian ini berfokus pada estimasi *standard error* menggunakan MVCR untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap tingkat pengangguran terbuka di Sulawesi Selatan.

1.2 Batasan Masalah

...jar pembahasan pada penelitian ini lebih terstruktur sebagai

...gunakan adalah data tingkat pengangguran terbuka di Sulawesi ...n kurun waktu dari tahun 2020 sampai 2022.

... difokuskan pada analisis regresi panel dengan pendekatan ...tap pengaruh individu yaitu kabupaten/kota di Sulawesi Selatan

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Memperoleh hasil estimasi parameter regresi data panel model terbaik dengan metode Matriks Variansi *Cluster Robust* pada kasus tingkat pengangguran terbuka di Sulawesi Selatan.
2. Membandingankan nilai *standard error* hasil estimasi parameter regresi data panel efek tetap dengan metode Matriks Variansi *Cluster Robust* pada kasus tingkat pengangguran terbuka di Sulawesi Selatan.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Sebagai sumber informasi bagi Dinas Sosail, Dinas Tenaga Kerja dan Transmigrasi, serta pihak pemerintah terkait dalam menentukan kebijakan guna mengurangi tingkat pengangguran terbuka di Sulawesi Selatan.
2. Menambah wawasan bagi pembaca tentang pengaplikasian estimasi parameter model regresi pada data yang bersifat heteroskedastisitas dan autokorelasi dengan metode Matriks Variansi *Cluster Robust.*

1.5 Teori

1.5.1 Regresi Linier Berganda

Secara umum model regresi linier berganda (Judge dan G., 1998) dapat ditulis sebagai berikut:

$$y_i = \beta_0 + X_{1i}\beta + X_{2i}\beta + \cdots + X_{Ki}\beta + u_i, \qquad i = 1,2,\ldots,N;\, k = 1,2,\ldots,K \tag{1}$$

y_i adalah variabel terikat; β_0 adalah konstanta; X_i adalah variabel bebas pada pengamatan ke-i; β_k adalah parameter koefisien regresi; dan u_i adalah galat pada pengamatan ke-i yang berdistribusi normal dengan rataan nol dan variansi σ^2 atau $u_i \sim N(0, \sigma^2)$. Model regresi linier berganda, dapat dijabarkan sebagai berikut:

$$\begin{aligned} y_1 &= \beta_0 + X_{1,1}\beta + X_{2,1}\beta + \cdots + X_{Ki}\beta + u_1 \\ y_2 &= \beta_0 + X_{1,2}\beta + X_{2,2}\beta + \cdots + X_{Ki}\beta + u_2 \\ &\vdots \\ y_N &= \beta_0 + X_{1,N}\beta + X_{2,N}\beta + \cdots + X_{Ki}\beta + u_N \end{aligned}$$

sehingga dapat dinyatakan dalam bentuk matriks berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \cdots & X_{K1} \\ 1 & X_{1,2} & X_{2,2} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1,N} & X_{2,N} & \cdots & X_{K,N} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

Persamaan (1) secara sederhana dapat ditulis sebagai berikut:

$$\vec{\boldsymbol{y}} = \boldsymbol{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{u}} \tag{2}$$

a Panel

ta panel adalah analisis regresi dengan struktur data panel. Hsiao (2022) menjelaskan bahwa regresi data panel gamatan yang terdiri dari dua tipe data yaitu data runtun waktu a runtun waktu adalah data dari pengamatan yang dilakukan pada satu individu pada runtun waktu yang dapat berubah-ubah seiring berjalannya waktu dengan periode waktu yang sama. Sedangkan data silang adalah data dari pengamatan yang dilakukan pada individu yang berbeda pada waktu yang sama.

Sembodo (2013) berpendapat bahwa regresi data panel merupakan sekumpulan teknik untuk memodelkan pengaruh variabel bebas terhadap variabel terikat pada data panel. Secara umum, Persamaan model regresi panel dapat dinyatakan sebagai berikut (Cameron dan Trivedi, 2005):

$$\boldsymbol{y}_{it} = \beta_0 + \boldsymbol{X}'_{it}\boldsymbol{\beta} + u_{it}, \qquad i = 1,2,\dots,N;\ t = 1,2,\dots,T \tag{3}$$

y_{it} adalah variabel terikat dari unit individu ke-i dan periode waktu ke-t; β_{0it} adalah konstanta; $\boldsymbol{X}'_{it}$ adalah variabel bebas dari unit individu ke-i dan unit waktu ke-t; $\boldsymbol{\beta}$ adalah koefisien regresi variabel bebas ke-k berukuran $K \times 1$; dan u_{it} adalah komponen galat regresi dari individu ke-i dan periode waktu ke-t. Hsiao (2022) menjelaskan bahwa galat pada model regresi data panel memiliki asumsi yaitu berdistribusi identik dan independen terhadap individu (i) dan waktu (t) dengan rataan 0 dan variansi σ^2.

Tabel 1. Struktur Data Penel Secara Umum

Individu	Tahun	Variabel				
(i)	(t)	(Y_{it})	(X_{1it})	(X_{2it})	$\cdots$	(X_{Kit})
1	1	$Y_{1,1}$	$X_{1,1,1}$	$X_{2,1,1}$	$\cdots$	$X_{K,1,1}$
	2	$Y_{1,2}$	$X_{1,1,2}$	$X_{2,1,2}$		$X_{K,1,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	T	$Y_{1,T}$	$X_{1,1,T}$	$X_{2,1,T}$		$X_{K,1,T}$
2	1	$Y_{2,1}$	$X_{1,2,1}$	$X_{2,2,1}$	$\cdots$	$X_{K,1,2}$
	2	$Y_{2,2}$	$X_{1,2,2}$	$X_{2,2,2}$		$X_{K,2,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	T	$Y_{2,T}$	$X_{1,2,T}$	$X_{2,2,T}$		$X_{K,2,T}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
N	1	$Y_{N,1}$	$X_{1,N,1}$	$X_{2,N,1}$	$\cdots$	$X_{K,N,1}$
	2	$Y_{N,2}$	$X_{1,N,2}$	$X_{2,N,2}$		$X_{K,N,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	T	$Y_{N,T}$	$X_{1,N,T}$	$X_{2,N,T}$		$X_{K,N,T}$

Penggunaan data panel memiliki residual yang terdiri dari dua komponen, yaitu komponen umum dan komponen khusus. Komponen umum mewakili efek yang ... individu dalam waktu yang berbeda, sedangkan komponen ... yang berbeda pada setiap individu dalam waktu yang berbeda. ...atis ditunjukkan pada Persamaan (4).

$$u_{it} = e + v_i + w_t \tag{4}$$

...en residual bersifat umum, yang mewakili efek yang sama pada ...n waktu yang berbeda; v_i merupakan komponen spesifik dari individu, yang mewakili efek yang berbeda pada setiap individu disatu waktu; dan w_t merupakan komponen spesifik dari deret waktu, yang mewakili efek yang berbeda pada satu individu pada waktu yang berbeda (Ariefianto & Trinugroho, 2020).

Apabila diasumsikan tidak terdapat pengaruh spesifik dari unit individu maupun waktu, maka estimasi Persamaan (3) dapat dilakukan dengan metode OLS yang umum digunakan. Namun, apabila diyakini bahwa terdapat pengaruh spesifik baik pada unit individu maupun waktu, maka pemodelan residual harus dilakukan secara eksplisit. Jika asumsi bahwa residual hanya terdiri dari komponen umum tidak dapat diterima, maka ada dua alternatif model yang dapat digunakan, yaitu:

a. Model Efek Tetap

Model Efek Tetap (MET) merupakan model yang mengasumsikan bahwa antar unit individu atau unit waktu memiliki perilaku yang berbeda. Perilaku tersebut terlihat dari nilai konstanta berbeda untuk setiap unit individu atau waktu, tetapi nilai koefisien regresi tetap untuk unit individu maupun waktu (Gujarati, 2003). Astuti dan Maruddani (2009) memaparkan teknik estimasi model regresi data panel dengan model efek tetap individu menggunakan pendekatan estimasi *Least Square Dummy Variable* (LSDV) sebagai berikut.

$$y_{it} = \sum_{j=1}^{J} \beta_{0j} D_{jt} + \sum_{k=1}^{K} \boldsymbol{X}'_{it}\boldsymbol{\beta} + u_{it} \qquad (5)$$

D_{jt} merupakan variabel *dummy* yang bernilai satu untuk pengamatan yang sama dengan individu ke-*j* dan bernilai nol untuk pengamatan individu lain.

$$D_{jt} = \begin{cases} 1 & \text{jika } j = i \\ 0 & \text{jika } j \neq i \end{cases}$$

Apabila dinyatakan dalam bentuk matriks, maka dapat dituliskan sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{D_N}\beta_0 + \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{u} \qquad (6)$$

$\boldsymbol{y}$ merupakan vektor variabel terikut berukuran $NT \times 1$; $\boldsymbol{X}'$ merupakan vektor variabel bebas berukuran $NT \times K$; $\boldsymbol{D_N}$ merupakan matriks variabel *dummy* individu berukuran $NT \times N$; β_0 merupakan vektor konstanta untuk keberagaman individu berukuran $N \times 1$; $\boldsymbol{\beta}$ merupakan vektor koefisien regresi berukuran $K \times 1$; dan $\boldsymbol{u}$ merupakan vektor galat berukuran $NT \times 1$.

b. Model Efek Acak

Model Efek Acak (MEA) merupakan model yang mengasumsikan pengaruh spesifik unit individu atau unit waktu masuk ke dalam komponen galat model sehingga pengaruh tersebut bersifat acak dengan sebaran dan keragaman tertentu. Oleh karena itu, konstanta antar unit individu atau waktu model ini bersifat acak (*random*), sedangkan koefisien regresi bersifat tetap untuk setiap unit individu maupun waktu

$$y_{it} = \beta_0 + \boldsymbol{X}'_{kit}\boldsymbol{\beta} + w_{it} \qquad (7)$$

ngan μ_i merupakan pengaruh spesifik unit individu ke-i yang nilai rataan nol dan ragam sebesar σ_μ^2, dan λ_t merupakan ke-t yang bersifat acak dengan nilai rataan 0 dan ragam sebesar

1.5.3 Penentuan Model Regresi Data Panel

Terdapat dua pengujian yang dilakukan untuk mengetahui metode regresi data panel yang sesuai dalam memodelkan data. Adapun dua pengujian tersebut adalah sebagat berikut.

a. Uji Chow

Greene (2012) menjelaskan bahwa uji Chow merupakan uji yang dilakukan untuk memilih antara model efek umum atau model efek tetap untuk mengestimasi data panel. Hipotesis yang digunakan dalam uji Chow adalah (Baltagi, 2021):

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_N = 0$ (Model yang sesuai adalah MEU)

H_1 : Ada satu $\beta_i \neq 0$; $i = 1,2,\ldots,N$ (Model yang sesuai adalah MET)

Statistik uji Chow sebagai berikut (Baltagi, 2021):

$$F = \frac{(R^2_{MEU} - R^2_{MET})/(N-1)}{R^2_{MET}/(NT-N-K)} \sim F_{(\alpha,N-1,NT-N-K)} \tag{8}$$

R^2_{MEU} merupakan R-*square* Model Efek Umum; R^2_{MET} merupakan *R square* Model Efek Tetap; *N* merupakan jumlah unit individu; *T* merupakan jumlah unit waktu; dan *K* merupakan jumlah variabel bebsas. Jika $F_{hitung} > F_{(\alpha,N-1,NT-N-K)}$ atau nilai *p-value* $< \alpha$, maka tolak hipotesis awal (H_0) sehingga dapat disimpulkan bahwa model regresi data panel menggunakan MET lebih baik daripada menggunakan MEU.

b. Uji Hausman

Uji Hausman adalah pengujian untuk memilih model terbaik antara model efek tetap dan model efek acak. Pengujian ini dilakukan menggunakan uji wald dengan hipotesis sebagai berikut (Baltagi, 2021).

$H_0 : Korelasi(x_{it}, u_{it}) = 0$ (Model yang sesuai adalah MEA)

$H_1 : Korelasi(x_{it}, u_{it}) \neq 0$ (Model yang sesuai adalah MET)

Statistik uji:

$$W = \{(q)^T[Var(q)]^{-1}(q)\} \sim \chi^2_{(\alpha,K)} \tag{9}$$

$q = \hat{\beta}_{MET} - \hat{\beta}_{MEA}$; $\hat{\beta}_{MET}$ merupakan vektor estimasi koefisien regresi Model Efek Tetap; dan $\hat{\beta}_{MEA}$ merupakan vektor estimasi koefisien regresi Model Efek Acak. Jika nilai $W > \chi^2_{(\alpha,K)}$ dengan *K* merupakan banyaknya variabel bebas atau *p-value* < α, maka tolak hipotesis awal (H_0) sehingga dapat disimpulkan bahwa model regresi data panel menggunakan MET lebih baik daripada menggunakan MEA.

1.5.4 Pengujian Asumsi Regresi

Yudiaatmaja (2013) menjelaskan bahwa model regresi data panel dapat disebut sebagai model yang baik jika model tersebut memenuhi kriteria *Best, Linear, and*
BLUE). Kriteria ini dapat dicapai bila memenuhi asumsi klasik.
cakup uji normalitas, uji multikolinieritas, uji heterokedastisitas,

alitas
kan untuk menguji galat pada model regresi menyebar secara
alah satu metode untuk uji normalitas adalah uji *Jarque-Bera*,
g dilakukan sebagai berikut:

H_0 : Galat berdistribusi normal
H_1 : Galat tidak berdistribusi normal
Berikut persamaan untuk statistik uji *Jarque-Bera* (Gujarati dan Porter, 2009).

$$JB = \frac{n}{6}\left(Skeeness^2 + \frac{(Kurtosis - 3)^2}{4}\right) \sim \chi^2_{(\alpha,2)} \tag{10}$$

Jika statistik hitung JB lebih kecil dari nilai kritis *chi-square* dengan tingkat signifikansi α dan derajat bebas dua $\left(JB < \chi^2_{(\alpha,2)}\right)$ atau *p-value* $> \alpha$, maka terima hipotesis awal (H_0) sehingga dapat disimpulkan bahwa galat berdistribusi normal.

b. Asumsi Multikolinieritas

Multikolinieritas merupakan kondisi adanya hubungan linier yang kuat antarvariabel bebas dalam suatu model regresi. Multikolinieritas dapat dilihat dari nilai *Variance Inflation Factors* (VIF). Rumus VIF adalah (Freund dkk., 2006):

$$VIF_j = \frac{1}{1 - R_j^2} \tag{11}$$

R_j^2 adalah koefisien determinasi dari variabel bebas x_j yang diregresikan terhadap variabel bebas lainnya. Jika nilai VIF ≤ 10, tidak terdapat multikolinieritas. Sebaliknya jika nilai VIF > 10 maka terjadi multikolinieritas.

c. Asumsi Homoskedastisitas

Uji ini digunakan untuk mengetahui apakah matriks struktur variansi galat bersifat homoskesdastik atau heteroskesdastisitas. Uji heteroskesdastisitas yang digunakan adalah uji Breusch Pagan dengan menggunakan hipotesis sebagai berikut (Greene, 2012):

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_N^2$ (Struktur variansi galat homoskesdastik)
H_1 : Ada satu $\sigma_i^2 \neq \sigma_{i^*}^2$ dengan $i \neq i^*$ (Struktur variansi galat heteroskesdastisitas)
Statistik uji Breusch Pagan sebagai berikut:

$$BP = \frac{1}{2}[\boldsymbol{f}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{f}] \sim \chi^2_{(\alpha,K)} \tag{12}$$

dengan elemen vektor f adalah:

$$f_i = \left(\frac{u_i^2}{\sigma^2} - 1\right)$$

u_i^2 adalah galat pengamatan ke-i, $i = 1,2,\dots,N$, σ^2 adalah $\frac{u'u}{n}$, $\boldsymbol{Z}$ adalah matriks berukuran $N \times (K+1)$ dari pengamatan dengan elemen kolom pertama yakni vektor satu, dan K adalah jumlah variabel bebas. Kriteria pengujian, yaitu jika nilai BP $> \chi^2_{(\alpha,K)}$ dan nilai *p-value* $< \alpha$, maka tolak hipotesis awal (H_0) sehingga dapat disimpulkan bahwa struktur variansi galat bersifat heterokesdastisitas.

korelasi
enjelaskan bahwa autokorelasi menandakan adanya korelasi
atan satu dengan yang lainnya. Uji yang digunakan untuk
aan autokorelasi adalah uji Breusch-Godfrey juga dikenal
Multiplier (LM) dengan hipotesis, yaitu (Gujarati, 2003):
dak terdapat autokorelasi)
rdapat autokorelasi)

Statistik uji Breusch-Godfrey sebagai berikut:

$$\lambda_{LM} = T \sum_{i=1}^{N-1} \sum_{j=1+1}^{N} \hat{\rho}_{ij}^2 \sim \chi^2_{(\alpha,K)} \tag{13}$$

$\hat{\rho}_{ij}$ adalah estimasi sederhana korelasi galat *pair-wise* dengan rumus sebagai berikut:

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{\frac{1}{2}}} \tag{14}$$

Kriteria pengujian, yaitu jika nilai $\lambda_{LM} > \chi^2_{(\alpha,K)}$ dan nilai *p-value* < α, maka tolak hipotesis awal (H_0) sehingga dapat disimpulkan bahwa terdapat autokorelasi.

1.5.5 Metode *Ordinary Least Squares*

Metode *Ordinary Least Squares* (OLS) adalah regresi dengan menggunakan metode kuadrat kesalahan terkecil yang sederhana. Estimator ini memiliki beberapa asumsi yang harus dipenuhi agar memperoleh hasil regresi yang tak bias. Metode estimator *least square* pada prinsipnya menentukan nilai parameter dengan meminimumkan jumlah kuadrat galat (Ekananda, 2015). Secara umum, persamaan OLS dapat dituliskan seperti Persamaan (2). Estimator *least square* dari $\boldsymbol{\beta}$ merupakan estimator yang meminimumkan jumlah kuadrat galat seperti pada Persamaan **(**15).

$$\begin{aligned} \boldsymbol{u}'\boldsymbol{u} &= (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} - (\boldsymbol{X\beta})'\boldsymbol{y} + (\boldsymbol{X\beta})'\boldsymbol{X\beta} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \end{aligned} \tag{15}$$

Kemudian dilakukan turunan pertama $\boldsymbol{u}'\boldsymbol{u}$ terhadap $\boldsymbol{\beta}$ yang disamadengankan nol untuk meminimumkan jumlah kuadrat galat pada persmaan (15).

$$\begin{aligned} \left.\frac{\partial(\boldsymbol{u}'\boldsymbol{u})}{\partial(\boldsymbol{\beta})}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \frac{\partial(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta})}{\partial(\boldsymbol{\beta})} &= 0 \\ -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} &= 0 \\ \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y} \\ \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \end{aligned} \tag{16}$$

Selanjutnya akan ditunjukkan jika $\hat{\boldsymbol{\beta}}$ merupakan estimasi parameter linier tak bias dari $\boldsymbol{\beta}$, yaitu sebagai berikut.

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X\beta} \\ &= \boldsymbol{\beta} \end{aligned} \tag{17}$$

Sehingga terbukti bahwa $\hat{\boldsymbol{\beta}}$ merupakan estimasi linier tak bias dari $\boldsymbol{\beta}$, yang mana ...an yang disebut dengan parameter $\boldsymbol{\beta}$ secara kuadrat terkecil ... Kemudian untuk matriks variansi kovariansi dari $\hat{\boldsymbol{\beta}}$ diperoleh ...sikan Persamaan (2) pada $\hat{\boldsymbol{\beta}}$, sehingga diperoleh sebagai ...

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta} + \boldsymbol{u}) \end{aligned}$$

$$
\begin{aligned}
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{X\beta}) + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{u}) \\
&= (\boldsymbol{X'X})^{-1}(\boldsymbol{X'X})\boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u} \\
&= \boldsymbol{I\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u} \\
\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\boldsymbol{X'X})^{-1}\boldsymbol{X'u} \qquad (18)
\end{aligned}
$$

Berdasarkan asumsi $E(\boldsymbol{uu'}) = \sigma^2\boldsymbol{I}$, sehingg diperoleh matriks variansi kovariansi, yaitu sebagai berikut.

$$
\begin{aligned}
\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}) &= E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] \\
&= E[((\boldsymbol{X'X})^{-1}\boldsymbol{X'u})((\boldsymbol{X'X})^{-1}\boldsymbol{X'u})'] \\
&= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'u}\,\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{u'}] \\
&= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'uu'X}(\boldsymbol{X'X})^{-1}] \\
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}E(\boldsymbol{uu'})\boldsymbol{X}(\boldsymbol{X'X})^{-1} \\
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{IX}(\boldsymbol{X'X})^{-1} \\
&= \sigma^2(\boldsymbol{X'X})^{-1}(\boldsymbol{X'X})\boldsymbol{I}(\boldsymbol{X'X})^{-1} \\
&= \sigma^2(\boldsymbol{X'X})^{-1} \qquad (19)
\end{aligned}
$$

1.5.6 *Standard Error*

Standard error adalah standar deviasi dari distribusi sampling suatu statistik (Cahyono, 2018). *Standard error* merujuk pada estimasi standar deviasi berdasarkan sejumlah n sampel yang digunakan untuk menghitung suatu nilai estimator. Metode yang digunakan untuk mengestimasi model dilandasi pada prinsip meminimalkan galat. Oleh karena itu, ketepatan dari nilai estimasi sangat ditentukan oleh *standard error* dari masing-masing estimator. *Standard error* estimasi OLS dapat dihitung dengan rumus sebagai berikut (Gujarati, 2003).

$$\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X'X})^{-1} \qquad (20)$$

Sehinga diperoleh nilai *standard error* sebagai berikut:

$$
\begin{aligned}
se(\widehat{\boldsymbol{\beta}}) &= \sqrt{\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}})} \\
se(\widehat{\boldsymbol{\beta}}) &= \sqrt{\sigma^2(\boldsymbol{X'X})^{-1}} \qquad (21)
\end{aligned}
$$

Oleh karena σ^2 merupakan penyimpangan yang terjadi dalam populasi dan nilainya tidak diketahui, maka σ dapat diestimasi sebagai berikut:

$$\hat{\sigma}^2 = \frac{\sum_i^n \hat{u}_i^2}{N-K} \qquad (22)$$

$\hat{\sigma}^2$ adalah estimator OLS dari σ^2 yang sebenarnya tidak diketahui; $N-K$ merupakan derajat bebas, $\sum \hat{u}_i^2$ adalah jumlah galat yang dikuadratkan.

1.5.7 Matriks Variansi *Cluster Robust*

... *luster Robust* (MVCR) merupakan pengembangan dari ... *onsistent Covariance Matrix Estimator* (HCCME) yang telah ... yang dikelompokkan oleh Liang dan Zeger (1986), yang ... imator MVCR untuk memperhitungkan heteroskedastisitas dan ... ata panel. Matriks ini juga telah diterapkan dalam pengaturan ... eh Arellano (1987) untuk mengestimasi model regresi linier dengan data panel. Berikut adalah rumus MVCR untuk model regresi linier pada data panel menurut Arellano (1987) sebagai berikut:

$$\boldsymbol{y}_g = \boldsymbol{X}'_g\boldsymbol{\beta} + \boldsymbol{u}_g \qquad (g = 1,2,\dots,G) \tag{23}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_G \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_G \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_G \end{bmatrix}$$

G adalah jumlah *cluster*, N_g adalah jumlah pengamatan dalam *cluster* ke-g, $\boldsymbol{y}_g$ adalah vektor pengamatan berukuran $N_g \times 1$ pada variabel terikat, $\boldsymbol{X}$ adalah matriks pengamatan berukuran $N_g \times k$ pada variabel bebas, $\boldsymbol{\beta}$ adalah vektor parameter yang diestimasi berukuran $K \times 1$, dan $\boldsymbol{u}_g$ adalah vektor berukuran $N_g \times 1$ yang berisi nilai galat untuk pengamatan dalam *cluster* ke-g. Estimator *ordinary least squares* (OLS) untuk model ini adalah:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \left(\sum_{g=1}^{G}\sum_{i=1}^{N_g}\boldsymbol{x}_{ig}\boldsymbol{x}'_{ig}\right)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\boldsymbol{x}_{ig}\boldsymbol{y}_{ig} \\ &= \left(\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{X}_g\right)^{-1}\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{y}_g \end{aligned} \tag{24}$$

Galat OLS adalah $\widehat{\boldsymbol{u}}_{ig} = \boldsymbol{y}_{ig} - \boldsymbol{x}'_{ig}\widehat{\boldsymbol{\beta}}$ pada notasi level individu dan $\widehat{\boldsymbol{u}}_g = \boldsymbol{y}_g - \boldsymbol{X}_g\widehat{\boldsymbol{\beta}}$ pada notasi level *cluster*, dengan asumsi bahwa tidak ada hubungan (korelasi) antara galat dari pengamatan di *cluster* yang berbeda (Hansen, 2020). Model regresi linier dengan asumsi ini sebagai berikut.

$$E[\boldsymbol{u}_g|\boldsymbol{X}_g] = 0 \tag{25}$$

Persamaan (25) menyatakan bahwa model regresi linier telah dispesifikasikan dengan benar. Pada model regresi *cluster*, hal ini berarti bahwa semua efek interaksi dalam *cluster* telah diperhitungkan dalam spesifikasi regressor individual $\boldsymbol{x}_{ig}$. Berdasarkan Persamaan (25), diperoleh rata-rata estimator OLS dengan mensubstitusikan Persamaan (23) ke Persamaan (24) sebagai berikut.

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{X}_g\right)^{-1}\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{u}_g$$

Rata-rata dari $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ dengan kondisi semua variabel bebas (regresor) dalam model regresi adalah:

$$\begin{aligned} E[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|X] &= \left(\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{X}_g\right)^{-1}\sum_{g=1}^{G}\boldsymbol{X}'_g E[\boldsymbol{u}_g|\boldsymbol{X}] \\ &= \left(\sum_{g=1}^{G}\boldsymbol{X}'_g\boldsymbol{X}_g\right)^{-1}\sum_{g=1}^{G}\boldsymbol{X}'_g E[\boldsymbol{u}_g|\boldsymbol{X}_g] \end{aligned}$$

$$= \quad \mathbf{0}$$

model regresi linier *cluster* $E[\widehat{\boldsymbol{\beta}}|\boldsymbol{X}] = \boldsymbol{\beta}$ dengan matriks kovarians dari $\widehat{\boldsymbol{\beta}}$.

$\boldsymbol{\Sigma}_g = E[\boldsymbol{u}_g \boldsymbol{u}_g' | \boldsymbol{X}_g]$ menyatakan matriks kovarians bersyarat ke-N_g dari galat dalam *cluster* ke-g, karena pengamatan bersifat independen di seluruh *cluster*, sehingga:

$$\begin{aligned} \operatorname{var}\left[\left(\textstyle\sum_{g=1}^{G} \boldsymbol{X}_g' \boldsymbol{u}_g\right)|\boldsymbol{X}\right] &= \sum_{g=1}^{G} \operatorname{var}[\boldsymbol{X}_g' \boldsymbol{u}_g | \boldsymbol{X}_g] \\ &= \sum_{g=1}^{G} \boldsymbol{X}_g' \, E[\boldsymbol{u}_g \boldsymbol{u}_g' | \boldsymbol{X}_g] \boldsymbol{X}_g \\ &= \sum_{g=1}^{G} \boldsymbol{X}_g' \, \boldsymbol{\Sigma}_g \boldsymbol{X}_g \\ &\overset{\text{def}}{=} \boldsymbol{\Omega} \end{aligned} \tag{26}$$

Secara umum, matriks variansi bersyarat pada $\boldsymbol{X}$ adalah sebagai berikut.

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}|\boldsymbol{X}] = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{\Omega} (\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{27}$$

Arellano (1987) mengusulkan sebuah metode pengestimasi matriks kovariansi yang merupakan perluasan dari metode pengestimasi White. Inti dari metode pengestimasi kovariansi White adalah bahwa kesalahan kuadrat u_i^2 tidak bias terhadap $E[\boldsymbol{u}_i^2|\boldsymbol{x}_i] = \sigma_i^2$. Demikian pula dengan ketergantungan cluster, matriks $\boldsymbol{u}_g \boldsymbol{u}_g'$ tidak bias terhadap $E[\boldsymbol{u}_g \boldsymbol{u}_g' | \boldsymbol{X}_g] = \boldsymbol{\Sigma}_g$, ini berarti bahwa pengestimasi tidak bias untuk Persamaan (27) adalah $\widetilde{\boldsymbol{\Omega}} = \sum_{g=1}^{G} \boldsymbol{X}_g' \boldsymbol{u}_g \boldsymbol{u}_g' \boldsymbol{X}_g$. Namun, estimator ini tidak dapat dihitung secara langsung karena *error term* u_i^2 tidak diketahui. Sebagai gantinya, Arellano mengusulkan untuk mengganti nilai galat dengan residual OLS untuk mendapatkan estimator:

$$\widehat{\boldsymbol{\Omega}} = \sum_{g=1}^{G} \boldsymbol{X}_g' \widehat{\boldsymbol{u}}_g \widehat{\boldsymbol{u}}_g' \boldsymbol{X}_g$$

$\widehat{\boldsymbol{\Omega}}$ adalah penjumlahan matriks variansi kovariansi galat untuk *cluster* g yang berukuran $N_g \; x \; N_g$. Sehingga, estimasi matriks variansi *cluster robust* adalah:

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}] = c \, (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \boldsymbol{X}_g' \widehat{\boldsymbol{u}}_g \widehat{\boldsymbol{u}}_g' \boldsymbol{X}_g \right) (\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{28}$$

c adalah penyesuaian sampel terbatas yang mungkin.

$$c = \left(\frac{G}{G-1}\right)\left(\frac{N-1}{N-K}\right)$$

Fakt... G ... oleh Hansen (2007) untuk mengoreksi estimasi jumlah *cluster* ...kan kinerja estimasi saat jumlah *cluster* terbatas. Faktor $\frac{N-1}{N-K}$ ...neralisasi yang lebih fleksibel dengan fungsi yang sama, yaitu ...imasi. Kemudian dengan memasukkan efek *cluster* pada fungsi ...onal melalui koefisien diperoleh model *Cluster-Specific Fixed* ...ut.

$$y_{jg} = \boldsymbol{X}'_{jg}\boldsymbol{\beta} + a_g + u_{jg} \qquad (j = 1,2,\dots,N_g, \qquad g = 1,2,\dots,G) \tag{29}$$

y_{jg} adalah nilai variabel terikat untuk pengamatan ke-j dalam *cluster* ke-g, a_g adalah efek *cluster* ke-g, $\boldsymbol{X}'_{jg}$ adalah vektor independen variabel untuk pengamatan ke-j dalam *cluster* ke-g, $\boldsymbol{\beta}$ merupakan vektor parameter yang diestimasi, dan u_{jg} vektor galat untuk pengamatan ke-j dalam *cluster* ke-g.

Efek *cluster* dalam model ini tidak hanya dipertimbangkan sebagai variabel tambahan, tetapi efek *cluster* terkandung dalam perubahan konstanta untuk setiap *cluster*. Sehingga, model *Cluster-Specific Fixed Effects* memungkinkan untuk memperhitungkan variasi antara *cluster* dalam nilai rata-rata kondisional, tanpa harus secara langsung mengidentifikasi dan memodelkan efek *cluster* tersebut secara terpisah (Cameron dan Trivedi, 2005).

1.5.8 Pengujian Parameter Model Regresi

Pengujian parameter model regresi dilakukan untuk mengetahui hubungan antara variabel terikat dengan variabel bebas. Terdapat dua pengujian yang harus dilakukan, yaitu pengujian secara simultan dan pengujian secara parsial.

a. Koefisien Determinasi

Nilai koefisien determinasi (R^2) ini mencerminkan seberapa besar variasi dari variabel terikat dapat diterangkan oleh variabel bebas. Bila nilai koefisien determinasi sama dengan 0 $(R^2 = 0)$, artinya variasi dari variabel terikat tidak dapat diterangkan oleh variabel bebas sama sekali. Sementara bila $R^2 = 1$, artinya variasi variabel terikat secara keseluruhan dapat diterangkan oleh variabel bebas, dengan kata lain $R^2 = 1$, maka semua pengamatan berada tepat pada garis regresi. Sehungga baik atau buruknya suatu persamaan regresi ditentukan oleh R^2-nya yang mempunyai nilai antara nol dan satu (Nachrowi, 2006). Nilai koefisien determinasi dapat diperoleh dari:

$$R^2 = 1 - \frac{\boldsymbol{u}'\boldsymbol{u}}{(\boldsymbol{y} - \boldsymbol{y})'(\boldsymbol{y} - \boldsymbol{y})} \tag{30}$$

b. Uji Simultan

Uji simultan adalah metode yang dilakukan untuk mengetahui pengaruh variabel bebas secara bersama-sama terhadap variabel terikat dengan menggunakan uji F. Hipotesis yang digunakan adalah sebagai berikut (Gujarati, 2003):

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

H_1: Ada satu $\beta_K \neq 0$, dengan $k = 1,2,\dots,K$

... nakan adalah:

$$F = \frac{R^2/(N + K - 1)}{(1 - R^2)/(NT - N - K)} \tag{31}$$

... determinasi, *N* merupakan jumlah unit individu; *T* merupakan ... n *K* merupakan jumlah variabel bebas. Kriteria pengujian, yaitu ... $F_{(\alpha,N+K-1,NT-T-K)}$ yang berarti variabel bebas secara bersama- ... rhadap variabel terikat.

c. Uji Parsial

Uji parsial digunakan untuk mengetahui bahwa variabel bebas X_k tidak mempengaruhi variabel terikat Y dengan asumsi variabel bebas lainnya konstan, yang berarti $\beta_k = 0$. Uji parsial dilakukan dengan menggunakan uji t dengan hipotesis yang digunakan sebagai berikut:

$H_0: \beta_k = 0$

H_1: Ada satu $\beta_k \neq 0$; $k = 1,2, \dots, K$

Statistik uji yang digunakan adalah:

$$t_{hitung} = \frac{\hat{\beta}_k - \beta_k}{SE\left(\hat{\beta}_k\right)} \tag{32}$$

kriteria pengujian, yaitu tolak H_0 jika $\left|t_{hitung}\right| > t_{\left(\frac{\alpha}{2}, NT-K\right)}$. yang berarti parameter tersebut signifikan secara statistik pada taraf signifikansi sebesar α.

1.5.9 Tingkat Pengangguran Terbuka

Menurut Badan Pusat Statistik (2022) pengangguran terbuka adalah penduduk yang tidak bekerja dan sedang mencari pekerjaan, atau mempersiapkan usaha baru, atau merasa tidak mungkin mendapat pekerjaan, atau sudah menerima pekerjaan tetapi belum mulai bekerja. Tingkat pengangguran terbuka (TPT) merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja. Secara umum, perhitungan TPT adalah sebagai berikut:

$$\mathrm{TPT} = \frac{\text{Jumlah pengangguran}}{\text{Jumlah Angkatan Kerja}} \times 100\% \tag{33}$$