

BAB III

METODOLOGI PENELITIAN

A. Tujuan Penelitian

Berdasarkan masalah yang telah dirumuskan oleh peneliti, maka tujuan dari penelitian ini adalah untuk mendapatkan pengetahuan berdasarkan data dan fakta yang tepat (benar, valid) dan dapat dipercaya (dapat diandalkan/realible) tentang faktor-faktor yang mempengaruhi ketimpangan distribusi pendapatan antar propinsi di Indonesia.

B. Objek Penelitian

Penelitian ini dilaksanakan dengan mengambil data PDRB per kapita dan jumlah penduduk untuk menghitung ketimpangan distribusi pendapatan dengan menggunakan indeks williamson dan juga tingkat pendidikan, serta realisasi *Foreign Direct Investment* antar propinsi di Indonesia periode 2007-2010. Waktu ini dipilih karena merupakan interval waktu yang paling baik antara sebelum dan sesudah krisis finansial tahun 2008.

C. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode *ex pos facto*, yang merupakan suatu penelitian yang dilakukan untuk meneliti peristiwa yang telah terjadi dan kemudian meruntut ke belakang untuk mengetahui faktor yang dapat menimbulkan kejadian tersebut. Metode ini

digunakan untuk memperoleh data sekunder.⁴⁴ Sedangkan pendekatan yang digunakan adalah pendekatan korelasional bertujuan untuk mengetahui seberapa besar pengaruh antara variabel-variabel yang diteliti yaitu ketimpangan distribusi pendapatan beserta faktor-faktor yang mempengaruhinya.

D. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data tahunan per provinsi PDRB per kapita dan jumlah penduduk untuk menghitung ketimpangan distribusi pendapatan dengan menggunakan indeks williamson dan juga tingkat pendidikan, serta realisasi *Foreign Direct Investment* yang tersedia di Badan Pusat Statistik.

Teknik pengambilan data dalam penelitian ini adalah dengan menggunakan data panel. Data panel adalah gabungan antara data *time series* (antar waktu) dan *cross section* (antar individu/ruang). Data yang digunakan dengan menggunakan *cross section* dari 30 provinsi di Indonesia dan *time series* selama 4 tahun dari tahun 2007-2010 dengan menghilangkan 3 propinsi lain, yaitu Sulawesi Barat, Gorontalo, dan Maluku. Alasan menghilangkan 3 propinsi yang disebutkan diatas karena ketiadaan data jumlah penduduk dan PDRB kota untuk menghitung indeks ketimpangan williamson di Sulawesi Barat, serta ketiadaan realisasi *Foreign Direct Investment* untuk tiga propinsi diatas selama tiga tahun berturut-turut, yaitu tahun 2007-2009. Sehingga, jumlah data secara keseluruhan

⁴⁴ Sugiyono, *Metode Penelitian Bisnis* (Jakarta : Alfabeta, 2004), p. 7.

dengan menggunakan *cross section* dan *time series* dalam bentuk data panel menjadi sebanyak 120 data analisis.

E. Operasionalisasi Variabel Penelitian

a. Ketimpangan Distribusi Pendapatan

1. Definisi Konseptual

Ketimpangan distribusi pendapatan adalah perbedaan pembagian pendapatan antara sekelompok masyarakat dengan sekelompok masyarakat lain. Sedangkan ketimpangan distribusi pendapatan antar regional merupakan perbedaan faktor *endowment* atau faktor yang terdapat dalam masing-masing daerah.

2. Definisi Operasional

Ketimpangan distribusi pendapatan regional merupakan perbedaan faktor *endowment* yang dimiliki oleh setiap daerah. Ketimpangan distribusi pendapatan antar regional dihitung berdasarkan indeks Williamson, dengan rumus sebagai berikut yaitu:

$$IW = \frac{\sqrt{\sum (Y_i - Y)^2 f_i/n}}{Y}$$

Dimana :

IW = Indeks Williamson
 Y_i = PDRB perkapita di kota i
 Y = PDRB perkapita di Provinsi
 f_i = Jumlah penduduk di kota i
 n = Jumlah penduduk di Provinsi

b. Tingkat Pendidikan

1. Definisi Konseptual

Tingkat pendidikan adalah tingkat pendidikan tertinggi yang dicapai penduduk suatu negara. Menurut BPS (2000), tingkat pendidikan tertinggi yang ditamatkan adalah mereka yang meninggalkan sekolah setelah mengikuti pelajaran pada kelas tertinggi suatu tingkatan sekolah sampai akhir dengan mendapatkan ijazah baik dari sekolah negeri ataupun swasta.

2. Definisi Operasional

Tingkat pendidikan adalah jenjang pendidikan tertinggi yang ditamatkan oleh penduduk. Jenjang itu terbagi menjadi jenjang pendidikan dasar, menengah pertama, menengah atas dan perguruan tinggi. Data tingkat pendidikan diambil merupakan nilai rata-rata lama sekolah 30 propinsi di Indonesia yang diambil di Badan Pusat Statistik.

c. Foreign Direct Investment

1. Definisi Konseptual

Foreign Direct Investment adalah penanaman modal yang dilakukan oleh perusahaan atau individu dari negara asal ke negara lainnya sebagai partisipasi langsung atas laba dari modal yang ia tanam dan tetap mengikuti peraturan yang berlaku dari negara penerima modal.

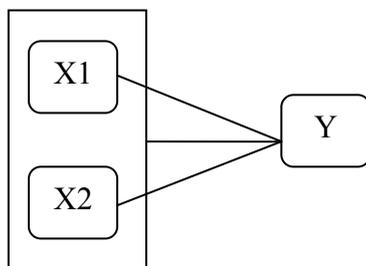
2. Definisi Operasional

Variabel FDI_{it} didefinisikan sebagai *share* aliran FDI terhadap PDRB tahun konstan 2000. *Proxy* dari jumlah FDI yang masuk ke propinsi i tahun t , yaitu diperoleh dari :

$$FDI_{it} = \frac{FDIFlow (US\$) \times RER (Real Exchange Rate Rp/US\$)}{PDRB \text{ Harga Konstan } t2000 (Rp)} \times 100\%$$

F. Konstelasi Hubungan Antara Variabel

Konstelasi pengaruh antar variabel dalam penelitian ini bertujuan untuk memberikan arah atau gambaran dari penelitian ini, yang dapat digambarkan sebagai berikut :



Keterangan :

X_1 : Tingkat Pendidikan

X_2 : Investasi

Y : Ketimpangan Distribusi Pendapatan

→ : Pengaruh

G. Teknik Analisis Data

Data yang digunakan dalam analisis ini berupa data panel. Data panel (*panel pooled data*) merupakan gabungan data *cross-section* dan data *time-series*. Dengan kata lain, data panel merupakan unit-unit individu yang sama yang diamati dalam kurun waktu tertentu. Jika kita memiliki T periode waktu ($t = 1, 2, \dots, T$) dan N jumlah individu ($i = 1, 2, \dots, N$), maka dengan data panel kita akan memiliki total unit observasi sebanyak NT . Jika jumlah unit waktu sama untuk

setiap individu, maka data disebut *balanced panel*. Jika sebaliknya, yakni jumlah unit waktu berbeda untuk setiap individu, maka disebut *unbalanced panel*. Penggunaan data panel pada dasarnya merupakan solusi akan ketidaktersediaan data *time-series* yang cukup panjang untuk kepentingan analisis ekonometrika.

Gujarati (2001) berdasarkan uraian dari Baltagi, keunggulan penggunaan data panel dibanding data runtun waktu dan data lintas sektor adalah :

1. Estimasi data panel dapat menunjukkan adanya heterogenitas dalam tiap unit.
2. Dengan data panel, data lebih informatif, mengurangi kolinieritas antara variabel, meningkatkan derajat kebebasan dan lebih efisien.
3. Data panel cocok digunakan untuk menggambarkan adanya dinamika perubahan.
4. Data panel dapat lebih mampu mendeteksi dan mengukur dampak.
5. Data panel bisa digunakan untuk studi dengan model yang lebih lengkap.
6. Data panel dapat meminimumkan bias yang mungkin dihasilkan dalam regresi.

1. Estimasi model

Dalam data panel, terdapat tiga spesifikasi model yang mungkin digunakan, yakni model *common effect*, *fixed effect*, dan *random effect*. Peneliti dianjurkan untuk memilih salah satu model yang akan dipakai di dalam penelitiannya. Ketiga model tersebut, yaitu:

a) **Model *Common Effect***

Model *common effect* atau *pooled regression* merupakan model regresi data panel yang paling sederhana. Model ini pada dasarnya mengabaikan struktur panel dari data, sehingga diasumsikan bahwa perilaku antar individu sama dalam berbagai kurun waktu atau dengan kata lain pengaruh spesifik dari masing-masing individu diabaikan atau dianggap tidak ada. Dengan demikian, akan dihasilkan sebuah persamaan regresi yang sama untuk setiap unit *cross-section*. Persamaan regresi untuk model *common effect* dapat dituliskan sebagai berikut:

$$Y_{it} = \alpha + \beta X_{it} + u_{it} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T$$

Keterangan : Y = variabel dependen

α = koefisien regresi

X = variabel independen

β = estimasi parameter (koefisien)

u = *error term*

N = jumlah (individu)

T = jumlah periode waktu.

Berdasarkan asumsi struktur matriks varians-covarians residual, maka pada model *common effect*, terdapat empat metode estimasi yang dapat digunakan, yaitu:

- 1) *Ordinary Least Square (OLS)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat homoskedastik dan tidak ada *cross sectional correlation*,
- 2) *Generalized Least Square (GLS)/Weighted Least Square (WLS): Cross Sectional Weight*, jika struktur matriks varians-kovarians residualnyadiasumsikan bersifat heteroskedastik dan tidak ada *cross sectional correlation*,
- 3) *Feasible Generalized Least Square (FGLS) / Seemingly Uncorrelated Regression (SUR)* atau *Maximum Likelihood Estimator (MLE)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada *cross sectional correlation*,
- 4) *Feasible Generalized Least Square (FGLS)* dengan proses *auto regressive (AR)* pada *error term*-nya, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada korelasi antar waktu pada residualnya.

b) Model *Fixed Effect*

Jika model *common effect* cenderung mengabaikan struktur panel dari data dan pengaruh spesifik masing-masing individu, maka model *fixed effect* adalah sebaliknya. Pada model ini, terdapat efek spesifik individu α_i dan diasumsikan berkorelasi dengan variabel penjelas yang teramati X_{it} . Ekananda (2005) menyatakan bahwa berdasarkan asumsi struktur matriks varians-kovarians residual, maka

pada model *fixed effect*, terdapat tiga metode estimasi yang dapat digunakan, yaitu:

- 1) *Ordinary Least Square (OLS/LSDV)*, jika struktur matriks varians-covarians residualnya diasumsikan bersifat homoskedastik dan tidak ada *cross sectional correlation*,
- 2) *Weighted Least Square (WLS)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan tidak ada *cross sectional correlation*,
- 3) *Seemingly Uncorrelated Regression (SUR)*, jika struktur matriks varians-covarians residualnya diasumsikan bersifat heteroskedastik dan ada *cross sectional correlation*.

c) Model *Random Effect*

Pada model *random effect*, efek spesifik dari masing-masing individu α_i diperlakukan sebagai bagian dari komponen *error* yang bersifat acak dan tidak berkorelasi dengan variabel penjelas yang teramati X_{it} . Dengan demikian, persamaan model *random effect* dapat dituliskan sebagai berikut

$$Y_{it} = \alpha + \beta X_{it} + w_{it}; i = 1, \dots, N; t = 1, \dots, T$$

dimana $w_{it} = \alpha + u_{it}$; $E(w_{it}) = 0$; $E(w_{it}^2) = \sigma^2 + \sigma_a^2$; $E(w_{it}w_{jt-l}) = 0$; $i \neq j$;

Meskipun komponen error w_{it} bersifat homokedastik, nyatanya terdapat korelasi antara w_{it} dan w_{it-s} (*equicorrelation*). Karena itu, metode OLS tidak bisa digunakan untuk mendapatkan estimator yang efisien bagi model *random effect*. Metode yang tepat

untuk mengestimasi model *random effect* adalah *Generalized Least Squares (GLS)* dengan asumsi homokedastik dan tidak ada *cross-sectional correlation*.

2. Pemilihan model terbaik

Untuk mengetahui model estimasi data panel terbaik, diperlukan pengujian signifikansi antar model sebagai berikut.:

a) Signifikansi *fixed effect* (common vs fixed)

Ho : *Common effect* lebih baik dari *fixed effect*

H1 : *Fixed Effect* lebih baik dari *common effect*

Alpha : 5%

Ketentuan : tolak Ho jika nilai p-value < alpha atau dengan membandingkan besar F_{hitung} dibandingkan dengan F_{tabel} . Jika $F_{hitung} > F_{tabel}$, maka Ho ditolak, maka *fixed effect* lebih baik dibandingkan dengan *common effect*.

b) Signifikans *random effect* (common vs random)

Ho : model estimasi *common effect* lebih baik dibandingkan dengan *random effect*

H1 : model estimasi *random effect* lebih baik dibandingkan dengan *common effect*

Alpha : 5%

Ketentuan : Tolak Ho jika $LM > \text{Chi-Square}$ dengan derajat bebas 1 ($df=1$)

c) Uji Hausman (fixed vs random)

Ho : model estimasi *random effect* lebih baik dari *fixed effect*

H1 : model estimasi *fixed effect* lebih baik dari *random effect*

Alpha : 5%

Ketentuan : tolak Ho jika p-value < alpha. Selain melihat nilai p-value < alpha, dapat dilihat juga dengan membandingkan nilai chi-square hitung dengan chi-square table. Jika chi-square hitung > dari chi-square table, maka Ho ditolak berarti *fixed effect* lebih baik.

Sementara itu, Judge *et al.* dalam Gujarati (2003) memberikan sejumlah pertimbangan terkait pilihan apakah menggunakan model *fixed effect* (FE) ataukah model *random effect* (RE). Pertimbangan-pertimbangan itu adalah sebagai berikut:

1. Jika jumlah data *time series* (T) besar dan jumlah data *cross-section* (N) kecil, ada kemungkinan perbedaan nilai parameter yang diestimasi dengan FE dan RE cukup kecil. Karena itu, pilihan ditentukan berdasarkan kemudahan perhitungan. Dalam hal ini, adalah model FE.
2. Ketika N besar dan T kecil estimasi kedua metode dapat berbeda secara signifikan. Pada kondisi seperti ini, pilihan ditentukan berdasarkan keyakinan apakah individu yang diobservasi merupakan sampel acak yang diambil dari populasi tertentu atau tidak. Jika observasi bukan merupakan sampel acak, maka digunakan model FE. Jika sebaliknya, maka digunakan model RE.

3. Jika efek individu tidak teramati α_i berkorelasi dengan satu atau lebih variabel bebas, maka estimasi dengan RE bias, sedangkan estimasi dengan FE tidak bias.
4. Jika N besar dan T kecil, serta semua asumsi yang disyaratkan oleh model RE terpenuhi, maka estimasi dengan menggunakan RE lebih efisien dibanding estimasi dengan FE.

Dalam penelitian ini, penentuan apakah model FE atau RE yang akan digunakan selain didasarkan pada sejumlah pertimbangan yang telah disebutkan, juga akan didasarkan pada kriteria ekonomi (*make sense* secara ekonomi). Dalam hal ini, adalah kesesuaian tanda hasil estimasi koefisien regresi setiap variabel di dalam model dengan teori dan kewajaran besaran nilai koefisien hasil estimasi tersebut.

Setelah menentukan spesifikasi model yang akan digunakan, tahapan selanjutnya adalah memilih metode estimasi (estimator) yang tepat sesuai dengan struktur varian kovarian residual. Konsekuensi yang muncul ketika membangun model regresi dengan data panel adalah bertambahnya komponen residual, karena adanya dimensi *cross-section* dan *time-series* pada data. Kondisi ini menyebabkan matriks varian kovarian residual menjadi sedikit lebih kompleks bila dibandingkan dengan model regresi klasik yang hanya menggunakan data *cross-section* atau data *time-series*.

Pada model regresi klasik, pelanggaran terhadap asumsi klasik terkait residual, seperti heterokedastisitas dan autokerelasi merupakan

masalah serius yang mengakibatkan penduga parameter regresi yang diestimasi dengan OLS tidak lagi bersifat BLUE (*best linier unbiased estimator*). Tindakan yang biasa dilakukan untuk mengatasi masalah ini adalah dengan melakukan penghitungan *robust standard error*. Dalam pemodelan regresi dengan data panel, terjadinya pelanggaran asumsi regresi linier klasik pada residual adalah hal yang sangat sulit dihindari, dan tidak seperti pada regresi klasik, pelanggaran dapat diakomodasi untuk menentukan metode estimasi terbaik bagi spesifikasi model yang digunakan.

Terdapat beberapa kemungkinan struktur varian kovarian residual yang mungkin terjadi pada model regresi data panel. Berbagai kemungkinan yang dibahas pada bagian ini adalah yang biasa dijumpai pada estimasi model dengan *common effects* dan *fixed effect*. Karena itu, metode-metode estimasi yang dapat digunakan terkait struktur varian kovarian residual yang dipaparkan pada bagian ini hanya akan diterapkan pada model yang diestimasi dengan *common effects* atau *fixed effect*.

a. Pemilihan Estimator Struktur Homoskedastik atau Heteroskedastik dengan Uji *Lagrange Multiplier (LM)*

Pada pengujian ini, hipotesis nul (H_0) yang digunakan adalah bahwa struktur varians-covarians residual bersifat homoskedastik. Sementara hipotesis alternatif (H_1) adalah bahwa struktur varians-covarians residual bersifat heteroskedastik.

Secara matematis, statistik uji yang digunakan dapat dirumuskan sebagai berikut :

$$LM = \frac{T}{2} \sum_{i=1}^n \left[\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right]^2$$

Di mana T adalah jumlah periode waktu, n adalah jumlah individu, adalah varians residual persamaan ke- i pada kondisi homoskedastik, dan adalah *Sum Square Residual (SSR)* persamaan *system* pada kondisi homoskedastik.

Statistik uji LM ini mengikuti distribusi statistik *chi-square* dengan derajat bebas sebanyak $n-1$. Jika nilai statistik LM lebih besar dari nilai kritis statistik *chi-square*, maka hipotesis nul ditolak, yang berarti struktur varians-covarians residual bersifat heteroskedastik.

b. Pemilihan Estimator Struktur Heteroskedastik dan Ada *Cross-sectional Correlation*

Pengujian ini dilakukan apabila hasil pengujian LM menunjukkan bahwa struktur varians-covarians residual bersifat heteroskedastik. Pada pengujian ini, hipotesis nul (H_0) yang digunakan adalah bahwa struktur varians-covarians residual bersifat heteroskedastik dan tidak ada *cross sectional correlation*. Sementara hipotesis alternatifnya (H_1) adalah bahwa struktur varians-covarians residual bersifat heteroskedastik dan ada *cross sectional correlation (Seemingly Uncorrelated Regression/SUR)*.

Secara matematis, statistik uji yang digunakan dapat dirumuskan sebagai berikut:

$$\lambda LM = T \sum_{i=2}^n \sum_{j=1}^{i-1} r_{ij}^2$$

Di mana T adalah jumlah periode waktu, n adalah jumlah individu, dan r_{ij} adalah *residual correlation coefficient* antara persamaan ke- i dan ke- j .

Statistik uji LM ini mengikuti distribusi statistik *chi-square* dengan derajat bebas sebanyak $n(n-1)/2$. Jika nilai statistik LM lebih besar dari nilai kritis statistik *chi-square*, maka hipotesis nul akan ditolak, yang berarti struktur varians-covarians residual bersifat heteroskedastik dan ada *cross sectional correlation* (*Seemingly Uncorrelated Regression/SUR*).

Pengujian statistik juga dilakukan untuk mengetahui apakah model regresi non linier merupakan model yang tepat untuk menggambarkan hubungan antar variabel dan apakah ada hubungan yang signifikan diantara variabel-variabel dependen dengan variabel-variabel penjelas (seperti : uji statistik t dan uji statistik F) selain itu kita bisa melihat nilai hasil estimasi untuk R^2 (koefisien determinasi).

3. Pengujian kriteria statistik

a) Uji signifikansi simultan (uji f)

Uji F adalah uji signifikansi yang digunakan untuk menguji koefisien regresi peubah bebas secara keseluruhan atau simultan. Selain itu, uji F

juga dapat digunakan untuk mengetahui apakah model regresi dapat digunakan untuk memprediksi variabel terikat atau tidak. Nilai F dapat dihitung dengan rumus: ⁴⁵

$$F = \frac{R^2 / (n-1)}{(1-R^2) / (nT-n-k)}$$

Keterangan :

R^2 = Koefisien determinasi

K = Jumlah variabel

n = Jumlah sampel

T = Jumlah unit waktu

Hasil yang diperoleh kemudian dibandingkan dengan tabel F sebagai Fkritis, dengan ketentuan taraf signifikan (α) adalah 0,05. Dalam hal ini perlu ditentukan hipotesis nol dan hipotesis tandingnya:

Ho: $\beta_1 = \beta_2 = \beta_3 = 0$ (model regresi tidak berarti atau tidak signifikan)

Hi: $\beta_1 \neq \beta_2 \neq \beta_3 \neq 0$ (model regresi berarti atau signifikan)

b) Uji signifikansi parameter individual (uji t)

Uji t adalah uji signifikan yang digunakan untuk menguji koefisien regresi peubah bebas satu demi satu. Dengan demikian, bagi setiap nilai koefisien regresi dapat dihitung nilai t-nya. Sebelum melakukan pengujian, biasanya dibuat hipotesis terlebih dahulu.

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

⁴⁵ Damodar N. Gujarati, *Dasar-dasar Ekonometrika Jilid 2*, Jakarta: Erlangga, 2007, p.69

Keterangan: $i = 1, 2$)

Nilai T dapat dihitung dengan rumus: ⁴⁶

$$t = \frac{\beta_i}{SE(\beta_i)}$$

Keterangan :

T = adalah nilai t hitung

b_i = koefisien variabel i

SE (b_i) = *standard error* variabel

Hasil yang diperoleh kemudian dibandingkan dengan tabel t sebagai t kritis, dengan ketentuan taraf signifikan (α) adalah 0,05 dan derajat kebebasan ($n - K$).

Kriteria pengujian:

- 1) Jika $|t_{\text{hitung}}| > t_{\text{tabel}}$, maka koefisien regresi dikatakan signifikan, artinya variabel bebas X_i mempunyai pengaruh yang cukup berarti terhadap variabel terikat Y.
- 2) Jika $|t_{\text{hitung}}| < t_{\text{tabel}}$, maka koefisien regresi dikatakan tidak signifikan.
- 3) Jika $|t_{\text{hitung}}| = t_{\text{tabel}}$, maka tidak dapat ditarik kesimpulan

⁴⁶ Nachrowi Djalal, *Penggunaan Teknik Ekonometri*, Jakarta: PT. RajaGrafindo Persada, 2002, p.25

c) Uji koefisien determinasi (uji R-Square)

R^2 digunakan untuk mengukur kebaikan atau kesesuaian suatu model persamaan regresi. Besaran R^2 dihitung dengan rumus:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

dan R^2 *adjusted* dihitung dengan rumus:

$$\overline{R^2} = 1 - (1 - R^2) \frac{nT - 1}{nT - n - k}$$

Di mana: ESS: jumlah kuadrat yang dijelaskan

RSS: jumlah kuadrat residual

TSS: jumlah kuadrat total

n : jumlah observasi

T : jumlah periode waktu

k : banyaknya variabel bebas tanpa intersep

Adjusted R² digunakan karena sudah menghilangkan pengaruh penambahan variabel bebas dalam model, karena nilai R^2 akan terus naik seiring dengan penambahan variabel bebas. Karena itu kita harus berhati-hati dalam menggunakan nilai R^2 ketika menilai kebaikan dan kesesuaian suatu model persamaan regresi. Penggunaan *adjusted R²* sudah memperhitungkan jumlah derajat bebas

4. Uji Asumsi Klasik

Untuk membangun persamaan regresi panel yang terbaik dari kriteria ekonometrika, perlu dilakukan penyelidikan dan penanganan adanya masalah-masalah yang berkaitan dengan pelanggaran asumsi dasar. Berikut ini adalah asumsi-asumsi yang diperlukan dalam analisis regresi:

a. Normalitas

Pengujian normalitas dilakukan untuk mengetahui apakah residual berdistribusi normal atau tidak. Berdasarkan dalil limit pusat (*central limit theorem*), ada kecenderungan *residual* yang terjadi sebenarnya menyebar secara normal. Jika residual ε , merupakan jumlah *residual* dari beberapa sumber maka apapun sebaran peluang masing-masing *residual* itu, akan mendekati sebaran normal bila komponen *residual*-nya semakin banyak. Pemeriksaan kenormalan terhadap residual dapat dilakukan menggunakan plot persentil-persentil (P-P Plot). Jika plot mengikuti garis lurus, maka residual mengikuti sebaran normal (Drapper dalam Firmansyah, 2009).

Pengujian normalitas dilakukan untuk mengetahui apakah residual berdistribusi normal atau tidak. Berdasarkan dalil limit pusat (*central limit theorem*), ada kecenderungan residual yang terjadi sebenarnya menyebar secara normal. Jika residual merupakan jumlah residual dari beberapa sumber, maka apapun sebaran peluang masing-masing residual itu, akan mendekati sebaran normal bila komponen residual semakin banyak. Salah satu metode yang dapat digunakan untuk menguji Normalitas adalah *Jarque-Bera test*. Uji statistik ini dapat dihitung dengan rumus berikut:

$$JB = n \left[\frac{\mu_3^2}{6\mu_2^3} + \frac{(\frac{\mu_4}{\mu_2} - 3)^2}{24} \right]$$

di mana:

n = jumlah sampel

μ_2 = varians

μ_3 = skewness

μ_4 = kurtosis

Jarque-Bera test mempunyai distribusi *chi square* dengan derajat bebas dua. Jika hasil *Jarque-Bera test* lebih besar dari nilai *chi square* pada $\alpha=5$ persen, maka tolak hipotesis nul yang berarti tidak berdistribusi normal. Jika hasil *Jarque- Bera test* lebih kecil dari nilai *chi square* pada $\alpha=5$ persen, maka terima hipotesis nul yang berarti *error term* berdistribusi normal.

b. Heteroskedastisitas

Asumsi dalam model regresi adalah nilai residual memiliki nilai rata-rata nol, residual memiliki varian yang konstan serta residual suatu observasi tidak saling berhubungan dengan residual lainnya sehingga menghasilkan estimator yang BLUE.

Apabila asumsi tidak terpenuhi, yang terpengaruh hanyalah slope estimator dan ini tidak membawa konsekuensi serius dalam analisis ekonometris. Sedangkan apabila asumsi residual memiliki varian yang konstan serta residual suatu observasi tidak saling berhubungan dengan residual lainnya sehingga menghasilkan estimator yang BLUE ini dilanggar, maka akan membawa dampak serius bagi prediksi dengan model yang dibangun.

Dalam kenyataannya, nilai residual sulit memiliki varian yang konstan. Hal ini sering terjadi pada data yang bersifat *cross section* dibanding data *time series*. Untuk mengidentifikasi ada tidaknya

masalah *heterokedastisitas*, ada beberapa metode yang dapat digunakan dalam program *Eviews 6.0* ini. Dalam hal ini peneliti menggunakan Uji *Lagrange Multiplier* (LM). Untuk menghilangkan masalah ini peneliti menggunakan penimbang *white cross-section* pada *fixed effect* model.

c. Autokorelasi

Autokorelasi adalah hubungan antara residual satu observasi dengan residual observasi lainnya. Autokorelasi lebih mudah timbul pada data yang bersifat *time series*, karena berdasarkan sifatnya, data masa sekarang dipengaruhi oleh data pada masa-masa sebelumnya. Meskipun demikian, tetap dimungkinkan autokorelasi dijumpai pada data yang bersifat *cross section*. Autokorelasi dapat berbentuk otokorelasi positif dan otokorelasi negatif. Dalam analisis *time series*, lebih besar kemungkinan terjadi otokorelasi positif, karena variabel yang dianalisis biasanya mengandung kecenderungan meningkat.