

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **A. Tujuan Penelitian**

Tujuan penelitian ini adalah untuk memperoleh data yang sah, benar dan dapat dipercaya (dapat diandalkan atau reliable), tentang :

- a. Pengaruh antara tingkat pendidikan (variabel bebas) terhadap kemiskinan (variabel terikat) penduduk di Indonesia tahun 2006-2010.
- b. Pengaruh antara tingkat pengangguran (variabel bebas) terhadap kemiskinan (variabel terikat) penduduk di Indonesia tahun 2006-2010.
- c. Pengaruh antara tingkat pendidikan dan tingkat pengangguran terhadap kemiskinan penduduk di Indonesia tahun 2006-2010.

#### **B. Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan di Indonesia karena wilayah ini ditemukan fenomena tentang masalah kemiskinan, juga agar peneliti lebih mengetahui keadaan sosial di Indonesia karena merupakan tempat tinggal peneliti.

Waktu penelitian dilaksanakan 1 bulan pada bulan April 2012 dengan alasan pada waktu itu paling luas untuk melakukan penelitian karena waktu tersebut dianggap lebih efektif, sehingga peneliti dapat lebih memfokuskan diri pada pelaksanaan penelitian.

### C. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode *Ekspos Facto* yaitu usaha penelitian untuk mendapatkan keterangan-keterangan yang jelas dengan menggunakan data yang sudah terjadi. Metode ini digunakan dengan alasan karena menggunakan data yang tersedia dengan menggunakan pendekatan korelasi sesuai dengan tujuan penelitian yaitu untuk mengetahui pengaruh antara tingkat pendidikan dan tingkat pengangguran dengan kemiskinan penduduk di Indonesia.

Analisis data panel adalah suatu metode mengenai gabungan dari data antar waktu (*time series*) dengan data antar individu/daerah (*cross section*) dengan menggunakan program Microsoft Excel 2007, SPSS dan Eviews6. Kegiatan pengolahan data dengan menggunakan *Microsoft Excel 2007* mencakup pembuatan tabel dan grafik untuk kepentingan analisis; Pengolahan data dengan *SPSS* untuk mendeteksi apakah terdapat gangguan multikolinietas yang terjadi pada variabel yang masuk ke dalam model regresi data panel terbaik dan pengujian asumsi normalitas; *Eviews6* dilakukan untuk mengestimasi parameter-parameter model regresi dan statistik-statistik yang digunakan untuk mengetahui apakah variabel bebas memiliki pengaruh yang signifikan terhadap variabel terikat serta uji-uji formal menyangkut pemilihan metode estimasi serta spesifikasi model terbaik.

Dengan menggunakan analisis data panel untuk mengetahui pengaruh antara variabel tingkat pendidikan (PD) dan variabel tingkat pengangguran

(PG) terhadap variabel kemiskinan (MS), model yang akan digunakan untuk mengetahui kemiskinan di Indonesia yaitu :

$$MS = f ( PD, PG ) \quad (3.1)$$

$$MS = \beta_0 + \beta_1 PD_{it} + \beta_2 PG_{it} + E_{it} \quad (3.2)$$

Dimana :

$MS$  = variabel kemiskinan.

$PD$  = variabel tingkat pendidikan.

$PG$  = Variabel tingkat pengangguran.

$i$  = *cross section*.

$t$  = *time series*.

$\beta_0$  = konstanta.

$\beta_1, \beta_2$  = koefisien.

$E$  = *error*.

#### **D. Jenis dan Sumber Data**

Data sekunder merupakan data yang bukan diusahakan sendiri pengumpulannya oleh peneliti, tetapi mengambil dari Badan Pusat Statistik, dokumen-dokumen perusahaan atau organisasi, surat kabar dan majalah, ataupun publikasi lainnya. Data sekunder yang digunakan adalah penggabungan dari deret berkala (*time series*) dari tahun 2006-2010 dan deret lintang (*cross section*) sebanyak 33 propinsi di Indonesia yang menghasilkan 165 observasi. Data yang diperlukan adalah:

1. Data persentase penduduk miskin untuk masing-masing provinsi di Indonesia tahun 2006 - 2010.

2. Data rata-rata lama sekolah sebagai proxy tingkat pendidikan untuk masing-masing provinsi di Indonesia tahun 2006-2010.
3. Data tingkat pengangguran terbuka sebagai proxy tingkat pengangguran untuk masing-masing provinsi di Indonesia tahun 2006-2010.

Data yang digunakan untuk mencapai tujuan dalam penelitian ini sepenuhnya diperoleh melalui studi pustaka sebagai metode pengumpulan datanya, sehingga tidak diperlukan teknik sampling serta kuesioner. Sebagai pendukung, digunakan buku referensi, jurnal, surat kabar, serta dari browsing website internet yang terkait dengan masalah kemiskinan.

## **E. Operasionalisasi Variabel Penelitian**

### **1. Tingkat Pendidikan ( $X_1$ )**

#### **a. Definisi Konseptual**

Tingkat Pendidikan adalah urutan pendidikan (proses untuk mendapatkan pengetahuan dan pemahaman lebih tinggi) formal mulai dari pendidikan dasar sampai dengan pendidikan tinggi dengan memperhitungkan tingkat pendidikan yang sedang diduduki dan pendidikan tertinggi yang ditamatkan.

#### **b. Definisi Operasional**

Tingkat Pendidikan adalah kegiatan individu dalam pembelajaran yang diukur dari rata-rata lama sekolah yakni perbandingan banyaknya penduduk usia 15 tahun ke atas yang menjalani pendidikan terhadap pendidikan yang terakhir ditamatkan

SD selama 6 tahun, SLTP selama 3 tahun, SLTA selama 3 tahun, DIII selama 3 tahun, S1 selama 4 tahun, S2 selama 2 tahun dan S3 selama 3 tahun yang diambil dari Badan Pusat Statistik.

## **2. Tingkat Pengangguran ( $X_2$ )**

### **a. Definisi Konseptual**

pengangguran adalah penduduk yang sedang mencari pekerjaan pada suatu tingkat upah tertentu, tetapi tidak dapat memperoleh pekerjaan yang diinginkannya, meliputi penduduk yang sedang mencari pekerjaan, penduduk yang sedang mempersiapkan suatu usaha, penduduk yang merasa tidak mungkin mendapatkan pekerjaan, penduduk yang sudah punya pekerjaan tetapi belum mulai bekerja.

### **b. Definisi Operasional**

Pengangguran ialah angkatan kerja produktif dengan batasan usia 15 tahun keatas yang sedang mencari pekerjaan, yang mempersiapkan usaha, yang tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan (sebelumnya dikategorikan sebagai bukan angkatan kerja), dan yang sudah punya pekerjaan tetapi belum mulai bekerja (sebelumnya dikategorikan bekerja), dan pada waktu yang bersamaan mereka tak bekerja yang diambil dari Badan Pusat Statistik.

### 3. Kemiskinan (Y)

#### a. Definisi Konseptual

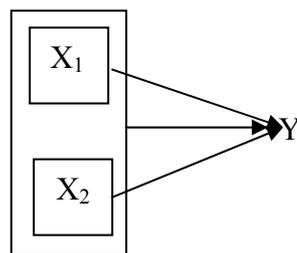
Kemiskinan adalah sebuah kondisi yang berada di bawah nilai standar kebutuhan minimum, baik untuk makanan dan non makanan yang disebut garis kemiskinan (*poverty line*) atau batas kemiskinan (*poverty threshold*) dari kebutuhan yang bersifat materil.

#### b. Definisi Operasional

Kemiskinan adalah suatu kondisi seseorang atau kelompok masyarakat yang mengalami berbagai kekurangan (ketidakberdayaan) yakni dengan memiliki pendapatan dibawah Rp 7.000 per hari untuk memperoleh makanan dengan kandungan kalori kurang dari 2.100 per harinya.

### F. Konstelasi Hubungan Antar Variabel

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



Keterangan :

X<sub>1</sub> : Tingkat Pendidikan (variabel bebas 1)

X<sub>2</sub> : Tingkat Pengangguran (variabel bebas 2)

Y : Kemiskinan (variabel terikat)

## G. Teknik Analisis Data

### 1. Metode Analisis

#### a. Analisis Data Panel

Analisis dengan menggunakan panel data adalah kombinasi antara deret waktu (*time-series data*) dan kerat lintang (*cross-section data*). Gujarati menyatakan bahwa untuk menggambarkan data panel secara singkat, misalkan pada data *cross section*, nilai dari satu variabel atau lebih dikumpulkan untuk beberapa unit sampel pada suatu waktu. Dalam data panel, unit *cross section* yang sama di survey dalam beberapa waktu. Dalam model panel data, persamaan model dengan menggunakan data *cross-section* dapat ditulis sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i ; i = 1, 2, \dots, N \quad (3.3)$$

dimana N adalah banyaknya data *cross-section*

Sedangkan persamaan model dengan *time-series* adalah :

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t ; t = 1, 2, \dots, T \quad (3.4)$$

dimana T adalah banyaknya *data time-series*

Mengingat data panel merupakan gabungan dari *time-series* dan *cross-section*, maka model dapat ditulis dengan :

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \epsilon_{it} \quad (3.5)$$

$i = 1, 2, \dots, N ; t = 1, 2, \dots, T$

dimana :

N = banyaknya observasi

$T$  = banyaknya waktu

$N \times T$  = banyaknya data panel

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*. Dalam penelitian ini, penulis membatasi pembahasan pada data panel yang bersifat *balanced panel* dan jumlah unit individu lebih banyak dibandingkan jumlah unit waktu ( $N > T$ ). Penggunaan data panel pada dasarnya merupakan solusi akan ketidakterersediaan data *time-series* yang cukup panjang untuk kepentingan analisis ekonometrika.

Menurut Hsiao dalam Greene keunggulan penggunaan data panel dibandingkan deret waktu dan kerat lintang adalah:

- a) Dapat memberikan peneliti jumlah pengamatan yang besar, meningkatkan *degrees of freedom* (derajat kebebasan), data memiliki variabilitas yang besar dan mengurangi kolinearitas antara variabel penjelas, dimana dapat menghasilkan ekonometri yang efisien.
- b) Dengan panel data, data lebih informatif, lebih bervariasi, yang tidak dapat diberikan hanya oleh data *cross section* dan *time series* saja.
- c) Panel data dapat memberikan penyelesaian yang lebih baik dalam inferensi perubahan dinamis dibandingkan data *cross section*.

## **b. Estimasi Model**

Dalam mengestimasi model regresi data panel terdapat tiga spesifikasi model yang mungkin digunakan, yakni model *common effects*, *fixed effect*, dan *random effect*.

### **1) Model Common Effect**

Model *common effects* 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*. Sesuatu yang secara realistis tentunya kurang dapat diterima. Karena itu, model ini sangat jarang digunakan dalam analisis data panel.

### **2) 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  $\alpha_i$  dan diasumsikan berkorelasi dengan variabel penjelas yang teramati  $X_{it}$ .

Ekananda menyatakan bahwa berdasarkan asumsi struktur matriks varians-kovarians residual, maka pada model *fixed effect*, terdapat 3 metode estimasi yang dapat digunakan, yaitu:

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

### 3) Model Pendekatan Efek Acak (*Random Effect*).

Pendekatan ini mengasumsikan *unobservable* individual *effects* ( $u_{it}$ ) tidak berkorelasi dengan *regressor* (X) atau dengan kata lain  $u_{it}$  diasumsikan bersifat random. Sebelum model diestimasi dengan model yang tepat, terlebih dahulu dilakukan uji spesifikasi apakah *fixed effect* atau *random effect* atau keduanya memberikan hasil yang sama.

#### c. Penyeleksian Model Estimasi Data Panel

##### 1) Pengujian Signifikansi *Common Effects* atau *Fixed Effects*

Signifikasi model *fixed effect* dapat dilakukan dengan uji statistik F. Uji F digunakan untuk mengetahui apakah teknik

regresi data panel dengan *fixed effect* lebih baik dari model regresi data panel tanpa variabel *dummy* (*common effect*) dengan melihat *residual sum squares* (*RSS*).

$$F = \frac{(RSS_1 - RSS_2)/(n-1)}{RSS_2 / (nT - n - k)} \quad (3.6)$$

Di mana,  $n$  adalah jumlah individu,  $T$  adalah periode waktu,  $k$  adalah parameter dalam model *fixed effect*,  $RSS_1$  dan  $RSS_2$  masing-masing merupakan *residual sum of squares* teknik tanpa variabel *dummy* dan teknik *fixed effect* dengan variabel *dummy*. Nilai statistik  $F$  hitung akan mengikuti distribusi statistik  $F$  dengan derajat bebas sebanyak  $(n-1)$  untuk numerator dan  $(nT - n - k)$  untuk *denominator*. Jika nilai statistik  $F$  hitung lebih besar dari  $F$  tabel, maka hipotesis nul akan ditolak, yang berarti koefisien intersep dan slope adalah sama tidak berlaku, sehingga teknik regresi data panel dengan *fixed effect* lebih baik dari *common effect*.

## 2) Pengujian Signifikansi *Common Effects* atau *Random Effects*

Breush dan Pagan telah mengembangkan pengujian *Lagrange Multiplier* untuk mengetahui signifikansi dari *random effects* berdasarkan residual dari OLS (*common effects*). Secara matematis, statistik uji untuk *LM test* (*Lagrange Multiplier*) dapat dituliskan sebagai berikut:

$$LM = \frac{nT}{2(T-1)} \left[ \frac{\sum_{i=1}^n (\sum_{t=1}^T e_{it})^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 = \frac{nT}{2(T-1)} \left[ \frac{\sum_{i=1}^n [T \bar{e}_i]^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \quad (3.7)$$

Di bawah hipotesis nul, LM mengikuti sebaran *chi-square* dengan derajat bebas satu. Jika hasil LM statistik lebih besar dari nilai kritis statistik *chi-square*, maka hipotesis nul akan ditolak, yang berarti estimasi yang tepat untuk regresi data panel adalah metode *random effect* dibandingkan metode *common effects*.

### 3) Pengujian Signifikansi Fixed Effects atau Random Effects

Setelah menguji signifikansi antara *common effects* atau *fixed effects* serta *common effects* atau *random effects*, maka selanjutnya jika terbukti *fixed effects* dan *random effects* sama-sama lebih baik dari *common effects* adalah melakukan pengujian signifikansi *fixed effects* atau *random effects*.

Uji ini dilakukan dengan membandingkan  $\beta_{FE}$  dan  $\beta_{RE}$  untuk *subset* dari koefisien variabel-variabel yang bervariasi antar unit waktu (*time-varying variables*). Secara matematis dengan menggunakan notasi matriks, statistik uji Hausman ( $H$ ) dapat dituliskan sebagai berikut:

$$H = \left( \hat{\beta}_{FE} - \hat{\beta}_{RE} \right)' \left[ \text{var} \left( \hat{\beta}_{FE} \right) - \text{var} \left( \hat{\beta}_{RE} \right) \right]^{-1} \left( \hat{\beta}_{FE} - \hat{\beta}_{RE} \right) \quad (3.8)$$

Di bawah hipotesis nul, statistik uji ini mengikuti sebaran *chi-square* dengan derajat bebas  $M$ , di mana  $M$  adalah jumlah variabel penjelas yang nilainya bervariasi antar unit waktu di dalam model.

Hipotesis nul pada uji Hausman adalah efek spesifik individu tidak berkorelasi dengan peregresi atau dengan kata lain model *random effect* lebih baik bila dibandingkan dengan model *fixed effect*. Di bawah hipotesis nul, pendugaan parameter dengan menggunakan *random effect* adalah konsisten dan efisien, sedangkan pendugaan dengan *fixed effect* meskipun tetap konsisten, tetapi tidak lagi efisien. Di bawah hipotesis alternatif, estimasi dengan *random effect* menjadi tidak konsisten, sebaliknya estimasi dengan *fixed effect* tetap konsisten. Jika nilai statistik Hausman lebih besar daripada nilai kritis statistik *chi-square*, maka hipotesis nul akan ditolak, yang berarti estimasi yang tepat untuk regresi data panel adalah metode *fixed effect* dari pada metode *random effect*.

Gujarati 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, pilhan 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  $\alpha_i$  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.

## 2. Pengujian Asumsi Klasik

Menurut Greene “uji asumsi klasik perlu dilakukan karena dalam model regresi perlu memperhatikan adanya penyimpangan-penyimpangan atas asumsi klasik, karena pada hakekatnya jika asumsi klasik tidak dipenuhi maka variabel-variabel yang menjelaskan akan menjadi tidak efisien”<sup>46</sup>. Konsekuensi yang muncul ketika membangun model regresi dengan data panel adalah bertambahnya komponen residual, karena adanya dimensi *cross-section* dan *time-series* pada

---

<sup>46</sup> William H. Greene, *Econometric Analysis* (New York : New York University, 2002), p. 307.

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 autokorelasi merupakan masalah serius yang mengakibatkan penduga parameter regresi yang diestimasi dengan OLS tidak lagi bersifat BLUE (*best linear unbiased estimator*). Tindakan yang biasa dilakukan untuk mengatasi masalah ini adalah dengan melakukan penghitungan *robust standard error*.

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. Uji Heterokedastisitas (*Lagrange Multiplier*)**

Heterokedastisitas adalah keadaan dimana terjadi kesamaan varian residual untuk semua pengamatan pada model regresi. Uji heterokedastisitas digunakan untuk mengetahui ada atau tidaknya ketidaksamaan varian dari residual pada model regresi

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 :

$$LM = \frac{T}{2} \sum_{i=1}^n \left[ \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right]^2 \quad (3.9)$$

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. Uji Autokorelasi**

Autokorelasi adalah keadaan dimana terjadinya korelasi antar residual pada suatu pengamatan dengan pengamatan lain pada model regresi. Uji autokorelasi digunakan untuk mengetahui ada atau tidaknya korelasi yang terjadi antara residual pada suatu pengamatan dengan pengamatan lain pada model regresi.

Pengujian ini dilakukan apabila hasil pengujian LM pada poin (2.1) 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:

$$\lambda LM = T \sum_{i=2}^n \sum_{j=1}^{i-1} r_{ij}^2 \quad (3.10)$$

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*).

### c. Uji Normalitas

Uji normalitas bertujuan untuk menguji apakah dalam model regresi, variabel terikat dan variabel bebas, keduanya mempunyai

distribusi normal ataukah tidak. Model regresi yang baik adalah yang mempunyai distribusi normal atau mendekati normal. Ada beberapa metode untuk mengetahui normal atau tidak gangguan ( $\mu$ ) antara lain J-B test dan metode grafik. Penelitian ini akan menggunakan metode J-B test yang dilakukan dengan menghitung skweness dan kurtosis, apabila J-B hitung  $<$  nilai  $X^2$  (Chi Square) tabel, maka nilai residual berdistribusi normal. Model untuk mengetahui uji normalitas adalah:

$$JB = n \left[ \frac{\mu_3^2}{6\mu_2^3} + \frac{(\frac{\mu_4}{\mu_2} - 3)^2}{24} \right] \quad (3.10)$$

di mana:

$n$  = jumlah sampel

$2$  = varians

$3$  = skewness

$4$  = 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.

#### d. Uji Multikolinieritas

Multikolinieritas berarti adanya korelasi antar variabel bebas, yang terjadi karena variabel-variabel bebas tersebut memiliki hubungan pada populasi atau hanya pada sampel. Cara mendeteksi adanya multikolinieritas :

1. Dengan memeriksa *simple pairwise* (Pearson) *correlation* antar variabel independen. Batas nilai yang disarankan sebagai indikasi kolinieritas serius berbeda-beda (0,8 menurut Berry dan Felman, 1985; dan 0,9 menurut Griffith dan Amerhein, 1997). Nash dan Bradford menyebutkan bahwa suatu variabel independen berkorelasi tinggi dengan variabel independen lainnya jika  $r$  lebih dari 0,85.

2. Uji formal untuk mendeteksi keberadaan dari multikolinieritas yaitu dengan melihat nilai *Variance Inflation Factor* (VIF). Nilai VIF ini menunjukkan bagaimana varians dari sebuah estimator akan meningkat akibat adanya multikolinieritas. Nilai VIF diperoleh dengan formula berikut:

$$VIF_k = \frac{1}{1-R_k^2} \quad (3.11)$$

dimana  $k = 1, 2, \dots, p - 1$  dan  $R_k^2$  merupakan koefisien determinasi dari regresi berganda ketika  $X_k$  diregresikan dengan  $p-2$  variabel lainnya dalam model. Jika nilai VIF lebih dari 10, maka hal tersebut dapat berindikasi bahwa multikolinieritas bersifat serius dan akan memengaruhi estimasi yang menggunakan OLS

karena meskipun estimator tetap bersifat *unbiased* namun sudah tidak lagi memiliki varians yang minimum. Selain itu, keberadaan multikolinieritas juga akan membuat estimator bersifat sensitif untuk perubahan yang kecil pada data, sehingga akan mengakibatkan kesalahan (*misleading*) dalam menginterpretasikan suatu model regresi. Cara mengatasi adanya multikolinieritas antara lain melepas satu atau lebih variabel yang memiliki korelasi yang tinggi, mentransformasi model, atau memperbesar jumlah sampel.

### 3. Pengujian Kriteria Statistik

#### a. Uji Signifikansi Simultan (Uji-F)

Uji statistik F pada dasarnya menunjukkan apakah semua variabel bebas yang dimasukkan dalam model mempunyai pengaruh secara bersama-sama terhadap variabel terikat. Hipotesis yang digunakan:

1.  $H_0 : b_1, b_2 = 0$  semua variabel independen tidak mampu mempengaruhi variabel dependen secara bersama-sama.
2.  $H_1 : b_1, b_2 \neq 0$  semua variabel independen mampu mempengaruhi variabel dependen secara bersama-sama.

Nilai F hitung dirumuskan sebagai berikut:

$$F = \frac{R^2 / (n + k - 1)}{1 - R^2 / (nT - n - k)} \quad (3.12)$$

dimana:

$k$  = jumlah parameter yang diestimasi termasuk konstanta

$N$  = jumlah observasi

Pada tingkat signifikansi 5 persen dengan kriteria pengujian yang digunakan sebagai berikut:

- a)  $H_0$  diterima dan  $H_1$  ditolak apabila  $F$  hitung  $<$   $F$  tabel, yang artinya variabel penjelas secara bersama-sama tidak mempengaruhi variabel yang dijelaskan secara signifikan.
- b)  $H_0$  ditolak dan  $H_1$  diterima apabila  $F$  hitung  $>$   $F$  tabel, yang artinya variabel penjelas secara bersama-sama mempengaruhi variabel yang dijelaskan secara signifikan.

#### **b. Uji Signifikansi Parameter Individual (Uji-t)**

Uji  $t$  digunakan untuk mengetahui apakah variabel-variabel bebas secara parsial berpengaruh signifikan terhadap variabel tak bebasnya. Hipotesis pengujian:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Nilai  $t$  hitung dapat dicari dengan rumus:

$$T_{hitung} = \frac{\beta_i}{se(\beta_i)} \quad (3.13)$$

$\beta_i$  adalah nilai penduga parameter ke- $i$ ,  $se(\beta_i)$  adalah simpangan baku dari nilai penduga parameter ke- $i$ .

Pada tingkat signifikansi 5 persen dengan pengujian yang digunakan adalah sebagai berikut:

- a) Jika  $t\text{-hitung} > t\text{-tabel}$  maka  $H_0$  ditolak, artinya salah satu variabel independen mempengaruhi variabel dependen secara signifikan.
- b) Jika  $t\text{-hitung} < t\text{-tabel}$  maka  $H_0$  diterima, artinya salah satu variabel independen tidak mempengaruhi variabel dependen secara signifikan.

**c. Uji Koefisien Determinasi (Uji R-squared)**

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

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \quad (3.13)$$

Dan  $R^2$  *adjusted* dihitung dengan rumus:

$$\overline{R^2} = 1 - (1 - R^2) \frac{nT - 1}{nT - n - k} \quad (3.14)$$

*ESS*: Jumlah kuadrat yang dijelaskan

*RSS*: Jumlah kuadrat residual.

*TSS*: Jumlah kuadrat total.

*n*: Jumlah observasi (negara)

*T*: Jumlah periode waktu.

*k*: Banyaknya variabel bebas tanpa intersep.

*Adjusted R<sup>2</sup>* 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*  $R^2$  sudah memperhitungkan jumlah derajat bebas.