

## **BAB III**

### **METODOLOGI PENELITIAN**

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

Berdasarkan masalah-masalah yang peneliti rumuskan, tujuan penelitian ini adalah untuk mendapatkan pengetahuan yang tepat (sahih, benar dan valid) dan dapat dipercaya (dapat diandalkan) tentang:

1. Menghitung dan menganalisis besarnya pengaruh impor barang modal terhadap pertumbuhan ekonomi di Negara ASEAN.
2. Menghitung dan menganalisis besarnya pengaruh penyerapan tenaga kerja terhadap pertumbuhan ekonomi di Negara ASEAN.

#### **B. Objek dan Ruang Lingkup Penelitian**

Objek penelitian ini adalah impor barang modal dan penyerapan tenaga kerja terhadap pertumbuhan ekonomi di Negara ASEAN. Data yang digunakan dalam penelitian ini adalah data sekunder yang tersedia di Bank Dunia, Badan Pusat Statistik, dan lembaga statistik lainnya. Setiap variabel dari masing-masing negara digunakan data dalam jangka waktu 7 tahun, yaitu mulai dari tahun 2010 – 2016.

Ruang lingkup penelitian ini adalah untuk mengkaji pengaruh dari impor barang modal dan penyerapan tenaga kerja terhadap pertumbuhan ekonomi di Negara ASEAN. Penelitian ini dilakukan pada bulan Desember 2016 – Agustus 2017 karena merupakan waktu yang paling efektif bagi peneliti untuk

melaksanakan penelitian, sehingga peneliti dapat fokus pada saat penelitian. Selain itu, peneliti juga memiliki keterbatasan waktu sebagaimana yang telah ditetapkan dalam jadwal akademik. Tenaga dan materi yang terbatas juga merupakan salah satu keterbatasan yang dimiliki oleh peneliti.

### **C. Variabel Operasional Penelitian**

Variabel operasional penelitian ini diperlukan untuk memahami jenis dan indikator dari variable-variabel yang terkait dalam penelitian ini. Selain itu, proses ini dimaksudkan untuk menentukan skala pengukuran dari masing-masing variabel sehingga pengujian hipotesis dengan alat bantu statistik dapat dilakukan secara komprehensif.

#### **1. Pertumbuhan Ekonomi (Varibel Y)**

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

Pertumbuhan ekonomi merupakan proses kenaikan output perkapita dalam jangka panjang yang menyebabkan barang dan jasa yang diproduksi akan bertambah dan kemakmuran masyarakat meningkat.

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

Pertumbuhan ekonomi merupakan suatu kondisi perekonomian Negara untuk menuju keadaan yang lebih baik melalui peningkatan pendapatan nasional. Data pertumbuhan ekonomi yang digunakan ialah data pertumbuhan ekonomi di Neagara ASEAN. Data pertumbuhan ekonomi yang digunakan dalam penelitian ini memiliki jangka waktu 7 tahun, yaitu mulai dari tahun 2010-2016 untuk tujuh negara ASEAN.

## **2. Impor Barang Modal (X1)**

### **a. Definisi Konseptual**

Impor barang modal adalah suatu kegiatan perdagangan internasional dengan cara memasukkan barang yang dilakukan oleh perorangan atau perusahaan yang bergerak dibidang ekspor impor yang digunakan untuk keperluan kelancaran atau kelangsungan suatu kegiatan produksi.

### **b. Definisi Operasional**

Impor barang modal adalah sejumlah peralatan berat seperti mesin yang digunakan sebagai faktor input untuk memproduksi barang yang didatangkan dari luar negeri. Data yang digunakan dalam penelitian ini diperoleh dari Bank Dunia selama tahun 2010 sampai dengan tahun 2016.

## **3. Penyerapan Tenaga Kerja (X2)**

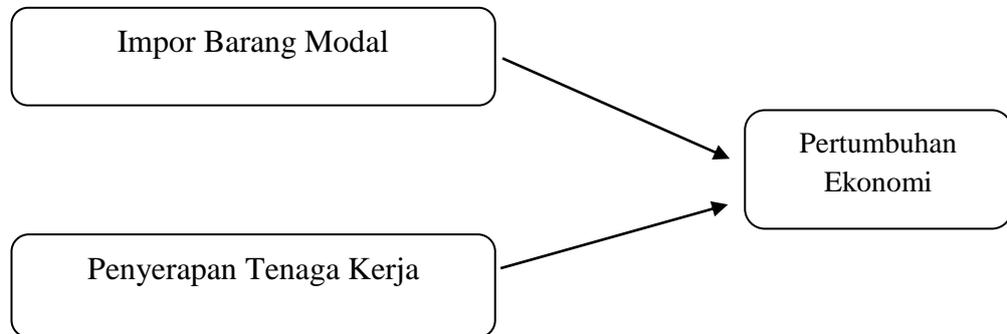
### **a. Definisi Konseptual**

Penyerapan tenaga kerja merupakan jumlah tenaga kerja yang terserap pada suatu sektor dalam waktu tertentu.

### **b. Definisi Operasional**

Penyerapan tenaga kerja ialah menghimpun tenaga kerja pada suatu sektor tertentu. Dalam penelitian ini, data diperoleh dari Bank Dunia selama 7 tahun di Negara ASEAN pada tahun 2010-2016.

Konstelasi pengaruh antar variable dapat digambarkan sebagai berikut:



Keterangan:

Impor Barang Modal : Variabel Independen X1

Penyerapan Tenaga Kerja : Variabel Independen X2

Pertumbuhan Ekonomi : Variabel Dependen Y

—————> : Arah Pengaruh

#### D. Jenis dan Sumber Data

Penelitian ini menggunakan data skunder dari setiap variabel, yaitu variabel Impor barang modal dan penyerapan ketenaga kerjaan. Menurut Istijanto (2009: 38) yang dimaksud dengan data skunder adalah data yang telah dikumpulkan oleh pihak lain bukan oleh periset itu sendiri, yang digunakan untuk tujuan yang lain. Data yang digunakan adalah data panel, yaitu kombinasi antara *cross section* dan *time series* (Gujarati, 2004: 636). Data panel yang dijadikan sampel dalam penelitian ini terdiri dari sepuluh negara ASEAN. Sedangkan rentang waktu yang digunakan dalam penelitian ini adalah 7 tahun, yaitu mulai dari tahun 2010-2016. Data yang digunakan dalam penelitian ini diperoleh dari

berbagai sumber, yaitu situs resmi dari Bank Dunia dan Badan Pusat Statistik, Serta situs-situs resmi lainnya.

### **E. Metode Pengumpulan Data**

Metode pengumpulan data adalah cara atau langkah yang digunakan oleh peneliti untuk mengumpulkan data yang dibutuhkan dalam melakukan penelitian. Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode *ekspos facto*. *Ekspos facto* adalah pencarian empiris yang sistematis Dimana peneliti tidak dapat mengendalikan variabel bebasnya Karena peristiwa ini telah terjadi atau sifatnya tidak dapat dimanipulasi. Cara menerapkan metode penelitian ini dengan menganalisis peristiwa-peristiwa yang terjadi dari tahun-tahun sebelumnya untuk mengetahui faktor-faktor yang dapat menimbulkan kejadian tersebut (Umar Husein, 2009: 28).

Metode ini bermanfaat untuk mencari dan menggambarkan hubungan antara dua variabel atau lebih serta mengukur seberapa besar hubungan antar variabel yang dipilih untuk diteliti. Metode ini dipilih karena sesuai untuk mendapatkan informasi yang bersangkutan dengan status gejala saat penelitian dilakukan.

### **F. Teknik Analisis Data**

Teknik Analisis yang digunakan dalam penelitian ini adalah analisis regresi linear berganda (*multiple linear regression*). Hal ini dikarenakan variabel yang digunakan dalam penelitian ini terdiri dari empat variabel, yaitu satu variabel dependen, dan tiga variabel independen. Langkah pertama yang dilakukan dalam teknik analisis regresi adalah mendeteksi gejala asumsi klasik dari data-data yang

digunakan dalam penelitian. Kemudian, karena data yang digunakan adalah data panel, maka diperlukan penentuan model estimasi yang tepat dengan melakukan beberapa uji. Terakhir dilakukanlah uji hipotesis dengan menggunakan uji t, uji F, dan mencari koefisien determinasi. Berikut ini adalah penjabaran dari langkah-langkah tersebut.

### **1. Deteksi Gejala Asumsi Klasik**

Model regresi data panel dapat dikatakan sebagai model yang baik, apabila memenuhi empat kriteria berikut: Best, Linear, Unbiased, dan Estimator (Fridayana Yudiantmaja, 2013). Keempat kriteria tersebut biasa disingkat dengan *BLUE*. Apabila model persamaan yang terbentuk tidak memenuhi kriteria *BLUE*, maka model persamaan tersebut diragukan dapat menghasilkan nilai-nilai prediksi yang akurat. Untuk itu perlu dilakukannya deteksi gejala asumsi klasik untuk mengetahui apakah model persamaan tersebut telah memenuhi kriteria *BLUE*. Hal ini dikarenakan model persamaan telah memenuhi kriteria *BLUE* apabila telah memenuhi asumsi klasik. Deteksi gejala asumsi klasik ini mencakup deteksi normalitas, deteksi linearitas, deteksi heterokedastisitas deteksi multikolinearitas, dan deteksi autokorelasi. Apabila model persamaan yang dideteksi telah bebas dari lima asumsi tersebut, maka dapat dikatakan model persamaan tersebut akan menjadi estimator yang baik (Agus Widarjono, 2007).

### a. Deteksi Normalitas

Deteksi normalitas juga dapat dilakukan dengan menggunakan uji *Jarque-Bera*, yaitu dengan mendeteksi normalitas pada residualnya yang dihasilkan dari model persamaan regresi linear yang digunakan. Uji *Jarque-Bera* ini menggunakan hipotesis sebagai berikut:

$H_0$ : Residual berdistribusi normal

$H_a$ : Residual tidak berdistribusi normal

Kriteria uji:  $H_0$  ditolak jika nilai  $JB > \text{chi square-tabel } (a,k)$  artinya residual tidak berdistribusi normal, dan jika sebaliknya maka residual berdistribusi normal. Selain melihat hasil dari nilai  $JB$ , dapat juga dengan melihat nilai probabilitas dari  $JB$ . Apabila  $P\text{-value}$  dari  $JB < 0,05$ , maka  $H_0$  ditolak artinya residual tidak berdistribusi normal, jika sebaliknya maka  $H_0$  diterima artinya residual berdistribusi normal.

### b. Deteksi Linearitas

Deteksi linearitas digunakan untuk mendeteksi apakah terdapat hubungan yang linear antara variabel dependen ( $Y$ ) dan independen ( $X$ ). Untuk mengetahui apakah terhadap hubungan yang linear tersebut, maka dapat dilakukan dengan *Ramsey Reset Test*. Menurut Ramsey bahwa jika tidak terdapat nonlinearitas maka berbagai transformasi nonlinear  $f_t = (\tilde{X}_t' \hat{\theta})$  tidak memberikan manfaat untuk menyatakan  $y_t$  (Kim T.H, Lee Y.S, et al, 2004).

Hipotesis yang digunakan untuk melakukan *Ramsey Reset Test* adalah sebagai berikut:

$H_0$  = terdapat hubungan linear antara variabel X dan Y

$H_a$  = terdapat hubungan non-linear antara variabel X dan Y

Berdasarkan hipotesis tersebut, maka  $H_0$  ditolak jika F-statistik > F-tabel, yang berarti hubungan antara variabel X dan Y bersifat non-linear. Jika sebaliknya, maka hubungan antara variabel X dan Y bersifat linear.

### c. Deteksi Heterokedastisitas

Deteksi heterokedastisitas bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan varians dari satu pengamatan ke pengamatan lainnya. Model regresi dikatakan baik apabila tidak terjadi heterokedastisitas, artinya adanya ketetapan atau konstan antara varians dari satu pengamatan ke pengamatan lainnya (Homokedastisitas). Hipotesis yang digunakan untuk mendeteksi heterokedastisitas berdasarkan uji *White* adalah sebagai berikut:

$H_0$  = (struktur *variance-covariance residual* homokedastik)

$H_a$  = (struktur *variance-covariance residual* heterokedastik)

Berdasarkan hipotesis tersebut, maka kriteria pengambilan kesimpulan yakni jika nilai probabilitas (*p-value*) dari Chi Square > 0,05, maka  $H_0$  diterima, artinya varians error bersifat homokedastik. Jika sebaliknya, maka  $H_0$  ditolak, yang berarti varians error bersifat heterokedastik.

#### d. Deteksi Multikolinieritas

Deteksi multikolinieritas bertujuan untuk mendeteksi apakah antara variabel independen (variabel bebas) terdapat korelasi. Sehingga sulit untuk memisahkan pengaruh antara variabel-variabel itu secara individu terhadap variabel terikat. Model regresi dikatakan baik apabila tidak ada korelasi antar variabel independen. Keberadaan multikolinieritas menyebabkan standar error cenderung semakin besar. Meningkatnya tingkat korelasi antar variabel, menyebabkan standar error semakin sensitif terhadap perubahan data.

Untuk mengetahui ada atau tidaknya multikolinieritas dapat dilihat dari nilai *Variance Inflation Factor (VIF)*. *VIF* merupakan suatu jumlah yang dapat menggambarkan bahwa variabel bebas dapat dijelaskan oleh variabel bebas lainnya. Batas *VIF* adalah 10, jika nilai *VIF* diatas 10 maka terjadi multikolinieritas. Menghitung *VIF (Variance Inflation Factor)* secara manual dapat dilakukan dengan menggunakan rumus:

[3.11]

$$VIF = \frac{1}{(1 - R_i^2)}$$

Keterangan:

$R_i^2$  = koefisien determinasi dari *auxiliary regression*

Menurut Gujarati (2004: 359) tingginya koefisien korelasi antar variabel bebas merupakan salah satu indikator dari adanya multikolinieritas antar

variabel bebas. Jika terjadi koefisien korelasi lebih dari 0,80 maka dapat dipastikan terdapat multikolinearitas antar variabel bebas.

#### e. Deteksi Autokorelasi

Deteksi autokorelasi bertujuan untuk mendeteksi apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode  $t$  dengan kesalahan pengganggu pada periode  $t-1$  (tahun sebelumnya) (Imam Ghazali, 2004: 110). Model regresi yang baik adalah tidak ada terjadi autokorelasi. Cara memprediksi dalam suatu model regresi terdapat autokorelasi atau tidak dapat dengan cara uji *Durbin-Watson* (*DW test*). Rumus statistik  $d$  *Durbin-Watson* sebagai berikut:

Uji *Durbin-Watson* akan menghasilkan nilai *Durbin-Watson* (DW) dan dari nilai *Durbin-Watson* tersebut dapat menentukan keputusan apakah terdapat autokorelasi atau tidak dengan melihat tabel berikut:

[3.12]

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e^2 t}$$

**Tabel 3.1**  
**Range *Durbin-Watson* untuk Autokorelasi**

Durbin-Watson	Kesimpulan
Kurang dari 1,10	Ada autokorelasi

1,10 – 1,54	Tanpa kesimpulan
1,55 – 2,46	Tidak ada autokorelasi
2,46 – 2,90	Tanpa kesimpulan
Lebih dari 2,91	Ada autokorelasi

Sumber: Muhammad Firdaus (2004: 101)

## 2. Penentuan Model Estimasi Regresi Data Panel

Model estimasi yang digunakan dalam penelitian ini adalah model estimasi regresi dengan menggunakan data panel. Data panel merupakan kombinasi dari data *cross-section* dan *time series*. Sehingga, secara umum persamaan regresi dengan menggunakan data panel dapat dituliskan sebagai berikut:

[3.1]

$$Y_{it} = \beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it}$$

Dimana  $Y_i$  merupakan variabel dependen yang nilainya dipengaruhi oleh tiga variabel *explanatory* yang dinotasikan dengan  $X_{2it}, X_{3it}, X_{4it}$ . Penggunaan notasi  $i$  pada variabel dependen dan variabel *explanatory* menunjukkan adanya nilai tersendiri dari setiap unit *cross-section*. Sedangkan notasi  $t$  menunjukkan adanya nilai tersendiri dari setiap unit *time series*.  $\beta_1, \beta_2, \beta_3$ , dan  $\beta_4$  merupakan parameter tetap yang nilainya tidak diketahui. Untuk memperoleh nilai dari keempat parameter tersebut perlu dilakukan perhitungan terlebih dahulu berdasarkan data-data yang diperoleh dari variabel dependen dan variabel *explanatory*. Dalam literature istilah yang digunakan untuk menyebut  $\beta_1$  adalah *intercept* dan  $\beta_2, \beta_3$ , dan  $\beta_4$  adalah *slope coefficient*. Besarnya nilai *intercept* sama dengan  $Y_{it}$  ketika  $X_{it}$  sama dengan nol. Selain

itu, variabel dependent juga bisa dipengaruhi oleh variabel lain yang nilainya tidak diketahui positif atau negatif dan berada di luar sistem. Variabel tersebut secara teknis dinamakan sebagai dengan *stochastic disturbance* atau *stochastic error term* (Gujarati, 2004: 44). Pada persamaan tersebut, variabel semacam ini dinotasikan dengan  $\mu_{it}$ , yang tidak diketahui nilainya dan merupakan variabel acak yang membawa pengaruh positif atau negatif terhadap  $Y_{it}$ .

Berdasarkan persamaan 3.1 di atas dapat dikatakan bahwa terdapat dua komponen yang dapat mempengaruhi besarnya nilai  $Y_{it}$ . Pertama,  $\beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it}$ , yang berarti besarnya nilai  $Y_{it}$  ditentukan oleh  $X_{it}$ . Komponen pertama ini disebut sebagai komponen sistemik atau deterministik. Kedua,  $\mu_{it}$  yang merupakan variabel acak yang bersifat *disturbance* (*pengganggu*) dan nilainya pun tidak diketahui positif atau negatif. Selain itu variabel ini berada di luar sistem persamaan dan tidak dapat diobservasi.

Menurut Gujarati (2004: 640) terdapat beberapa model yang digunakan untuk memodel estimasi regresi data panel. Alat untuk mengestimasi tersebut didasarkan pada asumsi berdasarkan *intercept*, *slope coefficient*, dan *error term*. Sehingga diperoleh beberapa kemungkinan di antaranya adalah:

1. Diasumsikan bahwa *intercept* dan *slope coefficient* konstan antar *time series* dan *cross-section*, serta *error term* meliputi perbedaan baik dalam waktu *time series* dan *cross-section*.

2. Diasumsikan bahwa *slope coefficient* konstan, akan tetapi *intercept* berbeda untuk setiap *cross-section*.
3. Diasumsikan bahwa *slope coefficient* konstan, akan tetapi *intercept* berbeda untuk setiap *cross-section* antar waktu.
4. Diasumsikan bahwa semua koefisien baik *intercept* dan *slope coefficient* berbeda untuk setiap *cross-section*.
5. Diasumsikan bahwa semua koefisien baik *intercept* dan *slope coefficient* berbeda untuk setiap *cross-section* antar *time series*.

Untuk memodel estimasi regresi dengan menggunakan data panel dapat dilakukan dengan tiga pendekatan, antara lain:

**a. Model *Common Effects***

Metode paling sederhana yang digunakan untuk memodel estimasi regresi dengan menggunakan data panel adalah model *common effects*. Pada dasarnya model *common effects* sama dengan model estimasi *Ordinary Least Square* (OLS), yaitu estimasi yang dilakukan dengan menggunakan pendekatan kuadrat terkecil. Namun data yang digunakan bukan data *time series* atau *cross-section* saja, melainkan data panel yang diterapkan dalam bentuk *pooled* (kombinasi antara *cross-section* dan *time series*). Pada model estimasi regresi data panel ini, semua koefisien diasumsikan konstan, baik itu *intercept* ataupun *slope coefficient*-nya pada setiap unit *cross section* yang dijadikan sampel. Adapun persamaan regresi dalam model *common effects* dapat ditulis sebagai berikut:

[3.2]

$$Y_{it} = \beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it}$$

Dimana  $i$  menunjukkan *cross-section* (individu) dan  $t$  menunjukkan periode waktunya. Dengan asumsi komponen *error* dalam pengolahan kuadrat terkecil biasa, proses estimasi secara terpisah untuk setiap unit *cross-section* dapat dilakukan.

### **b. Model *Fixed Effects***

Model estimasi regresi data panel ini memiliki asumsi bahwa *intercept* berbeda dari setiap *cross-section* dan konstan dari setiap *time series*. Sedangkan *coefficient slope*-nya konstan dari setiap *cross-section* dan *time series*. Untuk menjelaskan asumsi tersebut kita dapat menuliskan model sebagai berikut:

[3.3]

$$Y_{it} = \beta_{1i} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it}$$

Untuk *intercept* ditambahkan dengan notasi  $i$  untuk menggambarkan bahwa nilai *intercept* dari setiap *cross-section* berbeda-beda. Perbedaan tersebut dapat mengacu pada faktor-faktor lain yang mempengaruhi besarnya nilai dari  $Y_{it}$  ketika variabel *explanatory* sama dengan nol. Sebagai contoh, di Indonesia penerimaan pajak bumi dan bangunan (PBB) bersifat otonom, yang besarnya tidak tergantung pada fluktuasi pendapatan. Namun hal ini dapat berbeda bagi negara lain. Dalam beberapa literatur model estimasi ini dikenal

sebagai model *fixed effects*. Istilah *fixed effects* mengacu pada fakta bahwa, meskipun *intercept* berbeda pada setiap *cross-section*, namun tidak bagi *time series*. Artinya *time series* dalam model ini bersifat *invariant*. Disisi lain, model *fixed effects* berasumsi bahwa *slope coefficient* tidaklah berbeda pada setiap *cross-section* atau *time series*. Bagaimana sebenarnya kita dapat membiarkan *intercept* berbeda-beda dari setiap *cross-section*? Kita dapat dengan mudah melakukan hal tersebut dengan menggunakan teknik variabel *dummy*. Penggunaan variabel *dummy* dalam mengmodel estimasi regresi data panel ini menyebabkan model ini sering disebut sebagai *Least Square Dummy Variable (LSDV)*. Dengan penggunaan variabel *dummy* dalam mengmodel estimasi regresi ini, kita dapat menuliskan persamaan regresi sebagai berikut:

[3.4]

$$Y_{it} = \alpha_1 + a_2 D_{2i} + a_3 D_{3i} + a_4 D_{4i} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it}$$

Dimana variabel *dummy* pada persamaan tersebut dinotasikan dengan *D* dan tambahan notasi *i* menggambarkan variasi nilai dari setiap *cross-section*. Nilai untuk variabel *dummy* berupa angka 0 dan 1. Angka 0 menggambarkan mengindikasikan apa yang tidak dimiliki dari suatu atribut. Sedangkan angka 1 mengindikasikan apa yang dimiliki dari suatu atribut.

### c. Model *Random Effect*

Keputusan untuk memasukan variabel *dummy* dalam model *fixed effects* memiliki konsekuensi berkurangnya *degree of freedom* yang akhirnya dapat mengurangi efisiensi dari parameter yang diestimasi. Oleh karena itu, dalam model data panel dikenal pendekatan yang ketiga yaitu model *random effects* (Baltagi, 2005). Model *random effects* disebut juga dengan model *error component*. karena di dalam model ini parameter yang berbeda antar unit *cross-section* maupun antar *timer series* dimasukkan ke dalam *error term*. Dengan menggunakan model *random effects*, maka dapat menghemat pemakaian derajat kebebasan dan tidak mengurangi jumlahnya seperti yang dilakukan oleh mode *fixed effects*. Hal ini berimplikasi pada parameter yang merupakan hasil estimasi akan menjadi semakin efisien dan model yang dihasilkan semakin baik. Untuk persamaan regresi dari model *random effects* dapat dimulai dari persamaan berikut:

[3.5]

$$Y_{it} = \beta_{1i} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it}$$

Dengan memperlakukan  $\beta_{1i}$  sebagai *fixed*, kemudian diasumsikan bahwa *intercept* memiliki nilai rata-rata sebesar  $\beta_1$ . Sedangkan nilai rata-rata dari setiap *cross-section* dapat dituliskan sebagai berikut:

[3.6]

$$\beta_{1i} = \beta_1 + \varepsilon_i \quad i = 1, 2, \dots, N$$

Dimana  $\varepsilon_i$  adalah *random error term* dengan nilai rata-rata sama dengan nol dan merupakan variasi dari  $\sigma_\varepsilon^2$ . Secara esensial, dapat dikatakan bahwa semua *cross-section* memiliki nilai rata-rata yang sama untuk *intercept*, yaitu sebesar  $\beta_1$ . Sedangkan perbedaan nilai *intercept* dari setiap unit *cross-section* direfleksikan dalam *error term*  $\varepsilon_i$ . Apabila persamaan 3.5 dan 3.6 disubstitusikan, maka akan diperoleh persamaan regresi sebagai berikut:

[3.7]

$$\begin{aligned} Y_{it} &= \beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \mu_{it} + \varepsilon_i \\ &= \beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + \beta_4 \cdot X_{4it} + \omega_{it} \\ \omega_{it} &= \mu_{it} + \varepsilon_i \end{aligned}$$

Berdasarkan persamaan di atas *error term* kini dinotasikan dengan  $\omega_{it}$ , yang terdiri dari dua komponen, yaitu  $\varepsilon_i$ , yang merupakan *cross-section error component*, artinya pada komponen  $\varepsilon_i$  ini terdapat perbedaan nilai *intercept* dari setiap unit *cross-section* yang direfleksikan oleh komponen  $\varepsilon_i$ . Sedangkan komponen  $\mu_{it}$  merupakan kombinasi antara *time series* dan *cross-section* dari *error component*, artinya terdapat perbedaan nilai *intercept* dari setiap unit *time series* dan *cross-section* yang direfleksikan oleh komponen  $\mu_{it}$ .

Perbedaan utama antara model *fixed effects* dan model *random effects* adalah pada perlakuan *intercept*. Pada model *fixed effects* setiap unit *cross-section* memiliki nilai *intercept* tersendiri yang *fixed*. Sedangkan pada model *random effects* setiap unit *cross-section* memiliki nilai *intercept* tersendiri

yang direfleksikan oleh *error term*  $\varepsilon_i$ . Sedangkan nilai *intercept* rata-rata dari seluruh *cross-section* direfleksikan oleh  $\beta_1$ .

Menurut Gujarati (2004: 650-651) dasar pemilihan antara model *fixed effects* dan *random effects* adalah sebagai berikut:

- f. Jika  $T$  (jumlah data *time series*) besar dan  $N$  (jumlah data dari *cross-section*) kecil, maka akan menunjukkan bahwa tidak ada perbedaan nilai parameter yang diestimasi oleh model *fixed effects* dan *random effects*. Pemilihan model terbaik dilakukan berdasarkan kemudahan penghitungan. Maka dalam hal ini, model *fixed effects* lebih baik daripada *random effects*.
- g. Ketika  $N$  besar dan  $T$  kecil, estimasi yang diperoleh dari kedua metode akan memiliki perbedaan yang signifikan. Jadi, apabila kita meyakini bahwa unit *cross-section* yang kita pilih dalam penelitian diambil secara acak maka model *random effects* harus digunakan. Sebaliknya, apabila kita meyakini bahwa unit *cross-section* yang kita pilih dalam penelitian tidak diambil secara acak maka kita harus menggunakan model *fixed effects*.
- h. Jika *error component*  $\varepsilon_i$  berkorelasi dengan variabel independen, maka parameter yang diperoleh dengan model *random effects* akan bias sementara parameter yang diperoleh dengan menggunakan model *fixed effects* tidak bias.

- i. Apabila  $N$  besar dan  $T$  kecil, dan apabila asumsi yang mendasari *random effects* dapat terpenuhi, maka model *random effects* akan lebih efisien dari model *fixed effects*.

Untuk memilih model mana yang paling tepat digunakan untuk pengolahan data panel, maka terdapat beberapa pengujian yang dapat dilakukan, antara lain:

**a. Chow Test**

*Chow Test* adalah pengujian untuk memilih apakah model yang digunakan model *common effects* atau *fixed effects*. Dalam pengujian ini dilakukan dengan hipotesis sebagai berikut (Baltagi, 2005) :

$H_0: a_1 = a_2 = \dots = a_n = 0$  (efek dari unit *cross-section* secara keseluruhan tidak berarti)

$H_a: a_i \neq 0; i = 1, 2, \dots, n$  (efek dari salah satu atau lebih unit *cross-section* berarti)

Statistik uji yang digunakan merupakan uji F, yaitu:

[3.8]

$$F - \text{statistik} = \frac{[RRSS - URSS]/(n - 1)}{URSS/(nT - n - k)}$$

Keterangan:

$n$  = Jumlah unit *cross-section*

$T$  = Jumlah periode waktu (*time series*)

$K$  = Jumlah variabel independen

$RSS$  = restricted residual sums of squares yang berasal dari model *common effects*

$URSS$  = unrestricted residual sums of squares yang berasal dari model *fixed effects*

Sedangkan F-tabel diperoleh dari:

[3.9]

$$F - \text{tabel} = \{ \alpha: df(n - 1, nt - n - k) \}$$

Keterangan:

$\alpha$  : Tingkat signifikansi yang dipakai (alfa)

$n$  : Jumlah unit *cross-section*

$nt$  : Jumlah unit *cross-section* dikali jumlah *time series*

$k$  : Jumlah variabel independen

Dasar penolakan terhadap hipotesis di atas adalah dengan membandingkan perhitungan F-statistik dan F-tabel. Apa bila hasil F-statistik lebih besar dari F-tabel, maka  $H_0$  ditolak, yang berarti model *fixed effects* yang paling baik untuk digunakan dalam mengmodel estimasi regresi data panel. Sebaliknya, apabila F-statistik lebih kecil dari F-tabel, maka  $H_0$  diterima, yang berarti

model *common effects* yang paling baik untuk digunakan dalam memodel estimasi regresi data panel (Wing Wahyu Winarno, 2009). Selain dengan membandingkan F-tabel dan F-statistik, dapat juga dilakukan dengan membandingkan antar nilai probabilitas dari F-statistik dan *alpha* (0,05). Apabila nilai probabilitas dari F-statistik  $> 0,05$ , maka  $H_0$  diterima yang artinya model *common effects* yang paling baik untuk digunakan. Jika sebaliknya, maka  $H_0$  ditolak yang artinya model *fixed effects* yang paling baik digunakan.

#### **b. Hausman Test**

*Hausman Test* adalah pengujian statistik sebagai dasar pertimbangan kita dalam memilih apakah menggunakan *model fixed effects* atau *random effects*. Uji ini bekerja dengan menguji apakah terdapat hubungan antara *error component* dengan satu atau lebih variabel independen dalam suatu model. Hipotesis awalnya adalah tidak terdapat hubungan antara *error component* dengan variabel independen. Menurut Baltagi hipotesis dari uji Hausman adalah sebagai berikut (Baltagi, 2005: 310) :

$H_0$ : Korelasi  $(X_{it}, \mu_{it}) = 0$  (efek *cross-section* tidak berhubungan dengan *error component*)

$H_a$ : Korelasi  $(X_{it}, \mu_{it}) \neq 0$  (efek *cross-section* berhubungan dengan *error component*)

Statistik uji yang digunakan adalah uji *chi square*. Jika nilai *chi square*-statistik  $>$  *chi square*-tabel ( $\alpha, k$ ) atau nilai *p-value* kurang dari taraf

signifikansi yang ditentukan, maka hipotesis awal ( $H_0$ ) ditolak sehingga model yang terpilih adalah model *fixed effects*. Hal ini dilakukan untuk mengetahui apakah terdapat efek random di dalam data panel (Dedi Rosadi, 2011: 264).

Dalam perhitungan uji Hausman diperlukan asumsi bahwa banyaknya kategori *cross-section* lebih besar dibandingkan jumlah variabel independen (termasuk konstanta) dalam model. Lebih lanjut, dalam estimasi uji Hausman diperlukan estimasi variansi *cross-section* yang positif, yang tidak selalu dapat dipenuhi oleh model. Apabila kondisi-kondisi ini tidak dapat dipenuhi, maka hanya dapat digunakan model *fixed effects*.

### c. *Lagrange Multiplier Test*

*Lagrange Multiplier Test* digunakan untuk menguji model apakah yang terbaik untuk digunakan dalam penelitian, yaitu untuk menguji model *common effects* dan model *random effects*. Hipotesis yang digunakan dalam uji Lagrange Multiplier adalah sebagai berikut:

$H_0$  = Model *Random Effects*

$H_a$  = Model *Common Effects*

Untuk dapat menentukan jawaban dari hipotesis di atas, maka diperlukanlah perhitungan LM-statistik nya. Perhitungan LM-statistik dapat dituliskan dengan rumus sebagai berikut:

[3.10]

$$\text{LM - statistik} = \frac{nT}{2(T-1)} \left[ \frac{T^2 \sum \bar{e}^2}{\sum e^2} - 1 \right]^2$$

Keterangan:

$n$  = jumlah *cross-section*

$T$  = jumlah *time-series*

$\sum \bar{e}^2$  = jumlah rata-rata kuadrat residual

$\sum e^2$  = jumlah residual kuadrat

Nilai LM-statistik akan dibandingkan dengan nilai *Chi Square*-tabel dengan derajat kebebasan (*degree of freedom*) sebanyak jumlah variabel independen (bebas) dan alpha atau tingkat signifikansi sebesar 5%. Apabila nilai LM-statistik > *Chi Square*-tabel, maka  $H_0$  di terima, yang artinya model yang dipilih adalah model *random effects*, jika sebaliknya maka  $H_0$  ditolak, yang artinya model yang dipilih adalah model *common effects*.

### 3. Uji Hipotesis

Uji hipotesis ini diperlukan untuk menguji apakah koefisien regresi yang didapat signifikan. Maksud dari signifikan di sini adalah suatu nilai koefisien regresi yang secara statistik tidak sama dengan nol. Jika *slope coefficient* sama dengan nol, berarti tidak dapat dikatakan bahwa terdapat cukup bukti untuk menyatakan variabel independen berpengaruh terhadap variabel dependen. Maka dari itu, semua koefisien yang terdapat pada persamaan regresi harus di

uji. Terdapat dua jenis uji hipotesis yang dapat dilakukan untuk menguji koefisien regresi, yaitu uji t dan uji F. Uji t digunakan untuk mengetahui secara parsial apakah variabel-variabel independen berpengaruh signifikan terhadap variabel dependen. Sedangkan uji F digunakan untuk secara simultan/keseluruhan apakah variabel-variabel independen berpengaruh signifikan terhadap variabel dependen.

### 1. Uji t (Parsial)

Uji parsial digunakan untuk mengetahui pengaruh masing-masing variabel independen terhadap variabel dependen secara parsial (Imam Ghazali, 2004: 98). Pengujian dapat dilakukan dengan menyusun hipotesis sebagai berikut:

Hipotesis statistik untuk *intercept*:

Hipotesis statistik untuk variabel Pendapatan per kapita:

$H_0: \beta_2 = 0$ , artinya secara parsial tidak ada pengaruh dari pendapatan per kapita terhadap penerimaan pajak.

$H_a: \beta_2 \neq 0$ , artinya secara parsial ada pengaruh dari pendapatan per kapita terhadap penerimaan pajak.

Hipotesis statistik untuk variabel kontribusi sektor industri:

$H_0: \beta_3 = 0$ , artinya secara parsial tidak ada pengaruh dari kontribusi sektor industri terhadap penerimaan pajak.

$H_a: \beta_3 \neq 0$ , artinya secara parsial ada pengaruh dari kontribusi sektor industri terhadap penerimaan pajak.

Hipotesis statistik untuk variabel inflasi:

$H_0: \beta_4 = 0$ , artinya secara parsial tidak ada pengaruh dari inflasi terhadap penerimaan pajak.

$H_a: \beta_4 \neq 0$ , artinya secara parsial ada pengaruh dari inflasi terhadap penerimaan pajak.

Hipotesis statistik untuk variabel tingkat keterbukaan ekonomi:

$H_0: \beta_4 = 0$ , artinya secara parsial tidak ada pengaruh dari tingkat keterbukaan ekonomi terhadap penerimaan pajak.

$H_a: \beta_4 \neq 0$ , artinya secara parsial ada pengaruh dari tingkat keterbukaan ekonomi terhadap penerimaan pajak.

Dasar pengambilan keputusan, apabila angka probabilitas signifikansi  $> 0.05$  maka  $H_0$  diterima dan  $H_a$  ditolak. Artinya variabel independen secara parsial tidak berpengaruh terhadap variabel dependen. Namun, apabila angka probabilitas signifikansi  $< 0,05$  maka  $H_0$  ditolak dan  $H_a$  diterima. Artinya variabel independen secara parsial berpengaruh terhadap variabel dependen. Dasar pengambilan keputusan juga dapat dilakukan dengan membandingkan nilai t-statistik dengan t-tabel.  $H_0$  diterima jika t-tabel  $>$  t-statistik dan ditolak jika t-tabel  $<$  t-statistik. Nilai t-statistik dapat dicari dengan menggunakan rumus sebagai berikut

[3.13]

$$t - \text{statistik} = \frac{\beta_i - \beta_0}{SE(\beta_i)}$$

$$i = 1, 2, 3, \dots, n = \text{parameter}$$

$0 =$  Hipotesis awal = nol

Keterangan:

$\beta_i$  = nilai parameter (*intercept* dan *slope coefficient*)

$\beta_0$  = Hipotesis awal yang diuji nilainya sama dengan nol

$SE$  = Standar eror setiap parameter (*intercept* dan *slope coefficient*)

## 2. Uji F (Uji Keandalan Parameter Secara Simultan)

Untuk menguji keberartian regresi dalam penelitian ini digunakan Uji statistik F dengan Tabel Anova. Uji statistik F pada umumnya menunjukkan apakah semua variabel independen atau bebas yang dimasukkan ke dalam model mempunyai pengaruh bersama-sama terhadap variabel dependen atau terikat. Pengujian dapat dilakukan dengan menyusun hipotesis terlebih dahulu sebagai berikut:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_a : \beta_2 \neq \beta_3 \neq \beta_4 \neq 0$$

Kriteria pengujian, apabila nilai signifikansi  $< 0,05$  maka  $H_0$  ditolak dan  $H_a$  diterima. Artinya semua variabel independe atau bebas secara simultan

berpengaruh terhadap variabel dependen atau terikat. Begitu juga sebaliknya, apabila nilai signifikansi  $>0,05$  maka  $H_0$  diterima dan  $H_a$  ditolak. Artinya semua variabel independen atau bebas secara simultan tidak berpengaruh terhadap variabel dependen atau terikat. Selain itu dapat digunakan kriteria lain pada pengujian keberartian regresi, yaitu apabila  $F\text{-tabel} > F\text{-statistik}$  maka  $H_0$  diterima dan apabila  $F\text{-tabel} < F\text{-statistik}$  maka  $H_0$  ditolak. Nilai dari  $F\text{-statistik}$  dapat dicari dengan menggunakan rumus sebagai berikut:

[3.14]

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

Keterangan:

$R^2$  = koefisien determinasi (residual)

$k$  = jumlah variabel independen ditambah intercept dari suatu model persamaan

$n$  = jumlah sampel

### 3. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi (*Goodness of fit*) dilakukan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen (Imam Ghazali, 2004: 97). Nilai  $R^2$  menunjukkan seberapa baik model yang dibuat mendekati fenomena dependen seharusnya. Rumus menghitungnya adalah dengan terlebih dahulu mencari nilai  $R$  atau koefisien korelasi:

[3.15]

$$R^2 = \frac{\beta_1 \sum X_1 Y + \beta_2 \sum X_2 Y + \beta_3 \sum X_3 Y}{\sum Y^2}$$

Nilai dari koefisien determinan adalah 0 sampai 1. Jika  $R^2 = 0$ , hal tersebut menunjukkan variasi dari variabel terikat tidak dapat diterangkan oleh variabel bebas. Namun jika  $R^2 = 1$ , maka variasi dari variabel terikat dapat dijelaskan oleh variabel bebas.

Kelemahan mendasar pada koefisien determinasi yaitu bias terhadap jumlah variabel independen yang masuk ke dalam model. Setiap penambahan satu variabel independen yang belum tentu berpengaruh signifikan atau tidak terhadap variabel dependen, maka nilai  $R^2$  pasti akan meningkat. Oleh sebab itu, digunakan nilai *adjusted*  $R^2$  yang dapat naik turun apabila ada penambahan variabel independen ke dalam model.