

BAB III

METODOLOGI PENELITIAN

A. Tujuan Penelitian

Tujuan penelitian ini adalah untuk mendapatkan pengetahuan yang tepat, valid, dan dapat dipercaya tentang:

1. Pengaruh pengeluaran pemerintah sektor pendidikan terhadap tingkat kemiskinan di Pulau Jawa.
2. Pengaruh pengeluaran pemerintah sektor kesehatan terhadap tingkat kemiskinan di Pulau Jawa.
3. Pengaruh pertumbuhan ekonomi terhadap tingkat kemiskinan di Pulau Jawa.
4. Pengaruh pengeluaran pemerintah sektor publik dan pertumbuhan ekonomi terhadap tingkat kemiskinan di Pulau Jawa.

B. Objek Penelitian

Penelitian ini dilaksanakan dengan mengambil data total pengeluaran pemerintah sektor publik, pertumbuhan ekonomi dan kemiskinan selama tahun 2004-2011 di provinsi-provinsi di Pulau Jawa, yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur dan Banten. Data tersebut diperoleh dari website Kementerian Keuangan bagian Direktorat Jendral Perimbangan Keuangan dan Badan Pusat Statistik (BPS).

Penelitian ini dilaksanakan pada bulan Mei 2012. Waktu ini dipilih dengan alasan pada periode ini merupakan periode terbaik peneliti untuk melakukan penelitian sehingga lebih fokus.

C. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini menggunakan regresi data panel untuk mengetahui seberapa besar pengaruh antara variabel-variabel yang diteliti yaitu tingkat kemiskinan sebagai variabel terikat (Dependen), pengeluaran pemerintah pada sektor publik yang diteliti dalam penelitian ini adalah pengeluaran pemerintah sektor pendidikan sebagai variabel bebas pertama (Independen) dan pengeluaran pemerintah sektor kesehatan sebagai variabel bebas kedua (Independen), dan pertumbuhan ekonomi sebagai variabel bebas ketiga (Independen).

D. Jenis dan Sumber Data

Dalam penelitian ini data yang digunakan berupa data sekunder yaitu data tahunan pengeluaran pemerintah sektor publik dan pertumbuhan ekonomi dan data tahunan kemiskinan. Jenis data yang digunakan adalah gabungan antara data runtut waktu *time series* dan data deret lintang *cross section* atau disebut juga dengan data panel. Data *time series* adalah data yang dikumpulkan dari waktu ke waktu terhadap suatu individu, sedangkan data *cross section* adalah data yang dikumpulkan dalam suatu waktu terhadap

banyak individu dalam suatu penelitian, dalam penelitian ini yang dimaksud adalah beberapa provinsi di Pulau Jawa.

Provinsi yang akan diteliti di Pulau Jawa dalam penelitian ini adalah DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, dan Banten yang dalam rentang waktu tahun 2004 sampai dengan tahun 2011. Dengan demikian data panel berjumlah 48 data.

Data sekunder adalah bentuk data yang sudah tersedia yaitu berupa data yang dipublikasikan oleh suatu badan atau instansi. Data sekunder tersebut diperoleh dari sumber-sumber seperti catatan atau laporan yang dipublikasikan oleh Kementrian Keuangan bagian Direktorat Jendral Perimbangan Keuangan (DJPk) dan Badan Pusat Statistik (BPS).

E. Operasionalisasi Variabel Penelitian

a. Tingkat Kemiskinan

1. Definisi Konseptual

Tingkat kemiskinan adalah sejumlah penduduk yang berada di bawah garis kemiskinan di mana mereka tidak dapat memenuhi kebutuhan dasar seperti makanan, pakaian, tempat berlindung, pendidikan dan kesehatan.

2. Definisi Operasional

Data kemiskinan dalam penelitian ini diukur dengan menggunakan data persentase penduduk miskin menurut BPS. Persentase penduduk

miskin diperoleh dengan cara membagi jumlah penduduk miskin dengan jumlah penduduk di suatu wilayah.

Data ini didapat oleh BPS melalui Survei Sosial Ekonomi Nasional (Susenas). Data persentase penduduk miskin yang digunakan berdasarkan data masing-masing provinsi di Pulau Jawa pada tahun 2004-2011.

b. Pengeluaran Pemerintah Sektor Pendidikan

1. Definisi Konseptual

Pengeluaran pemerintah sektor pendidikan adalah jumlah pengeluaran pemerintah daerah yang dialokasikan untuk pendidikan. Seperti program pendidikan Wajib Belajar Sembilan Tahun, dana BOS (Bantuan Operasional Sekolah) dan penyediaan sarana dan prasarana pendidikan yang bermutu.

2. Definisi Operasional

Data pengeluaran pemerintah sektor pendidikan merupakan besarnya pengeluaran pemerintah untuk pendidikan yang diperoleh dari laporan APBD (Anggaran Pendapatan dan Belanja Daerah) berdasarkan fungsi yang dipublikasikan oleh Kementrian Keuangan bagian Direktorat Jendral Perimbangan Keuangan (DJPk) dan dialokasikan untuk provinsi-provinsi di Pulau Jawa pada tahun 2004-2011.

c. Pengeluaran Pemerintah Sektor Kesehatan

1. Definisi Konseptual

Pengeluaran pemerintah sektor kesehatan adalah jumlah pengeluaran pemerintah daerah yang dialokasikan untuk kesehatan. Seperti pelayanan kesehatan, pencegahan penyakit dan pembangunan berbagai fasilitas kesehatan.

2. Definisi Operasional

Data pengeluaran pemerintah sektor kesehatan merupakan besarnya pengeluaran pemerintah untuk kesehatan yang diperoleh dari laporan APBD (Anggaran Pendapatan dan Belanja Daerah) berdasarkan fungsi yang dipublikasikan oleh Kementrian Keuangan bagian Direktorat Jendral Perimbangan Keuangan (DJPK) dan dialokasikan untuk provinsi-provinsi di Pulau Jawa pada tahun 2004-2011.

d. Pertumbuhan Ekonomi

1. Definisi Konseptual

Pertumbuhan ekonomi adalah suatu proses di mana terjadi kenaikan output total secara terus menerus suatu negara dalam jangka panjang dan dijadikan sebagai pengukur keberhasilan pembangunan ekonomi suatu negara.

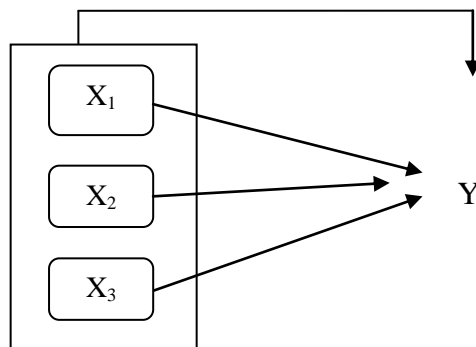
2. Definisi Operasional

Data pertumbuhan Ekonomi dalam penelitian ini diukur dengan menggunakan data laju pertumbuhan Produk Domestik Regional

Bruto (PDRB) atas harga konstan tahun 2000. Data ini diperoleh dari laporan yang dipublikasi oleh BPS berdasarkan data masing-masing provinsi di Pulau Jawa pada tahun 2004 sampai dengan 2011.

F. Konstelasi Pengaruh Antar 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 : Pengeluaran Pemerintah Sektor Pendidikan

X_2 : Pengeluaran Pemerintah Sektor Kesehatan

X_3 : Pertumbuhan Ekonomi

Y : Tingkat Kemiskinan

→ : Arah pengaruh

G. Teknik Analisis Data

1. Panel Data

Ada beberapa macam data yang digunakan dalam analisis ekonometrika, berupa *time series*, data *cross section*, atau data panel. Data panel 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 jumlah unit waktu sama untuk setiap individu, maka data ini disebut *balanced panel*. Namun jika terjadi sebaliknya, yakni jumlah unit waktu berbeda dengan untuk setiap individu, maka data ini disebut *unbalanced panel*. Penggunaan data panel pada dasarnya merupakan solusi karena ketidaktersediaan data *time series* yang cukup panjang untuk analisis ekonometrika.

Beberapa keunggulan menggunakan data panel menurut Baltagi, dibanding data runtut 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 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 meminimalkan bias yang mungkin dihasilkan dalam regresi.⁶⁶

Data panel merupakan gabungan dari data *cross section* dan data *time series*, dalam hal ini modelnya dapat dituliskan sebagai berikut:

$$Y_{it} = \alpha + \beta X_{1it} + \beta X_{2it} + \beta X_{3it} + \varepsilon_{it} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (3.1)$$

dimana :

N = banyaknya observasi

T = banyaknya waktu

⁶⁶ Jonni. J Manurung *et al.*, *Ekonometrika Teori dan Aplikasi* (Jakarta: PT. Elex Media Komputindo, 2005), p. 213-214

$N \times T$ = banyaknya data panel

Untuk mengestimasi parameter model dengan data panel, terdapat beberapa teknik yang ditawarkan, yaitu:

a. Model *Comon Effect*

Model *common effects* 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. Persamaan regresi untuk model *common effect* dapat dituliskan sebagai berikut:

$$Y_{it} = \alpha + \beta X_{it} + \beta X_{2it} + \beta X_{3it} + \varepsilon_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (3.2)$$

Di mana:

Y	= variabel dependen
α	= konstanta
X	= variabel independen
β	= koefisien regresi
ε_{it}	= <i>error term</i>
N	= jumlah (individu)
T	= jumlah periode waktu

Ekananda dalam Boy Azef menyatakan bahwa berdasarkan asumsi struktur matriks varians-kovarians residual, maka pada model *common effects*, terdapat 4 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 residualnya diasumsikan 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 *autoregressive (AR)* pada *error term*-nya, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada korelasi antar waktu pada residualnya.⁶⁷

b. Model Efek Tetap (*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 dalam Boy Azef menyatakan bahwa berdasarkan asumsi struktur matriks varians-kovarians residual, maka pada model *fixed effect*, terdapat 3 metode estimasi yang dapat digunakan, yaitu:

1. *Ordinary Least Square (OLS/LSDV)*, jika struktur matriks varians-kovarians residualnya diasumsikan bersifat homoskedastik dan tidak ada *cross sectional correlation*,

⁶⁷ Boy Azef, *Analisis Total Pengeluaran Wisatawan Mancanegara Berdasarkan Negara Tempat Tinggal Serta Variabel-Variabel Yang Mempengaruhinya*, Skripsi (Jakarta: STIS, 2011), p. 9

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-kovarians residualnya diasumsikan bersifat heteroskedastik dan ada *cross sectional correlation*.⁶⁸

c. Model *Random Effects*

Model berikutnya yaitu *random effect*, efek spesifik masing-masing individu α_i diperlakukan sebagai bagian dari *error* yang bersifat acak dan tidak berkorelasi dengan variabel penjelas yang teramati X_{it} .

Dalam model ini, parameter yang berbeda antar individu maupun antar waktu dimasukkan ke dalam *error*, karena hal inilah model ini sering juga disebut sebagai *error component model* (ECM). Bentuk model *random effects* dapat dijelaskan dengan persamaan dibawah ini sebagai berikut:

$$Y_{it} = \alpha + \beta X_{1it} + \beta X_{2it} + \beta X_{3it} + \varepsilon_{it} \quad \varepsilon_{it} = u_i + v_t + w_{it} \quad (3.3)$$

Di mana:

$$u_i \sim N(0, \delta U^2) = \text{error component cross section}$$

$$v_t \sim N(0, \delta V^2) = \text{error component time series}$$

$$w_{it} \sim N(0, \delta W^2) = \text{error component combinations}$$

Melihat persamaan di atas, maka dapat dinyatakan bahwa model *random effects* menganggap efek rata-rata dari data *cross-section* dan *time-series* direpresentasikan dalam *intercept*. Sedangkan deviasi efek

⁶⁸ *Ibid.*, p.10

secara random untuk data *time-series* direpresentasikan dalam v_t dan deviasi untuk data *cross-section* dinyatakan dalam u_i .

Telah diketahui bahwa: $\varepsilon_{it} = u_i + v_t + w_{it}$. Dengan demikian varians dari *error* tersebut dapat dituliskan dengan:

$$\text{Var}(\varepsilon_{it}) = \delta U^2 + \delta V^2 + \delta W^2 \quad (3.4)$$

Hal ini tentunya berbeda dengan model OLS yang diterapkan pada panel data, sebagaimana telah dijelaskan di atas, yang mempunyai varian *error* sebesar:

$$\text{Var}(\varepsilon_{it}) = \delta W^2 \quad (3.5)$$

Dengan demikian, *random effects* bisa diestimasi dengan OLS bila $\delta U^2 = \delta V^2 = 0$. Kalau tidak demikian, model *random effects* diestimasi dengan metode *Generalized Least Square* (GLS).

Asumsi yang digunakan dalam model *random effects* ini adalah *error* secara individual tidak saling berkorelasi, begitu pula dengan *error* kombinasinya. Penggunaan pendekatan *random effects* dapat menghemat derajat kebebasan dan tidak mengurangi jumlahnya seperti pada pendekatan *fixed effects*. Hal ini berimplikasi pada parameter hasil estimasi akan menjadi efisien. Semakin efisien maka model akan semakin baik.

Adapun model persamaan dalam penelitian ini adalah

$$Y_{it} = \alpha + \beta X_{1it} + \beta X_{2it} + \beta X_{3it} + \varepsilon_{it} \quad (3.6)$$

Keterangan:

Y_{it} = Tingkat Kemiskinan

α	= Konstanta
$X1_{it}$	= Pengeluaran pemerintah sektor pendidikan
$X2_{it}$	= Pengeluaran pemerintah sektor kesehatan
$X3_{it}$	= Pertumbuhan Ekonomi
β	= Parameter (Koefisien)
ε_{it}	= <i>error term</i>

Terkait dengan beberapa pilihan teknik untuk permodelan panel data, sebelum model diestimasi dengan model yang tepat, terlebih dahulu dilakukan uji spesifikasi apakah *common effect*, *fixed effect* atau *random effect* memberikan hasil yang sama. Penyeleksian model estimasi data panel antara lain:

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

Signifikansi 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*).

$$F = \frac{(RSS_1 - RSS_2) / (n-1)}{RSS_2 / (nT - n - k)} \quad (3.7)$$

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*. 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 null akan

ditolak, yang berarti koefisien intersep dan slope adalah sama tidak berlaku, sehingga teknik regresi data panel dengan *fixed effects* lebih baik dari *common effects*.

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

Untuk pengujian signifikansi model *random effect* dapat digunakan uji *Lagrange Multiplier* (LM) 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:

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n (\sum_{t=1}^T e_{it})^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right] \quad (3.8)$$

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

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \left[\text{var}(\hat{\beta}_{FE}) - \text{var}(\hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (3.9)$$

Di bawah hipotesis null, 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 null 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 null, 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 null akan ditolak, yang berarti estimasi yang tepat untuk regresi data panel adalah metode *fixed effect* dari pada metode *random effect*.

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, 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 *ai* 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.⁶⁹

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

⁶⁹ Damodar Gujarati, *Basic Econometrics, Fourth Edition* (New York: The McGraw-Hill Companies, 2004). p. 650

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 heteroskedastisitas dan autokorelasi 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*. Menurut Hidayat, estimasi *fixed effect* diasumsikan bebas dari autokorelasi, maka pengujian autokorelasi dapat diabaikan.⁷⁰

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

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

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

$$LM = \frac{T}{2} \sum_{i=1}^n \left[\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right]^2 \quad (3.10)$$

di mana T adalah jumlah periode waktu, n adalah jumlah individu, $\hat{\sigma}_i^2$ adalah varians residual persamaan ke- i pada kondisi homoskedastik, dan $\hat{\sigma}^2$ adalah *Sum Square Residual (SSR)* persamaan sistem 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 null

⁷⁰ Paidi Hidayat dan Wahyu Ario Pratomo. *Pedoman Praktis Eviews dalam Ekonometrika*. (Medan-USU Press, 2010). p. 20

ditolak, yang berarti struktur varians-kovarians residual bersifat heteroskedastik.

2. Uji Hipotesis

a. Uji Statistik t

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

Hipotesis pengujian:

$$H_0 : \beta_i \leq 0$$

$$H_1 : \beta_i > 0$$

Statistik uji yang digunakan adalah statistik uji t -student. Adapun formulanya adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \quad (3.11)$$

adalah nilai penduga parameter ke- i , adalah simpangan baku dari nilai penduga parameter ke- i .

Hipotesis null ditolak jika $t_{hitung} > t_{\alpha/2(nT-n-k-1)}$. Keputusan ini dapat juga didasarkan pada perbandingan nilai p -value dengan tingkat signifikansinya (α). Hipotesis null ditolak jika nilai p -value lebih kecil dari (α). Hal ini berarti secara parsial variabel bebas ke- i signifikan mempengaruhi variabel tidak bebasnya dengan tingkat kepercayaan sebesar $(1-\alpha) \times 100$ persen.

b. Uji Statistik F

Uji statistik F merupakan pengujian koefisien regresi secara keseluruhan. Pengujian ini menunjukkan apakah semua variabel bebas yang dimasukkan kedalam model mempunyai pengaruh secara bersama-sama terhadap variabel terikat.

Hipotesis pengujian yang digunakan adalah:

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

H_1 : paling sedikit salah satu nilai $\beta_i \neq 0$, dengan $i = 1, 2, \dots, k$.

Statistik uji F dihitung dengan formula sebagai berikut:

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

Hipotesis null ditolak jika $F_{hitung} > F_{\alpha; (n-k-1, nT-n-k)}$ yang berarti bahwa minimal ada satu variabel bebas yang signifikan berpengaruh terhadap variabel tidak bebasnya. Keputusan ini dapat juga didasarkan pada perbandingan nilai p -value dengan tingkat signifikansinya. Hipotesis null ditolak jika nilai p -value lebih kecil dari tingkat signifikansi (α).

c. Koefisien Determinasi (R^2)

R^2 digunakan untuk mengukur kebaikan atau kesesuaian suatu model persamaan regresi. Semakin besar R^2 maka semakin besar pula pengaruh antara variabel terikat dengan satu atau banyak variabel bebas. Besaran R^2 dihitung dengan rumus:

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

Sedangkan 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 (provinsi)

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.

3. 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 atau tidak. Hal tersebut didasarkan pada asumsi bahwa faktor kesalahan (residual) didistribusikan secara normal. Salah satu metode yang dapat digunakan untuk menguji normalitas adalah *Jarque-Bera test*. Uji statistik ini dapat dihitung dengan rumus berikut ini:

$$JB = \frac{N - k}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right] \quad (3.14)$$

di mana

N = jumlah sampel

S = skewness

K = kurtosis

k = banyaknya koefisien yang digunakan

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 null yang berarti tidak berdistribusi normal. Jika hasil *Jarque-Bera test* lebih kecil dari nilai *chi square* pada $\alpha=5$ persen, maka terima hipotesis null yang berarti *error term* berdistribusi normal.

b. 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 kolinieritas:

- Dengan melihat *Pearson Correlation Matrix* antar variabel independen, menurut Nash dan Bradford (2001) menyebutkan bahwa suatu variabel independen berkorelasi tinggi dengan variabel independen lainnya jika r lebih dari 0,85. Menurut Judge *et.al* dalam Gujarati, jika melebihi 0,8 maka ada kemungkinan terjadi kolinearitas yang serius.⁷¹
- 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:

$$VIF_k = \frac{1}{1 - R_k^2} \quad (3.15)$$

di mana $k = 1, 2, \dots, p - 1$ 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. Cara mengatasi adanya multikolinieritas antara lain melepas satu atau lebih variabel yang memiliki korelasi yang tinggi, mentransformasi model, atau memperbesar jumlah sampel.⁷²

⁷¹ Damodar Gujarati, *op. cit.*, p.359

⁷² *Ibid.*, p.363

c. Heteroskedastisitas

Heteroskedastisitas berarti adanya kesalahan residual yang diamati tidak memiliki varians yang konstan. Asumsi dalam model regresi adalah nilai residual memiliki (1) nilai rata-rata nol, (2) residual memiliki varian yang konstan serta (3) residual suatu observasi tidak saling berhubungan dengan residual lainnya sehingga menghasilkan estimator yang BLUE.

Apabila ketiga asumsi di atas dilanggar dan tidak menghasilkan estimator yang BLUE, 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 heteroskedastisitas, secara sistematis, statistik uji yang digunakan dapat dirumuskan sebagai berikut:

$$LM = \frac{T}{2} \sum_{i=1}^n \left[\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right]^2 \quad (3.16)$$

Hasil uji LM harus dibandingkan dengan nilai *chi-square* tabel dengan derajat (n-1) dengan ketentuan H_0 ditolak jika nilai LM lebih besar dari *chi-square* yang berarti model yang terbentuk mengandung masalah heteroskedastisitas.