

BAB III

METODE PENELITIAN

A. Objek atau Subjek Penelitian

1. Objek Penelitian

Penelitian ini menggunakan data upah, jumlah penduduk, pendidikan dan investasi di Provinsi Banten pada tahun 2010–2016. Dan penelitian ini dilakukan ada beberapa Kabupaten/Kota di Provinsi Banten, yaitu :

1. Kabupaten Pandenglang
2. Kabupaten Lebak
3. Kabupaten Tangerang
4. Kabupaten Serang
5. Kota Tangerang
6. Kota Cilegon
7. Kota Serang
8. Kota Tangerang Selatan

B. Jenis Data

Penelitian ini menggunakan analisis kuantitatif dan data sekunder berupa data *time series* dan *cross section* dalam bentuk data tahunan selama periode 2010 – 2016. Data dalam penelitian ini diperoleh dari Badan Pusat Statistik Provinsi Banten.

C. Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini dengan melakukan pencatatan secara langsung berupa data *time serie* dan *cross series* dari tahun 2010 sampai dengan tahun 2016 yang diperoleh secara langsung dari Badan Pusat Statistik Provinsi Banten.

D. Definisi Operasional Variabel Penelitian

2. Variabel Penelitian dan Definisi Operasional

Variabel adalah objek penelitian atau apa yang menjadi titik perhatian dalam suatu penelitian (Arikunto, 2010).

a. Variabel Dependen

Variabel dependen yang digunakan dalam penelitian ini adalah pengangguran terdidik, yaitu seseorang yang termasuk angkatan kerja tetapi belum mendapatkan pekerjaan dan sedang mencari pekerjaan dilihat dari tingkat pendidikan lulusan SLTA/Kejuruan, Diploma, dan Sarjana yang dihitung dalam satuan jiwa di Provinsi Banten pada tahun 2010-2016 dan data Badan Pusat Statistik Provinsi Banten.

b. Variabel Independen

Variabel independen yang digunakan dalam penelitian ini adalah :

1. Upah

Tingkat upah dilihat dari besarnya upah yang diterima oleh tenaga kerja selama satu bulan bekerja atas dasar upah minimum regional (UMR) yang dihitung dalam satuan rupiah di Provinsi Banten pada tahun 2010-2016 dan data yang diperoleh melalui Badan Pusat Statistik (BPS) Provinsi Banten.

2. Jumlah Penduduk

Tingkat jumlah penduduk dapat dilihat dari yang jumlah penduduk yang setiap tahun nya meningkat dari tahun ke tahun dalam satuan jiwa di Provinsi Banten pada tahun 2010-2016 dan data yang diperoleh melalui Badan Pusat Statistik Provinsi Banten.

3. Pendidikan

Pengangguran terdidik dipengaruhi salah satunya adalah pendidikan. Pengangguran terdidik dapat dihitung dari tingkat pendidikannya di Provinsi Banten pada tahun 2010–2016 dan data yang diperoleh melalui Badan Pusat Statistika (BPS) Provinsi Banten.

4. Investasi

Tingkat investasi dapat dihitung dari PDB masing masing daerah tersebut pada tahun 2010–2016 dan data yang diperoleh melalui Badan Pusat Statistik Provinsi Banten.

E. Uji Hipotesis dan Analisis Data

Dalam penelitian ini, penulis memilih metode regresi data panel yang digunakan untuk menganalisis data. Analisis regresi data panel digunakan untuk melihat sejauh mana pengaruh variabel–variabel bebas yang digunakan dalam meneliti Pengaruh upah, jumlah penduduk, pendidikan dan investasi terhadap Pengangguran Terdidik di Provinsi Banten tahun 2010–2016.

Data panel merupakan sebuah gabungan data runtut waktu (*time series*) dengan data silang (*cross section*). Widarjono (2009), mengatakan bahwa sebuah observasi yang menggunakan data panel mempunyai beberapa keuntungan. Pertama, data panel yang merupakan gabungan dua data *time series* dan *cross section* mampu menyediakan data lebih banyak sehingga akan lebih menghasilkan *degree of freedom* yang lebih besar. Kedua, penggabungan

data *time series* dan *cross section* dapat mengatasi masalah yang muncul ketika ada masalah pengurangan variabel (*omitted-variabel*) (Basuki dan Yuliadi, 2015).

Guharati (2004), mengatakan bahwa metode data panel merupakan suatu metode yang digunakan untuk melakukan analisis empiric dengan perilaku data yang lebih dinamis. Adapun kelebihan yang diperoleh dari penggunaan lebih banyak data, sehingga dapat memberikan informasi yang lebih lengkap. Kedua data panel mampu mengurangi kolinieritas variabel. Ketiga, dapat menguji dan membangun model perilaku yang lebih kompleks. Keempat, dengan menggabungkan informasi dari data *time series* dan *cross section* dapat mengatasi masalah yang timbul karena adanya masalah penghilangan variabel (*omitted variabel*). Kelima, data panel lebih mampu mendeteksi dan mengukur efek yang secara sederhana tidak mampu dilakukan oleh data *time series* murni maupun *cross section* murni. Keenam, data panel dapat meminimalkan bias yang dihasilkan oleh agregat individu, karena data yang diobservasi lebih banyak.

F. Model Regresi Data Panel

Dari beberapa variabel yang digunakan dalam penelitian ini maka dapat dibuat model penelitian sebagai berikut :

PT : f(upah, jumlah penduduk, pendidikan dan investasi).....[3.1]

$$PT_{it} = \alpha + \beta_1 Upah_{it} + \beta_2 Jp_{it} + \beta_3 Pendidikan_{it} + \beta_4 Investasi_{it} \dots\dots\dots [3.2]$$

Adanya perbedaan satuan dan besaran variabel bebas dalam persamaan menyebabkan persamaan regresi harus dibuat dengan model logaritma-linear (log). Sehingga model persamaan regresinya menjadi sebagai berikut:

$$\text{LogPT}_{it} = \beta_1 \text{Log}(\text{Upah})_{it} + \beta_2 \text{Log}(\text{Jp})_{it} + \beta_3 \text{Log}(\text{Pendidikan})_{it} + \beta_4 \text{Log}(\text{investasi})_{it} \dots \dots \dots [3.3]$$

Keterangan:

LogPT = Variabel dependen (Pengangguran Terdidik)

α = Kostanta

$\text{Log}\beta_{1234}$ = Koefisien variabel 1,2,3,4

Log Upah = Upah

Log Jp = Jumlah Penduduk

Log Pend = Pendidikan

Log Inves = Investasi

i = Kab/kota Provinsi Banten

t = 2010, 2011, 2012, 2013, 2014, 2015,2016

G. Metode Estimasi Model Regresi Panel

Dalam metode estimasi model regresi dengan menggunakan data panel dapat dilakukan melalui tiga pendekatan, antara lain :

a. Model *Polled Least Square (Common Effect)*

Model ini merupakan teknik regresi yang paling sederhana untuk mengestimasi parameter data panel dengan mengkombinasikan data time series dan cross section. Model dikatakan sama dengan metode *Ordinary Least Square (OLS)* karena hanya menggabungkan

data *time series* dan *cross section* tanpa melihat perbedaan antar waktu dan individu sehingga dapat karena menggunakan kuadrat terkecil. Adapun persamaan regresi dalam model *coomon effect* dapat ditulis sebagai berikut :

$$Y_{it} = \alpha + X_{it} + \epsilon_{it} \dots \dots \dots [3.4]$$

Dimana :

- i = Kab. Pandenglang, Kab. Lebak, Kab. Tangerang, Kab. Serang, Kota
Tangerang, Kota Cilegon, Kota Serang, Kota Tangerang Selatan
- t = 2010, 2011, 2012, 2013, 2014, 2015, 2016

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 Pendekatan Efek Tetap (*Fixed Effect*)

Pendekatan model *Fixed Effect* menggunakan variabel dummy untuk menangkap adanya perbedaan intersep antar individu dan mengasumsikan bahwa intersep dari setiap individu adalah berbeda sedangkan slope antar individu adalah tetap (sama). Pada metode *Fixed Effect* estimasi dapat dilakukan 2 cara, yaitu tanpa pembobot (*no weight*) atau *Least Square Dummy Variabel* (LSDV) dan dengan pembobot (*cross section weight*) atau *General Least Square* (GLS). Tujuan dilakukannya pembobotan adalah untuk mengurangi heterogenitas antar unit *cross section* (Gujarati, 200). Pemelihan model antara *Common Effect* dengan *Fixed Effect* dapat dilakukan dengan pengujian *Likelihood Test Radio* dengan ketentuan apabila nilai probabilitas yang

dihasilkan signifikan dengan alpha maka dapat diambil keputusan dengan menggunakan *Fixed Effect Model* (Gujarati, 2006).

c. Model Pendekatan Efek Acak (Random Effect)

Pendekatan model ini mengasumsikan bahwa parameter – parameter yang berbeda antar daerah maupun antar waktu dimasukan ke dalam error. Karena hal inilah, model efek acak juga disebut model komponen eror (*error component model*).

Dengan menggunakan model efek acak ini, maka dapat menghemat pemakaian derajat kebebasan dan tidak mengurangi jumlahnya seperti yang dilakukan pada model efek tetap. Hal ini berimplikasi parameter yang merupakan hasil estimasi akan jadi semakin efisien. Keputusan penggunaan model efek tetap ataupun acak ditentukan dengan menggunakan uji hausman. Dengan ketentuan apabila probabilitas yang dihasilkan signifikan dengan alpha maka dapat digunakan model *Fixed Effect* namun apabila sebaliknya maka dapat memilih salah satu yang terbaik antara *Fixed Effect* dengan *Random Effect*. Dengan demikian, persamaan model *Random Effect* yaitu :

$$Y_{it} = \alpha + X_{it}\beta + W_{it} \dots \dots \dots [3.5]$$

Keterangan :

i = Kab. Pandenglang, Kab. Lebak, Kab. Tangerang, Kab. Serang, Kota

Tangerang, Kota Cilegon, Kota Serang, Kota Tangerang Selatan

t = 2010, 2011, 2012, 2013, 2014, 2015, 2016

Dimana :

$$W_{it} = \epsilon_{it} + \mu_i; E(W_{it}) = 0; E(W_{it}^2) = \alpha^2 + \alpha_{\mu}^2; \dots \dots \dots [3.6]$$

$$E(W_{it}), (W_{jt-1}) = 0, i \neq j, E(\mu_i, \epsilon_{it}) = 0;$$

$$E(\epsilon_{i,}, \epsilon_{is}) = E(\epsilon_{it,}, \epsilon_{jt}) = E(\epsilon_{it,}, \epsilon_{is}) = 0$$

Meskipun komponen error W_t bersifat homoskedastik, nyatanya terdapat korelasi antara W_t dan W_{it-s} (*equicorrelation*), yakni :

$$\text{Corr}(W_{it}, W_{i(t-1)}) = \frac{\alpha_{\mu}^2}{\alpha^2 + \alpha_{\mu}^2} \dots \dots \dots [3.7]$$

Karena itu, metode OLS tidak bias digunakan untuk mendapatkan estimator yang efisien bagi model *Random Effects*. Model yang tepat untuk estimasi model *Random Effects* adalah *Generalized Least Square* (GLS) dengan asumsi homokedastik dan tidak ada *cross sectional correlation*.

H. Pemilihan Model

Untuk menganalisis pengaruh pertumbuhan ekonomi, upah, jumlah penduduk, inflasi terhadap pengangguran terdidik digunakan regresi data panel menggabungkan antara data *time series* dengan *cross section*. Prosedur regresi data panel tersebut adalah dengan memilih model yang paling tepat dengan cara:

1. Uji Chow Test

Uji Chow yaitu pengujian untuk menentukan model *Fixed Effect* atau *Random Effect* yang paling tepat digunakan dalam mengestimasi data panel. Hipotesis dalam uji chow adalah :

Ho : Menggunakan common effect model

H1 : Menggunakan fixed effect model

Dengan pengujian parameter jika $F\text{-tabel} > F\text{-hitung}$ maka H_0 diterima dan H_1 ditolak. Sebaliknya, jika $F\text{-hitung} > F\text{-tabel}$ maka H_0 ditolak dan H_1 diterima (Widarjono, 2009).

Perhitungan F-statistic didapat dari uji Chow dengan rumus (Baltagi, 2005) :

$$\frac{\frac{(SSE_1 - SSE_2)}{(n - 1)}}{\frac{SSE_2}{(nt - k)}} \dots\dots\dots [3.8]$$

Dimana :

SSE_1 = Sum Square Error dari model Common Effect

SSE_2 = Sum Square Error dari model Fixed Effect

n = Jumlah Kabupaten/Kota (cross section)

nt = Jumlah cross section x jumlah time series

k = Jumlah variabel independen

2. Uji Hausman

Hausman telah mengembangkan suatu uji untuk memilih apakah metode *Fixed Effect* dan metode *Random Effect* lebih baik dari metode *Common Effect*. Uji Spesifikasi Hausman membandingkan model *fixed effect* dan *random* dibawah hipotesis nol yang berarti bahwa efek individual tidak berkorelasi dengan regresi dalam model.

H_0 = Random Effect Model

H_1 = Fixed Effect Model

Hausman test ini menggunakan nilai chi-square sehingga keputusan pemilihan metode data panel ini dapat ditentukan secara statistik. Dengan asumsi bahwa error cara individual tidak saling berkorelasi begitu juga error kombinasinya.

Statistik hausman menggunakan nilai *Chi square* statistik. Jika hasil uji hausman test signifikan maka metode yang digunakan dalam pengolahan data panel adalah *Fixed Effect Model* (Widarjono, 2009).

Jika test hausman tidak menunjukkan perbedaan yang signifikan ($p > 0,05$) itu mencerminkan bahwa efek random estimator tidak aman bebas dari bias dan karenanya lebih dianjurkan kepada estimasi *fixed effect* daripada efek estimator tetap (Widarjono, 2009).

3. Uji Kualitas Data

Dengan pemakaian metode *Ordinary Least Squared* (OLS), untuk menghasilkan nilai parameter model penduga yang lebih tepat, maka diperlukan pendeteksian apakah model tersebut menyimpang dari asumsi klasik atau tidak, deteksi tersebut terdiri dari :

a. Uji Multikolinearitas

Multikolinearitas merupakan sebuah pengujian yang dilakukan pada regresi linear yang menggunakan lebih dari satu variabel bebas. Uji ini bertujuan untuk mengetahui apakah dalam regresi ini ditemukan adanya korelasi antar variabel independen. Jika terjadi korelasi maka terdapat problem multikolinieritas.

Menurut Chatterjee dan Price dalam Nachrowi (2002), adanya korelasi antar variabel – variabel bebas menjadikan interpretasi koefisien – koefisien regresi menjadi tidak benar lagi. Meskipun demikian, tidak berarti korelasi yang terjadi antara variabel–variabel bebas tidak diperbolehkan, hanya kolinieritas yang sempurna (*perfect collinierity*) saja yang tidak diperbolehkan, yaitu terjadi korelasi linear antara sesama variabel bebasnya. Sedangkan untuk sifat kolinier yang hampir sempurna (hubungan tidak bersifat linier atau korelasi mendekati nol) masih diperbolehkan atau tidak termasuk dalam pelanggaran asumsi.

Salah satu cara untuk mengetahui multikolinearitas dalam model yaitu dengan melihat koefisien korelasi hasil output komputer. Jika terdapat koefisien korelasi yang lebih besar dari 0,9; maka dapat dikatakan bahwa dalam model terdapat gejala multikolinearitas.

Untuk mengatasi masalah multikolinieritas, satu variabel independen yang memiliki korelasi dengan variabel independen lain harus dihapus. Dalam hal metode GLS, model ini sudah diantisipasi dari multikolinieritas.

b. Uji Heterokedastisitas

Uji heterokedastisitas digunakan untuk melihat apakah residual dari model yang terbentuk memiliki varians yang konstan atau tidak. Suatu model yang baik adalah model yang memiliki varians dari setiap gangguan atau residualnya konstan. Heterokedastisitas adalah keadaan dimana asumsi tersebut tidak tercapai, dengan kata lain dimana adalah eror dan adalah varians dari eror yang berbeda tiap periode waktu.

Suatu model regresi dikatakan terkena heterokedastisitas apabila terjadi ketidaksamaan varians dari residual dari suatu pengamatan ke pengamatan yang lain. Jika varians dari residual dari satu pengamatan ke pengamatan yang lain akan tetap, maka disebut heteroskedastisitas. Jika varians berbeda disebut heteroskedastisitas.

Adanya sifat heteroskedastisitas ini dapat membuat penaksiran dalam model bersifat tidak efisien. Umumnya masalah heteroskedastisitas lebih biasa terjadi pada data cross section dibandingkan dengan time series (Gujarati, 2006). Untuk mendeteksi masalah heteroskedastisitas dalam model, penulis menggunakan uji park yang sering digunakan dalam beberapa referensi. Dalam metodenya, Park menyarankan suatu bentuk fungsi spesifik diantara varian kesalahan σ_{ui}^2 dan variabel bebas yang dinyatakan sebagai berikut :

$$\sigma_{ui}^2 = \alpha X_i^\beta \dots\dots\dots[3.9]$$

Persamaan dijadikan linier dalam bentuk persamaan log sehingga menjadi :

$$\text{Ln}\sigma_{ui}^2 = \alpha + \beta \text{Ln } X_i + v_i \dots\dots\dots[4.0]$$

Karena varian kesalahan (σ_{ui}^2) tidak teramati, maka digunakan e_i^2 sebagai penggantinya.

Sehingga persamaan menjadi :

$$\text{Lne}_i^2 = \alpha + \beta \text{Ln } X_i + v_i \dots\dots\dots[4.1]$$

Apabila koefisien parameter β dari persamaan regresi tersebut signifikan secara statistik, berarti dalam data terdapat masalah heteroskedastisitas. Sebaiknya, jika β tidak signifikan, maka asumsi heteroskedastisitas pada data dapat diterima.

Model regresi yang baik adalah adanya heteroskedastisitas. Deteksi adanya heteroskedastisitas salah satunya dengan cara melihat hasil regresi data panel yang telah melakukan pengujian resid, kemudian dilihatnya hasil probabilitas. Jika nilai probabilitas dari masing-masing variabel tersebut $>0,05$ maka dapat dikatakan bahwa model tersebut lolos dari uji heteroskedastisitas.

4. Uji Statistik Analisis Regresi

Uji signifikan merupakan prosedur yang digunakan untuk menguji kesalahan atau kebenaran dari hasil hipotesis nol dari sampel.

a. Koefisien Determinasi (R^2)

Suatu model mempunyai kebaikan dan kelemahan jika diterapkan dalam masalah yang berbeda. Untuk mengukur kebaikan suatu model (*Goodnes Of Fit*) digunakan koefisien determinasi (R^2). Nilai koefisien determinasi merupakan suatu ukuran yang menunjukkan besar sumbangan dari variabel independen terhadap variabel dependen, atau dengan kata lain koefisien

determinasi menunjukkan variasi turunan Y yang diterangkan oleh pengaruh linier X. Menurut Gujarati (2003) dalam buku Ghozali (2009) menyimpulkan bahwa “Nilai koefisien determinasi antara 0 dan 1. Nilai koefisien determinasi yang mendekati 0 (nol) berarti kemampuan semua variabel independen dalam menjelaskan variabel dependen amat terbatas, sedangkan nilai koefisien determinasi yang mendekati 1(satu) berarti variabel-variabel independen hampir memberikan informasi yang dijelaskan untuk memprediksi variasi variabel dependen”.

b. Uji Parsial (t-Statistik)

Uji t-Statistik pada dasarnya untuk menunjukkan seberapa jauh pengaruh satu variabel bebas secara individu dalam menerangkan variasi variabel dependen dengan membandingkan t-hitung dengan t-tabel. Menurut Ghozali (2009) menyimpulkan pada tingkat $\alpha = 5\%$ maka kriteria pengujian yang digunakan sebagai berikut :

- Jika $t\text{-hitung} < t\text{-tabel}$ maka H_0 diterima dan H_1 ditolak, yang artinya salah satu variabel bebas (independen) tidak mempengaruhi variabel terikat (dependen) secara signifikan.
- Jika $t\text{-hitung} > t\text{-tabel}$ maka H_0 ditolak dan H_1 diterima, yang artinya salah satu variabel bebas (independen) mempengaruhi variabel terikat (dependen) secara signifikan.

c. Uji F-statistik

Uji F-statistik dilakukan untuk melihat seberapa besar pengaruh variabel independen secara keseluruhan atau bersama – sama terhadap variabel dependen. Untuk pengujian ini dilakukan hipotesa sebagai berikut (Ghozali, 2009) :

- $H_0 : \beta_1 = \beta_2 = 0$, artinya secara bersama-sama tidak ada pengaruh variabel independen terhadap variabel dependen.

- $H_A : \beta_1 \neq \beta_2 \neq 0$, artinya secara bersama-sama ada pengaruh variabel independen terhadap variabel dependen.

Pengujian ini dilakukan dengan membandingkan nilai F-hitung dengan F-tabel. Jika $F\text{-hitung} > F\text{-tabel}$ maka H_0 ditolak, yang berarti variabel independen secara bersama-sama mempengaruhi variabel dependen, jika sebaliknya maka H_0 diterima (Ghozali, 2009).