BAB III

MODEL REGRESI DATA PANEL

Pada bab ini akan dikemukakan dua pendekatan dari model regresi data panel, yaitu pendekatan *fixed effect* dan pendekatan *random effect* yang merupakan ide pokok dari tugas akhir ini. Selain itu, akan dibahas pula pengujian *Hausman* untuk mengetahui model terbaik yang dihasilkan oleh kedua pendekatan tersebut.

3.1 Model Regresi Data Panel

Model regresi data panel merupakan suatu model regresi yang observasi datanya didasarkan pada data panel. Secara umum, model regresi data panel dituliskan sebagai:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it} \qquad \ldots(3.1)$$

dengan y_{it} merupakan nilai variabel respon pada unit observasi ke-i dan waktu ke-t, x_{it} merupakan nilai variabel prediktor pada unit observasi ke-i dan waktu ke-t , α merupakan parameter *intersep* atau titik potong antara sumbu tegak Y dan garis fungsi linear nilai y_{it} , β merupakan koefisien *slope* atau koefisien arah atau koefisien kemiringan, dan ε_{it} merupakan kekeliruan atau galat atau komponen *error* pada unit observasi ke-i dan waktu ke-t.

Penaksiran untuk persamaan (3.1) dapat dilakukan melalui dua pendekatan, yaitu pendekatan *fixed effect* dan *random effect*. Berikut ini akan dikemukakan konsep-konsep dari kedua pendekatan tersebut.

3.1.1 Pendekatan *Fixed Effect*

Model regresi data panel yang menggunakan pendekatan *fixed effect* dinamakan *Fixed Effect Model (FEM)* yang juga sering disebut model *Least Square Dummy Variable (LSDV)*. FEM atau LSDV merupakan model yang mengasumsikan koefisien *slope* konstan tetapi *intersep* bervariasi antar anggota panel. Istilah "*fixed effect*" ini dikarenakan fakta bahwa meskipun *intersep* berbeda antar anggota panel, namun antar waktu tetap sama. Kasus seperti ini dinamakan *time invariant*.

Asumsi pada FEM ini jelas hampir sesuai dengan realita sebenarnya, karena karakteristik antar masing-masing anggota panel kemungkinan akan berbeda. Misalkan anggota panelnya adalah perusahaan, maka budaya perusahaan, gaya manajerial perusahaan, dan sebagainya akan berbeda antara satu perusahaan dengan perusahaan lainnya.

Pada asumsi ini, persamaan (3.1) dapat dituliskan sebagai berikut:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \qquad \ldots(3.2)$$

Harus diperhatikan bahwa pada persamaan (3.2) ditambahkan indeks i pada titik potong α. Indeks tersebut digunakan untuk menyatakan bahwa kasus ini merupakan kasus *time invariant*. Berbeda halnya bila indeks yang ditambahkan adalah it (α_{it}), maka kasusnya akan disebut dengan *time variant*.

Berikut ini merupakan FEM dengan kasus *time invariant*, yaitu:

$$y_{it} = \alpha D + \beta x_{it} + \varepsilon_{it} \qquad \ldots(3.3)$$

Perlu diperhatikan bahwa banyaknya *D* pada persamaan tersebut dapat dijelaskan dengan pernyataan berikut, "jika suatu variabel kualitatif mempunyai *n* kategori, maka hanya *n* -1 variabel boneka yang perlu diperkenalkan dalam model regresi, sedangkan satu variabel yang tidak diperkenalkan, rata-ratanya akan menjadi *intersep* atau titik potong dalam model".

Penaksiran parameter untuk model ini sangatlah sederhana, yakni dengan menggunakan metode *Ordinary Least Square (OLS)*. Pada bab ini tidak dibahas penaksiran parameter dengan metode OLS, karena penulis telah membahasnya pada bab sebelumnya.

3.1.2 Pendekatan *Random Effect*

Penambahan variabel boneka kedalam FEM bertujuan untuk memudahkan dalam penggunaan model tersebut. Hal ini dikarenakan variabel boneka dapat mewakili ketidaktahuan kita tentang model yang sebenarnya. Namun harus diingat bahwa penggunaan FEM akan membawa konsekuensi terhadap berkurangnya derajat kebebasan (*degree of freedom*) yang pada akhirnya akan mengurangi efisiensi parameter. Masalah inilah yang menjadi pendorong berkembangnya pendekatan selanjutnya, yaitu pendekatan *random effect*. Model regresi data panel yang menggunakan pendekatan ini dikenal dengan istilah *Random Effect Model (REM)*.

Ide dasar dari REM adalah menguraikan *intersep* pada persamaan (3.2), yaitu:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \qquad \ldots(3.4)$$

Dalam hal ini, α_i tidak lagi tetap (*fixed*). Sebagai gantinya, α_i diasumsikan sebagai variabel *random* dengan nilai rata-rata *(mean value)* α. Berikut ini adalah penjabaran *intersep* untuk masing-masing unit:

$$\alpha_i = \alpha + u_i \quad , i = 1, 2, 3, \ldots, N \qquad \ldots(3.5)$$

dimana u_i adalah komponen *error* acak dengan rata-rata nol dan varians σ_u^2.

Dengan mensubstitusikan persamaan (3.5) ke persamaan (3.4), akan diperoleh persamaan berikut:

$$y_{it} = \alpha + \beta x_{it} + u_i + \varepsilon_{it}$$

$$= \alpha + \beta x_{it} + w_{it} \qquad \ldots(3.6)$$

dimana

$$w_{it} = u_i + \varepsilon_{it} \qquad \ldots(3.7)$$

Komponen *error* w_{it} terdiri dari dua komponen, yaitu u_i yang merupakan komponen *error* masing-masing unit *cross section* dan ε_{it} yang merupakan kombinasi komponen *error time series* dan *cross section*. Karena terdiri dari dua (lebih) komponen *error*, maka REM juga dikenal dengan istilah *Error Components Model (ECM).*

Berikut ini beberapa asumsi yang berkaitan dengtan ECM, yakni:

$u_i \sim N(0, \sigma_u^2)$

$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$

$$E(u_i \varepsilon_{it}) = 0 \quad E(u_i u_j) = 0 \quad (i \neq j)$$

$$E(\varepsilon_{it}\varepsilon_{is}) = E(\varepsilon_{it}\varepsilon_{jt}) = E(\varepsilon_{it}\varepsilon_{js}) = 0 \quad (i \neq j; t \neq s)$$

Hal tersebut berarti bahwa komponen *error* tidak berkorelasi satu sama lain dan tidak ada autokorelasi antara unit *cross section* maupun unit *time series*.

Yang perlu diperhatikan, terdapat perbedaan penting antara FEM dan REM. Pada FEM, setiap unit *cross section* mempunyai nilai *intersep* tetap dari seluruh observasi N, sedangkan pada REM nilai *intersep* α menyatakan nilai rata-rata semua *intersep cross section* dan komponen *error* u_i menyatakan deviasi *intersep* unit *cross section* dari nilai rata-rata. Komponen *error* ini tidak dapat diamati secara langsung, sehingga dikenal dengan istilah *unobservable or latent variable*.

Penaksiran parameter untuk model ini tidak lagi menggunakan metode OLS karena metode ini tidak dapat menghasilkan penaksir yang efisien di bawah asumsi REM. Metode yang tepat untuk menaksir REM adalah *Generalized Least Square (GLS)*. Metode ini telah dijabarkan pada bab sebelumnya.

3.2 Uji Signifikansi *Fixed Effect Model* atau *Random Effect Model*

Setelah mendapatkan dua model yang signifikan melalui dua pendekatan, selanjutnya harus dipilih model mana yang paling sesuai untuk data yang dimiliki. Hal paling mendasar adalah dengan melihat korelasi antara komponen *error* spesifik *cross section* u_i dengan regresor atau variabel prediktor x. Jika diasumsikan u_i dan x tidak berkorelasi (*uncorrelated*), maka REM yang sesuai. Lain halnya jika u_i dan x berkorelasi, maka FEM lah yang paling sesuai.

Judge (Gujarati, 2003:650) mengemukakan beberapa pertimbangan pokok dalam memilih FEM dan REM, yaitu:

1. Jika jumlah data *time series* (T) besar dan jumlah unit *cross section* (N) kecil, maka nilai taksiran parameter antara FEM dan REM tidak berbeda secara signifikan. Akibatnya, pilihan didasarkan pada kemudahan perhitungan, yakni FEM.
2. Pada saat N besar dan T kecil, maka nilai taksiran parameternya berbeda secara signifikan. Perlu diperhatikan unit *cross section* pada sampel. Bila diyakini unit sampelnya tidak acak, maka FEM tepat untuk dipilih, namun bila unit sampelnya acak maka REM lebih tepat.
3. Jika komponen *error* u_i berkorelasi dengan satu atau lebih regresor, maka penaksir REM adalah penaksir yang bias dan penaksir FEM adalah penaksir yang tidak bias.
4. Jika N besar dan T kecil serta asumsi REM dipenuhi, maka penaksir REM lebih efisien dari pada penaksir FEM.

Pengujian formal untuk mengetahui ada atau tidaknya perbedaan nilai taksiran antar kedua model tersebut dikembangkan oleh Hausman. Pengujian ini selanjutnya dikenal dengan istilah *Hausman's Specification Test*, yang didasarkan pada ide bahwa model regresi dengan menggunakan penaksir OLS pada asumsi kedua dari pendekatan *fixed effect* dan model regresi dengan menggunakan penaksir GLS efisien, sedangkan model regresi dengan menggunakan penaksir

OLS tanpa variabel boneka pada asumsi "semua koefisien konstan antar waktu dan anggota panel" tidak efisien. Namun karena dalam tugas akhir ini tidak dibahas asumsi pertama dari pendekatan *fixed effect*, maka metode OLS dengan asumsi tersebut diabaikan. Berdasarkan hal tersebut, hipotesis nolnya (H_0) adalah nilai taksiran keduanya tidak berbeda sehingga pengujiannya dapat dilakukan berdasarkan perbedaan taksiran tersebut. Unsur penting untuk uji Hausman adalah matriks kovarian dari perbedaan vektor, $\left[\hat{\beta}-\hat{\beta}_{GLS}\right]$, yakni:

$$Var\left[\hat{\beta}-\hat{\beta}_{GLS}\right]=Var\left[\hat{\beta}\right]+Var\left[\hat{\beta}_{GLS}\right]-2Cov\left[\hat{\beta},\hat{\beta}_{GLS}\right] \qquad \ldots(3.8)$$

Hasil pokok dari uji Hausman adalah bahwa perbedaan kovarian dari penaksir yang efisien dengan penaksir yang tidak efisien adalah nol. Hal ini dapat dituliskan seperti berikut:

$$Cov\left[\left(\hat{\beta}-\hat{\beta}_{GLS}\right),\hat{\beta}_{GLS}\right]=Cov\left[\hat{\beta},\hat{\beta}_{GLS}\right]-Var\left[\hat{\beta}_{GLS}\right]=0 \qquad \ldots(3.9)$$

dengan kata lain,

$$Cov\left[\hat{\beta},\hat{\beta}_{GLS}\right]=Var\left[\hat{\beta}_{GLS}\right] \qquad ...(3.10)$$

Dengan mensubstitusikan persamaan (3.9) dan (3.10) ke dalam persamaan (3.8), maka akan diperoleh matriks kovarian berikut:

$$\begin{aligned}
Var\left[\hat{\beta}-\hat{\beta}_{GLS}\right]&=Var\left[\hat{\beta}\right]+Var\left[\hat{\beta}_{GLS}\right]-2Cov\left[\hat{\beta},\hat{\beta}_{GLS}\right]\\
&=Var\left[\hat{\beta}\right]+Var\left[\hat{\beta}_{GLS}\right]-2Var\left[\hat{\beta}_{GLS}\right]\\
&=Var\left[\hat{\beta}\right]-Var\left[\hat{\beta}_{GLS}\right]\\
&=Var\left[\hat{q}\right] \qquad \ldots(3.11)
\end{aligned}$$

Dari persamaan (3.11), dapat didefinisikan bahwa $\hat{q} = \left[\hat{\beta} - \hat{\beta}_{GLS}\right]$. Selanjutnya uji Hausman akan mengikuti distribusi Chi kuadrat dengan kriteria dari Wald, seperti berikut:

$$m = \hat{q}' Var\left(\hat{q}\right)^{-1} \hat{q} \qquad \ldots(3.12)$$

Statistik uji Hausman mengikuti distribusi Chi kuadrat dengan derajat kebebasan sebanyak *k* (jumlah variabel prediktor). Jika nilai statistik ujinya lebih besar dari nilai kritisnya, maka model yang tepat adalah model *fixed effect* (FEM), sedangkan jika nilai statistik ujinya lebih kecil dari nilai kritisnya, maka model yang tepat adalah model *random effect* (REM).

Pengujian hipotesis untuk uji Hausman tersebut dapat dituliskan secara matematis, seperti berikut:

- Perumusan Hipotesis

  H_0: Nilai taksiran parameter antara FEM dan REM tidak berbeda secara signifikan

  H_1: Nilai taksiran parameter antara FEM dan REM berbeda secara signifikan

- Besaran yang Diperlukan
  - $\hat{q} = \left[\hat{\beta} - \hat{\beta}_{GLS}\right]$
  - $Var\left[\hat{q}\right] = Var\left[\hat{\beta} - \hat{\beta}_{GLS}\right] = Var\left[\hat{\beta}\right] + Var\left[\hat{\beta}_{GLS}\right] - 2Cov\left[\hat{\beta}, \hat{\beta}_{GLS}\right]$

$$= Var\left[\hat{\beta}\right] + Var\left[\hat{\beta}_{GLS}\right] - 2Var\left[\hat{\beta}_{GLS}\right]$$

$$= Var\left[\hat{\beta}\right] - Var\left[\hat{\beta}_{GLS}\right]$$

- Statistik Uji

  $$m = \hat{q}' Var\left(\hat{q}\right)^{-1} \hat{q}$$

- Kriteria Pengujian

  Dengan taraf signifikansi α, tolak H_0 jika m < $\chi^2_{1-\alpha;k}$

- Kesimpulan

  Interpretasi dari ditolak atau diterimanya H_0.

Jika ternyata H_0 ditolak, artinya nilai taksiran kedua model berbeda secara signifikan. Selanjutnya penentuan model terbaik yang akan dipilih didasarkan pada kriteria Judge (Gujarati, 2003:650) seperti yang telah dibahas pada halaman 36.