

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Metode Pengumpulan Data

##### 3.1.1 Sumber Data

Data yang digunakan merupakan data pergerakan harga saham sebagai proksi untuk variabel dependen (*return* saham). Sedangkan untuk variabel independen yakni, variabel fundamental (*book-to-market ratio*, *debt-to-equity ratio*, dan *size*) dan variabel makroekonomi (inflasi, BI Rate, GDP, dan indeks sektoral), masing-masing berasal dari laporan tahunan yang dipublikasikan perusahaan pertambangan serta pemaparan statisrik moneter Indonesia yang dipublikasikan oleh Bank Indonesia dalam situsnya. Data dan informasi dikumpulkan dari data sekunder yang secara tidak langsung melalui berbagai sumber seperti situs [www.finance.yahoo.com](http://www.finance.yahoo.com), [www.idx.co.id](http://www.idx.co.id), [www.bi.go.id](http://www.bi.go.id), situs masing-masing perusahaan, situs-situs internet lainnya, ataupun dari artikel, majalah, dan jurnal.

##### 3.1.2 Obyek Penelitian

Penelitian ini melihat pengaruh variabel-variabel fundamental dan makroekonomi terhadap *return* saham perusahaan pertambangan yang terdaftar di Bursa Efek Indonesia dari tahun 2005-2008. Alasan digunakannya sektor pertambangan dalam penelitian ini karena berbagai penelitian dan analisis baik oleh akademis dan praktisi menemukan bahwa saham-saham sektor pertambangan memiliki volatilitas yang tinggi dalam pergerakan harga sahamnya. Ini menyebabkan saham pertambangan dapat menghasilkan *return* yang sangat tinggi, namun juga berisiko untuk mengalami kerugian besar seperti pada saat krisis keuangan global yang melanda perekonomian dunia.

### 3.1.3 Periode Penelitian

Penelitian ini akan mengambil periode tahun 2004-2008. Ini dikarenakan peneliti ingin melihat apakah hasil penelitian kali ini menghasilkan hasil yang berbeda dari sebelumnya karena adanya krisis keuangan global yang terjadi mulai akhir 2007 hingga 2008 yang terlihat dari menurunnya harga saham-saham pertambangan secara drastis.

### 3.1.4 Penentuan Sampel Penelitian

Pengambilan sampel pada penelitian ini dilakukan dengan teknik *purposive sampling*, yaitu pengambilan sampel berdasarkan kriteria tertentu. Adapun yang menjadi kriteria dalam pemilihan sampel adalah:

1. Perusahaan yang digunakan merupakan perusahaan-perusahaan yang bergerak di sektor pertambangan dengan subsektor batubara, minyak dan gas bumi, logam dan mineral lainnya, dan batu-batuan yang terdaftar di Bursa Efek Indonesia.
2. Perusahaan tidak pernah keluar dari bursa (*delisting*) selama periode penelitian kali ini
3. Data dan laporan keuangan lengkap dan dapat diakses
4. Ekuitas dan Pendapatan (*Earning*) positif. Hal ini dikarenakan variabel yang negatif tidak memiliki interpretasi yang berarti.

Berdasarkan kriteria di atas telah ditentukan sampel sejumlah 14 perusahaan, yakni:

**Tabel 3-1**  
**Emiten Sampel Perusahaan Pertambangan**

No.	Kode	Emiten	Listing
1	ANTM	Aneka Tambang (Persero) Tbk	27-Nov-97
2	APEX	Apexindo Pratama Duta Tbk	10-Jul-02
3	ATPK	ATPK Resource Tbk	17-Apr-02
4	BUMI	Bumi Resource Tbk	30-Jul-90
5	CITA	Cita Mineral Investindo Tbk	20-Mar-02
6	CNKO	Central Korporindo International Tbk	20-Nov-01
7	ENRG	Energi Mega Persada Tbk	7-Jun-04
8	INCO	International Nickel Indonesia Tbk	16-May-90
9	KKGI	Resource Alam Indonesia Tbk	1-Jul-91
10	MEDC	Medco Energi International Tbk	12-Oct-94
11	PGAS	Perusahaan Gas Negara Tbk	16-Jul-97
12	PTBA	Tambang Batubara Bukit Asam Tbk	23-Dec-02
13	PTRO	Petrosea	21-May-90
14	TINS	Timah Tbk	19-Oct-95

Jumlah sampel *cross section* pada penelitian ini adalah 14 perusahaan ( $n=14$ ), dan *time series* nya sebanyak 5 tahun ( $T=5$ , 2004 s/d 2008), sehingga total observasi dalam penelitian ini adalah  $nT=70$ .

### 3.1.5 Jenis Data

Jenis data dalam penelitian ini merupakan data sekunder berupa bukti, catatan atau laporan historis yang telah tersusun di dalam arsip (data dokumenter) yang dipublikasikan dan yang tidak dipublikasikan. Data sekunder yang digunakan dalam penelitian ini berasal dari sumber eksternal, yaitu laporan keuangan tahunan periode 2004-2008 dalam *database osiris*.

Untuk variabel fundamental, jenis data yang digunakan adalah data *cross section* yang dikarakteristikan dengan adanya beberapa observasi terhadap

beberapa unit tetap. Sedangkan untuk variable makroekonomi, data yang digunakan adalah data *time series* indikator-indikator makroekonomi bulanan yang dipublikasikan Bank Indonesia.

### 3.2 Hipotesa Penelitian

Hipotesa umum yang akan digunakan dalam penelitian ini mengacu pada penelitian sebelumnya yang dilakukan oleh Bhandari (1988) yang berjudul “*Debt/Equity Ratio and Expected Common Stock returns: Empirical Evidence*” dan Barber & Lyon (1997) yang berjudul “*Firm Size, Book-to-Market Ratio, and Security Returns*”. Penelitian ini berusaha melihat pengaruh variabel-variabel fundamental dan variabel-variabel makroekonomi saham pertambangan, sehingga secara umum hipotesa dalam penelitian ini adalah:

**Tabel 3-2**  
**Estimasi Arah Hubungan Variabel Independen**

Variabel	Estimasi Arah	Penjelasan
BMR	Positif (+)	BMR yang tinggi dapat menjadi resiko, karena nilai pasar yang terlalu rendah dibandingkan dengan nilai buku bisa mengindikasikan perusahaan tersebut bermasalah dan akan bangkrut. Pasar akan memandang perusahaan dengan BMR yang tinggi, lebih berisiko dibandingkan perusahaan dengan BMR yang rendah. Dengan demikian tingkat pengembalian perusahaan dengan BMR yang tinggi, juga akan lebih tinggi sesuai dengan risiko yang terdapat pada perusahaan tersebut.
DER	Positif (+)	Semakin tinggi DER menunjukkan tingginya ketergantungan permodalan perusahaan terhadap pihak luar sehingga beban perusahaan juga semakin berat. Tentunya hal ini akan mengurangi porsi hak pemegang saham (dalam bentuk deviden). Tingginya DER selanjutnya akan mempengaruhi minat investor terhadap saham perusahaan tertentu, karena investor pasti lebih tertarik pada saham yang tidak menanggung terlalu banyak beban hutang sehingga tingginya DER akan meningkatkan imbal hasil saham sebagai kompensasi atas risiko yang ditanggung investor
SIZE	Negatif (-)	<i>Size</i> merupakan pendekatan untuk ukuran perusahaan, sehingga semakin kecil suatu perusahaan seharusnya risiko semakin besar. Berkaitan dengan konsep ‘ <i>high risk high return</i> ’, karena risiko yang dimiliki oleh perusahaan kecil lebih besar maka imbal hasil yang dihasilkan seharusnya lebih besar dari perusahaan dengan ukuran yang lebih besar ( <i>small-firm effect</i> )
INF	Negatif (-)	Perusahaan-perusahaan pertambangan, yang terkena pengaruh inflasi

		akan mengalami pembengkakan dalam biaya-biaya dan nantinya akan menurunkan nilai sekarang ( <i>present value</i> ) perusahaan. Jika dihadapkan pada kondisi seperti ini tentunya akan semakin sedikit investor yang mau menanamkan modalnya di perusahaan-perusahaan
SBI	Negatif (-)	Suku bunga seperti SBI umumnya memiliki hubungan negatif atau terbalik dengan <i>return</i> saham. Hal ini diakibatkan kenaikan tingkat suku bunga akan mengurangi <i>present value</i> dari dividen saham karena naiknya biaya investasi yang nantinya akan mengurangi minat investor terhadap saham tersebut
PDB	Positif (+)	Peningkatan pada sektor pertambangan seharusnya memiliki hubungan positif dengan pergerakan dan imbal hasil saham. Ini dikarenakan prospek pertambangan yang baik dalam skala nasional akan berdampak baik juga pada perusahaan-perusahaan yang bergerak di sektor pertambangan. Di mana kondisi ini akan menggambarkan prospek ke depan dengan dilakukannya ekspansi usaha oleh perusahaan-perusahaan pertambangan

### 3.2.1 Hipotesa Variabel Fundamental

Untuk variabel fundamental:

1. Hubungan antara *Book-to-Market Ratio* dengan imbal hasil saham

$H_0$  : Tidak terdapat pengaruh yang signifikan dari *Book-to-Market Ratio* terhadap imbal hasil saham pertambangan

$H_1$  : Terdapat pengaruh yang signifikan dari *Book-to-Market Ratio* terhadap imbal hasil saham pertambangan.

2. Hubungan antara *Debt-to-Equity Ratio* dengan imbal hasil saham

$H_0$  : Tidak terdapat pengaruh yang signifikan dari *Debt-to-Equity Ratio* terhadap imbal hasil saham pertambangan

$H_1$  : Terdapat pengaruh yang signifikan dari *Debt-to-Equity Ratio* terhadap imbal hasil saham pertambangan

3. Hubungan antara *Size* dengan imbal hasil saham pertambangan

$H_0$  : Tidak terdapat pengaruh yang signifikan dari *Size* terhadap imbal hasil saham pertambangan

$H_1$  : Terdapat pengaruh yang signifikan dari *Size* terhadap imbal hasil saham pertambangan

### 3.2.2 Hipotesa Variabel Penelitian

Untuk variabel makroekonomi:

4. Hubungan antara laju inflasi dengan imbal hasil saham

$H_0$  : Tidak terdapat pengaruh yang signifikan dari laju inflasi terhadap imbal hasil saham pertambangan

$H_1$  : Terdapat pengaruh yang signifikan dari laju inflasi terhadap imbal hasil saham pertambangan

5. Hubungan antara SBI dengan imbal hasil saham

$H_0$  : Tidak terdapat pengaruh yang signifikan dari SBI terhadap imbal hasil saham pertambangan

$H_1$  : Terdapat pengaruh yang signifikan dari SBI terhadap imbal hasil saham pertambangan

6. Hubungan antara pertumbuhan PDB sektor pertambangan dengan imbal hasil saham

$H_0$  : Tidak terdapat pengaruh yang signifikan dari pertumbuhan sektor pertambangan terhadap imbal hasil saham

$H_1$  : Terdapat pengaruh yang signifikan dari pertumbuhan sektor pertambangan terhadap imbal hasil saham

3.3. Variabel dan Model Penelitian

3.3.1 Variabel Penelitian

Penelitian ini menggunakan dua model yang berbeda atas variabel fundamental dan variabel makroekonomi. Secara umum, Terdapat dua jenis variabel dalam penelitian ini yaitu variabel independen atau bebas dan variabel dependen atau terikat. Variabel terikat adalah variabel yang dipengaruhi oleh variabel bebas dan variabel bebas adalah variabel yang menjelaskan atau mempengaruhi variabel lain.

1. Variabel Dependen

Variabel dependen yang digunakan dalam kedua model penelitian kali ini adalah *return* saham dari masing-masing saham. Perhitungan *return* ini menggunakan perhitungan  *Holding Period Yield (HPY)*:

HPY = (Ending Value of Investment / Beginning Value of Investment) - 1 .....(3.1)

Dengan demikian perhitungan *return* saham dapat ditulis:

R\_it = (P\_it - P\_it-1) / P\_it-1 .....(3.2)

Di mana:

- R\_it = imbal hasil saham i pada periode t
P\_it = harga saham i pada periode t
P\_it-1 = harga saham i pada periode t-1

2. Variabel Independen (faktor fundamental)

Perhitungan variabel bebas mengikuti model Mukhreji, Dhatts, dan Kim(1997) dimana seluruh variabel independen dijadikan bentuk logaritma natural (ln). Penggunaan logaritma natural bisa menangkap efek

*leverage* dan mengarah kepada intepretasi yang mudah dari hubungan antara peran *leverage* dan *book to market equity* pada *average return*.

**a. Book-to-Market Value of Equity Ratio (BMR)**

Merupakan perbandingan antara nilai buku dari saham yang terdapat dalam laporan keuangan dan harga pasar saham yang terdapat di Bursa Efek Indonesia.

$$\text{BMR} = \text{Book Value of Equity} / \text{Market Value of Equity} \dots\dots(3.3)$$

**b. Debt-to-Equity Ratio (DER)**

Merupakan perbandingan antara nilai buku dari hutang dan nilai buku dari saham.

$$\text{DER} = \text{Total Debt} / \text{Total Equity} \dots\dots\dots(3.4)$$

**c. Ukuran Perusahaan (Size)**

Merupakan kapitalisasi pasar saham dalam BEI yang didapatkan dari jumlah saham yang beredar di pasar dikalikan dengan harga saham yang berlaku.

$$\text{SIZE} = \text{Outstanding Share} \times \text{Price per Share} \dots\dots\dots(3.5)$$

**d. Variabel Makroekonomi Inflasi**

Merupakan tingkat kenaikan harga-harga secara umum dalam perekonomian Indonesia yang diukur berdasarkan Indeks Harga konsumen (IHK) dan dipublikasikan secara periodic oleh Bank Indonesia.

**e. Variabel Tingkat Suku Bunga SBI**

Merupakan surat berharga yang dikeluarkan oleh Bank Indonesia sebagai pengakuan utang berjangka waktu pendek (1-3 bulan) dengan sistem diskonto/bunga. Tingkat suku bunga yang berlaku pada setiap penjualan SBI ditentukan oleh mekanisme pasar berdasarkan sistem



lelang. Sejak awal Juli 2005, BI menggunakan mekanisme "BI rate" (suku bunga BI), yaitu BI mengumumkan target suku bunga SBI yang diinginkan BI untuk pelelangan pada masa periode tertentu.

#### f. Variabel Pertumbuhan PDB Sektor Pertambangan

Pertumbuhan PDB Pertambangan didapat dari perhitungan PDB riil sektor pertambangan Indonesia yang dipublikasikan oleh Bank Indonesia dan Biro Pusat Statistik.

### 3.3.2 Model Penelitian

Penelitian ini menggunakan dua model. Model pertama menggunakan model data panel untuk mengetahui pengaruh variabel-variabel fundamental dan makroekonomi terhadap *return* saham pertambangan. Estimasi untuk model pertama tersebut adalah seperti yang telah disebutkan di bab satu:

$$r_{it} = \alpha_{it} + \beta_1 BMR_{it} + \beta_2 DER_{it} + \beta_3 SIZE_{it} + \beta_4 INF_t + \beta_5 SBI_t + \beta_6 PDB_t + \varepsilon_{it} \quad (3.6)$$

Di mana:

$r_{it}$  = imbal hasil saham perusahaan  $i$  pada periode  $t$

$\alpha_{it}$  = *intercept*

$\beta_1 - \beta_6$  = koefisien regresi

$BMR_{it}$  = *book-to-market value of equity ratio* perusahaan  $i$  pada periode  $t$

$DER_{it}$  = *debt-to-equity ratio* perusahaan  $i$  pada periode  $t$

$SIZE_{it}$  = *size* perusahaan  $i$  pada periode  $t$

$INF$  = laju inflasi Indonesia

$SBI$  = Tingkat Suku Bunga SBI

$PDB$  = Pertumbuhan Produk Domestik Bruto Sektor Pertambangan

$\varepsilon_{it}$  = *error term*

### 3.4 Tahap Pengolahan Data

Penelitian ini menggunakan dua model. Model pertama yang melihat pengaruh variabel fundamental terhadap *return* saham pertambangan akan menggunakan data panel. Menurut Gujarati (2003) Data panel sendiri merupakan gabungan antara data time series dan cross-section dengan jumlah unit data yang sama. Sebagai hasilnya data panel memungkinkan banyaknya observasi dalam setiap individual dalam sampel tersebut. Data panel sangat berguna karena memungkinkan peneliti untuk menghasilkan efek ekonomi yang tidak didapat dengan hanya menggunakan *cross section* maupun *time series*.

Baltagi (1998) dalam bukunya menjelaskan beberapa keuntungan dari penggunaan data panel antara lain:

- a. Data panel berkaitan dengan individual, perusahaan, negara, dan sebagainya dalam periode tertentu, maka tidak ada batasan terhadap heterogenitas dari unit ini.
- b. Dengan mengkombinasikan data *time series* dan *cross section*, data panel memberikan data yang lebih informatif, lebih bervariasi, sedikit kolinearitas antar variabel, lebih banyak derajat kebebasan (*degree of freedom*) dan lebih efisien.
- c. Dengan mempelajari penelitian *cross section* yang repetitif, data panel merupakan pendekatan yang lebih baik untuk mempelajari dinamika perubahan.
- d. Data panel bisa mendeteksi lebih baik dan mengukur efek yang tidak bisa diobservasi dengan data *cross section* atau data *time series* saja.
- e. Data panel membuat kita mempelajari model *behavioral* yang lebih kompleks.
- f. Data panel dapat meminimalkan bias yang dihasilkan oleh agregasi individu atau perusahaan karena unit data lebih banyak

Walaupun demikian, menurut Baltagi, data panel bukan merupakan metode yang sempurna dan tetap mempunyai beberapa kelemahan. Diantaranya adalah masalah desain dan pengkoleksian data, kemungkinan terjadinya distorsi dan kesalahan pengukuran, masalah *selectivity*, dimensi seri waktu yang lebih pendek, dan dependensi *cross section*.

Gujarati (2003) menyebutkan data panel sebagai *pooled data* (pengelompokan penelitian *time series* dan *cross-section*), data mikropanel, data longitudinal, analisis kejadian historis, atau analisis kohort. Secara umum, data panel dapat diestimasi dengan persamaan:

$$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + e_{it} \dots \dots \dots (3.7)$$

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

**Di mana:**

$Y_{it}$  : variabel dependen untuk setiap unit individu  $i$  pada periode  $t$

$\beta_1$  : *intercept* persamaan regresi

$\beta_j$  : koefisien persamaan regresi untuk setiap variabel independen  $j$

$X_{jit}$  : variabel independen untuk setiap unit individu  $i$  pada periode  $t$

$e_{it}$  : residual persamaan regresi (*error term*)

Pada persamaan diatas,  $i$  merepresentasikan komponen *cross section* dan  $t$  merepresentasikan komponen *time series*. Dengan kata lain,  $i$  menunjukkan unit-unit observasi pada setiap periode, sedangkan  $t$  menunjukkan periode atau waktu penelitian. Data panel memberikan kemungkinan untuk meneliti perubahan pada kelompok observasi dari waktu ke waktu, sehingga metode ini relevan untuk menguji apakah variabel-variabel fundamental seperti, *book-to-market ratio*, struktur modal, dan *size* dari perusahaan dapat mempengaruhi *return* saham

perusahaan-perusahaan tersebut, dimana dibutuhkan banyak unit observasi dalam lebih dari satu periode untuk mengetahui bagaimana hubungan tersebut.

### 3.4.1 Pendekatan Data Panel

Pengolahan data panel dapat dilakukan dengan tiga pendekatan, yaitu:

#### 1. *Pooled Least Square*/Pendekatan kuadrat terkecil

Metode *Pooled Least Square* merupakan pendekatan yang paling sederhana dalam mengestimasi data panel. Pada pendekatan ini tidak memperhatikan dimensi waktu dan individual, dan pendekatan ini mengasumsikan perilaku data antar perusahaan sama dalam berbagai kurun waktu. Sesuai dengan namanya yaitu *pooled* yang berarti dalam metode ini digunakan data panel dan *least square* yang berarti metode ini meminimumkan jumlah eror kuadrat.

Meminumkan eror kuadrat kemungkinan besar jika dijumlahkan akan bernilai nol dan jika eror hanya dijumlahkan saja tanpa dikuadratkan maka akan terjadi “ketidakadilan” karena nilai eror yang besar dan kecil disamaratakan. Dengan adanya asumsi tersebut maka regresi data panel yang akan dihasilkan sebuah perusahaan akan berlaku untuk semua perusahaan.

Persamaan untuk *Pooled Least Square* dapat ditulis dalam bentuk persamaan berikut :

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \dots \dots \dots (3.8)$$

Di mana:

$i$  = unit *cross section* (individu)

$t$  = periode waktunya.

Ini akan mengakibatkan hasil dimana terdapat T persamaan yang sama (individu sama, waktu berbeda) dan terdapat N persamaan yang sama untuk setiap T observasi (periode waktu sama, individu berbeda). Ini diakibatkan oleh metode *pooled least square* memiliki asumsi bahwa baik *intercept* dan *slope* dari persamaan regresi dianggap konstan baik untuk antar individu dan antar waktu.

## 2. *Fixed Effect*/Pendekatan efek tetap

Salah satu cara untuk memasukkan unsur individualitas dari masing-masing unit *cross-section* adalah dengan membiarkan *intercept* berbeda untuk setiap individu tetapi dengan tetap berasumsi bahwa koefisien *slope* konstan. Sementara itu, cara untuk membuat *intercept* berbeda bagi setiap individu adalah dengan menggunakan variabel *dummy* (Gujarati, 2003). Pendekatan ini menggunakan variabel boneka (*dummy*), sehingga disebut juga *Least Square Dummy Variabel Model* atau disebut juga *Covariance Model*. Dengan demikian, kita dapat memodelkan konstanta yang berbeda tersebut dengan menggunakan variabel *dummy* untuk setiap individu. Misalkan, sebuah penelitian memiliki sepuluh individu, maka model *fixed effect* dapat dituliskan:

$$y_{it} = \beta_{11}D_{1i} + \beta_{12}D_{2i} + \dots + \beta_{1,10}D_{10i} + \beta_2X_{2i} + \beta_3X_{3i} + e_{it} \dots\dots\dots(3.9)$$

$$D_{1i} \begin{cases} 1 \text{ jika } i = 1 \\ 0 \text{ selain } i = 1 \end{cases} \quad D_{2i} \begin{cases} 1 \text{ jika } i = 2 \\ 0 \text{ selain } i = 2 \end{cases} \quad D_{3i} \begin{cases} 1 \text{ jika } i = 3 \\ 0 \text{ selain } i = 3 \end{cases}$$

Pada persamaan diatas ditambahkan sebanyak  $(N-1) + (T-1)$  variabel boneka ke dalam model dan menghilangkan dua sisanya untuk menghindari kolinearitas sempurna antar variabel penjelas. Dengan menggunakan pendekatan ini akan terjadi *degree of freedom* sebesar  $NT - 2 - (N-1) - (T-1)$ , atau sebesar  $NT - N - T$ .

Kelemahan dari model *fixed effects* adalah terkadang variabel *dummy* yang ditambahkan tersebut tidak memiliki informasi penuh dalam menjelaskan model aslinya

## 3. *Random Effect*/Pendekatan efek acak

Dalam model efek acak, parameter-parameter yang berbeda antar daerah maupun antar waktu dimasukkan ke dalam *error*. Karena hal ini lah, model efek acak sering juga disebut model komponen *error* (*error component model/ECM*). Pada model ini gangguan diasumsikan bersifat acak untuk seluruh populasi. Ide dasar dari *random effect* adalah:

$$y_{it} = \bar{\beta}_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + \varepsilon_{it} \dots \dots \dots (3.10)$$

Dengan *error term*.

$$\varepsilon_{it} = \mu_i + v_t + w_{it} \dots \dots \dots (3.11)$$

Di mana:

$u_i \sim N(0, \delta_u^2)$  = komponen *cross section error*

$v_t \sim N(0, \delta_v^2)$  = komponen *time series error*

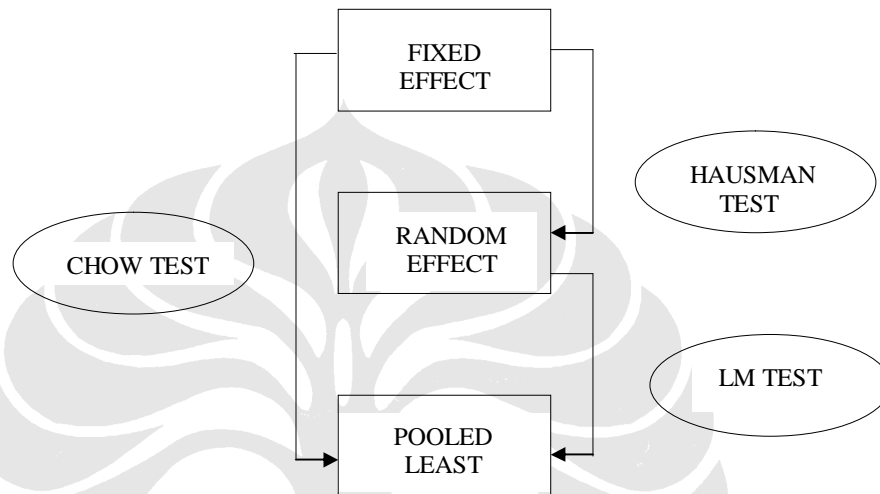
$w_{it} \sim N(0, \delta_w^2)$  = komponen *error kombinasi*

*Error* secara individual dan eror secara kombinasi diasumsikan tidak berkorelasi. Berbeda dengan *fixed effect*, model ini mengasumsikan bahwa sampel diambil secara acak pada setiap periode, sehingga diasumsikan  $\mu_i$  (dan  $V_t$  pada kasus dipertimbangkannya heterogenitas antar waktu) mengikuti distribusi normal. Model ini memberikan keuntungan dari segi efisiensi jumlah variabel dibandingkan dengan *fixed effect*, sehingga dapat meningkatkan efisiensi model. Kelebihannya dibandingkan dengan *pooled least square*, model ini dapat menangkap heterogenitas antar individu yang tercermin dari residual *intercept* untuk masing-masing individu. Tetapi dalam *random effects* juga terdapat kelemahan, kelemahan dari teknik *random effects* ini adalah adanya korelasi antara *error term* dengan variabel *independent*.

### 3.4.2 Pengujian Pemilihan Pendekatan Data Panel

Dalam bukunya, Baltagi (2003) menyebutkan bahwa untuk menentukan model terbaik, pengujian statistik harus dilakukan. Skema pengujian statistik tersebut adalah sebagai berikut:

**Gambar 3-1**  
**Skema Pengujian Pemilihan Metode Data Panel**



Sumber : Syahril, Syarif. Pengolahan Data Panel. Disampaikan sebagai bahan pelatihan Ekonometrika LABKOM Departemen Ilmu Ekonomi Fakultas Ekonomi Universitas Indonesia

### 1. Chow Test

Untuk mengetahui mana yang paling baik antara *intercept* dan *slope* yang harus konstan (metode *pooled least square*) atau *slopenya* saja yang konstan (metode *fixed effects model*) maka digunakanlah *Chow test*. Pengujian ini seringkali disebut juga dengan pengujian *F-Statistic*. Dasar penolakan terhadap hipotesa nol tersebut adalah dengan menggunakan F-Statistik seperti yang dirumuskan oleh *Chow*: (Baltagi, 2005)

$$F_{stat} = \frac{(R_{fe}^2 - R_{re}^2)/(N-1)}{(1 - R_{fe}^2)/(NT - N - K)} \dots\dots\dots(3.12)$$

Di mana:

$R_{fe}^2$  : *Retricted Residual Sum Square*

$R_{re}^2$  : *Unrestricted Residual Sum Square*

N : jumlah data *cross section*

T : jumlah data *time series*

K : jumlah parameter

Metode ini menggunakan *F Test* untuk melihat signifikasni parameter regresi secara keseluruhan, dengan demikian hipotesa yang diuji adalah:

$H_0$  : parameter-parameter *dummy variable* tidak signifikan dalam menjelaskan variabel dependen (  $\beta_2 = \beta_3 = K = \beta_n = 0$  )

$H_1$  : parameter-parameter *dummy variable* signifikan dalam menjelaskan variabel dependen (  $\beta_2 \neq 0, \beta_3 \neq 0, K, \beta_n \neq 0$  )

Jika satu atau lebih parameter *dummy variable* tidak signifikan, maka dapat dinyatakan bahwa tidak terdapat *intercept* yang tetap untuk masing-masing individu dan estimasi *fixed effect* menjadi tidak efisien. Sehingga ketika  $H_0$  tidak ditolak, model yang digunakan adalah *pooled least square* atau *random effect*. Sebaliknya, ketika  $H_0$  ditolak (parameter-parameter *dummy variable* signifikan), model yang digunakan adalah *fixed effeect*.  $H_0$  ditolak jika nilai *chow test* lebih besar daripada nilai *F* statistik dengan *degree of freedom* sebanyak  $N-1$  dan  $NT-N-K$ , atau jika nilai probabilitas lebih kecil daripada 0.05.

## 2. *Haussman Test*

Tes ini digunakan untuk melihat adanya korelasi antara residual pada *random effect* ( $\mu_i$ ) dengan masing-masing variabel independen. Jika terdapat korelasi



antara residual *cross section* dengan salah satu variabel independen, maka hasil estimasi *random effect* tidak lagi konsisten. Koefisien pada *random effect* konsisten dan efisien jika tidak terdapat korelasi, namun tidak konsisten jika terdapat korelasi. Di sisi lain, koefisien *fixed effect* konsisten pada kedua kondisi tersebut, namun tidak efisien jika dibandingkan dengan *random effect* ketika tidak terdapat korelasi. Jika hasil statistik menunjukkan tidak korelasi antara residual *cross section* dengan variabel independen, maka *random effect* lebih baik untuk digunakan karena memberikan hasil estimasi yang lebih efisien., dilihat dari jumlah parameter yang tetap meskipun jumlah individu yang diobservasi bertambah. Sebaliknya jika ternyata terdapat korelasi antara  $\mu_i$  dengan  $X_{it}$ , maka *fixed effect* lebih baik digunakan. *Haussman Test* dengan dua variabel didefinisikan sebagai berikut.

$$H = \frac{(\beta_{FEM} - \beta_{REM})^2}{Var(\beta_{FEM}) - Var(\beta_{REM})} \sim \chi_1^2 \dots \dots \dots (3.13)$$

Di mana:

fixed = koefisien parameter pada *fixed effect*

random = koefisien parameter pada *random effect*

Var( rand) = Varians dari koefisien parameter pada *fixed* dan *random*

Pada pengujian ini,  $H_0$  ditolak jika nilai  $H$  lebih besar daripada nilai *chi square* dengan *degree of freedom* sebanyak 1, atau jika nilai probabilitas lebih kecil daripada 0.05. Hipotesa yang diuji adalah:

$H_0$  : tidak terdapat korelasi antara residual *cross section* dengan salah satu variabel independen [ $E(\mu_i | X_{it}) = 0$ ]

$H_1$  : terdapat korelasi antara residual *cross section* dengan salah satu variabel independen [ $E(\mu_i | X_{it}) \neq 0$ ]

### 3. The Breusch-Pagan LM Test

Semakin kecil varians dari residual pada *random effect* ( $\sigma_u^2$ ), maka estimasi REM akan semakin mendekati OLS. Tujuan dari pengujian ini adalah untuk mengetahui apakah varians tersebut nol atau berbeda signifikan dari nol atau dengan kata lain memilih antara pendekatan *random effect* dan *pooled least square*. Hipotesis yang diuji adalah:

$$H_0 : \sigma_u^2 = 0 / \text{random effect}$$

$$H_1 : \sigma_u^2 > 0 / \text{fixed effect}$$

Dasar penolakan terhadap  $H_0$  dengan menggunakan statistik LM yang mengikuti distribusi dari *Chi Square*. Adapun perhitungan untuk statistic LM adalah sebagai berikut : (Baltagi, 2005)

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

$$LM = \frac{nT}{2(t-1)} \left[ \frac{\sum_{i=1}^n (T - \bar{e}_i)^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \dots\dots\dots(3.14)$$

$$LM = \frac{NT}{2(T-1)} \left[ \frac{T^2 \sum_{i=1}^n \bar{e}_i^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \approx \chi^2_{under H_0}$$

#### 3.4.3 Pengujian Pelanggaran Asumsi Klasik

Istilah klasik dalam ekonometrika digunakan untuk menunjukan serangkaian asumsi-asumsi dasar yang dibutuhkan untuk menjaga agar *Ordinary Least Square* (OLS) dapat menghasilkan estimator yang paling baik pada model-model regresi. Terdapat beberapa asumsi *error* pada regresi, yaitu:

1.  $E(u_i) = 0$  ; nilai rata-rata dari error adalah nol
2.  $\text{var}(u_i) = \sigma^2 < \infty$  ; varians dari error bersifat konstan dan finite untuk setiap  $x_t$
3.  $\text{cov}(u_i, u_j) = 0$  ; error bersifat independen secara statistik

4.  $\text{cov}(u_i, x_i) = 0$  ; tidak ada hubungan antara error dengan  $x$
5.  $u_i \sim N(0, \sigma^2)$  ;  $u_i$  memiliki distribusi normal

Jika errors hasil regresi memenuhi syarat 1 sampai 4 maka dapat dikatakan parameter yang diestimasi telah memiliki karakteristik BLUE (*best linear unbiased estimators*). Best; OLS estimator memiliki *minimum variance*, Linear; parameter yang diestimasi bersifat linear, Unbiased; nilai sesungguhnya dari parameter akan sama dengan nilai estimasinya. Karakteristik dari estimator yang bersifat BLUE adalah:

1. *Consistency*; kemungkinan nilai estimasi akan berbeda jauh dengan nilai parameter populasi akan mendekati nol apabila jumlah sampel ditambah
2. *Unbiasedness*: secara rata-rata nilai estimasi akan mendekati nilai parameter populasi
3. *Efficiency*: tidak ada estimator lain yang memiliki varians lebih kecil

Dalam prakteknya seringkali model regresi yang kita gunakan tidak bisa memenuhi semua asumsi di atas. Pembahasan kali ini akan melihat pelanggaran asumsi apa saja yang mungkin terjadi, konsekuensinya, cara mendeteksinya dan cara untuk memperbaiki.

Untuk asumsi yang pertama,  $E(u_i) = 0$  biasanya terpenuhi pada model regresi yang memasukkan intercept. Jika kita tidak memasukkan intercept pada persamaan regresi dan ternyata rata-rata error tidak sama dengan nol maka konsekuensinya:

- 1)  $R^2$  bisa negatif; rata-rata sampel bisa memberikan penjelasan lebih baik daripada variabel independen
- 2) Akan terjadi bias yang cukup besar pada *slope* yang diestimasi.

Untuk mengetahui apakah suatu model bersifat BLUE maka perlu dilakukan beberapa pengujian:

## 1. Heteroskedastisitas

Pada asumsi yang kedua,  $\text{var}(u_i) = \sigma^2 < \infty$  variasi dari *error* diharapkan konstan untuk setiap observasi. Asumsi ini disebut juga *homoscedasticity*. Ketika variasi dari *error* tidak konstan maka kita akan menemui *heteroscedasticity* di dalam *error*. Konsekuensi adanya heteroskedastisitas dalam suatu penelitian adalah:

1. Estimator yang dihasilkan tetap konsisten, tetapi tidak lagi efisien karena ada estimator lain yang memiliki varians lebih kecil daripada estimator yang memiliki eror heteroskedastik.
2. *Standard error* yang dihitung dari OLS yang memiliki error heteroskedastik tidak lagi akurat. Hal ini menyebabkan inferensi (uji hipotesis) yang menggunakan standard error ini tidak akurat.

Cara mendeteksi terjadinya *heteroscedasticity* bisa dilakukan dengan metode informal maupun uji formal. Pengamatan informal dilakukan dengan metode grafis dengan cara mem-plot residual kuadrat dengan atau dengan mem-plot residual kuadrat dengan salah satu variabel independen. Pengamatan formal dapat dilakukan dengan menggunakan *white test*. Kerangka uji *white test* adalah sebagai berikut:

Model yang akan diestimasi regresi adalah:

$$r_{it} = \alpha_{it} + \beta_1 BMR_{it} + \beta_2 DER_{it} + \beta_3 LEQ_{it} + \beta_4 INF + \beta_5 SBI + \beta_6 PDB + \varepsilon_{it} \quad (3.15)$$

Dan

$$r_{it} = \alpha_{it} + \varepsilon_{it} \dots\dots\dots(3.16)$$

Selanjutnya akan dilakukan *auxiliary regression*. Model:

$$r_{it} = \alpha_1 + \alpha_2 BMR_{it} + \alpha_3 DER_{it} + \alpha_4 LEQ_{it} + \alpha_5 BMR_{it}^2 + \alpha_6 DER_{it}^2 + \alpha_7 LEQ_{it}^2 + \alpha_8 BMR_{it} DER_{it} LEQ_{it} + \varepsilon_{it}$$

.....(3.17)

Auxiliary regression untuk makroekonomi:

$$r = \alpha_1 + \alpha_2 INF + \alpha_3 SBI + \alpha_4 PDB + \alpha_5 INF^2 + \alpha_6 SBI^2 + \alpha_7 PDB^2 + \alpha_8 INF, SBI, PDB + \varepsilon_{it}$$

.....(3.18)

Rumusan Hipotesa pada *white test* adalah:

H<sub>0</sub> : tidak ada heteroskedastisitas

H<sub>1</sub> : ada heteroskedastisitas

Nilai yang dibandingkan adalah antara nilai tabel dari *chi-square* dengan df sama dengan jumlah *regressors* (intercept dikeluarkan) dengan *sample size* (n) dikalikan R<sup>2</sup> dari *auxiliary regression*. Jika hasil penghitungan pada persamaan di atas melebihi nilai *chi-square* tabel maka kita menolak H<sub>0</sub> dan menyimpulkan bahwa ada heteroskedastisitas. Persamaannya dapat dirumuskan:

$$n.R^2 : \chi_{df}^2 \dots\dots\dots(3.19)$$

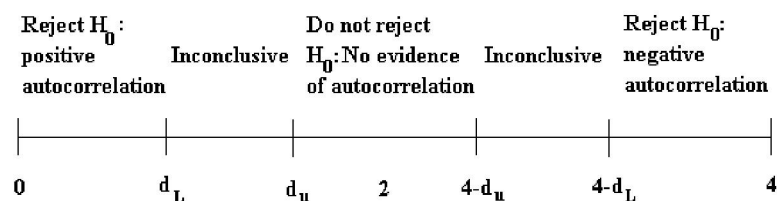
Proses remedial untuk heteroskedastisitas, yaitu untuk memperbaiki distribusi *error* yang *heteroscedastic* ada beberapa cara yang dapat dilakukan, di mana tergantung pada pola *heteroscedasticity* yang kita anggap terjadi. Pada *White test* asumsi yang digunakan adalah variasi dari eror atau residual mengikuti pola *regressors*, kuadrat dari *regressors* dan hasil perkalian dari *regressors*. Nyatanya masih ada beberapa bentuk variasi eror yang mungkin terjadi tetapi tidak bisa diidentifikasi. Cara-cara yang dapat dilakukan untuk memperbaiki distribusi *error* yang *heteroscedastic* adalah, pertama, jika kita yakin tentang pola heteroskedastisitas yang terjadi pada data set, kita dapat membagi keseluruhan model kita dengan pola data tersebut dan menerapkan *Weighted Least Square*. Kedua, dengan merubah *standard error* dari OLS dengan *White heteroscedasticity consistent coefficient variance*. Cara yang lain adalah dengan mengubah variabel menjadi log, karena ada kemungkinan variabilitas data akan membuat banyak *outliers* dan transformasi data diharapkan mampu memperkecil *range* data.

## 2. Autokorelasi

Pada asumsi ketiga,  $\text{cov}(\mu_i, \mu_j) = 0$ , menyatakan bahwa kovarians antar error bernilai nol. Karena dalam penelitian yang menggunakan data cross-section dan time series seperti penelitian ini, eror antar observasi yang satu dengan lainnya tidak boleh saling berkorelasi. Adanya korelasi antar eror menyebabkan autokorelasi (*autocorrelation*). Konsekuensi adanya autokorelasi sama dengan konsekuensi adanya heteroskedastisitas, yaitu:

1. Estimator yang dihasilkan tetap konsisten, tetapi tidak lagi efisien karena ada estimator lain yang memiliki varians lebih kecil daripada estimator yang memiliki eror heteroskedastik
2. *Standard error* yang dihitung dari OLS yang memiliki eror heteroskedastik tidak lagi akurat. Hal ini menyebabkan inferensi (uji hipotesis) yang menggunakan standard error ini tidak akurat.

Cara untuk mendeteksi adanya autokorelasi adalah dengan menggunakan uji *Durbin-Watson* (uji-DW). Uji-DW menguji autokorelasi order pertama (antara error sekarang dengan error satu periode kebelakang). Hipotesis null-nya adalah tidak ada autokorelasi. *DW-table* dapat dicari dengan  $T$ =jumlah observasi, dan  $k$ =jumlah variabel independen (tidak termasuk intercept).



Adapun beberapa syarat yang harus dipenuhi untuk melakukan uji-DW:

1. Harus ada intersep pada persamaan regresi
2. Variabel independen harus bersifat *non-stochastic*
3. Tidak ada lag dari variabel dependen pada persamaan regresi

Pengujian DW dilakukan dengan hipotesis sebagai berikut:

$H_0$  :  $\rho = 0$  (tidak terdapat autokorelasi)

$H_1$  :  $\rho \neq 0$  (terdapat autokorelasi)

Nilai statistik DW kira-kira sama dengan  $2(1 - \rho)$ . Karena  $\rho$  adalah autokorelasi, nilai  $DW$  akan terletak di antara -1 dan 1 atau  $-1 \leq DW \leq 1$ . Dengan mensubstitusikan batasan ini terhadap nilai  $0 \leq DW \leq 4$ , maka dalam uji statistik Durbin Watson, implikasi dari nilai DW adalah:

1.  $\rho = 0$ ,  $DW = 2$ . Artinya tidak ada autokorelasi di antara residual atau dengan kata lain hipotesis nol ditolak jika nilai DW mendekati 2
2.  $\rho = 1$ ,  $DW = 0$ . Artinya, residual berkorelasi secara positif antar yang satu dengan yang lain. Atau terdapat *positive autocorrelation* dalam error.
3.  $\rho = -1$ ,  $DW = 4$ . Artinya, residual berkorelasi secara negatif antar yang satu dengan yang lain. Atau terdapat *negative autocorrelation* dalam error
4. Ada wilayah *inconclusive* pada table DW yang berarti hasil penghitungan tidak dapat menghasilkan keputusan ada atau tidaknya autokorelasi.

Jika terbukti terdapat autokorelasi, maka langkah remedial untuk mengatasinya adalah dengan melakukan penyesuaian pada *standard error* untuk meningkatkan akurasi uji hipotesis yang kita lakukan untuk parameter yang diestimasi. Untuk melakukan perubahan *standar error*, dapat digunakan aplikasi program Eviews uji residual, dengan *correlogram* Q-stat.

### 3. Multikolinearitas

Multikolinearitas merupakan situasi dimana terjadinya hubungan linear antar variabel independen. Di mana asumsi dalam regresi adalah tidak adanya korelasi antar variabel independen. Jika tidak adanya hubungan antar variabel, maka dapat dikatakan orthogonal satu sama lainnya. Jika variabel tersebut orthogonal satu dengan lainnya, maka penambahan atau pengurangan variabel

dari persamaan regresi tidak akan menyebabkan nilai koefisien variabel lainnya berubah. konsekuensi dari adanya multikolinearitas adalah:

1. Jika terdapat multikolinearitas, parameter yang diestimasi akan bersifat BLUE tetapi estimator akan memiliki varians dan *standard error* yang besar sehingga uji hipotesis kurang akurat. Karena *standard error* yang besar maka interval pengujian akan besar sehingga hipotesa nol akan sering ditolak.
2. T-stat akan banyak yang tidak signifikan walaupun  $R^2$  tinggi
3. Estimator OLS akan sensitif terhadap perubahan kecil pada data

Untuk langkah-langkah pendeteksian atas multikolinearitas dapat dilakukan dengan beberapa cara:

1. Menghitung koefisien korelasi antar variabel independen
2. Melihat apabila  $R^2$  tinggi tetapi tidak ada atau sedikit t-stat yang signifikan
3. Melakukan regresi antar satu variabel independen dengan variabel independen lainnya. Jika terdapat hasil regresi yang  $R^2$  tinggi maka ada kemungkinan multikolinearitas.

Bila ternyata terdapat multikolinearitas, langkah remedial untuk memperbaikinya dapat dilakukan dengan:

1. Melakukan transformasi terhadap variabel yang berkorelasi tinggi menjadi rasio dan memasukkan rasio tersebut kedalam model sebagai variabel penjelasnya. Hal ini juga tidak dapat dilakukan jika terdapat teori dasar yang menjadi landasan penyertaan variabel tersebut
2. Menghilangkan salah satu variabel yang saling berkorelasi. Hal ini tidak dapat dilakukan jika terdapat teori dasar yang menjadi landasan penyertaan variabel di dalam model. Dan juga jika variabel yang akan dihilangkan relevan terhadap proses perolehan variabel terikat di dalam model.



3. Membiarkan variabel yang berkorelasi, karena terkadang keberadaan multikolinieritas tidak menurunkan rasio T dari variabel yang mungkin signifikan tanpa adanya multikolinieritas. Menurut Blanchard (1967) multikolinieritas juga disebut dengan *force majeure/God's will*, sehingga sangat sulit untuk dihindari.

