

BAB III

DATA DAN METODOLOGI

3.1 Spesifikasi Model

Penelitian ini menggunakan model yang sama dengan model yang digunakan oleh Fan, Hazzel, dan Thorat dalam penelitian yang berjudul *Government Spending, Growth And Poverty: An Analysis Of Interlinkages In Rural India* yang dipublikasikan pada tahun 1998. Pada penelitian dilakukan beberapa penyesuaian karena tidak tersedianya data dan perbedaan karakteristik antara India dengan Indonesia

Model asli yang digunakan pada penelitian Fan, Hazel, dan Thorat di India pada tahun 1998 adalah :

$$P = f(TFP, WAGE, NAGEMPLY, TT, LANDN, POP, RAIN, T) \quad (3.1)$$

$$TFP = f(RDE, RDE_{.1}, \dots, RDE_{.i}, IR, LITE, ROADS, RAIN, T) \quad (3.2)$$

$$WAGE = f(TFP, ROADS, LITE, HELE, HELE_{.1}, \dots, HELE_{.l}, T) \quad (3.3)$$

$$NAGEMPLY = f(GERDEV, ROADS, LITE, GCSSL, PVELE, T) \quad (3.4)$$

$$PUIR = f(IRE, IRE_{.1}, \dots, IRE_{.j}, PVELE, ATT, T) \quad (3.5)$$

$$PRIR = f(PUIR, PVELE, ATT, T) \quad (3.6)$$

$$ROADS = f(ROADE_{.1}, \dots, ROADE_{.k}, T) \quad (3.7)$$

$$LITE = f(EDE, EDE_{.1}, \dots, EDE_{.m}, T) \quad (3.8)$$

$$PUELE = f(PWRE, PWRE_{.1}, \dots, EDE_{.m}, T) \quad (3.9)$$

$$LANDN = f(TFP, T) \quad (3.10)$$

$$TT = f(TFP, TFP_n, WAPI, T) \quad (3.11)$$

Keterangan :

POP = tingkat pertumbuhan penduduk di desa

WAPI = world agricultural price index

IRE = pengeluaran pemerintah untuk irigasi

RDE = pengeluaran pemerintah untuk penelitian agrikultural

ROADE = pengeluaran pemerintah untuk jalan di desa

EDE = pengeluaran pemerintah untuk pendidikan di desa

PWRE = pengeluaran pemerintah untuk listrik di desa

GCSSL = pengeluaran pemerintah untuk konservasi tanah dan air

HELE = pengeluaran pemerintah untuk kesehatan

GERDEV = pengeluaran pemerintah untuk pengembangan masyarakat dan desa

RAIN = tingkat curah hujan

T = *time trend*

ATT = *moving 5-year average of the terms of trade*

P = persentase kemiskinan di desa (*headcount poverty*)

LITE = tingkat melek huruf

ROADS = kepadatan jalan di desa

IR = persentase sawah yang diirigasi

PUIR = persentase sawah yang beririgasi publik

PRIR = persentase sawah yang beririgasi privat

PVELE = persentase penduduk desa yang memiliki listrik

WAGE = tingkat upah buruh petani

NAEMPLY = persentase pekerja nonagrikultur pada total pekerja di desa

TFP = *total factor productivity* (Tornqvist-Theil Index)

TFPn = *total factor productivity* pada tingkat nasional (Tornqvist-Theil Index)

LANDN = Persentase rumah tangga yang tidak memiliki tanah

TT = nilai tukar petani (*terms of trade*)

Pada penelitian yang dilakukan oleh mereka setiap variabel diubah menjadi *geometric annual growth rate in logarithm*, yaitu $dx = \ln(x_t/x_{t-n})/n$, dimana karena

menggunakan n adalah 1, maka setiap variabel diubah menjadi turunan pertama dalam logaritma.

Sedangkan model yang digunakan dalam penelitian ini adalah sebagai berikut :

$$Illiterate_{it} = \beta_1 + \varphi_1 Exp_pendidikan_{it} + \varepsilon_1 \quad (3.12)$$

$$Jalan_{it} = \beta_2 + \psi_1 Exp_perhubungan_{it} + \varepsilon_2 \quad (3.13)$$

$$Emponagri_{it} = \beta_3 + \sigma_1 Jalan_{it} + \sigma_2 Illiterate_{it} + \sigma_3 Listrik_{it} + \varepsilon_3 \quad (3.14)$$

$$Produktivitas_{it} = \beta_4 + \gamma_1 Sawah_irigasi_{it} + \gamma_2 Illiterate_{it} + \gamma_3 Jalan_{it} + \gamma_4 Hujan_{it} + \varepsilon_4 \quad (3.15)$$

$$Upah_buruhtani_{it} = \beta_5 + \theta_1 Produktivitas_{it} + \theta_2 Jalan_{it} + \theta_3 Illiterate_{it} + \theta_4 Kesehatan_{it} + \varepsilon_5 \quad (3.16)$$

$$TT_{it} = \beta_6 + \omega_1 TFP_{it} + \omega_2 Produktivitas_{it} + \varepsilon_6 \quad (3.17)$$

$$Pov_{it} = \beta_7 + \alpha_1 Produktivitas_{it} + \alpha_2 Upah_buruhtani_{it} + \alpha_3 Emponagri_{it} + \alpha_4 TT_{it} + \alpha_5 Pop_growth_{it} + \alpha_6 Hujan_{it} + \varepsilon_7 \quad (3.18)$$

Keterangan:

Pov = persentase kemiskinan di desa

Produktivitas = tingkat produktivitas lahan sawah (quintal/hektare)

Upah_buruhtani = upah buruh tani harian

Emponagri = persentase pekerja nonagrikultur di desa

TT = terms of trade (nilai tukar petani)

Pop_growth = pertumbuhan populasi desa

Hujan = tingkat curah hujan (mm)

Sawah_irigasi = persentase sawah beririgasi

Illiterate = persentase buta huruf di desa

Jalan = persentase desa dengan permukaan jalan terluas aspal

Exp_kesehatan = pengeluaran untuk kesehatan di desa

Exp_pendidikan = pengeluaran untuk pendidikan di desa

Exp_perhubungan = pengeluaran untuk perhubungan di desa

Listrik = persentase rumah tangga di desa yang telah tersambung dengan listrik

TFP = total faktor produktivitas tingkat nasional(Tornqvist-Theil index)

Model di atas diregresikan dengan menggunakan metode rekursif dan data panel. Selain itu, digunakan juga fungsi *double log*. Hal ini bertujuan agar data menjadi lebih linear dan mempunyai nominal yang sama. Penjelasan tentang data panel, model rekursif, dan pengujian model akan dijelaskan lebih lanjut pada bagian berikutnya.

3.2 Data

Data yang digunakan dalam penelitian ini berasal dari berbagai sumber. Data tersebut merupakan data pada tingkat propinsi dan dimulai dari tahun 1993 hingga 2005. Data propinsi yang digunakan adalah propinsi Sumatera Selatan, Sumatera Utara, Sumatera Barat, Lampung, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Selatan, dan Sulawesi Selatan.

Berikut ini adalah penjelasan jenis data, sumber data, dan keterangan lainnya.

- *Poverty* atau kemiskinan
Angka kemiskinan dengan metode *headcount poverty* di desa pada setiap propinsi didapatkan dari Biro Pusat Statistik (BPS). Data tahun 1993-1999 merupakan data tiga tahunan. Sedangkan data mulai tahun 2000 adalah data tahunan. Terdapat perbedaan antara cara penghitungan dengan metode sebelum dengan setelah 1998. Pada data ini, hanya tahun 1993 yang menggunakan “Metode 1993”. Angka ini adalah dalam bentuk persentase.

- Buta huruf
Data ini didapatkan dari BPS (dalam buku Statistik Kesejahteraan) dalam bentuk persentase dan hanya mencakup orang yang telah berusia 10 tahun ke atas.
- Upah buruh tani
Data ini merupakan hasil kalkulasi dari BPS. Pada data yang sebenarnya, dapat dilihat bahwa upah buruh tani terbagi atas 3 macam, yaitu upah mencangkul, merambat, dan menanam, di mana upah tertinggi pada umumnya terjadi pada saat mencangkul. Data upah ini merupakan data upah yang diterima buruh tani perhari secara rata-rata berdasarkan ketiga macam upah buruh tani. Pada tahun 2004, rata-rata mereka bekerja yang semula dari 4-6 jam kerja berubah menjadi 5-6 jam kerja. Angka ini telah mengalami penyesuaian dengan indeks harga konsumen dengan tahun dasar 2000. Angka ini dalam bentuk rupiah.
- Curah hujan
Data curah hujan berasal dari Badan Meteorologi dan Geofisika (BMG). Data ini merupakan data bulanan. Untuk Propinsi Sumatera Utara, data berasal dari stasiun Medan. Untuk Sumatera Barat, Sumatera Selatan, dan Lampung berturut-turut berasal dari Padang, Palembang, dan Tanjung Karang. Untuk Propinsi Jawa Barat, Jawa Tengah dan Jawa Timur, stasiun BMG berturut-turut adalah Dermaga, Ahmad Yani (untuk tahun 193 dan 1994 berasal dari stasiun Simongan), dan Juanda. Pada Propinsi Kalimantan Selatan stasiun BMG yang ada adalah di Banjarmasin. Sedangkan untuk Sulawesi Selatan adalah Ujung Pandang. Angka merupakan dalam mm (milimeter).
- Produktivitas
Produktivitas ini merupakan data hasil produktivitas lahan yang ditanami oleh padi. Baik yang berasal dari sawah irigasi maupun ladang. Angka ini adalah dalam ton/hektar, berupa data propinsi, serta didapatkan dari BPS.
- Jalan
Data tentang jalan didapatkan dari Potensi Desa (Podes) dari tahun 1993. Podes sendiri keluar setiap 3 tahun. Jenis jalan yang diambil adalah jalan yang beraspal. Data awal adalah banyaknya desa yang memiliki

permukaan jalan terluas berupa aspal. Kemudian data tersebut dibagi dengan jumlah desa definitif pada setiap propinsi. Data ini dalam bentuk persentase.

- Pekerja nonagrikultur

Data ini merupakan data persentase dari orang yang bekerja pada bidang nonagrikultur pada daerah pedesaan. Format data ini berubah pada tahun 2004. Pada data sebelum tahun 2004, orang yang termasuk dalam kategori ini adalah orang yang berumur di atas 10 tahun. Mulai pada 2004 terjadi perubahan format menjadi orang yang berumur di atas 15 tahun. Data ini berasal dari BPS.

- Pengeluaran untuk pendidikan dan kesehatan

Data ini berasal dari dua sumber, yaitu BPS dan Departemen Keuangan. Data ini telah disesuaikan dengan Indeks Harga Konsumen tahun 2000 dan dalam bentuk rupiah.. Untuk beberapa propinsi, pengeluaran untuk sektor ini pada APBD sebelum tahun 2002 adalah tergabung dalam sektor kehutanan. Namun, setelah tahun 2002 beberapa propinsi ada yang mulai menggunakan format baru, di mana sektor kehutanan terpisah dengan sektor pertanian. Data ini telah disesuaikan dengan Indeks Harga Konsumen (IHK) dengan tahun dasar 2000.

- Pengeluaran untuk perhubungan desa

Data ini merupakan data hasil penghitungan BPS pada tingkat desa dan bukanlah data riil seluruh pengeluaran untuk perhubungan di desa pada setiap propinsi. Data ini dihitung dengan cara pengambilan sampel.

- *Terms of trade* (nilai tukar) petani

Data ini merupakan hasil bagi dari nilai yang diterima oleh petani dan nilai yang dibayarkan oleh petani. Pada awalnya, tahun dasar untuk propinsi di Pulau Jawa adalah 1983 sedangkan untuk propinsi di luar Pulau Jawa adalah 1987. Data mengalami perubahan tahun dasar menjadi 1993 pada tahun 1997. Pada akhirnya semua nilai *terms of trade* ini disesuaikan untuk memiliki tahun dasar pada tahun 1993. Data ini berasal dari BPS.

- Total faktor produktivitas agrikultur
Data ini merupakan hasil penelitian dari Keith O Fuglie (n.d). Akan tetapi, data ini hanya tersedia sampai tahun 2000 sehingga untuk tahun selanjutnya dilakukan *forecasting*. Data ini merupakan data pada tingkat nasional.
- Elektrifikasi
Data ini adalah persentase dari jumlah rumah tangga pada daerah pedesaan di setiap propinsi di Indonesia yang telah memiliki sambungan terhadap listrik. Data ini dikeluarkan oleh BPS dalam buku Statistik Kesejahteraan.
- Pertumbuhan populasi desa
Data ini merupakan hasil kalkulasi berdasarkan jumlah orang yang tinggal di daerah pedesaan pada setiap propinsi. Data ini dalam bentuk persentase.

Seluruh data pengeluaran pemerintah kecuali untuk perhubungan desa disesuaikan dengan persentase orang yang tinggal di desa pada masing-masing propinsi.

3.3 Metode Pengolahan Data

3.3.1 Rekursif

Rekursif merupakan salah satu metode persamaan simultan. Ketika *Ordinary Least Square (OLS)* diaplikasikan dalam konteks persamaan simultan, model yang didapatkan dapat berupa model *recursive*, *triangular*, dan *causal*. Untuk melihat karakteristik model ini, perhatikan tiga persamaan berikut ini:

$$Y_{1t} = \beta_{10} + \gamma_{11} X_{1t} + \gamma_{12} X_{2t} + u_{1t} \quad (3.19)$$

$$Y_{2t} = \beta_{20} + \beta_{21} Y_{1t} + \gamma_{21} X_{1t} + \gamma_{22} X_{2t} + u_{2t} \quad (3.20)$$

$$Y_{3t} = \beta_{30} + \beta_{31} Y_{1t} + \beta_{32} Y_{2t} + \gamma_{31} X_{1t} + \gamma_{32} X_{2t} + u_{3t} \quad (3.21)$$

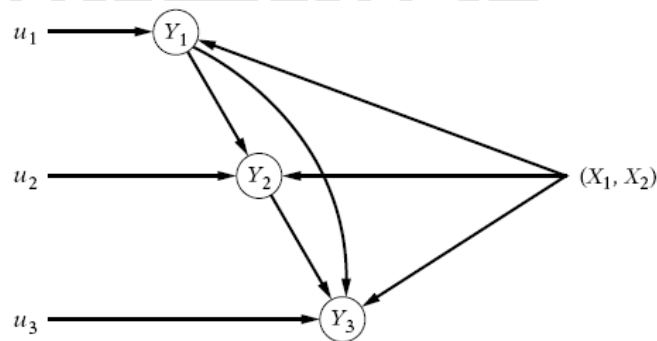
dimana, Y dan X merupakan variabel endogen dan eksogen. Gangguan yang terjadi adalah:

$$\text{cov}(u_{1t}, u_{2t}) = \text{cov}(u_{1t}, u_{3t}) = \text{cov}(u_{2t}, u_{3t}) = 0 \quad (3.22)$$

dengan gangguan dalam persamaan yang berbeda pada periode yang sama adalah tidak berkorelasi (asumsi: *zero contemporaneous correlation*).

Persamaan (3.9), yaitu persamaan untuk Y_{2t} . Karena variable eksogen yang terletak pada sebelah kanan diasumsikan tidak mempunyai korelasi dengan *error* u_{1t} , persamaan ini memenuhi asumsi kritis OLS klasik sehingga OLS dapat diaplikasikan pada persamaan ini. Persamaan kedua mengandung variabel endogen Y_{1t} sebagai *explanatory variabel* bersama dengan X *nonstochastic*. OLS juga dapat diaplikasikan pada persamaan ini dengan mengasumsikan bahwa Y_{1t} dan u_{2t} tidak berkorelasi. OLS juga dapat diaplikasikan pada persamaan ketiga dengan mengasumsikan bahwa Y_1 dan Y_2 tidak berkorelasi u_3 .

Dalam sistem rekursif, OLS dapat diaplikasikan secara terpisah pada masing-masing persamaan. Sebenarnya, tidak ada permasalahan persamaan simultan dalam situasi ini. Dari struktur sistem, terlihat jelas bahwa tidak ada saling ketergantungan antara variabel endogen. Ketika Y_1 mempengaruhi Y_2 , Y_2 tidak mempengaruhi Y_1 . Y_1 dan Y_2 mempengaruhi Y_3 sedangkan Y_3 tidak mempengaruhi Y_1 dan Y_2 . Dalam kata lain, masing-masing persamaan memperlihatkan *unilateral causal dependence*, atau dikenal juga dengan nama model kausal. Berikut ini adalah gambaran dari model kausal tersebut, sedangkan gambaran keseluruhan model terdapat pada lampiran:



Gambar 3.1 Hubungan Rekursif

Sumber : Basic Econometrics, Gujarati

3.3.2 Panel

Data panel (*pooled data*) merupakan sekumpulan data yang berisi data dari sampel individu (yaitu beberapa propinsi pada penelitian ini) pada sebuah periode

waktu tertentu. Dengan kata lain, data panel merupakan gabungan antara data deret waktu (*time series*) dengan data kerat lintang (*cross section*). Simbol yang digunakan adalah t untuk deret waktu observasi dan n untuk unit *cross section* yang diobservasi. Proses pembentukan data panel adalah dengan mengkombinasikan unit deret waktu dengan kerat lintang sehingga terbentuklah suatu kumpulan data. Proses tersebut sendiri disebut dengan *pooling*. Data panel dapat diolah jika memiliki kriteria $t > 1$ dan $n > 1$. Jika $t = 1$ dan $n \geq 1$, data disebut dengan deret waktu murni, sedangkan jika $t \geq 1$ dan $n = 1$, disebut dengan kerat lintang murni. Jika jumlah periode observasi sama banyaknya untuk setiap unit *cross section*, dinamakan dengan *balanced panel*. Sebaliknya juga berlaku, yaitu apabila jumlah periode observasi pada setiap *cross section* tidak sama dengan periode observasi, disebut dengan *unbalanced panel*.

Terdapat beberapa keuntungan yang didapatkan jika menggunakan data panel ini. Pertama, dapat mendalami efek-efek ekonomi yang tidak dapat diperoleh jika hanya menggunakan data *time series* maupun *cross section*. Kedua, bila jumlah data dan observasi yang meningkat, peningkatan ini akan menghasilkan kenaikan derajat kebebasan (*degree of freedom*) sehingga variasi koefisien menjadi efisien dan nilai koefisien menjadi lebih stabil (Hsiao, 1986). Ketiga, dengan mengakomodasi semua informasi yang terkait dengan variabel-variabel kerat lintang dan deret waktu, data panel mampu menurunkan masalah *omitted variable*, terutama jika harus menghilangkan variabel yang relevan. Bersamaan dengan itu, masalah kesalahan spesifikasi pun dapat dikurangi/hilangkan.

Beberapa hal di atas sesuai dengan apa yang dikemukakan oleh Baltagi (2001) yang menyatakan beberapa manfaat dari menggunakan data panel, antara lain:

1. Mampu mengontrol heterogenitas individu.
2. Memberikan informasi lebih banyak dan lebih bervariasi dibandingkan dengan hanya menggunakan data deret waktu atau deret lintang. Data panel juga mengurangi kolinearitas antar variabel, meningkatkan *degree of freedom*, dan meningkatkan efisiensi.
3. Sangat baik untuk mempelajari model perilaku (*behavioral model*) yang lebih kompleks.

4. Sangat baik untuk digunakan dalam studi perubahan yang dinamik (*study of dynamics adjustment*).
5. Dapat mendeteksi dan mengukur efek dengan lebih baik dibandingkan dengan data deret waktu murni dan kerat lintang murni.

Dalam mengestimasi data panel terdapat tiga cara yang dapat digunakan, yaitu *pooled least square*, *fixed effect*, dan *random effect*.

3.3.2.1 Pooled least square (OLS)

Proses estimasi dilakukan dengan menggunakan metode kuadrat terkecil biasa, yaitu :

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \quad (3.23)$$

untuk $i=1,2,3,\dots,N$ dan $t=1,2,3,\dots,T$

N = jumlah kerat lintang

T =jumlah time series

Metode ini merupakan metode yang paling sederhana. Namun, hasil yang diberikan kurang memadai karena setiap observasi dianggap berdiri sendiri. Proses estimasi dapat dilakukan untuk setiap unit kerat lintang dengan asumsi komponen error pada data panel ini sama dengan komponen error dalam pengolahan kuadrat terkecil biasa (OLS).

Untuk periode $t=1$ akan diperoleh persamaan regresi kerat lintang sebagai berikut:

$$Y_{i1} = \alpha + \beta X_{i1} + \varepsilon_{i1} \quad (3.24)$$

Persamaan di atas berimplikasi pada diperolehnya persamaan sebanyak T yang sama. Begitu juga sebaliknya, dapat diperoleh persamaan deret waktu sebanyak N persamaan untuk setiap T observasi. Namun, untuk mendapatkan parameter α dan β yang konstan dan efisien, dibutuhkan bentuk regresi yang lebih besar dengan melibatkan observasi sebanyak N dikalikan T observasi.

Metode ini tidak memperhatikan perbedaan yang mungkin akan timbul akibat dimensi ruang dan waktu. Model ini mengasumsikan bahwa *intercept* dan *slope* koefisien dari dua variabel adalah sama untuk semua unit kerat lintang. Karena terdapat kemungkinan atas ketidaksesuaian asumsi ini, model ini mungkin akan mendistorsi deskripsi dari hubungan X dan Y yang sebenarnya.

3.3.2.2 Fixed Effect Model

Untuk memungkinkan terjadinya perubahan dalam *intercept* dari unit kerat lintang dan deret waktu, digunakan variabel boneka (*dummy variable*) sehingga akan terjadi perbedaan nilai parameter, baik atas unit kerat lintang maupun deret waktu. Pendekatan yang paling sering dilakukan adalah dengan membiarkan *intercept* bervariasi antarunit kerat lintang namun tetap menghasilkan asumsi bahwa koefisien variabel adalah konstan untuk unit kerat lintang. Pendekatan ini dikenal dengan sebutan model efek tetap (*fixed effect model/fem*). Pendekatan ini dapat ditulis dengan persamaan sebagai berikut :

$$Y_{it} = \alpha + bX_{it} + g_2W_{2t} + g_3W_{3t} + \dots + g_NW_{Nt} + d_2Z_{i2} + d_3Z_{i3} + \dots + d_TZ_{iT} + e_{it} \quad (3.25)$$

di mana, $W_{it}=1$ untuk individu ke- i , $i=2,\dots,N$

$=0$ untuk sebaliknya

$Z_{it}=1$ untuk periode ke- t , $t=2,\dots,T$

$=0$ untuk sebaliknya

Dari persamaan di atas, terlihat bahwa telah ditambahkan sebanyak $(N-1)+(T-1)$ variabel boneka ke dalam model dan menghilangkan dua sisanya untuk menghindari kolinearitas sempurna antar variabel independen. Dengan menggunakan pendekatan ini, akan didapatkan *degree of freedom* sebesar $NT-2-(N-1)-(T-1)$, atau sebesar $NT-N-T$.

Penggunaan model *fixed effect* di atas dapat dilakukan jika memiliki sedikit unit kerat lintang. Namun jika unit kerat lintang tergolong besar, penggunaan model *fixed effect* akan mengurangi derajat kebebasan yang pada akhirnya akan mengurangi efisiensi dari parameter yang diestimasi.

3.3.2.3 Random Effect Model

Metode ini mengasumsikan bahwa komponen error individu tidak berkorelasi satu sama lainnya dan komponen *error* antarwaktu dan kerat lintang juga tidak berkorelasi (*no autocorrelation*) (Pyndick, 1998). Dalam model ini parameter-parameter yang berbeda antardaerah maupun antarwaktu dimasukkan ke dalam

error. Hal ini dilakukan untuk meningkatkan efisiensi proses pendugaaan OLS.

Bentuk model ini adalah :

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \quad (3.26)$$

$$E_{it} = u_i + v_t + w_{it} \quad (3.27)$$

di mana:

U_i =komponen error antar kerat lintang

V_t =komponen error deret waktu

W_{it} =komponen error kombinasi

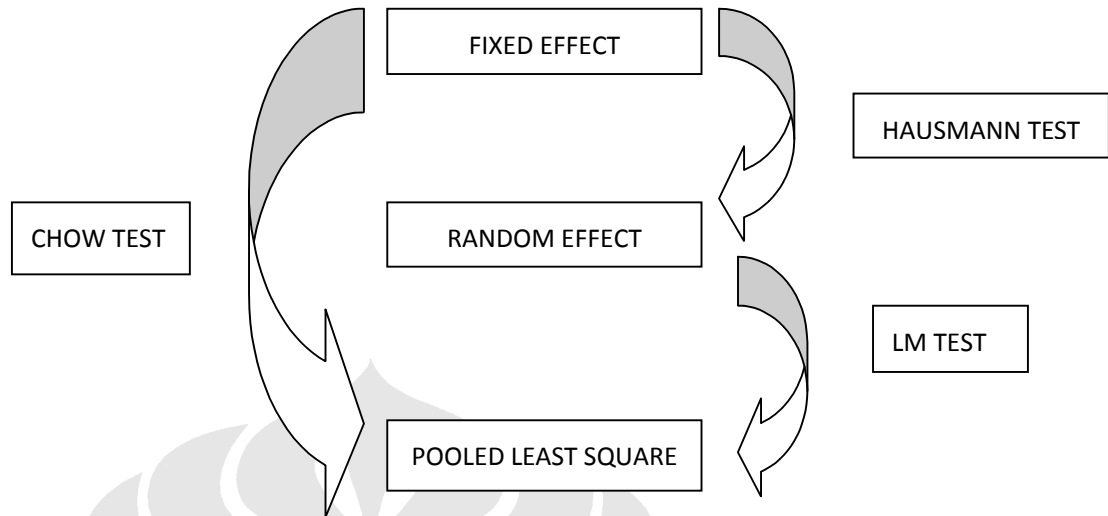
3.3.2.4 Pemilihan metode estimasi

Dari penjelasan di atas diketahui bahwa terdapat tiga pendekatan dalam metode data panel. Pada bagian ini akan dijelaskan pemilihan pendekatan yang paling tepat diantara ketiganya. Pemilihan ini bertujuan agar pendekatan yang dipilih cocok dengan tujuan penelitian dan karakteristik data sehingga proses estimasi memberikan hasil yang lebih tepat.

Metode OLS terlalu sederhana untuk mendeskripsikan fenomena yang terdapat dalam data sehingga pilihan selanjutnya adalah memilih di antara *fixed effect model* dengan *random effect model*. Pemilihan antara dua model tersebut dapat dilakukan secara teoritis. Jika diasumsikan bahwa error kerat lintang tidak berkorelasi dengan variabel X , *random effect model* yang dipilih. Jika diasumsikan bahwa error kerat lintang berkorelasi dengan X (*error* mempunyai pengaruh tetap/dianggap sebagai bagian dari *intercept*), *fixed effect model* yang dipilih.

Jika secara teoritis tidak dapat ditentukan model mana yang akan dipilih, kriteria dasar pemilihan model selanjutnya dapat didasarkan pada sampel penelitian. Jika data diambil dari sampel individu atas suatu populasi yang besar secara acak, *random effect model* yang dipilih. Namun jika sampel merupakan seluruh populasi yang dipilih, *fixed effect model* merupakan metode yang tepat (Hsiao, 1986).

Pengujian secara formal untuk menentukan model mana yang lebih baik digunakan dilakukan berdasarkan pengujian statistik. Serangkaian pengujian statistik yang dapat dilakukan terdiri atas beberapa langkah. Hal ini dijelaskan pada gambar dibawah ini:



Gambar 3.2 Tes Data Panel

Dari gambar di atas, dapat dilihat bahwa secara statistik terdapat tiga pengujian yang dapat digunakan untuk menentukan metode apa yang akan dipilih. Ketiga pengujian itu adalah :

1. Chow Test

Chow test merupakan pengujian F statistik yang berfungsi dalam mendapatkan model yang paling tepat di antara model *fixed effect* atau *pooled least square*. Pengujian ini menggunakan hipotesis sebagai berikut:

Ho= model Pooled Least Square (*restricted*)

H1= model Fixed Effect (*unrestricted*)

Jika nilai chow statistik (f-statistik) lebih besar daripada F tabel, Ho ditolak dan model yang dipilih adalah model *fixed effect*. Hal ini berlaku sebaliknya. Besaran nilai Chow didapat dari perhitungan dibawah ini :

$$\frac{(RRSS - URSS)/(N - 1)}{URSS/(NT - N - K)} \quad (3.28)$$

di mana:

RRSS= *Restricted residual sum square*

URSS= *Unrestricted residual sum square*

N= jumlah kerat lintang

T= jumlah deret waktu

K= jumlah peubah bebas

2. Hausman Test

Pengujian ini dilakukan untuk menentukan apakah model *fixed effect* atau model *random effect* yang dipilih. Pengujian ini dilakukan dengan hipotesis sebagai berikut:

H₀= model *random effect*

H₁=model *fixed effect*

Model dari Hausman test sendiri adalah:

$$H = [\hat{\beta}_{FE} - \hat{\beta}_{RE}] \Sigma^{-1} [\hat{\beta}_{FE} - \hat{\beta}_{RE}] \quad (3.29)$$

$$\Sigma = \text{Var}[\hat{\beta}_{FE}] - \text{Var}[\hat{\beta}_{RE}] = \text{Var}[\hat{\beta}_{FE} - \hat{\beta}_{RE}] \quad (3.30)$$

Dasar penolakan H₀ adalah dengan menggunakan pertimbangan statistik *chi square*. Jika *chi square stat* > *chi square table* (p-value < 0), H₀ ditolak (model menggunakan *fixed effect*), dan sebaliknya.

Namun ada pula cara yang lebih sederhana untuk menentukan apakah model yang digunakan antara model *fixed effect* atau model *random effect*, yaitu :

- a. Bila T (banyaknya unit *time series*) besar sedangkan N (jumlah unit *cross section*) kecil, hasil dari *fixed effect* dan *random effect* tidak jauh berbeda sehingga dapat dipilih pendekatan yang lebih mudah, yaitu *fixed effect model*
- b. Bila N besar dan T kecil, hasil estimasi kedua pendekatan tersebut akan berbeda jauh. Jadi apabila diyakini bahwa unit *cross section* yang dipilih dalam penelitian diambil secara acak, model *random effect* harus digunakan. Sebaliknya, apabila diyakini bahwa unit *cross section* yang dipilih dalam penelitian tidak diambil secara acak, digunakan model *fixed effect*.
- c. Apabila komponen error individual berkorelasi dengan variabel bebas X, parameter yang diperoleh dengan menggunakan model *random effect* akan bias sementara parameter yang diperoleh dengan model *fixed effect* tidak bias.

- d. Apabila N besar dan T kecil, dan apabila asumsi yang mendasari model *random effect* dapat dipenuhi, model *random effect* akan lebih efisien dibandingkan dengan model *fixed effect*.

3.4 Pengujian Model

Untuk mengetahui tepat atau tidaknya suatu model, perlu diadakan pengujian melalui kriteria ekonomi dan kriteria statistik.

3.4.1 Kriteria Ekonomi

Pengujian model dengan kriteria ekonomi berarti pada model tersebut dilakukan pengujian berdasarkan teori-teori ekonomi yang ada. Jika hasil dari pembuktian model menunjukkan adanya kesesuaian dengan teori ekonomi, model tersebut dinyatakan benar secara teori ekonomi, begitu pula sebaliknya. Jika tidak sesuai dengan teori ekonomi yang ada, diperlukan penelitian dan pembahasan lebih lanjut mengenai penyebab tentang berbedanya model dari kriteria ekonomi yang berlaku. Jika memang pada akhirnya didapatkan perbedaan antara kenyataan dengan teori bertentangan, hal ini mungkin diakibatkan oleh adanya perbedaan karakteristik dari variabel-variabel yang ada pada model atau bahkan mampu menciptakan suatu teori tersendiri untuk karakteristik yang ada.

3.4.2 Kriteria Statistik

a. Uji Signifikansi

Uji signifikansi dilakukan untuk memeriksa apakah koefisien regresi yang didapat signifikan (berbeda nyata). Maksud dari signifikan di sini adalah nilainya tidak sama dengan nol. Terdapat dua jenis uji signifikansi terhadap koefisien korelasi yang dapat dilakukan, yaitu Uji- F dan Uji- t .

1. Uji- F

Uji ini dilakukan untuk menguji signifikansi koefisien regresi secara bersamaan. Tabel yang digunakan dalam uji ini adalah Tabel ANOVA:

Tabel 3.3 Tabel ANOVA

Sumber	Sum of Squares	Df	Mean Squares	F hitung
Regresi	SSR	K	MSR=SSR/k	F=MSR/MSE
<i>Error</i>	SSE	n-k-1	MSE=SSE/(n-k-1)	
Total	SST	n-1		

Sumber: Nachrowi (2006)

di mana, df adalah *degree of freedom*, k adalah jumlah variabel bebas (koefisien slope), dan n jumlah observasi (sampel). Jika F hitung > F tabel, H₀ ditolak, atau dengan kata lain paling tidak ada satu pun *slope* regresi yang signifikan.

- Uji-t

Uji-t digunakan untuk menguji koefisien regresi secara individu. Rumus t hitung adalah:

$$t = b_j / \text{standard error } b_j \quad (3.31)$$

Nilai t hitung dibandingkan dengan nilai tabel. Jika t hitung > t tabel, H₀ ditolak, berarti koefisien regresi signifikan. Pada E-views, variabel dinilai signifikan bila probabilitas t-statistik di bawah 0,05.

- Koefisien Determinasi (R^2 atau Adjusted R^2)

Koefisien determinasi (*Goodness of Fit*) merupakan suatu ukuran untuk menentukan apakah model yang regresi yang didapatkan telah cukup baik atau belum. Angka koefisien determinasi ini menunjukkan seberapa dekat garis regresi yang terestimasi dengan yang sesungguhnya. Perbedaan antara R^2 dengan *Adjusted R²* adalah R^2 tidak akan pernah menurun nilainya jika terjadi penambahan variabel yang tidak signifikan sedangkan *Adjusted R²* akan menurun nilainya.

b. Penyimpangan Asumsi Klasik

Untuk mengupayakan hasil model yang efisien, *feasible*, dan konsisten, maka diperlukan pendeteksian terhadap pelanggaran asumsi model, yaitu gangguan antara waktu (*time related disturbance*), gangguan antarindividu, dan gangguan akibat keduanya. Agar model yang digunakan dalam model ini *feasible* dan efektif, perlu diperhatikan beberapa pelanggaran atas asumsi dasar, antara lain :

1. Multikolinearitas

Pelanggaran ini muncul jika diantara satu variabel independen memiliki korelasi yang tinggi dengan variabel independen lain sehingga sulit untuk memisahkan efek satu variabel independen dari efek variabel independen lainnya. Hal ini terjadi karena distribusi parameter regresi sangat peka terhadap korelasi antarvariabel bebas dan besaran *error* regresi. Kepekaan ini tercermin dari bentuk varians *error* regresi yang sangat tinggi sehingga nilai *t* hitung menjadi kecil dan akibatnya hipotesa nol biasanya diterima karena besarnya *error* dugaan. *T*-hitung yang semakin kecil menyebabkan menurunnya signifikansi dari *t*. Kemungkinan kedua, ada koefisien yang *overestimates* dan ada koefisien yang *underestimates*. Hal ini berarti bahwa nilai koefisien regresi bukan nilai yang sebenarnya.

Pelanggaran ini menjadi masalah jika tujuan melakukan regresi adalah untuk menafsirkan koefisien regresi. Indikasi dari terdapatnya multikolinearitas dalam suatu model adalah:

- a. Terdapat nilai R^2 yang sangat tinggi dan nilai *F* hitung yang signifikan tetapi nilai sebagian besar *t* hitung tidak signifikan.
- b. Korelasi sederhana yang relatif tinggi (>0.8) antara satu atau lebih pasangan variabel independen. Jika koefisien korelasi kurang dari 0.8 berarti belum terjadi multikolinearitas. Akan tetapi jika lebih dari 0.9 berarti terjadi kolinearitas berganda yang merupakan masalah serius.
- c. Dengan melakukan regresi bantuan, dengan cara meregresi masing-masing variabel bebas dengan variabel bebas lainnya. Apabila nilai R^2 tinggi, terdapat indikasi ketergantungan linear yang hampir pasti diantara variabel *X*.

- d. Cara lain dalam pengujian multikolinearitas adalah dengan menggunakan Variance Inflation Factor (VIF). VIF adalah versi skala dari koefisien korelasi berganda antara variabel j dengan variabel independen lainnya. Cara penghitungannya adalah :

$$VIF_j = \frac{1}{(1 - R_j^2)} \quad (3.32)$$

di mana R_j adalah koefisien korelasi berganda.

Jika R_j sama dengan 0, berarti tidak ada korelasi antara X_j dengan variabel bebas yang lainnya) sehingga VIF_j sama dengan 1. Neter, Wasserman, dan Kutner merekomendasikan untuk melihat pada nilai VIF yang terbesar. Jika nilainya lebih besar dari 10, diduga terdapat masalah autokolinearitas.

Pemecahan masalah multikolinearitas dilakukan dengan cara mengurangi variabel independen dalam model, mengubah bentuk model, dan menambah data/memilih sampel baru.

2. Heterokedastisitas

Pelanggaran asumsi ini terjadi apabila varians dari *error* berubah. Hal ini biasa terjadi pada data *cross section* dan tidak terjadi pada data *time series* karena perubahan pada variabel dependen dan perubahan dalam satu atau lebih variabel independen memiliki kemungkinan yang sama besar. Dampak dari adanya heterokedastisitas adalah pendugaan kuadrat terkecil memberikan bobot lebih besar pada observasi yang memiliki varians *error* lebih besar dibandingkan dengan observasi yang memiliki varians *error* lebih kecil. Karena pembobotan implisit ini, koefisien kuadrat terkecil biasa adalah tidak bias dan konsisten, namun tidak efisien, yaitu varians dugaan bukanlah varians minimum. Di samping itu, varians dugaan dari parameter dugaan adalah penduga-penduga yang bias dari varians yang sebenarnya. Untuk menguji ada atau tidaknya heterokedastisitas, digunakan uji Breusch pagan. Akan tetapi terdapat solusi pemecahan masalah heterokedastisitas yang cukup sederhana, yaitu dengan menggunakan koreksi standar *error* (White Heterocedasticity Consistent

Standard Error & Covariance) dan *weighted least square*, yaitu membobotkan setiap variabel dengan varians yang tidak konstan. Tujuannya adalah agar varians menjadi konstan. Selain itu juga dengan melakukan transformasi model dalam bentuk logaritma natural. Secara lebih lanjut, White (1980) menderivasi sebuah matriks perkiraan kovarian konsisten pada heteroskedastisitas (*heteroskedasticity consistent covariance matrix estimator*) yang memberikan perkiraan yang benar pada koefisien kovarian dalam hadirnya heteroskedastisitas pada bentuk yang tidak diketahui. Matriks kovarian White dilambangkan dengan:

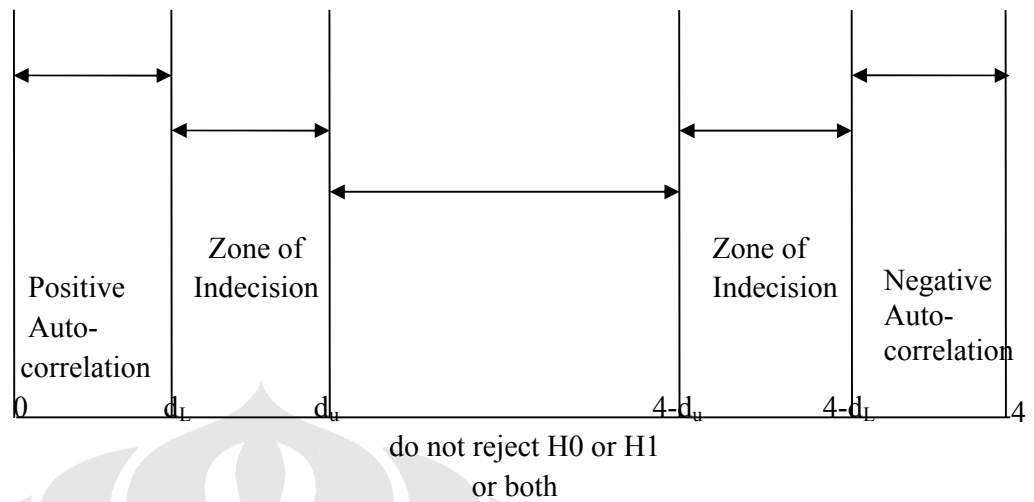
$$\hat{\Sigma}_W = \frac{T}{T-k} (X'X)^{-1} \left(\sum_{t=1}^T u_t^2 x_t x_t' \right) (X'X)^{-1} \quad (3.33)$$

di mana, T merupakan jumlah observasi, k merupakan jumlah regresor, dan u_t merupakan residual kuadrat terkecil.

3. Autokorelasi

Pelanggaran asumsi ini terjadi jika *error* dari observasi yang berbeda berkorelasi, dengan kata lain terjadi korelasi *error* antar waktu. Autokorelasi terjadi jika *error* dari periode waktu yang berbeda adalah berkorelasi. Hal ini biasa terjadi pada jenis data *time series*. Korelasi serial tidak mempengaruhi ketidakbiasaan atau konsistensi penduga kuadrat terkecil biasa, tetapi berpengaruh terhadap efisiensi.

Untuk mengetahui keberadaan autokorelasi dalam suatu model regresi yang menggunakan data *time series*, digunakan uji Durbin Watson d. Kriteria ada atau tidaknya autokorelasi dapat dilihat pada gambar berikut ini:



Gambar 3.4 Kriteria Durbin Watson d

Sumber : Gujarati

Penjelasan dari gambar di atas adalah bahwa suatu model dikatakan tidak memiliki autokorelasi apabila nilai Durbin Watson hasil perhitungan berada diantara d_U sampai $4-d_L$. Suatu model dikatakan memiliki autokorelasi negatif jika nilai DW hasil perhitungan berada pada area lebih besar dari $4-d_L$. Selanjutnya, suatu model dikatakan mempunyai autokorelasi positif jika nilai DW berada pada area $0-d_L$. Autokorelasi tidak dapat dijelaskan keberadaannya apabila nilainya berada pada daerah $4-d_U$ sampai $4-d_L$.

Model yang merupakan model autoregresif sebaiknya menggunakan Durbin Watson h . Karena pada umumnya jika model ini menggunakan Durbin Watson d , angka yang didapatkan akan cenderung mendekati 2.

Cara Penghitungan Durbin Watson h adalah:

$$h = \left(1 - \frac{1}{2}d\right) \sqrt{\frac{T}{1 - T \cdot \widehat{\text{Var}}(\hat{\beta}_1)}}, \quad (3.34)$$

Model tersebut dikatakan tidak memiliki autokorelasi jika hasilnya lebih kecil dari F -value. Akan tetapi terdapat batasan dalam menggunakan Durbin Watson h , yaitu jumlah observasi minimal 30 dan $(1-T \cdot \widehat{\text{Var}}(\hat{\beta}_1))$ lebih besar daripada 0.

3.4.3 Lag

Penggunaan *lag* dalam model ini ditujukan untuk beberapa variabel. Variabel tersebut adalah pengeluaran untuk bidang kesehatan, pendidikan, dan perhubungan. Kegunaan dari *lag* adalah untuk menjelaskan bahwa dibutuhkan waktu agar variabel independen berpengaruh terhadap variabel dependen. Kriteria dalam pemilihan *lag* pada umumnya untuk data *time series* adalah dengan melihat Akaike dan Schwartz Criteria (AIC dan SIC), di mana *lag* dipilih berdasarkan angka AIC dan SIC terendah. Karena model ini menggunakan data panel, penelitian ini hanya akan melihat Adjusted R^2 dan signifikansi pada t-test untuk variabel yang memiliki lag tersebut. *Lag* terbaik adalah pada saat model berada dalam kondisi terbaik (Adjusted R^2 tinggi, t-test signifikan, dan angka Durbin Watson tes menunjukkan tidak adanya autokorelasi).

3.4.4 Interpolasi

Interpolasi merupakan sebuah cara untuk mengetahui nilai sebuah titik di antara dua atau lebih titik pada data. Banyak cara yang dilakukan untuk melakukan interpolasi. Sebagian besar meliputi pencocokan beberapa fungsi data dan evaluasi fungsi titik yang diinginkan. Hal ini tidak berarti ada pengecualian terhadap metode statistik dalam menghitung data yang diinterpolasi.

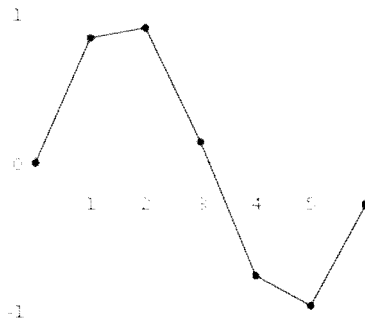
Cara yang paling sederhana adalah dengan menghitung rata-rata aritmatik (*arithmetic mean*) dari nilai dua titik yang berdekatan untuk menemukan nilai tengahnya. Cara ini akan memberikan hasil yang sama dalam bentuk evaluasi fungsi linear pada titik tengah. Perhatikan fungsi berikut ini:

$$f(x_k) = y_k, \quad k = 1, \dots, n \quad (3.35)$$

Pada fungsi terdapat sejumlah n angka yang berurutan, x_k , dan untuk setiap x_k , terdapat sebuah y_k . Pasangan x_k, y_k disebut dengan titik data, dan f merupakan *interpolant* untuk titik data.

Interpolasi linear

Cara paling sederhana dalam melakukan interpolasi adalah interpolasi linear (disebut juga dengan *lerp*). Sebagai contoh, penentuan nilai $f(2.5)$ dari gambar berikut:



Gambar 3.5 Pola Interpolasi Linear

Sumber: www.wikipedia.com

Secara umum, interpolasi linear dari dua titik data, (x_a, y_a) dan (x_b, y_b) akan menghasilkan *interpolant* (x, y) yang dicari dengan menggunakan rumus:

$$y = y_a + (x - x_a) \frac{(y_b - y_a)}{(x_b - x_a)} \quad (3.36)$$

Interpolasi linear mudah dan cepat digunakan, tetapi tidak terlalu tepat. Kekurangan lainnya adalah *interpolant* tidak terdiferensiasi pada titik x_k . Berikut ini adalah contoh dari ketidaktepatan nilai *interpolant*. Fungsi yang ingin diinterpolasi dilambangkan dengan g , nilai x berada antara x_a and x_b , dan g didiferensiasi dua kali secara berkelanjutan. *Error* interpolasi linear yang terjadi adalah:

$$|f(x) - g(x)| \leq C(x_b - x_a)^2 \quad \text{where} \quad C = \frac{1}{8} \max_{y \in [x_a, x_b]} |g''| \quad (3.37)$$

Dengan kata lain, *error* yang didapatkan sama dengan kuadrat jarak antara titik data. *Error* dari metode lain, termasuk *polynomial interpolation* and *spline interpolation* adalah sama dengan jarak antara dua titik dengan nilai lebih besar. Metode ini akan menghasilkan *interpolant* yang lebih halus.

Adapun data yang merupakan hasil interpolasi adalah :

1. Jalan di desa menggunakan interpolasi karena data tiga tahunan
2. Kemiskinan pada sebelum tahun 2000 merupakan data tiga tahunan. Data aslinya adalah pada tahun 1993, 1996, dan 1999, di mana metode

penghitungannya adalah merupakan metode awal, bukan merupakan metode baru yang dipublikasikan pada tahun 1998

3. Data curah hujan tahun 1996 menggunakan interpolasi
4. Persentase sawah beririgasi tahun 2004
5. Persentase elektrifikasi tahun 2005
6. Pengeluaran untuk pendidikan dan kesehatan Propinsi Sumatera Barat pada tahun 2003
7. Jumlah penduduk tiap propinsi hasil interpolasi karena data dari tahun 1990, 1995, 2000, 2005 (Susenas)

3.4.5 Forecasting

Salah satu cara untuk melakukan forecasting adalah dengan metode Autoregressive Moving Average (ARMA). Proses ini dapat dilakukan melalui dua cara, yaitu dengan memberikan *autoregressive* (AR) dan *moving average* (MA). Penentuan angka yang tepat untuk nilai AR dan MA diperlukan *trial and error*. Kriteria untuk menentukan AR dan MA yang terbaik adalah dengan menggunakan tingkat R^2 tertinggi dan AIC dan SIC yang terendah. Pembentukan model AR adalah:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \mu_t + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t \quad (3.38)$$

di mana δ adalah rata-rata dari Y dan u_t adalah *uncorrelated random error term* dengan rata-rata adalah 0 dan varians yang konstan. Tingkatan dari AR dilihat dari p disebut juga dengan AR(p).

Sedangkan untuk MA, modelnya adalah

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3.39)$$

di mana μ adalah konstan dan u adalah *white noise stochastic error term*. Tingkatan dari MA dilihat pada q disebut juga dengan MA(q).

ARMA akan digunakan jika datanya adalah stasioner. Jika tidak stasioner maka digunakan Autoregressive Integrated Moving Average (ARIMA). Penggunaan ARIMA adalah penggunaan ARMA dengan penurunan pertama ($d=1$) maupun kedua ($d=2$). Proses ARIMA tersebut menjadi ARIMA (p,d,q).

Untuk pengujian data apakah stasioner atau tidak digunakan *unit root test*, yaitu Augmented Dickey Fuller Tet (ADF Test). Model dari ADF Test adalah

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_1 \sum \Delta Y_{t-1} + \varepsilon_t \quad (3.40)$$

Jika probabilitasnya kurang dari 5% berarti bahwa data tersebut stasioner. Data yang mengalami *forecasting* adalah TFP Indonesia untuk data dari tahun 2000 hingga 2005.

