

BAHAN AJAR/MODUL

EKONOMETRIKA



**PROGRAM STUDI DOKTOR ILMU EKONOMI
JURUSAN ILMU EKONOMI
FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS MULAWARMAN**

I. Ekonometrika dan Kegunaannya: Peranan Teori Ekonomi, Statistik dan Matematika

Sesungguhnya ekonometrika sering dinterpretasikan sama dengan “*Economic Measurement*.” Meskipun benar pengukuran adalah sesuatu yang penting dalam ekonometrika, namun cakupan ekonometrika lebih luas dari sekedar pengukuran variable dan hubungan ekonomi. Untuk itu, ekonometrika umumnya di definisikan sebagai ilmu sosial dimana peralatan teori ekonomi, matematika dan inferensi statistik secara bersamaan diaplikasikan pada berbagai analisis tentang fenomena ekonomi.

Ekonometrika sebagai kajian (ilmu ekonomi) terbagi dua tipe: *theoretical econometrics* dan *applied econometrics*. Tipe pertama terfokus pada pengembangan metode yang tepat dalam mengukur hubungan ekonomi sesuai dengan kaidah model ekonometrik. Di sini, ekonometrika akan sangat berat terkait dengan matematikal stastistik. Tipe kedua menggunakan berbagai peralatan teori ekonometrika yang dihasilkan tipe pertama untuk mengkaji bidang ilmu ekonomi tertentu seperti aplikasi metode estimasi fungsi produksi, fungsi konsumsi (demand), fungsi supply, fungsi investasi, fungsi permintaan uang dan lainnya.

Intrigalator mengemukakan 3 (tiga) kegunaan penting ekonometrika. **Pertama**, untuk analisis **structural** (analisis kuantitatif tentang hubungan berbagai fenomena ekonomi). Apakah satu atau beberapa variable ekonomi (independent variable) benar memiliki hubungan yang signifikan (berarti) atau bukan hanya sekedar secara insendital mempengaruhi suatu variable ekonomi tertentu yang ingin diamati (dependent variable). Di sini, analisis ekonometrika dapat memberikan perbandingan secara structural berupa suatu besaran koefisien dan sifat (tanda) hubungan suatu variable (negatif/positif) yang diperoleh dari data empirik. Sebagai contoh, ketika secara teoritis Keynes mengemukakan bahwa konsumsi secara proporsional dipengaruhi oleh tingkat pendapatan dengan proporsi antara 0 dan 1 ($0 < mpc < 1$). Maka suatu analisis ekonometrika misalnya dapat memberikan hasil estimasi structural, sebagai suatu pemberian, dari kajian data empirik bahwa konsumsi secara signifikan benar dipengaruhi oleh pendapatan yaitu apabila pendapatan mengalami kenaikan 1 % maka akan menyebabkan konsumsi masyarakat akan meningkat pula sebesar rata-rata 0.8 persen.

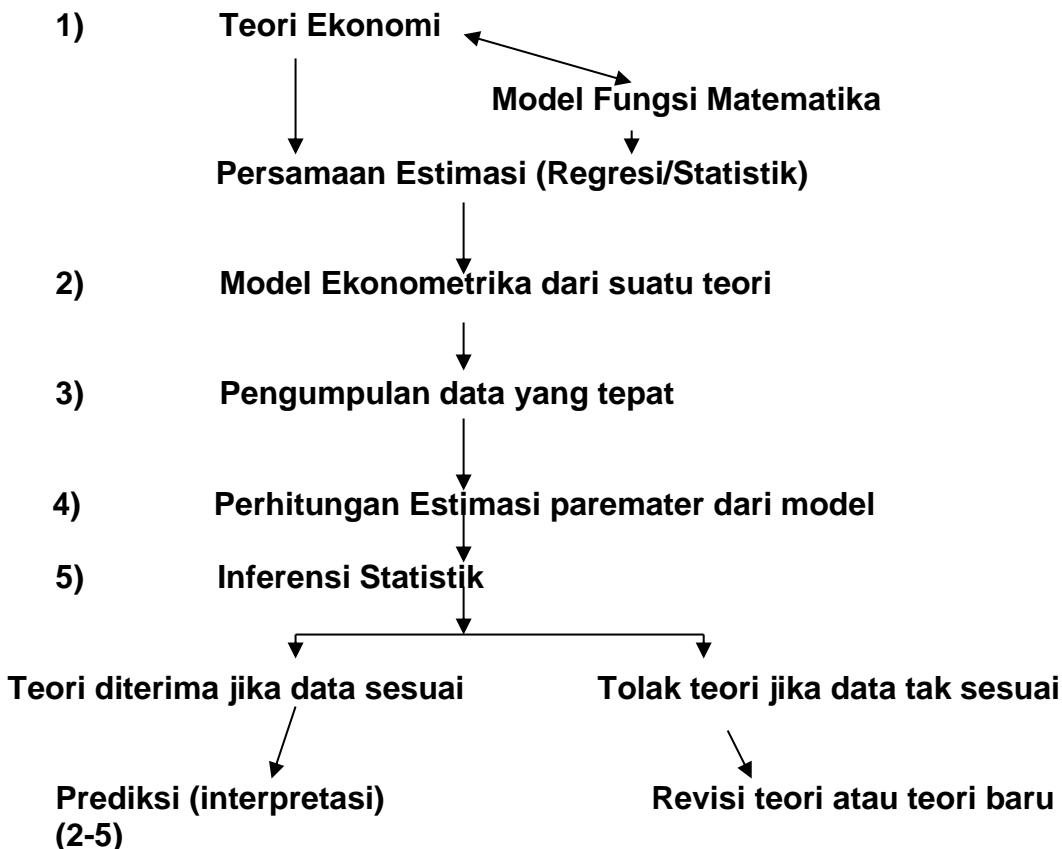
Kedua, untuk **forecasting** yaitu dari hasil analisis ekonometrika akan memberikan koefisien korelasi atau koefisien determinasi (R-square). Apabila R-square tinggi akan berarti bahwa variable independent (variable penjelas) yang digunakan akan memberikan suatu tingkat *predictable power* yang tinggi pula sehingga hasil persamaan estimasi layak untuk digunakan sebagai persamaan Forecasting (peramalan). **Ketiga**, untuk **Policy choosing** yaitu apabila tingkat inflasi misalnya dinyatakan bahwa

dipengaruhi oleh tingkat indeks harga impor, kurs dan jumlah uang beredar dan tingkat rata-rata tariff impor. Maka hasil analisis ekonometrika akan dapat memberikan sifat pengaruh (positif/negatif) variable independent tersebut beserta besaran koefisien masing-masing. Hasil seperti ini tentu akan menyediakan pilihan untuk *policy* kontrol (evaluasi) terhadap tingkat inflasi domestik dengan mengamati variable independent mana yang memiliki pengaruh besar beserta magnitudnya (positif atau negatif) terhadap inflasi domestik tersebut.

Kemudian terdapat beberapa catatan penting dalam memahami suatu hasil analisis ekonometrika sehingga menarik untuk dicatat di sini. **Pertama**, hasil estimasi ekonometrika bukanlah suatu "kebenaran mutlak" dan sangat perlu dipahami bahwa ia dibangun dari suatu model *undeterministic* melalui uji kebenaran yang sangat tergantung pada suatu inferensi statistik dengan menerapkan **metode regresi**. Di sini mungkin perlu dipahami juga adanya perbedaan antara persamaan estimasi ekonometrik (*statistical dependency*) yang undeterministik dengan formulasi **fungsi matematika** (*functional dependency*) yang sifatnya deterministic yaitu adanya penetuan hubungan yang eksak antar variabel. Namun perlu disadari pula bahwa model ekonometrik biasanya harus diawali dengan suatu rumusan fungsi matematika yang eksak berdasarkan pada logika yang ada dalam **teori ekonomi**. Hubungan fungsional matematika yang eksak ini tampak kemudian dapat berubah menjadi *undeterministic* ketika pernyataan logika teori ekonomi dalam bentuk fungsi matematika itu akan diuji dengan empirikal data yang mana data tersebut pada dasarnya merupakan *sampling* dari suatu populasi tertentu. **Kedua**, metode regresi terutama diperuntukkan untuk menunjukkan hubungan (korelasi) yang mana tentu bukan merupakan satu-satunya metode statistik untuk sekedar menunjukkan korelasi antara variable. Metode lain untuk mengamati korelasi adalah *chi-square* misalnya, namun hal ini tergolong statistik non-parametrik sehingga hanya mampu menunjukkan ada tidaknya (signifikansi) korelasi tanpa disertai suatu besaran koefisien dan sifat (tanda) dari hubungan yang terjadi antar variable. Sedangkan analisis regresi tergolong statistik (inferensial) parametric sehingga untuk memahaminya perlu pula pengetahuan tentang bentuk (teori) distribusi (terutama distribusi normal dan asumsi lineritas). Hal lain adalah bahwa regresi sebagai suatu bentuk estimasi statistik lebih menekankan pada *point estimation* dibandingkan dengan *interval estimation*. **Ketiga**, bagaimanapun akuratnya suatu model dan hasil analisis ekonometrika (dimulai dari pernyataan hubungan fungsional matematika sampai pada perhitungan persamaan estimasi regresi dan inferensi statistik), apabila tanpa sebelumnya didasari dengan **logika teori ekonomi yang benar**, maka ia tampak perlu kembali untuk dipertanyakan. Meskipun perlu disadari pula bahwa hasil prediksi ekonometrika dengan suatu data yang tepat, bukannya tidak boleh untuk bertentangan dengan teori ekonomi yang ada.

Untuk jelasnya, prosedur penelitian ekonomi dengan menggunakan model analisis ekonometrika dapat digambarkan melalui skema berikut:

Skema 1. Prosedur Model Analisis Ekonometrika



II. Analisis Regresi antara Dua Variabel (Satu Variable Independent): Methode Ordinary Least Square (OLS)

Ide dasar yang melatar belakangi analisis regresi adalah adanya suatu hubungan statistik (undeterministik) antara variable dependent dan satu atau lebih variable independen. Tujuan dari analisis seperti ini adalah untuk mengestimasi (memprediksi) nilai mean (nilai rata-rata) dependent variable dengan nilai tertentu (*fixed value*) dari satu atau beberapa variable independent (pada bahasan sessi ini dibatasi untuk hanya satu varibel independent). Dengan demikian, analisis regresi adalah suatu studi tentang satu variable dependent (Y) atas satu atau beberapa variable penjelas (X) terutama untuk mengestimasi nilai rata-rata (populasi) dari variable dependent tersebut dari suatu nilai tertentu (fixed and diketahui) sebagai suatu sampling (dapat diulang) dari variable penjelas.

Jadi, *fitted* garis regresi tidak lain adalah suatu nilai ekspektasi rata-rata Y kondisional atas nilai X, atau $E(Y|X)$. Sebagai ilustrasi, mari kita simak Tabel 1 berikut:

Tabel 1. Konsumsi (Y) Menurut kelompok Pendapatan (X) perminggu, dalam US\$

Y (pengeluaran Konsumsi)	Kelompok pendapatan (Xi)									
	80	100	120	140	160	180	200	220	240	260
55	65				10	11	12	135	13	150
60	70	79	80	2	0	0	0	137	7	152
65	74				10	11	13	140	14	175
70	80	84	93	7	5	6	152	5	178	
75	85				11	12	14	157	15	180
...	88	90	95	0	0	0	0	160	5	185
...	...			10	11	13	14	162	16	191
		94	3	6	0	4			5	
			10	11	13	14			17	
		98	8	8	5	5			5	
		11	12	14			18	
		3	5	0			9	
			11	
			5							
total	325	462	455	707	678	750	685	1043	966	1211

Dari Tabel 1, sebagai ilustrasi, diketahui suatu total populasi 60 RT yang dibagi menurut 10 kelompok pendapatan (Xi), diasumsikan kita akan mengestimasi mean (populasi) tingkat konsumsi perminggu (Y) dengan diketahui tingkat pendapatan kelompok RT (X). Dengan demikian, untuk kelompok pendapatan \$80 (tertentu $X= 80$) terdapat 5 RT yang memiliki *range* konsumsi dari 55 sampai \$75 perminggu. Sedangkan untuk $X= \$240$, terdapat 6 RT dengan *range* konsumsi perminggu antara \$137 sampai \$189. Dengan kata lain, terdapat distribusi konsumsi perminggu (Y) korespondensi dengan tingkat tertentu pendapatan RT (X), yaitu terdapat suatu distribusi kondisional Y dengan tertentu nilai X. Selanjutnya, terdapat kondisional probability Y, yaitu $p(Y|X)$ dimana untuk kasus $X=\$80$, terdapat 5 nilai Y: 55, 60, 65, 70 dan \$75. Maka probility diantara siapa saja memiliki tingkat konsumsi Y diantara 5 RT adalah 1/5 atau misalnya $p(Y=55 | X=80) = 1/5$. hal yang sama untuk $P(Y=150 | X=260) = 1/7$. kemudian untuk setiap distribusi kondisional probability Y, kita dapat menghitung rata-rata nilai Y yang kondisional atas nilai X atau $E(Y|X)$ atau lihat pada Table 2.

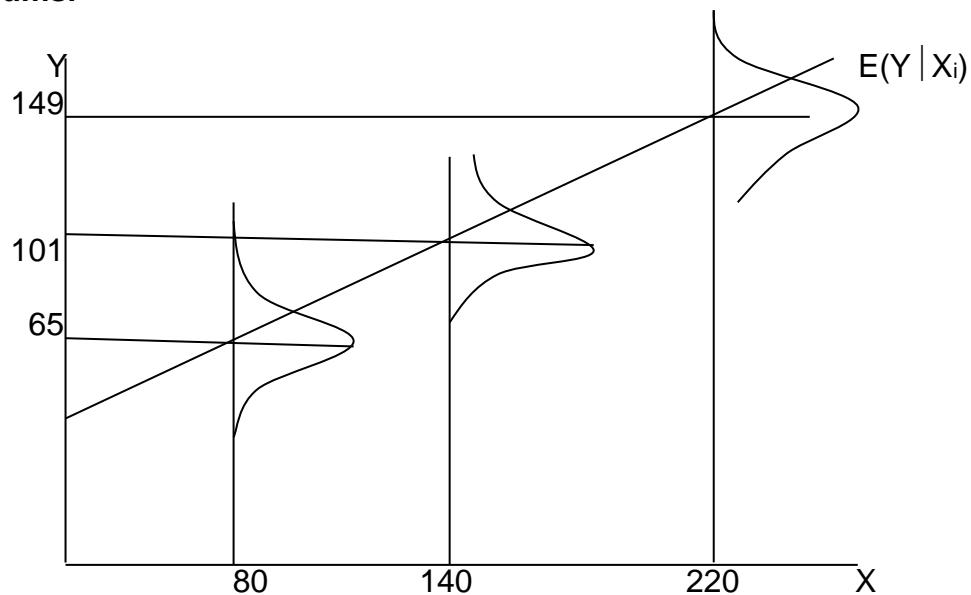
Tabel 2. Kondisional Probility $p(Y | X)$ untuk data dari Tabel 1

$p(Y X)$	Kelompok pendapatan (X_i)									
	80	100	120	140	160	180	200	220	240	260
kondisional probaility	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	...	1/6	1/7	1/6	1/6	1/7	1/6	1/6	1/7
	1/7	1/7	...	1/7
Kondisional mean dari $Y, E(Y X)$.	65	77	89	101	113	125	137	149	161	173

Catatan: $E(Y | X=80) = 55(1/5) + 60(1/5) + 65(1/5) + 70(1/5) + 75(1/5) = 65$.

Secara grafik, garis *fitted regression* (Populasi) dapat dilihat pada Skema 2 berikut:

Skema 2. Garis Fitted Regresi Populasi dari Studi Pengeluaran Konsumsi



Perlu dicatat bahwa ilustrasi di atas hanya merupakan contoh hipotetis. Sesungguhnya *Population Regression Function* (PRF) adalah tidak *observable*, sebab apa yang diperoleh (predictor) dari suatu model ekonometrika tidak lain adalah *Sample Regression Function* (SRF) yang

diperuntukkan untuk mengestimasi suatu garis PRF (Populasi). Untuk itu konsep statistik tentang estimasi dari suatu sampling proses sebagai *predictor* populasi sangat penting dan perlu untuk dipahami.

2.1. Beberapa Catatan tentang Konsep PRF

Merujuk pada contoh hipotetik sebelumnya (terutama pada contoh Skema 2), maka terdapat dua postulat dasar untuk PRF:

- a) Dalam suatu observasi populasi yang berasosiasi dengan suatu proses sampling, terdapat *probability density function (pdf)* Y untuk setiap level X (terdapat hubungan statistik antara Y dan X yang memiliki *mean* dan *variance*).
- b) Mean dari pdf Y memiliki hubungan fungsional dengan X , yaitu $E(Y | X) = \beta_1 + \beta_2 X_i$; dimana $(Y | X_i)$ adalah *random variable*.

Dengan demikian PRF dapat dinyatakan sebagai berikut: $E(Y | X_i) = f(X_i)$. Apabila fungsi $f(\cdot)$ liner, maka $E(Y | X_i) = \beta_1 + \beta_2 X_i$; dimana β_1 dan β_2 adalah parameter. Perlu dicatat bahwa istilah parameter di sini adalah sesuatu nilai yang diperoleh dari populasi, sedangkan statistic (biasanya dengan symbol b_i) adalah sesuatu nilai yang diperoleh dari Sampling untuk maksud sebagai perediktor parameter (populasi); jadi b_1 adalah estimator untuk β_1 , sedangkan b_2 adalah predictor untuk β_2 . Catatan lain adalah tentang lineritas yang mana dimaksudkan di sini adalah liner pada parameter bukan pada variable; jadi bentuk berikut ini adalah juga liner pada parameter, miskipun non-liner pada variable, $E(Y | X_i) = \beta_1 + \beta_2 1/X_i$.

- **Spesifikasi Stokastik PRF: Asumsi Regresi Liner Sederhana tentang μ_i**

Model regresi liner sederhana dapat dinyatakan sebagai berikut:

$$Y_i = \beta_1 + \beta_2 X_i^2 + \mu_i; \text{ dan } E(Y | X_i) = \beta_1 + \beta_2 X_i^2. \text{ maka}$$

$\mu_i = Y_i - E(Y | X_i)$ atau $Y_i = E(Y | X_i) + \mu_i$; dimana μ_i adalah stokastik *error terms* (ganguan stokastik) dan β_1 adalah konstanta dan $\beta_2 = \partial Y / \partial X$ adalah slope dari persamaan regresi (PRF) liner sederhana tersebut.

Asumsi regresi sederhana tentang μ_i untuk metode OLS memenuhi BLUE adalah sebagai berikut:

- a) $E(\mu_i | X_i) = 0 \longrightarrow E(Y | X_i) = \beta_1 + \beta_2 X_i^2$
- c) $Cov(\mu_i, \mu_j) = 0$ untuk semua $i \neq j$. Ini berarti tidak ada serial correlation (autocorrelation) antara μ_i dan μ_j . Dengan kata lain, antara μ_i dan μ_j adalah statistically (stochastically) independent (catatan Covariance sama dengan nol tidak selalu berarti statistically independent, sebaliknya Statistically independent berarti covariance adalah nol).
- d) $Var(\mu_i) = \sigma^2$ untuk semua i . Asumsi ini disebut *homoscedasticity* yaitu pdf Y untuk setiap tingkat X memiliki variance yang sama.
- e) X adalah non-random yaitu memiliki nilai tertentu atau $Cov((\mu_i, X) = 0)$.
- f) Tidak terdapat *specification error* yaitu: menambah variable yang tidak relevan atau mengeluarkan variable yang relevan.

Apabila kelima asumsi di atas dapat terpenuhi maka estimator yang diperoleh melalui metode OLS adalah BLUE.

2.2. Beberapa Catatan tentang Konsep SRF: Sifat dari Hasil Fitted Garis Regresi OLS

Seperti dikemukakan sebelumnya PRF adalah *unobservable* dan yang dapat diamati hanyalah SRF yang mana merupakan suatu cara terbaik untuk mendekati garis regresi PRF.

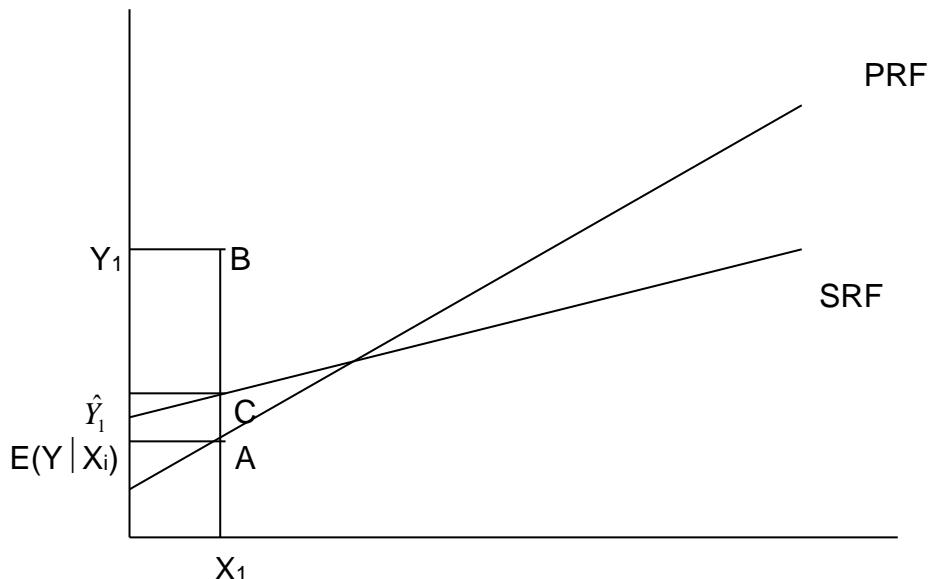
Andaikan PRF dinyatakan $Y_i = \beta_1 + \beta_2 X_i^2 + \mu_i$ dan

SRF dinyatakan $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$ dan $e_i = Y_i - \hat{Y}_i$; maka

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

dimana $\hat{\beta}_1$ dan $\hat{\beta}_2$ adalah estimator β_1 dan β_2 ; sedangkan e_i adalah residual term sebagai estimator μ_i (*stochastic disturbances*). Sekali lagi yang dapat diamati hanyalah SRF sedangkan PRF adalah *unobservable*. Untuk ilustrasi antara fitted garis regresi SRF dan PRF dapat disimak pada Skema 3 berikut.

Skema 3. Ilustrasi Perbandingan antara Garis Regresi PRF dan SRF



AB = Stochastic disturbance $\longrightarrow \mu_1$

AC = AB - BC adalah residual term $\longrightarrow e_1$

Metode estimasi regresi (SRF) dengan OLS pada intinya adalah untuk menemukan garis SRF terbaik untuk mendekati PRF yang tidak dapat diamati. Dengan kata lain, menemukan SRF yang meminimumkan AC (e_1), dimana sekali lagi titik A tidak dapat diamati yang dapat diketahui dari data yang ada adalah titik B (nilai Y_1 dengan tertentu nilai X_1) seperti pada Skema 3 di atas. Metode OLS merupakan suatu cara terbaik yaitu dengan mencoba meminimumkan nilai dari penjumlahan kuadrat e_i atau disebut dengan RSS (*residual sum square*).

$$\begin{aligned} \text{RSS} &= \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \end{aligned}$$

dengan demikian $\text{RSS} = f(\hat{\beta}_1, \hat{\beta}_2)$ dapat diminimumkan yaitu:

$$\partial \text{RSS} / \partial \hat{\beta}_1 = 0; \text{ dan}$$

$$\partial \text{RSS} / \partial \hat{\beta}_2 = 0;$$

dari proses di atas dapat diperoleh persamaan normal dan estimator OLS yang BLUE, $\hat{\beta}_1$ dan $\hat{\beta}_2$ sebagai berikut:

$$\sum Y_i = n \hat{\beta}_1 + \hat{\beta}_2 \sum X_i; \text{ dan}$$

$$\sum X_i Y_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

Dari penyelesaian dua persamaan normal di atas dapat diperoleh:

$$\hat{\beta}_2 = (n \sum X_i Y_i - \sum X_i \sum Y_i) / (n \sum X_i^2 - (\sum X_i)^2); \text{ dan}$$

$$\hat{\beta}_1 = (\sum X_i^2 \sum Y_i - \sum X_i (\sum X_i Y_i)) / (n \sum X_i^2 - (\sum X_i)^2)$$

$$\text{atau } \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}; \text{ dimana } \bar{Y} = Y/N \text{ dan } \bar{X} = X/N$$

- **Sifat-Sifat dari Hasil Fitted Garis Regresi (SRF)**

- a) Garis SRF melalui sample mean dari X dan Y (\bar{X}, \bar{Y}).
- b) Mean dari estimated Y adalah sama dengan \bar{Y} .
- c) Rata-rata (mean) residual term adalah nol
- d) Residual term e_i uncorrelated dengan X_i .
- e) estimated Y (\hat{Y}) uncorrelated dengan residual term e_i .

Referensi:

Gujarati, D., (1984), “Basic Econometric,” International Student Edition, Singapore: McGraw-Hill International Book Company. **Chap. 1-5.**

Ekonometrika

Juliansyah Roy

Program S3 Ilmu Ekonomi

FEB Unmul

Ekonometrika



ilmu sosial dimana peralatan teori ekonomi, matematika dan inferensi statistik secara bersamaan diaplikasikan pada berbagai analisis tentang fenomena ekonomi

PURPOSE OF ECONOMETRICS

1. Specification of Econometric Model or Structural Analysis

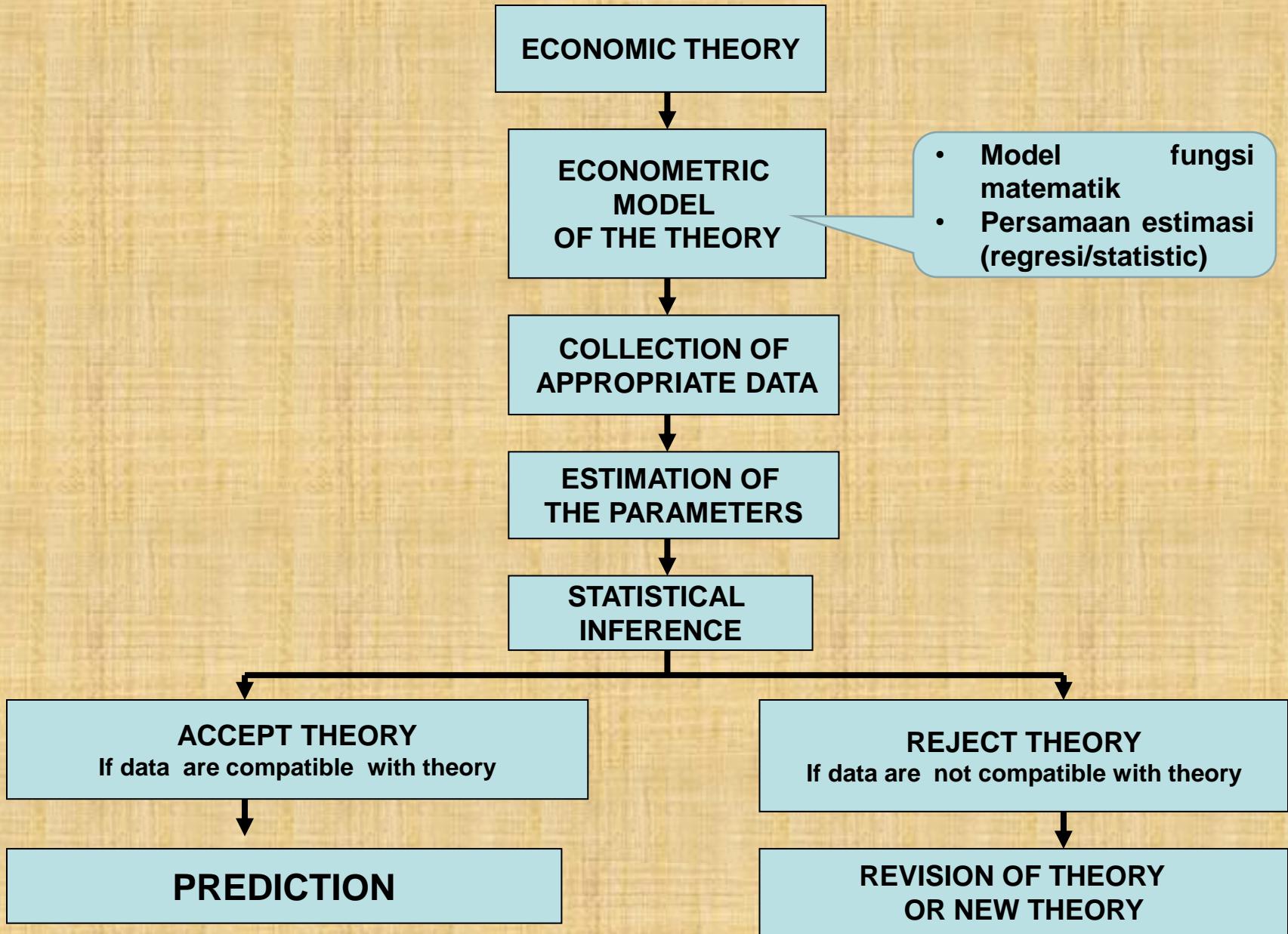
$Y = \alpha + \beta X + u$ model
equation...(1)

$Y = 75 + 0,75 X$ structural model...(2)

2. Forecasting

3. Using the Models for Policy Purposes

METODOLOGI EKONOMETRIKA



Catatan penting memahami hasil analisa ekonomtrika

1. Hasil estimasi ekonometrika bukanlah suatu “kebenaran mutlak” dan sangat perlu dipahami bahwa ia dibangun dari suatu model *undeterministic* melalui uji kebenaran yang sangat tergantung pada suatu inferensi statistik dengan menerapkan **metode regressi**.
2. Metode regresi terutama diperuntukkan untuk menunjukkan hubungan (korelasi) yang mana tentu bukan merupakan satu-satunya metode statistik untuk sekedar menunjukkan korelasi antara variable.

Catatan penting memahami hasil analisa ekonometrika

3. bagaimanapun akuratnya suatu model dan hasil analisis ekonometrika (dimulai dari pernyataan hubungan fungsional matematika sampai pada perhitungan persamaan estimasi regresi dan inferensi statistik), apabila tanpa sebelumnya didasari dengan **logika teori ekonomi yang benar**, maka ia tampak perlu kembali untuk dipertanyakan.

REGRESI LINIER

- Ide dasar yang melatar belakangi analisis regresi adalah adanya suatu hubungan statistik (undeterministik) antara variable dependent dan satu atau lebih variable independen.
- Tujuan dari analisis seperti ini adalah untuk mengestimasi (memprediksi) nilai mean (nilai rata-rata) dependent variable dengan nilai tertentu (*fixed value*) dari satu atau beberapa variable independent (pada bahasan sessi ini dibatasi untuk hanya satu varibel independent).
- Dengan demikian, **analisis regresi adalah suatu studi tentang satu variable dependent (Y) atas satu atau beberapa variable penjelas (X)** terutama untuk mengestimasi nilai rata-rata (populasi) dari variable dependent tersebut dari suatu nilai tertentu (*fixed* dan *diketahui*) sebagai suatu sampling (dapat diulang) dari variable penjelas.

Jadi, *fitted* garis regresi tidak lain adalah suatu nilai ekspektasi rata-rata Y kondisional atas nilai X, atau $E(Y|X)$. Sebagai ilustrasi, mari kita simak Tabel 1 berikut:

Tabel 1. Konsumsi (Y) Menurut kelompok Pendapatan (X) per minggu, dalam US\$

Y (pengeluaran Konsumsi)	Kelompok pendapatan (X_i)									
	80	100	120	140	160	180	200	220	240	260
55	65	79	80	102	110	120	135	137	150	
60	70	84	93	107	115	136	137	145	152	
65	74	90	95	110	120	140	140	155	175	
70	80	94	103	116	130	144	152	165	178	
75	85	98	108	118	135	145	157	175	180	
...	88	113	125	140	160	189	185	
...	115	162	...	191	
Total	325	462	455	707	678	750	685	1043	966	1211

- Dari Tabel 1, sebagai ilustrasi, diketahui suatu total populasi 60 RT yang dibagi menurut 10 kelompok pendapatan (X_i), diasumsikan kita akan mengestimasi mean (populasi) tingkat konsumsi perminggu (Y) dengan diketahui tingkat pendapatan kelompok RT (X)
- Dengan kata lain, terdapat distribusi konsumsi perminggu (Y) korespondensi dengan tingkat tertentu pendapatan RT (X), yaitu terdapat suatu distribusi kondisional Y dengan tertentu nilai X . Selanjutnya, terdapat kondisional probability Y , yaitu $p(Y|X)$ dimana untuk kasus $X=\$80$, terdapat 5 nilai Y : 55, 60, 65, 70 dan \$75. Maka probility tingkat konsumsi Y diantara 5 RT adalah $1/5$ atau $p(Y=55 | X=80) = 1/5$.

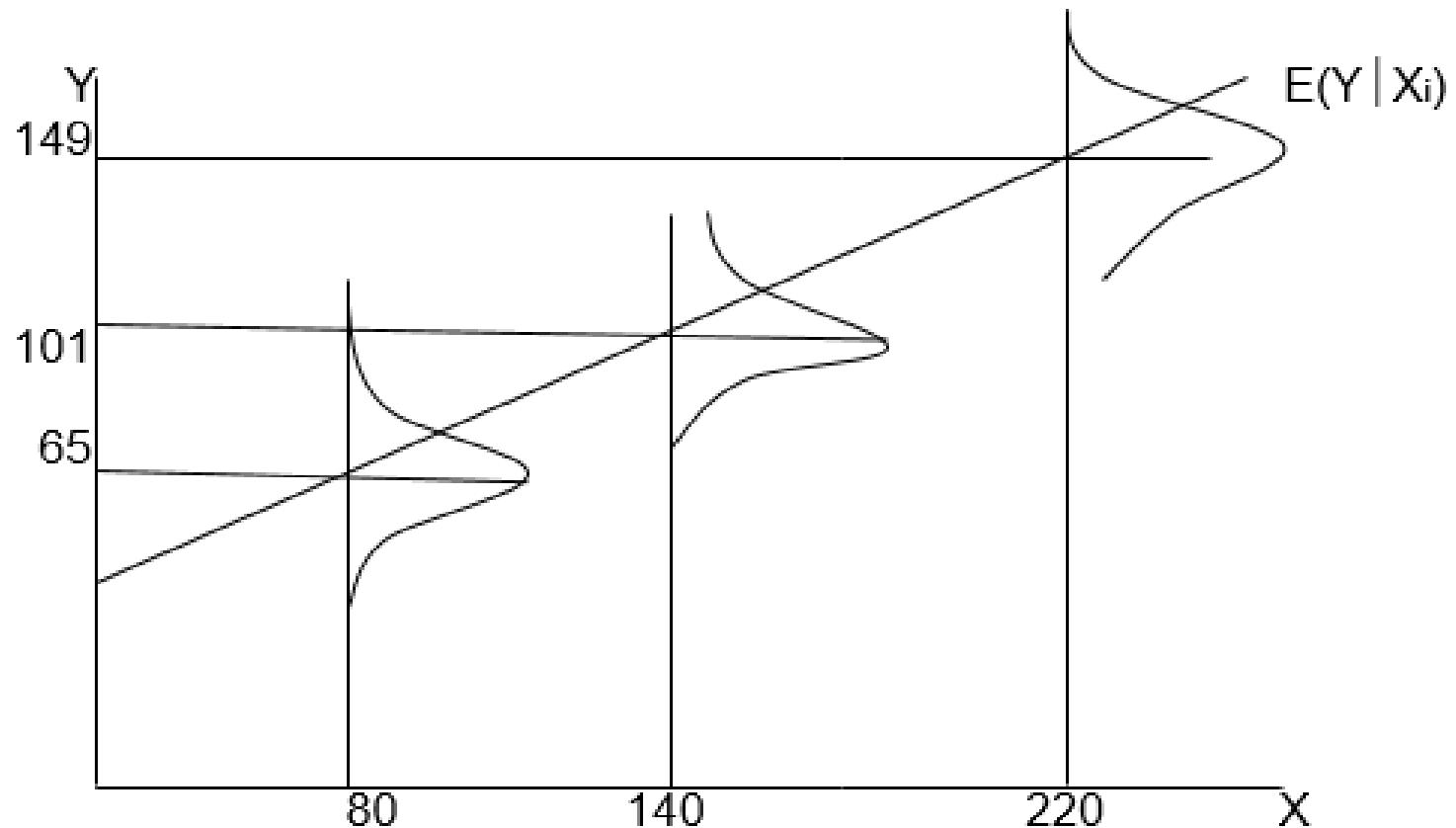
Tabel 2. Kondisional Probability $p(Y|X)$ untuk data dari Tabel 1

$p(Y X)$	Kelompok pendapatan (X_i)									
	80	100	120	140	160	180	200	220	240	260
kondisional probaility	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	...	1/6	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/7	1/7	...	1/7
Kondisional mean dari $Y, E(Y X)$.	65	77	89	101	113	125	137	149	161	173

Catatan: $E(Y|X=80) = 55(1/5) + 60(1/5) + 65(1/5) + 70(1/5) + 75(1/5) = 65$. \square

Secara grafik, garis *fitted regression* (Populasi) dapat dilihat pada Skema 2 berikut:

Skema 2. Garis Fitted Regresi Populasi dari Studi Pengeluaran Konsumsi



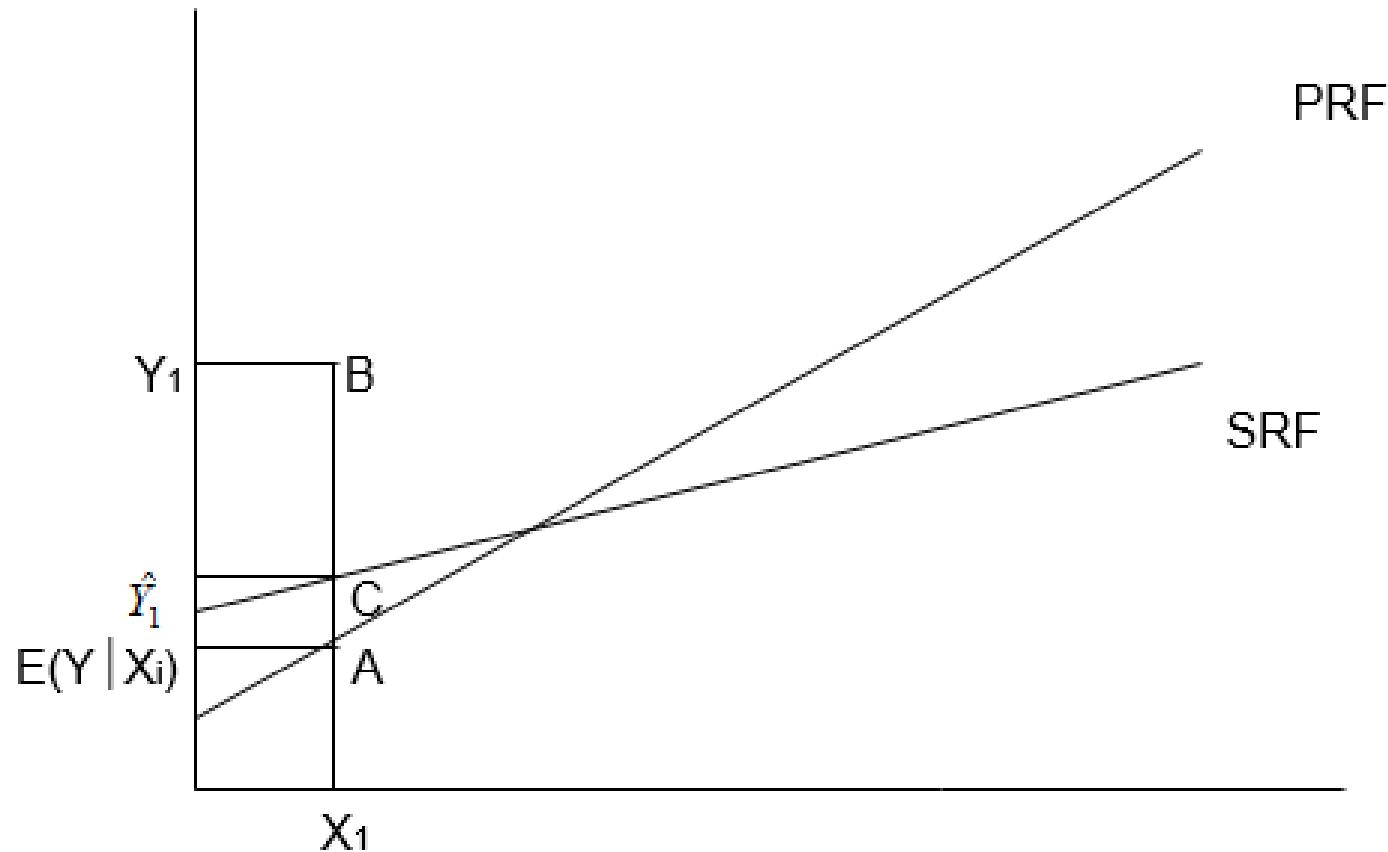
Catatan

- Perlu dicatat bahwa ilustrasi di atas hanya merupakan contoh hipotetis. Sesungguhnya *Population Regression Function* (PRF) adalah tidak *observable*, sebab apa yang diperoleh (*predictor*) dari suatu model ekonometrika tidak lain adalah *Sample Regression Function* (SRF) yang diperuntukkan untuk mengestimasi suatu garis PRF (Populasi). Untuk itu konsep statistik tentang estimasi dari suatu sampling proses sebagai *predictor* populasi sangat penting dan perlu untuk dipahami.

Merujuk pada contoh hipotetik sebelumnya (terutama pada contoh Skema 2), maka terdapat dua postulat dasar untuk PRF:

- Dalam suatu observasi populasi yang berasosiasi dengan suatu proses sampling, terdapat *probability density function (pdf)* Y untuk setiap level X (terdapat hubungan statistik antara Y dan X yang memiliki *mean* dan *variance*).
- Mean dari pdf Y memiliki hubungan fungsional dengan X , yaitu $E(Y|X) = \beta_1 + \beta_2 X_i$; dimana $(Y|X_i)$ adalah *random variable*.

Skema 3. Ilustrasi Perbandingan antara Garis Regresi PRF dan SRF



AB = Stochastic disturbance $\longrightarrow \mu_1$

AC = AB - BC adalah residual term $\longrightarrow e_1$

BENTUK-BENTUK MODEL REGRESI LINIER

1. Model Regresi Linier Klassik (Model Lin-Lin)

- Model dimana baik variabel Y maupun variabel X dalam bentuk linier

$$Y = b_0 + b_1 X + u \quad \dots \dots \dots \quad (3)$$

2. Model Log-Log (Double Log)

- adalah hasil transformasi model tidak linier menjadi model linier.

3. Model Semi Log :

a. Model log-lin

Model dimana variabel Y dalam bentuk log sedangkan variabel X dalam bentuk linier

$$\ln Y = b_1 + b_2 X + u \quad \dots \dots \dots \quad (5)$$

$b_2 = \frac{\text{perubahan relatif dalam Y}}{\text{perubahan absolut dalam X}}$

b. Model Lin-Log

adalah model dimana variabel Y dalam bentuk linier sedangkan variabel X dalam bentuk logaritma

$\beta_1 = \frac{\text{Perubahan absolut dalam } Y}{\text{Perubahan relatif dalam } X}$

3. Model Reciprocal

Model ini menunjukkan bahwa nilai variabel Y adalah kebalikan dari nilai variabel X

$$Y = \beta_0 + \beta_1 1/X + u \dots \dots \dots (7)$$

Kalau X sangat besar maka rata-rata nilai Y mendekati β_0

Contoh:

$$Q = f(\text{AFC})$$

$$Q = \beta_0 + \beta_1 1/\text{AFC} + u$$

REGRESI DAN KORELASI

- Adalah dua metode analisis hubungan antar variabel yang sangat erat keterkaitannya dan tak dapat dipisahkan
- Regresi Sederhana → Model Regresi antar dua variabel
- Korelasi sederhana → Korelasi antar dua variabel
- Regresi Berganda → Regresi Antara tiga variabel atau lebih.
- Korelasi Berganda → Korelasi antara tiga variabel atau lebih

Model Regresi Linier Sederhana

$$\hat{Y} = \alpha + \beta X + \epsilon \quad \dots \dots \dots (8)$$

- Dimana :
- Y = variabel dependen
- X = variabel independen
- ϵ = random error
- α = intercept
- β = slope garis regresi atau koefisien regresi
- Estimasi parameter dilakukan dengan metode OLS
-

METODE PANGKAT DUA (OLS)

- Metode OLS mengestimasi model/garis regresi dengan jalan meminimalkan jumlah dari kuadrat kesalahan setiap observasi terhadap garis tersebut.
- Regresi bertujuan mengestimasi fungsi regresi populasi berdasarkan fungsi regresi sampel.

Gauss-Markov Theorem

Setiap estimator harus memenuhi kriteria BLUE :

- Best = yang terbaik
- Linier = linier function of a random variable of Y
- Unbiased = Rata-rata nilai harapan parameter, $E(b_i)$ sama dengan nilai sebenarnya b_i .
- Efficient estimator = memiliki varians minimum.

Asumsi dasar metode OLS untuk model regresi linier :

1. Linier dalam parameter
2. Nilai rata-rata kesalahan/penyimpangan adalah nol. $E(u_i|x_i) = 0$
3. Varians kesalahan sama untuk setiap periode (setiap x_i) $\text{var.}(u_i|x_i) = \delta^2$
4. Tidak ada autokorelasi antar kesalahan
 $\text{Cov.}(u_i, u_j) = 0$
5. Tidak ada multikolinearitas antar variabel bebas
6. Ada variabilitas dalam nilai x
7. Jumlah Observasi harus lebih besar dari jumlah parameter yang diestimasi

Penggunaan Metode OLS

$$\hat{Y} = \alpha + \beta X + \varepsilon \quad \dots \dots \dots \quad (9)$$

$$\varepsilon = (Y - \hat{Y}) \quad \dots \dots \dots \quad (10)$$

$$\sum \varepsilon^2 = \sum (Y - \alpha - \beta X)^2 \quad \dots \dots \dots \quad (11)$$

Dalam hal ini :

$$\sum \varepsilon^2 = f(\alpha, \beta) \quad \dots \dots \dots \quad (12)$$

- Minimum $\sum \varepsilon^2$ tercapai jika diferensial $\sum \varepsilon^2$ terhadap parameter α dan β sama dengan nol.

$$\frac{\partial \sum \varepsilon^2}{\partial \alpha} = -2 \sum (y - \alpha - Bx) = 0 \quad \dots \dots \dots \quad (13)$$

$$\frac{\partial \sum \varepsilon^2}{\partial \beta} = -2 \sum (y - \alpha - Bx)(X) = 0 \quad \dots \dots \dots \quad (14)$$

- Hasil diferensial tersebut diperoleh dua persamaan berikut :

$$\sum Y = n\alpha + \beta \sum X \quad \dots\dots\dots(15)$$

$$\sum YX = \alpha \sum X + \beta \sum X^2 \quad \dots\dots\dots(16)$$

- Dari persamaan (15) dan (16) diperoleh rumus untuk α dan β sebagai berikut :

- $$\alpha = \frac{\sum X^2 \sum Y - \sum X \sum YX}{n \sum X^2 - (\sum X)^2} \quad \dots\dots (17)$$

- $$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \quad \dots\dots (18)$$

Dengan Metode Pendek diperoleh
Rumus Pendek Sbb :

$$\beta = \frac{\sum xy}{\sum x^2} \quad \dots \dots \dots \quad (19)$$

$$\alpha = \bar{Y} - \beta \bar{X} \quad \dots \dots \dots \quad (20)$$

Dimana :

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

Goodness of Fit suatu model

- Ketepatan RFS sampel untuk menaksir RFP diukur dari Goodness of Fit-nya.
- Nilai statistik t, nilai F dan R^2
- Signifikan, jika nilai statistiknya berada di daerah kritis (daerah penolakan H_0) dan sebaliknya

Koefisien Determinasi (r^2)

- $\sum (Y - \hat{Y})^2 = \sum (\hat{Y} - Y)^2 + \sum \varepsilon^2 \dots \dots \dots (21)$

$$1 = \frac{\sum (\hat{Y} - Y)^2}{\sum (Y - \hat{Y})^2} + \frac{\sum \varepsilon^2}{\sum (Y - \hat{Y})^2} \dots \dots \dots (22)$$

$$\frac{\sum \hat{y}^2}{\sum y^2} + \frac{\sum \varepsilon^2}{\sum y^2} \dots \dots \dots (23)$$

$$= \frac{\sum \hat{y}^2}{\sum y^2} - \frac{\sum \varepsilon^2}{\sum y^2} \dots \dots \dots (24)$$

$$\bullet \quad r^2 = 1 - \frac{\sum \varepsilon^2}{\sum y^2} \dots\dots\dots\dots\dots(25)$$

$$\sum \varepsilon^2 = \sum y^2 - \beta^2 \sum x^2 \dots\dots\dots\dots\dots(26)$$

Dari dua persamaan terakhir diperoleh :

$$r^2 = \frac{(\sum xy)^2}{(\sum x^2)(\sum y^2)} \dots\dots\dots\dots\dots(27)$$

Variance dan Standar Error Parameter

- $\text{Var}(\beta) = \sigma^2 \cdot \frac{1}{(\sum x^2)}$ dimana $\sigma^2 = \frac{\sum \varepsilon^2}{(n - 2)}$ (28)
- Standar error (β) :
- $Se(\beta) = \sqrt{\text{Var}(\beta)}$ (29)
- $\text{Var}(\alpha) = \frac{\sigma^2 (\sum X^2)}{n (\sum x^2)}$ dan $Se(\alpha) = \sqrt{\text{Var}(\alpha)}$ (30)

Pengujian Hipotesis Signifikansi Parameter

- Untuk model regresi $\hat{Y} = \alpha + \beta X$, maka uji t ditentukan :
- $t(\beta) = \frac{\beta}{Se(\beta)}$ dan $t(\alpha) = \frac{\alpha}{Se(\alpha)}$ (40)
- H_0 diterima untuk tiap parameter jika nilai :
 - $t(\beta) < t_{\text{tabel}}$ dan $t(\alpha) < t_{\text{tabel}}$
 -
- H_1 diterima untuk tiap parameter jika nilai :
 - $t(\beta) > t_{\text{tabel}}$ dan $t(\alpha) > t_{\text{tabel}}$

Analisis Kasus

N	Y	X	X ²	XY	y	x	xy	x ²	y ²
1	69	9	81	621	+ 6	0	0	0	36
2	76	12	144	912	+ 13	+ 3	39	9	169
3	52	6	36	312	-11	- 3	33	9	121
4	56	10	100	560	- 7	+ 1	-7	1	49
5	57	9	81	513	- 6	0	0	0	36
6	77	10	100	770	+ 14	+ 1	14	1	196
7	58	7	49	406	- 5	- 2	10	4	25
8	55	8	64	440	- 8	-1	8	1	64
9	67	12	144	804	+ 4	+ 3	12	9	16
10	53	6	36	318	-10	-3	30	9	100
11	72	11	121	792	+ 9	+ 2	18	4	81
12	64	8	64	512	+ 1	-1	-1	1	1
Σ	756	108	1020	6960	0	0	156	48	894

Multiple Linier Regression

- 1. Model Regresi Linier dengan Tiga Variabel

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Estimasi model struktur dilakukan dengan melalui estimasi parameter-parameter β_0 , β_1 dan β_2 menggunakan metode OLS

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\varepsilon = (Y - \hat{Y})$$

$$\begin{aligned}\sum \varepsilon^2 &= \sum (Y - \hat{Y})^2 \\ &= \sum (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)^2\end{aligned}$$

Minimumkan nilai $\sum \varepsilon^2$ maka $d(\sum \varepsilon^2) = 0$

Sedangkan :

$$d(\sum \varepsilon^2) = \partial(\sum \varepsilon^2) / \partial \beta_0 + \partial(\sum \varepsilon^2) / \partial \beta_1 + \partial(\sum \varepsilon^2) / \partial \beta_2$$

- $\partial(\sum \varepsilon^2) / \partial \beta_0 = 2 \sum (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)(-1) = 0 \dots (1)$
-
- $\partial(\sum \varepsilon^2) / \partial \beta_1 = 2 \sum (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)(-X_1) = 0 \dots (2)$
-
- $\partial(\sum \varepsilon^2) / \partial \beta_2 = 2 \sum (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)(-X_2) = 0 \dots (3)$

- $\sum (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$
- $\sum X_1(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$
- $\sum X_2(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$

- $$\Sigma Y = n\beta_0 + \beta_1 \Sigma X_1 + \beta_2 \Sigma X_2)$$
- $\Sigma X_1 Y = \beta_0 \Sigma X_1 + \beta_1 \Sigma X_1^2 + \beta_2 \Sigma X_1 X_2)$
 - $\Sigma X_2 Y = \beta_0 \Sigma X_2 + \beta_1 \Sigma X_1 X_2 + \beta_2 \Sigma X_2^2)$
 - Tiga persamaan normal tersebut diselesaikan dengan pendekatan matriks diperoleh β_0 , β_1 dan β_2 :

- $$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \end{bmatrix}$$

-
- $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}$
- $(B_{ij}) = (X'X)^{-1} X'Y$
- Menyelesaikan matriks diatas diperoleh rumus untuk β_0 , β_1 dan β_2 .

Metode Deviasi

- $y = Y - \bar{Y} ; \quad x = X - \bar{X}$
- $\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
- $\varepsilon = y - \hat{y}$
- $\sum \varepsilon = \sum (y - \hat{y})$
- $\sum \varepsilon^2 = \sum (y - \beta_1 x_1 + \beta_2 x_2)^2$
- Minimumkan $\sum \varepsilon^2$, maka $d(\sum \varepsilon^2) = 0$
- $d(\sum \varepsilon^2) = \partial(\sum \varepsilon^2) / \partial \beta_1 + \partial(\sum \varepsilon^2) / \partial \beta_2 = 0$

- Dari turunan parsil $\partial(\sum \varepsilon^2) / \partial \beta_1$ dan $\partial(\sum \varepsilon^2) / \partial \beta_2$, diperoleh persamaan normal :
- $\Sigma x_1 y = \beta_1 \Sigma x_1^2 + \beta_2 \Sigma x_1 x_2$
- $\Sigma x_2 y = \beta_1 \Sigma x_1 x_2 + \beta_2 \Sigma x_2^2$
- Bentuk matriks dari dua persamaan normal tersebut :
- $$\begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 \\ \Sigma x_1 x_2 & \Sigma x_2^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \Sigma x_1 y \\ \Sigma x_2 y \end{bmatrix}$$

- Penyelesaian terhadap matriks tersebut diperoleh rumus untuk β_1 , dan β_2 :
-
- $$\frac{(\sum x_2^2)(\sum x_1y) - (\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$
- $$\beta_1 = \frac{(\sum x_2^2)(\sum x_1y) - (\sum x_1x_2)(\sum x_1y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$
- $$\beta_2 = \frac{(\sum x_1^2)(\sum x_2y) - (\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

Coefficient of Multiple Determination (R^2_{y,x_1x_2})

- Menunjukkan proporsi variasi perubahan variabel dependen yang ditentukan oleh variabel independen secara simultan

- $R^2_{y,x_1x_2} = \Sigma \hat{y} / \Sigma y = 1 - \Sigma \varepsilon^2 / \Sigma y$

- $\Sigma \varepsilon^2 = \Sigma y^2 - \beta_1 \Sigma x_1 y - \beta_2 \Sigma x_2 y$

- $$= \frac{\Sigma y^2 - (\Sigma y^2 - \beta_1 \Sigma x_1 y - \beta_2 \Sigma x_2 y)}{\Sigma y^2}$$

- $$= \frac{\beta_1 \Sigma x_1 y + \beta_2 \Sigma x_2 y}{\Sigma y^2}$$

The Adjusted Coefficient of Multiple Determination

The adjusted \bar{R}^2 ($= \underline{R}^2$) adalah R^2 yang telah dilakukan penyesuaian dengan derajat bebasnya

$$\underline{\bar{R}^2} = 1 - \frac{\sum \epsilon^2 / (k-1)}{\sum y^2 / (n-k)}$$

n = jumlah observasi
 k = jumlah variabel

Atau :

$$\underline{\bar{R}^2} = 1 - (1 - R^2) \frac{k-1}{n-k}$$

Coeficient of Multiple Correlation ($R_{y.x_1x_2}$)

- Angka yang menunjukkan arah dan kuatnya hubungan simultan antara variabel dependen dan variabel independen dalam model
- $R_{y.x_1x_2} = \sqrt{R^2_{y.x_1x_2}}$

Partial Correlation Coefficient

Menunjukkan korelasi antara dua variabel jika variabel lainnya konstan.

$r_{y1.2}$ = partial correlation coeff. between
y and x_1 , holding x_2 constant

$$r_{y1} - r_{y2} \cdot r_{12}$$

$$r_{y1.2} = \frac{r_{y1} - r_{y2} \cdot r_{12}}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}}$$

Relationship between R², Simple and Partial Correlation Coefficients

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2 r_{y1.} r_{y2.} r_{12}}{1 - r_{12}^2}$$

$$R^2 = r_{y1}^2 + (1 - r_{y1}^2) r_{y2.1}^2$$

$$R^2 = r_{y2}^2 + (1 - r_{y2.}) r_{12.3}^2$$

Variance dan Standar Error Parameter

$$\text{Var} (\beta_1) = \frac{\sum x_2^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \sigma^2$$

$$se (\beta_1) = \sqrt{\text{Var} (\beta_1)}$$

$$\text{Var} (\beta_2) = \frac{\sum x_1^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \sigma^2$$

$$se (\beta_2) = \sqrt{\text{Var} (\beta_2)}$$

$$\sigma^2 = \frac{\sum \varepsilon^2}{n - k} = \frac{\sum y^2 - \beta_1 \sum x_1 y - \beta_2 \sum x_2 y}{n - k}$$

Test of Significance of the Parameter

1) Testing About Individual Significance of Partial Regression Coefficient

A) The Student's test of the null hypothesis

$$t(\beta_i) = \frac{\beta_i}{Se(\beta_i)}$$

Kalau $t(\beta_i) <$ nilai t – tabel, pada d.f = n-k, H_0 diterima, H_1 ditolak. Sebaliknya kalau $t(\beta_i) >$ nilai t – tabel, H_0 ditolak dan H_1 diterima.

B) Standard Error Test

Menekankan pada pengujian dua ekor yang menggunakan 5 % level of significance

- ▶ Kalau $Se(\beta_i) > \frac{1}{2} \beta_i$, maka H_0 diterima, H_1 ditolak.
- ▶ Kalau $Se(\beta_i) < \frac{1}{2} \beta_i$, maka H_0 ditolak, H_1 ditolak.
- ▶ The smaller the standard errors, the stronger is the evidence that the estimates are statistically significant.

2) Testing the Overall Significance of an observed Multiple Regression

$$\begin{aligned} TSS &= ESS + RSS \\ \sum y^2 &= \sum \hat{y}^2 + \sum \varepsilon^2 \end{aligned}$$

$$F = \frac{ESS/(k-1)}{RSS/(n-k)}$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

Derajat bebas = d.f. = $(k-1)/(n-k)$

Worksheet for the regression

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

n	Y	X ₁	X ₂	y	x ₁	x ₂	y ²	x ₁ ²	x ₂ ²	yx ₁	yx ₂	X1x2 0
1	100	5	1000	20	-1	200	400	1	40000	-20	40000	20
2	75	7	600	-5	1	-200	25	1	40	-5	1	20
3	80	6	1200	0	0	400	0	0	160	0	0	0
4	70	6	500	-10	0	-300	100	0	90	0	3	0
5	50	8	300	-30	2	-500	900	4	250	-60	15	100
6	65	7	400	-15	1	-400	225	1	160	-15	6	40
7	90	5	1300	10	-1	500	100	1	250	-10	5	50
8	100	4	1100	20	-2	300	400	4	90	-40	6	60
9	110	3	1300	30	-3	500	900	9	250	-90	15	150
10	60	9	300	-20	3	-500	400	9	250	-60	10	150
Σ	800	60	8000	0	0	0	3450	30	1580	-300	65	590

- $$\beta_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$
- $$(1580.000)(-300) - (-5900)(65.000)$$
- $$\beta_1 = \frac{(30)(1580.000) - (-5900)^2}{(30)(1580.000) - (-5900)^2} = -7,19$$
- $$\beta_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$
- $$(30)(65.000) - (-5900)(-300)$$
- $$\beta_2 = \frac{(30)(1580.000) - (-5900)^2}{(30)(1580.000) - (-5900)^2} = 0,014$$

$$\beta_0 = 80 - (-7,19)(-300) - (0,0143)(800) = 111,69$$

- $(-7,19)(-300)+(0,0143)(65.000)$
- $R^2 = \frac{3450}{365,7} = 0,894$
- $\sum \varepsilon^2 = \sum y^2 - \beta_1 \sum x_1 y - \beta_2 \sum x_2 y = 365,7$
- $\sigma^2 = 365,7 / (10-3) = 52,24$
- $Var(\beta_1) = (1580.000 / 12.590.000)(52,24) = 6,53$
- $Se(\beta_1) = \sqrt{6,53} = 2,55$
- $Var(\beta_2) = (30 / 12.590.000)(52,24) = 0,0001$
- $Se(\beta_2) = \sqrt{0,0001} = 0,01$
- $t(\beta_1) = -2,82 ; t(\beta_2) = 1,28 ; t(0,05) = 2,365$

Regresi non Linear

Regresi Linear Sederhana belajar sendiri

Regresi Non Linear

1. Sederhana (satu variabel)

a. Parabola kuadratik

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \varepsilon_t$$

b. Parabola kubik

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3 X_t^3 + \varepsilon_t$$

c. Eksponen

$$Y = ab^X e^u$$

$$Y = ae^{bX} e^u$$

d. Invers Eksponensial

$$e^Y = ab^X e^u$$

e. Geometrik

$$Y = aX^b e^u$$

f. Logistik

$$Y = e^{a+bX+u}$$

Atau

g. Hiperbola

$$Y = \frac{1}{a + bX + \mu}$$

$$\frac{1}{Y} = a + bX + \mu$$

Bentuk Regresi Non Linear.

Bentuk Logaritma (log-Log), $\ln Y = \ln A + b \ln X$

Bentuk Lin-Log, $Y = \ln A + b \ln X$

Bentuk Log-Lin, $\ln Y = a + bX$

Model Reciprocal

Kurva Philips

Kurva Engel

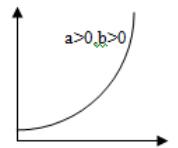
Model linier, log linier, semilog dan hiperbola dapat ditentukan elastisitasnya masing-masing seperti pada Tabel 1 [2].

Tabel 1 : Model Kurva dan Elastisitas

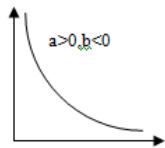
No	Model Kurva	Elastisitas
1	Linier $y_i = a + bx$	$bx/(a + bx)$
2	Log Liner $\ln y = a + b \ln x$	b , untuk semua x
3	Semilog $y = a + b \ln x$	$b/(a + b \ln x)$
4	Hiperbola $y = a + b/x$	$-(b/(ax + b))$

Sumber : David L. Clements (1984)

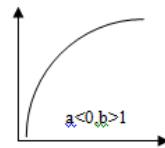
Bentuk log linier $y=ax^b$ atau $\ln y=a+b\ln x$, $b \neq 1$, dengan variasi bentuk seperti dalam Gambar 3, 4, dan 5



Kebutuhan Pokok
(Necessary goods)
Gambar 3



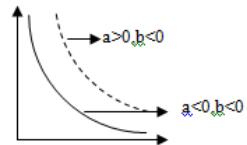
Kurang dibutuhkan
(Inferior goods)
Gambar 4



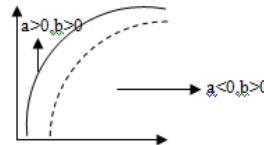
Kebutuhan Mewah
(luxury goods)
Gambar 5

❖ Bentuk Persamaan Semilog

Bentuk semilog $y = a + b \ln x$, $b \neq 0$, $x > 0$ dan $x \neq 1$, mempunyai variasi bentuk kurva Gambar 6 dan 7



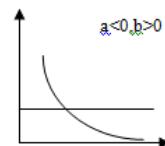
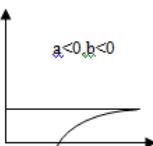
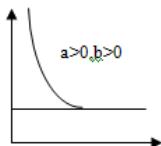
Kurang dibutuhkan
Gambar 6



Kebutuhan pokok/mewah
Gambar 7

❖ Bentuk Persamaan Hiperbola

Bentuk hiperbola $y = a + b/x$, atau $y = a - b/x$, dengan variasi bentuk kurva Gambar 8, 9, dan 10



2. Ganda (lebih dari satu variabel bebas)

a. Fungsi Cobb Douglas

$$Y = A \prod_{i=1}^n X_i^{b_i} e^u ,$$

Atau

$$\ln Y = \ln A + b_i \sum_{i=1}^n \ln X_i + \mu$$

Untuk dua variabel bebas dinyatakan sebagai

$$Y = AK^\alpha L^\beta e^u , \quad \alpha + \beta = 1 \quad (\text{bentuk asli fungsi Cobb Douglas})$$

b. Fungsi CES (Constant elasticity of substitution)

Fungsi produksi CES dikembangkan oleh Arrow, Chenery, Minhas dan Solow yang pada tahun 1961 menulis artikel dengan judul: Capital-labor Substitution and economic efficiency. Fungsi produksi CES dinyatakan

$$Y = A[\alpha L^{-\beta} + (1-\alpha)K^{-\beta}]^{-1/\beta}$$

dimana:

Q = output/ jumlah produksi

K = kapital

L = tenaga kerja

α adalah parameter yang menunjukkan bobot atau intensitas dari modal dan tenaga kerja $0 < \alpha < 1$, A adalah parameter efisiensi dan β adalah parameter substitusi dan bukan elastisitas substitusi ($=\sigma$), karena

$$\beta = \frac{1}{(1+\sigma)} , \quad \text{dan} \quad -1 < \beta < \infty$$

a. Fungsi Translog (transcendental logarithmic function)

$$Y = A \prod_{i=1}^n X_i^{b_i} e^{b_{i+1}} (0,5) \sum_{i=1}^n \sum_{j=1}^m \ln X_i \ln X_j$$

Atau
Atau

$$\ln Y = \ln A + b_i \sum_{i=1}^n \ln X_i + (0,5)b_{i+1} \sum_{i=1}^n \sum_{j=1}^m \ln X_i \ln X_j$$

Untuk dua variable bebas diperoleh:

$$Q = B K^{\beta_1} L^{\beta_2} e^{0,5 \{(\beta_3 (\ln K)^2 + \beta_4 (\ln L)^2 + \beta_5 (\ln K \ln L)\}}$$

Atau

$$\ln Q = \ln B + \beta_1 \ln K + \beta_2 \ln L + 0,5 \beta_3 (\ln K)^2 + 0,5 \beta_4 (\ln L)^2 + 0,5 \beta_5 (\ln K) (\ln L)$$

(Intriligator, 1980; 280)

Soal Latihan :

Soal-soal

1. Apa yang dimaksud dengan fungsi Cobb-Douglas, jelaskan munculnya nama tersebut. Jelaskan sejarah perkembangan fungsi produksi sampai dengan Fungsi produksi Translog (transcendental logarithmic function)
2. Apa yang dimaksud dengan fung CES. Bandingkan dengan fungsi Cobb-Douglas sehingga nampak kelebihan dan kekurangan masing-masing.
3. Diketahui fungsi Cobb-Douglas

$$Y = AX_1^\alpha X_2^\beta, \text{ tunjukkan bahwa } \alpha \text{ dan } \beta \text{ adalah elastisitas untuk } X_1 \text{ dan } X_2$$

4. Sebuah perusahaan memproduksi mentega, Y (kg) dengan tenaga kerja terampil, X (orang) sebagai berikut:

X	0	1	2	3	4	5	6	7	8
Y	3	9	14	16	18	17	15	12	6

- a. Buat persamaan regresi non liner dari fungsi :
 - i). $Y = AX_i^{b_i} e^u$
 - ii). $Y = a_0 + a_1 X + a_2 X^2 + e$
- b. Uji pengaruh tenaga kerja terhadap produksi pada taraf keyakinan $\alpha = 0,05$
- c. Tentukan nilai Y pada $X = 20$ dan juga $X = 6,5$.
- d. Gambarkan kedua fungsi yang diperoleh.

Regresi Berganda dgn Variabel Dummy

A. Regresi Linear

Ganda (lebih dari satu variable bebas)

Regresi Liner Berganda (dua Variabel bebas) dengan Variabel Boneka

Bentuk:

1. No Interaction

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 D + e$$

2. Interaction

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 D + \beta_4 DX_1 + \beta_5 DX_2 + e$$

Dimana:

C_t Konsumsi pada tahun ke t

Y_t adalah pendapatan pada tahun ke t

P_t adalah harga pada tahun ke t

D adalah Otonomi Daerah (D=0 sebelum OTDA dan D= 1 masa OTDA)

Sebelum OTDA D= 0

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 P_t + e$$

Masa OTDA D= 1

$$C_t = (\beta_0 + \beta_3) + \beta_1 Y_t + \beta_2 P_t + e$$

B. Dengan Interaksi

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 P_t + \beta_3 D + \beta_4 DY_t + \beta_5 DY_t + e$$

Sebelum OTDA, D = 0

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 P_t + e$$

Masa OTDA, D= 1

$$C_t = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) Y_t + (\beta_2 + \beta_5) P_t + e$$

Contoh Data Hipotesis:

C	Y	P	D	DY	DP
86	98	5,33	0	0	0
96	104	5,5	0	0	0
103	179	5	0	0	0
116	156	6	0	0	0
132	176	6,5	0	0	0
242	298	7	1	298	7
248	354	7,12	1	354	7,12
258	318	7,32	1	318	7,32
267	338	7,61	1	338	7,61
269	396	8,75	1	396	8,75
272	571	8,79	1	571	8,79

A. Regresi non Linear dengan Vriabel Dummy

1). Satu variabel Bebas

$$Q = aL^{\beta_1} e^{\beta_2 D + \beta_3 \ln X + u}$$

Atau

$$\ln Q = \ln a + \beta_1 \ln L + \beta_2 D + \beta_3 D \ln L + u$$

D=0

$$\ln Q = \ln a + \beta_1 \ln L + u$$

atau

$$Q = aL^{\beta_1} e^u$$

Untuk D=1

$$\ln Q = \ln a + \beta_1 \ln L + \beta_2 + \beta_3 \ln L + u$$

$$\ln Q = (\ln a + \beta_2) + (\beta_1 + \beta_3) \ln L + u$$

$$Q = (ae^{\beta_2})L^{(\beta_1 + \beta_3)}e^u$$

2). Dua Variabel Bebas

$$Q = aK^{\beta_1}L^{\beta_2}e^u$$

Dimana:

Q = Produksi

K = Kapital

L = Tenaga Kerja

Contoh Data Hipotetis:

Y	N	U	lnY	LnU	D	DlnN	D lnU
86	98	42					
96	104	42					
103	179	43.5					
116	156	44					
132	176	44					
242	298	44.5					
248	354	44.5					
258	318	45					
267	338	45					
269	396	50					
272	571	50					

Tentukan persamaan regresinya.

C. Persamaan Regresi dengan varaiel dummy (lebih dari satu)

1. Satu variabel bebas (variabel benaran)

$$Y = f(X, D) \dots \dots \dots \dots \dots \dots \dots \quad (1)$$

Persamaan (1) dapat dibuat dalam bentuk persamaan regresi linear berganda (dengan variabel boneka) dengan formulasi:

$$Y = a_0 + a_1X + a_2D_1 + a_3D_2 + a_4D_3 + a_5D_1X + a_6D_2X + a_7D_3X + \varepsilon \dots \dots \dots \dots \dots \dots \dots \quad (2)$$

Dimana:

Y = konsumsi listrik di Kalimantan Timur (MWh)

X = PDRB dengan harga berlaku (trilyun rupiah)

D_i = Dummy variable (variabel boneka), $i = 1,2,3,4$

Kelompok

Konsumsi listrik Rumah tangga

Konsumsi listrik Usaha

Konsumsi listrik untuk Industri

Konsumsi listrik untuk Umum

1. Persamaan Regresi untuk Rumah Tangga

$D_1 = 1$ adalah konsumsi listrik Rumah tangga

0 yang lain

$$Y = (a_0 + a_2) + (a_1 + a_5) X_1 + \varepsilon \quad \dots \dots \dots \quad (3)$$

2. Persamaan Regresi untuk Usaha

$D_2 = 1$ adalah konsumsi listrik usaha

0 yang lain

$$Y = (a_0 + a_3) + (a_1 + a_6) X_1 + \varepsilon \quad \dots \dots \dots \quad (4)$$

3. Persamaan Regresi untuk Industri

$D_3 = 1$ adalah konsumsi listrik untuk industri

0 yang lain

$$Y = (a_0 + a_4) + (a_1 + a_7) X_1 + \varepsilon \quad \dots \dots \dots \quad (5)$$

4. Persamaan Regresi untuk Umum

$D_1 = 0$

$D_2 = 0$

$D_3 = 0$

$$Y = a_0 + a_1 X_1 + \varepsilon \quad \dots \dots \dots \quad (6)$$

Persamaan Regresi

2. Dua Varaibel bebas (benaran)

(Statistical Program for Social Science) diperoleh Persamaan Regresi Linear Berganda dengan variabel boneka sebagai

$$\begin{aligned}
 Y &= 0,0125 + 0,0191 X_1 + 0,00026X_2 - 0,265 D_1 - 0,093 D_2 - 0,446 D_3 \\
 t_h : & (0,064) (0,219) (0,784) (-0,957) (-3,335) (-1,610) \\
 \text{Sig} : & (0,949) (0,829) (0,442) (0,350) (0,741) (0,123) \\
 & + 0,183 D_1X_1 + 0,042 D_2X_1 + 0,238D_3X_1 - 0,0029 D_1X_2 - 0,0013 D_2X_2 \\
 & (1,483) (0,338) (1,927) (6,108) (2,723) \\
 & (0,154) (0,739) (0,068) (0,000) (0,013) \\
 & - 0,000012 D_3X_2 (7) \\
 & (-0,026) \\
 & (0,979)
 \end{aligned}$$

Angka-angka dalam kurung, atas menunjukkan t hitung dan bawah signifikansi.

$$Y = 0,056 + 0,0003 X_2 - 0,309 D_1 - 0,472 D_3 + 0,202 D_1X_1$$

$$\begin{aligned}
 t_h : & (6,889) (2,086) (-1,701) (-3,211) (2,502) \\
 \text{Sig} : & (0,000) (0,048) (0,102) (0,004) (0,020) \\
 & + 0,247 D_3X_1 + 0,0029 D_1X_2 + 0,0014 D_2X_2 (8) \\
 & (4,035) (8,357) (11,253) \\
 & (0,000) (0,000) (0,000)
 \end{aligned}$$

(Angka-angka dalam kurung, atas menunjukkan t hitung dan bawah signifikansi).

$$F_{\text{hitung}} = 349,626 \quad \text{Sig.} = 0,0000$$

$$R^2 = 0,990$$

$$R = 0,995$$

1. Persamaan Regresi untuk Rumah Tangga

$D_1 = 1$ adalah konsumsi listrik rumah tangga
 0 yang lain

$$\begin{aligned}
 Y &= 0,365 + 0,202 X_1 + 0,0032 X_2 (9) \\
 \text{Sig.} & (0,102) (0,020) (0,048)
 \end{aligned}$$

2. Persamaan Regresi untuk Usaha

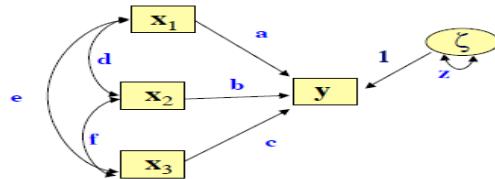
Kuliah 3

Sumber: 1. Giorgio Russolillo
2. Saleh

Path Analysis (Dummy, Mediation, Moderating, Laten Variables)

Tracing rule #4: an example

You want to estimate $\text{cov}(y, x_1)$ from this model:



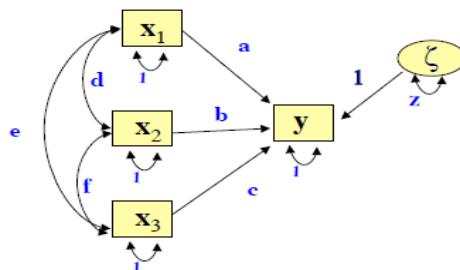
Permissible paths are a, ec, db , but the path $adfc$ is not allowed

Obtaining a numerical solution for model parameters

The multiple regression model (standardized variables) :

$$y = ax_1 + bx_2 + cx_3 + \zeta$$

can be “drawn” by using a Path Diagram:

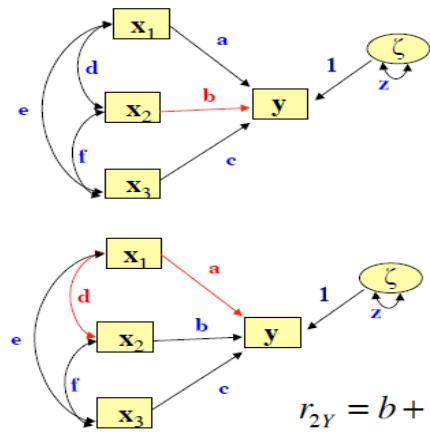


DATA: correlation matrix				
	X1	X2	X3	Y
X1	1.00	0.2	0.24	0.7
X2	0.20	1.0	0.30	0.8
X3	0.24	0.3	1.00	0.3
Y	0.70	0.8	0.30	1.0
				f

- a, b and c are standardized partial correlation coefficients (path coefficients) to estimate
- z is the residual variance to estimate
- d, e and f are covariances (correlations) between the exogenous variables.

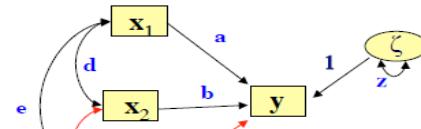
Obtaining a numerical solution for path coefficients (2)

Writing the set of paths for $r(x_2, y)$



DATA: correlation matrix

	X1	X2	X3	Y
X1	1.00	0.2	0.24	0.7
X2	0.20	1.0	0.30	0.8
X3	0.24	0.3	1.00	0.3
Y	0.70	0.8	0.30	1.0



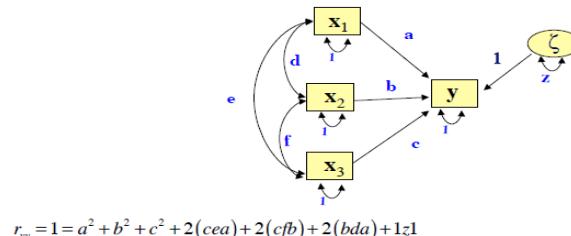
$$r_{2Y} = b + da + fc \rightarrow 0.80 = b + 0.2a + 0.3c$$

$$r_{1Y} = a + ec + db \rightarrow 0.70 = a + 0.24c + 0.2b$$

$$r_{3Y} = c + fb + ea \rightarrow 0.30 = c + 0.3b + 0.24a$$

$$\begin{cases} 0.70 = a + 0.24c + 0.2b \\ 0.80 = b + 0.2a + 0.3c \\ 0.30 = c + 0.3b + 0.24a \end{cases} \Rightarrow \begin{cases} a = 0.57 \\ b = 0.70 \\ c = -0.05 \end{cases}$$

Variance of the error



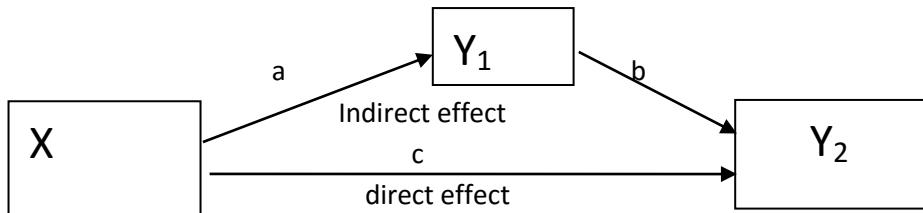
$$r_{yy} = 1 = a^2 + b^2 + c^2 + 2(cea) + 2(cfb) + 2(bda) + 1z1$$

$$\begin{aligned} z &= r_{yy} - [1a1 + 1b1 + 1c1 + 2(cea) + 2(cfb) + 2(bda)] \\ &= 1 - [a^2 + b^2 + c^2 + 2(cea) + 2(cfb) + 2(bda)] \\ &= 1 - [0.57^2 + 0.70^2 + (-0.05)^2 + 2(-0.05)(0.24)(0.57) + 2(-0.05)(0.30)(0.70) + 2(0.70)(0.20)(0.57)] = \\ &= 1 - 0.946 = 0.054 \end{aligned}$$

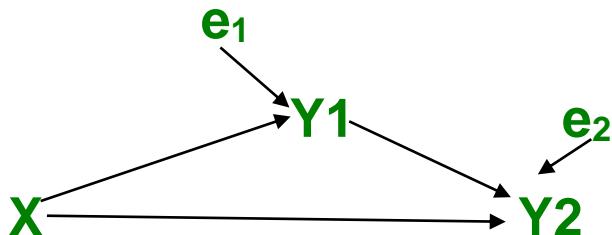
Mediation Model

Direct, Indirect and Total Effects

Conceptual model



analysis model



Unstandardized coefficient

$$Y_1 = a_0 + aX + e_1 \quad \text{a= direct effect}$$

$$Y_2 = c_0 + cX + bY_1 + e_2$$

atau

$$Y_2 = c_0 + cX + b(a_0 + aX + e_1) + e_2$$

$$Y_2 = c_0 + cX + b(a_0 + aX + e_1) + e_2$$

$$Y_2 = c_0 + cX + ba_0 + abX + be_1 + e_2$$

$$Y_2 = (c_0 + ba_0) + (c + ab)X + (be_1 + e_2)$$

ab= indirect effect

(c+ab)= indirect effect

Standardized coefficient

$$Y_1 = aX + e_1$$

$$Y_2 = cX + b(aX + e_1) + e_2$$

$$Y_2 = (c + ab)X + (be_1 + e_2)$$

Dummy with Interaction (Moderator Variables)

Basic Definitions

A moderator variable is a variable involved in an interaction with another variable in the model such that the effect of the other variable depends upon the value of the moderator variable, i.e., the effect of the other variable changes depending on the value of the moderator.

Moderated mediation occurs when a moderator variable interacts with a mediator variable such that the value of the indirect effect changes depending on the value of the moderator variable. This is known as a conditional indirect effect, i.e., the value of the indirect effect is conditional on the value of the moderator variable.

- **Moderator variables** - "In general terms, a moderator is a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable. Specifically within a correlational analysis framework, a moderator is a third variable that affects the zero-order correlation between two other variables. ... In the more familiar analysis of variance (ANOVA) terms, a basic moderator effect can be represented as an interaction between a focal independent variable and a factor that specifies the appropriate conditions for its operation."

Another way to think about this issue is that a moderator variable is one that influences the strength of a relationship between two other variables, and a mediator variable is one that explains the relationship between the two other variables.

We begin with a linear causal relationship in which the variable **X** is presumed to cause the variable **Y**. A moderator variable **M** is a variable that alters the strength of the causal relationship. So for instance, psychotherapy may reduce depression more for men than for women, and so we would say that gender (**M**) moderates the causal effect of psychotherapy (**X**) on depression (**Y**). Most moderator analysis measure the causal relationship between **X** and **Y** by using a regression coefficient. Although classically, moderation implies a weakening of a causal effect, a moderator can amplify or even reverse that effect. Complete moderation would occur in the case in which the causal effect of **X** on **Y** would go to zero when **M** took on a particular value. The reader might consult papers by Kraemer and colleagues (2001; 2002) for a related but somewhat different approach to defining and testing of moderators. Frazier, Tix, and Barron (2004) provide a very good introduction to the topic of moderation and Marsh, Hau, Wen, Nagengast, and Morin. (2011) for a more detailed discussion of the topic.

A moderation analysis is an exercise of external validity in that the question is how universal is the causal effect.

A key part of moderation is the measurement of **X** to **Y** causal relationship for different values of **M**. We refer to the effect of **X** on **Y** for a given value of **M** as the *simple effect **X** on **Y***.

Deciding which variable is the moderator depends in large part on the researcher's interest. For the earlier example in which gender moderates the effect of psychotherapy, if one was a gender researcher, one might say that psychotherapy moderates the effect of gender.

Causal Assumptions

If **X** is a randomized variable, there is no causal ambiguity. Uncertainties arise when **X** is not randomized. If **X** is not manipulated, then the direction of causation must be assumed based on theory or common sense. As shown in Judd and Kenny (2010), it is even possible that the moderator effect can reverse if the direction of causation is flipped (presuming that **Y** causes **X** instead of vice versa). In a moderator analysis, if **X** is not manipulated, the researcher needs to justify the choice of causal direction.

Timing of Measurement

Ideally the moderator should be measured prior to variable **X** being measured. So if **X** is manipulated, then **M** should be measured prior to **X** being manipulated. Of course,

if **M** is a variable that does not change (e.g., race), the timing of its measurement is less problematic. It is possible, but quite complicated, but **M** can be both a mediator and a moderator (see Kraemer et al. (2001) for a different point of view).

Moderator and Causal Variable Relationship

If **X** is a manipulated variable, in principal, there should be no relationship between **X** and **M**. If **X** is not randomized, it might be correlated with **M**. Unlike mediation, there is no need for **X** and **M** to be correlated and that correlation has no special interpretation. However, if **X** and **M** are too highly correlated, there can be estimation problems.

Measurement of Moderation

Generally, moderator effects are indicated by the interaction of **X** and **M** in explaining **Y**. The following multiple regression equation is estimated:

$$Y = i + aX + bM + cXM + E \quad (1)$$

The interaction of **X** and **M** or coefficient **c** measures the moderation effect. Note that path **a** measures the *simple effect* of **X**, sometimes called the *main effect* of **X**, when **M** equals zero. As will be seen, the test of moderation is not always operationalized by the product term **XM**. Given Equation 1, the effect of **X** on **Y** is **a** + **cM**. Thus, the effect of **X** on **Y** depends on the value of **M**. It is noted that the effect of **X** on **Y** equals zero when **M** equals **-a/c**, which may or may not be a plausible value of **M**.

Alternative Interpretations of Moderator Effects

Finding that **c** is statistically significant does not prove moderator effects. One major worry is non-additivity. Consider the case in which the relationship between **X** and **Y** is nonlinear. For instance, **X** is income and **Y** is work motivation. Imagine that the relationship between the two is nonlinear such that if **X** is small the relationship is larger than when **X** is large. If age were tested as a moderator the income-motivation relationship, then because younger workers make less money, we would find the “moderator” effect, that the income-motivation relationship is stronger for younger than for older workers.

Another worry is the actual moderator may not be the moderator but some other variable with which the moderator correlates. For instance, if we find that gender is a moderator, the real moderator might be height, masculinity-femininity, expectations of others, or income. Unless the moderator is a manipulated variable, we do not know whether it is a “true” moderator or just a “proxy” moderator.

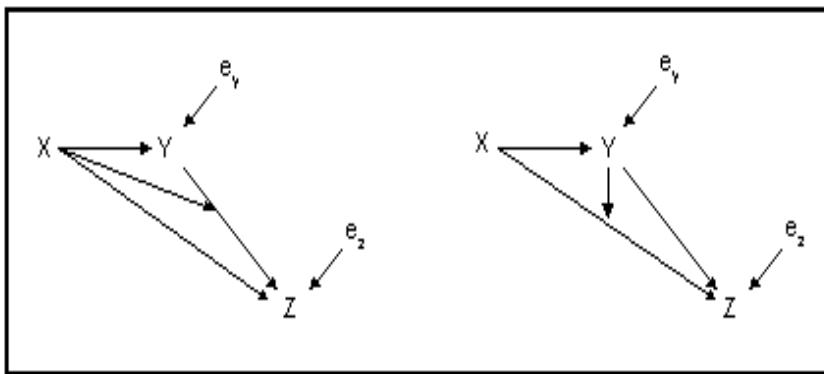
Level of Measurement of the Variables

It is presumed here and throughout that the outcome variable is measured at the interval level of measurement. Should the outcome be a dichotomy, logistic regression would need to be used (Hayes & Matthes, 2009).

The remainder of the page is organized around the levels of measurement of the moderator and the causal variable. The causal variable, **X**, can either be categorical (typically a dichotomy) or a continuous variable. So for instance, **X** might be psychotherapy versus no psychotherapy (a dichotomy) or it might be the amount of psychotherapy (none, one month, two months, or six months; a continuous variable). Much in the same way, the moderator or **M** can be either categorical (e.g., gender) or continuous (e.g., age). Readers are encouraged to read the next two sections, even if they are more interested in one of the other cases, as many key concepts in mediation are discussed there.

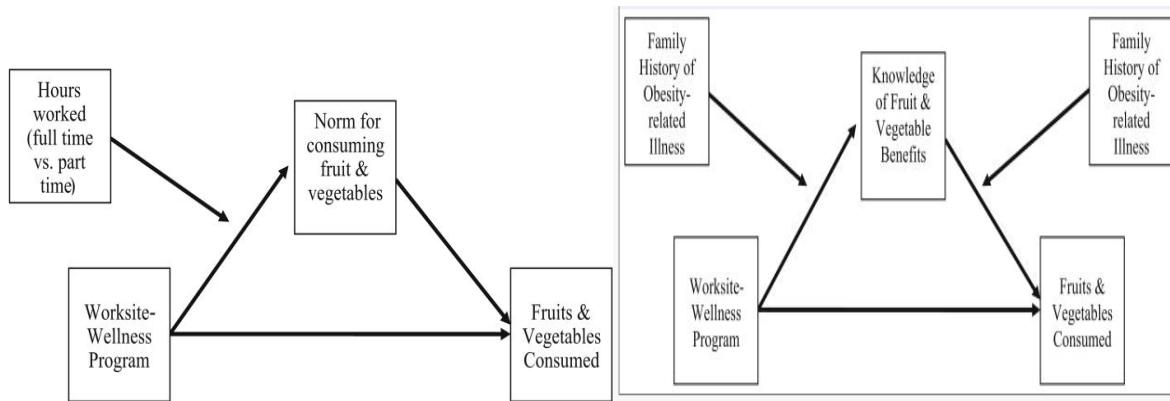
Interaction

A statistical interaction is said to exist when the effect of an independent variable on a dependent variable differs across levels of a third (or control) variable. Suppose, for example, the variable **Z** has two levels. We calculate the association between **X** and **Y** for **Z=1**, and separately calculate the association between **X** and **Y** for **Z=2**. If the two "parts" of the association between **X** and **Y**, controlling **Z**, differ, then statistical interaction exists. There is no single standard way of representing interaction in causal diagrams. One method shows an arrow going from the "control" variable to the effect connecting an independent to a dependent variable. Here are two examples:

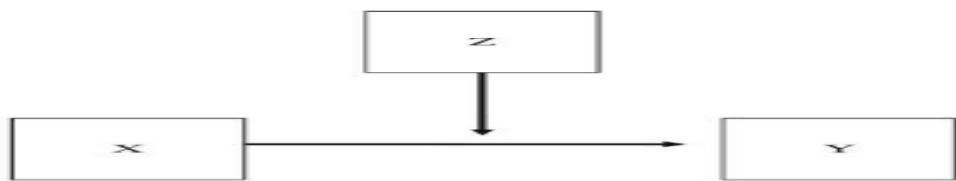


In the panel at the left, the hypothesis states that (in addition to direct and indirect effects) the effect of Y on Z differs across, or depends on the level of X. In the panel at the right, the hypothesis is that the effect of X on Z depends on the level of Y. Note that, in both examples, direct effects of each of the independent variables are shown in addition to their interaction. This is not strictly necessary, as effects may be hypothesized to exist only in the presence of other variables. It is more common for theories to propose that X and Y each have effects on Z independently of one another, and that, in addition there is an effect in common. This form of interaction is termed a "hierarchical" interaction.

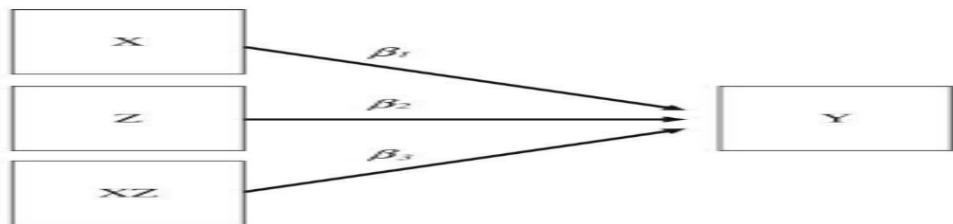
Conceptual diagram for moderation of an indirect effect example



Model 1

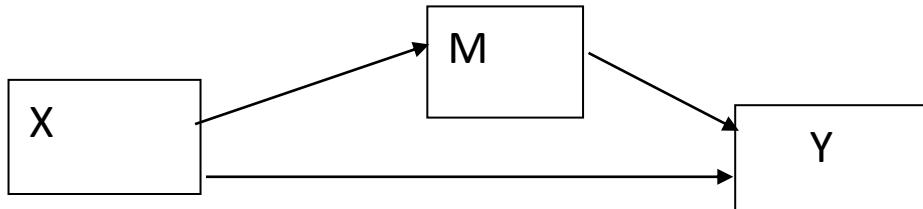


Model Analysis

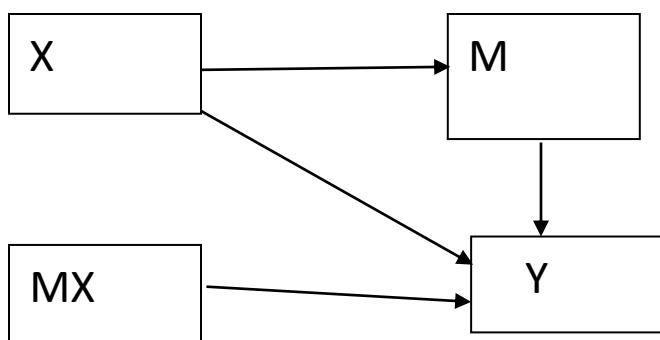


$$Y = i_5 + \beta_1 X + \beta_2 Z + \beta_3 XZ + e_5$$

Model 2



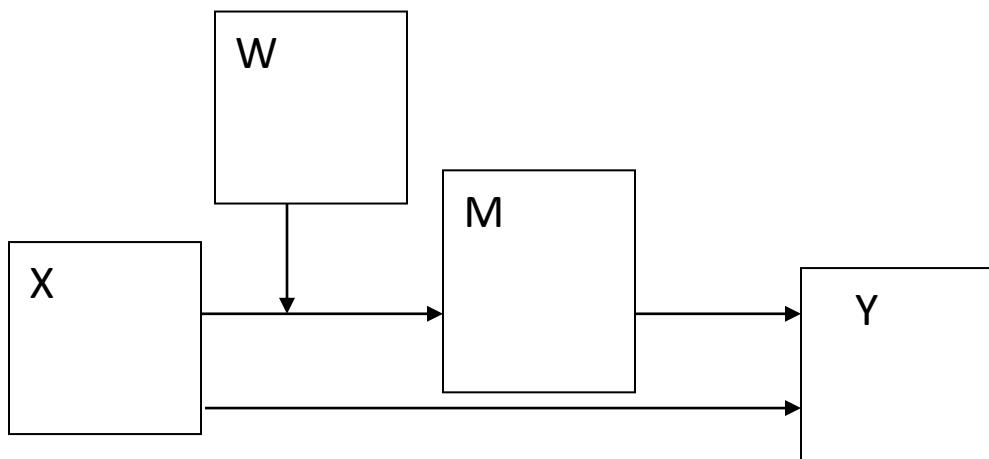
Model analysis



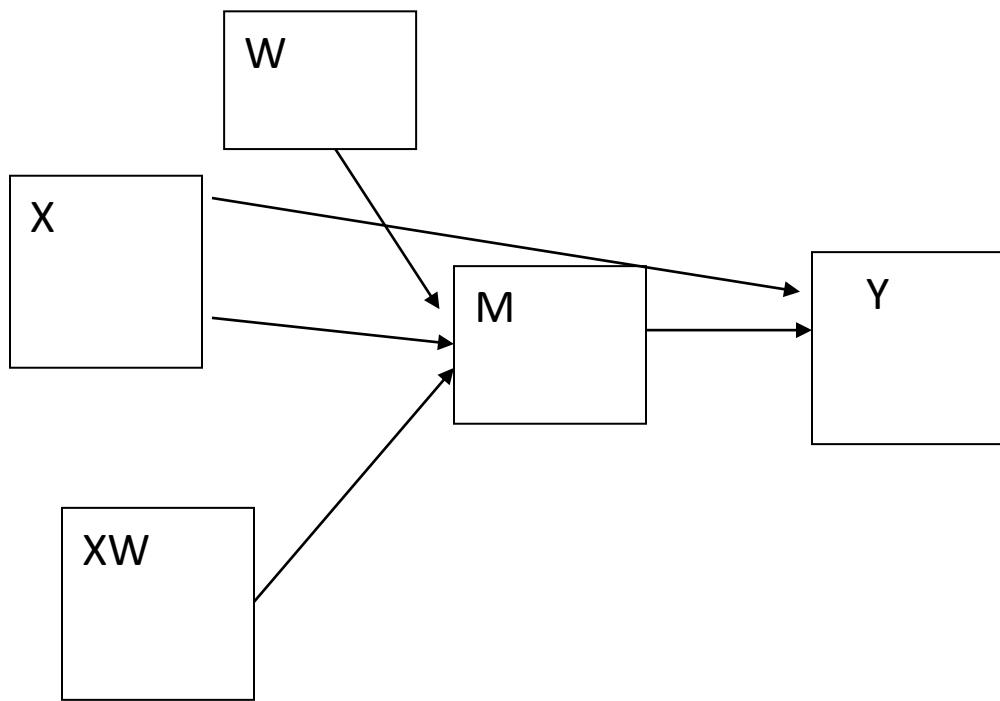
$$M = a_0 + a_1 X$$

$$Y = b_0 + b_1 M + b_2 X + b_3 MX$$

Model 3



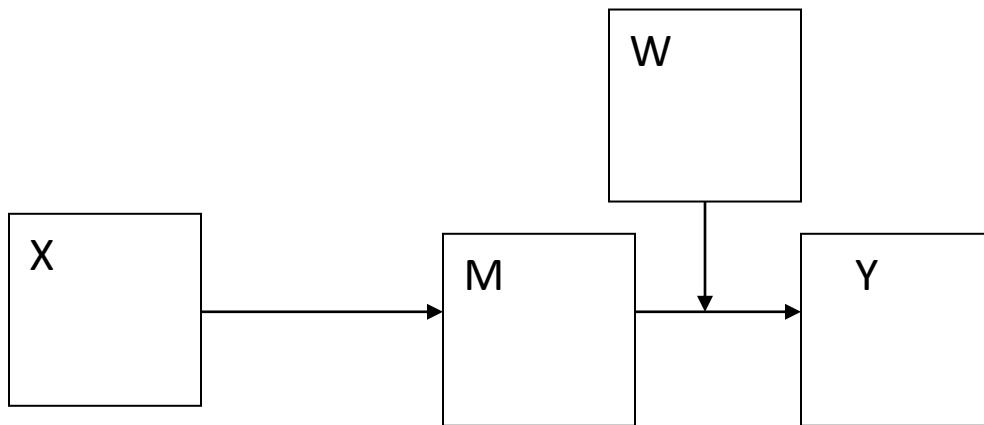
Atau dengan model analisis:



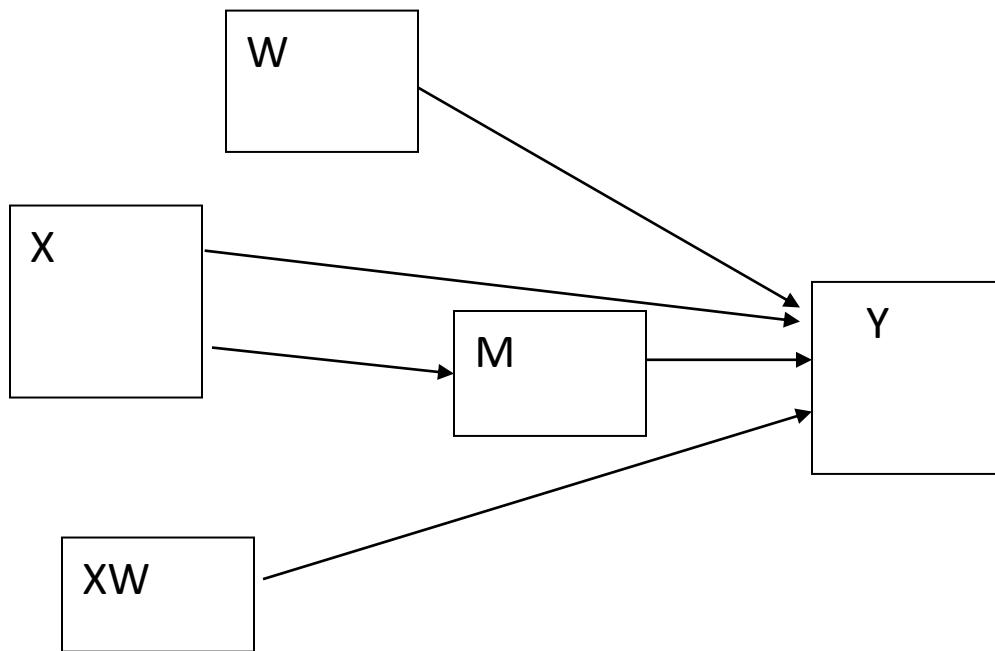
$$M = a_0 + a_1X + a_2W + a_3 XW$$

$$Y = b_0 + b_1M + b_2X$$

Model 4



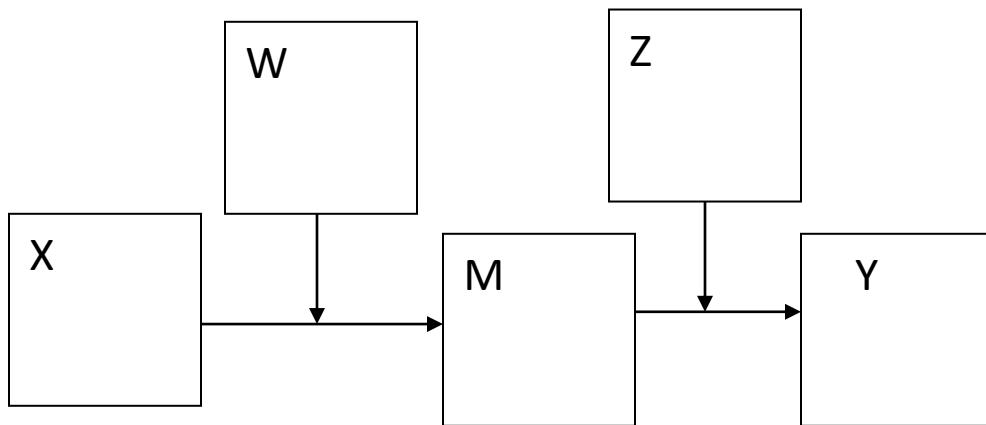
Atau dengan model analisis:



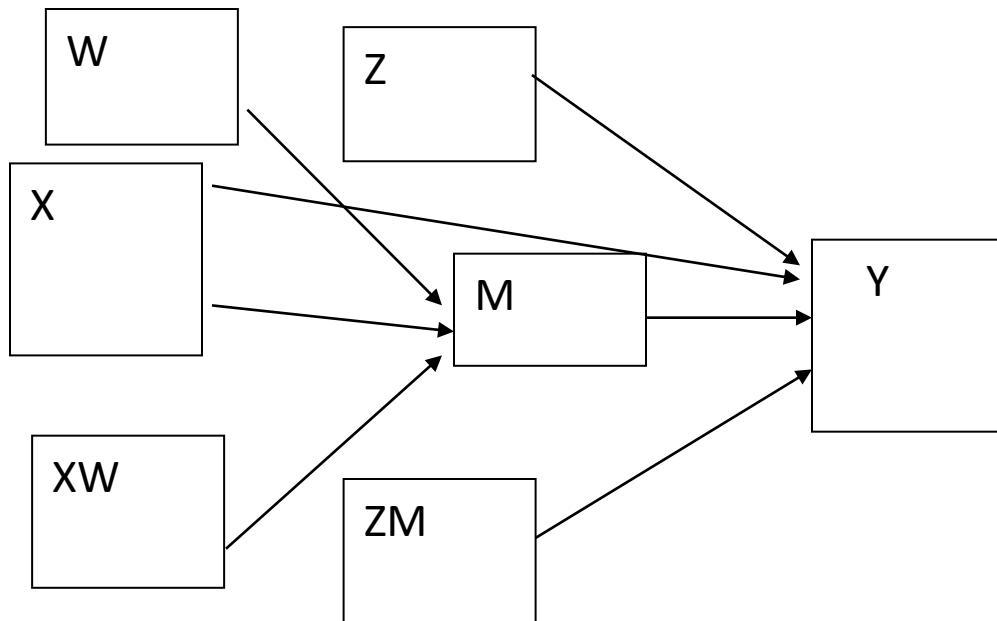
$$M = a_0 + a_1X$$

$$Y = b_0 + b_1M + b_2X + b_3W + b_4XW$$

Model 5



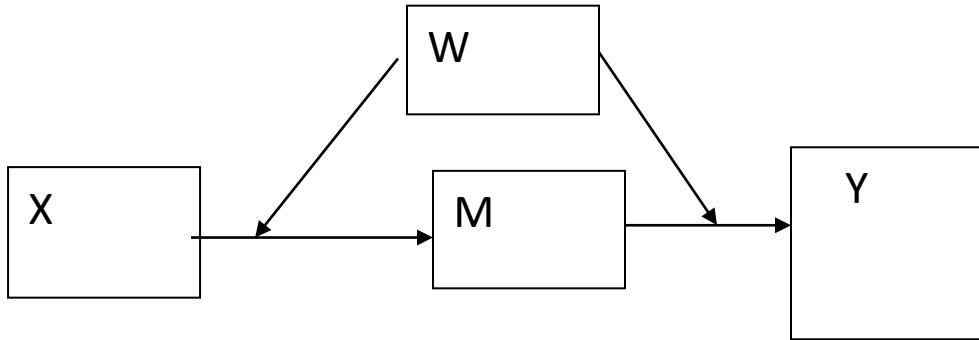
Atau dengan model analisis:



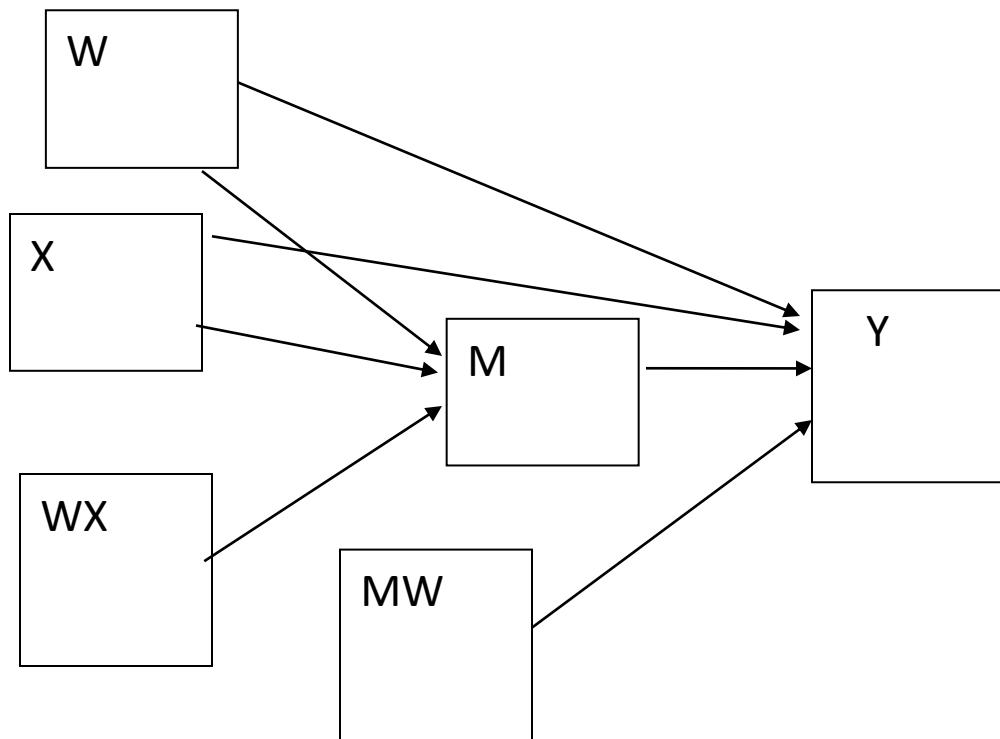
$$M = a_0 + a_1X + a_3XW$$

$$Y = b_0 + b_1M + b_2X + b_3W + b_4XW + b_5Z + b_6MZ$$

Model 6



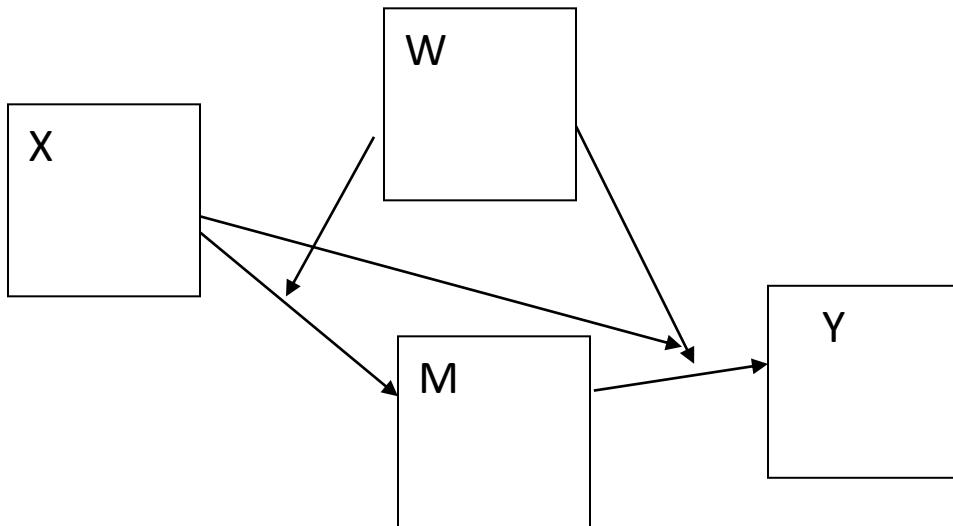
Atau dengan model analisis:



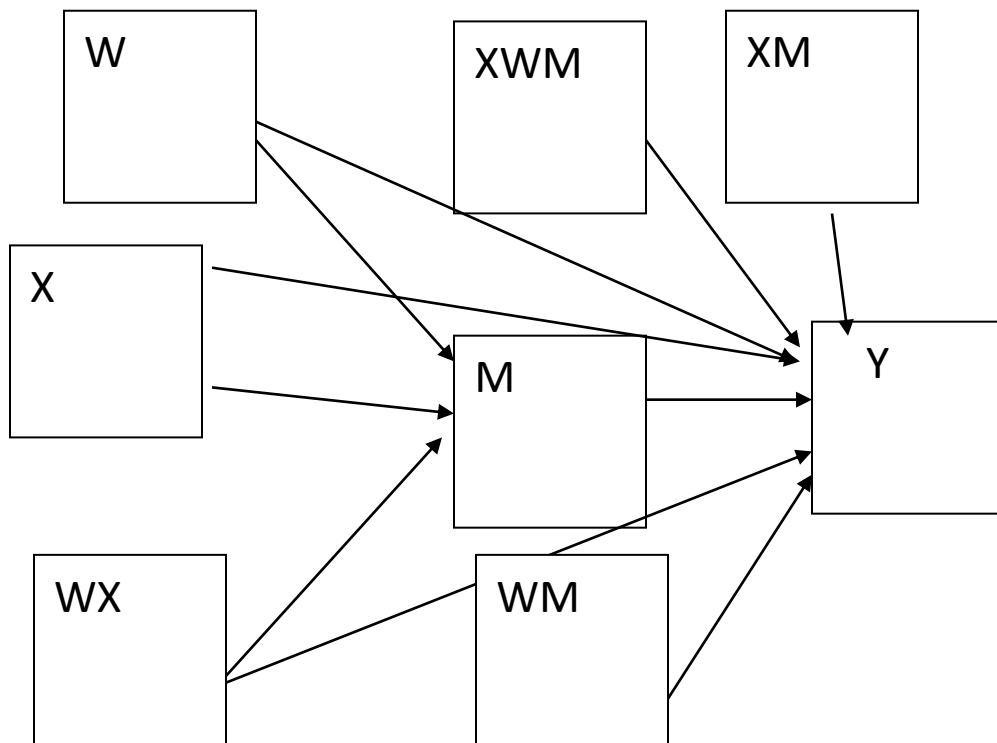
$$M = a_0 + a_1X + a_2W + a_3WX$$

$$Y = b_0 + b_1M + b_2X + b_3W + b_4WX + b_5MW$$

Model 7



Atau dengan model analisis:

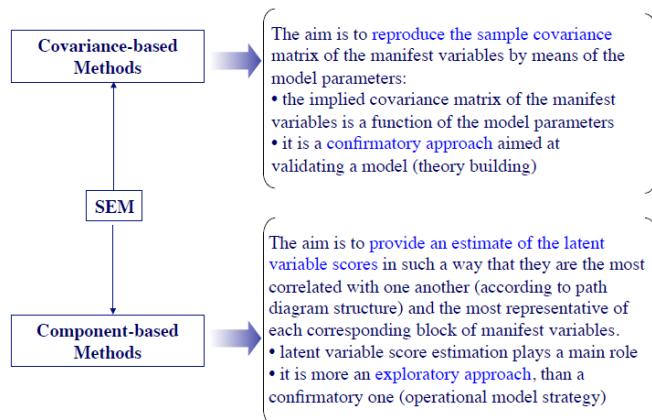


$$M = a_0 + a_1X + a_2W + a_3XW$$

$$Y = b_0 + b_1X + b_2W + b_3W + b_4M + b_5MW + b_6MX + b_7WXM$$

Path Modelling With Latent Variable

Two families of methods



Path model with latent variables

x = manifest variable associated to exogenous latent variables

y = manifest variable associated to endogenous latent variables

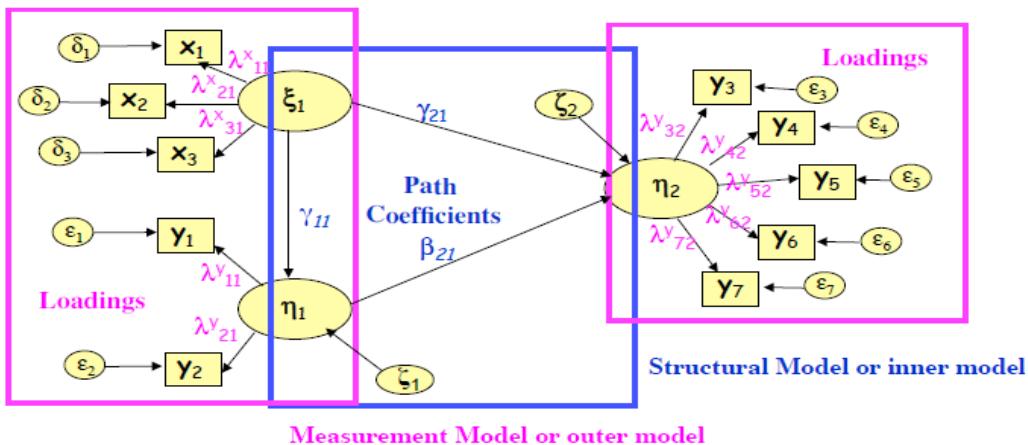
ξ = exogenous latent variable

η = endogenous latent variable

δ = measurement error associated to exogenous latent variables

ε = measurement error associated to endogenous latent variables

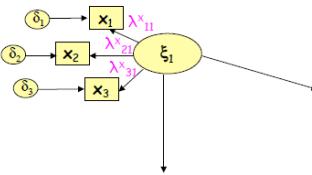
ζ = structural error



The measurement (outer) model

For the exogenous MVs

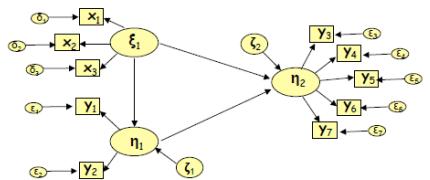
$$\begin{aligned} x_1 &= \lambda_{11}^x \xi_1 + \delta_1 \\ x_2 &= \lambda_{21}^x \xi_1 + \delta_2 \\ x_3 &= \lambda_{31}^x \xi_1 + \delta_3 \end{aligned}$$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \Lambda_x = \begin{bmatrix} \lambda_{11}^x \\ \lambda_{21}^x \\ \lambda_{31}^x \end{bmatrix} \quad \xi = [\xi_1] \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

$$\mathbf{x} = \Lambda_x \xi + \delta$$

The structural Equation Model

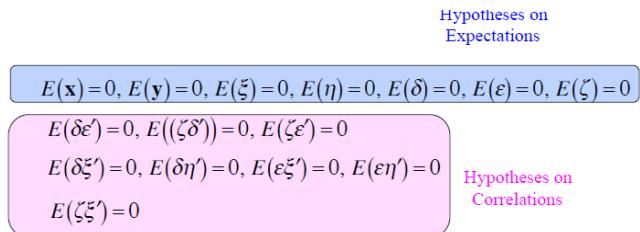


$$\begin{aligned} \mathbf{x} &= \Lambda_x \xi + \delta \\ \mathbf{y} &= \Lambda_y \eta + \varepsilon \end{aligned} \quad \left. \begin{array}{l} \text{Measurement Models} \\ \text{Structural Model} \end{array} \right\} \eta = \Gamma \xi + \mathbf{B} \eta + \zeta \quad \Leftrightarrow \quad \eta = (\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)$$

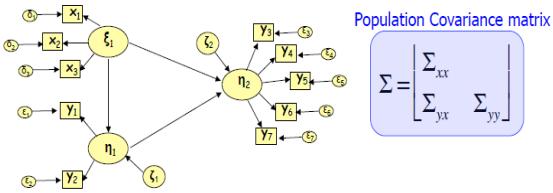
Model Assumption

Assuming that:

- i) the MVs, the LV and the errors (both in structural and measurement models) are centered
- ii) Two errors of different type (structural, exogenous measurement and endogenous measurement) do not covariate
- iii) Measurement errors and LVs do not covariate
- iv) The covariance between structural error and exogenous LVs is equal to zero



Analyzing Covariance



Population Covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

We can write the covariance matrix among the MVs in terms of model parameters (**implied covariance matrix**)

$$C = \Sigma(\Omega) = \Sigma(\Gamma, \mathbf{B}, \Lambda_x, \Lambda_y, \Phi, \Psi, \Theta_\delta, \Theta_\varepsilon)$$

↓
Path Coefficients Loadings Exog. LV Covariance Structural Error Covariance Measurement Error Covariance

Population Covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

Empirical covariance matrix

$$S = \begin{bmatrix} S_{xx} & S_{yx} \\ S_{yx} & S_{yy} \end{bmatrix}$$

"Implied" covariance matrix

$$C = \Sigma(\Omega) = \begin{bmatrix} \Lambda_x \Phi \Lambda'_x + \Theta_\delta & \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi' \Lambda'_x] \\ \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi' \Lambda'_x] & \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y + \Theta_\varepsilon] \end{bmatrix}$$

Implied Covariance matrix $\Sigma(\Omega)$

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \rightarrow \text{Covariance matrix of the population}$$

It can be rewritten as a **function of model parameters**

$$\Sigma(\Omega) = \begin{bmatrix} \Lambda_x \Phi \Lambda'_x + \Theta_\delta & \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y + \Theta_\varepsilon] \\ \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi' \Lambda'_x] & \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y + \Theta_\varepsilon] \end{bmatrix}$$

Covariance matrix of the population implied by the model

The sub-matrix Σ_{XX}

Outer model for exogenous blocks :

$$\mathbf{x} = \Lambda_x \xi + \delta$$

The covariance matrix of the exogenous MVs can be rewritten in terms of model parameters:

$$\begin{aligned}\Sigma_{XX}(\Omega) &= E(\mathbf{x}\mathbf{x}') = E\left[\left(\Lambda_x \xi + \delta\right)\left(\Lambda_x \xi + \delta\right)'\right] = \\ &= E\left[\left(\Lambda_x \xi + \delta\right)\left(\xi' \Lambda'_x + \delta'\right)\right] = \\ &= \Lambda_x E(\xi \xi') \Lambda'_x + \Lambda_x E(\xi \delta') + E(\delta \xi') \Lambda'_x + E(\delta \delta') \\ \text{let } & \quad \begin{matrix} E(\xi \xi') = \Phi \\ E(\delta \delta') = \Theta_\delta \end{matrix} \quad \rightarrow \quad \Sigma_{XX}(\Omega) = \Lambda_x \Phi \Lambda'_x + \Theta_\delta\end{aligned}$$

The sub-matrix Σ_{YY}

Outer model for the endogenous blocks :

$$\mathbf{y} = \Lambda_y \eta + \varepsilon$$

The covariance matrix of the exogenous MVs can be rewritten in terms of model parameters:

$$\begin{aligned}\Sigma_{YY}(\Omega) &= E(\mathbf{y}\mathbf{y}') = E\left[\left(\Lambda_y \eta + \varepsilon\right)\left(\Lambda_y \eta + \varepsilon\right)'\right] = \\ &= E\left[\left(\Lambda_y \eta + \varepsilon\right)\left(\eta' \Lambda'_y + \varepsilon'\right)\right] = \\ &= E\left[\Lambda_y \eta \eta' \Lambda'_y + \varepsilon \eta' \Lambda'_y + \Lambda_y \eta \varepsilon' + \varepsilon \varepsilon'\right] = \\ &= \Lambda_y E(\eta \eta') \Lambda'_y + \Lambda_y E(\varepsilon \eta') + E(\eta \varepsilon') \Lambda'_y + E(\varepsilon \varepsilon')\end{aligned}$$

This matrix is a function of other parameters

$$\text{let } E(\varepsilon \varepsilon') = \Theta_\varepsilon \quad \rightarrow \quad \Sigma_{YY}(\Omega) = \Lambda_y E(\eta \eta') \Lambda'_y + \Theta_\varepsilon$$

The sub-matrix Σ_{YY}

Covariance matrix of the endogenous LVs:

$$\eta = (\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)$$

$$\begin{aligned}\Sigma_{\eta\eta} &= E(\eta \eta') = E\left[\left[(\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)\right]\left[(\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)\right]'\right] = \\ &= E\left[\left[(\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)\right]\left[(\xi \Gamma' + \zeta')(\mathbf{I} - \mathbf{B})^{-1'}\right]\right] = \\ &= (\mathbf{I} - \mathbf{B})^{-1} E(\Gamma \xi \xi' \Gamma' + \zeta \xi' \Gamma' + \Gamma \xi \zeta' + \zeta \zeta') (\mathbf{I} - \mathbf{B})^{-1'} = \\ &= (\mathbf{I} - \mathbf{B})^{-1} [\Gamma E(\xi \xi') \Gamma' + E(\zeta \xi') \Gamma' + \Gamma E(\xi \zeta') + E(\zeta \zeta')] (\mathbf{I} - \mathbf{B})^{-1'} \\ \text{let } & \quad E(\zeta \zeta') = \Psi \quad \rightarrow \quad \Sigma_{\eta\eta}(\Omega) = (\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1'}\end{aligned}$$

The sub-matrix Σ_{YY}

The covariance matrix of the endogenous MVs can now be obtained as a function of model parameters:

$$\Sigma_{YY}(\Omega) = \Lambda_y E(\eta \eta') \Lambda'_y + \Theta_\varepsilon$$



$$\Sigma_{YY}(\Omega) = \Lambda_y \left[(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1'} \right] \Lambda'_y + \Theta_\varepsilon$$

The sub-matrix Σ_{XY}

Measurement model:

$$\mathbf{x} = \Lambda_x \xi + \delta \quad \text{and} \quad \mathbf{y} = \Lambda_y \eta + \varepsilon$$

The cross-covariance matrix between exogenous and endogenous MVs can be written as

$$\begin{aligned}\Sigma_{XY}(\Omega) &= E(\mathbf{x}\mathbf{y}') = E\left[\left(\Lambda_x \xi + \delta\right)\left(\Lambda_y \eta + \varepsilon\right)'\right] = \\ &= E\left[\left(\Lambda_x \xi + \delta\right)\left(\eta' \Lambda'_y + \varepsilon'\right)\right] = \\ &= \Lambda_x E(\xi \eta') \Lambda'_y + \Lambda_x E(\xi \varepsilon') + E(\delta \eta') \Lambda'_y + E(\delta \varepsilon')\end{aligned}$$

$$\Sigma_{XY}(\Omega) = \Lambda_x E(\xi \eta') \Lambda'_y$$

The measurement errors are uncorrelated among them and with the exogenous LVs

The sub-matrix Σ_{XY}

$$\begin{aligned}\eta &= (\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta) \\ \xi' &= (\xi' \Gamma' + \zeta')(\mathbf{I} - \mathbf{B})^{-1} \\ \xi \eta' &= (\xi \xi' \Gamma' + \xi \zeta')(\mathbf{I} - \mathbf{B})^{-1} \\ E(\xi \eta') &= \Phi \Gamma' (\mathbf{I} - \mathbf{B})^{-1}\end{aligned}$$

$$\Sigma_{XY}(\Omega) = \Lambda_x \Phi \Gamma' (\mathbf{I} - \mathbf{B})^{-1} \Lambda'_y$$

Cross-covariance matrix between endogenous and exogenous LVs expressed as a function of model parameters

Model Identification

Model identification in SEM

T-rule: A SEM is identified if the covariance matrix may be uniquely decomposed in function of the model parameters

→ if its $DF \geq 0$, that is if the number of covariances is larger than the number of parameters to be estimated

$$DF = \left[\frac{1}{2} (P+Q)(P+Q+1) - t \right]$$

of MVs in the model:
 $P = \# \text{ of exogenous MVs}$
 $Q = \# \text{ of endogenous MVs}$

of parameters to be estimated

$\frac{1}{2}(P+Q)(P+Q+1)$ → Number of unique elements in the Σ matrix

Necessary (but not sufficient) condition

Model Identification in SEM

Two-Step rule:

→ In the first step the analyst treats the model as a confirmatory factor analysis (both x and y variables are viewed as exogenous, relations between the LVs are expressed in terms of covariance)

→ The second step examines the latent variable equation of the original model and treat it as though it were a structural equation of observed variables

N.B. Those are sufficient (but not necessary) conditions

Model fit and Validation

χ^2 test

Chi-square Test - Global Validation Tests

$H_0: \Sigma = C \rightarrow$ Perfect fit
 $H_1: \Sigma \neq C$

Test Statistic

$$(N-1)F \xrightarrow{N \rightarrow \infty} \chi^2_{DF}$$

N : number of observations
 F : Discrepancy function
 DF : Degrees of freedom of the model

Decision Rule:

The model is accepted if p-value ≥ 0.05 (We cannot reject the hypothesis H_0)

N.B.: For a fixed level of differences in covariance matrices, the estimate of the Chi-square increases with N

- The power (i.e. the probability of rejecting a false H_0) depends on the sample size. If the sample size is important, this test may lead to reject the model even if the data fit well the model!
- We cannot use this test to compare model estimated on different sample size

χ^2 difference test

A restricted Model A is said to be nested within a Model B with more parameters and less degrees of freedom than Model A, if Model A can be derived from Model B by introducing some restrictions
 Given two nested models A and B with $df_A > df_B$, under the null hypothesis of equal fit for both models:

$$\chi^2_{DF_A} - \chi^2_{DF_B} / DF_A - DF_B \sim \chi^2_{DF_A - DF_B}$$

- If the χ^2 difference is significant, that is if $\Delta\chi^2_{obs} > \chi^2_{(DF_A - DF_B), 1-\alpha}$ the null hypothesis of equal fit for both models is rejected at the risk level α and Model B should be retained
- if the χ^2 difference is nonsignificant, the null hypothesis of equal fit for both models cannot be rejected and the restricted Model A should be favored

Test of Close-Fit

RMSEA (Root mean square error of approximation) (Steiger and Lind)

$$RMSEA = \sqrt{\frac{F_0}{DF}}$$

where $F_0 = \log|C| + tr(\Sigma C^{-1}) - \log|\Sigma| - (P+Q)$

- The introduction of degrees of freedom in the index allows for parsimonious models.

Test of Close-Fit

RMSEA is estimated as:

$$RMSEA_{estimated} = \sqrt{\frac{F}{DF} - \frac{1}{N-1}}$$

where $F = \log|C| + tr(\Sigma C^{-1}) - \log|\Sigma| - (P+Q)$

- The model is accepted if $RMSEA \leq 0.05$ or, at least, less than 0.08
- A confidence interval is usually provided
- It depends very little on N
- A 90% confidence interval around the point estimate enables an assessment of the precision of the RMSEA estimate. The lower boundary of the confidence interval should contain zero for exact fit and be $< .05$ for close fit.

Goodness of Fit index

This index was initially devised by Joreskog and Sorbom (1984) for ML and ULS estimation. It has then been generalised to other estimation criteria.

$$GFI = 1 - \frac{\text{tr} \left[\left((\mathbf{S} - \mathbf{C}) \mathbf{W}^{-1} \right)^2 \right]}{\text{tr} \left[\left(\mathbf{S} \mathbf{W}^{-1} \right)^2 \right]}$$

Adjusted GFI

This index take into account the df of the models by adjusting the GFI by a ratio of the DF used in the model and the max DF available

$$AGFI = 1 - \left(1 - GFI \right) \frac{(P+Q)(P+Q+1)/2}{DF}$$

→ If a model is able to explain any true covariance between the observed variables, then $GFI = 1$

The model is accepted if GFI is at least 0.95

→ If AGFI equals the unity, the model shows a perfect fit.
N.B.: It is not bounded below by zero.

The model is accepted if AGFI is at least 0.9

Bentler Comparative Fit Index

$$CFI = \frac{\left[(N-1)F_{IND} - DF_{IND} \right] - \left[(N-1)F - DF \right]}{\left[(N-1)F_{IND} - DF_{IND} \right]}$$

Fit function that would results if all parameters were zero
(minimum fit function value of the INDEPENDENCE (NULL) Model)

Degree of freedom of the INDEPENDENCE Model

Bentler-Bonnet Non-Normed Fit Index

(Also known as Tucker-Lewis index (TLI or rho₂)

$$NNFI = \frac{\frac{F_{IND}}{DF_{IND}} - \frac{F}{DF}}{\frac{F_{IND}}{DF_{IND}} - \frac{1}{n-1}}$$

It is bounded between 0 and 1, with higher value indicating better fit

The model is accepted if CFI is at least 0.9

→ It is usually (not necessarily) between 0 and 1, with higher value indicating better fit

The model is accepted if TLI is at least 0.95

Residual-based overall badness-of-fit indexes

Root Mean Square Residual index

$$RMR = \sqrt{\frac{\sum_{p=1}^P \sum_{j=1}^i (s_{pj} - \hat{s}_{pj})^2}{2(P+Q)(P+Q+1)}}$$

Good fit if RMR < 0.05

Standardized Root Mean Square Residual index

$$SRMR = \sqrt{\frac{\sum_{p=1}^P \sum_{j=1}^i (r_{pj} - \hat{r}_{pj})^2}{2(P+Q)(P+Q+1)}}$$

Good fit if SRMR < 0.1

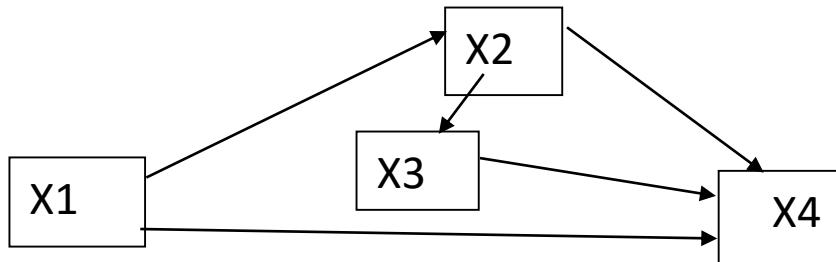
$$\text{where } r_{pj} = \frac{s_{pj}}{s_{pp} \times s_{jj}} \text{ and } \hat{r}_{pj} = \frac{\hat{s}_{pj}}{s_{pp} \times s_{jj}}$$

Soal ini harus dikerjakan dan dikumpul dalam waktu satu minggu, jawaban dikirim ke
Salehmire1957@gmail.com

1. What is the difference between Regression Analysis and Path Analysis ?
2. Practical difference between SEM and Path Analysis?
3. Explain the assumption of path analysis.
4. Given the correlation, $r =$

	X1	X2	X3	X4
X1	1			
X2	0.41	1		
X3	0.32	0.43	1	
X4	0.35	0.32	0.45	1

And path diagram



- a Find: p_{X1X2} , p_{X1X4} , p_{X2X3} , p_{X2X4} , p_{X3X4}
- b. Find total effect of X4 and describe on table
- c. Write the regression equations of the diagram path
- d. If X2 is moderator variable, draw the diagram and write the regression equations
5. What are difference between standardized and unstandardized regressions in the path model.
6. OLS and maximum likelihood methods can be used to predict the path coefficient. What is the difference of the methods in the path analysis and how to do it.

Dynamic Econometrics

AUTOREGRESSIVE
AND
DISTRIBUTED-LAG
MODELS

Dynamic Econometric Models:

A. Autoregressive Model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_k Y_{t-k} + e_t$$

(With lagged dependent variable(s) on the RHS)

B. Distributed-lag Model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + e_t$$

(Without lagged dependent variables on the RHS)

Where β_0 is known as the short run multiplier, or impact multiplier, because it gives the change in the mean value of Y following a unit change of X in the same time period. If the change of X is maintained at the same level thereafter, then, $(\beta_0 + \beta_1)$ gives the change in the mean value of Y in the next period, $(\beta_0 + \beta_1 + \beta_2)$ in the following period, and so on. These partial sums are called interim, or intermediate, multiplier. Finally, after k periods, that is

$$\sum_{i=0}^k \beta_i = \beta_0 + \beta_1 + \beta_2 + \dots + \beta_k = B, \text{ therefore } \Sigma \beta_i \text{ is called the long run multiplier or}$$

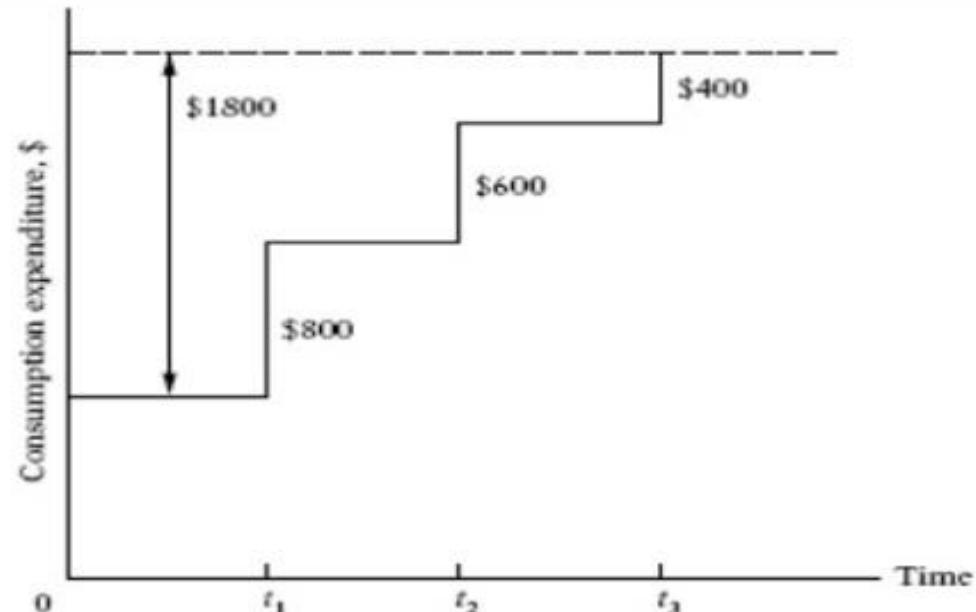
total multiplier, or distributed-lag multiplier. If define the standardized $\beta_i^* = \beta_i / \Sigma \beta_i$, then it gives the proportion of the long run, or total, impact felt by a certain period of time. In order for the distributed lag model to make sense, the lag coefficients must tend to zero as $k \rightarrow \infty$. This is not to say that β_2 is smaller than β_1 ; it only means that the impact of X_{t-k} on Y must eventually become small as k gets large.

For example: a consumption function regression is written as

$$Y = \alpha + 0.4X_t + 0.3 X_{t-1} + 0.2 X_{t-2} + 0.1 X_{t-3} \dots + e_t$$

The role of lag time

Consumption function



$$Y_t = \text{constant} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2}$$

SHORT RUN

$$Y_t = \text{constant} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2}$$

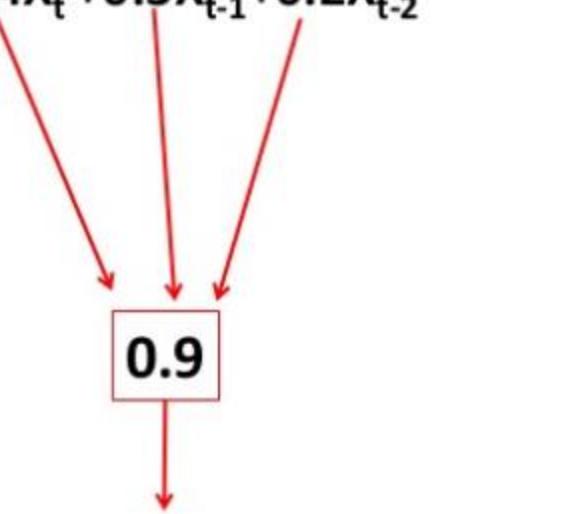
The consumption of Current year of increment



LONG RUN



$$Y_t = \text{constant} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2}$$

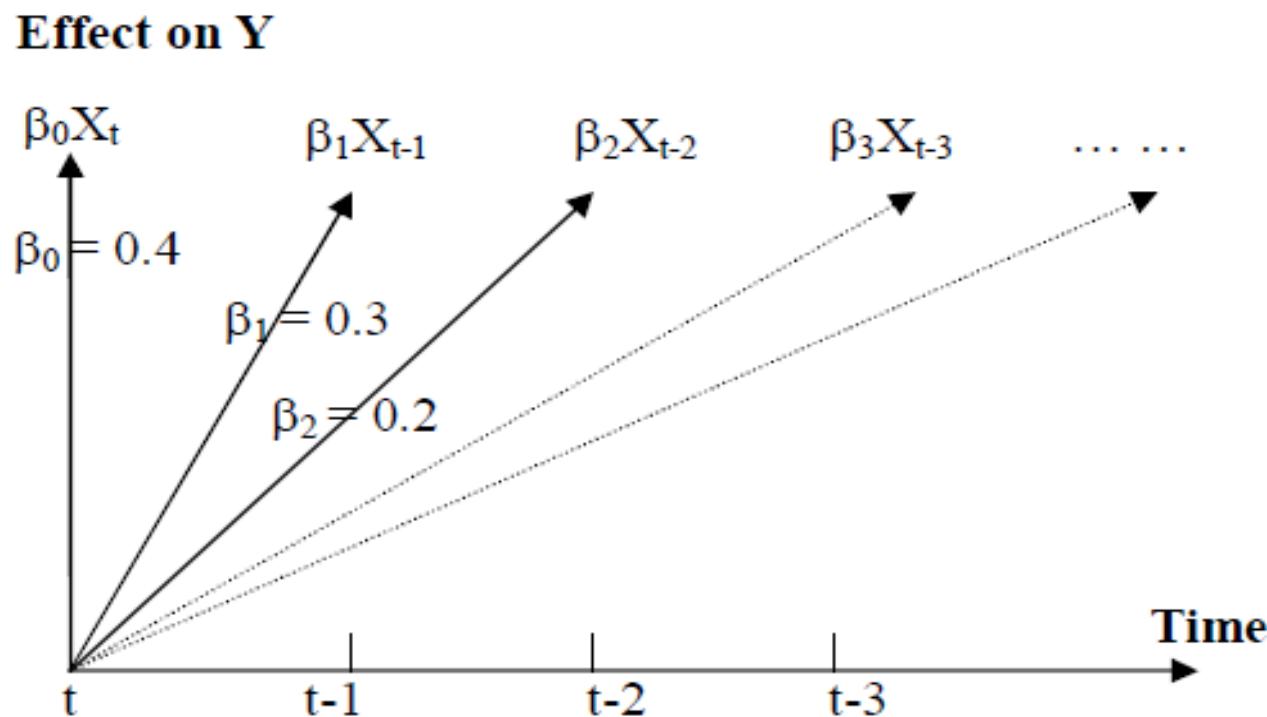


Sum of the consumption of over all period

Example:

$$Y = \alpha + 0.4X_t + 0.3 X_{t-1} + 0.2 X_{t-2} + 0.1 X_{t-3} \dots + e_t$$

Then the effect of a unit change of X at time t on Y and its subsequent time periods can be shown as the follow diagram:



Example : Money creation process, inflation process due to money supply, productivity growth due to expenditure or investment.

Period	New Deposits	New Loans	Required Reserve (10%)
1	\$1000	\$900	\$100
2	900	810	90
3	810	729	81
4	729	656.1	72.9
...			
Total	\$10,000	\$9,000	\$1,000

In general: Suppose the distributed-lag model is

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + e_t$$

The basic idea: The long run responses of Y to a change in X are different from the immediate and short-run responses. And suppose the expect value in different periods is A. Then $E(X_t) = E(X_{t-1}) = E(X_{t-2}) = A$

continued

In general: Suppose the distributed-lag model is

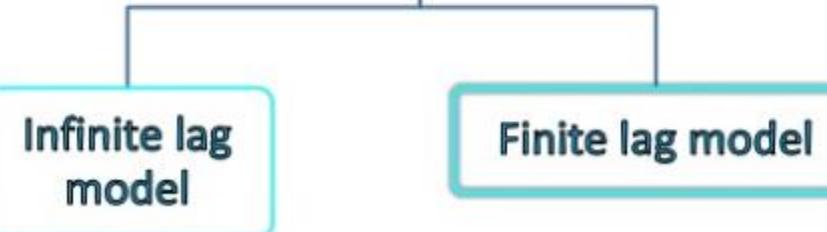
$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + e_t$$

The basic idea: The long run responses of Y to a change in X are different from the immediate and short-run responses. And suppose the expect value in different periods is A. Then $E(X_t) = E(X_{t-1}) = E(X_{t-2}) = A$

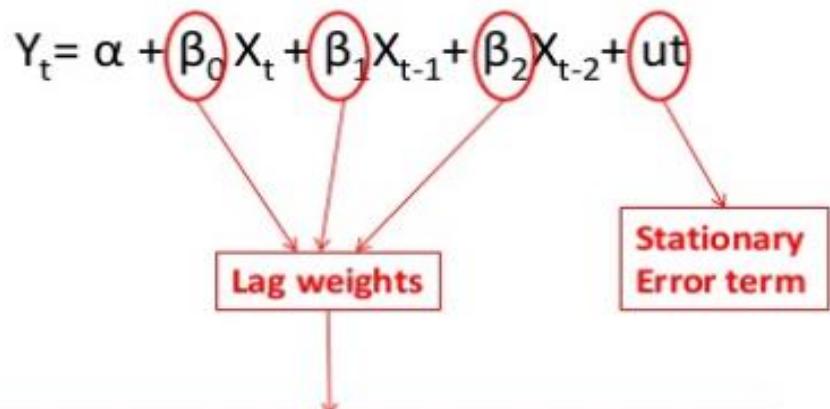
$$\begin{aligned}E(Y_t) &= \alpha + \beta_0 E(X_t) + \beta_1 E(X_{t-1}) + \beta_2 E(X_{t-2}) + \dots + E(e_t) \\&= \alpha + \beta_0 A + \beta_1 A + \beta_2 A + \dots + 0 \\&= \alpha + \sum \beta_i A\end{aligned}$$

This gives the constant long run corresponding to $X = A$, and $E(Y)$ will remain at this level unit when X changes again.

Distributed lag model



Estimation of distributed lag models



They define the pattern of how x affects y over time.

Infinite lag model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + u_t$$

Finite lag model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + u_t$$

"k is specified"

Application on Distributed Lag Model

Dependent Variable: PPCE

Method: Least Squares

Date: 04/30/15 Time: 23:14

Sample (adjusted): 1971 1999

Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1949.865	343.7191	-5.672845	0.0000*
PPDI	1.142786	0.190810	5.989137	0.0000*
PPDI(-1)	-0.144730	0.192419	-0.752161	0.4587
R-squared	0.990875	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion Hannan-Quinn criter. Durbin-Watson stat	16063.90 2910.503 14.26514 14.40658 14.30944 0.659906	
Adjusted R-squared	0.990173			
S.E. of regression	288.5250			
Sum squared resid	2164414.			
Log likelihood	-203.8445			
F-statistic	1411.614			
Prob(F-statistic)	0.000000			

Short run

long run = 0.99797

No +ve autocorrelation

Since we are using the 1 past lag value so number of observation reduced from 30 to 29

continued

DURBIN-WATSON d STATISTIC: SIGNIFICANCE POINTS OF

n	$K' = 1$		$K' = 2$		$K' = 3$		$K' = 4$		$K' = 5$	
	d_L	d_U								
30	1.133	1.263	1.070	1.339	1.005	1.421	0.941	1.511	0.877	1.606

TABLE 12.6 DURBIN-WATSON d TEST: DECISION RULES

Null hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \leq d \leq d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \leq d \leq 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$

Interpretation Summary

- Short run $\rightarrow 1.142786$ or a unit change in PPDI on average the PPCE will increase up to 1.142 units
- Long run $\rightarrow 1.142786 + (-0.144730) \rightarrow 0.99805$
- Durbin Watson \rightarrow no positive autocorrelation
- Observation \rightarrow 1 past lag value include so observation reduce from 30 to 29.
- R square define up to 99.08% effect of PPDI on PPCE.
- % of a total impact of a unit change in PPDI on PPCE :
 - 1st year: $1.1427 / 0.99805 \rightarrow 11.4\%$
 - 2nd year: $0.14473 / 0.9980 \rightarrow 11.44\%$

Ad HOC

Comparison of past lag 3 and 4

Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2289.241	394.2584	-5.806448	0.0000
PPDI	1.093976	0.193514	5.653206	0.0000
PPDI(-1)	0.067515	0.285623	0.236377	0.8153
PPDI(-2)	-0.151476	0.290981	-0.520571	0.6079
PPDI(-3)	0.003731	0.204352	0.018259	0.9856

Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2218.414	424.7393	-5.223002	0.0000
PPDI	1.014922	0.210843	4.813639	0.0001
PPDI(-1)	0.116158	0.293245	0.396113	0.6962
PPDI(-2)	-0.120322	0.292454	-0.411422	0.6851
PPDI(-3)	-0.273918	0.295097	-0.928230	0.3644
PPDI(-4)	0.280040	0.209413	1.337262	0.1961

Final Ad Hoc result

Equation: UNTITLED Workfile: TABLE 17.2::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: PPCE

Method: Least Squares

Date: 05/03/15 Time: 15:54

Sample (adjusted): 1974 1999

Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2218.414	424.7393	-5.223002	0.0000
PPDI	1.014922	0.210843	4.813639	0.0001
PPDI(-1)	0.116158	0.293245	0.396113	0.6962
PPDI(-2)	-0.120322	0.292454	-0.411422	0.6851
PPDI(-3)	-0.273918	0.295097	-0.928230	0.3644
PPDI(-4)	0.280040	0.209413	1.337262	0.1961

R-squared	0.990947	Mean dependent var	16519.12
Adjusted R-squared	0.988684	S.D. dependent var	2717.050
S.E. of regression	289.0351	Akaike info criterion	14.37015
Sum squared resid	1670826.	Schwarz criterion	14.66048
Log likelihood	-180.8119	Hannan-Quinn criter.	14.45375
F-statistic	437.8390	Durbin-Watson stat	0.510738
Prob(F-statistic)	0.000000		

Drawback of Ad Hoc Estimation

- Maximum length of the lags.
- Fewer degree of freedom left
- Data mining (big data)
- Multicollinearity

Koyck approach

Adjustment of speed

slow

$$\beta_0 = \beta_k \lambda^k$$

($k=0,1,\dots$) ($0 < \lambda < 1$)

(The higher the value of λ the lower the speed of adjustment, and the lower the value of λ the greater the speed of adjustment)

fast

Koyck transformation

Distributed lag model

Transformed

Autoregressive

$$1) Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \dots + u_t \quad (\text{For Infinite lag})$$

Koyck lag by one period :

$$2) Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_0 \lambda X_{t-2} + \beta_0 \lambda^2 X_{t-3} \dots + u_{t-1}$$

Multiply by λ :

$$3) \lambda Y_{t-1} = \lambda \alpha + \lambda \beta_0 X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \beta_0 \lambda^3 X_{t-3} \dots + u_{t-1}$$

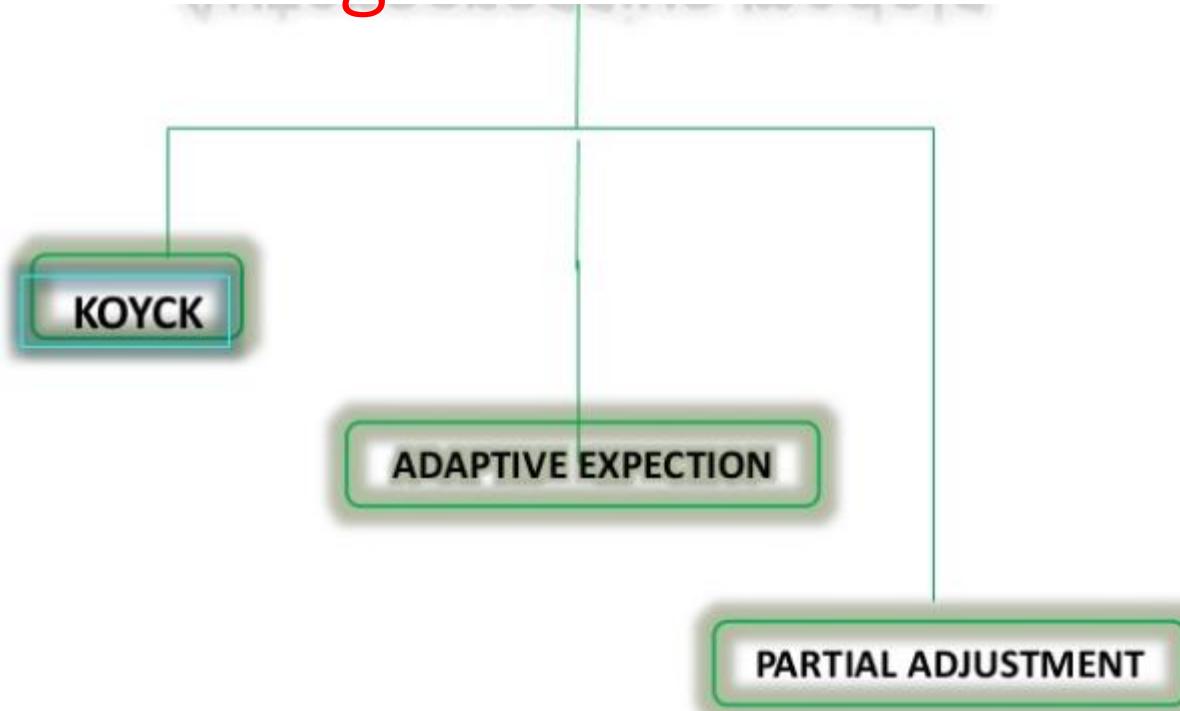
Subtracting the 1st and 3rd equation :

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

We get koyck transformed equation

Estimation of Autoregressive Model

Autoregressive Model



Koyck Autoregressive Model

Dependent Variable: PPCE

Method: Least Squares

Date: 05/01/15 Time: 04:07

Sample (adjusted): 1971 1999

Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1242.169	402.5785	-3.085533	0.0048
PPDI	0.603376	0.150261	4.015516	0.0004
PPCE(-1)	0.410687	0.154618	2.656138	0.0133

R-squared	0.992666	Mean dependent var	16063.90
Adjusted R-squared	0.992102	S.D. dependent var	2910.503
S.E. of regression	258.6577	Akaike info criterion	14.04659
Sum squared resid	1739499.	Schwarz criterion	14.18803
Log likelihood	-200.6755	Hannan-Quinn criter.	14.09088
F-statistic	1759.610	Durbin-Watson stat	1.005665
Prob(F-statistic)	0.000000		

$$Y_t = \underline{\alpha(1 - \lambda)} + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

$$PPCE_t = -1242.169 + 0.6033PPDI_t + 0.4106PCE_{t-1} + v_t$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1242.169	402.5785	-3.085533	0.0048
PPDI	0.603376	0.150261	4.015516	0.0004
PPCE(-1)	0.410687	0.154618	2.656138	0.0133

➤ Rate of adjustment → $1 - \lambda \rightarrow 1 - 0.4106$

RATE of adjustment → 0.589

➤ Short run : current year consumption of increment
Short run → 0.6033

➤ Long run:

Long run = coeff $\alpha_{t-1} / 1 - \lambda \rightarrow 1.0237$

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

$$PPCE_t = -1242.169 + 0.6033PPDI_t + 0.4106PCE_{t-1} + v_t$$

Median lag: median lag(time required) for 1st half

➤ Median lag = $\log(2)/\log \lambda \rightarrow \log(2)/\log (0.4106) \rightarrow 0.776$

(on average with a unit change in income the median lag(time required) for 1st half is 0.776)

Mean lag: effect of change in independent to be felt on dependent variable

➤ Mean lag = $\lambda / 1 - \lambda \rightarrow 0.4106 / 0.5894 \rightarrow 0.6966$

(on average ,for the effect of change in ppdi to be felt on ppce)

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

$$PPCE_t = -1242.169 + 0.6033PPDI_t + 0.4106PCE_{t-1} + v_t$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1242.169	402.5785	-3.085533	0.0048
PPDI	0.603376	0.150261	4.015516	0.0004
PPCE(-1)	0.410687	0.154618	2.656138	0.0133
R-squared	0.992666	Mean dependent var	16063.90	
Adjusted R-squared	0.992102	S.D. dependent var	2910.503	
S.E. of regression	258.6577	Akaike info criterion	14.04659	
Sum squared resid	1739499.	Schwarz criterion	14.18803	
Log likelihood	-200.6755	Hannan-Quinn criter.	14.09088	
F-statistic	1759.610	Durbin-Watson stat	1.005665	

➤ Durbin h test $\rightarrow (1-d/2)[n/1-n\{\text{var}(\alpha_{t-1})\}]^{1/2}$

Durbin h test $\rightarrow (0.4972) [(30/1-30(0.239)]^{1/2}$

Durbin h test $\rightarrow 5.1191$

(Durbin h ~ Norm distribution , As h value exceed ± 3 , so the probability is highly significant)

Decision:

There is a positive autocorrelation ,

Soal

1. Misalkan diketahui persamaan :

$$M_t = \alpha \ddot{Y}_t^{\beta_1} \ddot{R}_t^{\beta_2} e^{st}$$

M_t = permintaan untuk real cash balance

\ddot{Y}_t = pendapatan real yang diharapkan

\ddot{R}_t = tingkat bunga yang diharapkan

Nilai harapan dirumuskan

dimana $\ddot{Y}_t = \delta_1 Y_t + (1 - \delta_1) \ddot{Y}_{t-1}$ dan $\ddot{R}_t = \delta_2 R_t + (1 - \delta_2) YR_{t-1}$

1. Apa persoalan estimasi yang dihadapi
2. Tentukan nilai observasi sebenarnya untuk M_t
2. Buat suatu kasus yang membahas dinamis distribusi lag. Waktu lag minimal adalah 15 tahun. Anda dapat menggunakan data hipotetis, jika tidak memiliki data autentik. Buat dan tentukan hubungan antar variabel untuk menentukan persamaan dinamis distribusi lag dengan menggunakan metode Koyck dan metode Almon. Bandingkan

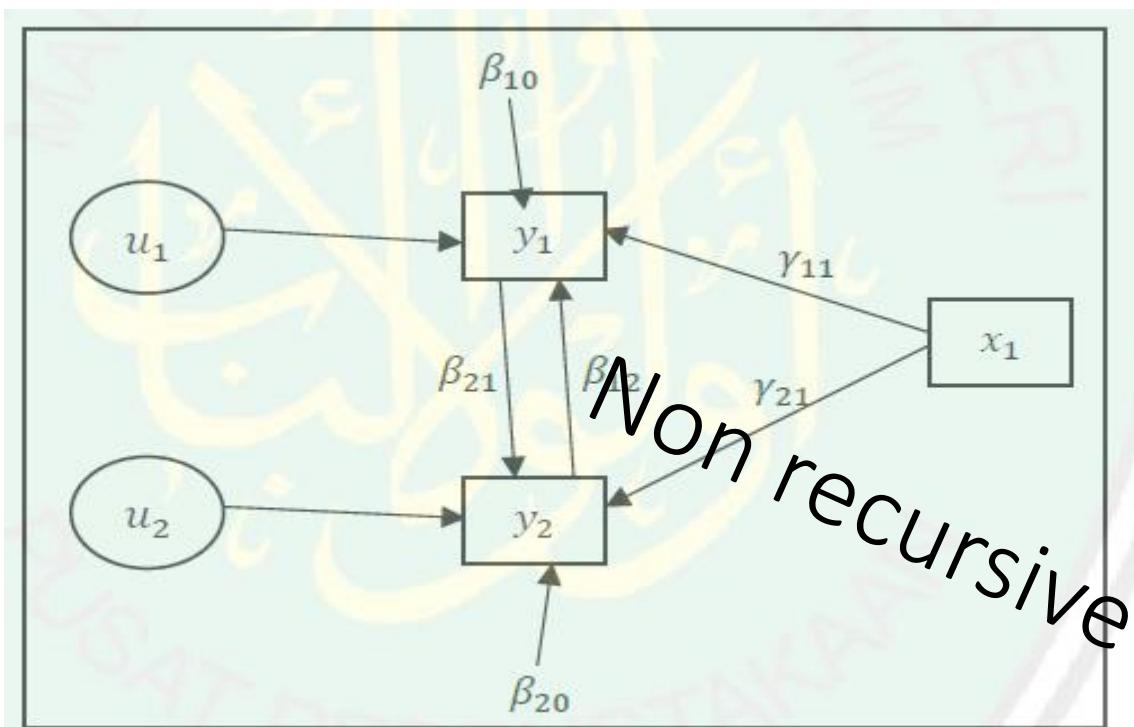
Sumber:

1. J. Supranto
2. Ammara Aftab

SEM = Simultaneous equation model (1)

SEM = Structural equation modelling (2)

An Example of SEM (1)



SEM Model

$$\beta_{11}y_{1t} + \beta_{12}y_{2t} + \dots + \beta_{1M}y_{Mt} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \dots + \gamma_{1K}x_{Kt} = u_{1t}$$

$$\mathbf{B}_{(MxM)} \mathbf{Y}_{(Mx1)} + \boldsymbol{\Gamma}_{(MxK)} \mathbf{X}_{(Kx1)} = \mathbf{U}_{(Mx1)}$$

$$\beta_{21}y_{1t} + \beta_{22}y_{2t} + \dots + \beta_{2M}y_{Mt} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + \dots + \gamma_{2K}x_{Kt} = u_{2t}$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\beta_{M1}y_{1t} + \beta_{M2}y_{2t} + \dots + \beta_{MM}y_{Mt} + \gamma_{M1}x_{1t} + \gamma_{M2}x_{2t} + \dots + \gamma_{MK}x_{Kt} =$$

$$u_{Mt}$$

$$\mathbf{Y} = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Kt} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Mt} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{21} & \cdots & \beta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{M1} & \beta_{M1} & \cdots & \beta_{MM} \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{21} & \cdots & \gamma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{M1} & \gamma_{M1} & \cdots & \gamma_{MK} \end{bmatrix}$$

$$\mathbf{B}_{(MxM)} \mathbf{Y}_{(Mx1)} + \boldsymbol{\Gamma}_{(MxK)} \mathbf{X}_{(Kx1)} = \mathbf{U}_{(Mx1)}$$

$$\mathbf{Y}_{(Mx1)} + \mathbf{B}_{(MxM)}^{-1} \boldsymbol{\Gamma}_{(MxK)} \mathbf{X}_{(Kx1)} = \mathbf{B}_{(MxM)}^{-1} \mathbf{U}_{(Mx1)}$$

Sehingga

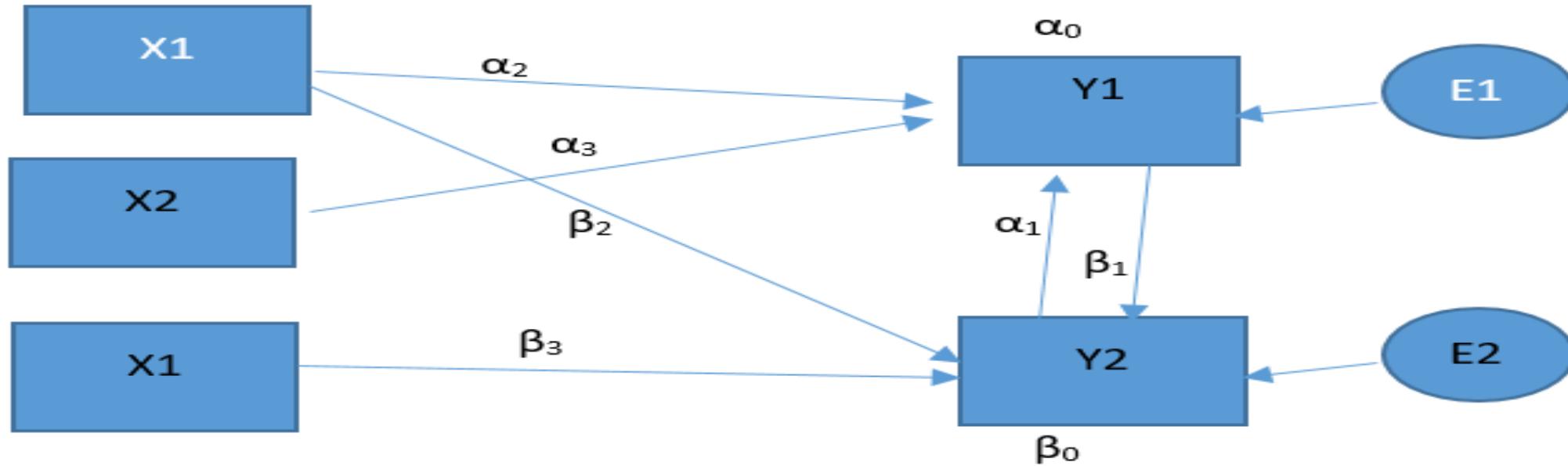
$$\mathbf{Y} = \boldsymbol{\Gamma} \mathbf{X} + \mathbf{V}$$

$$E(U_t) = 0, \quad t = 1, 2, 3, \dots, M$$

$$Var(U_t) = \Sigma, \quad t = 1, 2, 3, \dots, M$$

$$Cov(U_t, U_s) = E(U_t U_s^T) = 0, \quad \forall t \neq s$$

An Example of SEM System



Y_1 = growth

X_1 = domestic Investment

X_2 = foreign Investment

Y_2 = foreign debt

X_3 = Net export

The principle of identification

1. $K - k > m - 1$ and rank matrix A is $M - 1$, over identified, TSLS
2. $K - k = m - 1$ and rank matrix A is $M - 1$, Exactly identified, ILS
3. If $K - k \geq m - 1$ and rank matrix A is $M - 1$ less then $M - 1$,
identified, (TSLS, ILS)
4. $K - k < m - 1$, not identified

Example

Demand and supply models

Fungsi Permintaan

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \varepsilon_{1t}, \quad \alpha_1 < 0$$

Fungsi Penawaran

$$Q_t^s = \beta_0 + \beta_1 P_t + \varepsilon_{2t}, \quad \beta_1 > 0$$

Equilibrium

$$Q_t^d = Q_t^s$$

Example 1:

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \mu_{1t}$$

$$Q_t^s = \beta_0 + \beta_1 P_t + \mu_{2t}$$

$$Q_t^d = Q_t^s$$

$$\alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \mu_{1t} = \beta_0 + \beta_1 P_t + \mu_{2t}$$

$$(\alpha_1 - \beta_1)P_t = (\beta_0 - \alpha_0) - \alpha_2 Y_t + (\mu_{2t} - \mu_{1t})$$

$$P_t = \frac{(\beta_0 - \alpha_0)}{(\alpha_1 - \beta_1)} - \frac{\alpha_2}{(\alpha_1 - \beta_1)} Y_t + \frac{(\mu_{2t} - \mu_{1t})}{(\alpha_1 - \beta_1)}$$

atau

$$P_t = \frac{(\alpha_0 - \beta_0)}{(\beta_1 - \alpha_1)} + \frac{\alpha_2}{(\beta_1 - \alpha_1)} Y_t + \frac{(\mu_{1t} - \mu_{2t})}{(\beta_1 - \alpha_1)}$$

$$P_t = \pi_1 + \pi_2 Y_t + v_{1t}$$

Dimana:

$$\pi_1 = \frac{(\alpha_0 - \beta_0)}{(\beta_1 - \alpha_1)} ; \quad \pi_2 = \frac{\alpha_2}{(\beta_1 - \alpha_1)} ; \quad v_{1t} = \frac{\mu_{1t} - \mu_{2t}}{(\beta_1 - \alpha_1)}$$

Contoh2:

Diketahui suatu model persamaan simultan adalah sebagai berikut :

$$Q_d = \alpha_0 + \alpha_1 P + \alpha_2 X + v \quad \dots \dots \dots \quad (1.13)$$

$$Q_s = \beta_0 + \beta_1 P + \beta_2 P_I + u \quad \dots \dots \dots \quad (1.14)$$

Dimana:

Q_d = Jumlah barang yang diminta

Q_s = Jumlah barang yang ditawarkan

P = harga barang

X = Income

P_I = harga Input

- Persamaan *reduce form-nya* adalah sebagai berikut :

$$P = \Pi_0 + \Pi_1 X + \Pi_2 P_I + \Omega_1 \quad \dots \dots \dots \quad (1.15)$$

$$Q = \Pi_3 + \Pi_4 X + \Pi_5 P_I + \Phi_2 \quad \dots \dots \dots \quad (1.16)$$

Persamaan *Reduce Form* dapat dicari dengan langkah sebagai berikut:

Selesaikan persamaan

$$Q_d = Q_s \dots \dots \dots \dots \dots \dots \dots \quad (1.17)$$

$$\alpha_0 + \alpha_1 P + \alpha_2 X + v = \beta_0 + \beta_1 P + \beta_2 Pl + u$$

$$\alpha_1 P - \beta_1 P = \beta_0 - \alpha_0 - \alpha_2 X + \beta_2 Pl + u - v$$

$$P = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{u - v}{\alpha_1 - \beta_1} \right)$$

$$P = \Pi_0 + \Pi_1 X + \Pi_3 Pl + \Omega$$

- Kemudian substitusikan persamaan P diatas dengan salah satu persamaan Q, misalnya dengan Qd
- $Q_d = \alpha_0 + \alpha_1 P + \alpha_2 X + v$

$$Q_d = \alpha_0 + \alpha_1 \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{u - v}{\alpha_1 - \beta_1} \right) + \alpha_2 X + v$$

$$Q_d = \alpha_0 + \left(\frac{\alpha_1 \beta_0 - \alpha_1 \alpha_0}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_1 \alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{\alpha_1 u - \alpha_1 v}{\alpha_1 - \beta_1} \right) + \alpha_2 X + v$$

$$Q_d = \alpha_0 + \left(\frac{\alpha_1 \beta_0 - \alpha_1 \alpha_0}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_1 \alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{\alpha_1 u - \alpha_1 v}{\alpha_1 - \beta_1} \right) + \alpha_2 X + v$$

- Lalu samakan semua penyebutnya dengan $\alpha_1 - \beta$

$$Q_d = \left(\frac{\alpha_0\alpha_1 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \right) + \left(\frac{\alpha_1\beta_0 - \alpha_1\alpha_0}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_1\alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{\alpha_1u - \alpha_1v}{\alpha_1 - \beta_1} \right)$$

$$+ \left(\frac{\alpha_1\alpha_2 - \beta_1\alpha_2}{\alpha_1 - \beta_1} \right) X + \left(\frac{\alpha_1v - \beta_1v}{\alpha_1 - \beta_1} \right)$$

$$Q_d = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \right) - \left(\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \right) X + \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \right) Pl + \left(\frac{\alpha_1u - \beta_1v}{\alpha_1 - \beta_1} \right)$$

$$Q_d = \Pi_3 + \Pi_4 X + \Pi_5 Pl + \Phi$$

The reduce form we find 6 coefficients:

$\Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4$ and Π_5 that will be used for predicting 6 structural coefficients :

$\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ dan β_2

Indirect Least Squares (ILS)

ILS method is carried out by using OLS in reduced form

Assumptions by using ILS:

1. *the structural equation must be exactly identified.*
2. *residual variable of the reduced form must fulfill all classical assumption of OLS.*

Otherwise the estimation coefficient will be biased

ILS Application

Production Index (Q)	Price inde (P)	Per capita Income (Y)
93	99	1883
92	100	1909
92	103	1969
96	106	2015
93	106	2126
99	103	2239
95	105	2335
100	100	2403
103	101	2486
104	97	2534
101	100	2610
112	107	2683
113	115	2779
119	164	2945
110	212	2846

Sumber: Supranto, 1995

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \mu_{1t}$$

$$Q_t^s = \beta_0 + \beta_1 P_t + \mu_{2t}$$

$$Q_t^d = Q_t^s$$

$$\alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + \mu_{1t} = \beta_0 + \beta_1 P_t + \mu_{2t}$$

$$(\alpha_1 - \beta_1)P_t = (\beta_0 - \alpha_0) - \alpha_2 Y_t + (\mu_{2t} - \mu_{1t})$$

$$P_t = \frac{(\beta_0 - \alpha_0)}{(\alpha_1 - \beta_1)} - \frac{\alpha_2}{(\alpha_1 - \beta_1)} Y_t + \frac{(\mu_{2t} - \mu_{1t})}{(\alpha_1 - \beta_1)}$$

atau

$$P_t = \frac{(\alpha_0 - \beta_0)}{(\beta_1 - \alpha_1)} + \frac{\alpha_2}{(\beta_1 - \alpha_1)} Y_t + \frac{(\mu_{1t} - \mu_{2t})}{(\beta_1 - \alpha_1)}$$

$$P_t = \pi_1 + \pi_2 Y_t + v_{1t}$$

$$Q_t^s = \beta_0 + \beta_1 (\pi_1 + \pi_2 Y_t + v_{1t}) + \mu_{2t}$$

$$Q_t^s = \beta_0 + \beta_1 \pi_1 + \beta_1 \pi_2 Y_t + \alpha_1 v_{1t} + \mu_{1t}$$

$$Q_t^s = (\beta_0 + \beta_1 \pi_1) + \beta_1 \pi_2 Y_t + (\beta_1 v_{1t} + \mu_{1t})$$

$$Q_t^s = \pi_3 + \pi_4 Y_t + \alpha_1 v_{2t}$$

$$Q_t^s = \beta_0 + \beta_1 P_t + \mu_{2t}$$

$$Q_t^s = (\beta_0 + \beta_1 \pi_1) + \beta_1 \pi_2 Y_t + (\beta_1 v_{1t} + \mu_{1t})$$

$$Q_t^s = \pi_3 + \pi_4 Y_t + \alpha_1 v_{2t}$$

$$\beta_0 + \beta_1 \pi_1 = \pi_3 \quad \text{sehingga} \quad \beta_0 = \pi_3 - \beta_1 \pi_1$$

dan

$$\beta_1 \pi_2 = \pi_4, \text{ sehingga}$$

$$\beta_1 = \frac{\pi_4}{\pi_2}$$

$$\hat{Q} = 51,2911 + 0,4318 P$$

TSLS

$$y_1 = \beta_{10} + \beta_{12}y_2 + \beta_{13}y_3 + \gamma_{11}x_1 + \gamma_{12}x_2 + \dots + \gamma_{1k}x_k + e_1$$

$$y_2 = \beta_{20} + \beta_{21}y_1 + \beta_{23}y_3 + \gamma_{21}x_1 + \gamma_{22}x_2 + \dots + \gamma_{2k}x_k + e_2$$

$$y_3 = \beta_{30} + \beta_{31}y_1 + \beta_{32}y_2 + \gamma_{31}x_1 + \gamma_{32}x_2 + \dots + \gamma_{3k}x_k + e_3$$

Step 1 , to establish reduced form

$$y_1 = \pi_{10} + \pi_{11}x_1 + \pi_{12}x_2 + \dots + \pi_{1k}x_k + v_1$$

$$y_2 = \pi_{20} + \pi_{21}x_1 + \pi_{22}x_2 + \dots + \pi_{2k}x_k + v_2$$

$$y_3 = \pi_{30} + \pi_{31}x_1 + \pi_{32}x_2 + \dots + \pi_{3k}x_k + v_3$$

\hat{y}_1 and \hat{y}_2

Step 2 , to estimate equation 2 and equation 3

$$\hat{y}_1 = \hat{\pi}_{10} + \hat{\pi}_{11}x_1 + \hat{\pi}_{12}x_2 + \cdots + \hat{\pi}_{1k}x_k$$

$$\hat{y}_2 = \hat{\pi}_{20} + \hat{\pi}_{21}x_1 + \hat{\pi}_{22}x_2 + \cdots + \hat{\pi}_{2k}x_k$$

$$\hat{y}_3 = \hat{\pi}_{30} + \hat{\pi}_{13}x_1 + \hat{\pi}_{23}x_2 + \cdots + \hat{\pi}_{3k}x_k$$

Step 3 , substitute the value of \hat{y}_1 and \hat{y}_2 in y1 equation

$$y_1 = \beta_{10} + \beta_{11}\hat{y}_1 + \beta_{12}\hat{y}_2 + \gamma_{11}x_1 + \gamma_{12}x_2 + \dots + \gamma_{1k}x_k$$

, substitute the value of \hat{y}_1 and \hat{y}_3 in y2 equation

$$y_2 = \beta_{20} + \beta_{21}\hat{y}_1 + \beta_{23}\hat{y}_3 + \gamma_{21}x_1 + \gamma_{22}x_2 + \dots + \gamma_{2k}x_k$$

, substitute the value of \hat{y}_1 and \hat{y}_2 in y3 equation

$$y_3 = \beta_{30} + \beta_{31}\hat{y}_1 + \beta_{32}\hat{y}_2 + \gamma_{31}x_1 + \gamma_{32}x_2 + \dots + \gamma_{3k}x_k$$

TSLS

$$\text{Supply: } q_t = \alpha_2 p_t + \varepsilon_t$$

$$\text{Demand: } q_t = \beta_2 p_t + \beta_3 y_t + \beta_4 w_t + u_t$$

Persamaan *reduced form*:

$$q_t = \pi_{12} y_t + \pi_{13} w_t + v_{1t} \quad (\text{i})$$

$$p_t = \pi_{12} y_t + \pi_{13} w_t + v_{1t} \quad (\text{ii})$$

Tahapan Metode Two-Stage Least Squares (2SLS):

1. Pers *reduced form* p_t (ii) diduga menggunakan OLS. Dugaan p_t akan bebas dgn ε_t dan u_t (dlm contoh besar)
2. Pers struktural *supply* diduga dgn mengganti p_t dgn dugaan p_t dari tahap pertama. Penggunaan OLS dlm tahap kedua ini menghasilkan penduga parameter *supply* α_2 yg konsisten

Note: Dugaan p_t merupakan peubah instrumen utk menggantikan p_t dlm pers *supply*

Consider a supply and demand model for truffles:

$$\text{Demand: } Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + e_i^d$$

$$\text{Supply: } Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + e_i^s$$

The rule for identifying an equation is:

- In a system of M equations at least $M - 1$ variables must be omitted from each equation in order for it to be identified
 - In the demand equation the variable PF is not included; thus the necessary $M - 1 = 1$ variable is omitted
 - In the supply equation both PS and DI are absent; more than enough to satisfy the identification condition

The reduced-form equations are:

$$Q_i = \pi_{11} + \pi_{21} PS_i + \pi_{31} DI_i + \pi_{41} PF_i + v_{i1}$$

$$P_i = \pi_{12} + \pi_{22} PS_i + \pi_{32} DI_i + \pi_{42} PF_i + v_{i2}$$

OBS	P	Q	PS	DI	PF
1	29.64	19.89	19.97	2.103	10.52
2	40.23	13.04	18.04	2.043	19.67
3	34.71	19.61	22.36	1.870	13.74
4	41.43	17.13	20.87	1.525	17.95
5	53.37	22.55	19.79	2.709	13.71

	Summary Statistics				
	Mean	18.46	22.02	3.53	22.75
	Std. Dev.	4.61	4.08	1.04	5.33

Reduce form of quantity of Q

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.8951	3.2434	2.4342	0.0221
PS	0.6564	0.1425	4.6051	0.0001
DI	2.1672	0.7005	3.0938	0.0047
PF	-0.5070	0.1213	-4.1809	0.0003

TSLS estimates for demand

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.2795	5.5439	-0.7719	0.4471
P	-0.3745	0.1648	-2.2729	0.0315
PS	1.2960	0.3552	3.6488	0.0012
DI	5.0140	2.2836	2.1957	0.0372

Reduce form of quantity of P

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-32.5124	7.9842	-4.0721	0.0004
PS	1.7081	0.3509	4.8682	0.0000
DI	7.6025	1.7243	4.4089	0.0002
PF	1.3539	0.2985	4.5356	0.0001

TSLS estimates for supply

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.0328	1.2231	16.3785	0.0000
P	0.3380	0.0249	13.5629	0.0000
PF	-1.0009	0.0825	-12.1281	0.0000

Order Condition

$K - k = m-1$, exactly identified

$K - k > m-1$, over identified

$K - k < m-1$, not identified

$$y_{1t} = \beta_{10} + \beta_{12}y_{2t} + \beta_{13}y_{3t} + \gamma_{11}x_{1t} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{23}y_{3t} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + u_{2t}$$
$$y_{3t} = \beta_{30} + \beta_{31}y_{1t} + \gamma_{31}x_{1t} + \gamma_{32}x_{2t} + u_{3t}$$
$$y_{4t} = \beta_{40} + \beta_{41}y_{1t} + \beta_{42}y_{2t} + \gamma_{43}x_{3t} + u_{4t}$$

No. Persamaan	$(K - k)$	$(m - 1)$	Keterangan
1	$3 - 1 = 2$	$3 - 1 = 2$	Tepat Teridentifikasi
2	$3 - 2 = 1$	$2 - 1 = 1$	Tepat Teridentifikasi
3	$3 - 2 = 1$	$2 - 1 = 1$	Tepat Teridentifikasi
4	$3 - 1 = 2$	$3 - 1 = 2$	Tepat Teridentifikasi

Rank condition

$$y_{1t} = \beta_{10} + \beta_{12}y_{2t} + \beta_{13}y_{3t} + \gamma_{11}x_{1t} + u_{1t}$$

$$y_{2t} = \beta_{20} + \beta_{23}y_{3t} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + u_{2t}$$

$$y_{3t} = \beta_{30} + \beta_{31}y_{1t} + \gamma_{31}x_{1t} + \gamma_{32}x_{2t} + u_{3t}$$

$$y_{4t} = \beta_{40} + \beta_{41}y_{1t} + \beta_{42}y_{2t} + \gamma_{43}x_{3t} + u_{4t}$$

Koefisien-koefisien								
No.	C	y_1	y_2	y_3	y_4	x_1	x_2	x_3
1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

No.	$(K - k)$	$(m - 1)$	Identifikasi <i>Order</i>	Identifikasi Rank	Kesimpulan
1	$3 - 1 = 2$	$3 - 1 = 2$	Tepat Teridentifikasi	$Rank < M - 1 = 3$	Tidak Teridentifikasi
2	$3 - 2 = 1$	$2 - 1 = 1$	Tepat Teridentifikasi	$Rank < M - 1 = 3$	Tidak Teridentifikasi
3	$3 - 2 = 1$	$2 - 1 = 1$	Tepat Teridentifikasi	$Rank < M - 1 = 3$	Tidak Teridentifikasi
4	$3 - 1 = 2$	$3 - 1 = 2$	Tepat Teridentifikasi	$Rank = M - 1 = 3$	Tepat Teridentifikasi

Rank

$$A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

$$\det(A) = (0)(-\gamma_{32})(-\gamma_{43}) + (-\gamma_{22})(0)(1) + (0)(0)(0) - (1)(-\gamma_{32})(0) - (0)(0)(0) - (-\gamma_{43})(0)(-\gamma_{22})$$

$$= 0 + 0 + 0 - 0 - 0$$

$$= 0$$

$$B = \begin{bmatrix} -\beta_{13} & -\gamma_{11} & 0 \\ -\beta_{23} & -\gamma_{21} & -\gamma_{22} \\ 1 & -\gamma_{31} & -\gamma_{32} \end{bmatrix}$$

$$\begin{aligned} \det(B) &= (-\beta_{13})(-\gamma_{21})(-\gamma_{32}) + (-\gamma_{11})(-\gamma_{22})(1) + (0)(-\beta_{23})(-\gamma_{31}) - (1)(-\gamma_{21})(0) - (-\gamma_{31})(-\gamma_{22})(-\beta_{13}) - (-\gamma_{32})(-\beta_{23})(-\gamma_{11}) \\ &= -\beta_{13}\gamma_{21}\gamma_{32} + \gamma_{11}\gamma_{22}1 + \gamma_{31}\gamma_{22}\beta_{13} + \gamma_{32}\beta_{23}\gamma_{11} \\ &\neq 0 \end{aligned}$$

Make a relationship by putting arrows

