INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -11

Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

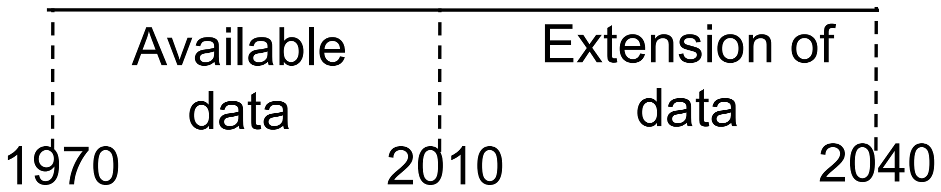
Summary of the previous lecture

- Introduction to time series analysis
  - Realization ; Ensemble
  - Stationarity
  - Auto covariance Function
  - Auto corrleation and correlogram

DATA EXTENSION & FORECASTING

Data Extension & Forecasting

e.g., Stream flow records for reservoir planning



Available data

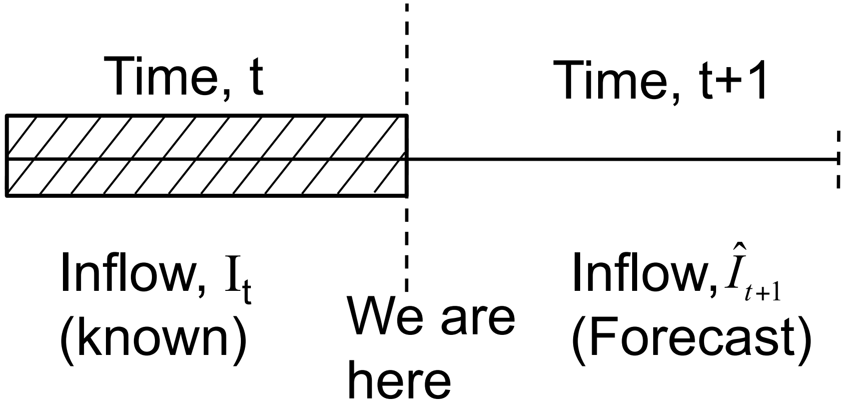
Extension of data

1970

2010

2040

Data forecasting



Time, t

Time, t+1

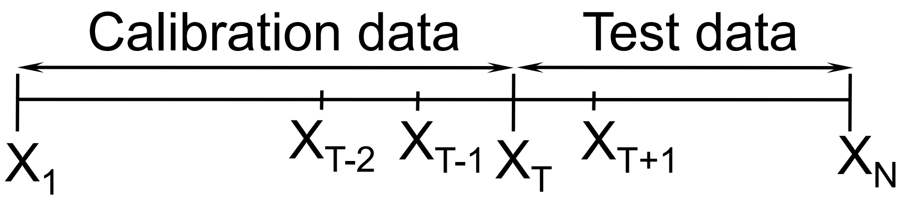
Inflow, I_t
(known)

We are here

Inflow, $\hat{I}_{t+1}$
(Forecast)

$$\hat{I}_{t+1} = f\left[I_t, I_{t-1}, \ldots\ldots\ldots.\right]$$

Data Extension & Forecasting



Calibration data Test data

X_1 X_{T-2} X_{T-1} X_T X_{T+1} X_N

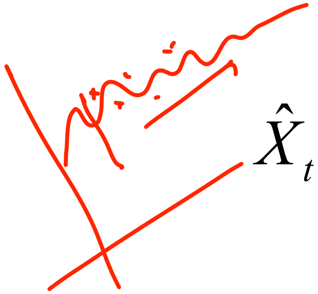
Use first ‘T’ values to build the model; rest of data to validate it

F_{T+1}, F_{T+2} ……… F_N : forecasts obtained from the model

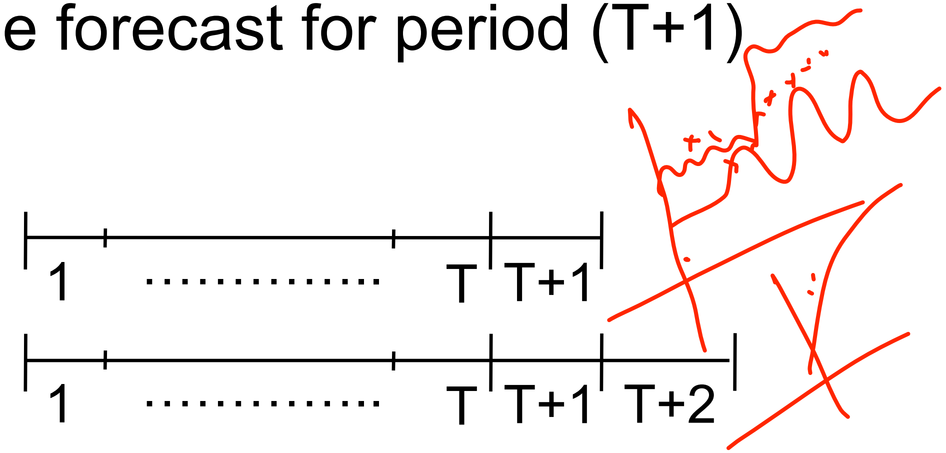
$$\left.\begin{array}{l} (X_{T+1} - F_{T+1}) \\ (X_{T+2} - F_{T+2}) \\ . \\ . \\ (X_N - F_N) \end{array}\right\} \text{Forecast errors}$$

Data Extension & Forecasting

Method of simple averages: take the average of all the data up to period ‘T’ as the forecast for period (T+1)



$$\hat{X}_{t+1} = F_{T+1} = \frac{\sum_{t=1}^{T} X_t}{T}$$



$$\hat{X}_{t+2} = F_{T+2} = \frac{\sum_{t=1}^{T+1} X_t}{T+1}$$ and so on

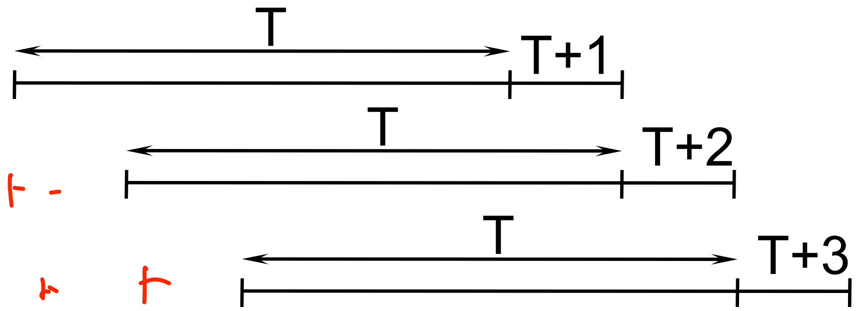
For series with jumps, trends and periodicities, this is not a good procedure

Example-1

Data	Forecast
105	-
115	110
103	107.67
108	107.75
120	110.2
97	108
110	108.28
121	109.87
117	110.67
79	107.5

Data Extension & Forecasting

Method of Moving Averages (MA)



- As a new observation becomes available, new average is computed by dropping the oldest observation and including the newest one.
- No. of data points used for computing the average remains the same
- Uses the latest ‘T’ periods of known data

Example-2

Data	MA (3)
105	-
115	-
103	-
108	107.67
120	108.67
97	110.33
110	108.33
121	109
117	109.33
79	116

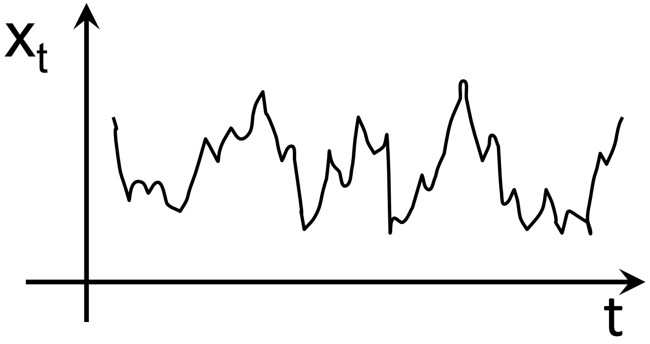
Example-2

Data	MA (3)	MA (3 x 3)
105	-	
115	-	
103	-	
108	107.67	
120	108.67	
97	110.33	108.89
110	108.33	109.11
121	109	109.22
117	109.33	108.89
79	116	111.44

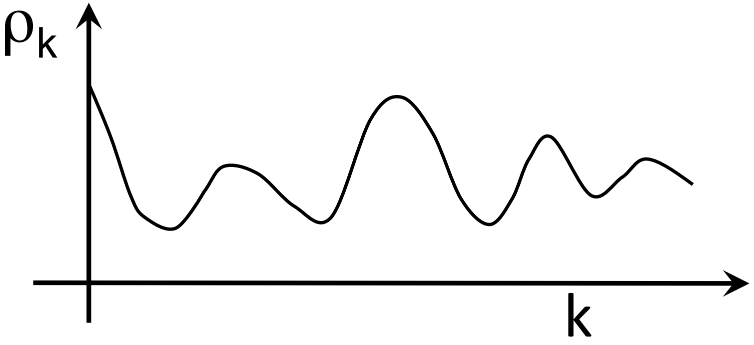
Data Generation – Uncorrelated Data

Purely random stochastic process:

Plot the time series

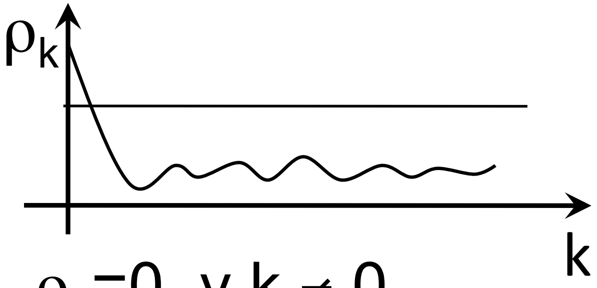


Plot the correlogram



Data Generation – Uncorrelated Data

If the correlogram indicates that the time series is purely random



$\rho_k=0, \forall\ k \neq 0$

Mainly used for flood peaks, storm intensities, short duration rainfall etc .

Not useful for stream flows, seasonal rainfall, and such long time processes.

- X_t , X_{t-k} are independent
- Distribution of X_t is known
- Generate X_t using data generation technique to follow given distribution with parameters estimated from sample

$F(y) = R_u$

Solve for y

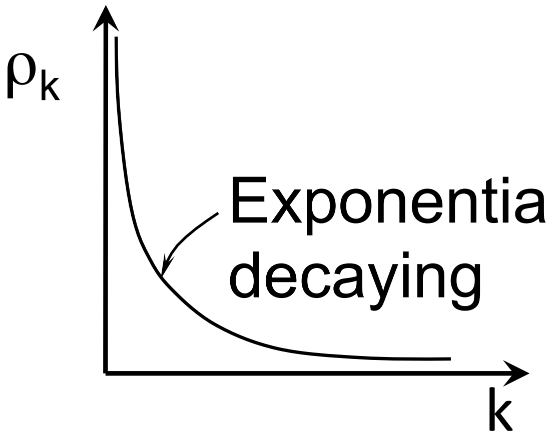
Data Generation – Serially Correlated Data

- Most hydrologic time series exhibit serial dependence
  e.g., X(t) correlated with X(t-τ)

$$\rho_k \cong (\rho_1)^k$$

$$\rho_k \to 0,\ k \to \infty$$

$$P[X_t \mid X_{t-1}, X_{t-2} \cdots - X_0] = P[X_t \mid X_{t-1}]$$



First order Markov process

Data Generation – Serially Correlated Data

First order Markov process:

Random component

$$X_{t+1} = \underbrace{\mu_x + \rho_1 (X_t - \mu_x)}_{\text{Deterministic component}} + \varepsilon_{t+1}$$

$\varepsilon \sim$ Mean 0 and variance σ_ε^2

This model is stationary w.r.t both mean and variance

Data Generation – Serially Correlated Data

$$E[X_{t+1}] = E[\mu_x + \rho_1 (X_t - \mu_x) + \varepsilon_{t+1}]$$

$$= E[\mu_x] + \rho_1\{E[X_t] - E[\mu_x]\} + E[\varepsilon_{t+1}]$$

$$= \mu_x + \rho_1(\mu_x - \mu_x) + 0$$

$$= \mu_x$$

$$\sigma_X^2 = E\left[X^2\right] - \left(E\left[X\right]\right)^2$$

$$= E\left[\left(\mu_x + \rho_1\left(X_t - \mu_x\right) + \varepsilon_{t+1}\right)^2\right] - \left(E\left[X_{t+1}\right]\right)^2$$

Data Generation – Serially Correlated Data

$$\sigma_X^2 = E\left[\mu_x^2 + \rho_1^2\left(X_t - \mu_x\right)^2 + \varepsilon_{t+1}^2 + 2\mu_x\rho_1\left(X_t - \mu_x\right) + \right.$$

$$\left. + 2\varepsilon_{t+1}\rho_1\left(X_t - \mu_x\right) + 2\mu_x\varepsilon_{t+1}\right] - \left(E\left[X_{t+1}\right]\right)^2$$

$$= E\left[\mu_x^2\right] + \rho_1^2 E\left[\left(X_t - \mu_x\right)^2\right] + E\left[\varepsilon_{t+1}^2\right] + 2\mu_x\rho_1 E\left[\left(X_t - \mu_x\right)\right]$$

$$+ 2\rho_1 E\left[\varepsilon_{t+1}\right]E\left[\left(X_t - \mu_x\right)\right] + 2\mu_x E\left[\varepsilon_{t+1}\right] - \left(E\left[X_{t+1}\right]\right)^2$$

$$= \mu_x^2 + \rho_1^2 E\left[\left(X_t - \mu_x\right)^2\right] + E\left[\varepsilon_{t+1}^2\right] + 0 + 0 + 0 - \mu_x^2$$

$$= \rho_1^2\sigma_X^2 + \sigma_\varepsilon^2$$

$$\sigma_\varepsilon^2 = \rho_1^2\left(1 - \sigma_X^2\right)$$

Data Generation – Serially Correlated Data

If X ~ N(μ_x, σ_x^2) then ε ~ N(0, σ_ε^2)

If {u_t} ~ N(0, 1) , {$u_t\sigma_\varepsilon$} $\left(\text{i.e., } u_t\sigma_x\sqrt{1-\rho_1^2}\right)$ is N(0, σ_ε^2)

$$X_{t+1} = \mu_x + \rho_1\left(X_t - \mu_x\right) + u_{t+1}\sigma_x\sqrt{1-\rho_1^2}$$

Standard normal deviate

First order stationary Markov model
Or
Thomas Fiering model (Stationary)

Data Generation – Serially Correlated Data

To generate data using First order Markov model,

$$X_{t+1} = \mu_x + \rho_1 \left(X_t - \mu_x \right) + u_{t+1} \sigma_x \sqrt{1 - \rho_1^2}$$

•Known sample estimates of μ_x, σ_x, ρ_1

•Assume X_1 (may be assumed to be μ_x)

•Generate values X_2, $X_{3,}$ $X_{4,}$ X_5 ……

•Generate a large set of values and discard first 50-100 values to ensure that the effect of initial value dies down

•Negative value: retain it for generating next value, set it to zero, in applications.

Example-3

Consider the annual stream flow data (in cumecs) at a river for 29 years

S.No.	Data	S.No.	Data	S.No.	Data
1	1093.31	11	1042.33	21	1444.97
2	1636.87	12	1492.13	22	1203.08
3	1485.67	13	1205.90	23	910.73
4	1579.51	14	1245.77	24	883.59
5	1443.00	15	1197.81	25	970.98
6	1327.40	16	1754.55	26	1001.92
7	1108.70	17	1108.56	27	1434.91
8	928.10	18	957.64	28	1635.00
9	840.83	19	1425.80	29	1875.78
10	1447.03	20	1128.62		

Example-3 (contd.)

For the data,

$\mu_x = 1269$, $\sigma_x = 281$, $\rho_1 = 0.255$

$$c_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{n}$$

$$r_1 = \frac{c_1}{c_0}; \qquad c_0 = \sigma_x^2$$

Assume $X_1 = \mu_x = 1269$~~.33~~

$$X_2 = \mu_x + \rho_1 (X_1 - \mu_x) + u_{t+1}\sigma_x\sqrt{1-\rho_1^2}$$

$$= 1269 + 0.255(1269 - 1269) + (-0.464)281\sqrt{1-0.255^2}$$

$$= 1143$$

x_2 , μ_x

$$X_3 = 1269 + 0.255(1143 - 1269) + (0.335)281\sqrt{1-0.255^2}$$

$$= 1328$$

Example-3 (contd.)

$$X_4 = 1269 + 0.255\left(1328 - 1269\right) + \left(-0.051\right)281\sqrt{1-0.255^2}$$

$$= 1270$$

$$X_5 = 1269 + 0.255\left(1270 - 1269\right) + \left(1.226\right)281\sqrt{1-0.255^2}$$

$$= 1602$$

Data Generation – Serially Correlated Data

First order Markov model with non-stationarity:

- First order stationary Markov model assumes that the process is stationary in mean, variance and lag-one auto correlation
- The model is generalized to account for non-stationarity (mainly due to seasonality/periodicity) in hydrologic data to some extent
- A main application of this model is in generating the monthly stream flows with pronounced seasonality.
- Periodicity may affect not only the mean, but all the moments of data including the serial correlations.

Data Generation – Serially Correlated Data

$$X_{j+1} = \mu_x + \rho_1 \left(X_j - \mu_x \right) + t_{j+1} \sigma_x \sqrt{1 - \rho_1^2}$$

Stationary First order Markov Model

First order Markov model with non-stationarity, for stream flow generation:

$$X_{i,j+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j} \left(X_{ij} - \mu_j \right) + t_{i,j+1} \sigma_{j+1} \sqrt{1 - \rho_j^2}$$

ρ_j is serial correlation between flows of jth month and j+1th month.

$t_{i,\, j+1} \sim N(0, 1)$

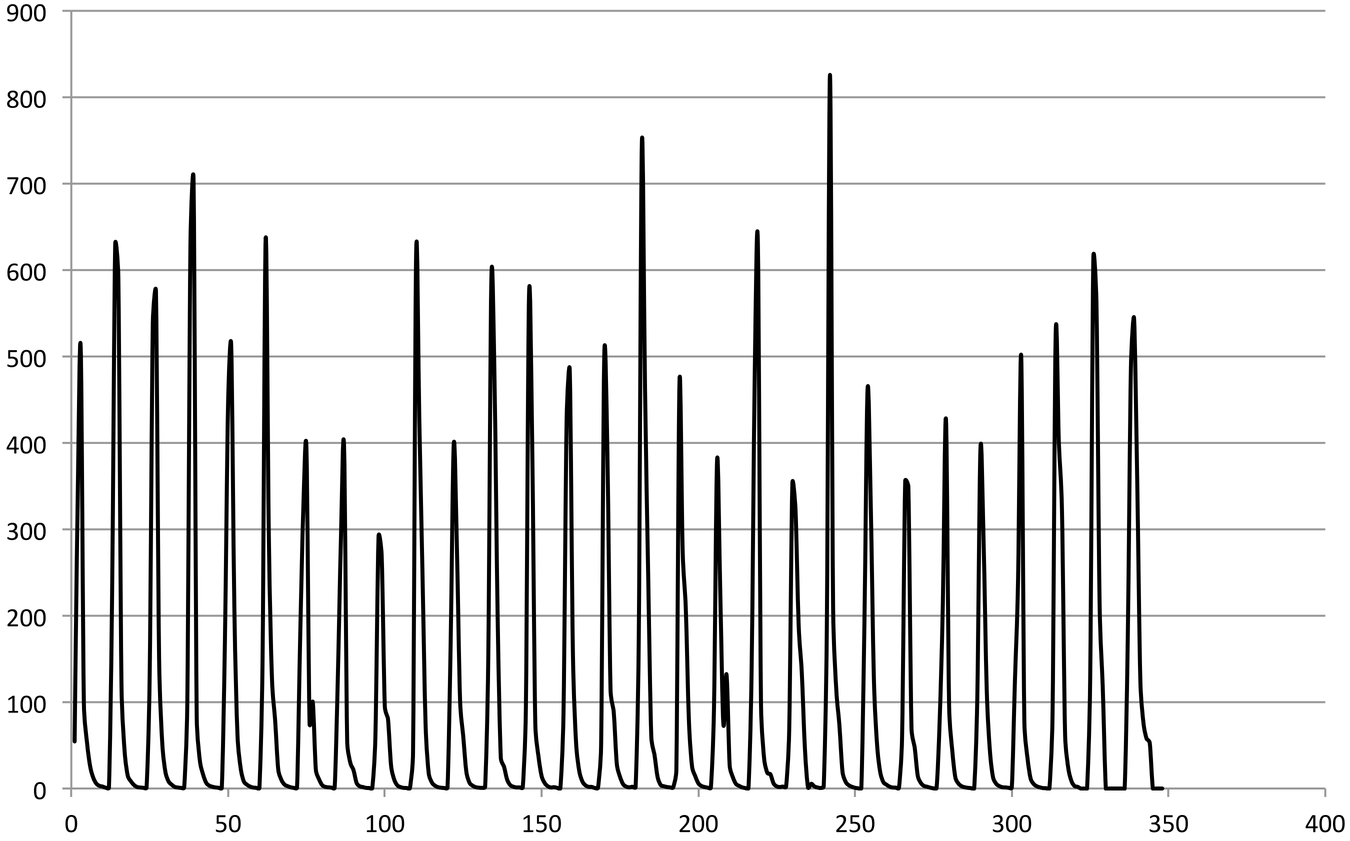
Example-4

The monthly stream flow (in cumec) for a river is available for 29 years (only 12 years data is given here)

SL. NO.	YEAR	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY
1	1979-80	54.60	325.40	509.50	99.40	53.50	25.80	12.50	5.60	3.10	2.20	0.90	0.81
2	1980-81	220.78	629.16	591.32	120.33	43.33	14.83	8.41	4.05	1.73	1.12	0.85	0.96
3	1981-82	131.30	538.89	574.21	151.06	53.03	19.49	8.38	4.51	1.89	1.11	0.74	1.06
4	1982-83	100.19	630.02	702.07	83.29	32.45	16.60	6.80	3.33	2.03	1.23	0.85	0.65
5	1983-84	171.30	444.30	512.30	211.00	62.40	24.00	8.40	4.50	2.30	1.10	0.80	0.60
6	1984-85	147.80	636.20	293.50	127.70	79.70	22.10	10.10	4.60	2.70	1.40	0.70	0.90
7	1985-86	174.50	323.30	393.20	75.40	100.60	21.80	10.90	4.00	1.90	1.40	1.00	0.70
8	1986-87	126.40	288.30	395.30	54.40	29.80	21.40	6.40	2.60	1.70	0.70	0.60	0.50
9	1987-88	60.50	291.00	269.60	95.09	80.84	26.39	10.37	3.68	1.65	0.71	0.62	0.38
10	1988-89	40.95	620.00	427.60	251.80	74.73	17.71	7.05	3.33	1.51	0.87	0.59	0.90
11	1989-90	167.10	398.80	277.80	102.70	61.10	19.54	6.79	3.33	1.52	0.96	0.77	1.93
12	1990-91	150.80	591.50	471.20	197.00	35.67	25.62	10.52	4.02	2.10	1.22	1.32	1.16

Example-4 (contd.)

Time series of monthly stream flow for 29 years



Example-4 (contd.)

S.No.	Month	Mean	Stdev.	Lag-1 correlation
1	JUN	117.49	52.24	0.348
2	JUL	474.50	150.18	0.154
3	AUG	421.39	126.53	0.169
4	SEP	145.94	77.65	0.365
5	OCT	66.61	30.67	0.490
6	NOV	22.99	13.26	0.798
7	DEC	10.30	9.82	0.955
8	JAN	5.55	9.16	-0.385
9	FEB	1.91	0.74	0.733
10	MAR	1.09	0.54	0.654
11	APR	0.76	0.51	0.676
12	MAY	0.80	0.60	-0.005

Example-4 (contd.)

Assume $X_1 = \mu_1 = 117.49$;

$\sigma_1 = 52.24$, $\rho_1 = 0.348$

$\mu_2 = 474.5$, $\sigma_2 = 150.18$,

$$X_{1,2} = \mu_2 + \rho_1 \frac{\sigma_2}{\sigma_1}\left(X_{1,1} - \mu_1\right) + t_{1,2}\sigma_2\sqrt{1-\rho_1^2}$$

$$= 474.5 + 0.348\frac{150.18}{52.24}\left(117.49 - 117.49\right)$$

$$+ 0.335 * 150.18\sqrt{1-0.348^2}$$

= 521.67

Example-4 (contd.)

$X_{1,2}$ =521.67, μ_2 = 474.5; σ_2 = 150.18, ρ_2 = 0.154

μ_3 = 421.39, σ_3 = 126.53,

$$X_{1,3} = 421.39 + 0.154\frac{126.53}{150.18}\left(521.67 - 474.5\right)$$

$$+ 0.377 * 126.53\sqrt{1 - 0.154^2}$$

= 474.64

Example-4 (contd.)

$X_{1,3}$ =474.64, μ_3 = 421.39; σ_3 = 126.53, ρ_3 = 0.169

μ_4 = 145.94, σ_4 = 77.65,

$$X_{1,4} = 145.94 + 0.169\frac{77.65}{126.53}\left(474.64 - 421.39\right)$$

$$+ 0.379 * 77.65\sqrt{1 - 0.169^2}$$

= 180.45