



Chương 14

CHUỖI THỜI GIAN VÀ

DỰ BÁO TRÊN CHUỖI THỜI GIAN

Ths. Nguyễn Tiến Dũng

Viện Kinh tế và Quản lý, Trường ĐH Bách khoa Hà Nội

Email: dung.nguyentien3@hust.edu.vn

MỤC TIÊU CỦA CHƯƠNG

- Sau khi học xong chương này, người học sẽ
 - Phát biểu được chuỗi thời gian là gì
 - Phân biệt được các khái niệm và các cách tiếp cận trong dự báo
 - Thực hiện được các phương pháp dự báo dựa trên chuỗi thời gian: lượng tăng giảm tuyệt đối, tốc độ phát triển bình quân
 - Thực hiện được một số phương pháp dự báo theo mô hình nhân

CÁC NỘI DUNG CHÍNH

- 14.1 CHUỖI THỜI GIAN
- 14.2 CÁC PHƯƠNG PHÁP DỰ BÁO DỰA TRÊN CHUỖI THỜI GIAN
- 14.3 DỰ BÁO BẰNG MÔ HÌNH NHÂN

14.1 CHUỖI THỜI GIAN

- 14.1.1 Khái niệm
- 14.1.2 Các đại lượng mô tả chuỗi thời gian

14.1.1 Khái niệm

- Time-series data
- Chuỗi các giá trị của một chỉ tiêu NC (đại lượng) được sắp xếp theo thứ tự thời gian
 - $Y = \{Y_1, Y_2, Y_3, \dots Y_n\}$
- Chuỗi số thời kỳ:
 - DL thu thập trong kỳ
 - Có tính cộng: cộng các thời kỳ khác nhau với nhau được
 - TD
- Chuỗi số thời điểm
 - DL thu thập tại một thời điểm
 - Không cộng lại với nhau để đưa ra con số tích lũy được
 - TD

14.1.2 Các đại lượng mô tả chuỗi thời gian

● 14.1.2.1 Giá trị TB

- Chuỗi thời kỳ

- Chuỗi thời điểm

- Nếu khoảng cách giữa các thời điểm bằng nhau

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$\bar{Y} = \frac{1}{n-1} (0,5Y_1 + Y_2 + Y_3 + \dots + Y_{n-1} + 0,5Y_n)$$

- Nếu khoảng cách giữa các thời điểm không bằng nhau, nhưng thời gian NC là liên tục

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i t_i}{\sum_{i=1}^n t_i}$$

- 14.1.2.2 Lượng tăng giảm tuyệt đối
 - Lượng tăng giảm tuyệt đối liên hoàn
 - Lượng tăng giảm tuyệt đối định gốc
 - Lượng tăng giảm tuyệt đối TB

$$\delta_i = Y_i - Y_{i-1} \quad (i=\overline{2,n})$$

$$\Delta_i = Y_i - Y_1 \quad (i=\overline{2,n})$$

$$\Delta_n = \sum_{i=2}^n \delta_i$$

$$\bar{\delta} = \frac{1}{n-1} \sum_{i=2}^n \delta_i = \frac{\Delta_n}{n-1}$$

14.1.2.3 Tốc độ phát triển

- Tốc độ phát triển liên hoàn
- Tốc độ phát triển định gốc
- Liên hệ giữa tốc độ phát triển liên hoàn và tốc độ phát triển định gốc
- Tốc độ phát triển TB

$$t_i = \frac{Y_i}{Y_{i-1}}$$

$$T_i = \frac{Y_i}{Y_1}$$

$$T_n = \frac{Y_n}{Y_1} = \prod_{i=2}^n t_i$$

$$\bar{t} = \sqrt[n-1]{\prod_{i=2}^n t_i} = \sqrt[n-1]{T_n}$$

14.1.2.4 Tốc độ tăng trưởng

- Tốc độ tăng trưởng liên hoàn

$$a_i = \frac{Y_i - Y_{i-1}}{Y_{i-1}} = t_i - 1$$

- Tốc độ tăng trưởng định gốc $A_i = \frac{Y_i - Y_1}{Y_1} = \frac{\Delta_i}{Y_1} = T_i - 1$

- Tốc độ tăng trưởng TB

$$\bar{a} = \bar{t} - 1$$

14.1.2.5 Trị tuyệt đối của 1% tăng trưởng liên hoàn

- Phản ánh 1% tăng giảm của 2 thời kỳ đứng liền nhau tương ứng với số tuyệt đối là bao nhiêu.

$$g_i = \frac{\delta_i}{a_i \cdot 100} = \frac{Y_i - Y_{i-1}}{\frac{Y_i - Y_{i-1}}{Y_{i-1}} \times 100} = \frac{Y_{i-1}}{100}$$

14.2 DỰ BÁO DỰA TRÊN CHUỖI THỜI GIAN

- Hoạch định tốt → Thành công cao
- Dự báo → hoạch định (lập kế hoạch)
- Các cách tiếp cận trong DB
 - Cách tiếp cận định tính: phỏng vấn sâu, thảo luận nhóm đối với chuyên gia và khách hàng
 - Cách tiếp cận định lượng:
 - Sử dụng X để dự báo Y
 - Sử dụng các GT quá khứ của Y để dự báo các GT tương lai của Y
 - Các điều kiện và giả định để DB định lượng
 - Có sẵn DL quá khứ
 - Có thể lượng hoá DL quá khứ
 - Các quy luật quá khứ sẽ tiếp diễn trong tương lai

14.2.1 Một số vấn đề liên quan đến dự báo

- 14.2.1.1 Thời đoạn DB
 - Là tần suất thời gian mà DL phục vụ dự báo được thu thập, như ngày, tuần, tháng, quý, năm.
- 14.2.1.2 Tầm xa DB
 - DB tức thì: dưới 1 tháng
 - DB ngắn hạn: từ 1 đến 3 tháng
 - DB trung hạn: từ 3 tháng đến hơn 1 năm.
 - DB dài hạn: từ 2 năm trở lên

14.2.1.3 Các chỉ tiêu đánh giá mức độ phù hợp của mô hình dự báo

- Sai số tuyệt đối TB – MAE (Mean Absolute Error)

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n}$$

$$Y = \{Y_1; Y_2; \dots, Y_n\}$$

$$F = \{F_1; F_2; \dots, F_n\}$$

$$e_i = Y_i - F_i$$

- Sai số phần trăm tuyệt đối TB – MAPE (Mean Absolute Percent Error)

$$MPAE = \frac{\sum_{i=1}^n |e_i| / Y_i}{n}$$

- Sai số bình phương TB – MSE (Mean Square Error) và Căn bậc hai của sai số bình phương TB

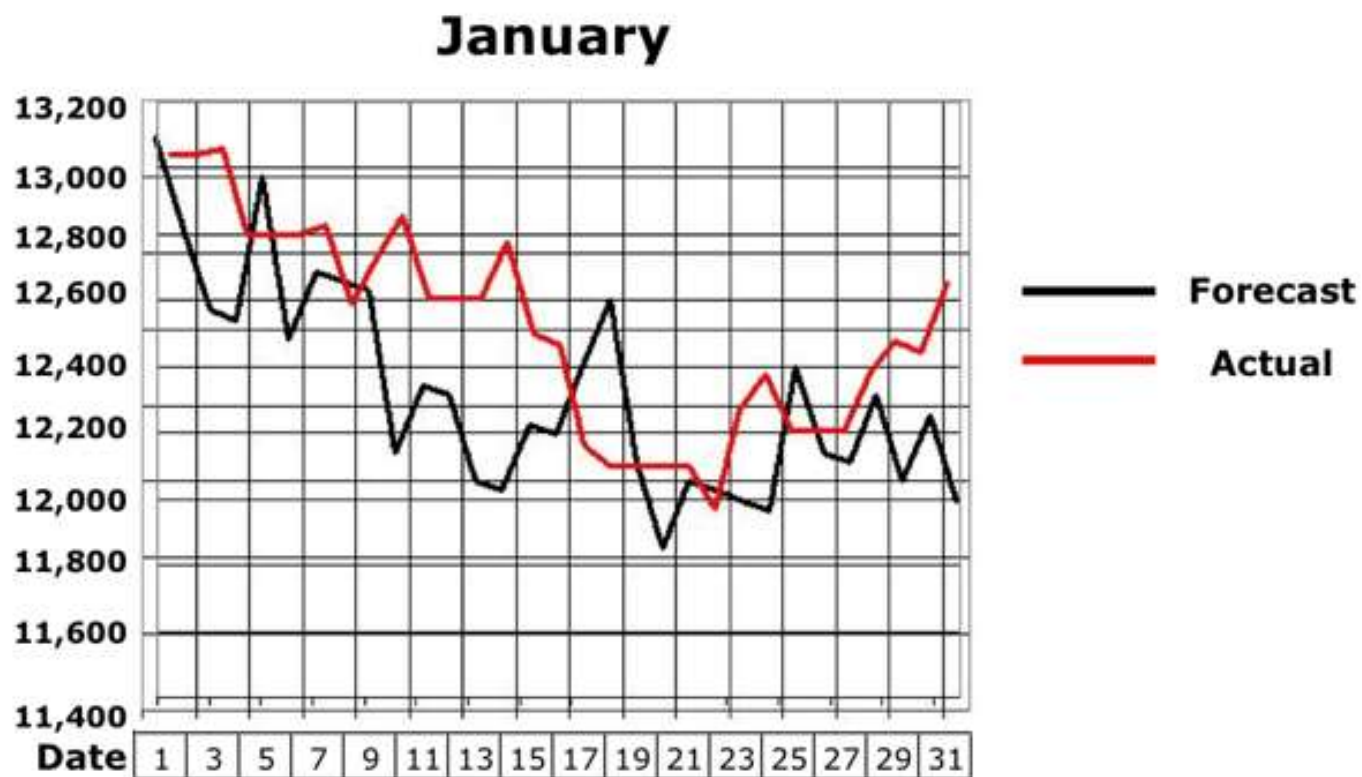
$$MSE = \frac{\sum_{i=1}^n e_i^2}{n}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}}$$

- Chỉ số U

$$U = \frac{\text{RMSE của mô hình dự báo đang sử dụng}}{\text{RMSE của mô hình dự báo ngây thơ (naive)}}$$

- Đánh giá trực quan bằng đồ thị



14.2.2 Các phương pháp DB đơn giản

- 14.2.2.1 Dự báo dựa vào lượng tăng trưởng tuyệt đối TB
- 14.2.2.2 Dự báo dựa vào tốc độ phát triển TB
- 14.2.2.3 Dự báo bằng phương pháp TB trượt (moving average)

14.2.2.1 Dự báo dựa vào lượng tăng trưởng tuyệt đối TB

- L: tầm xa dự báo ($L = 1, 2, 3, \dots$)
- F_{t+L} : giá trị dự báo ở thời gian $t+L$
- $\bar{\delta}$: lượng tăng trưởng tuyệt đối TB
- TD

$$F_{n+L} = Y_n + \bar{\delta} \cdot L$$

t	1	2	3	4
Y	100	118	121	?
delta	-	18	3	

$$\bar{\delta} = \frac{1}{2}(18 + 3) = 10,5$$

$$F_4 = F_{3+1} = Y_3 + 1 \cdot \bar{\delta} = 121 + 10,5 = 131,5$$

14.2.2.2 Dự báo dựa vào tốc độ phát triển trung bình

t	1	2	3	4
Y	100	118	121	?

$$F_{n+L} = Y_n \cdot (\bar{t})^L$$

$$\bar{t} = \sqrt{Y_3 / Y_1} = \sqrt{121 / 100} = 1,1$$

$$F_4 = F_{3+1} = Y_3 \cdot \bar{t} = 121 \times 1,1 = 133,1$$

14.2.2.3 Dự báo bằng phương pháp trung bình trượt

- Số điểm lấy TB:
 - $m = 2k+1$ hoặc $m = 2k$
 - Nếu m lẻ, không phải trung tâm hoá
 - Nếu m chẵn, phải trung tâm hoá
 - Chọn m bằng bao nhiêu?
 - Dãy số có mức độ biến động ít, chọn m nhỏ (TD, $m=3$)
 - Dãy số có mức độ biến động nhiều, chọn m lớn hơn ($m = 5, 7 \dots$)
 - Phương pháp “Trial-and-error”: thử các giá trị m khác nhau, phương pháp nào có MSE nhỏ nhất thì chọn.
 - m càng lớn, đường dự báo càng trơn

14.2.2.4 Mô hình ngoại suy xu thế

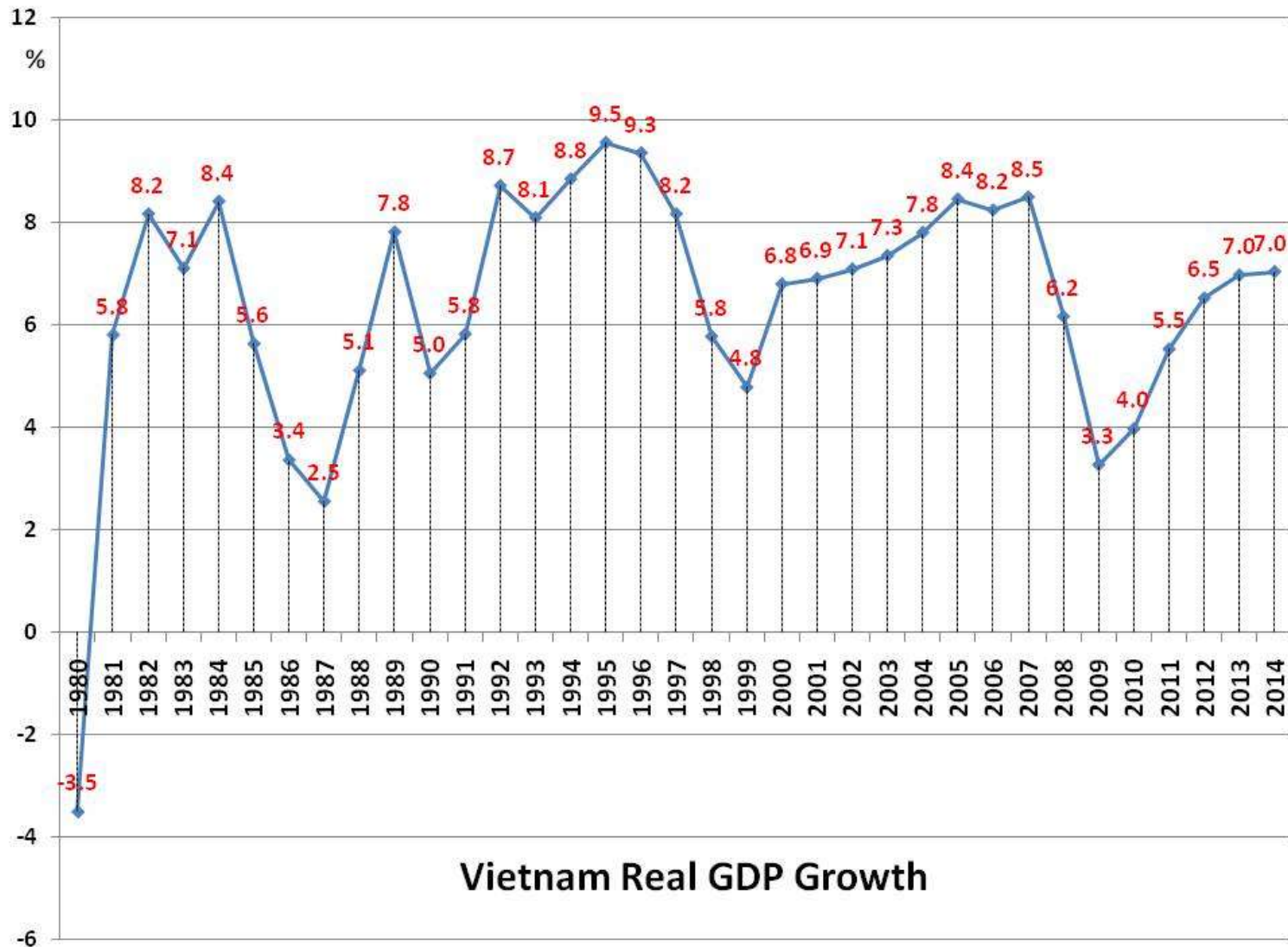
- Sử dụng các mô hình hồi quy tuyến tính đơn biến và đa biến để dự báo

$$\hat{Y} = b_0 + b_1 X$$

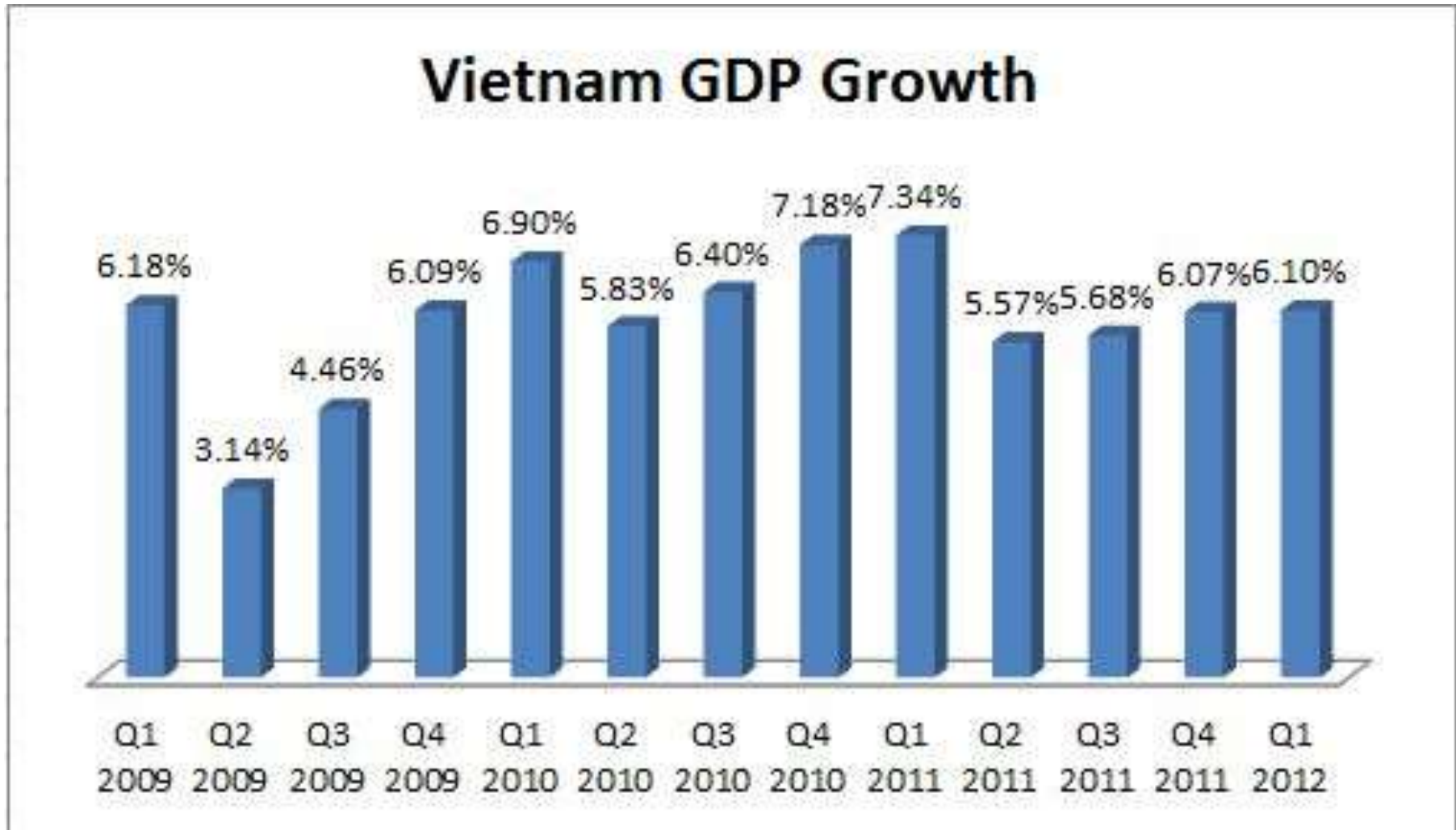
14.3 DỰ BÁO BẰNG MÔ HÌNH NHÂN

- Mô hình nhân (Multiplication Model)
- Chuỗi số liệu theo thời gian: $Y = \{Y_1, Y_2, \dots, Y_n\}$
- Các thành phần có thể có mặt:
 - TP xu thế (Trend) T_i
 - TP chu kỳ dài hạn (Cyclical) C_i
 - TP mùa vụ (Seasonal) S_i
 - TP bất thường (Erratic) E_i
- Mô hình nhân: $Y_i = T_i \cdot C_i \cdot S_i \cdot E_i$
- Quy trình dự báo theo mô hình nhân
 - Nhận diện các thành phần của chuỗi
 - Tách riêng các thành phần
 - Lắp ghép chúng lại để có giá trị dự báo mong muốn

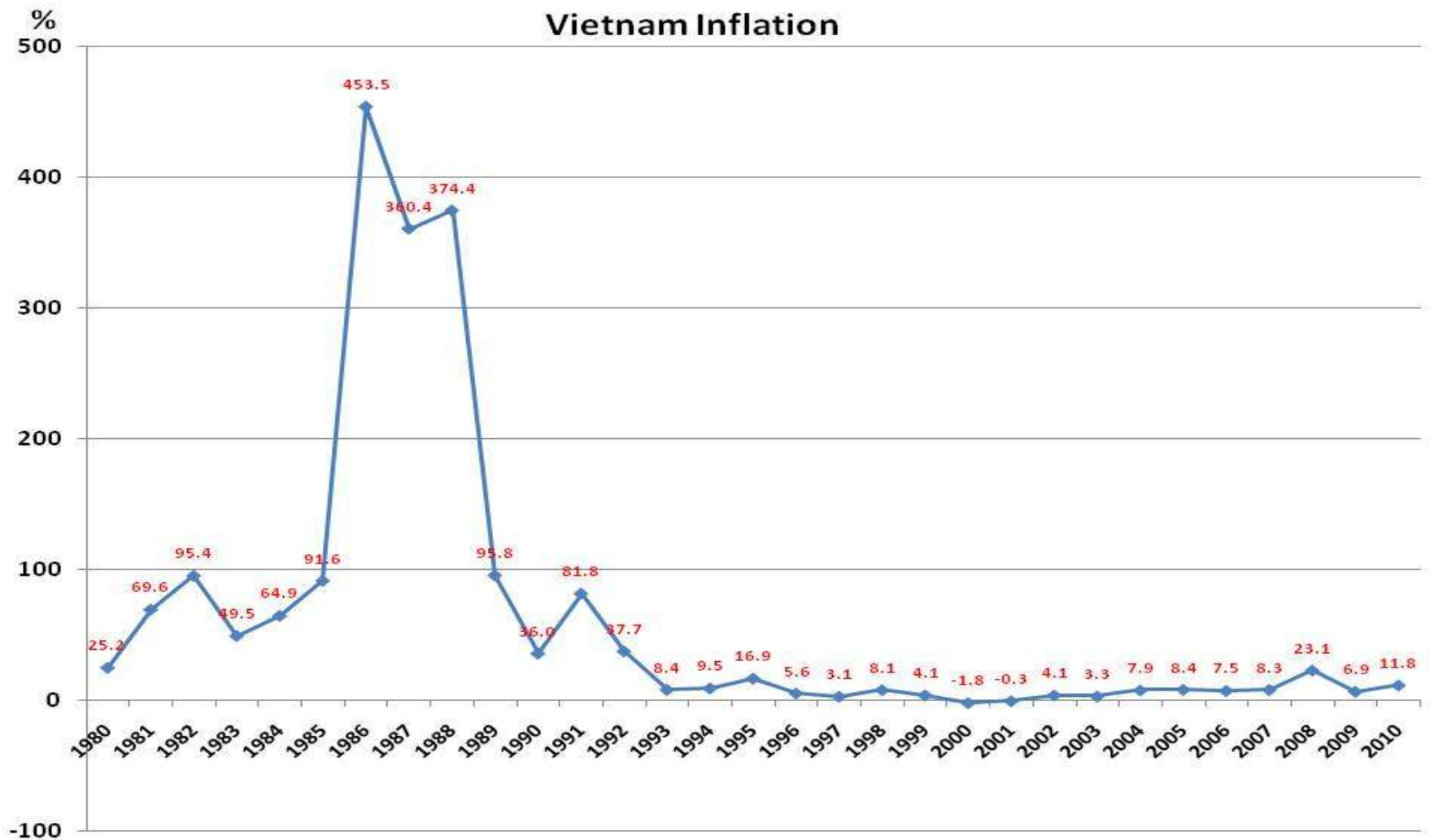
Dãy DL có thành phần xu hướng và chu kỳ



Dãy dữ liệu có thành phần xu hướng (T), chu kỳ (C) và mùa vụ (S)



Tỷ lệ lạm phát 1980-2010



Quy trình dự báo theo PP Holt-Winter



- B1: Tách thành phần mùa vụ và bất thường ra khỏi chuỗi dữ liệu bằng phương pháp trung bình trượt trung tâm hoá (Centered Moving Average)
- Nếu DL theo quý, chọn số điểm lấy TB trượt là $m = 4$, rồi trung tâm hoá
- MA: Moving Average
- CMA: Centered Moving Average

$$MA_{t-0,5} = \frac{Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1}}{4}$$

$$MA_{t+0,5} = \frac{Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}}{4}$$

$$CMA_t = MA_{t-0,5} + MA_{t+0,5}$$

$$t \geq 3$$

- Nếu dữ liệu thu thập theo tháng, chọn số điểm lấy TB trượt là 12, rồi trung tâm hoá

$$MA_{t-0,5} = \frac{Y_{t-6} + Y_{t-5} + Y_{t-4} + Y_{t-3} + Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2} + Y_{t+3} + Y_{t+4} + Y_{t+5}}{12}$$

$$MA_{t+0,5} = \frac{Y_{t-5} + Y_{t-4} + Y_{t-3} + Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2} + Y_{t+3} + Y_{t+4} + Y_{t+5} + Y_{t+6}}{12}$$

$$CMA_t = MA_{t-0,5} + MA_{t+0,5}$$

$$t \geq 7$$

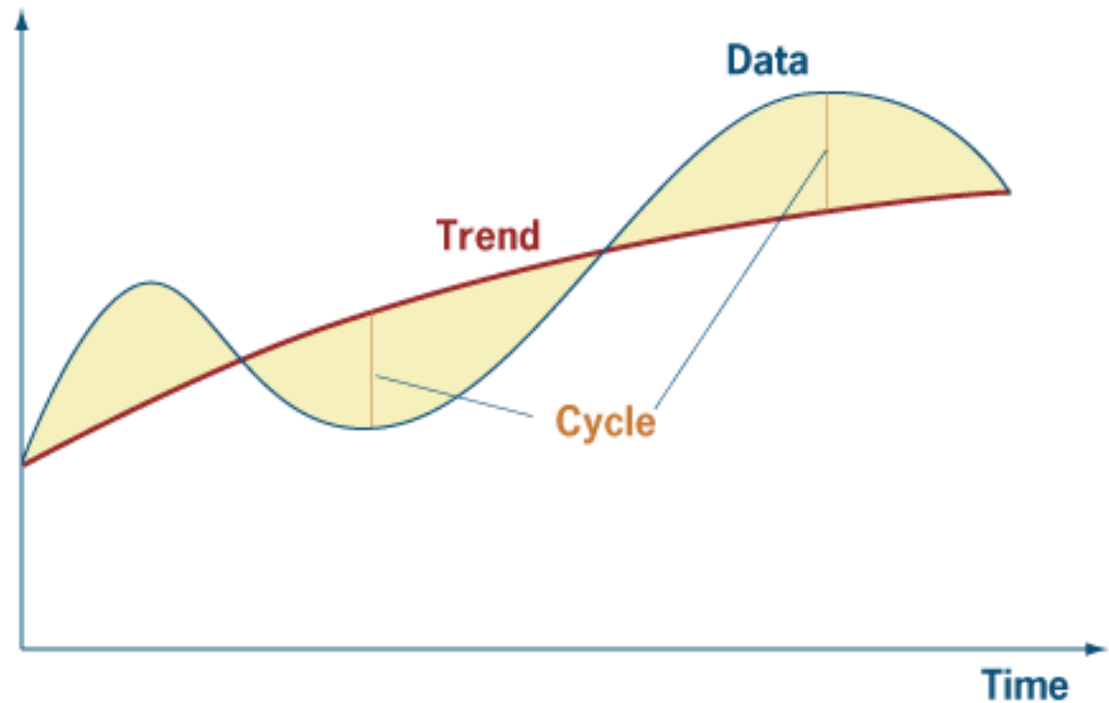
TD: Tính TB trượt trung tâm hoá CMA 4 điểm

TT	Y_t	MA_t	CMA_t
1	Y_1		-
2	Y_2	$MA_{2,5}$	-
3	Y_3	$MA_{3,5}$	CMA_3
4	Y_4	$MA_{4,5}$	CMA_4
5	Y_5	$MA_{5,5}$	CMA_5
6	Y_6	$MA_{6,5}$	CMA_6
7	Y_7	$MA_{7,5}$	CMA_7
8	Y_8	$MA_{8,5}$	CMA_8
9	Y_9	$MA_{9,5}$	CMA_9
10	Y_{10}	$MA_{10,5}$	CMA_{10}
11	Y_{11}		-
12	Y_{12}		-

Lọc thành phần mùa vụ S_t và bất thường E_t

- $S_t \cdot E_t = Y_t / CMA_t$

Figure 3. Removing the Trend Component from the Data

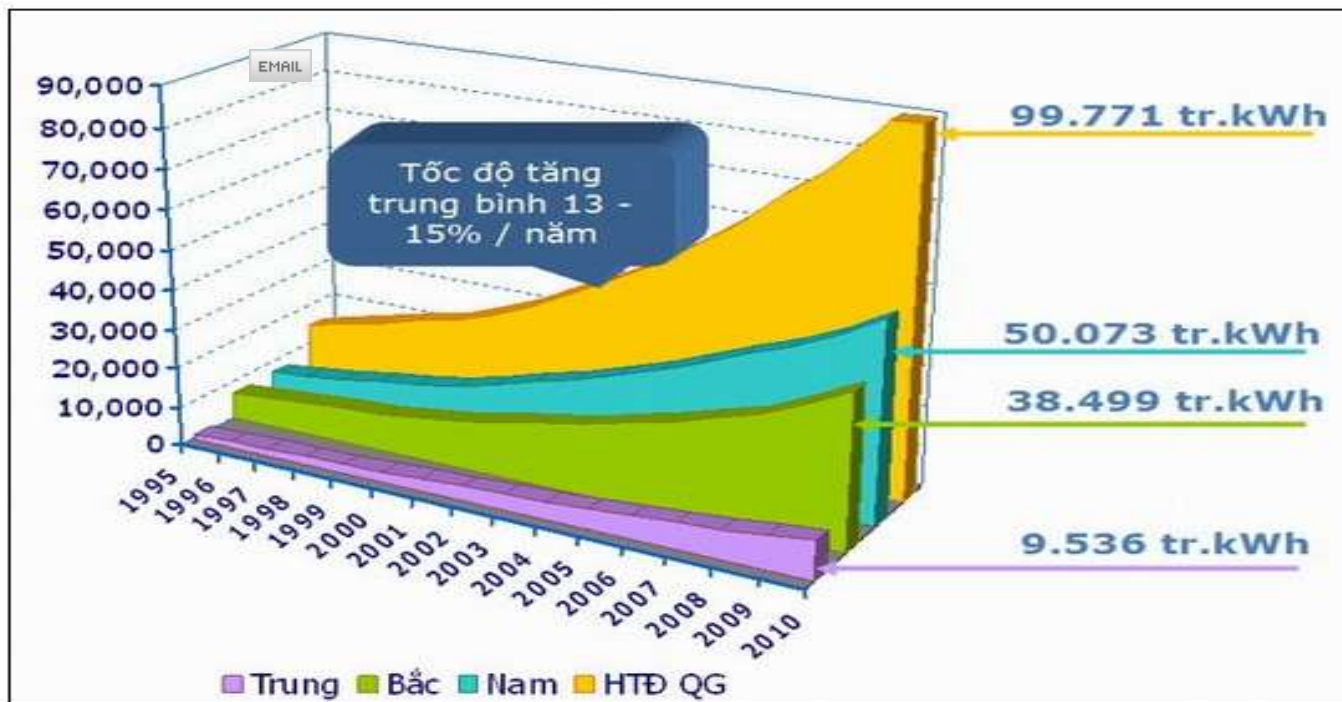


- Thí dụ: Bảng 14.8 Trang 441
- File Excel

14.4 DỰ BÁO BẰNG HÀM TĂNG TRƯỞNG MŨ

- Chuỗi thời gian có tốc độ tăng trưởng hầu như không đổi qua các giai đoạn
- TD:
 - Quy hoạch điện quốc gia
 - Chuỗi nhà hàng Western Steakhouses 1978-1992 – Trang 444

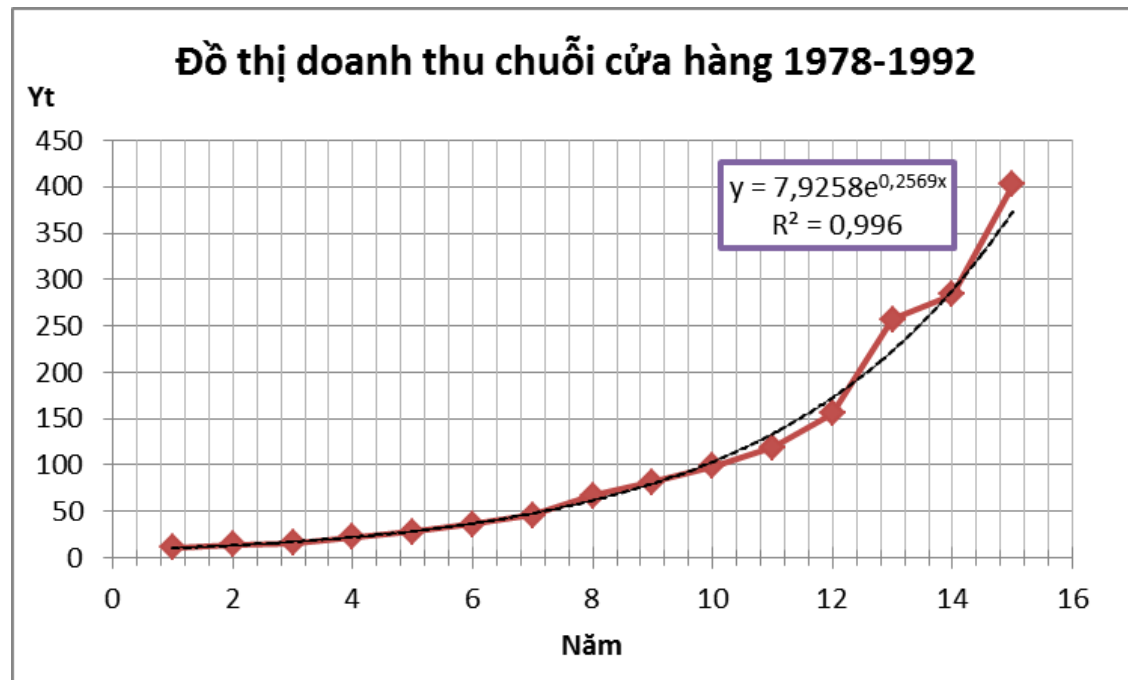
$$Y_t = a.e^{bt}$$



Phê duyệt tổng sơ đồ VII - Cơ hội đầu tư vào ngành điện

Trong bối cảnh nhu cầu phụ tải ngày càng tăng cao, theo quy hoạch điện VII, dự báo đến năm 2015 nhu cầu điện năng là $194 \div 210$ tỷ kWh; năm 2020 là $330 \div 362$ tỷ kWh và năm 2030 là $695 \div 834$ tỷ kWh. Tổng vốn đầu tư cho toàn ngành từ 2011-2030 lên đến 123,8 tỷ USD. Tuy nhiên, với giá bán điện hiện tại còn thấp, nền kinh tế vĩ mô còn nhiều bất cập, các yếu tố đầu vào (lạm phát, tỷ giá, chi phí đầu tư...) đang có xu hướng tăng cao, gây nhiều khó khăn cho việc huy động vốn trung và dài hạn. Do đó, việc giải bài toán giá bán điện đầu ra luôn là áp lực đối với EVN cũng như tất cả các đơn vị phát điện

Năm	t	Yt	Yt/Yt-1
1978	1	11	-
1979	2	14	1,273
1980	3	16	1,143
1981	4	22	1,375
1982	5	28	1,273
1983	6	36	1,286
1984	7	46	1,278
1985	8	67	1,457
1986	9	82	1,224
1987	10	99	1,207
1988	11	119	1,202
1989	12	156	1,311
1990	13	257	1,647
1991	14	284	1,105
1992	15	403	1,419
1993	16	?	
1994	17	?	



14.5 DỰ BÁO BẰNG SAN BẰNG HÀM SỐ MŨ

- 14.5.1 San bằng hàm mũ đơn giản
- 14.5.2 Phương pháp Holt
- 14.5.3 Phương pháp Holt-Winter

14.5.1 Phương pháp san bằng mũ đơn giản

- Exponential Smoothing Method

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$$

- F_{t+1} là giá trị dự báo ở giai đoạn $t+1$
- Y_t là giá trị thực tế ở giai đoạn t , $t = 1, 2, 3, \dots, n$
- α là hệ số làm trơn, $0 < \alpha < 1$.
- Giá trị dự báo: $F_{n+1} = \alpha Y_n + (1 - \alpha)F_n$.
- Phạm vi áp dụng
 - Dãy DL không có thành phần xu hướng và mùa vụ

- Chọn α
 - α càng gần 1, dãy DL dự báo càng phản ánh rõ những thay đổi gần nhất của dãy DL gốc (càng ít trơn)
 - α càng gần 0, dãy DL dự báo càng trơn hơn, và ít phản ánh những thay đổi gần nhất của dãy DL gốc. Như vậy, với dãy DL gốc có nhiều biến đổi bất thường, nên chọn α nhỏ.
 - Để tìm ra α tối ưu, cần thử α sao cho MSE là min. Có thể dùng hàm Solver của Excel.

14.5.2 Phương pháp Holt

- Áp dụng: dãy DL có tính xu hướng

$$L_1 = Y_1$$

$$b_1 = Y_2 - Y_1$$

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta).b_{t-1}$$

$$F_{t+m} = L_t + m.b_t$$

14.5.3 Phương pháp Holt-Winter

- Phạm vi áp dụng: Dãy DL có tính xu hướng và mùa vụ
- Các công thức tính

$$L_t = \alpha(Y_t / S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$$

$$S_t = \gamma(Y_t / L_t) + (1 - \gamma)S_{t-s}$$

$$F_{t+m} = (L_t + m.b_t).S_{t-s+m}$$

- s là số giai đoạn trong một vòng thời vụ (đối với DL quý thì s=4; với DL tháng thì s=12)
- L_t là đại diện cho mức độ của chuỗi thời gian
- b_t là thành phần đại diện cho xu hướng
- S_t là thành phần mùa vụ
- F_{t+m} là giá trị dự báo cho m thời đoạn về sau

Phương pháp Holt-Winter: TD Trang 465

Năm	Quý	Thời đoạn t	Doanh số (1.000USD)	PP đơn giản			PP Holt-Winter					alpha	beta	gamma
				F(t)	e(t)	e(t)^2	L(t)	b(t)	S(t)	F(t)	e(t)^2	0,8220156	0,0552789	0
1	1	1	362	362,00		-			0,953			0,4642482		
	2	2	385	362,00	23,00	529,00			1,013					
	3	3	432	372,68	59,32	3.519,13			1,137					
	4	4	341	400,22	- 59,22	3.506,77	380,000	9,750	0,897					
2	1	5	382	372,73	9,27	86,00	398,993	10,261	0,953	371,288	114,74			
	2	6	409	377,03	31,97	1.021,98	404,679	10,008	1,013	414,639	31,80			
	3	7	498	391,87	106,13	11.262,98	433,896	11,070	1,137	471,434	705,77			
	4	8	387	441,14	- 54,14	2.931,37	433,700	10,447	0,897	399,299	151,26			
3	1	9	473	416,01	56,99	3.248,23	487,198	12,827	0,953	423,109	2.489,12			
	2	10	513	442,47	70,53	4.975,08	505,214	13,114	1,013	506,604	40,91			
	3	11	582	475,21	106,79	11.403,86	513,081	12,824	1,137	589,257	52,66			
	4	12	474	524,79	- 50,79	2.579,39	527,801	12,929	0,897	471,930	4,29			
4	1	13	544	501,21	42,79	1.831,02	565,653	14,306	0,953	515,116	834,31			
	2	14	582	521,07	60,93	3.711,86	575,424	14,056	1,013	587,590	31,25			
	3	15	681	549,36	131,64	17.329,27	597,328	14,489	1,137	670,145	117,84			
	4	16	557	610,47	- 53,47	2.859,39	619,122	14,893	0,897	549,026	63,59			
5	1	17	628	585,65	42,35	1.793,66	654,739	16,039	0,953	603,983	576,81			
	2	18	707	605,31	101,69	10.340,85	693,005	17,268	1,013	679,604	750,53			
	3	19	773	652,52	120,48	14.515,57	685,350	15,890	1,137	807,468	1.188,06			
	4	20	592	708,45	- 116,45	13.561,14	667,099	14,003	0,897	629,271	1.389,09			

Bài tập về nhà

- 1, 2, 3, 6, 8
- Bài 8: Dùng Excel, không có lời giải