

SỬ DỤNG MÔ HÌNH KINH TẾ LƯỢNG TRONG THỐNG KÊ ĐỂ XỬ LÝ SỐ LIỆU CHO BÀI BÁO KHOA HỌC

TS.Trần Doãn Phú¹, ThS.Nguyễn Thị Hảo¹,
ThS. Nguyễn Hữu Việt¹, Nguyễn Thị Phương Uyên²

¹Trường Đại học Thành Đô

Tel: 0915891934; Email: tdphu@thanhdowni.edu.vn

²Trường THCS Thị trấn Trại Trôi

TÓM TẮT

Ứng dụng các công cụ thống kê để xử lý dữ liệu trong một bài báo khoa học đã được phát triển từ khá lâu. Hiện nay, các tạp chí uy tín trong và ngoài nước đều đánh giá cao các bài báo sử dụng các phần mềm thống kê tin cậy. Trong nghiên cứu khoa học không có mô hình kinh tế lượng nào tốt nhất hay dở nhất, quan trọng là khi sử dụng mô hình kinh tế lượng giúp chúng ta dễ dàng và minh bạch hơn trong quá trình xử lý số liệu. Bài viết này chia sẻ hai mô hình kinh tế lượng thông dụng giúp các nhà nghiên cứu thuận tiện hơn khi viết bài báo khoa học.

Từ khóa: Kinh tế lượng, nghiên cứu khoa học, thống kê dữ liệu, bài báo khoa học, xử lý số liệu.

ABSTRACT

The application of statistical tools for data processing in scientific papers has been established for a long time. Currently, both domestic and international prestigious journals highly value articles that use reliable statistical software. In scientific research, the use of econometric models simplifies and clarifies the data processing procedure. This article introduces two common econometric models to make statistics more convenient for researchers when preparing scientific articles.

Keywords: *Econometrics, scientific research, data statistics, scientific articles, data processing.*

1. GIỚI THIỆU

Trong những năm gần đây, hoạt động nghiên cứu khoa học của các trường đại học ở Việt Nam đã có những chuyển biến nhất định, số lượng bài báo đăng trên các tạp chí uy tín trong nước và quốc tế tăng cả về số lượng và chất lượng. Tuy nhiên, số lượng công bố khoa học vẫn chưa thực sự tương xứng với số lượng giảng viên của các trường. Một trong những nguyên nhân của tình trạng này có thể do đội ngũ giảng viên

chưa có nhiều kinh nghiệm trong việc viết và đăng bài báo khoa học. Viết bài báo khoa học là một quá trình nghiên cứu mất rất nhiều công sức và trí tuệ. Để viết được một bài báo khoa học và được chấp nhận đăng trên các tạp chí trong và ngoài nước, các nhà nghiên cứu cần có những kỹ năng viết, phân tích số liệu, suy luận và tư duy nhất định. Đặc biệt, đối với những nhà khoa học trẻ, viết bài báo khoa học là một thách thức không hề dễ dàng, luôn gặp rất nhiều khó khăn trong việc thống kê và xử lý số liệu. Một

trong những lý do một số bài báo khoa học của các tác giả Việt Nam khó được đăng trên các tạp chí uy tín nước ngoài là vì số liệu không được xử lý bằng những phần mềm thống kê tin cậy. Ngay ở trong nước, những bài báo có sử dụng những phần mềm thống kê đều có sức thuyết phục hơn.

Có rất nhiều định nghĩa khác nhau về kinh tế lượng, tuy nhiên có thể đưa ra một định nghĩa đơn giản về kinh tế lượng “*Kinh tế lượng là việc sử dụng các phương pháp thống kê toán học trong kinh tế*” [1]. “*Một cách đầy đủ và chi tiết hơn thì Kinh tế lượng là khoa học nghiên cứu những vấn đề thực nghiệm của các quy luật kinh tế, là sự kết hợp chặt chẽ giữa các số liệu thực tế, lý thuyết kinh tế và công cụ toán học không thể thiếu được là lý thuyết xác suất thống kê kết hợp với các phần mềm vi tính hỗ trợ, nhằm lượng hóa các quy luật kinh tế nói riêng và thực tiễn nói chung thông qua những mô hình toán học phù hợp với thực tế, ước lượng các tham số, phân tích, đánh giá và dự báo các chỉ tiêu kinh tế, xã hội*” [2]. Kinh tế lượng vì thế còn được áp dụng trong các lĩnh vực khoa học kỹ thuật, môi trường, dân số, giáo dục, v.v.... Bài viết này, tác giả chia sẻ một trong những mô hình kinh tế lượng thông dụng giúp các nhà nghiên cứu trở thuận tiện hơn khi viết bài báo khoa học là phương pháp san mũ Holt - Winters [3].

2. DỰ BÁO SAN MŨ ĐƠN GIẢN (Holt-Winters)

Phương pháp san mũ đơn giản không chỉ giúp chúng ta loại bỏ yếu tố ngẫu nhiên mà còn giúp chúng ta dự báo giá trị trong tương lai. Phương pháp này là cơ sở khoa học để các nhà nghiên cứu sau khi đã có số liệu nghiên cứu thực tế

tiếp tục phân tích, dự đoán tình hình trong thời gian tiếp theo để đưa ra các đề xuất, giải pháp phù hợp.

2.1. Dự báo chuỗi thời gian có yếu tố xu thế (mô hình Holt-Winters No Seasonal)

Xét một chuỗi thời gian Y_t (GDP, lãi suất tiền gửi tiết kiệm, giá cổ phiếu, ...) có yếu tố xu thế. Mục đích không chỉ là ước lượng giá trị hiện thời của chuỗi mà còn ước lượng phần xu thế. Phần xu thế ở thời kỳ t được hiểu là chênh lệch giữa hai giá trị Y_t và Y_{t-1} .

Ký hiệu T_t là ước lượng phần xu thế ở thời kỳ t . Khi đó ước lượng của Y thời kỳ $t - 1$ gồm hai phần: ước lượng phần hệ thống \hat{Y}_{t-1} và ước lượng phần xu thế T_{t-1} . Do đó ước lượng giá trị của Y tại $t - 1$ là: $\hat{Y}_{t-1} + T_{t-1}$. Theo công thức đệ quy san mũ đơn giản ta có:

$$\hat{Y}_t = \alpha Y_t + (1 - \alpha)(\hat{Y}_{t-1} + T_{t-1}) \quad (1.1)$$

α là hằng số san mũ $0 \leq \alpha \leq 1$.

Vì T cũng là chuỗi thời gian nên theo phương pháp san mũ đơn giản ta cũng có:

$$T_t = \beta(\hat{Y}_t - \hat{Y}_{t-1}) + (1 - \beta)T_{t-1} \quad (1.2)$$

β là hằng số san mũ, $0 \leq \beta \leq 1$.

Để tính toán bắt đầu ta cho $T_2 = Y_2 - Y_1$ và $\hat{Y}_2 = Y_2$. Với hai công thức đệ quy (1.1), (1.2) và hai hằng số san mũ α và β ta có thể tính được \hat{Y}_n và T_n .

Sau khi tính được \hat{Y}_n và T_n ta có công thức dự báo cho thời kỳ $n + h$:

$$\hat{Y}_{n+h} = \hat{Y}_n + hT_n$$

Ví dụ 1: Từ số liệu về số học sinh ở Mỹ từ năm 1985 đến 1991, mô hình Holt-Winter được tính với $\alpha = 0,7$; $\beta = 0,6$, tính toán như sau:

Bảng 1. Số học sinh ở Mỹ từ năm 1985 đến 1991

STT	Y	\hat{y}	T	STT	Y	\hat{y}	T
1(1985)	50,74			8(1992)	46,64	46,77	-0,97
2(1986)	50,43	50,43	-0,31	9(1993)	45,95	45,90	-0,91
3(1987)	50,05	50,07	-0,34	10(1994)	45,20	45,14	-0,82
4(1988)	49,79	49,77	-0,31	11(1995)	44,74	44,61	-0,64
5(1989)	49,48	49,47	-0,31	12(1996)	44,23	44,15	-0,53
6(1990)	48,72	48,85	-0,49	13(1997)	43,92	43,83	-0,41
7(1991)	47,64	47,86	-0,80				

Sử dụng công thức dự báo cho thời kỳ $n + h$:

$$\hat{Y}_{n+h} = \hat{Y}_n + hT_n = 43,83 + h*(-0,41)$$

Ta có thể dự báo, chẳng hạn số lượng học sinh năm 2000 (ta có $h = 3$), ở Mỹ là:

$$\hat{Y}_{n+h} = \hat{Y}_n + hT_n = 43,83 + 3*(-0,41) = 42,6$$

Ví dụ 2: Dự báo kết quả chạy bằng phần mềm EviewsTa có bảng thống kê lượt khách quốc tế, khách nội địa và tổng thu từ du lịch các năm 2000 đến năm 2018 như sau:

Bảng 2. Thống kê lượt khách quốc tế, khách nội địa và tổng thu từ du lịch của Việt Nam các năm 2000 đến năm 2018

Năm	Khách QT (triệu lượt khách)	Tốc độ tăng trưởng (%)	Khách nội địa (triệu lượt khách)	Tốc độ tăng trưởng (%)	Tổng thu từ DL (nghìn tỷ đồng)	Tốc độ tăng trưởng (%)
2000	2,140		11,200		17,40	
2001	2,331	8,9	11,700	4,5	20,50	17,8
2002	2,628	12,8	13,000	11,1	23,00	12,2
2003	2,430	-7,6	13,500	3,8	22,00	-4,3
2004	2,928	20,5	14,500	7,4	26,00	18,2
2005	3,468	18,4	16,100	11,0	30,00	15,4
2006	3,583	3,0	17,500	8,7	51,00	70,0
2007	4,172	16,0	19,200	9,7	56,00	9,8
2008	4,254	0,6	20,500	6,8	60,00	7,1
2009	3,772	-10,9	25,000	22,0	68,00	13,3
2010	5,050	34,8	28,000	12,0	96,00	41,2
2011	6,014	19,1	30,000	7,1	130,00	35,4

Năm	Khách QT (triệu lượt khách)	Tốc độ tăng trưởng (%)	Khách nội địa (triệu lượt khách)	Tốc độ tăng trưởng (%)	Tổng thu từ DL (nghìn tỷ đồng)	Tốc độ tăng trưởng (%)
2012	6,848	10,8	32,500	8,3	160,00	23,1
2013	7,572	10,6	35,000	7,7	200,00	25,0
2014	7,874	4,0	38,500	10,0	230,00	15,0
2015	7,944	0,9	57,000	48,0	337,83	46,9
2016	10,013	26,0	62,000	8,8	400,00	18,4
2017	12,922	29,1	73,200	18,1	510,90	27,5
2018	15,498	19,9	80,000	9,3	620,00	21,4

(Nguồn Tổng cục Thống kê và Tổng cục Du lịch Việt Nam)

Ta sẽ dùng mô hình Holt-Winters để dự báo số lượng khách quốc tế, số lượng khách du lịch nội địa và tổng doanh thu từ du lịch trong những năm tới.

Dự báo số lượng khách quốc tế:

Date: 08/03/19 Time: 15:57		
Sample: 2000 2018		
Included observations: 19		
Method: Holt-Winters No Seasonal		
Original Series: SLKQT		
Forecast Series: SLKQTSM		
Parameters:	Alpha	1,0000
	Beta	0,7900
Sum of Squared Residuals		9,792341
Root Mean Squared Error		0,717904
End of Period Levels:		Mean 15,49800
		Trend 2,591015

Kết quả trên cho chúng ta số lượng khách du lịch quốc tế dự báo của năm 2019 là $15,49800 + 2,591015 = 18,08902$ triệu lượt khách, còn của năm 2020 là $15,49800 + 2 * 2,591015 = 20,68003$ triệu lượt khách.

Dự báo số khách du lịch nội địa:

Date: 08/03/19 Time: 16:13		
Sample: 2000 2018		
Included observations: 19		
Method: Holt-Winters No Seasonal		
Original Series: SLKND		
Forecast Series: SLKNDSM		
Parameters:	Alpha	0,8000
	Beta	0,5000
Sum of Squared Residuals		275,8082
Root Mean Squared Error		3,810016
End of Period Levels:		Mean 80,48614
		Trend 8,698290

Dự báo cho năm 2019 là: $80,48614 + 8,698290 = 89,18443$, còn năm 2020 là: $80,48614 + 2 * 8,698290 = 97,88272$ triệu lượt khách.

Dự báo tổng doanh thu từ du lịch:

Date: 08/03/19 Time: 16:26		
Sample: 2000 2018		
Included observations: 19		
Method: Holt-Winters No Seasonal		
Original Series: TONGTHU		
Forecast Series: TONGTHSM		
Parameters:	Alpha	0,7100
	Beta	1,0000
Sum of Squared Residuals		7616,386
Root Mean Squared Error		20,02155
End of Period Levels:		Mean 615,8667
		Trend 112,7528

Dự báo tổng doanh thu từ du lịch năm 2019 là: $615,8667 + 112,7528 = 728,6195$ nghìn tỷ đồng, còn năm 2020 là: $615,8667 + 2 * 112,7528 = 841,3723$ nghìn tỷ đồng.

2.2. Dự báo chuỗi thời gian có yếu tố xu thế và yếu tố thời vụ (mô hình Holt-Winters Multiplicative Seasonal)

Nhiều chuỗi thời gian trong kinh tế chứa đựng yếu tố thời vụ khá mạnh. Các chuỗi thời gian theo quý, tháng, tuần, ... thường mang tính thời vụ cao.

Ta ký hiệu Y_t , \hat{Y}_t , T_t tương ứng là giá trị quan sát, giá trị ước lượng và xu thế của chuỗi cần nghiên cứu ở thời kỳ t , F_t là yếu tố thời vụ. Nếu chuỗi có s thời kỳ trong một năm thì yếu tố thời vụ của năm trước sẽ là F_{t-s} . Yếu tố thời vụ tỷ lệ với trung bình của chuỗi ở mỗi thời kỳ. Do đó mô hình Holt-winters được cải tiến như sau:

$$\hat{Y}_t = \alpha \frac{Y_t}{F_{t-1}} + (1-\alpha)(\hat{Y}_{t-1} + T_{t-1})$$

$$T_t = \beta(\hat{Y}_t - \hat{Y}_{t-1}) + (1-\beta)T_{t-1}$$

$$F_t = \gamma \frac{Y_t}{\hat{Y}_t} + (1-\gamma)F_{t-1}$$

Dự báo ở thời kỳ $n+h$:

$$\hat{Y}_{n+h} = \begin{cases} (\hat{Y}_n + hT_n)F_{n+h-s} & , h=1,2,\dots,s \\ (\hat{Y}_n + hT_n)F_{n+h-2s} & , h=s+1,s+2,\dots,2s \end{cases}$$

Ví dụ: Bảng sau đây cho tiền lãi trên một đơn vị vốn trong một quý. Dưới đây chúng ta sẽ phân tích và dự báo các yếu tố hình thành chuỗi này với $\alpha = \beta = 0,5$; $\gamma = 0,3$.

Chuỗi số liệu của chúng ta là theo quý, nên $s = 4$. Để tính được các chuỗi theo công thức truy toán, ta cần các giá trị xuất phát của các chuỗi này. Trước hết ta tính các trung bình trượt trung tâm cho từng quý:

Kết quả tính toán được cho trong bảng tính toán sau đây:

$$Y_t^* = (Y_{t-2} + 2(Y_{t-1} + Y_t + Y_{t+1}) + Y_{t+2})/8$$

$$Y_3^* = (0,3 + 2(0,46+0,345+0,91) + 0,33)/8 = 0,5075$$

Vì chúng ta cần có các $F_{t-s} = F_{t-4}$, nên cần 4 giá trị xuất phát với 4 quý khác nhau:

$$F_3 = Y_3 / Y_3^* = 0,345 / 0,507$$

$$F_4 = Y_4 / Y_4^* = 0,91 / 0,522$$

$$F_5 = Y_5 / Y_5^* = 0,33 / 0,544$$

$$F_6 = Y_6 / Y_6^* = 0,545 / 0,952.$$

Từ đây tính được $F_7 = F_{4+3} = 0,3 * Y_7 / Y_7^* + 0,7 * F_{7-4} = 0,6925$.

Tương tự ta tính được $F_8 = 1,7029$; $F_9 = 0,6437$; $F_{10} = 0,9495$.

Giá trị xuất phát của của chuỗi \hat{Y}_t là :

$$\hat{Y}_{10} = Y_{10}^* = (1,04 + 2(0,495 + 0,68 + 0,545) + 1,285)/8 = 0,721 ;$$

$$T_{10} = Y_{10}^* - Y_9^* = 0,721 - 0,677 = 0,043$$

$$(Y_9^* = (0,44 + 2(1,04 + 0,495 + 0,68) + 0,545)/8 = 0,677).$$

Từ quan sát 11 trở đi ta mới tính được các giá trị của các chuỗi tương ứng theo công thức đệ quy.

$$\hat{Y}_{11} = 0,5Y_{11}/F_{11-4} + 0,5(\hat{Y}_{11-1} + T_{11-1})$$

$$= 0,5(0,545/0,6925) + 0,5(0,7206 + 0,043) = 0,7757;$$

$$T_{11} = 0,5(\hat{Y}_{11} - \hat{Y}_{10}) + 0,5T_{11-1}$$

$$= 0,5(0,7757 - 0,7206) + 0,5 * 0,0438 = 0,049;$$

$$F_{11} = 0,3Y_{11} / \hat{Y}_{11} + 0,7F_{11-4}$$

$$= 0,3(0,545/0,7755) + 0,7 * 0,6925 = 0,6955.$$

Thời gian		Y	\hat{Y}	T	F
2011	QI	0,3			
	QII	0,46			
	QIII	0,345	0,5075		0,6798
	QIV	0,91	0,5219		1,7437
2012	QI	0,33	0,5444		0,,6062
	QII	0,545	0,5725		0,9520
	QIII	0,44	0,6094		0,6925
	QIV	1,04	0,6469		1,7029
2013	QI	0,495	0,6769		0,6437
	QII	0,68	0,7206	0,0438	0,9495
	QIII	0,545	0,7757	0,049	0,6955
	QIV	1,285	0,7899	0,032	1,6801
2014	QI	0,55	0,8380	0,040	0,6475
	QII	0,87	0,8971	0,050	0,9555
	QIII	0,66	0,9478	0,050	0,6958
	QIV	1,58	0,9692	0,036	1,6652
2015	QI	0,59	0,9580	0,012	0,6380
	QII	0,99	1,0032	0,029	0,9649
	QIII	0,83	1,1124	0,069	0,7109
	QIV	1,73	1,1102	0,033	1,6331
2016	QI	0,61	1,0498	-0,013	0,6209
	QII	1,05	1,0622	-0,001	0,9720
	QIII	0,92	1,1780	0,058	0,7319
	QIV	2,01	1,2423	0,061	1,6358
2017	QI	0,7	1,2154	0,017	0,6074
	QII	1,23	1,2489	0,025	0,9759
	QIII	1,06	1,3612	0,069	0,7460
	QIV	2,32	1,4241	0,066	1,6338
2018	QI	0,82	1,4200	0,031	0,5984
	QII	1,41	1,4479	0,029	0,9753
	QIII	1,25	1,5765	0,079	0,7600
	QIV	2,73	1,6632	0,083	1,6361

Như vậy, ta có $\hat{Y}_{32} = 1,6632$; $T_{32} = 0,083$ và hệ số thời vụ:

Quý I/2018: 0,5984;

Quý II/2018: 0,9753;

Quý III/2018: 0,7600;

Quý IV/2018: 1,6381.

Chúng ta thử dự báo tiền lãi cho một đơn vị vốn quý III năm 2019:

$$\hat{Y}_{32+3} = (\hat{Y}_n + hT_n)F_{n+h-s} = (1,6632 + 3*0,083)*0,76 = 1,4533.$$

Trong thực tế ta không phải tính toán thủ công như trên mà phần mềm Eviews giúp chúng ta ước lượng \hat{Y}_n ; T_n và các hệ số thời vụ kể cả tự động hóa chọn α ; β và γ để sai số là nhỏ nhất.

3. KẾT LUẬN

Viết học thuật nói chung, viết bài báo khoa học nói riêng không phải chỉ dựa trên ý tưởng và kinh nghiệm của một tác giả. Nội dung của bài báo khoa học

còn sử dụng các số liệu và kết quả nghiên cứu của các nguồn khác như sách, báo, trang web, v.v. của các tác giả khác trong cùng lĩnh vực hoặc các lĩnh vực có liên quan. Các số liệu này có thể được sử dụng để hỗ trợ các ý tưởng của tác giả hoặc tác giả có thể phân tích hoặc bàn luận. Vì vậy, khi viết bài báo khoa học và muốn bài báo đó đạt độ tin cậy nhất định thì việc sử dụng các công cụ và mô hình kinh tế lượng để thống kê và xử lý số liệu gần như là yêu cầu bắt buộc, đòi hỏi tác giả phải thực hiện cho đúng quy cách và phù hợp với yêu cầu trong nghiên cứu khoa học.

Kiến thức kinh tế lượng rộng và khó, tác giả đã chọn ra mô hình thông dụng trên để giới thiệu với các nhà nghiên cứu trẻ. Hy vọng có thể giúp các nhà nghiên cứu trẻ thuận tiện hơn trong việc thống kê, xử lý số liệu khi viết bài báo khoa học.

TÀI LIỆU TRÍCH DẪN

- [1] http://eldata11.topica.edu.vn/HocLieu/v1.0/STA301/Giao%20trinh/03_TVU_S TA301_Bai1_v1.00131012140.pdf, truy cập ngày 14/10/2022.
- [2] Trần Kim Thanh, Nguyễn Văn Phong, Nguyễn Trung Đông (2015), Bài giảng kinh tế lượng, khoa Cơ bản, bộ môn Toán - Thống kê, Trường ĐH Tài chính Marketing.
- [3] Nguyễn Quang Đông (2009), Kinh tế lượng chương trình nâng cao, khoa Toán kinh tế, ĐH Kinh tế quốc dân, NXB Khoa học và kỹ thuật.
- [4] <https://mosl.vn/mo-hinh-hoi-quy-tuyen-tinh/>, truy cập ngày 14/10/2022.
- [5] https://bvag.com.vn/wp-content/uploads/2013/01/k2_attachments_PHAN-TICH-HOI-QUY-TUYEN-TINH-DON-GIAN.pdf, truy cập ngày 14/10/2022.