

통계패키지에서의 시계열 분석방법의 비교연구

김 수 화, 김 승 희, 조 신 섭¹⁾

요 약

각종 통계패키지 내에 수용되어 있는 시계열 분석방법은 패키지의 특성이나 기능에 따라 다소 차이가 있다. 본 논문에서는 일반적으로 많이 사용되고 있는 8종류의 통계패키지 (EXECUSTAT, MINITAB, RATS, SAS, SCA, S-PLUS, SPSS, TSP)에서 시계열 분석이 어떻게 이루어지는지를 비교 검토하였다. 지수평활법과 ARIMA 모형에 의한 분석방법을 중심으로 비교하였으며, 아울러 사용자 관점에서 편리하고 보다 효율적인 패키지가 갖추어야 할 기능들을 제시하였다.

1장. 서론

현재 시중에는 많은 종류의 통계패키지가 나와 있으며 이들을 이용하여 비교적 쉽게 시계열 자료를 분석할 수 있다. 시계열 분석의 목적은 시간에 따라 관측된 자료를 분석함으로써 주어진 자료의 성격과 구조를 이해하고 모형화하며 그것을 바탕으로 미래의 값을 예측하는데 있다. 시계열 분석은 자료의 형태와 성질에 따라 적합한 분석방법이 다르며 또한 분석자의 능력에 따라 분석결과가 달라질 수 있다(Makridakis et al., 1982). 또한 패키지를 사용하여 자료를 분석할 때는 분석과정이 패키지에서 제공되는 기능들과 절차에 영향을 받게된다.

본 논문에서는 시계열 분석시 영향을 미칠 수 있는 점들을 고려하여 각종 패키지가 가지고 있는 기능들과 분석절차를 살펴보고 비교하여 시계열 자료를 분석하고자 할 경우 패키지의 선택,사용에 도움을 주는데 그 목적이 있다. 통계패키지의 사용과 관련된 국내 논문으로는 패키지의 선택에 관한 연구(김병천, 1987)와 EDA 기능을 중심으로한 패키지 비교연구(허명희·정진환, 1990)가 있었으며 본 논문에서는 시계열 분석에 초점을 맞추어 비교 연구하였다.

여러가지 시계열 예측 방법가운데서 가장 많이 이용되는 것은 지수평활법과 ARIMA 모형에 의한 방법이다. 이 두가지 방법을 중심으로 각 패키지에서 시계열 분석이 어떻게 수용되어 있는지를 비교 검토하고 사용자 관점에서 효율적인 패키지의 기능을 제시하였다.

본 논문에서 고려된 패키지는 PC Dos version 의 EXECUSTAT(ver3.0), MINITAB(release7.0 standard version), RATS(ver4.0), SAS(ver6.0), SCA(ver3.0), S-PLUS(ver2.0), SPSS(ver4.0.1), TSP(ver7.0)이다. 최근에는 위의 패키지 중 일부가 Windows version으로 바뀌어 그래픽 기능이 강화되고 기능이 향상되었으나, 이 논문에서는 현재 널리 사용되고있는 Dos version 만을 비교하였다. 고려의 대상인 8종류의 통계패키지에 대해 간략히 살펴보면 다음과 같다. (김병천, 1987 ; 허명희·정진환, 1990 ; Yukiewicz, 1993)

EXECUSTAT은 메뉴 방식으로 처리되며 그래픽 기능이 우수하다. MINITAB은 간단한 명령어들로 작업이 처리되며 사용하기가 쉬워 교육용으로 많이 이용되고 있다. RATS(Regression Analysis of Time Series)는 시계열 분석과 회귀분석이 중심으로 계량경제학과 경영학 분야에서 많이 사용되고 있다. SAS(Statistical Analysis System)는 그래픽과 자료편집 기능이 뛰어나며 현재 가장 널리 사용되는 프로

1) (151-742) 서울특별시 관악구 신림동 산 56-1, 서울대학교 계산통계학과.

그램이다. SCA(Scientific Computing Associates)는 다양한 통계처리를 제공하며 주 내용은 시계열 분석과 품질관리이다. S-PLUS 는 그래픽에 의한 자료탐색 분석능력이 우수하다. SPSS(Statistical Package for the Social Sciences)는 대화형식으로 되어 있어 매우 편리하다. TSP(Time Series Processor)는 계량경제학 분야의 분석에 많이 쓰이며 사용하기에 간편하다. 각 통계패키지에서 가능한 입출력 기능과 운영환경, 운영방식등 일반적 특성에 대해 정리하면 표 1과 같다(Yurkiewicz,1993).

먼저 2장에서는 지수평활법에 의한 예측방법을 비교해보고, 3장에서는 ARIMA 모형을 이용한 예측방법을 비교해 보았다. 마지막으로 4장에서는 시계열 분석에 있어서 패키지가 가져야할 성질에 대해 언급하였다.

표 1. 통계패키지의 일반적특성 비교

패키지 종류		EXECUSTAT	MINITAB	RATS	SAS	SCA	S-PLUS	SPSS	TSP
1 일반적 특성									
운영 환경	IBM	●	●	●	●	●	●	●	●
	Macintosh		●	●		●		●	●
	Other		UNIX, VAX/VMS	UNIX, Main	VAX, UNIX, Main	Main, Work	UNIX, Windows	over 50 platforms	
운영 방식	Automatic	●			●				
	Semi-Automatic	●		●			●	●	●
	Manual		●	●		●	●	●	
자료 입력	Reads following spreadsheet files	Lotus, Sylk,WK1, DBase,DIF	Lotus	Lotus	Lotus	Lotus	Lotus, EXcel, numerous	Lotus, EXcel	Lotus
	Reads ASCII files	●	●	●	●	●	●	●	●
	Saves ASCII file of forecasting session	●	●	●	●	●	●	●	●
출력	Exports to a spreadsheet file	●	●	●			●	●	●
	Can export graphics	●	●	●	●	●	●		●

● : 있음 Main : Mainframes Work : Workstation

☞ Automatic : software suggests technique & parameters.

Semi-Automatic : user choose technique.

Manual : user choose technique & parameters.

2장. 지수평활법에 의한 예측

2.1 지수평활법

평활법이란 시계열의 미래값을 예측하기 위해 과거의 관측값들에 이동평균, 가중이동평균 등의 방법을 적용하여 시계열의 불규칙 변동을 제거하는 방법이다. 평활법은 이론적인 배경은 부족하나 직관적이고 계산이 간편하기 때문에 가장 오랫동안 사용되어 온 시계열 예측방법이다. 지수평활법은 가중이동평균방법의 한 형태로서 과거의 값들을 지수적으로 가중평균해서 예측값을 구한다.

지수평활법에는 계절성의 존재여부와 추세의 형태에 따라 다음과 같은 방법들이 있다.

1) 계절성이 없는 경우

a. 단순지수평활법(simple exponential smoothing)

시계열이 추세를 갖지 않고 일정한 수준만 유지하는 경우에 상수모형 $Z_t = b_0 + e_t$ 을 가정해서 적합시키는 예측방법이다(e_t 는 오차항을 나타냄).

b. 이중지수평활법(double exponential smoothing)

시계열이 직선형태의 추세를 갖는 경우에 직선모형 $Z_t = b_0 + b_1t + e_t$ 을 가정하여 적합시키는 예측방법으로 Brown의 일모수 방법과 Holt의 이모수 방법이 있다.

c. 삼중지수평활법(triple exponential smoothing)

시계열이 이차곡선 형태의 추세를 갖는 경우에 이차모형 $Z_t = b_0 + b_1t + b_2t^2 + e_t$ 을 가정하여 적합시키는 예측방법이다.

2) 계절성이 있는 경우

a. 승법계절지수평활법(multiplicative seasonal exponential smoothing)

Winters에 의해 제안된 방법으로 관측된 시계열이 선형추세와 시간에 따라 변동폭이 커지는 승법계절변동을 갖는 경우에 승법모형 $Z_t = (b_0 + b_1t + b_2t^2)S(t) + e_t$ 을 가정하여 적용하는 예측방법이다($S(t)$ 는 계절변동을 나타냄).

b. 가법계절지수평활법(additive seasonal exponential smoothing)

Winters에 의해 제안된 방법으로 시계열이 선형추세와 시간에 따라 변동폭이 일정한 가법계절변동을 갖는 경우에 가법모형 $Z_t = (b_0 + b_1t + b_2t^2) + S(t) + e_t$ 을 가정하여 적합시키는 예측방법이다.

2.2 통계패키지에서의 지수평활법의 비교

주어진 시계열에 지수평활법을 적용할 때는 시계열도표를 통해 시계열의 추세성과 계절성을 조사한 후 2.1절에서 언급한 여러가지 지수평활법 중에서 어떤 방법을 사용할 지를 결정한다. 지수평활법이 선택된 후 적당한 평활계수(smoothing constant), 예측하고자 하는 예측값의 개수등을 제시함으로써 예측을 하게된다. MINITAB과 S-PLUS는 지수평활법을 수용하지 않으므로 나머지 6종류의 패키지에 대해서만 비교해 보았다.

1) 사용 가능한 지수평활법의 종류

표 2.1과 표 2.2에서 알 수 있듯이, 단순지수평활법과 승법계절지수평활법은 모든 패키지에서 수용되어 있고 Holt의 지수평활법과 가법계절지수평활법은 각각 SAS와 EXECUSTAT를 제외한 나머지 패키지에 수용되어 있다. 이중지수평활법은 EXECUSTAT, SAS, SCA, TSP에서만 사용할 수 있다. 또한 삼중지수평활법은 SAS와 EXECUSTAT에서만 가능하고 지수추세(exponential trend)에 대한 지수평활법은 SPSS와 RATS에서만 가능하다. 특히 SCA에서는 계절지수(seasonal indicator)와 조화함수(harmonic functions)를 이용한 일반적인 지수평활법이 가능하고 SPSS에서는 감폭하는(damped) 추세에 대한 지수평활법이 가능하다. 대부분의 패키지에서는 사용할 지수평활법을 제시해야만 예측을 하게 되는데 RATS의 경우에는 TREND와 SEASONAL 명령문에 select라는 option을 줌으로써 예측오차제곱합을 최소로 하는 모형에 의해 예측을 한다. 따라서 관측된 시계열의 추세성이나 계절성 파악이 어려울때는 RATS를 사용하면 편리할 것이다.

표 2.1. 각 패키지에서 수용하고 있는 지수평활법의 종류 비교
(계절성이 없는 경우)

패키지 종류		EXECUSTAT	RATS	SAS	SCA	SPSS	TSP
추세의 형태	모수						
없음 (단순지수평활법)	1	●	●	●	●	●	●
선형 (이중지수평활법)	1	●		●	●		●
선형 (Holt 지수평활법)	2	●	●		●	●	●
곡선	1	●		●			
지수	2		●			●	
감폭	3					●	

● : 있음

※ MINITAB, S-PLUS에서는 지수평활법을 수용하지 않음.

2) 평활계수

평활계수는 사용자가 지정해주거나 또는 프로그램 내에서 자동으로 선택할 수도 있다. 평활계수를 지정하지 않는 경우, 각 패키지의 처리방법을 살펴보면 다음과 같다. RATS, SAS, SPSS에서는 각각 default 값인 $0.3, 1 - (0.8)^{1/trend}$ (trend의 값은 단순지수평활법의 경우 1, 이중지수평활법 2, 삼중지수평활법 3), 0.1을 이용하여 예측한다. EXECUSTAT에서는 RMSE(Root Mean Squared Error)를 최소화하는 값을,

표 2.2. 각 패키지에서 수용하고 있는 지수평활법의 종류비교
(계절성이 있는 경우)

패키지		EXECUSTAT	RATS	SAS	SCA	SPSS	TSP
지수평활법	모수						
계절성, 추세							
가법, 없음	2		●	●		●	
가법, 선형	3		●	●	●	●	●
가법, 곡선	3			●			
가법, 지수	3		●			●	
가법, 감폭	3					●	
승법, 없음	2		●	●		●	
승법, 선형	3	●	●	●	●	●	●
승법, 곡선	3			●			
승법, 지수	3		●			●	
승법, 감폭	3					●	

● : 있음
 ※ MINITAB, S-PLUS 에서는 지수평활법을 수용하지 않음.

TSP와 RATS에서는 예측오차제곱합을 최소화하는 값을 구하여 사용한다. 그러나 SCA에서는 default 값이 없으므로 반드시 평활계수를 지정해 주어야 한다.

평활계수를 사용자가 지정하는 경우 일반적으로 사전 정보와 사용자의 경험에 의해 주관적으로 결정한다. 그러나 사전정보가 없을 경우에는 대부분 한 시점 후의 예측오차 (one-step-ahead forecast errors) 제곱합을 최소로 하는 평활계수를 선택한다. EXECUSTAT, RATS, TSP에는 이러한 최적의 평활계수를 찾는 루틴이 있어서 편리 하지만 SAS, SCA, SPSS에는 이러한 루틴이 없으므로 사용자가 평활계수를 바꾸어가며 한 시점 후의 예측오차의 제곱합을 계산해서 최소가 되는 평활계수를 찾아야하는 어려움이 있다.

최적화를 위해 RATS에서는 simplex 방법을 사용하나 이 방법에 의한 결과가 최적이라는 보장이 없다는 점이 지적될 수 있다.

3) 예측값의 개수, 예측 시작시점 지정

각 패키지에서 예측값의 개수와 예측 시작시점(time origin)의 지정은 필수적이라 할 수 있다. SAS와 SCA에서는 예측값의 개수를 지정하지 않으면 각각 default 값인 12, 24개의 예측값을 구해준다. 그러나

EXECUSTAT, RATS, TSP에서는 예측값의 개수를 반드시 지정해주어야 한다. 예측 시작시점은 SAS를 제외하고 모두 지정해줄 수 있다.

4) 예측구간 한계값과 MSE(Mean squared error) 출력

모든 패키지에서 예측결과로서 실제값과 예측값을 기본적으로 제공하고 있고 EXECUSTAT, SAS의 경우는 예측구간 한계값도 제공한다. RATS, SAS, SPSS, TSP에서는 예측오차를 저장할 수 있지만 EXECUSTAT와 SCA에서는 저장할 수 없다. 최적의 평활계수를 찾을 때 필요한 MSE 출력에 있어서 SCA에서는 MSE, EXECUSTAT에서는 RMSE(root mean squared error), MAE(mean absolute error), MAPE(mean absolute percentage error), ME(mean error), MPE(mean percentage error), TSP에서는 SSE, RMSE 그리고 SPSS에서는 SSE를 출력해 준다. 그러나 RATS와 SAS에서는 출력해 주지 않는다. SAS에서는 MSE대신에 자료의 오차항의 표준편차를 추정해준다.

2.3 사용자 관점에서 바람직한 패키지의 성질

패키지가 수용하고 있어야 할 지수평활법의 종류로는 일반적으로 많이 사용하는 단순지수평활법, 이중지수평활법, Holt의 이중지수평활법, Winters의 가법지수평활법, Winters의 승법지수평활법 정도이면 충분할 것 같다. 지수평활법에서 가장 중요한 것은 적당한 평활계수를 결정하는 것이다. 따라서 패키지 자체에서 최적의 평활계수를 찾아주는 루틴을 제공한다면 자료에 대한 사전정보가 없을 경우나 경험이 부족한 사용자에게 매우 유용할 것이다. 아울러 사용자가 특정한 평활계수를 선택했을 때 그에 해당하는 예측오차제곱합을 계산해 주는 것이 필요하다. 물론 예측값의 개수와 예측 시작시점의 지정도 가능해야 할 것이다. 예측 결과로는 실제값, 예측값, 예측오차, MSE 또는 SSE 정도면 충분하겠다.

3장. ARIMA 모형에 의한 예측

3.1 자기회귀누적이동평균 (AutoRegressive Integrated Moving Average : ARIMA) 모형

시계열 분석에서 가장 많이 사용되는 것이 다음과 같은 ARIMA 모형이다.

$$\text{ARIMA}(p,d,q) : \phi(B)(1-B)^d Z_t = \theta(B)e_t$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p : \text{AR 부분}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q : \text{MA 부분}$$

d : 차분차수

일반적으로 많이 사용되고 있는 시계열 모형화 방법은 Box와 Jenkins(1976)가 제안한 방법이다. Box와 Jenkins 방법은 모형의 식별(model identification), 모형의 추정(model estimation), 모형의 검진(model diagnostic checking)의 세 단계를 반복해서 모형을 선택한다.

모형의 식별단계에서는 시계열도표(Time series plot), 표본자기상관함수(Sample AutoCorrelation Function: SACF), 표본부분자기상관함수(Sample Partial AutoCorrelation Function: SPACF)등을 이용해서 AR 차수 p , 차분차수 d , MA 차수 q 를 결정한다. 즉, 관측된 시계열의 정상성 여부를 조사한 후 비정상적인 경우에는 변수 변환이나 차분법을 사용해서 정상 시계열을 만든 후에 SACF와 SPACF를 통해 p , q 를

결정한다. 또한 모형식별에 사용되는 통계량으로는 AIC와 SBC가 있으며 AIC와 SBC를 최소화 하는 p, q 를 모형의 차수로 결정한다.

모형의 추정 단계에서 일반적으로 많이 쓰이는 모수의 추정 방법으로는 최우추정법(maximum likelihood estimation), 비조건부 최소제곱추정법(unconditional least squares estimation), 조건부 최소제곱추정법(conditional least squares estimation)이 있으며 이외에도 베이즈추정법(Bayesian estimation)이 있다.

모형의 검진단계에서는 추정된 모수의 유의성 검정과 잔차분석을 통해서 잠정적으로 선택된 모형의 적합 정도를 진단하게 된다. 추정된 모형이 관측된 시계열에 적합하면 추정 후의 잔차가 순수 오차의 성질인 백색잡음과정의 성질을 가져야 한다. 따라서 잔차의 시계열도표, 잔차들의 SACF와 SPACF를 보고 모형의 적합 정도를 판단하게 된다. 또한 Ljung과 Box(1978)에 의해 제안된 통계량 Q 를 사용하여 퍼트맨토우(Portmanteau)검정을 함으로써 모형의 적합 정도를 진단할 수 있다.

$$Q = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_a^2(j)}{n-j} \sim \chi^2(k-p-q),$$

$\hat{\rho}_a(j)$: 시차 j 에서의 잔차계열 $\{a_t\}$ 의 SACF

위에서 언급한 세 단계를 반복해서 최종 모형이 결정되면 그 모형을 이용해서 최소평균제곱오차(Minimum Mean Squared Error: MMSE)방법에 의해 예측을 한다.

3.2 통계패키지에서 ARIMA 모형 비교

1) 모형의 식별단계

표 3에서 알 수 있듯이 모든 패키지에서 기본적으로 SACF와 SPACF를 제공하고 있다. 특히 RATS와 SAS에서는 표본역자기상관함수(Sample Inverse AutoCorrelation Function: SIACF)를, SCA에서는 SIACF와 확장된 자기상관함수(Extended AutoCorrelation Function: EACF)를 제공하고 있다.

패키지에서 ACF를 계산할 때는 구하고자 하는 lag의 수를 지정한다. 지정하지 않은 경우에는 각 패키지의 default 값으로 지정된다. 각 패키지의 default 값은 표 3에 나타나 있다. EXECUSTAT와 TSP에서는 사용자가 반드시 지정해야 한다.

각 패키지에서는 ACF의 처리 결과로 ACF 도표를 그려주며 MINITAB을 제외한 모든 패키지가 ACF의 유의성 검정을 위한 표준오차한계(two standard error limit)를 표시해준다. RATS에서는 CORRELATE와 GRAPH명령문을 사용하여 표시는 할 수 있지만 많은 번거로움이 따른다.

Q 통계량은 MINITAB, SAS, SCA, SPSS, TSP에서는 제공되고 EXECUSTAT, RATS, S-PLUS에서는 제공되지 않는다. AIC와 SBC통계량은 SAS와 SPSS에서는 제공되고 S-PLUS에서는 AIC통계량만 제공한다. EXECUSTAT, MINITAB, RATS, SCA, TSP에서는 AIC와 SBC통계량 어느것도 제공되지 않는다.

2) 모형의 추정단계

각 패키지에서 사용하는 모수추정법에 대해 살펴보면 표 4와 같다. 대부분의 패키지에서는 여러 모수 추정방법 중에서 한가지만 택해서 추정하고 있다. SAS에서는 최우추정법, 조건부 최소제곱추정법, 비조건부 최소제곱추정법을, SCA에서는 최우추정법, 조건부 최소제곱추정법 중에서 사용자가 선택할 수 있다.

표 3. 각 패키지에서 모형식별에 관련된 기능 비교

모형식별 \ 패키지종류	패키지종류								
	EXECUSTAT	MINITAB	RATS	SAS	SCA	S-PLUS	SPSS	TSP	
자기상관함수 (ACF)	●	●	●	●	●	●	●	●	
역자기상관함수 (IACF)			●	●	●				
부분자기상관함수 (PACF)	●	●	●	●	●	●	●	●	
확장된 자기상관함수 (EACF)					●				
자기상관함수 계산시 필요한 lag수 지정 (default)	지정	$\sqrt{n}+10$	$n / 4$	24	36	$\log(n)$	16	지정	
표준오차한계표시	●		●	●	●	●	●	●	
ACF 각 lag 에 대한 Ljung-Box Q 통계량		●	●	●	●		●	●	
상관함수 계산하는 독립된 procedure	●	●	●	●	●	●	●		
AIC 통계량			●	●		●	●		
SBC 통계량				●			●		

● : 있음 n : 표본의 개수
 ☞ 지정 : default 값이 없음.
 반드시 지정해야 함.

3) 모형의 검진단계

모형의 적합성을 검진하기 위해서는 모수의 유의성 검정과 잔차분석을 하게 된다. 잔차분석은 잔차의 SACF, SPACF, 잔차의 시계열도표, 퍼트맨토우검정, Durbin-Watson 통계량을 이용하여 가능하며 각 패키지에서 이들의 수용여부를 알아보면 표 5와 같다. 각 패키지에서 잔차의 ACF를 계산할때는 잔차를 변수에 저장한 후에 계산하는 경우와 변수에 저장하지 않고 추정이나 예측 혹은 검진에 관련된 절차에서 하나의 option을 줌으로서 계산 가능한 경우가 있다. MINITAB, RATS, SCA, SPSS, TSP에서는 잔차를 저장한 후에 계산하는 경우이고 SAS는 option에 의해서만 가능하며 EXECUSTAT와 S-PLUS는 둘다 가능한 경우다. 모든 패키지에서 잔차에 대한 퍼트맨토우 검정과 잔차도표를 제공하고 있고 모수의 유의성 검정을 위해서 모수에 대한 t 값과 유의확률도 제공하고 있다. Durbin-Watson통계량은 RATS와 TSP에서만 제공된다.

표 4. 각 패키지에서 수용하고 있는 모수추정방법 비교

패키지 종류 모형추정		EXECUSTAT	MINITAB	RATS	SAS	SCA	S-PLUS	SPSS	TSP
		모수추정법							
모수추정법	최우추정법 (maximum least squares)				●	●	●	●	
	조건부최소제곱법 (conditional least squares)			●	D ●	D ●			●
	비조건부최소제곱법 (unconditional least squares)	●	●		●				

● : 있음 D : default

표 5. 각 패키지에서 모형검진에 관련된 기능 비교

패키지 종류 모형검진		EXECUSTAT	MINITAB	RATS	SAS	SCA	S-PLUS	SPSS	TSP
		잔차의 SACF	●	●	●	●	●	●	●
잔차의 SIACF			●	●	●				
잔차의 SPACF	●	●	●	●	●	●	●	●	
잔차의 ACF 에 대한 퍼트맨토우검정	●	●	●	●	●	●	●	●	
잔차도표	●	●	●	●	●	●	●	●	
Durbin - Watson 통계량			●					●	

● : 있음

표 6. 각 패키지에서 예측에 관련된 기능 비교

패키지 종류 예측	EXECUSTAT	MINITAB	RATS	SAS	SCA	S-PLUS	SPSS	TSP
예측시점지정	●	●	●	●	●	●		●
예측값의 개수 지정 (default)	지정	지정	지정	24	24	지정	지정	지정
예측구간한계값		●		●		●	●	
표준오차				●	●	●	●	●
예측값과 예측구간한계값의 도표	●	●		●		●	●	

● : 있음
 □ 지정 : default 값이 없음.
 반드시 지정해야 함.

4) 예측단계

예측을 하기 위해서는 최종적으로 결정된 모형을 제시하고 예측 시작시점과 예측하고자 하는 예측값의 개수를 지정한다. 그러면 패키지는 예측 결과로 예측값과 예측구간 한계값, 표준오차 등을 제공해주고 도표도 제공해 준다. 각 패키지의 수용형태를 살펴보면 표 6과 같다.

모든 패키지에서 예측시점을 지정할 수 있고 예측값의 개수를 지정할 수 있다. 예측개수의 지정에 있어서 SAS와 SCA에서는 사용자가 지정하지 않을 수도 있는데 이때는 SAS와 SCA 모두 default 값인 24 개를 예측하게 된다. SAS와 SCA 이외의 경우에는 사용자가 지정해야 한다.

예측 결과에서 예측구간 한계값은 EXECUSTAT, RATS, SCA, TSP에서는 제공되지 않고 MINITAB, SAS, S-PLUS, SPSS에서는 제공되고 있다. 예측에 따른 표준오차는 EXECUSTAT, MINITAB, RATS에서만 제공되지 않고 나머지 패키지에서는 제공된다.

3.3 사용자관점에서 바람직한 패키지의 성질

ARIMA 모형에 의해 예측할 경우 모형화과정은 단계별로, 즉 모형식별, 모형추정, 모형검진 단계들로 구성되어야 할 것이다. 각 단계가 가져야 할 기능들을 살펴보면 먼저 모형식별 단계에서는 기본적으로 SACF, SPACF를 처리할 수 있어야 하고 백색잡음과정에 대한 검정이 가능해야 한다. 물론 ACF의 도표도 제공해야 하는데, 이 때는 표준오차의 한계가 표시되어 있어 ACF도 검정할 수 있어야 한다. 또한 원 시계열 뿐만 아니라 차분된 시계열이나 대수변환된 시계열의 모형식별이 option만으로 가능해야 효율적일 것이다. 그리고 모형식별 단계에서 필요한 기능들을 따로 처리할 수 있는 독립된 절차가 존재하여 차분한 시

계열이나 대수변환한 시계열에 대해, 특정한 ACF를 필요로 할 때마다 구할 수 있어야 효율적일 것이다.

모형추정 단계에서는 모수추정법으로 최우추정법, 조건부 최소제곱추정법 그리고 비조건부 최소제곱추정법이 가능하여 사용자가 선택해서 사용할 수 있어야 할 것이며 적어도 비조건부 최소제곱추정법은 제공되어야 할 것이다.

모형검진 단계에서는 모수의 유의성 검정과 잔차분석을 할 수 있는 기능들, 즉 모수에 대한 유의확률, 잔차의 SACF, SPACF 처리가 가능해야 하고 잔차의 시계열도표, 잔차에 대한 퍼트맨토우 검정이 이루어져야 할 것이다.

예측 단계에서는 먼저, 예측 시작시점의 지정과 예측값의 개수 지정이 가능해야 하고 처리 결과로 예측값과 예측구간 한계값을 제공해야 할 것이다. 아울러 예측값과 예측구간 한계값을 동시에 시간축에 대해 그린 도표를 제공함으로써 시각상으로 예측의 정확도가 확인되도록 하는것이 바람직 할 것이다.

주어진 자료를 모형수립을 위한 부분(modeling sample)과 검증을 위한 부분(validation sample)으로 나누어 선택한 모형의 예측력을 비교할 수 있도록 해 주는 기능도 주어져야 할 것이다.

4장. 결론

시계열 분석에 사용되는 8종류의 패키지를 비교해 본 결과 어느 특정한 패키지가 좋다고 단정하기는 어렵다. 패키지에 따라 각각 장·단점이 있으며, 패키지 운용상의 편리함과 기능의 다양성이 서로 상충하는 면이 있다. 저자들의 경험에 의해 지수평활법에 의한 예측과 ARIMA 모형에 의한 예측 방법에 국한해서 살펴보면 다음과 같다.

1) 지수평활법에 의한 예측을 할 경우는 최적의 평활계수를 찾는 루틴을 가진 EXECUSTAT, RATS, TSP를 사용하면 편리하다. 특히 계절성이 있는 자료의 경우 EXECUSTAT, TSP보다는 다양한 경우의 처리가 가능한 RATS를 추천하고 싶다. SAS와 SPSS의 경우 평활계수를 찾는 루틴은 없지만 여러가지 경우에 대한 지수평활법이 가능하다는 장점이 있다.

2) ARIMA 모형에 의한 예측을 하는 경우에는 MINITAB이 사용하기에 가장 편리하다. 그러나 MINITAB의 경우 ACF의 유의 수준표시가 없어 차수결정에 불편함을 주고 또한 원하는 차수만의 모형을 지정하기 곤란하다는 문제점이 있다. 모형화의 과정을 단계별로 원한다면 RATS, SAS, SCA, TSP를 사용하는 것이 좋다. EXECUSTAT, TSP, SPSS는 메뉴로 운용되기 때문에 사용자가 패키지의 문법적인 형식을 알지 못하더라도 쉽게 사용할 수 있어 편리하다. S-PLUS의 뛰어난 그래픽 기능을 활용하는 것도 좋을 것이다. 여러가지 기능의 다양성과 운용상의 편리함을 고려할 때, SAS가 가장 바람직한 것 같다.

패키지에서 지수평활법에 의한 예측, ARIMA 모형에 의한 예측을 할때 사용자 관점에서 바람직한 패키지는 다음과 같은 성질을 가져야 할 것이다. 지수평활법에 있어서는 무엇보다 중요한 최적의 평활계수를 찾는 루틴이 존재해야 한다. 분석해야할 자료의 수가 많은 경우 각각의 자료에 대해 사용자가 최적의 평활계수를 찾아 지정한다는 것은 무리이다. ARIMA 모형에서는 각 단계별로 필요로 하는 기능, 즉 모형식별에서는 여러가지 ACF들을 처리하는 절차가 존재해야 하고 모형의 검진 단계에서는 잔차를 변수에 저장해서 다시 여러가지 잔차분석을 하는게 아니라 하나의 option 만으로 가능해 져야 할 것이다. 그리고 예측단계에서는 예측의 정확성의 조사를 위해 예측값과 예측구간 한계값의 도표가 제공되어야 할 것이다.

시계열 자료를 분석할 때 사용자에게 가장 어려운 부분이 바로 모형식별 단계이다. 모형식별에 도움을

주는 여러가지 통계량들이 제공되고 있으나 이것에 충분히 익숙하지 않은 일반 사용자들에게는 많은 어려움이 따른다. 그러므로 가급적 사용자의 개입을 최소로 하는 모형선택 방법이 패키지에서 제공되어야 할 것이다. 또한 숙련된 분석자들을 위해서는 다양한 기능들이 사용자에게 의해 선택될 수 있어야 할 것이다. 실제로 예측이 주 목적일때 분석자는 여러가지 예측방법에 의해 얻은 결과를 서로 비교해보고 이중 한가지를 선택하거나 여러 예측결과들의 혼합 형태로 예측을 하게 된다. 그러므로 각각의 예측방법에 의한 결과들을 한데 모아서 비교할 수 있게 하는 기능도 필요할 것이다. 다른 통계분석에서보다 시계열 분석에서 더욱 중요시되는 것이 바로 그래픽기능이다. 시계열도표를 통해 분석자는 많은 정보를 얻는다. 기존의 패키지에서 제공되는 시계열도표는 정교하지 않을 뿐더러 명령어 사용이 어렵다. 이러한 것들이 앞으로 더욱 개선되고 연구되어야 할 과제이다.

참 고 문 헌

- [1] 김병천(1987). 개인용 컴퓨터에서의 통계패키지의 선택과 활용, 『응용통계연구』, 제 1권 1호, 75-90.
- [2] 허명희·정진환(1990). 탐색적 데이터분석 (EDA) 기능에 관한 통계패키지 프로그램의 비교 검토, 『응용통계연구』, 제 3권 2호, 17-25.
- [3] Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis : forecasting and control, 2nd ed.*, Holden-Day, San Francisco.
- [4] Ljung, G.M. and Box, G.E.P. (1978). On measure of lack of fit in time series models, *Biometrika*, Vol. 65, 297-303.
- [5] Makridakis, S. et al. (1982). The accuracy of Extrapolation(Time Series) methods: results of a forecasting competition, *Journal of Forecasting*, Vol. 1, 111-153.
- [6] *Micro TSP setup Guide*, version 7.0 (1990). Quantitative Micro Software, Irvine, California.
- [7] *RATS User's Manual*, version 4.0 (1992). Estima, Evanston.
- [8] *SAS/ETS User's guide*, version 6, first edition (1988). SAS Institute Inc, Cary, NC.
- [9] *S-PLUS for DOS User's Manual* Vol. 1, version 2.0 (1991). Statistical Sciences, Inc. Seattle, Washington.
- [10] *S-PLUS for DOS User's Manual* Vol. 2, version 2.0 (1991). Statistical Sciences, Inc. Seattle, Washington.
- [11] *Student edition of EXECUSTAT 3.0 from the developers of STATGRAPHICS* (1993), Duxbury Press. Belmont, California.
- [12] *The SCA stistical system reference manual for forecasting and time series*, version III (1986). Scientific Computing Associates, Dekalb, Illinois.
- [13] *The SCA stistical system reference manual for fundamental capabilities* (1991). Scientific Computing Associates, Dekalb, Illinois.
- [14] Yurkiewicz, J. (1993). Forecasting Software Clearing Up A cloudy Picture, *OR/MS Today*, 64-75.