

시계열 및 예측모델 선택과정에서 스펙트럼의 이용

(The use of spectral analysis in choosing time series and forecasting models)

전 덕 빈*

Abstract

A spectrum analysis method is presented with an example as an aid to Box and Jenkins' model identification procedure, where the theoretical spectrum of ARMA model and its confidence intervals derived by chi-square distribution are compared. An APL (A Programming Language) program for the method is developed for the 16-bit personal computer.

1. 서 론

시계열 분석 및 예측(Time series analysis and forecasting)시 시계열 자료의 underlying process와 관계된 이론의 부족 또는 상황이나 상식에 근거한 지식의 부족으로 모형설정(Model building)이 불가능한 경우, Box-Jenkins[5]의 ARIMA(Autoregressive-integrated moving average) 모형으로의 분류가 도움이 될 수 있다. Box-Jenkins의 시계열 분석법은 관계된 시계열 자료의 샘플 autocorrelation을 추정하여, 그와 유사한 이론적 autocorrelation 함수를 갖는 ARI-

MA 모형을 선택하고, 그 모형의 특성을 활용하여 분석하고 예측한다. 그러나 샘플 autocorrelation과 이론적 autocorrelation의 비교를 통한 모형의 선택과정에서 통계를 비롯한 제 이론들의 이해뿐 아니라 전문가의 실제적인 경험을 요한다. Box-Jenkins의 시계열 분석법이 위에서 언급한 상황하에서 강력한 수단 중의 하나이면서도 실제적으로 산업체 및 그밖의 현장에서 널리 사용되지 못하는 이유가 바로 그점에 있다. 이러한 모형선택과정(Model identification procedure)에서의 난점을 극복하기 위하여 Akaike[1], Gray, Kelley & McIntire[6],

* 경희대학교 공과대학 산업공학과

본 논문은 1987년도 문교부 자유공모과제 학술연구조성비에 의하여 연구되었음.

Woodward & Gray[13] 등등이 여러가지 방법들을 제시하여 왔다.

본 연구에서는 스펙트럼 분석(Spectrum analysis)을 통하여 Box-Jenkins 분석법중 초기 단계의 모형선택 과정과 최종 단계의 모형의 적합여부 검사과정(Diagnostic checking procedure)에서 도움이 될 수 있는 방법을 제시하며 16비트 퍼스널 컴퓨터에서 편리하게 사용할 수 있는 프로그램 "Spectra"를 소개한다. 통계이론 중심의 현존하는 시계열 분석 및 예측과 관련된 소프트웨어 중에는 간혹 fast Fourier

transform과 스펙트럼 분석과 관계된 기본적인 기능들을 갖고 있는 것들이 있다(Mahmoud외 3인 [7,8], Beaumont[2] 참조). "Spectra"에서는 그밖에 스펙트럼의 추정치와 Box-Jenkins의 ARMA(Autoregressive and moving average) 모형들의 이론적 스펙트럼, 그와 관련된 신뢰구간 등을 유도하며 그 결과를 그래프로 처리하여 실제 사용이 용이하다. "Spectra"는 컴퓨터 언어 APL(A Programming Language)으로서 작성되었으며, 그 언어의 이해없이도 사용자가 대화식으로 편리하게 사용할 수 있다.

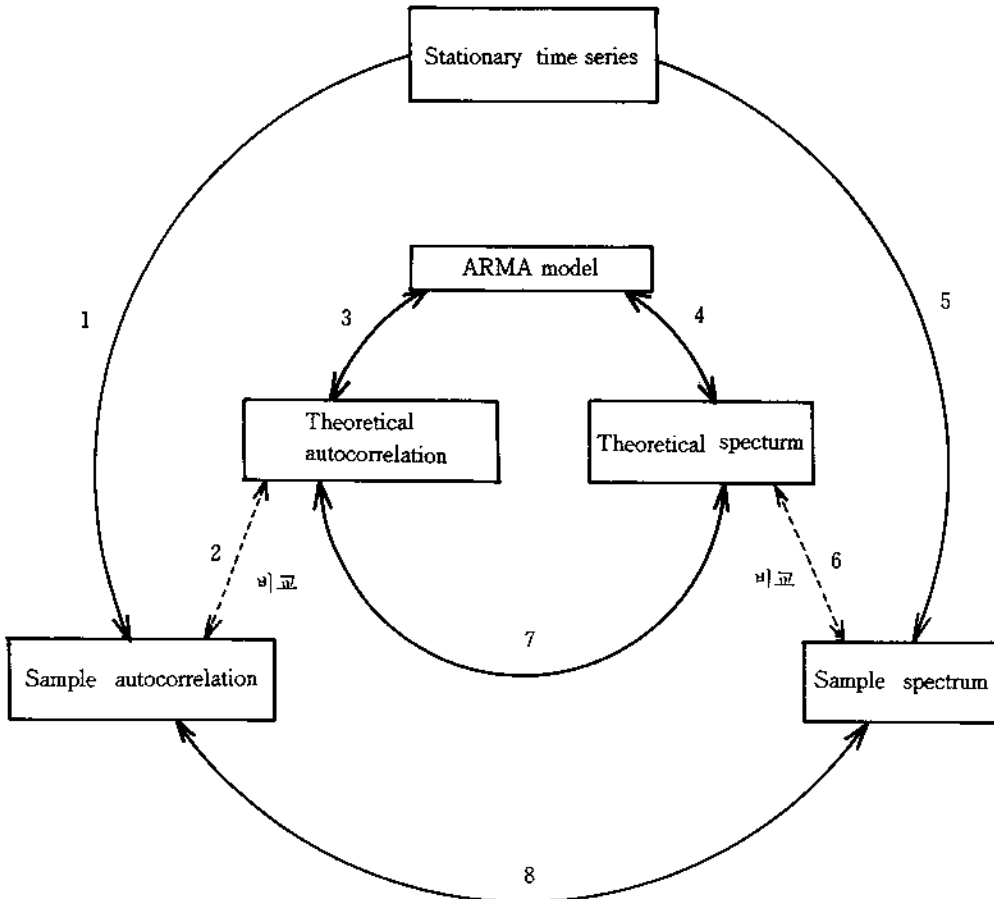


그림1. ARMA 모형의 Autocorrelation 및 Spectrum 과의 관계

2. 모형선택 과정에서의 스펙트럼 분석법의 활용

시계열의 특성은 두가지 각도에서 관찰 가능하다. 시간의 관점에서 접근하는 방법(Time domain approach)의 하나로서 일정시간 간격을 갖는 stationary time series간의 상관관계 즉 autocorrelation을 관찰, 분석할 수 있다. 또한 stationary time series의 주기(Frequency)가 서로 다른 sine 및 cosine 곡선으로 구성되어 있다고 가정하여, 어떤 주기를 갖는 곡선들이 그 시계열의 특성을 주도하는가를 연구하는 주기의 관점에서 접근하는 방법(Frequency domain approach)을 사용하여 분석이 가능하다. 일반적으로 전자의 방법을 사회과학 분야에서 많이 사용하며, 후자의 방법은 공학 분야에서 과거부터 널리 사용되어 왔다. 이 두가지 방법중 어느 한 방법이 다른 방법보다 우월하다고 주장하기는 곤란하며, 분석하고자 하는 대상 시계열의 특성에 따라 어떤 방법이 다른 방법보다 더 많은 도움을 줄 수 있는 있다. 즉 두가지 방법은 보완관계에 있으므로 병행하여 사용하는 것이 바람직하다. 이러한 점을 고려하여 다음과 같은 모형선택 과정의 보완을 제시한다.

Box-Jenkins의 분석법과 동일하게 첫째 분석하고자 하는 시계열을 differencing하여 stationary time series로 변환하고, 둘째 autocorrelation 추정치를 분석하여 유사한 이론적인 autocorrelation 함수를 갖는 후보모형들을 비교 선택하고, 셋째 선택된 모형에 관계되는 모수(Parameter)들을 추정하고, 넷째 모형의 적합 여부를 검사하고, 이러한 단계들을 반복하여 적합한 모형들을 선택한다. 이 과정은 그림1에서의 경로 1, 2, 3에 해당한다. 이러한 과정들을 통하여 선택된 ARMA 모형들의 이론적 스펙트럼과 시계열 자료로부터의 스펙트럼의 추정치 및 신뢰구간을 비교하여 즉 그림1의 경로 4, 5, 6에 해당되는 과정을 추가하여 모형을 선택한다.

3. Spectra의 구성 및 절차

이론적 autocorrelation과 스펙트럼과의 관계를 활용하여 ARMA 모형의 autocorrelation 함수로부터 이론적 스펙트럼의 유도가 가능하지만(그림1의 경로 3과 7), 경로 4에 의하여 ARMA 모형의 이론적 스펙트럼으로부터 직접 유도하였다(Box-Jenkins [5] 참조). 즉 ARMA(p, q) 모형의 이론적 스펙트럼은 주기를 f 로 표시하고 Box-Jenkins의 표기법을 사용하면 다음과 같다.

$$2\sigma_a^2 \frac{|1 - \theta_1 e^{-i2\pi f} \dots - \theta_q e^{-i2\pi qf}|^2}{|1 - \phi_1 e^{-i2\pi f} \dots - \phi_p e^{-i2\pi pf}|^2},$$

$$0 \leq f \leq 0.5 \quad (식1)$$

또한 스펙트럼 추정치 역시 그림 1의 경로 1과 8을 사용하여 샘플 autocorrelation 으로부터 유도가 가능하지만, 경로 5을 통하여 직접 시계열 자료로부터 유도하였다. 이 과정에서 데이터의 시작과 끝부분의 각 10퍼센트에 split cosine bell taper를 적용한 후, fast Fourier transform을 사용 계산하였으며, 0과 0.5사이의 주기값들의 원하는 간격을 사용자가 명시하거나, 또는 적절한 간격을 프로그램 내에서 제공하도록 하였다. 이론적 스펙트럼의 신뢰구간은 정규분포를 사용하여 근사적으로 유도가 가능하지만, 보다 정확한 chi-square 분포를 사용하여 계산하였다. 전체 스펙트럼을 위한 simultaneous confidence band는 대단히 긴 시계열에서만 그 효과가 나타나므로 Woodroffe와 Van Ness[12]의 결과를 사용하지는 않았다(Bloomfield[3] 참조). 이 과정에서 사용자에게 의한 smoothing bandwidth의 명시가 가능하며, BMDP Statistical Soft ware[4]에서와 같이 세가지 값중 선택도 가능하도록 하였다.

Spectra는 384K RAM, 그래픽스 어댑터 및 그에 적합한 모니터를 갖춘 16비트 퍼스널 컴퓨터에서 사용 가능하며 단계별 내용, 입력자료, 출력결과 및 기능은 표1과 같다.

표1. Spectra의 구성 및 특성

단 계	내 용	입력자료	출력결과	기 능
1	ARMA 모형의 명시	- Differencing order - Stationary mean - Autoregressive parameters - Moving average parameters - Variance of noise	- 입력 자료들의 확인을 위한 출력	- 교정 - 차후 사용을 위한 저장 - 시계열 자료의 스펙트럼 관찰시 생략
2	시계열 자료입력	- 시계열 자료	- 자료의 순서와 값들의 확인을 위한 출력	- 재입력 및 개별 교정 - 차후 사용을 위한 저장 - 이론적 스펙트럼 관찰시 생략
3	샘플 스펙트럼 계산	- 주기간 간격수 - bandwidth	- 주기별 스펙트럼 추정치 - 95% 신뢰구간 - chi-square 분포의 degrees of freedom	- 필요시 적합한 주기간 간격수 제공 - 3가지 bandwidth중 선택 - 주기간 간격수 및 bandwidth 변화 - 출력 결과 저장
4	ARMA 모형의 이론적 스펙트럼 계산	- 주기간 간격수	- 주기별 이론적 스펙트럼	- 필요시 적합한 주기간 간격수 제공 - 출력 결과 저장
5	이론적 스펙트럼 및 샘플 스펙트럼 비교	- 단계 3,4의 결과 자동 입력	- 주기별 이론적 스펙트럼 및 신뢰구간 - 주기별 이론적 스펙트럼의 신뢰구간 내 존재여부	- 단계 3,4의 주기간 간격수 비교 및 교정
6	이론적 스펙트럼 및 신뢰구간 그래프 작성	- 단계 5의 결과 자동 입력	- 이론적 스펙트럼 및 신뢰구간의 그래프 - 로그 스케일 그래프	- 모니터 및 프린터 출력

*사용자 편의를 위한 可 또는 否 형태의 입력자료는 모두 생략하였음.

*필요시 특정 단계의 생략 또는 소급 가능

4. 실험계산 예

Box-Jenkins [5]는 그의 책에서 예로서 사용한 시계열 D(Chemical process viscosity readings : every hour)를 위한 모형으로서 다음 두가지 형태를 제시하였다.

$$\text{모형 1 : } Z_t - 0.87 Z_{t-1} = 1.17 + a_t, \sigma_a^2 = 0.090$$

$$\text{모형 2 : } \nabla Z_t = a_t - 0.06a_{t-1}, \sigma_a^2 = 0.096$$

그림2와 3은 (0, 0.5)간의 주기를 256개의 간격으로 나누어서 smoothing bandwidth=0.0176, degrees of freedom=9.8의 chi-square 분포를 사용한 모형1의 Z_t 와 모형2의 ∇Z_t 의 이론적 스펙트럼과 시계열 D의 전 자료를 사용한 95% 신뢰구간이다. 그림에서 중간에 위치한 유연한 곡선들이 각 모형들의 이론적 스펙트럼을 표시한다.

명확한 관찰을 위하여 그림 3에서는 스펙트럼과 신뢰구간의 실제값을, 그림 2에서는 그 값들의 밑(base)이 10인 로그를 취한 값을 나타내었다. 참고로 257개 주기별 이론적 스펙트럼 중 모형1의 경우에는 16개가, 모형 2의 경우에는 30개가 신뢰구간을 벗어나고 있다.

그림2의 경우 전체적으로 거의 완벽하게 모형 1의 주기별 이론적 스펙트럼들이 신뢰구간내에 존재하며, 그림3의 경우(0, 0.28)내의 낮은 주기들에서 모형2의 이론적 스펙트럼들이 신뢰구간을 지속적으로 벗어나고 있다. 모형의 적합성 여부를 현재 제시하고 있는 스펙트럼 분석 결과만으로서 결정할 수는 없다. 그 보다는 다른 criteria들과 비교 검토하여 결론을 짓는 것이 바람직하다. 그러나 본 연구에서 제시한 방법이 추가됨으로서 모형의 적합성 정도의 이해 및 결정에 도움이 될 수 있다.

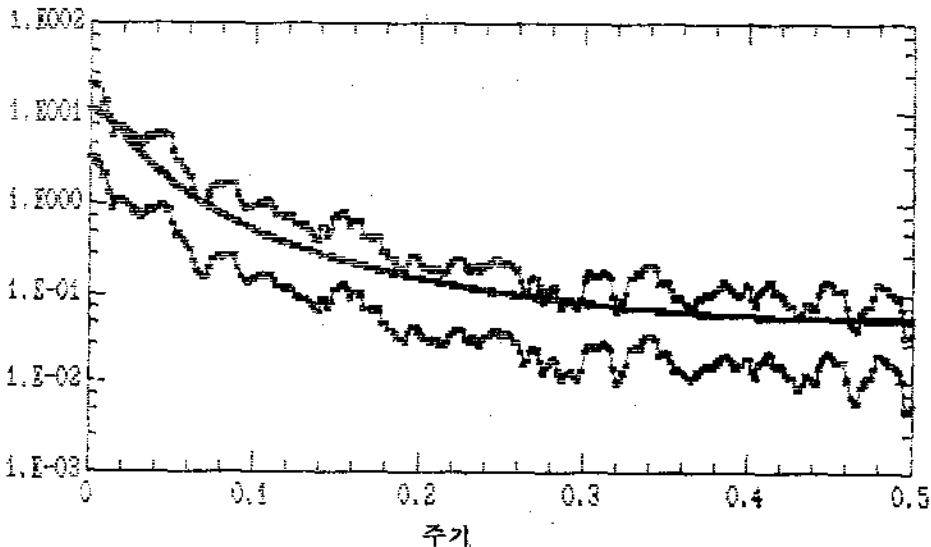
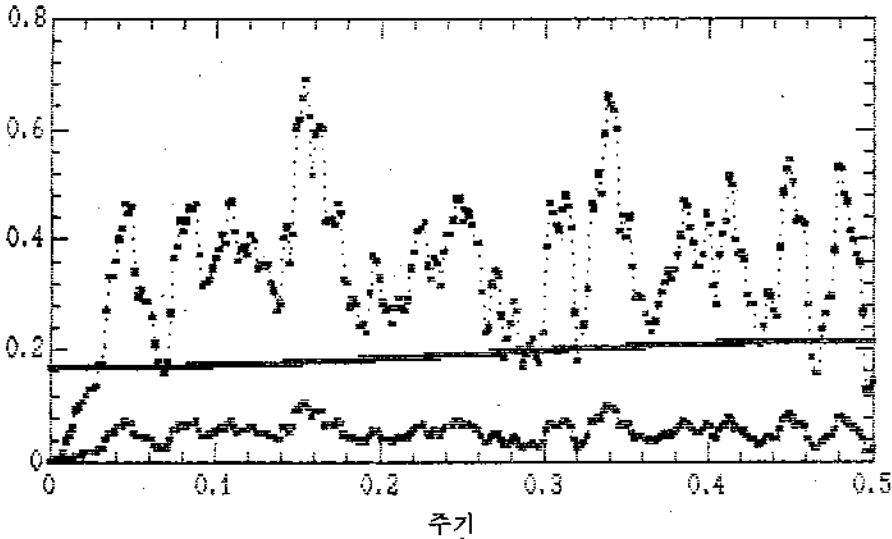


그림2. 모형1의 주기별 이론적 스펙트럼과 95% 신뢰구간의 로그를 취한 그래프

그림3. 모형2의 ∇Z_t 의 주기별 이론적 스펙트럼과
95% 신뢰구간의 그래프



Reference

1. Akaike, H., "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, pp. 716-723 Dec. 1974.
2. Beaumont, C., "Software Reviews", *Journal of Forecasting*, Vol. 6, No. 1, pp. 67-74, 1987.
3. Bloomfield, P., *Fourier Analysis of Time Series: An Introduction*, John Wiley & Sons, 1976.
4. *BMDP Statistical Software*, University of California Press, Berkeley, 1983.
5. Box, G.E.P. and Jenkins, G.M., *Time Series Analysis: forecasting and control*, 2nd ed., Holden-Day, Oakland, 1976.
6. Gray, H.L., Kelly, G.D., and McIntire, D.D., "A New Approach to ARMA Modeling", *Communications in Statistics*, Vol. B7, pp. 1-77, 1978.
7. Mahmoud, E., Rice, G., McGee, V.E., and Beaumont, C., "Mainframe Multipurpose Forecasting Software: A Survey", *Journal of Forecasting*, Vol. 5, No. 2, pp. 127-137, 1986.
8. Mahmoud, E., Rice, G., McGee, V.E., and Beaumont, C., "Mainframe Specific Purpose Forecasting Software: A Survey", *Journal of Forecasting* Vol. 5, No. 1, pp. 75-83, 1986.
9. Marija, J.N., *SPSS/PC+ for the IBM PC/XT/AT*, SPSS Inc., 1986.
10. *SAS Introductory Guide for Personal Computer*, Version 6, SAS Institute Inc., 1985.
11. *STATGRAPHICS Statistical Graphics System*, STSC Inc., 1985.
12. Woodroffe, M.B. and Van Ness, J.W., "The Maximum Deviation of Sample Spectral Densities", *Annals of Mathematical Statistics*, Vol. 38, pp. 1558-1569, 1987.
13. Woodwad, W.A., and Gray, H.L., "On the Relationship Between the S Array and the Box-Jenkins Method of ARMA Identification", *Journal of the American Statistical Association*, Vol. 76, No. 375, pp. 579-587, Sep. 1981.