

시계열분석에 대한 R 패키지의 활용*

김희재¹

요 약

시계열자료에 대한 분석 및 예측을 위한 통계패키지는 SAS, SPSS, MINITAB 그리고 S-plus 등을 일반적으로 사용한다. 상기의 패키지들은 서로 여러 가지 장단점이 있을 수 있으나 공통된 점은 고가의 유료 패키지라는 점이다. 본 연구에서는 무료 공개 소프트웨어인 R 패키지를 활용하여 시계열자료에 대한 기본적인 분석과 ARIMA 모형 및 승법계절 ARIMA 모형에 대한 분석 및 예측이 가능함을 보이며 또한 SAS, SPSS 그리고 MINITAB을 사용하는 경우에 대한 비교표를 제공하여 시계열분석을 위한 사용자에게 여러 가지 상황에 따르는 선택의 폭을 넓히며 R 패키지를 이용하여 시계열분석에 대한 연구개발에 일조토록 하는데 연구의 목적을 두었다.

주요용어 : 시계열분석, R 패키지, ARIMA 모형, 승법계절 ARIMA 모형.

1. 서론

R 시스템의 library들은 1989년 New Zealand의 Auckland 대학교 통계학과와 Robert Gentleman과 Ross Ihaka 교수에 의해 처음으로 만들어졌으며, 1991년부터 GPL(general public license) 즉, 개발의 처음과 종료까지 무료 공개 소프트웨어로 등록되어 R재단의 CRAN(comprehensive R archive network) 홈페이지를 통해서 무료로 배포되고 있다. 현재 Version 2.12.0 (2010-10-15)까지 개발되어 있으며, 특별한 관심을 갖는 것은 자료의 조직화, 시각화, 분석을 목적으로 개발된 객체지향형 프로그램 언어인 기존의 S-plus가 갖고 있는 환경, 사용법이 거의 일치하면서도 무료이며 특히 소스가 공개되어 있기 때문에 사용자의 사용목적에 따라 R 자체의 수정도 가능하다는 점을 들 수 있다. R은 S-plus에 비해 자유로운 개발환경과 수많은 library의 제공으로 여러 연구 분야에 두루 사용되고 있으며 특히 생물학의 DNA microarray 분석 분야에서는 표준 시스템으로 사용되고 있다. R은 UNIX, Windows, 그리고 Macintosh 환경 등 대부분의 컴퓨팅 환경을 지원하고 또한 R은 객체지향형 프로그램 언어이므로 JAVA, Python, C, C++, FORTRAN 등의 언어와 인터페이스가 가능하고 각종 DBMS의 데이터에 접근할 수도 있으며 최근에는 SPSS-R integration package를 SPSS에 탑재하여 SPSS 사용 환경에서 R의 다양한 library 들을 활용하고 있다.

R 패키지를 활용한 기존의 연구에는 샘플링 검사에 대한 R 시스템의 활용에 관한 고찰(김희재, 2003)과 R을 이용한 Weibull 수명분포의 모수추정(김희재, 2005) 등을 들 수 있으며, 시계열분석에 관한 최근의 연구에는 주택 가격과 거시경제 변수간의 시계열분석(김연형 등, 2006), 금융시계열의

*이 논문은 2010학년도 경성대학교 학술연구비지원에 의하여 연구되었음.

¹608-736 부산광역시 남구 대연동 110-1, 경성대학교 통계정보학과 교수. E-mail : hjkim@ks.ac.kr

[접수 2010년 12월 23일; 수정 2011년 1월 23일; 게재확정 2011년 1월 26일]

상관행렬 속성분해를 위한 새로운 방법 연구(엄철준 등, 2007) 그리고 스펙트럼 분석을 이용한 시계열 자료의 패턴 분류(전세봄 등, 2009) 등을 들 수 있다.

2. 시계열분석에 대한 R 패키지의 활용

시계열자료는 분석 목적에 따라 진동수 영역에서의 분석법과 시간영역에서의 분석법으로 대별될 수 있으며 시간영역에서의 분석법으로는 평활법, 분해법 그리고 확률과정에 의한 분석법 등으로 대별될 수 있다. 확률과정에 의한 분석법은 흔히 Box-Jenkins의 ARIMA 모형 및 ARIMA 모형의 활용으로 시계열자료를 분석하고 예측한다. 주어진 시계열자료를 Box-Jenkins 모형으로 분석 및 예측하는 경우 일반적으로 ARIMA 모형 또는 승법계절 ARIMA 모형으로 분석 및 예측을 하며 또한 분석에 사용되는 도구(패키지)로는 흔히 SAS, SPSS, MINITAB 및 S-plus 등이 사용되고 있으나 R 패키지를 활용하는 경우는 거의 전무한 것이 작금의 실정이다.

본 연구에서는 R 패키지를 활용하여 ARIMA 모형에 대한 자기회귀함수(ACF: auto correlation function)와 편자기상관함수(PACF: partial autocorrelation function)에 대한 모의시계열의 생성과 그래픽적 고찰을 기타의 패키지들과 비교하며 또한 1970년 1사분기부터 2010년 2사분기까지의 우리나라의 국내총생산(GDP) 분기별 시계열자료(단위는 10억)에 대하여 승법계절 ARIMA 모형의 수립과 예측 및 적합측도의 절차를 각 단계별로 기타의 패키지들과 비교하였다.

2.1. 시계열자료의 생성과 시계열그림

주어진 시계열자료를 시계열속성을 갖는 자료로 변환하는 R 패키지(이하 R)의 함수는 ts()이며 이를 활용한 R script는 다음과 같이 작성될 수 있다.

《R script 1》 시계열자료의 생성과 시계열그림

```
gdp <- scan("e:/gdp.txt")
(gdp.ts <- ts(gdp, start=c(1970, 1), frequency=4))
plot.ts(gdp.ts)
```

SAS나 SPSS 혹은 MINITAB 등에서 각각의 명령문은 전체 프로그램의 일부분으로서 수행되지만 R은 객체지향형 프로그램 언어이므로 독립된 함수(함수객체)가 각 단계별로 이들을 수행한다. 또한 상용되는 기타 패키지들의 key word와의 관계는 Table 1과 같이 요약 정리할 수 있다.

Table 1. Generating time series and picture of time series

Package	Generating time series	Picture of time series
R	ts(gdp, start=c(1970, 1), frequency=4)	plot.ts(gdp.ts)
SAS	date=intnx('qtr','01jan1970'd, _n_-1); format date yyq. ;	proc gplot ; plot gdp*date=1;
SPSS	date quarter 1 4 year 1970.	tsplot var=gdp /id=date_ .
MINITAB	index	tsplot 'Gdp' ; index; connect .

2.2. 시계열모형 구축을 위한 시계열자료의 부분선택

시계열속성을 갖는 시계열자료로부터 시계열모형 구축을 위한 시계열자료의 부분선택 기능의 R 함수는 `window()`이며 이를 활용한 R script는 다음과 같이 작성될 수 있다.

《R script 2》 시계열모형 구축을 위한 시계열자료의 부분선택

```
(zt <- window(gdp.ts, c(1970,1), c(2008,2)))  
plot.ts(zt)
```

SAS에서는 data 문에서 if 명령문으로 시계열자료를 선택하여 구축할 수 있으며, SPSS에서는 use 명령문으로 시계열자료를 부분 선택할 수 있다. MINITAB에서는 copy 명령문과 use 부명령문으로 부분 선택할 수 있다. 시계열자료에 대한 부분선택 수행의 패키지별 key word를 요약 정리하면 Table 2와 같다.

Table 2. Extracting the subset of the time series

Package	Extracting the subset of the time series
R	<code>window(gdp.ts, c(1970,1), (2008,2))</code>
SAS	<code>if _n_ < 155;</code>
SPSS	<code>use year 1970 quarter 1 thru year 2008 quarter 2.</code>
MINITAB	<code>copy c1 into c2 ; use 1: 154 .</code>

2.3. 시계열의 분산과 추세 및 계절추세의 안정화

시계열자료에 대한 분산의 안정화 필요여부는 일반적으로 시계열 그래프를 사용하여 판별하며, 분산의 안정화 필요시에는 흔히 log 변환 등을 이용하여 분산을 안정화시키고 추세 및 계절추세의 안정화는 일반차분 및 계절차분을 통하여 시계열자료를 안정화시키는데 시계열의 안정화에 관한 R script는 다음과 같이 작성될 수 있다.

《R script 3》 시계열의 분산과 추세 및 계절추세의 안정화

```
lnzt <- log(zt)  
dif1.lnzt <- diff(lnzt)  
dif41.lnzt <- diff(dif1.lnzt, differences=4)  
plot.ts(dif41.lnzt)
```

SAS에서는 시계열 그래프 활용 이외에 매크로 %logtest를 이용하여 로그변환 전후의 AIC 또는 SBC 값의 비교에 의하여 분산안정화 필요여부를 결정할 수 있다. 분산과 추세 및 계절추세의 안정화 수행에 대한 패키지별 key word를 요약 정리하면 Table 3과 같다.

Table 3. Stabilization of variance, trend and seasonal trend

Package	Variance	Trend	Seasonal trend
R	<code>log(zt)</code>	<code>diff(zt)</code>	<code>diff(zt, differences=#)</code>
SAS	<code>log(zt)</code>	<code>diff#(zt)</code>	<code>diff#(zt)</code>
SPSS	<code>ln(zt)</code>	<code>diff(zt)</code>	<code>sdif(zt)</code>
MINITAB	<code>loge(zt)</code>	<code>differ #</code>	<code>differ #</code>

분산의 안정화는 자연대수 변환 이외에도 역수나 제곱근 등의 변환도 가능하며 Table 3에서 #가 4인 경우는 분기별, 12인 경우는 월별 시계열 자료인 경우에 해당되며 #가 1인 경우는 생략이 가능하다.

2.4. 모의 시계열의 생성

시계열 Z_t 가 비정상 자기회귀-이동평균과정인 $ARIMA(p, d, q)$ 모형을 따르는 경우의 일반적인 형태는 B 를 후진연산자, δ 를 상수항, $(1-B)^d Z_t$ 를 Z_t 의 d 차분 시계열이라고 할 때 아래와 같이 나타낼 수 있다.

$$\phi(B)(1-B)^d Z_t = \delta + \theta(B)\epsilon_t \quad (2.1)$$

여기서, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.

$ARIMA(p, d, q)$ 모형에 대한 모의 시계열의 생성은 시계열분석의 학습이나 연구를 위하여 매우 중요하다고 할 수 있는데, SAS나 MINITAB에서는 $d=0$ 인 경우 즉, $ARIMA(p, q)$ 모형인 경우는 사용자가 명령문을 직접 작성하여 용이하게 생성할 수 있으나 d 가 1 이상인 $ARIMA(p, d, q)$ 모형에 대한 모의 시계열의 생성은 다소 난해하다. 특히 SPSS에서 명령문을 직접 작성하여 $ARIMA(p, d, q)$ 모형에 대한 모의 시계열을 생성하는 것은 매우 난해하다.

R에서는 시스템에서 제공하는 모의 시계열 생성함수 `arima.sim()` 함수를 이용하여 $ARIMA(p, d, q)$ 모형에 대한 모의 시계열의 생성이 가능하다. R에서는 또한 이론적인 자기상관함수(ACF)와 편자기상관함수(PACF)에 대하여도 시스템에서 제공하는 `ARMAacf()` 함수의 이용이 가능하므로 이를 활용하면 이론적인 자기상관함수(ACF)와 편자기상관함수(PACF)의 그래프도 쉽게 출력할 수가 있으므로 시계열분석의 학습에 매우 유용하게 이용될 수 있다.

《R script 4》 $ARIMA(p, d, q)$ 모형에 대한 모의 시계열의 생성

```
zt1=arima.sim(n=100, sd=sqrt(0.18), list(order=c(1,1,1), ar=0.9, ma=-0.23))
```

```
zt2=arima.sim(n=63, sd=sqrt(0.18), list(ar=c(0.89,-0.49),ma=c(-0.28, 0.25)))
```

```
plot.ts(zt1); plot.ts(zt2)
```

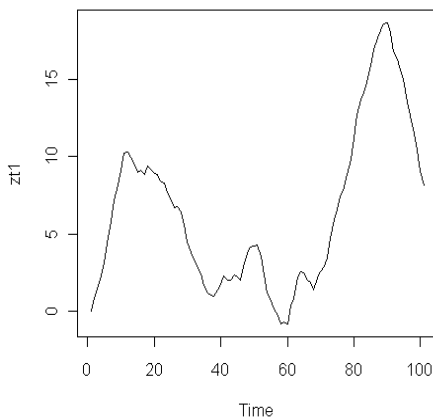


Figure 1.1. $ARIMA(0.9,1,-0.23)$

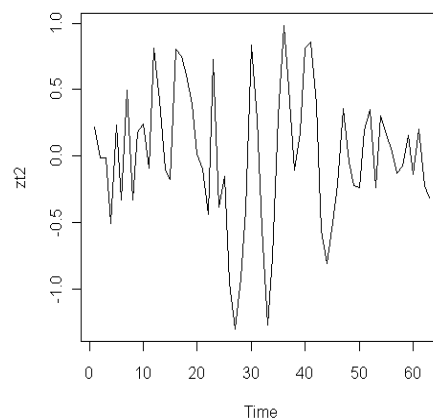


Figure 1.2. $ARIMA(-0.49,0,0.25)$

2.5. 모형의 식별

모형 식별을 위한 통계량은 여러 가지가 있을 수 있는데 상용적으로는 흔히 자기상관함수(ACF)와 편자기상관함수(PACF)를 이용하여 식별한다. 기타의 식별 통계량으로 역자기상관함수(IACF) 등을 이용할 수 있는데, 각 패키지별 모형의 식별기능에 따른 비교는 통계 패키지에서의 시계열 분석방법의 비교(김수화 등, 1994)를 참고하면 가능하다. R 패키지를 이용한 자기상관함수(ACF)와 편자기상관함수(PACF)에 관한 R script는 다음과 같이 작성될 수 있다.

《R script 5》 자기상관함수(ACF)와 편자기상관함수(PACF)

```
acf(diff(lnzt), 48); pacf(diff(lnzt), 48)
acf(diff(diff(lnzt), 4), 48); pacf(diff(diff(lnzt), 4), 48)
```

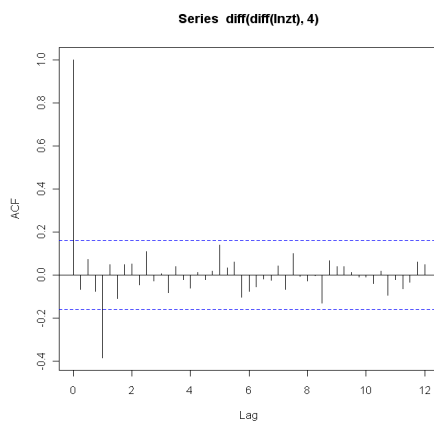


Figure 2.1. ACF for ARIMA(p,d,q)

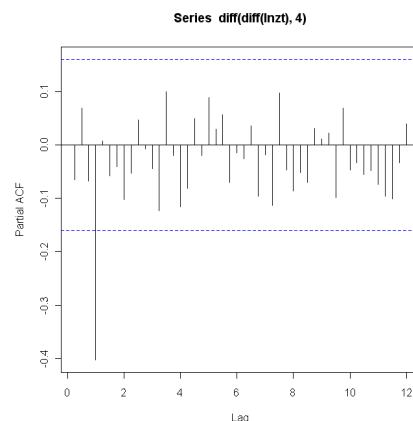


Figure 2.2. PACF for ARIMA(p,d,q)

SAS, SPSS 및 MINITAB에서는 ACF와 PACF는 그래프와 함께 각 시차에 따르는 상관계수 및 편상관계수도 함께 출력된다. R에서 시차에 따르는 상관계수 및 편상관계수의 출력은 좌우 괄호를 추가하여 이룰데면, (acf(diff(lnzt), 48))을 시행하면 각 시차에 따르는 상관계수가 출력된다. 시계열 자료에 대한 자기상관함수(ACF)와 편자기상관함수(PACF)에 대한 각 패키지별 key word를 요약 정리하면 Table 5와 같다.

2.6. 모형의 적합

모형 식별의 결과로부터 모형의 적합(모수추정)을 위한 R script는 arima()함수를 이용하여 구할 수 있으며 대수변환된 시계열이 $ARIMA(0,1,0)(1,1,1)_4$ 로 식별된 경우 모형의 적합을 위한 R script는 다음과 같다.

《R Script 6》 모형의 적합을 위한 R script

```
arima(lnzt, order=c(0,1,0), seasonal=list(order=c(1,1,1), period=4),
      nclude.mean = TRUE, method = c("CSS-ML", "ML", "CSS"))
```

Table 5. ACF and PACF

Package	ACF and PACF
R	acf(diff(diff(lnzt), 4), 48); pacf(diff(diff(lnzt), 4), 48)
SAS	proc arima; identify var=lnzt(1,4) nlag=48;
SPSS	acf vars=dif121.lnzt/pacf. or acf vars=zt/diff=1/sdiff=1/ln.
MINITAB	acf 48 c4 and pacf 48 c4

R에서는 `arima()` 함수의 출력 결과로부터 일반 및 계절 요인에 대한 계수와 이들의 표준오차 그리고 대수변환된 시계열에 대한 분산과 대수우도(log likelihood) 및 AIC 통계량 값을 구할 수 있다. 또한 `arima()` 함수에 사용된 옵션의 의미는 R에서 도움말을 참조하면 가능하며 추정방법이 default인 경우는 최대가능도법(ML)이 적용된다. 모형식별 통계량으로 SAS에서는 AIC 이외에 SBS 통계량, SPSS에서는 BIC 통계량 등을 추가로 더 구할 수 있다. 모형의 적합에 대한 패키지별 key word를 요약 정리하면 Table 6과 같다.

Table 6. Model Estimation

Package	Model Estimation
R	arima(lnzt, order=c(0,1,0), seasonal=list(order=c(1,1,1), period=4)
SAS	proc arima; identify var=lnzt(1, 4); estimate p=(4) q=(4);
SPSS	arima variables = zt /model = (0,1,0) (1,1,1) noconstant ln.
MINITAB	arima (0,1,0)*(1,1,1) 4 'lnzt' resid c6 fitted c7

2.7. 모형의 진단

추정된 시계열 모형에 대한 적합도 검정은 잔차들의 백색잡음과정에 대한 검정으로 잔차들의 모든 표본자기상관계수(residual SACF)를 이용하는 포트맨토(Portmanteau) 검정과 Box와 Pierce(1970)의 Q^* 통계량 등을 이용한 검정 그리고 Ljung과 Box(1978)의 수정된 Q^* 통계량을 이용한 검정 등을 들 수 있다. R 패키지에서는 `tsdiag()` 함수를 사용하여 추정된 시계열 모형에 대한 적합도 검정이 가능하며, 이에 대한 수행은 R script 7과 같이 작성될 수 있으며, 이를 실행한 경우의 그래프 출력 결과는 Figure 3과 같다.

《R script 7》 모형의 진단 위한 R script와 출력 결과의 예

```
tsdiag(zt.fit, gof.lag=48)
```

SAS와 SPSS에서는 모형의 추정을 위한 명령문의 실행시에 포트맨토(Portmanteau)의 잔차에 대한 자기상관 검정(autocorrelation check of residuals)과 잔차진단이 포함 되어 출력되며, MINITAB에서는 모형 적합후의 잔차에 대하여 ACF와 PACF를 실행함으로써 잔차진단의 통계량인 수정된 Box-Pierce (Ljung-Box) χ^2 통계량을 구할 수 있다(Table 7).

Table 7. Model Diagnostics

Package	Model Diagnostics
R	tsdiag(zt.fit, gof.lag=48)
SAS	proc arima; identify var=lnzt(1, 4); estimate p=(4) q=(4);
SPSS	arima variables=zt /model=(0,1,0) (1,1,1) noconstant ln.
MINITAB	arima (0,1,0)*(1,1,1) 4 'lnzt' resid c6, acf 36 c6, pacf 36 c6

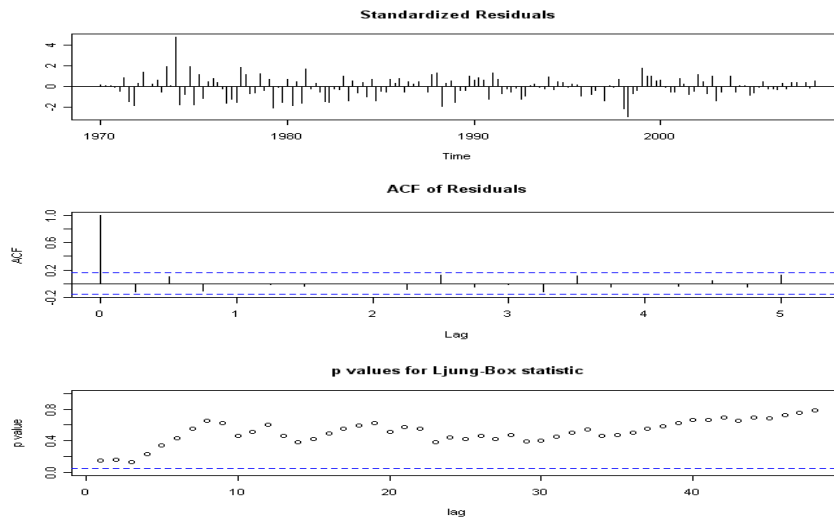


Figure 3. Diagnostic Display for the ARIMA(p,d,q) Model

2.8. 예측값 및 예측의 적합척도

시계열분석에서 예측은 예측오차(forecast error)가 작을수록 예측이 정확하게 이루어졌다고 볼 수 있다. 일반적으로는 바람직한 예측값을 구하는 방법은 여러 가지가 있으나 평균제곱오차(MSE)를 최소로 하는 값을 예측값으로 흔히 사용한다.

예측의 정도를 평가하는 척도는 평균제곱오차(MSE), 제곱근 평균제곱오차(RMSE), 퍼센트오차(PE), 평균 퍼센트오차(MPE), 평균 절대퍼센트오차(MAPE) 그리고 평균절대오차(MAE) 등이 있을 수 있으나, 일반적으로는 RMSE 또는 MAPE를 흔히 사용한다. 대수변환된 시계열 z_t 가 $ARIMA(0,1,0)(1,1,1)_4$ 로 식별된 경우 예측값 및 예측오차를 구하는 R 프로그램은 R Script 8과 같이 작성될 수 있으며, 예측의 적합척도는 각 척도의 정의에 따르는 간단한 연산에 의하여 쉽게 구할 수 있다.

《R script 8》 예측값과 예측오차를 구하는 R script

```
( lnzt.fit <- arima( lnzt, order=c(0,1,0), seasonal=list(order=c(1,1,1), period=4),
  include.mean = TRUE, method = c("CSS-ML", "ML", "CSS")) )
( lnzt.pr <- predict(lnzt.fit, n.ahead=8) )
( pred <- exp(lnzt.pr$pred) )
( se <- exp(lnzt.pr$se) )
```

SAS에서는 arima procedure 안에서 identify문과 estimate문이 선행된 후에 forecast 문을 사용하며, MINITAB에서는 arima 명령문의 부명령문으로서 forecast문을 사용한다. 특히 SPSS에서는 arima 명령문의 사용이전에 predict 문이 선행되어야 가능하다. 각 패키지별 예측값과 예측오차를 구하는 key word를 요약정리하면 Table 8과 같다.

Table 8. Predictions and Prediction Error

Package	Predictions and Prediction Error
R	predict(lnzt.fit, n.ahead=8)
SAS	proc arima; identify var=lnzt(1, 4); estimate p=(4) q=(4); forecast lead=8 out=lnzt_out
SPSS	predict thru year 2008 quarter 2. arima variables = zt /model = (0,1,0) (1,1,1) noconstant ln.
MINITAB	arima (1,1,1)*(0,1,1) 4 'lnzt'; forecast 12, save in c10 c11 c12.

3. R 패키지를 활용하는 사례연구

1970년 1사분기부터 2010년 2사분기까지의 우리나라의 국내총생산(GDP) 분기별 시계열자료(단위는 10억)를 2008년 2사분기까지의 시계열자료를 사용하여 승법계절 ARIMA 모형을 구축하고, 이를 이용하여 2008년 3사분기부터 2010년 2사분기의 예측값과 적합척도를 구하기 위한 R 프로그램은 R script 9와 같이 작성될 수 있다. 단, 국내총생산(GDP) 분기별 자료는 통계청 홈페이지에서 쉽게 구할 수 있다.

《R script 9》 GDP자료의 시계열분석을 위한 R script

```
# 원시 시계열 & 시계열 그래프
gdp <- scan("e:/gdp.txt")
(gdp.ts <- ts(gdp, start=c(1970, 1), frequency=4)); plot.ts(gdp.ts)
# 모형구축용 시계열
(z1 <- window(gdp.ts, c(1970, 1), c(2008,2)))
# 분산과 추세 및 계절추세의 안정화
lnzt <- log(z1); dif1.lnzt <- diff(lnzt); plot.ts(dif1.lnzt)
dif41.lnzt <- diff(dif1.lnzt, differences=4)
# 모형의 식별
acf(dif1.lnzt, 32); pacf(dif1.lnzt, 32)
acf(dif41.lnzt, 4, 32); pacf(dif41.lnzt, 4, 32)
# 모형의 적합과 결과 및 진단
(lnzt.fit <- arima(lnzt, order=c(0,1,0), seasonal=list(order=c(1,1,1), period=4)) )
tsdiag(lnzt.fit, gof.lag=32)
# 예측 및 예측의 적합척도
(lngdp.pr <- predict(lnzt.fit, n.ahead=8))
(pred <- exp(lngdp.pr$pred))
(p.err <- pred-actual); sse=p.err^2; mse=mean(sse); rmse=sqrt(mean(sse));
pe=(p.err/actual)*100 ; mpe=mean(pe); ape=(abs(pe)); mape=mean(ape)
cat("RMSE= ", rmse, "MAPE=", mape, "\n")
# 실제값 및 예측값 시계열 그래프
actual <- gdp[155:162]; quarter = 98:154 ; quarter1=155:162
```



```

legends <- c("o = predict value", " - = real value")
plot(quarter, zt[98:154], type="l", ylim=c(50000, 330000),
     xlim=c(98, 162), ylab="Gdp in Korea")
legend(100, 300000, legend=legends )
title(main="우리나라 분기별 GDP, 단위 10억, 1970.1-2010.2")
lines(quarter1, pred, col="black", type="o")
lines(quarter1, actual, col="black", lty="dashed")
abline(v=154.5, lty="dotted")

```

《R script 9》의 실행결과 가운데 핵심부분을 요약 정리하면 다음과 같다.

첫째: GDP자료에 적합된 승법계절 ARIMA모형

$$(1 - 0.1616B^4)(1 - B)(1 - B^4)\text{Ingdp} = (1 - 0.2750B^4)\epsilon_t, \quad (2.2)$$

$$a_t \sim iid WN(0, \sigma_a^2), \quad \sigma_a^2 = 74079.$$

둘째: 대수우도 및 AIC 식별 통계량

$$\log \text{likelihood} = 325.18, \text{aic} = -644.36 \quad (2.3)$$

셋째: 예측의 적합척도

$$RMSE = \sqrt{\frac{1}{8} \sum_{l=1}^8 \hat{e}_n(l)^2} = 17348.19, \quad (2.4.1)$$

$$MAPE = \frac{1}{8} |\hat{e}_n(l)/z_t(n+l)| \times 100 = 6.18247. \quad (2.4.2)$$

넷째: 2008년 3분기에서 2010년 2분기까지의 예측값과 예측오차

Table 9. GDP in Korea, 1970.1-2010.2, billion Won

분기별 (Time)	실제값 (Real)	예측값 (Predictions)	예측오차 (Prediction Error)
2008.3	262615	267851.9	5236.857
2008.4	261318	284168.9	22850.906
2009.1	237855	261019.6	23164.594
2009.2	264414	281701.9	17287.865
2009.3	275367	288497.2	13130.204
2009.4	285424	306337.2	20913.189
2010.1	265624	281250.8	15626.788
2010.2	290817	303803.9	12986.882

다섯째: 2008년 3분기에서 2010년 2분기까지의 실제값과 예측값의 시계열 그래프(Figure 4)

본 연구에서는 시계열자료에 대한 분석을 상용되는 패키지인 SAS, SPSS, MINITAB을 사용하지 않고, 무료 공개 소프트웨어인 R 패키지를 활용하여 ARIMA 모형 및 승법계절 ARIMA 모형에 대한 분석이 가능함을 사례연구를 통하여 분석 및 예측이 가능함을 보였다.

시계열자료에 대한 기타의 모형 특히 경제시계열모형에서 자주 언급되는 자기회귀 시차분포모

형(ADL 모형), 자기회귀 조건부 이분산 모형(ARCH 및 GARCH 모형), 오차수정모형 등과 시계열 분석에서 다루는 기타의 분야에도 R 패키지의 활용이 더욱 연구되어야 할 과제로 사료된다.

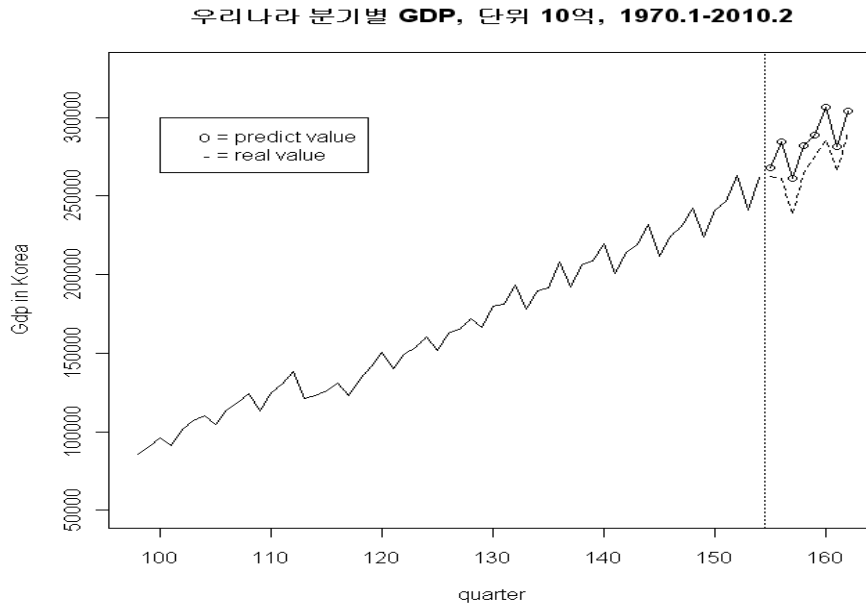


Figure 4. GDP in Korea, 1970.1-2010.2, Unit: a billion Won

참고문헌

- 김수화, 김승희, 조신섭 (1994). 통계패키지에서의 시계열 분석방법의 비교연구, *한국통계학회논문집*, 1(1), 119-130.
- 김연형, 정영숙 (2006). 주택가격과 거시경제 변수간의 시계열분석, *Journal of the Korean Data Analysis Society*, 8(6), 2383-2398.
- 김희재 (2003). 샘플링 검사에 대한 R시스템의 활용에 관한 고찰, *Journal of the Korean Data Analysis Society*, 5(3), 601-616.
- 김희재 (2005). R을 이용한 Weibull 수명분포의 모수추정, 경성대학교기초과학 연구소, *기초과학연구논문집*, 17, 21-30.
- 엄철준, 오갑진, 박종원 (2007). 금융시계열의 상관행렬 속성분해를 위한 새로운 방법연구, *Journal of the Korean Data Analysis Society*, 9(4), 1833-1848.
- 전세봄, 최보승, 김성용 (2009). 스펙트럼 분석을 이용한 시계열 자료의 패턴 분류, *Journal of the Korean Data Analysis Society*, 11(1B), 349-359.
- Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelation on auto-regressive integrated moving average time series models, *Journal of the American Statistical Associations*, 65, 1509-1526.
- Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika*, 65, 297-303.

The Application of Time Series Analysis under R Environment^{*}

*Hee Jae Kim*¹

Abstract

The statistical package SAS, SPSS and MINITAB are commonly used on time series analysis and prediction but they are very expensive. In this study, the multivariate seasonal ARIMA model is fitted and predict values are obtained for the Gross Domestic Product (GDP) of South Korea from 1970 Q2 to 2010 Q2 by taking advantage of the free R package. In addition, the aim of this study is to broaden the choice of packages and to contribute to the research and development of time series analysis by providing a comparison chart of key word and option according to analysis module among the above packages.

Keywords : time series analysis, R package, ARIMA model, multivariate seasonal ARIMA model, GDP of South Korea.

^{*}This research was supported by 2010 Research Grant of Kyungsung University.

¹Professor, Department of Informational Statistics, Kyungsung University, Busan 608-736, Korea.
E-mail : hjkim@ks.ac.kr

[Received 23 December 2010; Revised 23 January 2011; Accepted 26 January 2011]