

## 졸업논문

제 목 : 시계열분석을 통한 환율과 해외여행 수요 간의 관계 분석

지도교수 : 강 기 훈

2012 년      10 월      30 일

한 국 외 국 어 대 학 교                      자 연 과 학 대 학                      통 계 학 과

학 번 : 2 0 0 7 0 2 5 8 7                      성 명 : 이 종 훈 ㉠

학 번 : 2 0 0 9 0 2 6 7 1                      성 명 : 이 정 민 ㉠

학 번 : 2 0 0 9 0 2 9 0 3                      성 명 : 임 슬 기 ㉠

시계열분석을 통한  
환율과 해외여행 수요 간의 관계 분석

by

Jong-hoon Lee

Jung-min Lee

Seul-ki Lim

Under the direction

of

Professor Ki-hoon Kang

2012. 10. 30

A thesis submitted and approved by the committee of the Department of  
Statistics of Hankuk University of Foreign Studies in partial fulfillment of  
the requirements for the degree of Bachelor

Thesis Committee : \_\_\_\_\_

DEPARTMENT OF STATISTICS

## < 목 차 >

### 제1장 서 론

1.1 연구의 배경 .....	5
1.2 연구의 목적 .....	5

### 제2장 시계열 분석

2.1 시계열 .....	6
2.2 시계열 자료 분석 순서 .....	10
2.3 시계열 분석의 기본 개념 .....	12

### 제3장 모형의 적합

3.1 자료의 개요 .....	13
3.2 시계열자료의 정상성 판단 .....	16
3.3 환율 자료 .....	17
(1) 변환 및 차분 .....	17
(2) 모형 식별 .....	19
(3) 모수 추정 .....	21
(4) 모형의 적합성 검정 .....	22
3.4 내국인 출국자 수 자료 .....	23
(1) 변환 및 차분 .....	23
(2) 모형 식별 .....	25
(3) 모수 추정 .....	26
(4) 모형의 적합성 검정 .....	27

### 제4장 예측

4.1 환율 예측 .....	29
4.2 내국인 출국자 수 예측 .....	30

### 제5장 상관분석

5.1 환율과 내국인 출국자 수 자료의 상관분석 .....	31
----------------------------------	----

### 제6장 전이함수모형

6.1 전이함수 .....	33
(1) 전이함수와 충격반응함수 .....	33

(2) 전이함수의 계수와 모수 .....	34
6.2 전이함수모형의 설정 .....	35
(1) 사전백색화 .....	35
(2) 잡음모형의 설정 .....	35
(3) 최종 전이함수의 결정 .....	36
6.3 전이함수모형의 진단 .....	37
6.4 전이함수모형의 예측 .....	37
※ 참고문헌 .....	38
※ 전이함수모형의 예측 .....	39

# 1. 서 론

## 1.1 연구의 배경

최근 들어 우리나라 안에서 가까운 나라 일본, 중국 등을 비롯하여 멀리 미국, 유럽 등 다양한 나라로의 해외여행이 증가하고 있기는 하다. 외국과의 교류가 활발하게 이루어지고 점점 더 개방 되고 있어 내국인의 해외 진출 및 여행규모는 과거 대비 증가 추세가 나타나고 있다. 최근 체감 적으로 내국인의 해외여행 수요가 증가하고 있다고 느끼는 것이 일반적이다. 이러한 현상에 대해 환율 상승과 하락이 관광 수지에 얼마나 영향을 미치고, 과연 환율과 해외여행간의 뚜렷한 상관관계가 있는지에 대해 알아보고자 한다. 만일 환율과 해외여행 사이에 상관관계가 존재한다면 환율은 관광수지, 관광객의 입·출국을 결정하는 주요 변수로 인식되면서 환율 변동이 해외여행 전반에 미치는 효과에 관한 의미가 크다고 할 수 있다.

이러한 연구과정을 통해 본 논문은 첫째, 환율 변동과 내국인 해외여행자 수의 증감 각각의 자료를 이용해 지난 12년간 어떤 추세를 보였는지, 두 자료의 시계열 모형을 확인 할 수 있다. 둘째, 전이함수모형을 적합 시켜 두 항목 간의 관계를 파악하여 환율 변동이 내국인 해외여행자 수의 증감에 영향을 미칠 것이라는 추론을 할 수 있다.

## 1.2 연구 목적

우리나라는 외국과의 교류가 활발하게 이루어지고 점점 더 개방 되고 있어 내국인의 해외 진출 및 여행규모가 확대되는 동시에 해외방문객의 국내유입 증가추세가 뚜렷이 나타나고 있다. 환율은 관광수지, 국내외 관광객의 입·출국을 결정하는 주요 변수로 인식되면서 환율 변동이 해외여행 전반에 미치는 효과에 관한 의미가 크다고 할 수 있다. 하지만 최근 체감 적으로 내국인의 해외여행 수요가 증가하고 있다고 느끼는 것이 일반적이다. 이러한 현상에 대해 환율 상승과 하락이 해외여행 수요에 얼마나 영향을 미치고, 과연 환율과 해외여행간의 뚜렷한 상관관계가 있는지에 대해 알아보고자 한다.

## 2. 시계열 분석

### 2.1 시계열

시계열이란 시간의 흐름에 따라 일정한 간격으로 관측하여 기록된 자료를 말한다. 예를 들면 증권시장의 종합주가지수는 거래가 이루어질 때마다 변하는데, 매일 증권 거래 마감 시간의 종합주가지수를 일정기간 동안 관측하여 기록하면 이것이 바로 일별 종합주가지수 시계열이 된다. 이 외에도 특정 소비자의 월별 판매량 혹은 연도별 농작물의 생산량 등을 관측하여 기록한 자료 역시 시계열이다. 시계열 분석(time series analysis)이라고 하는 것은 이런 시계열을 해석하고 이해하는 데 쓰이는 여러 가지 방법을 연구하는 분야이다. 예컨대, 이런 시계열이 어떤 법칙에서 생성되어서 나오느냐는 기본적인 질문을 이해하는 것이 궁극적인 목표라고 할 수 있다. 시계열 분석이란 관측된 과거의 자료를 분석하여 법칙성을 발견하고, 이를 모형화하여 추정하는 것을 뜻한다. 시계열 예측(time series prediction)이라고 하는 것은 주어진 시계열을 보고 수학적인 모델을 만들어서 미래에 일어날 것들을 예측하는 것을 뜻하는 말이다.

#### ▶시계열 분석의 목적

##### 1) 자료의 기술

자료를 시간의 흐름에 따라 그림으로 표현하고, 자료의 특성을 파악하여 법칙성을 발견한다. 예를 들면 대체적인 추세, 계절성, 특이점, 변화점등을 탐지한다.

##### 2) 예측(prediction/forecasting)

시계열 분석의 가장 중요한 목적은 예측이다. 만약 어느 증권투자자가 주가를 성공적으로 예측할 수 있다면 주가의 매매시기를 잘 선택하여 성공적인 증권투자를 할 수 있을 것이다. 백화점의 경우 판매량을 어느 정도 정확히 예측할 수 있다면 물품 및 자금의 유통을 원활히 관리할 수 있으므로 합리적인 경영을 할 수 있을 것이다.

##### 3) 설명(explanation)

어떤 시스템의 시계열적 형태를 설명하기 위한 과학적 노력을 말한다. 예를 들면 경제시계열의 분석을 통하여 경기주기(business cycle)가 존재한다는 것을 입증한다든지, 어느 제조 공정에서 입력과 출력 사이에는 어떠한 함수관계가 있는지를 알아내어 시스템 안에 존재하는 공학적 원리에 대한 이해를 하고자 하는 경우이다.

#### 4) 제어(control)

미래에 대한 예측이 가능해지면 그 시스템에 어떤 인위적인 조작을 가함으로써 시계열 실현 값이 우리가 원하는 목표에 부응하도록 유도할 수 있다. 예를 들어 경기의 지나친 활성화나 침체화가 우려되면 정부가 중앙은행을 통하여 경제에 정책적으로 개입하여 안정시키려는 노력을 할 수 있다. 그 결과로 최근에는 과거에 비해 경기 주기 폭이 상대적으로 작아졌다는 이야기가 있다. 또 다른 대표적인 예는 통신공학 및 자동제어, 화학공학의 많은 공학적 응용의 경우가 있다. 공학 시스템의 출력을 목표 범위 내로 유지하기 위하여 입력과 출력 사이의 관계를 이용하여 입력 변수를 조작하면 출력 값을 제어할 수 있다. 이 중에서 가장 중요한 것은 자료의 기술 및 예측이다. 거의 대부분의 자료에서 80% 정도의 정보는 기술통계를 통하여 얻을 수 있다. 정확한 예측을 하기 위해서는 우선 좋은 자료가 필요하고 객관적이고 과학적인 예측기법과 평가가 필요하다.

#### ▶시계열의 변동요인

시계열 자료의 구성요소는 일반적으로 추세변동, 순환변동, 계절변동, 불규칙변동의 네 가지 요인으로 구성되어 있다.

##### 1) 추세변동(장기변동)

추세변동(trend movement)은 시계열이 장기간에 걸쳐 점진적으로 지속하는 상승적인 또는 하강적인 변화 상태를 나타내는 경향을 뜻한다. 기본적인 변동경향이므로 시계열의 단기 변동의 중간을 따라 진행되는 현상은 경향변동을 포함하고 있으므로 이 경향변동을 측정할 필요가 있다. 이러한 측정을 한다는 것은 통계학적으로 말하면 시계열에 적합 하는 경향선(trend line) 또는 추세선을 긋는 것이다. 경향선은 시간적으로 배열된 경과도표상에서 각 점을 가장 가깝게 통과하는 선을 긋는 것이며 이 선은 직선으로 될 때도 있고 또는 곡선으로 되는 경우도 있다. 이것이 통계학에서 말하는 소위 선 맞추기(curve fitting)의 문제가 된다. 이와 같이 시계열의 경향변동을 선 맞추기로 만들면 장기간에 걸쳐 형성된 현상이 장래에 어떠한 경향으로 발전할 것인가 하는 예측을 할 수 있게 된다. 경향변동의 계산 또는 경향선을 결정하는 방법으로서 목측법, 최소제곱법 및 이동 평균법이 있다. 목측법(eye-measurement)은 시계열의 경과도표에서 각 점을 가장 가깝게 통과하는 소위 평균선으로 생각되는 직선 또는 곡선을 눈짐작으로 그리는 방법이다. 이 방법은 어디까지나 주관적인 판단에 의하여 그려지므로 불합리한 것이다. 이것을 더욱 정확하게 그리는 방법은 최소제곱법을 이용하는 것이다. 최소제곱법에 의해서 시계열의 경향변동을 계산하는 경우에는 먼저 그 장기변동의 형태에 적당한 경향선의 형을 선정하지 않으면 안 된다. 그러므로 시

계열을 그림으로 표현하여 목적에 의해서 될 수 있는 한 간단한 곡선을 그려보는 것이 필요하다.

## 2) 순환변동(경기변동)

순환변동(cyclical movement)은 주기가 일정하지 않는 상하 반복하는 운동으로 1순환을 보통 3년에서 10년으로 생각하는 것이다. 경제이론에서는 경기변동에 해당한다. 주기가 일정하지 않은 순환변동 (cyclical movement)은 경향변동이나 계절변동과는 달리 시계열에서 이것을 직접 검출하는 적당한 방법이 없다. 그러므로 일반적으로 주어진 시계열의 변화에서 경향변동과 계절변동을 제거하여 얻어지는 잔차계열을 순환변동으로 생각한다. 그러나 이 잔차계열에는 불규칙변동요인이 남아 있으므로 정확히 말하면 이것은 순환불규칙변동의 계열이다.

## 3) 계절변동

계절변동(seasonal movement)은 일정한 주기를 갖는 상하 반복의 규칙적인 변동으로 1년을 변동의 주기로 한다. 분기별 혹은 월별 자료에서 나타난다. 예를 들면, 여름에 맥주소비량이 많은 것은 1년을 1기로 하는 규칙적인 변동이며, 매주 일요일에 영화관이 만원이 되는 것은 1주를 1기로 하는 규칙적인 변동이다. 이와 같은 것에 의하여 1년 주기, 1개월 주기, 매주의 주기, 1일의 주기가 있다. 그러나 일반적으로 계절변동은 년 내의 각월에 생기는 계절적·주기적인 변동을 의미한다. 계절변동을 시계열 변동에서 분리하여 표시하는 방법에 다음 2가지 방법이 있다.

### ① 계절고정값

1년 각 월의 고정적인 변동을 시계열과 같은 단위의 숫자로 표시한 것을 말한다. 경향변동이나 부조적 변동을 상반하지 않는 계열의 경우에만 사용된다. 예컨대 기상 통계의 각월의 평균기온, 강우량 등의 평균값의 계열 등이 이에 해당된다.

### ② 계절지수(seasonal index)

각 월의 계절변동이 년 평균을 100으로 하는 지수의 형식으로 표시되는 것을 말한다.

계절변동은 그 자체를 분석의 목적으로 계산하는 경우도 있으나 계절변동의 주된 목적은 시계열을 계절변동에 관해서 수정하고 계절변동의 영향을 소거한 계열로서 표시하는데 있다.

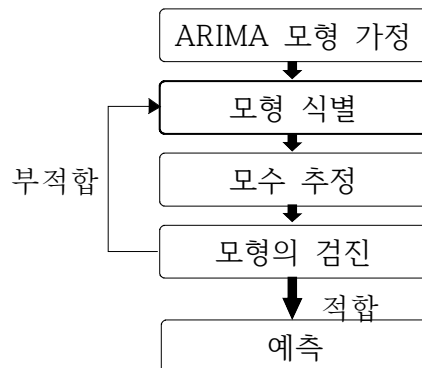


#### 4) 불규칙변동(잔차변동)

불규칙변동(irregular movement)은 위의 세 가지 변동 이외의 나머지 변동부분으로 우연적 혹은 불규칙적으로 나타나는 변동이다.

## 2.2 시계열 자료 분석 순서

일반적인 시계열 분석은 다음과 같은 그림으로 나타낼 수 있다.



ARIMA 모형을 가정하기 위해 가장 먼저 시계열 그림 고찰을 해야 한다. 시계열 그림을 고찰하는 순서는 ① 자료의 이상점(outlier)과 개입(intervention)을 살핀다. ② 정상성 만족의 여부와 계절성을 살핀다. ③ 개략적인 추세의 유무를 판단한다. ④ 자료의 분산이 일정한지 아닌지를 판단한다. 어느 주식의 시계열 자료에서 주가가 폭락한 시점이 있을 때 주가가 폭락한 이유를 아는 경우에는 개입이라고 하고 이유를 모를 경우에는 이상점이라 한다. 이상점으로 판단될 경우에는 자료를 제거하고, 개입으로 판단될 경우에는 회귀분석을 한다. 정상 시계열은 어떤 시계열 자료의 평균이 있고 기준을 만족하는 범위 내에서 자료가 존재할 경우를 정상 시계열이라 하고 기준을 벗어나면 비정상 시계열이라 한다. 예를 들어 주식, 환율과 같은 자료가 비정상 시계열이다. 또, 추세가 있거나 자료의 분산이 일정하지 않은 경우 비정상 시계열이다. 얼마나 자세히 시계열 자료의 그림을 살펴보느냐가 시계열 분석을 얼마나 정확하게 하느냐에 영향을 미치게 된다.

시계열 그림 고찰 후에는 비정상 시계열을 변환(Transformation) 및 차분(Difference)을 해서 정상 시계열로 바꾼 후 분석을 한다. 분산이 일정하지 않은 경우 변환을 하는데 로그변환, Box-Cox변환 등을 하고 추세가 있거나 가상의 평균을 지나는 횟수가 매우 작을 경우 차분을 하는데 이 과정은 Dickey-Fuller Test로 검정을 한다.

시계열 그림 고찰 과 변환 및 차분 과정은 시계열 분석의 전처리 과정이다. 전처리 과정이 끝나면 Box-Jenkinson의 분석 3단계를 해야 한다. 첫 번째로 모형 식별(Model identification)을 한다. 이는 주어진 시계열 자료가 어떤 확률 모형에서 나왔는지를 식별하는 단계이다. 주로 사용되는 통계량은 자기상관함수(Autocorrelation function, ACF)와 부분자기상관함수(Partial autocorrelation function, PACF)이다. 자료로부터 표본 자기상관함수와 표본 부분자기상관함수를 추정하고, 추정된 자기상관함수와 표본 부분자기상관함수를 이론적인 ACF와 PACF를 비교함으로써 모형 식별을 하게 된다.

두 번째로는 모수 추정을 한다. 모형 속에 포함된 모수를 추정하는 단계인데 추정방법으

로는 최소제곱추정법(Least Squares Estimation), 최대우도추정법(Maximum Likelihood Estimation), 적률추정법(Method of Moment Estimation) 등이 있다. 자료의 수가 많은 경우 이론적으로 어떤 추정방법을 사용하여도 같은 결론을 얻게 되는 것으로 알려져 있다. 분석에서는 최대우도추정법과 최소제곱추정법이 사용되고 이론전개에서는 적률추정법이 주로 사용된다.

세 번째로 모형의 검진을 한다. 이는 모형이 잘 적합 되었는지를 살펴보는 과정으로 먼저 잔차의 자기상관함수(ACF)와 부분자기상관함수(PACF)를 구하여 일차적인 판단을 내린 후 수정된 Q-통계량 또는  $\chi^2$ -통계량이라 부르기도 하는 퍼트멘토우 통계량에 의한 유의확률의 값이 0.05보다 크면 클수록 모형이 잘 적합 되었음을 말해준다.

Box-Jenkinson의 분석 3단계가 끝나면 예측(Forecast)를 한다. 예측은 시계열 분석에서 가장 중요하다. 모형을 찾아낸 후 주어진 모형을 바탕으로 예측을 하는데 최소평균제곱오차 예측(Minimum mean square error forecast)을 이용하여 예측오차의 분산을 최소로 해 주는 예측 값으로 예측을 하는 방법이다. 자료의 수가 t개가 있고 1시점까지 예측한다고 가정하고, t+1 시점의 예측값을  $\hat{Z}_{t+1}$  이라 할 때 최소평균제곱오차예측은 다음과 같다.

$$\hat{Z}_{t+1} = E[Z_{t+1}|Z_t, Z_{t-1}, \dots, Z_1]$$

ARIMA(Auotoregressive Integrated Moving-Average) 모형은 추세가 있는 경우(계절성 포함) 이를 제거하여 정상적(stationary) 데이터로 변환한 후, AR, MA, ARMA 모형중에서 가장 적합한 모형을 선택할 수 있게 해준다. 적절한 모형의 선택 즉 AR, MA, ARMA 모형중에서 가장 적합한 모형의 선택은 주로 자기상관계수와 부분자기상관계수를 분석하여 이루어진다. 적합성 검정은 기본적으로 잔차분석에 의하여 잔차가 패턴을 가지고 있지 않음이 판단되면 Overfitting을 이용하여 AIC, SBC를 비교하여 더욱 적합한 모형을 찾아낸다.

#### ▶ ARIMA 모형의 일반식.

$$ARIMA(p,d,q) \times (P,D,Q)_L$$

$$= \Phi_p(B)\phi_p(B)(1-B^L)^p(1-B)dZ_t = \theta_0 + \Theta_Q(B^L)\theta_q(B)a_t$$

$$\Phi_p(B^L) = (1 - \Phi_1 B^L - \dots - \Phi_p B^{LP}) \quad : \text{Seasonal AR operator}$$

$$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p) \quad : \text{AR operator}$$

$$\Theta_Q(B) = (1 - \Theta_1 B^L - \dots - \Theta_Q B^{L-Q}) \quad : \text{Seasonal MA operator}$$

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q) \quad : \text{MA operator}$$

$$a_t \quad : \text{White noise process}$$

모든 형태의 단변량 시계열을 AR, MA, ARMA, ARIMA, 계절ARIMA 모형에 적합하여 효과적으로 시계열을 분석한다.

## 2.3 시계열 분석의 기본 개념

### ▶ 정상성(Stationarity)

정상성의 의미는 자료가 시점에 상관없이 시계열의 특성이 일정하다는 것이다. 정상성의 정의는 평균이 일정하고 분산이 일정하며 공분산  $Cov(Z_t, Z_s)$ 는 단지 시차(lag)에만 의존한다는 것이다. 정상성을 만족하는지 만족하지 않는지는 정상성 조건 세 가지를 살펴보아야 한다. 즉 평균, 분산 그리고 자기공분산을 구하였을 때 그 결과에 시점을 나타내는 첨자 't'가 나타나지 않으면 주어진 시계열은 정상성을 만족하게 된다.

### ▶ 시계열의 예

백색잡음과정(White noise process)는 대표적인 정상 시계열이며 백색잡음과정  $\{a_t\}$ 는 독립이고 같은 분포를 하며 평균이 '0'이고 분산이  $\sigma_a^2$ 인 확률 변수이다. 자기공분산은 다음과 같다.

$$Cov(a_t, a_{t-k}) = \begin{cases} \sigma_a^2, k=0 \\ 0, k \geq 1 \end{cases}$$

확률보행과정(Random walk process)는 대표적인 비정상 시계열이며 과거의 오차가 현재, 미래에도 영향을 준다. 확률보행과정의 모형은

$$Z_t = Z_{t-1} + a_t$$

이며, 여기서  $a_t$ 는 백색잡음이다.

### 3. 모형의 적합

#### 3.1 자료의 개요

환율과 내국인 해외여행 수요 간의 변동의 관계를 분석하기 위해서 e-나라지표 사이트에 있는 해외여행자수 지표와 환율 지표를 사용하였다. e-나라지표의 소개에 의하면 e-나라지표는 국정통계정보시스템으로서 국정 전 분야에 걸친 각종 국정통계에 대해 정확한 통계와 추이, 생생한 통계의미 분석 등을 제공함으로써 국민, 전문가, 정책결정자들 모두에게 국정을 알기 쉽게 설명하고, 모니터링 할 수 있도록 할 뿐만 아니라 필요한 국정통계정보를 손쉽게 접근할 수 있다. 지표탐색을 이용하여 2000년~2011년 환율과 내국인 출국자 수에 대한 지표를 얻었다.

먼저 2000년부터 2011년 까지 월별 환율 자료 중 일부인 2000년 1월부터 12월까지의 환율 자료를 <표 3.1>에 제시하였다.

<표 3.1> 2000년 1월~12월 환율 자료

	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000
	01월	02월	03월	04월	05월	06월	07월	08월	09월	10월	11월	12월
원/ 환율 <sup>1)</sup>	1,123	1,131	1,106	1,109	1,129	1,115	1,116	1,108	1,115	1,139	1,214	1,264
달	.20	.00	.00	.10	.40	.00	.70	.80	.10	.00	.30	.50
러												
절상률 2)(%)	1.3	0.6	2.9	2.6	0.8	2.1	1.9	2.6	2.1	-0.1	-6.3	-10

출처 : 기획재정부 (통화별 환율 조사통계)

다음의 <표 3.2>는 2000년부터 2011년까지의 내국인 출국자 수<sup>3)</sup>의 자료 중 2000년 1월부터 12월까지의 출국자 수이다.

<표 3.2> 2000년 1월~12월 내국인 출국자 수 자료

	계	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
2000년	5,508	454	397	398	407	458	453	589	571	430	447	440	463

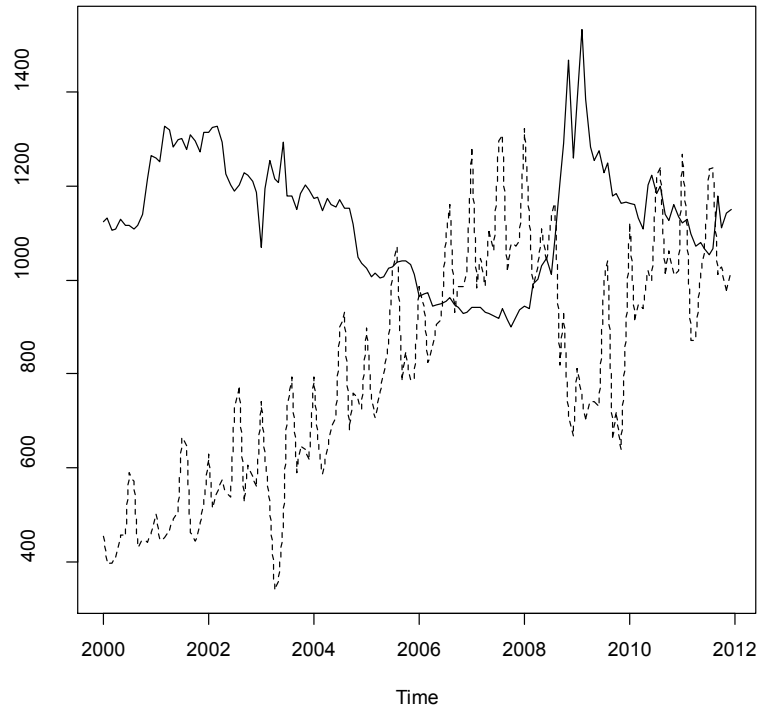
출처 : 한국관광공사 「내국인 출국」

1) 환율은 서울외환시장 종가(15:00현재, '05.2월 이전은 16:00현재, '03년 이전은 16:30현재).

2) 절상율은 전년말비 변동율.

3) 내국인 출국자 수는 총 내국인 출국자에서 WTO(세계관광기구) 기준에 의거하여 재외국민 부분을 제외하고 별도 집계되는 승무원을 산입한 수치이다.

2000년부터 2011년까지의 환율과 내국인 출국자 수 자료를 이용하여 분석에 사용하려고 한다. 먼저 12년간의 환율과 내국인 출국자 수의 수치를 <그림 3.1>에 그래프로 그려 보았다.



<그림 3.1> 환율과 내국인 출국자 수 시계열자료의 그래프

실선은 환율을 나타낸 것이고, 점선은 내국인 출국자 수를 의미한다. 그래프를 보면 전반적으로 환율이 하락하면 내국인 출국자수는 증가하며, 환율이 상승하면 내국인 출국자수는 하락함을 알 수 있다. 외국이 우리나라 환율에 미치는 영향은 대부분 외국경제문제, 전쟁, 정치, 투기자본, 기상이변으로 곡물가 변동, 국제정치문제로 유가 파동 등으로 인해 영향을 받는다. 그리고 우리나라가 해외의존도가 상당히 높기 때문에 시시각각 영향을 받고 있는 상황이다.

기획재정부의 지표 해석을 참고하면 2000년 IT버블 붕괴 사건과 2001년 911 테러 사건으로 큰 상승이 있었고, 1999~2001년 보합권 유지로 원/달러 환율이 등락을 거듭하다가 2002년부터 점진적으로 하락하였으며 글로벌 달러약세가 전개되고, 외환위기를 완전히 극복하는 과정에서 수출호조, 외국인자금유입 등으로 달러공급이 많아지면서 추세적으로 하락하는 경향이 있다.

2006~2008년 초반까지 미국 금융파동과 한국외환보유고 바닥으로 스와프정책추진으로 환율에 영향을 받았고, 미국에서 서브프라임모기지 문제가 불거지기 시작해서 파장을 일으켰다.

2007년 환율이 전년말대비 6.3원 소폭 상승하였는데 그동안 과도한 원화절상에 대한 반작용, 미국 서브프라임 위험에 따른 글로벌 안전자산 선호 현상 등으로 상승 반전을 의미한다.

2008년 초반까지는 리만브라더스가 파산에 이르는 글로벌 금융위기로 인해 국제금융시장에 불확실성 증가 및 신용위험 증가로 국제금융시장에 유동성이 부족하여 안전자산 선호 현상에 따른 달러값이 폭등하고 원화는 900원대에 이른다. 그 후 경상수지 적자 전환, 외국인 주식 순매도 및 글로벌 신용경색에 따른 세계경제 침체 등으로 전년말대비 323.4원 큰 폭으로 상승하였다.

2009년 3월 이후 글로벌 금융위기가 해소 국면을 보이면서 경상수지 흑자, 외국인 주식 순매수 등으로 전년 대비 하락하였다.

2010년 이란제제정책 동참으로 이란 진출기업 철수, 계약해지 등으로 환율에 영향을 주었고, 천안함, 연평도 사태 등으로 일시적으로 급등세를 보였으나, 양호한 경제 펀더멘탈 등을 반영하여 전년 대비 하락하였다.

### 3.2 시계열 자료의 정상성 판단

정상시계열의 성질은 평균과 분산이 시간의 변화에도 안정되고, 이웃하는 확률변수 사이의 공분산이 시차에만 의존하는 것을 말한다.

시계열의 분산이 시점에 따라 변하면 로그변환이나 제곱근변환 등과 같은 변환을 통하여 분산을 안정시킨다. 「경제 및 금융자료를 위한 시계열분석. 김해경, 이명숙 저」를 참고하면 분산은 안정되어 있지만 평균이 안정되지 않은 평균비정상계열에 대해서는 차분을 통하여 평균을 정상화한다고 제시되어 있다. 정상시계열인지 비정상시계열인지는 ACF를 통해 판단할 수 있다. 정상시계열은 ACF가 빠르게 0으로 감소하며 뚜렷한 추세를 보이지 않는다. 비정상시계열은 ACF가 천천히 감소하며 뚜렷한 직선추세 즉, 추세요인 또는 계절요인이 있는 시계열이 대표적이다. 환율 자료에 대한 ACF는 천천히 감소하는 직선의 추세를 보이고 있으므로 비정상시계열이라 할 수 있다. 이러한 비정상시계열을 정상시계열로 바꾸는 방법을 차분이라고 한다.

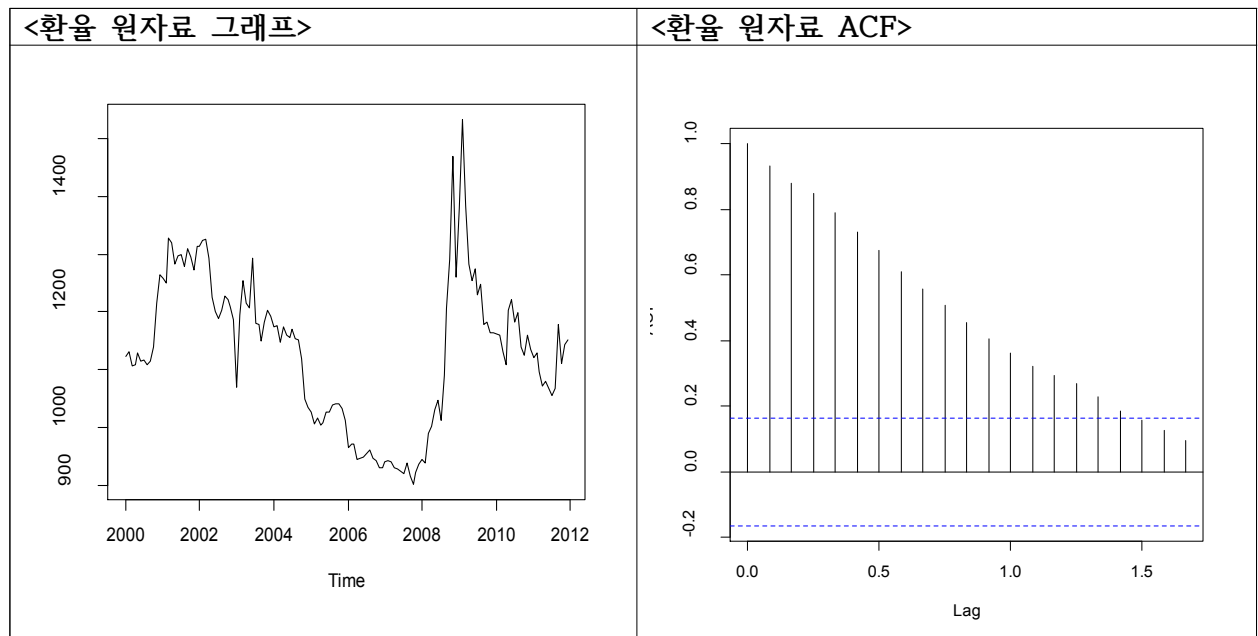
로그차분이란 원래의 자료에 로그변환을 취한 것의 차분으로 원 자료의 변화율을 나타낸다. 로그변환을 취한 자료에 각각 100을 곱하여 차분을 구하면 원 자료의 백분율 변화를 구할 수 있다. 로그변환을 취한 자료의 차분이 원 자료의 %변화율과 거의 일치한다는 면에서 로그변환의 유용성을 잘 활용할 수 있다. 특히 인구, GNP, 환율 등 경제 또는 금융시계열 자료는 대부분의 경우 시간에 대한 지수함수 형태의 추세를 지닌다. 이러한 자료에 자연로그 변환을 취하면 시간에 대해 지수함수 형태인 추세를 선형함수 형태의 추세로 변환이 가능하다.



### 3.3 환율자료

#### 3.3.1 변환 및 차분

환율 자료의 그래프와 ACF를 그려 보았다.



<그림 3.2> 환율 원자료 그래프와 ACF

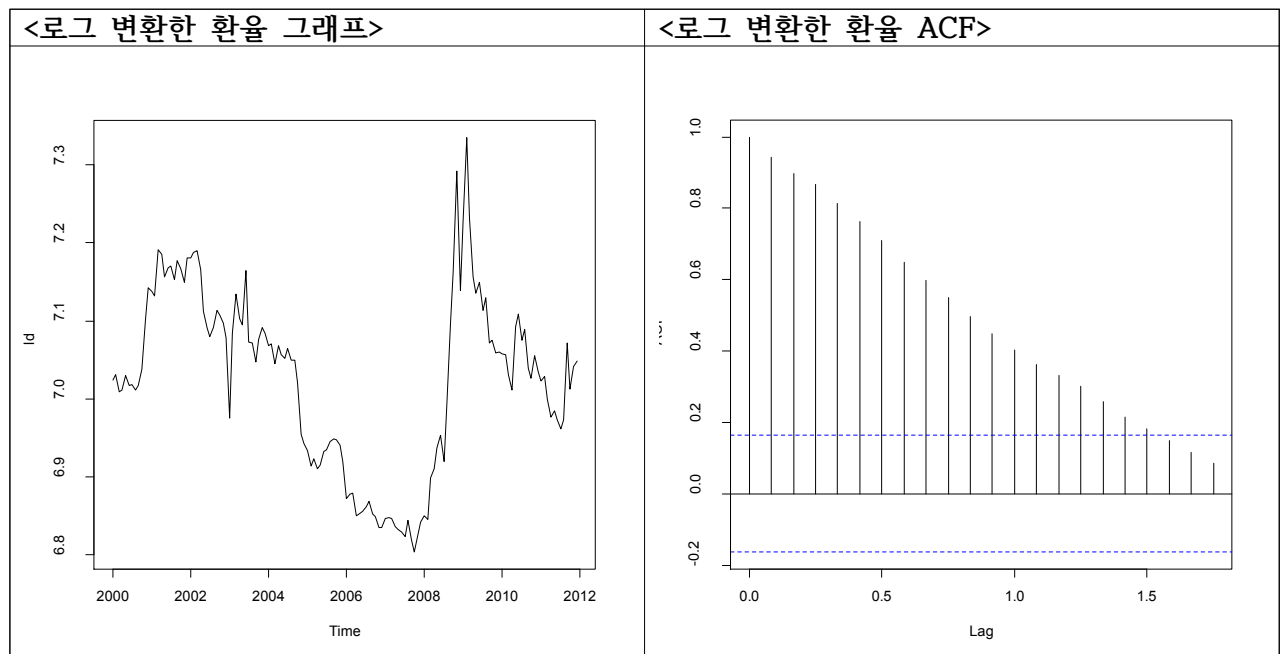
<그림 3.2>에서 볼 수 있듯이 평균과 분산이 일정하지 않고 ACF가 0으로 가까워지는 속도가 매우 느린 걸로 보아 환율 자료는 비정상 시계열 자료임을 예상 할 수 있다. 확실한 검증을 위해 정상인지 비정상인지 검정하는 Dickey-Fuller Test를 시행해 보고 <표 3.3>에 제시하였다. DF-test의 귀무가설이 비정상 시계열이고 대립가설이 정상 시계열, p-value가 유의수준 0.05보다 크면 귀무가설을 기각 할 수 없으므로 자료가 비정상 시계열이라고 할 수 있다.

<표 3.3> 환율 시계열자료에 대한 Dickey-Fuller Test

Augmented Dickey-Fuller Test		
data : ER		
Dickey-Fuller=-2.2724	Lag order=5	p-value=0.463
alternative hypothesis : stationary		

환율 자료를 DF-test 한 결과 p-value가 0.463으로 0.05보다 크므로 비정상성 시계열임을 알 수 있다.

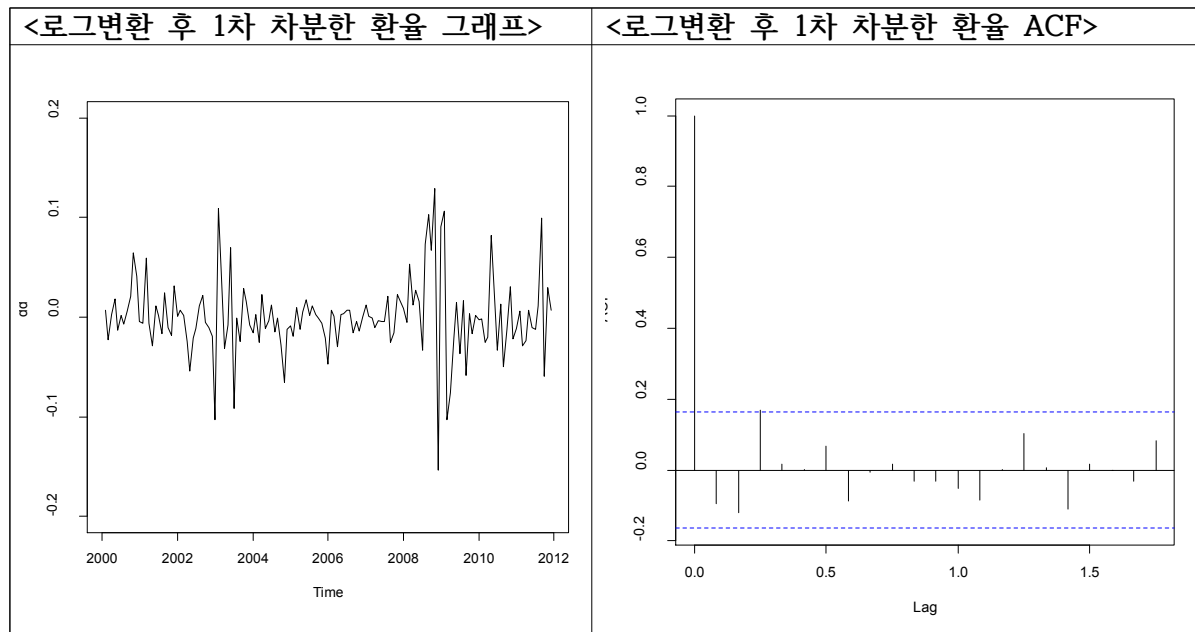
비정상 시계열을 정상 시계열로 만들기 위해 먼저 비정상 시계열의 분산안정화를 위해 로그변환을 해 보았다. 분산의 변동이 큰 경우는 로그변환이 제곱근변환보다 효과적이므로 로그변환을 주었다.



<그림 3.3> 로그 변환한 환율 그래프와 ACF

<그림 3.3>을 보면 로그변환 한 시계열 자료의 ACF는 시차가 커질수록 서서히 감소하는 형태를 띠기 때문에 이동평균모형으로 예상되며 비정상시계열을 정상시계열로 변환하기 위해서 일반차분을 한다.

다음의 <그림 3.4>는 1차 차분한 환율 그래프와 ACF이다.



**<그림 3.4> 로그변환 후 차분한 환율 그래프와 ACF**

1차 차분을 한 결과 원 자료에 비해 평균이 일정해 진 것을 볼 수 있고, 1차 차분한 자료의 ACF를 보면 0으로 빠르게 감소하는 형태를 나타내므로 정상 시계열이 된다. 따라서 환율 자료를 1차 차분 및 로그변환을 하여 비정상성을 없애 변환된 자료로 분석에 이용하겠다.

### 3.3.2 모형 식별

주어진 시계열 자료가 어떤 확률 모형에서 나왔는지 식별하는 단계이다. 시계열 분석에 사용되는 확률모형은 여러 가지가 있으며 각 모형마다 각기 다른 특성을 가지고 있다. 자료를 그림으로 나타냈을 때 그 특성이 바로 나타나지 않기 때문에 모형 식별을 위하여 자료로부터 모형을 식별할 수 있는 여러 통계량 들을 구해야 하며 이들 통계량들을 이용하여 모형식별이 가능하게 된다. 제일 적합한 모형을 찾기 위해서 여러 추정 값들을 사용해 보도록 하겠다.

먼저 환율 자료를 통계프로그램 R을 이용하여 나올 수 있는 모형들에서 각각의 AIC값을 찾아보았다. 그 결과 아래와 같은 AIC값들을 얻을 수 있었다.

<표 3.4> 환율 시계열자료의 AIC

모형	df	AIC
ARIMA(0,1,0)	1	-407.2241
ARIMA(1,1,0)	2	-443.7060
ARIMA(0,1,1)	2	-511.4282
ARIMA(1,1,1)	3	-510.5106
ARIMA(0,1,2)	3	-510.7845
ARIMA(2,1,0)	3	-482.7295
ARIMA(1,1,2)	4	-507.7024
ARIMA(2,1,1)	4	-510.6906
ARIMA(2,1,2)	5	-510.5458

<표 3.4>에서 AIC값이 가장 작은 모형이 적합한 모형으로 ARIMA(0,1,1)이 가장 적합한 모형임을 알 수 있다.

### 3.3.3 모수 추정

모형이 식별되었으면 그 모형에 포함된 모수들을 추정하여야 한다. 주로 사용되는 추정 방법으로는 최소제곱 추정법, 최대우도 추정법, 적률 추정법등이 있다. 적률 추정 법은 이론 전 개가 쉽다는 장점이 있는 반면에 효율성이 떨어지는 단점이 있다. 추정에서는 일반적으로 최소제곱 추정법, 최대우도 추정법이 사용되며 이들의 초기 값으로 적률 추정법에서 얻어진 값이 사용된다. 검정은 일반적인 t-검정이 사용된다.

다음은 R 패키지 중 최적의 시계열 모형을 찾아주는 함수 auto.arima의 기본값으로 정해진 최대우도 추정법을 이용하여 추정한 결과를 summary한 것이다.  $Z_t$ 는 t시점에서의 자료이며  $a_t$ 는 평균이 0, 분산이  $\sigma_a^2$ 인 백색잡음이다.

<표 3.5> 환율 시계열자료의 모형 추정치

Coefficients
ma1
-0.1552
$Z_t = Z_{t-1} + a_t - (-0.1552)a_{t-1}$
$(1 - B)Z_t = (1 + 0.1552B)a_t$
$W_t = (1 + 0.1552B)a_t$

auto.arima 함수에 대한 summary를 R로 입력하여 나온 <표 3.5>의 결과에서 보면 한 번 차분한 환율 모형인 ARIMA(0,1,1)의 ma 1차 계수를 얻을 수 있었다. 따라서 환율 시계열 모형식은

$$W_t = (1 + 0.1552B)a_t$$

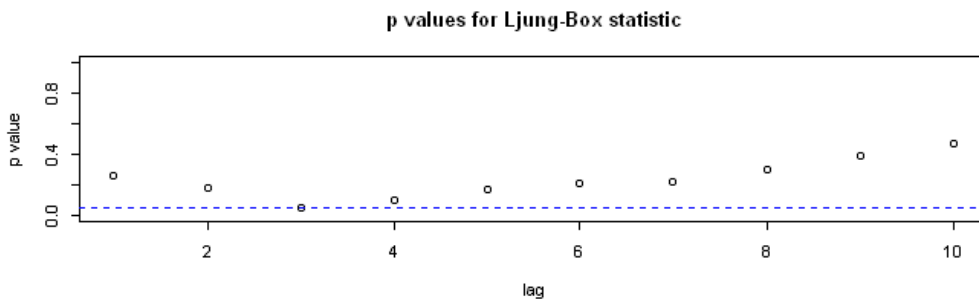
이다.

### 3.3.4 모형의 적합성 검정

모수 추정을 통하여 얻은 예측 모형의 잔차는 시험 모형을 평가하고 진단하는데 주요한 정보를 제공한다. 잔차 계열의 자기상관계수도 모든 시차에서 0이 되어 백색잡음임이 확인된다면 추정된 시험 모형은 적절한 것으로 판단되는 것이다.

#### ▶ 퍼트멘토우 검정(Box and Pierce portmanteau test)

퍼트멘토우 검정은 「경제 및 금융자료를 위한 시계열 분석. 김해경, 이명숙 저」에 의하면 표본자기상관계수의 집단을 기초로 잔차 계열이 백색잡음과 같이 어떤 추세나 종속성이 없이 완전히 랜덤한가를 검정하여 적합된 모형이 좋은지 나쁜지를 판단하는 검정이다. 여기서 P-value값이 크면 클수록 모형이 잘 선택되고 분석이 잘 된 것으로 판단하면 된다. 퍼트멘토우 검정을 하여 모형에 주어지는 오차항의 추정 값인 잔차들의 자기상관계수가 모든 시차에서 0이 되는지에 대한 유의성 검정을 하기에 앞서 <그림 3.5>는 잔차들의 diagnostic plots을 그린 결과이다.



<그림 3.5> 환율 자료 Diagnostic plots

백색잡음(White noise)은 대표적 정상 시계열이며 평균이 0이고 분산이  $\sigma_a^2$ 인 확률 변수로

$$a_t \sim iid(0, \sigma_a^2)$$

이다. <그림 3.5>에서 점선은 유의수준을 나타내는 0.05 이고, 작은 동그라미는 각 시차에서의 P-value로 이 값을 보면 모두 0.05보다 크므로 잔차가 백색잡음의 성질을 갖기 때문에 모형의 식별과 추정이 잘 되었다고 볼 수 있다.

다음으로 <표 3.6>는 Ljung-Box Test라고도 불리는 퍼트멘토우 검정의 판별을 위하여 자기상관함수는 표본 크기의 약 25%의 시차에서 계산하는 것이 효과적이므로 환율 자료의 크기인 144의 약 25%인 시차 0에서 36까지에서 계산을 해 보았다.

<표 3.6> 환율 자료 Ljung-Box Test

Lag	Chi	DF	p-value
6	8.1959	6	0.2241
12	9.8276	12	0.6311
18	14.1999	18	0.716
24	15.8988	24	0.8917
30	19.4767	30	0.9295
36	23.3486	36	0.9487

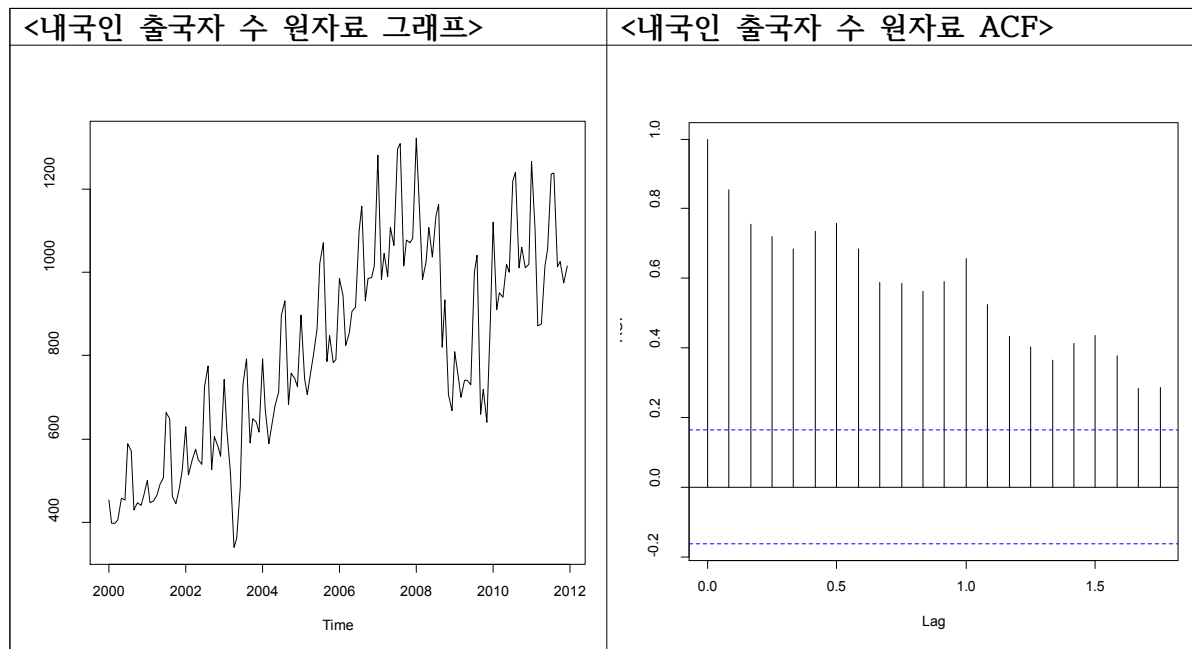
$H_0$  : 적합된 모형이 타당하다. (모든 시차에서 자기상관계수가 0이다.)

위의 결과에서 보면 시차 1에서 6까지의 자기상관계수를 기초로 하는 카이제곱검정통계량의 값은 8.1959이고 자유도는 6, 그때의 p값은 0.2241로 이 시차들에서 자기상관계수가 모두 0이라는 사실을 나타내고 있다. 시차 1에서 12까지, 시차 1에서 18까지 그리고 시차 1에서 36까지에 대한 검정을 통해서도 같은 결과를 얻는다. 결국 모형 선택과 모수추정이 만족할 만한 수준임을 확인할 수 있다. 모든 p-value가 유의수준 0.05보다 훨씬 크므로 적합된 ARIMA(0,1,1) 모형이 타당하다는 귀무가설을 기각할 수 없다. 따라서 적합된 ARIMA(0,1,1) 모형이 타당하다고 할 만한 충분한 근거가 있다. 퍼트멘토우 결과도 같은 결과를 말해줌으로써 모형 선택과 모수추정이 만족할 만한 수준임을 확인할 수 있다.

### 3.4 내국인 출국자 수 자료

#### 3.4.1 변환 및 차분

내국인 출국자 수 자료의 그래프와 ACF를 그려 보았다.



<그림 3.6> 내국인 출국자 수 시계열자료의 그래프와 ACF

<그림 3.6>의 그래프에서 볼 수 있듯이 내국인 출국자 수의 시계열 자료는 평균이 일정하지 않고, 추세가 있는 것처럼 보이므로 계절차분이 필요하다고 판단된다. 한편 분산은 다소 일정한 편이나 로그변환을 통해서 분산이 더 일정해지는지 확인해 볼 것이다. ACF가 전반적으로 천천히 감소하는 직선의 추세를 보이고 있으며 일정하게 감소하는 형태가 아닌 증가했다가 감소하는 일정 주기가 반복되는 모습을 띄고 있다. 계절성이란 절기나 분기 또는 월, 일별로 관측되는 시계열자료에서 계절적 패턴을 갖는다. 즉 계절이 규칙적으로 변화하고 반복되면서 이에 따른 변화가 나타나는 것이다. 따라서 내국인 출국자 수 자료는 비정상 계절성 시계열 자료임을 예상 할 수 있다.

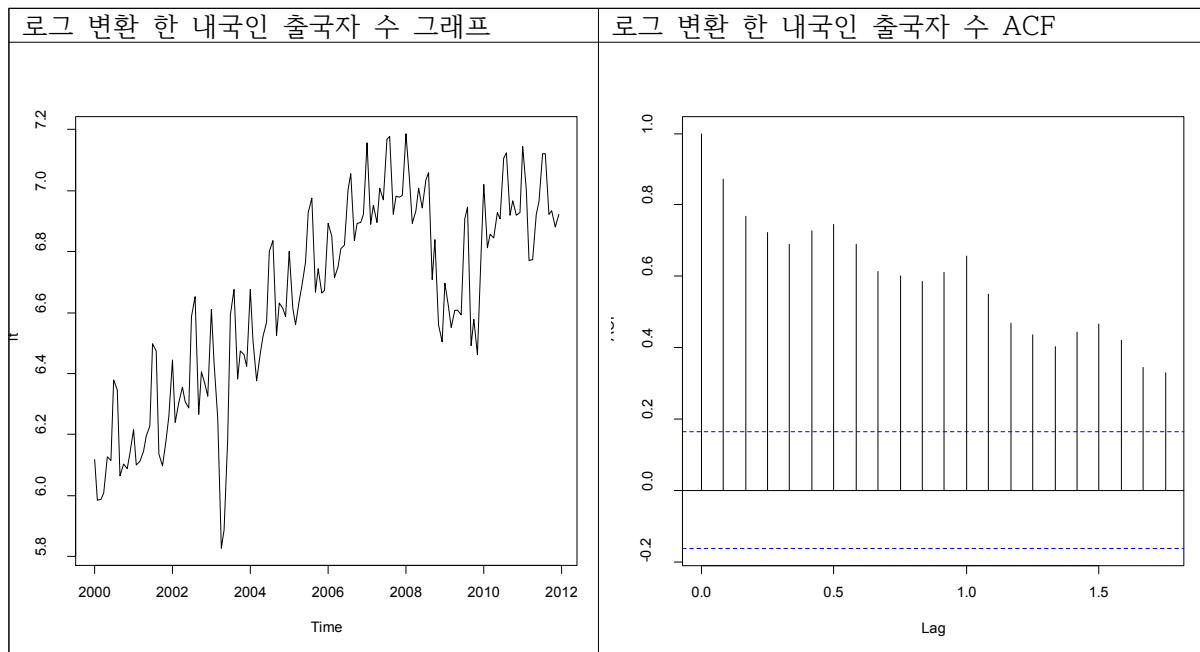
먼저 확실한 검증을 위해 정상인지 비정상인지 검정하는 Dickey-Fuller Test를 시행해 보았다. DF-test의 귀무가설이 비정상 시계열이고 대립가설이 정상 시계열로, p-value가 유의수준 0.05보다 크면 귀무가설을 기각 할 수 없으므로 자료가 비정상 시계열이라고 할 수 있다.

<표 3.7>는 내국인 출국자 수 자료가 정상시계열인지 비정상시계열인지 알아보기 위해 DF-test를 시행한 결과이다.

<표 3.7> 내국인 출국자 수 시계열자료에 대한 Dickey-Fuller Test.

Augmented Dickey-Fuller Test		
data : trip		
Dickey-Fuller=-1.6831	Lag order=5	p-value=0.7084
alternative hypothesis : stationary		

<표 3.7>의 DF-test 결과, p-value가 0.7084로 유의수준 0.05보다 크기 때문에 비정상 시계열이라고 할 수 있다. 먼저 비정상시계열의 분산의 안정화를 위해 로그변환을 해서 자료의 그래프와 ACF를 <그림 3.7>에서 확인해 보았다.

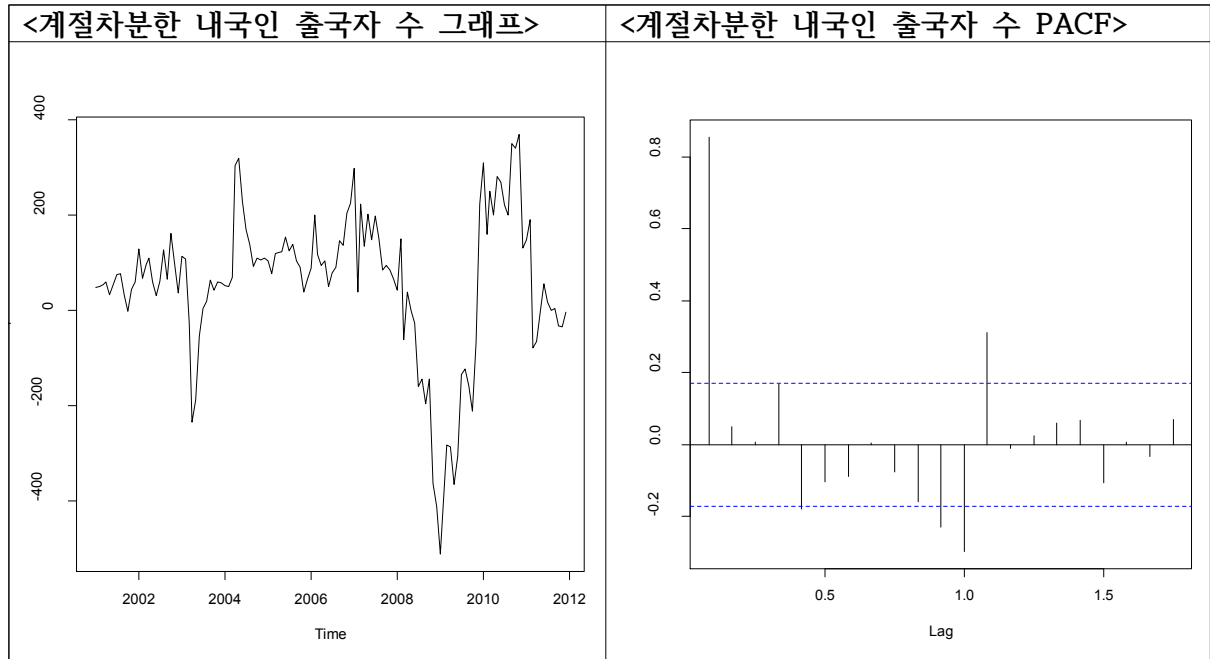


<그림 3.7> 로그 변환한 내국인 출국자 수 시계열자료 그래프와 ACF

<그림 3.7>에서 볼 수 있듯이 로그변환을 한 결과, 원시계열자료의 그래프와 ACF간의 큰 차이가 없기 때문에 로그변환을 하지 않은 원시계열자료를 사용할 것이다.

내국인 출국자 수 자료의 그래프와 ACF를 보면 변동이 다소 일정한 것으로 보아 분산이 일정하다고 할 수 있지만, 평균은 일정하지 않으며 추세가 있고 자기상관함수가 빠른 속도로 감소하지 않기 때문에 계절 차분이 필요하다고 판단된다. <그림 3.8>은 내국인 출국자 수 시계열 자료를 계절 차분한 그래프와 PACF이다.





<그림 3.8> 계절차분한 내국인 출국자 수 시계열자료의 그래프와 PACF

계절차분을 한 결과 원 자료에 비해 평균이 다소 일정해 진 것을 볼 수 있으며, PACF를 보면 12를 주기로 부분자기상관함수가 튀어나오고 시차가 증가함에 따라 튀어나온 부분이 감소하는 것을 볼 수 있다. 따라서 내국인 출국자 수 자료는 주기가 12인 승법 계절 ARIMA 모형임을 예상할 수 있고, 원자료를 계절 차분하여 비정상성을 없애 변환된 자료로 분석에 이용하겠다.

### 3.4.2 모형 식별

<표 3.8>은 내국인 출국자수 자료의 모형 식별을 위해 가능한 모형의 AIC를 구한 결과이다.

<표 3.8> 내국인 출국자 수 시계열자료의 AIC

모형	계절모형	df	AIC
ARIMA(1,0,0)	ARIMA(0,1,0)	2	1534.047
ARIMA(1,0,0)	ARIMA(0,1,1)	3	1440.667
ARIMA(1,0,0)	ARIMA(0,1,2)	4	1405.889
ARIMA(1,0,0)	ARIMA(1,1,0)	3	1481.031
ARIMA(1,0,0)	ARIMA(1,1,1)	4	1422.447
ARIMA(1,0,0)	ARIMA(1,1,2)	5	1406.781
ARIMA(1,0,0)	ARIMA(2,1,0)	4	1433.032
ARIMA(1,0,0)	ARIMA(2,1,1)	5	1406.461
ARIMA(1,0,0)	ARIMA(2,1,2)	6	1407.051
ARIMA(1,0,1)	ARIMA(0,1,0)	3	1535.991
ARIMA(1,0,1)	ARIMA(0,1,1)	4	1442.182
ARIMA(1,0,1)	ARIMA(0,1,2)	5	1405.114
ARIMA(1,0,1)	ARIMA(1,1,0)	4	1482.228
ARIMA(1,0,1)	ARIMA(1,1,1)	5	1423.098
ARIMA(1,0,1)	ARIMA(1,1,2)	6	1407.113
ARIMA(1,0,1)	ARIMA(2,1,0)	5	1434.077
ARIMA(1,0,1)	ARIMA(2,1,1)	6	1406.847
ARIMA(1,0,1)	ARIMA(2,1,2)	7	1406.847
ARIMA(1,0,2)	ARIMA(0,1,0)	4	1537.937
ARIMA(1,0,2)	ARIMA(0,1,1)	5	1444.104
ARIMA(1,0,2)	ARIMA(0,1,2)	6	1407.095
ARIMA(1,0,2)	ARIMA(1,1,0)	5	1484.146
ARIMA(1,0,2)	ARIMA(1,1,1)	6	1425.094
ARIMA(1,0,2)	ARIMA(1,1,2)	7	1409.093
ARIMA(1,0,2)	ARIMA(2,1,0)	6	1435.466
ARIMA(1,0,2)	ARIMA(2,1,1)	7	1408.832
ARIMA(1,0,2)	ARIMA(2,1,2)	8	1408.826

<표 3.8>에서 AIC값이 가장 작은 모형이 적합한 모형으로 ARIMA(1,0,1)(0,1,2)[12]가 가장 적합한 모형임을 알 수 있다.

### 3.4.3 모수 추정

<표 3.9>은 R 패키지 중 최적의 시계열 모형을 찾아주는 함수 auto.arima의 기본값으로 정해진 최대우도 추정법을 이용하여 모수를 추정한 결과를 summary한 것이다.

<표 3.9> 출국자 수 모수 추정

Coefficients			
ar1	ma1	sma1	sma2
0.9716	-0.1647	-0.8301	0.1385
$(1 - B^{12})(1 - \phi B)Z_t = (1 - \theta B)(1 - \Theta B^{12} - \Theta B^{24})a_t$			
$(1 - B^{12})(1 - 0.9716B)Z_t = (1 + 0.1647B)(1 + 0.8301B^{12} + 0.1385B^{24})a_t$			

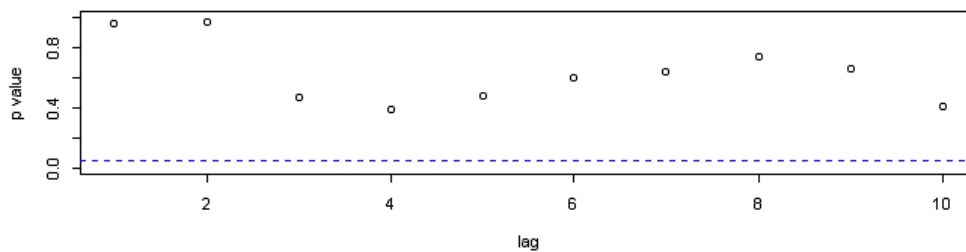
환율 자료와 마찬가지로 auto.arima라는 함수를 이용하여 예측 그래프를 얻었고 auto.arima 함수에 대한 summary를 R로 입력하여 나온 결과 출국자 수 모형인 ARIMA(1,0,1)(0,1,2)[12]의 ar(1), ma(1), 계절ma(1), 계절ma(2)의 계수를 얻을 수 있었다. 따라서 출국자 수 시계열 모형식은

$$(1 - B^{12})(1 - 0.9716B)Z_t = (1 + 0.1647B)(1 + 0.8301B^{12} + 0.1385B^{24})a_t$$

이다.

#### 3.4.4 모형의 적합성 검정

모수 추정을 통하여 얻은 예측 모형의 잔차는 시험 모형을 평가하고 진단하는데 주요한 정보를 제공한다. 잔차 계열의 자기상관계수도 모든 시차에서 0이 되어 백색잡음임이 확인된다면 추정된 시험 모형은 적절한 것으로 판단되는 것이다. 퍼트멘토우 검정을 하여 모형에 주어지는 오차항의 추정 값인 잔차들의 자기상관계수가 모든 시차에서 0이 되는지에 대한 유의성 검정을 하기에 앞서 <그림 3.9>은 잔차들의 diagnostic plots을 그린 결과이다.



<그림 3.9> 내국인 출국자 수의 잔차에 대한 P-values for Ljung-Box Statistic

<그림 3.9>에서 점선은 유의수준을 나타내는 0.05 이고, 작은 동그라미는 각 시차에서의 P-value로 이 값을 보면 모두 0.05보다 크므로 내국인 출국자 수 시계열자료의 잔차가 백색잡음의 성질을 갖기 때문에 모형의 식별과 추정이 잘 되었다고 볼 수 있다.

다음으로 <표 3.10>은 Ljung-Box Test라고도 불리는 퍼트멘토우 검정의 판별을 위하여 자기상관함수는 표본 크기의 약 25%의 시차에서 계산하는 것이 효과적이므로 내국인 출국자 수 자료의 크기인 144의 약 25%인 시차 0에서 36까지에서 계산을 해 보았다.

<표 3.10> 내국인 출국자 수 자료 Ljung-Box Test

Lag	Chi	DF	p-value
6	4.4271	6	0.6191
12	11.1263	12	0.5181
18	16.5406	18	0.5549
24	27.0486	24	0.3022
30	28.9991	30	0.5176
36	36.5679	36	0.4423

$H_0$  : 적합된 모형이 타당하다.

출국자 수 자료도 마찬가지로 위의 결과에서 보면 P-value 값이 유의수준 0.05보다 크므로 적합된 ARiMA(1,0,1)(0,1,2)[12] 모형이 타당하다는 귀무가설을 기각할 수 없다. 따라서 적합된 ARiMA(1,0,1)(0,1,2)[12] 모형이 타당하다고 할 만한 충분한 근거가 있다.

퍼트멘토우 결과도 같은 결과를 말해줌으로써 모형 선택과 모수추정이 만족할 만한 수준임을 확인 할 수 있다.

## 4. 예측

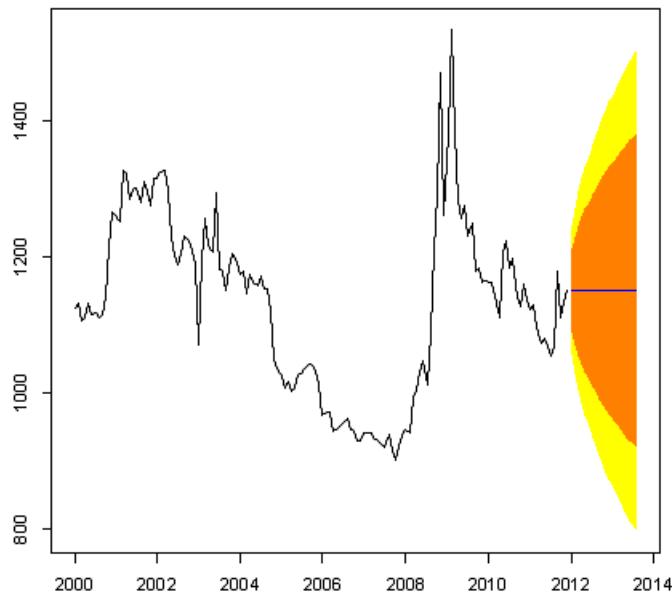
### 4.1 환율 예측

시계열 모형이 선택되면 마지막 단계로 주어진 시계열의 미래 값을 결정하는 예측의 절차가 요구된다. <표 3.5>의 결과에서 알 수 있듯이 환율 시계열자료의 추정된 모형은

$$W_t = A_t - 0.1552A_{t-1}$$

이다. 원시계열의 미래 값들을 추정하는데 사용할 수 있는 예측모형을 보여주는 R의 forecast 패키지를 이용해서 t=145, 146, 147 등에서 미래 값을 예측해 보았다.

<그림 4.1>은 R의 Forecast 패키지를 통해 얻어진 환율 시계열자료의 예측 그래프이다.



<그림 4.1> 환율 시계열자료의 예측

<그림 4.1>에서 검은 선은 2000년부터 2011년 까지 주어진 자료를 통해서 적합한 모형인 ARIMA(0,1,1)을 의미한다. 파란 선은 2012년 후의 자료는 주어지지 않았기에 R의 패키지가 예상해서 얻어진 값을 의미한다. 환율은 뚜렷한 주기 없이 정해지기 때문에 예측을 한 결과 평균값으로 구해진 것으로 보인다. 그래프의 우측에 보이는 주황색 구간은 예상해서 얻어진 값의 80% 신뢰구간을 노란색 구간은 예상해서 얻어진 값의 95% 신뢰구간을 의미한다. 따라서 80%, 95% 신뢰 구간을 예상 포인트로 계획할 수 있다.

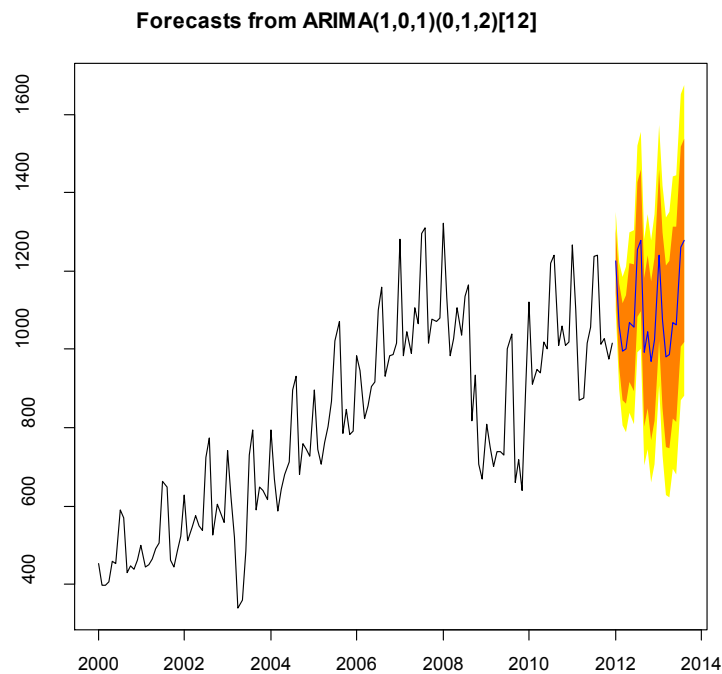
## 4.2 내국인 출국자 수 예측

<표 3.9>의 결과에 따라 내국인 출국자 수 시계열자료에 대해 추정된 모형은

$$(1 - B^{12})(1 - 0.9716B)Z_t = (1 + 0.1647B)(1 + 0.8301B^{12} + 0.1385B^{24})a_t$$

이다. 원시계열의 미래 값들을 추정하는데 사용할 수 있는 예측모형을 보여주는 R의 forecast 패키지를 이용해서 t=145, 146, 147 등에서 미래 값을 예측해 보았다.

R의 Forecast 패키지를 통해 얻어진 그래프이다.



<그림 4.2> 내국인 출국자 수 시계열자료의 예측

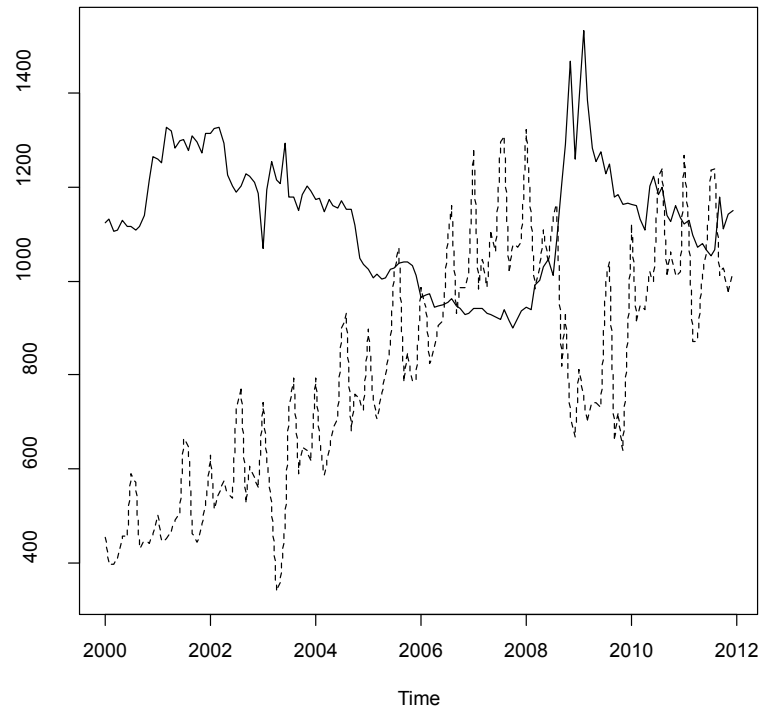
<그림 4.2>과 같이 forecast 패키지를 이용하여 내국인 출국자 수 시계열 자료에 대해 예측을 한 결과 우선 원자료의 시계열모형이 ARIMA(1,0,1)(0,1,2)[12]로 나왔고 2012년 이후 내국인 출국자 수를 예측하는 그래프 또한 그려졌다. 위 그래프의 검은 선은 2000년부터 2011년 까지 주어진 자료를 통해서 적합한 모형인 ARIMA(1,0,1)(0,1,2)[12]를 의미한다. 파란 선은 2012년 후의 자료는 주어지지 않았기에 R의 패키지가 예상해서 얻어진 값을 의미한다. 그래프의 우측에 보이는 주황색 구간은 예상해서 얻어진 값의 80% 신뢰구간을 노란색 구간은 95% 신뢰구간을 의미한다.

## 5. 상관분석

### 5.1 환율과 내국인 출국자 수 자료의 상관분석

환율과 내국인 출국자 수 자료에 대해서 상관분석을 해 보았다.

다음의 그림은 3장의 <그림 3.1>에 해당하는 두 자료를 한 그래프 상에 그린 결과이다.



환율이 높아지면 내국인 출국자 수가 낮아지는 경향이 있는 것으로 보이기 때문에 두 자료가 음의 상관관계가 있다고 생각하여 대립가설을 상관계수가 0보다 작다는 단측 검정을 하였다.

R의 cor.test 함수를 이용했고 함수 옵션에 단측 검정임을 설정했다. 상관분석을 한 결과는 <표 5.1>과 같다.

<표 5.1> 환율과 내국인 출국자 수 자료간의 상관분석

Pearson's product-moment correlation ( data : exchange rate and trip)				
$H_0$ : 피어슨 상관계수가 0 이다. vs $H_1$ : 피어슨 상관계수가 0보다 작다.				
t	df	p-value	95% 신뢰구간	sample estimates
-8.3645	142	2.55e-14	(-1, -0.4743945 )	-0.5745225

<표 5.1>에서 보면 95% 신뢰구간이 (-1, -0.4743945)로 상관계수 추정치인 -0.5745225를 포함하고 있으므로 귀무가설을 기각 한다. 또한 가설에 대한 p-value가  $2.55 \times 10^{-14}$ 로 유의수준 0.05보다 작으므로 귀무가설을 기각 할 근거가 있다. 여기서 p-value는 R 패키지에서 단측 검정에 따른 p값이다. 따라서 상관계수가 0보다 작다고 할 수 있다. 피어슨 상관계수가 -0.5745225로 환율 자료와 내국인 출국자 수 자료 간에 음의 상관관계가 있다고 할 근거가 있다.



## 6. 전이함수모형

전이함수모형은 박스-젠킨스에 의하여 고안된 예측모형으로 한 시계열의 미래값을 예측하는 데 자신의 과거나 현재의 값은 물론이고 이 시계열과 인과관계를 갖고 선도하는 다른 시계열의 과거나 현재 그리고 미래의 예측값까지도 예측에 이용함으로써 예측 효과를 높이는 모형이다. 실제로 전이함수모형은 비교적 적은 개수의 모수를 갖는 선형시스템의 모형으로 예측오차를 크게 줄이는 모형으로 알려져 있다.

이 장의 내용은 「경제 및 금융자료를 위한 시계열 분석. 김해경·이명숙 저」를 참고한 것이다.

### 6.1 전이함수

#### 6.1.1 전이함수와 충격반응함수

두 시계열  $\{X_t\}$ 와  $\{Y_t\}$  사이에 종속관계가 존재하고 그 중 한 시계열  $\{X_t\}$ 가  $\{Y_t\}$ 를 선도할 때,  $\{X_t\}$ 를 입력시계열(input time series) 그리고  $\{Y_t\}$ 를 출력시계열(output time series)이라 한다. 주어진 두 시계열 사이에 존재하는 시간에 따라 변하는 종속관계를 적은 개수의 모수를 통하여 모형화하는 방법과 시계열의 분석을 한다.

입력변수  $X_t$ 가 출력변수  $Y_t$ 에 주는 영향을 선형변환을 통하여 입력시계열  $\{X_t\}$ 을 출력시계열  $\{Y_t\}$ 로 전환하는 선형시스템(linear system) 모형

$$\begin{aligned} Y_t &= \nu_0 X_t + \nu_1 X_{t-1} + \dots \\ &= \nu(B)X_t, \quad \nu(B) = \nu_0 B^0 + \nu_1 B^1 + \dots \end{aligned} \quad (6.1)$$

( $B$ 는 후진연산자)이 효과적이다. 이때  $\nu(B) = \nu_0 B^0 + \nu_1 B^1 + \dots$ 를 전이함수(transfer function)라 하고, 그 계수  $\nu_0, \nu_1, \dots$ 를 가중값으로 하여 출력계열  $Y_t$ 에 전달됨을 뜻한다. 특히  $\sum_{i=0}^{\infty} |\nu_i| < \infty$ 이면 모형(6.1)을 안정적이라 한다. 모형 (6.1)을 결정하는 일은 가중값인 무한개의 계수를 정하는 것이다. 그러나 유한개의 정보로 무한개의 계수를 추정하기는 실제로 불가능하므로 이러한 문제를 해결하는 데는 박스-젠킨스가 고안한 전이함수모형 방법이 효과적이다. 그러나 이 방법은 시간영역방법으로 입력계열이 일변수인 경우에만 적용된다. 먼저  $\nu(B)$ 를

$$\nu(B) = \frac{\omega(B)B^b}{\delta(B)} \quad (6.2)$$

로 놓고,  $\omega(B) = \omega_0 - \omega_1 B$ ,  $\delta(B) = 1 - \delta B$ ,  $b = 0$ 인 경우,

$$\begin{aligned}
\nu(B) &= \frac{\omega(B)}{\delta(B)} B^b = \frac{\omega_0 - \omega_1 B}{1 - \delta B} \\
&= (\omega_0 - \omega_1 B)(1 + \delta B + \delta^2 B^2 + \dots) \\
&= \omega_0 + (\delta\omega_0 - \omega_1)B + (\delta\omega_0 - \omega_1)\delta B^2 + \dots
\end{aligned}$$

으로,  $\nu_0 = \omega_0$ ,  $\nu_1 = \omega_0\delta - \omega_1$ ,  $\nu_k = \delta^{k-1}\nu_1$  ( $k \geq 2$ )이다. 여기서  $b$ 는 두 변수 사이의 시차를 나타내는 모수이다.

한편,  $Y_t$ 는  $X_t$ 의 현재와 과거의 값만으로 결정되지 않고 보통 잡음(noise)이라고 부르는 확률선분  $N_t$ 와의 합으로 나타난다. 다시 말해

$$Y_t = \nu(B)X_t + N_t = \frac{\omega(B)B^b}{\delta(B)}X_t + N_t \quad (6.3)$$

로 가정된다. 여기서  $N_t$ 는 입력변수  $X_t$ 와는 독립 또는 무상관인 자기회귀이동과정(ARMA)으로 가정된다. 모형 (6.3)을 전이함수잡음모형(transfer function-noise models) 또는 간단히 전이함수모형(transfer function models)이라 한다. 또 이 모형을 ARMAX 모형이라고도 한다. 이 모형은 과거나 현재의  $X_t$ 값이 현재나 미래의  $Y_t$ 의 값에 영향을 주지만 그 역은 성립하지 않는 일방적 인과관계를 설명하고 있다.

### 6.1.2 전이함수의 계수와 모수

전이함수모형의 분석은 모형(6.3)에 주어진 전이함수의 설정과 그 모수의 추정, 잡음과정  $N_t$ 의 모형설정과 그 모수의 추정 그리고 최종적으로 얻은 모형을 통하여 출력계열의 미래 값들을 예측하는 것을 목적으로 한다. 따라서 입·출력 시계열 자료를 기초로 하여 관련되는 두 다항식  $\omega(B)$ 와  $\delta(B)$ 의 차수와 계수 그리고 입력계열이 출력계열에 영향을 주는 최초의 시점을 나타내는 모수  $b$ 의 값이 결정되어야 한다.

## 6.2 전이함수모형의 설정

### 6.2.1 사전백색화

$\{X_t\}$ 를 미리 백색잡음계열로 만드는 사전백색화(prewhitening) 방법을 이용하는데 먼저  $\{X_t\}$ 와  $\{Y_t\}$ 를 정상시계열이라고 가정한다. 1단계는 입력계열의 사전백색화로 입력시계열  $X_t$ 에 ARMA(또는 ARIMA)모형을 적합하여 백색잡음인 잔차계열을 만든다. 즉

$$\begin{aligned}\phi_x(B)X_t &= \theta_x(B)\alpha_t \\ \alpha_t &= \theta_x^{-1}(B)\phi_x(B)X_t\end{aligned}$$

을 만족하는 백색잡음계열  $\{\alpha_t\}$ 을 만든다.

2단계는 출력계열의 변환으로 출력시계열  $Y_t$ 를 위에서 얻은 같은 ARMA모형을 통하여 변환한다. 즉  $\beta_t = \theta_x^{-1}(B)\phi_x(B)Y_t$  을 만족하는 변환계열  $\{\beta_t\}$ 을 만든다. 이 변환계열은 백색잡음계열일 필요는 없고 일반적으로 자기종속된 계열이라 할 수 있다.

3단계는 충격반응가중값의 추정으로 변환된 두 계열  $\{\alpha_t\}$ 와  $\{\beta_t\}$ 를 통하여 모형 (6.3)에 대한 전이함수의 충격반응가중값  $\nu_h$ 들을 구한다.  $\{\alpha_t\}$ 와  $\{\beta_t\}$ 의 전이변환함수는 본래의 두 시계열  $\{X_t\}$ 와  $\{Y_t\}$ 의 전이변환함수와 같다는 사실을 이용하고,  $\{\alpha_t\}$ 이 백색잡음인 성질을 이용하면 전이함수의 계수  $\nu_h$ 의 추정값은

$$\hat{\nu}_h = \frac{s_\beta}{s_\alpha} r_{\alpha\beta}(h) \quad (6.4)$$

이다. 여기서  $r_{\alpha\beta}(h)$ 는  $\{\alpha_t\}$ 와  $\{\beta_t\}$ 의 표본교차상관이고,  $s_\alpha^2$ 와  $s_\beta^2$ 은 각각의 표본분산이다.

4단계는 전이함수의 모수추정이다. 시차 h에 대한 표본상관계수  $r_{\alpha\beta}(h)$ 을 계산하여 가중값  $\nu_h$ 를 (6.4)를 통하여 추정한다. 추정된 충격반응가중값  $\hat{\nu}_h$ 들을 기초로 하여 차수 (s, r, b)를 결정하고 계수비교식을 통하여

$$\hat{\nu}(B) = \frac{\hat{\omega}(B)B^b}{\hat{\delta}(B)}$$

얻는다.

### 6.2.2 잡음모형의 설정

전이함수모형이 일단 결정되면, (6.3)의 잡음성분의 모형을 설정한다.

1단계는 잡음계열의 추정이다. 추정값  $\hat{\omega}(B)$ 와  $\hat{\delta}(B)$ 을 이용하여 잡음계열  $\{N_t\}$ 의 추정계열  $\{\hat{N}_t\}$ 을 만든다.

2단계는 잡음계열의 판별로 잡음계열의 추정계열  $\{\hat{N}_t\}$ 의 표본자기상관함수와 표본편자기상관함수를 기초로 하여 잡음계열  $\{N_t\}$ 의 적절한 ARMA모형을 정한다.

### 6.2.3 최종 전이함수의 결정

전이함수모형과 잡음모형이 결정되면, 최종적인 전이확률모형은

$$Y_t = \frac{\omega(B)}{\delta(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} \epsilon_t \quad (6.5)$$

로 주어진다.

### 6.3 전이함수모형의 진단

전이함수모형을 판별한 후 모형이 갖는 모수들을 추정하고 나면 이 결과를 예측이나 제어 등에 사용하기에 앞서 이 모형의 적절성에 대한 진단이 요구된다.

모형이 적절하지 않은 경우로는 전이함수모형만이 적절하지 않은 경우, 잡음모형만이 적절하지 않은 경우, 전이모형과 잡음모형 모두가 적절하지 않은 경우 등으로 구분된다.

전이함수모형의 진단에서는 계산된 사전백색화계열  $\{\alpha_t\}$ 와 추정된 오차계열  $\{\hat{\epsilon}_t\}$ 에 대한 검토를 해야 한다. 적절한 전이함수모형을 위해서는  $\{\hat{\epsilon}_t\}$ 에 자기상관이 존재하지 않아야 하며, 또  $\{\alpha_t\}$ 와  $\{\hat{\epsilon}_t\}$ 에 교차상관이 없어야 한다.

잔차  $\{\hat{\epsilon}_t\}$ 의 자기종속에 대한 유의성 검정과  $\{\alpha_t\}$ 와  $\{\hat{\epsilon}_t\}$ 의 교차종속에 대한 유의성 검정에 박스-젠킨스 방법에서와 같이 바틀렛 검정(Bartlette test)이나 퍼트멘토우(Portmanteau test)을 이용할 수 있다.

### 6.4 전이함수모형의 예측

판별과 추정 그리고 진단의 절차를 통하여 적절한 전이확률모형이 확정되면, 이 모형은 출력계열  $\{Y_t\}$  자신의 과거의 자료는 물론 입력계열  $\{X_t\}$ 의 자료를 기초로 하여 출력계열의 미래의 값을 예측하는 데 이용된다. 이 예측모형은 출력변수 하나만을 기초로 하는 일변수의 예측에서보다 예측오차를 훨씬 감소시키며, 특히 입력변수가 출력변수의 선도인자가 되는 경우에는 더 효과적임이 밝혀졌다.

## ※ 참고문헌

ARIMA 모형을 이용한 시계열 분석 - 신기일 저

시계열 분석 - 신기일 저

시계열 연도별 사건 (출처 : <http://railroadbrothers.tistory.com/4>)

경제 및 금융자료를 위한 시계열 분석. 김해경, 이명숙 저

## ※ 부록

## 환율 시계열그림 ##

```
e<-read.csv("exchange.csv", header=F, sep=",")
```

```
e<-unlist(e)
```

```
er<-as.numeric(e)
```

```
ER<-ts(er, start=2000, frequency=12)
```

```
plot(ER,ylab="Exchange Rate", main="Time Series of Exchange Rate")
```

## 출국자수 시계열 ##

```
t<-read.csv("trip.csv", header=F, sep=",")
```

```
t<-unlist(t)
```

```
tr<-as.numeric(t)
```

```
trip<-ts(tr, start=2000, frequency=12)
```

```
plot(trip, ylab="# of departure", main="departure # of local residents")
```

```
ts.plot(ER, trip, lty=c(1,2),lwd=1, ylab="Figure", main="Time Series between  
Exchange Rate and Departure # of local residents")
```

## 환율 모형식별화 ##

```
ld<-log(ER) # 로그변환
```

```
plot(ld)
```

```
acf(ld)
```

```
dd<-diff(ld) # 로그변환한 거 차분
```

```
plot(dd)
```

```
acf(dd)
```

```
shapiro.test(ER) # test for normality
```

```
###Now let's take a look at the ACF and PACF of dlj###
```

```
par(mfrow=c(2,1)) # The power of accurate observation is commonly  
called cynicism
```

```
                #      by those who have not got it. - George Bernard  
Shaw
```

```
acf(ER, 20)      # ACF to lag 20 - no graph shown... keep reading
```

```
pacf(ER, 20)     # PACF to lag 20 - no graph shown... keep reading
```

```
pacf(dd)
```

```
acf(dd)
```

```
Trip<- time(trip)-2000 # not necessary to "center" time, but the results  
look nicer
```

```
reg<-lm(log(ER)~Trip, na.action=NULL) # run the regression without an  
intercept
```

```
#-- the na.action statement is to retain time series attributes
```

```
summary(reg)
```

```
## ER=7.069426-0.007792*Trip / R-squared : 0.0554 (이 추정된 회귀식은  
55.4% 정도 설명한다) ##
```

```
#ARIMA모형식별#
```

```
a <-arima(dER, order = c(0,1,0))
```

```
a1<-arima(dER, order = c(1,1,0))
```



```

a2<-arima(dER, order = c(0,1,1))
a3<-arima(dER, order = c(1,1,1))
a4<-arima(dER, order = c(0,1,2))
a5<-arima(dER, order = c(2,1,0))
a6<-arima(dER, order = c(1,1,2))
a7<-arima(dER, order = c(2,1,1))
a8<-arima(dER, order = c(2,1,2))          # 모형의식별 aic값이 제일작은
arima(0,1,1) 백색잡음인 모형이라고 볼수있다.
AIC(a,a1,a2,a3,a4,a5,a6,a7,a8)

```

## 출국자수 모형 식별화 ##

```

lt<-log(trip)
plot(lt)
acf(lt)

```

```

dd<-diff(log(trip)) ##로그변환 후 차분
plot(dd)
plot(trip)
dtrip<-diff(trip) # 원자료 1차차분
plot(dtrip)
shapiro.test(trip) # test for normality

```

```

par(mfrow=c(2,1))          # set up the graphics
hist(dtrip, prob=TRUE, 12)  # histogram
lines(density(dtrip), col='red') # smooth it - ?density for details
qqnorm(dtrip)               # normal Q-Q plot
qqline(dtrip, col='red', ylim=c(-3,3)) # add a line
?qqline

```

```
par(mfrow=c(2,1)) # The power of accurate observation is commonly  
called cynicism
```

```
# by those who have not got it. - George Bernard
```

```
Shaw
```

```
acf(trip, 20) # ACF to lag 20 - no graph shown... keep reading
```

```
pacf(dtrip, 20)
```

```
#계절차분
```

```
plot(trip)
```

```
acf(trip)
```

```
pacf(trip)
```

```
logs<-log(trip)
```

```
plot(logs)
```

```
acf(logs)
```

```
pacf(logs)
```

```
strip<- diff(trip, lag=12, differences=1)
```

```
plot(strip)
```

```
acf(strip)
```

```
pacf(strip)
```

```
#모형
```

```
b1<-arima(strip, order = c(0,0,0), seasonal = list(order=c(0,1,0)))
```

```
b2<-arima(strip, order = c(0,0,0), seasonal = list(order=c(0,1,1)))
```

```
b3<-arima(strip, order = c(0,0,0), seasonal = list(order=c(0,1,2)))
```

```
b4<-arima(strip, order = c(0,0,0), seasonal = list(order=c(1,1,0)))
```

```
b5<-arima(strip, order = c(0,0,0), seasonal = list(order=c(2,1,0)))
```

```
b6<-arima(strip, order = c(0,0,0), seasonal = list(order=c(1,1,1)))
```

```
b7<-arima(strip, order = c(0,0,0), seasonal = list(order=c(1,1,2)))
```

```
b8<-arima(strip, order = c(0,0,0), seasonal = list(order=c(2,1,1)))
```

```
b9<-arima(strip, order = c(0,0,0), seasonal = list(order=c(2,1,2)))
```

```
b10<-arima(strip, order = c(0,0,1), seasonal = list(order=c(0,1,0)))
```

```
b11<-arima(strip, order = c(0,0,1), seasonal = list(order=c(0,1,1)))
```

```
b12<-arima(strip, order = c(0,0,1), seasonal = list(order=c(0,1,2)))
```

```
b13<-arima(strip, order = c(0,0,1), seasonal = list(order=c(1,1,0)))
```

```
b14<-arima(strip, order = c(0,0,1), seasonal = list(order=c(2,1,0)))
```

```
b15<-arima(strip, order = c(0,0,1), seasonal = list(order=c(1,1,1)))
```

```
b16<-arima(strip, order = c(0,0,1), seasonal = list(order=c(1,1,2)))
```

```
b17<-arima(strip, order = c(0,0,1), seasonal = list(order=c(2,1,1)))
```

```
b18<-arima(strip, order = c(0,0,1), seasonal = list(order=c(2,1,2)))
```

```
b19<-arima(strip, order = c(0,0,2), seasonal = list(order=c(0,1,0)))
```

```
b20<-arima(strip, order = c(0,0,2), seasonal = list(order=c(0,1,1)))
```

```
b21<-arima(strip, order = c(0,0,2), seasonal = list(order=c(0,1,2)))
```

```
b22<-arima(strip, order = c(0,0,2), seasonal = list(order=c(1,1,0)))
```

```
b23<-arima(strip, order = c(0,0,2), seasonal = list(order=c(2,1,0)))
```

```
b24<-arima(strip, order = c(0,0,2), seasonal = list(order=c(1,1,1)))
```

```
b25<-arima(strip, order = c(0,0,2), seasonal = list(order=c(1,1,2)))
```

```
b26<-arima(strip, order = c(0,0,2), seasonal = list(order=c(2,1,1)))
```

```
b27<-arima(strip, order = c(0,0,2), seasonal = list(order=c(2,1,2)))
```

```
AIC(b1,b2,b3,b4,b5,b6,b8,b9,b10,b11,b12,b13,b14,b15,b17,b18,b19,b20,b21,b22,  
,b23,b24,b25)
```

```
# auto.arima
```

```
fit<-auto.arima(ER, d = NA, D = NA, max.p = 5, max.q = 5,  
  max.P = 2, max.Q = 2, max.order = 5,  
  start.p=2, start.q=2, start.P=2, start.Q=2,  
  stationary = FALSE, ic = c("aic"),  
  stepwise=TRUE, trace=FALSE,  
  approximation=length(ER)>100 | frequency(ER)>12, xreg=NULL,  
  test=c("kpss","adf","pp"), allowdrift=TRUE)
```

```
plot(forecast(fit,h=20))  
c(fit)
```

```
?auto.arima  
fit<-auto.arima(ER)  
fit <- auto.arima(WWWusage)  
plot(forecast(fit,h=20))
```

```
summary(fit)
```

```
fit1<-auto.arima(trip, d = NA, D = NA, max.p = 5, max.q = 5,  
                 max.P = 2, max.Q = 2, max.order = 5,  
                 start.p=2, start.q=2, start.P=2, start.Q=2,  
                 stationary = FALSE, ic = c("aic"),  
                 stepwise=TRUE, trace=FALSE,  
                 approximation=length(trip)>100 | frequency(trip)>12, xreg=NULL,  
                 test=c("kpss","adf","pp"), allowdrift=TRUE)
```

```
plot(forecast(fit1,h=20))
```

```
f<-arima(dER,order=c(0,1,1))  
tsdiag(f) #produces a diagnostic plot of a fitted time series model
```

```
?Box.test  
Box.test(f$residuals, lag=6)  
Box.test(f$residuals, lag=12)  
Box.test(f$residuals, lag=18)  
Box.test(f$residuals, lag=24)  
Box.test(f$residuals, lag=30)  
Box.test(f$residuals, lag=36)
```

```
st<-arima(strip, order = c(1,0,1), seasonal = list(order=c(0,1,2), period=12))  
tsdiag(st)
```

```
?arima
```

```
Box.test(st$residuals,lag=12)
```

```
cor.test(ER,trip, alternative="less")
```

```
##DF-test : 정상인지 비정상인지 ###
```

```
adf.test(ER)
```

```
adf.test(trip)
```