Box-Jenkins 豫測技法 紹介

한국 과학 기술원 경영과학과(工博) 朴 成 桂
한국 과학 기술원 경 영 과 학 과 全 泰 俊

1. 서 론

계획은 경영 및 관리에서 가장 중요한 부분이며 이러한 계획의 수립에 있어서 첫째로 필요한 기능은 미래 상황을올 바로 예측하는 것이다. 한 회사의 수요 예측은 생산·판매 등의 계획을 수립하고 여러 대안들을 선택하고 평가하는데 가장 기본이 되며 국가 경기변동의 정확한 예측은 국가 경제 정책을 성공적으로 이끄는 첩경이라 할 수 있다.

예측에는 그 예측 기간에 따라 장기 및 단기 예측(long and short range forecasting)으로 구분되며 현대와 같이 모든 것이 급격히 변화하는 상황에서는 단기 예측의 중요성이 크다고 하겠다. 특히 예측치에 영향을 미치는 변수를 찾기 어렵거나 변수를 찾았더라도 그 상관 관계를 밝히기 어려운 경우에 더욱 정확한 예측치를 구할 수 있는 것이 단기 예측 기법이다.

종래의 단기 예측 방법으로는 직관에 의하거나 외국의 비슷한 상황을 거친 기관의 실측치를 그대로 미래의 예측치로 사용한 경우가 많았으며 이는 주로 변수에 관한 정보가 부족하거나 너무 많은 변수가 영향을 미치고 있기 때문이었다. 이러한 경우에 정확한 예측치를 구하기 위한 과학적인 방법이 시계열 분석(Time Series Analysis)이다. 시계열 분석법은 예측하고자 하는 변수의 시간에 따른 형태에 대한 분석을 통해서 예측치를 구하는 방법으로 그 기본 가정은 과거 시계열의 형태가 그 특성을 잃지 않고 미래에도 반복된다는 것이다.

이중 Box-Jenkins 시계열 분석법은 이산적 혹은 연속적 시계열(discrete or continuous time series)을 모형화하고 예측치를 구하는데 사용하며 자동회귀모형(Autoregressive model)과 이동평균 모형(moving average model)을 일관성 있게 통합함으로써 모든 시계열을 모형화 할 수 있을 뿐만 아니라 모형에 관계된 계수의 수의 최소화할 수 있게 하였다.

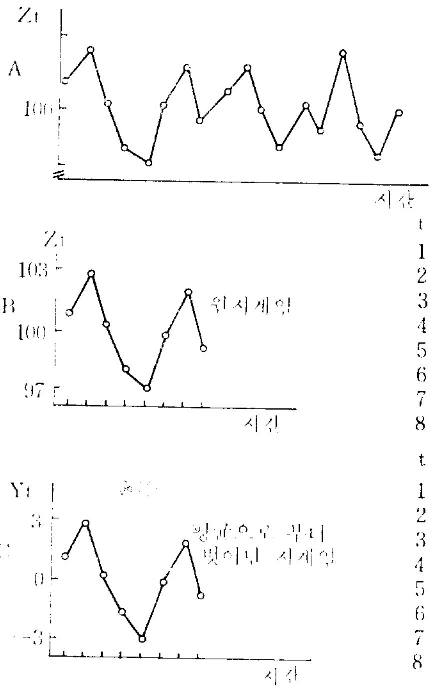
Box-Jenkins 방법의 특정은 일반적인 통계적인 예측 방법이 모형을 세우기 위해서 도표, 지수 등을 분석한 후 경험에 의해서 특정한 모형을 결정해야 하는 데 비해서 Box-Jenkins 방법은 모형을 결정하지 않고 최소의 가능한 모형으로 부터 시작하여 부적당한 부분을 제거시켜 나가는 과정에서 최적 모형에 이르게 하는 점에 있다. 다시 말하면 Box-Jenkins 방법은 경험과 직관을 최소화 함으로서 모형 확립을 위한 시행착오 과정을 최대한으로 제거한 합리적인 접근 방법이라 하겠다.

2. Box-Jenkins 모형

시계열이란 특정 변수의 시간에 의한 일련의 관측치를 말하며 시계열 분석이란 시간에 따른 관측치의 상호 관계를 밝혀 모형화 함으로서 미래 시계열을 예측하고자 하는 체계적인 방법을 뜻한다. 시계열은 관측되는 상태에 따라서 연속적 시계열(continuous time series)와 이산적 시계열(discrete time series)로 구분할 수 있으며

연속적인 시계열의 경우에는 관측시간의 간격을 알맞게 선택하면 근사적인 이산적인 시계열로 볼 수 있으므로 일반적으로 분석 대상이 되는 시계열은 동일한 시간 간격에 따라 관측된 이산적 시계열에 국한한다.

또한 시계열은 특성에 따라 안정적인 시계열(stationary time series)와 불안정적 시계열(nonstationary time series)로 구분할 수 있다.



t	Z_t
1	101
2	103
3	100
4	98
5	97
6	100
7	102
8	99

t	Y_t
1	1
2	3
3	0
4	-2
5	-3
6	0
7	2
8	-1

그림 2.1 안정적 시계열

안정적 시계열은 그림 2.1과 같이 일정한 평균치를 중심으로 평행을 이루고 일정한 분산을 가지며 공분산(covariance)은 실측치간의 시간차에 의해서만 영향을 받는 상태를 말한다. 다시 말하자면 시계열을 생성시키는 과정(process)이나 매개 변수의 값들이 시간의 흐름에 따라 변하지 않는 시계열을 말한다. 시계열 Z_{t+j}, $j=0, 1, \cdots, N$의 분포 함수를 $P(Z_t, Z_{t+1}, \cdots, Z_{t+N})$이라 하면 안정적 시계열에서는 (2-1)식이 성립한다.

$$P(Z_t, Z_{t+1}, \cdots, Z_{t+N}) = P(Z_{t+M}, Z_{t+M+1}, \cdots, Z_{t+M+N}) \cdots\cdots (2\text{-}1)$$

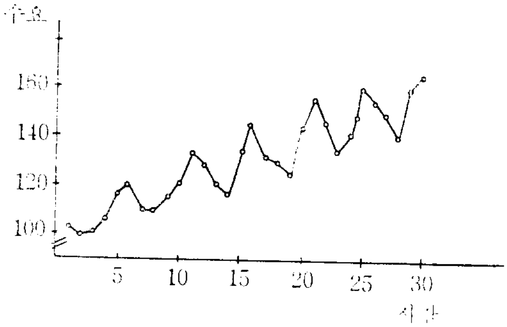


그림 2.2 불안정적 시계열

불안정적 시계열은 그림 2.2와 같이 시간 경과에 따라 평균이나 분산이 변하는 것으로서 예를 들어 연간 계절 수요 형태는 연중 평균 수요 비율이 변하기 때문에 불안정적 시계열로 볼 수 있다. 이러한 경우에는 관측치간의 계산차이(differencing)을 취하거나 변환(transformation)을 시켜서 안정적인 시계열로의 변환이 필요하다.

(1) 안정적 시계열 모형

안정적 시계열에서 일단 예측치(Z_t)를 세우면 예측오차($Z_t-\hat{Z}_t$)는 a_t로 정의하며, 이는 평균이 0이고 분산 σ^2을 가지며 독립적으로 정규분포를 따르는 순수오차(white noise)이라고 가정한다.

1) 자동회귀 모형(Autoregressive Model: AR Model)

자동회귀 모형에서 시계열의 현재 값은 현재 관측치를 설명하여 주는 시계열의 이전 값들과 설명하지 못하는 부분 a_t의 선형 결합으로써 표시할 수 있다.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t \qquad \cdots\cdots\cdots\cdots\cdots\cdots (2.2)$$

여기에서 Y_t는 관측치와 평균치 간의 차($Z_t-\mu$)를 나타내며 ϕ_p는 p번째 전의 기간에서의 가중치를 나타낸다. 방정식(2.2)는 차수가 p인 자동회귀 과정 $AR(p)$ 모형이라 불리는데 그 이유는 회귀 분석에서 종속변수 Y_t와 독립변수 $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}$와의 관계를 짓기 때문이다.

이를 역변환 연산(backshift operator: B)을 사용하여 표현하면

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)Z_t=a_t.$$

여기서 $B^m Z_t=Z_{t-m}$ 이다.

2) 이동 평균 모형(Moeing Average Model; MA Model)

이동 평균 모형은 시계열의 현재값을 이전의 예측오차의 선형결합으로 표시할 수 있다는 가정에 근거를 둔 것이다.

$$Y_t=a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}-\cdots-\theta_q a_{t-q} \cdots(2.3)$$

이를 q 차의 이동평균 모형 $MA(q)$이라고 부른다. 여기서 θ_q는 q 번째 이전의 예측 오차에 대한 가중치를 나타낸다. 이 모형은 지수 평활 모형(exponential smoothing model)의 일반형으로 볼 수 있으며 가중치의 특성에 의해서 그 차이가 구분된다. 즉 지수 평활 모형에서는 과거 자료보다 최근 자료가 현재의 관측치에 더 큰 영향을 준다고 보는데 반하여 이동 평균 모형에서는 시계열의 중심경향이 시간의 경과와 더불어 변한다는 데 있다.

$MA(q)$ 모형은 다음과 같이 표현될 수 있다.

$$Y_t=(1-\theta_1 B-\cdots-\theta_2 B_q)\ a_t$$

3) 자동회귀—이동 평균 모형(ARMA Model)

자동회귀 모형과 이동 평균 모형을 확장시켜서 두개의 모형을 결합할 경우, 이를 차수 p, q의 자동회귀—이동 평균 모형 또는 ARMA(p, q) 모형이라고 하며 식 (2.4)와 같이 정의된다.

$$\phi_p(B)Y_t=\theta_q(B)a_t \cdots\cdots\cdots\cdots(2.4)$$

$\phi_p(B)$와 $\theta_q(B)$는 차수가 $p.q$인 역변환 연산의 다항식이다.

$$\phi_p(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$$

$$\theta_q(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$$

(2) 불안정적 시계열 모형(ARIMA Model)

실제로 당면하게 되는 대부분의 시계열은 불안정적 시계열이다. 즉 $\phi(B)=0$의 근(root)이 단위원 상이나 그 내부에 존재하는 경우이다. 근이 단위원 내부에 존재하는 경우에는 Z_t가 무한대로 급속히 증가하는 폭발적 불안정성(explosive nonstationary)을 나타내므로 고려 대상에서 제외한다. 근이 단위원 상에 존재하는 경우에는 그 불안정적 시계열은 특정 수준에서 자유로이 이탈하여 변화하지만 부분부분들은 본질적으로 동일한 형태로 움직이는 성질 즉 동질성(homogeneity)을 가진다.

이러한 시계열은 관측치의 차이계산을 통해 안정적인 시계열로 변환시킬 수 있다. 동질성을 가지는 불안정적 시계열 모형은 다음과 같이 정의된다.

$$\phi_p(B)\nabla^d Y_t=\theta_q(B)a_t \cdots\cdots\cdots\cdots(2.5)$$

여기서 $\nabla(=1-B)$는 차이 변환 연산(difference operator)이다. 식 (2.5)는 차수가 p, d, p인 혼합 자동회귀·이동 평균 모형(Autoregressive Integrated Moving Average) 또는 ARIMA(p, d, q)라 불린다.

불안정적 시계열 중에서 시계열의 상대적 변화가 동질적인 불안정성을 보이는 경우나 시계열의 수준이 증가함에 따라 분산이 증가하는 경우에는 다음과 같은 변환(log transform)을 통하여 안정적인 시계열로 바꿀 수 있다.

$$Y_t=\begin{cases}\dfrac{(Z_t+m)^\lambda-1}{\lambda} & (\lambda\neq 0)\\ \log\ Z_t & (\lambda=0)\end{cases} \cdots\cdots\cdots(2.6)$$

이러한 상황에서, 원래 자료에 자연 대수(log transform)을 취하여 동질성을 유도할 수 있고 대수를 취한 값들에 계산차이를 취하여 모형을 확립할 수 있다.

(3) 계절적 시계열 모형(Seasonal Model)

안정적 시계열 및 불안정적 시계열을 나타내는 데 유용한 모형들을 확장시켜서 계절주기 s인 시계열에 대한 일반적인 모형은 다음과 같이 정의된다.

$$\Phi_P(B^s)\nabla_s{}^D Z_t=\Theta_Q(B^s)\varepsilon_t$$

여기서 $\nabla_s{}^D=(1-B_s)^D$

$$\Phi_P(B^s)=1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps}$$
$$\Theta_Q(B^s)=1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs} \quad (2.7)$$

이를 ARIMA$(P, D, Q)_s$ 모형이라고 부른다. ε_t에는 계절내의 관측치들의 관계가 포함되어 있으므로 이것을 추가하면

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t=\theta_q(B)\Theta_Q(B)a_t \quad (2.8)$$

식 (2.8)은 차수 $(p, d, q)\times(P, D, Q)_s$인 곱의 계절적 모형(multiplicative seasonal model)이라고 불린다.

3. Box-Jenkins 모형화 절차

시계열의 모형화는 모형의 확립(model identification), 매개변수의 추정(parameter estimation), 모형의 적합성 검토(Diagnostic checking)의 3단계의 절차를 반복적으로 수행하며 그림 3.1은 그 모형화 절차를 그림으로 표현한 것이다.

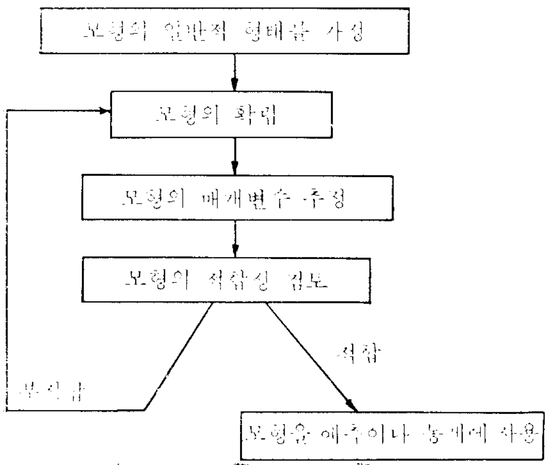


그림 **3.1** Box-Jenkins 모형화 절차

(1) 모형의 확립

모형의 확립의 과정이란 이론적인 ARIMA 모형의 특성을 이용하여 잠정적인 모형을 정하는 것으로 다음의 두 단계가 필요하다.

가. 불안정적 시계열을 안정적 시계열로 변환시키기 위해 차이 계산의 차수 결정

나. 자동 회귀의 자수 p와 이동 평균의 차수 q를 정하여 적합한 ARIMA 모형의 선택

이론적인 ARIMA 모형의 특징을 가타내는 것으로는 다음에 설명할 자동 상관 관계 함수(Autocorrelation Function, ACF), 편자동 상관관계 함수(Partial Antocorrelation Function, PACF)가 있다. 자동 상관 관계는 시계열 $\{Z_t\}$에서 관측치 상호간의 연관성을 나타내며 다음과 같이 정의된다.

$$r_k=\frac{\sum_{t=1}^{n-k}(z_t-\bar{z})(z_{t+k}-\bar{z})}{\sum_{t=1}^{n}(z_t-\bar{z})^2} \quad k=0, 1, 2, \cdots, K \quad (3.1)$$

여기서 r_k는 시간차 k를 독립변수로 하는 자동 상관관계 함수가 된다. 실질적으로 자동 상관 함수의 유용한 추정치를 얻기 위해 적어도 50개의 관측치가 필요하며 추정한 자동 상관 관계 $r_k(k=0, 1, 2, \cdots, K)$의 수효는 관측치 갯수의 1/4 즉 $n/4$ 보다 크지 않아야 한다.

편 자동 상관 관계는 ϕ_{kk}로 표시하는데 이는 k차의 자동 회귀 과정에서 j번째 계수를 뜻하며 이는 Yule-Walker 방정식이라 불리우는 식 (3.2)를 만족한다.

$$\begin{pmatrix} 1 & r_1 & r_2 \cdots\cdots r_{k-1} \\ r_1 & 1 & r_1 \cdots\cdots r_{k-2} \\ \vdots & & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} \cdots 1 \end{pmatrix}\begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix}=\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix} \cdots (3.2)$$

이로부터 $\phi_{11}=r_1$

$$\phi_{22}=\frac{\begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 \\ r_1 & 1 \end{vmatrix}}=\frac{r_2-r_1^2}{1-r_1^2}$$

$$\phi_{33}=\frac{\begin{vmatrix} 1 & r_1 & r_1 \\ r_1 & 1 & r_2 \\ r_2 & r_1 & r_3 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{vmatrix}}$$

이와 같은 방법으로 ϕ_{kk}를 풀 수 있는데 이를 k를 독립변수로 하는 편 자동 상관 관계 함수라고 한다. 자동 상관 관계 함수와 편 자동 상관 관계 함수는 모형에 따라서 독특한 형태를 가지고 있으며 표 3.1은 그것을 요약해 놓은 것이다.

표 3.1 모형에 따른 자동 및 편 자동 상관관계 함수의 형태

모형 \ 함수	자동상관관계함수	편 자동 상관 관계 함수
불안정적 시계열 모형	k에 따라서 빨리 감소하지 않는다.	
안정적 AR(P)모형	지수함수나 sine 함수 형태	p번째에서 cut off가 일어남
안정적 MA(q)모형	q번째에서 cut off가 일어남	지수 함수나 sine 함수 형태
안정적 ARMP(pq) 모형	p-q 번째 이후에서 지수 함수나 sine 함수 형태	p-q 번째 이후에서 지수 함수나 sine 함수 형태

이상과 같은 이론적인 모형들의 형태를 이용하여 모형을 정립하기 위해서는 먼저 표본 자동 상관 관계 함수와 표본 편 자동 상관 관계 함수를 구하고 이것이 이론적인 형태와 같은 모형을 잠정적인 모형으로 선택하는 것이다. 불안정적 시계열 모형인 경우에는 계산 차이나 자연 대수 변환을 통하여 안정적 시계열로 바꾼 후에 형태를 비교한다. 표 3.2에서 실제 모형 확립 과정의 예를 보였다.

(2) 매개 변수의 추정

일단 잠정적인 모형이 확립되면 다음 단계는 최소 자승법을 사용하여 미지의 계수를 추정하는 것이다. 일반적으로 ARIMA 모형에서 계수를 추정할 때 다음의 식을 최소화하는 ϕ와 θ의 값을 구한다.

$$S(\hat{\phi},\hat{\theta})=\sum_t (z_t-\hat{z}_t)^2$$

여기에서 $z_t-\hat{z}_t$는 실제 관측치와 추정한 관측치 간의 차이인 오차(residual error)를 타나내며 $\hat{\phi}$와 $\hat{\theta}$는 각기 자동회귀 모형과 이동 평균 모형의 계수를 나타낸다. 비선형 최소 자승 계산법(nonlinear least square algorithm)은 최소값에 이를 때까지 $S(\phi,\theta)$로 정의된 반응 표면(response surface)을 찾는 것이다. 이때 $\hat{\phi}$, $\hat{\theta}$의 계수의 초기값 결정이 반응 표면을 찾는데 중요하다. p와 q가 2보다 작은 경우에는 γ_k와 $\hat{\phi}$, $\hat{\theta}$ 간의 관계를 이용하여 계산된 도표를 사용하여 초기치를 구할 수 있으며 2보다 큰 경우에는 여러개의 다른 초기값을 가지고 시도해 볼 수 있다.

(3) 모형의 적합성 검토

Box-Jenkins 방법의 최종 단계는 앞에서 세운 모형이 적합한가를 검토하는 것이다. 모형 검토는 예측 오차 수열 $\hat{a}_t(=z_t-\hat{z}_t)$의 표본 자동 상관 관계를 조사하며 모형이 적합할 때 a_t는 0주위에서 무작위 분포를 이룬다. 이러한 모형의 적합성을 검토하기 위해 다음과 같이 Chi-Square 검정을 한다.

$$Q=n\sum_{k=1}^{K} r_k^2(\hat{a})$$

여기서, n=전체 관측 갯수—최대 back order 수

K=계산한 표본 자동 상관 관계의 수

$r_k(\hat{a})$=시간차 k에서 $\hat{a}_t$ 시계열의 표본 자동 상관 관계

검정 통계량 Q의 값이 $\chi^2(K-p-q)$보다 클 때이는 현재 모형이 부적당하며 모형을 개선하여야 함을 말해 준다. 반면에 Q의 값이 작은 때는 최종 모형을 예측 및 통제에 사용할 수 있다.

(4) 예 측

시계열의 정확한 모형을 일단 세우면, 이러한 모형으로 부터 지금까지의 자료와 예측기간 T를 사용하여 예측 방정식을 세울 수 있다. 예를 들어 다음과 같이 모형을 세우고 계수를 추정하였다면

$$(1-0.5B-0.4B^2)(1-B)Z_t=(1-0.25B)a_t$$

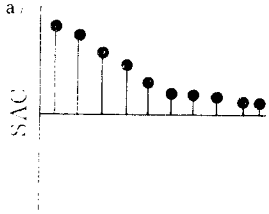
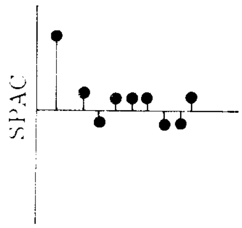
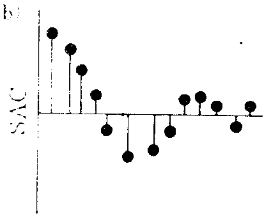
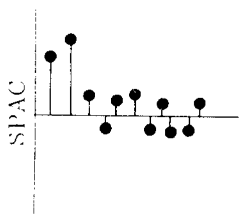
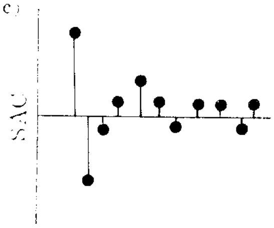
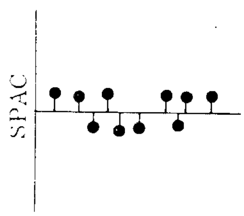
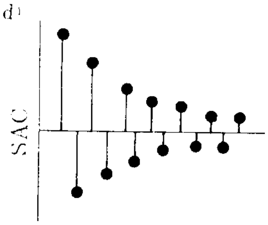
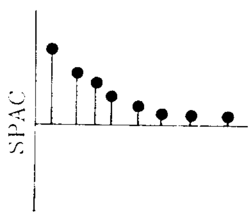
현재의 시점 t에서 앞으로 l번째 기간을 예측할 때 예측 방정식은 다음과 같이 된다.

$$\hat{Z}_t(l)=1.5[Z_{t+l-1}]-0.2[Z_{t+l-2}]-0.3[Z_{t+l-3}]+[a_{t+l}]-0.25[a_{t+l-1}]$$

여기서 $[Z_{t+k}]=\begin{cases}\hat{Z}_t(k), & k>0\\ Z_t, & k\le 0.\end{cases}$

$[a_{t+k}]=\begin{cases}0, & k>0\\ a_{t+k}, & k\le 0.\end{cases}$

표 3.2 모형 확립 과정의 예

표본자동 상관관계	편자동 상관관계	관계	잠정적모형
a) 		ARMA(1,0)	$(1-\phi_1 B)Y_t = a_t$
b) 		ARMA(2,0)	$(1-\phi_1 B-\phi_2 B^2)Y_t = a_t$
c) 		ARMA(0,2)	$Y_t = (1-\theta_1 B-\theta_2 B^2)a_t$
d) 		ARMA(? ?) 잠정적 ARMA(1,0)	$(1-\theta_1 B)Y_t = a_t$

4. 전이 함수 모형

(1) 전이 함수 모형

그림 4.1에서 보는 바와 같이 동적인 시스템으로부터, 입력 X와 출력 Y를 동일한 시간 간격으로 뽑아 관측치의 쌍(X_t, Y_t)를 얻을 수 있다고 할 때, 전이 함수 모형은 입력 시계열 X_t와 출력 시계열 Y_t와의 관계를 나타내는 것으로 이 모형을 세우기 위해서는 단일 변량 시계열 모형에서와 같은 방법을 사용할 수 있다. 즉 모수를 추정하고, 모형의 적합성을 검토하는 과정을 만족스러운 모형을 얻을 때까지 반복하며 이러한 모형을 예측시에 적용시킨다.

이산적인 동적 시스템은 다음과 같은 일반적인 선형 차등 방정식으로 표시할 수 있다.

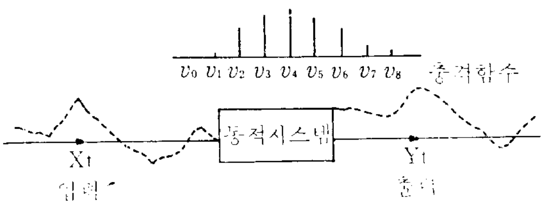


그림 4.1 동적 시스템

$$(1+\xi_1\nabla+\cdots+\xi_q\nabla^r)Y_t$$
$$=g(1+\eta_1\nabla+\cdots+\eta_s\nabla^s)X_{t-b}\cdots\cdots\cdots\cdots(4.1)$$

(4.1)식은 차수(r, s)의 전이함수 모형이라 불리우며 역변환 연산 $B=1-\nabla$에 의해 다음과 같이 표시할 수 있다.

$$(1-\delta_1B+\cdots-\delta_rB^r)Y_t$$
$$=(w_o-w_1B-\cdots-w_sB^s)X_{t-b}\quad\cdots\cdots\cdots(4.2)$$

또는 $\delta(B)Y_t=w(B)X_{t-b}$

여기서 $Y_t=\delta^{-1}(B)w(B)B^bX_t$ 이므로 $v(B)=\delta^{-1}(B)\ w(B)B^b$ 를 전이함수라 한다.

$\delta(B)v(B)=w(B)B^b$ 이므로

$$\begin{cases} v_j=0 & j<b \\ v_j=\delta_1v_{j-1}+\delta_2v_{j-2}+\cdots+\delta_rv_{j-r}+w_o & j=b \\ v_j=\delta_1v_{j-1}+\delta_2v_{j-2}+\cdots+\delta_rv_{j-r}-w_{j-b} & \\ \qquad\qquad j=b+1,\ b+2,\ \cdots,\ b+s & \\ v_j=v_{j-1}+\delta_2v_{j-2}+\cdots+\delta_rv_{j-r} & j>b+s \end{cases}$$

여기서 v_j는 충격함수(impulse function)의 가중치를 나타낸다. 계단함수(step function)의 가중치 v_j의 함수를 $V(B)$로 표시하면

$$V(B)=V_o+V_1B+V_2B^2+\cdots$$
$$=v_o+(v_o+v_1)B+(v_o+v_1+v_2)B^2+\cdots$$

$$v(B)=(1-B)V(B)$$

이 된다.

현실적으로 시스템이 잡음에 의해 영향받는 경우가 있는데, 이는 전이 함수 모형에 의해 예측된 출력을 N_t만큼 변화시킨다. 이때 잡음 N_t가 ARIMA(p, d, q) 과정으로 가정한다면 결합 전이함수 —잡음 모형은 다음과 같이 표시할 수 있다.

$$Y_t=\delta^{-1}(B)w(B)X_{t-b}+\phi^{-1}(B)\theta(B)a_t\quad(4.3)$$

전이함수 모형의 식별에 기본이 되는 도구로는 충격 함수와 계단 함수 외에 자동 상관 관계 함수, 편자동 상관 관계 함수와 Cross 상관관계 함수 등이 있으며 이러한 함수들은 모형 식별 과정에서 여러 가지의 계산 차이를 한 시계열에 대해 계산되어 진다. 자동 상관관계와 편자동상관 관계는 단일 변량 시계열의 경우와 같으며 시간차 k에서 Cross 상관관계 함수는 시계열에서 입력 x_t와 출력 y_t간의 상관관계를 나타내는 것으로 다음과 같이 계산할 수 있다.

$$r_{xy}(k)=\frac{c_{xy}(k)}{s_xs_y}=\frac{c_{xy}(k)}{\sqrt{c_{xx}(0)}\sqrt{c_{yy}(0)}}\cdots\cdots(4.4)$$

여기서 $c_{xy}(k)$는 시간차 k에서 Cross공분산계수의 추정치를 나타낸다. 즉

$$c_{xy}(k)=\begin{cases}\frac{1}{n}\sum_{t=1}^{n-k}(x_t-\bar{x})(y_{t+k}-\bar{y}) & k=0,1,2,\cdots \\ \frac{1}{n}\sum_{t=1}^{n-k}(y_t-\bar{y})(x_{t+k}-\bar{x}) & k=0,-1,-2,\cdots\end{cases}$$
$$\cdots\cdots\cdots\cdots(4.5)$$

다른 상관관계와 마찬가지로 Cross상관관계 함수도 유용한 추정치를 얻기 위해서는 적어도 50쌍의 관측치가 필요하다.

(2) 외생 변수(Intervention of exogeneons Vairable)의 영향이 있는 시계열 모형

불안정적 시계열의 수준변화는 잡음 모형(Noise model)만으로는 그 형태를 알 수 없고 외생변수, 예를 들면 법률의 공포, 상품의 선전 판매 가격의 변화와 같은 외생 변수의 영향에 의한 것일 수도 있다. 시계열 Z_t에 외생 변수의 영향이 있을때 식 (4.6)과 같이 나타낼 수 있다.

$$Y_t=\sum_{j=1}^{k}\frac{w_j(B)}{\delta_j(B)}\xi_{jt}+N_t(\phi,\theta) \cdots\cdots\cdots\cdots (4.6)$$

여기서, Y_t : 관측치 Z_t를 변환시킨 시계열

ξ_{jt} : 시간 t에서 j번째 외생 변수의 영향 유무를 나타내는 변수.

$N_t(\phi,\theta)$: 외생변수의 영향을 받지 않는 ARIMA 모형

5. Box-Jenkins Package의 사용

(1) Box-Jenkins Package 개요

Box-Jenkins 시계열 분석법을 근거로 컴퓨터 프로그램화한 Package가 다양하게 개발되었는데 여기서는 그중에서 KAIST 전산개발 센터에 설치된 Box-Jenkins Package를 소개하겠다. Package에는 단일 변량 시계열 모형과 단일입력 전이 함수 모형에 관한 것이 함께 수록되어 있으며 4개의 주로 프로그램(iDEN 1, iDEN 2, iDEN 3, iDEN 4)과 40개의 Subroutine으로 구성되어 있다. 주 프로그램의 기능을 소개하면

- iDEN1 : 이 주 프로그램은 단일 변량 시계열 모형(Univaniate Time Series Model)의 모형 확립을 위한 것으로 시계열의 각 계산차이에 대한 자동 상관관계 함수(ACF)와 편자동 상관관계 함수(PACF) 및 chi-square 값을 계산하고 plot 한다.
- iDEN2 : 단일 변량 시계열 모형의 매개변수를 추정하고, 모형의 적합성을 검토하며 미래의 예측을 수행한다.
- iDEN3 : 이 주 프로그램은 단일 입력 전이함수 모형(Single input transfer function model)을 세우기 위한 것으로 시계열의 각 계산 차이에 대한 자동 상관 관계 함수(ACF), 편 자동 상관관계 함수(PACF), cross 상관관계 함수 및 chi-square 값을 계산하고 plot 한다.
- iDEN4 : 전이함수 모형의 매개 변수를 추정하고 모형의 적합성을 검토하며 예측을 수행한다.

따라서 일반 사용자가 package를 사용할 때

5,000	5,657	6,132	7,411
4,965	6,010	6,111	7,233
4,496	6,109	5,948	6,958
4,491	6,052	6,056	6,960
4,455	6,391	6,342	6,927
4,585	6,798	6,626	6,814
4,724	6,740	6,591	6,757
4,951	6,778	6,302	6,765
4,917	7,005	6,321	6,870
4,888	7,045	5,837	6,954
5,087	7,279	5,572	6,551
5,082	7,367	5,744	6,022
5,039	6,934	6,005	5,975
5,054	6,506	6,239	6,052
4,940	6,374	6,523	6,033
4,913	6,066	6,652	6,040
4,871	6,102	6,585	5,944
4,901	6,204	6,622	5,543
4,864	6,138	6,754	5,416
4,750	5,938	6,712	5,571
4,856	5,781	6,675	5,571
4,959	5,813	6,882	5,627
5,004	5,811	7,011	5,679
5,415	5,818	7,140	5,455
5,550	5,982	7,197	5,443

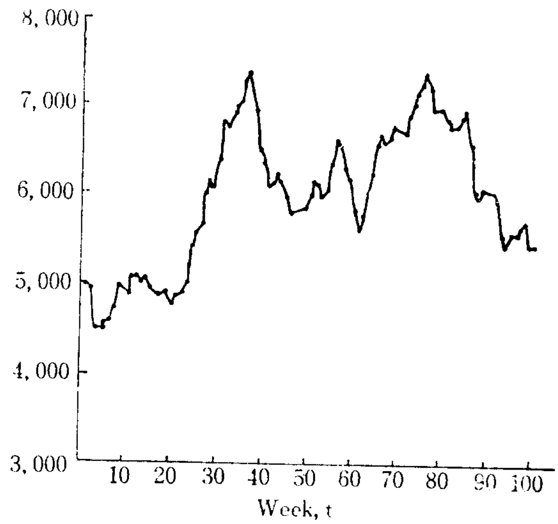
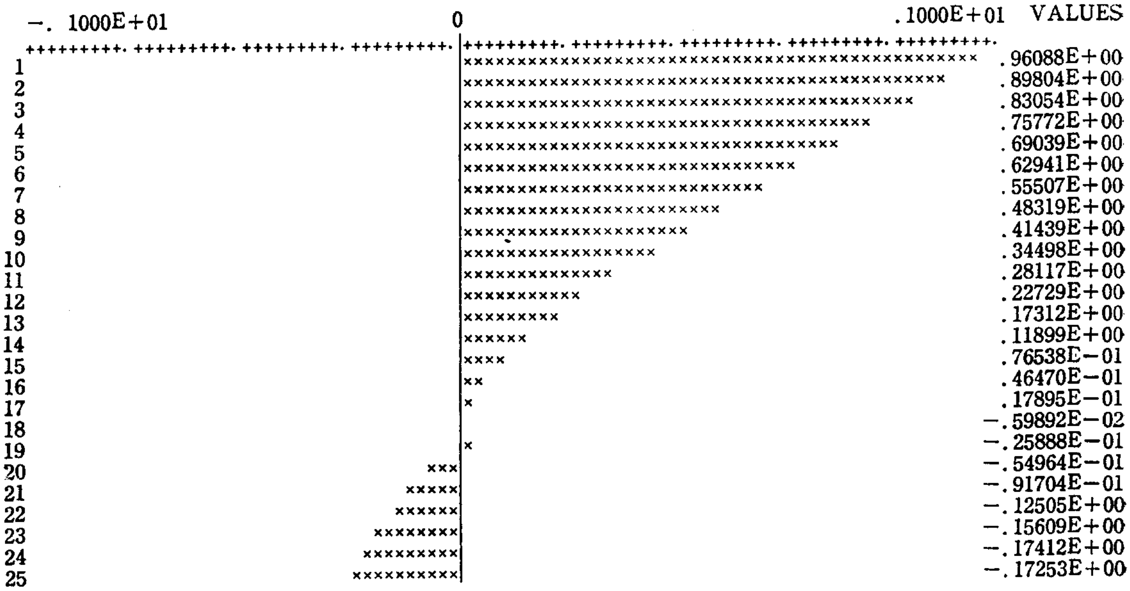


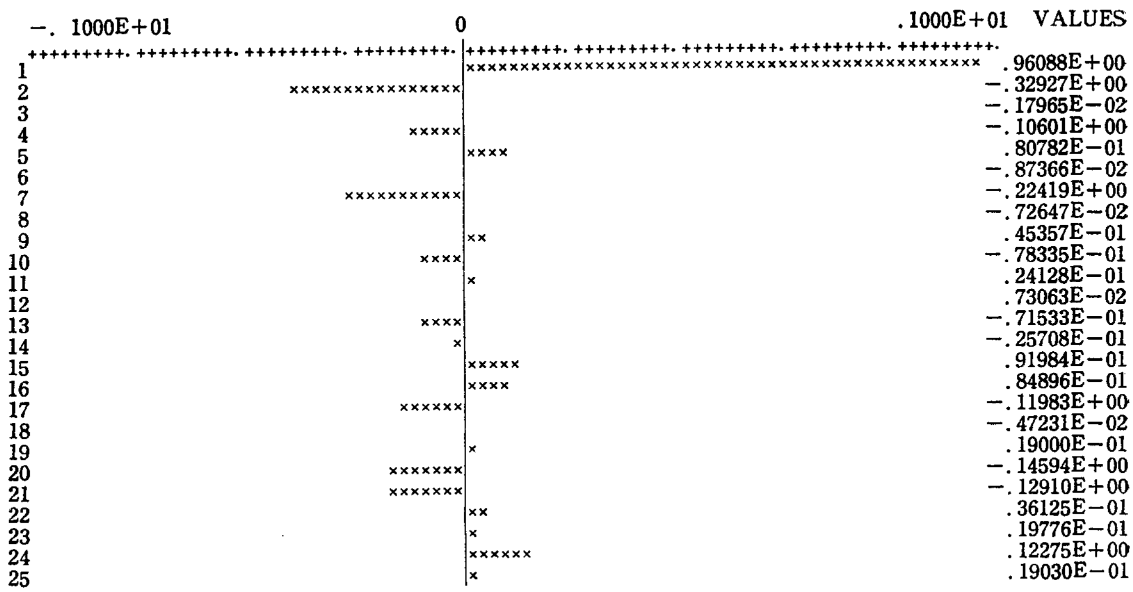
그림 5.1 Plastic Container의 주별 수요

그림 5.2 iDEN1의 결과

WEEKLY DEMAND FOR PLASTIC CONTAINER GRAPH
GRAPH OF OBSERVED SERIES ACF
GRAPH INTERVAL IS .2000E-01



WEEKLY DEMAND FOR PLASTIC CONTAINER
GRAPH OF OBSERVED SERIES PACF
GRAPH INTERVAL IS .2000E-01



WEEKLY DEMAND FOR PLASTIC CONTAINER
GRAPH OF DIFFERENCE 1 ACF
GRAPH INTERVAL IS .2000E−01

```
   −. 1000E+01                                0                                                    .1000E+01   VALUES
   +++++++++. +++++++++. +++++++++. +++++++++.|+++++++++. +++++++++. +++++++++. +++++++++. +++++++++.
 1                                            |xxxxxxxxxxxxxxxxxxxx                                  .40506E+00
 2                                           x|                                                     −.18099E−01
 3                                            |xxxx                                                  .81600E−01
 4                                        xxxx|                                                     −.84616E−01
 5                                       xxxxx|                                                     −.10641E+00
 6                                            |xxxxxxxxxx                                            .19475E+00
 7                                            |xxxxxxx                                               .14519E+00
 8                                         xxx|                                                     −.54426E−01
 9                                            |x                                                     .14114E−01
10                                          xx|x                                                    −.41956E−01
11                                   xxxxxxxxx|                                                     −.17293E+00
12                                           x|                                                     −.20609E−01
13                                            |                                                      .97093E−02
14                                   xxxxxxxxx|                                                     −.15243E+00
15                                    xxxxxxxx|                                                     −.15992E+00
16                                         xxx|                                                     −.50218E−01
17                                       xxxxx|                                                     −.10265E+00
18                                        xxxx|                                                     −.78501E−01
19                                            |xxxxxxx                                               .13384E+00
20                                            |xxxxxxxxx                                             .17852E+00
21                                            |x                                                    −.27533F−01
22                                          xx|                                                     −.30438E−01
23                                         xxx|                                                     −.69993E−01
24                                  xxxxxxxxxx|                                                     −.19160E+00
25                                          xx|                                                     −.36561E−01
```

WEEKLY DEMAND FOR PLASTIC CONTAINER
GRAPH OF DIFFERENCE 1 PACF
GRAPH INTERVAL IS .2000E−01

```
   −. 1000E+01                                0                                                    .1000E+01   VALUES
   +++++++++. +++++++++. +++++++++. +++++++++.|+++++++++. +++++++++. +++++++++. +++++++++. +++++++++.
 1                                            |xxxxxxxxxxxxxxxxxxxx                                  .40506E+00
 2                                 xxxxxxxxxxx|                                                     −.21792E+00
 3                                            |xxxxxxxxxxx                                           .22456E+00
 4                             xxxxxxxxxxxxxxx|                                                     −.29865E+00
 5                                            |xxxxxxxx                                              .15678E+00
 6                                            |xxxxxxxxx                                             .17221E+00
 7                                         xxx|                                                     −.51958E−01
 8                                          xx|                                                     −.44952E−01
 9                                            |x                                                     .17438E−01
10                                        xxxx|                                                     −.81069E−01
11                                          xx|                                                     −.45150E−01
12                                            |xx                                                    .30718E−01
13                                       xxxxx|                                                     −.95970E−01
14                                        xxxx|                                                     −.89031E−01
15                                      xxxxxx|                                                     −.11578E+00
16                                            |xxx                                                   .59420E−01
17                                        xxxx|                                                     −.76734E−01
18                                            |                                                       13204E−02
19                                            |xxxxxxx                                               .14999E+00
20                                            |xxxxxx                                                .12168E+00
21                                    xxxxxxxx|                                                     −.14700E+00
22                                            |xxx                                                   .61230E−01
23                                   xxxxxxxxx|                                                     −.15430E−00
24                                           x|                                                     −.19599E−01
25                                            |                                                     −.90323E−02
```

에는 iDEN1 이나 iDEN3 를 사용하여 먼저 모형에 대한 가정을 한 다음에 iDEN2 나 iDEN4 를 수행하여 추정과 적합성 검정 및 예측을 수행하는 2가지 단계를 밟아야 한다.

그림 5.3 iDEN2 의 결과

```
TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .33937E+02
SHOULD BE COMPARED WITH A CHI-SGUARE VARIABLE WITH 23 DEGREES OF FREEDOM
THE ESTIMATED RESIDUALS- MODEL 1
GRAPH OF OOSERVED SERIES ACF
GRAPH INTERVAL IS .2000E-01
-. 1000E+01                     0                        .1000E+01   VALUES
 +++++++++.+++++++++.+++++++++.+++++++++.|+++++++++.+++++++++.+++++++++.+++++++++.+++++++++.
 1                                    xxx|                                  -.53761E-01
 2                                    xxx|                                  -.58278E-01
 3                                       |xxxxxxx                            .13008E+00
 4                                    xxx|                                  -.53153E-01
 5                              xxxxxxxxx|                                  -.18608E+00
 6                                       |xxxxxxxxxxxx                       .23536E+00
 7                                       |xxxx                               .82546E-01
 8                                 xxxxxx|                                  -.12465E+00
 9                                       |xxx                                .53402E-01
10                                       |xxx                                .33539E-01
11                            xxxxxxxxxxx|                                  -.21209E+00
12                                       |xxx                                .69605E-01
13                                       |xx                                 .25676E-01
14                                 xxxxxx|                                  -.10732E+00
15                                xxxxxxx|                                  -.14043E+00
16                                       |xxx                                .69108E-01
17                                  xxxxx|                                  -.10510E+00
18                                     xx|                                  -.46099E-01
19                                       |xx                                 .42527E-01
20                                       |xxxxxxxxxxx                        .21317E+00
21                               xxxxxxxx|                                  -.15024E+00
22                                       |xxx                                .51471E-01
23                                      x|                                  -.23372E-01
24                               xxxxxxxx|                                  -.15529E+00
```

SUMMARY OF MODEL 1

DATA-Z=WEEKLY DEMAND FOR PLASTIC CONIAINER   100 DESERVATIONS
DIFFERENCING ON Z-1 OF ORDER 1

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	ESTIMATED VALUE	95 PER SENT LOWER LIMIT	95 PER SENT UPPER LIMIT
1	MOVING AVERAGE 1	1	-.6765E+00	-.82553E+00	-.52751E+00

OTHER INPORMATION AND RESLUTS

RESIDUAL SUM OF SQUARES	.25170E+01	98 D.F.	RESIDUAL MEAN SQUARE	.25633E-01
NUMBER OF RESIDUALS	99		RESIDUAL STANDARD ERROR	.16026E+00

NUMBER OF TIME ORIGINS FOR FORECASTS= 1
NUMBER OF FORECASTS AT EACH TIME ORIGIN=20
FORECAST TIME ORINGINS ARE T= 100
MODEL 1 FORECASTS AT BASE PERIOD 100 WITH 95 PERCENT CONFIDENCE LIMITS

PERIODS AHEAD	LO. CONF. LIMIT	FORECAST	UP. CONF. LIMIT	ACTUAL, IF KNOWN
1	.5219501E+01	.5533547E+01	.5947592E+01	
2	.4920496E+01	.5533547E+01	.6146597E+01	
3	.4725440E+01	.5533547E+01	.6341653E+01	
4	.4569056E+01	.5533547E+01	.6498037E+01	
5	.4434707E+01	.5533547E+01	.6632386E+01	
6	.4315083E+01	.5533547E+01	.6752010E+01	
7	.4206196E+01	.5533547E+01	.6860897E+01	
8	.4105588E+01	.5533547E+01	.6961505E+01	
9	.4011617E+01	.5533547E+01	.7055476E+01	
10	.3923119E+01	.5533547E+01	.7143974E+01	
11	.3839238E+01	.5533547E+01	.7227855E+01	
12	.3759318E+01	.5533547E+01	.7307775E+01	
13	.36828465+01	.5533547E+01	.7384247E+01	
14	.3609410E+01	.5533547E+01	.7457683E+01	
15	.3539677E+01	.5533547E+01	.7528416E+01	
16	.3470367E+01	.5533547E+01	.7596726E+01	
17	.3404247E+01	.5533547E+01	.7662846E+01	
18	.3340119E+01	.5533547E+01	.7762974E+01	
19	.3277814E+01	.5533547E+01	.7789279E+01	
20	.3217184E+01	.5533547E+01	.7849909E+01	

(2) Package 사용예

Box-Jenkins Package의 사용 실례를 보이기 위해서 plastic contamer의 주별 수요 자료 100개를 이용하였다. 그림 5-1에서 보는 바와 같이 시계열이 불안정성을 보이고 있음을 알 수 있다.

그림 5.2는 iDEN1의 결과를 나타내고 있으며 원 시계열의 ACF는 빨리 줄어들지 않으므로 불안정적 시계열로 생각할 수 있다. 1차의 계산 차이에서 ACF는 안정성을 보이며 lag 1에서 cut off가 일어나고 PACF는 감소하는 sine 함수 형태를 가지므로 ARIMA(0, 1, 1) 모형을 잠정적인 모형으로 하였다.

그림 5.3은 iDEN2의 결과를 나타내고 있으며 iDEN1에서 가정한 ARIMA(0, 1, 1) 모형의 결과이다. 매개변수의 추정에서 $\theta_1=-0.68$ 이므로 이 모형은 다음과 같이 표시된다.

$$Z_t - Z_{t-1} = a_t + 0.68a_{t-1}$$

적합성 검정에서 Q값(33.94)가 $\chi^2(23, 0.05)$의 값인 35.17보다 작으므로 통과한다. 또한 이후의 20주에 대해서 예측치를 계산하였다.

6. 결　론

Box-Jenkins 시계열 분석법은 변수에 관한 정보가 부족하거나 너무 많은 변수가 영향을 미치고 있는 경우에도 과학적인 예측치를 구할수 있는 단기예측 방법이다.

Box-Jenkins 모형은 자동회귀 모형(Autoregressive Model), 이동 평균 모형(Moving average Model), 계질적 시계열 모형을 통합한 일반적인 모형이기 때문에 특별한 불안정성을 보이지 않는 경우에는 모두 모형화 할 수 있으며 모형에 관계된 계수의 수를 최소화 하면서 만족스러운 모형을 찾을 수 있다.

Box-Jenkins 예측 방법은 모형 선정, 매게 변수 추정, 적합성 검정의 3단계를 반복으로 수행함으로써 최적 모형에 이르게 하게 하고 있기 때문에 최소의 가능한 모형으로 부터 시작하여 부적당한 부분을 제거시켜 나감으로써 시행 착

오의 과정을 최소화 할 수 있다.

일반 사용자가 Box-Jenkins 시계열 분석법을 쉽게 사용할 수 있도록 Box-Jenkins Package가 개발되었으며 여기서는 KAIST 전산 개발 센터에 설치된 Package를 소개에고 그 사용예를 보였다.

〈참 고 문 헌〉

1. Box, G.E.P., and Jenkins, G.M., *Time Series Analysis, Forecasting and control*, Holden-Day, Inc., San Francisco, 1970.
2. Mabert, V.A., "An Introduction to Short Term Forecasting Using the Box-Jenkins Methodology," *AIIE Transaction*, 1975.
3. Montgomery, D.C., and Johnson, L.A., *Forecasting and Time Series Analysis*, New York, Mc Graw-Hill, pp. 198~240, 1976.
4. Pack, D.J., "Appendix to Computer Programs for the Analysis of Univariate Time Series Models and Single input Transfer Function Models Using the Methods of Box and Jenkins," The Data Center, College of Administration Science, The Ohio State University, 1974.
5. 박성주, 안철우, *Box-Jenkins* 시계열 분석(KAIST Soffware 안내 시리즈 No. 1), 한국 과학기술원 전산 개발센터, 1979.