

동적 모형에 의한 예측치의 정도 향상에 관한 연구
-A Study on increasing the fitness of forecasts
using Dynamic Model -

윤 석 환*

Yoon, Seok-Hwan

윤 상 원**

Yoon, Sang-Weon

신 용 백***

Shin, Yong-Back

ABSTRACT

We develop a dynamic demand forecasting model compared to regression analysis model and AutoRegressive Integrated Moving Average(ARIMA) model. The dynamic model can apply to the current dynamic data to forecasts through introducing state equation.

A multiple regression model and ARIMA model using given data are designed via the model analysis. The forecasting fitness evaluation between the designed models and the dynamic model is compared with the criterion of sum of squared error.

1.서론

정확한 예측치를 구하기 위해서는, 기반으로 하고 있는 과거의 데이터에 가장 적합한 모델을 설정하는 것이 가장 중요하다. 일반적으로 많이 적용되고 있는 예측 모형은 회귀분석 모형과 시계열 분석 모형 등을 들 수 있다[1,2,3,6]. 이러한 모형들은 과거의 데이터를 기반으로 하여 미래 특정 시점에서의 예측치를 구하려는 것이므로 통계학적 관점에서 접근하는 방식이라고 할 수 있다.

그러나, 이러한 예측 기법을 적용하여 미래 데이터를 예측하는 것은 과거의 데이터가 생성된 당시의 주변 여건, 예를 들면 경제적 환경, 사회적 환경 등이 지금까지 계속 유지된다고 하는 가정을 두고 접근하는 것이기 때문에, 이러한 가정이 달라지게 되면 이에 의한 예측치의 정도(fitness)가 다소 떨어진다고 볼 수 있다.

이러한 점을 보완하기 위해서는 일정한 시간대 개념을 도입하여 가장 최근의 데이터에 가장 많은 가중치를 주어 계산하는 방법이 있고, 과거의 데이터에 대해 정성적 변수(Qualitative Variable)를 도입하여 계산하는 방법이 있다. 또한, 시계열 데이터에 맞는 적절한 동적 상태 공간 모형을 설계하여 예측에 활용하는 방법도 있을 수 있다.

본 논문에서는 주어진 데이터에 적합한 동적 상태 공간 모형을 정의하고 이를 적용하여 예측치의 정도를 향상시킬 수 있는 방안의 타당성을 사례분석을 통하여 확인하는 데에 그 목적이 있다.

*한국전자통신연구소 책임연구원
**영동공과대학교 산업공학과 교수
***아주대학교 산업공학과 교수

2. 동적 모형(Dynamic Model)

선형 동적 시스템 모형(Linear Dynamic System Model)의 일반적인 모형은 다음과 같이 표현된다[4,5].

$$X(k+1)=AX(k)+B*U(k)+W(k) \text{ -----(1)}$$

$$Y(k+1)=C*X(k+1)+V(k+1) \text{ -----(2)}$$

단, X(k): k시점의 상태 변수

U(k): k시점의 제어 입력 변수

W(k): $N(0, \sigma_w^2)$ 을 따르는 프로세스 노이즈

Y(k): k시점의 측정 변수

V(k): $N(0, \sigma_v^2)$ 을 따르는 측정 노이즈

위 모형에서 식(1)은 상태 방정식(state equation)으로서, n번째 차수의 추계적 시스템의 방정식으로 표현되며, 식(2)는 추계적 모형의 측정 방정식이 된다.

A, C는 물리적인 값으로 미리 주어지며, 시스템의 상황에 따라 시변, 시불변으로 처리된다. W(k), V(k)는 랜덤 프로세스 통계량으로서 사전에 알고 있는 정보이다. U(k)는 연속형 제어 입력 변수로서 기계·전자식 연속제어 변수로 처리되며, 이산형 추계적 모형에서는 고려되지 않는 변수이나 U(k)를 제거하여 표현된 모형은 전형적인 칼만 모형이 되며, 주어진 시스템의 상황에 따라 여러가지 모형 설정이 가능하게 된다.

3. 모형별 비교 분석

우리나라의 연간 강수량은 세계 평균보다 높으나 1인당 수자원은 세계 평균에 훨씬 못 미친다. 올해 물 공급량은 수요량에 비해 여유량이 약 7%정도로 예상하고 있으며 2006년에는 부족량이 약 5%로써 전망되고 있다. 수자원이 미래의 중요 자원으로 평가되는 현시점에서, 정확한 예측을 통한 효율적인 사전관리 및 공급정책은 중대한 사항으로 고려될 수 있다. 이러한 관점에서 본 연구에서는 특정 연구 대상 지역 및 기간의 실증적 자료를 적용하여 동적 용수(상수도) 수요 예측 모형을 개발하고자 한다.

본 논문을 위한 사례로서 제시된 모형은 $Y(t)=P(t)*R(t)*U(t)*T(t)$ (단, Y(t):t기의 총수요, P(t):t기의 인구 추정치, R(t):t기의 단위 급수량, U(t):t기의 보급율, T(t):t기의 총일수)이다. 이 모형에서 해당지역의 장래인구 추정은 과거 20년간의 자료로 최소자승법에 의한 방법을 제시하고 있으며, 급수 보급율과 단위 급수량은 해당 지역의 성격, 발전상황 등의 비슷한 타지역의 실적과 도시계획상의 장래 발전계획에 따라 적용되도록 되어있다. 또한 제시된 모형은 생활 수준의 향상과 더불어 물의 사용량이 점점 증가할 수 있는데 용수의 수요를 인구자료로만 설명되며, 용수의 예측오차에는 인구 추정의 불확실성까지 겹쳐 예측오차가 증폭될 우려가 존재하게 된다.

본 연구에서는 예측력이 좋은 수요 예측 모형을 설정하기 위해 먼저 다중회귀분석 및 Box-Jenkins ARIMA 모형을 이용하여 수요예측 모형을 설계하였다. 마지막으로 최적제어 이론의 상태 공간 모형이 응용된 동적 수요예측 모형을 개발한다. 다중 회귀 분석 모형의 설명 변수로는 급수인구의외 비급수 인구, 기온, 용수의 톤당 가격, 소비자 물가지수, 소득을 도입하여 사용된다. 종속 변수로는 단순히 관련 기관에서 공급한 공급량이 아니라 추정의 정확성을 더 높이기 위해, 실제 그 지역의 소비량 즉, 공급량(1-누수율)=실제판매량을 계산하여 적용하였다.

3-1. 다중회귀분석 모형에 의한 분석

용수의 수요는 실질요금, 총인구, 급수인구, 소득, 기온의 변수에 의존한다고 가정하여 다음과 같은 수요함수가 구성된다.

$$Y(t)=f(\text{Price}(t)/\text{CPI}(t), \text{TPOP}(t), \text{Ben}(t), \text{Income}(t), \text{Temp}(t)) \text{-----}(3)$$

- 단, Y(t) : t월의 용수수요(실제 판매량)
- Price(t) : t월의 톤당 용수가격(원)
- CPI(t) : t월의 소비자 물가지수(1985=100을 기준)
- TPOP(t) : t월의 D지역 총인구
- Ben(t) : t월의 급수인구
- Temp(t) : t월의 평균기온
- Income(t) : t월의 전국 월평균 소득

여기에서 총인구는 반기별로 파악되기 때문에 그 사이의 데이터는 선형으로 증가한 것으로 계산된다. 급수인구는 상수도 보급율(인구기준)을 총인구에 곱해서 얻어지며, 실질요금 수준은 톤당가격을 소비자 물가지수(CPI)로 디플레이트하여 사용된다. 식(3)을 선형모형화하면 다음과 같다.

$$Y(t)=\beta_0+\beta_1(\text{Price}(t)/\text{CPI}(t))+\beta_2\text{TPOP}(t)+\beta_3\text{Ben}(t)+\beta_4\text{Temp}(t)+\beta_5\text{Income}(t)+\varepsilon(t) \text{-----} (4)$$

식(4)에 의한 결과(STATGRAPHICS 통계패키지 사용)는 <표.1>, <표.2>, <표.3>과 같다.

<표.1> 모형 fitting 결과1

Model fitting results for : Water.dat

Independent variable	coefficient	std. error	t-value	sig.lev
CONSTAT	3050.367285	682.189421	4.4714	0.00
DATA.Unit_pr_/DATA.CPI	-3595.896209	330.60644	-10.8767	0.00
DATA.T_Pop_	-3.617174	1.477575	-2.4480	0.01
DATA.Benefic_	12.504382	1.927073	6.4888	0.00
DATA.Temperat_	19.433671	2.517135	7.7206	0.00
DATA.Income	0.812754	0.468171	1.7360	0.08

R-SQ.(ADJ.) = 0.9592 SE= 257.368011 MAE= 202.899609 Durbwat= 1.28
 Previously: 0.0000 0.000000 0.000000 0.00
 114 obsrvations fitted, forecast(s) computed for 0 missing val. of dep. var.

<표.2> 상관계수 행렬

Correlation matrix for coefficient estimates

	CONSTAT	DATA.Unit_pr_/DA	DATA.T_Pop	DATA.Benefi
CONSTAT	1.0000	-.9179	-.1388	-.00
DATA.Unit_pr_/DA	-.9179	1.0000	.2314	-.18
DATA.T_Pop_	-.1388	.2314	1.0000	-.96
DATA.Benefic_	-.0029	-.1827	-.9630	1.00
DATA.Temperat_	.0270	-.0466	-.0533	.02
DATA.Income	-.1089	.4066	.2514	-.43

	DATA.Temperat_	DATA.Income
CONSTAT	.0270	-.1089
DATA.Unit_pr_/DA	-.0466	.4066
DATA.T_Pop_	-.0533	.2514
DATA.Benefic_	.0261	-.4387
DATA.Temperat_	1.0000	.0511
DATA.Income	.0511	1.0000

<표.3> 95%신뢰구간

95 percent confidence intervals for coefficient estimates

	Estimate	Standard error	Lower Limit	Upper Lim
CONSTAT	3050.37	682.189	1697.85	4402.
DATA.Unit_pr_/DA	-3595.90	330.606	-4251.36	-2940.
DATA.T_Pop_	-3.61717	1.47758	-6.54664	-0.687
DATA.Benefic_	12.5044	1.92707	8.68373	16.32
DATA.Temperat_	19.4337	2.51714	14.4432	24.42
DATA.Income	0.81275	0.46817	-0.11545	1.740

위 표에서 알 수 있듯이 상수항과 $\beta_5(\text{Income})$ 에 대한 t-value가 비교적 낮은 편이며, 상수항과 실질요금, 급수인구와 총인구간의 상관계수가 각각 -0.9179, -0.9630으로 대단히 크게 나타난다. β_5 에 대한 신뢰구간이 0을 포함하고 있기때문에 식(4)에서 제거된다. 급수인구와 총인구사이에 공선성(Collinearity)가 있는 것으로 보고 다음과 같은 새로운 모형이 구성되어진다.

$$Y(t) = \beta_1(\text{price}(t)/\text{CPI}(t)) + \beta_3\text{Ben}(t) + \beta_4\text{Temp}(t) + \epsilon(t) \quad (5)$$

식(5)에 의한 결과는 <표.4>와 같다.

<표.4> 모형 fitting 결과2

Model fitting results for : Water.dat

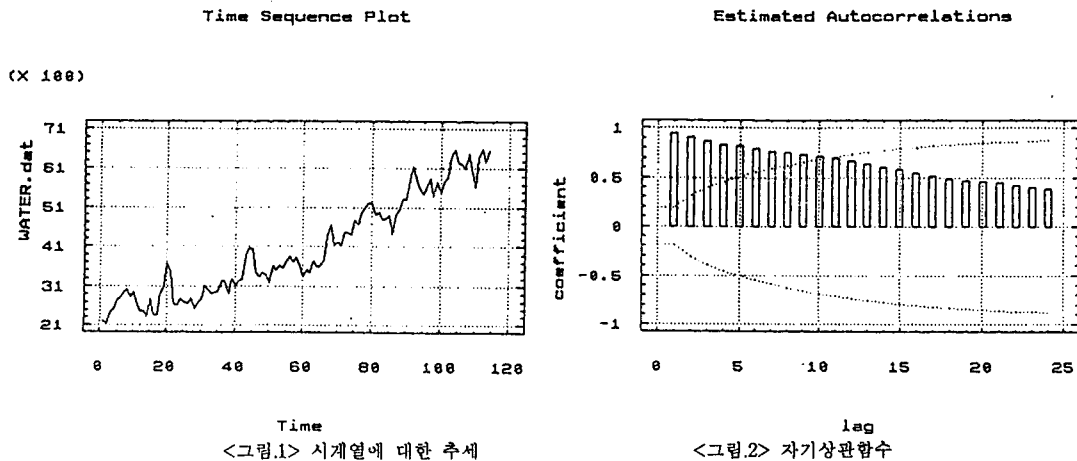
Independent variable	coefficient	std. error	t-value	sig.lev
DATA.Unit_pr_/DATA.CPI	-2441.140108	91.589955	-26.6529	0.00
DATA.Benefic_	10.210414	0.188648	54.1243	0.00
DATA.Temperat_	18.422699	2.805768	6.5660	0.00

R-SQ. (ADJ.) = 0.9955 SE= 288.019916 MAE= 234.567325 Durbwat= 0.95
 Previously: 0.0000 0.000000 0.000000 0.00
 114 obsrvations fitted, forecast(s) computed for 0 missing val. of dep. var.

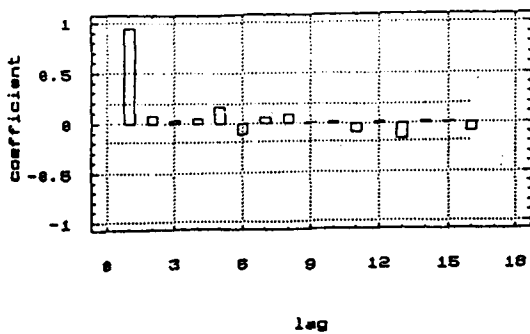
<표.4>에서 나타난 바와 같이 R²값이 식(4)의 95.92%보다 높은 99.55%로 설명력이 한층 더 증대된다. 모든β에 대한 t값이 높게 나왔으며 특히 급수인구의 계수(β₃)의 t값은 식(4)의 6.48보다 훨씬 높은 54.12로 t-test에 의한 신뢰도가 아주 향상된 것으로 나타난다. 또한 가격의 탄력성은 -2441.14로서 t-value도 -26.65로 잘 추정 된것으로 판단되며, 소득 수준의 증가가 용수의 소비를 상승시킨다는 가정에서는 설명력을 가지지 못하는 것으로 나타난다. 잔차들의 분포에 있어서도 0을 주위로 분포되어 있으며 정규분포에 가깝다고 할 수 있다. 따라서 다중 선형회귀모형인 식(5)로서 최종 설정하여 분석하게 된다.

3-2. 시계열 분석모형에 의한 분석

원래의 시계열 데이터(Y(t),t=1~114)가 어떤 경향을 가지고 있는지 알아보기 위해 시계열에 대한 추세, 자기상관함수 및 편자기상관함수를 그림으로 나타내어 보면 각각 <그림.1>, <그림.2>, <그림.3>과 같다.

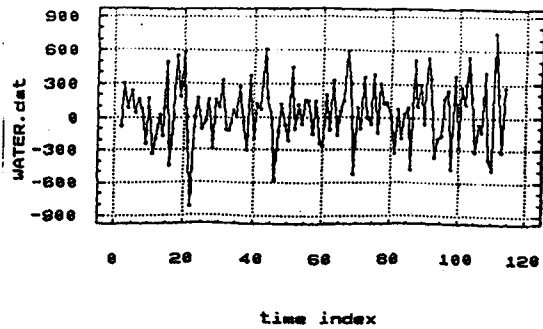


Estimated Partial Autocorrelations



<그림.3> 편자기상관함수

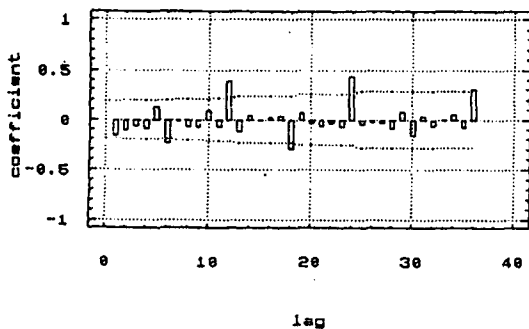
1 Nonseasonal Differences



<그림.4> 1차 차분후의 추세

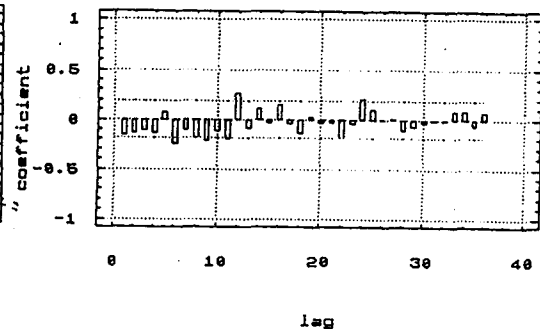
위 <그림.1>을 보면 용수관매량은 매년 증가하는 추세에 있으며, <그림.2>의 자기상관함수도 선형으로 조금씩 감소되게 나타나고 있다. <그림.3>의 편자기상관함수도 첫번째 시차 lag에서만 아주 큰값을 가지며 나머지는 모두 0에 가깝다. 위 분석 결과에 따라 본래의 데이터는 비정상성을 알 수 있다. 비정상성을 제거하기 위해 1차 차분(first differencing) $\{Y(t)-Y(t-1) = (1-B)Y(t)\}$ 을 실시하여 이에 대한 추세 및 상관함수를 보면 <그림.4>, <그림.5>, <그림.6>과 같다.

Estimated Autocorrelations
for 1 Nonseasonal Differences



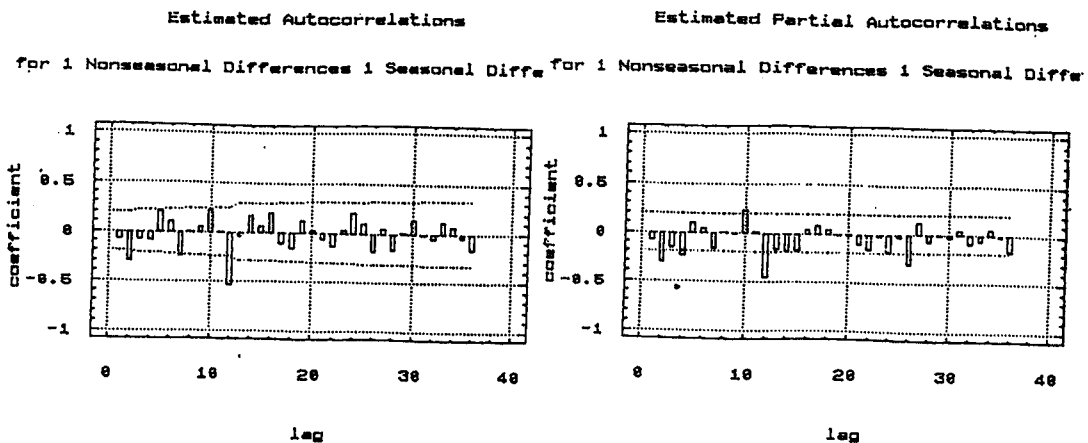
<그림.5> 자기상관함수

Estimated Partial Autocorrelations
for 1 Nonseasonal Differences



<그림.6> 편자기상관함수

<그림.4>를 보면 1차 차분후의 시계열은 어느 정도 정상성을 보여주고 있으며 <그림.5>의 자기상관계수를 보면 시차(lag) 12와 24에서 높은 양의 값을 가지므로 계절성(seasonality)이 존재한다는 것을 알 수 있다. 특히 자기 상관 계수값 $r_{12}=0.48$, $r_{24}=0.44$ ($r_{12}>r_{24}$)으로 보아 길이가 12개월인 계절성을 갖는다. 계절성을 제거하기 위해 1차 차분한 $(1-B)Y(t)$ 에 다시 기간이 12인 1차 차분 $(1-B)(1-B^{12})Y(t)$ 를 구해서 자기상관함수 및 편자기상관함수를 구해보면 <그림.7>과 <그림.8>에 나타난 바와 같다.



<그림.7> 자기상관함수

<그림.8> 편자기상관함수

<그림.7> 자기상관함수

<그림.8> 편자기상관함수

<그림.7>에서 자기상관계수가 시차(lag) 12에서 크게 나타나고 편자기상관계수도 0에 가깝지만 일정한 패턴을 가지면서 진동한다. 이것은 ARMA(1,1)의 요소가 존재한다고 볼 수 있으므로 원래의 시계열 $\{Y(t)\}$ 는 ARIMA(1,1,1)(1,1,0)¹² 모형으로 추정될 수 있다.

이것을 일반 모형식으로 표현하면 식(6)과 같고 구체적인 식은 식(7)와 같이 나타낼 수 있다.

$$(1-\phi B)(1-\phi B^{12})(1-B)(1-B^{12})Y(t)=(1-\theta B)a(t) \text{ -----(6)}$$

$$Y(t)=(1+\phi)Y(t-1)-\phi Y(t-2)+(1+\phi)Y(t-12)-(1+\phi+\phi\phi)Y(t-13) \\ +(\phi+\phi\phi)Y(t-14)-\phi Y(t-24)+(\phi+\phi\phi)Y(t-25)-\phi\phi Y(t-26) \\ +a(t)-\theta a(t-1) \text{ -----(7)}$$

단, $a(t)$: 가우시안 백색 잡음

ϕ : AR 과정의 모수

ϕ : SAR 과정의 모수

θ : MA 과정의 모수

잠정적으로 정한 모형 식(7)에 모수 ($\phi, \phi, \theta, \sigma_a^2$)들을 추정하기 위한 분석 결과는 <표.5>와 같으며 그에 따른 최종 추정된 모형은 식(8)과 같다.

<표.5> ARIMA (1,1,1)(1,1,0)¹²의 모수추정 결과

```
Iteration 2: RSS = 5.76047E6   b = 0.0163603 -0.588723 0.220051
Iteration 3: RSS = 5.52143E6   b = 0.186411 -0.633677 0.459612
Iteration 4: RSS = 5.28922E6   b = 0.444538 -0.640406 0.780683
Iteration 5: RSS = 5.2729E6    b = 0.482008 -0.632073 0.826739
Final:      RSS = 5.26796E6    . . . stopped on criterion 2
```

 Summary of Fitted Model for : Water.dat

Parameter	Estimate	Std.error	T-value	P-value
AR(1)	.51139	.14965	3.41727	.00092
SAR(12)	-.62853	.08695	-7.22832	.00000
MA(1)	.85062	.09613	8.84845	.00000

Model fitted to differences of order 1

Model fitted to seasonal differences of order 1 with seasonal length = 12

Estimated white noise variance = 53754.7 with 98 degrees of freedom.

Estimated white noise standard deviation (std err) = 231.851

Chi-square test statistic on first 20 residual autocorrelation = 14.9272

with probability of a larger value given white noise = 0.600722

Backforecasting : no

Number of iterations performed : 6

$$\begin{aligned}
 Y(t) = & 1.5113Y(t-1) + 0.6285Y(t-2) + 0.3714Y(t-12) - 0.5614Y(t-13) \\
 & + 0.1900Y(t-14) + 0.6285Y(t-24) - 0.9499Y(t-25) + 0.3214Y(t-26) \\
 & - 0.85062a(t-1) + a(t) \text{ ----- (8)}
 \end{aligned}$$

추정된 모형의 검증은 잔차분석방법을 적용시켰으며 그 결과는 <표.6>과 <그림.9>, <그림.10>, <그림.11>, <그림.12>와 같다.

분석결과를 보면 잔차들은 평균이 0에 가깝고 정규분포에 가깝게 나타나고 있음을 알 수 있다. 잔차의 자기상관계수가 모든 시차에서 0에 가깝게 나타나고 있으며 모두 95%의 신뢰구간 안에 들어가 있다. 또한 잔차의 편자기상관계수도 역시 0에 근접하고 있으며 잔차들의 integrated periodgram 역시 잔차들이 평균이 0인 정규분포에 근접하고 있음을 알 수 있다.

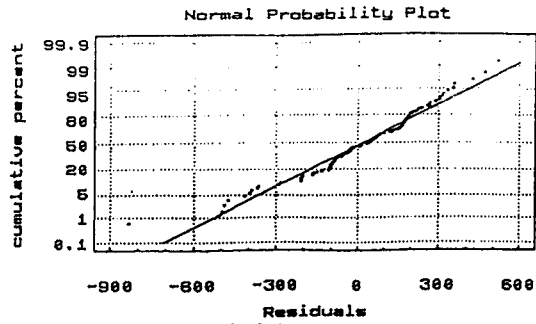
<표.5>에서 잔차들에 대한 자기상관함수의 시차 20까지의 chi-square test statistic $\chi^2 = 14.92$ 으로 잔차들의 자기상관함수가 0이 아니라는 가설은 받아들이기 힘들다.

결국 위와 같은 분석결과로서 앞에서 추정된 ARIMA(1,1,1)(1,1,0)모형은 타당성이 있다고 할 수 있다.

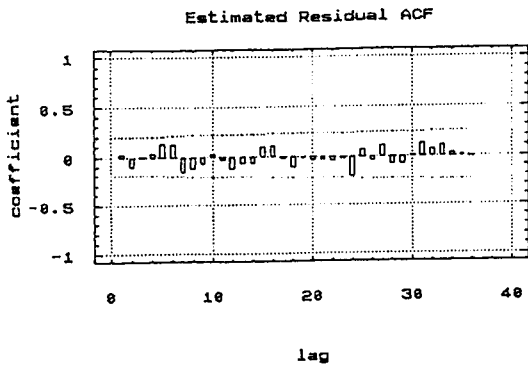
<표.6> 잔차 요약

Residual Summary

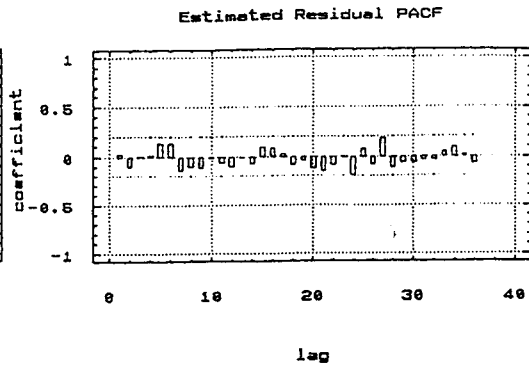
Number of Observation	= 101
Residual Average	= 20.897
Residual Variance	= 53754.7
Residual Std. Error	= 231.851
Coeff. of skewness	= -0.733804
Standardized Value	= -3.01069
Coeff. of kurtosis	= 1.28731
Standardized Value	= 2.64082



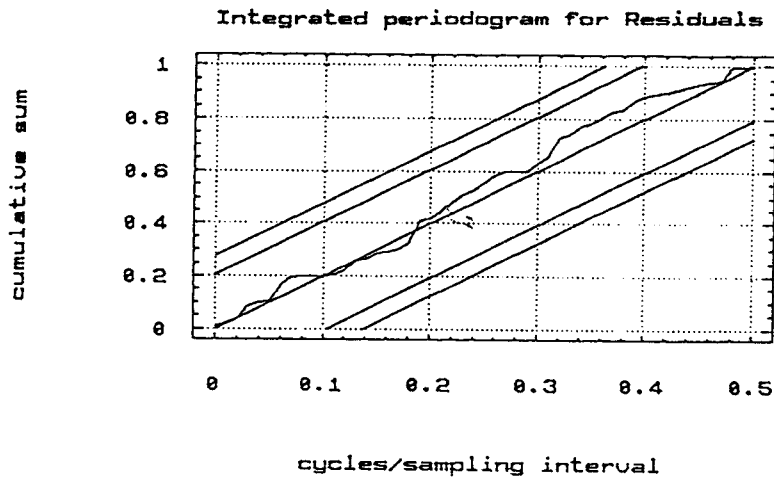
<그림.9> 잔차의 분포



<그림.10> 잔차의 자기상관함수



<그림.11> 잔차의 편자기상관함수



<그림. 12> 잔차의 integrated periodogram

3-3 동적 모형에 의한 분석

본 연구에서는 식(1)의 A 및 C 행렬을 주어진 용수 시계열 데이터에 맞는 적절한 동적 상태 공간 모형의 설계를 통해 새로운 예측 모형을 개발하였다.

설계된 모형은 (9), (10)식에 제시된다.

$$X(k) = \begin{bmatrix} a & 1-a \\ 0 & 1 \end{bmatrix} X(k-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(k-1) \text{-----(9)}$$

단, $X(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$, $a: 0 < a < 1$,

$x_1(k)$: k시점에서의 판측치, $x_2(k)$: k시점에서의 추정치, $w(k) \sim N(0, \sigma_w^2)$

$$y(k) = [1, 0] X(k) + v(k) \text{-----(10)}$$

단, $v(k) \sim N(0, \sigma_v^2)$

식(9), 식(10)에서 A, C행렬이 결정되면 그에 따른 동적 상태공간 모형 알고리즘에 의해 분석된다. 전체 알고리즘 구성도는 동일하지만 주어진 데이터에 맞는 내부 A행렬(시스템 모수), C행렬(판측 모수)의 구성도에 따라 여러가지 다양한 예측 모형 설정이 가능하다.

특히 본 연구에서는 식(9)에서 $\hat{X}(0)$ (초기 추정치값)를 각각 $\hat{x}_1(0)$ 은 0, $\hat{x}_2(0)$ 는 전체 데이터의 평균값으로 설정된다. P_0 (초기에러공분산)는 임의의 일정한 값, $w(k), v(k)$ 는 각각 $N(0, \sigma_w^2), N(0, \sigma_v^2)$ 을 따르는 가정하에 σ_w^2, σ_v^2 을 임의의 일정값으로 둔다. 그때 a값을 다양하게 변화시켜가면서 전체 데이터에 가장 근접된 예측효율을 갖는 동적모형이 설계된다. 모형 설계를 위해 a값 및 초기값들은 다음과 같이 설정된다.

$$\hat{x}_1(0)=0, \hat{x}_2(0)=4069.05, P_0(1,1)=0.01, P_0(1,2)=0$$

$$P_0(2,1)=0, P_0(2,2)=0.02, \sigma_w^2=0.01, \sigma_v^2=0.001, \text{와 같고}$$

a는 0.001, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99의 8가지로 나누어 분석한다. 각모형의 비교기준으로 마지막 18개월에 대한 예측 오차제곱합의 크기로 비교 되며 이를 정리한 결과는 <표.7>, <표.8>, <표.9>와 같고, 제곱합의 크기를 종합하여 정리한 결과는 <표.10>과 같다.

<표.7> 동적 모형별 오차비교

모형 (A)	$A = \begin{bmatrix} 0.001 & 0.999 \\ 0 & 1 \end{bmatrix}$			$A = \begin{bmatrix} 0.01 & 0.99 \\ 0 & 1 \end{bmatrix}$			$A = \begin{bmatrix} 0.1 & 0.9 \\ 0 & 1 \end{bmatrix}$		
	실측치	추정치	예측 오차 제곱	실측치	추정치	예측 오차 제곱	실측치	추정치	예측 오차 제곱
1	5856	5833	529	5856	5835	441	5856	5852	16
2	5384	5414	900	5384	5411	729	5384	5386	4
3	5759	5701	3364	5759	5703	3136	5759	5718	1681
4	5483	5491	64	5483	5489	36	5483	5478	25
5	5760	5737	529	5760	5736	576	5760	5753	49
6	5877	5846	961	5877	5846	961	5877	5856	441
7	6417	6353	4096	6417	6357	3600	6417	6391	676
8	6593	6572	441	6593	6572	441	6593	6592	1
9	6280	6296	256	6280	6295	225	6280	6279	1
10	6219	6215	16	6219	6214	25	6219	6206	169
11	6086	6091	25	6086	6090	16	6086	6081	25
12	6486	6432	2916	6486	6435	2601	6486	6456	900
1	6103	6109	36	6103	6107	16	6103	6089	196
2	5627	5642	225	5627	5638	121	5627	5605	484
3	6385	6307	6084	6385	6311	5476	6385	6350	1225
4	6595	6568	729	6595	6570	625	6595	6593	4
5	6282	6296	196	6282	6294	144	6282	6280	4
6	6559	6508	2601	6559	6510	2401	6559	6521	1444

<표.8> 동적 모형별 오차비교

모형 (A)	$A = \begin{bmatrix} 0.3 & 0.7 \\ 0 & 1 \end{bmatrix}$			$A = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix}$			$A = \begin{bmatrix} 0.7 & 0.3 \\ 0 & 1 \end{bmatrix}$		
	실측치	추정치	예측 오차 제곱	실측치	추정치	예측 오차 제곱	실측치	추정치	예측 오차 제곱
1	5856	5885	841	5856	5898	1764	5856	5851	25
2	5384	5355	841	5384	5371	169	5384	5738	2916
3	5759	5729	900	5759	5710	2401	5759	5677	6724
4	5483	5467	256	5483	5474	81	5483	5493	100
5	5760	5770	100	5760	5766	36	5760	5735	625
6	5877	5877	0	5877	5896	361	5877	5893	256
7	6417	6450	1089	6417	6479	3844	6417	6454	1369
8	6593	6639	2116	6593	6692	9801	6593	6733	19600
9	6280	6263	289	6280	6290	100	6280	6395	13225
10	6219	6186	1089	6219	6168	2601	6219	6199	400
11	6086	6062	576	6086	6043	1849	6086	6035	2601
12	6486	6485	1	6486	6479	49	6486	6412	5476
1	6103	6071	1024	6103	6088	225	6103	6127	576
2	5627	5544	6889	5627	5516	12321	5627	5562	4225
3	6385	6395	100	6385	6364	441	6385	6241	20736
4	6595	6649	2916	6595	6695	10000	6595	6668	5329
5	6282	6267	225	6282	6303	441	6282	6393	12321
6	6559	6531	784	6559	6525	1156	6559	6533	676

<표.9> 동적 모형별 오차 비교

모형 (A)	$A = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix}$			$A = \begin{bmatrix} 0.99 & 0.01 \\ 0 & 1 \end{bmatrix}$		
	실측치	추정치	예측오차 제곱	실측치	추정치	예측오차 제곱
t						
1	5856	5724	17424	5856	5833	529
2	5384	5498	12996	5384	5779	156025
3	5759	5614	21025	5759	5811	2704
4	5483	5514	961	5483	5771	82944
5	5760	5658	10404	5760	5802	1764
6	5877	5809	4624	5877	5847	900
7	6417	6258	25281	6417	6014	162409
8	6593	6615	484	6593	6201	153664
9	6280	6546	70756	6280	6280	0
10	6219	6412	37249	6219	6325	11236
11	6086	6221	18225	6086	6326	57600
12	6486	6370	13456	6486	6413	5329
1	6103	6192	7921	6103	6389	81796
2	5627	5761	17956	5627	5246	145161
3	6385	6065	102400	6385	6272	12769
4	6595	6417	31684	6595	6402	37249
5	6282	6389	11449	6282	6407	15625
6	6559	6526	1089	6559	6469	8100

<표.10> 종합정리결과

구분	a							
	0.001	0.01	0.1	0.3	0.5	0.7	0.9	0.99
예측오차제곱합	23968	21570	7345	20036	47640	97180	405384	935804

<표.10>결과를 보면, a값을 0.1 기준에서 0.2간격으로 0.9까지 분석하고, 최소값 0.001과 최대값 0.99의 값도 분석되었다. 주어진 시계열 데이터의 속성에 따라 최소예측오차 제곱합을 갖는 A값이 결정 된다고 할때 본 연구의 용수(상수도)데이터에 맞는 모형은 최소예측오차제곱합 7,345의 값을 갖는 a=0.1이 선택된다. 따라서 최종 동적 예측모형식은 다음과 같이 설정된다.

$$X(k) = \begin{bmatrix} 0.1 & 0.9 \\ 0 & 1 \end{bmatrix} X(k-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(k-1) \text{ -----(11)}$$

$$y(k) = [1 \ 0] X(k) + v(k) \text{ -----(12)}$$

4. 결과 비교

모형의 비교기준으로 마지막 18개월 데이터에 대한 예측 오차 제곱합의 크기로 비교되며 이를 정리한 결과는 <표.11>과 같다.

<표.11> 모형별 오차 비교

모형 t	회귀 모형			ARIMA(1,1,1)(1,1,0) ¹² 모형			동적모형 $A = \begin{bmatrix} 0.1 & 0.9 \\ 0 & 1 \end{bmatrix}$		
	실측치	추정치	예측오차 제곱	실측치	추정치	예측오차 제곱	실측치	추정치	예측오차 제곱
1	5856	5150	498436	5856	5643	45369	5856	5852	16
2	5384	5292	8464	5384	5582	39204	5384	5386	4
3	5759	5443	99856	5759	5840	6561	5759	5718	1681
4	5483	5611	16384	5483	5744	68121	5483	5478	25
5	5760	5774	196	5760	5884	15376	5760	5753	49
6	5877	5893	256	5877	5915	1444	5877	5856	441
7	6417	5985	186624	6417	6215	40804	6417	6391	676
8	6593	6017	331776	6593	6520	5329	6593	6592	1
9	6280	5993	82369	6280	6214	4356	6280	6279	1
10	6219	5874	119025	6219	6217	4	6219	6206	169
11	6086	5773	97969	6086	6015	5041	6086	6081	25
12	6486	5611	765625	6486	6133	124609	6486	6456	900
1	6103	5624	229441	6103	6508	164025	6103	6089	196
2	5627	5674	2209	5627	5677	2500	5627	5605	484
3	6385	5836	301401	6385	6138	61009	6385	6350	1225
4	6595	6635	1600	6595	6300	87025	6595	6593	4
5	6282	6154	16384	6282	6772	240100	6282	6280	4
6	6559	6259	90000	6559	6371	35344	6559	6521	1444
예측 오차 제곱합=2848015			예측 오차 제곱합=946221			예측 오차 제곱합=7345			

<표.11>에서와 같이 18개 데이터의 예측오차 제곱합은 회귀모형이 2,848,015로서 ARIMA 모형의 946,221보다 약 3.01배 정도 크며, 동적모형의 7,345보다 약 387.75배 정도 크게 나타나고 있다. 특히 ARIMA모형은 동적 모형보다 약 128.82배 정도 크며, 예측오차에 대한 표준편차는 각각 397.8, 229.3, 20.2로서 동적 모형이 2개의 모형보다 예측력이 더 좋다고 할 수 있다.

한편, 동적모형에서 가장 낮은 예측력을 보여주는 a=0.99에서 예측오차 제곱합 935,804은 세 모형중 중간정도의 예측력을 갖는 ARIMA모형의 946,221보다 약 1.01배의 차이가 발생하고 있으며, 회귀모형에서는 약 3.04배의 차이를 나타내고 있다.

또한, 회귀모형에서는 95% 신뢰구간 밖으로 18개의 실측 데이터 중 4개가, ARIMA모형에서는 1개의 실측 데이터가 하한선을 약간 벗어난 반면 동적 모형에서는 한계선 내에 모두 존재함을 알 수 있다.

따라서, 주어진 용수(상수도)데이터를 예측하는데 있어서 동적 모형이 우수함을 알 수 있음은 물론 계절성과 비정상성을 갖는 시계열 데이터의 정도 높은 예측력을 갖는 모형설정이 중요하다라는 관점에서 본 연구에서 제시한 동적 모형은 더욱 실용적이고 유효한 예측 모형이라 평가할 수 있다.

5. 결론

본 연구에서는 용수(상수도) 수요예측의 실증적 연구에 관한 내용으로서 기존에 많이 사용

되고 있는 수요예측기법인 회귀분석과 시계열 분석기법을 사용하여 주어진 시계열 데이터에 맞는 모형설정 및 검증을 실시하였다. 또한 제어공학등에 일반적으로 적용되는 동적모형을 응용하여 용수 수요예측을 위한 상태공간 모형을 새로이 구축하였다. 구축된 동적 모형에 의한 분석결과 회귀분석 및 시계열 분석보다 우수한 예측력을 갖는 모형으로 나타났다.

또한, 동적 모형은 시시각각의 관측치와 예측치를 비교하여 필터를 개량해 나가면서 오차를 점점 줄일수 있는 방법으로서, 반드시 시스템의 장래를 관측해서 얻어지는 확률과정이 정상적이 아니더라도 응용할 수 있는 특징을 갖고 있다는 관점에서, 예측력이 더욱 향상된 동적예측 모형의 연구는 향후 연구과제로서 설정될 수 있겠다.

참 고 문 헌

- [1] 김연형, 시계열 분석과 예측, 자유아카데미, pp. 363-510, 1990.
- [2] 김치홍, 예측과학, 성균관대학교출판부, pp. 107-125, 1985.
- [3] 박성현, 회귀분석, 대영사, pp. 171-196, 1990.
- [4] Bozic, S.M., Digital and Kalman Filtering; An Introduction to Discrete Time Filtering and Optimum Linear Estimation, Edward Arnold, pp. 113-120, 1979.
- [5] Lewis, F.L., Optimal Estimation, John Wiley & Sons, pp. 67-80, 1986.
- [6] O. C. Montgomery, Forecasting & Time Series Analysis, McGraw-Hill, pp. 241-290, 1990.
- [7] Stsc, Statgraphics, pp. 19-142, 1987.