

Prediction of Electricity Sales by Time Series Modelling

Young Sook Son^{a,1}

^aDepartment of Statistics, Chonnam National University

(Received February 10, 2014; Revised March 31, 2014; Accepted May 20, 2014)

Abstract

An accurate prediction of electricity supply and demand is important for daily life, industrial activities, and national management. In this paper electricity sales is predicted by time series modelling. Real data analysis shows the transfer function model with cooling and heating days as an input time series and a pulse function as an intervention variable outperforms other time series models for the root mean square error and the mean absolute percentage error.

Keywords: Electricity sales, Winters seasonal exponential smoothing, AR error regression model, multiplicative seasonal ARIMA model, transfer function model, intervention variable, pulse function, cooling and heating days.

1. 서론

전력산업이 국가에 의해서 관리되는 기간산업인 것은 현대사회의 가정, 산업, 국가가 전력 자원에 전적으로 의존하여 운영되기 때문이다. 전력통계정보시스템(<http://www.kpx.or.kr/epsis/>)에 의하면 2012년 우리나라 전력사용량은 주거용, 업무용, 산업용이 각각 14%, 33%, 53%로 구성되어 있다. 전력이 끊겼던 때를 생각해 보자. 전력이 끊기면 가정에서는 전력 뿐만 아니라 도시가스도 사용할 수 없게 되어 취사마저 불가능하고 각종 상업 및 산업활동은 위축되거나 정지된다. 우리나라는 최근 수년 동안 전력수급난을 겪고 있다. 특히 혹서기와 혹한기에는 냉난방기 등 전력사용을 자제하자는 캠페인을 벌이곤 한다. 정확한 전력수급예측은 국민들의 일상적 생활 유지, 산업 활동, 국가 경영에 있어서 매우 중요하다. 미래의 전력수요에 대한 정확한 예측을 통하여 미리 전력공급을 준비해야 한다. 전력수급 예측에서 가장 핵심적인 예측은 최대전력수요와 전력판매량 예측이다. 정부는 전력수급을 안정시키기 위하여 2년마다 향후 15년의 전력수급계획을 담은 전력수급기본계획을 확정한다. 이 계획은 2002년 처음 발표를 시작하여 가장 최근 지식경제부 (Ministry of Knowledge Economy, 2013)는 제 6차 전력수급기본계획(2013년~2027년)을 발표하였다. 이 기본계획에는 2012년까지의 실적치를 기준으로 2013년부터 2027년까지의 년도별 최대전력수요와 전력판매량의 예측치를 담고 있다.

최대전력수요(electricity peak demand)는 어느 일정기간 동안의 1시간 평균전력이 최대인 전력수요 값을 말한다. 산정기간에 따라 일일, 일주간, 일개월, 연간 최대전력수요 등으로 구분하며, 요일별, 계절

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2011-0022864).

¹Department of Statistics, Chonnam National university, 300 Yongbong-Dong, Buk-Gu, Kwangju 500-757, Korea. E-mail: ysson@jnu.ac.kr

별, 기후조건, 기타 전력소비의 형태 등에 따라 발생시간대가 다르다. 여름철 최대전력수요는 냉방기기의 가동이 많은 15시를 전후한 낮 시간대 주로 발생하며, 겨울철에는 21시를 전후로 한 야간시간대에 발생한다.

전력판매량(electricity sales)은 가정, 사무실, 공장, 학교, 관공서 등 각 수용가에서 사용한 전력량을 모두 합한 것으로서 사업자와 전기위탁 판매량을 포함하여 수용가에 판매된 전체 전력량을 말한다. 판매자 입장에서 전력판매량은 소비자 입장에서 전력소비량과 같은 의미이다.

월별 최대전력수요 예측에 관한 연구로서 Nam 등 (2008)은 다중회귀모형에 의해, Song (2007)은 퍼지 최소제곱 선형회귀모형에 의해 예측하였다. Yoon 등 (2009)은 Winters의 계절지수평활법과 두 종류의 분해시계열법에 의해 예측하였다. 그 중 하나는 추세성분과 계절성분을 시간의 함수로 추정하여 분해해내고 남은 불규칙 성분을 ARMA 오차로 모형화한 분해법과 또 다른 분해시계열 방법으로 추세성분은 전력소비량으로, 계절성분은 월별 평균 기온 및 열대야 일수로 추정하여 분해해내고 남은 불규칙성분을 ARMA 오차로 모형화하였다. Lee 등 (2013)은 일별 최대전력수요 예측을 위하여 승법계절 ARIMA모형, Winters 계절지수평활법, 두 가지 계절 주기로 평활한 Winters 이중계절지수평활법 (Taylor, 2003), AR 오차모형에 이분산을 수용하는 AR-GARCH모형, 그리고 평균기온을 고려하여 회귀모형에 의해 사전조정된 오차에 ARIMA모형을 사용하였다. 전력거래소 (Korea Power Exchange, 2010)는 전력판매량, 기온, 명절 효과, 하계 휴가 효과를 설명변수로 하는 중회귀모형에 의하여 최대전력수요 예측을 하였다.

최대전력수요 예측에 비해 전력판매량 예측에 관한 연구는 거의 없는 편이다. 전력거래소 (Korea Power Exchange, 2010)는 기온, GDP, 전력 가격, 산업생산지수를 설명변수로 하는 중회귀모형에 의하여 월별 전력판매량 예측을 하였다.

시계열자료를 회귀모형으로 적합시키는 많은 사례에서 잔차들은 자기상관관계를 갖게 되어 랜덤오차를 가정하는 회귀모형의 가정에 위배된다. 또한 추세성분과 계절성분을 먼저 분해해내고 남은 잔차에 ARMA모형 적합을 하는 것과 같은 2단계 적합과정을 거치지 않고 바로 ARMA오차를 갖는 회귀모형, 즉 전이함수모형 적합을 시계열분석을 위한 통계소프트웨어인 SAS Institute Inc. (2008)의 SAS/ETS, PROC ARIMA에 의해서 수행할 수 있다.

본 연구에서는 시계열모형에 의해 우리나라 월별 전력판매량(전력소비량)을 예측하였다. 2000년 1월부터 2011년 12월까지의 전력판매량에 대해 시계열모형 적합을 한 후, 2012년 1월부터 2013년 12월까지 2년 동안 우리나라 월별 전력판매량을 예측하였다. 전력판매량 시계열은 직선 추세성분과 계절성분을 가지므로 이들 성분과 함께 오차들간의 자기상관 관계를 수용할 수 있는 시계열모형들을 적합시켰다. 이러한 목적을 위하여 사용된 모형은 시간의 함수로 추세성분 및 계절성분을 나타낸 AR오차 회귀모형, 승법계절 ARIMA모형, 기온과 관련된 냉방도일과 난방도일을 합한 냉난방도일 시계열을 입력시계열로 갖고 이상점을 개입변수로 추가한 전이함수 모형이다. 또한, 오차들 간의 자기상관관계를 수용하는 구성요소는 없지만 자동화 프로그램으로 간단히 사용가능한 Winters 계절지수평활법도 함께 적용해봤다.

2. 시계열모형

2.1. Winters 계절지수평활법

일반적으로 지수평활법은 분석 도중에 전문가의 판단을 필요로 하지 않는 방법으로 특히 온라인상에서도 실시간 분석이 가능한 자동화 방법으로 유용하다. 지수평활법이 하나의 평활상수를 사용하는 반면에 Winters (1960)의 계절지수평활법은 시계열의 직선추세, 직선추세의 기울기, 그리고 계절성분을 갱신하기 위한 3개의 서로 다른 평활상수를 사용하는 지수평활에 의해 예측한다. 시계열 계절주기의 폭이 시

간의 흐름에 따라 일정하지 않고 점점 더 커지는 경우 적용할 수 있는 승법계절지수평활모형은 직선추세 $T_t = \beta_0 + \beta_{1,t}t$ 를 가정하는 경우 시점 n 에서 l 시차 후 미래값 Y_{n+l} 은 식 (2.1)과 같다.

$$Y_{n+l} = T_{n+l} \cdot S_{n+l} + I_{n+l} = (T_n + \beta_{1,n} \cdot l) \cdot S_{n+l} + I_{n+l}, \quad (2.1)$$

여기서 T_{n+l} 은 추세성분, S_{n+l} 은 계절주기 s 를 가지는 계절성분, I_{n+l} 은 불규칙성분으로서 랜덤오차에 해당한다. 미래값 Y_{n+l} 의 예측값 $\hat{Y}_n(l)$ 에 대한 예측식은 (2.2)와 같이 정의된다.

$$\hat{Y}_n(l) = (\hat{T}_n + \hat{\beta}_{1,n} \cdot l) \cdot \hat{S}_{n+l-ks}, \quad l = (k-1)s + 1, (k-1)s + 2, \dots, ks, \quad k = 1, 2, \dots \quad (2.2)$$

시점 $n+1$ 이 도래하였을 때 예측식 (2.2)에서 \hat{T}_n , $\hat{\beta}_{1,n}$, 그리고 \hat{S}_{n+l-ks} 는 식 (2.3), (2.4), (2.5)에 의해 갱신된다. 이때 $\omega_1, \omega_2, \omega_3$ 는 0과 1 사이의 값을 가지는 평활상수이다.

$$\hat{T}_{n+1} = \omega_1 \frac{Y_{n+1}}{\hat{S}_{n+1-s}} + (1 - \omega_1) (\hat{T}_n + \hat{\beta}_{1,n}), \quad (2.3)$$

$$\hat{\beta}_{1,n+1} = \omega_2 (\hat{T}_{n+1} - \hat{T}_n) + (1 - \omega_2) \hat{\beta}_{1,n}, \quad (2.4)$$

$$\hat{S}_{n+1} = \omega_3 \frac{Y_{n+1}}{\hat{T}_{n+1}} + (1 - \omega_3) \hat{S}_{n+1-s}. \quad (2.5)$$

Winters의 계절지수평활법에 대한 보다 자세한 내용은 Cho와 Son (2011) 혹은 Montgomery와 Johnson (1976)을 참고하기 바란다.

2.2. AR오차 회귀모형

직선 추세성분은 시간의 선형 함수로, 계절주기 s 를 갖는 월별 계절성분은 시간의 지시함수로 표현한 AR(autoregressive)오차 회귀모형은 식 (2.6)과 같다. 식 (2.7)은 AR(p)오차를 나타낸다.

$$Y_t = \beta_1 t + \sum_{i=1}^s \beta_{s,i} I_{t,i} + a_t, \quad I_{t,i} = \begin{cases} 1, & t = i(\text{mod } s), \\ 0, & \text{그 외의 경우,} \end{cases} \quad (2.6)$$

여기서

$$a_t = \phi_1 a_{t-1} + \phi_2 a_{t-2} + \dots + \phi_p a_{t-p} + \epsilon_t, \quad (2.7)$$

$\{\epsilon_t\}$ 는 평균 0, 분산 σ_ϵ^2 을 갖는 백색잡음과정(white noise process)이다.

2.3. 승법계절 ARIMA모형

Box와 Jenkins (1976)의 계절주기 s 를 갖는 승법계절 ARIMA(aotoregressive integrated moving average)모형 ARIMA(p, d, q)(P, D, Q) $_s$ 는 식 (2.8)과 같이 정의된다. 이때 B 는 $B^j Z_t = Z_{t-j}$ 를 만족하는 후진작용소이다.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t = \delta + \theta(B)\Theta(B^s)\epsilon_t, \quad (2.8)$$

여기서 $\{\epsilon_t\}$ 는 평균 0, 분산 σ_ϵ^2 을 가지는 백색잡음과정이고,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p, \quad (2.9)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q, \quad (2.10)$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \Phi_3 B^{3s} - \dots - \Phi_P B^{Ps},$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \Theta_3 B^{3s} - \dots - \Theta_Q B^{Qs}.$$

2.4. 전이함수모형

Box와 Jenkins (1976)의 전이함수(transfer function)모형은 입력시계열을 갖는 ARMA오차 회귀모형이라는 의미에서 ARMAX모형이라고도 부른다. 특히, 이상점과 같은 개입효과도 입력변수로 사용하여 전이함수모형을 표현할 수 있다 (Wei, 1990). k 개의 입력시계열, $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ 와 m 개의 이상점, $I_{1,t}, I_{2,t}, \dots, I_{m,t}$ 을 갖는 경우, ARMA(p, q)오차를 갖는 전이함수모형은 식 (2.11)과 같이 정의된다.

$$Y_t = \sum_{i=1}^k \frac{\omega_{i,s}(B)}{\delta_{i,r}(B)} B^{b_i} X_{i,t} + \sum_{j=1}^m \frac{\eta_{j,u}(B)}{\nu_{j,z}(B)} B^{b_j} I_{j,t} + \frac{\theta(B)}{\phi(B)} \epsilon_t, \quad (2.11)$$

여기서 $\{\epsilon_t\}$ 는 평균 0, 분산 σ_ϵ^2 을 가지는 백색잡음과정이고, $\phi(B)$ 와 $\theta(B)$ 는 식 (2.9)와 식 (2.10)에서 각각 정의되며,

$$\begin{aligned} \omega_{i,s}(B) &= \omega_{i,0} - \omega_{i,1}B - \omega_{i,2}B^2 - \dots - \omega_{i,s}B^s, \\ \delta_{i,r}(B) &= \delta_{i,0} - \delta_{i,1}B - \delta_{i,2}B^2 - \dots - \delta_{i,r}B^r, \\ \eta_{j,u}(B) &= \eta_{j,0} - \eta_{j,1}B - \eta_{j,2}B^2 - \dots - \eta_{j,u}B^u, \\ \nu_{j,z}(B) &= \nu_{j,0} - \nu_{j,1}B - \nu_{j,2}B^2 - \dots - \nu_{j,z}B^z. \end{aligned}$$

식 (2.11)에서 $I_{j,t}$ 는 개입변수(intervention variable)로서 어떤 사건이 T 시점에서 발생하여 그 효과가 T 시점에만 영향을 미치는 경우 식 (2.12)와 같은 펄스함수(pulse function)로 정의된다.

$$I_{j,t} = \begin{cases} 1, & t = T, \\ 0, & t \neq T. \end{cases} \quad (2.12)$$

3. 자료분석

본 절에서는 제 2절에서 소개된 시계열모형을 적합시켜 국내 전력판매량을 예측하였다. 모든 시계열 분석은 SAS/ETS를 사용하여 수행하였다. 특히 Dickey-Fuller 검정을 위하여 DFTEST macro program을, Winters 계절지수평활을 위하여 PROC FORECAST를, AR오차 회귀모형 분석을 위하여 PROC AUTOREG를, 승법계절 ARIMA모형 및 전이함수모형 분석을 위하여 PROC ARIMA를 사용하였다.

3.1. 자료

전이함수모형에서는 출력시계열인 전력판매량을 예측하기 위하여 입력시계열이 필요하다. 본 연구에서 처음에 고려하였던 입력시계열은 전력수용가수(단위: 천호), 전력가격(단위: 원/Kwh), 산업생산지수, 서울지역 월평균기온(단위: °C), 냉방도일(단위: 일수), 난방도일(단위: 일수)이었다. 냉(난)방도일은 매월 우리나라 일평균기온 18°C를 기준으로 이보다 높으면(낮으면) 18°C와의 기온 차이를 누적시킨 값이다 (산업생산지수는 국가통계포털(KOSIS, <http://kosis.kr/>)에서, 그 외의 모든 자료는 국가에너지통계정보시스템(KESIS, <http://www.kesis.net>)에서 얻을 수 있다.). 그러나 모든 시계열이 전력판매량의 예측을 위한 입력시계열로 유용하지 않았다. 최종적으로 냉방도일과 난방도일을 합한 냉난방도일 시계열을 만들어 사용하였다. Figure 3.1에서 기온(temperature) 및 냉난방도일(cooling and heating days)은 계절성분만을 가짐을 알 수 있다. 기온보다는 냉난방도일의 계절성분 모양이 전력판매량(electricity sales)과 유사한 패턴을 보인다. 이러한 이유로 원래 냉방도일 및 난방도일 시계열이 각각

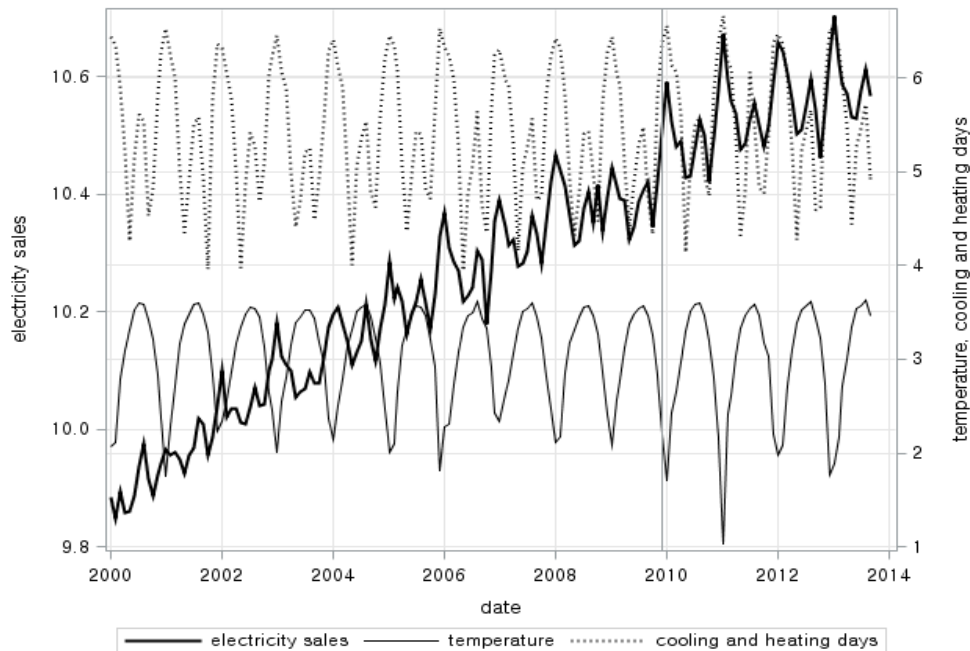


Figure 3.1. Time series plot of electricity sales, temperature, cooling and heating days(log transformed data)

생산되지만 본 연구에서는 두 시계열을 합한 냉난방도일 시계열을 사용하였다. 모형적합을 위하여 사용된 훈련용자료(training data)는 2000년 1월부터 2011년 12월까지의 월별 시계열이며, 예측검증을 위한 검증용자료(test data)로 2012년 1월부터 2013년 12월까지 2년 동안의 월별 시계열자료를 사용하였다. 모든 자료는 로그변환하여 모형적합에 사용하였으며 예측력 평가를 할 때는 원자료로 변환하여 평가하였다.

3.2. 시계열자료의 정상화

ARIMA모형이나 전이함수모형에 의한 시계열분석은 정상시계열을 대상으로 한다. 따라서 비정상성을 가지는 시계열은 정상시계열로 변환 후 분석이 진행된다. 특히 시계열의 차분(differencing)에 의하여 정상시계열이 된다면 그 시계열은 단위근을 갖는다고 말한다. Dickey와 Fuller (1979), Dickey 등 (1984)에 의한 Dickey-Fuller 검정은 “ H_1 : 시계열은 정상이다.”에 대하여 “ H_0 : 시계열은 1시차(혹은 계절 시차)에서 단위근을 갖는다. 즉, 시계열은 직선추세(혹은 계절추세)를 갖는 비정상 시계열이다.”를 검정한다. 또한 그래프 방법으로서 시계열의 표본자기상관함수(SACF; sample autocorrelation function)가 시차 1에서는 1에 가까운 큰 값이지만 시차가 커질수록 서서히 감소하고, 표본편자기상관함수(SPACF; sample partial autocorrelation function)는 시차 1에서 1에 가까운 큰 값을 가지고 그 외 다른 시차에서는 거의 0에 가까운 값을 보이면 1차 차분을 필요로 한다고 판단할 수 있다. 계절 주기가 s 일 때 계절차분을 필요로 하는 경우는 시계열의 SACF가 시차 s 에서는 1에 가까운 큰 값이지만 시차가 $2s, 3s, \dots$ 로 갈수록 서서히 작아지며, SPACF는 시차 s 에서 1에 가까운 큰 값을 가지고 그 외 다른 시차 $2s, 3s, \dots$ 에서는 거의 0에 가까운 값을 보이면 s 차 차분, 즉, 계절차분을 필요로 한다고 판단할 수 있다.

3.3. 모형적합도검정

모형이 추정된 후에는 모형에 대해 설정하였던 가정이 만족되는지를 검토하는 모형적합도검정을 실시해야 한다. 제 2절에서 소개된 각 시계열모형에 대한 기본 가정인 “ $H_0 : \{\epsilon_t\}$ 가 백색잡음과정이다”는 “ $H_0 : \rho_\epsilon(1) = \rho_\epsilon(2) = \dots = \rho_\epsilon(K) = 0$ ”와 같은 의미이다. 여기서 $\rho_\epsilon(k)$ 는 $\{\epsilon_t\}$ 들의 k -시차 자기상관함수(ACF; autocorrelation function)이다. 이 가설을 검정하기 위한 포트맨토우검정(Portmanteau test)은 다음과 같은 검정통계량을 사용한다 (Wei, 1990).

$$Q = n(n+2) \sum_{k=1}^K (n-k)^{-1} \hat{\rho}_\epsilon^2(k),$$

여기서 n 은 자료의 수, $\hat{\rho}_\epsilon^2(k)$ 는 잔차 $\{\hat{\epsilon}_t\}$ 들의 k -시차 SACF, 그리고 K 는 검정에서 고려해야할 ACF의 개수이다. H_0 하에서 Q 통계량은 근사적으로 $\chi^2(K)$ 분포를 따른다.

3.4. 모형평가

모형적합도검정에서 적합모형으로 판정되면 예측모형으로 사용할 수 있다. 각 예측모형에 대하여 적합 및 예측력 비교를 위한 척도로 사용된 제공근평균제곱오차(RMSE; root mean square error)와 평균절대백분위오차(MAPE; mean absolute percentage error)는 각각 다음과 같이 정의된다.

$$\text{RMSE} = \sqrt{\frac{1}{l} \sum_t (Y_t - \hat{Y}_t)^2},$$

$$\text{MAPE} = \frac{1}{l} \sum_t \text{APE}_t = \frac{1}{l} \sum_t \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100,$$

여기서 l 은 예측(적합)기간의 길이, Y_t 는 실제 관측자료, \hat{Y}_t 는 예측(적합)값이며, APE_t 는 특정 시간 t 에서의 절대백분위오차(APE; absolute percentage error)이다. APE, MAPE의 단위는 %로서 실제 자료에 대한 예측(적합)오차의 백분율을 의미한다.

3.5. 시계열모형의 적합

3.5.1. Winters 계절지수평활법 전력판매량(Y_t)에 대하여 식 (2.3), (2.4), (2.5)의 승법계절지수평활법을 적용할 때 필요한 평활상수 $\omega_1, \omega_2, \omega_3$ 는 1시차 예측오차의 제곱합인 $\text{SSE}(\omega_1, \omega_2, \omega_3) = \sum_{t=2}^n (Y_t - \hat{Y}_{t-1}(1))^2$ 을 최소화 해주는 값인 $\omega_1 = 0.2, \omega_2 = 0.1, \omega_3 = 0.6$ 을 사용하였다.

3.5.2. AR오차 회귀모형 전력판매량은 직선추세를 가지므로 시간의 직선 함수로, 그리고 계절주기 $s = 12$ 를 갖는 월별 계절성분은 시간의 지시함수로 표현한 AR오차 회귀모형 추정식은 식 (3.1)과 같다. 모든 모수 유의성검정에서 p -value < 0.0001 으로서 모수들은 매우 유의하다.

$$\begin{aligned} \hat{Y}_t = & 0.004724t + 9.9775I_{t,1} + 9.9284I_{t,2} + 9.9121I_{t,3} + 9.8926I_{t,4} + 9.8441I_{t,5} + 9.8539I_{t,6} \\ & + 9.8790I_{t,7} + 9.9157I_{t,8} + 9.8795I_{t,9} + 9.8371I_{t,10} + 9.8771I_{t,11} + 9.9317I_{t,12} + \hat{a}_t, \end{aligned} \quad (3.1)$$

여기서

$$\hat{a}_t = 0.5498\hat{a}_{t-1} + 0.1561\hat{a}_{t-11}, \quad \hat{\sigma}_\epsilon^2 = 0.000892.$$

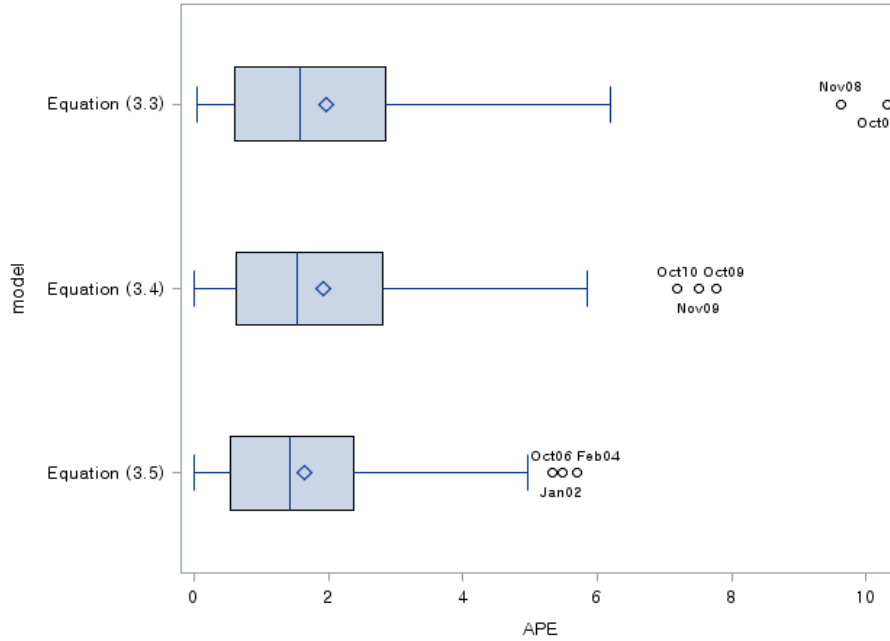


Figure 3.2. Box plot of APE for transfer function models

3.5.3. 승법계절 ARIMA모형 Dickey-Fuller 검정, 그리고 SACF와 SPACF의 그래프 판단에 의해 전력판매량은 1차 및 12차분을 통하여 정상시계열을 얻을 수 있다. 이제 $(1 - B)(1 - B^{12})Y_t$ 에 대한 $ARIMA(p, 1, q)(P, 1, Q)_{12}$ 모형의 적합에서 모수절약의 법칙을 따르기로 한다. 따라서 $p, q, P, Q \leq 2$ 인 작은 차수들 중에서 모수의 유의성검정 및 잔차분석 결과를 고려하여 최종적으로 추정모형 $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ 모형은 식 (3.2)와 같다. 모든 모수 유의성검정에서 p -value < 0.0001 이므로 모수들은 매우 유의하다.

$$(1 - B)(1 - B^{12})Y_t = (1 - 0.5397B)(1 - 0.5992B^{12})\epsilon_t, \quad \hat{\sigma}_\epsilon^2 = 0.000752. \quad (3.2)$$

3.5.4. 전이함수모형 전이함수모형에 사용된 입력변수는 냉난방도일(X_t)이다. 냉난방도일은 Dickey-Fuller 검정, 그리고 SACF와 SPACF의 그래프 판단에 의해 12차분을 통하여 정상시계열을 얻을 수 있다. 모형식 (2.11)에서 펄스함수, $I_{j,t}$ 를 사용하지 않고 오직 한개의 입력변수인 냉난방도일을 사용한 전이함수모형의 추정식은 식 (3.3)과 같다. 추정모수 아래 괄호 안의 수치는 모수 유의성검정에 대한 p -value이다.

$$(1 - B)(1 - B^{12})Y_t = \underset{(0.0053)}{0.0272}(1 - B^{12})X_t + (1 - \underset{(<0.0001)}{0.5141}B)(1 - \underset{(<0.0001)}{0.6001}B^{12})\epsilon_t, \\ \hat{\sigma}_\epsilon^2 = 0.000713. \quad (3.3)$$

Figure 3.2의 첫 번째 상자그림은 전이함수모형 식 (3.3)에 대한 잔차분석 결과로서 APE의 분포를 보여준다. 2008년 10월 및 11월에 APE가 각각 10.3144% 및 9.6189%에 이른다. 이러한 수치는 전체 적합기간에서 APE의 평균이 1.9719%, 표준편차가 1.7499%, 그리고 5% 및 95% 분위수가 각각 0.1738%

및 5.4143%인 기술통계분석 결과에 비추어 이례적이다 할 수 있다. 2008년 10월 및 11월에 발생하는 두개의 펄스함수를 추가한 전이함수모형의 추정에서 2008년 11월의 펄스함수에 상응하는 모수추정치로 인하여 추정치들의 상관행렬이 비정칙행렬이 되어 추정 결과를 신뢰할 수 없다. 따라서 2008년 10월에 발생하는 하나의 펄스함수만 사용하여 추정한 전이함수모형은 식 (3.4)와 같다. 추정모수 아래 괄호 안의 수치는 모수 유의성검정에 대한 p -value이다.

$$(1 - B)(1 - B^{12})Y_t = \underset{(0.0026)}{0.0284}(1 - B^{12})X_t + \underset{(0.0153)}{0.0553} - \underset{(0.0011)}{0.0750}B I_t + (1 - \underset{(<0.0001)}{0.5099} B)(1 - \underset{(<0.0001)}{0.5052} B^{12})\epsilon_t, \quad \hat{\sigma}_\epsilon^2 = 0.000671, \quad (3.4)$$

여기서

$$I_t = \begin{cases} 1, & t = 2008, \text{October}, \\ 0, & t \neq 2008, \text{October}. \end{cases}$$

Figure 3.2의 두 번째 상자그림은 전이함수모형 식 (3.4)에 대한 APE의 분포로서 전이함수모형 식 (3.3)에 대한 APE의 분포보다 산포가 축소되어 있음을 보여준다. 그러나 여전히 2009년 10월, 11월, 그리고 2010년 10월에서 APE가 각각 7.7632%, 7.5206%, 7.1963%에 이른다. 이제 기존의 2008년 10월, 그리고 새롭게 2009년 10월, 11월, 그리고 2010년 10월에 발생하는 모두 4개의 펄스함수로 구성된 전이함수모형의 추정에서 모수 유의성검정 등을 통하여 5% 유의수준에서 유의한 변수들로 구성된 전이함수모형은 식 (3.5)와 같다.

$$(1 - B)(1 - B^{12})Y_t = \underset{(0.0005)}{0.0281}(1 - B^{12})X_t + \underset{(<0.0001)}{0.1149} - \underset{(<0.0001)}{0.1554} B I_{1,t} - \underset{(<0.0001)}{0.1153} I_{2,t} + \underset{(<0.0001)}{0.1594} I_{3,t} + (1 - \underset{(<0.0001)}{0.4301} B)(1 - \underset{(<0.0001)}{0.4738} B^{12})\epsilon_t, \quad \hat{\sigma}_\epsilon^2 = 0.000464, \quad (3.5)$$

여기서

$$I_{1,t} = \begin{cases} 1, & t = 2008, \text{October}, \\ 0, & t \neq 2008, \text{October}. \end{cases} \quad I_{2,t} = \begin{cases} 1, & t = 2009, \text{October}, \\ 0, & t \neq 2009, \text{October}. \end{cases} \quad I_{3,t} = \begin{cases} 1, & t = 2009, \text{November}, \\ 0, & t \neq 2009, \text{November}. \end{cases}$$

Figure 3.2의 세 번째 상자그림은 전이함수모형 식 (3.5)를 적용한 후의 APE 분포로서 추정도차가 많이 줄어 들었음을 확인할 수 있다. 이때 APE의 평균은 1.6539%, 표준편차는 1.2829%, 그리고 5% 및 95% 분위수가 각각 0.05537% 및 3.7429%이다. 세 번째 상자그림에서 보여주는 2002년 1월, 2004년 2월, 2006년 10월의 펄스함수를 추가한 모형 추정에서는 모수의 유의성검정 결과 추가된 3개의 펄스함수 모두 5% 유의수준 하에서 유의하지 않았다.

3.6. 잔차분석

지금까지 추정된 각 모형에 대한 모형적합도검정 즉, 포트맨토우검정 결과는 Table 3.1과 같다. Winters 계절지수평활법에 의한 적합 후 잔차들은 자기상관관계가 매우 큼을 알 수 있다. 그 외의 모형들에 의한 적합 후 잔차들은 독립이라고 볼 수 있다.

3.7. 예측모형의 비교

전이함수모형에 의한 예측을 위해서는 입력시계열인 냉난방도일(X_t)의 2012년 1월부터 2013년 12월까지 24개월 예측값들이 필요하다. 냉난방도일은 모수절약의 법칙, 모수 유의성검정 및 잔차분석 등을 통

Table 3.1. p -value for goodness-of-fit test at time lag K

model	equation	K					
		6	12	18	24	30	36
Winters exponential smoothing		0.0051	0.0059	0.0320	0.0009	0.0014	0.0042
AR error regression	(3.1)	0.7655	0.5543	0.7814	0.3010	0.1862	0.2217
multiplicative seasonal ARIMA	(3.2)	0.6453	0.6678	0.8623	0.1377	0.1500	0.1965
transfer function	(3.3)	0.9372	0.8553	0.9685	0.3261	0.3764	0.3988
transfer function with a pulse	(3.4)	0.8182	0.8772	0.9067	0.2071	0.2943	0.3612
transfer function with 3 pulses	(3.5)	0.3827	0.7336	0.8079	0.4446	0.3812	0.4127

Table 3.2. RMSE for the fitted period and the prediction period

model	equation	fitted period	prediction period		
		2000–2011	2012	2013	2012–2013
Winters exponential smoothing		890	816	672	747
AR error regression	(3.1)	679	1921	3272	2683
multiplicative seasonal ARIMA	(3.2)	817	880	1420	1181
transfer function	(3.3)	801	790	731	761
transfer function with a pulse	(3.4)	776	826	336	630
transfer function with 3 pulses	(3.5)	608	743	378	589

Table 3.3. MAPE for the fitted period and the prediction period

model	equation	fitted period	prediction period		
		2000–2011	2012	2013	2012–2013
Winters exponential smoothing		2.2216	1.5486	1.4643	1.5065
AR error regression	(3.1)	1.5605	4.4284	8.0483	6.2384
multiplicative seasonal ARIMA	(3.2)	2.0587	1.7087	3.4609	2.5848
transfer function	(3.3)	1.9719	1.4077	1.6417	1.5247
transfer function with a pulse	(3.4)	1.9199	1.6020	0.7401	1.1711
transfer function with 3 pulses	(3.5)	1.6539	1.4251	0.8249	1.1250

하여 다음과 같은 ARIMA(0, 0, 0)(0, 1, 1)₁₂ 모형에 의하여 예측하였다. 추정모수 아래 괄호 안의 수치는 모수 유의성검정에 대한 p -value이다.

$$(1 - B^{12})X_t = (1 - \underset{(<0.0001)}{0.7800} B^{12})\epsilon_t, \quad \hat{\sigma}_\epsilon^2 = 0.039303.$$

Table 3.2와 Table 3.3은 모형 적합기간과 예측기간 동안의 RMSE와 MAPE를 각각 보여준다. AR오차 회귀모형은 예측기간보다 적합기간에서 더 정확하여 과적합되어 있음을 알 수 있다. AR오차 회귀 모형 적합 후 잔차에 대한 이분산 검정결과는 강한 이분산의 징후를 보인다. 그러나 이분산 모형 AR-GARCH모형에서 대부분 차수의 적합에는 실패하였다. Winters 계절지수평활법은 2년 예측기간 동안 MAPE가 1.5065%의 좋은 결과를 보여준다. 이상점을 수용하는 전이함수모형 식 (3.4)와 (3.5)는 2년 예측기간 동안 MAPE가 각각 1.1711% 및 1.1250%로서 매우 우수하며, 특히 2013년의 예측오차는 각각 0.7401% 및 0.8249%로서 매우 정확함을 알 수 있다. RMSE도 MAPE와 비슷한 결과를 보여준다.

Figure 3.3에서 세로축의 Difference(APE)는 펄스함수를 포함하지 않은 전이함수모형 식 (3.3)을 사용했을 때의 APE에서 3개의 펄스함수를 포함한 전이함수모형 식 (3.5)를 사용했을 때의 APE를 뺀 수치이다. 지면 관계상 생략된 2000년 1월부터 2004년 12월까지의 Difference(APE)는 2005년 1월부터 2008년 9월까지의 패턴과 거의 비슷하다. 즉, 두개의 모형은 펄스함수를 사용하기 전까지는 성능에 차

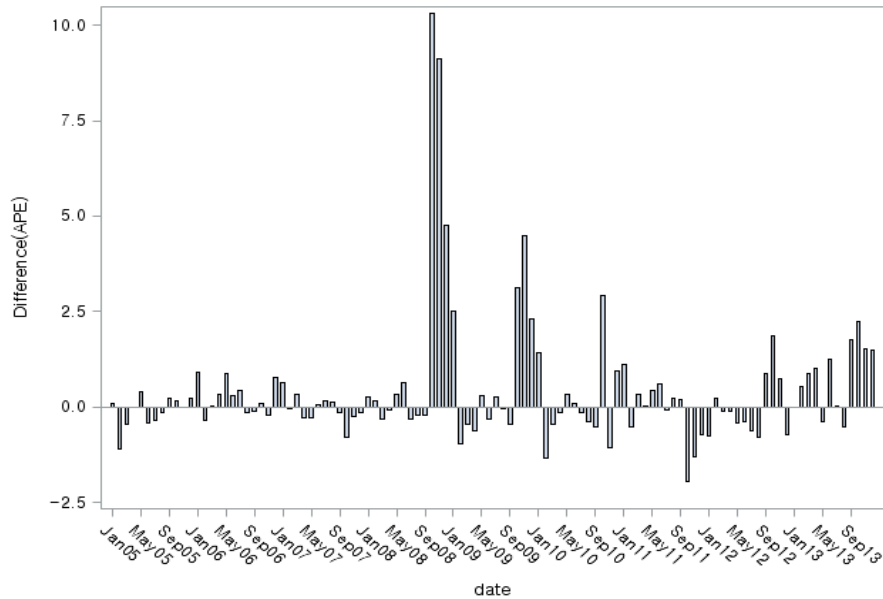


Figure 3.3. Difference of APE between the transfer function model equation (3.3) and (3.5)

이가 거의 없다. 그러나 적합구간에서 펄스함수를 사용한 2008년 10월부터 4개월, 2009년 10월부터 4개월동안 APE의 감소가 두드러진다. 예측구간에서 2011년 10월 이후 4개월은 오히려 펄스함수를 사용한 전이함수 모형의 성능이 떨어졌으나 2012년 10월부터 성능이 더 좋아지기 시작하여 특히 2013년에는 2개월을 제외하고 전 구간에서 보다 우수하였다. 이는 Figure 3.1에서 전체적인 전력판매량의 추세를 볼때 2008년과 2009년 후반부에 발생한 전력판매량의 침체가 다시 2012년 및 2013년 후반부에 재현되어 보이는 것과 연관지어 볼 수 있다.

4. 결과

전력수급 예측의 핵심은 최대전력수요 예측과 전력판매량 예측에 있다. 본 연구에서는 전력판매량의 예측을 다루었다. 전력을 판매하는 한국전력 입장에서의 전력판매량은 전력을 구매하여 소비하는 국민 혹은 산업체의 입장에서의 전력소비량과 같은 의미이다. 주식 투자자가 미래의 전력판매량을 정확하게 예측할 수 있다면 한국전력의 주식을 거래하는 데 도움이 될 것이다. 전력소비량 예측이 정확하지 못하면 전력난으로 국민들의 일상 생활이 불편할 뿐만 아니라 산업 활동이 침체된다. 정부의 입장에서는 발전 설비, 수요관리, 그리고 전력구입 비용 등이 급증한다. 전력소비량의 정확한 예측이 이루어진다면 정부는 예측치에 기반한 전력수급계획 하에 시간이 걸리더라도 저비용의 기저설비를 건설하여 미래를 대비할 것이다. 그러나 예측이 부정확하여 실제 전력소비량이 많아지면 정부는 고비용일지라도 건설기간이 짧은 발전소를 지어 전력공급을 하려 할 것이며, 전력감축을 위하여 절전 산업체에 지급하는 전력부하관리 지원금의 지출, 그리고 전력부족분을 충당하기 위하여 공기관과 민간 발전사들에게서 사들이는 전력구입액이 급증할 수 밖에 없다.

본 연구에서는 우리나라 전력수급 예측의 핵심이라 할 수 있는 전력판매량(전력소비량)에 대해서 다양한 시계열모형에 의한 예측을 시도하였다. 시계열자료에 대하여 회귀모형이나 평활법을 적용하는 경우

에 잔차들에는 여전히 자기상관관계가 남아 있는 경우가 대부분이다. 그것은 모형식이 자료의 자기상관관계를 수용하는 구조가 아니기 때문이다. 잔차에 남아있는 자기상관관계를 확률모형에서 표현해줄 수 있다면 보다 정확한 예측을 할 수 있을 것이다. 이것이 자기상관관계를 확률모형에서 수용하는 시계열모형을 사용하여야 하는 이유이다.

전이함수모형의 사용에서 핵심은 출력시계열을 잘 설명해줄 수 있는 입력시계열의 선택이다. 월별 냉(난)방도일은 우리나라 일평균기온 18℃를 기준으로 이보다 높으면(낮으면) 18℃와의 기온 차이를 한 달 동안 누적시킨 월별 시계열이다. 즉, 냉(난)방도일이 높은 달은 많은 전력을 필요로 한다. 따라서 냉방도일과 난방도일을 합한 시계열인 냉난방도일은 단순한 기온에 비해서 매우 우수한 입력변수 역할을 함을 자료분석에서 알 수 있다. 미국의 서브프라임모기지 사태로 촉발된 2008년 9월의 세계금융위기로 인하여 우리나라의 대부분 경제시계열은 2008년 9월 이후 수개월 동안 정상적인 시계열패턴에서 벗어남을 볼 수 있다. 실제 자료분석을 통하여 입력시계열로서 냉난방도일, 그리고 2008년 10월, 혹은 추가로 2009년 9월, 10월에 발생하는 개입변수로 펄스함수를 사용한 전이함수모형이 다른 시계열모형에 비해서 RMSE 및 MAPE의 의미에서 더 우수하였다.

References

- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, Prentice Hall.
- Cho, S. and Son, Y. S. (2011). *Time Series Analysis using SAS/ETS*, Yulgokbooks.
- Dickey, D. A. and Fuller, W. A. (1979). Distribution of estimation for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.
- Dickey, D. A., Hasza, D. P. and Fuller, W. A. (1984). Testing for unit roots in seasonal time series, *Journal of the American Statistical Association*, **79**, 355–367.
- Korea Power Exchange (2010). *Prediction of Electricity Supply and Demand for the Next Two Years(2010.07~2010.12)*, Report, Korea Power Exchange.
- Lee, J., Sohn, H. and Kim, S. (2013). Daily peak load forecasting for electricity demand by time series models, *The Korean Journal of Applied Statistics*, **26**, 349–360.
- Ministry of Knowledge Economy (2013). *The 6th Basic Plan for Long-term Electricity Supply and Demand (2013~2027)*, MKE Notice 2013-63, Ministry of Knowledge Economy.
- Montgomery, D. C. and Johnson, L. A. (1976). *Forecasting and Time Series Analysis*, McGraw-Hill, Inc.
- Nam, B. W., Song, K. B., Kim, K. H. and Cha, J. M. (2008). The spatial electric load forecasting algorithm using multiple regression analysis method, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, **22**, 63–70.
- SAS Institute Inc. (2008). *SAS/ETS 9.2 User's Guide*, SAS Institute Inc. Cary, NC.
- Song, K. B. (2007). Various models of Fuzzy Least-Squares Linear Regression for load forecasting, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, **21**, 61–67.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799–805.
- Wei, W. W. S. (1990). *Time Series Analysis*, Addison-Wesley Publishing Company.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages, *Management Science*, **6**, 324–342.
- Yoon, S. H., Lee, Y. S. and Park, J. (2009). Statistical modeling for forecasting maximum electricity demand in Korea, *Communications of the Korean Statistical Society*, **16**, 127–135.

시계열모형에 의한 전력판매량 예측

손영숙^{a,1}

^a전남대학교 통계학과

(2014년 2월 10일 접수, 2014년 3월 31일 수정, 2014년 5월 20일 채택)

요약

전력수급의 정확한 예측은 국민들의 일상적 생활 유지, 산업활동, 그리고 국가경쟁을 위하여 매우 중요하다. 본 연구에서는 시계열모형화에 의해 전력판매량을 예측한다. 실제 자료분석을 통하여 입력시계열로서 냉난방도일과 개입 변수로 펄스함수를 사용한 전이함수모형이 다른 시계열모형에 비해서 제공근평균제곱오차 및 평균절대오차의 의미에서 더 우수하였다.

주요용어: 전력판매량, Winters 계절지수평활법, AR오차 회귀모형, 승법계절 ARIMA모형, 전이함수 모형, 개입변수, 펄스함수, 냉난방도일.

본 연구는 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2011-0022864).

¹(500-757) 광주광역시 북구 용봉동 300번지, 전남대학교 통계학과. E-mail: ysson@jnu.ac.kr