

A Study on Demand Forecasting for KTX Passengers by using Time Series Models

In-Joo Kim^a · Hueng-goo Sohn^b · Sahm Kim^{b,1}

^aDepartment of Hotel Tourism Service, Dae-Duk College

^bDepartment of Applied Statistics, Chung-Ang University

(Received November 18, 2014; Revised November 22, 2014; Accepted November 22, 2014)

Abstract

Since the introduction of KTX (Korea Train eXpress) in Korea railway market, number of passengers using KTX has been greatly increased in the market. Thus, demand forecasting for KTX passengers has been played a important role in the train operation and management.

In this paper, we study several time series models and compare the models based on considering special days and others.

We used the MAPE (Mean Absolute Percentage Errors) to compare the performance between the models and we showed that the Reg-AR-GARCH model outperformed other models in short-term period such as one month. In the longer periods, the Reg-ARMA model showed best forecasting accuracy compared with other models.

Keywords: Taylor's Holt-Winters model, Reg-ARMA model, Reg-AR-GARCH model, MAPE, KTX Passenger.

1. 서론

우리나라의 철도여객 시장은 KTX에 등장에 따라 국내 여객시장은 여객기에 비하여 저렴하고, 접근이 용이한 KTX시장의 이용변동이 이루어졌다. 최근 3년간(2011년~2013년) KTX 수송인원은 8.98%의 증가율을 보이며 국내 여객시장에서 가장 보편적인 여객방법으로 자리매김하였다. 이러한 KTX의 증가율이 나타남에 따라 효율적인 철도운영을 위해서는 열차 배차 의사결정의 기초자료인 철도수요에 대한 예측이 필요하다.

철도수요의 예측방법으로 과거 시계열 적인 요소를 고려한 회귀모형 및 패턴인식 방법 중 단변량 시계열 모형이 고려되어 왔다. 이러한 단변량 시계열 방법은 단기간 내에 수요 예측이 가능할 뿐만 아니라 예측의 정확도도 다른 방법에 비해 우수하다. 그러나 명절과 휴일, 금요일의 급격한 변동에 대한 적용이 어렵다. 이에 따라 단변량 시계열 분석 결과를 바로 철도수요예측에 활용할 수 없는 현실이다.

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 (2014-027173).

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr

시계열모형을 이용한 철도수요예측의 국내 연구는 Lee 등 (1998)에서 데이터 처리방법(Data Granulization)을 이용하여 정적모형과 동적모형으로 나누어 연구되었으며, Oh 등 (2001)는 X-12-ARIMA를 통한 추세분해를 통하여 열차수송에 미치는 영향을 판단하였다. Choi 등 (2004)는 Holt-Winters 모형을 적용하여 시계열의 계절성과 추세를 수요예측에 반영하였다. 그러나 Holt-Winters 방법은 단변량 시계열 기법으로서 타 독립변수를 고려하지 못하는 단점이있다. Kim 등 (2011b)은 주중 여객수와 주말 여객수에 대한 월별 자료를 이용하여 개입 ARIMA 모형을 적용하였다. 이에 대하여 자료는 월별 자료를 이용하여 일자별 세부적인 변동에 대하여는 고려하지 않았다. Kim 등 (2011a)은 KTX 단기 수요 예측을 위하여 군집 분석을 통하여 통행 행태에 대하여 분석을 실시하였다. 국외 연구는 Tsai 등 (2005)이 대만철도 1999, 2000년 수송실적으로 신경망모형을 구축하여 수요를 예측 하였으며, 요일별, 월별 특성을 반영하여 비선형 자기회귀 신경망(nonlinear autoregressive neural network)을 제안하였고, Marco Lippi (2013)가 SVR모형과 ARIMA모형을 이용한 역별 열차수송에 대한 연구를 실시하였으며, Wei 등 (2013)은 Holt-Winters 모형을 이용한 철도 이용객 예측을 실시하였다.

본 연구는 KTX 수요를 예측하기 위한 방법으로 일별 자료를 이용한 Reg-ARIMA 모형을 적용하고자 한다. Reg-ARIMA모형은 단변량 분석인 ARIMA 모형과 독립변수를 이용한 회귀분석의 방법의 결합 형식으로, 독립변수로 명절과 주말요소 변수(금,토,일), 명절과 주말요소를 제외한 휴일 변수를 사용한다.

본 연구는 2장에서 제안된 모형을 이용한 연구방법에 대해 설명하며, 3장에서는 분석 데이터에 대한 설명과 모형별 예측력에 대해 비교해보고, 4장에서는 결론으로 끝맺음 할 것이다.

2. 시계열모형의 소개

2.1. Taylor의 Holt-Winters 지수평활모형

Holt-Winters 지수평활법은 Holt (1957)의 이중 모수 지수평활법을 Winters (1960)가 확장시킨 방법으로 계절형 시계열 예측에 사용되는 방법이다. 지수 평활법은 계산이 쉽고 필요정보의 양이 최소화 할 수 있는 장점이 있어 유용하다. Y_t 는 고려할 t 시점의 일별 KTX 이용객수로 계절주기를 s 라고 하면 L_t 는 t 시점의 평활수준을 나타낸다. T_t 는 t 시점의 평활가법추세를 의미하며, S_t 는 t 시점의 계절지표를 의미한다. α 는 수준모, β 는 추세에 대한 모수, γ 는 계절평활 모수이다. F_{t+h} 는 h 시점 후에 예측한 값을 의미하며 이는 식 (2.1)과 같이 적용된다.

$$\begin{aligned} L_t &= \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\ S_t &= \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}, \\ F_{t+h} &= L_t + T_t \times h + S_{t+h-s}. \end{aligned} \quad (2.1)$$

이중 계절성을 적용시킨 계절형 Holt-Winters 지수평활법은 식 (2.2)와 같이 적용된다. 여기서 Y_t 는 고려할 t 시점의 일별 KTX 이용객수로 계절주기를 s_1 과 s_2 라 하면 L_t 는 t 시점의 평활수준을 나타내며, T_t 는 t 시점의 평활 가법추세를 의미하며, S_t 과 D_t 는 t 시점의 계절주기 s_1 과 s_2 의 계절지표를 의미한

다.

$$\begin{aligned}
 L_t &= \alpha(Y_t - S_{t-s_1} - D_{t-s_2}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\
 T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\
 S_t &= \gamma(Y_t - L_t - D_{t-s_2}) + (1 - \gamma)S_{t-s_1}, \\
 D_t &= \delta(Y_t - L_t - S_{t-s_1}) + (1 - \delta)D_{t-s_2}, \\
 F_{t+h} &= L_t + T_t \times h + S_{t+h-s_1} + D_{t+h-s_2}.
 \end{aligned} \tag{2.2}$$

식 (2.2)의 초기값은 식 (2.3)의 계산을 통하여 나타낸다.

$$\begin{aligned}
 L_{s_1} &= \frac{1}{s_1} \sum_{t=1}^{s_1} Y_t, \\
 L_{s_2} &= \frac{1}{s_2} \sum_{t=1}^{s_2} Y_t, \\
 T_{s_1} &= \frac{1}{s_1^2} \left(\sum_{t=s_1+1}^{2s_1} Y_t - \sum_{t=1}^{s_1} Y_t \right), \\
 T_{s_2} &= \frac{1}{s_2^2} \left(\sum_{t=s_2+1}^{2s_2} Y_t - \sum_{t=1}^{s_2} Y_t \right), \\
 S_1 &= Y_1 - L_{s_1}, \dots, S_{s_1} = Y_{s_1} - L_{s_1}, \\
 D_1 &= Y_1 - L_{s_2}, \dots, D_{s_2} = Y_{s_2} - L_{s_2}.
 \end{aligned} \tag{2.3}$$

평활 모수 $\alpha, \beta, \gamma, \delta$ 의 추정은 이 m 개의 모형추정자료를 통하여 추정가능하다. 모형추정단계 중 1차 시점 예측에서 SSE(Sum of Squared Error)를 최소화하는 과정을 통하여 계산되며, SSE는 식 (2.4)에 의하여 계산된다.

$$\text{SSE} = \sum_{t=2}^m (Y_t - F_t)^2. \tag{2.4}$$

이에 대하여 Taylor의 이중계절 Holt-winters 지수평활법은 Taylor (2003)에 의하여 제안되었다. 수정 Holt-Winters 지수평활법은 식 (2.2)에서 F_{t+h} 를 식 (2.5)과 같이 변형하여 정의한다.

$$F_{t+h} = L_t + T_t \times h + S_{t+h-s_1} + D_{t+h-s_2} + \phi^h [Y_t - L_{t-1} - T_{t-1} - S_{t-s_1} - D_{t-s_2}]. \tag{2.5}$$

ϕ 는 1차 상관계수에 조정을 나타내는 값으로 평활 모수 $\alpha, \beta, \gamma, \delta, \phi$ 는 1차 예측단계에서 SSE(Sum of Squared Error)를 최소화하는 과정을 통하여 계산된다.

2.2. 계절형 ARIMA(SARIMA) 모형

Box와 Jenkins (1994)의 계절형 자기회귀누적이동평균(autoregressive integrated moving average; ARIMA) 모형은 다른 설명변수의 도입이 없어도 교란항과 과거치만으로 시계열에 적합한 모형을 설정할 수 있도록 하는 가장 기본적인 방법으로, 단기예측과 계절형 변동을 포함하는 시계열의 분석에 효과적이다. Box의 이중 계절형 자기회귀누적이동평균의 기본 형태를 보면 (2.6)의 모형과 같다.

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Pi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}Y_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Psi_{Q_2}(B^{s_2})\epsilon_t. \tag{2.6}$$

여기서 B 는 $B^k Y_t = Y_{t-k}$ 에 연산을 맡는 후진연산자이고, $\phi_p(B)$ 와 $\theta_q(B)$ 는 p 차 AR과 q 차 MA를 의미한다. $\Phi_{P_1}(B^{s_1})$, $\Pi_{P_2}(B^{s_2})$, $\Theta_{Q_1}(B^{s_1})$, $\Psi_{Q_2}(B^{s_2})$ 는 s_1, s_2 에 따르는 P_1, P_2, Q_1, Q_2 차 계절성 AR과 계절성 MA를 의미한다. d, D_1, D_2 는 차분의 차수를 말하며, ϵ_t 는 백색잡음으로 $E(\epsilon_t) = 0$ 이고 $\text{Var}(\epsilon_t) = \sigma^2$ 이다. 다항식 $\phi_p(B) = 0$, $\Phi_{P_1}(B^{s_1}) = 0$, $\Pi_{P_2}(B^{s_2}) = 0$, $\theta_q(B) = 0$, $\Theta_{Q_1}(B^{s_1}) = 0$, $\Psi_{Q_2}(B^{s_2}) = 0$ 의 근은 단위원 밖에 위치한다.

2.3. Reg-ARIMA 모형

Reg-ARIMA 모형은 기존 ARIMA 모형에 회귀모형의 독립변수 효과를 고려한 모형으로 regression 과 ARIMA를 결합한 모형이다. k 개의 독립변수를 고려한 시계열 $\{Y_t | t = 1, 2, \dots, T\}$ 가 $\text{ARIMA}(p, d, q) \times (P, D, Q)_s$ 를 따를 때의 모형은 식 (2.7)와 같이 나타난다.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D \left(Y_t - \sum_{i=1}^k \beta_i x_{ti} \right) = \theta_q(B)\Theta_Q(B^s)\epsilon_t. \quad (2.7)$$

각 요소들의 의미는 Section 2.2 동일하며, β_i 들은 각각 독립변수 x_{ti} 의 계수를 의미하며, 본 논문에서는 독립변수로 명절 변수, 주말과 관련한 금, 토, 일요일 변수, 공휴일 변수를 고려하였다.

2.4. GARCH 모형

오차항의 분산이 시간에 따른 변동이 발생할 경우, 이분산성을 고려한 모형을 적용하여야 한다. 이분산성을 고려한 ARCH(autoregressive conditionally heteroscedasticity) 모형은 Engle (2001)에 의해 제안되었고, Bollerslev (1986)는 현재 분산이 과거 분산의 영향을 받는 형태로 확장시킨 일반화된 형태인 GARCH(generalized autoregressive conditionally heteroscedasticity) 모형을 제안하였다. 이에 GARCH(q^*, p^*) 모형은 식 (2.8)와 같이 나타난다.

$$\begin{aligned} Y_t &= \mu_t + \epsilon_t, \\ \epsilon_t &= \sqrt{h_t}e_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{p^*} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q^*} \beta_i h_{t-i}. \end{aligned} \quad (2.8)$$

이 때, $e_t \sim WN(0, \sigma^2)$ 이며, 각 모수들은 $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, $\alpha_i + \beta_i < 1$ 의 조건을 갖는다. 또한, p^* 은 ARCH 부분의 차수이고, q^* 은 GARCH 부분의 차수를 뜻한다. 만약 q^* 이 0일 경우, 즉, GARCH($0, p^*$)모형은 ARCH(p^*)과 같다.

2.5. REG-SAR-GARCH 모형

Seasonal AR-GARCH 모형은 기존 GARCH 모형에서 평균모형 μ_t 가 SAR(p)(P) $_s$ 모형일 때를 의미하며 SAR(p)(P) $_s$ -GARCH(q^*, p^*)은 식 (2.9)와 같이 나타난다.

$$\begin{aligned} \Phi(B)\phi(B)(Y_t - \mu) &= \epsilon_t, \\ \epsilon_t &= \sqrt{h_t}e_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{p^*} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q^*} \beta_i h_{t-i}. \end{aligned} \quad (2.9)$$

이 때, $\phi(B)$ 는 후진연산자 B에 대한 다항식으로 $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ 를 의미하고, $\Phi(B)$ 는 후진연산자 B에 대한 다항식으로 $\Phi(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$ 이다. 분산모형은 GARCH 모형에서와 같이 $e_t \sim WN(0, \sigma^2)$ 이며, 각 모수들은 $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, $\alpha_i + \beta_i < 1$ 의 조건을 갖는다.

k 개의 독립변수 효과를 고려한 Reg(k)-SAR(p)-GARCH(q^*, p^*)모형은 식 (2.10)과 같다.

$$\begin{aligned} \Phi(B)\phi(B) \left(Y_t - \sum_{i=1}^k \gamma_i x_{it} \right) &= \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} e_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{p^*} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q^*} \beta_i h_{t-i}. \end{aligned} \tag{2.10}$$

3. 자료분석

본 논문에서 이용한 데이터는 2010년 10월 1일부터 2013년 8월 31일까지 KTX의 지역별 이용 여객수 데이터이다. 이 중 2010년 10월 1일부터 2013년 3월 31일까지의 자료를 훈련 자료(training data)로 모형을 적합하는데 사용하였고, 2013년 4월 1일부터 2013년 6월 30일까지 3개월의 데이터를 검정 자료(test data)로 모형의 성능을 평가하는데 사용하였다.

KTX 이용객에 대한 원 데이터의 시계열 그래프를 보면 Figure 3.1과 같다. Figure 3.1로부터 부산과 울산, 동대구역의 일별 이용객 수에서 부산과 동대구역에 비하여 울산역에 경우는 2개 지역에 비하여 적은 이용객 수로 나타난다. 이에 대하여 2012년 4월 2일부터 4월 29일까지의 4주간의 월요일부터 일요일까지를 나타낸 Figure 3.2를 살펴보면 그래프는 월요일부터 시작하는 그래프로써 데이터가 1년을 주기적으로 유사한 패턴을 나타내는 것을 알 수 있다. 금요일을 제외한 평일은 금, 토, 일요일에 비하여 낮은 수치로 나타나며, 이와 유사한 명절 변수와 공휴일 변수를 개별적인 더미변수로 설정하여 연휴효과를 모형에 반영하도록 한다.

3.1. 모형 적합

먼저 명절의 효과를 고려하는 변수로서 KTX에서 지정한 명절 승차권 사용 기간으로 지정하며, 주말 변수는 주말요일로 산정된 금, 토, 일에 대하여 1로 지정을 하며 나머지는 0으로 한다. 공휴일에 대한 변수로는 국가 공휴일을 1로 지시하였으며 나머지는 0으로 지정하였다. 이 때, 지정 되는 변수들에 대하여 중복되는 일자에 대하여 명절, 주말, 공휴일 순으로 우선적으로 지정을 한다.

$$\begin{aligned} \text{공휴일} &= \begin{cases} 1, & \text{if 명절 기간,} \\ 0, & \text{if 명절 기간이 아닌 경우,} \end{cases} & \text{금요일} &= \begin{cases} 1, & \text{if 금요일,} \\ 0, & \text{if 금요일이 아닌 경우,} \end{cases} \\ \text{토요일} &= \begin{cases} 1, & \text{if 토요일,} \\ 0, & \text{if 토요일이 아닌 경우,} \end{cases} & \text{일요일} &= \begin{cases} 1, & \text{if 일요일,} \\ 0, & \text{if 일요일이 아닌 경우,} \end{cases} \\ \text{공휴일} &= \begin{cases} 1, & \text{if 공휴일,} \\ 0, & \text{if 공휴일이 아닌 경우.} \end{cases} \end{aligned}$$

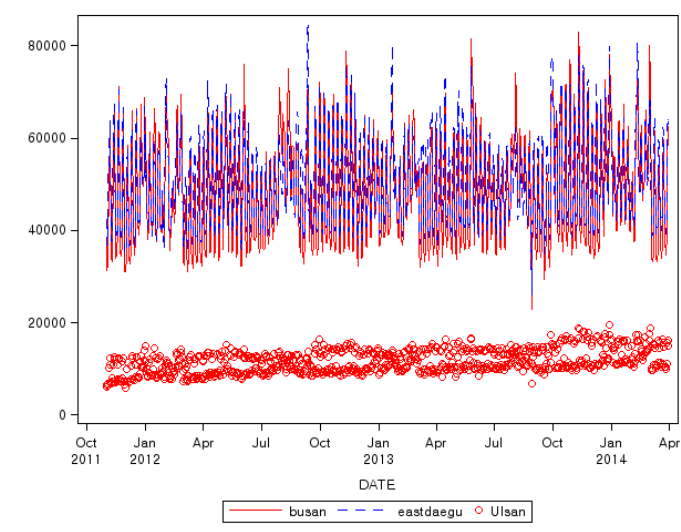


Figure 3.1. Time plot for KTX Using customer in Busan, East Daegu and Ulsan stations

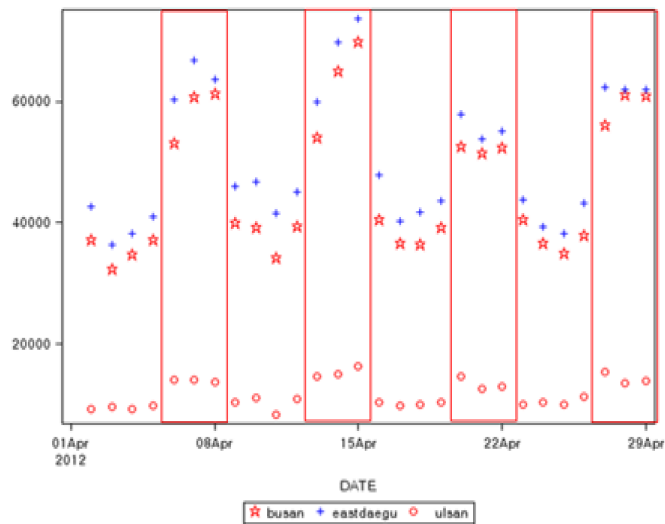


Figure 3.2. Time plot for KTX Using customer in 4 weeks

본 논문에서 고려하는 변수로는 Taylor가 제안한 Holt-Winters 지수평활법과 명절변수와 주말변수, 공휴일 변수를 고려한 계절형 Reg-ARIMA 모형, 그리고 계절형 Reg-AR-GARCH 모형을 통하여 모형을 사용한다.

Holt-Winters 지수평활법은 여객수만을 이용한 모형으로서 본 논문에서는 2010년 10월 1일부터 2013년 3월 31간의 자료를 이용하여 부산, 동대구, 울산역에서의 여객수를 이용한 결과를 산출하였으며 이에 대한 모수 추정치는 Table 3.1과 같이 나타났다.

ARIAM모형에서 독립변수로는 명절, 주말, 공휴일을 고려하였다. 모형 선정은 AIC(Akaike's in-

Table 3.1. Estimates of Taylor's Holt-Winters

<i>Parameter</i>	<i>Busan</i>	<i>Ulsan</i>	<i>EastDaegu</i>
<i>Level</i> (α)	0.03565	0.00112	0.01359
<i>Trend</i> (β)	0.00015	0.25630	0.02013
<i>Seasonal1</i> (γ)	0.02303	0.26593	0.00000
<i>Seasonal2</i> (δ)	0.19512	0.28397	0.50399
<i>Autoregressive</i> (ϕ)	0.60814	0.57629	0.35638

Table 3.2. Estimates of $REG - ARIMA(1, 0, 0)(1, 0, 1)_7(1, 0, 0)_{364}$ in Busan

모수	추정값	표준오차	p-value
μ	40111.0	1224.5	< .0001
Θ_7	0.54895	0.06073	< .0001
ϕ_1	0.71095	0.02413	< .0001
Φ_7	0.81348	0.04208	< .0001
Φ_{364}	0.29487	0.04551	< .0001
β_{Holiday}	8853.0	793.8	< .0001
β_{Friday}	14366.7	849.9	< .0001
β_{Saturday}	21331.2	958.5	< .0001
β_{Sunday}	20608.1	843.2	< .0001
β_{Festival}	20642.6	1212.7	< .0001

Table 3.3. Estimates of $REG - ARIMA(1, 0, 0)(1, 0, 1)_7(1, 0, 0)_{364}$ in East Daegu

모수	추정값	표준오차	p-value
μ	42587.7	854.8	< .0001
Θ_7	0.58952	0.05651	< .0001
ϕ_1	0.52812	0.02911	< .0001
Φ_7	0.83980	0.03801	< .0001
Φ_{364}	0.31977	0.04581	< .0001
β_{Holiday}	4120.5	853.3	< .0001
β_{Friday}	15982.6	940.3	< .0001
β_{Saturday}	19332.9	1034.8	< .0001
β_{Sunday}	17338.2	925.3	< .0001
β_{Festival}	23468.3	1109.2	< .0001

formation criterion)가 최소인 모형으로 선택하였다. AIC를 고려한 최적 모형은 3지역 모두 동일하게 $ARIMA(1, 0, 0) \times (1, 0, 1)_7 \times (1, 0, 0)_{364}$ 로 나타났다. 잔차의 독립성을 보여주는 포트만트 검정(Portmanteau test) 결과 0.01보다 커 잔차들 사이에 독립성을 만족하였고, 추정된 모든 모수의 p-value는 0.05보다 작아 유의하게 나타났다. 이 때 각 지역에 따라 부산역 이용 여객수의 모수 추정치는 Table 3.2, 동대구역의 모수추정치는 Table 3.3, 울산역의 모수 추정치는 Table 3.4와 같다.

AR-GARCH모형에서도 동일한 독립변수를 사용하여 AIC가 최소인 모형을 선택한 결과 부산역에서는 모든 독립 변수가 고려된 $AR(1)(4)_7$ -GARCH(0, 1)모형으로 나타났으며, 동대구역은 모든 독립 변수가 고려된 $AR(2)(4)_7$ -GARCH(0, 1)모형으로 나타났다. 울산역은 공휴일 변수가 독립변수에서 제외된 모형으로 $AR(3)(3)_7$ -GARCH(0, 1)모형으로 선택되었다. 이에 대하여 모수추정치는 Table 3.5 ~ Table 3.7과 같다.

Table 3.4. Estimates of $REG - ARIMA(1, 0, 0)(1, 0, 1)_7(1, 0, 0)_{364}$ in Ulsan

모수	추정값	표준오차	p-value
μ	7631.0	405.9	< .0001
Θ_7	0.77471	0.02479	< .0001
ϕ_1	0.50719	0.02956	< .0001
Φ_7	0.99147	0.00694	< .0001
Φ_{364}	0.37589	0.04804	< .0001
β_{Holiday}	819.6	203.7	< .0001
β_{Friday}	3169.1	350.8	< .0001
β_{Saturday}	3926.4	375.7	< .0001
β_{Sunday}	4339.8	334.8	< .0001
β_{Festival}	3514.7	285.2	< .0001

Table 3.5. Estimates of $REG - AR(1)(4)_7 - GARCH(0, 1)$ in Busan

모수	추정값	표준오차	p-value
μ	40187.0	1095.0	< .0001
ϕ_1	-0.72860	0.02464	< .0001
Φ_7	-0.32310	0.04324	< .0001
Φ_8	0.25510	0.04157	< .0001
Φ_{14}	-0.17530	0.03944	< .0001
Φ_{15}	0.15000	0.04234	0.0004
Φ_{28}	-0.12850	0.03753	0.0006
α_0	154072	0.17012	< .0001
α_1	0.20410	0.04380	< .0001
β_{Holiday}	8456.0	631.9388	< .0001
β_{Friday}	14439.0	1092.0	< .0001
β_{Saturday}	21167.0	1050.0	< .0001
β_{Sunday}	20303.0	922.3	< .0001
β_{Festival}	21824.0	1160.0	< .0001

Table 3.6. Estimates of $REG - AR(2)(4)_7 - GARCH(0, 1)$ in East Daegu

모수	추정값	표준오차	p-value
μ	42832.0	1097.3	< .0001
ϕ_1	-0.60932	0.04272	< .0001
ϕ_2	0.07243	0.03353	0.0304
Φ_7	-0.33408	0.03352	< .0001
Φ_8	0.20748	0.03868	< .0001
Φ_{14}	-0.17178	0.04257	< .0001
Φ_{15}	0.12653	0.03688	< .0001
Φ_{28}	-0.06612	0.03378	0.0499
α_0	154416.8	0.21708	< .0001
α_1	0.16052	0.04832	0.0009
β_{Holiday}	3913.2	652.6	< .0001
β_{Friday}	15576.1	1002.3	< .0001
β_{Saturday}	18704.3	1112.1	< .0001
β_{Sunday}	16561.8	966.6	< .0001
β_{Festival}	23710.3	854.1	< .0001

Table 3.7. Estimates of $REG - AR(3)(3)_7 - GARCH(0, 1)$ in Uusan

모수	추정값	표준오차	p-value
μ	9158.2	394.9	< .0001
ϕ_1	-0.58320	0.03558	< .0001
Φ_3	-0.08211	0.02767	0.0030
Φ_7	-0.31729	0.04071	< .0001
Φ_{14}	-0.19068	0.04482	< .0001
Φ_{15}	0.10631	0.03860	0.0059
Φ_{21}	-0.10272	0.03782	0.0066
α_0	892627	0.28219	< .0001
α_1	0.25349	0.04709	< .0001
β_{Friday}	3515.1	239.9	< .0001
β_{Saturday}	3721.0	260.7	< .0001
β_{Sunday}	4187.2	250.2	< .0001
β_{Festival}	3648.1	310.1	< .0001

Table 3.8. MAPE of the models

MAPE	부산			동대구			울산		
	HW	ARIMA	AR-GARCH	HW	ARIMA	AR-GARCH	HW	ARIMA	AR-GARCH
1MONTH	6.399	5.319	4.592	9.167	5.453	4.839	8.249	5.258	4.832
2MONTH	7.612	5.359	5.631	10.734	5.473	5.813	8.588	5.292	6.180
3MONTH	7.642	5.451	6.256	13.332	5.609	6.212	8.725	5.452	6.396

3.2. 모형의 성능비교

본 논문에서는 모형의 예측력 비교를 위해 MAPE(Mean Absolute Percentage Error)를 사용했으며, 이는 (3.1)과 같이 정의된다.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - F_t}{Y_t} \right| \times 100(\%). \tag{3.1}$$

Table 3.8은 각 모형별로 Multistep-Ahead Forecasting 방법으로 2013년 4월부터 6월에 대하여 예측을 실시하였다. 2013년 4월 1일부터 2013년 6월 30일에 해당하는 MAPE 값으로 예측후 1달 간의 예측은 REG-AR-GARCH 모형의 MAPE값이 부산, 동대구, 울산역에서 3가지 모형 중 가장 작게 나타난다. 그러나 1달 이후의 예측에서는 REG-ARIMA 모형이 가장 작은 MAPE값이 가장 작아 장기예측에 경우는 AR-GARCH보다 REG-ARIMA모형이 적합한 것으로 나타난다. Taylor가 제안한 Holt-winters 모형은 단변량 모형으로서 주말변수변수와 휴일 변수를 고려하지 않는데, MAPE 값이 모든 모형들 중 가장 크게 나타나 적합하지 않은 모형으로 판단된다. 이에 따라 KTX 열차 이용객 수 예측에서는 모형에서 주말 변수와 휴일 변수를 고려한 모형이 예측에서 더 적합한 것으로 판단된다. Figure 3.3에서 Figure 3.5는 실제값과 각 모형별 예측값을 나타낸 그래프이다.

4. 결론

본 논문에서는 시계열 모형을 이용하여 부산, 동대구, 울산역에서의 KTX 이용 여객수에 대해 3개월간의 예측을 실시하였다. 시계열 자료의 분석과 예측에 있어서 단변량 모형인 Holt-Winters 모형과 주말

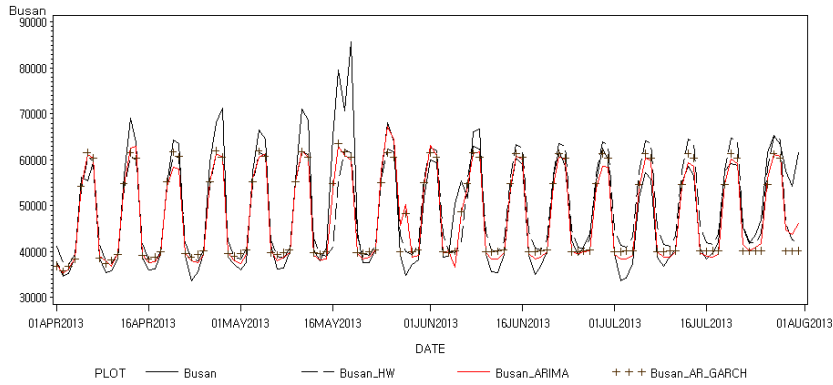


Figure 3.3. Time plot for each model in Busan station

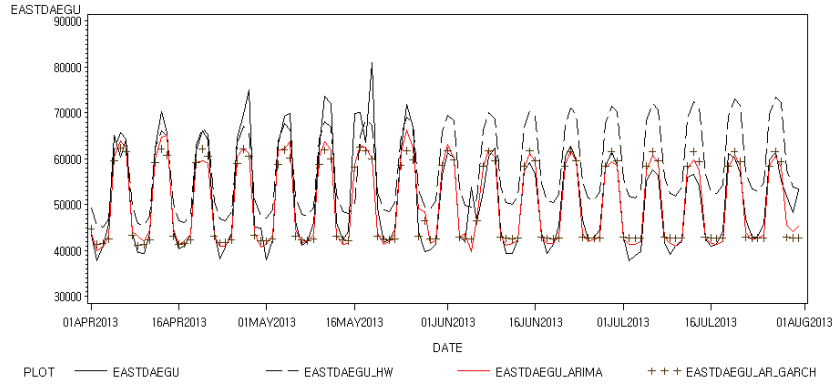


Figure 3.4. Time plot for each model in East Daegu station

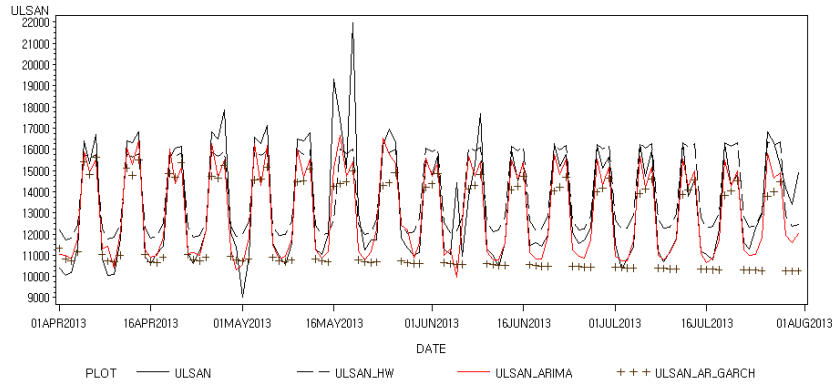


Figure 3.5. Time plot for each model in Ulsan station

변수와 공휴일 변수를 고려한 Reg-SARIMA 모형과 Reg-SAR-GARCH 모형을 사용하여 KTX 이용 여객수에 대한 예측을 시행하였다. 이 때, 1달간의 예측에서는 Reg-AR-GARCH 모형이 3 가지 모형 중 예측의 결과로서 MAPE값이 가장 작은 수치로 나타나 적합한 모형으로 판단되었다. 이는 KTX 이용 여객들이 주말이나 휴일에 따라 이용객수의 추세를 고려함과 동시에 여객들의 열차이용의 이분산성

을 고려하여 단기적인 예측에서 설명력을 높인 것으로 판단된다. 그러나 2개월 이후의 예측에 있어서는 Reg-SARIMA 모형이 가장 작은 MAPE값으로 적합하게 나타나는데, 이는 GARCH 모형에 경우 분산성을 고려하여 단기적인 예측에 적합하지만, 중·장기적인 예측에서는 Reg-ARIMA 모형이 더 적합하다고 할 수 있다.

부산역, 동대구역, 울산역 모두 주말변수와 공휴일 변수, 명절변수를 고려한 Reg-SARIMA, Reg-SAR-GARCH 모형이 이용 여객수만을 변수로 고려한 Taylor가 제안한 수정된 Holt-Winters 방법보다 낮은 MAPE를 보여준다. 이에 따라 국내 KTX 여객수 예측에서는 휴일에 대한 고려가 중요한 것으로 파악된다.

본 논문에서는 휴일에 대하여 주말, 공휴일, 명절 변수를 이용하여 KTX 여객수에 대한 예측을 실시하였는데, 휴일 변수에 대한 추가적인 연구로 휴일 전날과 다음날의 가중된 영향을 고려한 변수의 개발이 필요할 것이며, ARIMA 모형과 GARCH모형을 결합한 REG-ARIMA-GARCH 모형을 적용한다면 더 좋은 예측 모형을 제안할 수 있을 것으로 사료된다.

References

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). Time series analysis: forecasting and control, *Princeton-Hall International*.
- Choi, T. S. and Kim, S. H. (2004). An Empirical Comparison among Initialization Methods of Holt-Winters Model for Railway Passenger Demand Forecast, *The Korean Society for Railway*, **7**, 94-98.
- Engle, R. F. (2001). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987-1007.
- Holt, C. C. (1957). Forecasting trends and seasonals by exponentially weighted moving averages, *Office of Naval Research, Research Memorandum, Carnegie Institute of Technology*, **52**.
- Kim, H. S. and Yun, D. H. (2011a). Travel Behavior Analysis for Short-Term Railroad Passenger Demand Forecasting if KTX, *In Proceedings of the Conference of the Korean Society for Railway*, 1282-1289.
- Kim, K. H., Kim, H. S., Lee, S. D., Lee, H. G. and Yoon, K. M. (2011b). KTX Passenger Demand Forecast with Intervention ARIMA Model, *In Proceedings of the Conference of the Korean Society for Railway*, **14**, 1282-1289.
- Lee, D. K., Hong, T. H., Kim, H. B. and Woo, K. B. (1998). Study on the Demand Prediction for Transportation System Utilizing Data Granulization, *In Proceedings of the Conference of the Korean Society for Railway*, **5**, 211-218.
- Lippi, M., Bertini, M. and Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning, *Intelligent Transportation Systems, IEEE Transactions on*, **14**, 871-882.
- Oh, S. M. and Kim, D. H. (2001). A Study on the Seasonal Decomposition of the Railway Passenger Demand, *In Proceedings of the Conference of the Korean Society for Railway*, **10**, 111-116.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799-805.
- Tsai, T., Lee, C. and Wei, C. (2005). Design of dynamic neural networks to forecast short-term railway passenger demand, *Journal of the Eastern Asia Society for Transportation Studies*, **6**, 1651-1666.
- Wei, Z. M. and Song, H. C. (2013). Prediction Scheme of Railway Passenger Flow Based on Multiplicative Holt-Winters Model, *Applied Mechanics and Materials*, **416**, 1949-1953.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages, *Management Science*, **6**, 324-342.

시계열 모형을 이용한 KTX 여객 수요예측 연구

김인주^a · 손흥구^b · 김삼용^{b,1}

^a대덕대학교 호텔서비스학과, ^b중앙대학교 응용통계학과

(2014년 11월 18일 접수, 2014년 11월 22일 수정, 2014년 11월 22일 채택)

요약

KTX에 등장에 따라 국내 여객시장은 KTX 시장을 중심으로 변화가 이루어졌다. 이에 따라 KTX 이용 여객의 수요예측은 열차 운영에 있어서 매우 중대한 사안이다. 본 논문에서는 여러 시계열 모형의 비교를 통해 KTX 이용 여객의 수요와 연관이 있는 요일과 공휴일, 명절을 어떠한 형태로 고려할 것인지 연구하였다.

모형 간 예측력을 비교하기 위하여 Mean Absolute Percentage Errors (MAPE)를 사용하였으며, 1달간의 단기간 예측에 있어서 변동성을 고려해줄 수 있는 Reg-AR-GARCH 모형이 우수한 예측력을 나타냈으며, 1달을 초과한 기간의 예측에서는 Reg-ARMA 모형이 우수한 예측력을 나타냈다.

주요용어: Taylor's Holt-Winters 모형, Reg-ARMA 모형, Reg-AR-GARCH 모형, MAPE, KTX 여객수.

이 논문은 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2014-027173).

¹교신저자: (156-756) 서울 동작구 흑석동 221, 중앙대학교 응용통계학과, 교수. E-mail: sahm@cau.ac.kr