

KTX passenger demand forecast with multiple intervention seasonal ARIMA models

Hyoyoung Cha^a · Yoonsik Oh^b · Jiwoo Song^a · Taewook Lee^{a,1}

^aDepartment of Statistics, Hankuk University of Foreign Studies;

^bResearch Center, Korea Railroad

(Received November 15, 2018; Revised December 8, 2018; Accepted December 8, 2018)

Abstract

This study proposed a multiple intervention time series model to predict KTX passenger demand. In order to revise the research of Kim and Kim (*Korean Society for Railway*, **14**, 470–476, 2011) considering only the intervention of the second phase of Gyeong-bu before November of 2011, we adopted multiple intervention seasonal ARIMA models to model the time series data with additional interventions which occurred after November of 2011. Through the data analysis, it was confirmed that the effects of various interventions such as Gyeong-bu and Ho-nam 2 phase, outbreak of MERS and national holidays, which affected the KTX transportation demand, are successfully explained and the prediction accuracy could be quite improved significantly.

Keywords: ARIMA, forecasting, intervention, KTX, seasonal ARIMA

1. 서론

일반적으로 시계열 자료 분석의 목적은 자료에 내재되어 있는 다양한 시계열적 성질 및 특성을 효과적으로 설명하는 동시에 미래를 가능한 정확히 예측하는 데 있다. 가장 일반적인 시계열 모형으로 자료의 자기상관성을 분석하는 ARIMA 모형과 계절적인 특성을 함께 분석하는 계절형 ARIMA 모형이 있다. 그러나 시계열 자료에 영향을 미치는 외생변수 또는 개입변수 등이 존재하는 경우에 ARIMA 계열의 모형들은 적절한 모형이 아니므로, 이를 반영하기 위해 전이함수모형 또는 개입 ARIMA 모형 등을 고려한다. 시계열 자료에서 개입은 시계열 자료에 영향을 미칠 수 있는 사건을 의미하며, 이들을 입력변수 또는 설명변수로 반영하여 분석하는 모형을 개입 시계열 모형이라고 한다. 예를 들어 KTX 수송실적에 영향을 미치는 개입으로 경부고속철도 2단계 개통, 호남고속도로 2단계 개통과 같은 정책의 변화, 그리고 SARS 및 MERS와 같은 국가감염병 등의 사건을 들 수 있다. 이러한 개입이 발생하는 경우에 실제로 시계열 자료의 패턴에 영향을 미치는 지를 밝혀내고, 영향을 미친다면 어떠한 형태인지를 모형화하여 그 효과를 추정하고 이를 이용하여 좀 더 정확한 미래 예측을 하는 것이 개입 시계열 모형 분석의 목적이라고 볼 수 있다. 개입시계열 모형으로 자료를 분석한 선행연구로서 KTX 수송수요를 예측한 Kim과

This work was supported by Hankuk University of Foreign Studies Research Fund of 2018.

¹Corresponding author: Department of Statistics, Hankuk University of Foreign Studies, 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si, Gyeonggi-do 17035, Korea. E-mail: twlee@hufs.ac.kr

Kim (2011), Kim과 Jo (2014), 한국의 공무원 인력규모를 분석한 Hong과 Lee (2016)와 관광객 수 예측을 분석한 Huh와 Kim (2001), Kim과 Seong (2011) 및 Song (2016) 등이 있다.

미래의 수송계획 및 열차운영계획을 효율적으로 수립하기 위해서는 정확한 단기수송 수요예측이 매우 필요하다. KTX 수송수요 예측을 위해 시계열 분석을 이용한 선행 연구로서 Choi와 Kim (2004) 및 Kim과 Kim (2011) 등이 있다. Choi와 Kim (2004)은 Holt-Winters 모형을 적용하였는데 KTX 수송 수요 시계열 자료에 영향을 미치는 다양한 개입 효과를 해석하고 이를 미래 수송수요 예측에 반영할 수 없다는 한계가 있다. Kim과 Kim (2011) 등은 이를 극복하기 위하여 개입 계절형 ARIMA 모형을 도입하여 경부 2단계 개통 개입 효과를 적절히 설명하였다. 본 연구는 2011년 이전의 자료로서 경부 2단계 개통 개입만 고려한 Kim과 Kim (2011)의 연구를 수정 보완하기 위해 다양한 개입이 발생하고 있는 2011년 이후의 시계열 자료를 효과적으로 모델링하는 한편 KTX 수송수요를 정확히 예측하기 위한 방법으로 다중개입 계절형 ARIMA 모형을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 연구 자료의 소개와 일반적인 다중개입 계절형 ARIMA 모형을 살펴 보고, 본 연구의 KTX 수송수요 자료에 발생하는 다양한 개입을 설명한 후 이를 반영하기 위한 구체적인 개입 형태를 제안한다. 3절에서는 KTX 수송실적 자료에 적합한 다중개입 계절형 ARIMA 모형 식별과 추정을 통해 다양한 개입 효과를 분석하고, 모형 검증을 통해 예측력을 살펴 본다. 끝으로 4절에서는 연구결과를 정리하고, 결론을 도출한다.

2. 연구자료 및 다중개입 계절형 ARIMA 모형

2.1. 연구자료

본 연구에서 이용한 자료는 2006년 1월 이후부터 2016년 11월까지의 KTX 전체 월별 일평균 통행량(통행/일)이다. Choi와 Kim (2004) 및 Kim과 Kim (2011)의 선행 연구 등을 참고하여 연구 자료를 검토한 결과 다음과 같은 분석방향을 설정하였다.

첫째, 경부축과 호남축을 구분하여 예측한다. Kim과 Kim (2011)의 선행 연구에서는 경부축과 호남축의 수송실적을 합하여 총량 자료를 구축한 후 이를 분석하였으나, 본 연구에서는 경부축과 호남축으로 구분하여 각각 분석하고자 한다. 이는 경부축과 호남축에 영향을 미친 개입과 시계열적 특성이 서로 다르기 때문에 이러한 개별적인 성질을 반영하여 분석하는 것이 경부축과 호남축 자료의 해석력을 높이는 동시에 궁극적으로 예측의 정확도를 향상시킬 수 있다고 보여진다. 둘째, KTX 수송수요를 주중과 주말로 구분하여 예측한다. 이는 Kim과 Kim (2011) 등의 선행 연구에서 KTX 수송실적 자료를 주중(월-목)과 주말(금-일)로 구분하여 분석하여 예측 정확도를 향상시킬 수 있었음을 참고하였다.

셋째, 경부 및 호남고속철도 2단계 개통, KTX 운행을 대폭 확대한 운송력 증대 및 메르스 발병 등의 개입을 반영한 다중개입 계절형 ARIMA 모형을 도입한다. Figure 2.1의 경부축 및 호남축 KTX 수송실적 시계열도 모두 전체적으로 완만하게 상승하는 추세와 더불어 몇 차례 추세가 급격히 변화하고 있으며, 계절성을 포함한 시계열적인 특성이 혼합된 형태를 나타내고 있다. 경부축은 대략 2010년 11월 경부고속철도 2단계 개통을 기준으로 수송실적이 급격히 증가하는 한편 추세에도 변화가 있으며, 2012년 11월 KTX 운행을 대폭 확대한 운송력 증대 이후로 추세가 다시 변하고 있어 자료에 비정상성이 존재하는 것으로 보인다. 호남축은 2012년 11월 운송력 증대로 수송실적이 일정부분 증가하였으며, 2015년 4월 호남 2단계 개통 이후 추세가 급격히 변하고 있어 역시 자료에 비정상성이 존재한다. 경부축과 호남축 시계열 자료에 국소적 불연속 선형 추세 등의 비정상성이 있다고 판단하여 수송수요 시계열 자료를 정상시계열로 변환하기 위해 1차 차분을 실시하여 추세가 제거된 정상시계열로 변환한 후 계절형 시계열 모형을 도입하여 분석할 수도 있으나, 월별 수송실적 시계열도의 국소적 불연속 선형 추세의 원인이

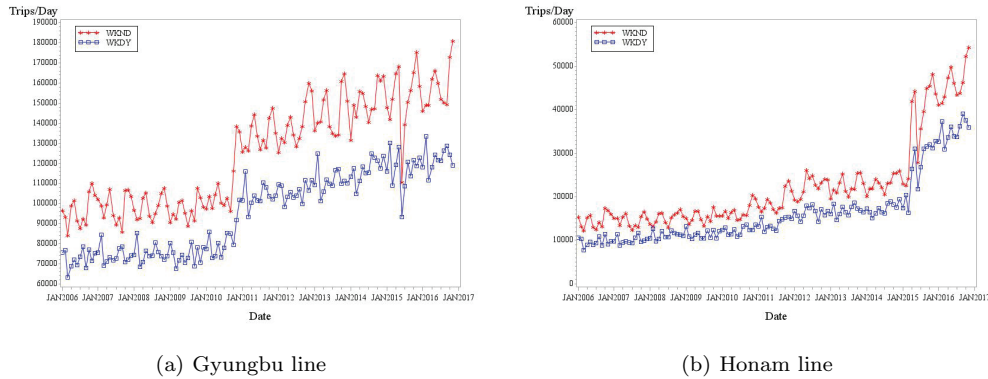


Figure 2.1. Gyungbu line (a) and Honam line (b) monthly average daily trips.

2010년 11월 경부 2단계 개통, 2011년 11월 운송력 증대와 2015년 호남 2단계 개통 등의 개입임이 자명하므로 1차 차분을 통해 국소적 불연속 선형 추세를 제거하기 보다는 다양한 개입 효과를 적절히 설명할 수 있는 다중개입 계절형 시계열 모형을 도입하는 것이 자료의 해석과 예측 정확도를 향상시킬 수 있다고 보여진다.

끝으로 모형식별 및 추정, 검증 및 예측에 적합한 분석기간을 설정한다. 모형식별 및 추정 기간은 2006년 1월부터 2015년 12월까지의 10년 동안의 자료로서 총 120개월이며, 모형 검증은 2016년 1월부터 2016년 11월까지의 11개월이다. 참고로 2016년 12월은 SRT의 개통으로 인해 KTX 수송실적이 일시적으로 크게 감소하여 검증기간에 포함하지 않았음을 밝힌다.

2.2. 다중개입 계절형 ARIMA 모형

다중개입 시계열 분석은 시계열 자료가 장기간에 걸쳐 관측되는 경우에 제어할 수 없는 여러 가지 외부적인 사건들이 발생하면 이에 대한 영향을 모형화한 후 이의 효과를 추정하고 이를 통해 미래를 예측하는 분석법이다. 다중개입 시계열 모형은 개입이 포함되지 않은 시계열에 가해지는 여러 가지 외부적인 충격의 영향이 가법적인 형태로 미친다고 가정하여 다음의 형태로 모형화한다.

$$Y_t = \sum_{i=1}^k \frac{w_i(B)}{\delta_i(B)} B^{b_i} I_{t,i} + Z_t, \tag{2.1}$$

여기서 B 는 후행연산자(backshift operator), $I_{t,i}$ 는 개입변수, b_i 는 개입효과의 지연시차, $w_i(B)$ 와 $\delta_i(B)$ 는 각각 개입의 초기 기대효과와 개입의 지속효과를 나타내며 다음과 같이 정의한다.

$$w_i(B) = w_{i0} - w_{i1}B - \dots - w_{is_i}B^{s_i},$$

$$\delta_i(B) = 1 - \delta_{i1}B - \dots - \delta_{ir_i}B^{r_i}.$$

위 식에서 개입변수 $I_{t,i}$ 는 펄스함수 $I_{t,i} = P_{t,i}^{(T_i)}$ 또는 계단함수 $I_{t,i} = S_{t,i}^{(T_i)}$ 의 형태로 나뉘며, 구체적인 표현은 아래와 같다.

$$P_{t,i}^{(T_i)} = \begin{cases} 0, & t = T_i, \\ 1, & t \neq T_i, \end{cases} \quad S_{t,i}^{(T_i)} = \begin{cases} 0, & t < T_i, \\ 1, & t \geq T_i, \end{cases}$$

여기서 T_i 는 i 번째 개입의 발생 시점을 의미하며, 펄스함수는 개입의 효과가 발생시점 T_i 에만 영향을 미치는 경우에 사용하며, 계단함수는 개입의 효과가 T_i 시점에 발생하여 그 효과가 T_i 시점 이후에도 지속적으로 영향을 미치는 경우에 사용한다. 다양한 형태의 개입모형은 Box와 Tiao (1975), Cryer와 Chan (2008) 및 Cho 등 (2016)에 자세히 소개되어 있다.

한편, 식 (2.1)의 오차 Z_t 는 계절형 ARIMA 모형을 가정하여 아래 형태의 승법계절모형으로 표현하며, 간단히 $ARIMA(p, d, q)(P, D, Q)_s$ 로 나타낸다.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \delta + \theta_q(B)\Theta_Q(B^s)\epsilon_t.$$

여기서, s 는 계절주기, $\phi_p(B)$ 와 $\Phi_P(B)$ 는 각각 비계절형 및 계절형 AR연산자이며, $\theta_q(B)$ 와 $\Theta_Q(B^s)$ 는 각각 비계절형 및 계절형 MA연산자, d 와 D 는 각각 비계절형 및 계절형 차분 차수, p 와 P 는 각각 비계절형 및 계절형 AR차수, q 와 Q 는 각각 비계절형 및 계절형 MA차수, δ 는 평균과 관련된 상수, ϵ_t 는 백색잡음이다.

2.3. 경부축과 호남축 다중개입 정의 및 모형 설정

경부축과 호남축에 영향을 미치는 개입을 요약하여 정리하면 다음과 같다. 먼저 경부축과 호남축에 모두 영향을 미치는 공통 개입으로서 2010년 11월 경부 2단계 개통, 2012년 11월 운송력 증대, 2015년 6월과 7월에 발병한 메르스, 그리고 2006년 이후의 설 또는 추석 명절이 있다. 참고로 경부 2단계 개통은 경부축 뿐만 아니라 호남축의 철도 공급에도 영향을 미쳐 수송실적을 변화 시키기 때문에 공통 개입으로 정의한다. 설 또는 추석 명절 개입과 2015년 6월과 7월에 발생한 메르스 개입은 개입의 효과가 발생 시점에만 영향을 미치는 경우이므로 분석을 위해 펄스함수를 사용하고, 개입의 효과가 바로 반영되어 지연 효과가 없으며 그 크기가 모두 일정한 경우를 가정한다. 참고로 설추석 명절 개입은 주중에만 효과를 미친다고 보아 주중 자료에만 반영하였으며, 명절에 평일이 1-2일이 포함되어 있는 경우와 3일 이상이 포함되어 있는 경우로 구분하여 개입 변수를 설정하였다. Figure 2.1의 시계열도를 보면 메르스 개입의 효과가 2015년 6월과 7월에 서로 다르다고 볼 수 있기 때문에 이를 구분하여 개입 변수를 설정하였다. 경부 2단계 개통, 운송력 증대 및 호남 2단계 개통 개입은 발생한 시점 이후에 지속적으로 영향을 미친다고 판단하여 계단함수를 사용하고, 마찬가지로 개입의 효과가 바로 반영되어 지연 효과가 없으며 그 크기가 일정한 경우와 선형으로 증가하는 경우를 동시에 가정하여 개입 모형을 설정하였다. 단, 호남축 수송실적 자료에 대한 2011년 11월부터 2015년 3월 기간동안의 운송력 증대 개입 효과는 Figure 2.1의 시계열도를 고려해 볼 때 선형으로 증가하지 않는 것으로 판단하여 선형 증가 효과는 포함하지 않았다.

경부축 주중 수송수요 자료의 다중개입 계절형 ARIMA 모형은 위 정의된 다양한 개입 효과를 반영하여 다음과 같이 표현한다.

$$Y_t = w_{1,10}S_t^{(10)} + \frac{w_{2,10}}{1-B}S_t^{(10)} + w_{1,12}S_t^{(12)} + \frac{w_{2,12}}{1-B}S_t^{(12)} + w_{M_6}P_t^{M_6} + w_{M_7}P_t^{M_7} + w_{H_{12}}P_t^{H_{12}} + w_{H_3}P_t^{H_3} + Z_t \quad (2.2)$$

여기서, 각각의 변수와 계수에 대한 설명은 다음과 같다. $S_t^{(10)}$ 는 2010년 11월 경부 2단계 개통 이후부터 2012년 11월 운송력 증대 이전까지의 계단함수, $S_t^{(12)}$ 는 2012년 11월 이후 운송력 증대 계단함수, $P_t^{M_6}$ 는 2015년 6월 메르스 펄스함수, $P_t^{M_7}$ 는 2015년 7월 메르스 펄스함수, $P_t^{H_{12}}$ 는 2006년 이후 평일이 1-2일 포함된 명절 펄스함수, $P_t^{H_3}$ 는 2006년 이후 평일이 3일 이상 포함된 명절 펄스함수이다. $w_{1,10}$ 는 2010년 11월 경부 2단계 개통 이후부터 2012년 11월 운송력 증대 이전까지의 일정 증가 효

과, $w_{2,10}$ 는 동 기간 동안의 선형 증가 효과, $w_{1,12}$ 는 2012년 11월 이후 운송력 증대 일정 증가 효과, $w_{2,12}$ 는 동 기간 동안의 선형 증가 효과, w_{M_6} 는 2015년 6월 메르스 일정 증가 효과, w_{M_7} 는 2015년 7월 메르스 일정 증가 효과, $w_{H_{12}}$ 는 2006년 이후 평일이 1-2일 포함된 명절 일정 증가 효과, w_{H_3} 는 2006년 이후 평일이 3일 이상 포함된 명절 일정 증가 효과이다. 경부축 주말 수송실적 자료의 다중개입 계절형 ARIMA 모형은 경부축 주중 모형 (2.2)에서 명절 변수를 고려하지 않았으며, 다음과 같이 표현한다.

$$Y_t = w_{1,10}S_t^{(10)} + \frac{w_{2,10}}{1-B}S_t^{(10)} + w_{1,12}S_t^{(12)} + \frac{w_{2,12}}{1-B}S_t^{(12)} + w_{M_6}P_t^{M_6} + w_{M_7}P_t^{M_7} + Z_t. \quad (2.3)$$

호남축 주중 수송실적 자료의 모형은 경부축 주중 모형인 식 (2.2)에 호남 2단계 개통 변수를 추가하였으며, 다음과 같이 표현한다.

$$Y_t = w_{1,10}S_t^{(10)} + \frac{w_{2,10}}{1-B}S_t^{(10)} + w_{1,12}S_t^{(12)} + w_{1,15}S_t^{(15)} + \frac{w_{2,15}}{1-B}S_t^{(15)} + w_{M_6}P_t^{M_6} + w_{M_7}P_t^{M_7} + w_{H_{12}}P_t^{H_{12}} + w_{H_3}P_t^{H_3} + Z_t \quad (2.4)$$

여기서, $S_t^{(15)}$ 는 2015년 4월 이후 호남 2단계 개통 계단함수이며, $w_{1,15}$ 와 $w_{2,15}$ 는 각각 2015년 4월 이후 호남 2단계 개통 일정 증가와 선형 증가 효과이다. $S_t^{(12)}$ 는 경부축에서와는 달리 2012년 11월 이후 운송력 증대 이후부터 2015년 4월 이후 호남 2단계 개통 이전까지의 계단함수이며, $w_{1,12}$ 와 $w_{2,12}$ 는 2012년 11월 운송력 증대 이후부터 2015년 4월 이후 호남 2단계 개통 이전까지의 일정 증가 효과와 선형 증가 효과를 의미한다. 호남축 주말 수송실적 자료의 모형은 식 (2.4)에서 명절 변수를 고려하지 않았으며, 다음과 같이 표현한다.

$$Y_t = w_{1,10}S_t^{(10)} + \frac{w_{2,10}}{1-B}S_t^{(10)} + w_{1,12}S_t^{(12)} + w_{1,15}S_t^{(15)} + \frac{w_{2,15}}{1-B}S_t^{(15)} + w_{M_6}P_t^{M_6} + w_{M_7}P_t^{M_7} + Z_t. \quad (2.5)$$

경부축과 호남축의 주중과 주말 수송실적 모형식 (2.2)-(2.5)에서 오차 Z_t 에 대한 모형은 다음과 같다.

$$Z_t = \frac{\theta(B)\Theta(B^s)}{(1-B^s)^D\phi(B)\Phi(B)}\epsilon_t.$$

여기서 경부축 및 호남축 수송실적 자료의 국소적 불연속 선형 추세의 비정상성을 다양한 개입 효과로 설명하기 때문에 비계절형 차분은 고려하지 않으나, 계절 차분은 개입 효과로 설명하기 어렵기 때문에 오차 시계열 모형에 포함하였다.

3. 다중개입 계절형 ARIMA 모형 추정 및 검증

3.1. 모형식별 및 추정

다중개입 시계열 모형의 모형 식별은 시계열 Z_t 가 $ARIMA(p, d, q)(P, D, Q)_s$ 모형을 따른다는 가정 하에, 차분 계수인 d 와 D , ARMA 모형 차수인 p 와 P , 계절 ARMA 모형 차수인 q 와 Q 를 결정하는 단계이다. 이를 위해 일반적으로 개입이 있기 전의 시계열 자료를 이용하여 모형화 하지만 본 논문에서는 전통적인 모형 식별 방법 대신 다음과 같이 모형 식별을 진행하였다.

첫째, 비계절형 차분연산자의 차수는 $d = 0$ 을 고려하였다. 시계열 모형의 1차 차분은 확률적 추세 변화가 있는 시계열 자료의 경우에 사용하는 것이 일반적인 데, 본 논문에서 분석하고 있는 KTX 자료는 확률적 추세 변화보다는 개입으로 인한 결정적 추세 변화가 있는 것으로 판단된다. 따라서 개입으로 인한 결정적 추세 변화를 1차 차분으로 분석하기 보다는 개입 분석을 활용하는 것이 바람직하다고 판단하였

Table 3.1. Fitted models for Gyeong-bu and Ho-nam KTX Lines based on AIC

| | Gyeong-bu | | | | Ho-nam | | | |
|------------------|----------------------------------|-----------------|----------------------------------|-----------------|----------------------------------|-----------------|----------------------------------|-----------------|
| | Weekday | | Weekend | | Weekday | | Weekend | |
| | Estimates | <i>p</i> -value | Estimates | <i>p</i> -value | Estimates | <i>p</i> -value | Estimates | <i>p</i> -value |
| $w_{1,10}$ | 20000 | <0.0001 | 30747 | <0.0001 | 1177 | 0.1433 | 1632 | 0.0565 |
| $w_{2,10}$ | 373 | 0.1056 | 341 | 0.0628 | 183 | 0.0018 | 266 | <0.0001 |
| $w_{1,12}$ | 33189 | <0.0001 | 43622 | <0.0001 | 6484 | <0.0001 | 7725 | <0.0001 |
| $w_{2,12}$ | 295 | 0.0033 | 369 | <0.0001 | . | . | . | . |
| $w_{1,15}$ | . | . | . | . | 16018 | <0.0001 | 24145 | <0.0001 |
| $w_{2,15}$ | . | . | . | . | 514 | <0.0001 | 479 | <0.0001 |
| w_{M_6} | -25140 | <0.0001 | -43175 | <0.0001 | -7098 | <0.0001 | -11828 | <0.0001 |
| w_{M_7} | -9218 | 0.0009 | -8566 | 0.0514 | -2374 | 0.0034 | -3897 | 0.0014 |
| $w_{H_{12}}$ | 5772 | <0.0001 | . | . | 1498 | <0.0001 | . | . |
| w_{H_3} | 8953 | <0.0001 | . | . | 2673 | <0.0001 | . | . |
| AR ₁ | 0.9388 | <0.0001 | 0.3683 | 0.0001 | 0.4973 | <0.0001 | . | . |
| AR ₂ | . | . | . | . | 0.2763 | 0.0080 | . | . |
| MA ₁ | 0.6195 | <0.0001 | . | . | . | . | -0.4740 | <0.0001 |
| MA ₂ | . | . | . | . | . | . | -0.3732 | 0.0001 |
| SAR ₁ | . | . | -0.2687 | 0.0856 | . | . | . | . |
| SMA ₁ | 0.7740 | <0.0001 | 0.5332 | 0.0003 | 0.7493 | <0.0001 | 0.6832 | <0.0001 |
| SARIMA | (1, 0, 1)(0, 1, 1) ₁₂ | | (1, 0, 0)(1, 1, 1) ₁₂ | | (2, 0, 0)(0, 1, 1) ₁₂ | | (0, 0, 2)(0, 1, 1) ₁₂ | |
| AIC | 2040.3 | | 2132.3 | | 1792.3 | | 1863.4 | |

AIC = Akaike information criteria.

다. 둘째, 계절형 차분연산자의 차수는 $D = 1$ 을 고려하였다. 이는 본 연구의 선행 연구라고 할 수 있는 Kim과 Kim (2011)에서도 이미 동일한 자료에 대해 1차 계절차분을 시행하였을 뿐만 아니라, 다중개입으로 분석할 수 있는 비계절형 결정적 추세 변화와는 달리 계절형 결정적 추세 변화는 그 존재에 대해 설명하기 어려울뿐만 아니라 분석도 매우 복잡하기 때문이다. 셋째, 자기상관도와 부분자기상관도를 이용하여 비계절형 ARMA 모형 차수와 계절형 ARMA 모형 차수를 결정하는 대신, 각각의 비계절형 및 계절형 ARMA 차수의 상한을 상정한 후 모든 가능한 조합의 모형 중에서 Akaike information criteria (AIC)값을 최소화하는 차수를 결정하였다. 이는 자기상관도와 부분자기상관도를 이용하는 방법은 객관적인 방법이라고 볼 수 없음을 고려하였다. 본 연구에서는 $0 \leq p, q \leq 2$ 와 $0 \leq P, Q \leq 1$ 를 가정하였다. 마지막으로 상수항은 Kim과 Kim (2011) 등에서의와 같이 통계적으로 유의하지 않으므로 모형에서 제외하기로 결정하였다.

경부축과 호남축의 주중과 주말 수송실적 다중개입 ARIMA 시계열 모형 (2.2)–(2.5)의 식별 및 모수 추정 결과는 Table 3.1에서 확인할 수 있다. 대부분의 추정값이 유의수준 1%에서 유의하여 모형 적합이 매우 잘 이루어지고 있음을 확인할 수 있다. Table 3.1에서 몇 가지 중요한 해석을 하자면 다음과 같다.

첫째, 경부축 주중 및 주말 모형에서는 경부 2단계 개통으로 개입 이전 대비 각각 약 20,000명과 약 30,747명이 일정 증가하였고, 약 373명과 340명 매월 선형 증가하는 것으로 나타났다. 둘째, 호남축 주중 및 주말 모형에서는 호남 2단계 개통으로 개입 2010년 11월 경부 2단계 개통 개입 이전 대비 각각 약 16,018명과 24,145명이 일정 증가하였고, 약 514명과 479명이 매월 선형 증가하였다. 이를 2012년 10월 운송력 증대 개입 이전으로 비교하면 $w_{1,15} + w_{2,15} - w_{2,12}$ 값으로 계산하여 약 10,048명과 16,899명이 일정 증가하였고, 매월 선형 증가량은 이전 해석과 동일하다. 둘째, 메르스의 영향으로 경부축 주중 및 주말 모형에서 2015년 6월에 각각 약 25,140명과 약 43,174명이 7월에는 9,218명과

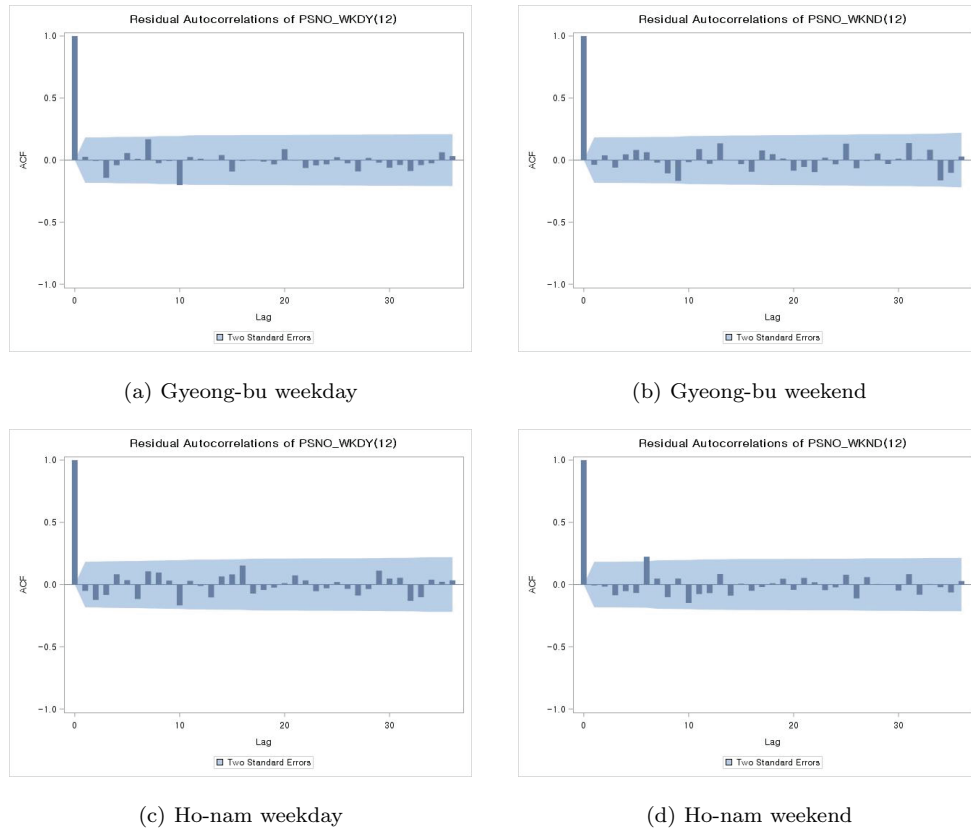


Figure 3.1. Plot of residual ACF.

Table 3.2. Residual analysis of multiple intervention seasonal ARIMA models for Gyeong-bu and Ho-nam line

| Lines | To lag | Weekday | | | Weekend | | |
|-----------|--------|----------|----|---------------|----------|----|---------------|
| | | χ^2 | DF | $Pr > \chi^2$ | χ^2 | DF | $Pr > \chi^2$ |
| Gyeong-bu | 6 | 3.13 | 3 | 0.372 | 3.08 | 3 | 0.379 |
| | 12 | 12.07 | 9 | 0.210 | 8.40 | 9 | 0.495 |
| | 18 | 13.34 | 15 | 0.576 | 14.26 | 15 | 0.506 |
| Ho-nam | 6 | 4.33 | 3 | 0.228 | 8.25 | 3 | 0.041 |
| | 12 | 11.64 | 9 | 0.234 | 13.24 | 9 | 0.152 |
| | 18 | 19.39 | 15 | 0.197 | 15.63 | 15 | 0.407 |

8,565명이 일시 감소하였고, 호남축 주중 및 주말 모형에서 2015년 6월에 각각 약 7,098명과 11,828명이, 7월에는 약 2,374명과 3,897명이 일시 감소하였다. 끝으로 평일이 1-2일이 포함되는 명절의 경우에 경부축과 호남축 주중 수송수요가 각각 5,772명과 1,498명이며, 평일이 3일 포함되는 경우에는 각각 8,953명과 2,673명이 일시적으로 증가하였다.

Table 3.1의 적합된 시계열 모형에 대해 잔차분석을 실시하여 Figure 3.1와 Table 3.2에 나타내었다. Figure 3.1의 자기상관성 도표와 Table 3.2에 있는 적합된 각 모형의 잔차에 대한 룰-박스 검정통계량 값과 유의확률에 의하면 경부축의 주중과 주말, 호남축 주중 모형 모두 적합이 잘 된 것으로 보인다. 호

Table 3.3. Performances of multiple intervention seasonal ARIMA models for Gyoung-bu and Ho-nam Line

| Lines | Weekday | | | Weekend | | |
|-----------|---------|--------|------|---------|--------|------|
| | RMSE | MAE | MAPE | RMSE | MAE | MAPE |
| Gyoung-bu | 2257.6 | 1673.9 | 1.34 | 3942.5 | 3397.8 | 2.17 |
| Ho-nam | 1619.6 | 1332.1 | 3.83 | 1792.3 | 1700.0 | 3.66 |

RMSE = root mean squared error; MAE = mean absolute error; MAPE = mean absolute percent error.

남측 주말의 경우 Figure 3.1에서 6시차의 자기상관성이 유의하였고, 이로 인해 Table 3.2에서 자유도가 3인 경우의 릉박스 검정에서 잔차의 자기상관성이 존재한다는 대립가설이 채택되었으나, 이를 제외하면 큰 문제가 없으므로 마찬가지로 적합이 적절하게 이루어진 것으로 보인다.

3.2. 모형검증

Table 3.1의 적합된 다중개입 계절형 ARIMA 모형의 예측력 검증을 위해 2016년 1월부터 2016년 11월 간의 예측치와 실측치를 제공근평균제곱오차(root mean squared error; RMSE), 평균절대오차(mean absolute error; MAE), 평균절대백분율 오차(mean absolute percent error; MAPE)로 비교하였다. 표본구간 내 예측력 검정을 실시한 결과 MAPE 기준으로 경부축 주중과 주말은 각각 1.34%, 2.17%, 호남축 주중과 주말은 각각 3.83%, 3.66%로 나타났다. 경부축에서는 RMSE, MAE, MAPE 모두 주중에 비해서는 주말이 더 크게 나타나고 있으며, 호남축에서는 RMSE, MAE는 주중이, MAPE는 주말이 더 크게 나타나고 있다. 경부축보다는 호남축의 수송수요가 안정이 덜 되어 있어 변동성이 상대적으로 크기 때문에 예측오차가 높게 나타난 것으로 보인다.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^T (\hat{Y}_t - Y_t)^2}, \quad \text{MAE} = \frac{1}{n} \sum_{t=1}^T |\hat{Y}_t - Y_t|, \quad \text{MAPE} = \frac{1}{n} \sum_{t=1}^T \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right| \times 100.$$

4. 결론

본 연구는 KTX 수송 수요를 예측하기 위한 방법으로 다중개입 시계열 모형을 제안하였다. 구체적으로 2010년 11월 경부 2단계 개통, 2015년 4월 호남 2단계 개통, 2015년 6월과 7월의 메르스 사태 등의 개입 효과를 분석하기 위하여 다중개입 계절형 ARIMA 모형을 도입하였다. KTX 수송수요 자료 분석의 중요한 결과는 다음과 같다. 경부축과 호남축의 주중 및 주말 수송수요 다중개입 계절형 ARIMA 모형에서 대부분의 모수 추정 결과 통계적으로 매우 유의하여 다양한 개입 효과를 성공적으로 설명하였다. 또한 릉박스 검정과 잔차 ACF 그림을 통해 모형 적합이 매우 적절하게 이루어졌음을 확인하였다. 끝으로 2016년 1월부터 11월까지 예측을 하여 모형 검증을 한 결과 MAPE 기준으로 경부축 주중과 주말이 각각 1.34%와 2.17%로서 예측의 정확도가 매우 높았으며, 호남축 주중과 주말도 각각 3.83%와 3.66%로 예측오차가 높지 않음을 확인하였다. 요약하자면 다중개입 계절형 ARIMA 모형을 도입하여 KTX 수송수요에 영향을 주었던 다양한 개입의 효과를 효과적으로 해석하는 한편, 이를 통해 예측의 정확성을 높일 수 있음을 확인하였다.

References

- Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with application to economic and environmental problems, *Journal of the American Statistical Association*, **70**, 70–79.

- Cho, S., Sohn, Y. and Seong, B. (2016). *Time Series Analysis using SAS/ETS*, Yulgok Books, Seoul.
- Choi, T. and Kim, S. (2004). An empirical comparison among initialization methods of Holt-Winters model for railway passenger demand forecast, *Journal of the Korean Society for Railway*, **7**, 9–13.
- Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis: With Applications in R*, Springer-Verlag, New York.
- Hong, C. and Lee, G. (2016). Study of factors affecting the size of government employment in Korea, *Korean Public Personnel Administration Review*, **15**, 29–60.
- Huh, H. J. and Kim, H. C. (2001). Forecasting demand for Jeju-bound tourist: an application of intervention method, *Journal of Tourism Sciences*, **25**, 27–42.
- Kim, K. and Kim, H. (2011). KTX passenger demand forecast with intervention ARIMA model, *Korean Society for Railway*, **14**, 470–476.
- Kim, S. and Seong, B. (2011). Intervention analysis of Korea tourism data, *Korean Journal of Applied Statistics*, **24**, 735–743.
- Kim, Y. and Jo, J. (2014). A study on demand forecasting for KTX passengers by using time series models, *Korean Journal of Applied Statistics*, **27**, 1257–1268.
- Song, J. (2016). A study on demand forecasting for Jeju-bound tourists by travel purpose using seasonal ARIMA-intervention, *Journal of the Korean Data & Information Science Society*, **27**, 725–732.

다중개입 계절형 ARIMA 모형을 이용한 KTX 수송수요 예측

차효영^a · 오윤식^b · 송지우^a · 이태욱^{a,1}

^a한국외국어대학교 통계학과, ^b한국철도공사연구원

(2018년 11월 15일 접수, 2018년 12월 8일 수정, 2018년 12월 8일 채택)

요약

본 연구는 KTX 수송수요를 예측하기 위한 방법으로 다중개입 시계열 모형을 제안하였다. 구체적으로 2011년 이전의 자료로서 경부 2단계 개통 개입만 고려한 Kim과 Kim (*Korean Society for Railway*, 14, 470–476, 2011)의 연구를 수정 보완하기 위해 다양한 개입이 추가적으로 발생하고 있는 2011년 이후의 시계열 자료를 효과적으로 모델링하는 한편 KTX 수송수요를 정확히 예측하기 위한 방법으로 다중개입 계절형 ARIMA 모형을 도입하였다. 자료 분석을 통해 KTX 수송수요에 영향을 주었던 경부 및 호남 2단계 개통, 메르스 발병과 설추석 명절 등 다양한 개입의 효과를 효과적으로 해석하는 한편, 이를 통해 예측의 정확성을 높일 수 있음을 확인하였다.

주요용어: 고속철도, 개입모형, 계절시계열, 시계열, 예측

이 논문은 2018학년도 한국외국어대학교 교내 학술연구비의 지원에 의하여 이루어진 것임.

¹교신저자: (17035) 경기도 용문인시 처인구 모현읍 외대로 81, 한국외국어대학교 통계학과.

E-mail: twlee@hufs.ac.kr