

개입 ARIMA 모형을 이용한 KTX 수요예측

KTX Passenger Demand Forecast with Intervention ARIMA Model

김관형¹ · 김한수[†]

Kwanhyung Kim · Hansoo Kim

Abstract This study proposed the intervention ARIMA model as a way to forecast the KTX passenger demand. The second phase of the Gyeongbu high-speed rail project and the financial crisis in 2008 were analyzed in order to determine the effect of time series on the opening of a new line and economic impact. As a result, the financial crisis showed that there is no statistically significant impact, but the second phase of the Gyeongbu high-speed rail project showed that the weekday trips increased about 17,000 trips/day and the weekend trips increased about 26,000 trips/day. This study is meaningful in that the intervention explained the phenomena affecting the time series of KTX trip and analyzed the impact on intervention of time series quantitatively. The developed model can be used to forecast the outline of the overall KTX demand and to validate the KTX O/D forecasting demand.

Keywords : Intervention ARIMA, Seasonal ARIMA, KTX Passenger Demand, Time Series Model, Kyongbu HSR 2 Phase

초 록 본 연구는 KTX 수요를 예측하기 위한 방법으로 개입 ARIMA 모형을 제안하였다. 신선개통과 경제충격으로 인한 시계열의 영향 여부를 파악하기 위해 경부고속철도 2단계 개통과 2008년 금융위기를 분석하였다. 분석결과 금융위기는 통계적으로 유의미한 영향이 없는 것으로 나타났으나, 경부고속철도 2단계는 주중 통행량 17,000 통행/일, 주말 통행량 26,000 통행/일 정도 증가한 것으로 나타났다. 본 연구는 개입이 통행량 시계열에 영향을 미치는 현상을 파악하고, 시계열 자료에 대한 개입효과를 계량적으로 분석했다는 점에서 의의가 있다. 개발된 모형은 KTX 전체 수요를 개략적으로 예측하는데 활용될 수 있으며, KTX O/D별 예측치를 검증하는데 활용이 가능하다.

주요어 : 개입 ARIMA, 계절형 ARIMA, KTX 수요, 시계열 모형, 경부고속철도 2단계

1. 서 론

효율적인 철도운행을 위해서는 각종 의사결정의 기초자료인 철도수요를 예측하는 것이 필요하다. 철도수요의 예측방법은 크게 과거 수송실적에 기반한 패턴인식방법과 열차운행횟수, 운임 등의 정책 변화를 고려할 수 있는 회귀모형, 직접수요모형, 전통적 4단계 모형으로 구분할 수 있다. 본 연구는 패턴인식방법 중 시계열 모형에 대한 것이다. 시계열 모형은 적은 자료만으로도 빠른 시간 내에 장래 수요를 예측할 수 있을 뿐만 아니라 예측의 정확도도 다른 방법에 비해 우수하다. 그러나 정책 변화를 반영하기 어렵기 때문에 장래 여건 변화에 따른 수요변화를 분석할 수 없는 단점이 있다. 그럼에도 불구하고 시계열 모형의 예측결과는 다른 방법에 의한 예측결과의 적정성을 판단할 수 있기 때문에 활용도가 높다.

시계열 모형을 이용한 철도수요예측의 국내 연구는 Lee et al. [1], Oh와 Hong [2], Oh와 Kim [3], Choi와 Kim [4]가 있

으나, 초기 연구들로 철도수요예측에 활용하기 어렵다. 이들 연구 중 Choi와 Kim [4]는 Holt-Winters 모형을 적용하여 시계열의 계절성과 추세성을 수요예측에 반영하였다. 국외 연구는 Tsai et al. [5]가 요일별, 월별 특성을 반영한 비선형 자기회귀 신경망(non-linear autoregressive neural network)을 제안하였으며, 대만철도 1999년~2000년 수송실적으로 신경망모형을 구축하여 장래 수요를 예측 하였다. 기존 연구들이 시계열의 계절적 특성을 반영하기 위한 시도를 주목할 필요가 있다.

본 연구는 KTX 수요를 예측하기 위한 방법으로 개입 ARIMA 모형(intervention ARIMA model)을 제안한다. 개입 ARIMA 모형은 시계열 자료에 영향을 미칠 가능성이 있는 사건인 개입을 반영할 수 있는 시계열 모형이다. 이 모형은 최근 개통된 경부고속철도 2단계와 같은 신선 건설 영향을 반영하여 장래를 예측하는 것이 가능하다. 본 연구는 2장에서 연구자료와 ARIMA 모형을 이용한 연구방법에 대해 설명한다. 3장에서는 KTX 수요예측에 적합한 개입 ARIMA 모형의 식별과 추정단계를 통해 경부고속철도 2단계 개통과 2008년 금융위기로 인한 개입영향을 분석한다. 모형 검증단계에서는 계절형 ARIMA 모형(seasonal ARIMA model)과 개입 ARIMA 모형을 이용한 예측결과를 설명하고, 실측치

[†]교신저자 : 한국철도공사 연구원
E-mail : hansman@korail.com

¹한국철도공사 대전충남본부

와 비교를 통해 개입 ARIMA 모형의 예측력에 대해 논한다. 마지막으로 연구결과를 정리하고, 향후 연구과제를 제시한다.

2. 연구자료 및 ARIMA 모형

2.1 연구자료

본 연구에서 이용한 자료는 2004년 4월 KTX 개통 이후 2011년 7월까지의 KTX 전체 월별 일평균 통행량(통행/일)이다. 연구자료를 검토한 결과 다음과 같은 분석방향을 설정하였다. 첫째, KTX 통행량을 주중과 주말로 구분하여 예측한다. Kim과 Yun [6]에서 제시된 바와 같이 KTX 통행량은 주중(월~목)과 주말(금~일)로 구분되는 특징이 있다. 이러한 구분은 주중과 주말 통행량 변화의 차이를 반영할 수 있으므로 장래 예측의 정확도를 향상시킬 수 있다. KTX 월별 일평균 통행량을 주중과 주말 통행량으로 구분하여 Fig. 1에 표현하였다. 2004년 4월 개통 후 지속적으로 증가하다 2006년부터 점증적 추세를 보이고 있으며, 2010년 11월 경부고속철도 2단계 개통과 2010년 12월 경전선 KTX 직결로 크게 증가하였다. 연도별 월별 일평균 통행량에 대한 요약 통계는 Table 1과 같다. 경부고속철도 2단계 개통으로 주중 통행량 보다 주말 통행량 증가가 큰 것을 확인할 수 있다.

둘째, KTX 통행량의 월별 계절변동을 반영한다. Fig. 2는 KTX 통행량이 안정된 기간(2006년~2009년) 동안의 월별 주중과 주말 일평균 통행량이다. 주중 기준으로는 2월, 8월이 높은 반면 3월이 낮으며, 주말 기준으로는 4-5월, 10-12월이 높은 반면 3월, 7월이 낮다. KTX 통행량은 주중과 주말 통행량 모두 계절적인 주기를 가지며, 주중과 주말 통행량의 월별 변동에 차이가 존재하기 때문에 이러한 특성을 반영하는 것이 필요하다. Choi와 Kim [4]는 시계열을 구성하는 성분들이 결정적이며, 서로 독립이라는 가정 하에 Holt-Winters 모형을 적용하였다. 그러나 대부분의 시계열은 그 구성성분이 확률적이거나 다른 성분들과 상관이 있는 경우가 많기 때문에 본 연구에서는 이를 반영할 수 있는 계절형 ARIMA 모형을 적용한다.

셋째, 경부고속철도 2단계 등과 같은 개입을 반영한다. 시계열 자료는 비교적 장기간에 걸쳐서 자료가 관측되기 때문에 신규 철도노선 건설, 경제충격 등과 같은 제어할 수 없는 여러 가지 사건이 발생한다면 기존에 관측된 시계열 자료와는 다른 패턴을 갖는 자료가 관측될 것이다. 시계열 자료가 만일 분석에 영향을 미칠 수 있는 사건 등에 의해 영향을 받는 경우 이를 분석에 반영하지 않는다면 그 분석결과를 신뢰하기 어려울 것이다. 이와 같이 시계열 자료에 영향을 미칠 가능성이 있는 사건들을 개입이라고 부르며 이들을 입력변수로 사용하는 시계열 모형을 개입모형이라고 한다. Box와 Tiao [7]은 고속도로 개통 및 자동차 배기가스의 발생을 규제하기 위한 여러 가지 법안 통과 등의 일련의 사건들이 Los Angeles 지역의 오존발생량에 어떠한 영향을 주었는지를 분석하기 위해 처음으로 개입모형을 이용한 분석을 시도하였다[8]. 본 연구에서는 신규 철도노선 건설과 경

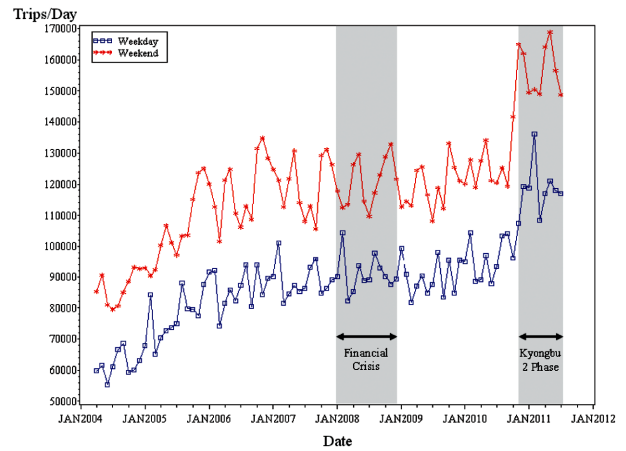


Fig. 1 Monthly average daily trips (APR2004 to JUL2011)

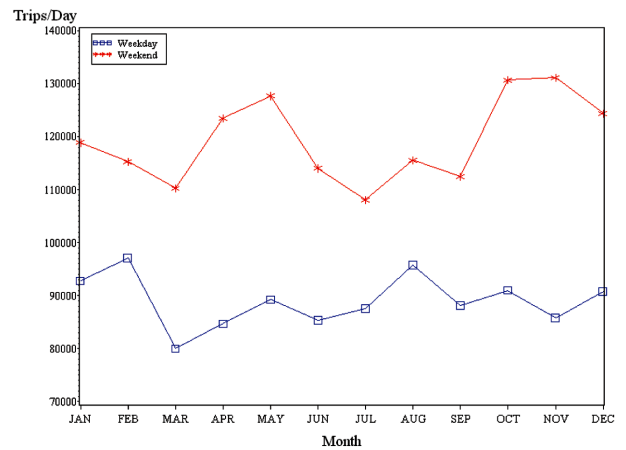


Fig. 2 Average of Monthly average daily trips (2006 to 2009)

제충격으로 인한 개입을 분석하여 본다. 우선 Fig. 1의 오른쪽 음영과 같이 경부고속철도 2단계 개통으로 인한 통행량 증가가 시계열의 구조변화를 시사할 정도의 통계적 유의성을 가지는지 분석한다. 경제충격으로 인한 개입은 2007년부터 2009년까지 진행된 서브 프라임 모기지 사태에 의한 금융위기를 대상으로 한다. 금융위기는 2008년이 정점으로 Fig. 1의 왼쪽 음영에 해당한다. 국제여행객수요는 경제충격에 영향을 받는 것으로 분석되었으나[9], 국내여행객수요에 대한 연구사례는 찾아볼 수 없다. 경제충격으로 인한 KTX 수요변화 여부를 확인하는 것은 향후 발생할 경제충격을 반영하는 것이 필요한지 여부를 판단할 수 있는 근거가 될 것이다. 본 연구는 이러한 분석결과를 바탕으로 장래 예측을 실시하여 그 결과에 대한 의미를 논의할 것이다.

마지막으로 모형식별, 추정 및 검증에 적합한 분석기간을 설정한다. 본 연구는 개입발생 후 단기간 내 장래 예측을 해야 하는 필요에 적합하도록 Fig. 4에 표시한 바와 같이 분석기간을 설정하였다. 모형식별은 개입이 존재하기 전의 자료를 이용해야 하며, 충분한 관측개수가 필요하다. 모형추정은 개입이 발생한 이후의 자료를 포함해야 한다. 모형식별

Table 1 Summary statistics of monthly average daily trips on weekday and weekend

Year	Weekday				Weekend			
	Mean	Standard deviation	Min.	Max.	Mean	Standard deviation	Min.	Max.
2004	61,707	3,987	55,286	68,617	86,293	5,203	79,607	93,089
2005	76,813	7,358	65,238	87,957	104,360	11,617	90,372	125,219
2006	86,430	6,098	74,216	93,911	117,856	10,597	101,720	134,914
2007	88,806	5,549	81,639	101,235	119,945	8,947	105,692	131,254
2008	90,995	5,809	82,389	104,454	120,695	7,555	109,824	132,933
2009	89,970	5,948	81,857	99,492	118,875	7,258	108,277	133,193
2010	98,835	9,281	87,792	119,308	132,010	16,199	119,003	165,021
2011	119,472	8,343	108,387	136,123	155,361	8,218	148,837	168,991

Unit: trips per day

과 추정을 위해 각각 79개월, 84개월을 분석기간으로 설정하였다. 모형검증기간은 현재 기준으로 수송실적이 충분히 확보되지 않았기 때문에 4개월을 분석기간으로 설정하였다.

- 모형식별 : 2004년 4월~2010년 10월 자료 (79개월)
- 모형추정 : 2004년 4월~2011년 3월 자료 (84개월)
(경부2단계 후 5개월)
- 모형검증 : 2011년 4월~2011년 7월 자료 (4개월)

2.2 ARIMA 모형

시계열이 일정한 계절주기를 가지고 변할 때 사용하는 계절형 ARIMA 모형은 시계열의 구성성분들이 확률적이거나 다른 성분들과 상관이 있는 경우에 사용할 수 있다. 시계열 Z_t 가 계절주기가 s 인 ARIMA(p,d,q)(P,D,Q)s 과정을 따를 때 식 (1) 형태의 모형으로 표현한다. 여기서, B 는 후행연산자(backshift operator), $\phi(B)$ 와 $\Phi(B^s)$ 는 각각 비계절형 및 계절형 AR연산자, $\alpha(B)$ 와 $\alpha(B^s)$ 는 각각 비계절형 및 계절형 MA 연산자, 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 평균에 관련된 상수이다.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2) \tag{1}$$

개입모형은 개입이 포함되지 않은 시계열 Z_t 가 ARIMA(p,d,q)(P,D,Q)s 모형을 따를 때, 외부적인 충격이 가법적인 형태로 영향을 미친다고 가정하여, 식 (2) 형태의 모형으로 표현한다. 여기서, b 는 개입효과 지연시차, $\omega(B)$ 는 개입의 초기 기대효과, $\alpha(B)$ 는 개입의 지속효과, I_t 는 개입변수를 의미한다. 개입변수들은 어떤 사건의 발생이 지속되는 기간에 따라 펄스함수(pulse function)와 계단함수(step function)의 두 가지 형태를 따른다. 펄스함수 $P_t^{(T)}$ 는 어떤 사건이 t 시점에서 발생하여 그 효과가 t 시점에만 영향을 미치는 경우에 사용하며, 식 (3)과 같이 정의한다. 반면 계단함수 $S_t^{(T)}$ 는 신

규 노선건설과 같이 어떤 사건이 t 시점에서 발생하여 그 효과가 발생시점 이후로 지속적으로 영향을 미치는 경우에 사용하며, 식 (4)와 같이 정의한다. Box와 Tiao [7]은 개입효과가 개입변수에 반영되는 형태 5가지를 제시하였으며, 본 연구는 그 중 개입효과가 즉시 반영되며, 그 크기가 ω 로 일정한 경우를 가정하여 식 (5)와 같이 개입모형을 설정하였다.

$$Y_t = \nu(B)B^b I_t + Z_t = \frac{\omega(B)}{\alpha(B)} B^b I_t + Z_t, \quad I_t = P_t^{(T)} \quad \text{또는} \quad I_t = S_t^{(T)} \tag{2}$$

$$P_t^{(T)} = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases} \tag{3}$$

$$S_t^{(T)} = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \tag{4}$$

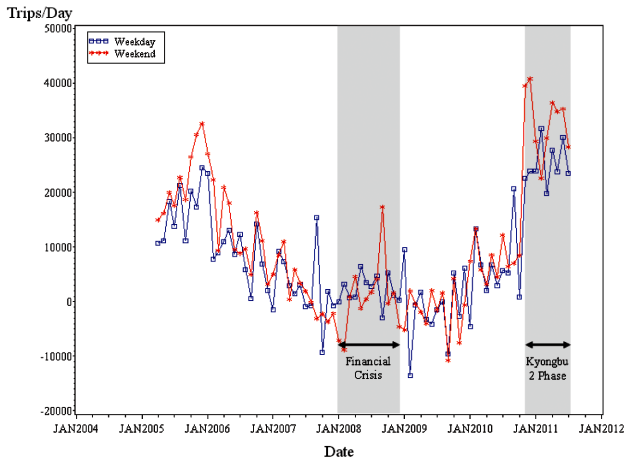
$$Y_t = \omega I_t + Z_t, \quad I_t = P_t^{(T)} \quad \text{또는} \quad I_t = S_t^{(T)} \tag{5}$$

3. 개입 ARIMA 모형 추정 및 검증

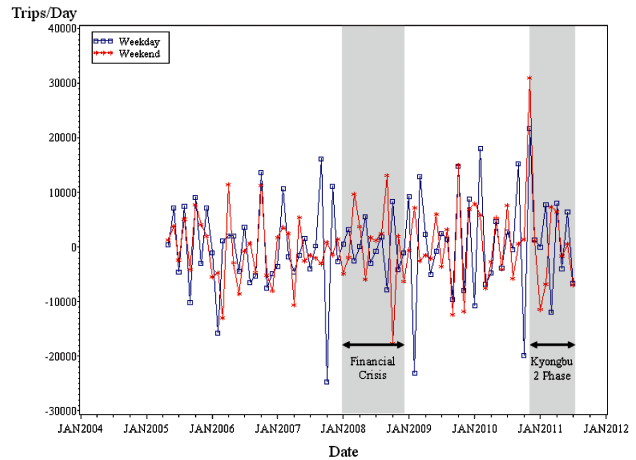
3.1 모형 식별

모형 식별단계는 ARIMA(p,d,q)(P,D,Q)s에서 차분 차수 $d+Ds$, AR 차수 $p+Ps$, MA 차수 $q+Qs$ 를 결정하는 단계이다. 첫 번째 단계는 시계열 그림과 표본상관도표(sample correlogram)를 그려 비정상이라고 판단되는 시계열을 정상시계열로 변환한다. Fig. 1에서 본 바와 같이 시계열의 분산이 시간의 흐름에 따라 분산이 일정하므로 로그변환 등의 분산안정화 변환은 불필요할 것으로 판단된다. 그러나 Fig. 2와 같이 계절성이 존재하기 때문에 계절 차분을 하며, Fig. 3의 a와 같이 계절 차분된 시계열에 추세가 존재하기 때문에 1차 차분을 한다. 계절 및 1차 차분을 한 결과 Fig. 3의 b와 같이 정상시계열로 변환 하였다.

두 번째 단계는 ARIMA 모형의 차수를 결정한다. 차수 결



(a) Trips after first difference at lag 12



(b) Trips after first difference at lag 1 then lag 12

Fig. 3 Monthly average daily trips after first difference (APR2004 to JUL2011)

Table 2 Model Identification for first differenced at lag 1 then at lag 12 series

	Weekday Models				Weekend Models			
	With constant		Without constant		With constant		Without constant	
	$(0,1,1)(1,1,0)_{12}$	$(0,1,1)(0,1,1)_{12}$	$(0,1,1)(1,1,0)_{12}$	$(0,1,1)(0,1,1)_{12}$	$(0,1,1)(1,1,0)_{12}$	$(0,1,1)(0,1,1)_{12}$	$(0,1,1)(1,1,0)_{12}$	$(0,1,1)(0,1,1)_{12}$
MA(1)	0.61050 (6.02)***	0.56503 (5.59)***	0.60975 (6.04)***	0.56533 (5.63)***	0.24228 (2.03)**	0.27204 (2.33)**	0.24101 (2.03)**	0.27203 (2.34)**
AR(12)	-0.38667 (-2.78)***	-	-0.38878 (-2.82)***	-	-0.44967 (-3.06)***	-	-0.45119 (-3.10)***	-
MA(12)	-	0.76402 (8.08)***	-	0.76511 (8.42)***	-	0.65509 (6.04)***	-	0.65572 (6.19)***
δ	95.41923 (0.41)	12.25112 (0.10)	-	-	95.01017 (0.22)	17.43183 (0.07)	-	-
AIC	1452.365	1436.537	1450.541	1434.547	1456.898	1448.349	1454.947	1446.354
SBC	1459.153	1443.325	1455.067	1439.073	1463.686	1455.137	1459.472	1450.879

Note : 1. Values in parenthesis are t-statistics. ** at 5% significance level, *** at 1% significance level.
2. Estimated models using monthly average daily trips from APR2004 to OCT2010.

정방법은 표본자기상관함수(Sample Auto-Correlation Function; SACF)와 표본부분자기상관함수(Sample Partial Auto-Correlations Function; SPACF)를 이용하여 Table 2와 같이 ARIMA(0,1,1)(0,1,1)₁₂, ARIMA (0,1,1)(1,1,0)₁₂ 모델을 식별하였으며, 모형선택 기준으로 사용되는 AIC(Akaike's Information Criterion)와 SBC(Schwartz's Bayesian Criterion) 통계량을 이용하여 ARIMA(0,1,1)(0,1,1)₁₂를 잠정모형으로 선택하였다. 마지막 단계는 시계열을 차분하였기 때문에 상수항 δ 를 모형에 포함시킬 것인지 여부를 결정해야 한다. 상수항 δ 를 모형에 포함시킬 경우 통계적으로 유의하지 않으므로 제외된 모형으로 최종 결정하였다.

3.2 모형 추정

본 연구는 2008년 금융위기와 경부고속철도 2단계 개통이 시계열에 영향을 주었는지를 파악하기 위해 다음과 같이

세 가지 모형을 구분하여 적용한다. 개입을 반영하기 위한 I08은 2008년 금융위기에 따른 변화가 있었는지를 검정하기 위한 더미변수로 2008년을 1, 나머지 기간은 0의 값을 가진다. 또한 I10은 경부고속철도 2단계 개통에 따른 변화가 있었는지를 검정하기 위한 더미변수로 2010년 11월 이후를 1, 나머지 기간은 0의 값을 가진다. 모형 추정은 SAS/ETS를 이용하였으며, 모형추정방법은 조건부 최소제곱추정법(conditional least squares estimation method)을 이용하였다.

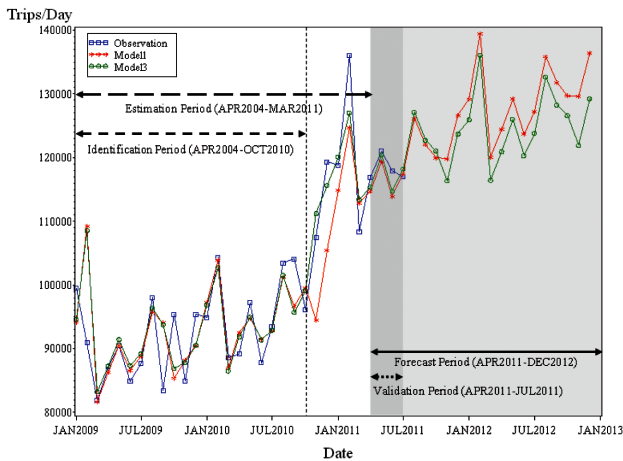
- 모형 1 : 계절형 ARIMA 모형-계절, 1차 차분모형
- 모형 2 : 개입 ARIMA 모형-계절, 1차 차분모형 (금융위기, 경부고속철도 2단계 개통)
- 모형 3 : 개입 ARIMA 모형-계절, 1차 차분모형 (경부고속철도 2단계 개통)

모형추정 결과는 Table 3에서 확인할 수 있듯이 AIC와 SBC 통계량이 최소인 경부고속철도 2단계 개통을 반영한 모

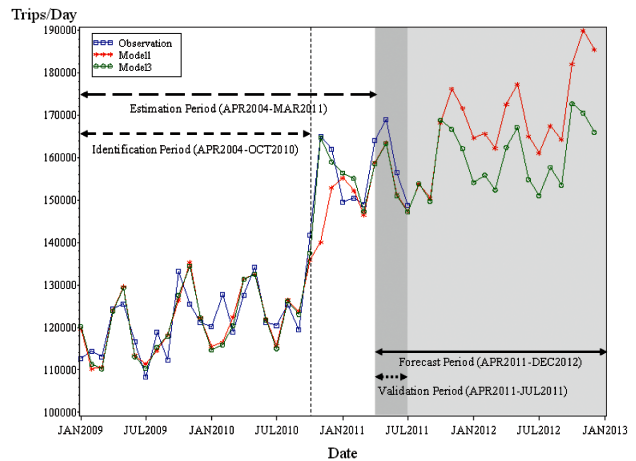
Table 3 Estimated models with and without intervention for first differenced at lag 1 then at lag 12 series

	Weekday Models			Weekend Models		
	SARIMA Model (w/o intervention)	Intervention ARIMA Models (with intervention)		SARIMA Model (w/o intervention)	Intervention ARIMA Models (with intervention)	
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MA(1)	0.56533 (5.63)***	0.68982 (7.65)***	0.68611 (7.64)***	0.27329 (2.35)**	0.44864 (3.94)***	0.44490 (3.97)***
MA(12)	0.76517 (8.42)***	0.71626 (7.59)***	0.71763 (7.68)***	0.65580 (6.20)***	0.54125 (4.76)***	0.54287 (4.82)***
I08	-	888.41294 (0.37)	-	-	568.54240 (0.20)	-
I10	-	17238.2 (4.38)***	17225.9 (4.38)***	-	26263.6 (5.19)***	26266.8 (5.22)***
AIC	1434.556	1423.293	1421.438	1446.521	1428.019	1426.063
SBC	1439.081	1432.343	1428.226	1451.046	1437.070	1432.851

Note : 1. Values in parenthesis are t-statistics. ** at 5% significance level, *** at 1% significance level.
 2. Estimated models using monthly average daily trips from APR2004 to MAR2011.



(a) Forecasting results of weekday models



(b) Forecasting results of weekend models

Fig. 4 Forecasting results (APR2011 to DEC2012)

형 3의 개입 ARIMA 모형이 가장 적합한 것으로 판명되었다. 모형 2에서 보는 바와 같이 2008년 금융위기는 영향을 미치지 않은 것으로 나타났다. I08의 계수 부호가 KTX 통행량이 증가하는 양(+)으로 나타났으며, 계수의 통계적 유의성도 없어 금융위기로 인해 통행량에 영향을 미쳤다고 보기 어렵다. 그러나 경부고속철도 2단계 개통은 KTX 통행량이 주중 17,000 통행/일, 주말 26,000 통행/일 정도 증가한 것으로 나타났으며, 유의수준 0.01에서 통계적으로 유의한 것으로 나타났다. 이러한 모형추정 결과 KTX의 장래 수요 예측 시 계절형 ARIMA 모형 보다는 개입 ARIMA 모형이 효과적일 것으로 판단된다.

3.3 모형 검증

계절형 ARIMA 모형과 개입 ARIMA 모형을 추정한 결

과 모형 1 보다 모형 3이 적합한 것으로 판명되었다. 이들 모형을 이용하여 Fig. 4의 음영부분에 해당하는 2011년 4월~2012년 12월에 대한 장래 예측을 실시하였다. 그 결과 2011년 4월~2011년 10월까지 큰 차이를 보이지 않다가, 2011년 11월 이후 예측치부터 차이가 커지는 것으로 나타났다. 모형 3에 비해 모형 1이 높은 예측결과를 제시하고 있다. 이러한 예측결과는 2010년 11월 이후의 통행량 증가를 일반적인 계절적 변동으로 반영하는지, 개입으로 인한 변동으로 반영하는지에 따라 차이가 발생하는 것이다. 모형 1은 2010년 11월 이후의 통행량 증가를 일반적인 계절적 변동으로 인식하였기 때문에 2011년과 2012년 11월에도 통행량이 크게 증가하는 계절적 변동이 존재할 것으로 예측하였다. 반면, 모형 3은 2010년 11월 이후의 통행량 증가를 개입으로 인한 변동으로 인식하였기 때문에 2011년과 2012년 11월에 급격

Table 4 Performances of ARIMA models with and without intervention

	Weekday Models		Weekend Models	
	Model 1 (without intervention)	Model 3 (with intervention)	Model 1 (without intervention)	Model 3 (with intervention)
RMSE	2,469	1,936	4,645	4,902
MAE	2,097	1,660	4,341	4,595
MAPE	1.77	1.41	2.68	2.84

한 증가가 존재하지 않을 것으로 예측하였다. 그러므로 모형 1의 비정상적인 증가 경향은 개입이 존재함에도 불구하고 이를 반영하지 못한 결과로 판단된다.

모형의 예측력 검증에 위해 수송실적이 확보된 2011년 4월~2011년 7월의 4개월간의 실측치와 모형 1과 모형 3의 예측치를 제곱근평균제곱오차(Root Mean Squared Error; RMSE), 평균절대오차(Mean Absolute Error; MAE), 평균절대백분율오차(Mean Absolute Percent Error; MAPE)로 비교한다. 표본구간 내 예측력 검증을 실시한 결과 주중모형은 모든 기준에서 모형 3이 낮은 것으로 나와 모형 3의 예측력이 상당히 신뢰할 수 있음을 알 수 있었다. MAPE가 1.41인 것은 직 전기 자료를 이용한 바로 다음 기의 예측은 약 1.41%의 오차가 있다는 것이다. 반면 주말모형은 모든 기준에서 모형 1이 낮은 것으로 나와 모형 추정결과와 앞에서 설명한 장래 예측결과 해석과 상반된다. 그러나 이는 예측력 검증을 위한 표본구간이 4개월로 짧고, 11월 이후의 실측치가 포함되지 않았기 때문에 나타나는 현상으로 이해할 수 있다. 즉, 2011년 11월 이후의 실측치까지 표본구간이 확대될 경우 모형 3의 예측력이 신뢰할 수 있음을 확인할 수 있을 것이다. 주말 모형 3의 MAPE는 2.84로 주중모형에 비해 높은 것을 알 수 있는데 이는 주말이 주중보다 변동성이 높기 때문에 예측오차가 상대적으로 높게 나타난 것이다.

KTX 통행량 시계열에 영향을 미치는지 분석하였다. 2008년 금융위기는 KTX 통행량 시계열에 의미 있는 영향을 미치지 못하는 것으로 나타난 반면 경부고속철도 2단계는 통계적으로 유의미한 영향으로 나타났다. 경부고속철도 2단계 개입효과는 주중 통행량 17,000 통행/일, 주말 통행량 26,000 통행/일 정도 증가한 것으로 나타났으며, 장래 예측에 있어서도 경부고속철도 2단계의 개입을 반영하지 않고는 적합한 예측결과를 산출하지 못할 것으로 판단되었다.

본 연구는 장래수요를 예측하는데 있어 빠른 시간 내에 적은 자료만으로도 예측이 가능한 시계열 모형을 이용하였다. 특히, 개입이 통행량 시계열에 영향을 미치는 현상을 파악하고, 시계열 자료에 대한 개입효과를 계량적으로 분석하였으며, 철도 통행량 예측에 이런 기법을 처음으로 적용했다는 점에서 의의가 있다. 이 모형은 KTX 전체 수요를 개략적으로 예측하는데 활용될 수 있으며, KTX O/D별 예측치를 검증하는데 활용이 가능하다. 그러나 시계열 모형의 특성 상 장래수요를 예측할 때 열차운행횟수, 운임, 운행시간 등의 변화를 반영할 수 없기 때문에 정책 활용에는 제약이 있다는 한계가 있다. 정책 활용과 철도수요의 정확성 향상을 위해서는 관련 자료를 축적할 수 있는 DB구축과 함께 과학적인 수요예측 기법에 대한 연구가 지속되어야 할 것으로 판단된다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{T}_t - T_t)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{T}_t - T_t| \quad (7)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{T}_t - T_t}{T_t} \right| \times 100 \quad (8)$$

4. 결 론

본 연구는 KTX 수요를 예측하기 위한 방법으로 개입 ARIMA 모형을 제안하였다. KTX 통행량은 2004년 4월 개통 이후 2006년까지 증가하다, 2010년 11월 경부고속철도 2단계 개통으로 인해 한 단계 증가하는 추세를 나타내고 있다. 개입 ARIMA 모형은 이와 같은 개입이 존재할 경우 효과적인 방법으로 사용될 수 있다. 본 연구에서는 경부고속철도 2단계의 신설 개통과 2008년 금융위기의 경제충격이

참고문헌

- [1] D. Lee, T. Hong, H. Kim, K. Woo (1998) Study on the Demand Prediction for Transportation System Utilizing Data Granulization, *Proceedings of the Conference of the Korean Society for Railway*, pp. 211-218.
- [2] S. Oh, S. Hong (2000) A study on the effect factors of the railway passenger demand forecasting by the disaggregate model, *Proceedings of the Conference of the Korean Institute of Electrical Engineers*, pp. 1445-1447.
- [3] S. Oh, D. Kim (2001) A Study on the Seasonal Decomposition of the Railway Passenger Demand, *Proceedings of the Conference of the Korean Society for Railway*, pp. 111-116.
- [4] T. Choi, S. Kim (2004) An Empirical Comparison among Initialization Methods of Holt-Winters Model for Railway Passenger Demand Forecast, *Journal of the Korean Society for Railway*, 7(1), pp. 9-13.
- [5] T. Tsai, C. Lee, C. Wei (2005) Design of dynamic neural networks to forecast short-term railway passenger demand, *Journal of the Eastern Asia Society for Transportation Studies*, 6,

- pp. 1651-1666.
- [6] H. Kim, D. Yun (2011) Travel Behavior Analysis for Short-term Railroad Passenger Demand Forecasting in KTX, *Proceedings of the Conference of the Korean Society for Railway*, pp. 1282-1289.
- [7] G. Box, G. Tiao (1975) Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association*, 70(349), pp. 70-79.
- [8] S. Cho, Y. Son (2010) Time Series Analysis using SAS/ETS, Yulgok Books, Seoul.
- [9] S. Mo (2011) Demand Pattern of the Global Passengers: Sea and Air Transport, *Journal of Korea Port Economic Association*, 27(1), pp. 1-11.
- 접수일(2011년 10월 5일), 수정일(2011년 10월 12일),
게재확정일(2011년 10월 26일)