

# 시계열 이상치 탐지 기법을 활용한 경부선 주요도시 철도 승객수의 이상치 탐색 연구

이지선 · 윤윤진\*

한국과학기술원 건설 및 환경공학과

## A Study on the Outliers Detection in the Number of Railway Passengers for the Gyeongbu Line From Seoul to Major Cities Using a Time Series Outlier Detection Technique

LEE, Jiseon · YOON, Yoonjin\*

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

\*Corresponding author: yoonjin@kaist.ac.kr

### Abstract

On April 1, 2004, KTX (Korea Train eXpress), the first HSR (High-Speed Rail) in Korea, was introduced to Gyeongbu Line. The introduction of the KTX service led to a change in the number of passengers for Gyeongbu Line. Previous studies have analyzed the pre and post-event changes of the intervening events by either simple statistics or intervention ARIMA analysis. However, the intervention ARIMA model has a limitation that several assumptions such as the occurrence time and the type of intervention events are necessary. To this end, this study analyzed the effects of intervention event on the number of passengers using the Gyeongbu line based on a time series outlier detection technique which can overcome limitations in the previous studies. The time series outlier detection technique can analyze the time, effect type and size of an intervention event without the assumption of the time and effect type of the intervention event. The data were collected from the Korea Transport Database (KTDB) for twelve years from 2003 to 2014 (144 months). The analysis results showed that the size of the influence type in the same intervention events was different across the major city routes, and the intervention event which could not be found by previous study methods was also found.

**Keywords:** Gyeongbu Line, intervention analysis, Korea Train eXpress (KTX), outlier detection, time series

### 초록

2004년 4월 1일, 국내 최초의 고속철도(HSR)인 KTX (Korea Train eXpress)가 경부선에 도입 되었다. KTX의 등장은 경부선을 이용하는 철도 승객들의 운송수단 선택 및 도시구간별 이용객 수 변화를 가져왔다. KTX의 등장과 같은 개입사건(Intervention events)의 영향은 개입사건 전후 변화를 단순 통계량으로 분석하거나 개입 ARIMA 모델을 통해 분석 되었다. 개입 ARIMA 모델은 개입사건의 발생 시점( $t$ )과 개입사건의 영향 형태(type) 등의 가정이 필요하다

J. Korean Soc. Transp.  
Vol.35, No.6, pp.469-480, December 2017  
<https://doi.org/10.7470/jkst.2017.35.6.469>

pISSN : 1229-1366  
eISSN : 2234-4217

Received: 30 December 2016

Revised: 28 February 2017

Accepted: 23 December 2017

Copyright ©  
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

는 한계가 있었으며, 본 연구에서는 기존 연구에서의 한계점을 보완할 수 있는 시계열 이상치 탐지(time series outlier detection)를 활용하였다. 일반적으로 개입사건의 발생시기는 잘 알려져 있지 않으므로 시계열 이상치 탐지를 통해 개입사건에의 발생 시기를 추정할 수 있다. 시계열 이상치 탐지기법을 활용하여 개입의 시점과 영향 형태에 관한 가정 없이 개입사건에 대한 영향을 분석할 수 있으며, 발생된 이상치의 시점을 개입사건의 시점, 이상치의 영향을 개입사건의 영향으로 가정하였다. 데이터는 KTDB (Korea Transport Database)로 부터 KTX가 도입되기 이전인 2003년부터 2014년까지 12년 동안의 경부선(4개의 주요 도시구간 합산)을 포함한 주요 도시구간 4개의 월별데이터를 수집하여 활용하였다. 경부선 도시 구간별 이상치를 탐지 하고 그 영향을 분석한 결과, 동일한 개입사건 임에도 그 영향의 형태의 정도가 도시구간마다 다르게 나타나거나 영향이 나타나지 않았으며, 기존 연구에서 분석되지 않은 개입사건을 찾을 수 있었다.

**주요어:** 경부선, 개입분석, KTX, 이상치 탐지, 시계열

## 서론

### 1. 연구의 배경 및 목적

2004년 4월, 국내 최초 고속철도(High Speed Rail, HSR)인 KTX (Korea Train eXpress)가 경부선에 도입되었다. 서울에서 대구까지 고속선(HSR line)을 신설하고, 대구에서 부산은 기존선로를 활용하는 형태로 1단계가 개통되었으며, 2010년 10월에 2단계 공사가 완료됨에 따라 서울-부산 전 구간이 고속선으로 연결되었다. KTX의 등장으로 서울에서 부산까지 KTX (2시간 20분)의 이동시간은 기존 무궁화호(5시간 30분)와 새마을호(4시간 20분) 대비 절반이상 단축 되었으며, 신규 편성된 KTX의 운행으로 무궁화호와 새마을호의 운행횟수가 감소되었다. 경부고속철도 등장과 같은 개입사건(Intervention event)은 경부선 주요 도시 간 철도 승객 수에 영향을 끼쳤으며 선행연구에서는 개입에 대한 영향을 개입사건 전후의 변화로 분석하였다.

Suh and Yang(2005)는 1단계 고속철도가 개통된 당해년도 3개월(2004.04-06)과 직전년도의 동일한 3개월(2003.04-06)의 철도 승객 수를 비교하여, KTX의 등장에 따른 경부선/호남선의 운송수단별 수요 변화를 분석하였다. 국내 고속철도 도입이후 경부선의 철도 수요가 25% 증가하였으나 항공노선의 경우 서울-대구 노선과 서울-부산 노선이 각각 70%와 28.6%의 감소를 보였다. Lee and Chang(2006)는 서울에서 KTX의 서비스를 제공하는 도시구간에 대한 운송수단별 승객 수 변화를 KTX의 도입 전(2003.04-2004.03)과 도입 이후(2004.04-2005.03)로 비교하였으며, 그 결과 기존의 무궁화호와 새마을호의 환승 없이 KTX의 이용으로 이동이 가능한 도시구간에서 철도 승객 비율이 높게 나타났고, KTX와 기존 열차와의 연계 서비스가 필요한 도시구간에서는 고속버스 승객 비율이 높게 나타났다.

개입영향의 정도는 개입사건 전후의 통계량 분석 뿐 아니라 개입 ARIMA 모형을 활용하여 개입효과와 형태를 시계열에 반영하여 분석할 수도 있다. Kim and Kim(2011)은 고속철도 2단계 개통과 같이 신선 개통과 2008년 금융위기와 같은 경제상황을 개입 변수로 하여 KTX의 수요를 예측하는 모형을 개발하였다. 2008년 금융위기는 통계적으로 유의미한 영향이 나타나지 않았으나 고속철도 2단계 개통으로는 승객수가 증가한 것으로 나타났다. 개입의 발생시점(t)과 영향 형태(type)를 알고 있는 경우에, 개입이 실제로 시계열 자료에 영향을 끼쳤는지 여부와 그 영향의 정도를 분석하여 시계열 모형에 반영하기 위하여 시계열 개입분석 기법이 도입되었다(Box and Tiao, 1975). 시계열 개입분석은 정책도입, 금융위기, 재난영향 및 의학 등의 연구에서 많이 활용되고 있으나 개입분석을 수행하기 위해서는 개입된 사건의 발생 시점과 영향의 형태에 대한 가정이 필요하다는 한계점이 있다(Kim et al., 2013; Park 2004; Park et al., 2015). 최근 경부선에는 고속철도 도입, 대구-부산 고속도로 개통, 저비용항공사(Low Cost Carrier, LCC) 등장, 2008년 경제 위기 등의 개입 사건들이 있었다.

본 연구에서는 시계열 이상치 탐지(time series outlier detection) 기법을 활용하여 개입에 대한 영향을 분석하고자 한다. 이상치 탐지는 기존의 데이터 분포와 비교해서 정상 데이터 분포에서 크게 벗어나는 데이터를 탐지하는 것을 말하며, 시계열 이상치 탐지는 특정 시간 주기 패턴을 파악하고 이에 벗어나는 이벤트가 발생하는 이상 현상을 탐지하는 것이다. 일반적으로 개입사건의 발생시기가 알려져 있지 않으므로 이상치 탐지를 통해 개입사건의 발생 시기를 추정할 수 있다. 시계열 이상치 탐지에서 발생한 이상치의 시점을 개입사건의 시점, 이상치의 영향을 개입사건의 영향으로 가정하였다. 시계열 이상치 탐지는 이전의 개입 ARIMA 분석에서 필요한 개입사건의 발생시점이나 영향 형태에 대한 가정 없이 개입사건의 영향을 분석할 수 있기 때문에 기존의 연구에서 분석하지 않았던 개입사건을 탐지할 수 있다. 본 연구에서는 경부 고속철도가 도입되기 이전의 2003년부터 2014년까지의 철도 승객 월별 데이터를 경부선의 주요 도시 구간인 서울-부산, 서울-대구, 서울-대전, 서울-천안을 중심으로 탐지된 이상치를 통하여 철도 수요에 직/간접적으로 영향을 주는 개입 사건과 그 영향을 분석하였다.

## 분석자료

경부선의 주요 도시 구간을 이용한 철도 승객 수를 Table 1과 같이 수집하였다. 데이터는 KTDB (Korea Transport Database)를 통해서 경부선의 주요 도시구간인 서울-부산, 서울-대구, 서울-대전 및 서울-천안 등의 4개 구간으로 한정하여 2003년부터 2014년까지의 월별 철도 승객 수 데이터를 수집하였다.

**Table 1. Descriptive data**

Data	Type	Source
Transportation type	Train	Korea Transport DataBase (KTDB) <a href="https://www.ktdb.go.kr/">https://www.ktdb.go.kr/</a>
	Conventional train -Samaeul (higher-grade) -Mugunghwa (lower-grade) Express Train (KTX)	
Route	Gyeongbu line	Seoul-Busan (442 km) Seoul-Daegu (326 km) Seoul-Daejeon (166 km) Seoul-Cheonan (102 km)

Table 1과 같이, 기차 승객은 무궁화호, 새마을호, 누리호(서울-천안 구간만 해당) 및 KTX를 이용한 승객 수를 경부선의 주요도시 구간별로 합산하여 산정하였다. KTDB는 출발역과 도착역 구간에 따른 OD (Origin & Destination)로 승객 수를 제공하고 있다. 본 연구에서 수행하고자 하는 도시 구간의 출/도착 도시에서 2개 이상의 역이 존재 하는 경우에는 한 도시구간에서 다수의 OD가 생기므로 관련 OD 별 승객 수를 합산 하여 분석에 활용하였다. 주요 도시구간별 포함되는 OD를 Table 2와 같이 정리하였다.

**Table 2. Origin and Destination pair from Seoul to major cities**

OD	Origin and destination pair
Seoul-Busan	HSR Seoul-Busan St.
	Conventional Seoul-Busan St., Yongsan-Busan St., Yeongdeungpo-Busan St., Seoul-Haeundae St., Yeongdeungpo-Haeundae St.
Seoul-Daegu	HSR Seoul-Dongdaegu St.
	Conventional Seoul-Daegu St., Yongsan-Daegu St., Yeongdeungpo-Daegu St., Seoul-Dongdaegu St., Yongsan-Dongdaegu St., Yeongdeungpo-Dongdaegu St.
Seoul-Daejeon	HSR Seoul-Daejeon St.
	Conventional Seoul-Daejeon St., Yongsan-Daejeon St., Yeongdeungpo-Daejeon St., Seoul-Seodaejeon St., Yongsan-Seodaejeon St., Yeongdeungpo-Seodaejeon St.
Seoul-Cheonan	HSR Seoul-Cheonan/Asan St.
	Conventional Seoul-Cheonan St., Yongsan-Cheonan St., Yeongdeungpo-Cheonan St.

## 연구방법

철도 승객 수의 변화에 대한 연구는 특정한 사건에 대한 개입의 영향성 여부 및 그 영향성의 정도를 분석하거나 미래 수요 예측 등을 하고자 할 때 활용된다. 일반적으로 개입사건이 발생한 시점 전후로의 통계량의 변화를 통하여 그 영향성의 여부와 영향의 정도를 도출한다. 앞서 선행연구에서도 KTX 1단계가 개통된 이후에 각각 당해 연도의 3개월 및 1년 등으로 기간을 정하고 직전년도와 동일한 기간 동안의 철도 승객 및 기타 운송수단 승객수의 변화를 활용하여 영향성을 분석하였다(Suh and Yang, 2005; Lee and Chang, 2006). 개입 사건의 전후 통계량 비교 이외에도 철도 승객 수의 시계열 자료를 통하여 개입 ARIMA 분석을 수행할 수 있다. 이는 개입사건에 대하여 시계열에 미치는 영향 및 영향정도를 알 수 있으며, 분석 기간 동안의 발생한 개입사건에 대한 영향성을 모형에 포함하여 미래 수요를 예측하는 방법이다(Kim and Kim, 2011). 개입 ARIMA 분석은 시계열 모형의 단점인 수요의 구조적 측정과 영향력 측정의 부재라는 문제점을 해결하기 위해 각 개입사건별 영향력 계수를 제시함으로써 시계열 모형이 가지고 있는 단점을 보완했다(Kim et al., 2012). 개입 ARIMA 분석은 개입이 발생했을 때, 개입 효과를 더미(Dummy) 시계열로 나타내어 분석하여 지시(Pulse)개입과 계단(Step)개입으로 나뉜다. 사건이 T 시점에서 발생하여 그 효과가 T 시점에만 영향을 미치는 경우가 지시개입이며, T 시점에서 발생하여 그 효과가 발생시점 이후로 지속적인 영향을 미치는 경우가 계단개입이다. 그러나, 개입 ARIMA의 분석 방법은 개입이 일어난 시점과 그 영향의 형태가 지시개입인지 계단개입인지에 대한 가정을 해야 하는 단점이 있다.

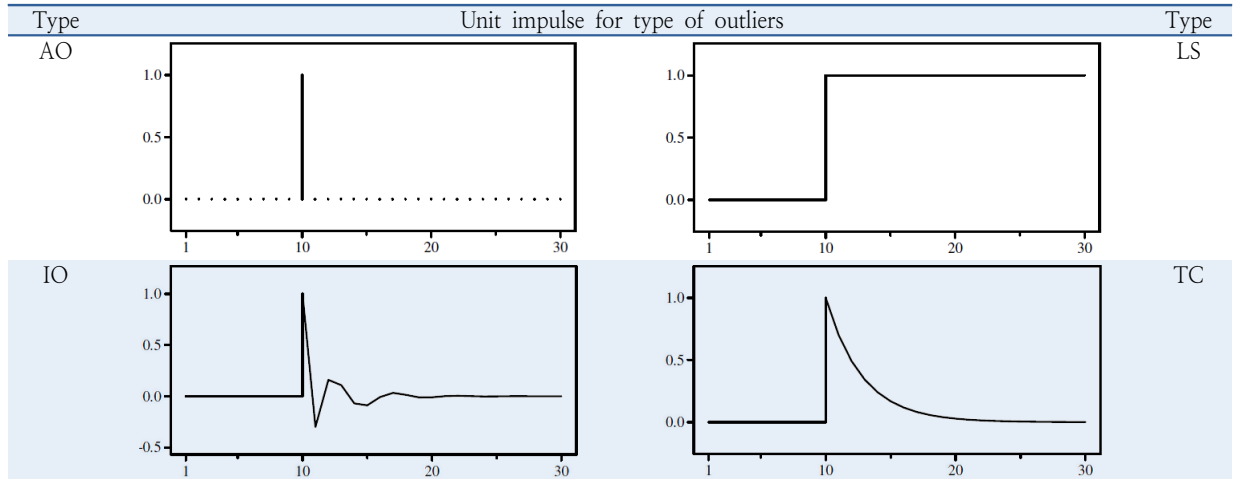
본 연구에서는 개입 ARIMA 분석이 가지고 있는 단점을 보완 할 수 있는 시계열 이상치 탐지방법을 활용하여 분석하였다. 본 분석 방법은 시계열에 영향을 미치는 사건의 발생 시점과 발생효과를 알 수 없는 상황에서 개입 사건의 시점과 영향성을 도출하고자 할 때 적합하다. 본 연구에서는 오픈소프트웨어인 R을 활용하여 시계열 이상치 탐지를 수행하였다(López, 2014, 2016; Chen and Liu, 1993). 분석의 결과로 나오는 이상치 형태는 Innovational Outlier (IO), Additive outlier (AO), Level Shift (LS), Temporary Change (TC) 등의 형태로 분류되며 다음과 같은 특징을 가진다.

- Additive Outlier (AO): T 시점에서만 이상치 효과가 더해진 형태이다. AO는 T 시점에서만 영향을 미치는 이상치로써 시계열을 관측하는 과정에서 발생하는 오차 또는 자료를 옮겨 적는 과정에서 발생하는 오차 등을 설명할 수 있다.
- Innovational outlier (IO): 한 시점의 관측 값에만 영향을 미치는 AO와는 다르게 이상치가 발생한 시점 이후에 지속적인 영향을 미치는 이상치를 말한다. 이상점 효과가 ARMA 구조를 통하여 T 시점 이후의 모든 관측값에 영향을 미치는 이상점이다.
- Level Shift (LS): T 시점 이후에 항구적 수준의 변화를 오게 하는 이상치이며, 한 시점에만 영향을 주는 AO나 T 시점 이후에 영향을 주나 점차적으로 줄어드는 IO와 TC와는 상이하다.
- Temporary Change (TC): 이상점에 의해 관측 값의 수준에 생긴 변화가 단기간 지속되고 변화의 크기도 점차적으로 줄어드는 경우의 이상치를 말한다. IO의 경우와 유사한 형태를 가지나 TC는 감쇠효과라 부르는  $\delta$ 에 따라 지수적으로 감소하며 지속되는 시간과 영향의 크기가  $\delta$ 의 크기에 의해 좌우된다( $\delta$ 는 0.7로 설정).

시계열 이상치 탐지방법은 분석할 데이터를 통하여 ARIMA 모델이 주어지면 가능한 모든 시간 지점에서 모든 유형별로 이상점 효과를 감지하고 통계량을 구하며, 이상치 감지는 임계값(critical value)보다 높은 통계치를 나타낸 시간 지점과 유형을 탐색한다. Chang and Tiao(1983)와 Kaiser and Maravall(1999)은 표본크기(sample size)에 따라서 3.0, 3.5, 4.0의 임계값을 사용하는 방법을 제안하였으며, Lopez(2016)은 표본크기가 50보다 작은 경우에는 3.0을 사용하고 표본의 크기가 450보다 큰 경우에는 4.0을 사용하였다. 본 연구에서 활용된 표본 크기는 11년 동안의 월별 데이터 144건이므로 임계값을 3.5로 사용하였다. 동일한 시점에 두 가지 유형의 이상치가 있을 경우에는 통

계 값이 큰 유형의 이상치를 유지하며 그 외의 이상치 유형은 제거된다. 주요 도시 4개 구간의 철도 승객 수 데이터를 시계열 이상치 탐지방법을 활용하여 이상치를 탐지하여 주요 도시 구간별로 비교하고, 이상치를 발생시킨 개입 사건을 조사·분석하였다.

**Table 3.** Type of outliers (López, 2016)

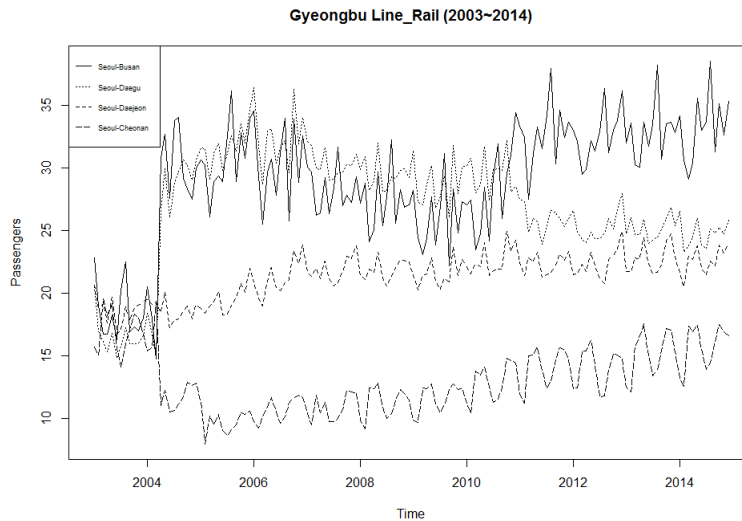


## 분석결과

### 1. 기초통계 분석 결과

#### 1) 구간별 이용객 수

Figure 1은 경부선의 주요 도시 4개 구간을 이용한 철도 승객수를 시계열 그래프로 나타낸 것이다. 서울-부산과 서울-대구 구간의 경우에는 2004년 4월 기준으로 승객수가 급격히 증가하였으나 서울-대전 구간은 KTX가 등장한 4월에도 큰 변화가 없었다. 서울-천안 구간의 경우에는 다른 구간과 다르게 승객 수가 줄어든 것으로 나타났다. 2010년 이후에 서울-부산 구간에서는 또 한 번의 증가가 보였으나, 서울-대구 구간은 감소의 형태를 보였다.



**Figure 1.** Number of railway passengers for the Gyeongbu line (2003-2014)

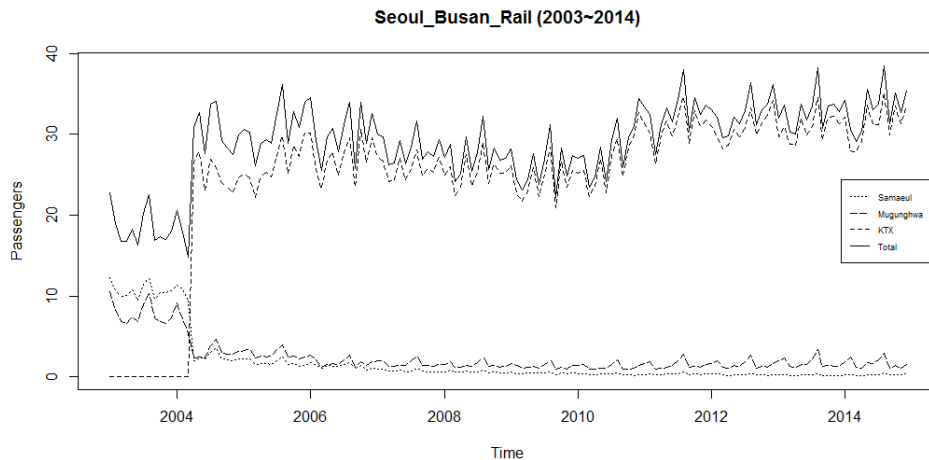
Table 4는 구간별 직전년도 대비 연간 철도 승객수의 증감률을 보여주고 있다. 2004년의 경우 서울-대전 구간을 제외하고는 승객수의 변화가 크게 증가하거나 감소한 것으로 나타났다. 특히나 서울-부산 구간 및 서울-대구 구간의 경우는 2003년 대비 2004년도의 이용객수가 각각 약 47%, 57%가 증가하였으며, 서울-천안 구간은 2004년과 2005년 모두 직전연도 대비 25% 감소 형태를 보였다. 하지만 2006년부터는 2012년을 제외하고는 서울-천안 구간의 승객수가 증가하는 것으로 나타났다. 서울-대구 노선의 경우 2011년도 승객수가 2010년도 대비 12.5% 감소한 것으로 나타났다.

**Table 4.** Number of railway passengers per year compared to previous year

Year	Seoul-Busan	Seoul-Daegu	Seoul-Daejeon	Seoul-Cheonan
2004	147.47%	157.28%	100.13%	76.64%
2005	112.92%	122.03%	104.35%	74.91%
2006	98.52%	102.23%	108.98%	110.00%
2007	93.25%	93.72%	103.12%	101.13%
2008	96.65%	97.02%	99.87%	103.92%
2009	95.26%	97.40%	98.66%	103.28%
2010	108.04%	103.15%	105.01%	110.84%
2011	116.28%	87.56%	98.69%	109.17%
2012	99.72%	97.73%	100.14%	99.79%
2013	100.58%	99.37%	101.65%	107.78%
2014	101.48%	98.42%	99.13%	103.77%

## 2) 철도 종류(Type)별 철도 이용객수

Figure 2는 구간별 철도 종류에 따른 승객 수의 변화를 시계열 그래프로 나타냈다. 철도 종류는 KTX, 새마을호, 무궁화호, 누리호(서울-천안 구간만 해당) 등으로 분류되며, 전체 철도 승객 수(구간별 모든 철도 종류의 승객 수의 합)도 포함 하였다. KTX가 등장한 이후에는 서울-천안 구간을 제외한 3개의 주요 도시 구간에서 KTX가 차지하는 승객 수의 비중이 높게 나타났다. 특히나 서울-부산 및 서울-대구 구간의 경우에는 KTX의 비중이 압도적으로 높게 나타났으며, 새마을호와 무궁화호는 KTX의 등장 이후 승객수가 급격히 줄어들었다. 그러나, 서울-천안의 구간은 KTX가 등장을 한 이후에도 무궁화호 승객수의 비중이 KTX의 승객 수보다 많았으며, 2009년 이후에는 KTX의 승객 수가 무궁화호 이용객 수보다 많아졌다.



**Figure 2.** Number of railway passengers by train type

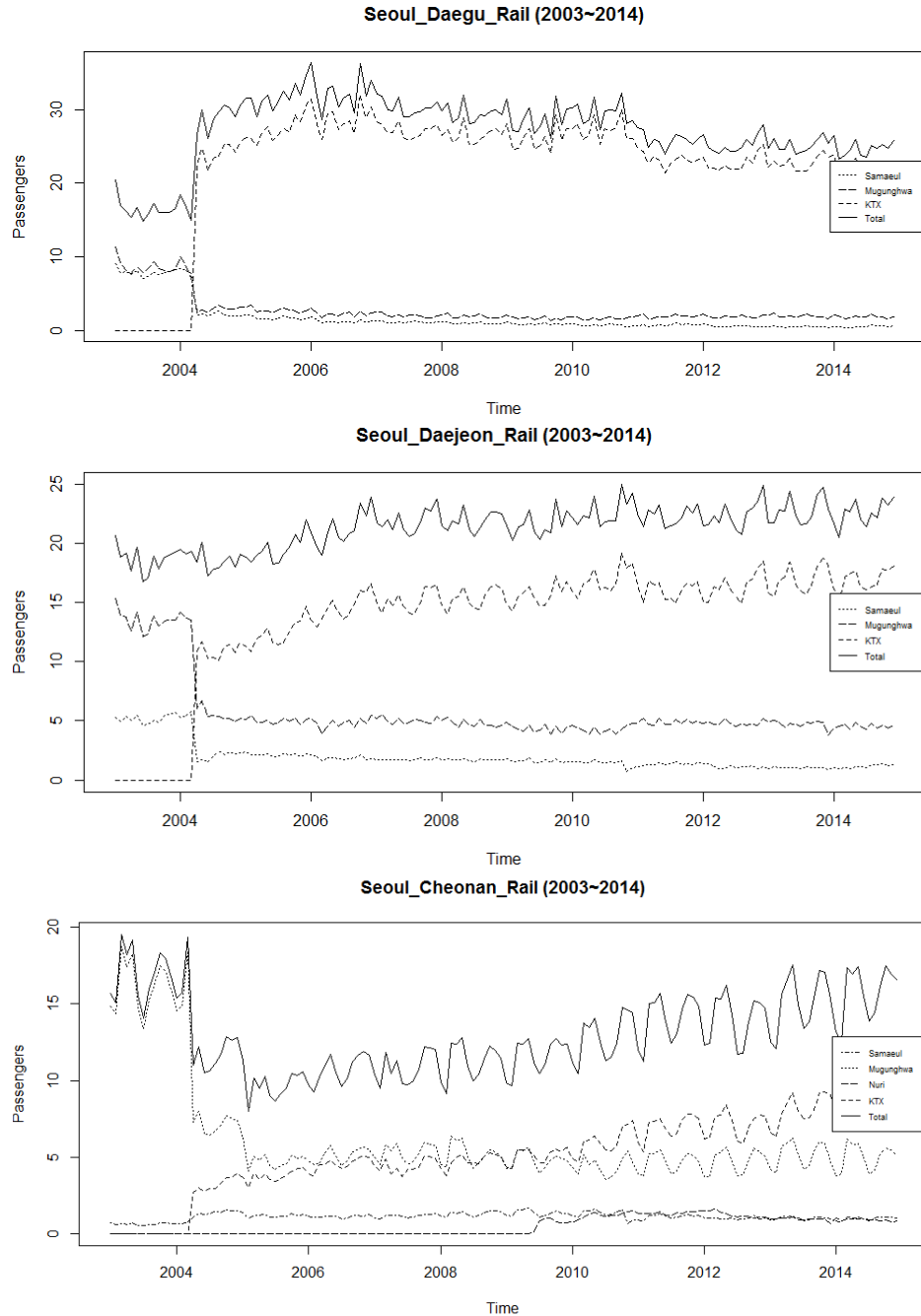


Figure 2. Number of railway passengers by train type (continued)

## 2. Time series outlier detection 분석결과

시계열 이상치 탐지는 개입에 의해 시계열의 변화가 있을 경우 발생되어 진다. 2003년부터 2014년까지 12년 동안 고속철도 도입, 대구-부산 구간 고속도로 개통, 서울-대구 구간 항공노선 폐지, 서울-부산 구간 LCC 등장, 2008년 경제 위기, 철도 파업(2006, 2013) 및 KTX 산천 도입 등의 개입 사건들이 있었다. 관련 개입사건들은 직간접적으로 경부선의 철도를 이용하는 승객에게 영향을 끼친다(Lim et al., 2008; Lee et al., 2013; Park and Kim, 2016). 경부선의 주요 도시구간의 승객 수 변화에 대하여 탐지된 이상치 형태, 이상치 시점, 이상치 영향 정도의 분석 결과

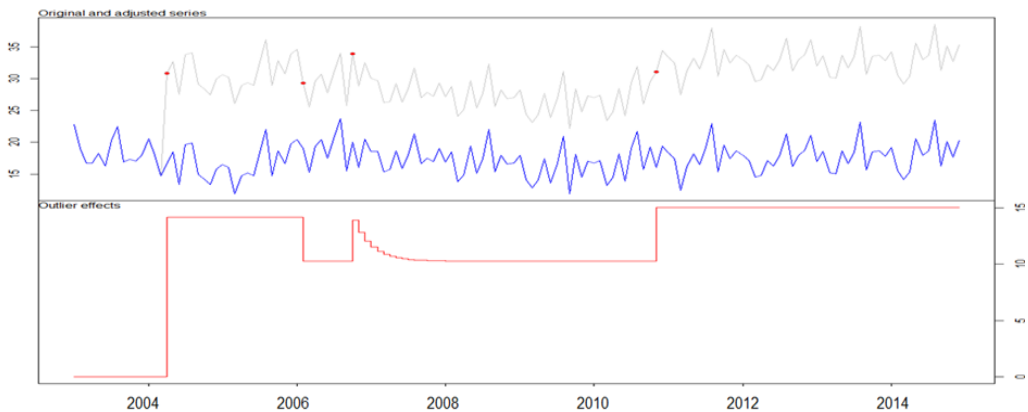
를 Table 5로 정리하였다. 계수(Coefficients) 값을 나타내는 단위는 가공한 데이터 단위로서, 본 연구에서는 승객수의 단위를 10,000명으로 설정하였다.

**Table 5. Outliers and intervention events**

OD	Type	Time	Coefficients	Tstat	Remark (intervention event)
Seoul-Busan	LS	2004.04	14.170	16.595	KTX 1st phase completed
	LS	2006.02	-3.906	-4.513	New expressway (Daegu-Busan)
	TC	2006.10	3.651	4.166	Long holiday
	LS	2010.11	4.768	6.150	KTX 2nd phased completed
Seoul-Daegu	LS	2004.04	12.044	16.289	KTX 1st phase completed
	LS	2006.01	3.814	4.097	-
	TC	2006.10	3.304	3.861	Long holiday
	LS	2010.11	-3.211	-4.560	KTX 2nd phased completed
Seoul-Daejeon	AO	2006.03	-1.923	-4.107	Strike (2006.3.1.-4)
	TC	2013.12	-1.950	-3.907	Strike (2013.12.9.-31)
Seoul-Cheonan	LS	2004.04	-6.875	-11.92	KTX 1st phase completed

Figure 3은 서울-부산 구간의 시계열 이상치 탐지 그래프이며, 상단에 나타난 실선과 점과 같이 나타난 회색 그래프(실제 승객 수)와 실선으로만 나타난 파란색 그래프(적용된 ARIMA 모형)의 차이(difference)를 하단의 빨간색 그래프(outlier)로 이상치 형태와 정도를 나타내고 있다.

서울-부산 구간은 2004년 4월(LS), 2006년 2월(LS), 2006년 10월(TC), 2010년 11월(LS)까지 총 4번의 이상치가 나타났다(Figure 3). 3번의 LS 중에서 2004년과 2010년 발생한 LS는 KTX의 1단계 및 2단계 개통 완료에 따라 줄어든 철도 이용시간 덕분에 승객수가 각각 14만 명과 4만 7천명이 증가하는 LS로 나타났다. 이와 반대로, 2006년 나타난 LS는 신대구 고속도로 개통에 따른 영향으로 약 3만 9천명의 승객 수가 줄어든 LS로 나타났다. 대구-부산의 육상 구간을 이어주는 신대구부산 고속도로가 2006년 2월 12일부터 운영이 시작되어 육상교통의 이동시간이 단축되었다. 그 결과, 서울-부산 구간의 고속버스 시간은 경부터미널 기준 50분, 동서울 터미널 기준 20분이 감소되어 고속버스 이용객이 증가하여 직간접적으로 경부선 철도 승객에게 영향을 끼친 것으로 예상된다.



**Figure 3. Seoul-Busan outliers**

2006년 10월에 나타난 TC는 해당 월에 포함하고 있는 긴 연휴기간에 따른 철도 이용객 수의 증가로 분석된다. 우리나라의 법정 공휴일 중에서 설날과 추석의 연휴가 가장 길다. 2003년부터 2012년까지의 추석과 설날의 연휴기간을 살펴본 결과, 2006년 10월의 경우 개천절(10월 3일)과 추석 연휴(10월 5-7일)가 징검다리 연휴로 나타났다. 많은



기업들이 개천절과 추석 연휴의 사이의 10월 4일도 같이 쉬면서 긴 연휴기간이 이어졌으며 여행 혹은 가족 친지를 만나기 위해 철도 승객이 증가했을 것으로 예상된다. 분석결과로 도출된 증가된 인원은 3만 6천명으로 나타났다.

서울-대구 구간의 경우에는 2006년에 나타난 이상치의 시점만 차이가 있을 뿐 서울-부산 구간에서 나타난 다른 모든 이상치의 시점과는 동일하나 그 영향의 형태와 정도가 다르게 나타났다(Figure 4).

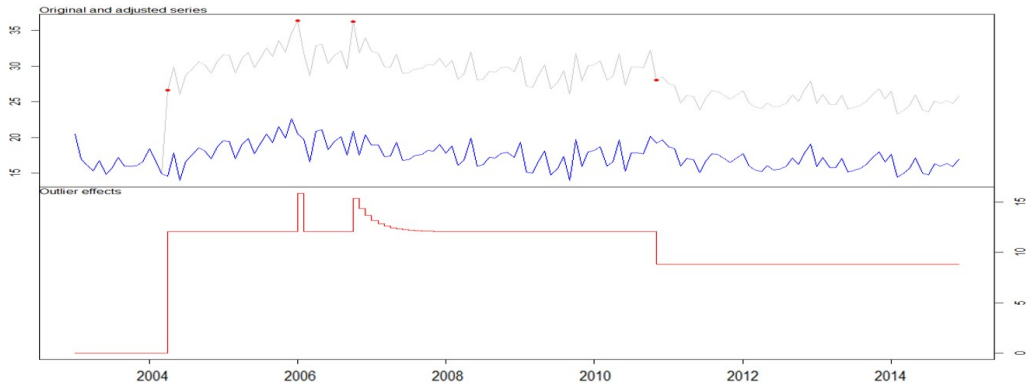


Figure 4. Seoul-Daegu outliers

먼저 2004년 KTX 1단계 개통에 따른 영향으로 12만 명이 증가하는 LS가 나타나고, 서울-부산 구간과는 다르게 2010년 KTX 2단계 개통으로는 3만 명이 감소하는 LS가 나타났다. 고속철도 2단계 공사가 완료되기 이전의 대구역은 서울-대구 구간까지 KTX를 이용하고 기존의 무궁화와 새마을호로 갈아타는 주요 환승역으로서의 기능을 수행하였으나, 2단계 공사가 완료된 이후에는 대구역에서 일반열차 환승 없이 KTX 이용만으로도 대구-부산 구간의 정차역까지 운행이 가능함에 따라 서울-대구 구간을 이용한 승객이 감소된 것으로 분석된다. Table 6은 KTX2단계 개통이 완료되기 이전에 기존 열차로 환승해야 하는 경주와 울산 지역에 대하여 2단계 공사가 완료되기 전 3개월과 이후 3개월 동안의 철도를 이용한 승객 수이다. 개통이 시작된 2010년 11월 기준으로 기존의 열차 승객 수는 줄어들거나 노선이 사라졌으며, 대구가 아닌 서울부터 이용한 KTX의 승객이 급격히 증가하였음을 알 수 있다.

Table 6. Number of passengers (Seoul-Gyeongju, Daegu-Gyeongju, Seoul-Ulsan, Daegu-Ulsan)

OD	Type	2010.08	2010.09	2010.10	2010.11	2011.12	2011.01
Seoul-Gyeongju	Conventional Rail	5,055	2,914	4,847	474	38	86
	KTX	-	-	-	45,168	48,584	43,311
Daegu-Gyeongju	Conventional Rail	7,012	6,040	6,782	129	194	170
	KTX	-	-	-	3,784	4,159	3,585
Seoul-Ulsan	Conventional Rail	3,816	3,041	2,283	-	-	-
	KTX	-	-	-	69,835	77,608	81,932
Daegu-Ulsan	Conventional Rail	6,010	5,901	5,813	-	-	-
	KTX	-	-	-	14,210	17,407	18,028

2006년 10월에 발생한 TC는 서울-부산 구간에서 나타난 이상치와 같이 징검다리 연휴로 인한 긴 휴일 기간에 따른 영향으로 분석되며, 그 영향으로 3만 3천명의 철도 승객이 증가하고 점차 감소할 것이라고 나타났다.

서울-대전 구간에는 2번의 이상치가 나타났으며, 2006년 3월에 AO, 2014년 12월에 TC이다(Figure 5). 두 개의 이상치 탐지가 발생한 이유로는 철도 파업으로 예상된다. 2006년 3월에는 4일(3/1-4), 2013년에는 23일(12/9-31) 동안 철도 조합의 파업이 발생했다. 2003년에 4일간 진행된 파업이지만 조합의 파업 참가율이 66%로 높았으며, 고속철도와 여객철도의 운행률이 각각 계획 운행대비 34%와 16%로 감축 운행되었다. 2013년에는 진행된 23일간 파

업 참가율은 30%이었으며, 고속열차와 일반여객 열차는 73%와 61%로 감축 되었다(국토교통부 보도자료). 서울-대전 구간은 서울-부산과 서울-대구 구간에서 탐지되었던 2004년 KTX 개통, 2006년 징검다리 휴일 및 KTX 2단계 개통시점에서 이상치가 나타나지 않았다. 이는 위의 3개의 사건이 서울-대전 구간에서 이상치가 나타날 만큼 시계열의 변화를 주지 않았음을 의미하며, 철도 승객수의 변화가 크지 않음을 시사한다.

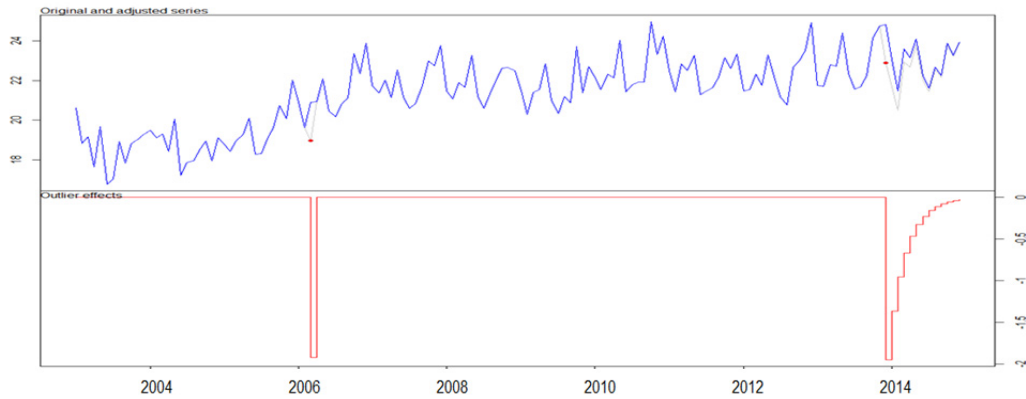


Figure 5. Seoul-Daejeon outliers

서울-천안 구간의 경우, 2004년 4월에 KTX가 등장한 시점에서 서울-부산과 서울-대구 구간에서 발생한 증가된 LS 형태와는 다르게 감소된 LS 형태로 나타났다(Figure 6). 천안에서 정차하는 KTX역은 기존의 새마을호와 무궁화호가 정차하는 천안역이 천안아산역이 신규 건설 되었다. 그 위치가 도심에서 멀리 떨어져 있어 접근성이 좋지 않았으며, 서울로부터의 거리가 102km 밖에 되지 않아 KTX의 탑승으로 인한 감소시간이 크지 않았다. 그러나, KTX의 이용 가격이 기존의 무궁화호와 대비하여 약 2배 이상 비쌌기 때문에 감소한 기존 열차 승객만큼 KTX의 승객이 증가하지 않아 전체 승객의 감소가 나타난 것으로 분석되며 그 영향으로 감소된 인원은 약 6만 8천명으로 나타났다.

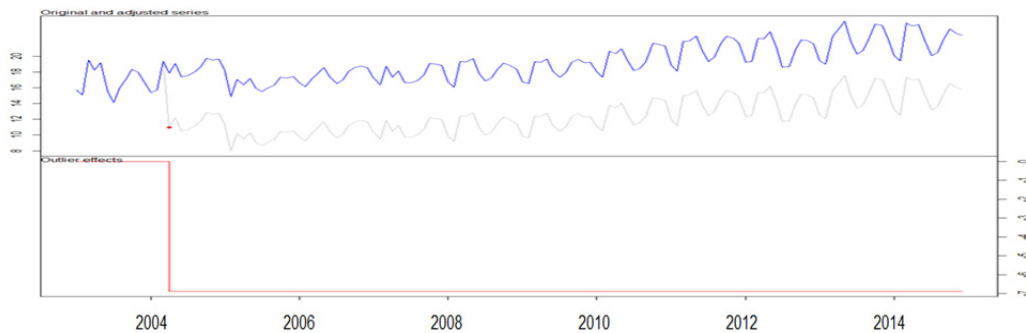


Figure 6. Seoul-Cheonan outliers

## 결론 및 한계점

기존 연구에서 수행된 개입사건에 대한 분석은, 발생한 시점 전후의 통계량 혹은 개입이 발생한 시점과 그 영향의 형태를 가정하고 개입사건에 대한 영향 분석을 수행하였다. 하지만 본 연구에서는 개입되는 외부요인의 영향을 시계열 이상치 탐지 방법으로 수행되었기 때문에, 개입이 일어나는 시점과 형태를 모르는 개입사건에 대하여 개입시점과 영향 형태에 대한 가정 없이 분석할 수 있었다. 특히, 국내 교통의 중추노선인 경부선을 중심으로 KTX의 도입

이전인 2003년부터 2014년도 까지 최근 12년간 주요 4개 도시 구간의 철도 승객수를 활용하여 분석하였다. 그 결과, 동일한 개입사건 임에도 그 영향의 형태와 정도가 도시 구간마다 다르게 나타나거나 영향이 나타나지 않았다. KTX가 처음 도입 되었던 2004년 4월의 경우, 서울-부산과 서울-대구 구간에서는 승객수가 증가한 반면에, 서울-부산 구간의 중간 지점인 대전은 승객 수의 변화가 없었고 서울-천안 구간은 승객수가 감소하는 결과로 나타났다. 대구-부산 구간의 신대구고속도로 개통은, 철도 승객에게 직접적이지는 않지만 육로 이동시간의 감축에 따른 간접적인 요인으로 서울-부산 구간의 철도 승객 감소라는 영향을 끼쳤다. 마지막으로, 분석된 이상치의 결과와 동향조사를 바탕으로 2006년 10월의 영향은 긴 연휴기간에 따른 개입이며, 서울-대전 구간에서 나타난 2개의 이상치는 철도 파업의 영향으로 이상치가 탐지 된 것으로 분석 된다.

그러나, 본 연구에서 수행된 시계열 이상치 탐지방법으로 밝혀낸 서울-대구 구간의 2006년 1월에 이상치 탐지 결과는 수집된 데이터와 동향조사를 수행했음에도 불구하고 개입의 영향을 찾기에는 한계가 있었다. 향후, 철도 운행 빈도, 가격 등의 정보 및 도시마다의 교통정책 등을 반영하여 철도 승객수를 변화시킬 수 있는 개입사건 및 요인에 대한 추가적인 정보를 통하여 개입의 영향성을 파악할 수 있을 것이며 별도의 추가 연구로 진행해야 할 것이다. 마지막으로, 수서 발 고속철도가 개통함(2016.12.09)에 따라 동일한 구간 내 두 개의 HSR이 운영되었을 때의 전후 통계량과 설문 결과를 분석한 선행 사례연구(Cascetta and Coppola, 2014)와 같이 국내 두 번째 HSR 운영자 등장과 같은 개입사건의 수요 영향 연구가 수행되어야 할 것이다.

## ACKNOWLEDGEMENT

This work was supported by the Brain Korea 21 Plus Project (Center for Creative SOC Infrastructure System Technology, 21A20132000003) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning.

**알림:** 본 논문은 제29회 KKHTCNN Symposium on Civil Engineering(2016.12.04)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

## REFERENCES

- Box G. E. P., Tiao G. C. (1975), Intervention Analysis with Applications to Economic and Environmental Problems, *Journal of the American Statistical Associations*, 70(349), 70-77.
- Cascetta E., Coppola P. (2014), Competition on Fast Track: An Analysis of the First Competitive Market for HSR Services, *Procedia-Social and Behavioral Sciences*, 111, 176-185.
- Chang I., Tiao G. C., Chen C. (1988), Estimation of Time Series Parameters in the Presence of Outliers, *Technometrics*, 30(2), 193-204.
- Chen C., Liu L. M. (1993), Joint Estimation of Model Parameters and Outlier Effects in Time Series, *Journal of the American Statistical Association*, 88(421), 284-297.
- Kaiser R., Maravall A. (1999), Seasonal Outliers in Time Series.
- Kim H. W., Lee D. H., Park H. S. (2013), The Impact of Gyeongbu High Speed Rail Construction on Regional Economic Growth, *KSCE Journal of Civil Engineering*, 17(6), 1206-1212.
- Kim K. H., Kim H. S. (2011), KTX Passenger Demand Forecast With Intervention ARIMA Model, *Journal of the Korean Society for Railway*, 14(5), 470-476.

- Kim M. S., Kim K. W., Sung S. P. (2012), A Study on the Air Travel Demand Forecasting Using Time Series ARIMA-Intervention Model, *Journal of the Korean Society for Aviation and Aeronautics*, 20(1), 63-74.
- Lee J. H., Chang J. (2006), Effects of High-speed Rail Service on Shares of Intercity Passenger Ridership in South Korea, *Transportation Research Record: Journal of the Transportation Research Board*, 1943, 31-42.
- Lee S.J., Change J.S., Lee S. J., Yoon Y. W. (2013), An Analysis on Speed-Distance Comparative Advantage of Transportation Modes After the Opening of Gyeongbu High-speed Railway Project Phase II, *Conference of the Korean Society for Railway*, 1162-1174.
- Lim S. J., Lim K. W., Lee Y. I., Kim K. H. (2008), A Study on Air and High Speed Rail Modal According to the Introduction of Low Cost Carrier Air Service, *J. Korean Soc. Transp.*, 26(4), Korean Society of Transportation, 51-61.
- López J., López M. J. (2016), Package 'tsoutliers'.
- López-de-Lacalle (2014), J. tsoutliers R Package for Detection of Outliers in Time Series.
- Park M. S., Eom J. K., Heo T. Y., Song J. (2015), Intervention Analysis of the Impact of Opening a New Railway Line on Passenger Ridership in Seoul, *KSCE Journal of Civil Engineering*, 1-11.
- Park M. S., Kim Y. (2016), The Impacts of High Speed Train on the Regional Economy of Korea, *Korean Journal of Applied Statistics*, 29(1), 13-25.
- Park S. K. (2004), The Impact of Terrorism on World Tourism, *Journal of Tourism Sciences*, 28(2), 77-94.
- Suh S., Yang K. Y. (2005), Customers' Reactions to the Introduction of High-speed Rail Service: Korean Train express, *Transportation Research Record: Journal of the Transportation Research Board*, 1916, 20-25.