계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구

A Study on Dynamic Change of Transportation Demand Using Seasonal ARIMA Model

이 재 민
(경북대학교 경제통상학부 전임강사)

권 용 재
(국민대학교 경영학부 전임강사)

목 차

Key Words : ARIMA, 계절성, 계절성을 감안한 ARIMA, 추정, 예측
ARIMA, Seasonality, Seasonal ARIMA, Estimation, Forecast

요 약

본 연구에서는 계절성(seasonality)을 감안한 적분된 자기회귀 이동평균 모형(ARIMA model)을 이용하여 우리나라 지역 간 철도의 동태적 변화과정을 추정하고 장래 통행수요를 예측하고자 하였다. 기존 국내연구에서 고려하지 않은 계절성 요인을 감안한 ARIMA 모형(Seasonal ARIMA model)과 월별 지역 간 철도 통행실적자료를 이용하여 교통수요 동태적 변화모형을 구축하였다. 구체적으로 2000년 1월부터 2008년 12월까지의 월별 수송인원 및 수송인-km 기준 지역 간 통행실적 자료를 이용하여 Box et al. (1994)에서 제시한 Seasonal ARIMA 모형을 적용하였으며 이에 따라 장래 지역 간 철도 통행수요를 예측하였다. 장래 통행수요 예측 결과에 따르면 수송인원 기준으로 2015년 및 2020년에는 2008년의 각각 약 1.36배와 1.71배 수준으로 산정되었다. 또한 수송인-km 기준으로 2015년과 2020년에는 2008년의 각각 약 1.25배와 1.78배 정도로 예측되었다.

This study is to estimate the dynamic change of the regional railway passenger traffic and, based on the estimated, to forecast the future regional railway passenger traffic by using the Seasonal ARIMA model. The existing studies using ARIMA failed to consider seasonality nor the monthly or the quarterly data. It was attempted in this study to use the monthly regional railway passenger traffic data to propose a model that estimates dynamic change of demand. The authors employed the Seasonal ARIMA model previously developed and used (1) the numbers of monthly passenger data and (2) the monthly passenger-km data. The test results showed that the numbers of passengers in 2015 and 2020 would increase by 36% and 71%, respectively, compared to those in 2008. The numbers of passenger-kms in 2015 and 2020 would increase by 25% and 78%, respectively, compared to those in 2008.

이 논문은 한국건설교통기술평가원에서 위탁시행한 교통체계효율화사업(09교통체계-지능06)의 지원에 의해 수행되었습니다.

Ⅰ. 서론

1. 연구의 배경 및 목적

최근 교통수요의 장래수요 예측은 매우 중요한 이슈로서 제기되고 있다. 고속도로, 도시철도 및 지역 간 철도의 장래 교통수요를 예측하고 이를 기반으로 통행요금 체계를 구축하는 것은 모든 교통산업에서 중요한 과제라 할 수 있다. 또한 다양한 연구에서 교통수요를 예측하는 모형을 구축하고 이에 따라 실제 교통수요를 예측하고 있다.

시계열 분석을 이용한 교통수요의 동태적 변화모형을 구축하는 것은 교통정책 수립에 큰 영향을 미치는 중요한 연구과제로 볼 수 있다. 예를 들어 고속도로 통행실적이 장래에 어떠한 형태를 띨 것인가를 분석하여 장래 교통수요를 예측할 수 있다면 한국도로공사의 장래 통행수입이 산정될 수 있으며 이를 기반으로 중앙정부 재정보조액, 통행요금 수립에 중요한 자료가 될 것이다.

그러나 우리나라 과거 교통수요 자료의 한계로 인하여 기존 시계열 분석을 이용한 교통수요 장래예측 모형은 큰 문제점이 존재한다[1). 첫째 우리나라의 교통수요 자료는 시계열이 매우 짧은 문제점을 가지고 있다. 대부분의 교통수요 변화 분석모형은 계량경제학적인 시계열 분석(time series analysis) 기법을 이용하고 있다. 이때 분석에 이용되는 시계열자료가 우리나라에서는 대부분 30 여년 전후의 자료가 대부분이며, 자료의 개수 문제가 시계열분석에서 항상 제기될 수 있는 부분이다.

둘째 30개 이상의 연도별 자료를 이용할 때 지나친 과거 자료 이용으로 장래 교통수요 예측에 오차를 유발할 수 있다는 점이다. 과거 1990년 중반까지는 대부분의 교통수단 실적들이 큰 폭으로 증가하고 있었다. 하지만 2000년대 이후 이러한 교통실적의 증가가 감소하고 있는 실정이며 이러한 경향은 향후 인구감소와 더불어 지속될 것으로 보인다.

본 연구에서는 이러한 우리나라 교통수요 자료상의 문제점을 보완하기 위하여 교통실적 월별자료와 계절성을 감안한 적분된 자기회귀 이동평균(Seasonal Autoregressive Integrated Moving Average, SARIMA) 모형을 이용하여 교통수요 동태적 변화모형을 구축하였다. 즉 2000년대 이후의 월별 교통실적 자료를 이용하여 과거 연구의 시계열 자료 개수 부족문제를 해결할 것이며 최근 자료를 이용하여 과거 자료 이용에 따른 왜곡을 해결하고자 하였다.

그리고 월별 자료 이용에 따른 계절성(seasonality) 문제를 해결하기 위하여 계절조정된 ARIMA 모형(Seasonal ARIMA, 이하 SARIMA)을 이용하였다. 계절성이란 어떤 시계열변수(time series variables)의 값이 주기적(periodically)인 특성을 지니는 경우를 의미한다. 예를 들어 항공수요가 여름휴가 및 성탄절 시즌에 피크를 이루며 국내총생산이 4/4분기에 가장 큰 값을 나타내는 경우를 의미한다. 따라서 월별 교통실적 자료를 이용할 때 계절성을 감안하지 않은 일반적인 ARIMA 모형을 이용하면 분석의 왜곡이 발생할 수 있다. 그러므로 계절성을 감안한 ARIMA 모형 즉 SARIMA 모형을 적용하여 교통수요 동태적 변화를 분석하여야 한다.

본 연구에서는 월별 교통실적 자료와 계절성을 감안한 ARIMA(SARIMA) 모형을 적용하여 교통수요의 동태적 변화를 추정하였다. 이를 기반으로 장래 교통수요 예측방법론을 구축하고자 하였다.

2. 연구의 범위 및 방법

본 연구에서는 철도여객부문의 월별 시계열 자료를 이용하였다. 구체적으로 2000년 1월부터 2008년 12월까지의 월별 지역 간 철도의 여객수송실적자료를 이용하여 자료의 개수문제와 지나친 과거 자료 이용에 따른 왜곡문제를 해결하고자 하였다.

이외에도 다양한 교통실적 자료(예를 들어 고속도로 및 수도권 전철 통행실적 등)에 Seasonal ARIMA 모형을 적용할 수 있다. 그러나 본 연구에서는 일단 지역 간 철도를 이용하여 분석하였다.

앞에서 지적한 것처럼 본 연구에서 이용되는 교통실적자료로는 지역 간 철도의 통행실적자료로서 월별 인기준 및 인-km기준 자료이다. 철도통계연보(2008)에 의하면 KTX, 새마을, 무궁화, 통근, 수도권 전철 등의 사업영역에서 수도권 전철을 제외한 부분을 지역 간 철도로 구분할 수 있다.

1) 교통수요 장래 예측은 시계열 분석 기법 외에도 O/D 자료를 이용한 4단계 분석기법이 널리 이용된다. O/D 자료를 이용한 4단계 분석 방법론은 장래 경제상황 변동 등을 감안할 수 있다는 점에서 시계열 분석 기법이 지니지 못한 장점이 있다. 본 연구에서는 기존 교통수요 분석방법 중에서 4단계 추정방법이 아니라 시계열 분석 방법에 연구의 초점을 맞출 것이다.

분석방법론으로는 시계열분석 기법을 이용하였다. 앞에서 지적한 것처럼 Seasonal Autoregressive Integrated Moving Average (SARIMA) 모형을 이용하였다. SARIMA 모형은 계절성(seasonality)을 감안한 ARIMA 모형으로 경제학 및 통계학에서 이용되고 있는 방법론이다. 특히 Box et al.(1994)에서 SARIMA 모형을 이용하여 항공수요를 분석하고 장래 수요를 예측하는 기법을 제시하고 있다.

기존 시계열 분석을 이용한 교통수요 연구들은 연도별 자료와 ARIMA 모형을 이용하거나 계절성을 감안하지 않은 월별 자료와 ARIMA 모형을 이용하였다. 하지만 본 연구는 계절성을 감안한 ARIMA(SARIMA) 모형을 적용하여 교통수요 분석에 널리 이용될 수 있는 사례를 제시하고자 하였다.

3. 기존 연구 고찰

앞에서 지적한 것처럼 월별 자료를 이용한 SARIMA 모형 적용의 대표적 문헌으로는 Box et al.(1994)을 들 수 있다. 여러 경제학 혹은 통계학 연구에서는 SARIMA 모형이 널리 이용되고 있지만 국내 교통부문에서는 실제 적용이 잘 되지 않고 있다.

다만 해외 연구 중 SARIMA 모형을 적용한 연구가 다수 있다. Dent & Swanson(1978)은 미국 철도화차 운송량의 장래 예측에서 SARIMA 모형을 적용하여 분석하였다. Lai & Lu(2005)는 9.11 테러 이후 미국의 항공수요가 어떠한 영향을 받고 있는 지를 SARIMA 모형을 이용하여 분석하였다.

한국교통연구원(2009)의 연구에서는 인천 지역 유료도로의 장래 교통수요를 ARIMA 모형을 이용하여 예측하였다. 이때 월별 자료를 이용하였지만 계절성을 감안한 SARIMA 모형을 적용한 것이 아니어서 한계가 존재한다.

이처럼 외국 문헌에는 월별 자료 이용에 따른 계절성을 감안한 ARIMA 모형을 적용한 사례가 다수 존재하지만 국내 문헌에는 월별 자료를 이용하면서도 이를 감안하지 않고서 분석하였다.

이외에도 장래 수요예측에서 다양한 경제변수가 고속도로 및 철도수송실적에 어떠한 영향을 받고 있는 지를 분석한 후에 여러 경제변수 장래 예측치를 이용하여 장래 교통수요를 예측한 다수의 연구가 존재한다. 이재민·한상용·이창운(2009)의 연구에서는 고속도로 및 서울지역 도시철도 통행실적이 유가 및 소득 등의 변수에 어떠한 영향을 받고 있는 지를 월별 자료와 자기회귀 시차모형(Autoregressive Distributed Lag model, ADL model)을 이용하여 분석하였다. 또한 문진수·이재민(2007)은 철도화물 수송실적이 사회경제변수에 어떠한 영향을 받는 지를 오차수정모형(Error Correction Model, ECM)을 이용하여 분석하였다.

자기회귀 시차모형 및 오차수정모형에서도 설명변수인 사회경제변수의 장래 예측치를 이용하여 교통수요 예측을 수행할 수 있다. 그러나 이러한 과정에서 사회경제변수의 장래 예측치를 이용해야 한다는 점에서 불확실성이 존재한다. 즉 소득, 유류가격 등의 불확실한 미래 예측치를 이용하여 장래 교통수요를 예측한다는 점에서 한계가 존재한다. 한편 자기회귀 시차모형 및 오차수정모형에서는 기간 동안 다른 내재적인 변화를 반영할 수 있으며 미래 계획된 내재적인 변화를 모형 안에 반영할 수 있다는 장점이 존재한다.

이에 반해 SARIMA 모형은 과거 및 현재의 교통수요 실적치를 이용하여 장래 교통수요를 예측한다는 점에서 자기회귀 시차모형 및 오차수정모형과 차별성이 존재한다. 따라서 소득 및 유류가격 변수와 같은 불확실한 미래 변수의 예측치를 적용할 필요가 없다. 그러나 자기회귀 시차모형이나 오차수정모형에서처럼 내재적인 변화(예: KTX 개통)를 모형 안에서 반영할 수 없고 장래 예측에서 계획된 내재적인 변화를 반영할 수 없다는 것이 단점이다.

따라서 SARIMA 모형과 자기회귀 시차모형 및 오차수정모형은 서로 보완이 가능한 모형이라 할 수 있다. 이러한 점에서 향후 장래 교통수요 예측에서 이 두유형의 모형을 모두 적용하여 서로 간의 교통수요를 비교하는 것이 적절할 것이다. 본 연구에서는 우선 국내 교통수요에 거의 적용된 적이 없는 SARIMA 모형을 적용하였다

〈표 1〉 시계열분석 모형의 장단점 비교

SARIMA	추정 및 예측이 용이	다른 설명변수 없이 추정 가능 미래 예측에서 다른 설명변수에 대한 정보가 필요없음
	장래 구조변화 반영이 어려움	장래 계획된 사회경제적 여건 변화를 반영하기 어려움
자기회귀 시차모형 및 오차수정 모형	추정 및 예측이 어려움	다른 설명변수에 대한 현재 정보 및 미래 정보가 필요
	장래 구조변화 반영 가능	장래 계획된 구조변화 반영 가능

는 점에서 의의가 있으며 총계 교통수요 예측 연구의 시금석이 될 것으로 기대한다.

Ⅱ. SARIMA 모형 고찰

1. SARIMA 모형

일반적인 적분된 자기회귀 이동평균(Autoregressive Integrated Moving Average, ARIMA) 모형은 다음과 같이 정의된다.

$$\phi_p(L)(1-L)^d y_t = \theta_q(L)u_t \tag{1}$$

식(1)은 ARIMA(p,d,q)로 정의되며 p는 자기회귀(Autoregressive, AR)항의 차수, q는 이동평균(Moving Average, MA)항의 차수, d는 단위근의 개수를 의미한다. ARIMA 모형은 일반적으로 불안정적인 시계열(non-stationary time series)을 대상으로 하는 시계열 모형으로 단위근을 감안한 모형이다. L은 시차변수(lag operator)이며, y는 교통수요 변수, u는 평균이 0이고 분산이 σ2인 교란항이다. ϕp(L)과 Θq(L)은 자기회귀항(AR term) 및 이동평균항(MA term)의 래그다항식을 의미한다. 대표적인 사례로서 ARIMA(1,1,1) 모형은 p, d, q의 차수가 모두 1인 경우이다. 즉

$$(1-\phi_1 L)(1-L)y_t = (1+\theta_1 L)u_t \tag{2}$$

ϕ_1 및 θ_1은 각각 1계 자기회귀항(AR(1) term)과 이동평균항(MA(1) term)의 계수이다. 이때 안정성(stationarity)을 위하여 $|\phi_1|<1$이며 가역성(invertibility)을 위하여 $|\theta_1|<1$을 가정하고 있다. 따라서 모형 추정에서 자기회귀항(AR term)과 이동평균항(MA term)의 계수추정치의 절대값이 1보다 작아야 하며 1에 근접한 값을 가진 경우를 배제해야 할 것이다.

만일 ARIMA 모형을 적용할 때 연도별 자료를 이용한다면 별 문제가 없지만 월별 및 분기별 자료를 이용하면 계절성 문제가 발생할 수 있다. 앞에서 지적한 것처럼 계절성이란 어떤 시계열변수(time series variables)의 값이 주기적(periodically)인 특성을 지니는 경우를 의미한다. 예를 들어 고속도로 통행수요가 여름휴가 시즌에 피크를 이루며 수도권 전철 통행수요가 학기가 시작하는 3월 및 9월에 증가하는 패턴을 보이는 경우를 의미한다.

계절성을 띤 월별 및 분기별 자료를 이용하면서 시계열분석을 수행하는 기법은 다음과 같은 3가지가 있다. 첫째 원자료(raw data)와 ARIMA 모형을 이용하여 분석하는 경우이며, 둘째 계절조정(seasonal adjustment)을 한 자료와 ARIMA 모형을 이용하는 방법이 있다. 마지막으로는 원자료(raw data)와 SARIMA 모형을 이용하여 분석하는 경우이다.

첫 번째 방식은 계절성을 고려하지 않았기 때문에 올바른 추정방식으로 볼 수 없으며 두 번째 방법 역시 원자료를 이용하지 않고 계절조정을 수행한 자료를 이용하여 원자료가 내포하고 있는 의미를 왜곡하는 단점이 존재한다. Judge et al.(1988)은 계절 조정된 자료를 이용하지 말고 원자료를 이용하여 계절성을 고려한 ARIMA 모형(Seasonal ARIMA)을 이용할 것을 제시하고 있다.[2)]

ARIMA 모형은 특정 연도의 연속적인 월별자료(between observations for successive months in a particular year) 간의 특성을 고려하는 모형이며 SARIMA 모형은 이외에도 연속 연도의 동일 시점(월) 자료(between the observations for the same month in successive years)의 특성을 모두 고려하는 모형이다. SARIMA 모형은 크게 두 부분으로 정의되며 첫 번째 부분은 ARIMA 부분으로 식(1)과 같다. 예를 들어, 첫 번째 부분은 2008년 12월 교통수요는 2008년 11월 자료와 관계가 있음을 의미한다. 계절성을 고려한 부분은 예를 들면 2008년 12월 교통수요가 2007년 12월 교통수요와 관련이 있음을 의미하며 다음과 같이 정의된다.

$$\Phi_P(L)(1-L^s)^D y_t = \Theta_Q(L)u_t \tag{3}$$

이때 s는 만약 분기별 자료라면 4가 될 것이고 월별자료라면 12가 될 것이다. 또한 D는 계절 단위근(Seasonal Unit Root)의 차수, P는 계절 자기회귀(Seasonal Autoregressive, SAR)항의 차수, Q는 계절 이동평균(Seasonal Moving Average, SMA)항의 차수이다. 또한 $\Phi_P(L)$과 $\Theta_Q(L)$은 각각 계절 자기회귀항

2) Judge et al.(1988), pp. 713~715.

(SAR term)과 계절 이동평균항(SMA term)의 다항식을 의미한다.

최종적으로 첫 번째 부분과 두 번째 부분을 모두 함께 고려하면 다음과 같이 정의된다.

$$\phi_p(L)\Phi_P(L)(1-L)^d(1-L^s)^D y_t = \theta_q(L)\Theta_Q(L)u_t \quad (4)$$

식(4)가 SARIMA 모형으로 차수는 (p,d,q)×(P,D,Q)로 정의된다. 본 연구에서 이용되는 월별 지역간 철도자료가 단위근 및 계절단위근이 1개이고 이외의 모든 차수가 1이라면 SARIMA(1,1,1)×(1,1,1)로 정의되며 다음과 같다.

$$(1-\phi_1)(1-\Phi_1)(1-L)(1-L^{12})y_t = (1+\theta_1)(1+\Theta_1)u_t \quad (5)$$

ϕ_1 및 θ_1은 각각 1계 자기회귀항(AR(1) term)과 이동평균항(MA(1) term)의 계수이며 Φ_1과 Θ_1은 각각 계절 자기회귀항(SAR(12) term)과 계절 이동평균항(SMA(12) term)의 계수이다. ARIMA 모형과 마찬가지로 AR, MA, SAR, SMA항의 계수추정치 절대값을 유의하여 분석하여야 한다.

2. SARIMA 모형 추정 절차

SARIMA 모형의 추정절차는 크게 다음과 같이 구분할 수 있다. 즉 모형구분(Identification), 추정(Estimation), 진단(Diagnostic Testing)으로 구분할 수 있다.

모형구분(Identification)에서는 SARIMA의 차수, 즉 (p,d,q)×(P,D,Q)를 결정하는 것이다. 이때 단위근(d) 및 계절 단위근(D)이 있는 지를 검정하기 위하여 단위근 검정을 수행하여야 한다. 또한 자기회귀(Autoregressive) 및 계절 자기회귀(Seasonal Autoregressive) 부분의 차수(order)와 이동평균(Moving Average) 및 계절 이동평균(Seasonal Moving Average) 부분의 차수(order)를 결정하여야 한다. 이를 위하여 원자료의 자기상관함수(autocorrelation function)및 편자기상관함수(partial autocorrelation function)를 포함하는 correlogram을 관찰하여야 한다.

어떤 변수(yt)의 자기상관함수(ρk)는 다음과 같이 정의된다.

$$\rho_k = Cov(y_t y_{t+k})/[Var(y_t)\,Var(y_{t+k})]^{1/2} \quad (6)$$

$$k = 0, 1, 2, \cdots$$

이때 k는 시차로 볼 수 있다. 또한 편자기상관함수(ϕk)는 개별 변수와 그의 시차변수 간의 회귀식의 회귀계수로서 다음과 같이 정의된다.

$$\phi_k = \rho_1 \quad \text{for } k=1 \quad (7)$$

$$= \frac{\rho_k - \sum_{j=1}^{k-1}\phi_{k-1,j}\rho_{k-j}}{1-\sum_{j=1}^{k-1}\phi_{k-1,j}\rho_j} \quad \text{for } k>1$$

$$\phi_{k,j} = \phi_{k-1,j} - \phi_k\phi_{k-1,k-j}$$

일반적으로 원시계열 자료가 p차 자기회귀모형(AR process)을 따른다면 자기상관계수가 지수적으로(exponentially) 혹은 기하학적으로(geometrically) 감소하는 추세를 보이며 p차 이후의 편자기상관계수는 급격히 감소(cut off)하는 추세를 보인다.

원시계열 자료가 q차 이동평균모형(MA process)을 따른다면 편자기상관계수가 지수적으로(exponentially) 혹은 기하학적으로(geometrically) 감소하는 추세를 보이며 q차 이후의 자기상관계수는 급격히 감소(cut off)하는 추세를 보인다. 만일 이에 해당되지 않는다면 이는 자기회귀 이동평균 모형(ARMA process)에 해당된다.

(p,d,q)×(P,D,Q)의 차수결정에서 무엇보다 중요한 것은 간결성(parsimonious)으로 가능한 한 차수를 작게 하는 것이 모형의 추정에 유리하다.

추정(estimation)에서는 MA term이 있기 때문에 최소자승법(OLS)을 이용할 수 없다. 통계 패키지에 따라 조건부 최우추정법(conditional MLE) 혹은 비선형자승법(NLS)이 이용될 수 있다. 이때 다양한 차수에 따라 모형을 추정하여 AIC(Akaike Information Criteria, AIC) 혹은 SC(Schwarz Criteria)를 최소화하는 모형을 선택할 수 있으며, 추정치들의 t-통계량도 고려하여 모형의 차수를 결정하여야 할 것이다.

진단(Diagnostic Testing)에서는 구분과 추정에 의한 결과가 적정한 지를 판단하는 부분이다. 이를 위해 선

택된 모형에 AR 혹은 MA term을 하나씩 추가하여 추정을 하고 그 결과를 비교할 수 있다. 또 다른 방법으로 추정에서 얻은 잔차항(residuals)이 자기상관(autocorrelation)이 존재하는 지를 Ljung-Box (1978)의 Q-통계량을 이용하여 검정할 수 있다. Q-통계량은 다음과 같이 정의된다.

$$Q = T(T+2)\sum_{j=1}^{k}\frac{\rho_j^2}{T-j} \sim \chi^2(k) \qquad (8)$$

이때 T는 전체 표본의 숫자이며 k는 잔차항의 래그항(lag term) 개수를 의미한다. 만일 적절한 모형이 채택되었다면 잔차항에 자기상관이 존재하지 않을 것이다. 따라서 자기상관이 존재하지 않는다는 귀무가설(null hypothesis)을 기각하지 않을 것이다. 즉 Q-통계량이 작거나 p-value가 0.05 혹은 0.1보다 클 것이다. 만일 적절하지 않은 모형이라면 잔차항에 자기상관이 존재하여 귀무가설을 기각하며 이때 Q-통계량이 크고 p-value가 작을 것이다.

Ⅲ. SARIMA 모형 추정

1. 자료 설명

본 연구에서 이용되는 자료는 지역 간 철도부문의 여객자료로서 철도부문 여객수송실적에서 수도권 전철을 제외한 수송실적 자료로서 수송인원과 수송인-km 자료를 이용할 것이다. 한국철도공사・한국철도시설공단의 『철도통계연보』에서 2000년 1월부터 2008년 12월까지의 통행실적을 구축하였다. 『철도통계연보』에서 여객수송실적 자료에서 수도권 전철을 제외한 자료가 일반구간계로 정리되어 있다.

2008년 12월 기준으로 지역 간 철도의 월별 수송인원과 수송인-km는 약 915만 명과 15억 159만 인-km

〈표 2〉 지역 간 철도 여객수송실적(2008년 12월 기준)

(단위: 인, 인-km)

수송인원	수송인-km
9,150,414	1,501,594,900

자료: 한국철도공사・한국철도시설공단, 『2008 철도통계연보』, 2009.

수준이다. 지역 간 철도의 수송인원과 인-km의 자기상관함수(Autocorrelation Function, AC)와 편자기상관함수(Partial Autocorrelation Function, PAC)의 형태를 이용하여 어떠한 process를 나타내는 지를 살펴볼 수 있다.

〈표 3〉에서 지역 간 철도 수송인원 및 인-km의 자기상관 및 편자기상관계수값을 살펴보면 자기회귀(autoregressive) 혹은 이동평균(moving average)의 전형적인 특성을 나타내지 않고 있다. 즉 특정 시차 이후에 자기상관계수나 편자기상관계수가 급격히 감소(cut off)하지도 않으며 기하하적으로 감소하는 형태도 띄지 않고 있어서 자기회귀 및 이동평균 과정을 따른다고 볼 수 없다. 또한 1계 차분 및 12계 계절차분의 correlogram에서도 특정 시차 이후에 자기상관계수나 편자기상관계수가 급격히 감소하지 않고 있으며, 지수적으로 감소하는 추세도 보이지 않는다. 그러므로 지역 간 철도 통행실적은 적분된(integrated) ARMA process를 따른다고 볼 수 있다.

2. 단위근 검정

단위근이 존재한다는 것은 원자료(수송인원 및 수송인-km)가 불안정적(non-stationary)이라는 것을 의미한다. 이러한 자료들은 안정적(stationary)인 자료로 변환될 때까지 차분(difference)하여 시계열분석에 이용한다. ARIMA 혹은 SARIMA는 변수를 안정적인 시계열로 변환한 후에 추정하는 모형으로 볼 수 있다. 원자료가 안정적인지 혹은 불안정적인지를 검증하는 방법으로 ADF test (Augmented Dickey Fuller test)를 이용할 수 있다.

ADF test를 이용하여 지역 간 철도 통행실적을 수송인원과 수송인-km로 구분하여 단위근 검정을 수행하였다. 수송인원은 모든 모형에서 p-value가 커서 단위근(unit root)이 있다는 귀무가설을 기각하지 못하였다. 수송인-km는 모형 1에서는 p-value 커서 단위근이 존재한다는 귀무가설을 기각하지 못하였지만 모형 2와 모형 3에서는 p-value가 작아서 단위근이 존재한다는 귀무가설을 기각하였다. 그러나 모든 모형에서 단위근이 존재하지 않는다는 가설을 기각하지 못함으로 단위근이 존재한다고 보아야 할 것이다.[3]

3) 일반적인 단위근 검정 외에도 계절 단위근(seasonal unit root) 검정을 수행하여야 하지만 현재의 자료가 월별 자료를 추출한 것으로 계절 단위근 검정을 수행하는 것은 그리 큰 의미가 없는 것으로 생각된다.

〈표 3〉 지역 간 철도 수송인원의 자기상관(AC) 및 편자기상관계수(PAC)

구분	원자료				1계 차분 및 12계 계절차분			
시차	AC	PAC	Q-Stat	P-value	AC	PAC	Q-Stat	P-value
1	0.095	0.095	0.997	0.318	-0.354	-0.354	12.293	0.000
2	0.111	0.103	2.389	0.303	-0.066	-0.219	12.726	0.002
3	0.314	0.301	13.567	0.004	0.022	-0.100	12.776	0.005
4	0.188	0.150	17.614	0.001	-0.014	-0.069	12.796	0.012
5	0.238	0.194	24.144	0.000	-0.081	-0.139	13.468	0.019
6	-0.068	-0.223	24.675	0.000	0.202	0.131	17.701	0.007
7	0.190	0.088	28.942	0.000	-0.203	-0.117	21.998	0.003
8	0.072	-0.098	29.551	0.000	0.063	-0.020	22.414	0.004
9	0.216	0.296	35.157	0.000	0.042	0.020	22.601	0.007
10	-0.022	-0.191	35.214	0.000	0.040	0.086	22.778	0.012
11	-0.116	-0.084	36.848	0.000	0.089	0.214	23.657	0.014
12	0.564	0.491	76.202	0.000	-0.372	-0.371	39.030	0.000
13	-0.112	-0.289	77.767	0.000	0.024	-0.240	39.094	0.000
14	-0.089	-0.140	78.770	0.000	0.168	-0.028	42.320	0.000
15	0.058	-0.116	79.196	0.000	-0.081	-0.073	43.082	0.000
16	0.019	-0.014	79.241	0.000	0.062	0.050	43.527	0.000
17	0.012	-0.112	79.262	0.000	0.113	0.105	45.047	0.000
18	-0.298	-0.157	90.953	0.000	-0.227	-0.004	51.218	0.000
19	0.016	0.044	90.989	0.000	0.193	0.092	55.714	0.000
20	-0.159	-0.096	94.413	0.000	-0.048	-0.058	56.002	0.000

주 : 1. 1계 차분 및 12계 차분은 금기변수에서 전기 및 12기 전변수를 모두 차분한 경우임.
2. 이후의 모든 통행실적 자료에 자연로그(natural log)를 취하여 분석하였음.

〈표 4〉 지역 간 철도 수송인-km의 자기상관 및 편자기상관계수

구분	원자료				1계 차분 및 12계 계절차분			
시차	AC	PAC	Q-Stat	P-value	AC	PAC	Q-Stat	P-value
1	0.313	0.313	10.862	0.001	-0.135	-0.135	1.798	0.180
2	0.005	-0.103	10.865	0.004	-0.377	-0.403	15.873	0.000
3	0.132	0.182	12.831	0.005	-0.005	-0.159	15.876	0.001
4	0.010	-0.109	12.843	0.012	0.058	-0.160	16.217	0.003
5	-0.017	0.042	12.876	0.025	0.021	-0.080	16.262	0.006
6	0.063	0.034	13.344	0.038	0.021	-0.030	16.308	0.012
7	0.010	-0.018	13.355	0.064	-0.099	-0.136	17.324	0.015
8	-0.034	-0.022	13.494	0.096	-0.025	-0.092	17.392	0.026
9	0.093	0.111	14.543	0.104	0.080	-0.045	18.082	0.034
10	-0.032	-0.124	14.667	0.145	0.164	0.155	20.990	0.021
11	0.038	0.150	14.841	0.190	0.015	0.156	21.016	0.033
12	0.405	0.356	35.092	0.000	-0.394	-0.279	38.219	0.000
13	0.102	-0.184	36.388	0.001	0.037	-0.055	38.377	0.000
14	-0.023	0.075	36.456	0.001	0.198	-0.092	42.849	0.000
15	0.053	-0.088	36.812	0.001	-0.016	-0.066	42.879	0.000
16	-0.051	-0.034	37.148	0.002	-0.031	-0.016	42.991	0.000
17	-0.038	0.037	37.334	0.003	-0.058	-0.105	43.383	0.000
18	0.062	0.007	37.840	0.004	-0.011	-0.053	43.397	0.001
19	0.013	0.000	37.862	0.006	0.105	-0.043	44.734	0.001
20	-0.115	-0.108	39.641	0.006	0.111	0.115	46.238	0.001

〈표 5〉 지역 간 철도 통행수요 단위근 검정 결과

구분		수송인원			수송인-km		
		모형 1	모형 2	모형 3	모형 1	모형 2	모형 3
원자료	통계량	-0.4417	-2.0456	-1.9467	-0.3070	-7.4116	-7.3968
	P-value	0.5207	0.2672	0.6222	0.5729	0.0000	0.0000
	시차	12	12	12	2	0	0
1계 차분	통계량	-3.7652	-3.7615	-3.8572	-12.5214	-12.4604	-12.3982
	P-value	0.0002	0.0046	0.0177	0.0000	0.0000	0.0000
	시차	11	11	11	1	1	1
1계 차분 및 12계 계절 차분	통계량	-9.8178	-9.7659	-9.7189	-11.0247	-10.9661	-10.9243
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	시차	1	1	1	1	1	1

주: 모형 1은 선형추세가 없는 (no deterministic trend) 경우, 모형 2는 상수항이 있는(intercept) 경우, 모형 3은 상수항과 시간추세(intercept & time trend)가 있는 경우를 의미함.

지역 간 철도 통행실적을 1계 차분 및 12계 계절차분을 하여 단위근 검정을 수행한 결과에 의하면 모두 p-value가 작아서 단위근이 존재하지 않는 것으로 나타났다. 이러한 분석에 입각할 때 지역 간 철도 통행수요 함수의 동태적 변화추정은 SARIMA 모형을 이용하여야 한다.

3. 모형 추정

SARIMA $(p,d,q)\times(P,D,Q)$ 모형에서 종속변수는 지역 간 철도 수송인원과 수송인-km의 자연로그 변수를 1계 차분 및 12계 계절 차분한 자료를 이용하였으며, 2000년 1월부터 2008년 12월까지의 월별 자료를 이용하였다.

SARIMA $(p,d,q)\times(P,D,Q)$ 모형을 추정할 때 중요한 것으로 모형 구분(identification) 문제가 중요한 관심사이다. 즉 $(p,d,q)\times(P,D,Q)$의 차수를 결정해야 하는데, d와 D는 ADF 검정에 의하여 1로 판명되었다. p와 q및 P와 Q의 차수를 결정하기 위해서는 위에서 제시한 AIC와 SC 통계량을 이용할 수 있는데 이와 함께 고려되어야 할 사항이 간결성(parsimonious)이다. 따라

〈표 6〉 SARIMA 모형 추정 결과(지역 간 철도 수송인원)

구분	(1,1,1)×(1,1,1)	(1,1,1)×(1,1,0)	(1,1,1)×(0,1,1)	(1,1,1)×(0,1,0)	(1,1,0)×(1,1,1)
	(1)	(2)	(3)	(4)	(5)
상수항	0.0000	0.0004	0.0002	0.0001	0.0004
	(0.0007)	(0.0015)	(0.0008)	(0.0020)	(0.0010)
AR(1)	-0.1249	-0.1546	-0.4190***	0.2090	-0.5225***
	(0.1367)	(0.1515)	(0.1242)	(0.1827)	(0.0683)
SAR(12)	-0.0220	-0.4857***	-	-	-0.0467
	(0.1058)	(0.1211)	-	-	0.1198
MA(1)	-0.5346***	-0.4849***	-0.1960	-0.6987***	-
	(0.1244)	(0.1596)	(0.1189)	(0.1305)	-
SMA(12)	-0.8673***	-	-0.8521***	-	-0.8368***
	(0.0502)	-	(0.0487)	-	(0.0550)
Adj R^2	0.5170	0.3783	0.5052	0.1844	0.4734
F-statistic	22.6759	17.4319	32.6499	11.5122	25.2762
관측치의 개수	82	82	94	94	82
AIC	-3.6586	-3.4177	-3.6531	-3.1636	-3.5837
SC	-3.5118	-3.3003	-3.5449	-3.0824	-3.4663
DW	1.8867	1.9672	1.9810	2.0015	2.0704

주: 1. ***, **, *는 1%, 5%, 10% 유의수준에서 통계적으로 유의함을 의미함.
2. 괄호안의 수치는 Newey-West 표준오차(standard error)임.
3. AIC 및 SC는 Akaike Information Criteria와 Schwarz Criteria 통계량을 의미하며 DW는 Durbin Waston 통계량을 의미함.

〈표 6 계속〉 SARIMA 모형 추정 결과(지역 간 철도 수송인원)

구분	(1,1,0)×(0,1,1)	(1,1,0)×(1,1,0)	(1,1,0)×(0,1,0)	(0,1,1)×(1,1,1)	(0,1,1)×(1,1,0)
	(6)	(7)	(8)	(9)	(10)
상수항	0.0006	0.0005	-0.0007	0.0003	0.0005
	(0.0008)	(0.0018)	(0.0031)	(0.0007)	(0.0014)
AR(1)	-0.4629***	-0.4919***	-0.3585***	-	-
	(0.0650)	(0.0738)	(0.0930)	-	-
SAR(12)	-	-0.4851***	-	-0.0159	-0.4706***
	-	(0.1353)	-	(0.1096)	(0.1175)
MA(1)	-	-	-	-0.6092***	-0.5779***
	-	-	-	(0.0855)	(0.0996)
SMA(12)	-0.8931***	-	-	-0.8615***	-
	(0.0463)	-	-	(0.0500)	-
Adj R^2	0.5044	0.3379	0.1176	0.5016	0.3783
F-statistic	48.3323	21.6652	13.3946	28.5114	25.9527
관측치의 개수	94	82	94	83	83
AIC	-3.6618	-3.3662	-3.0952	-3.6507	-3.4412
SC	-3.5807	-3.2782	-3.0411	-3.5341	-3.3537
DW	2.1992	2.1961	2.1443	1.9970	2.0826

〈표 6 계속〉 SARIMA 모형 추정 결과(지역 간 철도 수송인원)

구분	(0,1,1)×(0,1,1)	(0,1,1)×(0,1,0)	(0,1,0)×(1,1,1)	(0,1,0)×(1,1,0)	(0,1,0)×(0,1,1)
	(11)	(12)	(13)	(14)	(15)
상수항	0.0011	-0.0003	0.0008	0.0004	0.0015
	(0.0007)	(0.0023)	(0.0016)	(0.0023)	(0.0015)
AR(1)	-	-	-	-	-
	-	-	-	-	-
SAR(12)	-	-	0.0924	-0.3878***	-
	-	-	(0.1521)	(0.1334)	-
MA(1)	-0.6433***	-0.5384***	-	-	-
	(0.0471)	(0.0922)	-	-	-
SMA(12)	-0.7577***	-	-0.8104***	-	-0.6315***
	(0.0638)	-	(0.0775)	-	(0.0790)
Adj R^2	0.4974	0.1797	0.3008	0.1551	0.2783
F-statistic	47.5177	21.5928	18.6362	16.0521	37.2404
관측치의 개수	95	95	83	83	95
AIC	-3.6578	-3.1781	-3.3236	-3.1460	-3.3061
SC	-3.5771	-3.1243	-3.2362	-3.0877	-3.2523
DW	1.9315	1.8817	2.8452	2.8961	2.8475

서 AIC와 SC 통계량과 함께 p와 q및 P와 Q의 차수를 최소화하여 추정하는 것이 바람직하다. 또한 이외에도 당연히 개별 추정치의 t-통계량이 의미 있는 수치가 되어야 한다. d와 D의 차수를 1로 보았을 때 p와 q, 그리고 P와 Q의 가능한 차수, 즉 0과 1을 고려하면 모두 가능한 조합은 24로 16개의 모형을 고려할 수 있다. 이때 (0,1,0)×(0,1,0)[4]을 배제하면 모두 15가지 모형을 모두 고려해야 할 것이다.

우선 지역 간 철도 수송인원 기준으로 살펴보면 모형 1, 2, 4, 5, 9, 13은 개별 설명변수의 t-통계량이 작아

4) (0,1,0)×(0,1,0)은 AR, MA, SAR, SMA 항이 모두 존재하지 않은 경우로서 추정의 의미가 없음으로 고려할 필요가 없다.

〈표 7〉 SARIMA 모형 추정 결과(지역 간 철도 수송인-km)

구분	(1,1,1)×(1,1,1)	(1,1,1)×(1,1,0)	(1,1,1)×(0,1,1)	(1,1,1)×(0,1,0)	(1,1,0)×(1,1,1)
	(1)	(2)	(3)	(4)	(5)
상수항	0.0002	0.0005	-0.0003	0.0000	-0.0007
	(0.0021)	(0.0011)	(0.0013)	(0.0009)	(0.0049)
AR(1)	0.2338	0.3833**	0.8233***	0.4567***	-0.3475***
	(0.1652)	(0.1500)	(0.1745)	(0.0934)	(0.1041)
SAR(12)	-0.4443	-0.3795	-	-	-0.4869**
	(0.2697)	(0.2511)	-	-	(0.2319)
MA(1)	-0.7928***	-0.8879***	-0.9813***	-0.9805***	-
	(0.0774)	(0.0748)	(0.0093)	(0.0210)	-
SMA(12)	0.4277	-	-0.6788***	-	0.4148*
	(0.2848)	-	(0.1216)	-	(0.2387)
Adj R^2	0.4484	0.3947	0.3663	0.2400	0.3697
F-statistic	17.4645	18.6050	18.9192	15.6875	16.8388
관측치의 개수	82	82	94	94	82
AIC	-2.2851	-2.2036	-1.8961	-1.7246	-2.1632
SC	-2.1384	-2.0862	-1.7879	-1.6435	-2.0458
DW	1.9662	2.0047	2.0522	1.8226	2.1877

주: 1. ***, **, *는 1%, 5%, 10% 유의수준에서 통계적으로 유의함을 의미함
2. 괄호안의 수치는 Newey-West 표준오차(standard error)임
3. AIC 및 SC는 Akaike Information Criteria와 Schwarz Criteria 통계량을 의미하며 DW는 Durbin Waston 통계량을 의미함

〈표 7 계속〉 SARIMA 모형 추정 결과(지역 간 철도 수송인-km)

구분	(1,1,0)×(1,1,0)	(1,1,0)×(0,1,1)	(1,1,0)×(0,1,0)	(0,1,1)×(1,1,1)	(0,1,1)×(1,1,0)
	(6)	(7)	(8)	(9)	(10)
상수항	-0.0010	-0.0006	0.0005	-0.0002	-0.0006
	(0.0039)	(0.0031)	(0.0073)	(0.0032)	(0.0027)
AR(1)	-0.2995***	-0.1768	-0.1360	-	-
	(0.1057)	(0.1242)	(0.1004)	-	-
SAR(12)	-0.4181	-	-	-0.4586*	-0.4124
	(0.2543)	-	-	(0.2665)	(0.2710)
MA(1)	-	-	-	-0.6005***	-0.5800***
	-	-	-	(0.0990)	(0.0868)
SMA(12)	-	-0.8882***	-	0.4158	-
	-	(0.0453)	-	(0.2884)	-
Adj R^2	0.2868	0.4092	0.0079	0.4335	0.3536
F-statistic	17.2881	33.2070	1.7367	21.9198	23.4310
관측치의 개수	82	94	94	83	83
AIC	-2.0513	-1.9764	-1.4684	-2.2821	-2.1616
SC	-1.9632	-1.8953	-1.4143	-2.1655	-2.0742
DW	2.1715	2.0990	2.1022	1.8362	1.7762

서 통계적으로 유의하지 않은 모형이어서 배제할 수 있다. 모형 4, 5, 6, 7, 8, 10, 12, 13, 14, 15는 잔차항의 Q-통계량이 크게 나타나서 잔차항 간에 자기상관이 존재하지 않는다는 귀무가설을 기각하고 있어서 모형의 적절성에 문제가 있다.

모형 11을 이용하여 추정한 결과에 의하면 과도한 미래예측치가 산정되는 것으로 나타나서 모형 3, SARIMA (1,1,1)×(0,1,1)을 선택하였다. 모형 3은 AR(1)항과 SMA(12)이 1% 수준에서 통계적으로 유의하며 MA(1)항의 t-통계량이 약 -1.65로 산정되어 10% 유의수준에

〈표 7 계속〉 SARIMA 모형 추정 결과(지역 간 철도 수송인-km)

구분	(0,1,1)×(0,1,1)	(0,1,1)×(0,1,0)	(0,1,0)×(1,1,1)	(0,1,0)×(1,1,0)	(0,1,0)×(0,1,1)
	(11)	(12)	(13)	(14)	(15)
상수항	-0.0004	0.0007	-0.0009	-0.0011	-0.0007
	(0.0018)	(0.0036)	(0.0055)	(0.0044)	(0.0034)
AR(1)	-	-	-	-	-
	-	-	-	-	-
SAR(12)	-	-	-0.4777**	-0.4000*	-
	-	-	(0.2509)	(0.2384)	-
MA(1)	-0.6203***	-0.6965***	-	-	-
	(0.0661)	(0.0631)	-	-	-
SMA(12)	-0.8927***	-	0.3861	-	-0.8918***
	(0.0400)	-	(0.2656)	-	(0.0404)
Adj R^2	0.4665	0.0971	0.2935	0.2282	0.4038
F-statistic	42.0903	11.1096	18.0297	25.2442	64.6587
관측치의 개수	95	95	83	83	95
AIC	-2.0827	-1.5669	-2.0726	-1.9959	-1.9819
SC	-2.0021	-1.5132	-1.9852	-1.9377	-1.9281
DW	1.5856	1.4631	2.6739	2.5753	2.3062

서 단측검정에서는 한계적으로 유의하였다.

또한 AIC와 SC를 이용하여 모형의 적절성을 살펴보면 모형 3보다 AIC 및 SC가 더 적은 모형은 (6)과 (11)로 나타났다. 앞에서 지적한 것처럼 모형 6은 잔차항의 Q-통계량이 크게 산정되었으며 모형 11은 과도한 미래 예측치가 산정되어 배제할 수 있다. 따라서 모형 3을 가장 적절한 대안으로 선택할 수 있다.

지역 간 철도 수송인-km 기준으로 살펴보면 모형 1, 6, 7, 8, 9, 10, 13은 개별 설명변수의 t-통계량이 작아서 통계적으로 유의하지 않아서 적절한 모형이 아님을 알 수 있다. 모형 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15는 잔차항의 Q-통계량이 크게 나타나서 잔차항 간에 자기상관이 존재하지 않는다는 귀무가설을 기각하고 있어서 모형의 적절성에 문제가 있다.

모형 3과 모형 4는 추정치(MA(1) 항)의 값이 1에 근접하여 가역성(invertibility)을 충족시키지 못하여 채택할 수 없었다. 따라서 모형 2, SARIMA (1,1,1)×(1,1,0)을 선택할 수 있다. 모형 2의 AR(1)항 t-통계량이 약 -1.51로 나타나서 만족스러운 수준은 아니지만 다른 대안에 비해 적절하여 채택하였다.

그리고 모형 2보다 AIC와 SC가 작은 것은 모형 1과 모형 9인데 모두 개별 변수들의 t-통계량에 문제가 있어서 채택할 수 없었다.

4. 모형 진단

1) Q-통계량을 이용한 모형 진단

추정된 SARIMA 모형의 적절성 여부를 판단하기 위하여 모형 진단이 필요하다. 진단의 한 방법으로 추정에서 얻은 잔차항(residuals)이 자기상관(autocorrelation)이 존재하는 지를 Q-통계량을 이용하여 검정할 수 있다. 수송인원과 수송인-km의 경우에 각각 모형 3, SARIMA (1,1,1)×(0,1,1)과 모형 2, SARIMA (1,1,1)×(1,1,0)을 선택하여 추정한 후에 잔차항을 구할 수 있다. 이들 잔차항이 자기상관이 존재하는 지를 검정하기 위하여 Q-통계량을 산정하였다. 귀무가설은 잔차항 간에 자기상관이 존재하지 않는다는 것으로 Q-통계량이 작게 나타나서 귀무가설을 기각하지 않아야 모형 선택이 적절한 것으로 볼 수 있다.

〈표 8〉에서 수송인원의 경우에 시차 6까지는 Q-통계량이 다소 크게 산정되어 잔차항 간에 자기상관이 존재하는 것으로 나타나지만 그 이후에는 자기상관이 존재하지 않는 것으로 나타났다[5].

또한 수송인-km의 경우에 잔차항 간의 자기상관이 존재하지 않는다는 것을 Q-통계량이 보여 주고 있다.

따라서 수송인-km의 경우, 모형 선택이 적절하였다

5) 시차 6까지는 Q-통계량의 p-value가 0.1(10%)보다 작아서 귀무가설(잔차항간에 자기상관이 존재하지 않는다)을 기각하고 있다. 그러나 시차 7 이후에는 p-value가 0.1보다 커서 귀무가설을 기각하지 못하고 있다.

〈표 8〉 SARIMA 모형 추정 이후 잔차항의 Q 검정 결과

구분	수송인원		수송인-km	
시차	Q-통계량	P-value	Q-통계량	P-value
1	0.0006	-	0.0186	-
2	4.8276	-	0.4106	-
3	5.4163	-	0.8759	-
4	5.4504	0.020	1.0749	0.300
5	5.9215	0.052	1.1488	0.563
6	6.4749	0.091	1.5705	0.666
7	7.5242	0.111	3.9543	0.412
8	7.5492	0.183	4.0986	0.535
9	7.5523	0.273	4.6201	0.593
10	7.5523	0.374	5.1147	0.646
11	8.0629	0.427	5.1333	0.743
12	8.0993	0.524	7.8682	0.547
13	8.2078	0.609	8.2352	0.606
14	8.3832	0.679	8.726	0.647
15	9.0911	0.695	8.9668	0.706

고 볼 수 있으며 수송인원의 경우에도 다소 미흡하지만 모형 선택이 적절하다고 볼 수 있다.

2) ARIMA 모형과 비교를 통한 진단

본 연구의 SARIMA 모형의 적정성 여부를 기존의 ARIMA 모형과의 비교를 통하여 수행할 수 있다. 계절조정을 하지 않은 자료를 이용한 ARIMA 모형과 계절조정을 한 ARIMA 모형이 비교대상이 될 수 있다. Judge et al. (1988)에서 지적하는 것처럼 계절조정을 한 이후에 ARIMA 모형을 적용하는 것은 원자료가 내포하고 있는 의미를 왜곡하여 적절치 않다고 하였다. 따라서 본 연구에서는 계절조정을 하지 않은 자료를 이용한 ARIMA 모형과 본 연구의 SARIMA 모형을 비교하고자 한다. 계절조정을 수행한 ARIMA 모형의 종속변수와 SARIMA 모형 및 계절조정을 하지 않은 ARIMA 모형의 종속변수와 상이하여 직접적인 비교가 어렵다. 예를 들어 종속변수가 상이한 모형 간에는 adjusted R^2, Akaike Information Criteria (AIC) 및 Schwarz Criteria (SC)를 직접 비교할 수 없다 (Wooldridge, 2000). 따라서 본 연구의 SARIMA 모형과 계절조정을 하지 않은 ARIMA 모형을 직접 비교하고자 한다.[6]

계절조정을 하지 않은 ARIMA 모형에서 여러 가지경

〈표 9〉 ARIMA 모형 추정 결과

구분	수송인원	수송인-km
상수항	-0.0001	0.0001
	(0.0010)	(0.0005)
AR(1)	-	0.3217***
	-	(0.1039)
MA(1)	-0.8588***	-0.9873***
	(0.0436)	(0.0129)
Adj R^2	0.4705	0.3154
F-statistic	95.1713***	25.1905***
관측치의 개수	107	106
AIC	-2.5724	-1.8077
SC	-2.5225	-1.7323
DW	2.1458	1.9186

주: 1. ***, **, *는 1%, 5%, 10% 유의수준에서 통계적으로 유의함을 의미함
2. 괄호안의 수치는 Newey-West 표준오차(standard error)
3. AIC 및 SC는 Akaike Information Criteria와 Schwarz Criteria 통계량을 의미하며 DW는 Durbin Waston 통계량을 의미함

우의 수가 존재하며 이에 따른 추정결과가 존재한다. 이때 개별 추정결과에서 Adjusted R^2와 AIC 및 SC를 비교하고 개별 추정치들의 t-통계량을 검토하였다. 검토결과에 따르면 수송인원과 수송인-km에서는 ARIMA (0,1,1)과 ARIMA(1,1,1)이 각각 가장 적절한 것으로 나타났다.

수송인원의 경우, SARIMA 모형 3의 추정결과와 ARIMA 모형 추정결과를 비교하면 SARIMA 모형이 더 적절한 것을 알 수 있다. SARIMA 모형의 AIC 및 SC가 ARIMA 모형의 그것보다 더 작게 나타났다. 또한 SARIMA 모형 Adjusted R^2가 ARIMA 모형의 그것보다 크게 산정되어 모형설명력이 더 뛰어나다고 볼 수 있다.[7]

수송인-km의 경우, SARIMA 모형 2의 추정결과와 ARIMA 모형 추정결과를 AIC와 SC 및 Adjusted R^2로 비교하였다. SARIMA 모형의 AIC 및 SC가 ARIMA 모형의 그것들보다 더 작게 나타났으며, SARIMA 모형 Adjusted R^2가 ARIMA 모형의 그것보다 큰 값을 띠고 있다. 이는 수송인-km에서도 SARIMA 모형의 설명력과 예측력이 뛰어나다는 증거이다. 따라서 본 연구에서 제시한 SARIMA 모형이 적절하다고 볼 수 있다.

6) 교통부문의 대부분 시계열분석이 계절조정을 수행하지 않고 ARIMA 모형을 적용한 경우가 다수이다 (한국교통연구원, 2009). 따라서 이러한 비교는 기존 연구 방법론과 본 연구 방법론과의 비교라고 할 수 있다.
7) AIC와 SC의 값이 작고 Adjusted R^2가 클수록 적절한 모형이라 할 수 있다.

3) 구조변화 진단

SARIMA 모형과 같은 선형 시계열 모형 사용 시에는 분석하고자 하는 시계열에 무시할 수 없는 비선형성이 있는지 확인해야 한다. 교통 시계열을 예로 들면 새로운 운송수단의 등장이나 도로망의 확장은 기존의 교통흐름에 급격한 변화를 줄 수 있다. 그러나 이러한 변화는 과거 데이터에서는 예측될 수 없는 것이다. 따라서 이러한 불연속적인 사건을 내포하는 시계열 자료를 분석할 때는 SARIMA와 같은 선형시계열 모형은 적합하지 않을 수 있다. 따라서 철도수송실적 시계열에 구조적인 변화가 있었는지를 확인하고자 하였다.

고속철도의 개통 등과 같은 명시적인 사건 외에도 존재할 수 있는 구조변화 요인을 종합적으로 고려하기 위해 Bai & Perron (1998, 2003)이 제안한 구조변화 검정을 본 연구에 적용할 수 있다.

다음과 같은 k개의 다중회귀분석을 가정하여 보자.

$$y_t = \beta_j + u_t, (t = T_{j-1}+1, \cdots, T_j,\ j = 1, \cdots, k) \qquad (9)$$

y_t는 분석하고자 하는 시계열 자료이고 β_j는 j번째 국면(regime)에서의 시계열의 평균값이다. u_t는 t기의 잔차항이다. 총 k개의 구조변화 시점들(structural break points)이 있고, 따라서 $k+1$개의 국면이 존재한다. $(T_1, T_2, \cdots, T_k)$는 구조변화 시점들의 집합이다.[8] Bai-Perron 구조변화검정의 목적은 구조변화가 존재하는지 여부와 존재한다면 구조변화 시점을 내생적으로 측정하는 것이다. 우선 식 (10)와 같이 각 k개의 분할에 대하여 모든 경우의 수를 고려하여 잔차제곱의 합을 계산한다.

$$S_T(T_1, T_2, \cdots, T_k) = \sum_{j=1}^{k+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - \beta_j)^2,\ (k = 1, \cdots, m) \qquad (10)$$

이 중 최소의 잔차제곱의 합을 내는 구조변화 시점들의 집합이, 주어진 구조변화시점의 수에서 최적의 구조변화 시점들의 집합이 된다. 이를 수식으로 표현하면 식 (11)과 같다.

$$(\hat{T}_1, \hat{T}_2, \cdots, \hat{T}_k) = \operatorname{argmin} S_T(T_1, T_2, \cdots, T_k),\ (k = 1, \cdots, m) \qquad (11)$$

위에서 측정된 $(\hat{T}_1, \hat{T}_2, \cdots, \hat{T}_k)$에 대응하는 개별 국면 하에서의 시계열 평균값인 $\hat{\beta}(\{\hat{T}_1, \hat{T}_2, \cdots, \hat{T}_k\})$을 얻을 수 있다. 이러한 과정을 통해 총 m개의 구조변화점과 그에 대응하는 시계열의 평균값을 구하게 된다.

다음으로 우리는 구조변화가 통계적으로 유의한 수준으로 존재하느냐에 대한 검정을 해야 한다. 이를 위해 supF 검정통계량과 비가중 이중극대값 검정통계량(unweighted double maximum statistic)을 사용하였다. supF 검정통계량은 k개의 구조변화점이 있다는 가정 하에 '시계열에 k개의 구조변화가 존재한다.'는 대립가설 하에서 '시계열에 구조변화가 없다'는 귀무가설을 검정하는 F-통계량이다. 귀무가설과 검정통계량은 각각 다음과 같다.

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_{k+1} \qquad (12)$$

$$\begin{aligned} & supF_T(\lambda_1, \cdots, \lambda_k; 1) \\ & = sup \frac{1}{T}\left(\frac{T-(k+1)}{k}\right)\left((R\hat{\delta})'(R\hat{V}(\hat{\delta})R')^{-1}(R\hat{\delta})\right) \end{aligned} \qquad (13)$$

식 (13)에서 $\lambda_i = T_i / T$는 구조변화점을 전체표본의 기간으로 나누어 0과 1사이에 존재하도록 표준화 한 변수이다. $\delta = (\delta_1, \delta_2, \cdots, \delta_{k+1})$이고 $(R\delta)' = (\delta'_1 - \delta'_2, \cdots, \delta'_k - \delta'_{k+1})$이다. 그리고 $\hat{V}(\hat{\delta})$는 계열상관과 이분산에 강건한 분산-공분산 행렬이다. supF 검정통계량을 이용하면 다양한 구조변화의 수(k) 하에서 시계열에 구조변화가 존재하는 지에 대한 여부를 검정할 수 있다.

supF 검정통계량과 달리 구조변화의 수를 명시적으로 지정하지 않고 구조변화가 존재하는 지 여부를 검정하는 방법도 있는데 이는 비가중 이중극대값 통계량을 통해 가능하다. 비가중 이중극대값 통계량은 여러 k값 하에서의 supF 검정통계량 중 최대값이며 통계량의 유의성은 별도의 통계표를 이용하여 검정된다.

마지막으로 시계열에 구조변화가 존재한다면 몇 번의 구조변화가 있는지를 통계적으로 측정해야 한다. 이를 위해 베이즈 정보기준(Bayesian Information Criteria)과 Liu et al. (1997)이 제안한 수정된 슈와르츠 기준(modified Schwarz Criteria)을 이용하여 최적의 구조변화의 횟수를 추정하였다.

지금까지 기술한 검정통계량을 수송인원과 수송인-km 시계열에 적용한 결과가 〈표 10〉에 기록되어 있다.

8) m이 최대 구조변화 시점의 수라면 $k = 1, \cdots, m$이다.

〈표 10〉 수송인원과 수송인-km 시계열에 대한 Bai & Perron 구조변화 검정 결과

구분		수송인원	수송인-km
SupF test	0 vs 1	1.2071 (7.0400)	1.4882 (7.0400)
	0 vs 2	2.7148 (6.2800)	5.3852 (6.2800)
	0 vs 3	2.2981 (5.2100)	4.3330 (5.2100)
	0 vs 4	1.6403 (4.4100)	3.2807 (4.4100)
	0 vs 5	1.4238 (3.4700)	1.4255 (3.4700)
UDmax test		2.7148 (7.4600)	5.3852 (7.4600)
BIC	0 breaks	11.4780	11.5832
	1 break	11.5560	11.6604
	2 breaks	11.6342	11.7258
	3 breaks	11.7096	11.7977
	4 breaks	11.7933	11.8812
	5 breaks	11.8779	11.9786
	선택된 구조변화의 수	0	0
LWZ	0 breaks	11.4874	11.5925
	1 break	11.6392	11.7435
	2 breaks	11.7915	11.8831
	3 breaks	11.9414	12.0295
	4 breaks	12.1000	12.1880
	5 breaks	12.2600	12.3607
	선택된 구조변화의 수	0	0

주: 1. UDmax test는 비가중 이중극대값 통계량을 이용한 구조변화 검정임.
2. SupF test와 UDmax test의 괄호 안의 숫자는 10% 유의수준의 임계치(critical value)임.
3. BIC: 베이즈 정보기준(Bayesian Information Criteria), LWZ: Liu et al.(1997)가 제안한 수정된 슈와르츠 기준

우선 supF 검정통계량을 보면 두 시계열 모두 한 번에서 다섯 번의 구조변화가 있었다는 대립가설들에 대해 구조변화가 없었다는 귀무가설이 10% 유의수준에서 기각되지 않았다. 다음으로 귀무가설에서 구조변화의 수가 명시되지 않는 비가중 이중극대값 통계량을 살펴보면 두 시계열에서 10% 유의수준에서 귀무가설이 기각되지 않았다. 이는 유의한 수준의 구조변화가 없었다는 통계적인 증거이다. 다음으로 베이즈 정보기준과 수정된 슈와르츠 정보기준으로 구조변화의 수를 추정해보았다. 베이즈 정보기준의 경우 수송인원과 수송인-km 모두 구조변화의 수가 0일 때 베이즈 정보기준 통계량이 가장 작았다. 수정된 슈와르츠 정보기준 역시 베이즈 정보기준과 동일하게 구조변화의 수가 0일 때 통계량이 가장 작았다.[9] 두 정보기준 모두 산출된 통계량이 작은 모형일수록 더 적합한 모형으로 판정되므로 결국 두 교통시계열에서 유의한 수준의 구조변화가 없었다고 판단할 수 있다. 즉 KTX 개통 등의 환경변화가 수송실적을 크게 변화시키지 않고 있는 것으로 파악되었다. 이러한 결과를 토대로 할 때 추가적인 자료 가공이나 분할 없이 원자료를 그대로 이용한 본 연구의 SARIMA 모형 결과가 적정하다고 볼 수 있다.

Ⅳ. 장래 통행수요 예측

1. 통행수요 예측방법 및 유의사항

수송인원 및 수송인-km 기준 자료를 이용하여 동태적 변화과정을 SARIMA 모형으로 추정하였다, 각각 SARIMA(1,1,1)×(0,1,1)과 SARIMA (1,1,1)×(1,1,0)으로 채택된 모형을 추정하였다. 이를 기반으로 장래 교통수요를 예측할 수 있는데 이의 기본 원리는 다음과 같다. 즉

$$\widehat{y_T}(k) = E(y_{T+k}|y_T, y_{T-1}, \cdots) \tag{14}$$

이때 $\widehat{y_T}(k)$는 예측시점이 T(본 연구에서는 2008년 12월)이고 k 시점 이후의 yT+k에 대한 예측치를 의미한다. 결국 과거 및 현재 정보, yT, yT-1 등을 이용하여 미래를 예측하는 것이며 이는 과거 및 현재 정보를 이용한 조건부 기대치로 해석할 수 있다.

이러한 원리에 입각하고 SARIMA 모형의 추정치를 이용하면 수송인원 및 수송인-km 기준 통행수요 예측치를 월별로 산정할 수 있다.[10] 산정된 월별 예측치를 이용하면 연도별 예측치를 구할 수 있다.

본 연구에서는 약 10 여년치(정확하게 9년치)의 실적자료를 이용하였기 때문에 10여년 정도의 장래자료를 예측하였다. 약 10 여년치 실적자료를 이용하여 추정하고 지나치게 먼 미래까지 예측한다면 다소 문제가 있을

9) 베이즈 정보기준 통계량은 수송인원이 11.4780, 수송인-km이 11.5832였고 수정된 슈와르츠 기준 통계량은 수송인원이 11.4874, 수송인-km가 11.5925였다.
10) 월별 통행수요 예측은 통계 패키지 E-Views를 이용하여 산정하였다.

수 있다. 현재의 제약된 정보를 이용하여 먼 미래를 예측하는 것은 왜곡된 예측치를 얻을 수 있기 때문이다. 따라서 SARIMA 모형에 의한 예측은 실체 모형에 이용된 수준(연도)의 예측을 하여야 할 것이다.

2. 통행수요 예측

SARIMA모형에 의거하여 지역 간 철도의 장래 교통수요를 예측하면 수송인원 및 수송인-km 기준으로 2010년에는 1억 3,212만 명과 189억 3,699만인-km 수준으로 예측되었다. 이는 2008년의 1억 1,310만 명과 186억 7,136만 인-km와 비교하여 각각 16.82%와 1.42% 증가한 수치이다.

2015년과 2020년의 수송인원은 각각 1억 5,409만 명과 1억 9,338만 명으로 산정되었으며, 이는 2008년의 1.36배와 1.71배에 해당하는 수치이다.

2015년과 2020년의 수송인-km는 각각 232억 4,543만 인-km와 333억 1,731만인-km로 산정되었으며 이는 2008년의 1.25배와 1.78배에 해당하는 수치이다.

수송인원에 비해 수송인-km는 초기에는 증가율이 작지만 시간이 흐름에 따라 증가율이 높아지고 있다.[11] 앞에서 지적한 것처럼 추정에 이용된 연도(개월 수)만큼을 장래예측기간으로 설정하여 과도한 장래 예측에 따른 예측오차를 감소시키고자 하였다.

〈표 11〉 통행수요 예측

구분	수송인원	수송인-km
2010	132,119,229	18,936,991,011
2011	135,451,675	19,486,412,204
2012	139,275,618	20,176,459,611
2013	143,627,690	21,020,850,980
2014	148,550,330	22,036,764,958
2015	154,092,476	23,245,431,745
2016	160,310,369	24,672,867,037
2017	167,268,498	26,350,802,532
2018	175,040,711	28,317,851,498
2019	183,711,502	30,620,972,463
2020	193,377,522	33,317,306,415
2008년 실적 대비 2020년 예측치 비중	1.71	1.78

본 연구의 SARIMA 모형에서 제시한 장래 수요 예측치의 적절성을 다른 연구와 비교하여 평가할 필요가 있다. 그렇지만 본 연구의 지역 간 철도자료와 매칭되는 다른 연구가 부재한 상황에서 직접적인 비교는 매우 어려운 실정이다. 하지만 2010년 12월에 발표된 『국가기간교통망계획 제2차 수정계획』에서 향후 철도의 여객수송분담률 변화를 설명하고 있다. 이에 따르면 2020년 수송인원 및 수송인-km 기준 철도여객 수송실적 예측치가 각각 2008년의 1.64배와 1.81배에 달할 것이라고 하였다. 이는 본 연구의 2008년 실적치 대비 2020년 예측치 비중과 매우 유사하다고 볼 수 있다. 이러한 점에서 본 연구의 SARIMA 모형에 의한 예측이 적절하다고 볼 수 있다.

Ⅴ. 결론 및 정책적 시사점

본 연구에서는 계절성을 감안한 ARIMA 모형(SARIMA 모형)을 이용하여 교통실적의 동태적 변화과정을 추정하고 이를 이용하여 장래 교통수요에 적용하는 모형을 구축하였다. 국외 연구에서는 교통실적의 동태적 변화과정을 SARIMA 모형으로 추정하고 장래 예측을 시도하는 다양한 문헌이 존재한다. 그러나 국내에서는 연도별 교통실적자료를 이용하여 ARIMA 모형으로 추정하고 장래를 예측하는 모형이 간혹 있을 뿐이다. 앞에서 지적한 것처럼 연도별 자료의 이용은 추정자료 개수의 부족문제를 초래하여 추정치의 통계적 유의성에 문제를 불러일으킬 수 있다. 그러므로 기간이 짧은 연도별 자료보다는 월별 및 분기별 자료를 이용하여 자료의 개수 문제를 해결하여야 한다. 또한 분기별 및 월별 자료를 이용하게 되면 계절성을 감안한 SARIMA 모형을 적용하여야 한다.

본 연구는 우리나라 교통실적 자료들의 축적기한이 짧은 문제를 해결하기 위하여 월별 교통실적자료를 이용하였으며 교통실적에서 나타나고 있는 계절성을 감안한 ARIMA 모형을 적용하여 동태적 변화과정을 추정하였고 이를 기반으로 장래 교통수요를 예측하였다. 외국에서는 이와 같은 시도가 다수 있었지만 우리나라에서는

11) 본 연구의 지역 간 철도 수송인원과 국가교통DB센터(www.ktdb.go.kr)의 지역 간 수단별 여객 기종점 통행량 자료의 철도부문 여객, 2008, 2011, 2016, 2021년 수송인원을 비교할 수 있다. 그러나 국가교통DB센터의 기종점 통행량은 지역 간 철도와 수도권 전철을 구분하지 않고 있었다. 예를 들어 국가교통DB센터의 2008년 수송예측인원이 7억 4,230만 명으로 철도통계연보의 지역 간 철도 수송인원 약 1억 1,310만 명 혹은 지역 간 철도 및 수도권 전철의 전체 수송인원 약 10억 1,898만 명과 부합되지 않고 있었다. 따라서 국가교통DB와의 단순비교는 어려운 실정이다. 또한 국가교통DB센터에서는 수송인-km 예측치는 제공하지 않고 있다.

본 연구 외에는 이러한 시도가 전무하였다는 점에서 학술적 기여가 있을 것이다.

본 연구에서는 지나친 과거 교통실적자료를 이용에 따른 장래 교통수요 왜곡을 막기 위하여 과거 9년치 자료를 이용하였다. 즉 2000년대 이전의 교통실적 자료들은 증가율이 큰 편이었지만 2000년대 이후에는 증가율이 완만하게 감소하고 있는 실정이다. 미래 우리나라의 교통수요 여건과 인구정체를 감안하면 증가추세가 완연한 지나친 과거 자료 이용은 과도한 미래 예측치를 낳을 수 있다. 따라서 본 연구에서는 과거 1980년대 및 1990년대 자료를 배제하고 2000년 이후의 월별자료를 이용한 것이다.

이러한 지나친 과거 자료와 최근 자료의 차이로 인하여 최근 9년치 자료만을 이용하여 모형을 추정하고 장래 교통수요를 예측하였다. 따라서 먼 장래(예를 들어 30년치)의 교통수요 예측은 하지 못하였다. 이는 본 연구방법론의 한계라기보다는 우리나라 교통실적 자료의 특성에 말미암은 것이다.

향후 본 연구방법론을 다양한 정책 및 실무분야(예를 들어 KDI 예비타당성 조사)에 적용하기 위해서는 좀 더 자료의 축적이 이루어져야 할 것이다. 이러한 점은 본 연구의 한계로 작용할 수 있다.

또한 본 연구방법론을 다양한 분야에서 적용하기 위해서는 자료축적과 이에 따른 장래 예측기간에 대한 연구가 좀 더 이루어져야 할 것이며 이는 연구자들의 향후 과제이기도 하다.

또한 본 연구는 다음과 같은 점에서 한계가 있다. 장래 교통수요 예측에서 계절성을 감안한 SAIRMA 모형을 적용하였지만 총계자료(aggregate data)를 이용하였고 구조변화가 배제된 특정 지역 혹은 노선에 대한 수요를 감안한 것은 아니라는 점이다. 특정 지역 및 노선에 대한 수요예측을 위해서는 해당 지역에 대한 자료를 축적하고 SARIMA 모형을 적용하여 분석해야 할 것이다. 개별 노선별로 SARIMA 모형을 적용하여 이에 대한 분석과 시사점을 제시하지 못한 것은 본 연구의 한계이다.

둘째 SARIMA 모형은 앞에서 지적한 것처럼 외생적인 구조변화를 반영하지 않는 점이다.[12] 이러한 점은 외생변수를 분석에 반영하는 자기회귀 시차모형 혹은 오차수정모형과 대비되는 점이다. 그러나 이들 모형과 달리 SARIMA 모형은 다른 설명변수에 대한 정보 없이도 장래 예측을 수행할 수 있다는 점에서 연구자에게 편리한 점이 존재한다.

그러므로 장래 교통수요 예측에서는 SARIMA 모형과 자기회귀 시차모형 및 오차수정모형의 상호 보완이 필요할 것이다. 즉 자기회귀 시차모형과 오차수정모형을 이용하여 고속철도망 구축 혹은 열차운영 효율화 같은 구조변화를 반영한 장래 예측을 수행하고 이에 대한 보완으로 SARIMA 모형을 적용한 미래 예측치와 비교할 수 있을 것이다. 즉 오차수정모형이나 자기회귀 시차모형이 불확실한 장래 설명변수(예:장래 소득 및 장래 유류가격)에 대한 정보가 필요하여 연구자 입장에서 실제 적용이 어렵다는 점을 SARIMA 모형이 어느 정도 보완할 수 있다.[13] 이러한 점에서 과거 연구들과 비교해 볼 때 본 연구에서 수행한 SARIMA 모형의 차별성이 있을 것이다.

알림 : 본 논문은 연구보고서, “교통정책 지원 및 분석시스템 개발”의 일부 내용을 수정 보완하여 작성한 것입니다.

참고문헌

1. 국토해양부(2010), “국가기간교통망계획 제2차 수정계획 2001~2020”.
2. 이재민 · 한상용 · 이창운(2009), “Oil Price and Travel Demand”, 한국교통연구원.
3. 문진수 · 이재민(2007), “철도화물운송증대를 위한 지원제도 개선방안”, 한국교통연구원.
4. 한국교통연구원(2009), “민자터널 운영관리 개선방안 수립 연구”, 인천광역시.
5. 한국철도공사 · 한국철도시설공단(각년도), “철도통계연보”, 각년도.
6. Bai, J. & P. Perron (1998) “Estimating and Testing Linear Models with Multiple Structural Changes”, Econometrica, 66, pp.47~78.

12) 본 연구는 이와 관련해서 생길 수 있는 잠재적인 문제를 피하기 위해 Bai & Perron (1998, 2003)이 제안한 구조변화 검정방법을 채택하여 검정하였고 그 결과 시계열에 통계적으로 유의한 수준의 구조변화가 존재하지 않음을 확인하였다.

13) 구체적인 방법을 예를 들면 복수의 예측치를 결합하는 방법을 통해 더 나은 예측치를 얻을 수도 있다. 이에 대해서는 Clemen & Winkler (1986)을 참조.

7. Bai, J. & P. Perron .(2003), "Computation and Analysis of Multiple Structural Change Models", Journal of Applied Econometrics, 18, pp.1~22.
8. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994), "Time Series Analysis: Forecasting and Control", Prentice Hall.
9. Clemen, R. T. and R. L. Winkler (1986), "Combining Economic Forecasts", Journal of Business and Economic Statistics, Vol. 4, No. 1, pp.39~46.
10. Dent, W. T. and J. A. Swanson (1978). "Forecasting with Limited Information: ARIMA Models of the Trailer on Flatcar Transportation Market", Journal of the American Statistical Association, Vol. 73, No. 362, pp.293~299.
11. Judge, G. J., R. C. Hill, W. E. Griffiths, H. Lutkepohl, and T. Lee (1998), "Introduction to the Theory and Practice of Econometrics, 2nd edition", Wiley,
12. Lai, S. L. and W. Lu (2005), "Impact analysis of September 11 on air travel demand in the USA", Journal of Air Transport Management, Vol. 11, pp.455~458.
13. Liu, J., S. Wu, and J. V. Zidek (1997), "On Segmented Multivariate Regression", Statistica Sinica, Vol. 7, pp.497~525.
14. Ljung, G. M., and G. E. P. Box (1978), "On a Measure of Lack of Fit in Time Series Models," Biometrika, Vol. 65, pp.297~303.
15. Wooldridge, J. M.(2000), "Introductory Econometrics", South-Western College Publishing, 2000.

♧ 주 작 성 자 : 이재민
♧ 교 신 저 자 : 이재민
♧ 논문투고일 : 2011. 1. 3
♧ 논문심사일 : 2011. 3. 27 (1차)
2011. 5. 25 (2차)
2011. 8. 4 (3차)
♧ 심사판정일 : 2011. 8. 4
♧ 반론접수기한 : 2012. 2. 28
♧ 3인 익명 심사필
♧ 1인 abstract 교정필