

SARIMA 모형을 이용한 우리나라 항만 컨테이너 물동량 예측

민경창¹ · 하헌구^{2*}

¹ 한국해양수산개발원 국제물류연구실, ² 인하대학교 물류전문대학원

Forecasting the Korea's Port Container Volumes With SARIMA Model

MIN, Kyung-Chang¹ · HA, Hun-Koo^{2*}

¹ International Logistics Research Department, Korea Maritime Institute, Seoul 121-915, Korea

² Graduate School of Logistics, Inha University, Incheon 402-751, Korea

Abstract

This paper develops a model to forecast container volumes of all Korean seaports using a Seasonal ARIMA (Autoregressive Integrated Moving Average) technique with the quarterly data from the year of 1994 to 2010. In order to verify forecasting accuracy of the SARIMA model, this paper compares the predicted volumes resulted from the SARIMA model with the actual volumes. Also, the forecasted volumes of the SARIMA model is compared to those of an ARIMA model to demonstrate the superiority as a forecasting model. The results showed the SARIMA Model has a high level of forecasting accuracy and is superior to the ARIMA model in terms of estimation accuracy. Most of the previous research regarding the container-volume forecasting of seaports have been focussed on long-term forecasting with mainly monthly and yearly volume data. Therefore, this paper suggests a new methodology that forecasts short-term demand with quarterly container volumes and demonstrates the superiority of the SARIMA model as a forecasting methodology.

본 연구는 SARIMA 모형을 활용하여 기존에 다루어지지 않았던 분기별 항만 컨테이너 물동량을 예측하였다. 구체적으로 모델 추정에 활용된 자료는 1994년 1사분기부터 2010년 4사분기까지 총 84분기동안의 국내 전체 항만 컨테이너 물동량 자료이다. 본 연구에서 추정된 예측 모형의 예측 정확도를 검증하기 위하여 2011년 1사분기부터 2013년 4사분기까지 물동량을 예측하여 실제 물동량과 비교하였다. 또한 기존에 널리 활용되고 있는 ARIMA 모형을 활용하여 추정된 예측 모형과의 비교를 통해 분기별 항만 물동량 예측에 있어서 SARIMA 모형의 상대적 우수성을 검증하였다. 기존에 항만 물동량을 예측하는 대부분의 연구는 주로 장기 예측에 초점이 맞추어져 있다. 또한 월별, 연도별 물동량 자료가 활용된 경우가 대부분이다. 분기별 항만 컨테이너 물동량 자료를 활용하여 단기 수요를 예측함과 동시에 SARIMA 모형의 우수성을 입증한 본 연구는 충분한 가치가 있다고 판단된다.

Keywords

ARIMA, container volume forecasting, container volume through sea port, demand forecasting, seasonal ARIMA, seasonal unit-root test, unit-root test

ARIMA, 컨테이너 물동량 예측, 항만 컨테이너 물동량, 계절성 ARIMA(SARIMA), 수요예측, 계절단위근 검정, 단위근 검정

* : Corresponding Author

hkha@inha.ac.kr, Phone: +82-32-860-8232, Fax: +82-32-860-8226

Received 8 May 2014, Accepted 3 October 2014

서론

1. 연구의 배경 및 목적

우리나라의 전체 수출입 물동량은 지난 20여년동안 연평균 약 5.3%씩 증가하여왔다¹⁾. 이 가운데 약 99.9% 가량의 화물이 여전히 해상을 통해 운송되고 있는 것으로 나타났다. 우리나라의 주요 수출 품목들이 반도체, 휴대폰 등 작은 부피의 고부가가치 화물들로 변화함에 따라 항공운송의 중요성이 전보다 중요해지고 있는 것은 분명한 사실이지만, 자동차, 가전제품 등이 여전히 우리나라의 주요 수출품으로의 역할을 수행하고 있고, 또한 전체 수송량의 99.9% 가량을 담당하고 있는 해상운송의 특성 상, 현재까지와 마찬가지로 향후에도 해상운송이 우리나라의 교역에서 높은 비중을 차지할 것은 매우 자명하다고 할 수 있다.

그동안 국내외에서는 원활한 해상운송을 위하여 항만 분야에 대한 많은 연구들이 다각도로 진행되어져 왔다. 이러한 연구 결과물들이 다양한 경로를 통해 해상운송의 선진화에 이바지 하여왔고, 또한 관련 정부정책과정에 반영되어왔다. 일반적으로 많은 비용과 시간이 소요되는 항만산업의 특성을 감안하였을 때, 정확한 수요예측에 대한 연구는 그 무엇보다 중요한 연구 분야라 할 수 있다. 항만은 중요한 국가 기간산업 가운데 하나로써 적절한 시점에 적절한 규모로의 증설 및 증축을 통해 항만시설의 부족, 혹은 과잉 공급현상을 사전에 방지해야 한다. 이러한 증설 및 증축 계획은 수요예측 결과를 바탕으로 세워지게 되는데 만약 잘못된 수요예측 결과를 바탕으로 항만계획이 수립되었을 경우, 매우 큰 경제적 손실을 초래하게 된다.

그동안 항만분야 수요예측과 관련하여 많은 연구들이 진행되어져왔다. 그러나 현재까지 진행되어진 대부분의 항만분야 수요예측은 장기예측에 초점이 맞추어져 왔다. 항만분야 수요예측이 주로 항만 내 하드웨어 확장 등을 위함인 예측의 주요 목적이었기 때문이다. 이러한 이유로 대부분의 항만분야 수요예측은 5년, 혹은 10년 단위의 장기예측이 주로 이루어졌고, 경우에 따라서 1년 단위로 장기간의 수요예측들이 진행되어왔다. 상대적으로 오랜 시간과 많은 비용이 소모되는 항만시설에 대한 투

자를 생각한다면 선행적으로 장기간의 수요를 예측한 후, 그 예측된 결과를 바탕으로 시설 부족 현상이 발생하기 전에 시설 투자가 선행되어야 하는 것은 너무나도 분명한 사실이다. 하지만 장기적인 관점이 아니라, 단기적인 관점에서의 수요예측 역시 항만운영에 있어서 매우 중요한 요소 가운데 하나이다. 장기 수요예측이 하드웨어 측면에서 도움이 된다면, 단기 수요예측은 항만의 소프트웨어, 즉 단기 항만 운영계획에 직접적으로 도움이 될 수 있기 때문이다. 가령, 단기 수요예측을 통해 인력투입 계획, 시설사용계획 등의 최적화를 통해 단기간의 항만 운영 효율성을 높일 수 있다. 또한 이러한 과정을 통해 해운업체와 같은 수요자들에게 보다 높은 만족도를 제공할 수도 있다. 이러한 장점에도 불구하고, 항만분야에 대한 장기 수요예측에 비해 단기 수요예측은 그동안 크게 주목받아오지 못하였다.

Yi(2013)은 부산항의 월별 물동량 자료(120개월)를 바탕으로 SARIMA(Seasonal Autoregressive Integrated Moving Average) 모형을 활용하여 부산항의 물동량을 추정하였다. 추정된 SARIMA 모형을 바탕으로 향후 물동량(96개월)을 예측하였으나, 예측한 물동량이 실제 물동량을 얼마나 정확하게 예측하였는지에 대한 검증이 이루어지지 않았다.

Kim et al.(2011)는 국내 항만의 월별 물동량 자료(252개월)를 바탕으로 SD(System Dynamics)기법과 SARIMA 모형을 활용, 향후 10년간의 물동량을 추정하여 국토해양부에서 추정한 예측자료와 비교하였다. 해당 연구에서는 SD 기법을 활용한 자료가 국토해양부 예측자료와 비슷하다는 판단 하에 SD기법이 SARIMA 기법보다 보다 효과적인 예측방법이라고 정의하였다.

Chen et al.(2010)은 대만 내 주요 항만들의 월별 물동량 자료(312개월)를 활용하여 유전자 프로그래밍 기법(Genetic programming), 분해분석법(Decomposition Approach), SARIMA 모형 각각의 예측 모형을 추정하였다²⁾. 그리고 추정된 각각의 예측 모형을 활용하여 예측한 12개월 기간에 대한 예측 정확도를 비교함으로써 월간유전자 프로그래밍 기법이 월별 물동량 예측에서 가장 우수한 예측 기법인 것으로 판단하였다.

Peng(2009)는 분해분석법, 회귀분석(regression mode

1) 무역협회, '우리나라의 수출입 물동량 현황 자료' 참고

2) 분해분석법 모형의 경우, 해당 연구에서 추정하지 않고 기존에 대만 내 항만관련 부처에서 추정하였던 모형을 활용

1)3), Grey 모형(Grey Model)4), 그리고 SARIMA 모형 간의 예측정확도 비교를 위하여, 대만 내 주요 항만들의 월별 물동량 자료(24개월)를 활용하여 예측모형을 추정 한 후, 각 모형을 활용하여 12개월에 대한 예측을 통해 각 모형간의 예측 정확도를 비교하였다. 그 결과, 회귀분석을 가장 적합한 예측모형인 것으로 제시하였다.

Shin et al.(2008)는 ARIMA모형과 Artificial Neural Network모형을 결합한 하이브리드 모형을 활용하여 부산항, 광양항, 인천항 울산항, 마산항의 수출입 컨테이너 물동량을 예측하였다. 예측모형 추정에는 1997년부터 2004년까지 총 96개월의 자료를 활용하였고, 2005년부터 2006년까지에 물동량을 예측하여 실제 물동량과의 비교를 통해 하이브리드 모형의 예측 우수성을 검증하였다.

Chou et al.(2007)은 1989년부터 2001년까지 13년간 대만의 수입 컨테이너 물동량을 바탕으로 회귀모형들 간5)의 모형의 적합도를 검증하였다. 해당 연구에서는 모형간의 적합도만을 비교하였고, 실제 모형을 통한 예측 물동량들과 실제 물동량간의 예측 정확도를 비교하지는 않았다.

Kim(2007)은 1980년 1월부터 2006년 12월까지 총 324개월 간의 국내 항만물동량 자료를 바탕으로 SARIMA 모형을 활용 2007년부터 2012년까지의 항만 물동량을 예측하였다. 하지만, 연구시점(2007) 상 예측한 물동량과 실제 물동량간의 비교가 불가하여 실제 예측 정확도를 검증하지는 못하였다.

지금까지 주로 진행되어온 단기 항만 수요예측은 주로 월별 데이터를 활용한 경우가 대부분이었다. 방법론 측면에서는 시계열 분석기법들과 회귀모형 등이 다양하게 활용되어 왔다. 하지만 대부분의 연구가 예측 모형을 추정한 이후, 주로 모형의 적합도를 바탕으로 우수성을 검증하였을 뿐 실제 추정한 예측 모형을 바탕으로 예측 이후 시점에 대한 물동량을 예측한 후 실제 물동량과의 예측정확도를 비교함으로써 예측 모형에 우수함을 검증하는 연구는 많이 진행되지 않았다. 또한 다양한 예측방법이 활용되었음에도 불구하고 근래 가장 많이 활용되어지고 있는 시계열 예측기법 가운데 하나인 ARIMA 모형과 SARIMA 모형에 대한 직접적인 예측 정확도 비교를 다른 연구 역시 찾아볼 수 없었다.

본 연구에서는 단기 항만 물동량 예측 모형을 추정하는데 있어 기존에 활용되어지지 않았던 분기별 항만물동량 자료를 활용하였다. 구체적인 기간으로는 1990년 1사분기부터 2010년 4사분기까지의 분기별 자료이고 자료는 국내 전체 항만 컨테이너 물동량 자료를 활용하였다. 또한 본 연구에서 항만 컨테이너 수요예측에 사용된 분석방법론은 시계열분석 기법(Time Series Analysis)이다. 그 중에서도 항만 컨테이너 물동량이 계절적(Seasonality)인 특성을 지닌 것을 감안하여, 계절성 혹은 주기성(Periodicity)을 지닌 시계열 자료를 예측하는데 적합한 Seasonal Autoregressive Integrated Moving Average 기법, 즉, SARIMA 방법론을 활용하였다.

앞서 언급한 바와 같이 그동안 항만 수요예측에 있어서, 장기 수요예측에 비하여 단기 수요예측은 그 중요성에 비해 상대적으로 많이 이루어지지 않았다. 또한 기존에 이루어진 단기 수요예측에 관한 연구들 가운데 분기별 자료를 활용하여 분기별 수요예측을 수행한 경우는 전무하다시피하다. 본 연구에서는 분기별 항만 컨테이너 물동량을 예측하는데 있어서 SARIMA 모형의 활용가능성 및 정확성을 검증하였다. 구체적으로는 국내 전체 항만의 컨테이너 물동량 자료를 활용하여 분기별 항만 물동량을 예측하여 SARIMA 모형의 활용가능성을 살펴보았다. 본 연구의 결과를 바탕으로 전체 항만의 물동량 자료에 대한 예측뿐만 아니라, 향후 개별 항만의 단기 물동량을 예측하는데 SARIMA 모형이 활용되기를 기대한다. 또한 본 연구에서 활용된 방법론 및 항만 분야에 대한 단기 수요예측 결과물이 단기적 차원에서의 개별 항만 운영 효율성 향상 및 전체 항만에 대한 단기 항만 정책 수립에 활용되어지기를 기대한다.

방법론

1. SARIMA 모형

SARIMA(Seasonal Autoregressive Integrated Moving Average) 모형은 대표적인 Box-Jenkins 방법론 가운데 하나인 ARIMA(Autoregressive Integrated Moving Average) 모형을 보완한 예측기법이다. SARIMA 방법론 역시 ARIMA 방법론과 동일하게 Y_t 는 설명변수

3) The trigonometric regression model, the regression model with seasonal dummy variables

4) The grey model, and the hybrid grey model

5) The traditional regression Model, and The modified regression model

X_n 에 의해 설명된다.’는 인과모형이 아니라 ‘ Y_t 는 설명 변수 X_n 이 아닌 자신의 과거 값 Y_{t-n} 및 확률적 오차항 ϵ_t 에 의해 설명된다.’는 시계열의 확률적 특성에 기반을 둔 예측기법이라 할 수 있다.

SARIMA(p,d,q)(P,D,Q)s 모형은 식(1)과 같다.

$$\phi_p(L)\Phi_P(L^s)(1-L)^d(1-L^s)^D Z_t = \delta + \theta_q(L)\Theta_Q(L^s)\epsilon_t \tag{1}$$

- Z_t : 원시계열 자료
- t : 시간을 나타내는 연산자
- ϵ_t : $N(0, \sigma^2)$ 을 따르는 오차항, 백색잡음
- L : 후향연산자
- P : SAR(Seasonal Autoregressive)항의 차수
- Q : SMA(Seasonal Moving Average)항의 차수
- D : 계절차분의 차수
- p : AR(Autoregressive)항의 차수
- q : MA(Moving Average)항의 차수
- d : 차분의 차수
- δ : 상수

식(1)에서 $\Phi_P(L^s)$ 과 $\Theta_Q(L^s)$ 는 각각 계절자기회귀(Seasonal Autoregressive : SAR)항과, 계절이동평균(Seasonal Moving Average: SMA)항을 의미하고, $\phi_p(L)$ 는 자기회귀(Autoregressive: AR)항, 그리고 $\theta_q(L)$ 는 이동평균(Moving Average: MA)항을 의미한다. SAR항과 SMA항, 그리고 AR항과 MA항은 각각 식(2)와 같이 나타낼 수 있다.

$$\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps} \tag{2}$$

$$\Theta_Q(L^s) = 1 - \theta_1 L^s - \theta_2 L^{2s} - \dots - \theta_Q L^{Qs}$$

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

SARIMA 모형은 앞서 언급한바와 같이 ARIMA 모형을 보완하여 개발된 기법이다. SARIMA을 이해하기 위해서는 ARIMA 모형과의 비교를 통해 어떠한 형태로 보완이 되었는지를 확인할 필요가 있다.

ARIMA(p,d,q) 모형은 식(3)과 같다.

$$\phi_p(L)(1-L)^d Z_t = \delta + \theta_q(L)\epsilon_t \tag{3}$$

식(1)과 식(3)를 비교하여 보면, SARIMA모형은 ARIMA모형에 SAR항과 SMA항이 추가된 형태인 것을 알 수 있다. 즉, SARIMA 모형은 기존 ARIMA 모형에서 SAR항과 SMA항을 추가함으로써, 시계열자료의 계절성, 혹은 주기성을 감안할 수 없는 ARIMA 모형의 한계점을 극복할 수 있는 것이다⁶⁾. 만약 SARIMA 모형에서 SAR항과 SMA항의 차수, 그리고 계절차분의 차수가 모두 ‘0’일 경우, 즉, P 와 Q , 그리고 D 의 값이 모두 ‘0’일 경우, 모델은 ARIMA(p,d,q)와 같아지게 된다.

시계열 자료를 SARIMA, ARIMA 기법과 같은 Box-Jenkins기법에 적용하기 위해서는 사전적으로 시계열자료가 안정적(Stationary)인지 여부를 반드시 확인해야 한다. 시계열자료가 안정적일 경우에만 Box-Jenkins 기법에 적용시킬 수 있는데, 이러한 조건을 시계열자료의 ‘안정성 조건’이라고 한다. 만약 시계열이 안정적이지 않다면 시계열의 평균과 분산 값이 시점의 변화에 따라 변화한다는 것을 의미하는데, 그럴 경우 그 예측 값이 유효하지 않기 때문이다. 그러므로 시계열 자료를 Sarima 기법에 적용하기에 앞서 반드시 안정성 여부를 확인해야 하고, 만약 시계열자료가 안정적이지 않다면, 차분(Difference)의 과정을 통해 반드시 자료를 안정적으로 변환시킨 후 모델에 적용해야만 한다.

2. 모형 추정 절차

예측하고자 하는, 즉 예측에 활용할 시계열 자료를 가지고 SARIMA 기법을 활용하여 장래 수요를 예측하기 위해서는 Figure 1에 나타난 바와 같이 식별(Identification), 추정(Estimation), 진단(Diagnostic), 예측(Forecasting)의 4단계를 거치게 된다.

식별(Identification) 단계에서는 주어진 시계열 자료를 바탕으로 시계열이 계절적 특성을 지니고 있는지, 시계열은 안정적인지 여부를 판단하게 된다. 시계열자료가 계절적 특성을 갖는지 여부를 판단하여 SARIMA 기법을 활용할지 여부를 판단하게 되고, 다음으로는 안

6) SARIMA 모형은 기존 ARIMA 모형에서 SAR항과 SMA항을 추가함으로써, 예측시점을 기준으로 가까운 시점의 과거 자료 뿐 아니라, 시계열 자료의 주기적인 특성을 감안하여 그 이전 주기의 자료를 추가적으로 활용할 수 있기 때문에, 최근의 정보와 이전 주기의 정보를 모두 반영하는 특성을 갖게 된다.

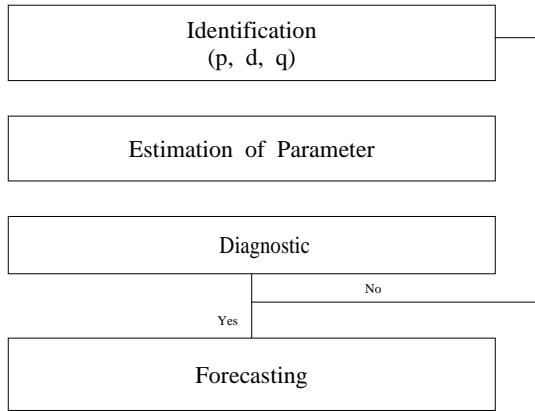


Figure 1. Flow chart of SARIMA

정성 여부를 판단하여 시계열 자료를 안정적으로 변환시켜야 할지를 결정하게 된다. 이 과정에서 SARIMA(p,d,q) (P,D,Q)s 모형 가운데 'd'와 'D', 그리고 's'의 값을 결정하게 된다. 이어서 시계열 자료의 상관도표(Correlogram)로 확인 가능한 자기상관함수(Autocorrelation Function: ACF)와 편자기상관함수(Partial Autocorrelation Function: PACF)를 이용하거나, 혹은 DF-GLS 등의 방법을 통해 나머지 'p'(AR의 차수)와 'q'(MA의 차수), 그리고 'P'(SAR의 차수)와 'Q'(SMA의 차수)의 값으로써 가능한 값들을 추정하게 된다.

추정(Estimation) 단계에서는 앞서 식별단계를 통해 알아낸 'p', 'd', 'q', 'P', 'D', 'Q'의 값들 가운데, 통계적으로 가장 적절한 값을 결정하게 된다. 이와 함께 'p', 'd', 'q', 'P', 'D', 'Q'의 값이 정해지면서 결정되는 AR, MA, SAR, SMA 각 항들에 대한 모수를 추정하게 된다. 각 항들의 모수를 추정하는 대표적인 추정법으로는 최대우도 추정법(Maximum Likelihood Estimation Method), 최소제곱 추정법(Least Squares Estimation Method)이 있고, 이외에 적률 추정법(Method of Moment Estimation), 비선형 추정법(Nonlinear Estimation Method) 등이 있다.

진단(Diagnostic)의 단계는 앞서 추정의 단계에서 추정된 모형에 대한 통계적 진단을 하게 되는 단계이다. 추정의 단계에서 모형 추정에 필요한 세부적인 요소들에 대한 통계적인 검증이 이루어졌다면, 본 단계에서는 앞서 추정된 모형 전체적인 관점에서 통계적으로 유의한지 여부와 모형이 과연 예측모형으로써 적절한지 여부를 판

단하게 된다. 진단의 단계에서는 주로 잔차분석(Residual Analysis), 과적합진단(Overfitting Diagnostics) 등을 통해 전체 모형을 진단하게 된다. Figure 1에 나타난 바와 같이 진단 과정에서 추정 모형이 적합하지 않은 것으로 나타나면 다시 식별의 단계로 돌아가 다시 모형을 추정하게 되고, 만약 적합한 것으로 나타난다면 다음 단계인 예측단계로 넘어가 예측모형으로써 활용되어지게 된다.

마지막 예측(Forecasting)은 수요예측의 최종 단계로서, 진단의 단계를 통해 선택된 최종모형을 활용하여 장래수요를 예측하게 된다. 이렇게 예측된 수요는 각 예측목적에 맞추어서 다양한 분야에 활용되어지게 될 것이다. 또한 예측기간이 지난 이후 예측치와 실측치간의 비교를 통해 예측모델의 정확도를 검증할 수 있고, 이를 바탕으로 향후 추가적인 예측에 그 결과를 반영하여 참고하게 된다.

모형 추정

본 연구에 사용된 자료는 국내 전체 항만의 분기별 컨테이너 물동량 자료로서, 해양수산부 해운항만물류정보센터(SPIDC)⁷⁾에서 제공하고 있는 월별 컨테이너 물동량 자료를 분기별 자료로 통합한 자료이다. 구체적인 기간으로는 1994년 1사분기부터 2010년 4사분기까지 총 68분기이다.

연구에 사용된 시계열 자료는 본격적인 분석 과정에 앞서 자연로그를 취한 후, 활용되었다.

1. 계절성 확인

앞서 SRAIMA 모형에 대한 설명에서 언급한 바와 같이 SARIMA 방법론은 계절성을 지닌 시계열 자료의 정확한 수요예측을 위해 개발되어진 방법론으로서 기존의 널리 활용되고 있는 시계열 방법론 가운데 하나인 ARIMA 모형에 계절성(주기적 특성)을 추가적으로 감안한 모형이다. 그러므로 만약 시계열 자료가 계절성을 보이지 않는다면 굳이 복잡한 SARIMA 방법론을 활용할 필요 없이 ARIMA 방법론을 적용하여 예측하면 된다.

본 연구에서 예측할 대상은 국내 전체 항만 컨테이너 물동량이다. 일반적으로 항공이나 해상화물의 수요는 계

7) www.spidc.go.kr

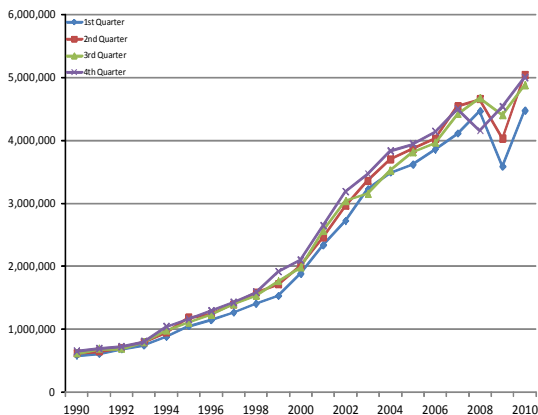


Figure 2. Comparison between each quarterly data

절적 특성을 지니고 있는 것으로 알려져 있다. SARIMA 모형을 적용하기에 앞서 예측 대상인 국내 전체 항만 컨테이너 물동량 시계열 자료가 계절성을 지니고 있는지 여부를 판단하였다.

Figure 2는 본 연구에서 예측모형에 추정에 사용된 전체 시계열의 분기별 자료를 년도별로 나타낸 도표이다. Figure 2를 살펴보면 국내 전체 항만 컨테이너 물동량은 각 년도별로 4사분기의 물동량이 가장 많고 1사분기의 물동량이 가장 적은 것으로 나타나고 있는 특성을 보이며 일정한 계절성을 나타내고 있다. 2008년의 경우에는 예외적으로 1사분기의 물동량이 4사분기의 물동량보다 높은 것으로 나타나고 있는데 이는 2008년 4분기에 발생한 글로벌 금융위기의 결과인 것으로 판단된다.

추가적으로 Figure 3는 본 연구에서 사용된 분기별 국내 항만 컨테이너 물동량 자료의 계절지수(Seasonal Index)를 나타낸 도표이다. Figure 3의 분기별 계절지수를 살펴보면, 앞서 Figure 2에서 살펴본 결과와 같이 1사분기와 4사분기의 물동량 차이가 약 10%이상 나타나고 있는 것으로 나타나고 있다. 이러한 계절지수를 바탕으로, 국내 항만 컨테이너 물동량은 1사분기 비수기이고 4사분기가 성수기인 계절적 특성을 지니고 있다고 판단되어진다.

그러므로 국내 전체 항만 컨테이너 물동량을 예측하는 본 연구의 특성 상, SARIMA 방법론을 예측기법으로 활용하는 것이 적합하다고 판단할 수 있다.

2. 안정성 검정

본 연구에서는 시계열 자료의 안정성 여부를 확인하

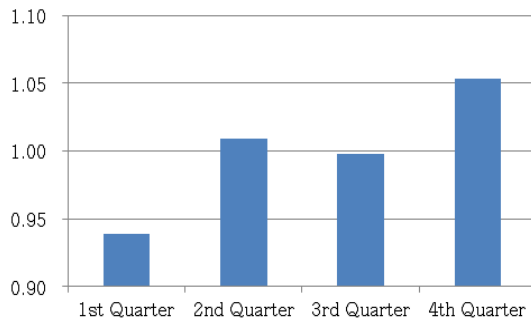


Figure 3. Seasonal index of time series data

기 위해 상관도표 분석과 단위근 검정(Unit root test)을 활용하였다. 또한 시계열 자료가 분기별 자료인 것을 감안하여 Hylleberg-Engle-Granger-Yoo(HEGY) 검정(1990)을 활용하여 시계열 자료의 계절 단위근 여부를 확인하였다. 본 연구에서 상관도표 분석과 더불어서 시계열 자료의 안정성 검정을 하는데 활용한 ADF(Augmented Dickey-Fuller) 검정은 가장 대표적인 단위근 검정기법 가운데 하나로써, 기존의 Dickey-Fuller 검정을 보완한 검정 기법이다. ADF 검정은 주어진 시계열 자료에 대하여 임의행보(Random walk)과정에서 표류경향(Drift)이 있을 경우와 없을 경우, 그리고 확정적(Deterministic) 추세와 확률적(Stochastic) 추세가 모두 존재할 경우에 대하여 검정을 수행하게 된다.

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \tag{4}$$

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \tag{5}$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + u_t \tag{6}$$

식(4)는 표류경향(Drift)이 없는 임의행보(Random walk)모형, 식(5)는 표류경향이 있는 임의행보모형, 그리고 식(6)은 확정적 추세가 확률적 추세가 모두 존재하는 임의행보모형을 의미한다. 시계열 자료의 안정성 조건을 만족시키기 위해서는 위의 3가지 경우에서 모두 시계열 자료가 안정적인 것으로 나타나야한다. 1가지의 경우에서라도 시계열 자료가 불안정적(단위근이 존재)인 것으로 나타난다면 차분 등의 과정을 통하여 자료를 반드시 안정적으로 변환시킨 후, 예측기법에 적용시켜야 한다.

Table 1. Correlogram of raw data

Lag	AC	PAC	Q	Prob.>Q	Autocorrelation			Partial Autocorr.		
					-1	0	1	-1	0	1
1	0.9679	0.9901	81.536	0.000						
2	0.9389	0.3356	159.19	0.000						
3	0.9061	-0.1159	232.42	0.000						
4	0.8786	0.3338	302.13	0.000						
5	0.8451	-0.3238	347.43	0.000						
6	0.8144	0.1378	428.85	0.000						
7	0.7847	-0.0446	486.62	0.000						
8	0.7584	0.0484	541.3	0.000						
9	0.7250	-0.1402	591.93	0.000						
10	0.6904	0.1333	638.46	0.000						
11	0.6523	0.0019	680.57	0.000						
12	0.6168	-0.0294	728.74	0.000						
13	0.5784	-0.2517	752.78	0.000						
14	0.5427	-0.0481	783.18	0.000						
15	0.5038	0.0955	809.76	0.000						

* After log-transform

Table 2. Result of ADF unit-root test (raw data)

	tau value	p > t	lags
Random Walk	2.948	0.9999	4
Random Walk with Drift	-1.543	0.5123	4
Random walk with Drift and Trend	-0.992	0.9452	4

* founded Lag value from DF-gls test (by STATA 11)

Table 3. Correlogram of regular and seasonal differencing

Lag	AC	PAC	Q	Prob.>Q	Autocorrelation			Partial Autocorr.		
					-1	0	1	-1	0	1
1	0.0626	0.0626	0.3213	0.5708						
2	-0.0334	-0.0394	0.41408	0.8130						
3	-0.1672	-0.1737	2.7668	0.4290						
4	-0.4077	-0.4598	16.949	0.0020						
5	0.0397	0.0633	17.086	0.0043						
6	0.0557	0.0595	17.358	0.0081						
7	0.0322	-0.1437	17.45	0.0147						
8	-0.0902	-0.4125	18.183	0.0199						
9	-0.0728	-0.0667	18.667	0.0282						
10	0.0242	0.0299	18.721	0.0439						
11	0.0886	0.0591	19.459	0.0533						
12	0.1075	-0.0538	20.564	0.0571						
13	0.0556	0.1583	20.893	0.0757						
14	-0.1253	-0.2019	22.41	0.0706						
15	-0.0011	0.0512	22.41	0.0975						

* After log-transform

Table 1은 분기별 국내 항만 컨테이너 물동량 자료의 상관도표이다. Table 1을 살펴보면 자기상관계수가 1차에서 매우 높은 값(0.9679)에서 시작해서 15차 시점까지 높은 값(0.5038)을 유지하고 있는 것으로 보아 시계열자료가 불안정적인 것으로 판단된다. Table 2는 같은

시계열 자료에 대한 ADF 검정 결과이다. DF-gls 검정을 통해 추정된 최적의 lag 값(4)을 활용하여 ADF 검정을 수행한 결과, 매우 높은 유의수준에서도 임의행보(Random walk) 모형에서 표류경향(Drift)이 있는 경우와 없는 경우, 그리고 확정적(Deterministic) 추세와

확률적(Stochastic) 추세가 있는 경우에서 모두 '시계열 자료가 단위근을 갖는다.'는, 즉, '시계열이 불안정적이다.' 라는 귀무가설을 기각할 수 없는 것으로 나타났다. 그러므로 아무런 변환을 하지 않은 원 시계열 자료, 즉 분기별 국내 항만 컨테이너 물동량 자료는 안정적이지 않다고 할 수 있다.

Table 3은 분기별 국내 항만 컨테이너 물동량 자료를 1계 분기 차분 및 1계 계절 차분을 통해 변화한 시계열 자료의 상관도표이다. Table 3을 살펴보면 자기상관이 3차 시차까지 0주변에서 머물고 있는 것으로 보아시계열 자료가 안정적인 것으로 판단할 수 있다. 추가적으로 ADF검정을 시행하기 위하여 최적의 Lag(시차)값을 추정하기 위하여 DF-gls 검사를 수행한 결과, Table 4에 나타난 바와 같이 최적의 Lag값이 3인 것으로 나타났다. 이렇게 추정된 lag 값을 바탕으로 ADF 검정을 시행한 결과인 Table 5를 살펴보면 임의행보모형에서표류경향이 있는 경우와 없는 경우, 그리고 확률적 추세와 확률적 추세가 있는 모든 경우에 있어서 유의수준 1% 이하에서 '시계열 자료가 불안정적이다.'라는 귀무가설을 기각하고 있음을 알 수 있다. 그러므로 1계 분기 차분 및 1계 계절 차분을 한 시계열 자료는 안정적인 것으로 판단할 수 있다.

본 연구에 활용되는 시계열 자료가 분기별 자료인 것을 감안하여 앞서 상관도표 및 ADF 검정결과 안정적으로 나타난 1계 분기 차분 및 1계 계절 차분된 시계열 자료에 대하여 추가적으로 계절 단위근이 존재하는지 여부를 판단하기 위해 HEGY 검정을 실시하였다. HEGY 검정을 수행하기 위해서는 우선 SARIMA 모형 내 계절성 부분을 다음과 같이 변환하여 진행하게 된다.

$$\begin{aligned} (1-L^4) &= (1-L)(1+L+L^2+L^3) & (7) \\ &= (1-L)(1+L)(1+L^2) \\ &= (1-L)(1+L)(1-Li)(1+Li) \end{aligned}$$

이와 같이 계절성 부분을 변환시킨 후, 진행하게 되는 HEGY 검정의 기본적인 형태는 식(8)과 같다.

$$\Phi(L)y_{4,t} = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + u_t \quad (8)$$

$$\begin{aligned} y_{1,t} &: (1+L+L^2+L^3)y_t \\ y_{2,t} &: -(1-L+L^2-L^3)y_t \\ y_{3,t} &: -(1-L^2)y_t \\ y_{4,t} &: (1-L^4)y_t \end{aligned}$$

시계열의 각 주기에 따른 단위근의 존재 여부는 식(8)의 $\pi_1, \pi_2, \pi_3, \pi_4$ 의 추정치들을 통해 판별할 수 있다. 만약 $\pi_1 = 0$ 이라면 비계절주기(영빈도(zero frequency), $L=1$)에서 단위근이 존재하고, $\pi_2 = 0$ 이라면 반년주기($L=-1$)에서 단위근이 존재한다고 판단할 수 있다. 또한 $\pi_3 = \pi_4 = 0$ 이라면 1년주기($L=\pm 1$)에서 단위근이 존재한다고 역시 추론이 가능하다.⁸⁾ 즉, $H_0: \pi_1 = 0, H_0: \pi_2 = 0, H_0: \pi_3 = \pi_4 = 0$ 에 대한 가설검정을 통하여 단위근의 존재여부를 확인하게 된다. 가설검정을 통해 만약 ' $\pi_1 \neq 0$ ', ' $\pi_2 \neq 0$ ', ' $\pi_3 \neq 0$ & $\pi_4 \neq 0$ '이라면, y_t 는 정상적인 시계열이라고 판단할 수 있다.

Table 6는 1계 분기 차분 및 4계 계절 차분된 시계열 자료의 HEGY검정 결과이다. HEGY검정 결과, 표

Table 4. Result of DF-gls test (data after regular and seasonal differencing were applied)

lags	tau value	1% critical value	5% critical value	10% critical value
3	-6.926	-3.660	-3.036	-2.746

* founded Lag value from DF-gls test (by STATA 11)

Table 5. Result of ADF unit-root test (data after regular and seasonal differencing were applied)

	tau value	p > t	lags
Random Walk	-7.580	0.000	3
Random Walk with Drift	-7.523	0.000	3
Random walk with Drift and Trend	-7.469	0.000	3

* founded Lag value from DF-gls test (by STATA 11)

8) S. Hylleberg, R.F. Engle and C.w.J. Granger, and B.S. Yoo, Seasonal Integration And Cointegration, Journal of Econometrics 44(1990) 215-238. North-Holland

Table 6. Result of HEGY test (data after regular and seasonal differencing were applied)

	Frequency			lags
	Zero ($t: \pi_1 = 0$)	Biannual ($t: \pi_2 = 0$)	Annual ($F: \pi_3 = 0 \cap \pi_4 = 0$)	
No-constant	-5.433 (0.0000)	-5.911 (0.0000)	47.280 (0.0000)	1
Constant	-5.400 (0.0000)	-5.872 (0.0000)	46.699 (0.0000)	1
Constant & Trend	-5.391 (0.0000)	-5.833 (0.0000)	46.229 (0.0000)	1

* founded Lag value, AIC and BIC from DF-gls test (by STATA 11)

류경향이 있는 경우와 없는 경우, 그리고 계절더미(추세)가 있는 경우에서 모두 1% 유의수준 이내에서 귀무가설이 기각되고 있음을 알 수 있다. 이러한 결과를 바탕으로 1계 분기 차분 및 4계 계절 차분된 시계열 자료는 계절 단위근을 갖지 않고 안정적이라고 판단할 수 있다.

3. 모형 추정

SARIMA(p,d,q)(P,D,Q)_s 모형의 최종형태는 'p', 'd', 'q', 'P', 'D', 'Q'의 값에 따라 결정되게 된다. 그러므로 SARIMA 모형의 추정은 곧 'p', 'd', 'q', 'P', 'D', 'Q'의 값을 추정해내는 것이라고 할 수 있다. 앞서 안전성 검정과정을 통해 1계 분기 차분 및 1계 계절 차분된 시계열 자료가 안정적인 것으로 판단하였다. 그러므로 'd'와 'D'의 값은 모두 '1'로 결정되었다. 다음으로는 AR항과 MA항, 그리고 SAR항과 SMA항의 차수인 'p', 'q', 'P', 'Q'의 값을 추정하여야 한다. 우리는 앞서 상관도표 분석과 DF-gls 검정을 통해 잠정적으로 'p=3', 'q=3', 'P=1', 그리고 'Q=1'의 값을 추정해낼 수 있었다. 이렇게 추정한 각 항의 잠정적인 값들과 더불어 '추정모형의 설명력이 아무리 우수하더라도 AR항과 MA항, 그리고 SAR항과 SMA항의 차수 값이 너무 클 경우, 모델의 실용성 및 활용성이 떨어지기 때문에 모형은 가능한 최대한 간결해야 한다.'는 간결성의 원칙을 감안하여 'p,q,P,Q' 각각의 범위를 다음과 같이 정의할 수 있다.⁹⁾

$$\begin{aligned}
 p &: 0 \leq p \leq 3 & (9) \\
 q &: 0 \leq q \leq 3 \\
 P &: 0 \leq P \leq 1 \\
 Q &: 0 \leq Q \leq 1
 \end{aligned}$$

이렇게 결정된 'p','q','P','Q' 각각의 범위를 바탕으로 'p','q','P','Q'의 최종적인 값을 추정하게 된다. 식(9)의 범위를 만족시킬 수 있는 모형은 총 63개가 존재한다.¹⁰⁾ 총 63개의 모형 가운데 최적 모형은 일반적으로 다음의 3가지 과정을 통해 결정하게 된다.

첫 번째로, 해당 SARIMA모형이 얼마나 시계열을 잘 설명하고 있는지 여부를 확인함과 동시에, AR항, MA항, SAR항 그리고 SMA항의 계수로 추정된 값들에 대한 유의성 검정을 실시해야한다. 이를 통해 각 모수들에 대한 통계적 유의성 여부와 더불어 모형 자체에 대한 적절성 여부를 판단하게 된다.

두 번째로, Akaike Information Criterion(AIC)와 Bayesian Information Criterion(BIC) 값을 살펴보아야 한다.

AIC와 BIC의 추정식은 식(10)과 같다.

$$\begin{aligned}
 AIC &= -2 \times \ln(\text{likelihood}) + 2 \times k & (10) \\
 BIC &= -2 \times \ln(\text{likelihood}) + \ln(N) \times k
 \end{aligned}$$

k : 모수의 수

N : 관측수

AIC와 BIC의 추정식을 살펴보면 AIC와 BIC값은 우도값, 즉 모형의 설명력이 높을수록 작아지고, 반대로 모형 내 모수의 수가 많아질수록, 즉 모형이 복잡해질수록 높아진다는 것을 알 수 있다. 그러므로 각 AIC와 BIC는 절대 값이 높을수록 적절한 모형이라 할 수 있다.¹¹⁾

마지막으로 감안해야 할 사항이 모델의 간결성이다. 앞서 'p','q','P','Q'값의 범위를 결정할 때 언급한바와 같

9) Min K. C., Jun Y. I., Ha H. K.(2013), Forecasting the Air Cargo Demand With Seasonal ARIMA Model: Focusing on ICN to EU Route, J. Korean Soc. Transp., 31(3), Korean Society of Transportation, 3-18.

10) 전체 64개 모형가운데 SARIMA(0,1,0)(0,1,0)₄ 모형은 제외

11) 모수의 개수가 많아진다는 것은 '항(term)'의 수가 증가함을 의미하고, '항'의 수가 증가할수록 '우도값(likelihood)값이 높아지는 성향이 있다. 그렇기 때문에 '모수의 수'와 '우도값' 사이에는 trade-off 관계가 형성된다. 이런 의미에서 볼 때, AIC와 BIC의 값은 이러한 trade-off 관계를 감안해서 모델을 선택할 수 있게 해준다.

이, 모델의 간결성은 곧 실용성과 연결되어진다. 추정 모델이 아무리 높은 예측정확도를 나타낸다 할지라도 모수가 많아지게 된다면 실제 이를 활용하는데 있어서 실용성이 떨어지게 된다. 그러므로 최종 예측 모델을 선택하는데 있어서 최대한 간결한 모델을 예측 모델로 선택하는 것을 원칙으로 한다.

Table 7은 앞선 상관도표 분석과 DF-gls검정 및 ADF 검정을 통해 추정해낸 잠정적 추정치를 바탕으로 추정된 모형인 Model(1)과 위의 조건 (9)을 만족하는 모형들 가운데 설명력과 계수의 유의성 검정¹²⁾, AIC와 BIC 값, 그리고 간결성 조건 등의 추정단계를 거쳐 추정한 Model(2)와 Model(3)을 나타낸 것이다.¹³⁾ 앞선 상관도표 분석과 DF-gls검정, 그리고 ADF검정을 통해 잠정적으로 예측한 'p=3', 'q=3', 'P=1', 'Q=1'을 토대로 추정한 Model(1)의 경우, 유의성 검정 결과 SAR(1) 항의 계수가 무의미 한 것으로 나타났다. 이에 반하여

Model(2)와 Model(3)의 경우에는 모수 값들이 유의수준 1% 수준에서 모두 유의미한 것으로 나타났다. AIC와 BIC의 값을 살펴보았을 경우, Model(2)의 값이 다른 두 모형과 비교하여 가장 우수한 것으로 나타났다. 간결성의 조건으로 판단하였을 경우에도 마찬가지로 Model(2)의 모수가 3개로서, Model(1) 8개 및 Model(3) 6개와 비교하여 가장 간결한 것으로 나타났다. 그러므로 국내 항만 컨테이너 물동량 예측을 위한 최적의 예측모델로서는 Model(2), 즉, SARIMA(1,1,1)(1,1,0)⁴ 모형이 가장 적합한 모형인 것으로 판단하였다.

4. 모형 검증

1) 잔차항의 자기상관여부 진단

앞서 모형의 식별과 추정단계를 통해 SARIMA(1,1,1)(1,1,0)⁴ 모형을 예측모형으로서 추정하였다. Table 8

Table 7. Estimated model

		Model(1)		Model(2)		Model(3)	
		SARIMA(3,1,3)(1,1,1) ⁴		SARIMA(1,1,1)(0,1,1) ⁴		SARIMA(2,1,3)(0,1,0) ⁴	
		coef	z-value	coef	z-value	coef	z-value
		(Std, Err.)	(p> z)	(Std, Err.)	(p> z)	(Std, Err.)	(p> z)
AR	AR(1)	1.007608	.	-0.9121545	-14.24	0.7968748	4.18
		.	(0.000)	(0.00640401)	(0.000)	(0.1908268)	(0.000)
	AR(2)	-1.207892	-34.49	-	-	-0.4479823	-2.70
		(0.0350244)	(0.000)			(0.1658325)	(0.007)
	AR(3)	0.7153351	38.16				
		(0.0187449)	(0.000)				
MA	MA(1)	-1.01903	-7.70	-0.9998414	5968.18	-0.9702911	-4.55
		(0.1322928)	(0.000)	(0.0001675)	(0.000)	(0.213175)	(0.000)
	MA(2)	1.20739	13.35	-	-	0.7504684	4.33
		(0.0904549)	(0.000)			(0.1735136)	(0.000)
	MA(3)	-.7380374	-4.37	-	-	-0.6408624	-3.80
		(0.1688766)	(0.000)			(0.1685012)	(0.000)
SAR	SAR(1)	.0867277	0.47	-	-	-	-
		(0.1859629)	(0.641)				
SMA	SMA(1)	-1.000001	-0.000014	-1.000052	-4083.27	-	-
		(0.00000727)	(0.000)	(0.0002449)	(0.000)		
Constant term		.0395336	8.15	-0.0008239	-0.92	-0.0005857	-0.36
		(0.0048507)	(0.000)	(0.0008946)	(0.357)	(0.0016239)	(0.718)
Number of obs		79		79		79	
Likelihood		135.9565		134.3932		127.654	
AIC		-255.913		-258.7863		-241.308	
BIC		-236.913		-246.9391		-224.7218	

12) 본 연구에서 Maximum Likelihood Estimation을 통해 모형이 추정되어졌는데, 이때 n>30이므로, 정규분포를 따른다고 가정하여 Z-test를 실시하였다.

13) 총 63개의 가능한 모델 중, Table 7에 제시된 Model(2)와 Model(3)을 제외하한 모든 모델이 Z-value 검정(유의수준 5% 수준 하)에 결과, 탈락되었다. Model(1)은 앞선 과정을 통해 예측한 잠정적 수치를 통한 모델이기 때문에 본문에서 제시함

Table 8. Q-value of SARIMA(1,1,1)(0,1,1)⁴ residual

Lag	Q	Prob>Q
1	0.1267	0.7219
2	0.20168	0.9041
3	3.1438	0.3700
4	4.7867	0.3099
5	4.8003	0.4407
6	4.8046	0.5691
7	4.8828	0.6743
8	4.9894	0.7587
9	5.1191	0.8238
10	5.8636	0.8266
11	7.0637	0.7939
12	7.0792	0.8523
13	7.6445	0.8660
14	8.7323	0.8478
15	10.3663	0.7963

은 앞서 추정한 예측모형인 SARIMA(1,1,1)(1,1,0)⁴ 모형의 잔차항이 자기상관(Autocorrelation)을 갖는지 여부를 판단하기 위한 잔차항의 Q통계량을 나타낸 표이다. 만약 잔차항에 자기상관이 존재한다면, 앞서 추정한 예측 모형인 SARIMA(1,1,1)(1,1,0)⁴ 모형은 통계적으로 유의하지 않다는 것을 의미하므로 Figure 1에 나타난 바와 같이 다시 예측모형을 추정해야만 한다.

Table 8에서 시차에 따른 Q-통계량을 살펴보면, 잔차항들 간에 자기상관이 존재하지 않는 것을 알 수 있다. 즉, SARIMA(1,1,1)(1,1,0)⁴ 모형의 추정치의 잔차항들 간에 자기상관이 존재 하지 않으므로 SARIMA(1,1,1)(1,1,0)⁴ 모형을 예측모형으로 활용하는데 통계적으로 문제가 없는 것으로 판단할 수 있다.

2) 예측정확도 검증

본 연구에서는 1990년 1사분기부터 2010년 4사분기까지의 국내 항만 컨테이너 물동량 자료를 바탕으로 예측모형을 추정하였다. 이렇게 추정한 SARIMA(1,1,1)(1,1,0)⁴ 모형을 활용하여 모델 추정에 활용된 기간인 1990년 1사분기부터 2010년 4사분기까지의 예측치를 추정함과 동시에, 예측시점을 2011년 1사분기로 가정하여 2011년 1사분기부터 2014년 4사분기까지의 물동량을 추가로 예측하였다. 이렇게 추정한 예측치를 실제 물동량 수치와의 비교를 통해 예측정확도를 살펴봄으로써 항만 컨테이너 물동량 예측에 있어서 SARIMA 모델의 활용가능성을 확인해 보았다. 또한 시계열 예측모형 가운데, 교통량 수요예측 분야에서 가장 많이 활용되고 있는 ARIMA 모형과의 예측 정확도 비교를 통해 예측모형

Table 9. ARIMA(1,1,0) model

		ARIMA	
		ARIMA(1,1,0)	
		coef	z-value
		(Std, Err.)	(p z)
AR	AR(1)	-0.3287157 (0.1558166)	-2.11 (0.035)
MA		-	-
Constant term		0.0258438 (0.0041839)	6.18 (0.000)
Number of obs		83	
Likelihood		130.2564	
AIC		-254.5128	
BIC		-247.2563	

간 상대적 정확도를 확인해 보았다.

Table 9는 ARIMA(1,1,0) 모형으로서, 본 연구에서 국내 항만 컨테이너 물동량 예측을 위해 추정한 ARIMA 모형이다. ARIMA(1,1,0) 모형을 앞서 예측 모형으로 선정한 SARIMA(1,1,1)(0,1,0)⁴ 모형 비교하였을 때, 우도값과 AIC값의 경우 SARIMA(1,1,1)(0,1,0)⁴ 모형이 보다 우수한 것으로 나타났고, BIC 값의 경우에는 ARIMA(1,1,0)모형이 근소하게 높은 것으로 나타났다. 일반적으로 모형이 시계열 자료를 얼마나 잘 설명하고 있는지를 보여주는 우도값에서 나타나는 차이에 비하여 상대적으로 AIC나 BIC 값의 차이가 작게 나타나는 것은 ARIMA모형은 계절차분과 더불어 계절항인 SAR항과 SMA항이 포함되지 않기 때문에 SARIMA모형에 비해 모수의 수가 적을 수밖에 없는 모형의 구조적 특성에 기인한 것으로 판단된다. 또한 AIC나 BIC의 추정식에서 우도값은 자연로그를 취한 후 반영되는 반면에, 모수의 수는 바로 식의 반영되기 때문에 우도값 보다 모수의 값이 더욱 큰 영향을 미치게 된다. 이러한 점을 감안하고 SARIMA(1,1,1)(0,1,0)⁴ 모형과 ARIMA(1,1,0) 모형을 비교하였을 때, SARIMA(1,1,1)(1,1,0)⁴ 모형이 항만 컨테이너 물동량 시계열 자료를 보다 잘 설명하고 있다고 판단된다. 그러나 어디까지나 지금까지의 판단과정은 추정된 모형 자체의 통계적 값들로 판단한 내용이기 때문에, 보다 정확한 판단을 위해서는 각 모형을 통해 예측된 예측치와 실제 물동량 수치와의 비교를 통해 두 모형의 예측 정확성을 확인해 보아야 한다.

Table 10은 2010년 1사분기부터 2013년 4사분기까지 국내 항만 컨테이너 물동량 실측치 자료와 SARIMA(1,1,1)(0,1,0)⁴ 모형과 ARIMA(1,1,0) 모형을 통한

Table 10. Comparison SARIMA(1,1,1)(0,1,1)⁴ model with ARIMA(1,1,0) model

	Actual Volume	SARIMA(1,1,1)(0,1,1) ⁴			ARIMA(1,1,0)		
		Forecasting	Accuracy	Error	Forecasting	Accuracy	Error
RMSE / MAPE (1990 1Q - 2010 4Q)		137,815 / 3.11%			166,664 / 3.70%		
2011 1Q	5,044,362	4,953,404	98.16%	-90,958	5,150,465	97.94%	106,103
2011 2Q	5,501,910	5,359,657	97.35%	-142,253	5,288,422	95.96%	-213,488
2011 3Q	5,459,730	5,281,029	96.62%	-178,701	5,425,830	99.38%	-33,900
2011 4Q	5,604,502	5,443,514	97.04%	-160,988	5,568,226	99.35%	-36,276
2012 1Q	5,422,932	5,356,627	98.76%	-66,305	5,713,896	94.91%	290,964
2012 2Q	5,836,830	5,751,372	98.51%	-85,458	5,863,522	99.54%	26,692
2012 3Q	5,584,079	5,677,855	98.35%	93,776	6,017,014	92.80%	432,935
2012 4Q	5,706,436	5,845,689	97.62%	139,253	6,174,543	92.42%	468,107
2013 1Q	5,616,606	5,745,534	97.76%	128,928	6,336,213	88.64%	719,607
2013 2Q	6,005,359	6,155,364	97.56%	150,005	6,502,098	92.36%	496,739
2013 3Q	5,810,654	6,072,598	95.69%	261,944	6,672,345	87.09%	861,691
2013 4Q	6,035,990	6,245,009	96.65%	209,019	6,847,030	88.15%	811,040
RMSE / MAPE (2011 1Q - 2013 4Q)		152,116 / 2.49%			475,755 / 5.95%		

*unit: teu

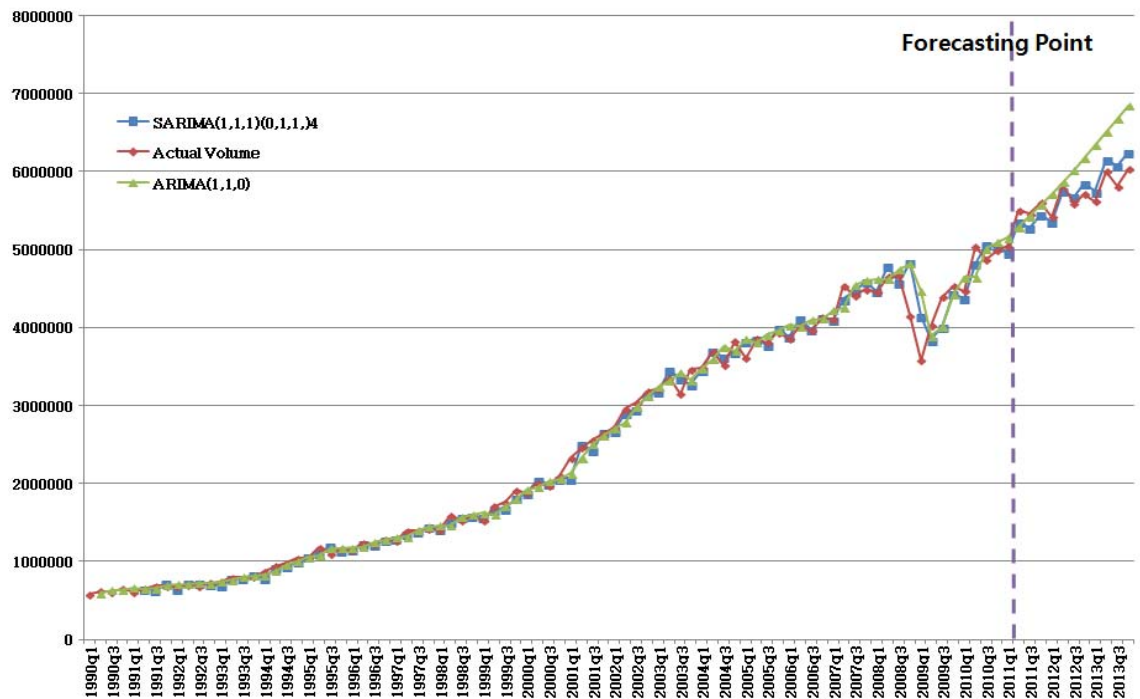


Figure 4. Comparison of forecasting accuracy SARIMA(1,1,1)(0,1,1)⁴ with ARIMA(1,1,0)

예측치를 비교하여 놓은 표이다. 물동량 예측모형에 활용된 예측기간인 1990년 1사분기부터 2010년 4사분기까지의 예측 모형별 RMSE 값을 살펴보면 SARIMA (1,1,1)(0,1,0)⁴ 모형의 RMSE 값이 137,815, ARIMA (1,1,0) 모형의 RMSE 값이 166,664로 나타났다. 또한 동기간 MAPE 값의 경우 역시 마찬가지로 SARIMA

(1,1,1)(0,1,0)⁴ 모형의 MAPE 값이 약 3.11%, ARIMA (1,1,0) 모형의 MAPE 값이 3.70%인 것으로 나타났다. 이를 바탕으로 SARIMA(1,1,1)(0,1,0)⁴ 모형이 모형 추정에 활용된 기간 즉, 1990년 1사분기부터 2010년 4사분기의 시계열 자료들을 보다 정확하게 예측하고 있는 것으로 나타났다. 모형추정에 활용된 기간이 아닌

실제 물동량을 예측한 기간, 즉, 2011년 1사분기부터 2013년 4사분기까지의 예측 결과를 살펴보면, SARIMA (1,1,1) (0,1,0)⁴ 모형의 RMSE 값과 MAPE 값은 152,116과 2.49%로 나타나, RMSE 값이 475,755, MAPE 값이 5.95%로 나타난 ARIMA(1,1,0) 모형에 비해 더욱 높은 예측 정확도를 갖는 것으로 나타났다. 특히 SARIMA(1,1,1)(0,1,0)⁴모형의 MAPE 값을 살펴보면 물동량 예측모형에 활용된 기간의 MAPE 값보다 실제 예측기간의 MAPE 값이 더욱 낮게 나타나며 오히려 실제 예측기간에 대한 예측정확도가 더욱 높은 것으로 나타났다. 또한 SARIMA(1,1,1)(0,1,0)⁴ 모형의 경우, 예측 시점이 멀어지는 경우에도 95%이상의 매우 높은 예측정확도를 계속하여 유지하고 있는 것을 알 수 있다. 반면 ARIMA(1,1,0) 모형은 단기간의 경우에는, SARIMA (1,1,1)(0,1,0)⁴ 모형과 예측과 마찬가지로 높은 예측정확도를 보였으나, 예측 시점이 멀어짐에 따라 예측정확도가 낮아지는 것으로 나타났다.

Figure 4는 앞서 추정한 예측모형인 SARIMA (1,1,1) (0,1,0)⁴ 모형과 ARIMA(1,1,0) 모형을 통해 예측된 각각의 예측 물동량과 실제 물동량을 시점에 따라 나타낸 그래프이다. 예측시작점인 2011년 1사분기부터의 예측 물동량을 살펴보면 SARIMA(1,1,1)(0,1,0)⁴ 모형의 경우, 계절성을 띄고 있는 실제 물동량의 특성을 정확하게 반영하며 물동량을 예측하고 있는 것을 알 수 있다. 반면에 ARIMA(1,1,0) 모형의 경우, 예측시점부터해서 물동량이 계속하여 증가하고 있는 것으로 예측하고 있다. 이는 ARIMA모형이 시계열 자료의 계절적 특성을 반영하지 못하는 단점이 그대로 나타나고 있는 모습이다. ARIMA 모형은 예측시점을 기준으로 가까운 시점의 과거 값들만을 미래 예측에 활용할 수 있기 때문에, 전년도 동분기와 같이 주기적인 값, 즉 계절성과 같은 주기적 특성을 예측과정에 반영할 수 없는 단점이 있다. 이에 비하여 SARIMA 모형은 최근의 과거 값 뿐만 아니라, 전년도 동분기의 값, 즉 주기적 특성을 반영할 수 있는 특성을 지니고 있어, 본 연구에서 예측하고자 하는 국내 전체 항만 컨테이너 물동량 같이 계절적 특성을 지니고 있는 시계열 자료의 예측에 있어 ARIMA 모형에 비하여 상대적으로 유용하게 활용할 수 있다. 이러한 특성으로 인해 국내 전체 항만 컨테이너 물동량을 추정하는 예측 모형으로서, SARIMA(1,1,1)(0,1,0)⁴ 모형이 ARIMA (1,1,0) 모형보다 보다 높은 예측 정확도를 보이는 것으로 판단된다.

4. 시사점

본 연구는 1990년 1사분기부터 2010년 4사분기까지의 분기별 국내 전체 컨테이너 물동량 자료를 바탕으로 SARIMA 모형을 활용, 분기별 국내 컨테이너 물동량 수요예측 모형을 구축하였다. 앞서 언급한 바와 같이 그동안의 항만수요예측관련 연구는 대부분 장기수요예측에 초점이 맞추어져 왔다. 또한 단기간의 수요예측에 있어서도 대부분 월별이나 연도별 시계열 자료를 활용하였고, 분기별 자료를 활용한 경우는 찾아보기 힘들다. 또한 대부분의 연구가 수요예측 모형을 추정한 이후, 예측에 활용된 기간의 시계열 자료와의 비교 및 모형의 적합도에 초점이 맞추어져 있어온 것에 반해, 본 연구는 예측에 활용된 기간의 시계열 자료와의 예측력 비교 및 모형의 적합도, 그리고 더 나아가 실제 예측시점 이후에 물동량을 직접 예측하고, 그 예측치를 실제 물동량 자료와 비교하여 봄으로써 예측모형의 예측정확도를 보다 실용적이고 다양한 측면에서 검증하여 보았다. 이와 더불어 기존에 널리 활용되어지고 있는 방법론 가운데 하나인 ARIMA 모형을 활용하여 추정한 예측 모형과의 비교를 통해 컨테이너 물동량 예측모형으로서의 SARIMA 모형의 우수성을 검증하여 보았다.

기존에 월별, 혹은 연도별 시계열 자료에 SARIMA 모형을 활용하여 항만 물동량 추정을 하였던 다양한 선행연구에서 SARIMA 모형에 예측 정확도가 다른 모형들에 비해 상대적으로 떨어지고 있는 것으로 나타났다. 하지만 분기별 시계열 자료를 활용한 본 연구에서는 SARIMA 모형의 예측정확도가 매우 우수한 것으로 나타났다. 이는 SARIMA 모형이 월별, 혹은 연도별 자료에 비하여 계절적 특성이 보다 분명하게 보이는 분기별 자료에 특화되어 있는 수요예측 방법론이라는 것을 의미한다. 그렇기 때문에 SARIMA 모형을 기존의 연구들과 같이 월별, 연도별 자료에 대한 수요예측에 활용하기보다는 SARIMA 모형의 강점을 최대한 활용하여 분기별 수요예측에 특화하여 사용한다면 항만 물동량 예측 방법론으로서 매우 유용하게 활용될 수 있을 것으로 기대된다.

하지만 본 연구에서 활용되어진 SARIMA 모형이 분기별 항만 물동량 예측에 있어서 매우 높은 예측정확도를 보였다 할지라도, SARIMA 모형이 분기별 예측에서 절대적으로 우수한 모델이라 정의할 수는 없다. 일반적으로 항만물동량이 계절성을 지니고 있다는 것은 모두가

알고 있는 사실이나, 모든 항만이 계절성을 지니고 있다고는 단언할 수는 없다. 각 개별항만의 특성에 따라 계절성을 보이는 항만이 있는가 하면, 그렇지 않은 항만도 분명히 존재한다. 그러므로 분기별 물동량 예측이라고 해서 단순히 SARIMA 모형을 활용하는 것 보다는, 해당 시계열 자료의 특성을 사전적으로 반드시 확인한 후, SARIMA 모형을 활용할지 여부에 대해 판단할 필요가 있다. 또한 본 논문에서 대조군으로 활용된 ARIMA 모형 이외에도 최근 광범위하게 사용되고 있는 ANN모형이나 GP모형 등과의 비교분석을 통해 SARIMA 모형의 예측정확도를 추가적으로 검증해 나갈 필요가 있다.

본 연구를 통해, SARIMA 모형이 향후 항만 물동량 예측을 하는데 있어서 유용한 방법론으로 활용되어질 만한 충분한 가치가 있음을 확인할 수 있었다. 앞으로 항만 물동량 예측에 있어서 SARIMA 모형이 보다 다양하고 널리 활용되어 질 것으로 기대된다.

결론

해상운송은 항공운송과 더불어 국가 간 교역에 주요 운송수단으로써 매우 큰 역할을 수행하여 왔으며, 특히 항만은 해상운송의 노드(node)로서 원활한 해상운송 활동을 유지하기 위한 절대적인 요소 가운데 하나로 역할을 수행하여왔다. 이러한 항만의 중요성을 인지하고 그동안 항만분야에 대하여 다양한 연구들이 진행되어져 왔고, 그 가운데 항만 물동량 수요예측이 매우 중요한 세부 연구 분야로서 다루어져 왔다.

그동안의 항만 물동량 수요예측 결과물이 많은 시간과 비용이 소비되는 항만 인프라 확장 등에 주로 이용되어오며, 장기수요 예측에 관한 연구가 중점적으로 이루어져 왔다. 하지만 항만 간 경쟁이 심화되고, 기능이 고도화됨에 따라 항만 소프트웨어 측면에서 활용이 가능한 항만 물동량에 대한 단기 수요 예측 역시 비중 있게 다루어져야 할 것이다.

본 연구는 SARIMA 모형을 활용하여, 기존에 다루어지지 않았던 분기별 국내 전체 항만 컨테이너 물동량을 예측함과 동시에 기존 ARIMA 모형을 통한 예측 모형과 비교함으로써 SARIMA 모형의 예측 정확도를 검증해 보았다. 모형 추정에 활용된 기간에 대한 모형 적합도 검정 및 RMSE값을 확인한 결과 SARIMA 모형이 ARIMA 모형에 비해 우수한 결과를 보였다. 또한 모형 추정 이후 실제 물동량을 예측하였던 2011년 1사분기부터 2013

년 4사분기까지 SARIMA 모형의 예측정확도를 보면 95.69%에서 98.76%까지 매우 높은 예측 정확도를 보이고 있는 것을 알 수 있다. 특히 주목할 만한 사항은 예측기간이 길어질수록 예측정확도가 떨어지는 것이 일반적인 수요예측 방법론의 한계인데, 본 연구에서 추천한 SARIMA 모형은 예측시점으로부터 12시점이 지난 2013년 4사분기의 예측정확도가 96.65%로 매우 높게 나타나는 것을 알 수 있다. 물론 본 연구에서 얻은 결과를 일반화하기에는 무리가 있으나, 최소한 SARIMA 모형이 분기별 물동량 예측에 있어 예측시점이 멀어지더라도 상대적으로 높은 예측정확도를 유지할 수 있다는 가능성이 있다는 것을 보이는 것만으로도 충분히 그 가치가 있다고 판단된다.

본 연구에서 활용한 SARIMA 모형은 대표적인 시계열 분석기법(Time series analysis) 가운데 하나로서, 미래 값이 다른 변수가 아닌 자기 자신의 과거 값에 의해 결정된다는 전제에서 출발하는 예측 방법론이다. 시계열 분석 기법은 현재까지의 추세가 미래에도 계속된다는 가정 하에서 미래를 예측하는 기법이다. 그렇기 때문에 만약 미래의 추세가 현재까지의 추세와 다를 것으로 예상될 경우, 시계열 분석기법을 통한 예측을 수행하는데 큰 제한사항이 따르게 된다. 또한 외부변수에 의한 외부효과를 전혀 반영하지 못한다는 단점이 있다. 그렇기 때문에 일반적으로 GDP, 국가 간 교역량 등과 밀접한 관련성을 갖는 항만 물동량의 특성상, 물동량을 추정하는데 있어서 시계열 분석기법 보다는 회귀분석과 같은 인과분석 모형이 더욱 적합하다는 의견도 적지 않은 것이 사실이다. 인과모형을 활용하여 예측모형을 추정할 경우, 각 결정변수들이 물동량에 미치는 각각의 영향을 파악할 수 있기 때문에 매우 실용적일 수 있으나, 미래의 수요예측을 위해서는 예측하고자 하는 미래 시점의 변수들에 대한 예측이 선행되어야 하는 단점 역시 존재한다. 그러므로 수요예측의 목적에 따라 인과모형과 시계열 분석기법이 적절하게 병행되어 활용되어질 필요성이 있다. 이러한 관점에서 본 연구에서 활용한 SARIMA 모형 역시 항만 물동량 예측기법의 하나로써, 적절히 활용되어지기를 기대한다.

마지막으로 본 연구를 통해 SARIMA 모형이 분기별 물동량을 예측하는데 있어 매우 높은 예측 정확도를 보이는 것을 확인할 수 있었다. 항만 물동량을 예측하는데 있어서 그 목적에 따라 장기 및 단기 수요예측이 함께 병행되어야 한다. 또한 단기 수요예측에서도 짧게는 월단

위 및 분기단위, 길게는 연단위의 수요예측에 대한 필요성 역시 갈수록 강조되어질 것이다. 이러한 시점에서 본 연구의 결과, 즉, SARIMA 모형이 분기별 항만 물동량 예측에서 매우 높은 정확도를 보이는 것을 제시하고 있는 본 연구의 결과는 향후 분기별 항만 물동량을 예측하는데 있어서 시사하는 바가 크다고 생각한다. 향후 분기별 항만 물동량 예측에 있어서 SARIMA 모형이 지금보다 더욱 다양하고 유용하게 활용되어지기를 기대한다.

ACKNOWLEDGEMENT

This work was supported by Inha University and National Research Foundation of Korea Grant Funded by the Korean Government (MOEHRD) (NRF-2011-413-B00008).

REFERENCES

Box G. E. P., Jenkins G. M., Reinsel G. C. (1976), *Time Series Analysis, Forecasting and Control*, Sanfrancisco; Holden Day.

Chen S. H., Chen J. N. (2010), *Forecasting Container Throughputs at Ports Using Genetic Programming, Expert Systems with Applications*, 37, 2054-2058.

Chou C. C., Chu C. W., Liang G. S. (2008), *A Modified Regression Model for Forecasting the Volumes of Taiwan's Import Containers, Mathematical and Computer Modelling*, 47, 797-807.

Dickey D. A., Fuller W. A. (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association*, 74(366), 427-431.

Hylleberg S., Engle R. F., Granger C. W. J., Yoo B. S. (1990), *Seasonal Integration and Cointegration, Journal of Econometrics* 44, 215-238.

Kim C. B. (2007), *Forecasting the Trading Volumes of Marine Transport and Ports Logistics Policy-Using Multiplicative Seasonal ARIMA Model, Journal of Korea Port Economic Association*, 23(1), 149-162.

Kim J. K. et al. (2011), *A Study on Forecasting Container Volume of Port Using SD and ARIMA, Journal of Navigation and Port Research International Edition*,

35(4), 343-349.

Lee J. M., Kwon Y. J. (2011), *A Study on Dynamic Change of Transportation Demand Using Seasonal ARIMA Model, J. Korean Soc. Transp.*, 29(5), Korean Society of Transportation, 139-155.

Min K. C., Jun Y. I., Ha H. K. (2013), *Forecasting the Air Cargo Demand With Seasonal ARIMA Model: Focusing on ICN to EU Route, J. Korean Soc. Transp.*, 31(3), Korean Society of Transportation, 3-18.

Peng W. Y., Chu C. W. (2009), *A Comparison of Univariate Methods for Forecasting Container Throughput Volumes, Mathematical and Computer Modelling*, 50, 1045-1057.

Shin C. H. et al. (2008), *A Study on the Forecast of Port Traffic Using Hybrid ARIMA-Neural Network Model, Journal of Navigation and Port Research*, 31(1), 81-88.

Wu S. (2010), *Lag Length Selection in DF-gls Unit Root Tests, Communications in Statistics - Simulation and Computation*, 39(8), 1590-1604.

Yi G. D. (2013), *Forecasting the Container Throughput of the Busan Port using a Seasonal Multiplicative ARIMA Model, Journal of Korea Port Economic Association*, 29(3), 1-23.

- ✉ 주 작성자 : 민경창
- ✉ 교신저자 : 하현구
- ✉ 논문투고일 : 2014. 5. 8
- ✉ 논문심사일 : 2014. 7. 2 (1차)
2014. 9. 14 (2차)
2014. 10. 3 (3차)
- ✉ 심사판정일 : 2014. 10. 3
- ✉ 반론접수기한 : 2015. 4. 30
- ✉ 3인 익명 심사필
- ✉ 1인 abstract 교정필