JOURNAL OF MARITIME RESEARCH

Vol XIV. No. III (2017) pp 8–15

ISSN: 1697-4040, www.jmr.unican.es

Estimation of Cargo Demand at Major Seaports in India

P. K. Sahu[1], G. Patil[2]

ARTICLE INFO

Article history:
Received 5 October 2017;
in revised form 24 October 2017;
accepted 25 November 2017.

Keywords:
Indian seaports, Cargo demand estimation, Regression, Time series.

ABSTRACT

Estimation of cargo demand is essential for augmenting port capacity, planning port facilities, making decision on improvement of port operational efficiency, etc., and these activities need large and irreversible investment. Future cargo flow projections also helps in making decisions on cargo rates. Therefore, accuracy in estimating the cargo demand at port locations is critical. Indian ports experienced 151% increase in cargo volume between fiscal years from 2002-03 to 2014-15. This paper analyses such growth systematically with classical regression and time series modeling techniques. The regression models are developed using macroeconomic conditions as causal variables. The models are estimated using quarterly and monthly cargo flow data for twelve major seaports in India. The analysis revealed that the time series models perform better in terms of prediction accuracy than the regression models. The average prediction error from the regression models varied from 6% to 20%, while the error associations with time series models are varying from 3. 8% to 12.6%. This study is intended to provide infrastructure planners with some guidance on short to medium term development of transport infrastructure requirements over the lifetime of port infrastructure, while planning for port connectivity roads within an urban transport network.

1. Introduction

India is a major maritime nation with a long coastline, spanning about 7516.6 kilometres, constituting 13 major (12 Government owned and 1 private) and 187 non-major (minor and intermediate) seaports. These ports are strategically located along the two coast lines: East coast and West coast, to facilitate national/international trade. The east coast and the west coast have 54 and 146 seaports, respectively (DBI, 2009; MoS, 2011; RWGPS, 2011). Although India's international trade in terms of value is less than one percent of the total world trade; the maritime transport system in India facilitates about 95% by volume and 77% by value of the nations overseas trade (DBI, 2009; MoS, 2011; ISI, 2010). The Indian ports which were handling 418.72 million tons of cargo in financial year (FY) 2002-03; more than has doubled to 1052.01 million tons of cargo during the FY 2014-15. The total cargo volume growth between 2002-03 and 2014-15 is about 151%. The total cargo volume at major ports grew from 313.55 million tons to 581.34 million tons during this period; whereas, the growth at non-major ports was from105.17 million tons to 470.67 million tons (MoS, 2011; RWGPS, 2011). The XIIth five-year: 2012-2017 plan projects the estimated investment requirement for the Indian marine infrastructure development is about 33 billion USD (RWGPS, 2011). There is little doubt that throughput cargo volume growth is one of the key factors which influences such large and irreversible investment for ports and its associate infrastructure development. Nonetheless, the allocation of such huge investment should be done systematically for Indian marine system and their associated infrastructure expansion in order, not only to decrease the total transportation cost but also help in fostering the smooth flow of traffic through the port. Because, the development of port systems is a continuous process based on the trade requirement and the future expected traffic volume; accurate projection of cargo throughput is critical for such developments.

The future traffic volumes at ports have also implications for the development of transport infrastructure in the regions neighbouring to ports. Cargo growth at seaports significantly

[1]Birla Institute of Technology and Science Pilani.
[2]Indian Institute of Technology Bombay.

affects the performance of transportation networks in surrounding neighbourhoods. For example, the truck trips generated at Mumbai Port has significant impact on the transportation systems in Mumbai Metropolitan region. To analyse the impact, it is needed to know the number of truck trips generated by a port. The number truck trips generated at a port depend on the cargo volume handled by the port. In other words, the demand estimation is needed not only to plan the port facilities, but also to plan the transportation system in the neighbouring area and in this regard the hinterland connectivity is a crucial element.

A few studies (Haralambides and Gujar, 2012; Panigrahi and Pradhan, 2012; Raghuram and Shukla, 2014; Sahu et al. 2014) are available on Indian port system, which mainly focus on the historical, social, environmental, and maritime policy aspects. No study was found that deals with the cargo demand estimation for Indian ports, except a recent study on cargo demand analysis at Mumbai port by Sahu and Patil (2013, 2014, 2015) and Patil and Sahu (2016a, b). This paper presents demand estimation models for all major ports in India using regression and time series approaches. The finding of this study will be useful for transport planners, Government policy, strategy makers, and port operators.

The paper is organized as follows: Literature review is discussed in the next section. The section following literature review briefly reports about the data used for this study. The next section describes the cargo demand model structures used for this study. Model estimation and validation are also discussed in the same section. The conclusions of this study are presented in the final section. .

2. Literature review

Cargo forecasting models can be grouped into five classes namely, the flow factoring method (FFM), the origin-destination (O-D) factoring method, the truck model, the four-step commodity model, and the economic activity model (Cohen et al., 2008). Out of these five model classes, the FFM estimates the cargo volume on roads, railroads, and ports intending short term forecasts. FFM is relatively simple and used by state Department of Transportation across the USA. The FFM relies on regression equations i.e., econometric analysis and time series analysis. The available sample of literature revealed that standard regression analysis is the widely used technique for cargo demand estimation. The Federal Highway Administration's Guidebook (USDOT, 1999) on State Wide Travel Forecasting focused on Autoregressive Integrated Moving Average (ARIMA) models and linear regression model to forecast truck volumes. Ballach and Tadi (1994) examined the use of regression procedure to build a truck trip generation models for computing truck trip rates with activity level and land use categories as explanatory variables. Seabrooke et al. (2003), predicted the cargo growth at the Hong Kong port by means of regression analysis.

Al-Deek (2001) investigated the use of neural networks with multiple regression analysis for developing cargo prediction models for the port of Miami, and the port of Jacksonville. Klodzinski and Al-Deek (2003) used both regression and Artificial Neural Network (ANN) methods to develop cargo forecasting models for major Florida seaports. These models were used to estimate daily truck trips in and out of the ports. The authors compared ANN with multiple regression models. They concluded that neural networks are more flexible and precise tool to predict truck trips generated from seaport cargo activity. Al-Deek et al. (2000), used time series models for predicting seasonal variations in cargo movements for the port of Miami. Schulze and Prinz (2009) used ARIMA model to predict container transhipment in Germany. Eddie et al. (2004)), used Error Correction Model (ECM) approach to forecast the cargo throughput for the port of Hong Kong. They suggested classical regression approach is valid in case of the dataset which is stationary without presenting a trend over time. This is because for independent trending variables, the direction of movement is same under the common trend, which creates an illusion of causal relationship. Therefore, they proposed an alternative approach namely ECM for cargo forecast. They applied a log-linear model to forecast the cargo throughput at the port of Hong Kong.

The literature on Indian seaports is scarce and is not focused on quantitative analysis of cargo demand. Recently, Sahu and Patil (2013, 2015) and Patil and Sahu used multivariate linear regression and time series technique to predict annual cargo demand at Mumbai port, India. It was found that macroeconomic variables such as Gross Domestic Product (GDP), crude oil production, etc. affect the cargo volume at Mumbai port. Patil and Sahu (2016b) proposed a dynamic demand model to estimate the cargo demand at all the major ports simultaneously.

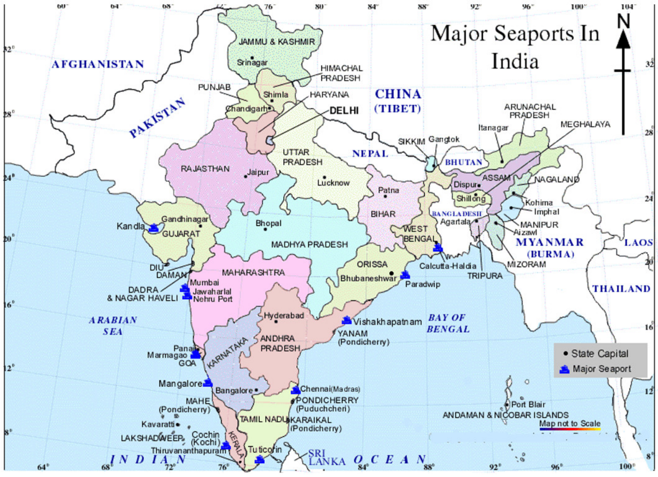
In general, the forecasting spectrum includes methods and techniques like regression, time series and, in recent times, artificial intelligence tools like neural network, fuzzy logic, can be deployed to estimate the future traffic (Chen et al., 2009; Cryer and Chan, 2008; Faghri and Hua, 1992; Middendrof et al., 1982). A structural review of available literature by Woo et al. (2011), on cargo demand estimation reported that about 13% of total published literature used regression techniques for seaport cargo data analysis. The same study also reported that the use of time series modeling towards seaport cargo demand estimation is limited in the existing literature. In this study, regression and time series modeling procedures are used to forecast short term cargo flow for the major ports in India. The models are compared based on the prediction accuracy to suggest the improved modeling approach for short to medium term cargo flow prediction at Indian ports.

3. Study data

According to the federal structure of Indian constitution, maritime transport is administered by both the central and the state Governments. The federal Ministry of Shipping administers the major ports and the non-major ports are administered by the respective coastal states.

However, this study is restricted to major ports only. Figure 1 shows a map of India with the locations of major ports. Monthly and quarterly cargo (inbound and outbound) data from

Figure 1: Thematic Representation of Major Port Locations



Source: www.mapsofindia.com

Table 1: Annual Cargo Voulme Handled at Major Ports (2002-2013)

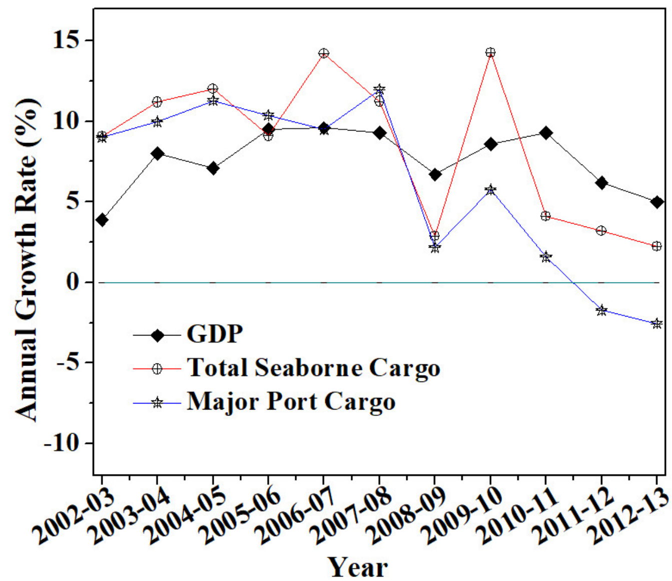
Sl. No.	Port Name	Port Code	Cargo Volume (in million tons)	
			2015 –2016	2002 – 2003
1	Kolkata	1001	50.20	35.80
2	Paradip	1002	76.39	23.90
3	Visakhapatnam	1003	57.03	46.01
4	Chennai	1004	50.06	33.69
5	Tuticorin	1005	36.85	13.29
6	Cochin	1006	22.10	13.02
7	New Mangalore	1007	35.58	21.43
8	Mormugao	1008	20.78	23.65
9	Mumbai	1009	61.11	26.80
10	JNPT	1010	64.03	26.84
11	Ennore	1011	32.21	8.49
12	Kandla	1012	100.05	40.63

2002-03 to second quarter of 2016-17 for 12 major ports were obtained from Centre for Monitoring Indian Economy (CMIE). Since Port Blair was added as the 13th major port in 2010, it is not considered in this study. For ease of the analysis, the ports are assigned four digit numbers starting from 1001 to 1012. Table 1 shows the cargo tonnage values for the years 2002-03 and 2015-16 along with the assigned numbers for all 12 study ports.

It may be observed from Table 1 that eleven out of twelve major ports reported growth in cargo traffic from the year 2002-03 to 2015-16. Ennore reported highest increase of 279%, followed by Paradip 220%, followed by Tuticorin 177%, followed by Kandla 146%, JNPT (139%), Mumbai (128%), New Mangalore (66%), etc. The only port that showed decline (-12%) in cargo movement is the Mormugao port. This is because the primary commodity handled by this port is iron ore; Government of India has significantly restricted the iron ore mining activities in India for the last two to three years.

Quarterly data were collected for several economic indicator variables to analyze their influence on estimating the quarterly cargo demand. These data include Gross Domestic Product (GDP), Coal Production (CLP), Cement Production (CMT), Crude Oil Production (CRLP), Fertilizer Production, Refinery Product Production (RFP), Steel Production, and . The

Figure 2: Growth in Indian Seaborne Trade vis-à-vis Growth in GDP



Source: Authors.

GDP values are in thousand billion rupees and the remaining variable values are in million tons. The growth in India's GDP, major port traffic, and total seaborne trade is presented in Figure 2. The figure shows the consistent rise in total cargo volume with the growth of GDP. Although, there is slowdown in major port activity for the last two years, most of the planners and policy makers are optimistic about cargo traffic growth (IPA, 2013; MoS, 2011)) due to the gradual growth in the Indian economy in the last one year while, some foresee slower prospects. In this paper, an attempt is made to establish the cargo volume association with GDP and associated other macroeconomic variables at all major port locations in India.

4. Cargo demand estimation

4.1. Regression models

The most widely used technique for estimating cargo flow is ordinary least square (OLS) regression. OLS regression (Makridakis, 2005; Affi et al., 2012) analysis is simple and can model factors associated with cargo flow, including the underlying causes. Regression models are developed for the 12 study ports to predict the short term cargo demand. Based on the assumption that cargo flows are driven by nation?s economy, macroeconomic variables are considered as explanatory variables. Cargo volume is regressed on the macroeconomic variables and their associations are estimated by the classical regression model. The models are developed using quarterly data.

4.1.1. Model structure

The scatter plot matrix among the dependent and independent variables suggested the following regression model structure to be adopted for modelling the cargo flow.

$$Y_i = \beta_0 + \beta_1 GDP + \beta_2 GDP^2 + \beta_3 CRLP + \beta_3 CLP + \beta_3 RFP + \beta_3 FRTP + \beta_3 CMT + \beta_3 STL + \varepsilon_i \quad (1)$$

Where, i = Inbound or Outbound cargo tonnage in million tons; E i= 0, gross domestic product in thousand billion Indian rupees; the remaining variables are in million tons

4.1.2. Model estimation

The regression analysis was started using the proposed model structure discussed in Equation 4.1. The outliers were removed from the dataset through scatter plot diagnostic tool and the removal was confirmed from the residual analysis. The removal process resulted with 52 valid data points; 85% of the data were used for modeling, keeping 15% data for model validation. The model calibration was done by using ordinary least square method. Based on performance, 24 models are selected to represent the inbound and outbound flow. The models along with the calibration results are presented in Tables 2 and 3. These models are selected based on the higher R^2, lower standard error and lower prediction error value. The model parameters: R^2, Adjusted R^2, F-statistics, and t-statistics, are also presented in Tables 2 and 3.

The performances of the developed models are good for almost all the ports. The positive sign of GDP indicates that there will be increase in port activities with the increase in gross domestic product. The proposed univariate model structure resulted with a maximum R^2 value: 87.3% (see M5-1003 in Table 2). The model candidate M5-1003 is the representative of Visakhapatnam port's inbound cargo flow. The corresponding adjusted R^2 value is 86.3% which indicates that over 86% of the quarterly variations in Visakhapatnam inbound cargo throughput can be explained by the GDP, coal production, and fertilizer production. In other words, if these explanatory variables values are known, quarterly inbound cargo volume can be estimated with mean error fewer than 14% for the Visakhapatnam port. The R^2 value is more than 70% for the remaining ports and similar conclusions may be drawn for both inbound and outbound cargo tonnage at each port. However, it may be observed that the R^2 value for Ennore port outbound cargo flow (see model M22-1011 in Table 3) is very low. No model was found to fit for this port with the variables considered as explanatory variables in the analysis. The possible reason could be the presence of missing data resulting with lower sample size. More research is required for missing data computation for fitting a suitable outbound quarterly model for Ennore port. These models were further investigated for homoskedasticity and normality. The residual plot and the normal probability plot confirmed the normality and constant variance of the datasets used in the study.

4.1.3. Model validation

The models were validated using 15% of the unused data and the average prediction errors were computed. As a sample of validation results, the estimated and actual cargo tonnage values are presented in Figure 3 and 4 for Chennai (1004) and

Table 2: Inbound Demand Estimation Regression Model Summary

Model Number	Constant	GDP	GDP^2	CRLP	CLP	RFP	FRTP	CMT	STL	R^2	Adj. R^2	F-value
M1-1001	1.402 (18.82)*	0.522 (7.21)	0.037 (8.62)	-	0.011 (2.68)	-	-	-	-	0.725	0.722	118.234
M3-1002	0.757 (3.22)	0.081 (3.95)	-	-	-	-	-0.757 (-5.67)	-	0.661 (6.76)	0.855	0.848	144.816
M5-1003	0.856 (13.94)	0.256 (9.32)	-	-	0.022 (4.82)	-	-0.261 (-4.34)	-	-	0.875	0.862	102.612
M7-1004	1.415 (11.03)	0.311 (7.73)	-	0.265 (3.53)	0.016 (3.29)	-	-	-	-	0.775	0.758	112.628
M9-1005	1.082 (11.54)	0.161 (9.36)	-	-	0.028 (3.36)	-	-	-	-	0.836	0.825	114.852
M11-1006	1.341 (14.27)	0.052 (4.26)	-	-	-	0.011 (3.08)	-	-	-	0.784	0.776	134.546
M13-1007	1.410 (9.66)	0.211 (8.37)	-	-	0.004 (3.38)	-	-	-	-	0.822	0.814	148.688
M15-1008	1.121 (13.48)	0.524 (4.56)	0.005 (5.88)	-	-	-	-	-	-	0.708	0.706	58.638
M17-1009	1.427 (9.22)	0.635 (5.55)	-	-0.164 (-3.61)	-	0.051 (7.76)	-	-	-	0.822	0.821	108.826
M19-1010	0.852 (9.38)	0.588 (13.38)	-	-	-	-	-	-	-	0.813	0.807	152.684
M21-1011	0.408 (4.16)	0.205 (2.58)	0.022 (4.17)	-	0.002 (6.44)	-	-	-	-	0.728	0.726	82.467
M23-1012	2.038 (5.96)	0.442 (4.94)	-	0.336 (3.19)	-	0.124 (3.46)	-	-	-	0.826	0.814	168.544

**(x.xx)= t-statistics value*

Table 3: Outbound Demand Estimation Regression Model Summary

Model Number	Constant	GDP	GDP^2	CRLP	CLP	RFP	FRTP	CMT	STL	R^2	Adj. R^2	F-value
M2-1001	1.134 (8.98)	0.412 (5.95)	0.032 (5.66)	-	-0.004 (-3.27)	0.040 (2.32)	-	-	-	0.698	0.696	104.243
M4-1002	1.263 (8.48)	0.246 (3.77)	-	-	0.004 (3.88)	-	0.336 (3.25)	-	0.284 (3.82)	0.852	0.849	184.483
M6-1003	1.238 (12.38)	0.205 (3.39)	-	-	-	-	0.247 (2.65)	-	0.155 (2.35)	0.722	0.718	112.522
M8-1004	1.002 (10.33)	0.329 (3.48)	0.017 (3.68)	0.003 (3.84)	-	-	-	-	-	0.686	0.677	78.126
M10-1005	0.888 (9.32)	0.045 (4.28)	0.008 (3.48)	-	-	-	-	-	-	0.758	0.746	98.224
M12-1006	0.565 (5.44)	0.018 (4.76)	-	-0.055 (-2.44)	-	0.017 (3.94)	-	-	-	0.712	0.701	86.136
M14-1007	0.155 (13.75)	0.102 (6.38)	0.008 (6.22)	-	-	-	-	-	-	0.738	0.715	124.826
M16-1008	0.181 (4.75)	0.105 (4.46)	-	-	0.124 (3.84)	-	-	-	0.682 (4.98)	0.736	0.732	162.628
M18-1009	0.791 (8.81)	0.042 (7.78)	-	0.212 (5.34)	-	-	-	-	-	0.756	0.752	108.482
M20-1010	1.362 (5.84)	0.316 (4.94)	-	-	-	-	-	0.002 (3.18)	-	0.757	0.748	93.208
M22-1011	0.004 (4.24)	0.086 (1.16)	0.005 (1.02)	-	-	-	-	-	-	0.201	0.165	6.246
M24-1012	1.261 (4.44)	0.181 (6.65)	-	0.279 (4.74)	-	0.058 (5.61)	-	-	-	0.844	0.842	136.928

**(x.xx)= t-statistics value*

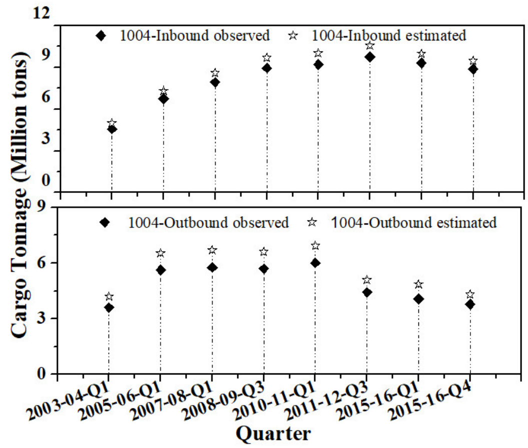
Kandla (1012) ports. Similar results were obtained for other ports. Fig. 5 shows the prediction error plot for both inbound and outbound cargo flow.

It can be seen from the Figure 5 that the error association with the regression prediction models varies from 6% to 20%. The outbound error is higher than the inbound error for all the ports. The possible reason could be the association between the inbound cargo and independent variables are stronger than that with outbound cargo. Additional representative variables may be considered for outbound and some conclusion can be drawn with more confidence. Conventionally, Ordinary Linear Square (OLS) regression models identify causal relationship, which is valid only for the stationary data series i.e., the series do not show any trend over a period of time. Although, the OLS regression model estimates the causal relationship, the comovement between the variables may be temporally inflated by virtue of time, creating a spurious relationship. Thus, it was planned to investigate the demand estimation through time series modeling approach with the use of monthly data. The subsequent sub-section discusses about the time series models developed in this study.

4.2. Time series models

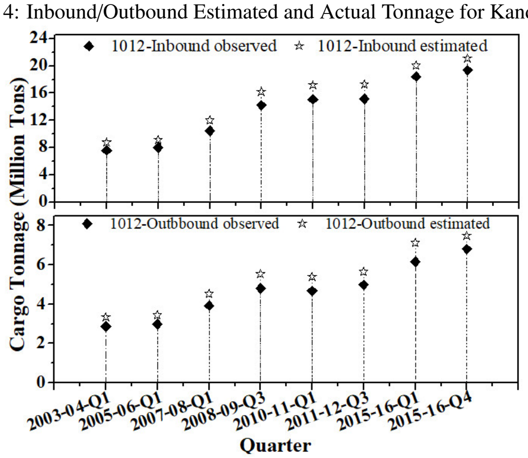
Time series models are used for short term predictions. The ARIMA modeling technique is used to estimate the cargo de-

Figure 3: Inbound/Outbound Estimated and Actual Tonnage for Chennai port



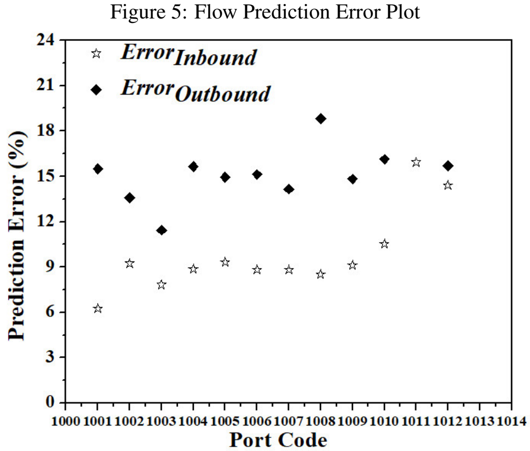
Source: Authors.

Figure 4: Inbound/Outbound Estimated and Actual Tonnage for Kandla port



Source: Authors.

Figure 5: Flow Prediction Error Plot



Source: Authors.

mand models. ARIMA, a regression based model introduced by Box and Jenkins (Box et al., 2008) is widely used to forecast univariate time series data. This model estimates future projections by regressing past values of the variable on itself and the current value with the error terms of the past values at different lag length. The model originated from autoregressive (AR) and moving average (MA) model with integration of order d. This study considered the seasonal component of the time series and the resulting model can be termed as SARIMA model.

4.2.1. Model structure

The time series modeling procedure involves with three main stages to build ARIMA or SARIMA model: 1) model identification; 2) model estimation and 3) model diagnostic to select the appropriate model. The SARIMA model structure (Box et al., 2008; Cryer and Chan, 2008) considered for the present study is given by Equation 2.

$$\phi_p(L)\Phi_p(L^S)(1-L)^d(1-L^S)^D Y_t = \theta_q(L)\Theta_Q(L^S).e_t \quad (2)$$

Where,

$$e_t \sim NID\,(0, \sigma^2)$$

Non Seasonal AR (p):

$$\phi_p(L) = 1 - \phi_1 L^1 - \phi_2 L^2 - ...... - \phi_p L^p ......(2.1)$$

Seasonal AR (P):

$$\Phi_P(L^S) = 1 - \phi_1 L^{1S} - \Phi_2 L^{2S} - ...... - \Phi_P L^{PS} ......(2.2)$$

Non Seasonal difference (d): $(1-L)^d$
Seasonal difference (D): $(1-L^S)^D$
L =Lag operator:

$$Y_{t-k} = L^k Y_t$$

d and D = Non-seasonal and seasonal order of differences.
Non Seasonal MA (q):

$$\theta_q(L) = 1 - \theta_1 L^1 - \theta_2 L^2 - ...... - \theta_q L^q ......(2.3)$$

Seasonal MA (Q):

$$\Theta_Q(L^S) = 1 - \Theta_1 L^1 S - \Theta_2 L^2 S - ...... - \Theta_Q L^Q S ......(2.4)$$

The model is truncated to SARIMA $(p, d, q)(P, D, Q)_S$. The SARIMA model reduces to pure ARIMA (p, d, q) when the time series do not have any seasonal effect.

4.2.2. Mode identification and estimation

The total number of available observations for monthly cargo flow are 161 for all ports except the Ennore port (125 observations for inbound and 101 observations for outbound). The last five data points were kept for validating the developed models. The time plot for cargo flow suggested that some of the ports have experienced seasonal flow. The stationary characteristics and the seasonality of the study data were examined using 1) graphical analysis i.e., time plot, Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) and, 2)

Table 4: Calibrated Inbound Flow Demand Estimation Time Series Models

Port Code	Model type	Model order (p,d,q) $(PDQ)_{12}$	AR 1 ϕ_1	AR 2 ϕ_2	SAR 12 Φ_1	SAR 24 Φ_2	MA 1 θ_1	SMA 12 Θ_1	Constant μ
1001	SARIMA	$0,1,1(0,1,1)_{12}$	--	--	--	--	0.793 (13.54)	0.841 (12.42)	0.118 (3.78)
1002	ARIMA	0,1,1	--	--	--	--	0.798 (14.43)	--	0.121 (3.74)
1003	ARIMA	1,1,1	0.359 (4.20)*	--	--	--	0.979 (51.52)	--	0.056 (9.02)
1004	ARIMA	0,1,1	--	--	--	--	0.746 (12.46)	--	0.208 (3.14)
1005	ARIMA	0,1,1	--	--	--	--	1.001 (61.64)	--	0.044 (18.24)
1006	ARIMA	0,1,1	--	--	--	--	0.922 (23.96)	--	0.117 (3.29)
1007	SARIMA	$0,1,1(0,1,1)_{12}$	--	--	--	--	0.990 (47.87)	0.863 (10.21)	0.002 (2.49)
1008	SARIMA	$0,1,1(0,1,1)_{12}$	--	--	--	--	0.810 (13.63)	0.858 (9.24)	0.108 (4.16)
1009	ARIMA	0,1,1	--	--	--	--	0.988 (49.41)	--	0.052 (15.80)
1010	ARIMA	1,1,1	0.461 (5.17)	--	--	--	0.977 (29.51)	--	0.007 (11.26)
1011	ARIMA	0,1,1	--	--	--	--	0.857 (18.43)	--	0.016 (2.53)
1012	ARIMA	2,1,1	-1.162 (-6.10)	-0.507 (-5.26)	--	--	-0.595 (-2.83)	--	0.351 (3.95)

t-statistics value

Table 5: Calibrated Outbound Flow Demand Estimation Time Series Models

Port code	Model type	Model order (p,d,q) $(PDQ)_{12}$	AR 1 ϕ_1	AR 2 ϕ_2	SAR 12 Φ_1	SAR 24 Φ_2	MA 1 θ_1	SMA 12 Θ_1	Constant μ
1001	SARIMA	$1,1,1(1,1,1)_{12}$	0.451 (4.07)*	--	-0.249 (-2.20)	--	0.913 (13.90)	0.836 (9.75)	0.048 (4.45)
1002	SARIMA	$1,2,1(1,2,1)_{12}$	-0.454 (-4.57)	--	-0.458 (-4.12)	--	0.914 (12.30)	0.825 (6.63)	0.013 (2.88)
1003	SARIMA	$1,1,1(1,1,1)_{12}$	0.505 (4.60)	--	-0.336 (-3.03)	--	0.931 (14.36)	0.856 (9.32)	0.015 (3.56)
1004	SARIMA	$1,2,1(1,2,1)_{12}$	-0.209 (-2.06)	--	-0.511 (-5.12)	--	0.998 (91.96)	0.890 (9.14)	0.041 (27.93)
1005	ARIMA	0,1,1	--	--	--	--	0.908 (20.68)	--	0.007 (2.27)
1006	ARIMA	0,1,1	--	--	--	--	0.636 (9.07)	--	0.001 (5.55)
1007	SARIMA	$0,1,1(0,1,1)_{12}$	--	--	--	--	0.776 (12.92)	0.871 (11.88)	0.056 (2.71)
1008	SARIMA	$2,3,1(2,3,1)_{12}$	-0.652 (-6.55)	-0.489 (-4.73)	-1.389 (-50.30)	-1.001 (-14.63)	1.025 (74.91)	0.751 (6.43)	0.718 (2.81)
1009	SARIMA	$1,2,1(1,2,1)_{12}$	-0.258 (-2.60)	--	-0.491 (-4.98)	--	1.005 (80.78)	0.902 (10.48)	0.002 (19.19)
1010	SARIMA	$1,1,1(1,1,1)_{12}$	0.504 (3.61)	--	-0.001 (-4.01)	--	0.841 (9.57)	0.849 (7.92)	0.135 (5.40)
1011	ARIMA	0,1,1	--	--	--	--	0.886 (21.80)	--	0.001 (4.55)
1012	SARIMA	$1,1,1(1,1,1)_{12}$	0.331 (2.41)	--	-0.241 (-2.12)	--	0.827 (10.33)	0.884 (9.64)	0.176 (4.12)

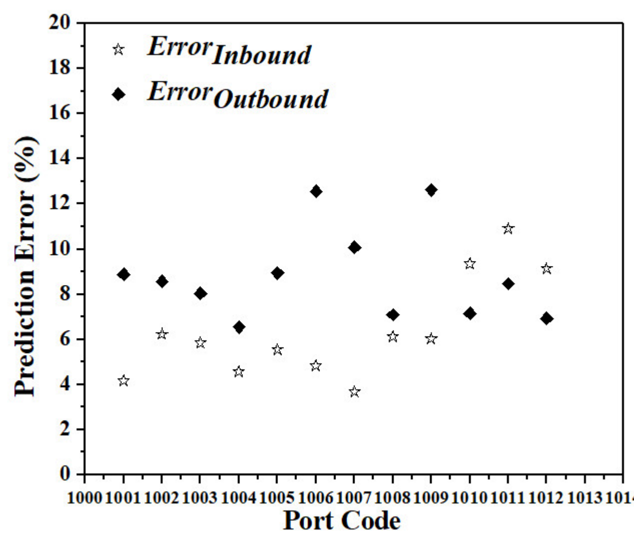
t-statistics value

unit root test i.e., Augmented Dickey-Fuller (ADF) test. The ACF, PACF plots and ADF test statistics values were compared with critical values at 0.05 significant levels and the conclusion is drawn that all the datasets for the 12 major ports are non-stationary. The datasets were differenced up to third order and the seasonality component is eliminated from the study data. The seasonality elimination was confirmed through ADF test. The ADF test results at stationary level are presented in Appendix I (see Table AI-1). The orders of the SARIMA/ARIMA models were determined from the resulting dataset and an array of models was developed to select the best models. The models were selected based on Mean Absolute Percentage Error (MAPE) value, t-statistics, and p-value. The model development exercise was accomplished using MiniTab 16.0. Tables 4 and 5 report the selected 24 ARIMA/SARIMA models for short term prediction of the monthly cargo flow for all the ports. The t-tests were used to assess the significance of individual coefficients for AR, SAR, MA and SMA component of the prediction model. A Portmanteau test called Modified Box-Pierce (Ljung-Box) Chi-Square test was carried out on the selected models and the statistical values are examined at lags 12, 24, 36 and 48 for residual analysis to check the appropriateness of the selected models. The reported demand estimation models were calibrated using maximum likelihood procedure. All the t-statistics values for all the parameters were found to be significant at 95% confidence level. From the residual analysis, it was found that the Ljung-Box Chi-Square statistics are not significant for each lag. All the p-values are well above 0.05 suggesting the acceptance of null hypothesis i.e., $O_i = E_i$ (observed value is equal to estimated value). This finding suggested that the residuals for all the selected models could be considered as white noise series. All these points discussed here confirmed the appropriateness of the order considered for the selected models.

4.2.3. Model validation

The models were validated using the last five months' data, which were not used during the model development process.

Figure 6: Error Plot on Flow Prediction



Source: Authors.

Table 6: Inbound Cargo Forecast for the Study Ports, September-December, 2016

	Inbound cargo (million tons)			
Port Code	September	October	November	December
1001	2.861	2.807	2.796	3.127
1002	4.609	5.593	5.538	5.406
1003	2.818	3.525	3.136	3.135
1004	2. 508	2.601	2.281	2.151
1005	1.839	2.319	2.273	2.341
1006	1.622	1.548	1.345	1.591
1007	1.922	2.138	2.049	2.416
1008	0.802	0.693	0.830	1.277
1009	4.083	3.948	3.950	3.953
1010	2.717	2.755	2.798	3.060
1011	2.435	2.215	2.194	2.090
1012	5.662	5.064	6.228	5.968

t-statistics value

Table 7: Inbound Cargo Forecast for the Study Ports, September-December, 2016

Outbound cargo (million tons)				
Port Code	September	October	November	December
1001	1.343	1.243	1.274	1.367
1002	1.836	2.007	2.066	1.927
1003	2.044	2.392	1.983	2.040
1004	1.532	1.517	1.597	1.432
1005	0.721	0.931	0.797	0.821
1006	0.372	0.324	0.283	0.335
1007	0.628	0.612	0.743	0.695
1008	1.534	2.058	2.221	2.594
1009	1.199	1.117	1.238	1.345
1010	2.462	2.391	2.456	2.607
1011	0.267	0.241	0.242	0.236
1012	2.678	2.328	2.814	2.743

**t-statistics value*

Using the models reported in Tables 4 and, 5; monthly cargo traffic flows were predicted for the last five months covering the 12 study ports. The estimated values were compared with the actual values and the average prediction errors were calculated. Fig. 6 shows the prediction error plot for the two cargo operations. The error ranged from 3.8% to 12.6%. Most of the ports (10 ports out of 12 ports) have prediction error values are within 8% limit for all the cargo operations, which is quite acceptable. This empirical evidence from the prediction error plot confirmed the validation of the models. It may also be noted that the time series models resulted in low prediction error as compared to the error obtained from the regression models using the same data.

It was discussed in sub-sub-section 4.1.3 that the average error varied from 6% to 20% while predicting the cargo flow using the calibrated regression models. However, the average prediction error ranged from 3.8% to 12.6% with the time series models. Most of the port cargo demands were predicted within error limit of 8% using the time series models. These predictions are better than that obtained by regression models.

5. Forecast generation

Time series models were used to generate the cargo forecast for inbound and outbound models. The models presented in Table 4 and 5 were recalibrated with the recent data (till April 2015) and the revised models were used to forecast cargo flow at all ports for the next 4 successive periods. The forecasts are given in Tables 6 and 7 for inbound and outbound cargo respectively.

Discussion and conclusions

This paper focused on empirical analysis on cargo demand forecast emerging from 12 major ports in India. Separate demand models were developed for inbound and outbound cargo movement. The models were developed using regression and time series techniques. These demand models can be used to estimate short to medium term cargo tonnage at Indian port locations. The results indicated the statistically significance of macroeconomic variables while estimating the seaborne cargo demand. The univariate regression models suggested that the amount of cargo movement increases with the increase in the gross domestic product of the nation over time. More than 70% of cargo movements at Indian port system are explained by the various economic drivers of the country. The seasonality variations of port cargo volume were captured through time series models. Overall, the time series models were found to exhibit better results in terms of prediction accuracy.

Based on forecast accuracy and reliability, SARIMA models may be preferred over OLS regression models. The validation results for the time series models revealed that forecasting error is within 8% in most of the cases with the maximum prediction error is 12.6%. The performances of inbound models were found to be better than the outbound models in terms of prediction accuracy. The limitation with the present modeling approach is that these models are not appropriate for long term prediction. This study is intended to provide infrastructure planners with some guidance on short to medium term development of transport infrastructure requirements for the ports. Also, it is expected that this study will be helpful to transportation planners, while planning for hinterland road network connectivity to port locations.

Acknowledgements

The authors wish to thank the Mumbai Port Trust Authority, Paradip Port Trust Authority and CMIE for the collection of the port cargo data and economic indicator data used in this study.

References

Afifi, A., May, S., and Clark, A.V. (2012) *Practical Multivariate Analysis*, CRC Press, Taylor and Francis Group Publications.

Al-Deek, M. H. (2001) Which method is better for developing cargo planning models at seaports ? neural networks or multiple regression? *Transportation Research Record* 1763: 90-97.

Al-Deek, M. H., Johnson, G., Mohamed, A., and El-Maghraby, A. (2000) Truck trip generation models for seaports with container and trailer operation. *Transportation Research Record* 1719: 1-9.

Balach, P., and R. R. Tadi (1994) Truck trip generation characteristics of nonresidential land uses. *ITE Journal* Vol. 64: 43-47.

Box, G.E.P., Jenkins G.M., and Reinsel G.C. (2008) *Time Series Analysis, Forecasting and Control*. New Jersey: John Wiley and Sons Inc. Publications.

Chen, C.F., Chang Y.H., and Chang Y.W. (2009) Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan, *Transportmetrica* 5: 125-140.

Cohen, H., Horowitz, A., and Pendyala, R. (2008) NCHRP Report 606: Forecasting statewide cargo toolkit. *Transportation Research Board, Washington, D.C., USA.*

Cryer, J.D., and Chan K.S. (2008) *Time Series Analysis: With Application in R*. New York: Springer Publications.

DBI. (2009) Indian Infrastructure Market: opportunities in the seaports sector, Department of Business and Innovation, State Govt. of Victoria, Australia, July 2009, *http://export.business.vic.gov.au/data/assets/pdf_file/0010/337-852/Indian-Infrastructure-Market-Opportunities-in-the-Sea-Ports-Sector.pdf*, last accessed on November, 2013.

Eddie C.M. H., William S., and Gordon K.C. W. (2004) Forecasting cargo throughput for the port of Hong Kong: Error correction model approach. *Journal of Urban Planning and Development* 130 (4): 195-203.

Faghri, A., and Hua, J. (1992) Evaluation of artificial neural network application in transportation engineering. *Transportation Research Record* 1358: 71-80.

Haralambides, H., and Gujar, G. (2012) On balancing supply chain efficiency and environmental impacts: An eco-DEA model applied to the dry port sector of India. *Maritime Economics and Logistics* 14: 122-137.

IPA. (2013) E-Magazine, Indian Port Association December 2013 Report, *www.ipa.nic.in/e-magazine.pdf*, last accessed on January 2014.

ISI. (2010) Indian Shipping Industry: A catalyst for growth, occasional paper no. 142, *http://www.eximbankindia.in/sites/default/files/Full%20OP/-op142.pdf*, last accessed February 2014.

Klodzinski, J., and Al-Deek, M. H. (2003) Transferability of an intermodal cargo transportation forecasting model to major Florida seaports. *Transportation Research Record* 1820: 36-45.

Makridakis, S., Wheelwright, C.S., and Hyndman, J.R. (2005) *Forecasting Methods and Applications.* John Wiley & Sons, Inc. Publications.

Middendrof, D.P., Jelavich, M., and Ellis, R.H. (1982) Development and application of statewide multimodal cargo forecasting procedures for Florida. *Transportation Research Record* 889: 7-14.

MoS. (2011) *Maritime Agenda: 2010-2020*. Ministry of Shipping, Government of India.

Panigrahi, K. J., and Pradhan A. (2012) Competitive maritime policies and strategic dimensions for commercial seaports in India. *Ocean and Costal Management* 62: 54-67.

Patil, G. R., and Sahu, P. (2016a). Estimation of cargo demand at Mumbai port using regression and time series models. *KSCE Journal of Civil Engineering*, Vol. 20(4), pp. 1514-1525, 2016, DOI: 10.1007/s12205-015-0386-0

Patil. G., and Sahu, P. (2016b). Simultaneous Dynamic Demand Estimation Models for Major Seaports in India., *Transportation Letters*, 2016, (DOI: 10.1080/19427867.2016.1203582)

Raghuram, G., and Shukla, N. (eds) (2014) Issues in PPPs in Ports in India, Public Private Partnership: *The Need of the Hour, Academic Reference Series*, Bloomsbury Publishing India Pvt. Ltd., New Delhi, 2014, ISBN: 978-93-82951-59-9.

RWGPS. (2011) Report of working group for port sector for the 12th five year plan (2012-2017), Ministry of Shipping, Government of India, *http://www.planningcommission.gov.in/aboutus/committee/-wrkgrp12/transport/report/wg_port.pdf*, last accessed on November, 2013.

Sahu, P., and Patil, G.R. (2013) Analysis and Modeling of Cargo Demand at Mumbai Port Using Regression and Time Series Techniques. *In Proceedings (CD-ROM) of* 92nd *Transportation Research Board Annual Meeting*, January 2013, Washington DC., USA.

Sahu, P, and Patil, G. (2014) Classification of Indian Major Seaports using Hierarchical Grouping Method and Their Demand Estimation Models., *In Proceedings of* 93rd *Transportation Research Board Annual Meeting*. January, 2014, Washington DC., USA.

Sahu, P., Sharma, S., and Patil, G. (2014) Classification of Indian seaports using hierarchical grouping method., *Journal of Maritime Research*, 11 (3), pp. 51-57.

Sculze, P. M. and Prinz A. (2009) Forecasting container transshipment in Germany, *Applied Economics*, 41.

Seabrooke, W., Hui, E.C.M., Lam, W.H.K., and Wong, G.K.C. (2003) Forecasting cargo growth and regional role of the port of Hong Kong. *Cities* 20 (1): 51 ? 64.

USDOT. (1999) *Guidebook on Statewide Travel Forecasting*. U.S.A. Department of Transportation, Federal Highway Administration.

Woo, S.-H., Petit, S.J., Kwak, D.-W., and Beresford, K.C.A. (2011), Seaport research: a structured literature review on methodological issues since the 1980s. *Transportation Research Part A: Policy and Practice*, 45: 667-685.