KSCE Journal of Civil Engineering (2016) 20(5):2022-2032

DOI 10.1007/s12205-015-0386-0

Transportation Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Estimation of Freight Demand at Mumbai Port using Regression and Time Series Models

Gopal R. Patil* and Prasanta K. Sahu**

Received July 14, 2014/Revised 1st: November 22, 2014, 2nd: February 28, 2015/Accepted July 8, 2015/Published Online October 23, 2015

Abstract

Forecasting future freight demand at a seaport is important for its planning and development. India has 13 major ports which handle 75% of the total seaport freight. Among the 13 major ports, Mumbai Port, ranked at number three in the country for the year 2013-14, handles about 11% of the total freight at major seaports in India. The focus of this paper is on developing inbound and outbound demand forecasting models for Mumbai Port. The models are developed using additive regression and time series techniques. In regression analysis economic indicators, Gross Domestic Product (GDP) and Crude Oil Production (CRLP) are found to be significant. The multivariate models performed better than the univariate models. The validation of time-series models resulted in error of less than 5%. Both multivariate regression and time-series models are used to forecast freight demand for the years 2014-15 through 2017-18. The regression models are producing more optimistic forecasts than the time series models. The elasticity analysis suggested that Mumbai's inbound freight will be growing almost with India's GDP growth rate, the outbound freight, however, will experience slower growth than that of inbound.

Keywords: *mumbai port, seaport freight demand estimation, regression analysis, time series models, elasticity*

1. Introduction

Seaports act as intermodal hubs for the movement of freight within national and international transportation networks. In recent times, liberal economic policies followed by advances in Information and Communications Technology (ICT) have had significant impact on seaport transportation and its associated infrastructures across the globe. Developing economies like India are also experiencing a new growth scenario in port transportation. The Indian economy is expected to grow at an average of 6 to 8% by Gross Domestic Product (GDP) in the near future. This growth rate estimates the transport requirement in the country to be increased by 7.5 to 10% with elasticity of 1.25 at the GDP growth in medium and long term range (Planning Commission, Govt. of India, 2011; Dept. of State Development, Business and Innovation, State Govt. Victoria, 2009). Indian ports handle around 95% of the total volume of the country's international trade and about 70% in terms of value. India has 13 major ports and 187 non-major (intermediate and minor) ports (Planning Commission, Govt. of India, 2011). Major ports handle 75% of the total seaborne freight. Mumbai Port is one of the major ports, and has experienced a significant growth in freight movements in the last decade. The total freight (outbound and inbound inclusive) handled by Mumbai Port was 26.71 million tons in 2001-2002, whereas the same in 2013-2014 was 59.19 million tons (MPT, 2014). It is ranked at number 3 in India for the year 2013-14, and it shares about 11% of the total freight throughput for all the major ports in India. The financial year in India is from 1st April to 31st March, thus two calendar years are mentioned for each financial year.

In order to handle the future demand efficiently, planners need to evaluate the existing infrastructure within and outside the seaport. If necessary, seaport capacity augmentation or land connectivity expansions will have to be initiated to meet the future demand. Since such projects are resource-and time-intensive and cannot be reversed, a systematic analysis and forecasting of freight demand is necessary. The quantity of seaborne freight in a country can be related to the macroeconomic indicators of the country (Seabrooke *et al.*, 2003). We developed univariate and multivariate regression models for demand estimation by treating macroeconomic indicators as explanatory variables. Additionally, time series models are developed which use past demand trend to estimate the future demand. The past inbound and outbound freight flow data are obtained from Mumbai Port (MbP) database and the economic indicators obtained from Centre for Monitoring Indian Economy (CMIE) database.

This paper is structured as follows: Section 2 discusses the

*Assistant Professor, Dept. of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India (Corresponding Author, E-mail: gpatil@iitb.ac.in)
**Assistant Professor, Dept. of Civil Engineering, Birla Institute of Technology and Science Pilani, Pilani 333031, India (E-mail: prasantsahu222@gmail.com)

literature review. A brief report on Mumbai port is presented in Section 3. Section 4 outlines the data collection process and preliminary analysis of the data. Development of additive regression and time series models is described in Section 5. Section 6 reports the validation and comparison of the developed models. Forecasting of freight using the developed time series and regression models is demonstrated in Section 7, and Section 8 concludes the paper.

2. Literature Review

Freight forecasting models can be grouped into five classes: 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; Chow *et al*., 2010). The FFM is relatively simple and is used by state Departments of Transportation across the USA. The FFM is based on regression equations (econometric analysis and time series analysis). The available sample of literature revealed that standard regression analysis is widely used for freight demand estimation. A structural review of available literature by Woo *et al.* (2011) on freight demand estimation reported that about 13% of the total published literature used regression techniques for seaport freight data analysis. The same study also reported that the use of time series modelling towards seaport freight demand estimation is very limited in the existing literature. The Federal Highway Administration's Guidebook (1999) on state-wide travel forecasting focused on Autoregressive Integrated Moving Average (ARIMA) models and linear regression model to forecast truck volumes. Balbach and Tadi (1994) examined the use of regression procedure to build a truck trip generation model for computing truck trip rates with several land use categories and levels of activities as independent variables. Seabrooke *et al.* (2003) predicted the freight growth at the Hong Kong port by means of regression analysis. They considered GDP as their explanatory variable while predicting the future freight flow for Hong Kong port. Coto-Millán *et al.* (2005) explained the determinants of marine exports and imports in Spain using log linear regression model structure. The explanatory variables considered were GDP, maritime transport service price, price of non-energy industrial goods, world income, etc.

In recent years, methods based on artificial intelligence, such as, neural network and fuzzy logic are also being used in demand forecasting. Al-Deek (2001) investigated the use of neural networks with multiple regression analysis for developing freight 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 freight forecasting models for major Florida seaports. These models were used to estimate daily truck trips in and out of the ports. The authors compared artificial neural networks with multiple regression models. It was concluded that neural networks were

Table 1. (a) Summary of Past Studies on Port Freight Demand Estimation

Researchers Port Name	Modeling Approach Result	Data Sources Data Type	Dependent variable	Independent variables	Port Area Quay Length	Freight handled 2013
Seabrooke *et al.*, (2003) Hong Kong	Regression $R^2 = 0.76$ to 0.97 Error = < 4%	Hong Kong Port Maritime Board, National Bureau of Statistics of China, Statistical Bureau of Guangdong Annual data	Inward and outward freight movements, 17 key commodities	Hong Kong GDP, Trade value of Import/Export/ Re-export, Population, Electricity demand, Expenditure on building and construction	201 hectares 5794mts	22.35 million TEUs
Hu*i et al.,* (2004) Hong Kong	Regression R^2 =0.96 Error = < 5%	Shipping and Cargo Section of Census and Statistics Department in Hong Kong, International Monetary Fund, Shenzhen Information and Statistics Bureau Quarterly data	Total freight throughput	China's total trade value, USA total trade value, No. of berths of container terminal, Cargo throughput at the ports of Shenzhen	271hectares, 7734mts	22.35 million TEUs
Lam et al., 2004 Hong Kong	Neural Network $R^2 = 0.95$ NN models perform better than regression models	Census and Statistics Department of the Hong Kong Special Administrative Region HKSAR government Annual data	Inward and outward freight movements, 37 key commodities	Trade value of imports/ exports/re-exports, population, electricity demand, and Hong Kong GDP	271hectares, 7734mts	22.35 million TEUs
Chen and Chen, 2010 Keelung Taichung Kaohsiung	Genetic Programming, Time series modeling Genetic programming predictions were 32-36% better than time series predictions Error = 4%	Respective port maritime board Monthly data	Container volume	Past container volume in terms of Twenty-foot Equivalent Unit (TEU)	Not available	Kaohsiung 9.94 million TEUs
Gosasang *et al.*, 2011 Bangkok	Regression and Neural Network $R^2 = 0.86$ for Regression model $R^2 = 0.95$ for NN Neural network models are better than regression models	Bank of Thailand; National Economic and Social Development Board; World Bank; Ministry of Interior and Energy Policy, and Planning Office Monthly data	Container volume	Thailand GDP, World GDP, Exchange rate, Population, inflation rate, interest rate, fuel price	381 hectares 1528mts	20.36 million tons

Table 1. (b) Summary of Past Studies on Port Freight Demand Estimation Contd

Researchers Port Name	Modeling approach Result	Data Sources Data Type	Dependent variable	Independent variables	Port Area Quay Length	Freight handled 2013
Liu and Park, 2011 Busan and Gwang-Yang, Shanghai, Shenzhen, Dalian, Qingdao	Regression $R^2 = 0.81$ Port tariff, transshipment, investment, hinterland export-import volume are the strongest factors	Respective port authorities, Government of Korea, Government of China Annual data	Container volume	Terminal storage facility, berth length, direct call liner, Hinterland GDP, Hinterland export –import volume, port's tariff, FTZ area, Government investment	Busan 84 hectares 4144mts; Gwang-Yang 14.5 hectares 2750mts Shanghai 361900hectares 20000mts; Shenzhen 5405mts, Dalian 1500hectares 41000mts	Shanghai 33.63 million TEUs Busan 17.69 million TEUs Shenzhen 23.28 million TEUs
Al-Deek, 2001 Miami and Jacksonville	Regression and Neural Network Neural network model worked better than neural network model	Data collected from site survey Daily data	Inbound and outbound heavy truck volume	Number of imported and exported containers	---	--
Klodzinski and Al-Deek, 2003 Major Florida ports	Neural network model	Truck Data collected from the field and Vessel freight data given by port authorities Daily data	Daily inbound and outbound truck volumes	Vessel freight data i.e. import/export volume	--	--
Langen et al., 2012 Hamburg – Le Havre range ports in France, Belgium, the Netherlands, Germany	European transport network model: Trans-Tools*; Expert judgment	Various port authorities Annual data	11 major commodity flows	Past commodity flows data	--	--

*More information about Trans-Tool model can be found on the website of the Joint Research center of the European commission: http://energy.jrc.ec.europa.eu/transtools
**Sources: Various; -- indicates that data was not available to us.

more versatile and accurate tools for modelling truck trips generated by seaport freight activity. Al-Deek *et al.* (2000) used time series models for predicting seasonal variations in freight movements for the port of Miami.

The key results obtained from the studies discussed here along with several other studies that have used regression, time series, or artificial intelligence to predict port freight for various ports around the world are presented in Table 1(a) and (b). Table 1 show that there are significant associations ($R^2 >= 0.76$) of macroeconomic indicators with port freight for several ports. The national GDP is the most preferred causal variable used in most of the studies. Some studies (Lam *et al*., 2004; Gossang *et al*., 2011; Al-Deek, 2000) reported that neural network prediction is better than regression models in estimating port freight demand. Most of the cited literature here focused on estimating the inbound (import) and outbound (export) freight flow separately using univariate regression approach. The use of Multivariate Linear Regression (MVLR) for seaport freight flow estimation is not found in the literature, although it has been used for demand estimation in some other areas. Literature on analysis and forecasting of freight demand in India is scarce, and we did not find any past study on the systematic analysis of freight demand at Mumbai Port.

Multivariate linear regression has been widely applied in several fields like engineering, marketing research, biometrics, econometrics, and many other related research fields to study relationships between multiple responses and a set of explanatory variables. The multivariate linear regression model has been found producing better forecast in situations, where the response variables are associated (Granian *et al.,* 2015; Akinbinu, 2010; Siddiqui *et al*., 2013; Tolosana-Delgado *et al*., 2010). In case of multivariate regression, explanatory variables also explain the covariance structure of response variables. This is one of the reasons that multivariate regression models appear to be better predictive models over univariate regressions, in situations, where response variables are associated (Rencher and Christensen, 2012; Timm, 2002). Also, the joint distribution of response variables reduces the residual variance in case of multivariate models, thus models produce more accurate predictions. In the present study, the two response variables are inbound and outbound freight, which are assumed to be associated. Therefore, the present study applied both and univariate and multivariate regression models to model the freight flow at Mumbai port.

3. Data Collection and Analysis

Mumbai Port, formerly known as Bombay Port, lies at Latitude 18º 56.3' N, Longitude 72º 45.9' E on the West coast of India, and has a natural deep water harbour of about 400 square kilometres. Mumbai port has a vast hinterland (see Fig. 1) covering the whole of Maharashtra and some parts of Madhya Pradesh, Gujarat, Rajasthan, Delhi, and Uttar Pradesh. This hinterland is very rich in agricultural and industrial resources. The entire hinterland has undergone large scale economic improvement which has helped in the rapid growth of this port. A dense network of roads and railways connects the port with its hinterland. The hinterland along with major trade routes for Mumbai port is shown in Fig. 1.

The port handles various types of freight such as liquid bulk, dry bulk, containers, general cargo, etc. The commodities include

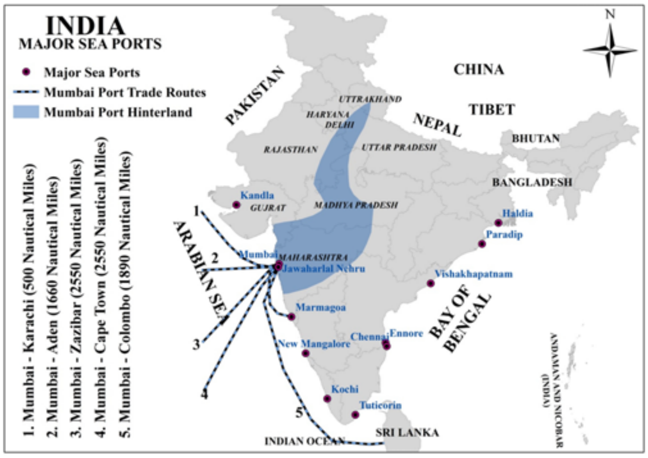

Fig. 1. Mumbai Port's Hinterland and Trade Routes

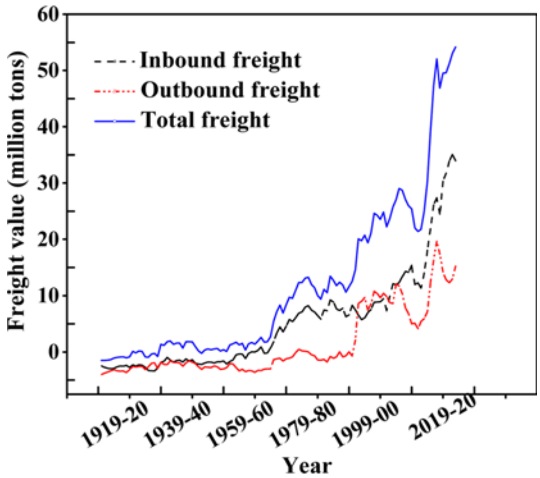

Fig. 2. Inbound, Outbound, Total Freight at Mumbai Port

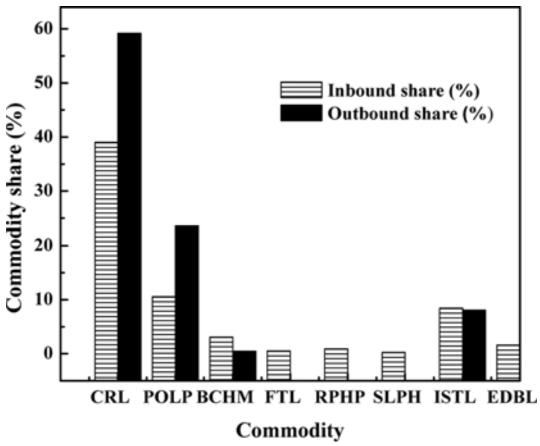

Fig. 3. Commodity Share at Mumbai Port

crude oil, Petroleum Oil and Lubricant (POL) products, iron and steel, fertilizers, edible oil, food grains, and sugar. The POL products include Liquefied Petroleum Gas (LPG), bitumen, lubes, Aviation Turbine Fuel (ATF), etc. The port is administered by the Mumbai Port Trust (MbPT).

The annual freight flow (inbound and outbound) data from the year 1900-01 to 2013-14 are obtained from MbPT. We also collected commodity wise inbound/outbound freight data from 2002-03 through 2013-14. The data for several economic indicator variables like Gross Domestic Product (GDP), Foreign Exchange Reserve (FER), Crude Oil Production (CRLP), and Personal Income (PI) are obtained from the CMIE database. The span of the received economic indicator data is from 1948-49 to 2013-14. In the database from CMIE, the values of economic indicators until year 2006-07 are in the base year 1999-00 prices and the remaining data are in 2004-05 base year prices. We converted all the values to a common base year 2004-05.

The inbound and outbound flow data are examined to analyse the freight growth pattern (Fig. 2). The graphs in the figure indicate that the inbound, outbound, and total freight operations are growing exponentially in recent decades. The inbound freight flow appeared to be stationary until the year 1953-54, and then started increasing rapidly. The outbound flow, however, did not increase much until the year 1980-81, after which it increased significantly. There was reduction in the overall freight activities during 1970-1978. This may have been due to the oil crisis of 1973, since crude oil (CRL) is the primary commodity handled at Mumbai Port (see Fig. 3). The inbound and outbound freight flow started falling from 1999-2000 onward, and Mumbai Port experienced the lowest freight flow in 2002-03 during the years from 2000 to date. This was the result of the recession in the global economy in the year 2000.

Figure 3 shows 12 years annual average percentage shares of various commodities handled by Mumbai Port. The principal outbound and inbound commodities include crude oil, POL products, iron and steel. Crude oil (CRL) stands first, with shares of 40% in the inbound commodities and 60% in the outbound commodities. Standing second, POL products (POLP) account for 11% of inbound and 25% of outbound freight. Iron and steel (ISTL) contribute equal shares in both inbound and outbound freight. The other commodities include bulk chemicals (BCHM), fertilizers (FTL), rock phosphate (RPHP), sulphur (SLPH), and edible oil (EDBL).

4. Port Freight Flow Models Development

We developed standard additive regression and time series models to estimate the freight flow. Statistical Package for Social Sciences 16.0 (SPSS 16.0) and Statistical Analysis Software 9.2 (SAS 9.2) are used for descriptive analysis and model development. Univariate and multivariate models are developed for inbound and outbound flow and are compared. In the former, independent models are developed for inbound and outbound demand. In the latter, both inbound and outbound freight flow models are estimated simultaneously. The inbound freight includes imported goods from various countries as well as freight from other Indian ports. Similarly, the outbound freight includes export and transport of goods to other Indian seaports. Therefore, it is believed that the inbound and outbound freight are related to each other. In such situations, where the response variables are related, multivariate regression models may result with better forecasting capability (Rencher and Christensen, 2012; Tim, 2002).

4.1 Regression Models

The central theme of the present study is to forecast the inbound and outbound freight. The forecasts are based on the assumption that the inbound/outbound freight is a linear function of national macroeconomic indicators. The data for various economic indicators were available from 1948-49 onwards. Thus, regression models are developed using freight flow and economic indicator data starting from the year 1948-49. Before developing regression models, outliers from the data sets are removed through scatter plot diagnostics, and the outlier removal is also confirmed from the residual analysis of the developed models. After the removal of outliers, the total number of observations for regression analysis is 64, out of which 10 (about 16%) are randomly selected and kept for model validation. The variables considered for the Port Freight Forecast (PFF) model development are given below:

Possible Explanatory Variables

- Gross Domestic Product (GDP)in trillion Indian Rupees
- Crude Oil Production (CRLP) in million tons
- Foreign Exchange Ratio (FER) in trillion Indian Rupees
- Personal Income (PI) in trillion Indian Rupees

Response Variables

- Inbound tonnage in million tons
- Outbound tonnage in million tons
- The structure of univariate regression model is as follows.

$$y = \beta_0 + \sum_{r=1}^{m} \beta_r x_r + \varepsilon \tag{1}$$

Where,
y = *Inbound or Outbound Tonnage*
x_r = *Economic indicator variable*
ε = Random error and $\varepsilon \sim N(0, \sigma^2)$ i.e., $E(\varepsilon) = 0$; $Cov(\varepsilon) = \sigma^2$

The structure of multivariate regression model is as follows (Timm, 2002).

$$[Y]_{n\times p} = [X]_{n\times q}[B]_{q\times p} + [\varepsilon]_{n\times p}$$

Where, $q = k + 1$

$$Y_{n\times p} = \begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}; \; B_{q\times p} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \cdots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \cdots & \cdots & \beta_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \cdots & \beta_{kp} \end{bmatrix}$$

$$X_{n\times q} = [1_m, x_1, x_2, \ldots\ldots, x_k] \quad \varepsilon_{n\times p} = [\varepsilon_1, \varepsilon_2, \ldots\ldots, \varepsilon_p]'$$

Where, ε = *Random error and ε's are independently normally distributed.*

$$E(\varepsilon_{(i)}) = 0 \quad Cov(\varepsilon_{(i)}, \varepsilon_{(j)}) = \sigma_{ij} I \quad i, j = 1, 2, \ldots\ldots\ldots\ldots, p$$

4.1.1 Estimation of Univariate Regression Models

We started regression analysis with one explanatory variable. We tried with the explanatory variables listed above in both linear and nonlinear forms. Different nonlinear forms are also tested. Based on the performance, we selected power regression models for both inbound and outbound freight estimation (Eqs. (2) and (3)).

Models M1 and M2:

$$Inbound = 4.232(GDP)^{0.516} \tag{2}$$

$$Outbound = 1.322(GDP)^{0.748} \tag{3}$$

Where, GDP: Gross Domestic Product in trillion Indian Rupees

Furthermore, Multiple Linear Regression models (MLR) are developed for inbound and outbound flow separately. Different combinations of explanatory variables are tried, it is found that variables GDP and CRLP are significant and give logical signs. The final models selected are presented below (Eqs.(4) and (5)).

Models M3 and M4:

$$Inbound = 8.441 + 0.588GDP - 0.048CRLP \tag{4}$$

$$Outbound = 3.388 + 0.074GDP + 0.318CRLP \tag{5}$$

Where,
GDP: Gross Domestic Product in trillion Indian Rupees
CRLP: Crude Oil Production in million tons

Various statistical parameters including R^2, F-value, t-statistics, and p-values for all univariate regression models are given in Table 2. The t-statistics and p-values indicate that all the variables selected are significant at 95% confidence level.

Table 2. Statistical Parameters for Univariate Models

Freight flow	Model	R^2	Adj. R^2	F-value	t-statistics (p-values)		
					β_0	GDP	CRLP
Inbound	M1	0.860	0.858	256.912	16.52(0.000)*	2.92(0.012)	
	M3	0.943	0.940	271.126	3.28(0.022)	14.29(0.000)	-3.82(0.002)
Outbound	M2	0.773	0.772	138.264	12.62(0.000)	2.54(0.014)	
	M4	0.827	0.825	72.721	2.23(0.043)	2.28(0.034)	4.51(0.000)

*The values in parenthesis are p-values

Although the multiple linear regression models have higher R^2, the single variable non-linear models also have good R^2. The R^2 values varied from 0.773 to 0.943 for all the models, suggesting that Mumbai port freight is closely associated with national economic indicators like GDP and crude oil production. In other words, more than 77% of freight flow variations at Mumbai port is explained by national GDP alone and more than 82% is explained by GDP and crude oil production. The accuracy of our models is comparable with models in other studies listed in Table 1.

It can be observed from Eqs. (4) and (5) that the increase in GDP results in increased inbound as well as outbound freight flows, but the impact of GDP change on the inbound freight is higher than that on the outbound. This can be explained as follows. Outbound freight mostly consists of crude oil (CRL) and petroleum products (POLP) (see Fig. 2), whereas the inbound freight comprises of several other commodities like fertilizer (FTL), bulk chemicals (BCHM), edible oil (EDBL), etc., in addition to CRL and POLP. CRL and POLP produced in Mumbai port's hinterland are not much influenced by India's GDP, whereas the inbound CRL and POLP are affected by it. For example, if GDP is higher, there will be higher demand for CRL and POLP products. Since, the hinterland supply of these products is limited; the higher demand should be met by sources outside the hinterland, which will increase the inbound flow. Moreover, higher GDP will increase the demand for other inbound commodities at Mumbai port. Therefore, the GDP has more impact on inbound freight than outbound at Mumbai port. CRLP has negative sign for inbound and positive for outbound freight flow. In other words, increase in crude oil production (CRLP) increases outbound but decreases inbound freight flow. This is because a major share of outbound CRLP at Mumbai port comes from oil reserves located on the western coast; the oil field is called Bombay High.

4.1.2 Residual Analysis and Variance Inflation Factor (VIF) for Univariate Regression Models

We further investigated the regression models for homoscedasticity, adjusted R^2, and normality. Figs. 4 and 5 present the scatter plots and the normality test results for residuals. The residuals in Figs. 4(a) and 5(a) are randomly scattered around 0 without any pattern, thus the homoscedasticity assumption is not violated. Normal probability plots of residuals for inbound and outbound flow are presented in Figs. 4(b) and 5(b) respectively. It can be noted that all the points are within the lower and upper bounds at 95% confidence level. This satisfies the normality assumption for the data used.

The variables selected for inclusion in the final models are GDP and CRLP. The VIF value for these variables in the final multivariate models is 2.688. The VIF value is lower than 10, which means the effect of multicollinearity among GDP and CRLP does not exist.

4.1.3 Estimation of Multivariate Regression Models

Multivariate linear regression (MVLR) models are developed using Statistical Analysis Software (SAS). The multivariate models are given by Eqs. (6) and (7). Table 3 reports various statistical parameters for these regression models. It can be noted that all the variables are significant at 95% confidence level.

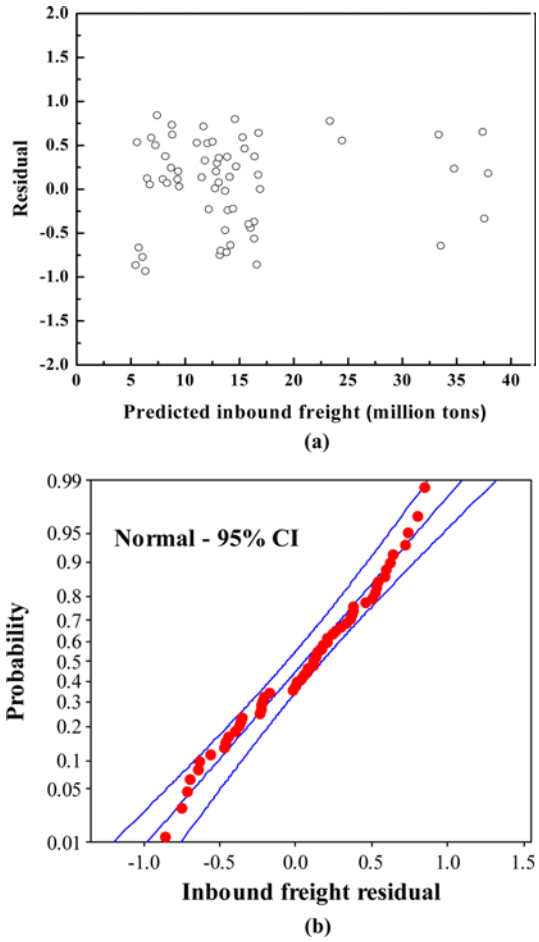

Fig. 4. Residual and Normality Plot for Inbound Freight: (a) Residual Scatter Plot, (b) Residual Normality Plot

Models M5 and M6:

$$Inbound = 5.135 + 0.622GDP + 0.055CRLP \quad (6)$$

$$Outbound = 2.502 + 0.089GDP + 0.412CRLP \quad (7)$$

Where,
GDP: Gross Domestic Product in trillion Indian Rupees
CRLP: Crude Oil Production in million tons

All the variables in the multivariate models have the same signs as for the univariate multiple linear regression models. The values of intercepts are smaller and the coefficients of GDP are higher in the multivariate models than in the univariate models. The values of R^2 are higher compared to that of univariate models. It is observed that R^2 and Adj. R^2 for the selected models are very close (i.e., the difference between R^2 and Adj. R^2 is < 0.05). The comparison of these values checks the over fitting of the models with the explanatory variables selected for modelling.

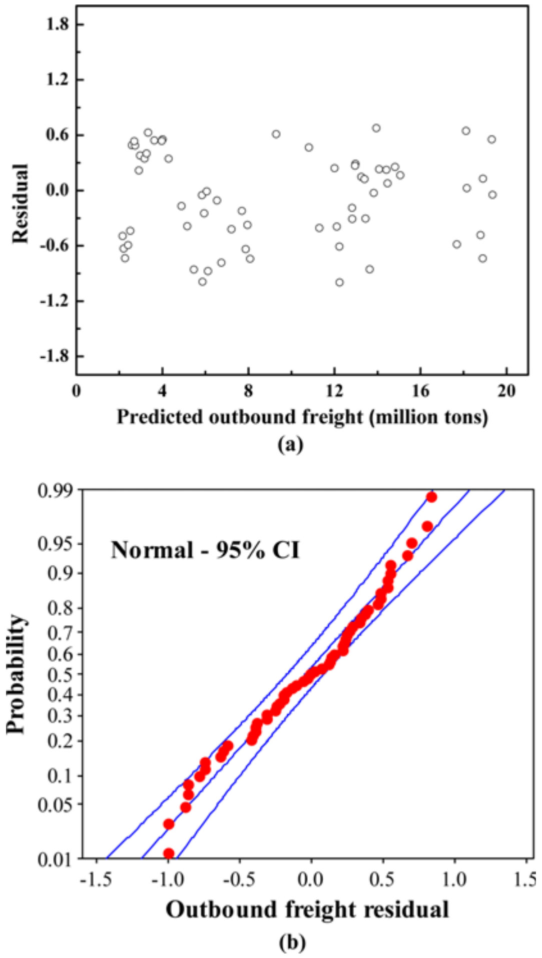

Fig. 5. Residual and Normality Plot for Outbound Freight: (a) Residual Scatter Plot, (b) Residual Normality Plot

4.2 Time Series Models

Time series analysis is another ideal approach to examine the effect of time on relationship among the time ordered variables (Box *et al*., 2008). In the past, some studies (Al-Deek *et al.*, 2000; Schulze and Prinz, 2009) have used single equation autoregressive integrated moving average (ARIMA) time series models to forecast seaport freight. The ARIMA model—also known as Box-Jenkins forecasting model—is a model in which forecasted values are obtained by regressing past values of the variable itself and the current value with the error terms of the past values at different lag length. In the present study, we used a times series regression model structure instead of a pure auto regression model to understand how much forecasting capability changes from a static model (regression model) to a time series model. The proposed model structure is given in Eq. (8).

$$Y_t = C + \beta_1 Y_{t-1} + \beta_2 GDP + \beta_3 CRLP + \varepsilon \tag{8}$$

Where, Y_t = Inbound/Outbound freight in million tons at time t. The remaining variables and parameters are self-explanatory. The same data sets as in case of regression models are chosen to estimate and validate the above time series regression model. The estimated models are presented in Eqs. (9) and (10). Calibrated coefficient values along with relevant statistical parameters are given in Table 4.

Models M7 and M8:

$$Inbound = 2.261 + 0.709\, Inbound_{t-1} + 0.204\, GDP - 0.018\, CRLP \tag{9}$$

$$Outbound = 0.664 + 0.839\, Outbound_{t-1} + 0.025\, GDP + 0.042\, CRLP \tag{10}$$

5. Validation and Comparison

We randomly selected 10 years (about 16% of the total data) freight flow data to validate all models. The validation data set is not used for the estimation of the models. The inbound and outbound flows are estimated using models M1 through M8 and the estimation errors are presented in Table 5. The errors for univariate regression models vary from about 13% to 23%, whereas the error for multivariate models lies between 6% and 12% at 95% confidence level. This observation is in line with our earlier observation made based on statistical parameters in section 4 that multivariate models are better. The time series regression models are also validated using the same validation data. The error from the time series models are significantly lower (i.e., 1% to 5%) than that from regression models. It can be seen from Table 1 that prediction capability of time series regression models is better or comparable with the prediction accuracy of models for other ports.

Table 3. Statistical Parameters for Multivariate Models

Freight flow	Model	R^2	Adj. R^2	F-value	t-statistics (p-values)		
					β_0	GDP	CRLP
Inbound	M5	0.972	0.970	474.284	3.16(0.021)*	2.93(0.005)	-2.42(0.025)
Outbound	M6	0.869	0.864	109.416	3.24(0.000)	3.61(0.001)	6.83(0.001)

*The values in parenthesis are p-values

Table 4. Time Series Regression Model Parameter Estimates

Freight flow	Model	R^2	Adj. R^2	F-value		t-statistics (p-values)		
					β_0	Y_{t-1}	GDP	CRLP
Inbound	M7	0.974	0.973	722.851	4.28(0.002)*	8.15 (0.000)	3.69(0.000)	-2.07(0.034)
Outbound	M8	0.932	0.930	270.532	2.69(0.026)	9.75(0.000)	2.32(0.023)	3.84(0.003)

*The values in parenthesis are p-values

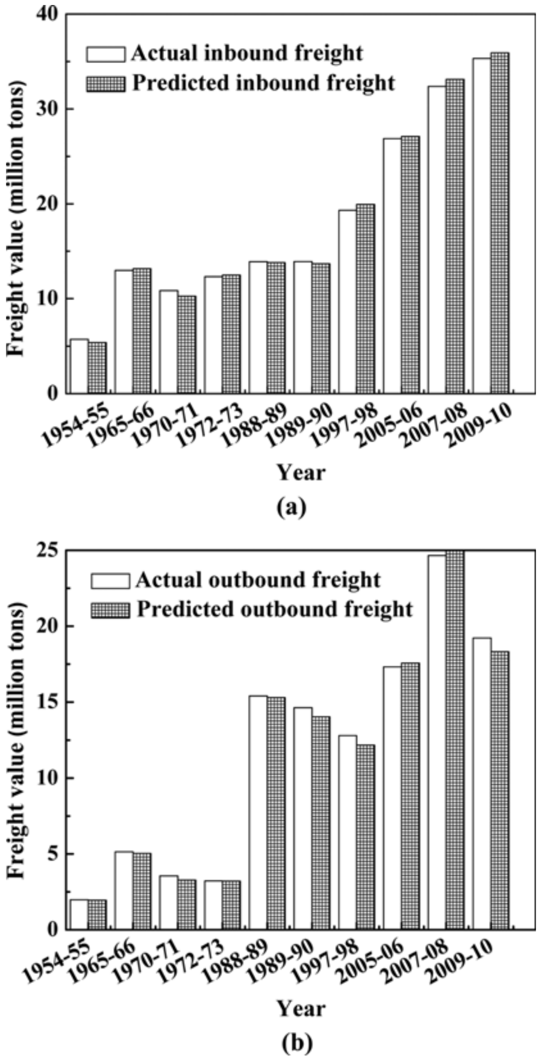

Fig. 6. Comparison of Actual and Estimated Freight using Time Series Models: (a) Inbound, (b) Outbound

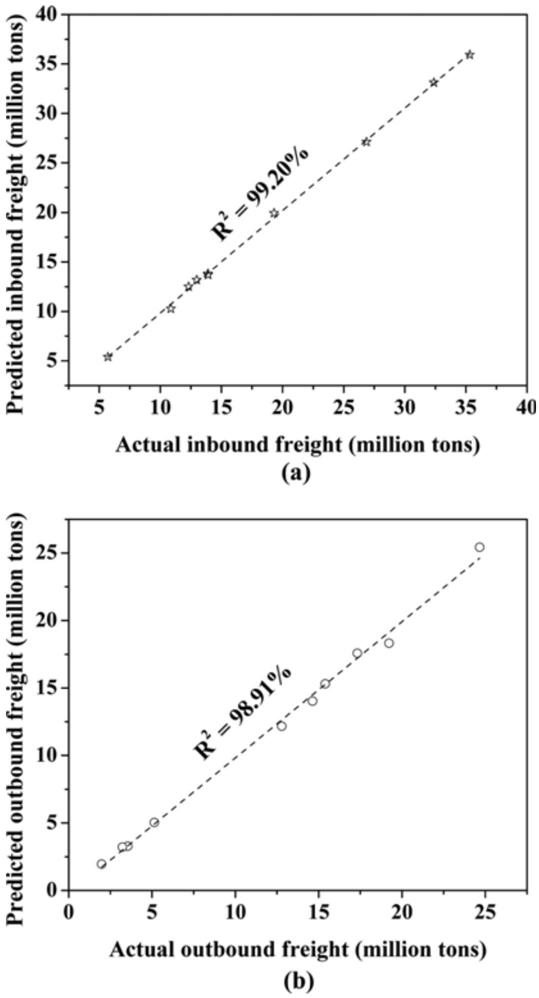

Fig. 7. Actual Versus Estimated Freight with Time Series Model: (a) Inbound, (b) Outbound

Figures 6(a) & (b) show the comparison of the actual and estimated values with time series regression models. Figs. 7(a) and 7(b) also show that the estimated values are close to actual values (R^2 for the line fitted between predicted and actual values is 0.992 for inbound and 0.989 for outbound freight flow). Therefore, we suggest using time series regression models to predict freight flow at Mumbai port.

6. Forecasting and Elasticity Analysis

6.1 Forecasting Freight Flow for Mumbai Port

Using the regression and time series models developed in section 5, we forecasted demand for years 2014-15 to 2017-18. We selected models M5, M6, M7, and M8 for forecasting the

Table 5. Prediction Errors (%) with Validation Data

Freight	Model	Modelling Approach	Mean Prediction Error (%)	Error Range (C. L. 95%)
Inbound	M1	Univariate regression (nonlinear)	12.72	9.48% - 15.96%
	M3	Univariate multiple regression	15.67	12.59% - 18.75%
	M5	Multivariate regression	8.56	6.05% - 11.07%
	M7	Time series regression	2.47	1.16% - 3.78%
Outbound	M2	Univariate regression (nonlinear)	18.92	15.36% - 22.48%
	M4	Univariate multiple regression	17.14	14.91% - 19.37%
	M6	Multivariate regression	9.42	6.79% - 12.05%
	M8	Time series regression	3.04	1.39% - 4.69%

Table 6. Forecasted Freight Demand (million tons)

Modelling Approach	Freight	Forecasted Freight			
		2014-15	2015-16	2016-17	2017-18
Multivariate Linear Regression	Inbound	39.20	42.43	45.98	49.71
	Outbound	26.17	25.70	25.63	26.86
	Total	65.37	68.13	71.61	76.57
Time Series Regression	Inbound	41.40	44.01	46.80	49.71
	Outbound	21.04	21.71	22.34	23.13
	Total	62.44	65.73	69.14	72.85

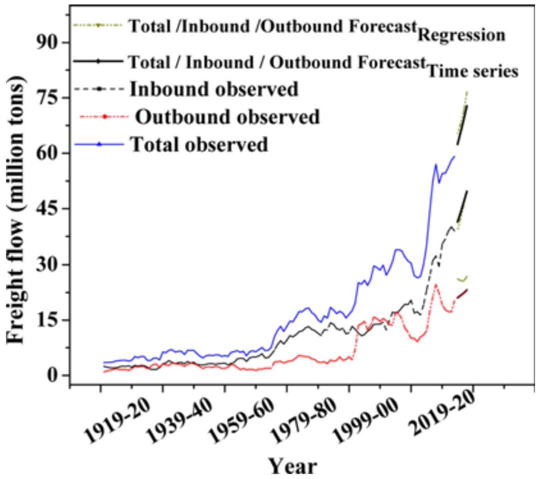

Fig. 8. Actual and Forecasted Freight Flow

Mumbai port freight. We used International Monetary Fund projected GDP values (IMF, 2014); for crude oil production we used the projection values published by Ministry of Petroleum and Natural Gas, Government of India (IPNG, 2013).The forecasted demand from M5 and M6 is presented in the upper part of Table 6. The forecast from time series regression models M7 and M8 are presented in the lower part of Table 6. The forecasted freight flow for time series models is superimposed over the past trend shown in Fig. 2; the resulting plot is given in Fig. 8.

The values in Table 6 indicate that the inbound flow in Mumbai port is continuing the steep upward trend through year 2017-18. Although the port will experience an increasing trend in outbound flow, the growth rate is predicted to be lower than that for the inbound flow. The multivariate regression models' forecasts are more optimistic than those of the time series models.

6.2 Elasticity Analysis

Elasticity can be defined as percentage change in freight flow for a 1% change in a decision attribute. In the present study the decision attributes are GDP and CRLP. We carried out elasticity analysis on models M5 and M6 to understand the implications of India's GDP or crude oil production on the value of inbound/ outbound freight of Mumbai port. The elasticity function is defined as follows as shown in Eq. (11).

$$Elasticity\left(\frac{y}{x}\right) = \frac{\partial yx}{\partial xy} \tag{11}$$

Where, y = inbound/outbound freight and x = GDP/ CRLP. The GDP elasticity of inbound freight is 0.854 and CRLP elasticity of outbound freight is 0.842. Similarly, the GDP elasticity of outbound freight is 0.248. The elasticity values are calculated by considering the average value of inbound/outbound freight for the last five years. The projected elasticities are 0.941 and 0.684 for inbound and outbound freight. The elasticity values clearly show that both the inbound and outbound freight are inelastic with respect to India's GDP growth and crude oil production. However, the GDP elasticity of inbound freight is closer to 1, which indicates the inbound freight will be growing almost with the same rate as India GDP. The GDP elasticity of outbound freight indicates that GDP has very low impact on outbound freight. This is already explained in sub-section 4.1.1.

7. Conclusions

The Mumbai Port is one of the important intermodal hubs for India's economy as it handles about 11% of the total seaport freight flow. In this study, we developed both univariate and multivariate regression and time series models to forecast inbound and outbound freight flow at Mumbai Port. The methodology for modelling the freight demand using the nation's economic indicators is successfully tested on Mumbai port. The variables that are found to be significant in the regression models are Gross Domestic Product (GDP) and crude oil production (CRLP). The multivariate additive regression models are found to have higher R^2 values and lower mean square error than univariate models. The validation of the models for 16% unused data also showed that the multivariate models perform better than univariate models. The R^2 values varied from 0.773 to 0.974 for all the models. This suggests Mumbai port freight is closely associated with national economic indicators like GDP and crude oil production. Instead of using a pure autoregressive model, we proposed time series regression models to predict Mumbai port freight flow. The time series regression results are more promising than the static (regression) models. The error from the time series models are significantly lower (i.e., 1%-5%) than that from regression models (i.e., 6%-12%). This clearly states that the forecasting capability improved by 58% to 83% at 95% confidence level from the static models to the time series models.

Both the time series models and multivariate linear regression models are used for forecasting freight flow for years 2014-15 through 2017-18. The models forecast that inbound and outbound flows will continue to increase; however, the outbound flow will experience a lower growth rate than the inbound flow. Moreover, the regression forecast is more optimistic than the time series models. The elasticity analysis of multivariate regression models suggest that Mumbai's inbound freight will be growing almost with India's GDP growth rate. However, the outbound freight growth will be slower than that of the inbound freight.

We believe this study will help policy makers to understand the interrelation of economic indicators and the magnitude of port activities. The forecasted freight flow will also help in the planning of the infrastructure requirements to handle the freight flow in the near future. It should be, however, noted that the models developed in the study have some limitations. A major policy change, although unlikely in the near future, can make the models less useful. For example, a major renovation of Mumbai Port or any other competing nearby ports (Jawaharlal Nehru Port, Mundra Port, etc.) can change the freight volume at Mumbai Port.

References

Affi, A., May, S., and Clark, A. V. (2012). *Practical multivariate analysis*, 5th ed., CRC Press, Taylor and Francis Group.

Akinbinu, V. A. (2010). "Prediction of fracture gradient from formation pressures and depth using correlation and stepwise multiple regression techniques." *Journal of Petroleum Science and Engineering*, Vol. 72, Issue 1-2, pp. 10-17, DOI: 10.1016/j.petrol.2010.02.003.

Al-Deek, M. H. (2001). "Which method is better for developing freight planning models at seaports – neural networks or multiple regressions?" *Transportation Research Record*, No. 1763, pp. 90-97, DOI: 10.3141/1763-14.

Al-Deek, M. H., Johnson, G., Mohamed, A., and EI-Maghraby, A. (2000). "Truck trip generation models for seaports with container and trailer operation." *Transportation Research Record*, No. 1719, pp. 1-9, DOI: 10.3141/1719-01.

Balach, P. and Tadi, R. R. (1994). "Truck trip generation characteristics of nonresidential land uses." *ITE Journal*, Vol. 64, No. 7, pp. 43-47.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). Time series analysis: forecasting and control, 4th ed., John Wiley & Sons, Inc.

Cambridge Systematics, Inc. (1997). *A guidebook for forecasting freight transportation demand*, NCHRP report no. 388, Transportation Research Board, National Research Council, Washington D.C., USA.

Chen, S. and Chen, J. (2010). "Forecast container throughputs at ports using genetic programming." *Expert Systems with Applications*, Vol. 37, Issue 3, pp. 2054-2058, DOI: 10.1016/j.eswa.2009.06.054.

Chow, J. Y. J., Yang, C. H., and Regan, A. C. (2010). "State-of-the art of freight forecast modeling: Lesions learned and the road ahead." Transportation, Vol. 37, No. 6, pp. 1011-1030, DOI: 10.1007/s11116-010-9281-1.

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

Department of Economic Affairs, Ministry of Finance (2010). *Position paper on the port sector*, DoEA, MoF, Government of India.

Faghri, A. and Hua, J. (1992). "Evaluation of artificial neural network application in transportation engineering." *Transportation Research Record*, No.1358, pp. 71-80.

Federal Highway Administration (1999). *Guidebook on statewide travel forecasting*, U.S. Department of Transportation.

Gosasang, V., Chandraprakaikul, W., and Kiattsin, S., (2011). "A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok port." *The Asian Journal of Shipping and Logistics*, Vol. 27, Issue 3, pp. 463-482, DOI: 10.1016/S2092-5212(11)80022-2.

Hamid, G., Tabatabaei, S. H., Asadi, H. H., and Carranza, E. J. M. (2015). "Multivariate regression analysis of lithogeochemical data to model subsurface mineralization: a case study from the Sari Gunay epithermal gold deposit, NW Iran." *Journal of Geochemical Exploration*, Vol. 148, pp. 249-258, DOI: 10.1016/j.gexplo.2014.10.009.

Hui, E. C. M. H., Seabrooke, W., and C, W. G. K. (2004). "Forecasting cargo throughput for the port of Hong Kong: error correction model approach." *Journal of Urban Planning and Development*, Vol. 130, Issue 4, pp. 195-203, DOI: 10.1061/(ASCE)0733-9488(2004)130:4(195).

IMF (2014). *World economic outlook: World economic and financial surveys*, International Monetary Fund (IMF), Washington D.C., USA.

IPNG (2013). *Indian petroleum and natural gas statistics 2013-2014*, Ministry of Petroleum and Natural Gas, Economics and Statistics Division, Government of India, New Delhi.

Klodzinski, J. and Al-Deek, M. H. (2003). "Transferability of an intermodal freight transportation forecasting model to major Florida seaports." *Transportation Research Record*, No. 1820, pp. 3645, DOI: 10.3141/1820-05.

Lam, W. H. K., Ng, P. L. P., Seabrooke, W., and Hui, E. C. M. (2004). "Forecast and reliability analysis of port cargo throughput in Hong Kong." *Journal of Planning and Development*, Vol. 130, Issue 3, pp. 133-134, DOI:10.1061/(ASCE) 0733-9488(2004)130:3(133).

Langen, P. W. de, Meijeren, J. van, and Tavasszy, L. A. (2012). "Combining models and commodity chain research for making long term projections of port throughput: an application to the Hamburg-Le Havre Range." *European Journal Transport Infrastructure Research (EJTIR)*, Vol. 12, Issue 3, pp. 310-331.

Liu, L. and Park, G. (2011). "Empirical analysis of influence factors to container throughput in Korea and China ports." *The Asian Journal of Shipping and Logistics*, Vol. 27, Issue 2, pp. 279-304, DOI: 10.1016/S2092-5212(11)80013-1.

Makridakis, S., Wheelwright, C. S., and Hyndman, J. R. (2005). *Forecasting methods and applications*, 3rd ed., John Wiley & Sons, Inc.

Middendrof, D. P., Jelavich, M., and Ellis, R. H. (1982). "Development and application of statewide multimodal freight forecasting procedures for Florida." *Transportation Research Record*, No. 889, pp. 7-14.

Mumbai Port Trust (2014). *Annual report 2013-14*, MPT, Mumbai, India.

Planning Commission (2011). *Report of working group for port sector for the 12th five year plan: 2012-2017*, [online] PC, Government of India. www.planningcommission.gov.in/aboutus/committee/wrkgrp12/transport/report/wg_port.pdf (Accessed on 28-08-2013).

Rencher, A. C. and Christensen, W. F. (2012). *Methods of multivariate analysis*, 3rd ed., John Wiley & Sons, Inc., Hoboken, NJ, USA.

Sculze, P. M. and Prinz A. (2009). "Forecasting container transshipment in Germany." *Applied Economics*, Vol. 41, No. 22, pp. 2809-2815, DOI: 10.1080/00036840802260932.

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*, Vol. 20, No. 1, pp. 51-64, DOI: 10.1016/S0264-2751(02)00097-5.

Siddiqui, F. I. and Syed Osman, S. B. A. B. (2013). "Simple and multiple regression models for relationship between electrical resistivity and various soil properties for soil characterization." Environmental Earth Sciences, Vol. 70, No.1, pp. 259-267, DOI: 10.1007/s12665-012-2122-0.

The Department of State Development, Business and Innovation (2009). Report on Indian infrastructure market opportunities in the seaport sector, [online], DSDBI, State Government of Victoria.http://export.business.vic.gov.au/data/assets/pdf_file/0010/337852/Indian-InfrastructureMarket-Opportunities-in-the-Sea-Ports-Sector.pdfm (Accessed on 28-09-2013).

Timm, N. H. (2002). *Applied multivariate analysis*, Springer, New York.

Tolosana-Delgado, R. and Eynatten, H. V. (2010). "Simplifying compositional multiple regression: Application to grain size controls on sediment geochemistry." *Computers and Geosciences*, Vol. 36, No. 5, pp. 577-589, DOI: 10.1016/j.cageo.2009.02.012.

Woo, S.-H., Pettit, S. J., Kwak, D.-W., and Beresford, A. K. C. (2011). "Seaport research: A structured literature review on methodological issues since the 1980s." *Transportation Research Part A*, Vol. 45 No. 5, pp. 667-685, DOI: 10.1016/j.tra.2011.04.01.