M.Phil Dissertation

# An Empirical Analysis of Rail Freight Demand in Pakistan

*Dissertation submitted to department of Econometrics & Statistics, Pakistan Institute of Development Economics, Islamabad in partial fulfillment of the requirements for the degree of Master of Philosophy in Econometrics*


**Submitted By:**

***Muhammad Afzal Khan***

**16/M.Phil-ETS/PIDE/2011**

**Supervised By:**

***Dr. Saqlain Raza,*** *Assistant Professor*

Department of Econometrics and Statistics



**Pakistan Institute of Development Economics, Islamabad,**

**March, 2015**

## Authorship Statement

I Muhammad Afzal Khan solemnly declare and affirm on oath that I myself have authored this M.Phil Thesis with my own work and mean and I have not used any further means except those I have explicitly mentioned in this thesis. All items copied from internet or other written sources have been properly mentioned in quotation marks with a reference to the source of citation.

**Muhammad Afzal Khan** s/o **Abdul Ghani Khan**

## Acknowledgment

Nothing could be possible without the help of Allah and affection of Prophet Muhammad (Peace Be Upon Him). I would like to express my deep gratitude to Dr. Saqlain Raza, my research supervisor for his patient guidance, enthusiastic encouragement and useful critiques to complete this research work. He was always there whenever I need his assistance and guidance, especially, during thesis writing. I also like to thank Dr. Hafsa Hina, Muhammad Ramzan for their advice and assistance in keeping my progress. My grateful thanks are also extended to Pakistan Railways for their assistance in data collection. A bundle of thanks to the students and teachers of Department of Econometrics and statistics of PIDE. My deepest appreciation and sincerest thanks to my cousins and Muhammad Rashid Khan, Nadeem Khan, Hamid Maqsood, Muhammad Babar, Muhammad Usman Virk, Muhammad Jawad Virk, Khalid Mehmood Khan, Muhammad Munir Khan, Ghulam Ahmed Samdani, Chaudhary Amjad Warraich, Mahboob ul Hassan for encouraging me at every stage of my study. Cordial and sincere obligations are rendered to my mother, whose hands always raised with unlimited affection by virtue of which I could reach at this position. My humble thanks to my dear brothers Altaf Khan, Sarfraz Khan and Khalid Usman Khan for their encouragement and unconditional support for the completion of my education. Would that my late father could see me at this stage of life.

**Muhammad Afzal Khan**

***<u>In The Name of Allah, The Most Gracious, The Most Merciful</u>***

*And if all the trees on the earth were pens and the sea (were ink wherewith to write), with seven seas behind it to add to its (supply), yet the words of Allah would not be axhausted. Verily, Allah is Al-Mighty, All-Wise.*

(Parah: 21, Surah Luqman: 31, Verse: 27)

*It is not for the sun to overtake the moon, nor does the night outstrip the day. They all float, each in an orbit.*

(Parah: 22, Surah: Yasin: 36, Verse: 40)

*(He is) the Lord of the two easts (places of sunrise during early summer and early winter) and Lord of the two Wests (places of sunset during early summer and early winter).*

*Then which of the Blessings of your Lord will you both (jinn and men) deny?*

(Parah: 27, Surah: Al- Rahman, Verse: 17-18)

*This Thesis is Dedicated to:*

***My World, My Mother***

# Contents

# List of Tables

# List of Figures

**Abstract**

Railways are usually considered as the life line of the country. Pakistan railways facing deficit from early 70's. Most importantly, rail freight earnings have declined gradually that in 2012-13 it downshifted to only 11% share whereas it remained more than 60% up to 70's. This study aims to examine the factors affecting rail freight demand in the short and long run by incorporating macroeconomic variables. For this purpose annual time series data for the period of 1981-2012 have been used and obtained from different soures. For analysis of short and long run impact of variables on rail freight demand growth, Autoregressive Distributed Lag approach (ARDL) to cointigration have been employed. The results showed that GDP and share of manufacturing sector to GDP has positive impact on rail freight demand in the short and long run. Furthermore, the study finds negative influence of freight rate and domestic oil prices in both the dynamics. The implications of the study that government may adopt policies for lowering oil prices, and promoting manfacturing sector and taking measures to reduce freight rates.

**Keywords:** Rail freight demand, freight rate, GDP, ARDL, Pakistan.

# Chapter 1

# Introduction:

Transport has reshaped human development from the beginning of human civilization whereas freight railway is very essential for economic growth. Efficient freight transport is an important component of economic growth. Existence of efficient freight enhances global development patterns and can be a boost or a barrier to economic growth in individual nations (World Bank, 2011). Freight movement is an essential and integral part of national development process because it boosts the full exploitation of factors of production and it is the basis for the network of distribution. The freight and passenger transport demand increases with the increase in economic development and population, rapid industrialization, urbanization, and agricultural development (Ramanthan, 2001). Many public decisions relating to taxes, subsidies and price regulations as well as carrier's decisions on optimal pricing require precise knowledge on demand for freight transport (Oum, 1979). Lack of mobility of critical industrial inputs can be a major obstacle to growth.

Decline in the rail freight started in 90's and at present touched the bottom line of its ever demand. The causes of failure to increase its demand contains decades of shocks as well as short term crisis. Therfore, it is important to find out short run and long run responsiveness of freight transport demand and its major determinants (Kulshereshtha et al., 2001). Our concern in this study is to determine different macroeconomic factors affecting rail freight demand growth because Pakistan railways have been earning a major share through freight traffic upto 1989.

In the earlier times cities grew up on natural bays, ports, rivers and lakes where transport facility was available. Romans built roads to unify and provide access to their far-flung empire but the industrial revolution generated new transport demands which required higher volume of coal, iron and other materials that lead to canal construction, extended water transportion and to early railway development (World Bank, 2011). Railways basically provides services in the form of passenger and freight in the world. These services have advantages over other modes of transport in the form of less traveling time, reduced air pollution, $CO_2$ emissions and less accidents. Rail freight services which is our research focussed is important to economic growth and efficient in many countries, regions and can move massive volumes of cargo over long distance effectively at feasible prices. There is general agreement that the rail freight industry should be efficient and market-responsive to serve the transport needs of shippers; financially sustainable, generating revenue for reinvestment and not imposing an unsustainable burden on the public purse; be safe and meet prevailing environmental standards. Many countries have gone further, aspiring to an increase in the proportion of freight that is carried by railways compared to road transport, for a number of reasons: better safety performance; less road congestion; less damage to roads; better energy efficiency and lower greenhouse gas emissions per tonne hauled (International Transport Forum, 2014). Moreover trains cause high fixed cost but rail freight traffic is only profitable if large trains moved over large distance. The high demand for freight merely exist between large cities (Kellner M. et al., 2012).

## 1.1 Pakistan Railways:

About two-third of all rail freight of the world carried in developing countries and over three quarters it is carried in China, India and Russia (Freight Transport for Development toolkit-Railway Transport, 2009). Pakistan railways has long history of services as first railway line from Karachi city to Kotri was opened for public traffic on May 13, 1861. Further developments were made gradually to build new stations from time to time but maximum work had done in the last quarter of 19th century. Pakistan railways consist on 7791 route kilometers and 11755 track kilometers ( Ministry of Pakistan Railways, 2012-13). Pakistan railways has been playing an important role for trading in neighboring countries for exportation of fertilizers, wheat, rice to Iran and Afghanistan (Ministry of railways, 1982-83).

The proposed budget for Pakistan railways is going to be increased every year but the output is getting decreased. It is evident from the budget numerals. It reveals that the actual total budget financed by the government for developmental and non-developmental purposes was 6652.503 million rupees in 2000-01 and 85113.573 million rupees in 2013-14 and the estimated budget for the year 2014-15 is 104566 million rupees (Finance Division for ministry of railways, Government of Pakistan).[1]These numerals do not include foreign loans taken by Pakistan railways. Number of employees in 1972-73 were 132938, in 2012-13 were 81880 but the cost of employees for the same fiscal years was 387.497 and 20557.285 million rupees respectively. The number of locomotives (engines) in 1972-73 were 993 and 493 in 2012-13. The number of passenger coaches in 1972-73 were 2057 and 1540 in 2012-13 and freight wagons were 37436 and 16635 for the respective years. The total expenditures to gross earnings of Pakistan railways in 1955-56 were 66%, in 1972-73 were 68.3%, in 2011-12 and 2012-13 increased to 203. 59% and 194.36% respectively ( Ministry of Railways, Government of Pakistan).[2]

There were two hundred freight stations and about 12000 personnel for providing services

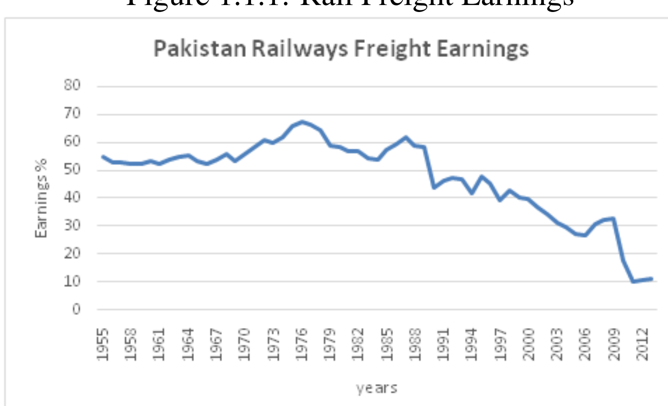
[1]The budget statistics have been collected from each years published book of Finance Division, Ministry of Finance, Government of Pakistan. The detail of financed budget for each project can be seen either it is for developmental purpose or non-developmental. I mentioned total final budget not the estimated except for the year 2014-15. Budget statistics were given in the form of rupees. For easy understanding I converted it into million rupees.

[2]All the figures mentioned are taken from the published year books of Ministry of railways, Government of Pakistan. The data for the fiscal years 1955-56, 1972-73, 2011-12, 2012-13 have been taken from year books of 1966-67, 1982-83 and 2012-13 respectively.

to the clients up to 2012-13. The freight services mainly concerned with two major sea ports Keamari and Bin Qasim, and dry ports as well. The major commodities handled were Chemical for Rayan Manufacturer, Petroleum Oil and Lubricants, wheat, coal, fertilizer, rock, phosphate, cement, sugar, oil-seed, wool, salt, rice, containers and goods for transit to Afghanistan (Ministry of railways year book, 1966-67 & 2010-11). Freight traffic earnings to gross earnings remained more than 50% from 1955-56 to 1989-90 and passenger earnings remained below 40% for the same period but in 1990-91 freight earnings declined to 43.8% from 58% of the previous year and passenger earnings rose up to 49.6% from 34.7%. In the period of 1955-56 to 1989-90, the highest earnings from freight services were 67.2% in 1976-77 and lowest were 52% in 1961-62 but highest earnings from passenger services were 38.8% 1983-84 and lowest were 27.1% in 1976-77. After 1989-90 freight traffic earnings decreased continuously as compared to increased passenger traffic earnings. So in 2012-13 freight earnings remained 11% and the rival rosed up to 75%. There can be seen a huge difference between both earnings ( Ministry of railways, 1991-92, 2004-05 & 2012-13). Pakistan Railways freight earnings can be seen in the following diagram:

### 1.1.1 Pakistan Raiways Freight Earnings:

Figure 1.1.1: Rail Freight Earnings



The above diagram showing that Pakistan railways freight earnings become lower with the passage of time. The reasons described by Pakistan Railways of lower earnings are shortage of locomotives and poor hauling capacity, irregular supply of fuel for locomotives, preference given to load high rated commodities and reduced rates for individual parties. Freight demand also affected by manufacturing sector, exports, imports, GDP, Domestic Diesel Oil prices, National Logistic Cell (NLC) and trucking industry. Manufacturing sector plays a vital role in increasing rail freight demand and this sector has been affected by declining share of agriculture sector because this sector provides raw material to manufacturing sector. More than 75% exports are comprised on manufactured goods but unfortunately real growth in manufactured exports bears a declining trend and continues to show a persistent decline. Exports and imports are mainely concerned with Port Bin Qasim, Keamari Port and other dry ports of the country. These exports and imports are mainely transacted by railways but over the time other modes of transport like NLC and trucking industry has also been used, as a result of this share of rail freight declined over time.

As our concern is to emprically analyse the rail freight demand in Pakistan. In this context econometric Studies differ widely about rail freight demand on the basis of functional form (linear or log-linear), the sectors studied (a particular sector or the whole economy), the sample used (cross-section or time series) and the variables used in the estimation (Agnoluccci and Bonilla, 2009). But to best of my knowledge we didn't found a relationship empirically investigated in either functional form. On the basis of these empirical evidences we are interested to analyze different factors which turned rail freight traffic earnings from peak to bottom level in Pakistan by using Autoregressive Distributed Lag Model.

## 1.2 Objective of the study:

As we mentioned the facts about railways that freight earnings contributed a considerable share to gross earnings of Pakistan railways were almost more than half up to 80's but hereafter, unfortunately it started declining and finally in the financial year 2012-13 rail freight earnings remained only up to 11% which was a huge decline. So on the basis of evidences, objective of our research is to find out the key determinants of rail freight demand decline in Pakistan. We shall observe the effect of GDP and share of manufacturing sector to GDP seperately. In this way, we shall be able to identify that whether the decline in the freight demand is associated with freight rate only or also other macroeconomic factors as well contributing in perturbing public interest. We are also interested in examining either there exist short and long run relationship between rail freight demand and its determinants affecting it.

## 1.3 Hypothesis:

It is the main concern of our study to identify potential factors affecting the rail freight demand positively or negatively. So to know that whether there exists a short run or long run relationships among the rail freight demand and its determinants we shall analyze two hypotheses.

$H_1$: There exist short run and long run impact of factors affecting rail freight demand growth including GDP;

$H_2$: Rail freight demand has positive relationship with share of manufacturing sector to GDP in the short and long run.

## 1.4 Significance of the Study:

Freight earnings have been playing a significant role in increasing the gross earnings of railways in the past. But, the slope of earnings has been decreasing since early 90's. Due to this, we are going to analyze different factors that possibly affect rail freight demand growth in Pakistan. From the literature it is evident that a number of empirical studies have been done for rail freight demand growth and its factors and on mode-choice behavior for passenger and freight but we didn't found a study done for rail freight demand growth in the case of Pakistan by incorporating the effect of macroeconomic activities. National Transport Research Centre (NTRC) studied descriptively the situation of Pakistan railways and other modes of trasport. To the best of my knowledge a purported relationship between rail freight demand and its factors has not been empirically elsewhere in the case of Pakistan.Our study is the primary empirical effort to analyze the long run behavior among the determinants of rail freight demand growth to other variables using Autoregressive Distributed Lag Model (ARDL). This approch has certain advantages over other techniques for analysing short and long run relationship between the vaiables like it can deal the problem of endogeniety, suitable for series either these are $I_{(0)}$ or $I_{(1)}$ or mixture of both, don't require pre-testing of unit root and appropriate for small sample data.

# Chapter 2

# Literature Review:

As discussed above that rail freight traffic contributed about more than half to gross earnings of Pakistan railways for a long period but now its share is at lowest level. Existing studies on rail freight demand can be divided into two groups these are time series studies and survey based studies but here our main focused literature will be on time series studies on non-bulk rail freight demand growth and its determinants. We didn't found estimates in literature for rail freight demand in Pakistan so there is need for an investigation to assess the long-run relationship between rail freight demand growth and its determinants.

Rao (1978) empirically investigated the demand for rail freight transport in Canada. The obective of this study was to construct a national rail forecasting system, which could help in measuring,a simultaneous framework, the impact of macroeconomic activities and intermodal competition on rail freight demand. For this purpose he had used data over the period of 1958-73. The proposed model was estimated by OLS and 2SLS because there were four jointly dependent variables in the specified models. He concluded in many ways like results suggest that commodity outputs are the main determinants of rail freight demand and output elasticity of rail freight demand is not significantly different from unity. In short run trucking industry also affects the rail freight demand especially for manufactured products. Rail freight demand in both volume and distance is not price elastic. Railway unit costs decreases with an increase in the average utilization of existing freight cars and these economies too differ among commodities. Finally for most of the commodities trucking and average carload resulted major determinants of the distance component of demand and price.

Oum (1979) conducted an analysis for the derived demand for freight transport and inter-modal competition in Canada. The purpose of the study was to formulate a demand model for inter-city freight transport as an intermediate input to production and distribution sectors of the economy and to estimate price elasticities of substitution between all modes like railway, waterway and highway. According to him previous studies attempted to estimate the demand functions on the basis of various ad hoc models. He also criticized Cobb- Douglas Production funtion that it is inappropriate for multi-model demand studies because it places a severe restrictions on the parameters of inter-model competition. Therfore he had analyzed price reponssiveness and various modes inter-modal competition on the basis of a derived demand model which allowed for free variations of parameters of inter-modal competition. For this purpose data used over the period of 1945-1974 and applied Maximum Likelihood method (Iterative Zellner Efficient Method) and iterative nonlinear least estimator. He concluded that the variations in three modes of transport explained very well and there is a significant competition between these modes of transport.

Levine (1981) estimated the impacts of railroad rates on prices, profitability and welfare in the rail industry whether the present railroad market power is sufficient to earn profits, prices and static dead-weight loss after the regulatory reforms by Inter-state Commerce Commission (ICC) in the Transportation Act of 1958. The data employed in the study was only confined to the year 1972 in which a sample of 349 markets defined by commodity group, mileage block and weight block were included. For estimation multinomial logit model for modal choice was used to describe the structure of demand for transport of manufactured commodities. In this model the market shares were explained in the truck, piggyback and rail by inter-modal differences in rates, speed, reliability and unobserved attributes of modes and shippers. He concluded that if the price of truck or piggyback service increases then the demand for rail transport also increase but if there is improvement in the services of piggyback and trucks occurs then rail transport resulted downward but trucks are not an alternative source for long hauls of minerals due to highway weight limitations and when rail-rates increases then there is possibility of decline in commodities loaded but could not moved to other modes. According to him if variations in inter-modal differences in attributes and rates in sample markets then demand elasticities would also vary across markets. Finally he concluded that the intensity of inter-railroad competition was an important determinant of the level of prices, profits and dead-weight loss under deregulation.

Winston (1981) did a disaggregated analysis to investigate mode choice behavior at individual level. He considered regulated and unregulated motor and rail freight models. The main objective of his study was to develop a theory of shipper or receiver behavior at disaggregated level which could be helpful to show the critical determinants of mode choice in freight transportation. For this purpose two types of data sets were used. One contained agricultural commodities and other included wide range of commodities covered the periods 1975-76 and 1976-77 respectively. On the basis of theoretical considerations for stochastic specification he had used probit model to control for dependence through random parameters. But due to the known aggregate choice probabilities he applied Weighted Exogenous Sample Maximum Likelihood (WESML) estimator which is consistent and asymptotically normal,it is not maximum likelihood estimator. Also Wald test had been used instead of the likelihood ratio test and Wald test is dependent on the consistency and asymptotic normality of the estimator. The results suggested that the models used for investigation provided an opportunity for maximum econometric specifications and disaggregated models showed more precise estimates than aggregated. In the end he concluded that price competition would lead to greater mode shifts than service competition and rail freight transport could capture a substantial amount of traffic by improving the transit time means that each mode had an opportunity to attract maximum traffic in particular markets either by service or price competition.

Lewis et al. (1982) attempted to estimate a translog transport demand model to investigate price and quality of service demand elasticities for rail and motor carrier shipments of assembled automobiles in United States and used time series data for the period of 1955-77. He criticized cross- section models of qualitative choice like logit and probit that these models have been used to generate own and cross price elasticities. They criticized Linear logit models that these are not appropriate for estimating the price responsivness of demand and simulating traffic allocation across modes. He applied Partial Adjustment Model to estimate the shares of rail and each equation system was estimated by Full Information Maximum Likelihood (FIML) method for both rail and motor car. The results suggested that due to lack of effective rail deregulation, truck deregulation would reduce the financial stability of railroads in the market of assembled automobiles. So effective rail deregulation like comparable low charges and increased speed then it could offset those effects of trucking deregulation. He concluded that the effective regulations in rail can significantly affect the modal shares.

Wilson (1984) investigated elasticities in demand as a result of mode choice of single rail car shortages and multiple car rates for grain transportation like wheat and barley in Dakota using monthly time series data from July 1973 to December 1982. He described the shortcomings of optimization models, models of modal choice, ad hoc and derived demand models. For estimation he applied duality approach for intermodal competition like Iterative Three Stages Least Squares (IT3SLS). This approach is attractive because its functional specification is consistent with neoclassical economic relationships. Further, hypotheses about changes in behavior of demand parameters can easily be incorporated into the model and tested. to account for the effect of simultanity, instrumental variable method had been used. In the first model, basic translog model which excluded the effect of both rail car shortages and multiple-car rates. Secondly the effects of rail car shortages analyzed and only the effects of multiple-car rates were included in the third model but the effects of both included in the fourth model. He concluded that the changes in output were rail intensive for wheat shipments to both Duluth and Minneapolis and truck intensive for barley only to Duluth. He also estimated the elasticities of substitution between rail and truck and found that elasticities were not significantly different from one. Other than it, own rate and cross rate Hicksian elasticities were estimated which showed that own rate elasticities for rail roads changed across movements and inelastic in all cases but only elastic to motor carrier for wheat and barley to Duluth and Minneapolis.

Inaba and Wallace (1989) addressed two issues of freight transportation demand analysis of simultaneity between quantity shipped and mode choices and other issues related to the effects of spatial price competition on demand for freight transportation. The data collected through questionnaire survey from all grain elevators for the period of last week of November to first two weeks of december, 1984. For survey seven mode of choices were questioned like truck, barge, single car rail, multiple car rail, unit train, truck or barge, truck or multiple rail car and seven destinations. Out of 329 firms only 183 responded properly and few of them reported that they used all modes of transportation. Logit and Switching regression techniques were used for estimation.Weighted Least Squares (WLS) method used to estimate the quantity shipped as a function of market area conditional on mode choice. He concluded that waiting cost, market boundary and unit train and barge dummy variables are statistically significant but loading cost and travel time cost not statistically significant at 0.05 level and all variables provided expected signs. The demand functions reported relatively inelastic rate. Finally the empirical model provided consistent estimates of unconditional freight demand.

Abdelwahab and Sargious (1992) analyzed demand for freight transport in a way that it combined the two decisions on demand for mode choice behavior about rail and truck and shipment size. They used suitable data for the period of years ended in the numbers 2 and 7 after every five years like 1967, 1972,1977 which was collected by Commodity Transportation Survey at disaggregated level other than it two more sources were also used. Mode choice attributes for rail and truck included were freight charges, mean and reliability of transmit time and for shipment atributes like commodity value, density, shipment size. For mode choice behavior, a binary probit model was formulated and two linear regression equations used to simulate the choices of shipment size by rail and truck and ultimately Maximum Likelihood and Two Stages Least Square Method was used. The results suggested that coefficient estimates of mode choice and shipment size equations supported the hypothesis that there were interdependence between the mode choice and shipment size decisions and biased results seen in case of single equation estimation of mode choice or shipment size.

Friedlaender (1992) assessed whether the equitable rates to captive coal shippers are compatible or not with the competitive rates of return in the rail industry due to the Staggers Act of 1980. There were a controversy over the nature of the regulation of rates on coal and other captive traffic which have no immediate access of other modes of transport. For estimation he used panel data from 1974-1986 on five class-I railroads of heavy coal carriers and for comparison he included non-coal, non-merged systems. He estimated the cost function and its associated input share equations by using Three Stages Least Squares Method. He concluded that there were a little difference in the returns to scale between coal and non-coal roads and the behavior of rates in the period of deregulation had not remained same. other than it adjustments that were made after the period of Staggers Act, showed that rate structure could not be the single source for railroads to achieve enough revenue but with the imposition on fuel tax, railroads could regain their traffic losses caused by trucks over the past decades.

Wardman et al. (1997) conducted an analysis to identify the degree of interaction between rail and car in inter-urban leisure travel market of Great Britain. He used data of summer 1990 which was collected in Trans-Pennine Rail Strategy Study of 1992 included actual choices for journeys exceeded about twenty-five miles and car users traveled across the Pennines were interviewed at roadside and train users were interviewed during their journey and it has 1080 useable observations out of which 848 observations related to the choice of alone travel and 232 related to group travel mode choice. He used functional form of Box-Cox transformation and multinomial logit model . The results showed that cross elasticities of demand for car with respect to rail service qualities were very small for inter urban leisure travel and reductions in rail journey time. He concluded that fare can be useful to reduce the demand for car but these measures were not sufficient and attractive option to increase the rail demand.

Chapin and Schmidt (1999) studied that do the mergers improve the technical efficiency and scale efficiency of United States Class-I railroad firms since deregulation by Staggers Act 1980. They used data from 1980-1993 and employed Data Envelopment Analysis (DEA) to measure the efficiency of each firm. The results suggested that with the arrival of mergers, tracks were operated more efficiently than previous but mergers reduce scale economies so overall they reduced total efficiency. Also the mergers increased the market power due to the fewer oligopoly. Before mergers larger firms had price above than cost which resulted in increased rail profits but causing deadweight loss. Finally they concluded that future merger firms should be carefully scrutinized that the claimed efficiency improvements from the merger exist in reality or not.

Ramanthan (2001) investigated the long run relationships between the variables of transport performance and other economic variables in India by using the cointigration and error correction models. He used time series data for the period of 1956-88. He concluded that PKM are likely to increase faster than GDP and much faster than urbanization and TKM was highly correlated with industrial growth and are likely to increase faster than index of industrial production but the performances of both passenger and freight were relatively inelastic to price changes. The error correction results showed that PKM and TKM adjusted to their respective long run equilibrium at a normal rate.

Kulshereshtha et al. (2001) analyzed the long-run structural relationships between Indian railways freight transport demand and economic variables. He used time series data 1960-

1995 and employed multivariate cointigrating Vector Auto Regressive (VAR) framework. Also Johansen's Likelihood approach used to have cointigration rank and to check the stability of co-integrating relationships he made two sub-samples (1962-85 and 1970-95) and then analysis done on these two samples as well as on the full sample. He found that at most one cointigrating relationship between GDP, TKM (Freight rate or real rate on per tonne-kilometer in paisa) for all samples and GDP resulted as the the major determinant of freight transport demand and vice-versa. He also concluded that low price elasticity below freight demand could increase the revenue of railways. GDP and TKM had a strong contemporaneous correlation and both adjusting to correct the disequilibrium within a period of three years. Finally, freight transport demand directly related to economic growth and industrialization and there is a need for large investments in wagons, locomotives, new tracks to increase the transport capacity.

Kremers et al. (2002) attempted to study the price elasticities of transport demand using meta-analysis. Data obtained from 24 previous studies about price elasticities. They empirically estimated the microeconomic, micro econometric and discrete choice models for this reason they introduced the idea of general equilibrium. They concluded that the microeconomic model as compared to micro econometric model showed a significant influence of estimating price elasticities and microeconomic models not only covered the economic factors which were important determinants of transport demand but also incorporated economic interactions. Also discrete choice models only take into account the mode choice and ignore the transport volume and micro econometric models reduced form equation relating to transport demand but skipped their economic interactions. Air transport resulted to be more sensitive as compared to all other modes of transport like rail, roads, etc. So for the estimation of price elasticity microeconomic model resulted most appropriate model. Meta regression results showed that it differs significantly for the estimation of price elasticity of transport demand whether the data collected from urban or national scale.

Bresson (2003) studied the main determinants of the demand for public transport in Great

Britain and France. English included 46 counties of urban and rural but French study included all public modes of transport used in urban travel and it includes bus, metro and rail transport data. Data used for England was 1987-1996 and for French 1986-1995. Both short run and long run elasticities were estimated and two functional forms like log-log and semi-log were used. They had two main objectives in this study firstly, to observe heterogeneity in each area and secondly, to compare estimated income, price and service elasticities in both panel countries using same set of variables. For estimation they applied fixed effect model, random effect model, Beysian approach used and for efficient estimation an instrumental variable method of Arrelano and Bond (1991) Generalized Method of Moments used. They concluded that all short-run and long-run elasticities were highly significant but income elasticity for France reported to zero and fare elasticity were similar for both countries. Instruments used found to be valid and public transport was relatively elastic to fare changes so subsidies in fare could play an important role in encouraging public transport, also service as important as fare because if service were of quality then people could adopt public transport.

Beuthe and Bouffioux (2008) estimated the relative importance and qualitative factors of the value for transport shippers. This paper contributed to a better understanding of the qualitative factors which could determine the transport mode choice like service quality including reliability of delivery, service frequency, risk of losses, transport speed or time and carriers flexibility in response to unexpected demands. The data collected through questionnaire from freight transport managers of Belgium. Out of 600 firms only 113 provided information which found useful for analysis and face to face interview also conducted. For estimation purpose, logit model was used on the basis of results he concluded that qualitative factors played an important role in mode choice transportation and their relative importance changed according to sub samples.

Shen et al. (2009) conducted an analysis for the role of road plus rail freight demand growth for the time series data of 1974-2006 at aggregate and disaggregated level in Great

Britain. For this purpose six econometric techniques were used like traditional Ordinary Least Square (OLS) method, Partial Adjustment (PA) model, Reduced Autoregressive Distributed Lag Model (Re-ADLM), VAR model, the Time Varying Parameter -TVP) model, the Structural Time Series Model (STSM), also the relative accuracy of alternative econometric model and forecasting performance of each model was observed and it was observed on the base of Mean Absolute Percentage Error (MAPE). The smaller the value of the MAPE indicated the most accurate forecasts. Different diagnostic tests were applied like Lagrange Multiplier (LM) test for serial correlation, the Jarque-Bera test for non-normality, the RESET test mis-specification and white test for heterscedasticity. He concluded that industrial production was an important determinant of both freight demand in Great Britain but it varied across different commodity groups because each commodity have different transport requirements and the magnitude of income elasticity also changed due to the use of different econometric models. For forecasting performance of each model he concluded that no model is best one but results showed that STSM best performed for short and medium term forecasts and PA and Re-ADLM perform well for longer horizons and TVP generally good forecasting tool for one-year ahead forecasting.

Growitsch (2009) examine a multi-country analysis of twenty seven European countries to investigate the performance of railways with particular focus on economies of scope. For this purpose fifty four railway companies were observed over the period of 2000-2004 and these firms divided into four groups integrated firms, infrastructure managers, passenger operators and freight operators. He applied two step approaches. In the first step estimated technical efficiency of integrated and non-integrated railways by using Data Envelopment Analysis (DEA) rather than Production Frontier Analysis (PFA) and secondly to determine whether separate or joint production is more efficient by applying DEA. He concluded that for most of the European railways companies, economies of scope existed.

Agnolucci and Bonilla (2009) used time series data from 1956-2003 to analyze road freight demand elasticities and decoupling in United Kingdom. According to them 64 percent of freight were moved through road. The main objective of their analysis was to estimate the aggregate relationship between road freight demand, economic activity and freight prices. For estimation general to specific modeling was used. Auto Regressive Distributive Lag Model with linear, stochastic and unconstrained error correction models were estimated for the sample of 1956-98. Pesaran, Smith and Shin (PSS) test rejected the long run relationship between the variables. When the models were estimated for three commodity group foods and tobacco, chemicals, iron and steel then it showed variability in parameters. Income elasticity resulted less than 0.8 which showed weak decoupling effect. Finally they concluded that stochastic trends were responsible for the lack of long run relationship between and its determinants.

Mitchell (2010) examined inter-capital freight demand in Australia using Full Information Maximum Likelihood (FIML) method and linear logit model for estimation of elasticities in mode choices like road, rail and sea because these three modes are substitutes across different length corridors between Sydney-Melbourne and Sydney-Brisbane. For this purpose he used data form 1972-2001. The translog cost function resulted that all modes were complements for freight transport on short distance corridors but cross price elasticities were statistically significant on long distance and linear logit models showed that there were concavity across all corridors. Overall road freight resulted relatively price inelastic but rail and sea demand were more price elastic.

Odggers et al. (2011) did an analysis about the relationships between yearly train boardings and different economic and demographic variables for metropolitan area of Melbourne, Australia. They used time series data from 1983-2009 . For the estimation purpose univariate and multivariate regression techniques were applied and analyzed on the basis of estimated sample regression equations of multivariate analysis. Three year moving average also used for comparison. The results showed that the variables average annual housing interest and estimated resident population have high adjusted $R^2$. The multivariate regression forecasting results showed that the variables real average price of a zone one ticket, real price per liter of unleaded petrol and estimated resident population indicated the demand would continue to be grown about 273.5 million with a range of 258.8-288.7 million in 2012-13 which is about 7.7 percent annual growth rate and stronger patronage annual growth.

Albert et al. (2013) attempted to analyse the determinants of passenger rail demand in Melbourne, Australia. For this they employed Engle-Granger Two step procedure for measuring long run elasticities and error correction method for short run elasticities for the period of 1979-2008. The determinants discussed for passenger rail demand were fare, population of the city, per capita income, fuel rate, the number of kilometers run by a rain, fatality as proxy to measure the passenger perception about rail's overall quality and vehicle price index. He concluded that passenger rail demand was highly inelastic to fare, population was resulted in highly significant to passenger rail demand and per capita income, fuel rate had positive impact on passenger boardings and statistically significant but fatality variable not statistically significant nor have expected sign of vehicle price index variable supported the hypothesis that passenger rail and private vehicles were substitute in the the long run. All other short run elasticities such as income and cross price elasticity were smaller than the long run elasticities except the population elasticity.

Albert et. al (2013) investigated the determinants of passenger rail demand growth in Perth city of Australia. Time series data used from 1983-2008. For estimation he applied Two step Engle-Granger cointegration technique to find the long run elasticities and error correction model to examine short run elasticities. He concluded that passenger rail demand is inelastic to price. Other variables behaved differntly in a way that city population had a significant impact on passenger rail demand but short run elasticities are smaller than long run and increase in per capita income had no statistical support that increase in income would leads to more passengers traveled. Finaly perception of passengers about the number of accidental deaths due to rail plays an important role in determining the choice of travel mode but this fatality variable was used as a proxy also number of kilometers run by rail had same impact.

Wang et al. (2013) analyzed the inter-state mode choice behavior for freight between truck and rail in Maryland, UK. He used logit and probit models for the comparison of modal behavior and for the verification of the differences of mode choice behavior among three zones in Maryland and a variety of variables were used. Analysis was done in two parts, the results of two parts suggested that transportation mileage ratio, the value of time for commodity, trade type, origin, and fuel cost play important roles in model choice for all shipments in the rest of the parts.

Wijeweera et al. (2014) empirically examined the impact of unannounced shocks in freight rate, international and business cycle on non-bulk freight rail demand growth in Australia covered the period 1970-2011. In this analysis he addressed the shortcomings of previous literature like aggregation data bias and endogeneity bias and determines the interrelationships between non-bulk freight demand and its determinants. He applied VAR, Impulse Response Function and Variance Decomposition techniques and different diagnostic tests. The concluded that freight rate proved a dominant impact on rail freight demand, also the volatility of Australian dollar through a significant impact on demand that means depreciation in dollar resulted to encourage exports but discourage imports as a result of it there will be more domestic production consumption. So macroeconomic conditions are favorable to bulk freight demand.

# Chapter 3

# Data and Methodology:

The aim of our study is to analyze determinants of rail freight demand in Pakistan. There is lot of literature available globally on rail freight demand but we didn't found significant literature regarding Pakistan which empirically analyzed factors affecting rail freight demand in the short and long run by using modern econometric techniques and by incorporating the effect of macroeconomic actitivies. Important determinants we have included in our analysis are freight rate, GDP, domestic diesel oil prices, share of manufacturing sector to GDP. For the long run analysis of rail freight demand we have applied Auto regressive Distributed Lag approach to cointegration. In this section we shall discuss economic relationship among variables, availability of data, source of data and steps to investigate the long run relationship.

## 3.1 Variables and Data Sources:

For studying the long run relationship among the variables we have considered time series data for the period from 1981 to 2012 because data from onward time period is not available. There are number of relationships among variables that have been analyzed in the literature but, in our study we have chosen only two types in accordance with the relevance to our study objectives.

### 3.1.1 Tonne- Kilometers:

Tonne-kilometers (TKM) are used as standard measure to show the rail freight demand. A tonne-kilometer[1] is a unit of measure of freight transport that can be obtained by multiplying number of Tonnes of Freight Carried (TFC) and Total Distance Covered (TDC) (in kilometers) in a given year. The more the TKM showed more rail freight demand.

$$TKM = TFC * TDC$$

It only considers the distance on the national territory of country to account for national, international and transit freight transport. Data for TKM has been obtained from yearly published books of Ministry of Railways. Originally, TKM is measured in thousand kilometers but, for our convenience, we have converted into million TKM.

### 3.1.2 Freight Rate (FR):

Theoretically it is considered that rail freight demand is a function of its own price and it is used as a major determinant of rail freight demand and there is significantly negative impact of freight rate on TKM over the long run (Wijiweera et al., 2014). A freight rate is a price at which a certain cargo is delivered from one point to another and it depends on the weight of the cargo, the distance of the delivery destination, form of the cargo, the mode of transport like train, truck, ship or aircraft. Many shipping services, especially air carriers, use dimensional weight for calculating the price, which takes into account both weight and volume of the cargo. We have collected data on freight rate from the year books of Ministry of railways. The variable was found mixed in paisa charged per tonne-kilometer and rupees charged per tonne-kilometer in different books. For standardization, we fixed the unit as rupees charged per tonne-kilometer.

[1] A tonne refers to metric tonnes and most of the railway systems in the world use this term to measure freight services, (Freight Transport for Development toolkit-Railway Transport, 2009)

It is of the view that only freight rate affects the rail freight demand but there are other macroeconomic variables which also influence the demand like GDP, share of manufacturing sector to GDP and domestic oil prices. Moreover, service characteristics which includes flexibility and reliability also have significant effect on mode choice for freight transport (Oum, 1979a; Lewis and Widup, 1982). We have restricted our analysis to macroeconomic variables only as these variables were not avaialable.

### 3.1.3 Gross Domestic Product (GDP):

GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets and degradation of natural resources. Data are in current local currency. Data have been obtained from World Bank. We have shifted the unit from rupees to million rupees for our convenience. There are mixed views about the relationship of GDP and rail freight demand. Some studies considers it a major determinant and some of them stated that:

Economic growth is the major determinant of freight transport demand in the long run and there is strong contemporaneous relationship between GDP and TKM that means when growth in economy occurs it will lead to increase rail freight demand so there is positive impact of GDP on TKM (Kulshreshtha et al., 2001). According to Bennathan et al. (1992), demand for Tonne kilometers for road, rail and water appears to have a higher (positive) elasticity with respect to GDP in the poor countries but the difference is not strictly significant. Herranz-Loncán, (2006) concludes that the contribution to railroads was lower to economic growth in Spain.

### 3.1.4 Domestic Diesel Oil Prices (DDOP):

When oil prices increase it has significant effect on transportation either it is rail, road, ship or air (Transport Economics and Management Systems Inc., 2008). In 2009-10, Pakistan railways had to bear expenses in the form of operation fuel only were 50.51% for total percentage of expenditures to gross earnings (Ministry of railways, 2010-11). The data for domestic oil prices have been obtained from Pakistan Bureau of Statistics. We have converted monthly data into annual by averaging monthly data. Unit for this variable used is rupees per liter.

### 3.1.5 Share of Manufacturing Sector to GDP (GDPMS):

Manufacturing sector is the third largest sector of the economy accounting for 14.7%, 18.3%, 18.5%, 15.7% and 15.2% share of GDP in 2000, 2005, 2010, 2011 and 2012 respectively. It includes Large Scale, Small Scale Manufacturing and Slaughtering included as a sub-category of this sector (Ministry of Finance, 2012-13). Most of the transaction of manufacturing sector had been done through Pakistan railways so it can be considered as an important determinant for affecting rail freight demand. The share of this sector in different countries is more than Pakistan e.g Thailand, China, Vietnam, Malaysia, Combodia (World Bank, 2007-08) As a result of rapid industrialization rail freight demand increases (Ramanthan, 2001; Kulsherestha et al., 2001). It is evident that GDP specifically from manufacturing sector is likely to affect the demand and it is important to analyze its affect. Data for this variable has been collected from Hand Book of Statistics of Pakistan and unit of analysis is million rupees.

## 3.2 Methodology:

Several methodologies have been used to analyze the demand for transportation in general and by mode. These include optimization models, models of modal choice, adhoc specified and estimated demand functions, derived demand model, Vector Auoregressive models and cointigration approach. Optimization models in the analysis of transport demand incorporate the interactions of commodity supply and demand conditions with transportation rates as well as constraints inherent in the system. The second type of transportation demand analysis is the estimation of a modal choice behavioral function (McFadden). This technique has been used in studies by Levin, Miklius et al., and Johnson (1976). Recently, Oum developed the theoretical assumptions underlying the use of linear logit models for transport demand studies (Spring 1979). The linear logit model imposes several rigid, a priori restrictions on estimated parameters and a structure of technology which may not be appropriate for use in the case of transport demand studies. A third type of demand analysis is specification of behavioral equations using adhoc models. These are characterized by regression models of shipments as a function of exogenous variables which are introduced without rigorous specification. These models are typically useful for forecasting but suffer in the analysis of price responsiveness of demand. The proper set of exogenous variables, and the functional form of the model, are both somewhat arbitrary. Coefficients estimated from these models are typically sensitive to the functional form and included exogenous variables. Estimation of derived demand models provides another methodology for analyzing modal demands for transportation. Friedlander and Spady applied these procedures to a cross section of shipments from U.S. manufacturers. Oum applied similar procedures in Canada using cross-sectional data (Autumn, 1979) and time series data (1978). The duality approach to analyzing intermodal competition in transportation is attractive because its functional specification is consistent with neoclassical economic relationships. Further, hypotheses about changes in behavior of demand parameters can easily be incorporated into the model and tested. All studies derive their econometric models from individual firm level behavior although they differ in their level of aggregation used to estimate the models and assumed role of mode choice decision maker in the firm's optimization problem.

Other than it traditional approaches of freight traffic growth models are often associated with several empirical problems such as the possible endogeneity of regressors as well as the non-stationarity of the variables. The standard classical methods such as the ordinary least squares (OLS) and hypotheses testing are based on the assumption that the time series are stationary. Since the distribution theory that applies to non-stationary series is different from the standard Gaussian asymptotic theory, use of classical estimation methods such as OLS for estimating relationships between non-stationary variables may give rise to misleading inferences or spurious regressions. The problems with estimation of single equation framework with integrated or non-stationary variables are: non-standard distributions of coefficient estimates, error process not being stationary, explanatory variables generated by processes that display autocorrelation, existence of more than one cointegrating vector and failure of weak exogeneity (Banerjee et al., 1986,1993). The remedy from problematic regressions with integrated variables is to test for cointegration and estimation of error-correction models to distinguish between short run and long run responses.

In the freight demand literature, most previous studies focussed on freight demand modelling, examining elasticities or modal choice based on either cross-section or time series data (see for example, the surveys by Zlatoper and Austrian, 1989; Graham and Glaister, 2004 and De Jong et al, 2004). Few studies of freight demand have employed the recent developments in multivariate dynamic econometric time series modelling, with notable exceptions being Bjorner (1999), Nag and Kulshrestha (2001) ) applied the cointegrating VAR model in modelling in Denmark and India respectively. Ramanathan (2001) applied Johansen and Juselious Cointegration approach (CI) and Error Correction Mechanism (ECM) in modelling and forecasting both passenger and freight transportation demand in India. But in our study we have used Autoregressive Distributed Lag Model for analysing short run and long run relationship among rail freight demand and its determinants . This technique is superior to other cointigrating approaches that it can apply on series either $I_{(0)}$, $I_{(1)}$OR mixture of $I_{(0)}$and $I_{(1)}$, suitable for finite sample data, can tackle problem of endogeniety and single equation model..

Econometrics is concerned with model building that typically begins with a statement of theoretical proposition that one variable is caused by other variables or some qualitative statement about the relationship between the variable and one or more covariates that are expected to be related. The next step is to convert this relationship into a set of equations with a view that it will answer some interesting questions about the variable of interest (Greene, 2010, p. 11). Our variable of interest is the rail freight demand which is represented by tonne-kilometers (TKM). Explanatory variables that possibly can affect rail freight demand growth are freight rate (FR), GDP, GDPMS and DOP. The data used for the period of 1981-2012. The functional form of economic relationship, can be defined as

$$TKM_t = f(FR_t, GDP_t, GDPMS_t, DDOP_t)$$

On the basis of the hypothesis we have developed the following economic relationship:

$$TKM_t = f(FR_t, GDP_t, DDOP_t)$$

$$TKM_t = f(FR_t, GDPMS_t, DDOP_t)$$

Where

$TKM$ = Tonne Kilometers (Dependent Variable)

and explanatory variables are

$FR$= Freight Rate

$DGP$= Gross Domestic Product

$DDOP$= Domestic Diesel Oil Prices

$GDPMS$= Share of GDP from Manufacturing Sector and 't' shows the time period. For analyzing the above economic relationship we have used level-log model because it better explains the effect of explanatory variables on dependent variable, (Wooldridge, 2012, p. 128). From the above relationship the econometric model can be formed in the following way:

$$TKM_t = \alpha_0 + \beta_1 LnFR_t + \beta_2 LnGDP_t + +\beta_3 LnDDOP_t + \mu_{1t}$$

$$TKM_t = \delta_0 + \lambda_1 LnFR_t + \lambda_2 LnGDPMS_t + +\lambda_3 LnDDOP_t + \mu_{2t}$$

Before examining the long run relationship between explanatory variables and the variable of interest, we have applied Augmented Dickey Fuller (Dickey, 1979) test to check the order of integration of the series. At the second stage, we have used Vector Auto Reggressive (VAR) Model to observe lag length of the series. The used criteria of observing lag length are Akaik Information Criteria (AIC), Schwartz Bayesian Criteria (SBC) and Hannan Criteria (HC). Third, we have used Auto Regressive Distibuted Lag (ARDL) approach. For examining the stability of parameters, we have applied CUSUM and CUSUM-squared test and different diagnostic tests like ARCH LM Test for Heteroscedasticity, Breusch Godfrey Lagrange Multiplier Test for Autocorrelation and Jarque-Bera Test for Normality.

### 3.2.1 Augmented Dickey Fuller Test:

ADF tests are most commonly used with the aim to check the stationarity level of the series. By stationarity, we mean that the average, variance and autocorrelation structure of the data do not change over time. Most economic variables that show strong economic trends like GDP, consumption and price levels are not stationary but stationarity can be achieved by differencing or some simple transformations (Greene, 2010, p. 942). According to Enders

(2010, p. 1) time series econometrics is concerned with the estimation of difference equations. It is now important to pre-test a series that either it is trend stationary or differenced stationary (Cochrane, 1991). If the series is stationary at level then it is denoted by $I_{(0)}$, read as integrated of order zero. If the series contains a unit root or becomes stationary at first difference then it is denoted by $I_{(1)}$ and $I_{(d)}$, read as integrated of order $d$ if it considered $d^{th}$ difference (Greene; 2011,p. 631, 645). Stationarity from the trended series can be achieved by using different measures like differencing and detrending.

There can be different forms in which unit root can occur. These forms can involve Random Walk model without drift term or Pure Random Walk:

$$Y_t = Y_{t-1} + \mu_t$$

Random Walk with drift Term:

$$Y_t = \delta_0 + Y_{t-1} + \mu_t$$

Deterministic Trend:

$$Y_t = \delta_0 + \delta_1 t + \mu_t$$

This type of unit root can be removed by using detrending. Due to this reason it is also known as trend stationary.

Random Walk with Drift and Deterministic Trend:

$$Y_t = \delta_0 + \delta_1 t + Y_{t-1} + \mu_t$$

With the help of differencing, a non-stationary series can be made stationary by an appropriate Autoregressive Moving Average (ARMA) Models but in the case of deterministic trend and a pure noise component, detrending will be used because differencing cannot convert it into an autoregressive process (Enders, 2010; p. 191).

ADF test is widely used to find out the unit root process. Let a $Y_t$ series has $n$ observations like $Y_1, Y_2, \ldots, Y_n$ which generate the model

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

which shows a random walk model without drift with $Y_0 = 0$, $\rho$ is a real number and $\varepsilon_t$ is white noise[2]. If the $|\rho| = 1$ then it shows that series has unit root or non-stationary and variance becomes $t\rho^2$ and if $|\rho| < 1$ then series will be stationary and if $|\rho| > 1$then series also contains unit root and variance increases exponentially as $t$ increases (Dickey and Fuller, 1979).

The assumption of DF test is that $\varepsilon_t$ is uncorrelated. Contrary to this, if correlation occurs then we move towards using Augmented Dickey Fuller (ADF) Test. In this test it is assumed that if $\varepsilon_t$ is correlated then add a lagged difference of the dependent variable until the autocorrelation is not been removed. For example in the case of Random Walk with Drift and Deterministic Trend model

$$\triangle Y_t = \delta_0 + \delta_1 t + \delta_2 Y_{t-1} + \sum_{i=1} \beta_i \triangle Y_{t-i} + \xi_t$$

ADF test has been preferred in the literature over Phillips-Perron (1988) and KPSS unit root tests because later are more sensitive about the time series (Gujrati and Porter; 2005; p. 754-760). ADF test is an extension of DF. The main problem in DF is that it donot take into account the autocorrelation prolem in residuals that is the residuals are not white noise. ADF test tackle this problem by introducing the lags of the dependent variable in the model. It is a parametric test but Phillips-Perron is a non-parametric test which tries to tackle autocorrelation by using Heteroscedastic Autocorrelated Standard Errors (HAC) and there are many questions on the power of the test. Although there is some concern about

[2]It means that $\varepsilon_t$is normally, independently and identically distributed with zero mean and constant variance OR A stationary process for which all autocorrelations are zero (Lutkepohl, 2004; p. 13)

the true functional form of ADF test but we manage to estimate its proper model then we can rely on this test.

### 3.2.2 Vector Autoregressive (VAR) Model:

The VAR models were introduced by Sims in 1980. According to him it is suitable to estimate large-scale macroeconomic models as unrestricted reduced form equations and assume all variables as endogenous. He applied this model on West Germany and US quarterly data after the post war period which was applicable to the series GNP, unemployment rate, price level and import price index (Sims, 1980). Sims (1980b) claimed that unrestricted VAR's can provide better information to understand macroeconomic relationship than the structural models. It is convenient way to model a several time series simultaneously (Davidson and McKinnon,1999; p. 614). As in univariate time series an autoregressive process of order $p$ or $AR_{(P)}$ is

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + \alpha_p Y_{t-p} + \psi_t$$

In the above model dependent variable is regressed on its own lags up to order $p$ but in VAR model lagged values of each variable will be appeared in each equation. To explain this let the series $y_{1t}, y_{2t}, \ldots, y_{pt}$ then

$$y_{1t} = \alpha_0 + \alpha_1 y_{1t-1} + \alpha_2 y_{2t-1} + \ldots\ldots\ldots\ldots\ldots + \alpha_p y_{pt-1} + \psi_{1t}$$

$$y_{2t} = \beta_0 + \beta_1 y_{1t-1} + \beta_2 y_{2t-1} + \ldots\ldots\ldots\ldots\ldots + \beta_p y_{pt-1} + \psi_{2t}$$

$$\ddots$$

$$y_{pt} = \gamma_0 + \gamma_1 y_{1t-1} + \gamma_2 y_{2t-1} + .............. + \gamma_p y_{pt-1} + \psi_{pt}$$

It can be written as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ . \\ . \\ y_{pt} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \\ . \\ . \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & . & . & . & \alpha_p \\ \beta_1 & \beta_2 & . & . & . & \beta_p \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \gamma_1 & \gamma_2 & . & . & . & \gamma_p \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ . \\ . \\ y_{pt-1} \end{bmatrix} + \begin{bmatrix} \psi_{1t} \\ \psi_{2t} \\ . \\ . \\ \psi_{pt} \end{bmatrix}$$

In vector form it can be written as

$$y_t = m + A_1 y_{t-1} + A_2 y_{t-2} + ................ + A_p y_{t-p} + \psi_t$$

where $m$ is vector of constants of order $k \times 1$, $A_i$are the $k \times k$ matrices of coefficients, $\psi_t$ is the vector of white noise error terms. There are two methods for the estimation of VAR models like the direct approach as above and the alternative is the reparametrization. In each VAR's above every variable is expressed as linear combination of its own lag and lagged values of all other variables in a group (Jhonston, **2010** , p. 287-295). For selecting the appropriate lag length testing based approach like Wald Statistic or information criterion approach like Akaik Information Criteria (AIC), Schwartz Bayesian Criteria can be use (Hansen,2013; p. 286-287).

## 3.3 Autoregressive Distributed Lag (ARDL) Model OR Bound Testing Approach:

This approach was introduced by Pesaran et al. (2001) which is based on his own paper in (1995,1999). It is also known as Bound Testing Approach to find out long run relationship between time series variables and it is a single equation model. It also finds the short-run and long-run dynamics. This approach is distinguished than others on the following grounds: In the first place the variables to be studied either are $I_{(0)}$ OR $I_{(1)}$ OR mixture of both; Secondly, it is appropriate for small sample data; Thirdly, it do not require the pre-testing of unit root; Fourthly, it considers all variable as endogenous; and finally, it is most appropriate than any other approach like Engle Granger which can be applied only for two variables, the multivariate cointigration techniques like Stock and Watson (1988), Johansen (1988), Johansen and Juselous (1990) are more appropriate for large samples but Johansen and Juselous approach also requires that all variables should be integrated of order one. But ARDL can be applied on either order of intigration. In spite of not requiring pre-testing of unit root, we have used this in our analysis to know the second differenced stationarity OR $I_{(2)}$because the assumption of this approach is that no any variable should be $I_{(2)}$. This approach avoids the problem of endogeniety and can simultaneously estimate the short run and long run estimates of the model and also removes the problems associated with serial correlation (Pesaran et al., 1999, 2001). The major draw back of this technique is that it collapses in the presence of second order difference. In this type of model the explanatory variables may include lagged values of the dependent variables and current and lagged values of the one or more regressors (Jhonston, 2010; p. 244). The general form of ARDL model with $n$ lags of $Y_t$and $m$ lags of $X_t$will be as follows:

$$Y_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i Y_{t-i} + \sum_{j=0}^{m} \beta_j X_{t-j} + \varepsilon_t$$

and the general form ARDL Equibrium Correction Mechanism (ECM) or Equibrium Correction Term (ECT) will be

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-i} + \alpha_2 X_{t-j} + \sum_{i=1}^{m} \delta_i \Delta Y_{t-i} + \sum_{j=0}^{m} \pi_j \Delta X_{t-j} + \varepsilon_{1t}$$

### 3.3.1 Diagnostic Tests:

The consequences of model mis-specification in regresssion analysis can be severe in terms of adverse effects on the sampling properties of both estimators and tests. Accordingly econometric literature places a good deal of emphasis on procedures for interrogating the quality of model's specification These procedures address the assumptions that may have been made about the distribution of the model's error term, and they also focus on the structural specification of the model, in terms of its functional form, the choice of regressors, and possible measurement errors. There is no doubt that diagnostic testing is now firmly established as a central topic in both econometric theory and practice, that test, test and test (DeBenedictis, 1996). So, for diagnostic checking following tests have been used Jarque Bera Test for normality, ARCH LM Test for Hetroscedasticity and Breauch Godfrey Lagrange Multiplier Test for Autocorrelation.

#### 3.3.1.1 Jarque Bera Test:

Statistical errors are common in scientific literature. The assumption of normality needs to be checked for many statistical procedures, namely parametric tests, because their validity depends on it. There are nearly 40 tests of normality are available in the literature. The effort of developing techniques to detect departures from normality was initiated by Pesaran (1895) who worked on skewness and kurtosis coefficients. These tests are differ in the characteristics of distributions. These tests also differ in the level at which they compare the empirical distribution with the normal distribution (Razal et al., 2011). This test introduced by Lomonicki (1961) and Jarque and Bera (1987) which based on skewness and Kurtosis of the distribution (Lutkepohl, 2004; p. 45). This test is used for checking the normality of the distribution of the error term of the model. It is the basic assumption that the distribution of $\varepsilon_t$ should be normally distributed with zero mean and constant variance. It is useful to obtain many exact results like confidence intervals and test statistics but normality is necessary to have results we use in multiple regressions (Greene, 2012; p. 276). The power of this test increases in case of increasing the size of the population for the distribution. The null and alternative hypotheses are as follows:

$H_0$ =The distribution of random elements is normal;

$H_1$ =The distribution of random elements is non-normal.

The test statistic to be used is

$$JB = \frac{N}{6}SK + \frac{N}{24}KU$$

With $\chi^2$distribution of $2p$ degrees of freedom. In the above formula $SK$ denotes Skewness and $KU$ shows Kurtosis. If calculated values of $JB \leq \chi^2_{2p}$, then accept $H_0$ and reject otherwise (Domanski, 2010). Jarque- Bera test is not able to detect normality in case of outliers *i.e* it is sensitive to outliers especially at left, right and symmetric contamination (Brys et al., 2004).

#### 3.3.1.2 Autoregressive Conditional Heteroscedasticity (ARCH) Lagrange Multiplier Test/Godfrey Lagrange Multiplier Test:

This test is based on the framework of Lagrange Multiplier (Breusch and Pagan, 1979) . There are many tests to detect hetroscedasticity. Gold-Feld Quadnt test which is applicable to only one regressor. White test is extremely general because it do not account for any assumptions about the nature of heteroskedasticity. At the same time, if the hypothesis about heteroskedasticity do not gets rejected, there is no way out. We have adopted ARCH LM Test and this test is more powerful in the absence of normality. This test is sensitive to the assumption of normality as well (Greene, 2012; p. 276).

#### 3.3.1.3 Lagrange Multiplier(LM) Test:

Most of the time series data show autocorrelation of the disturbances across periods due to omitted factors or inclusion of variables which are correlated across series (Greene, 2012, p. 903). The tests available for the detection of autocorrelation based on the principle that error terms are autocorrlated. Box and Pierce's Test and Lung's Refinement (1970) or Q-Test, The Durbin Watson Test (1970) and Lagrange Multiplier Tests are used for the detection of autocorrelation. We have preferred Lagrange Multiplier Test because of the limitations of other tests [3]. LM test was introduced by Breusch and Godfrey in 1978 (Greene, 2012; p. 923-924). The LM test is accurate for testing higher order autocorrelations in dynamic models. The one sided tests were introduced by Majumder and King (1989), Basak, Roise and Majumder (2005, 2008) but these tests only tested autocorrelations in linear regression models and are suitable for dynamic models (Roise et al., 2012). The null and alternative hypotheses are as follows:

$H_0$ : There is no autocorrelation.

$H_1$ : There is autocorrelation.

### 3.3.2 Stability Tests:

Since the earliest days of macroeconometric analysis, researchers have been concerned about the appropriateness of the assumption that model parameters remain constant over long periods of time. This concern is also central to the so-called the Lucas (1976) critique which has played a central role in shaping macroeconometric analysis in the last thirty years. Lucas (1976) emphasizes the fact that the decision models of economic agents are hard to describe in terms of stable parametrizations, simply because changes in policy may change these decision models and their respective parametrization. These arguments un-

[3]Durbin Watson test can only be applied on the first order autoregressive process and Box and Pierce's test becomes less powerful than LM test when the null hypothesis that there is no autocorrelation is rejected but asymptotically both LM and Box and Pierce's tests are equal.

derscore the importance of using structural stability tests as diagnostic checks for macroeconometric models (Boldea, 2011).

Most widely used tests for stability are CUSUM and CUSUM Square. This technique was introduced by Brown et al. (1975). CUSUM and CUSUM square test can be applied to measure the parameters constancy for single equation and these tests can also be applied for individual equations to vector models. These tests are valid for VAR and VECM models and efficient for single equation models. For parameters stability CHOW test can also be applied (Lutkepohl, 2004; p. 132). In our empirical work, we have applied these tests for checking the stability of the parameters. CUSUM test analyze the instability within the range of 0.05 percent and CUSUM Squared test performed on the squares of the residuals (Ahsen et al., 2011).

## 3.4 Empirical Model:

We have discussed about the methodology that we shall apply. On the basis of hypotheses stated above, two empirical ARDL Models for finding the long-run relationships and the short run dynamics will be estimated as under:

$$\Delta TKM_t = \alpha_0 + \alpha_1 TKM_{t-1} + \alpha_2 LnFR_{t-1} + \alpha_3 LnGDP_{t-1} + \alpha_4 LnDDOP_{t-1} + \sum_{i=1}^{m} \delta_{1i} \Delta TKM_{t-i}$$

$$+ \textstyle\sum_{i=0}^{n} \delta_{2i} \Delta LnFR_{t-i} + \sum_{i=0}^{n} \delta_{3i} \Delta LnGDP_{t-i} + \sum_{i=0}^{n} \delta_{4i} \Delta LnDDOP_{t-i} + \psi ECM_{t-i} + \varepsilon_{1t} \cdots (1)$$

$$\Delta TKM_t = \alpha_0 + \alpha_1 TKM_{t-1} + \alpha_2 LnFR_{t-1} + \alpha_3 LnGDPMS_{t-1} + \alpha_4 LnDDOP_{t-j} + \sum_{j=1}^{p} \pi_{1j} \Delta TKM_{t-j} +$$

$$\textstyle\sum_{j=0}^{q} \pi_{2j} \Delta LnFR_{t-j} + \sum_{j=0}^{q} \pi_{3j} \Delta LnGDPMS_{t-j} + \sum_{j=0}^{q} \pi_{4j} \Delta LnDDOP_{t-j} + \psi ECM_{t-j} + \varepsilon_{2t} \cdots (2)$$

In the models, lagged terms represent the long-run part and the differenced terms the dynamics in the short-run and $ECM_{t-i}$ is equilibrium correction mechanism orequilibrium correction term (ECT) which will be obtained from simple regression cointigrating model and $\psi$the speed of adjustment when a shock occurs. The number of lag difference will be observe by using information criterion like AIC and SBC.

## 3.5 Bound Testing Procedure:

ARDL or Bound Testing long-run relationship will be analyzed under the null and alternative hypothese:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

OR There exists no long-run relationship;

$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq 0$

OR There exists long-run relationship between the variables.

For evaluation about the joint significance of the lagged levels of the variables in a conditional unrestricted equilibrium error correction model Wald OR F-statistic can be used but F-statistic is non-standard irrespective of whether the variables are $I_{(0)}$ OR $I_{(1)}$OR mixed relationship. For this purpose Pesaran (2001) provided two sets of variables in which first presumes that all variables are $I_{(0)}$and other presumes values containing all variables are $I_{(1)}$. These sets of values form a band which includes all expected categorization of all orders like $I_{(0)}, I_{(1)}$OR even partially integrated. By using these values proposed by Pesaran (2001) if Wald OR $F_{calculated} < F_{critical}$ then $H_0$ will not be rejected and concluded that long-run relationship does not exist but if $F_{calculated} > F_{critical}$ then reject $H_0$ and concluded with the existence of long run relationship.

# Chapter 4

# Results Discussion and Analysis:

After discussing the data sources, we analyze the impact of different explanatory variables on rail freight demand on empirical grounds. To analyze these issues, we will give us a deep insight to draw some conclusion on the basis of empirical results of this research. The results are discussed as follows

## 4.1 Descriptive Statistics:

The descriptive statistics of the study are presented in the Table 4.1. Descriptive statistics consists of procedures used to summarize and describe the characteristics of a set of data. The table shows the averages values, standard deviation, skewness, kurtosis and J. Bera values of the selected variables.

Table 4.1: Descriptive Statistics of Variables

| *Variables* | *Mean* | *Max.* | *Min.* | *Std.Dev* | *Skeness* | *Kurtosis* | *J.Bera* | *Prob* |
|---|---|---|---|---|---|---|---|---|
| *TKM* | 5425.78 | 8363.91 | 402.48 | 1978.29 | -0.8543 | 3.7759 | 4.6947 | 0.09 |
| *FR* | 0.9206 | 4.1600 | 0.2120 | 0.8307 | 2.4736 | 9.7178 | 92.8074 | 0.00 |
| *GDP* | 4576066 | 20090862 | 278196 | 5443781 | 1.5343 | 4.3740 | 15.0729 | 0.00 |
| *DDOP* | 19.60091 | 92.11200 | 2.95000 | 24.15721 | 1.6911 | 4.8506 | 19.8188 | 0.00 |
| *GDPMS* | 701593 | 2678485 | 37446 | 817104.8 | 1.2859 | 3.3069 | 8.9447 | 0.01 |

The descriptive statistics i.e mean, median, maximum, minimum, standard deviation, skewness, kurtosis, Jarque Bera and probability values given in the table above. All the variable have been described after taking log except tonne kilometers (TKM) i.e we have used level log model. The table is almost self explanatory. The mean tells us about the average values of the data and median about the most middle observations of data. The standard deviation tells us about the deviation of the series from its mean for policy analysis. This deviation is not desirable. The skewness and kurtosis tells us that whether the series follows the normal distribution or not. Positive skewness means that the most of the observations lies to the right of its mean value while negative skewness means most of the observations lies to the left of its mean value. Similarly kurtosis tells us about the peakedness of the data that is whether the series is lepto- kurtic, platy-kurtic or meso-kurtic. In our case, skewness of all variables found to be greater than zero and positive which indicates the distribution is positively skewed. Kurtosis values of all the variables indicating platy-kurtic distribution due to values greter than 3. Jarque Bera values indicating that error terms are normally distributed with zero mean and constant variance. The values of Jarque-Bera showed errors are not nomally, independently and identically distributed.

## 4.2 Unit Root Test:

Before the estimation, the time series property of the data has been examined by using unit root test. The ARDL bounds test is based on the assumption that the variables should be $I_{(0)} or\ I_{(1)}$but the dependent variable should be integrated of order one so before applying this test we have determined the order of integration of all variables using ADF test. The objective of this test is to ensure that none of the variables should be $I_{(2)}$to avoid spurious results because in the presence of variables integrated of order two we cannot interpret the values of F-statistics provided by Pesaran et al. (2001).

In the table 4.2 it can be shown that all the variables contain unit root at level with intercept and with intercept and trend but all series become stationary at first difference showed by ***,** at 1% and 5%. As in our ARDL model can be applied either the series are $I_{(0)} or$ $I_{(1)} OR$ mixture of both and it is relatively more appropriate for small and finite sample data.

Table 4.2: Test of Non- Stationarity of variables

| Variables | With Intercept | | With Intercept and Trend | |
|---|---|---|---|---|
| | t-stat | p-value | t-stat | p-value |
| $TKM$ | -1.152010 | 0.6815 | -2.922644 | 0.1699 |
| $\Delta TKM$ | -3.564150 | 0.0129*** | -3.591010 | 0.0477** |
| $LnFR$ | 0.625933 | 0.9882 | -1.172639 | 0.8988 |
| $\Delta LnFR$ | -4.454638 | 0.0014 *** | -4.532680 | 0.0057*** |
| $LnGDP$ | 0.758353 | 0.9916 | -2.360460 | 0.3915 |
| $\Delta LnGDP$ | -6.063158 | 0.0000 *** | -6.224698 | 0.0001 *** |
| $LnDDOP$ | 2.391776 | 0.9999 | -1.279957 | 0.8743 |
| $\Delta LnDDOP$ | -4.921975 | 0.0004*** | -6.267835 | 0.0001*** |
| $LnGDPMS$ | -0.752735 | 0.8184 | -2.000649 | 0.5781 |
| $\Delta LnGDPMS$ | -4.565060 | 0.0010*** | -4.505341 | 0.0061 *** |

## 4.3 ARDL Models Estimation:

To estimate the "Autoregressive Distributed Lag Model" with both short and long run dynamics we have used general to specific approach on the basis of AIC and SBC by estimating the vector autoregressive model. Most preferably AIC has been used to select lag length which suggested maximum two lags for all the variables. All the variables chosen are in log form except the dependent variable (TKM) in both models. In the first step long run relationship is needed so in order to examine it "Bounds test" has been used and then applied different diagnostic and stability tests on the estimated model.

### 4.3.1 Results of ARDL Model-I:

For the analysis of short and long run relationships among the variables TKM, FR, GDP and DOP we have first estimated basic ARDL model after removing the insignificant lags and applied diagnostic tests like Breusch Godfrey LM Test for autocorrelation, ARCH LM Test for Heteroscedasticity and Jarque-Bera Test for normality. Stability tests for checking parameters constancy and then applied bounds tests to confirm long run relationship. After the confirmation of long run relationship we have estimated the Equilibrium Correction Model (ECM) and long run estimates. We have estimated the following ARDL model.

$$\Delta TKM_t = \alpha_0 + \alpha_1 TKM_{t-1} + \alpha_2 LnFR_{t-1} + \alpha_3 LnGDP_{t-1} + \alpha_4 LnDDOP_{t-1} + \sum_{i=1}^{m} \delta_{1i} \Delta TKM_{t-i} +$$

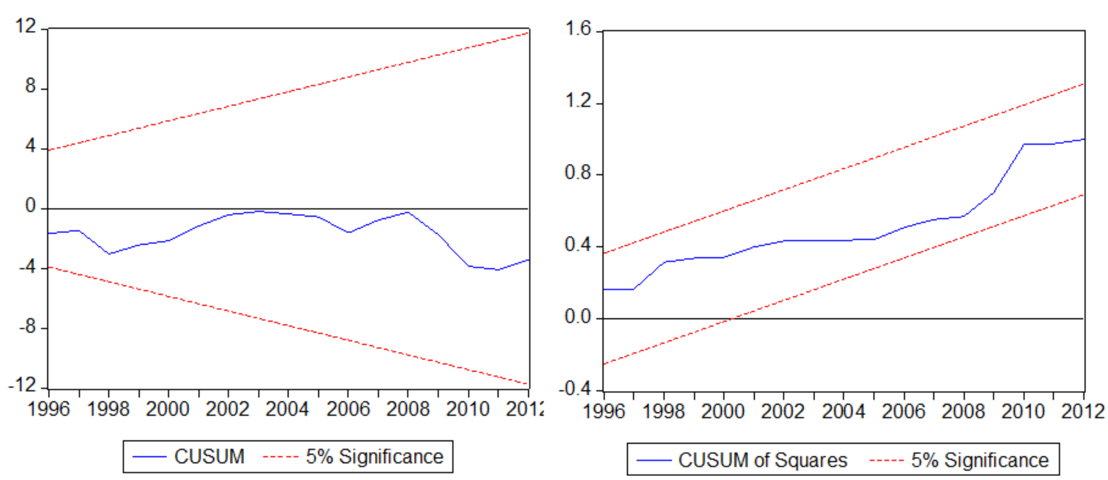
$$\sum_{i=0}^{n} \delta_{2i} \Delta LnFR_{t-i} + \sum_{i=0}^{n} \delta_{3i} \Delta LnGDP_{t-i} + \sum_{i=0}^{n} \delta_{4i} \Delta LnDDOP_{t-i} + \psi ECM_{t-i} + \varepsilon_{1t} \ldots\ldots\ldots\ldots (1)$$

The results of the basic ARDL Model-I can be seen in appendix table 5.1

#### 4.3.1.1 Stability Tests:

The following diagram showing results of stability tests applied on basic ARDL Model-I

Figure 4.3.1: CUSUM and CUSUM Squares Tests




The Plot of cumulative sum of recursive residual (CUSUM) and Cumulative sum of squares of recursive residual (CUSUMSQ) statistic indicate no evidence of mis-specification and structural instability for the estimation period.

#### 4.3.1.2 Bounds Test:

To determine the long run relationship among the variables we test the hypothesis that the coefficients of the lag level variables in the basic above ARDL model are equal to zero as:

$Hypothesis\ Testing:$

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq 0$

where

$\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$ showing the long run coefficients

First difference of the variables explain the short run effect of explanatory variables on the dependent variable. Based on the redundant variable test we obtained F-statistic equal to 7.59 provided by Pesaran (2001) is higher than the upper bound F-statistic unrestricted intercept and no trend tabulated value 4.35 $i.e\ F_{calculated} > F_{critical}$ null hypothesis gets rejected of no relationship among the variables and accepted the alternative having long run relationship. It can also be shown table 4.3.

Table 4.3: Cointigration Test

| | |
|---|---|
| $F-statistic\,(4,17)$ | 7.590 |
| $F-critical$ | 4.35 |
| $p-value$ | 0.0011 |

After the confirmation of long run relationship we have estimated the short and long run impacts of variables on rail freight demand growth.

#### 4.3.1.3 Long Run Estimates:

Long run cointigrating vectors (long-run coefficients) are obtained by normalizing the lagged level variables of ARDL-I model with the lagged coefficient of dependent variable and hence obtained the long-run elasticities. Results of the long run estimates can be seen in the following table 4.4.

Table 4.4: Long run Estimates after Normalization

| $Variables$ | $Long\ run\ effects$ | $Std.Error$ | $t-stat$ |
|---|---|---|---|
| $LnFR$ | $-9798.424$ | 1172.29 | $-8.3584$*** |
| $LnGDP$ | 5431.660 | 1327.444 | 4.0919*** |
| $LnDDOP$ | $-2995.414$ | 881.706 | $-3.3973$*** |

*** shows that the variables are significant at 1%

#### 4.3.1.4 Estimates of the Equilibrium Correction Mechanism:

For observing the short run impact of variables we shall estimate the following Equilibrium Correction Mechanism (ECM)

$$\Delta TKM_t = \sum_{i=1}^{m} \delta_{1i} \Delta TKM_{t-i} + \sum_{i=0}^{n} \delta_{2i} \Delta LnFR_{t-i} + \sum_{i=0}^{n} \delta_{3i} \Delta LnGDP_{t-i} +$$

$\sum_{i=0}^{n} \delta_{4i} \Delta LnDDOP_{t-i} + \psi ECM_{t-i} + \omega_t \cdots (1)$

Table 4.5: Short run Estimates Results

| $Variables$ | $Coefficients$ | $Std.\ Error$ | $t-stat$ | $p-value$ |
|---|---|---|---|---|
| $C$ | -1195.166 | 419.4545 | -2.849335*** | 0.0103 |
| $\Delta TKM_{t-1}$ | -0.813 | 0.238369 | 3.407394*** | 0.0030 |
| $\Delta TKM_{t-2}$ | -0.459 | 0.202201 | -2.272894** | 0.0348 |
| $\Delta LnFR$ | -3832.103 | 1904.894 | -2.011715* | 0.0587 |
| $\Delta LnFR_{t-1}$ | -1324.665 | 1923.708 | 0.688600** | 0.0494 |
| $\Delta LnFR_{t-2}$ | -3024.749 | 1855.123 | -1.630484* | 0.0595 |
| $\Delta LnGDP$ | 25131.22 | 7109.806 | 3.534727*** | 0.0022 |
| $\Delta LnDDOP$ | -4809.736 | 2956.439 | -1.626868** | 0.0202 |
| $ECM_{t-1}$ | -0.921091 | 0.390897 | -4.147109*** | 0.0005 |
| $R^2$ | | | 0.846187 | |
| $Adj\_R^2$ | | | 0.797212 | |
| $DW$ | | | 1.897504 | |
| $Diagnostic\ Tests$ | | | | |
| $Serial\ Correlation\ LM\ Test$ | | | | 0.6867[0.5166] |
| $ARCH\ Test$ | | | | 0.6275[0.4358] |
| $Jarque-Bera\ Test$ | | | | 4.5715[0.1017] |

***,**,* shows that the variables are significant at 1%, 5%, and 10% respectively.

p-values of diagnostic tests are shown in square brackets. Breusch Godfrey LM Serial Correlation Test, ARCH test have been concluded on the base of F-statistics and Jarque -Bera Normality Test on the base of Chi-Square test. The results of stability and normality tests of ECM are as follows:

The empirical results of Table 4.4 and 4.5 showed that all explanatory variables are significant and have expected alternating (positive, negative ) relationship with rail freight demand in short and long run. Short run is a period in which a firm can't change its fix factors. The coefficient of freight rate varied negatively with rail freight demand in all lagged changes and long run. The negative relationship between freight rate and rail freight demand is consistent with the findings of Wardman (1997), Lewis (1982), Levine (1981), Mitchell (2010) and Wijeweera (2014). This negative relationship may be due to the fact that demand always negatively effected by price. So high freight rates harms rail freight demand in both the dynamics.

The macroeconomic activity coefficient GDP provided direct evidence to effect rail freight demand. This relationship may be due to the fact that the increase in freight demand is always closely related to goods production component of GNP or GDP. Secondly, with the increase in GDP, exports will increase and ultimately it may effect rail freight demand. This relationship is broadly consistent with Wijeweera (2013), Kulsherestha (2001) and Bonilla (2009). It showed expected behavior that GDP has positive relationship with most of the economic variables.

The domestic oil prices used as to measure the macroeconomic activity effect. This variable has not been used commonly as the determinant for freight railway but it is an important variable because Pakistan railways uses diesel oil for transportation that's why it may have effect on rail freight demand. So, this variable is included in the model and after estimation it has come that it has significant and negative impact of rail freight demand in both short and long run. A significant increase in real fuel prices is likely to result in greater rate increases for the faster modes than for the slower ones and some corresponding shift of demand across modes. In our analysis, rail freight demand responded more to oil prices in lagged changes as compare to the long run estimates because increased fuel prices have ever immediate implication which increases costs and ultimately freight rates which discourages the clients and rail freight demand. Other than it railways requires more fuel to run than its major competitor like truck then increase in oil prices will not shift trucks transport demand to rail freight demand. It can be seen from the tables 4.4, 4.5.

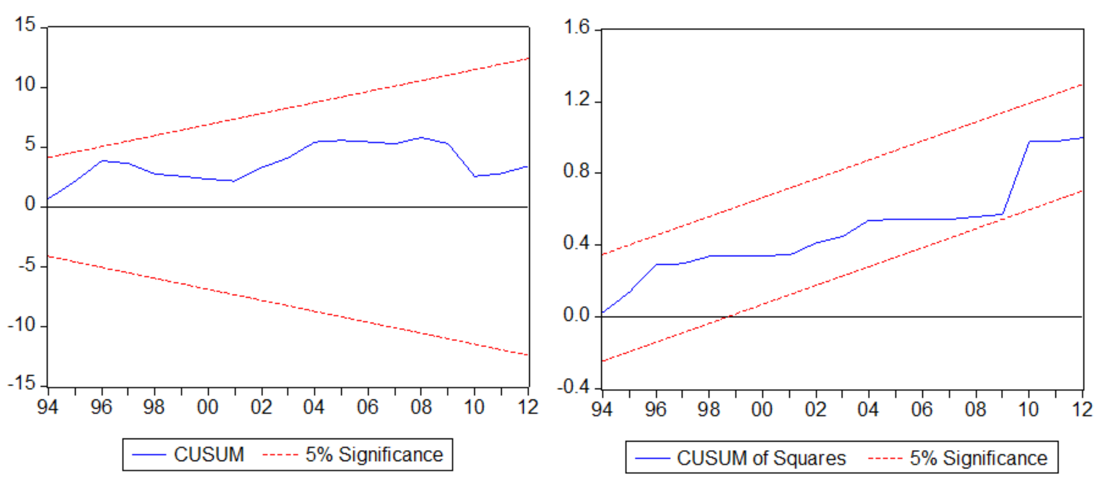
The empirical results of diagnostic tests are significant and reasonably well that is, the residuals of the estimated model are normally distributed, have no serial correlation and no heteroscedasticity.

The ECM shows the speed of adjustment to restoring the equilibrium in the short run. The lag of ECM indicates that how quickly or slowly the variables return to the equilibrium. In the analysis ECM is found to be (-0.921091) and highly significant which indicate that the deviation from mean is corrected by 92% within one year towards equilibrium and justify the Bannerjee (1998) condition for the existence of stable long run relationship between the variables. It represents that reversion to the mean is too quick. Other than it $R^2 and$ $Adjusted\_R^2$ of ARDL and ECM are high that shows the variations in the dependent variable is significantly explained by the independent variables included in the model means that the importance of keeping these variables in our model is significant.

#### 4.3.1.5 Stability Test on ECM:

After the estimation of ECM we have applied different diagnostic tests and stability tests. Figure 4.3.2 showed that parameters are stable. The dotted lines indicates critical bounds at 5% level of significance

Figure 4.3.2: CUSUM and CUSUM Squares Test



### 4.3.2 Results ARDL Model-II:

For the estimation of relationship among TMK, FR, GDPMS, DOP, we used general to specific approach on the basis of AIC and SBC by estimating VAR model. Both criteria's for lag selection indicated two lags for each variable. The following ARDL model have been estimated and its results are reported in table 4.6

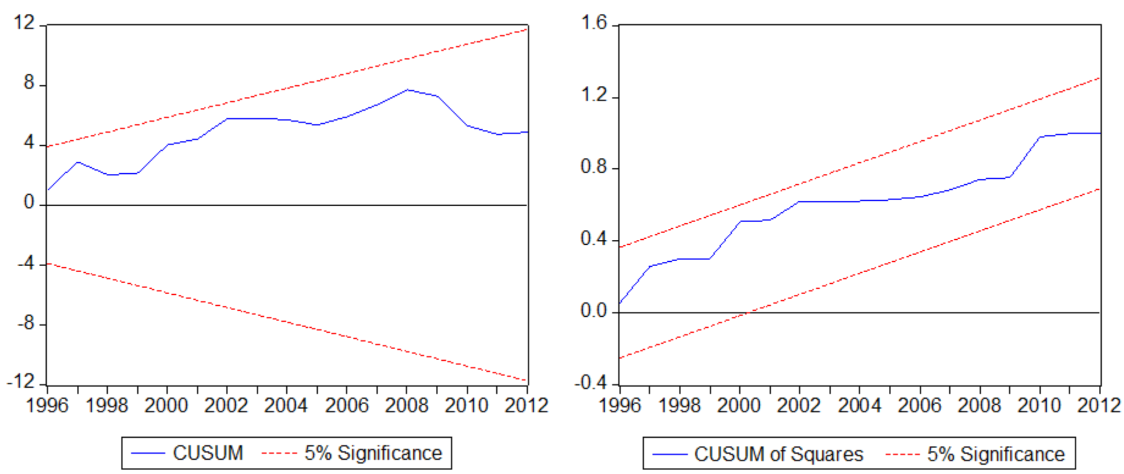
$$\Delta TKM_t = \alpha_0 + \alpha_1 TKM_{t-1} + \alpha_2 LnFR_{t-1} + \alpha_3 LnGDPMS_{t-1} + \alpha_4 LnDDOP_{t-1} + \sum_{j=1}^{p} \pi_{1j} \Delta TKM_{t-j} +$$

$$\sum_{j=0}^{q} \pi_{2j} \Delta LnFR_{t-j} + \sum_{j=0}^{q} \pi_{3j} \Delta LnGDPMS_{t-j} + \sum_{j=0}^{q} \pi_{4j} \Delta LnDDOP_{t-j} + \psi ECM_{t-j} + \eta_t \text{..........}(2)$$

#### 4.3.2.1 Stability Tests:

For diagnosting stability of the parameters we have applied CUSUM and CUSUM square test and the figure 4.4.1 represents parameters are stable.

Figure 4.3.3: CUSUM and CUSUM Squares Tests Results




#### 4.3.2.2 Bounds Test:

For verifying the existence of long run relationships among variables of the model we have applied bound test procedure as:

*Hypothesis Testing*

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

*i.e* There exists no long run relationship among variables

$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq 0$

*i.e* There exists long run relationship among variables

where

$\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$ are the long run coefficients of the model-2.

Based on the bounds values provided by Pesaran et al. (2001) we found that the computed F-statistics value 9.0667 is greater than the upper bound F-critical 4.35 which rejects the null hypothesis of non existence of long run relationship and accepted $H_1$ . It can be shown in the following table 4.6

Table 4.6: Cointegration Test

| | |
|---|---|
| $F-statistics\ (4,17)$ | 9.0667 |
| $F-critical$ | 4.35 |
| $p-value$ | 0.0004 |

To have the estimates of long run we normalize all explanatory variables by TKM.

#### 4.3.2.3 Long run estimates:

Table 4.7: Long run Estimates after Normalization

| $Variables$ | $Long\ run\ effects$ | $Std.Error$ | $t-stat$ |
|---|---|---|---|
| $LnFR$ | $-7771.084$ | 952.39 | $-8.1596$*** |
| $LnGDPMS$ | 3956.617 | 965.5473 | 4.0977*** |
| $LnDDOP$ | $-2249.165$ | 716.1455 | $-3.1407$*** |

***, shows that the variables are significant at 1%.

#### 4.3.2.4 Error Correction Mechanism (ECM):

The estimated ECM is for Model-2

$$\Delta TKM_t = \sum_{j=1}^{p} \pi_{1j} \Delta TKM_{t-j} + \sum_{j=0}^{q} \pi_{2j} \Delta LnFR_{t-j} + \sum_{j=0}^{q} \pi_{3j} \Delta LnGDPMS_{t-j} +$$

$\sum_{j=0}^{q} \pi_{4j} \Delta LnDDOP_{t-j} + \psi ECM_{t-j} + \nu_t$

Table 4.8: Error Correction Mechanism

| Variables | Coefficients | Std. Error | t-stat | p-value |
|---|---|---|---|---|
| $C$ | 151.6836 | 510.6079 | 0.297065* | 0.0696 |
| $\Delta TKM_{t-1}$ | -0.702592 | 0.271838 | 2.584600** | 0.0182 |
| $\Delta TKM_{t-2}$ | -0.483542 | 0.231934 | -2.084825** | 0.0508 |
| $\Delta LnFR$ | -2310.233 | 2090.493 | -1.105114** | 0.0429 |
| $\Delta LnFR_{t-1}$ | -405.7379 | 2191.262 | -0.185162** | 0.0351 |
| $\Delta LnGDPMS$ | 650.4238 | 7917.006 | 0.082155** | 0.0254 |
| $\Delta LDDOP$ | -2026.372 | 3585.515 | -0.565155* | 0.0786 |
| $\Delta LnDDOP_{t-2}$ | -2780.163 | 2822.237 | -0.985092 | 0.1370 |
| $ECM_{t-1}$ | -0.952374 | 0.392859 | -3.366023*** | 0.0032 |
| $R^2$ | | | | 0.8462 |
| $Adj-R^2$ | | | | 0.7551 |
| $DW$ | | | | 2.1245 |
| *Diagnostic Tests* | | | | |
| *Serial Correlation LM Test* | | | | 0.1903[0.8285] |
| *ARCH Test* | | | | 0.0008[0.9764] |
| *Jarque – Bera Test* | | | | 3.2412[0.1978] |

***,**,* shows that the variables are significant at 1%, 5%, and 10% respectively.

p-values of diagnostic tests are shown in square brackets. Breusch Godfrey LM Serial Correlation Test, ARCH Test are decided on the basis of F-statistics and Jarque -Bera Normality Test on the basis of Chi-Square test.

We have observed in tables 4.7 and 4.8 how rail freight demand responds to the other three variables lagged changes and long run. The coefficient of the freight rate is significant both in lagged changes as well as in the long run but it negatively effect rail freight demand. Our result also supports the empirical evidences of Wardman (1997), Lewis (1982), Levine (1981), Mitchell (2010) and Wijeweera (2014). All studies show that rail freight demand is elastic to prices. In the case of Pakistan railways it may be due to increased rates every year it declined the rail freight demand.

According to Ramanthan (2001), rail freight demand is highly correlated to industrial growth. The coefficient of share of manufacturing sector to GDP resulted as an important factor and provided direct evidence with rail freight demand in both dynamics. Our results supports the evidences of Shen (2009), Kulsherestha (2001), Ramanthan (2001). The possible justification of positive effect of manufacturing sector is that spatial distribution of industrial production is important for freight transport demand and for longer distances rail freight transport demand increases. Because Pakistan railways freight services mainly concerned with two major sea ports Keamari and Bin Qasim, and dry ports as well. The major commodities handling Chemical for Rayan Manufacturer, Petroleum Oil and Lubricants, wheat, coal, fertilizer, rock, phosphate, cement, sugar, oil-seed, wool, salt, containers and goods for transit to Afghanistan That's why it can be argued that well established manufacturing sector can increase the rail freight demand. So the results of the coefficient of GDPMS are consistent with economic relationship that with increase in manufacturing sector can increase rail freight demand.

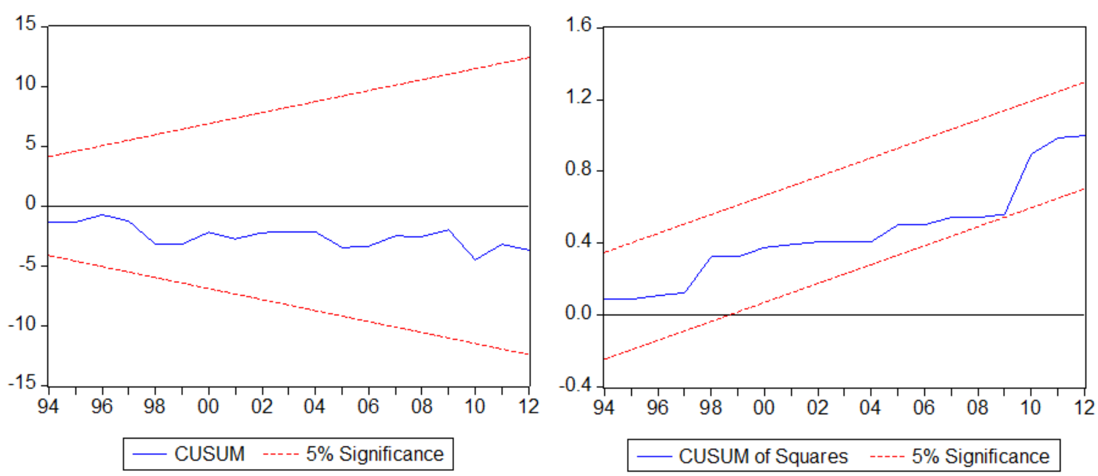
Domestic diesel oil prices (DDOP) variable showed a negative and significant influence on the dependent variable at all lagged changes and levels. So domestic oil prices and freight showed inverse relationship with rail freight demand growth. It can be due to when oil prices increases then it have immediate effect on on all the sectors of the economy. In the previous years oil prices remained high as a result it increases the costs of Pakistan railways which leads to increase in freight rates, ultimately rail freight demand declined. Other than it before 90's diesel oil prices have small variations but from the start of 90's increased slowly and about half of this decade rapidly which resulted in decreasing rail freight demand

The ECM shows the speed of adjustment to restoring the equilibrium in the short run. The lag of ECM indicates that how quickly or slowly the variables return to the equilibrium. The significance of lagged equilibrium correction term with negative sign confirms the convergence of long run equilibrium and justify Banerjee (1998) condition for the existence of stable long run relationship between the variables. In the estimation of ECM we found it (-0.952374) and significant and most of the short run coefficients are highly significant. The results of diagnostic tests of residuals showed significant i.e there is no autocorrelation, normally distributed and have no heteroscedasticity. Furthermore the stability tests indicates that parameters are stable. The value of $R^2$and $Adjusted_R^2$provided the desired results.

#### 4.3.2.5 Stability Tests:

In figure 4.3.4 CUSUM and CUSUM squares do not illustrate significant evidence of parameters instability that means regression coefficients are stable over the sample period.

Figure 4.3.4: CUSUM and CUSUM Squares Tests



## 4.4 Conclusions:

Objective of the study was to research macroeconomic determinants that are supposed to affect rail freight demand positively or negatively. Different studies have been conducted on rail freight segment but our research is to find out active determinants that can revive freight transportation and to suggest policies. For empirical investigation of determinants of rail freight demand we have used annual time series data over the period of 1981-2012. Statistics shows that freight services remained at the top in boosting the economy of Pakistan Railways. In order to ascertain a reliable link between macroeconomic activities and freight rail transport Autoregressive Distributed Lag (ARDL) model to cointigration has been used twice. The long run behavior of rail freight demand is expressed in terms of Tonne Kilometers (TKM). Behaviour of the series analyzed through Augmented Dickey Fuller (ADF) unit root test. Further Vector Autoregressive Model used to select the lag length of short run dynamics based on Akaike Information Criteria. After verification of data characteristics and its appropriateness, we have used ARDL. The major conclusions evidenced from the study are: GDP and share of manufacturing sector to GDP likely accounts for increasing freight rail demand; freight rate and domestic diesel oil prices may be reviewed so that to revive freight rail demand.

After the analysis, we further proceeded to enquire about econometric issues. Autocorrelation, heteroscedasticity and non-normality tests were performed to verify that the results estimated from ARDL model are not misleading and constancy of the parameters was also tested by using stability tests.

## 4.5 Policy Implications:

Our findings suggest that government may take steps to increase GDP and especially manufacturing sector. This can acheived through promoting agriculture sector and exports as these sectors highly influence the manufacturing sector to GDP. A dire need in reviewing policies to promote rail freight demand is suggested. Electricity and hybrid freight transport may combat the negative impacts of oil prices and fright rate on its demand.

## 4.6 Limitation of the study:

Transit time and service quality can attract demand for rail freight in the sense that if the goods transaction is within the time and services are better than other modes of transport. But due to unavailability of data of transit time, service quality in case of Pakistan, these variables are not included in this study but the previous literature guides that these have significant impact on rail freight demand growth.

## 4.7 Appendix:

### 4.7.1 ARDL Model-1:

Table 4.9: ARDL Model ( 2,2,2,2)

| *Variables* | *Coefficient* | *Std.Error* | *t − statistic* |
|---|---|---|---|
| $C$ | -54318.45 | 21414.58 | -2.536517** |
| $\Delta TKM_{t-1}$ | -1.239504 | 0.250096 | 4.956108*** |
| $\Delta TKM_{t-2}$ | -0.706937 | 0.296662 | 2.382968** |
| $\Delta LnFR$ | -6040.974 | 2328.799 | -2.594030** |
| $\Delta LnFR_{t-1}$ | 11076.71 | 3241.217 | 3.417453*** |
| $\Delta LnFR_{t-2}$ | 5003.255 | 2451.856 | 2.040600* |
| $\Delta LnGDP$ | 18422.47 | 5684.882 | 3.240608*** |
| $\Delta LnDDOP$ | -7764.646 | 3573.803 | -2.172656** |
| $TKM_{t-1}$ | -1.935149 | 0.384746 | -5.029681*** |
| $LnFR_{t-1}$ | -18961.41 | 5020.625 | -3.776704*** |
| $LnGDP_{t-1}$ | 10511.09 | 3858.439 | 2.724181*** |
| $LnDDOP_{t-1}$ | -5796.571 | 2344.894 | -2.471998** |
| $R^2$ | | | 0.857652 |
| $Adj\_R^2$ | | | 0.800838 |
| $DW$ | | | 2.136411 |
| *Diagnostic Tests* | | | |
| *SerialCorrelation LM Test* | | | 1.225162[0.3215] |
| *ARCH Test* | | | 1.667850[0.2079] |
| *Jarque − Bera − Test* | | | 1.030706[0.5973] |

The results of table 5.1 ***,**,* shows that the variables are significant at 1%, 5%, and 10% respectively.

p-values of diagnostic tests are shown in square brackets. Breusch Godfrey LM Serial Correlation Test, ARCH LM test have been concluded on the base of F-statistics and Jarque -Bera normality test on the base of Chi-Square test. The following table shows no auto-correlation, no heteroscedasticity, and residuals are normaly distributed. $R^2$ and $Adj\_R^2$

showing model is good fit. All the coefficients of short and long run are significant and supports the literature. So on the base of this model we have estimated ECM and long run estimates.HGF54

### 4.7.2 ARDL Model-II:

Table 4.10: ARDL Model (2,2,2,2)

| *Variables* | *Coefficient* | *Std.Error* | *t-stat* |
|---|---|---|---|
| $C$ | -27934.04 | 10631.27 | -2.627536*** |
| $\Delta TKM_{t-1}$ | -1.009281 | 0.218726 | 4.614357*** |
| $\Delta TKM_{t-2}$ | -0.739475 | 0.312137 | 2.369073** |
| $\Delta LnFR$ | -4632.728 | 2140.031 | -2.164795** |
| $\Delta LnFR_{t-1}$ | -9635.071 | 2613.938 | 3.686037*** |
| $\Delta LnGDPMS$ | 0.0018 | 6675.657 | 1.859946* |
| $\Delta LnDDOP$ | -12416.69 | 4306.530 | -2.883223*** |
| $\Delta LnDDOP_{t-2}$ | -6338.470 | 2709.456 | -2.339389** |
| $TKM_{t-1}$ | -1.819234 | 0.309608 | -5.875929*** |
| $LnFR_{t-1}$ | -14137.42 | 3145.271 | -4.494819*** |
| $LnGDPMS_{t-1}$ | 7198.012 | 2320.221 | 3.102296*** |
| $LnDDOP_{t-1}$ | -4091.757 | 1572.400 | -2.602236** |
| $R^2$ | | | 0.968381 |
| $Adj\_R^2$ | | | 0.818510 |
| $DW$ | | | 2.033355 |
| *Diagnostic Tests* | | | |
| *Serial Correlation LM Test* | | | 1.1936[0.3304] |
| *ARCH Test* | | | 0.6675[0.4214] |
| *Jarque – Bera Test* | | | 1.6876[0.4300] |

***,** and * shows that the variables are significant at 1%, 5%, and 10% respectively.

p-values of diagnostic tests are shown in square brackets. Breusch Godfrey LM Serial Correlation Test, ARCH Test have been concluded on the bases of F-statistics and normality test on the base of $\chi^2$. ARDL (2,2,2,2) shows we have estimated it with lag length of two.

# Bibliography


[1] Abdelwahab, W. and Sargious, M. (1992): “Modelling the demand for freight transport: a new approach”, *Journal of Transport Economics and Policy*, 26(1): 49-70.

[2] Asssociation of American Railroads. (2007):”National Rail Freight Infrastructure Capacity and Investment Study”, *Cambridge Systematics.*

[3] Banerjee, A., Dolado, J. and Mestre, R. (1998): “Error correction mechanism tests for cointegration in a single equation framework”, *Journal of time series analysis,* 19(3): 267-283.

[4] Beuthe, M. and Bouffioux, C. (2008): “Analysing qualitative attributes of freight transport from stated orders of preference experiment”, *Journal of Transport Economics and Policy*, 42(1): 105-128.

[5] Bresson, G., Dargay, J., Madre, J. L. and Pirotte, A. (2003): “The main determinants of the demand for public transport: a comparative analysis of England and France using shrinkage estimators”, *Transportation Research part A: policy and practice,* 37(7): 605-627.

[6] Boldea, O., and Hall, A. R. (2011):” Testing Structural Stability in Macroeconometric Models”, *Edward Elgar*.

[7] Chapin, A. and Schmidt, S. (1999): “Do mergers improve efficiency? Evidence from deregulated rail freight”, *Journal of Transport Economics and Policy*, 33(2): 147-162.

[8] Davidson, R. and MacKinnon, J. G. (2008): “Econometric Theory and Methods”, OXFORD University Press.

[9] Enderse, W. (2010): “Applied Econometric Time Series”, WILEY.

[10] Friedlaender, A. F. (1992): “Coal rates and revenue adequacy in a quasi-regulated rail industry”, *The RAND Journal of Economics*, 23(3): 376-394.

[11] Gujrati, D. M. and Porter, D. C. (2010): “Basic Econometrics”, *McGraw- Hill International Edition.*

[12] Greene, W. H. (2011): ”Econometric Analysis”, *7th edition, Prentice Hall.*

[13] Growitsch, C. and Wetzel, H. (2009): “Testing for economies of scope in European railways: an efficiency analysis”, *Journal of Transport Economics and Policy*, 43(1): 1-24.

[14] Inaba, F. S. and Wallace, N. E. (1989): “Spatial price competition and the demand for freight transportation”, *The Review of Economics and Statistics,* 71(4): 614-625.

[15] International Transport Forum (2014), “Peer review of railway freight development in Mexico”.

[16] Johnston, J. and Dinardo J. (1997): “Econometric Methods”, *4th edition, McGraw Hill, 185-295.*

[17] Kellner M., Boyson N. and Friedler N. (2011): “How to park freight trains on rail-rail transshipment yards: the train location problem ”, *Springer Verlag*, 34: 535-561.

[18] Kirchgassner, G., Wolters, J. and Hassalar, U. (2007): “Introduction to Modern Time Series Analysis”, *Springer*.

[19] Kulshereshtha, M., Nag, B. and Kulsherestha. M. (2001): "A multivariate cointigrating vector auto regressive model of freight transport demand: evidence from Indian railways", *Transportation Research Part A: Policy and Practice*, 35(1): 29-45.

[20] Levin, R. C. (1981): "Railroad rates, Profitability, and Welfare under Deregulation", *The Bell Journal of Economics, RAND Corporations,* 12(1): 1-26.

[21] Lewis, K. A., Widup, D. P. and Lewis, R. A. (1982): "Deregulation and rail-truck competition: evidence from a translog transport demand model for assembled automobiles", *Journal of Transport Economics and Policy*, 16(2): 139-149.

[22] Finance Division, "Current Expenditure and Development Expenditure Ministry of Railways 2000-2012", *Manager Printing Corporation of Pakistan Press Islamabad.*

[23] Ministry of Railways, "Year Books 1971-2012", *The Pakistan Railways Press Mughalpura Lahore.*

[24] Mitchell, D. (2010): "Australian intercapital freight demand: An econometric analysis", *Australasian Transport Research Forum 2010 Proceedings*, 29: 1-19.

[25] Oum, T. H. (1979): "Derived demand for freight transport and inter-modal competition in Canada", *Journal of Transport Economics and Policy*, 13(2): 149-168.

[26] Pesaran, M. H., Shin, Y. and Smith, R. J. (2001): "Bounds testing approaches to the analysis of level relationships", *Journal of applied econometrics,* 16(3): 289-326.

[27] Phillips, P. C. (1987): "Time series regression with a unit root". *Econometrica: Journal of the Econometric Society*, 55(2): 277-301.

[28] Rao, P. S. (1978): "Forecasting the demand for railway freight services", *Journal of Transport Economics and Policy*, 12(1): 7-26.

[29] Ramanthan, R. (2001): "The Long -run behaviour of transport performance in India: A cointegration approach", *Transportation Research Part A: Policy and Practice*, 35(4): 309-320.

[30] Said, S. E. and Dickey, D. A. (1984): "Testing for unit roots in autoregressive-moving average models of unknown order", *Biometrika*, 71(3): 599-607.

[31] Shen, S., Fowkes, T., Wightien, T. and Johanson, D. (2009): "Econometric modelling and forecasting of freight transport in Great Britain", *In proceedings of European transport conference*.

[32] Wardman, M., Toner, J. P. and Whelan, G. A. (1997): "Interactions between rail and car in the inter-urban leisure travel market in Great Britain", *Journal of Transport Economics and Policy*, 31(2): 163-181.

[33] Wang, Y., Ding, C., Liu, C. and Xie, B. (2013): "An analysis of Interstate freight mode choice between truck and rail: A case study of Maryland, United States", *Procedia-Social and Behavioral Sciences*, 96, 1239-1249.

[34] Wijeweera, A., To, H. and Charles, M. (2014): "An empirical analysis of Australian freight rail demand", *Economic Analysis and Policy,* 44(1): 21-29.

[35] Wijeweera, Charles, A., B. M. (2013): "Determinants of Passenger Rail Demand in Perth, Australia: A time Series Analysis", *Applied Econometrics and International Development*, 13(2): 221-234.

[36] Wilson, W. W. (1984): "Estimation of modal demand elasticities in grain transportation", *Western Journal of Agricultural Economics*, 9(2): 244-258.

[37] Winston, C. (1981): "A disaggregate model of the demand for intercity freight transportation", *Econometrica, Journal of the Econometric Society*, 49(4): 981-1006

[38] Woolridge, M. J. (2013): “Introductory Econometrics: A Modern Approach ”, *5th Edition, South Western, Cengage Learning, USA.*

[39] World Bank. (2011): “ Railway Reform: Toolkit for improving rail sector performance”, *Public Private Infrastructure Advisory Facility.*