<u>MPhil Dissertation</u>

An Empirical Analysis of Passenger Rail 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

Altaf Hussain

Reg.No. 19/MPhil-ETS/PIDE/2012

Supervised By:

Dr. Saqlain RAZA **Assistant Professor**

COMSATS Institute of Information Technology, Islamabad.

Pakistan Institute of Development Economics (PIDE), ISLAMABAD

IN THE NAME OFALLAH

The Most Beneficent, The Most Merciful

"To Allah belongs whatever is in the heavens and whatever is in the earth. Whether you show what is within yourselves or conceal it, Allah will bring you to account for it. Then He will forgive whom He wills and punish whom He wills, and Allah is over all things competent."

(Al-Baqarah, 2:284)

Authorship Statement

I, Altaf Hussain 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.

Altaf Hussain s/o Fazal-E-Huda

ACKNOWLEDGEMENTS

With the blessings of Almighty Allah I have completed this thesis. I am thankful to my supervisor, Dr. Saqlain Raza for his intellectual advice, guidance, and encouragement. The regular discussions from proposal writing till the final composition were very valuable and inspiring at every step of this research.

I also want to express my gratitude to Dr. Hafsa Hina, Dr. Anwer Hussain and Dr. Ehsan-ul-Haq Satti for their valuable comments and suggestion.

I am thankful to my parents whose prayers make difficulties of life into easiness. They always pray for my success, may God live them long. My special gratitude goes to my brothers for their unconditional support.

I am thankful to all my friends Kashif Ali, Azaz Ali Ather Bukhari, Maqsood Aslam and especially Muhammad Afzal Khan who always support me with these kind suggestions and advise.

ALTAF HUSSAIN

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARDL	Auto Regressive Distributed Lag Model
ANN	Artificial Neural Network
LR	Long Run
PKM	Passenger Kilometer
AR	Autoregressive
GDP	Gross Domestic Product
MA	Moving Average
VAR	Vector Autoregressive
ADB	Asian Development Bank
NTRC	National Transport Research Center
OLS	Ordinary Least Square
MARTA	Metropolitan Atlanta Rapid Transit Authority
JJ	Johansen- Juselius
EG	Engle Granger
LM	Lagrange Multiplier
ADF	Augmented Dickey Fuller
DF	Dickey Fuller
RGDP	Real Gross Domestic Product
ECM	Error Correction Mechanism
ARIMA	Autoregressive Integrated Moving Average
AIC	Akaike Information Criterion

SBC	Schwarz Information Criterion
OECD	Organization For Economic Co-operation and Development
DEA	Data Envelopment Analysis
CEM	Composed Error Mode
SARIMA	Seasonal Autoregressive Integrated Moving Average Model

ABSTRACT

This study identifies the major determinants of passenger rail demand along with forecasting performance of Pakistan Railways. The data set covers annual observations over the period from 1980 to 2012. To examine the association between passenger rail demand and its determinants in both dynamics, the ARDL model has been used. We have compared forecasting performance of univariate ARIMA and multivariate ARDL. The passenger kilometers (PKM) is our variable of interest whereas the explanatory variables are total population, GDP per capita, domestic diesel oil prices and fare. The findings of this study indicate that the GDP per capita and population plays pivotal role for increasing passenger rail demand whereas fare and domestic diesel oil diminishes the demand. The results postulated that ARDL is more precise than ARIMA setting for forecast. The performance of the predicted models is assessed based on mean absolute error, mean absolute percentage error and root mean square error. On the basis of results government may adopt policies for enhancing GDP, and generating alternative resources for reducing fare and diesel oil prices.

Key Words: ARDL, ARIMA, Forecasting, Passenger rail demand, econometric analysis

Chapter 1: Introduction

Rail transport had made valuable contribution in progress of many countries by expanding markets and exports (Rowstow; 1960), and enhancing its demand due to less cost, comfort ability, portability and availability. In Pakistan, passenger demand grows with the increase in population, economic growth, urbanization and industrialization.

Regardless of the betterment in transportation sector it is still inadequate to meet the demand of ever growing population of Pakistan. The transportation sector contributes 10 percent of GDP but the railway has been contributing with very meager shares. Government is continuously ignoring or spending a very tiny part of the assigned budget on the research and institutional talent building in the development of railway (Irfan et.al, 2012).

From decade road traffic, mutually passenger and freight have increased more rapidly than the country's economic progress. It needs to know the demand and contribution of railway in the economic development of Pakistan and to accommodate the burden of growing population.

Our research is based on analyzing the efficiency of different factors like GDP per capita (Gross Domestic Product), fare, total population and domestic diesel oil prices with respect to passenger kilometer[1]. Our analysis of these factors would not only reveal the outcomes in the form of their long & short run contributions but also help to determine the comparisons of ARDL forecasting performance along with ARIMA model.

[1]Passenger carried by railway is the number of travelers transported by rail periods kilometers travel. It is used as to measure passenger rail demand.

Up to the 70s Pakistan Railway was the main way of transportation in the country, but deviation of even now inadequate resources toward the extension of road network, the recital and situation of Pakistan Railway has fallen and its countrywide traffic share condensed from 73% to 4% for freight and 41% to 10% for passenger traffic (Choudhary et.al. 2007).

Pakistan Railway network which have consisted of 312 Km of metre gauge, 7,791 route-kilometers and 7,479 Km of broad gauge. In this network there are 589 stations, 293 Km of electrified sections and 1,043 Km of double-track sections (in total). Since 1982 the railways of Pakistan have not developed any new routes, and have eliminated light traffic branch line since the 1999. Pakistan railways in 2006-07 run engines with 544 and 1916 passenger trainers (NTRC, JICA[2], & Ministry of Communications, 2006). Over the era from 1997 to 2006, commuter travel by this network improved by 40 percent from 18-26 billion passenger kilometers per year (Pakistan Bureau of Statistics, 2008). During the same period, kilometer travelled by a single railway commuter greater than previously 277 to over 310 kilometer per passenger (JICA et al., 2006).

The trend of railroad travelling is on the increase in numerous countries as it is more environments friendly and cheap in the long term (Union Internationale des Chemins de Fer, 2006). In India a fastest increasing economy in South-East Asia 20 percent of the commuters were carried through railway network in 1992 (Harral and Sondhi, 2005). Railway had therefore an extra developmental character in those countries than in the central economies, which already had relatively efficient and competitive market structures at the beginning of the railway period (Coatsworth 1981; Kuntz Ficker 1999; Summerhill 2000, 2003; Dobado and Marrero 2005).

[2] Japan International Cooperation Agency

1.1 Objectives of the Study

- To check the short run as well as long run relationship between Passenger Kilometer and its determinants that affect its demand.
- To check the forecasting performance of ARDL model along with ARIMA model

1.2 Significance of the Study

This study gives to the existing literature in a way that it determines the cointegrating relationship between passenger rail demand and its determinants. It also analyzed the comparison of forecasting performance of ARDL and ARIMA model. According to my observation there is hardly any study about rail freight demand did by Khan and Raza (2015), but passenger rail demand and forecasting performance have not been analyzed in the context of Pakistan.

1.3 Hypothesis of the Study

H_1: There exist short as well as long run relationship of factors affecting passenger rail demand.

H_2: The forecasting performance of ARDL model and ARIMA model are same.

1.4 Organization of the study

After the introduction, Chapter 2 exemplifies theoretical literature on passenger rail demand. Chapter 3 defines the process of econometrics which involves unit root test, Cointegration to test the long run relationship and error correction (EC) model, ARIMA model and the forecasting performance of ARDL and ARIMA models have been discussed. Chapter 4 provides the estimation and empirical outcomes. The last Chapter denotes the conclusion and policy recommendations.

CHAPTER 2 LITERATURE REVIEW

Our focused literature is time series studies on passenger rail demand, its determinants and forecasting performance using econometric techniques. According to my knowledge there is hardly any study available for passenger rail demand in Pakistan. There is need to examine the short run and long run relationship between passenger rail demand, its determinants, and also the forecasting performance of univariate ARIMA and multivariate ARDL models.

Literature Review Related to Co-integration

Doi and Allen (1986) examined the monthly ridership for the urban rail fast transit line by using time series analysis. They used dual time series regression representations, for one is logarithmic and the other is linear form, for the estimation of these models. The model was standardized for a time period of around six and a half years (from 1978-1984). Results of this approach provide important implications for budgeting, train operation, system changes, policymaking of individual programs in pricing etc.

Owen and Phillips (1987) studied the features of railway passenger demand. They used time series (1973-84) and the tools of econometrics, studied the effect on rail demand and a number of economic aspects and also measured their relative importance. The magnitude of background noise and the view of obvious boundaries inherent in the data, the results attained revealed unusually high degree of consistency and precision.

McGeehan (1993) studied the railway expenditures and production growth in the case of republic of Ireland. He used the translog cost function and time series data from 1973 to 1983. The study findings tend to provision that “remarkable triumph” and suggested that real productivity achievements have been greater than a standard partial production index. Results showed that the cost function is non-homothetic and non-homogeneous.

Bentzen (1994) studied the demand of gasoline in Denmark by cointegration (Engle and Granger 1987) methods. He estimated the long as well as the short run elasticities of demand for gasoline. He used the Engle and Granger cointegration and error correction model for estimation and used time series data from 1948-91.The experiential evidence looks to propose comparatively minor price elasticity even in the long run it is inflexible to believe that the properties of this strategy will be very powerful on future gasoline consumption. The results showed that error correction model have expected signs for all parameters along with predictable magnitudes.

Oum and Yu (1994) studied the productive efficiency of railways in the OECD countries. They used the Data Envelopment Analysis (DEA) method and Tobit regression model by using a panel of nineteen railways from 1978 to 1989. They attempted to classify the effects of government interference and subsidization on creative efficiency of those railways which develop a high amount of trade from passenger services. Results show that efficiency measures may not be compared across railways without monitoring for the effects of the dissimilarities in working and market backgrounds.

Gathon and Pestieau (1995) investigated the occasion of European railways decomposing ability into its decision-making and its managerial components. They used stochastic or composed error model (CEM), and used data period from1961-88 which includes 19 European countries. The study found that procedural efficiency is affected by the environment and amount of government intervention and can be nurtured by increasing the self-sufficiency of the firm or by soothing the institutional restrictions to which it is subjected. Current application points to opportunities of future research. Limits outcome mainly from the data that has been used and that fail to encompass the excellence of services as well as a number of ecological features.

Eltony and Mutairi (1995) studied gasoline demand in Kuwait by using Engle and Granger cointegration techniques using data from 1970 to 1989. The empirical sign seems to propose relatively small price elasticity in the long run; it is hard to trust the sound effects of this policy in Kuwait will be very significant on future gasoline consumption. The results showed in the short as well as in the long run the price elasticity of gasoline demand is inelastic, and in the short run the gasoline demand with respect to income is inelastic whereas in the long run it is elastic. This study proposed that the demand of gasoline is higher with respect to income changes in the long run than as compared to the short run.

Wardman (1997) investigate inter urban demand of rail, challenge and elasticities in Great Britain. He used the nonlinear least squares method. The study purpose was to spread direct demand models in Great Britain as usually applied to review in effects of competition. The study concluded that British rail has corrects its forecasting methods to contain competitive effects. British rail has corrected its suggested forecasting technique to allow the GT (Generalized Time) and elasticities of price on the flows of non-London to differ according to the condition of competitive as represented by the comparative generalized costs of car and rail and of coach and rail.

Bollinger and Ihlanfeldt (1997) studied the effects of fast rail travel on economic development in Atlanta's MARTA. They used simultaneous model and data from 1980-90. The purpose of the study was to report our discoveries that in the station areas of Atlanta's MARTA rail route had the effects on employment and population. Results showed that Atlanta's MARTA has no obvious influence on employment or total population, but it has improved the employment arrangement in these regions in favor of the public region.

Savage (1997) investigated gage economies in United States (US) transit rail schemes. He used a three staged least square translog cost function to estimate for 13 heavy rail (subway) and 9 light rail (tramway), and used data from 1985-1991. The study objective was to examine the economies of both scheme size and density. The discoveries of this study deliver input to the continuing discussion about the competitive constricting or privatization of town transit schemes. The study does not directly discourse the issue of economies of scope among lines, the constant returns to system magnitude suggest that there would be insignificant cost disadvantage to contravention up firms into lesser component parts.

Andrikopoulos and Loizides (1998) studied the productivity growth and price structure in European railway mechanism. They used the translog model and annual data from 1969 to 1993. The aim of this study was to investigate the railway system cost-structure of the ten European union-countries and, later, to provide measure of economic efficiency and recognize its sources. Findings of this study were not fully conclusive yet provided definite features that may support the policy makers to adjust policies that may beneficial to advance the effectiveness of the railway schemes.

Cantos et.al. (1999) investigated the efficiency, productivity and technical modification in the railways of Europe. They used non-parametric (DEA) method and data used from 1970 to 1995. The goal of the study was to examine the current development of productivity in the European railways by breaking down its progression into modifications in efficiency and technical change.

Results show that the output growth is focused in the last period (1985-95), when the majority of the companies agreed to progressions of reforms.

Ramanathan (1999) studied the long as well as the short-run elasticities of gasoline demand in India by using cointegration methods. He used the cointegration (Engle Granger) and error correction procedures, and the time edge of the study is from 1972-1973 to 1993-1994.The result of this study showed that gasoline demand found inelastic to gasoline price fluctuations, equally in the long as well as in the short run. The empirical results display that a low elasticity of price even though in the long run which means that over pricing of gasoline as a policy mechanism is not probable to be very powerful on forthcoming gasoline demand in India.

Ramanathan and Parikh (1999) studied the transport of India in the situation of supportable development. In the previous few decades they provided a brief review of the Indian transportation sector. It is shown that the era has viewed a gradual transformation from rail-dominated transportation to road-dominated transportation. They used the cointegrating econometric (Engle and Granger, 1987) model, and data from 1980 to 1994. They pointed out the need for substituting personalized road transport by public transportation methods and by rail transport, refining the effectiveness of transportation modes, promoting further vigorously the electric vehicles and CNG, air travel exchange with great speed of train travel. As the great investment requirements and the environmentally benign features, this thing has identified the combined execution in Indian Railways as a significant policy choice towards maintainable transportation in the development of India.

Ramanathan (2001) employed cointegration (Engle and Granger) and error correction model to observe the steady state behavior (long run) of transport action in India. He used the annual data from 1956 to 1988. His study objects to discover the existence of two cointegrating relations with TKM and PKM as the response variables. The results depicted that in India passenger/kilometers (PKM) increases as gross domestic product (GDP) increases but the increase in (PKM) is faster than GDP and much faster than the increase in urbanization. The tonne-kilometers are highly associated to the industrial development and are more probable to increase faster than the development in the index of industrial production. The freight and passenger routines are moderately inflexible to price fluctuations. The error correction model displays that mutually tonne-kilometers and passenger modify to their long-run equilibrium at a moderate rate.

Beko (2003) examined the experiential analysis of Slovenia of the demand utilities for services of civic railway passenger transportation. He used the ordinary least square method (OLS) and monthly time series data from 1993 to 2002. In the study obtainable estimates of elasticity and the decisions derived from them, offer beneficial suggestions for setting the broad price policy for civic railway passenger transportation in Slovenia. Theoretical rise in average real prices leads to a percentage drop in the number of travelers travelling by rail that is lesser than the percentage rise in prices. In the short run, the predictable price elasticities imply that there is possibility of improving revenues of the railway operator by growing average real prices.

Haines and Margo (2006) investigated the local economic development and railroads of the United States. They used the data from 1850 to 1860. In the previous studies that rail access found to be positively correlated with rate of agricultural land at a fact in time, and have taken this correlation as indication that rail access chiefly profited agricultural land possessors in the way predicted by the Hekscher-Ohlin or Von Theunen models. In either model most of the estimated effects are small and the signs are not wholly consistent, and that the agriculture was the chief beneficiary of rail access under null hypothesis.

Lan and Lin (2006) studied the railways performance that yield freight and passenger services by dual stochastic distance function methodologies. This is stochastic consumption distance function (SCDF) model and stochastic input distance function (SIDF) model. They used 312 panel data composed of 39 railways over the era of 1995-2002. Preceding studies may have used input-oriented contrast (measuring the comparative inputs under the identical output level) or output-oriented contrast (measuring the comparative outputs under the identical input level) in measuring technical efficiency. It would be interesting to survey differences in technical efficiency and service usefulness resulting from these institutional and rearrangement changes in the rail region in a future study. Result findings indicate that technical railways' ineffectiveness and service inefficiency are negatively swayed by fraction of electrified lines, gross national income per capita and line compactness. Over-all, the railways in West Europe execute more efficiently and excellently than those in East Europe and Non-European areas.

Chen (2007) studied the rail transport demand by using dynamic econometric (Engle and Granger 1987) methods. He used the annual data of 1995-02. He applied numerous dynamic models and described rail transport because in the rail transport market different lags occur. For analyzing and forecasting demand these dynamic models provided powerful tools, but these models have some drawbacks that should be known. In non-stationary data problems of primary examination include fixed effects assumption, multicollinearity, and a danger of getting spurious regression results. Additional research is predictable to resolve all these difficult, testing non-stationarity including more explanatory variables, starting with removing outliners, and finally emerging a new model.

Akinboade et al. (2008) studied the gasoline demand in South Africa. They used the autoregressive distributed lag (ARDL) model and data used from 1978 to 2005. They analyzed that there exist long run relationships. The goals of this study was to test and construct an econometric model to classify the principal economic fundamentals that impact the performance of the gasoline motor consumption in South Africa. The study resulted increases in price will not deject the consumption of gasoline and that income increase will induce only small increases in the demand of gasoline. These outcomes could assist policy creators in dividing the distributional influence of their energy strategy in South Africa.

Odgers and Schijndel (2011) predicted yearly boardings train in Melbourne by using annual data. They used a univariate and multivariate linear regression by taking six independent demographic and economic variables of the period 1983-84 to 2009-10. From these variables two variables were lagged three months, and two variables were lagged by six months. Results indicate that both kinds of forecasts show an increase in train patronage from 2010-11 to 2012-13 that is twice the annual average rate of growing skilled in the preceding three years. Future research strategies and limitations of the study were also presented by the authors.

Tolulope and Taiwo (2013) studied economic growth and rail transport in Nigeria by taking data for the era 1970-2011 and using Johansen-Juselius and error correction (EC) methodology. The main goal of this study was to introduce the policy to convert the railway scheme in Nigeria from its current condition to the flexible, efficient and competitive way of transport so it could contribute in the country's development. They found long run association between the variables and the error correction term found to be significant in the EC (error correction) models. The regression results displayed inverse relationship between economic growth and rail transport.

They suggested that government should introduce policies by emphasizing on the development of subsectors of economy.

Wijeweera and Charles (2013) empirically examined the elements of passenger rail demand in Melbourne (Australia). They used the cointegration (EG) methodology to estimate the elasticities of the long-run, whereas an ECM is used to approximation the elasticities of short run with the help of annual data 1979-2008. They put a pilot effort to cultivate a time-series method that will be efficient for the analysis of passenger demand function of Australian urban rail travel. They found that the elasticity of demand in short-run is twice lower than the long-run. Inelasticity of the demand suggests that an increase in fare would not cause an important drip in boardings and hereafter results in an increase in total profits. Petrol price, city population, passenger's income and fare exert an affirmative impact on passenger rail demand.

Uma et al (2014) studied the impact of transportation network on economic development by using annual data for the era 1981 to 2009 in Nigeria. They used the Johansen rank test of cointegration to investigate the long run association through ordinary least squares method. The key objective was to study the effects of air, road, railways and water transport on economic development (measured by real gross domestic product) in Nigeria. The result shows that only road transport contributed significantly and positively on the real gross domestic product (RGDP). They recommended the sufficient allocation of resources for transportation to enhance the economic development.

Khan and Raza (2015) empirically analyzed the freight rail demand in Pakistan. They used ARDL model, and data used from 1981 to 2012. The result of this study showed that GDP and share of manufacturing sector to GDP has positive influence on rail freight demand in the short and long-run. The study also finds negative impact of freight rate and domestic diesel oil prices in both the dynamics (unpublished).

Shahbaz et al. (2015) investigated the causal relationship of macroeconomic variables and road transport carbon dioxide (CO_2) emissions in Tunisia. They used combined cointegration tests, which one is the autoregressive distributed lag model and used the time-series data from 1980 to 2012. The outcomes reveal that the presence of cointegration. The study arrange for new awareness to the policy producers for designing a broad energy, environment policies and transport for maintainable economic development in long run.

Literature Review Related to Forecasting

Rose (1986) examined the effects of service and price changes on travel ridership, and also investigated the extra dimension of short and long term elasticities by using time series ARIMA model, and used monthly time series data. In this study the effects of the prices of gas and rail services were significant, and the effects of transport fares were insignificant, which shows the elasticities of fare are zero in both of the short and long-term. Overall the results showed that ridership is not affected by change of fare rates and that changes of services do not affect ridership until twelve months after the change is executed.

Smith et.al (2002) studied the traffic flow forecasting which compare the parametric and nonparametric models. They used seasonal ARIMA model, and for that purpose they used data from 4 September to 30 November, 1996. In this study they examined the theoretical foundation of regression (nonparametric) and answer the problem of whether regression (nonparametric) is built on heuristically better forecast generation approaches method that the single interval traffic flow forecast performance of seasonal autoregressive integrated moving average models. Past studies indicates that the seasonal ARIMA models are statistically superior to basic applications of nonparametric regression. Furthermore, due to the overview of heuristic prediction generation methods this exemplifies that there are other occasions to further improvement recital of nonparametric regressions models.

Tsai et.al (2003) investigated the forecasting of short-term railway passenger demand by using artificial neural networks method, and used the daily data from 1999 to 2000. In this method the railway passenger demand forecasting was efficiently under numerous situations of train services. The models result can proposal comprehensive demand forecast for railway action scheduling, such as train planning and seat apportionments.

Whelan and Johnson (2004) investigated the modeling of the influence of alternative fare constructions on train congestion. They settled the PRAISE (Privatized Rail Services) rail operations model to comprise penalties for congestion based upon journey determination, degree of overcrowding and journey time. The aim of the study is to report on the growth of a simulation model to display the impact of crowding on rail demand and to exam alternative ticketing plans to deal with the difficulties of overcrowding. The strategies tested in this study established that it was easier to passenger prices out of the highest era rather than lure them away by dropping the rate off-peak and that unbiased revenue solutions can be establish using a mixture of fare rises in the peak and in the off-peak fare decreases. These policies are however restricted by the existing level of fares rule.

Butkevicius et al. (2004) estimated forecast of the dynamics of passenger transportation by the civic land transport. They used the multiple regression model, and time series data from 1996 to 2001. The paper was aimed to recognize major causes of transportation reduction common to railway and road transport (*i.e.* substantial fast growth of the number of personal motor cars, growth of tariffs, lack of comfort, low prestige of public transport, wear of transport facilities etc). The 2010 forecast showed that passenger haulage by public transport will reach 575m, whereas by the 2015 it will increase up to 893m, this means that transportation will rise by 1.6 times and 2.3 times, correspondingly, equated to passenger haulage in 2001. But, passenger transportation in 2015 will not attain the volume found in 2001.

Profillidis and Botzoris (2006) investigated the use of econometric models for the prediction of passenger demand in Greece. They used U-Theil Statistics for the three models for the forecast of passenger demand in Greece, and used the data from 1980-2000. It is necessary in transport planning to found a causal relationship among the demand of each transport mode and the parameters moving demand. The necessary suitable tests assured the validity of projected models, which can then be used for the prediction of future demand. Once checked the forecasting skill, the models can be used for the forecast of forthcoming demand and model separated for passenger demand in Greece.

Hamzacebi et al. (2009) studied the relationship of the iterative and the direct artificial neural network (ANN) forecast methods in the multi periodic time-series prediction. They forecast by using iterative and direct methods, and compared these results with other method (ARIMA and adaptive estimation techniques) which gives better result. As the result indicate that the ANN forecasting is better results as compare to the other methods.

Tsai et.al (2009) investigated the short term railway commuter demand predicting by the neural network which is based on temporal feature models. They used the data from 1999 to 2002. The objective of the study was to recover the predictive recital of neural networks (NN) in the request of short term railway commuter demand prediction. The study research query is whether intricate design of neural networks (NN) configurations in terms of railway data structures can raise prognostic performance in contrast with conventional multi-layer perceptron. The study results display that MTUNN (multiple temporal units neural network) and PENN (parallel ensemble neural network) overtake conservative MLP (multi-layer perceptron).

Celebi et al. (2009) investigated the prediction of light rail commuter demand by using artificial neural networks (ANN). They used multi-layer perception which is ideal due to verified success of resolving approximation problems, and for that purpose two year daily data used. The study results indicate that the neural network model predictions are somewhat more precise than the best autoregressive integrated moving average model. When model established before observations, the terms of error are probable to increase for an upcoming data set. For the applied purpose use, determining on the best autoregressive integrated moving average model for the upcoming forecast might be an interesting task itself.

Chiang et al. (2011) studied the prediction of ridership of the metropolitan transportation authority. Now a day the variation in gasoline price and the economic growth increase and decrease have made the organization of public transportation schemes particularly challenging. Precise forecasting of the ridership is necessary for the development and progression of transit services. They used for the precise forecasting models are autoregressive error correction, neural networks and ARIMA models, and data used from 2002 to 2004. Results show that the combination of three models forecasting is more accurate as compared to individually use. At the last, a scenario investigation is conducted to evaluate the influence of transit strategies on long-term ridership.

Dailydka et al. (2012) investigated the molding of selection of passenger trains on a track. The study aimed to describe the components of revenue and expenditure gained from the passenger shipping by local railway paths and to identify dependency thereof on stipulations of rolling-stock. The key factors affecting efficiency of passenger shipping by railways is minor increase in number of passengers which is subjected to the quantity of gained income. One of main causes to be considered seeking to shrinkage losses of passenger transportation by railway is lesser number of passengers which concludes the amount of gained revenue. Therefore all probable actions must be booked to increase the quantity of passengers travelling by railways, *i.e.* index of tenancy of passenger seats.

Cyprich et al. (2013) studied the predication of short term passenger demand by using the theory time-series analysis. The aim of the study was to analyse and identify the recital of forecasting model of commuter demand for the residential bus transport time series analysis, which fulfills the statistical importance of its parameters and randomness of its residuals. They used ARIMA model, and for that model monthly data 1/2000-12/2007 was used. The findings of this study will have progressive impact on increasing the predicting precision in the method of commuter demand prediction.

Dou et al. (2013) studied a train forwarding model based on fuzzy traveler demand forecasting for the period of holidays. They established the passenger change rate of fuzzy logical association based on time series theory and fuzzy theory, and data used on 2011. This study applied portfolio optimization, train operation adjustment theory and fuzzy set theory. Results showed that fuzzy passenger prediction is more precise than autoregressive integrated moving average model. With the rise in passenger flow during holidays, train forwarding is influenced by further factors and there are more restrictions to consider. Additional discussion is needed for model and further detailed examination of the railway line is still necessity in the case.

Wijeweera and Charles (2013) studied the contributing factor of the demand of passenger rail in Perth (Australia) by time-series analysis. They used the cointegration (Engle and Granger) methodology to evaluate the elasticities of the long-run, while an EC model is used to evaluate the elasticities of short-run with help of annual data 1983-2008. The results of Perth propose that Australian rail passengers are usually more reactive to price changes. Given that other studies propose that the perception of the transportation mode is a critical factor in determining the demand for any specific transport mode, it is hoped that improved perception variable will be obtainable for use in future studies.

Shitan et al. (2014) investigated the appropriate model for the Ampang Line light rail transit commuter ridership and to predict the commuter ridership in Malaysia. They used the SARIMA (Seasonal autoregressive integrated moving average) (2, 1, 0) $(0, 1, 0)_{12}$ model, and data used from January 2003 to December 2011. The prediction based on this model would be beneficial for the experts to design ahead and empower them to create policy conclusions. Hence, to rise the ridership of the LRT (light rail transit) possibly the number of trainers and their rate of recurrence could have to be improved. There could too be several other factors which could affect the ridership of light rail transit and these could be studied in forthcoming research.

Chapter 3 Data and Methodology

The goal of our study is to examine the determinants of passenger rail demand for Pakistan. A lot of literature available on passenger rail demand but we have not found a significant literature for Pakistan which empirically examined the factors affecting on passenger rail demand both in short as well as in long run and also the forecasting performance. Important determinants we have included in our analysis are GDP per capita, domestic diesel oil prices, fare and population. For the short run[3] and long-run[4] analysis we used ARDL approach and estimated univariate ARIMA and multivariate ARDL models are used for the forecasting performance. In this chapter we shall discuss economic relationship among variables, availability of data, sources of data and steps to examine the short and long-run association and forecasting performance.

3.1 Variables and Data Sources

The data used for this study consists of annual observations cover time period from 1980 to 2012 because data from onward period is not available. The passenger kilometers (PKM) is used as the response variable and the explanatory variables are population, GDP per capita, domestic diesel oil prices and fare.

3.1.1 Passenger Kilometers (PKM)

Passenger kilometers means the passenger travelled through railway by rail time in kilometers travelled. It only considers the distance on the national territory of country to account for national, international and transit passenger transport. A passenger kilometer is used as a unit of

[3] : At least one factor of production (labor, land, capital goods and entrepreneurship) is fixed

[4] : All factor of production (labor, land, capital goods and entrepreneurship) can be changed.

measurement of passenger rail demand which can be obtained by multiplying number of Total Passenger Travelled (TPT) and Total Kilometer Travelled (TKT) in a given year.

$$PKM = TPT * TKT$$

Where $TPT = Total \;\; Passenger \;\; Travelled$, and $TKT = Total \;\; Kilometers \;\; Travelled$

The more PKM shows more passengers rail demand. Data have been obtained from World development indicators. Unit is million passenger kilometer.

3.1.2 GDP Per Capita

GDP per capita is gross domestic product which is divided by the midyear population. GDP is the total gross value concluded by all inhabitant producers in the economy plus any product taxes and minus any subsidies not comprised in the value of the products. It is calculated without deductions for depreciation of fabricated possessions and degradation of natural resources. Data are in current local currency. Data have been obtained from World Development Indicators. There are different point of views about the association among passenger rail demand and GDP. Some studies showed major influence on passenger rail demand but some studies have the opinion of lower. Economic growth is the major determinant of passenger rail demand in the long run and there is strong contemporaneous relationship between GDP and PKM (Ramanathan 2001). Over the whole sample era, GDP only rose by a little over 11% (*i.e.* it rose by almost just 1%, on average, per annum). Thus, however the estimated long term average (median) elasticity with respect to GDP is around 1.5, the real long term growth in GDP was too small to have any great influence on passenger rail demand (Owen and Phillips, 1987).

3.1.3 Population

The population which takes into account all inhabitants irrespective of citizenship or legal status except for refugees not permanently settled in the country of asylum, which are generally considered portion of the population of their country of origin. Data have been obtained from World Development Indicators. According to different studies it has significant impact on passenger rail demand. The values shown are midyear estimates. According to Ramanthan and Parikh (1999) and Wijeweera and Charles (2013) it has significant and positive influence on passenger rail demand over the long run.

3.1.4 FARE

It is defined as the money paid for a journey on public transport. It is considered as major factor affecting rail passengers demand because it is charged by Pakistani Railway. Some of the data were available in paisa and some in rupees, so for our convenience we have standardized it into rupees. Some studies defined elastic but some inelastic relationship between passenger rail demand and fare. It exerts significant and negative impact on passenger rail demand in long run (Wijeweera and Charles, 2013). Owen and Phillips (1987) empirically examined that the volume of intercity rail travel is significantly affected by fare. Beko (2003) concluded that the increase of average real fares the number of passenger by rail decreases in terms of percentage by less than the fare actually increased. The recorded price inelasticity of demand leads us to conclude that returns of the railway worker rise when the average actual fare increases. Data have been obtained from different year book of ministry of railways.

3.1.5 Domestic Diesel Oil Prices (DDOP)

The increase of oil prices has significant influence on transportation either in road, rail, ship or air (Transport Economics and Management Systems Inc., 2008). In 2009-10, Pakistan Railways from 2009 to 2010 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.

3.2 Methodology

Numerous methods have been used to study the demand for transportation in general and by mode. These studies include vector autoregressive model, optimization models, models of modal choice, adhoc specified and estimated demand functions, derived demand model, univariate autoregressive model and cointegration approach. Optimization models in the analysis of transport demand integrate 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 model choice behavioral function (McFadden). This method has been used in studies by Levin, Miklius et al., and Johnson (1976). Oum developed the theoretical assumptions under the use of linear logit models for transport demand studies (Spring 1979) but these models are not suitable in case of transport demand studies because it imposes numerous rigid, a priori restrictions on estimated parameters and a structure of technology.

A third type of transportation demand analysis is specification of behavioral equations using adhoc models. These models are useful for forecasting but suffer in the analysis of price awareness of demand. The functional form of the model, and the suitable set of exogenous variables, is both somewhat arbitrary. The estimated Coefficients 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. Oum applied same procedures using time series data (1978) and cross-sectional data (Autumn, 1979) and duality approach to analyzing intermodal competition in transportation is attractive because its functional specification is consistent with neoclassical economic relationships.

There are several empirical problems in connection with traditional approaches of passenger rail demand specification *e.g.* non-stationarity of the variables and endogeneity associated with regressors. By another token, hypotheses testing OLS estimation technique considers the presence of stationarity associated with time series observations. Certain difficulties arise when we are estimating integrated or non-stationary variables under single equation model due to: error process not being stationary, independent variables generated by processes that depict serial correlation, non-standard distributions of coefficient estimates, existence of more than one cointegrating vector and failure of weak exogeneity (Banerjee et al., 1986, 1993). Therefore to investigate short run and long run association between integrated variables, cointegration test and estimation of ECM are available.

In the passenger rail demand literature, previous studies focused on passenger rail demand modeling, examining elasticities or modal choice based on either cross-section or time series data

(Surveys by Wardman 1997; Tolulope and Taiwo, 2013 and Wijeweera and Charles, 2013). Few studies of passenger rail demand forecasting have employed the recent developments in autoregressive integrated moving average model such as Rose, 1986; Chiang et.al, 2011; Cyprich et.al (2013) and Shitan et.al, 2014 respectively. Ramanathan (2001) applied Johansen and Juselius cointegration approach and Error Correction Mechanism in modeling and forecasting both for passenger and freight transportation demand in India. But in our study we have used ARDL Model for analyzing the possible association between passenger rail demand and its determinants in short-run and long-run which can tackle the problem of endogeneity of regressors, suitable for finite sample data and univariate ARIMA and multivariate ARDL model for forecasting performance of models.

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 passenger rail demand which is represented by passenger kilometers (PKM). Explanatory variables that possibly can affect passenger rail demand growth are fare, GDP per capita, total population and domestic diesel oil prices. The data used for the period of 1980-2012. The functional form of economic relationship can be defined as:

$$PKM_t = f(GDP_t, POP_t, FARE_t, DDOP_t)$$

From the functional relationship econometric model can be developed in the following way:

$$PKM_t = \beta_0 + \beta_1 GDP_t + \beta_2 POP_t + \beta_3 FARE_t + \beta_4 DDOP_t + v_t \qquad (3)$$

Cointegration is a comparatively modern estimation technique which is applied for the estimation of relationship of two or more than two variables in long run. Hence In mid 1980s such procedure was formalizing in time series analysis, and has been observed as the greatest significant development in current experiential modeling.

3.2.1 Augmented Dicky-Fuller (ADF) Test

Stationarity is the pre-requisite in econometric analysis of time-series data. A series will be considered white noise if and only if the mean is zero, the variance is constant and covariance is zero (Charemza and Deadman, 1992). If any of the white noise property doesn't holds then series is non-stationary. Such non stationarity is one of the contentious issues, while we are dealing with time series analysis.

ADF tests are commonly used to check the Stationarity level of the series. In case of non-stationary, series can be transformed into stationary by de-trending. A suitable way out of de-trending is by using first differences somewhat levels of the variables. Sometimes, the series is stationary by more than once. The non-stationary series which are transformed into stationary by differencing $'d'$ time is supposed to be integrated of order d as seen x_t is approximately $I_{(d)}$, (Engle and Granger, 1987). If the series becomes stationary at level, it is known to be integrated of order zero *i.e* $I_{(0)}$. Before proceeding with Cointegration analysis, there is need to check the possibility of unit-root in the series whether the given series is stationary or non-stationary. If the series is integrated at order 0, meaning the series is stationary at level then there is no need to do Cointegration analysis but if series are non–stationary then we can apply different Cointegration techniques. While we are dealing annual series then Augmented Dicky Fuller (ADF) test is the first choice of econometricians to check the possibility of unit roots (Charemza and Deadman, 1992). Several tests are also available for testing unit root like Phillips-Perron (PP), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) but ADF test has been preferred in the literature over PP and KPSS tests because later are more sensitive to time series (Gujrati and Porter; 2005: p. 754-760). ADF is the generalized form of Dicky Fuller (DF) test. Therefore ADF test involves followings regression:

$$\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (3.2.1)$$

Whereas, the variables 'y' is to be checked for the stationarity, Δ the first difference operator, ε_t the error term and k are the order of lags to check the possibility of autocorrelation in ε_t. The null (H_0) and alternative (H_1) hypothesis can be shown as:

H_0 : *There exist unit root in the series*

H_1 : *The given series is stationary.*

For the rejection of null hypothesis, probability value must be less than 0.05 or statistically significant (*e.g.* Fuller 1976; Guilkey and Schmidt 1989, MacKinnon 1991).

3.2.2 Cointegration

Nevertheless it is possible that a non-stationary series can be transformed into stationary if we form the linear combination of both stationary and non-stationary series. A certain linear combination of two and more variables becomes stationary. Overall, a specific linear combination of two or more $I_{(d)}$ variables may become an $I_{(b)}$ variable, $b \leq d$. Therefore the non-stationary series will be considered as cointegrated. The Cointegration relationship demonstrates the presence of long run equilibrium relationship among the variables. Hence the long-run linear association among the variables is as follows:

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots \dots \dots \dots + b_m x_{mt} + v_t \qquad (3.2.2)$$

Whereas y is the response variable and $x_1; x_2 \dots x_m$ are explanatory variables. When in the above equation, two variables seem to appear, and then it is necessary that both (y_t and x_{1t}) should be integrated at same orders. Whereas if there are two or more independent variables, then integration order of the response variable should not be greater than the order of any of the independent variables. It is also considered that either none or at least two independent variables integrated to the same order greater than the order of the response variable. The above equation can be estimated by using OLS estimation technique. However, if the Equation (3.2.2) identifies a long run association, then use of ADF test on the estimated residuals $'v_t'$ in Equation (3.2.2) should show that the errors have a lower order $'b'$ than the order $'d'$ of $x_1, x_2, \dots\dots\dots\dots x_m$ and $'y'$. In this case, the $b_1, b_2, \dots\dots\dots\dots b_m$ represent the long-run elasticities of the response variable $'y'$ with respect to the corresponding explanatory variables.

When the cointegrating relationship is established, then the final step of the cointegration analysis is the creation of an Error Correction Mechanism. Cointegrating variables are probable to re-establish themselves to their long-run equilibrium whenever there is a drift, and Error Correction Mechanism deals with short-run behavior. The Error Correction Mechanism is expected using the following linear equation:

$$\Delta y_t = c_0 + \sum_{i_1=0}^{n_1} c_{1i_1}\,\Delta x_{1(t-i_1)} + \sum_{i_2=0}^{n_2} c_{1i_1}\,\Delta x_{1(t-i_2)} + \cdots \ldots + \sum_{i_m=0}^{m} c_{1i_m}\,\Delta x_{1(t-i_m)} + \sum_{j=1}^{p} c_i\,\Delta y_{t-j} + u_t \qquad (3.2.2.1)$$

Where $n_1, n_2, \ldots\ldots\ldots n_m$ and $'p'$ are the lagged terms of their similar variables, chosen to make the residual term $'u_t'$, white noise.

Now first we check that these variables are stationary or non-stationary, if non-stationary then the order of integration, and then the long-run relationship exists or not.

3.2.3 Auto Regressive Distributed Lag (ARDL) Model

Autoregressive distributed lag modeling methodology established by Pesaran and Pesaran (1998), and Pesaran et al. (2001). This technique becomes popular due to its several advantages like it is a single equation cointegration procedure. Other cointegration procedures like Engle and Granger (1987) suitable only for two variables and Johansen and Juselius (1990) can only be applied on the series also which are integrated for the same order but more than two variables and correct for large sample dataset. ARDL method is appropriate irrespective of the independent variables that are integrated at the same order or not (Hamuda et al. 2013). It’s also appropriate for finite sample data. If the variable order of integration is $I_{(2)}$, then the ARDL procedures makes no sense, and the calculated F statistics, as created Pesaran et al. (2001) and Narayan (2005) are not valid for longer.

ARDL have unbiased results for small sample whereas Engle and Granger produces biased results for small samples (Mah, 2000: 240). Another advantage of the ARDL method is that evaluation is probable even when the regressors are endogenous, and is sufficient to concurrently accurate for residual autocorrelation (Tang, 2004, 2005). To study the existence of the long run association, a bounds test which is established on the Wald or F statistics was projected by Pesaran (2001). The asymptotic distribution of the F statistics is non-standard under the H_0 that there is no long run association among the dependent and independent variables and alternative as H_1 exists long run relationship, whereas the regressors are $I_{(0)}$ and or $I_{(1)}$. An ARDL model is depicted below;

$$y_t = \mu + \sum_{i=1}^{p} \beta_i \, y_{t-i} + \sum_{i=1}^{q} \gamma_i \, x_{t-i} + u_t$$

Where y_{t-i} and x_{t-i} are variables which are both stationary, and u_t has no autocorrelation (white noise).

3.2.3.1 Diagnostic Tests

For diagnostic checking behind tests will be used Jarque Bera Test for normality, Breusch Pagan/ Godfrey Lagrange Multiplier Test for Hetroscedasticity and Breauch Godfrey La grange Multiplier Test for Autocorrelation and CUSUM, CUSUM-Square tests for stability will also be applied.

3.2.3.1.1 Jarque Bera Test

This test presented by Lomnicki (1961) and Jarque and Bera (1987) which is based on the Kurtosis and skewness of the distribution, (Lutkepohl, 2004, p. 45). This test (JB) is used for checking the normality of the distribution of error term. It is the assumption that the distribution of μt should be distributed as normally with zero mean and constant variance and it is suitable. It is valuable to obtain more precise results like confidence interval and test statistics but normality is essential to have results we use in multiple regression, (Greene W. H.,2012,p.276). The power of this test increases in occasion of when size of the population for distribution rise. The null and alternative hypothesis is as follows:

H_0 = the distribution of random elements is normal

H_1 = the distribution of random elements is non-normal.

The test statistic of the test is:

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

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

3.2.3.1.2 Breusch-Pagan Lagrange Multiplier Test/ Godfrey Lagrange Multiplier Test

This test is based on the background of Lagrange Multiplier test, (Breusch and Pagan, 1979). There are numerous tests to detect hetroscedasticity. Gold-Feld Quadnt test which appropriate to only one regressor, White test which is extremely general because if we want to apply it there is no need to make any assumptions about the nature of hetroscedasticity which is a merits of this test but at the same time a shortcoming due to the reason that if we reject the null hypothesis that there is homoscedasticity then it did not tell what to do next this test is no constructive but here I will use Breusch-Pagan test and this test is more powerful in the absence of normality. This test is sensitive to the assumption of normality, (Greene W. H., 2012, p. 276).

3.2.3.1.3 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 W. H., 2012, p. 903). The test available for the detection of autocorrelation based on the principal 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 but here we will use Lagrange Multiplier test because it is superior to other tests in the sense that Durbin Watson test can only be applied on the first order autoregressive process and Box and Pierce's test become less powerful than LM test when the null Hypothesis (H_o) is rejected but asymptocally both LM and Box and Pierce's test are equal. LM test was introduced by Breusch and Gogfrey in 1978, (Greene W. H., 2012, P. 923-924). Both authors worked independently to develop LM Test against the autoregressive or moving average process, (Jhonston, p. 185). The LM test is accurate for testing higher order autocorrelations in dynamic models but it can only applied two tailed not on one tailed. 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 cannot in the dynamic models, (Roise R. et al, 2012) the null and alternative hypothesis are as follows:

H_0: There is no autocorrelation.

H_1: There is autocorrelation.

3.2.3.2 Stability Tests

Since the earliest days of macro econometric analysis, researchers have been concerned about the appropriateness of the assumption that model parameter remains 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 macro econometric analysis in the last thirty years. Lucas (1976) stressed the point that the verdict models of financial agents are hard to define in terms of stable parameterizations, just due to changes in strategy may change these verdict models and their respective parameterization. These points of view emphasize the prominence of using structural stability tests as diagnostic checks for macro econometric models (Boldea, 2011).

3.3 Empirical Model

Now take our empirical variable for the ARDL model. The cointegration association for the passenger kilometer equation is expected which used bounds test, the test established on the below Unrestricted Error Correction Model:

$$\Delta PKM_t = \alpha_0 + \sum_{i=0}^{p} \beta_i \Delta GDP_{t-i} + \sum_{i=0}^{p} \delta_i \Delta POP_{t-i} + \sum_{i=0}^{p} \sigma_i \Delta FARE_{t-i} + \sum_{i=1}^{p} \gamma_i \Delta PKM_{t-i} + \sum_{i=0}^{p} \varphi_i \Delta DDOP_{t-i} + \theta_6 * GDP_{t-1} + \theta_7 * DDOP_{t-1} + \theta_8 * POP_{t-1} + \theta_9 * FARE_{t-1} + \theta_{10} * PKM_{t-1} + \psi ECM_{t-j} + u_t \qquad (3.3)$$

In the above model lagged terms shows the long run part and difference terms showing short run dynamics and ECM (Error Correction Mechanism). ψ Indicate speed of adjustment when a shock occurs.

3.3.1 Bound Testing Procedure

For analyzing the long run relationship the following null and alternative hypothesis will be observed:

$H_0 : \theta_6 = \theta_7 = \theta_8 = \theta_9 = \theta_{10} = 0$ (No long run relationship exists)

$H_1 : \theta_6 \neq \theta_7 \neq \theta_8 \neq \theta_9 \neq \theta_{10} \neq 0$ (long run relationship exists)

The H_0 is tested by allowing for the UECM for passenger kilometer in equation (3.3) excluding the lagged variables PKM, GDP, POP, DDOP and FARE.

Narayan (2004) formulated two critical values which show $I_{(0)}$ for lower bound and $I_{(1)}$ for upper bound. For level of significance, if the calculated value which calculated by F-statistics is greater than $I_{(1)}$ upper bound then cointegration exists, when the calculated value is less than lower bound $I_{(0)}$ then the cointegration not present in the model and when the calculated value fall within the $I_{(0)}$ and $I_{(1)}$ then inconclusive.

To establish the ARDL model which is good fit, so conduct the stability and diagnostic tests. Diagnostic test studies normality of the model, autocorrelation, the functional form and heteroscedasticity. Where lags selection depends upon the Akaike Information Criterion and Schwarz Bayesian Criterion.

3.4 Forecasting

3.4.1 Autoregressive Integrated Moving Average Model

There are numerous causes why an Autoregressive Integrated Moving Average (ARIMA) model is better than to multivariate regression and joint time series analysis. The joint finding in multivariate regressions and time series investigation is that the error-term is associated with the values of their own lagged. This autocorrelation violates the regression theory assumption that the current value of the error does not depend on the previous error values.

The key problems related with autocorrelation are:

- Basic time-series analysis and regression are not so longer effective amongst the dissimilar linear estimators. Though, as the previous residuals which help for prediction of current residuals, so we take benefit from this type of information to well forecast of the response variable through autoregressive integrated moving average.
- Standard errors which calculated by using the time series analysis and regression are not accurate and generally understand. If there include lagged value of the response variables as a regressors, the estimations coefficient of regressors are not unbiased and also not consistent but these problem are fixed in ARIMA model.

The approval of the autoregressive integrated moving average model is in line for its statistical properties and its famous Box Jenkins procedure in the model structure progression. In addition, numerous exponential smoothing models can be executed by ARIMA models. While ARIMA models are relatively flexible which represent different kinds of time-series, that is autoregressive, moving average and combine both are autoregressive moving average (ARMA) series, their primary constraint is the pre-assumed linear form of the model. i.e., a linear relationship construction is presumed amongst the values of time-series and for that reason, no non-linear forms can be captured by the ARIMA model (Zhang g.p.,2001).

3.4.2 ARIMA Model

The autoregressive integrated moving average model, the forthcoming value of a variable is presumed to be a linear function of numerous previous observations and randomize errors. The below procedure that make the time-series has the form;

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (3.4.1)$$

Whereas ε_t is the error and y_t is the real value at t time period, correspondingly;

φ_i(i=1, 2, 3…, p) and θ_j (j=0, 1, 2… q) parameters of the model. q and p are integers and frequently represents orders of model. The error ε_t, presumed to be identically and independently distributed with zero mean and constant variance equation (3.4.1) requires several essential special cases of the autoregressive integrated moving average family of models. If the value of q is zero then equation (3.4.1) becomes an autoregressive (AR) model which p order. When p = 0, the model convert to a moving average (MA) model of q order. In the Box-Jenkins method comprises three iterative stages which one is the identification of model second is diagnostic checking and third is estimation of parameter. The first step the identification of model is that if a time-series is created from an autoregressive integrated moving average procedure; it should have some theoretical serial correlation properties. By identical patterns of the experiential serial correlation with the hypothetical ones, it is frequently probable to recognize one or numerous potential models for the given time-series. Box and Jenkins projected partial autocorrelation function and autocorrelation function for lag selection of autoregressive integrated moving average model.

The last stage of model structure is to check diagnostic test for model acceptability. This is fundamentally to check the assumptions of the model which error term are fulfilled. Numerous statistics of diagnostic test and residuals plots is used to examine the good fit of the data. If model is not accept, means not fulfilled the assumption of the theory than a new uncertain model should be known, which is again followed by the steps estimation of parameter and model confirmation. This three stage model structure procedure is usually repeated numerous times up to acceptable model is lastly nominated. Now the lastly model which selected can be used for the forecasting purposes. With the special effects of Box Jenkins, the ARIMA specification is most popular univariate methods in the predicting research and practice (Zhang g. p., 2001).

3.5 Comparison of ARDL forecasting with ARIMA

For the forecasting performance mean absolute error, root mean square error and mean absolute percent error were used. In our case we compare the forecasting performance of autoregressive distributed lag model with ARIMA model. The model which has less values of MAE, RMSE and MAPE the better forecasting performance as compare to the other. The formulae for these three are given below

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}\left|\mathrm{y}_{(\mathrm{t})} - \hat{\mathrm{y}}_{(\mathrm{t})}\right| \qquad (3.5.1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathrm{y}_{(\mathrm{t})} - \hat{\mathrm{y}}_{(\mathrm{t})}\right)^2} \qquad (3.5.2)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left(\left|\frac{y_t - \hat{y_t}}{y_t}\right|\right) \times 100\% \qquad (3.5.3)$$

Where number of observation is n, actual values are $y_{(t)}$ and the forecasted values are $\hat{y}_{(t)}$.

CHAPTER 4 EMPIRICAL ANALYSIS

On empirical ground, we investigate the effect of different explanatory variables on rail passenger kilometers demand. Therefore we will draw lucid and vivid crux based on empirical findings. The results are discussed as follows.

4.1 Descriptive statistics

The descriptive statistics of the study are mention in Table 4.1. Descriptive statistics depicts the characteristics of data set at a glance such as; standard deviation, kurtosis, Jarque Bera, values of the selected variables, skewness and the averages values.

Table 4.1: Descriptive Statistics of Variables

Variables	Mean	Median	Max	Min	Std.Dev	Skew[5]	Kurt[6]	J.Bera	p-value
PKM	19914.8	18980	25621	16385.13	2812.469	0.753	2.437	3.55	0.169
GDP	2.83E+4	16298.3	111891.5	2932.176	30097.35	1.4423	4.102	13.11	0.001
FARE	0.264882	0.2321	0.7785	0.0343	0.190058	0.979	3.305	5.40	0.067
DDOP	1.91E+01	7.1075	92.112	2.923	23.95335	1.733	5.001	22.08	0.000
POP	1.30E+08	1.30E+08	1.79E+08	79984297	30354119	-0.033	1.764	2.11	0.349

The mean average values of the data and median about the most middle value of the ascending or descending order of the data. The standard deviation tells us about the deviation of the series from its mean for policy analysis. The skewness and kurtosis shows whether the series follows the normal distribution or not. Positive skewness show that most of the observations lies to the

[5] Skewness
[6] Kurtosis

right of its mean value while negative skewness means that most of the observations lie to the left of its mean value. Similarly kurtosis show about the peakedness of the data that is whether the series is leptokurtic, platykurtic or mesokurtic (normal). In our case, skewness of all variables found to be positive except population which is negatively skewed. Kurtosis values of GDP, fare and domestic diesel oil prices indicating lepto-kurtic distribution due to values which greater than 3, whereas passenger kilometer and population is platy-kurtic which values has less than 3. Jarque Bera values indicating that error-terms are distributed normally with zero mean and constant variance.

After analyzing descriptive statistics we have taken log of all the series except fare such that we used level log model. Before drawing Cointegration analysis it is needed to check the presence of non-stationarity, if yes, then the order of integration of all variables will need to be undertaking. The stationarity of the series can be checked by unit root. Hence to check unit root process Augmented Dicky-Fuller (ADF) test is very common in practice.

4.2 Augmented Dicky-Fuller (ADF) Test

The results of the ADF unit root test shows that all the series were non-stationary at 5% level of significance except the total population. All of the series were become stationary after taking first difference. This study consisted of the followings variables *i.e.*, passenger kilometers, GDP per capita, fare and Domestic Diesel oil prices were integrated at order one *i.e.*, $I_{(1)}$, while total population is integrated of order zero, $I_{(0)}$.

Table 4.2 ADF Test Results

Variable	Level		1st Difference	
	T-statistics	p-value	T-statistics	p-value
PKM	-1.601	0.4707	-4.102	0.0033**
GDP	-2.335	0.4045	-6.148	0.000***
FARE	-1.816	0.3662	-4.725	0.0035**
POP	-3.642	0.0429**	-0.450	0.9805
DDOP	-1.8765	0.6397	-6.2845	0.0001***

Note: ***, **, *indicates, significance at 1, 5, 10 percent level of significance respectively.

The results depicted that all variables are $I_{(1)}$ except population which is $I_{(0)}$ and that they satisfy the conditions under which ARDL can be used. The condition that the dependent variable are integrated of order one which satisfied that the passenger kilometer are $I_{(1)}$ and other condition that the integration are not the same order which also satisfied that four variables are integrated of order one and one variable are order zero. The third condition that none of the variables are integrated of order two which also satisfied. Fourthly, our sample data is a finite sample so we can apply the Autoregressive Distributed Lag Model.

4.3 Co-integration

According to ADF (1979) test of stationary of variables at first difference *i.e.,* passenger kilometers, GDP per capita, domestic diesel oil prices and fare are integrated of order one, $I_{(1)}$, except total population which is integrated of order zero *i.e* $I_{(0)}$. In such a situation the long-run relationship between passenger kilometer and GDP per capita, domestic diesel oil prices, fare and population has been analyzed by ARDL Bound test. The results of final ARDL model are represented below:

Table 4.3: ARDL Model (3,3,0,1,3)

Variable	Coefficient	Std. Error	p-value
C	-14.66125	8.175793	0.0962
D(LPKM(-1))	0.629931	0.198317	0.0073
D(LPKM(-2))	0.454602	0.205610	0.0456
D(LPKM(-3))	0.548295	0.239766	0.0396
D(LGDP(-1))	0.922018	0.380512	0.0307
D(LGDP(-2))	0.727860	0.313490	0.0371
D(LGDP(-3))	0.636369	0.240251	0.0201
D(LPOP(-1))	1.447461	3.145472	0.0426
D(LDDOP(-1))	-0.723010	0.155478	0.0005
D(LDDOP(-2))	-0.469074	0.117289	0.0015
D(LDDOP(-3))	-0.457223	0.114400	0.0015
LPKM(-1)	-1.261883	0.221730	0.0001
LGDP(-1)	-0.975780	0.403210	0.0309
FARE(-1)	-0.415648	0.219792	0.0811
LPOP(-1)	2.908270	1.210282	0.0319
LDDOP(-1)	0.683052	0.149691	0.0005

The below figure show the results of stability tests applied on ARDL Model

Figure 4.1: CUSUM Test

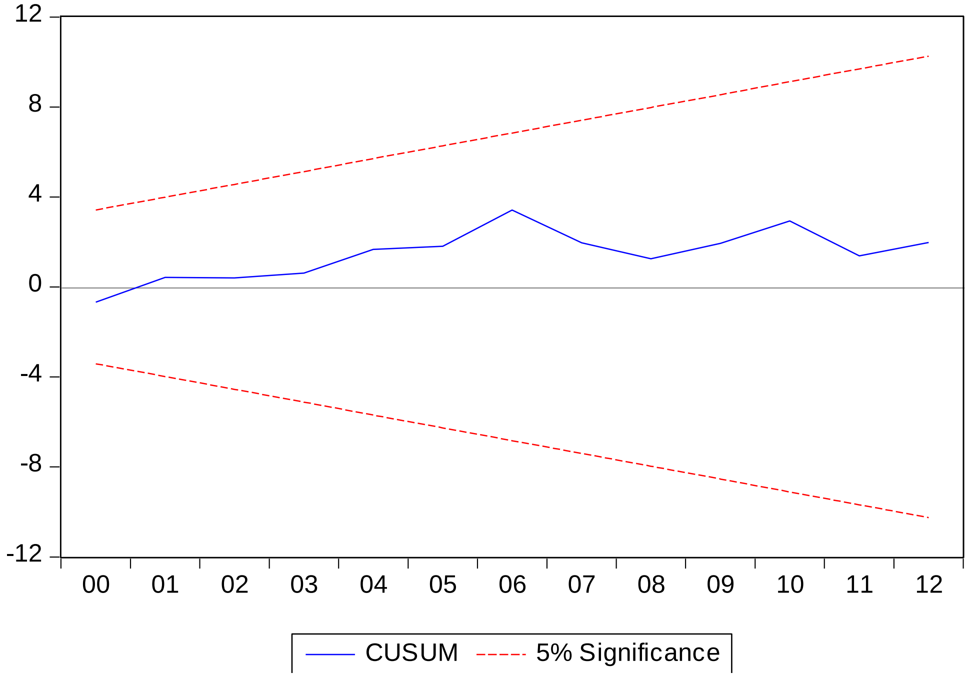
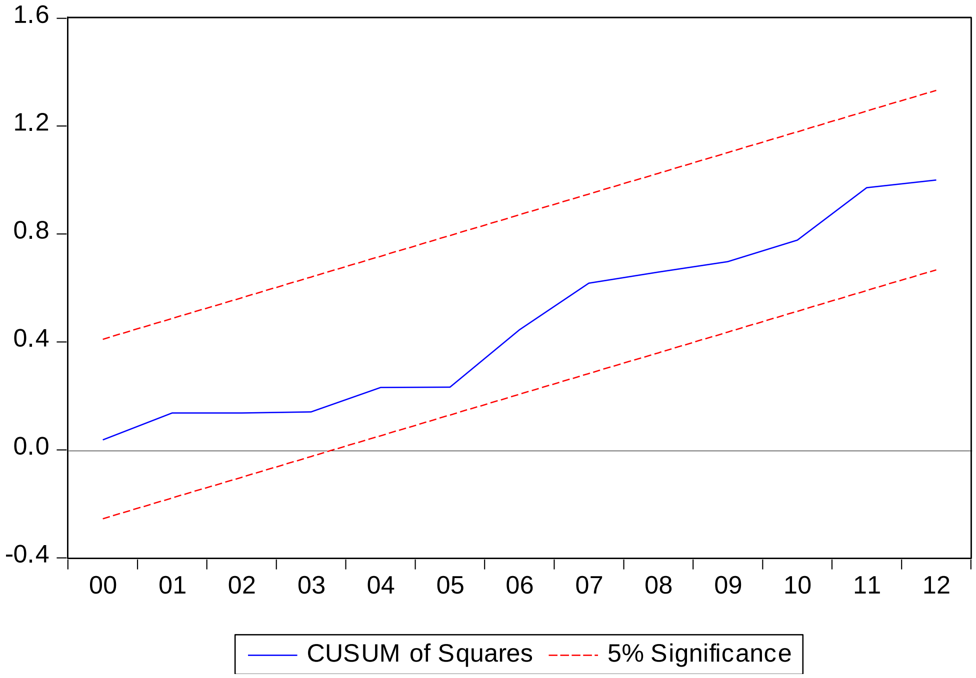


Figure 4.2: CUSUM Square Test



The plot of Cumulative sum of recursive residual (CUSUM) and Cumulative sum of squares of recursive residual (CUSUMSQ) statistics depicts the absence of misspecification and structural instability for the estimation period.

Table 4.4 below displayed the results of the co-integration test of the model employed in this study using the Bound test.

Table 4.4: Co-integration Test using ARDL

Equation	F-statistic	lower bound I(0)	upper bound I(1)	p-value	Remark
PKM	7.382091	2.86	4.01	0.0018	Co-integrated

The critical values which given in Pesaran table, case three (unrestricted intercept and no trend). The results displayed that the calculated value is greater than the upper bound value, so we do not reject the null hypothesis (there is no long-run relationship exists) and accept the alternative hypothesis (there is long-run relationship exists). The normalized vectors of variables (GDP per capita, Domestic Diesel oil prices, fares and total population) on passenger kilometer displayed long-run relationship in the model, which is shown in Table 4.5.

Table 4.5: Estimated Long-Run Coefficient ARDL Model

Variable	Coefficient	Std. Error	p-value
FARE	-0.288962	0.219438	0.0347
POP	3.037205	1.197847	0.0277
GDP	0.995516	0.391003	0.0272
DDOP	-0.621856	0.148583	0.0015

4.4 Error correction Model (ECM)

The ECM delivers a background for establishing links among the short-run and long-run methodologies to econometric modeling. The outcome of the passenger kilometers ECM is presented in Table 4.5

Table 4.6: Error Correction Mechanism outcomes

Variable	Coefficient	Standard error	p-value
C	-16.42974	8.142096	0.0687*
D(LPKM(-1))	0.744185	0.198846	0.0033**
D(LPKM(-2))	0.289312	0.220968	0.0171**
D(LPKM(-3))	0.480006	0.237367	0.0682*
D(LGDP(-1))	0.925571	0.366746	0.0283**
D(LGDP(-2))	0.668904	0.301522	0.0485*
D(LGDP(-3))	0.648074	0.228027	0.0160**
D(LPOP(-1))	4.502409	7.142924	0.0414*
D(LDDOP(-1))	-0.671236	0.159748	0.0015***
D(LDDOP(-2))	-0.415699	0.120739	0.0055**
D(LDDOP(-3))	-0.408209	0.113461	0.0042***
ECT(-1)	-0.837632	0.351766	0.0072**
R^2	0.855476	Adj-R^2	0.845259
DW	2.088699		

Serial correlation LM test		0.586639 [0.5761]
Breusch-Pagan-Godfrey		0.581533 [0.8426]
Jarque-Bera test		4.754038[0.120712]

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

It is shown in the table above that p-value for all variables are small at 5% level which indicating significant impact on passenger kilometer. The coefficient of fare varied negatively with passenger rail demand in all lagged changes and long run**.** The negative association between fare and passenger rail demand is consistent with the findings of findings of Doi and Allen (1986), Owen and Phillips (1987), Beko (2003) and Wijeweera and Charles (2013). This negative association may be due to the fact that demand always negatively influence by price. So high fares harms passenger rail demand in both the dynamics. At last increase in fare will lead to be a significant decrease in the passenger rail demand.

Passenger kilometers and population are positively associated; the coefficient of total population is statistically significant at practically any level. So, passenger rail demand also tends to be positivity associated with population progression. This association is mostly consistent with Ramanthan and Parikh (1999), Ramanthan (2001) and Wijeweera and Charles (2013). It show that population is positively associated with passenger rail demand. The population is steadily rising of the country so the passenger rail demand increase in the coming years. The higher the population the demand of railway passenger should be larger.

There is positive relationship among passenger kilometer and GDP Per capita. Moreover the slope coefficient of GDP per capita is found to be significant. Therefore the rise in GDP per capita would uplift passenger rail demand. The relationship between GDP and passenger rail demand are consistent with Owen and Phillips (1987), Ramanathan and Parikh (1999), Ramanathan (2001) and Wijeweera and Charles (2013). It indicated that GDP has a positive influence with most of the economic variables. Long run higher levels of income indicate that passenger rail demand is positively associated with GDP in Pakistan. However in short run a rapid rise in passenger rail demand is also associated with increase of GDP in Pakistan.

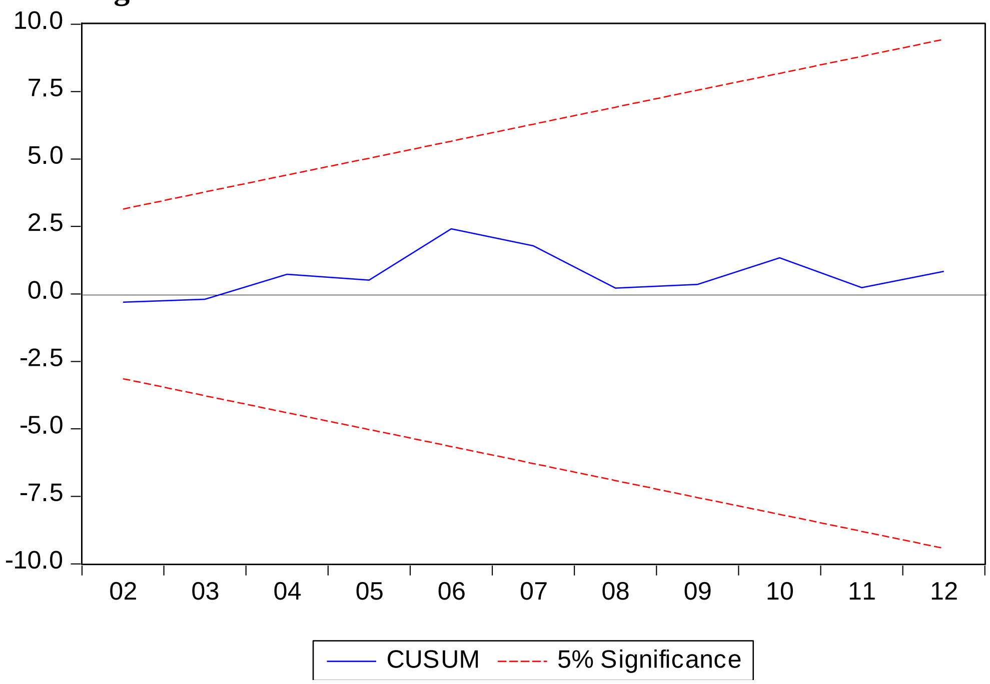
Domestic oil prices used to measure the macroeconomic activity effect. This variable has not been used usually as the determinant for the passenger rail but it is a significant variable because Pakistan railways uses diesel oil for transportation that's why it may have impact on passenger rail demand. So, this variable is comprised in the model has negative and significant influence on passenger rail demand at all lagged changes and levels. So fare and domestic diesel oil prices displayed inverse association with passenger rail demand growth. It can be due to when oil prices increases then it have immediate influence on all the sectors of the economy. In the prior year's oil prices remained high as a result it increases the costs of Pakistan railways which leads to increase in fare, eventually passenger rail demand declined. Further it before 90's has small variations in diesel oil prices but from the start of 90's increased gradually and about half of this decade quickly which resulted in decreasing passenger rail demand.

P-values of diagnostic tests are shown in square brackets. Breusch Godfrey LM Serial Correlation Test, Breusch-Pagan-Godfrey test have been concluded on the base of F-statistics and Jarque-Bera normality test based on the Chi-Square distribution. 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 estimation outcome in table 4.5 shows that ECT has negative and statistically significant, so the coefficient of the ECT of -0.837632 suggests reasonable adjustment process. About 84 percent of the disequilibrium of the previous period is adjusted back to the long run equilibrium in the current period. The significance of lagged equilibrium correction term with negative sign confirms the convergence of long run equilibrium and justifies Banerjee (1998) condition for the presence of stable long run association among the variables.

Plot of cumulative sum of recursive residuals of the ARDL (3, 3, 0, 1, 3) model to check for the stability of the estimated parameters:

Figure 4.3: CUSUM Test

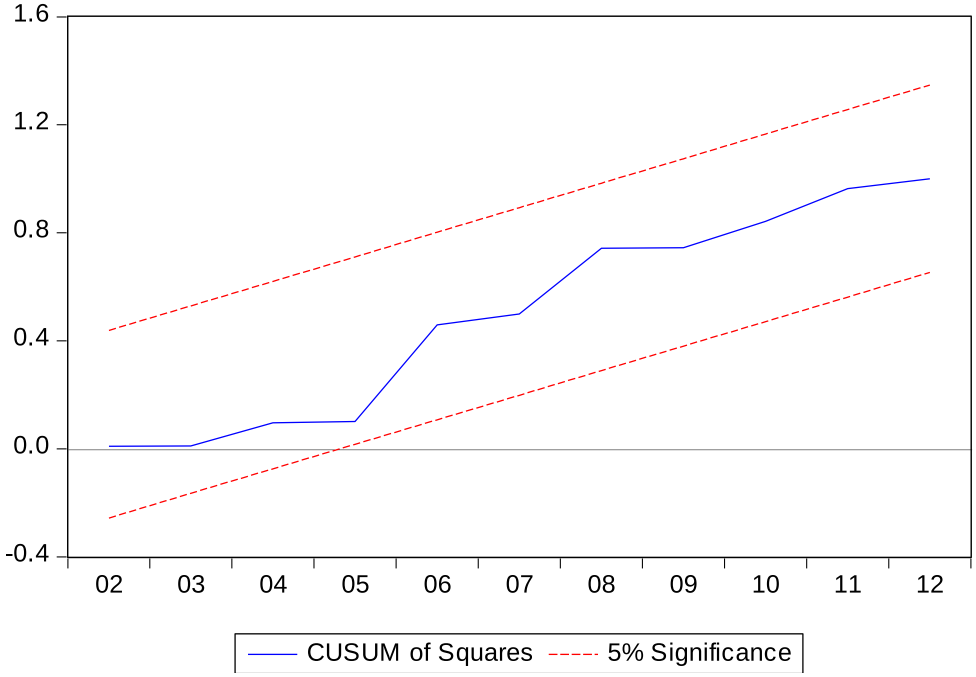


Straight lines represent critical bounds at 5 percent significance level.

The design of the stability test of the model is available in above figure (CUSUM). The CUSUM are designed against the critical bound at the 5 percent significance level. The results displayed that the model is sable in the meantime the critical bounds at 5 percent fell in between 5 percent lines.

The result showed that the model is stable since the critical bounds at 5 percent fell in between the two 5 percent lines. Also there is no autocorrelation in the model

Figure 4.4: CUSUM Square Test



4.5 Forecasting through ARIMA

The within sample forecast values of passenger kilometers to the railways sector in Pakistan, the ARIMA model is applied. On the first step we check autocorrelation function and partial autocorrelation function. Now on the level both the ACF and PACF probability values are insignificant then we take on first difference, so the probability values of both ACF and PACF are statistically significant. On the other step we apply ADF test to see the series is stationary or not, so we apply ADF test on the basis of this test the series is non-stationary at level and it has been made stationary after taking its first difference so table given below:

Table 4.7: ADF Test Results for PKM

Variable	At Level		1st Difference	
	T-Statistics	p-value	T-statistics	p-value
PKM	-1.601	0.4707	-4.102	0.0033**

Note: ***, **, *indicates, significance at 1, 5, 10 percent level of significance respectively

To define the proper order of autoregressive (AR) and moving average (MA), several equations have been tried in which the following equation was found appropriate, which yields the proper order of AR and MA. The outcomes are given below, which displayed that for forecasting passenger kilometers, the proper order of AR and MA are seven and five respectively.

Table: 4.8: ARIMA (7, 1, 5)

Dependent Variable: PKM				
Variable	Coefficient	Std. Error	T-Statistic	p-value
AR(7)	0.420724	0.199302	2.110986	0.0459
MA(5)	-0.870233	0.048374	-17.98954	0.0000
R-squared	0.541564	Mean dependent var		0.001845
Adjusted R-squared	0.521632	S.D. dependent var		0.026475

On the above derived results, the proper model for forecasting is ARIMA (7, 1, 5).

Now the forecasting comparison of ARDL model and ARIMA model which results below:

Table: 4.9 Comparisons of ARDL and ARIMA model

Prediction model	MAE	MAPE	RMSE
ARIMA	0.018223	0.61254	0.027535
ARDL	0.016324	0.374573	0.019547

The above table shows that the autoregressive distributed lag model forecasting has a superior prediction value as compared to the ARIMA model.

Based on the consequence which is derived above, the forecasting performance of ARDL model is superior to the ARIMA model. The values of mean absolute percent error; root mean square error and mean absolute error of autoregressive distributed lag (ARDL) model is less than autoregrssive integrated moving average (ARIMA) model. On the basis of these values the

ARDL model forecast is better than ARIMA model forecasting.

CHAPTER 5 CONCLUSION AND POLICY RECOMMENDATION

5.1 Conclusion

The ultimate objective of this research was to estimate the long run association among Passenger Kilometer and its determinants (*i.e.,* GDP per capita, Fare, Domestic Diesel oil prices and total Population), and the forecasting performance of ARDL model along with ARIMA model is the railways sector in case of Pakistan. The data set used in this study consists of annual observations covering the period from 1980 to 2012. The passenger kilometers (PKM) was taken as the response variable and the explanatory variables are total population, GDP per capita, Domestic diesel oil prices and Fare.

The results revealed by employing ADF technique, the Population was stationary at level and all others variables *i.e.,* GDP per capita, Fare, domestic diesel oil prices and Passenger Kilometer (PKM) were become stationary after first difference. To test the long-run relationship, the situation fit to ARDL Bound Test. The passenger kilometer model exhibited long-run relationship when the vectors of variables (GDP per capita, Fares, Domestic Diesel oil prices and Population) were normalized on passenger kilometer. The results postulated that the GDP per capita, domestic diesel oil prices, population and Fare have a long-run influence on passenger kilometers.

The results of ECM demonstrated negative sign and statistically significant, stating the coefficient of the ECM of -0.83 which suggest that about 83 percent of the disequilibrium of the preceding period is adjusted back to the long run equilibrium in the present period. Overall, the findings of this study demonstrated that the all determinants plays pivotal role in the performance of passenger kilometers in Pakistan.

In order estimate the forecasting performance values of passenger kilometers to the railways sector in Pakistan, the ARDL and ARIMA models were applied. We have applied ARDL and ARIMA models forecasting from 2000 to 2012. On the basis of mean absolute percent error, mean absolute error and root mean square error of ARDL model is superior as compare to ARIMA model.

Policy Implication

The findings of the study showed that GDP per capita and population has positive impact on passenger rail demand while fare and domestic diesel oil prices negatively. On the basis of results our policy recommendations are:

Government may focus on the policies that can encourage passengers to move towards railway sector because GDP per capita and population play a major role in increasing passenger rail demand. On the other hand government may provide diesel oil on subsidized prices or to generate alternate resources which can reduce the cost of Pakistan railways in the form of electric trains because Pakistan railway has to bear more than 50% expenditures in the form of fuel. If Pakistan railways charge lower rates than other modes of transport then it can also encourage masses to increase passenger rail demand. It is better to use multivariate ARDL model than univariate ARIMA model due to its better forecasting performance.

References

Akinboade, A.O., Ziramba, E., and Kumo, L. W. (2008). The demand for gasoline in South Africa: An empirical analysis using co-integration techniques. *Energy Economics, 30,* 3222-3229.

Ahmed, U. M., Muzib, M., and Roy, A. (2013). Price-wage spiral in Bangladesh: Evidence from ARDL bound testing approach. *International Journal of applied economics, 10(2),* 77-103.

Albert, W., and Michael, C. (2013). Determinants of Passenger Rail Demand in Perth, Australia: A Time Series Analysis. *Applied Econometrics and International Development, 13(2).*

Andrikopoulos, A. A., and Loizides, J. (1998). Cost structure and productivity growth in European railway systems. *Applied Economics,30,* 1625-1639.

Beko, J. (2003). Demand Functions for Services of Public Railway Passenger Transportation: An Empirical Analysis for Slovenia. *Managing Global Transitions, 1 (2),* 135-152.

Bentzen, J. (1994). An empirical analysis of gasoline demand in Denmark using cointegration techniques. *Energy Economics, 16(2),* 139-143.

Bhatti, A. J., and Razzak, A. J. (2010). Railway associated injuries in Pakistan. *International Journal of Injury Control and Safety Promotion, 17 (1),* 41-44.

Bollinger, R. C., and Ihlanfeldt, R. K. (1997). The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's MARTA. *Journal of Urban Economics, 42,* 179-204.

Butkevicius, J., Mazura, M., Ivankovas, V., and Mazura, S. (2004). Analysis and forecast of the dynamics of passenger transportation by public land transport. *Transport, 19 (1),* 3-8.

Cantos, P., Pastor, M. J., and Serrano, L. (1999). Productivity, efficiency and technical change in the European railways: A non-parametric approach. *Transportation, 26,* 337-357.

Celebi, D., Bolat, B., and Bayraktar, D. (2009). Light rail passenger demand forecasting by artificial neural networks. *Management Engineering, Istanbul Technical University, 34367, Turkey.*

Chen, N. (2007). Modeling Demand for Rail Transport with Dynamic Econometric Approaches. *International Review of Business Research Papers, 3(2),* 85-96.

Chiang, C. W., Russell, A. R., and Urban, L. T. (2011). Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A, 45,* 696-705.

Choudhary, A. M., Khan, N., Arshad, M., and Abbas, A. (2007). Analyzing Pakistan's Freight Transportation Infrastructure Using Porter's Framework and Forecasting Future Freight Demand using Time Series Models. *Proceedings of the 2nd WSEAS International Conference on URBAN PLANNING and TRANSPORTATION.*

Cyprich, O., Konecny, V., and Kiliannova, K. (2013). Short-term passenger demand forecasting using univariate time series theory. *Traffic and Transportation, 25(6),* 533-541.

Dailydka, S., Naimovicius, S., and Lingaitis, V. (2012). Modeling of Selection of Passenger Trains on A Route. *Contemporary Issues in Business, Management and Education.*

Doi, M., and Allen, B.W. (1986). A time series analysis of monthly ridership for an urban rail rapid transit line. *Transportation, 13,* 257-269.

Dou, F., Xu, J., Wang, L., and Jia, L. (2013). A train dispatching model based on fuzzy passenger demand forecasting during holidays. *Journal of Industrial Engineering and Management (JIEM), 6(1),* 320-335.

Ebi, K. U., Mbonu, B. O., and N Aniagolu, H. (2014). Does Transportation sector have any impact on economic development? A time series assessment with Nigeria in focus. *International Journal of Economics, commerce and Management, 2 (8).*

Eltony, M. N., and Al-Mutairi, N, H. (1995). Demand for gasoline in Kuwait An empirical analysis using cointegration techniques. *Energy Economics, 17(3),* 249-253.

Gathon, J. H., and Pestieau, P. (1995). Decomposing efficiency into its managerial and its regulatory components: The case of European railways. *European Journal of Operational Research, 80,* 500-507.

Haines, R. M., and Margo, A. R. (2006). Railroads and local economic development: The United States in the 1850S. *Neber working paper series.*

Hamuda, M. A., Sulikova, V., Gazda, V., and Horvath, D. (2013). ARDL investment model of Tunisia. *Theoretical and Applied Economics, 20(2),* 57-68.

Hamzacebi, C., Akay, D., and Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert System with Application,* 36, 3839-3844.

Irfan, M. S., Kee, H. M. D., and Shahbaz, S. (2012). Service Quality and Rail Transport in Pakistan: A Passenger Perspective. *World Applied Sciences Journal, 18 (3),* 361-369.

Khan, A. M,. and Raza, S. (2015) An empirical analysis of rail freight demand in Pakistan. *Pakistan institute of development economics (PIDE) Islamabad,*

Lan, W. L., and Lin, J. T. (2006). Performance Measurement for Railway Transport: Stochastic Distance Functions with Inefficiency and Ineffectiveness Effects. *Journal of Transport Economics and Policy, 40 (3),* 383-408.

Loncan, H. A. (2011). The Role Of Railways In Export-led Growth: The Case of Uruguay, 1870-1913. *Economic History of Developing Regions, 26 (2),* 1-32.

McGeehan, H. (1993). Railway Costs and Productivity Growth The Case of the Republic of Ireland, 1973-1983. *Journal of Transport Economics and Policy, 27(1),* 19-32.

Odgers, F.J., and Schijndel, V. A.L.(2011). Forecasting annual train boardings in Melbourne Using time series data. *Australasian Transport Research Forum.*

Odili, O. (2014). Exchange rate and balance of payment: An autoregressive distributed lag (ARDL) Econometric investigation on Nigeria. *IOSR Journal of economics and finance (IOSR-JEF), 4(6),* 21-30.

Oum, H. T., and Yu, C. (1994). Economic Efficiency of Railways and Implications for Public Policy A Comparative Study of the OECD Countries' Railways. *Journal of Transport Economics and Policy, 28 (2),* 121-138.

Owen, D. A., and Phillips, D. G. (1987). The characteristics of railway passenger demand: An Econometric Investigation. *Journal of Transport Economic and Policy, 21 (3),* 231-253.

Paul, P. B. (2014). Testing Export-Led Growth in Bangladesh: An ARDL Bounds Test Approach. *International Journal of Trade, Economics and Finance, 5(1).*

Profillidis, A. V., and Botzoris, N. G. (2006). Econometric models for the forecast of passenger demand in Greece. *Journal of Statistics and Management Systems, 9(1),* 37-54.

Ramanathan, R. (1999). Short- and long-run elasticities of gasoline demand in India: An empirical analysis using cointegration techniques. *Energy Economic*, 21, 321-330.

Ramanathan, R. (2001). The long-run behavior of transport performance in India: a cointegration approach. *Transportation research part A, 35,* 309-320.

Ramanathan, R., and Parikh, K. J. (1999). Transport sector in India: an analysis in the context of sustainable development. *Transport Policy, 6,* 35-45.

Rose, G. (1986). Transit passenger response: Short and long term elasticities using time series analysis. *Transportation, 13,* 131-141.

Shitan, M., Karmokar, K. P., and Lerd, Y. N. (2014). Time series modeling and forecasting of Ampang line passenger ridership in Malaysia. *Pak. J. Statist. 30(3),* 385-396.

Shittu, I. O., Yemitan, A. R., and Yaya, S. O. (2012). On autoregressive distributed lag, Cointegration and error correction model [An application to some Nigeria macroeconomic variables]. *Australian Journal of Business and Management Research, 2(8),* 56-62.

Shahbaz, M., Khraief, N., and Jemaa, B. M. M. (2015). On the causal nexus of road transport CO_2 emissions and macroeconomic variables in Tunisia: Evidence from combined cointegration tests. *Renewable and Sustainable Energy Reviews, 51,* 89-100.

Shrestha, B. M., and Chowdhury, K. (2005). ARDL modeling approach to testing the financial liberalization hypothesis. *University of Wollongong economics working paper series.*

Smith, L. B., Williams, M. B., and Oswald, K. R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C, 10,* 303-321.

Tolulope, O. A., and Taiwo, A. (2013). Rail Transport and Economic Growth in Nigeria (1970-2011). *Australian Journal of business and management Research, 3(05),* 18-24.

Transport Policy, 8, 77-83.

Tsai, H. T., Lee, K. C., and Wei, H. C. (2003). An artificial neural networks approach to forecast short-term railway passenger demand. *Journal of the Eastern Asia Society for Transportation* Studies, 5, 221-235.

Tsai, H. T., Lee, K. C., and Wei, H. C. (2009). Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications, 36,* 3728-3736.

Wardman, M. (1997). Inter-urban rail demand, elasticities and competition in Great Britain: Evidence from direct demand models. *Transpn Res.-E (logistics and Transpn Rev.), 33 (1),* 15-28.

Whelan, G., and Johnson, D. (2004). Modeling the impact of alternative fare structures on train overcrowding. *International Journal of Transport Management, 2,* 51-58.

Wijeweera, A., and Charles, M. (2013). An Empirical Analysis of the Determinants of Passenger Rail Demand in Melbourne, Australia. *Economic Analysis & Policy, 43(3),* 249-264.