Forecasting tax revenues using time series techniques – a case of Pakistan

Dalia Streimikiene, Rizwan Raheem Ahmed, Jolita Vveinhardt, Saghir Pervaiz Ghauri & Sarwar Zahid

Published online: 09 Mar 2018.

Submit your article to this journal

Article views: 1207

View Crossmark data

Routledge
Taylor & Francis Group

Forecasting tax revenues using time series techniques – a case of Pakistan

Dalia Streimikiene[a], Rizwan Raheem Ahmed[b], Jolita Vveinhardt[a], Saghir Pervaiz Ghauri[b] and Sarwar Zahid[c]

[a]Institute of Sport Science and Innovations, Kaunas, Lithuania; [b]Faculty of Managment Sciences, Indus University, Gulshan, Pakistan; [c]Department of Business Administration, Bahria University, Islamabad, Pakistan

ABSTRACT

The objective of this research was to forecast the tax revenue of Pakistan for the fiscal year 2016–17 using three different time series techniques and also to analyse the impact of indirect taxes on the working class. The study further analysed the efficiency of three different time series models such as the Autoregressive model (A.R. with seasonal dummies), Autoregressive Integrated Moving Average model (A.R.I.M.A.), and the Vector Autoregression (V.A.R.) model. In any economy, tax analysis and forecasting of revenues is of paramount importance to ensure the economic and fiscal policies. This study is important to identify significant variables affecting tax revenue specifically in Pakistan. The data used for this paper was from July 1985 to December 2016 (monthly) and focused on forecasting for 2017. For the forecasting of total tax revenue, we used components of tax revenues such as direct tax, sales tax, federal excise duty and customs duties. The results of this study revealed that among these models the A.R.I.M.A. model gives better-forecasted values for the total tax revenues of Pakistan. The results further demonstrated that major tax revenue is generated by indirect taxes, which cause more inflation that directly hits the working class of Pakistan.

1. Introduction

Taxation is not a popular terminology for individuals and business organisations, but it is a vital instrument for collecting revenues for government expenditure. As the collection of taxation increases, it allows the government to conceive maximum developmental projects for the public interest, and to improve the basic infrastructure of health, education, and the quality of life of common people. However, increase in taxation also increases the inflation level for goods and services, which make life more difficult to survive (Palacios & Harischandra, 2008; Poulson & Kaplan, 2008). Similarly, in business activities, local and foreign investors are reluctant to pour in any further investment in the country because of profit shrinkage. This situation triggers low output and decrease in overall G.D.P. of the

---

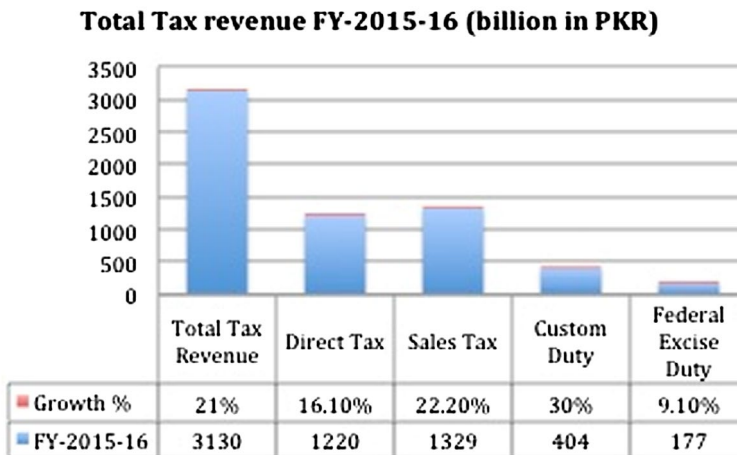**CONTACT** Dalia Streimikiene ✉ dalia@mail.lei.lt

country, which consequently also reduces the employment rate. The collection of taxation is now, a burning issue in Pakistan; therefore, the government is trying to document the whole economy, especially bringing services and agricultural sectors, and individual people into the tax net. The direct, and indirect taxation is an imperative question in the context of Pakistan because more than 60% of tax revenue is being collected through indirect taxes (Aamir et al., 2001). According to the Federal Board of Revenue (F.B.R.), the total tax revenue collection for the fiscal year (F.Y.) 2015–16 was Pakistani Rupee (P.K.R.) 3130 billion. a registered growth of 21% as compared to the previous fiscal year. The breakdown of total tax revenue and corresponding growth can be seen in Figure 1 (F.B.R., 2015–16).

According to the Federal Board of Revenue (F.B.R.) tax to G.D.P. ratio is a burning question in the Pakistani context because it is very low, between 8.5%–9.7% for last ten years as displayed in Figure 2 (F.B.R., 2015–16). It is even very low as compared to the other South Asian countries, however, the government set a target of 12.2% tax to G.D.P. ratio for the F.Y.-2016–17 (Federal Board of revenue, 2015–16).

This research is carried out in the context of future forecasting of these tax revenues for the fiscal year 2016–17. It is evident from Figure 1; the substantial amount of tax revenue is generated by indirect taxes. According to Myles (2000), the tax to G.D.P. ratio had significantly increased in developed economies, but in developing economies this ratio remained low. Chaudhry (2001) has proposed an optimal tax theory, and suggested an effective agricultural policy to increase tax net, and recommended, if local bodies were made to be responsible then tax revenue can be increased significantly.

1.1. Background of the study

According to the purchasing power, Pakistan is the 25th largest economy, and in regard to G.D.P., its rank is 38th. Pakistan has a population of more than 190 million making it the 6th largest country in the world. Besides the documented economy, there is a black economy estimated at 37% of its overall economy, which is not documented and not included in the tax net (Economic Survey of Pakistan, 2015–16). Pakistan is still a developing economy

Total Tax revenue FY-2015-16 (billion in PKR)

	Total Tax Revenue	Direct Tax	Sales Tax	Custom Duty	Federal Excise Duty
Growth %	21%	16.10%	22.20%	30%	9.10%
FY-2015-16	3130	1220	1329	404	177

Figure 1. Total tax revenue – F.Y.-2015–16. Source: Federal Board of Revenue (2015–16).
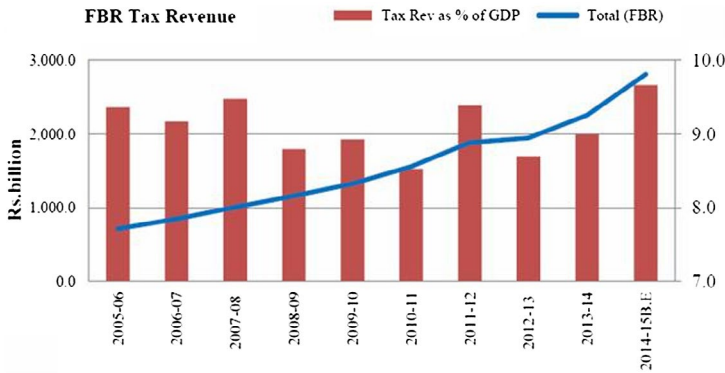
Figure 2. Total revenue to G.D.P. ratio. Source: Federal Board of Revenue (2015–16).

but it has tremendous potential to evolve as one of the largest economies in the world. However, political instability, and a vulnerable law and order situation for last 35 years have worsened the social life of the common man, this situation has had adverse affects on total tax revenue generation.

The Pakistani Rupee (P.K.R.) faces terrible pressure, and has been depreciating in relation to the U.S. dollar for the last several years; therefore, the current account deficit touches the surplus. However, the central bank of Pakistan tried to keep balance by lowering the discount rate, and buying U.S. dollars from the open market to safeguard exports in order to compete with the rival economies. According to the Economic survey of Pakistan (2015–16), the government has given sales tax exemptions of P.K.R.665 billion, in the current fiscal year, the government has already given sales exemption of P.K.R.478 billion in the first half.

The government has faced losses of P.K.R.65 billion taxes from the decrease in exports, and a further loss of P.K.R.18 billion is expected due to relief in tax under the fifth schedule of the sales tax act. The Federal Board of revenue also gave relief of P.K.R.83.6 billion in income tax, P.K.R.9.5 billion tax exemptions have been given to some specific industries, and P.K.R.2.5 billion losses because of inferior tax on capital gains in the stock market. Similarly, in the case of customs duty, the state provided the relief of P.K.R.103.1 billion, and P.K.R.27 billion losses incurred because of Pak-China free trade agreement. P.K.R.18.4 billion losses are expected in automobile imports' customs duty (Federal Board of revenue, 2015–16). Thus, in this way the government is going to incur significant tax revenue losses for the fiscal year 2016–17.

1.2. Objective of the research

The objective of this research is to forecast the total tax revenue for the fiscal year 2016–17, and analyse the impact of indirect taxes on the working class of Pakistan. We have selected three different time series models, namely the autoregressive (A.R.) model with seasonal dummies, the autoregressive integrated moving average (A.R.I.M.A.) model, and the vector Auto regression (V.A.R.) model for forecasting purposes. For the evaluation of the efficiency of these models, we used root mean squared error (R.M.S.E.) test. We have taken monthly data for the time period from July 1985 to December 2016. For the forecasting of total tax revenue, we use components of tax revenues such as direct tax, sales tax, federal excise

duty, and customs duties. For the multivariate model, we used the series such as large-scale manufacturing index (L.S.M.) – proxy for G.D.P. growth, consumer price index (C.P.I.), and international oil prices.

1.3. *Significance of this research*

The significance of this research is many fold, for example, we employed several econometric techniques to forecast the taken macroeconomic variables, this will be a significant addition to the literature, which will be helpful for the future research studies. We have incorporated the seasonal factors while we forecasted the total revenue of the country, this is an appropriate and suitable method in forecasting. We have examined the total revenue forecasting, the impact of direct tax, sales tax, customs duty, federal excise duty, large scale manufacturing (L.S.M.), consumer price index (C.P.I.), and international crude oil prices on total tax revenues in Pakistan. The results of this research provide the basis to formulate, and aligned the policies to the public, private institutions, and policy makers. Another important aspect of this research is the duration of the data, we have used quite substantial time period (31 years) from July 1985 to December 2016.

2. Literature review

The literature regarding determinants of tax revenue has gained considerable attention from financial experts, economist, and researchers worldwide. The findings of previous studies demonstrated the different effects in developed and developing economies. The developed economies have strong tax networks, which are well documented electronically. However, the developing countries are still in a transition phase to establish a strong and documented taxation system. Therefore, we have discussed the following previous literature specifically for the developing economies:

Himani (2016) studied the elements of tax collection in the case of India; he took tax revenue collection from direct and indirect means. He concluded that the tax collected through indirect means has generated more revenue as compared to direct taxes. Aamir et al. (2001) carried out a comparative study between India and Pakistan; they compared direct and indirect taxes for both countries. They concluded that Pakistan collected more revenuethrough indirect taxes, however, India generated more revenue through direct taxes. This study further indicated that indirect taxes extended the gaps between rich and poor, and further exploits the vulnerability of working class. Javid, Arif, and Arif (2011) examined the fiscal performance of distinctive developing economies of Asia. They concluded that trade openness, agriculture value addition, per capita G.D.P., and debt and population growth are significant elements of total revenue to G.D.P. ratio across developing economies. These studies have limited scope and did not incorporate the individual determinants of direct and indirect tax collection; thus, this gap is bridged in our research study, and we examined the impact of indirect taxes on the working class of Pakistan.

Eugene and Chineze (2016) studied the impact of taxation policies on the overall economic growth of Nigeria. The results of the study confirmed the positive impact of a tax on economic growth. The results of the study further demonstrated that there is a linear association between tax base and total tax revenue. They also concluded a positive and significant association between tax base, tax policies, and further concluded that there is

no significant association between economic growth and total tax revenue. Sunday (2015) also examined the tax policies of Nigeria, and its impact on inflation. He concluded that the taxation policies have a long-term association with the inflation rate, and consumption and property taxes also have a long-term positive association with inflation. The forecasting of tax revenue and other economic indicators were diagnosed through different forecasting models, and several studies proved the efficiency and superiority of these models (Lutkepohl, 1991, 1999; Sims, 1980; Waggoner & Zha, 1999; Watson, 1994)

Karagöz (2013) has examined the fiscal deficit of Turkey, he assessed cointegration through O.L.S., and concluded that the industrial and agricultural shares in G.D.P. affect the total tax revenue; he further concluded that foreign trade openness does not have any impact on the total tax revenue. Nanthakumar, Kogid, Sakami, and Muhamad (2011) examined the association between total tax revenue and the government spending in the case of Malaysia. The results of the study revealed that the total tax revenue is directly proportional to spending. Husain and Qasim (2007) examined the causal relationship between the government spending and the total tax revenue in Pakistan. The results of the study exhibited the unidirectional causality from spending to the total tax revenue. On similar topics some important research studies have been carried out that established the superiority of the forecasting models (Box & Jenkins, 1976; Nau, 2014; Tiao & Box, 1981; Zhang, 2013)

Daba (2015) investigated the elements of total tax revenues of Ethiopia, it is concluded from the study that F.D.I.s and per capita have a strong and significant impact on total tax collection. He further concluded that inflation and interest rates have an insignificant impact on total tax generation. According to Hassan and Tahmina (2012), the government of Bangladesh is borrowing significantly from government sources, which negatively affected the overall economy of the country. They finally concluded that internal borrowing hurts the economy and G.D.P. rate more as compared to foreign debts. These studies did not consider the individual tax revenue determinants, thus in our study we have examine the impact by employing direct and indirect factors for total tax revenue. Moreover, we predict the forecasting of next year's tax revenue by using three different methods; hence, our research is a further extension of these studies.

Patoli, Zarif, and Syed (2012) investigated the association between the inflation rate and tax revenue in the case of Pakistan. They concluded that total tax revenue and inflation have a positive correlation; they further concluded that any change in the inflation rate causes further increase in the taxation rate. According to Tanko (2015), personal income tax is the major source of total revenue for the Nigerian government. He concluded that the insufficiency of the internally collected tax hampers the overall economy of Nigeria. He further concluded that the greater dependency on the federal accounts and total expenses disrupt the economy, therefore, the government should craft the strategies to increase revenue creation. Rasheed (2006) has studied the relationship between the total tax revenue, and G.D.P., total tax collection and money supply, total tax and broad and narrow money supply, and total tax collection and C.P.I. etc. He used the cointegration technique and concluded that there is no strong relationship between the tax revenue growth and the public debt, tax revenue growth, and C.P.I., tax revenue growth, and investment and credit. Thus, we incorporate additional factors besides the considered variables of previous research; moreover, we forecast the tax revenue for the next year, hence, our study is an extension of these studies.

Mehrara, Pahlavani, and Elyasi (2011) investigated the association between the spending and the government revenue of 40 Asian economies. They recommended that the

governments of these countries should enhance their total revenue, and reduce their expenditure in order to control their budgetary deficits. Chaudhry and Munir (2010) examined the causes of low tax revenue in the case of Pakistan. The results of the study demonstrated that the trade openness, external aids, broad money, and political stability are the major causes of the tax collection. In the light of the results of these two studies, we have identified the gap in crucial variables, and incorporated these factors in our study to bridge this gap. Samuel (2014) has conducted a study for Nigerian tax revenue collection and took federally generated tax as the dependent variable. He has taken petroleum profit tax, value added tax, custom and excise duty, and company income tax as independent variables. The result of the study indicates that V.A.T. is significantly profitable to the Nigerian economy; using the same approach we have incorporated more variables to bridge the gap of this research. Das-Gupta (2011) examined the resilience of sales tax and revenue to G.D.P., and also investigated the impact of direct and indirect taxes on G.D.P. with and without V.A.T. He concluded that two thirds of the sample did not have any indirect effect on V.A.T. Kenny and Winer (2006) carried out a very important study in which they studied the tax structure of 100 countries. They concluded that the structure of tax generation changes with the change of political regimes. They further concluded that progress governments do not concentrate on the collection of revenues from individual income tax. In the same manner, the study also examines determinants of indirect taxes, and their impact on the economy and the wellbeing of individuals.

An important aspect of tax collection and tax evasion is based on the morality of individuals; several research studies have been carried out on this issue. These research studies have emphasised the role of religion, culture, and economic aspects on tax compliance (Dowling, 2014; Strielkowski & Čábelková, 2015). People are ready to pay taxes but governments have failed to fulfil the social contract with their inhabitants. The vulnerability of this situation has been reflected in developing countries, particularly where the governments have shown the deficiencies in order to provide social security. Thus, the morality does not depend on the individuals only but the governments are equally responsible for this fiasco (Čábelková & Strielkowski, 2013; Torgler & Schneider, 2007).

3. Data and methodology

3.1. Data collection

In our study, for forecasting purposes, we have used univariate examination to analyse a single variable (tax revenue) at a time for univariate variables such as the total tax revenue (T.R.), and considered its four important components, direct tax (D.T.), sales tax (S.T.), federal excise duty (F.E.D.), and customs duty (C.T.), but these are the determinants of tax revenue, thus we have employed univariate analysis (Bagshaw, 1987). On the other hand, we have selected three isolated variables to analyse together with tax revenue for potential interfaces, thus, we have used the multivariate model, and we have taken three variables such as large-scale manufacturing (L.S.M.), international oil prices (I.O.P.), and consumer price index (C.P.I.) (Bagshaw, 1987). The secondary data for the whole series are collected from the different issues of the Pakistan Bureau of Statistics (P.B.S.) monthly bulletin, however, the I.O.P. data is collected from the International Monetary Fund (I.M.F.) website shown in Table 1. We have considered a reasonable period of time for the forecasting purposes, we

Table 1. Data collection sources.

Variables	Frequency	Sources
Total Tax Revenue	Monthly	Pakistan Bureau of Statistics
Direct Taxes	Monthly	Pakistan Bureau of Statistics
Federal Excise Duty	Monthly	Pakistan Bureau of Statistics
Sales Tax	Monthly	Pakistan Bureau of Statistics
Custom Duty	Monthly	Pakistan Bureau of Statistics
Large Scale Manufacturing	Monthly	Pakistan Bureau of Statistics
Consumer Price Index	Monthly	Pakistan Bureau of Statistics
International Oil Price	Monthly	International Monetary Fund Website

Source: Authors' compilations.

took around 31 years' data for the time period from July 1985 to December 2016. The tax exemptions have already been incorporated in the final tax revenues, which were published by the Pakistan Bureau of statistics, thus, the impact of tax exemptions does not have any distortion while we carry out the undertaken research study.

3.2. Estimation and data analysis tools

As explained earlier, we have used time series data for the period from July 1985 to December 2016, and forecast the total tax revenue (T.R.) for Pakistan by taking its four components L.D.T., L.S.T., L.F.E.D., and L.C.D. The proposed methods used for the forecasting of total revenue are Autoregressive A.R. (1) with seasonal dummies and the A.R.I.M.A. model. For the multivariate analysis, we used the V.A.R. model for the forecasting of L.T.R. with L.L.S.M., L.C.P.I., L.I.O.P., and seasonal dummies as exogenous variables. These three forecasting methods are considered for the short run forecasting models as suggested by the cointegration analysis in last part of this section; moreover, the variables' causal relationship was tested through the Granger causality test.

The analysis of the undertaken study began with checking the unit root test in the series, for this purpose we employed the Augmented Dickey–Fuller (Dickey & Fuller, 1979, 1981) test. The tax revenues' four components were estimated through the autoregressive (A.R.) model with seasonal dummies (Doan, Litterman, & Sims, 1984; Litterman, 1986; Zhang, 2013). Similarly, we estimated the autoregressive integrated moving average (A.R.I.M.A.) model for all four components (Box & Jenkins, 1976; Nau, 2014; Tiao & Box, 1981; Zhang, 2013). Finally forecasting through the vector autoregressive (V.A.R.) model in a difference form with order L.T.R., L.L.S.M., L.C.P.I., and L.I.O.P. with seasonal dummies (Thomas, Litterman, & Sims, 1984; Tiao & Box, 1981; Zhang, 2013). For the causation and directionality we applied the Granger causality test to check the causality between the variables. Lastly, we employed forecasting by using models and finally compared the robustness of these models through root mean squared error (R.M.S.E.) test (Nau, 2014; Zhang, 2013).

3.3. Unit root test

The first objective of our estimation to check the stationarity of the data series because it is an essential way forward for any higher and sophisticated econometric modelling. For this purpose, we have different unit root tests, but the Augmented Dickey–Fuller (1979, 1981) is the most popular and widely used unit root test worldwide. The generalised equation form of the test is as given follows:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \sum_{i=1}^{n} \alpha \Delta y_t + e_t \tag{1}$$

where: in Equation (1), 't' is the time period, 'y' denotes the time series, 'n' is the optimum number of lags, α_o is a constant and 'e' is an error term.

3.4. Autoregressive model (A.R. model)

The random walk is one of the most important models in econometric time series modelling, this random walk model is also known as the autoregressive model or simply the A.R. (1) process. The following equation is the simplest form of the model:

$$y_t = y_{t-1} + u_t \tag{2}$$

Equation (2) is called the non-drift random walk; if we just add the constant value in above model then it will be converted into a random walk model with drift.

As we know, if the A.R. (p) procedure is stationary, in which it comprises to examine the roots of its characteristic modelling equation. If the $\varphi(L)^{-1}$ touches or converges to zero, then the A.R. (p) model is said to be a stationary model with lag operator notation and can be expressed as follows:

$$\begin{aligned} \varphi(L)y_t &= u_t \\ y_t &= \varphi(L)^{-1} u_t \end{aligned} \tag{3}$$

If we increase the lag length, then the autocorrelations decay to zero. If the A.R. (p) procedure is said to be stationary, the following are the roots of the characteristics equation:

$$1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0 \tag{4}$$

In Equation (4) all roots placed on the outer side of the circle, it is also > 1. It is important to know that if the roots of the random walk model lie on the unit circle then it is known as the non-stationary process of a random walk, and can be computed as follows:

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ y_t &= L y_t + u_t \\ y_t(1 - L) &= u_t \\ 1 - z &= 0 \\ z &= 1 \end{aligned} \tag{5}$$

In the above expression, '$1-z$' is known as the characteristic equation, and the root (z) lies on the unit circle. The other characteristics of the A.R. (p) model are the variance and the mean of A.R. (1) process, and can be expressed as follows:

$$E(y_t) = \frac{\mu}{1 - \varphi_1}, var(y_t) = \frac{\sigma^2}{(1 - \varphi_1^2)} \tag{6}$$

3.5. Autoregressive integrated moving average (A.R.I.M.A.) model

The autoregressive integrated moving average (A.R.I.M.A.) is a generalised form of the autoregressive moving average (A.R.M.A.) model in econometric time series modelling. These two models are known to be the best in forecasting for future values. The equation of the stationary A.R.M.A. (p, q) process can be described as a sequence of random variables (X_t). The equation can be written as follows:

$$X_t - \phi_1 X_{t-1} - L - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + L + \theta_q Z_{t-q} \tag{7}$$

In Equation (7), 'Z_t' is denoted for the sequences of uncorrelated random variables, which have zero mean and constant variance, and can be shown as follows:

$$\{Z_t\} \sim WN(0, \sigma^2), \tag{8}$$

In Equation (9), the 'X' process is known as the A.R.I.M.A. (p, d, q) procedure, if 'd' has a property of non-negative integers, such as, $(1-B)^d$, then 'X' will be known as a causal A.R.M.A. (p, q) method. Hence, the A.R.I.M.A. (p, d, q) processes satisfy the following form of a diferential equation:

$$\phi * (B) \equiv X_t \phi(B)(1 - B)^d X_t = \theta(B) Z_t, \{Z_t\} \sim WN(0, \sigma^2), \tag{9}$$

where: $\phi(z)$ and $\theta(z)$ are known as the polynomials of p and q degrees correspondingly, moreover, $\phi(z) \neq 0$ for $|z| \leq 1$. Since the $\phi*(z)$ having a '0' of order 'd' with $z = 1$.

As the procedure 'X_t' known as stationary if $d = 0$, and in this case the expression reduces to an A.R.M.A. (p, q) method. A fractionally integrated or a long memory process A.R.M.A., and autoregressive fractionally integrated moving average – A.R.F.I.M.A. (p, d, q) methods with $0 < |d| < 0.5$ is called a stationary process, in which autocorrelation function $\rho(k)$ decreasing slowly at lag k, as $k \to \infty$ that also fulfils the property of $\rho(k) \sim Ck^{2d-1}$. Thus, the A.R.F.I.M.A. method also gratifies the differential expression:

$$(1 - B)^d \phi(B) X_t = \theta(B) Z_t \tag{10}$$

where: $\{Z_t\} \sim WN(0, \sigma^2)$,
$$\phi(z) = 1 - \phi_1 z - \dots\dots - \phi_p z^p \text{ satisfying } \phi(z) \neq 0 \tag{11}$$

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \text{ satisfying } satisfying \theta(z) \neq 0 \tag{12}$$

For all the z values such as $|z| \leq 1$, here 'B' is known as the backward shift operator, and $(1-B)^d$ operator is described by the binomial extension as follows:

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_i B^j \text{ with } n_0 = 1 \text{ and } \pi_j = \underset{0 < k \leq j}{\Pi} \frac{k - 1 - d}{k} \text{ for} j = 0, 1, 2, \dots \tag{13}$$

Suppose 'Y_t' is denoted for the time series, and 'y_t' is known as the observed time series, further we also know that there is a steady decrease in the level of 'y_t'. Thus, we can obtain

a new series by differencing time series at lags 1, which is almost constant in its level, and we called it 'X_t'. For this 'X_t' we can tailor the A.R.M.A. (p, q) model.

3.6. Forecasting error

Since there is always error in forecasting, thus, in order to measure that error, and to ensure the utmost accuracy in forecasting we have used three techniques, namely root mean squared error (R.M.S.E.), which is known as the most reliable as compared to other two, i.e., mean absolute percentage error (M.A.P.E.), and mean absolute error (M.A.E.). The mathematical expressions for the three techniques are given as follows:

$$MAE = \frac{\sum_{i=1}^{n} |x_i - \hat{x}_i|}{n} \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n}} \tag{15}$$

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{n} \times 100\% \tag{16}$$

where: in Equation (16), 'n' is signified for the number of predicted values, and 'x_i' is known as the actual observed values, whereas, $\hat{x}_i$ is denoted for the predicted values.

3.7. Granger causality test

It is imperative to establish that whether one variable could forecast another factor or variable in time series analysis (Granger, 1969). Multiple linear regressions are behind the methodology of Granger causality technique. According to Goebel, Roebroeck, Kim, and Formisano (2003) by using the F-statistics on the basis of residuals factor several research studies have already been carried out. As pointed out by the Chen et al. (2009) the ratified trace coefficients could be applied to perform t-test at the level of group statistics. According to McFarlin, Kerr, Green, and Nitschke (2009), the negative trace coefficients are the restrained aftereffects. The pairwise Granger causality expression could be expressed as follows:

$$Y_t = \sum_{n=1}^{p} A_n X_{(t-p)} + \sum_{n=1}^{p} B_n Y_{(t-p)} + CZ_t + E_t \tag{17}$$

$$X_t = \sum_{n=1}^{p} A'_n Y'_{(t-p)} + \sum_{n=1}^{p} B'_n X'_{(t-p)} + C' Z_t + E'_t \tag{18}$$

In Equation (18), 'X' and 'Y' known as the two time series at distinctive time period 't', and 'X_{t-p}' and 'Y_{t-p}' are called two data time series at period 't_p'. Here 'p' used as a number of lagged orders of the time period. Similarly, in the above expression (A_n) and (A_n') are ratified trace coefficients then (B_n) and (B_n') are known as the autoregression model. In the above expressions, we considered 'Z_t' as co-variables at the time period of 't'. The value of F-statistics could be calculated for the normal Wald-test as follows:

$$F = \frac{(RSSR - RSSU)/m}{RSSU/(n-k)} \tag{19}$$

Equation (19) follows the $F_{m,n-k}$ distribution, in which $k=m+n+1$. If the estimated value of 'F' greater than the critical value of 'F', then we reject the null hypothesis and concluded that X_t causes Y_t.

3.8. Vector autoregression model (V.A.R. model)

For the analysis of the multivariate time series, the vector autoregression (V.A.R.) is the obvious choice because this model is very flexible and user-friendly. The vector autoregression (V.A.R.) model is the logical annex of univariate A.R. model for the dynamic multivariate data time series. This model is specifically useful for the description of the dynamic behaviour of time series and their forecasting. Sims (1980) was a famous econometrician, who made this model useful and used this effectively in his research. Moreover, Lutkepohl (1991, 1999), Watson (1994), Nau (2014), Zhang (2013), Tiao and Box (1981), and Waggoner and Zha (1999) are examples of the distinctive and technical use of this model.

We can derive the following form of the V.A.R. model, let $Y_t = (y_{1t}, y_{2t}, ..., y_{nt})$ signify for (n x 1) vectors of the time series variables, and fundamental 'p' lags for the vector autoregressive then the V.A.R. (p) model could be expressed as follows:

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + ... + \Pi_p Y_{t-p} + \varepsilon_t, \tag{20}$$

where: $t = 1, ..., T$

In the above expression, Π_i are the (n x n) matrices of coefficients, and 'ε_t' is as the (n x 1) non-observable mean white noise vector procedure, which is consecutively uncorrelated or independent with the time-invariant covariance matrix Σ. Thus, we can express the bivariate V.A.R. (2) model as follows:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi_{11}^2 & \pi_{12}^2 \\ \pi_{21}^2 & \pi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{21}$$

Or, it can also be written as follows:

$$y_{1t} = c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \pi_{11}^2 y_{1t-2} + \pi_{12}^2 y_{2t-2} + \varepsilon_{1t} \tag{22}$$

$$y_{2t} = c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \pi_{21}^2 y_{1t-1} + \pi_{22}^2 y_{2t-1} + \varepsilon_{2t} \tag{23}$$

In the above expressions cov ($\varepsilon_{1t}, \varepsilon_{2s}$) = σ_{12} for $t = s$, 0 otherwise, and it is important to note that each expression has the same regressors – lagged estimations of y_{1t} and y_{2t}. Thus, it is observed that the V.A.R. (p) technique is just an apparently dissimilar regression (S.U.R.)

model with lagged factors or variables, and in defined terms, it is same as regressors. The V.A.R. (p) can be written as follows for lag operator notation:

$$\Pi(L)Y_t = c + \varepsilon_t \tag{24}$$

In the above equation $\Pi(L) = \text{I}n - \Pi_1 L - ... - \Pi_p L^p$. The model V.A.R. ($p$) is considered to be a firm or stable if the roots placed the outer side of the multifaceted unit circle, which have modulus > 1, or evenly (equivalently). The eigenvalues of the companion matrix can be written as follows:

$$F = \begin{pmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_n \\ \text{I}_n & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \text{I}_n & 0 \end{pmatrix} \tag{25}$$

The above expression F-matrix has modulus < 1. Suppose that procedure has been rearranged in the unlimited historical values, thus the firm V.A.R. (p) procedure is considered to be a stationary, and therefore with the time variant variance, means, and autocovariance. Let the series 'Y_t' is reflected as covariance stationary then the unrestricted means can be expressed as follows:

$$\mu = \left(\text{I}_n - \Pi_1 - \cdots - \Pi_p \right)^{-1} c \tag{26}$$

Thus, the mean-adjusted shape of the model V.A.R. (p) is written as:

$$Y_t - \mu = \Pi_1 \left(Y_{t-1} - \mu \right) + \Pi_2 \left(Y_{t-2} - \mu \right) + \cdots + \Pi_p \left(Y_{t-p} - \mu \right) + \varepsilon_t \tag{27}$$

Since the fundamental V.A.R. (p) technique is so obstructive to show adequately the major characteristics of the data time series. Specifically, additional deterministic expressions, such as seasonal dummies or linear time trends may require showing the time series correctly; moreover, there is a requirement of stochastic exogenous variables. Hence, the generalised form of A.R. (p) model can be expressed with deterministic terminologies and exogenous factors as follows:

$$Y = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \Phi D_t + \Psi X_t + \varepsilon_t \tag{28}$$

In the above expression 'D_t' indicates (1 x 1) deterministic module matrices and 'X_t' signifies (m x 1) the exogenous factors matrix, whereas, Φ and Ψ are parameter matrices.

4. Estimations and results

4.1. Augmented Dickey–Fuller (A.D.F.)

The stationarity of the data is an essential condition to proceed further and employ any time series model. The augmented Dickey–Fuller (1979, 1981) is the most widely used technique to test the stationarity of the data series. A.D.F. test outcomes (Table 2) showed that L.T.R., L.D.T., L.F.E.D., L.S.T., L.C.U.S., L.L.S.M., L.I.O.P., and L.C.P.I. series have level unit root

Table 2. Stationarity of data series (A.D.F. test).

| | A.D.F. Test Statistics | | | |
| | At Level | | At Ist Difference | |
Variables	Value	P-Value	Value	P-Value
LTR	−0.0372	0.9536	−7.4088	0.0000
LDT	−0.5531	0.8774	−12.7051	0.0000
LFED	−1.5192	0.5229	−6.4887	0.0000
LST	−1.8023	0.3792	−5.4803	0.0000
LCUS	0.5526	0.9883	−6.7820	0.0000
LLSM	−0.9335	0.7770	−7.5315	0.0000
LCPI	−1.6212	0.4707	−13.8210	0.0000
LIOP	−0.4176	0.9032	−16.7488	0.0000

Notes: MacKinnon (1996) one-sided p-values: at 1% level: −3.4481, at 5% level: −2.8693, at 10% level: −2.5709.
Source: Authors' calculations.

so the series are transformed and checked at first difference where they become significant and become stationary, consequently these data time series are integrated of order 1 or I(1).

4.2. Graphical representation of stationarity

As exhibited by Figure 3, all the data series are non-stationary at the level, but at the first difference time series of all the eight variables became stationary. Hence, this is also confirmation of the A.D.F. test result that series are integrated of order one or I(1).

4.3. Revenue forecasting: A.R. model with seasonal dummies

Autoregressive of order two with seasonal dummies are estimated to four components of total revenue (T.R.), such as: Direct Tax (D.T.), Sales Tax (S.T.), Federal Excise Duty (F.E.D.), and Customs Duty (C.U.D.). By using this model, variables are forecast for a total of 6 months, that is from January 2017 to June 2017; the Root Mean Square Error (R.M.S.E.) is calculated as 0.3215 (for the detail calculation see Annexure 1). The total of all these forecast values (January–June 2017) for total revenues (T.R.) is presented in Table 3. The total forecast and actual value for FY-2016–17 is P.K.R.3382.95 billion (U.S.$32.22 billion). The actual revenue of P.K.R.1468.32 billion (U.S.$13.98 billion) for the first half of F.Y.-2016–17 (July–Decemeber 2016) has been taken from the Federal Board of Revenue Department of Pakistan. The total forecast revenue of P.K.R.1914.63 billion (U.S.$18.23 billion) for the last six months of FY-2016–17 (January–June 2017) have been forecastthrough the autoregressive seasonal dummies model. The results of the forecast numbers of last six-month are presented in Table 3.

4.4. Revenue forecasting: the A.R.I.M.A. model

The autoregressive integrated moving average (A.R.I.M.A.) model is used for forecasting purposes by taking four components of total revenue (T.R.), such as: Direct Tax (D.T.), Sales Tax (S.T.), Federal Excise Duty (F.E.D.), and Customs Duty (CU.D.). Using this model the variables forecast total revenue of P.K.R.1807.54 billion (U.S.$17.21 billion for the last 6 months of F.Y.-2016–17 that is from January 2017 to June 2017; the Root Mean Square
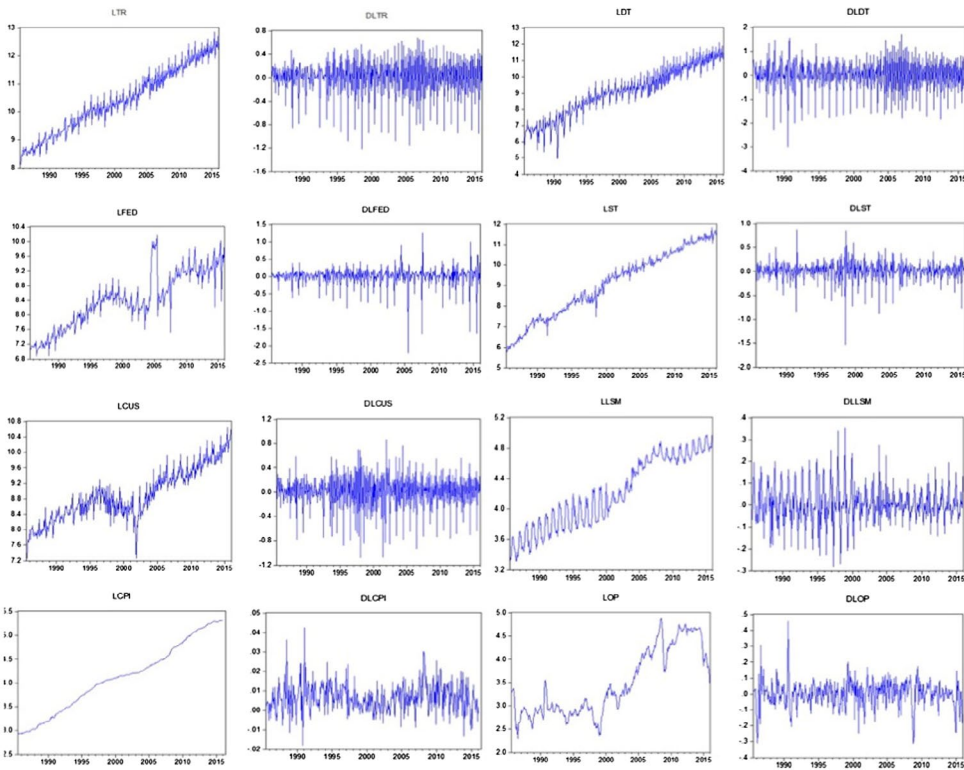
Figure 3. Stationary and non-stationary graphs of series. Source: Authors' calculations.

Error (R.M.S.E.) is calculated as 0.2235 (for the detailed calculation see Annexure 2). The

Table 3. Forecasted values of total revenue using A.R. model.

Period/Variables	D.T.: Direct Tax	S.T.: Sales Tax	F.E.D.: Federal Excise Duty	C.D.: Custom duty	T.R.: Total Revenues
Jan-17	102.56	102.00	19.68	41.04	265.28
Feb-17	88.17	97.79	19.12	38.81	243.90
Mar-17	133.43	112.54	21.26	45.28	312.51
Apr-17	105.59	111.18	22.18	43.31	282.26
May-17	119.41	123.00	23.09	46.88	312.38
Jun-17	256.03	151.34	27.94	63.00	498.30

Note: P.K.R.: Pakistani Rupee in billion.
Source: Authors' calculations.

second half (January–Junue 2017) forecast values of total revenues (T.R.) are presented in Table 4. The total forecast value for FY-2016–17 are P.K.R.3279.88 billion (US\$30.93 billion) which is less than the Federal Board of Revenue (F.B.R.) target set by the government of Pakistan is P.K.R.3521 billion (revised) (U.S.\$33.20 billion). Therefore, results of our study forecast the shortfall of P.K.R.241.12 billion (US\$2.27 billion) as estimated by using the A.R.I.M.A. model.

Table 4. Forecasted values of total revenue using A.R.I.M.A. model.

Period/Variables	D.T.: Direct Tax	S.T.: Sales Tax	F.E.D.: Federal Excise Duty	C.D.: Custom duty	T.R.: Total Revenues
Jan-17	80.44	93.03	14.81	37.01	225.28
Feb-17	86.21	98.24	18.57	34.61	237.63
Mar-17	144.43	106.52	20.10	44.55	315.60
Apr-17	90.38	110.61	14.39	38.30	253.68
May-17	111.49	118.95	23.05	38.42	291.91
Jun-17	235.17	158.10	34.42	55.74	483.44

Note: P.K.R.: Pakistani Rupee in billion.
Source: Authors' calculations.

4.5. Granger causality test

Before applying the Vector Autoregression (V.A.R.) model, it is better to find out if the direction of the variables, which are used in the V.A.R. model, have any causal relationship or not. The Granger Causality test applied between the pairs of L.T.R. with L.L.S.M. L.C.P.I., and L.I.O.P. in difference form with using lag 1 and lag 2. The results are given in Table 5, shows that at lag 2, L.L.S.M. and L.C.P.I. have the two-way causal relation, whereas L.I.O.P. has a one-way causal relationship to L.T.R. (bold figures). Which means that we can apply V.A.R. and the results obtained through V.A.R. may be the better than other models, which are used in this study.

4.6. Revenue forecasting: vector autoregression model (V.A.R.)

V.A.R. of the order D.(L.T.R.) D.(L.L.S.M.) D.(L.C.P.I.) D.(L.I.O.P.) is estimated using the lags 1 to 7 and 9 to14 with seasonal dummies as exogenous variables. Lags are selected by applying the V.A.R. lag exclusion Wald test (for detailed results see Annexure 3 and 4). Then forecast values are obtained from January 2017 to June 2017 that is P.K.R.1840.60 billion (U.S.$17.53 billion) as showed in Table 6, with R.M.S.E. 0.2354. Thus, the total revenues forecasted for F.Y.-2016–17 are P.K.R.3312.92 billion (U.S.$31.24 billion). According to the Federal Board of Revenue (F.B.R.) department, the actual total revenue collected in a first half of the fiscal year (July–December 2016) was P.K.R.1468.32 billion (U.S.$13.98 billion).

Table 5. Pairwise Granger causality test.

Null Hypothesis:	Obs.	F-Statistic	Prob.
Lags: 1			
D.(L.L.S.M.) does not Granger Cause D.(L.T.R.E.V.)	376	1.9811	0.1601
D.(L.T.R.E.V.) does not Granger Cause D.(L.L.S.M.)		7.1027	0.0080
D.(L.C.P.I.) does not Granger Cause D.(L.T.R.E.V.)	376	0.8322	0.3622
D.(L.T.R.E.V.) does not Granger Cause D.(L.C.P.I.)		5.0356	0.0254
D.(L.O.P.) does not Granger Cause D.(L.T.R.E.V.)	376	1.9759	0.1607
D.(L.T.R.E.V.) does not Granger Cause D.(L.O.P.)		0.3796	0.5382
Lags: 2			
D.(L.L.S.M.) does not Granger Cause D.(L.T.R.E.V.)	375	2.8017	0.0620
D.(L.T.R.E.V.) does not Granger Cause D.(L.L.S.M.)		4.0440	0.0183
D.(L.C.P.I.) does not Granger Cause D.(L.T.R.E.V.)	375	14.2482	0.0000
D.(L.T.R.E.V.) does not Granger Cause D.(L.C.P.I.)		3.6550	0.0268
D.(L.O.P.) does not Granger Cause D.(L.T.R.E.V.)	375	3.0473	0.0487
D.(L.T.R.E.V.) does not Granger Cause D.(L.O.P.)		0.6193	0.5389

Source: Authors' calculations.

Table 6. Forecast values of total tax revenue using V.A.R. model.

Period/Variable	T.R.: Total Revenues
Jan-17	229.94
Feb-17	250.06
Mar-17	311.37
Apr-17	260.41
May-17	286.91
Jun-17	501.89

Note: P.K.R.: Pakistani Rupee in billion.
Source: Authors' calculations.

Table 7. Total Revenue Forecasting Error.

Models/Forecasting Error	R.M.S.E.	M.A.E.	M.A.P.E.
A.R. Model with Seasonal dummies	0.3215	0.3127	69.23%
A.R.I.M.A. Model	0.2235	0.2128	38.75%
V.A.R. Model	0.2354	0.2245	45.85%

Source: Authors' calculations.

Therefore, the results of our study forecast the shortfall of P.K.R.208.08 billion (U.S.$1.96 billion) as estimated by using the V.A.R. model against the budgeted total revenue for F.Y.-2016–17.

4.7. Total revenue forecasting error

Since the R.M.S.E., M.A.E., and M.A.P.E. of the A.R.I.M.A. model is minimal for the A.R.I.M.A. model as shown by Table 7 among the other two time series models, which are used to forecast the total revenues for Pakistan, Now, we can say that the forecast value of total revenues for Pakistan for F.Y.-2016–17 is P.K.R.3279.88 billion (U.S.$31.20 billion), which is less than the Federal Board of Revenue (F.B.R.) target set by the government of Pakistan of P.K.R.3521 billion (revised) (U.S.$33.20 billion). Therefore, results of our study forecast a shortfall of P.K.R.241.12 billion (U.S.$2.27 billion) as estimated by using the A.R.I.M.A. model.

4.8. Johansen cointegration technique

We employed a test for cointegration, and for this purpose we employed the Johansen cointegration approach. Outcomes of Table 8 and Table 9 exhibited that there is no evidence of long run relationship amongst the variables because both Max-eigenvalues and Trace statistic values are less than the critical values, thus, it is established that there is only short run association exisiting amongst the variables. Therefore, the short run association has already been estimated in the three forecasting models discussed earlier.

4.9. Actual and forecast tax revenue (January 2017–June 2017)

Outcomes of Table 10 exhibited the actual and forecast tax revenue (January 2017–June 2017) values of the A.R. model with seasonal dummies, A.R.I.M.A. model and V.A.R. model. The individual components of the actual total tax revenue, and total tax revenue values

Table 8. Unrestricted cointegration rank test (trace).

Hypothesised		Trace	0.05	
No. of C.E.(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.0549	35.7515	47.8561	0.4090
At most 1	0.0300	15.0302	29.7971	0.7781
At most 2	0.0104	3.8645	15.4947	0.9142
At most 3	0.0001	0.0223	3.8415	0.8812

Note: Trace test indicates no cointegration at the 0.05 level.
Source: Authors' calculations.
* denotes rejection of the hypothesis at the 0.05 level.
***MacKinnon-Haug-Michelis (1999) p-values.*

Table 9. Unrestricted cointegration rank test (maximum eigenvalue).

Hypothesised		Max-Eigen	0.05	
No. of C.E.(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.0549	20.7213	27.5843	0.2934
At most 1	0.0300	11.1656	21.1316	0.6307
At most 2	0.0104	3.8422	14.2646	0.8756
At most 3	0.0001	0.0223	3.8415	0.8812

Note: Max-eigenvalue test indicates no cointegration at the 0.05 level.
* denotes rejection of the hypothesis at the 0.05 level.
**MacKinnon-Haug-Michelis (1999) *p*-values.
Source: Authors' calculations.

Table 10. Actual and forecast values of tax revenue (January 2017–June 2017).

Forecast Tax Revenue (Jan'17–June'17)	D.T.: Direct Tax	S.T.: Sales Tax	F.E.D.: Federal Excise Duty	C.D.: Custom duty	T.R.: Total Revenues
A.R. Model	805.19	697.85	133.27	278.32	1914.63
A.R.I.M.A. Model	748.12	685.45	125.34	248.63	1807.54
V.A.R. Model	–	–	–	–	1840.58
Actual Tax Revenue (Jan'17–June'17)	751.70	745.82	113.14	278.04	1888.71
Difference b/w Actual & Forecasted (Jan'17–June'17)	D.T.: Direct Tax	S.T.: Sales Tax	F.E.D.: Federal Excise Duty	C.D.: Custom duty	T.R.: Total Revenues
A.R. Model	−53.49	47.97	−20.13	−0.28	−25.92
A.R.I.M.A. Model	3.58	60.37	−12.2	29.41	81.17
V.A.R. Model	–	–	–	–	48.13

Source: Authors' calculations and Federal Board of Revenue report 2016–17.

are depicted in Table 10. The results show that the A.R. model with seasonal dummies are closer to the forecast values as far as the total revenue is concerned, however, the actual values of direct tax revenue are the best forecast values as compared to other models. The wide variation is observed in the sales tax and customs duty, most probably the reason behind the unexpected increase of sales tax and custom duty in Q4 2017. The government has increased G.S.T. on more than 40 items, similarly, increased customs duty on luxury goods, and other important items in Q4 2017 as well. Thus, the gap between the actual and forecast values of sales tax and customs duty is more noticeable. Moreover, the government has revised the total tax revenue target and re-adjusted from P.K.R.3621 billion to P.K.R.3521 billion in Q4 2017 (F.B.R., 2016–17). Hence, this deviation has occurred due to these two

components, otherwise, the A.R.I.M.A. model has best forecast the values of the direct tax and federal excise duty.

5. Discussions and conclusions

5.1. Discussions

According to the results, the A.R.I.M.A. model shows the best forecasting values of total tax revenue because the value of the R.M.S.E. is minimum as compared to the A.R. model with seasonal dummies and the V.A.R. model. The results of the study further showed that the major portion of tax revenue is generated by indirect taxes. The results are consistent with previous research (Aamir et al., 2001; Chaudhry & Munir, 2010; Himani, 2016), which studied the elements of tax collection in the case of India and Pakistan and concluded that in Pakistan, the ratio of tax collection is significant through indirect taxes. They concluded that the indirect taxes have widened the gap between the rich and poor, and further exploits the vulnerability of the working class. In our research, we used the total tax revenue as a dependent variable, and direct and indirect taxes as independent variables, previous studies also used the same variables (Aamir et al., 2001; Himani, 2016; Husain & Qasim, 2007; Patoli et al., 2012; Samuel, 2014). They concluded the same results that indirect taxes are more prominent compared with direct taxes except for the findings of Himani (2016) in which he concluded that the total tax revenue is more prominent and dominant than the direct taxes in the case of India. The results also showed unidirectional causalities and concluded that there is Granger causality from the total tax revenue to large-scale manufacturing, and the total tax revenue to consumer price index in lag 1. However, lag 2 revealed a bidirectional causality between the total tax revenue and large-scale manufacturing, and consumer price index, whereas, the crude oil prices Granger causes to the total tax revenue. Thus, these results concluded that excessive tax collection causes inflation, and because of inflation the government is imposing further taxes on the working class of Pakistan in the form of indirect taxes. Similarly, the government of Pakistan collects more taxes from the large-scale manufacturing sector or in other words the growth of L.S.M. increases the tax net, which further increases tax collections. These results are also consistent with previous research studies (Chaudhry & Munir, 2010; Daba, 2015; Daba & Mishra, 2014; Rasheed, 2006; Samuel, 2014; Sunday, 2015). The study further showed that there is a Granger cause from crude oil prices to the total tax revenue. Since Pakistan is an oil importing country, and taxes are 40% on petroleum products, thus, these results are very much aligned with the true picture or ground realities, and also consistent with the previous research studies such as Chen et al. (2009), Eugene and Chineze (2015), Hassan and Tahmina (2012), and Karagöz (2013).

5.2. Conclusions

The results of the study demonstrated the effectiveness of three different time series models, moreover, the precise results of forecasting, total tax revenue for the F.Y.-2016–17, which lay down the foundations for proper policy-making by the government of Pakistan. The results of this study revealed that among these models A.R.I.M.A. model gives better-forecast values for the total tax revenues of Pakistan, Now, we can say that the forecast value of

total revenues for Pakistan for F.Y.-2016–17 is P.K.R.3279.88 billion (U.S.$30.92 billion). Hence, the forecast tax revenue is less than the Federal Board of Revenue (F.B.R.) target set by the government of Pakistan that is P.K.R.3521 billion (U.S.$33.20 billion). Moreover, we have equated the actual and forecast tax revenues; the comparison revealed that the A.R. model with seasonal dummies have forecast the results closer to the actual results of total tax revenue. However, the actual values of direct tax revenue are the best-forecast values as compared to the other models. Wide variation is observed in the sales tax and custom duty, most probably the reason behind the unexpected increase of sales tax and custom duty in Q4 2017. The government has increased G.S.T. on more than 40 items, similarly, increased customs duty on luxury goods, and other important items in Q4 2017. Thus, the gap between the actual and forecast values of sales tax and customs duty is more noticeable. Hence, this deviation has occurred due to these two components, otherwise, the A.R.I.M.A. model has best forecast the values of the direct tax and federal excise duty. Therefore, the results of our study forecast the shortfall of P.K.R.341.12 billion (U.S.$2.27 billion) as estimated by using the A.R.I.M.A. model. Therefore, the government should take appropriate and immediate measures in order to fix this anticipated problem. A very important finding of our research is that the government is collecting three times more taxes from indirect means compared to direct taxes. This is quite unusual if we compare the developing and developed economies, where the major chunk of taxes is being generated from the direct taxes. This shows the weak tax collection system in Pakistan because the government has failed to broaden the tax net of income tax from the individuals, merchants, individual services, real estate, stock market, and the agriculture sector. Therefore, the government achieves its tax revenue budget by imposing indirect taxes in the form of G.S.T.; gain tax, levies tax, withholding tax, customs duty, federal excise duty etc. The tax to G.D.P. ratio is also very alarming, and was between 8.5% to 9.7% for the last ten years, it is not comparable to the developed economies, even very low compared to the other South Asian developing countries such as India, Bangladesh, Sri Lanka, and China etc. Therefore, there is a strong need to improve the tax collection system and the government should bring new tax reforms, and broaden the tax net.

5.3. Limitations of the study

The undertaken study has certain limitations, for instance, we did not incorporate exogenous factors like political instability, local and international economic turmoil, terrorism, and government policies, thus, future research studies may incorporate these exogenous factors. Moreover, these exogenous variables may cause structural breaks in the considered time series. Thus, this research study has not incorporated these structural breaks; this is another limitation of the study. Future studies may incorporate these structural breaks. Finally, other macroeconomic indicators such as inflation rate, exchange rate, and unemployment have not been incorporated in undertaken study; thus, it might be another potential area for future studies. In Pakistan the segregated data for total revenues is not available in monthly frequency that is why we use the total revenues such as direct tax, sales tax, custom duty and excise tax, this is another limitation of undertaken research. For the case of Pakistan the model established after trying all possible macroeconomic variables in relation with total revenues and the best model is estimated in this study (since G.D.P. in monthly or quarterly

is published by the statistical agency of Pakistan which is also one of the limitation of this study that is why we use L.S.M. as best proxy of G.D.P.).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aamir, M., Qayuum, A., Nasir, A., Hassain, S., Khan, K. I., & Butt (2001). Determinants of tax revenue: Comparative study of direct taxes and indirect taxes of Pakistan and India. *International Journal of Business and Social Sciences, 2*, 171–178.

Bagshaw, M. L. (1987). *Univariate and multivariate ARIMA versus vector* (Working Paper 8706). The Federal Reserve Bank of Cleve1and. Retrieved March 16, 2017, from https://fraser.stlouisfed.org/files/docs/historical/frbclev/wp/frbclv_wp1987-06.pdf

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco, CA: Holden-Day.

Čábelková, I., & Strielkowski, W. (2013). Is the level of taxation a product of culture? A cultural economics approach. *Society and Economy, 35*, 513–529.

Chaudhry, M. G. (2001). Theory of optimal taxation and current tax policy in Pakistan's agriculture. *Pakistan Development Review, 40*, 489–502.

Chaudhry, I. S., & Munir, F. (2010). Determinants of low tax revenue in Pakistan. *Pakistan Journal of Social Sciences, 30*, 439–452.

Chen, G., Hamilton, J. P., Thomason, M. E., Gotlib, I. H., Saad, Z. S., & Cox, R. W. (2009). Granger causality via vector auto-regression tuned for fMRI data analysis. *Proceedings of the International Society for Magnetic Resonance in Medicine, 17*, 1718.

Daba, T. A. (2015). *Determinants of tax revenue in Ethiopia* (Master's thesis). Addis Ababa University, Ethiopia. Retrieved March 6, 2017, from http://etd.aau.edu.et/bitstream/123456789/6819/1/Tesfaye%20Alemayehu.PDF

Daba, D., & Mishra, D. K. (2014). Tax reforms and tax revenues performance in Ethiopia. *Journal of Economics and Sustainable Development, 5*, 11–19.

Das-Gupta, A. (2011). *An assessment of the revenue impact of the state level VAT in India* (*Working Paper*). Goa Institute of Management, India. doi:10.2139/ssrn.1953165

Dickey, D. A., & Fuller, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*, 427–431. doi:10.1080/01621459.1979.10482531

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*, 1057–1072. doi:10.2307/1912517

Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews, 3*, 1–100. doi:10.1080/07474938408800053

Dowling, G. R. (2014). The curious case of corporate tax avoidance: Is it socially irresponsible? *Journal of Business Ethics, 124*, 173–184.

Economic Survey of Pakistan. (2015–16). *Overview of the Economy* (Various issues). Retrieved March 9, 2017, from www.finance.gov.pk/survey/chapters_16/Overview_of_the_Economy.pdf

Eugene, N., & Chineze, E. A. (2015). Productivity of the Nigerian tax system (1994–2013). *International Journal of Business Administration, 6*, 30–40. doi:10.5430/ijba.v6n4p30

Eugene, N., & Chineze, E. A. (2016). Effect of tax policy on economic growth in Nigeria (1994–2013). *International Journal of Business Administration, 7*, 50–58. doi:10.5430/ijba.v7n1p50

Federal Board of Revenue. (2015–16). *Finance and economic survey.* Retrieved March 15, 2017, from http://www.pkrevenue.com/finance/economic-survey-20152016-fbrs-tax-to-gdp-remained-narrow-despite-significant-revenue-growth/

Goebel, R., Roebroeck, A., Kim, D. S., & Formisano, E. (2003). Investigating directed cortical interactions in time–resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging, 21*, 1251–1261. doi:10.1016/j.mri.2003.08.026

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424–438. doi:10.2307/1912791

Hassan, H. M., & Tahmina, A. (2012). Impact of public debt burden of economic growth: Evidence from Bangladesh: A VAR apprach. *Journal of Finance and Banking, 10*, doi:10.2139/ssrn.2152592

Himani. (2016). Determinants of tax revenue in India. *International Journal of Research in Economics and Social Sciences, 6*, 161–170.

Husain, F., & Qasim, M. A. (2007). *The relationship between federal government revenue and expenditure in Pakistan* (Working Paper). Pakistan Institute of Development Economics.

Javid, A. Y., Arif, U., & Arif, A. (2011). Economic, political and institutional determinants of budget deficits volatility in selected Asian countries. *The Pakistan Development Review, 50*, 649–662.

Karagöz, K. (2013). Determinants of tax revenue: Does sectorial composition matter? *Journal of Finance, Accounting and Management, 4*, 50–63.

Kenny, L. W., & Winer, S. L. (2006). Tax systems in the world: An empirical investigation into the importance of tax bases, administration costs, scale and political regime. *International Tax and Public Finance, 13*, 181–215. doi:10.1007/s10797-006-3564-7

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregression – Five years of experience. *Journal of Business & Economic Statistics, 4*, 25–38.

Lutkepohl, H. (1991). *Introduction to multiple time series analysis*. Springer-Verlag, Berlin. doi:10.1007/978-3-662-02691-5

Lutkepohl, H. (1999). *Vector autoregressions* (Unpublished manuscript). Institut fur Statistik und Okonometrie, Humboldt-Universitat zu Berlin.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics, 11*, 601–618. doi:10.1002/(sici)1099-1255(199611)11:6<601::aid-jae417>3.0.co;2-t

MacKinnon, J. G., Haug, A. A., & Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics, 14*, 563–577.

McFarlin, D. R., Kerr, D. L., Green, D. E., & Nitschke J. B. (2009). The effect of controllability on context dependent granger causality in snake phobia fMRI data at 400-ms resolution. *Society for Neuroscience, 289*,15, Chicago, IL.

Mehrara, M., Pahlavani, M., & Elyasi, Y. (2011). Government revenue and government expenditure nexus in Asian countries: Panel cointegration and causality. *International Journal of Business and Social Science, 2*, 199–207.

Myles, G. (2000). Taxation and economic growth. fiscal studies. *The Journal of Institute for Fiscal Studies, 21*, 141–168.

Nanthakumar, L., Kogid, M., Sakami, M. N., & Muhamad, S. (2011). Tax revenue and government spending constraints: Empirical evidence from Malaysia. *China-USA Business Review, 10*, 779–784.

Nau, R. (2014). *Forecasting with moving averages* (Working Paper). Fuqua School of Business, Duke University. Retrieved April 16, 2017, from https://people.duke.edu/~rnau/Notes_on_forecasting_with_moving_averages–Robert_Nau.pdf

Palacios, M., & Harischandra, K. (2008). The impact of taxes on economic behavior. In J. Clemens (Ed.), *The impact and cost of taxation in Canada: The case for flat tax reform* (pp. 3–31). The Fraser Institute. Retrieved March 12, 2017, from http://www.fraserinstitute.org/researchnews/display.aspx?id=13518

Patoli, A. Q., Zarif, T., & Syed, N. A. (2012). Impact of inflation on taxes in Pakistan: An emperical study of 2000–2010 period. *Journal of Management and Social Sciences, 8*, 31–41.

Poulson, B. W., & Kaplan, J. G. (2008). State income taxes and economic growth. *Cato Journal, 28*, 53–71.

Rasheed, F. (2006). An analysis of the tax buoyancy rates in Pakistan. *Market Forces-Journal of Management, Business & Economics, 2*. Retrieved March 10, 2017, from http://www.pafkiet.edu.pk/marketforces/index.php/marketforces/article/view/107/109

Samuel, O. L. (2014). *The impact of value added tax on revenue generation in Nigeria* (Working Paper). Olabisi Onabanjo University, Nigeria. https://doi.org/10.2139/ssrn.2513207

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48*, 1–48. doi:10.2307/1912017

Strielkowski, W., & Čábelková, I. (2015). Religion, culture, and tax evasion: Evidence from the Czech Republic. *Religions, 6*, 657–669.

Sunday, C. N. (2015). The implications of tax revenue on the economic development of Nigeria. *Issues in Business Management and Economics, 3*, 74–80 doi:10.15739/IBME.2014.017

Tanko, M. (2015). *The impact of personal income tax on state governments revenue profile in Nigeria* (Working Paper). Kaduna State University, Nigeria. doi:10.2139/ssrn.2690402

Thomas, D., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews, 3*, 1–100.

Tiao, G. C., & Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association, 76*, 802–816.

Torgler, B., & Schneider, F. (2007). What shapes attitudes toward paying taxes? Evidence from multicultural European countries. *Social Science Quarterly, 88*, 443–470.

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics, 81*, 639–651. doi:10.1162/003465399558508

Watson, M. W. (1994). Vector autoregressions and cointegration. In R. F. Engle & D. McFadden (Eds.), *Handbook of econometrics* (Volume IV). Amsterdam: Elsevier Science. doi:10.1016/s1573-4412(05)80016-9

Zhang, H. (2013). *Modeling and forecasting regional G.D.P. in Sweden using autoregressive models* (Master's thesis in Micro Data Analysis). School for Technology and Business Studies, Dalarna University.

Annexure 1.

A.R. model with seasonal dummies

Dependent Variable: D.(L.C.U.S.)
Method: Least Squares
Date: 06/13/17 Time: 07:58
Sample (adjusted): 1985M10 2017M04
Included observations: 379 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.1180	0.0157	7.4919	0.0000
D(LCUS(−1))	−0.6193	0.0497	−12.4690	0.0000
D(LCUS(−2))	−0.3235	0.0364	−8.8865	0.0000
@SEAS(1)	−0.1190	0.0348	−3.4230	0.0007
@SEAS(2)	−0.1756	0.0339	−5.1787	0.0000
@SEAS(4)	−0.1078	0.0365	−2.9553	0.0033
@SEAS(6)	0.2303	0.0329	6.9955	0.0000
@SEAS(7)	−0.5326	0.0404	−13.1814	0.0000
@SEAS(8)	−0.2796	0.0423	−6.6150	0.0000
@SEAS(10)	−0.0738	0.0370	−1.9931	0.0470
@SEAS(11)	−0.1598	0.0340	−4.6981	0.0000
R-squared	0.7410	Mean dependent var		0.0082
Adjusted R-squared	0.7339	S.D. dependent var		0.3151
S.E. of regression	0.1626	Akaike info criterion		−0.7670
Sum squared resid.	9.7240	Schwarz criterion		−0.6527
Log likelihood	156.35	Hannan-Quinn criter.		−0.7217
F-statistic	105.26	Durbin-Watson stat		2.0404
Prob.(F-statistic)	0.0000			

Dependent Variable: D.(L.S.T.)
Method: Least Squares
Date: 06/13/17 Time: 07:59
Sample (adjusted): 1985M10 2017M04
Included observations: 379 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0528	0.0119	4.4474	0.0000
D(LST(−1))	−0.6896	0.0477	−14.4523	0.0000
D(LST(−2))	−0.3325	0.0408	−8.1441	0.0000

Dependent Variable: D.(L.D.T.)
Method: Least Squares
Date: 06/13/17 Time: 07:58
Sample (adjusted): 1985M10 2017M04
Included observations: 379 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.3948	0.0398	9.9083	0.0000
D(LDT(−1))	−0.6235	0.0395	−15.7813	0.0000
D(LDT(−2))	−0.4883	0.0404	−12.0903	0.0000
@SEAS(1)	−0.5185	0.0658	−7.8772	0.0000
@SEAS(2)	−0.5102	0.0726	−7.0231	0.0000
@SEAS(3)	−0.3540	0.0716	−4.9480	0.0000
@SEAS(4)	−0.4521	0.0654	−6.9131	0.0000
@SEAS(5)	−0.2119	0.0688	−3.0802	0.0022
@SEAS(6)	0.3286	0.0671	4.8950	0.0000
@SEAS(7)	−1.4985	0.0679	−22.0791	0.0000
@SEAS(8)	−0.6656	0.1001	−6.6519	0.0000
@SEAS(9)	−0.2378	0.0946	−2.5128	0.0124
@SEAS(11)	−0.2772	0.0715	−3.8776	0.0001
R-squared	0.8316	Mean dependent var		0.0126
Adjusted R-squared	0.8261	S.D. dependent var		0.7195
S.E. of regression	0.3001	Akaike info criterion		0.4640
Sum squared resid	32.9526	Schwarz criterion		0.5991
Log likelihood	−74.93	Hannan-Quinn criter.		0.5176
F-statistic	150.60	Durbin-Watson stat		2.0240
Prob.(F-statistic)	0.0000			

Dependent Variable: D.(L.F.E.D.)
Method: Least Squares
Date: 06/13/17 Time: 07:59
Sample (adjusted): 1985M10 2017M04
Included observations: 379 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0746	0.0134	5.5709	0.0000
D(LFED(−1))	−0.3516	0.0373	−9.4238	0.0000
D(LFED(−2))	−0.1900	0.0369	−5.1428	0.0000

(Continued)

Annexure 1. (*Continued*)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SEAS(1)	-0.0582	0.0285	-2.0386	0.0422
@SEAS(2)	-0.1068	0.0287	-3.7190	0.0002
@SEAS(5)	0.0871	0.0292	2.9841	0.0030
@SEAS(6)	0.2216	0.0293	7.5700	0.0000
@SEAS(7)	-0.3178	0.0313	-10.1488	0.0000
@SEAS(8)	-0.1511	0.0343	-4.4042	0.0000
@SEAS(10)	0.0614	0.0299	2.0522	0.0409

R-squared	0.6233	Mean dependent var	0.0151
Adjusted R-squared	0.6141	S.D. dependent var	0.2335
S.E. of regression	0.1451	Akaike info criterion	-0.9974
Sum squared resid	7.7638	Schwarz criterion	-0.8935
Log likelihood	199.01	Hannan-Quinn criter.	-0.9562
F-statistic	67.84	Durbin-Watson stat	2.1163
Prob(F-statistic)	0.0000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SEAS(1)	-0.1039	0.0399	-2.6050	0.0096
@SEAS(2)	-0.0943	0.0399	-2.3661	0.0185
@SEAS(6)	0.1387	0.0408	3.4009	0.0007
@SEAS(7)	-0.7138	0.0414	-17.2315	0.0000

R-squared	0.5765	Mean dependent var	0.0070
Adjusted R-squared	0.5697	S.D. dependent var	0.3234
S.E. of regression	0.2121	Akaike info criterion	-0.2451
Sum squared resid	16.7377	Schwarz criterion	-0.1723
Log likelihood	53.44	Hannan-Quinn criter.	-0.2162
F-statistic	84.41	Durbin-Watson stat	2.0826
Prob(F-statistic)	0.0000		

Source: Authors' calculations.

Annexure 2.

A.R.I.M.A. model

Dependent Variable: D.(L.C.U.S.)
Method: A.R.M.A. Maximum Likelihood (OPG - BHHH)
Date: 06/13/17 Time: 08:05
Sample: 1985M08 2017M04
Included observations: 381
Convergence achieved after 56 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0082	0.0032	2.5593	0.0109
AR(1)	-0.2321	0.0730	-3.1800	0.0016
AR(2)	-0.1736	0.0773	-2.2441	0.0254
AR(4)	-0.2146	0.0657	-3.2666	0.0012
AR(5)	-0.1438	0.0576	-2.4968	0.0130
AR(8)	-0.0600	0.0583	-1.0290	0.3041
AR(11)	-0.0469	0.0398	-1.1771	0.2399

Dependent Variable: D.(L.D.T.)
Method: A.R.M.A. Maximum Likelihood (OPG - BHHH)
Date: 06/13/17 Time: 08:06
Sample: 1985M08 2017M04
Included observations: 381
Convergence achieved after 187 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0130	0.0033	3.9652	0.0001
AR(1)	-0.0093	0.0310	-0.3003	0.7641
AR(2)	-0.0447	0.0360	-1.2404	0.2156
AR(4)	-0.0177	0.0349	-0.5070	0.6125
AR(5)	-0.0152	0.0343	-0.4425	0.6584
AR(7)	-0.0197	0.0379	-0.5195	0.6038
AR(8)	-0.0139	0.0340	-0.4080	0.6835

(*Continued*)

Annexure 2. (Continued)

	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0.4955	0.0424	11.6741	0.0000
MA(1)	−0.4240	0.0848	−5.0007	0.0000
MA(2)	0.0154	0.0856	0.1801	0.8572
MA(3)	−0.0918	0.0686	−1.3371	0.1820
MA(4)	0.1415	0.0831	1.7026	0.0895
MA(6)	−0.2138	0.0534	−4.0038	0.0001
SIGMASQ	0.0309	0.0018	17.3072	0.0000

R-squared	0.6870	Mean dependent var	0.0087
Adjusted R-squared	0.6759	S.D. dependent var	0.3147
S.E. of regression	0.1792	Akaike info criterion	−0.5483
Sum squared resid	11.7799	Schwarz criterion	−0.4034
Log likelihood	118.46	Hannan-Quinn criter.	−0.4908
F-statistic	61.96	Durbin-Watson stat	1.9965
Prob(F-statistic)	0.0000		

Dependent Variable: D.(L.S.T.)
Method: A.R.M.A. Maximum Likelihood (OPG - BHHH)
Date: 06/13/17 Time: 08:06
Sample: 1985M08 2017M04
Included observations: 381
Convergence not achieved after 500 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0150	0.0018	8.2398	0.0000
AR(1)	−0.1316	0.0723	−1.8203	0.0695
AR(5)	0.0190	0.0331	0.5756	0.5653
AR(6)	−0.0243	0.0311	−0.7824	0.4345
AR(7)	0.0197	0.0335	0.5866	0.5579
AR(9)	−0.0135	0.0343	−0.3949	0.6932

	Coefficient	Std. Error	t-Statistic	Prob.
AR(9)	−0.0129	0.0405	−0.3189	0.7500
AR(10)	−0.0291	0.0382	−0.7618	0.4467
AR(11)	−0.0001	0.0401	−0.0025	0.9980
AR(12)	0.8503	0.0283	30.0942	0.0000
MA(1)	−0.7940	0.0520	−15.2701	0.0000
MA(2)	−0.1052	0.0717	−1.4670	0.1433
MA(3)	0.0967	0.0773	1.2504	0.2120
MA(4)	−0.1616	0.0599	−2.6957	0.0074
MA(6)	0.1958	0.0600	3.2652	0.0012
MA(7)	−0.1862	0.0783	−2.3766	0.0180
MA(8)	0.0673	0.0804	0.8363	0.4036
MA(9)	0.0742	0.0810	0.9164	0.3601
MA(10)	−0.1240	0.0809	−1.5336	0.1260
MA(11)	−0.1701	0.0621	−2.7383	0.0065
MA(13)	0.1763	0.0466	3.7858	0.0002
SIGMASQ	0.0654	0.0033	19.6958	0.0000

R-squared	0.8730	Mean dependent var	0.0153
Adjusted R-squared	0.8652	S.D. dependent var	0.7185
S.E. of regression	0.2638	Akaike info criterion	0.2710
Sum squared resid	24.9207	Schwarz criterion	0.5090
Log likelihood	−28.62	Hannan-Quinn criter.	0.3654
F-statistic	111.84	Durbin-Watson stat	1.9592
Prob(F-statistic)	0.0000		

Dependent Variable: D.(L.F.E.D.)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 06/13/17 Time: 08:06
Sample: 1985M08 2017M04
Included observations: 381
Convergence not achieved after 500 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0073	0.0124	0.5891	0.5562
AR(1)	−0.8153	0.0618	−13.2030	0.0000
AR(2)	−0.4158	0.0574	−7.2463	0.0000
AR(3)	−0.1464	0.0396	−3.6980	0.0003
AR(5)	0.0715	0.0396	1.8065	0.0717
AR(6)	0.0506	0.0345	1.4654	0.1437

(Continued)

Annexure 2. (*Continued*)

	Coefficient	Std. Error	t-Statistic	Prob.
AR(10)	−0.0230	0.0361	−0.6357	0.5254
AR(11)	0.0081	0.0377	0.2144	0.8304
AR(12)	0.9349	0.0296	31.6145	0.0000
AR(13)	0.1309	0.0642	2.0375	0.0423
MA(1)	−0.6257	0.1147	−5.4564	0.0000
MA(5)	−0.0950	0.0884	−1.0753	0.2829
MA(6)	0.2164	0.0779	2.7793	0.0057
MA(7)	−0.2002	0.0794	−2.5194	0.0122
MA(9)	0.0837	0.0705	1.1863	0.2363
MA(10)	−0.0330	0.0763	−0.4330	0.6653
MA(11)	−0.0968	0.0926	−1.0445	0.2969
MA(12)	−0.6208	0.0842	−7.3699	0.0000
MA(13)	0.3720	0.0639	5.8229	0.0000
SIGMASQ	0.0189	0.0017	10.9502	0.0000
R-squared	0.6522	Mean dependent var		0.0149
Adjusted R-squared	0.6339	S.D. dependent var		0.2332
S.E. of regression	0.1411	Akaike info criterion		−0.9992
Sum squared resid	7.1863	Schwarz criterion		−0.7922
Log likelihood	210.34	Hannan-Quinn criter.		−0.9170
F-statistic	35.63	Durbin-Watson stat		1.9877
Prob(F-statistic)	0.0000			

	Coefficient	Std. Error	t-Statistic	Prob.
AR(10)	−0.2291	0.0421	−5.4479	0.0000
AR(11)	−0.7003	0.0526	−13.3176	0.0000
MA(1)	0.4576	0.0661	6.9258	0.0000
MA(6)	0.0837	0.0651	1.2861	0.1992
MA(10)	0.0830	0.0680	1.2209	0.2229
MA(11)	0.6190	0.0687	9.0052	0.0000
SIGMASQ	0.0538	0.0020	26.9283	0.0000
R-squared	0.4818	Mean dependent var		0.0072
Adjusted R-squared	0.4650	S.D. dependent var		0.3225
S.E. of regression	0.2359	Akaike info criterion		0.0145
Sum squared resid	20.4847	Schwarz criterion		0.1490
Log likelihood	10.24	Hannan-Quinn criter.		0.0678
F-statistic	28.52	Durbin-Watson stat		1.8669
Prob(F-statistic)	0.0000			

Source: Authors' calculations.

Annexure 3.

Vector Autoregression Estimates
Date: 06/13/17 Time: 08:10
Sample (adjusted): 1986M10 2017M04
Included observations: 367 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	D.(L.T.R.E.V.)	D.(L.L.S.M.)	D.(L.C.P.I.)	D.(L.O.P.)
D(LTREV(-1))	−0.8410	−0.0640	0.0035	−0.0404
	−0.0562	−0.0308	−0.0040	−0.0465
	[−14.9748]	[−2.07712]	[0.89455]	[−0.86958]
D(LTREV(-2))	−0.5883	−0.0752	0.0062	−0.0217
	−0.0731	−0.0401	−0.0052	−0.0605
	[−8.04586]	[−1.87597]	[1.20063]	[−0.35902]
D(LTREV(-3))	−0.2937	−0.0207	0.0087	0.0119

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates

	(0.0770)	(0.0422)	(0.0054)	(0.0638)
	[−3.81282]	[−0.49019]	[1.59208]	[0.18645]
D(LTREV(−4))	−0.2929	0.0071	0.0062	0.0126
	(0.0761)	(0.0417)	(0.0054)	(0.0630)
	[−3.84751]	[0.16963]	[1.14541]	[0.19993]
D(LTREV(−5))	−0.3663	−0.0539	0.0043	−0.0671
	(0.0758)	(0.0416)	(0.0054)	(0.0628)
	[−4.83078]	[−1.29632]	[0.79867]	[−1.06843]
D(LTREV(−6))	−0.1371	−0.0627	−0.0031	−0.0614
	(0.0710)	(0.0390)	(0.0050)	(0.0588)
	[−1.92933]	[−1.60913]	[−0.61948]	[−1.04368]
D(LTREV(−7))	−0.0868	−0.0239	0.0005	0.0060
	(0.0550)	(0.0302)	(0.0039)	(0.0455)
	[−1.57758]	[−0.79087]	[0.13018]	[0.13133]
D(LTREV(−9))	−0.0193	0.0166	−0.0027	0.0227
	(0.0550)	(0.0301)	(0.0039)	(0.0455)
	[−0.35114]	[0.55094]	[−0.70049]	[0.49927]
D(LTREV(−10))	−0.0801	0.0201	−0.0060	0.0072
	(0.0697)	(0.0382)	(0.0049)	(0.0577)
	[−1.14906]	[0.52636]	[−1.20965]	[0.12420]
D(LTREV(−11))	−0.2172	−0.0358	−0.0056	0.0664
	(0.0741)	(0.0406)	(0.0052)	(0.0614)
	[−2.93062]	[−0.88196]	[−1.06771]	[1.08281]
D(LTREV(−12))	−0.0170	−0.0578	−0.0073	0.0510
	(0.0749)	(0.0411)	(0.0053)	(0.0620)
	[−0.22716]	[−1.40852]	[−1.38329]	[0.82308]
D(LTREV(−13))	0.0000	−0.0403	−0.0040	−0.0198
	(0.0701)	(0.0384)	(0.0050)	(0.0580)
	[−4.9e-05]	[−1.04801]	[−0.80568]	[−0.34192]
D(LTREV(−14))	−0.0814	0.0329	−0.0080	−0.0104
	(0.0546)	(0.0300)	(0.0039)	(0.0452)
	[−1.48947]	[1.09814]	[−2.07825]	[−0.22968]
D(LLSM(−1))	0.1814	−0.3448	0.0020	0.0240
	(0.1068)	(0.0586)	(0.0075)	(0.0884)
	[1.69822]	[−5.88836]	[0.26380]	[0.27155]
D(LLSM(−2))	0.2170	−0.2142	−0.0007	0.0387
	(0.1087)	(0.0596)	(0.0077)	(0.0900)
	[1.99594]	[−3.59317]	[−0.08869]	[0.42961]

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates

D(LLSM(-3))	0.2803	-0.1084	0.0091	0.0627
	-0.1049	-0.0575	-0.0074	-0.0869
	[2.67135]	[-1.88336]	[1.23063]	[0.72163]
D(LLSM(-4))	0.2459	-0.1835	0.0026	0.0071
	-0.1050	-0.0576	-0.0074	-0.0869
	[2.34257]	[-3.18828]	[0.35396]	[0.08156]
D(LLSM(-5))	0.2150	-0.1556	-0.0018	-0.0966
	-0.1115	-0.0612	-0.0079	-0.0923
	[1.92729]	[-2.54409]	[-0.22969]	[-1.04589]
D(LLSM(-6))	0.0800	-0.1988	-0.0004	-0.0093
	-0.1076	-0.0590	-0.0076	-0.0891
	[0.74302]	[-3.36928]	[-0.04901]	[-0.10410]
D(LLSM(-7))	0.2657	-0.1730	0.0072	0.0375
	-0.1047	-0.0574	-0.0074	-0.0866
	[2.53866]	[-3.01419]	[0.97086]	[0.43293]
D(LLSM(-9))	0.3064	0.0016	0.0214	0.1364
	-0.1043	-0.0572	-0.0074	-0.0864
	[2.93637]	[0.02828]	[2.90252]	[1.57927]
D(LLSM(-10))	0.1285	-0.0692	0.0022	0.1759
	-0.1073	-0.0588	-0.0076	-0.0888
	[1.19773]	[-1.17706]	[0.29673]	[1.98120]
D(LLSM(-11))	0.3032	0.0494	0.0115	0.0621
	-0.1112	-0.0610	-0.0079	-0.0921
	[2.72606]	[0.80956]	[1.45824]	[0.67399]
D(LLSM(-12))	0.0987	0.4058	0.0031	-0.0272
	-0.1060	-0.0581	-0.0075	-0.0878
	[0.93059]	[6.98129]	[0.40792]	[-0.30947]
D(LLSM(-13))	-0.1745	0.1559	0.0087	-0.0279
	-0.1138	-0.0624	-0.0080	-0.0942
	[-1.53403]	[2.49929]	[1.08923]	[-0.29671]
D(LLSM(-14))	0.1247	0.1130	0.0057	0.1380
	-0.1054	-0.0578	-0.0074	-0.0872
	[1.18299]	[1.95586]	[0.77081]	[1.58243]
D(LCPI(-1))	1.0870	-0.3171	0.0904	0.2098
	-0.8148	-0.4468	-0.0575	-0.6744
	[1.33399]	[-0.70973]	[1.57121]	[0.31113]
D(LCPI(-2))	1.3332	-0.3449	0.0893	0.3583
	-0.8065	-0.4422	-0.0569	-0.6675
	[1.65308]	[-0.78003]	[1.56883]	[0.53677]

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates

D(LCPI(-3))	0.1781	0.2670	0.0674	0.3611
	-0.8158	-0.4473	-0.0576	-0.6752
	[0.21827]	[0.59692]	[1.16988]	[0.53488]
D(LCPI(-4))	-0.1841	-0.0434	0.0645	-0.2390
	-0.8113	-0.4449	-0.0573	-0.6715
	[-0.22687]	[-0.09757]	[1.12575]	[-0.35586]
D(LCPI(-5))	0.2653	-0.3125	-0.0382	0.3479
	-0.8108	-0.4446	-0.0572	-0.6711
	[0.32722]	[-0.70288]	[-0.66686]	[0.51835]
D(LCPI(-6))	-0.3426	-0.1051	0.0121	-0.8702
	-0.8058	-0.4418	-0.0569	-0.6669
	[-0.42520]	[-0.23779]	[0.21349]	[-1.30483]
D(LCPI(-7))	1.1674	-0.3967	0.0314	-0.0694
	-0.8024	-0.4400	-0.0566	-0.6641
	[1.45499]	[-0.90172]	[0.55449]	[-0.10450]
D(LCPI(-9))	1.2660	0.5138	0.0360	-0.7778
	-0.8057	-0.4418	-0.0569	-0.6668
	[1.57139]	[1.16315]	[0.63284]	[-1.16641]
D(LCPI(-10))	0.2827	0.2544	0.0093	0.5376
	-0.7997	-0.4385	-0.0565	-0.6619
	[0.35345]	[0.58017]	[0.16526]	[0.81219]
D(LCPI(-11))	1.3151	-0.2411	0.1269	-0.4034
	-0.7997	-0.4385	-0.0565	-0.6619
	[1.64445]	[-0.54982]	[2.24724]	[-0.60938]
D(LCPI(-12))	-1.6542	0.3718	0.0097	-0.1993
	-0.8026	-0.4401	-0.0567	-0.6643
	[-2.06095]	[0.84490]	[0.17154]	[-0.29998]
D(LCPI(-13))	-1.5180	-1.0299	0.0350	-0.3764
	-0.7958	-0.4363	-0.0562	-0.6586
	[-1.90760]	[-2.36029]	[0.62251]	[-0.57146]
D(LCPI(-14))	1.3066	-0.0347	0.0839	0.2433
	-0.7994	-0.4383	-0.0564	-0.6616
	[1.63453]	[-0.07912]	[1.48764]	[0.36766]
D(LOP(-1))	-0.0061	-0.0765	0.0068	0.2912
	-0.0701	-0.0384	-0.0050	-0.0580
	[-0.08735]	[-1.99122]	[1.37296]	[5.02167]
D(LOP(-2))	0.2199	0.0626	-0.0059	0.0010
	-0.0713	-0.0391	-0.0050	-0.0590

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates

	[3.08495]	[1.60215]	[−1.17424]	[0.01647]
D(LOP(-3))	−0.0118	0.0582	0.0028	0.0005
	−0.0720	−0.0395	−0.0051	−0.0596
	[−0.16378]	[1.47400]	[0.54313]	[0.00846]
D(LOP(-4))	0.0549	0.0048	−0.0023	−0.0635
	−0.0713	−0.0391	−0.0050	−0.0590
	[0.76924]	[0.12215]	[−0.44835]	[−1.07644]
D(LOP(-5))	−0.0567	−0.0241	0.0063	−0.0261
	−0.0714	−0.0392	−0.0050	−0.0591
	[−0.79399]	[−0.61478]	[1.24064]	[−0.44174]
D(LOP(-6))	−0.0150	−0.0106	0.0124	−0.1112
	−0.0711	−0.0390	−0.0050	−0.0588
	[−0.21083]	[−0.27308]	[2.47955]	[−1.88973]
D(LOP(-7))	0.0126	0.0466	0.0044	0.0456
	−0.0697	−0.0382	−0.0049	−0.0576
	[0.18148]	[1.22071]	[0.89701]	[0.79170]
D(LOP(-9))	−0.0423	−0.0190	−0.0021	−0.0903
	−0.0674	−0.0370	−0.0048	−0.0558
	[−0.62732]	[−0.51510]	[−0.43762]	[−1.61838]
D(LOP(-10))	−0.0816	−0.0428	0.0053	0.0866
	−0.0703	−0.0385	−0.0050	−0.0582
	[−1.16077]	[−1.11162]	[1.06217]	[1.48962]
D(LOP(-11))	0.0838	0.0376	−0.0021	0.0011
	−0.0712	−0.0390	−0.0050	−0.0589
	[1.17728]	[0.96387]	[−0.41823]	[0.01915]
D(LOP(-12))	−0.0075	−0.0593	−0.0033	−0.0389
	−0.0715	−0.0392	−0.0050	−0.0591
	[−0.10495]	[−1.51304]	[−0.66009]	[−0.65853]
D(LOP(-13))	0.0765	0.0396	0.0087	−0.0870
	−0.0713	−0.0391	−0.0050	−0.0590
	[1.07243]	[1.01405]	[1.71874]	[−1.47382]
D(LOP(-14))	−0.1990	0.0339	0.0015	−0.0226
	−0.0675	−0.0370	−0.0048	−0.0559
	[−2.94691]	[0.91541]	[0.32017]	[−0.40430]
C	−0.0188	0.0762	−0.0045	−0.0714
	−0.0458	−0.0251	−0.0032	−0.0379
	[−0.40982]	[3.03200]	[−1.38625]	[−1.88435]
@SEAS(1)	0.0297	−0.0529	0.0037	0.0896
	−0.0601	−0.0330	−0.0042	−0.0498

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates

	[0.49466]	[0.87517]	[1.80198]
@SEAS(2)	-0.0450	0.0072	0.1014
	-0.0674	-0.0048	-0.0558
	[-0.66753]	[1.50366]	[1.81683]
@SEAS(3)	0.0552	0.0107	0.0905
	-0.0612	-0.0043	-0.0506
	[0.90315]	[2.46864]	[1.78770]
@SEAS(4)	0.0921	0.0145	0.1324
	-0.0649	-0.0046	-0.0537
	[1.42063]	[3.16401]	[2.46723]
@SEAS(5)	0.1399	0.0040	0.1015
	-0.0636	-0.0045	-0.0526
	[2.19910]	[0.88091]	[1.92840]
@SEAS(6)	0.2884	0.0084	0.1066
	-0.0617	-0.0044	-0.0511
	[4.67276]	[1.92127]	[2.08622]
@SEAS(7)	-0.3604	0.0075	0.1342
	-0.0652	-0.0046	-0.0540
	[-5.52495]	[1.61945]	[2.48555]
@SEAS(8)	-0.1103	0.0100	0.0324
	-0.0695	-0.0049	-0.0575
	[-1.58865]	[2.04934]	[0.56296]
@SEAS(9)	-0.0075	-0.0009	0.0104
	-0.0709	-0.0050	-0.0587
	[-0.10582]	[-0.17389]	[0.17693]
@SEAS(10)	0.1246	0.0094	0.0841
	-0.0668	-0.0047	-0.0553
	[1.86453]	[2.00151]	[1.52023]
@SEAS(11)	0.0644	0.0065	0.0641
	-0.0606	-0.0043	-0.0502
	[1.06230]	[1.52475]	[1.27918]
R-squared	0.9397	0.3844	0.2640
Adj. R-squared	0.9271	0.2564	0.1110
Sum sq. resids	2.8018	0.0140	1.9193
S.E. equation	0.0962	0.0068	0.0796
F-statistic	74.89	3.00	1.73
Log likelihood	373.83	1346.71	443.25
Akaike AIC	-1.6885	-6.9902	-2.0667

(*Continued*)

Annexure 3. (*Continued*)

Vector Autoregression Estimates				
Schwarz SC	−1.0074	−2.2092	−6.3092	−1.3857
Mean dependent	0.0105	0.0050	0.0066	0.0036
S.D. dependent	0.3562	0.0902	0.0079	0.0844
Determinant resid covariance (dof adj.)		7.0E-12		
Determinant resid covariance		3.3E-12		
Log likelihood		2.8E+03		
Akaike information criterion		−1.4E+01		
Schwarz criterion		−1.1E+01		

Source: Authors' calculations

Annexure 4.

V.A.R. Lag Exclusion Wald Tests
Date: 12/18/17 Time: 08:39
Sample: 1985M07 2017M06
Chi-squared test statistics for lag exclusion:
Numbers in [] are p-values

	D.(L.T.R.E.V.)	D.(L.L.S.M.)	D.(L.C.P.I.)	D.(L.O.P.)	Joint
Lag 1	211.7907	65.6307	4.3038	26.9363	313.5390
	[0.000000]	[1.90e-13]	[0.366453]	[2.05e-05]	[0.000000]
Lag 2	74.0491	31.7541	5.4962	0.7315	110.2031
	[3.22e-15]	[2.15e-06]	[0.240068]	[0.947386]	[4.44e-16]
Lag 3	16.5489	18.4682	5.4385	0.6911	43.3855
	[0.002365]	[0.000999]	[0.245184]	[0.952418]	[0.000244]
Lag 4	19.6419	24.0344	2.5535	1.3473	49.2117
	[0.000588]	[7.86e-05]	[0.635083]	[0.853308]	[3.06e-05]
Lag 5	26.6986	18.1871	3.1570	3.5521	49.4773
	[2.29e-05]	[0.001134]	[0.531905]	[0.470005]	[2.78e-05]
Lag 6	5.4432	27.8797	5.5395	6.4844	46.2675
	[0.244770]	[1.32e-05]	[0.236276]	[0.165779]	[8.85e-05]
Lag 7	8.5727	21.2361	3.0304	1.0028	35.9541
	[0.072713]	[0.000284]	[0.552758]	[0.909373]	[0.002936]

(*Continued*)

Annexure 4. (Continued)

Lag 8	2.4636 [0.651173]	14.9926 [0.004717]	4.1180 [0.390277]	21.1115 [0.174264]
Lag 9	7.2260 [0.124415]	8.4881 [0.075248]	6.2889 [0.178588]	28.7224 **[0.025877]**
Lag 10	5.0823 [0.278958]	11.2472 [0.023922]	2.2635 [0.687428]	24.3923 **[0.081279]**
Lag 11	20.7123 [0.000361]	2.1578 [0.706767]	6.4064 [0.170787]	33.2750 **[0.006794]**
Lag 12	5.7750 [0.216595]	16.3306 [0.002606]	2.1480 [0.708552]	28.3715 **[0.028535]**
Lag 13	5.0538 [0.281818]	11.0883 [0.025589]	5.1703 [0.270266]	25.9347 [0.054954]
Lag 14	18.1278 [0.001165]	3.2306 [0.520001]	4.6171 [0.328887]	28.8451 **[0.025002]**
Lag 15	5.3652 [0.251833]	3.2812 [0.511923]	3.6864 [0.450105]	15.9026 [0.459777]
df	**4**	**4**	**4**	**16**

Note: P-values < 0.05 for joint are selected as V.A.R. lags.
Source: Authors' calculations.