3

Data and Methodology

Present chapter includes the important source of data, and discussed the problems and limitations of its availability. Further this chapter throws lights on statistical tools employed for the analysis and its constraints.

Section 1

3.1. Data sources and limitations

3.1.1. Data on Revenue and Expenditure in Rajasthan

The study is based on secondary data. Government of Rajasthan publishes revenue and expenditure data in different volumes of budget papers every year. Revenue receipts are published in volume 2 and expenditure on General Services is published volume 2B, expenditure on Social Services is published in volume 2C, expenditure on Economic Service is published in volume 2D. Capital expenditure is published in volume 3A and data related to Public Debt and Loans are published in volume 3B. Government Grants and investment are published in volume 4B. However, with current budget every year, previous year data on actual expenditure and revenue are published. For the study, actual revenue and expenditure figures have been used. The revenue and expenditure data is published by Statistical Abstract of Rajasthan; it is an annual publication of the Department of Economics and Statistics. The data culled out from different volumes of budget papers were counter checked with the statistical abstract. The data on Net State Domestic Product (NSDP) of Rajasthan was collected from the publication of The Central Statistical Office (CSO) of Ministry of Statistics and Programme Implementation (MoSPI). The NSDP is used at constant price at 2004-05. As mentioned above, the revenue and expenditure data is at current price and it has to be transformed into
constant price. For deflation, implicit deflator of NSDP of Rajasthan at 2004-05 prices was used. The study has used the same data set for testing the Keynesian and Wagnerian hypothesis of Public expenditure. In the Indian context, the limitation in using consumer expenditure data is that it is available only on quinquennial basis at the state level from the National Sample Survey Office Consumer Expenditure Surveys. Therefore, we are constrained to use the total expenditure to test the hypothesis of both Keynesian and Wagner’s. Therefore, the coefficient obtained from causality checks for Keynes and Wagner’s law.

3.1.2. Period of the Study

For the analysis, annual data for the period 1970-71 to 2013-14 was used. The choice of the period was influenced primarily by the availability of comparable data on expenditure and revenue, i. e., the data is available from 1970-71. The terminal year of the analysis was preferred to be 2013-14, which is again chosen on the basis of the availability of final data on NSDP of Rajasthan, expenditure and revenue data at disaggregated level.

3.2. Hypotheses

H₀₁: There is no causal relationship and unidirectional causality between public expenditure and economic growth in Rajasthan.

H₁₁:- There is causal relationship and unidirectional causality between public expenditure and economic growth in Rajasthan.

Explanation: Keynesian theory of public expenditure asserts that government expenditure causes economic growth. The causality between public expenditure and economic growth are found to be unidirectional from public expenditure to economic growth.

H₀₂: There is no causal and unidirectional relationship between economic growth and public expenditure.
H$_{12}$: There is causal and unidirectional relationship and causality between economic growth and the government expenditure.

**Explanation:** Wagner’s Law states that there is a positive relationship between economic growth and government expenditure. It is indicative of the fact that a higher economic growth enables governments to mobilise more revenue in terms of tax and non-tax revenue and expand more. It means, there is a unidirectional causality from economic growth to government expenditure.

H$_{03}$:- There is no causal Relationship between tax revenue and public expenditure.

H$_{13}$:- There is causal relationship between tax revenue and public expenditure;

**Explanation:** There are tested and proved hypothesis on the relationship between tax revenue and public spending by the government. Important and proved hypotheses related to tax and government spending are: (i) Spend and tax hypothesis, (ii) Fiscal Synchronization Hypothesis, (iii) Institutional Separation Hypothesis.

H$_{04}$:- There is no significant difference in public expenditure in Rajasthan between pre and post economic reform period.

H$_{14}$:- There is significant difference in public expenditure in Rajasthan between pre and post economic reform period.

**Explanation:** Economic reform phase is based on the theoretical postulate of curtailing government’s role in economic activities of the economy.

H$_{05}$:- There is no significant association between the size of government spending and different determinants of public expenditure in Rajasthan.

H$_{15}$:- There is significant association between the size of government spending and different determinants of public expenditure in Rajasthan.

**Explanation:** The size of public expenditure is not influenced by its determinants viz; tax revenue, NSDP, grants, borrowing, population etc. It implies that these variables have no significant influence on government expenditure in Rajasthan during the period of 1970-71 to 2013-14.
3.3. Methodology of the Study

The Methodology used in the study is different across the objectives. This section is sub-divided into two sections. The first section provided the detail discussion on different statistical tools, which has been used to analyze the trend and pattern of Growth rate of Government expenditure, government revenue, debt, NSDP etc. for the state of Rajasthan from period 1970-71 to 2013-14. In the second section, the study tests hypothesis put forwarded by different theoretical strands on the association between government expenditure, government revenue and economic growth. Moreover, the direction of causality is important from theoretical perspective. Rather complex statistical tools were used for the estimation of elasticity of different variables and its association.

3.3.1. Estimation of Growth Rates

The study used the simple ratio and percentages to examine the trend and pattern in government expenditure along with disaggregate and per capita analysis of government expenditure, government revenue and NSDP in Rajasthan for period 1970-71 to 2013-14. The NSDP implicit deflator\(^1\) at constant price 2004-05 has been used to convert the Nominal values of the whole data set into real values. The study used the absolute change \((Y_t - Y_{t-1})\), relative change \([(Y_t - Y_{t-1})/ Y_{t-1}]\) and percentage growth rate \([(Y_t - Y_{t-1})/ Y_{t-1}] \times 100\) to study the trends in Government expenditure and its components over the period of 44 years. In order to analyse the growth rate of government expenditure, revenue and NSDP of Rajasthan state the exponential

\[NSDP_{Deflator} = \left[ \frac{NoalNSDP}{RealNSDP} \right] \times 100\]

\(^1\) NSDP_{Deflator} = \left[ \frac{NoalNSDP}{RealNSDP} \right] \times 100
growth rate and Quadratic growth function has been used. The model specification of Exponential growth rate model can be specified as following –

\[ \log Y = a + \beta \ t + ut; \]  

(EQ. 3.1)

Where \( \log Y \) = log values of government expenditure; \( a \) = constant, \( \beta \) = Coefficient of time, \( ut \) = Error term. This model is also known as log-linear model or Semi-log model as the regressand of this model is in log form and the regressor is Time. However, the trend in the growth rate can be measured by fitting a quadratic model of the following form:

\[ \log Y = a + \beta t + \beta_1 t^2 + ut; \]  

(EQ. 3.2)

Where, \( \beta_1 \) is coefficient of time square or \( t^2 \). This model is also known as log-lin model with quadratic trend. In this model, the sign of \( \beta_1 \) is indicative of the trend in the growth rate as it is increasing at an increasing or decreasing rate.

### 3.3.2. Kinked Exponential Growth Function

On analyzing the Exponential growth rate and quadratic growth rate it has been observed that the long run rate of growth in government expenditure and NSDP was not uniform and has encountered a structural break. The hypothesis can be statistically tested and the year of structural change can be pinpointed. The method commonly followed to estimate the exponential trend in the growth of a variable can be expressed in the form of the following equations.

\[ y_t = y_0 e^{gt} \]  

(EQ. 3.3)

Logarithmic version in both sides of the equation gives

\[ \ln y_t = \ln y_0 + \hat{g} t + u_t \]  

(EQ. 3.4)
Where $y_t$ is variable for which ‘$g$’ has to be estimated and $y_o$ is the initial year and $u_t$ are the error term. The growth model assumes a constant rate of growth for the entire period of the analysis. Apparently, such an assumption deviates from reality because an economic variable like government expenditure or NSDP, which is subjected to frequent changes on account of several other macroeconomic variables or effected through other endogenous or exogenous variables. The discussion on the long-run movement of government expenditure has categorically pointed out that there exist different breaks. Often, the long term trend in the rate of growth in economic variables with such structural breaks is estimated by fitting a kinked exponential function developed by Boyce (Boyce, 1986: 385-391). It is assumed in such growth models that the break took place exactly in the year in which the policy change occurred. The kinked exponential function usually employed to work out the rate of growth with an assumed year of break in the long run movement of the government expenditure takes the following form:

\[
\ln Y_t = \hat{\alpha}_1 + \hat{\alpha}_1 (d_1t + d_2k) + \hat{\alpha}_2 (d_2t - d_2K) + u_t \quad (EQ. 3.5)
\]

where: $\ln Y_t$ is the logarithmic version of variable under consideration; $\hat{\alpha}_1$ – constant; $\hat{\alpha}_1$ - growth rate for the period 1 and $\hat{\alpha}_2$ are the growth rate for the period 2; $K$ is the break point; $d$ is dummy variable (1 and 0 as the case may be) and $u_t$ is the stochastic error term. This equation can be applied if there is only one break or kink in the variable. If the number of kinks extends to more than one, the above equation takes the following form:

\[
\ln Y_t = \hat{\alpha}_1 + \hat{\alpha}_1 (d_1t + d_2k) + \hat{\alpha}_2 (d_2t - d_2K) + \hat{\alpha}_3 (d_3t + d_3K) + \hat{\alpha}_4 (d_3t - d_3K) + \hat{\alpha}_5 (d_3t + d_4K) + \hat{\alpha}_6 (d_4t - d_4K) + \hat{\alpha}_7 (d_4t + d_5K) + \hat{\alpha}_8 (d_5t - d_5K) + u_t \quad (EQ. 3.6)
\]
As mentioned elsewhere, the estimation of long-term growth demands a perfect knowledge of the variables and the variables in question. Often such breaks are identified arbitrarily based on certain presumptions or with visual judgment or graphical presentation of data. However, such identified breakpoints need not always be statistically valid or the result forms the Chow test would yield a misleading result (Hansen, 2001: 117-128).

3.2.3. Cointegration and Causality Analysis

The objective of the study is to analyze the causality between government expenditure, revenue and NSDP for period 1970-71 to 2013-14. So the study used the necessary time series analysis tools viz; (i) Stationary Test; (ii) Cointegration Test; (iii) Causality test; (iv) Diagnostic test for time series analysis, to find the causality between the variables under consideration. The detail description of the time series methodology used for testing different hypothesis based on the association and causality between government expenditure and Economic growth and government expenditure and revenue are listed in this section.

3.3.4. Stationary Test (Unit Root)

The First Step of Time series data analysis is to test the stationary properties of the variables. The classical regression assumes that the time series variable \( \{X_t\}_{t=1}^{T} \) should be stationary or the mean, variance and covariance of the series should be independent of time. It is also known as covariance stationary or Weak Stationary\(^2\). If the data set is non-stationary, it means the mean (\( \mu \))

\[ \text{\emph{Weak stationary or second order stationary time series is a series which has the constant mean\{E(X_t)= \mu \}, variance \{ Var(X_t) = \sigma^2 \} and covariance (Cov( X_t, X_{t+k})= E\{X_t E(X_{t+k})\} \{ X_{t+k}- E(X_{t+k})\}) over time but if all the moments of all the degree (Mean, variance, covariance and higher order moments) of a time series is constant over time than it is Known as the Strictly stationary or strongly stationary. In other words if the joint distribution of a time series is constant then it is known as strictly stationary. Moreover, More formally, a time series will} \]

---

\(^2\) Week stationary or second order stationary time series is a series which has the constant mean\{E(X_t)= \mu \}, variance \{ Var(X_t) = \sigma^2 \} and covariance (Cov( X_t, X_{t+k})= E\{X_t E(X_{t+k})\} \{ X_{t+k}- E(X_{t+k})\}) over time but if all the moments of all the degree (Mean, variance, covariance and higher order moments) of a time series is constant over time than it is Known as the Strictly stationary or strongly stationary. In other words if the joint distribution of a time series is constant then it is known as strictly stationary. Moreover, More formally, a time series will
and variance ($\sigma^2$) of the data set is inconsistent or varies with respect to time and would yield spurious\(^3\) result of regression parameters. Stationary Property of time series variable is also important in order to forecast the series or to understand that what kind of the process we have to develop in our model to get the accurate prediction (Diebold and Kiillian, 1999). There are different methods to identify whether the time series under consideration is stationary or not viz; (i) Graphical method\(^4\) (ii) Correlogram\(^5\); and (iii) Unit root analysis\(^6\).

---

\(^3\) The problem of spurious regression was highlighted by Granger and Newbold, by the use of Monte-Carlo Technique in order to show the regression between two independent non-stationary time series with a coefficient that was statistically different from zero. In other words the problem of spurious regression occurred when two time series variables are highly correlated whereas there is no actual relationship between two time series variables and this high correlation is due to time trend in both the variables (Granger and Newbold, 1974).

\(^4\) In this method the series is plotted against time to identify weather the series is stable or not. In addition, it is an informal method which provides the picture of a series that the series has constant movement over time or not.

\(^5\) Correlogram or ACF plots the correlation coefficient (correlation of a series with its future or past values) of a time series lagged by one. If a Correlogram has very few significant spikes at very small lags or the spikes dies down quickly with lags then the series is considered as stationary.

\(^6\) Unit root test is widely popular as a test for stationary properties of the time series. According to Markov first order autoregressive model-

\[ Y_t = \rho Y_{t-1} + \mu_t \]  

Where; (i) -1$\leq$$\rho$$\leq$1; (ii) $\mu_t$ is a white noise error term with normal distribution of zero mean and unit variance.
As mention elsewhere the First two method, to identify the stationary properties of time series are the informal method, therefore one cannot rely on this method and every time series should be tested for its stationary properties with the use of a formal method or different Unit root tests. There are Different Unit root tests Viz; (i) Dickey-Fuller Test (DF Test); (ii) Augmented Dickey Fuller Test (ADF Test); (iii) Phillips-Perron Test (PP Test), (iv) Kwiatkowski–Phillips–Schmidt–Shin (KPSS Test). The study used the Augmented Dickey-Fuller and Phillips-Perron Test of Unit root to check the integration order of the variables in the study as these test are the prominent test to check the integration order of the time series variables.

3.3.5. Augmented Dickey Fuller Test of Unit Root (ADF Test)

It is a commonly used test to examine the stationary property of time series data (Dickey and Fuller, 1979). The Augmented Dickey fuller test is

If $\rho=1$ then this is the condition of Unit root or random walk model without drift or also known as non-stationary stochastic series. Now if we subtract the lagged value of $Y_t$ from both side of the equation (1) such as

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

then we will get

$$Y_t - Y_{t-1} = (\rho - 1)Y_{t-1} + \mu_t;$$

or

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

Where $\delta = (\rho - 1)$ and $\Delta$ is first difference Operator. The null hypothesis of DF test is that the value of $\delta=0$ (or $\rho=1$) which means that $\rho=1$ or it implies that the series under consideration is non-stationary or has a unit root (if the value of $\delta$ is negative or the value of $\rho$ is less than one then the series will be stationary). Here the ‘t’ value of the estimated coefficient of $Y_{t-1}$ follow the $\tau$ (tau) Statistics [$\tau = \delta/\sqrt{\text{Var}(\delta)}$] (Dickey and Fuller, 1976).
considered as a superior to the Dickey Fuller\textsuperscript{7} test as the Dickey-Fuller (DF) test is based on the assumption that the errors (residuals) are serially uncorrelated (Gujrati, 2007: PP 836). But the ADF test overcomes this limitation by with the assumption that the errors (residuals) may be serially correlated or ADF test “Augment” by specifying the lagged value\textsuperscript{8} of the dependent variable, i.e., $\Delta Y_t$ as one of the independent variables in the equation. The Augmented Dickey-Fuller (ADF) test is employed in three forms, viz., (i) ADF with drift or constant terms; (ii) ADF with drift and trend; (iii) and ADF without drift and trend. The equations of different forms of ADF test is specified below.

**ADF with drift or constant:**

$$\Delta \ln NSDP = \alpha + \delta \ln NSDP_{t-1} + \sum_{i=1}^{m=1} \theta_i \Delta \ln NSDP_{t-i} + \varepsilon_t$$  \hspace{1cm} (EQ. 3.7)

Where $\ln$- is Natural Logarithum; NSDP – Net State Domestic Product of Rajasthan at constant price (2004-05)

**ADF with drift and trend:**

$$\Delta \ln NSDP = \alpha + \beta t + \delta \ln NSDP_{t-1} + \sum_{i=1}^{m=1} \theta_i \Delta \ln NSDP_{t-i} + \varepsilon_t$$  \hspace{1cm} (EQ. 3.8)

\textsuperscript{7} Dickey Fuller test can be estimated with three different forms-

(i) $Y_t$ as a random walk: $\Delta Y_t = \delta Y_{t-1} + \mu_t$

(ii) $Y_t$ as a random walk with Drift: $\Delta Y_t = \alpha + \delta Y_{t-1} + \mu_t$

(iii) $Y_t$ as a random walk with Drift and Trend $\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \mu_t$

Where $\mu_t$ is IID White noise or it is identically and independently distributed with mean zero and variance unit.

\textsuperscript{8} In ADF Equations, each disturbance term is White noise and the selection of optimal lag length is important for the cointegration test. Lags are selected sequentially and with equal values such as 1, 2, 3 or 4 (Bird, 1971). The number of lagged difference terms is usually determined empirically or on the basis of different Criterion of optimal leg length viz; (i) Akaike Information Criterion (AIC) (ii) Schwarz Information Croterion (SIC); (iii) Hannan-Quinn Criterion (HQC); (iv) Modified Akaike Information Criterion; (v) Modified Schwarz Information Croterion; (vi) Modified Hannan-Quinn Criterion (vii) t-statistic. The lag having the minimum value of all these criterions is considered as an optimal lag.
ADF without drift and trend:

$$\Delta \ln NSDP = \delta \ln NSDP_{t-1} + \sum_{i=1}^{m} \theta_i \Delta \ln NSDP_{t-i} + \epsilon_i$$  \hspace{1cm} (EQ. 3.9)

Here $\Delta$ is first difference operator, $\delta$ is coefficient of proceeding observation. $\Delta \ln NSDP_{t-1}$ is difference leg term, $m$ is the number of lags, $\theta_i$ is the parameter to be estimated, $\epsilon_i$ is the disturbance term. In all the three variant equations of ADF tests, the null hypothesis is that $\delta = 0$ or NSDP for Rajasthan at constant price (2004-05) from 1970-71 to 2013-14 has unit root or the series is non-stationary. The alternative hypothesis is that the $\delta < 0$. It means the series is stationary. If the null hypothesis is rejected$^9$, then the series is stated to be stationary or Non-rejection of null hypothesis implies that there is presence of Unit Root i.e. series is non-stationary. If $\delta$ is statistically significant from zero, the computed $\tau$-statistics of the parameter $\delta$ is compared with the critical tabulated value of Mackinnon Table (1999). If the data series is non-stationary at level, then same procedure is repeated on the first difference of the variable or data series employing the following regression equation.

$$\Delta^2 \ln NSDP = \alpha + \beta t + \delta \ln NSDP_{t-1} + \sum_{i=1}^{m} \theta_i \Delta^2 \ln NSDP_{t-i} + \epsilon_i$$  \hspace{1cm} (EQ. 3.10)

### 3.3.6. Phillips- Perron Test of Unit Root (PP test)

Phillips-Perron test (1988) is an alternative test to the ADF test of unit root. The PP test is a non-Parametric test, against the ADF test which assumes that error terms are IID (error terms are white noise or identically and independently distributed (IID) with mean zero and variance unity). On the contrary, PP test does not assume any parametric assumptions for error term.

---

$^9$ If computed $\tau$ statistic is more negative than ADF critical values (tabulated value of Mackinnon, 1991) then we will reject the null of unit root means that the series under test is stationary. On the contrary, If computed $\tau$ statistic is less negative than ADF critical values (tabulated value of Mackinnon, 1991) then we will accept the null hypothesis of unit root or the series under consideration is non-stationary
which mean this test can be applied on wide data set having large number of observation. In other words it is based on the asymptomatic theory. The PP test takes care of the problem of serial correlation in the error terms without adding the lagged difference terms (Gujrati, 2007, PP 836). PP test adjust for the serial correlation problem by the using the Newey–West (1987) heteroskedasticity and autocorrelation consistent covariance matrix estimator. Another advantage of PP test is that in this test there is no need to specify the lag length of the series. Null Hypothesis of PP test is same as the ADF test.

3.3.7. Kwiatkowski–Phillips–Schmidt–Shin (KPSS Test)

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS Test) has been used to confirm the results of ADF and PP test of Unit root. The ARDL Model is very sensitive to the lag order 2. In other words, it cannot be applied in cases where the variable specified in the model is integrated of order 2. Moreover, the null hypothesis of the KPSS Test is also inverse to the hypothesis of KPSS and PP tests. The null hypothesis for the KPSS test assumes that the variable has no unit root or series is stationary. In KPSS Test, the null hypothesis is that the variable is stationary or it has no unit root problem, which is the inverse of the null hypothesis of the ADF Test.

3.3.8. Test for Cointegration

Once the order of integration of all the variables tested for Stationary, the next step is to test the stable relation between the variables for making the inference of future behavior of the variables. The concept of the cointegration was first introduced by Engle and Granger in 1987. According to them if two variables are cointegrated then they will share the common stochastic trend over time. The cointegration is a long run equilibrium relationship between two time series variables. If two time series variable have different trend then they cannot stay fixed in relation with each other in long run, implies that we cannot make the inference for future behavior of the variables or from the inferences
point of view our model will be invalid. If two variables are stationary at $I(0)$ then they must be cointegrated and in this case, we can make the inferences on the basis of simple OLS Method of Regression as regression will be non-spurious. On the other side, if two time series variables are not stationary at $I(0)$ but integrated at $I(1)$, can be cointegrated or there may be the possibility that the linear combination$^{10}$ of the two variables would be stationary, if it is then they must be cointegrated. In other words, two time series variables can be cointegrated, even if there is short run drift or the time series variables are individually non-stationary and the linear regression of these two variables will not be spurious. There are several methods to test the cointegration of time series variables viz; (i) Engle - Granger two step method (EGM), (ii) Engle Yahoo three step Method (EYM); (iii) The Johansen Maximum Likelihood Method (ML) Vector Autoregressive (VAR) Method; (iv) The Saikkonen Method; and (v) The cointegrating Regression Durbin Watson Statistic. There are a few other methods too. For the present study, Engle - Granger two step methods (EGM) and Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method has been used to find the cointegration between government expenditure, government revenue, and economic growth in Rajasthan for the period 1970-71 to 2013-14.

3.2.9. Engle-Granger Two Step Technique for Cointegration
The Engle-Granger Two Step Technique for cointegration involves the testing of the Unit Root of residuals. The first step of the Engle-Granger test is to test the stationary property of the variables and if all the variables are integrated of $I(1)$ or the integration order of all the variables is the same but these variables are not integrated at level $I(0)$, then the OLS is applied and obtain $\beta$. The second step is the estimation of the residual of the regression model and

---

$^{10}$ Linear combination means regression of $Y_t$ and $X_t$ variable yields a stationary residuals ($e_t$) series.
testing the unit root properties of the residual series. If the residual of the OLS model are stationary at level or they have no unit root, then the variables are said to be cointegrated or there is long run equilibrium relationship between those variables. In other words according to Engle-Granger Two Step method, if the government expenditure and NSDP are not stationary at \( I(0) \), the error term of regression of government expenditure on NSDP is tested and found stationary at \( I(0) \), the stochastic error term in these variables under consideration cancels each other out and the linear combination of these two time series variables are stated to stationary at \( I(0) \). In this case, it is stated that time series data of government expenditure and NSDP are cointegrated and, therefore, the regression of expenditure on income yields non-spurious results. The regression of two non-stationary variables at \( I(0) \) level, but integrated theoretically can be regressed. Such a regression is called Cointegrating Regression and the Regression Parameters are called Cointegrating Parameters. The present study estimated following equations for testing the cointegration between government expenditure and Net State Domestic product for the Rajasthan state economy.

Step -1

\[
\ln \text{GE} = \alpha + \beta \ln \text{NSDP} + \mu_t \tag{EQ. 3.11}
\]


Step-II: Testing of Unit Root properties of the residual series

\[
\mu = \alpha + \rho \mu_{t-1} + \epsilon_t \tag{EQ. 3.12}
\]

If residuals are found stationary (Eq3.12), it shows cointegration between GE and NSDP variables and therefore the regression of NSDP on GE may not produce spurious regression parameters. The Engle-Granger Two Step method is based on simple regression analysis and ADF test of Unit root, which are easy to compute, so it is preferable in comparison with other tests of cointegration, but still this method has some limitations.
Limitations of Engle-Granger Two Step Technique for Cointegration:
The Engle Granger cointegration test suffers from the following limitations: (i) There may be significant small sample biases in OLS estimation of cointegration (Banerjee et al, 1986); (ii) Conventional DF and ADF tests suffer from parameter instability (Hendry and Mizon, 1990); (iii) Limiting distributions for the DF and ADF tests are not well defined, implying that the power of the tests is low (Phillips and Ouliaris, 1990); (iv) The Engle Granger step method does not give any specification on dependent and independent variable in first step of the method. In other words, in case of two time series which series is considered as regression or regressand and why; (v) if there is the possibility of occurrence of an error in first step then it will lead to another error in the second step; and (vi) There may be the possibility of more than one long run relationship or multiple cointegrating vectors. The OLS estimates of the cointegrating vector cannot identify multiple long run relationship or test for number of cointegration vectors. The EG test is not robust as compared to Johansen cointegration test. Thus, it is necessary to complement the Engle-Granger Two Step procedures (EG) test with the Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method.

3.2.10. Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method for cointegration
Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method or The Johansen and Juselius Cointegration technique considered as a superior test for cointegration as it fulfills all the desirable statistical properties. The present study employed the Johansen Maximum Likelihood (ML) Vector Autoregressive (VAR) Method for testing the long run relationship between government expenditure, government revenue and economic growth for Rajasthan state for period 1970-71 to 2013-14. Johensen and Juselius (1990) suggested a cointegration estimation methodology to overcome the limitations of the Two-Step Approach. It is a full information maximum likelihood
approach (Johansen, 1988, and Johansen and Juselius, 1990). The basic assumption of the Johansen approach of cointegration is that all the variables under test should be \( I(1) \) (if the variable are not stationary at \( I(1) \) then the variables can be transformed with different transformation method\(^{11}\) to make the variable stationary at \( I(1) \)). The Johansen approach is based on maximum likelihood estimates of all the cointegrating vectors. The Johansen approach is based on likelihood ratio (LR) test to determine the number of cointegration vectors in the regression. Three tests statistics are suggested to determine ‘the number of cointegration vectors\(^{12}\)': the first is Johansen’s “trace” statistic method, the second is “maximum Eigen value” statistic method, and the third method chooses \( r \) to minimize an information criterion. The Johansen approach of cointegration examines the number of independent linear combinations (\( k \)) for an \( n \) time series variables. The Johansen and Juselius cointegration technique is based on the following VAR model with order \( \rho \):

\[
Y_t = A_0 + \sum_{i=1}^{\rho} \beta_i Y_{t-i} + e_t \tag{EQ. 3.13}
\]

Where \( Y_t \) is a \( n \times 1 \) vector of non-stationary \( I(1) \) variable, \( A_0 \) is an \( n \times 1 \) vector of constant, \( \rho \) is lag length, \( B_i \) is an \( n \times n \) matrix of Parameters, and \( e_t \) is \( n \times 1 \) vector of white noise errors. The above VAR model needs to turned in a vector error correction Model which is as follow-

\[
\Delta Y_t = A_0 + \sum_{i=1}^{\rho} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-\rho} + e_t \tag{EQ. 3.14}
\]

Where \( \Delta \) is the first difference operator,

---

\(^{11}\) There are different method of transform the non-stationary variable into stationary viz; (i) Detrend the variable if there is time trend in the variable; (ii) log transformation of the variable; (iii) Differencing the variable; etc. The present study used the log form of all the variables instead of real value of the series to make the variable stationary at lower order.

\(^{12}\) The number of cointegrating Vector means the cointegrating Rank.
Here $I$ is an $n \times n$ identity matrix. The cointegration between time series variables can be observed on the basis of rank ($r$) of the matrix $\Pi$ by its Eigen values \(^{13}\). The rank of a matrix is the number of characteristic roots that are different from zero. In this case the null hypothesis is that $\Pi=\alpha \beta'$. Here $\alpha$ and $\beta$ are $n \times n$ loading matrices of eigenvectors where $\beta$ matrix will gives the cointegration vectors and $\alpha$ provides the amount of cointegration for each equation in VECM. Here $\alpha$ is also known as the adjustment parameters. The main objective is to find the $r$ cointegrating vectors. The number of characteristic root can be found on the basis of trace Statistic and Maximum Eigen value statistics.

\[
\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{k} \ln(1 - \tilde{\lambda}_i) \quad \text{(EQ. 3.17)}
\]

\[
\lambda_{\text{max}}(r, r+1) = -T \ln(1 - \tilde{\lambda}_{r+1}) \quad \text{(EQ. 3.18)}
\]

Here $r$ is the number of cointegrating vectors; $T$ is the number of observations and $\lambda$ is the eigenvalue from the $\Pi$ matrix (decreasing order of characteristic roots). If the eigenvalue is significantly different from zero then it is an indication of cointegrating vector and each significant eigenvalue present a stationary relation. The null hypothesis of trace test is that the number of

\(^{13}\) If the rank($r$) of $\Pi$ matrix is zero then each element of $\Pi$ matrix will be equal to zero and it is the indication of absence of cointegration. If the rank($r$) \(\leq(n-1)\) then there are $(n-1)$ cointegration relationship exists in the model ($n$ is number of variables).
cointegrating vectors is less than \( r \) or equal to \( r \), against the alternative hypothesis of more than one \( r \) cointegrating vector. On the other side the null hypothesis of Maximum Eigen value statistics test based on the null hypothesis that the number of cointegrating vectors is less than or equal to \( r \), against the alternative hypothesis that the number of cointegrating vectors is \( r+1 \) (Enders, 2004; Brooks, 2008). On examining the cointegration, if there is a sign of cointegration among the variables then the causality can be examined by the vector error correction model. On the contrary, if there is no cointegration between the variables, then the VAR Model can be applied to the first difference of all the variables.

3.2.11 Autoregressive Distributed Lag Model (Bound test) for Cointegration
The study employed the autoregressive distributed lag model proposed by Pesaran et al. (2001), Pesaran and Pesaran (1997). To analyze cointegration between government expenditure and its different determines. As mention in the literature, the ARDL approach of cointegration is more suitable than the other cointegration approaches such as Engle and Granger (1987) and Johansen and Juselius (1980) (Sakib-Bin-Amin, 2011, Emran et al. 2007). The ARDL test has a nonstandard distribution and the main advantage of a test is that ARDL Test can be applied in the case when the two series are integrated of different order i.e. one is I(0) and another is I(1) (Odhiambo, 2009, Harvie and Pahlavani, 2006 ). In addition to it, it also avoids the pre-testing procedure of standard cointegration analysis (the classification of the variables into I(1) and I(0)) and also provides more robust estimation when the sample size is small (Sakib-Bin-Amin,2011, Narayan,2005). The cointegration under ARDL Model identified the long-run relationship between the variables by comparing the upper and lower bound values and the computed F-statistics for testing the significance of the lagged levels of the variables in the error correction (Pesaran and Pesaran, 1996). The Null hypothesis under the
ARDL Model is that there is no long-run relationship between the variables and the alternative hypothesis is that there is the long run relationship between the variables under the testing procedure. If the computed F-statistics exceeded the upper bound value, the null hypothesis of no long run association can be rejected and can be concluded that there is the long run relationship between the variables and if the calculated F-statistics will be below the lower bound value, the Null hypothesis of no cointegration cannot be rejected. The results of the ARDL Model will be inconclusive if the calculated F-statistics lie between the set of upper and lower bound values of Pesaran table. The ARDL also estimate the Error Correction term (By ECM). It provides the information about the speed of adjustment towards the equilibrium in the model. The Error Correction term should be negative, less than 1 and highly significant which is an evidence of the existence of a stable long-term relationship (Bannerjee, Dolado and Mestre, 1998).

The study estimated following Multivariate model of bound testing procedure for possible long run relationship between government expenditure and different determinants of government expenditure under the study.

\[
\Delta \ln GE_t = \alpha_0 + \sum_{j=1}^{n-1} \phi_j \Delta \ln GE_{t-j} + \sum_{j=1}^{n} \sigma_j \Delta \ln NSDP_{t-j} + \sum_{j=1}^{n} \lambda_j \Delta \ln Debt_{t-j} + \sum_{j=1}^{n} \varphi_j \Delta \ln Tax Rev_{t-j} + \\
\delta_1 \Delta \ln GE_{t-j} + \delta_2 \Delta \ln NSDP_{t-j} + \delta_3 \Delta \ln Debt_{t-j} + \delta_4 \Delta \ln Tax Rev_{t-j} + \epsilon_t
\]

(EQ. 3.19)

---

14 (The two critical bound values (lower and upper) in Pesaran and Pesaran (1997): one is estimated by assuming that all the variables under the ARDL model are I(0) and the another is calculated assuming that all variables are I(1)).

15 The model specification is same for all the versions except the change in the independent variable and dependent variable so here the study shows only for the first version model specification and it will be in similar way for rest of the versions.
\[ \Delta \ln NSDP_t = \alpha_0 + \sum_{i=1}^{n} \beta_i \Delta \ln NSDP_{t-i} + \sum_{i=1}^{n} \sigma_j \Delta \ln GE_{t-i} + \sum_{i=1}^{n} \theta_j \Delta \ln Debt_{t-i} + \sum_{i=1}^{n} \gamma_j \Delta \ln Tax Rev_{t-i} + \eta_1 \ln NSDP_{t-i} + \eta_2 \ln GE_{t-i} + \eta_3 \ln Debt_{t-i} + \eta_4 \ln Tax Rev_{t-i} + \mu_{it} \]

(EQ. 3.20)

Here \( \ln GE \) is the log value of government expenditure; \( \ln NSDP \) is the log value of Net state domestic product, \( \ln Debt \) is the total Debt, \( \ln Tax Rev \) is the total tax revenue, \( \Delta \) is the difference operator and \( \mu_{it}, \varepsilon_{it} \) are the disturbance terms. Both the equations test the joint significant of the lagged value of the variables with F- statistics (Wald test). Here the null hypothesis is \( H_0: \eta_1 = \eta_2 = \eta_3 = \eta_4 = 0 \) in Equation 3.20 and \( H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0 \) in equation 3.19 (or null hypothesis is that there is no cointegration between the variables) and the alternative hypothesis is \( H_1: \eta_1 \neq \eta_2 \neq \eta_3 \neq \eta_4 \neq 0 \) in Equation 3.20 and \( H_1: \delta_1 \neq \delta_2 \neq \delta_3 \neq \delta_4 \neq 0 \) in equation 3.19. In this case if the null hypothesis is rejected that the empirical analysis has the evidence of the long run relationship between both variable. The next step is to estimate the ARDL Model with the selection of the order of the ARDL(\( n, m \)) using the AIC criteria and the following model will be estimated:\(^{16}\)

\[ \ln GE = \alpha_0 + \sum_{i=1}^{n} \psi_1 \ln GE_{t-i} + \sum_{i=1}^{n} \psi_2 \ln NSDP_{t-i} + \sum_{i=1}^{n} \psi_3 \ln Debt_{t-i} + \sum_{i=1}^{n} \psi_4 \ln Tax Rev_{t-i} + \varepsilon_t \]

(EQ. 3.21)

Then in the final step of the bound test an error correction model has been estimated in order to find the short run dynamic parameters. In this step the lagged values of the variables will be replaced with the error correction term to estimate the speed of the adjustment of the dependent variables to independent variables.

---

\(^{16}\) The model specification is same for all these versions except the change in the independent variable and dependent variable so here the study shows only for the first version model specification and it will be in similar way for rest of the versions.
The following equation is to be estimated to find the short-run relation between the variables:\textsuperscript{17}
\[
\Delta \ln GEXP_t = \gamma + \sum_{i=1}^{r_1} \Gamma_i \Delta \ln GEXP_{t-i} + \sum_{i=2}^{r_2} \Omega_i \Delta \ln NSDP_{t-i} + \sum_{i=3}^{r_3} \Psi_i \Delta \ln Debt_{t-i} + \sum_{i=4}^{r_4} II_i \Delta \ln Tax\ Re_{t-i} + \Phi ECM_{t-4} + \varepsilon_t
\]
(EQ. 3.22)

Here $r_1, \Omega, \Psi_i$ and $II_i$ are the short-run dynamic coefficients and $\Phi$ is the speed of adjustment and ECM is the error correction term. If the value of $\Phi$ is equal to zero then there is no long-run relationship between the variables and if the value of $\Phi$ is lie between the -1 to 0 then there is partial adjustment. Moreover if the value of the speed of the adjustment is less than -1 then it will be the evidence of the over adjusted model in current period. On the contrary, a positive value of the speed of the adjustment revealed that the system will be moved away from the equilibrium in the long run.

\subsection*{3.2.12 Causality Using Vector Error Correction Model}

On testing the Cointegration between the variables, the next step is to test the short run as well as the long-run relationship between the variables. The study employed the Vector Error Correction model (Naraya et al., 2008; Aregbeyen2006) to test the short run and long run equilibrium between the variable under study. A Vector Error Correction Model is a restricted VAR model which is used with the non-stationary series but cointegrated with each other. The Vector Error Correction Model has cointegration relation built in it to restrict the long run behavior of the endogenous variables to converge to cointegrating relationship in the short run adjustment dynamics. The study specifies the following VECM Model for analysis:\textsuperscript{18}

\textsuperscript{17} The model specification is same for all the versions except the change in the independent variable and dependent variable so here the study shows only for the first version model specification and it will be in similar way for rest of the versions.

\textsuperscript{18} The rest of the Models four others Version are similarly specified.
\[ \Delta \ln GE_t = \alpha_1 + \sum_{i=1}^{m} \beta_i \Delta \ln GE_{t-i} + \sum_{i=0}^{n} \phi_i \Delta \ln NSDP_{t-i} + \delta_i ECT_{t-i} + \mu_t \]  
\text{(EQ. 3.23)}

\[ \Delta \ln NSDP_t = \alpha_2 + \sum_{i=1}^{m} \beta_2 \Delta \ln GE_{t-i} + \sum_{i=0}^{n} \phi_2 \Delta \ln NSDP_{t-i} + \delta_2 ECT_{t-i} + \epsilon_t \]  
\text{(EQ. 3.24)}

Where \( \Delta \) is difference operator, \( \alpha, \beta, \delta \) and are coefficients and \( \mu_t, \epsilon_t \) are disturbance terms and ECI\(_{t-1}\) and ECT\(_{t-1}\) are error correction terms\(^{19}\) of lagged one period. The inclusion of the error correction term introduces a long run relationship through Granger causality (Akpan, 2012). If the coefficient of error correction term is negative and statistically significant than it indicates that there is long run causality between two time series variables (between GE and NSDP). In equation (3.22) the statistical significance of \( \beta_1 \phi_1, \delta_1 \) (the coefficient of \( \ln NSDP_{t-1} \) and ECT\(_{t-1}\)) will represents that the \( \ln NSDP \) cause \( \ln GE \) (which will supported by the Wagner law) and in equation (3.23) the statistical significance of the \( \beta_2, \phi_2, \delta_2 \) represent that \( \ln GE \) cause \( \ln NSDP \) (which will supported by Keynesian Hypothesis) in the long run as well as in short run. The short run causality between two variables is tested by Wald F-statistic. If all the coefficients in equation 3.22 and 3.23 are statistically significant, it is suggested that there is bidirectional causality between the GE and NSDP. It confirms that either the Wagner or the Keynesian hypothesis of public expenditure is valid.

\(^{19}\) If two variables are I (1) but there is possibility that the linear combination of these variables would be I(0). It implies that even there is short run disequilibrium between two non-stationary variables but it is possible that in long run they have co-movement or are cointegrated. In this case the VECM model used the Lag value of Error Correction Term (\( \mu_t=GE_t-\alpha-\beta NSDP_t \)) as an independent variable to find the speed of adjustment at which the short run disequilibrium will adjust for long run equilibrium.
3.2.13. Causality Using Vector Autoregressive Model (VAR)

In the absence of cointegration between the variable, the vector error correction model could not be used for detecting short run relationship between variables (Ansari et al., 1997). If there is no cointegration among variables, it may be of interest to examine the short run relationship (Gemmell, 1990 and Manning and Adriacanos, 1993). In such a situation, Vector Autoregressive Model can be applied to find the short run causal relationship between the variable. The critical condition for applying VAR is that all the variables under consideration should be stationary and VAR Model is also free from the problem of identification of endogenous and exogenous variables in the system. In case if the variables are not stationary at their level in that case we can apply VAR on the first difference of all the variables of VAR system. The testing of causality through VAR analysis is known as the standard Granger causality test which gives the short run causal relationship between variables (Demirbas, 1999 and Aregbeyen, 2006). The following equations are estimated for VAR Model\textsuperscript{20}.

\[
\Delta \ln GE_i = \alpha_i + \sum_{i=1}^{m} \theta_i \Delta \ln NSDP_{i-1} + \sum_{j=1}^{m} \gamma_j \Delta \ln GE_{i-j} + \mu_{it}
\]

(EQ. 3.24)

\[
\Delta \ln NSDP_i = \beta_i + \sum_{i=1}^{m} \upsilon_i \Delta \ln GE_{i-1} + \sum_{j=1}^{m} \eta_j \Delta \ln NSDP_{i-j} + \mu_{2i}
\]

(EQ. 3.25)

Where \(\mu_{1t}\) and \(\mu_{2t}\) are uncorrelated error term, \(m\) is the maximum lag length.

Here, in equation (11) the null hypothesis is of \(\theta_1 = \theta_2 = \theta_3 = \theta_m = 0\) against the alternative hypothesis of \(\theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4 \neq \cdots \neq \theta_m \neq 0\) would test the use of standard F-statistics. The rejection of the null hypothesis would conclude that NSDP granger causes GE. In equation 11, the null hypothesis of \(\psi_1 = \psi_2 = \psi_3 = \psi_4 = \cdots \neq 0\) against the alternative hypothesis \(\psi_1 \neq \psi_2 \neq \psi_3 \neq \psi_4 \neq \cdots\)

\textsuperscript{20} The rest of the Models four others Version are similarly specified.
$\psi_m \neq 0$. If the null hypothesis is rejected, we conclude that GE causes NSDP. There are three possibilities of any Granger causality test (Gujarati, 2003.P.697) as the following:

1. When the one null hypothesis is accepted and the other is rejected, there exists unidirectional causality. In this case, the direction of the causality will be either from Net State Domestic product to Government expenditure or from Government Expenditure to Net State Domestic Product. It amounts to the fact that either Keynes or Wagner hypothesis will be valid.

2. When both null hypotheses are rejected, it amounts to bidirectional causality and it means, there is a feedback and bilateral causality. It means that the set of coefficients are statistically significant from zero in both the regressions (in this case neither Keynes nor Wagner hypothesis is valid).

3. When both the null hypotheses are accepted, it indicates that both variables are independent of each other. In this case, the set of coefficients are not statistically significant from zero (neither the variable Granger cause to other variable).

### 3.2.14 Causality Using Toda-Yamamoto Approach

Toda-Yamamoto approach (1995) is alternative approach to the Granger causality approach and that involves testing for Granger non-causality in level VARs irrespective of whether a series is I(0), I(1) or I(2), non-cointegrated or cointegrated (Karimi, 2009, Huang, 2006). The Toda-Yamamoto approach is augmented with extra leg estimation by the order of integration of the series. It employ a modified Wald test to the restriction on parameters of the VAR (k) (k is the lag length of the VAR model). The test has an asymptotic chi-square distribution with k degrees of freedom order in the limit when a VAR (k+dmax) is estimated. Three steps involved in the estimation of Toda and
Yamamoto causality procedure\textsuperscript{21}. The first step is to test maximum order of interaction (d max) of all the time series variables by utilizing the test like the augmented dickey fuller test, Phillips- Perron (PP) test\textsuperscript{22} and the Kwiatkowski, Phillips, Schmidt and Shin test\textsuperscript{23} (KPSS) etc. The second step of the Toda-Yamamoto approach is to determine the optimal lag length (p)\textsuperscript{24}. by the use of various lag length criterion like the Akaike’s information criterion (AIC) Schwarz information criterion (SC) final prediction error (FPE) and Hannan-Quinn (HQ) Information Criterion. The final step of this procedure involved the modified Wald procedure to test the VAR (k+dmax) model for causality where optimal leg length is equal to k = (p + dmax).

The Following equations have been formulated for the analysis of the Toda-Yamamoto test of Granger causality\textsuperscript{25} with multivariate VAR (K+dmax):

\begin{align*}
\ln GE_t &= \phi_1 + \sum_{i=1}^{\text{dmax}} \beta_1 \ln GE_{t-i} + \sum_{j=1}^{\text{dmax}} \beta_2 \ln GE_{t-j} + \sum_{i=1}^{\text{dmax}} \Pi_1 \ln NSDP_{t-i} + \sum_{j=1}^{\text{dmax}} \Pi_2 \ln NSDP_{t-j} + \sum_{i=1}^{\text{dmax}} \theta_1 \ln Debt_{t-i} + \\
&\sum_{j=1}^{\text{dmax}} \theta_2 \ln Debt_{t-j} + \sum_{i=1}^{\text{dmax}} \psi_1 \ln Tax \Re v_{t-i} + \sum_{j=1}^{\text{dmax}} \psi_2 \ln Tax \Re v_{t-j} + \epsilon_{t},
\end{align*}

(EQ. 3.26)

\begin{align*}
\ln NSDP_t &= \phi_1 + \sum_{i=1}^{\text{dmax}} \alpha_1 \ln NSDP_{t-i} + \sum_{j=1}^{\text{dmax}} \alpha_2 \ln NSDP_{t-j} + \sum_{i=1}^{\text{dmax}} \gamma_1 \ln GE_{t-i} + \sum_{j=1}^{\text{dmax}} \gamma_2 \ln GE_{t-j} + \sum_{i=1}^{\text{dmax}} \omega_1 \ln Debt_{t-i} + \\
&\sum_{j=1}^{\text{dmax}} \omega_2 \ln Debt_{t-j} + \sum_{i=1}^{\text{dmax}} \phi_1 \ln Tax \Re v_{t-i} + \sum_{j=1}^{\text{dmax}} \phi_2 \ln Tax \Re v_{t-j} + \epsilon_{t}.
\end{align*}

(EQ. 3.27)

\textsuperscript{21} Null hypothesis is non-stationary.

\textsuperscript{22} with null hypothesis is stationary.

\textsuperscript{23} all these tests are used to test the stationary property of the time series data.

\textsuperscript{24} The lag length (p) is always unknown and has to be determined in the process of the VAR, among the variables in the system (in levels).

\textsuperscript{25} The model specification is same for all the versions except the change in the independent variable and dependent variable so here the study shows only for the first version model specification and it will be in similar way for rest of the versions.
Where $K$ is the optimal lag length; $d_{\text{max}}$ is the maximum order of integration $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are white noise error term with zero mean, constant variance and no autocorrelation. $\ln\text{NSDP}$ is the log value of Net State Domestic Product, $\ln\text{GE}$ is the log value of government expenditure $\ln\text{Debt}$ is the total debt, $\ln\text{Tax Rev}$ is total tax revenue. The causality from NSDP to government expenditure will be tested by examine the null hypothesis of $\ln\text{NSDP}$ does not cause to $\ln\text{GEXP}$. $H_0: \Pi_{ii} = 0 \forall i$ in equation 3.26. Similarly in equation 3.27 the null hypothesis is $H_0: \gamma_{ii} = 0 \forall i$ (for causality from government expenditure to NSDP) and the same will be tested for other determinants of the equation for causality. If the analysis will reject the null hypothesis under equation 3.26 and accept the null hypothesis of equation 3.27, it will lead the conclusion that NSDP is causing to government expenditure and if one of the hypothesis (either hypothesis under equation 3.26 or hypothesis under equation 3.27 will be rejected then it will show the unidirectional causality between both the variables. Moreover if there is indication of the rejection of both the null hypothesis (under equation 3.26 and 3.27) then it revealed the bidirectional causality between both the variables whereas no causality exists between the variables if neither the hypothesis of equation 3.26 nor the hypothesis of equation 3.27 will be rejected.

3.2.15. Impulse Response Function

Impulse Response function (IRFs) has been used to analyse the dynamic relationship between the variables under VAR System. As the Impulse response functions gives the responses of one variable on account of a shock\textsuperscript{26} in another variable (either one or more than one variables) of VAR System (Lutkepohl, 2005). The Impulse Response Function gives the insight that how and to what extent a standard deviation shock in one variable affects

\textsuperscript{26} Here shocks refers as an external change
the current and future level of the other variable of VAR System over time. Widely used method of IRFs in the literature is the Generalised Impulse response Function (Pesaran and Shin, 1998, Koop et al., 1996), as it does not considered the order of the endogenous variables in VAR System. More precisely, Generalised Impulse response Function (GIRFs) is based on arbitrary orthogonalization of shocks. In other words, it measured the effect of a unit change in each variables innovation at time ‘t’ on all the other variables of VAR system at time ‘t+j’, assuming all other innovations as constant. The method for Impulse response function in literature is Cholesky decomposition (which orthogonalized the shocks). It is based on the condition that the ordering of the variables will affect the outcome and the magnitude of a shock. Moreover, this method imposes some restrictions on the VAR system which are based on the economic theory.

3.2.16. Variance Decomposition Analysis

Variance Decomposition (VDC) analysis proposed by Enders (1995), has been employed to analyze the dynamic relationship between the variables in terms of variance decomposition from the generalized approach. The VDC gives the information about the relative importance of random innovations (Narayan P. and Narayan S., 2004). With the use of VDC analysis, the total variance in a forecast error of one variable can be decompose into two parts: (i) The percentage of variance in the forecast error of a variable as explained by it own Innovation and; (ii) The percentage of variance explained by other variables in VAR System. The VDC and IRFs are known as innovation accounting in the literature.

27 It is assumed that the error term is isolated and it will not be change from one IRFs to other IRFs.
3.2.17. Structural Break tests

Structural Break occurs when there is sudden change in economic policy or in the period of War or upheavals in the economy. A stationary time series variable may be non-stationary in the presence of structural change or sudden shift in the variable which can be examined by Unit root test with structural break. In other words, the pattern of the series may be change on account of structural Break in the series (Perron 1989, 2005). The break in a time series variable can be removed by taking the log of the series and then the break tests and applied on that log form series which is known as break in the growth rate of the variable. To analyze the structural break in the growth rate of government expenditure and NSDP of Rajasthan state the Kinked Exponential function has been used On the contrary, when there is break in the estimated parameters of the two series in linear regression than the unit root test for structural break cannot be applied. There are two types of test to detect the structural break in the linear regression of two time series variables viz.; (i) Known Break analysis with the use of Chow break test; (ii) Unknown structural Break analysis with different statistical tests which are subdivided into two- first for Unknown Single Break (Quandt Andrew Test) and Second for Unknown Multiple Break (Bai- Perron Test). The present thesis tested both the Known as well as the Unknown break analysis in Government expenditure and Economic Growth for the state of Rajasthan from 1970-71 to 2013-14.

3.2.18. Chow test for Known Break

Chow test (1960) is a conventional test and widely used in literature for exogenous break analysis in the parameters of the regression analysis. The Chow test of Structural Break utilized the F-test to analyse whether a single

---

28 The kinked exponential growth function has been discussed in detailed in the present chapter in section 3.2.2.
regression is more efficient than the two separate regressions which divides the sample into two sub-samples (Hansen, 2001). In order to analyze significant structural break for Year 1991-92 (Economic Reform) in the data series, the study divide the data into two sub-periods to Pre and Post Reform. The Chow test estimates the parameters for both the sub-samples by simple linear regression and then compares the parameters of both the sub-samples by the following F-Test:

\[
F = \frac{[RSS_1 - (RSS_2 + RSS_3)]/k}{(RSS_2 + RSS_3)/(n_1 + n_2 - 2k)}
\]  
(EQ. 3.28)

Where RSS\(_1\) is the residual sum of squares of the estimated model for the entire period, RSS\(_2\) is the residual sum of squares of the estimated model for the 1st sub-period (Pre-Reform Period), RSS\(_3\) is the residual sum of squares of the estimated model for the 2nd sub-period (Post-Reform period), \(n_1\) = number of observation in 1st sub-sample, \(n_2\) = number of observation in 2nd sub-sample, \(K\) = number of explanatory variables in the model. The null hypothesis of the Chow Test is that there is no structural break. The estimated F-test is significant than the null hypothesis of no structural break can be rejected and can be conclude that there is break in the regression analysis of both the sub-periods under consideration. As mention elsewhere, there is several limitation of Chow Break Test. According to literature the chow test is based on the difference between the parameters of two time period with known break date, which is the major limitation of chow test. To perform the chow test one have to have an information about the break date of the series before go to analysis, for which there are two options, first is to take an arbitrary break date and second take the break data on the basis of characteristics of the data series. The application of chow breakpoint analysis probably leads to misleading results as the break date is endogenously determined which may not be the true break date for estimated equation. So it has as series problem that it will not analyze
the unknown and more than one break. To analyze the unknown break single and multiple breaks the study used the modern econometrics techniques.

3.2.19. Quandt-Andrews Test (Single Unknown Break Test)
Quandt-Andrews is testing the single endogenous structural break in the regression analysis of two time series. Quandt-Andrews utilized a single Chow’s Breakpoint Test at every observation between two periods \((t_0 \text{ and } t_1)\) and then summarized \(n\) test statistics obtained from the Chow’s Breakpoint into one test statistic within the trimmed data at both the ends of the series. The null hypothesis of Quandt Andrew test is that there is no break point between \(t_0\) and \(t_1\). Quandt (1958, 1960) proposed likelihood ratio test statistics for an unknown break also known as Supremum (Max)-Test (Break date decide on the basis of maximum value of Chow Test Statistics), Andrews (1993) developed the Wald and Lagrange Multiplier test statistics, Andrews and Ploberger (1994) Proposed Exponential (LR, Wald and LM) and Average (LR, Wald and LM) tests. The Average and Exponential test Statistics are based on all Chow statistics, but these tests do not gives the information on the break data of the series. Out of all the tests Statistic of Quandt-Andrews Test only Max Statistic gives the Break date where the actual endogenous break occurs. On the other side, the Max Statistics is based on the assumption that residuals from the regression must be homogeneous before and after the break and it is based on the alternative hypothesis of only one break against the Null hypothesis of no break.

3.2.20. Bai-Perron Test (Unknown Multiple Breaks)
In time series regression there is the possibility of more than one break in linear regression of two time series variables. Bai- perron test (1998, 2001) has been utilized to detect the multiple breaks in the linear regression of time series variable under study. As the Bai-Perron test is based on the null hypothesis that there is one break against the alternative hypothesis of more
than one break in the regression analysis of two time series variable. It is based on the assumption of consistent estimation to detect the Number and Location of the breakpoints and parameters. There are several methods to determine the number of breaks and the distance between the break points viz; Sequential L+1 breaks vs L Break, Global L Breaks vs none, L+1 breaks vs global L break etc. The global method detects for L Break against the no break. On the contrary, the Sequential test detects for L+1 break against L Break test. If the test reject the null of L breaks than the series will be subdivided into two sub-samples and the same procedure will be applied on the each sub-sample and it will be continuous till the L break hypothesis is accepted.

3.2.21. Diagnostic Tests

Once the causality will be confirmed between the variable the next important task of analysis is to test whether the fitted model is efficient or not to make the inferences for both the variable on the basis of findings of the model. The present study used various diagnostic tests to check the efficiency of the fitted model which are following-

3.2.22. JarqueBera (JB) Test of Normality - The JB test is used to test that the residuals of the model are normally distributed or not. It is based on the residuals of OLS regression\(^{29}\). The null hypothesis of JB test is that the residuals of the model are normally distributed. If the null hypothesis of the JB

\[^{29}\] The JB test Jointly test the Skewness and Kurtosis of residuals of OLS Regression that the value of the Skewness is equal to zero and Kurtosis is equal to 3 and used the following test statistic-

\[ JB = n \left[ \frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \]

where \( n \) is number of Observation, \( S \) and \( K \) are stands for Skewness and Kurtosis coefficient. In case of normal distribution the value of \( S \) and \( K \) should be 0 and 3 respectively and in that case the value of JB test statistic will also be zero (Gujarati, 2007, PP-151).
The ARCH test will be accepted then it is the indication that the residuals are normally distributed.

3.2.23. **Autoregressive Conditional Heteroscedasticity (ARCH) Test** - the ARCH test is used to test the properties of Homoscedasticity of the residuals series in time series analysis. The null hypothesis of the ARCH Test is that there is no ARCH effect against the alternative hypothesis of ARCH effect. If the null hypothesis of ARCH test will be accepted then it can be concluded that the model is free from the problem of Heteroscedasticity.

3.2.24. **Breusch-Godfrey (BG) test** - Breusch-Godfrey test is used to test the serial correlation problem in time series data, as BG test is based on Lagrange Multiplier Principle so it is also known as the LM test for serial correlation. The null hypothesis of LM test is that there is no serial correlation between the successive residuals of OLS regression.

3.2.25. **Cumulative Sum (CUSUM) and Cumulative Sum of Square (CUSUM Square) Test** - CUSUM and CUSUM Square tests are used to check the stability of parameters of the estimated models. In other words, these tests are used to test the consistency of parameters (Johnston and DiNardo, 1997). The CUSUM test is based on the cumulative sum of the errors in regression whereas the CUSUM Square is based on the squares of these errors. The test plot the Sum and Square of sum of Recursive residuals with 5 % critical lines, if the line of Sum and Squared of Sum will cross these two critical lines then is the indication of parametric instability of the model.

---

30 The ARCH test is applied on squared of residuals and it tests whether the variance of the residuals are constant over time or not if the variance of the residuals are not constant over time then is known as the problem of Heteroscedasticity.

31 The BG test is based on Auxiliary regression for residuals. The test statistic of LM test is Known as Obs*R-squared statistic, which is simply multiplication of R2 with the number of observation (T*R2) (Gujarati, 2007, PP.-484).