

CHAPTER- IV

DATA AND METHODOLOGY

4.1. Introduction

This chapter deals with the nature and source of data sets used in the study. Methodological issues are also being considered and explained. Explanations of some estimation procedures along with their implications are being incorporated in this chapter.

4.2 Nature, Period of Study and Source of Data

The study is based on time series datasets. Annual time-series data for *budget deficit* and *trade deficit* for the country of Maldives have been used. The base period is 1995 (1995 = 100). The datasets cover the period 1979-2003. The datasets have been collected from different Year Books of *International Financial Statistics*, published by the IMF.

4.3 Methodology:

The methodology used for the study is basically econometric and time series analysis. Time series methods encompasses both the '*Time Domain*' and '*Frequency Domain*' analyses. The time series methodology involves the tests for '*Stationarity*', '*Integrability*', '*Cointegration*', '*Vector Error Correction Modeling (VECM)*', application of '*Unrestricted*' and '*Restricted Vector Autoregression (VAR) Models*', '*Intervention Analysis*' through the applications of '*Impulse Response Functions*', as well as '*Variance Decomposition*', '*Spectral Analysis*' involving the study of '*Periodograms*', '*Autospectrum*', '*Cospectrum*' '*Coherency Spectrum*,' '*Gain Spectrum*', '*Phase Spectrum*' etc.

A brief features of some of the estimation techniques and methodological issues are being described below.

4.4 Stationarity and Unit Root Test :Relevance

Econometric analysis requires that the time series data-sets be 'stationary'. Non-stationary data contain unit roots. The main objective of unit root tests is to determine the degrees of integration of each individual time series. Various methods for unit root tests have been applied in this study. Some of these are being explained below.

4.5 Dickey-Fuller Unit Root Test

For the Dickey- Fuller tests, the relevant model is

$$y_t = \beta_0 + \beta_1 t + u_t \quad (4.1)$$

$$\text{where } u_t = \alpha u_{t-1} + \epsilon_t \quad (4.2)$$

Here ϵ_t is a covariance stationary process with zero mean. the reduced form for this model is,

$$y_t = \gamma + \delta t + \alpha y_{t-1} + \epsilon_t \quad (4.3)$$

$$\text{where } \gamma = \beta_0(1 - \alpha) \text{ and } \delta = \beta_1(1 - \alpha)$$

This equation is said to have a unit root if $\alpha = 1$ (in which case $\delta = 0$)

4.6 Augmented Dickey-Fuller Unit Root Test

In order to test for the existence of unit roots, and to determine the degree of differencing necessary to induce stationarity, we have applied the 'Augmented Dickey-Fuller test'. Dickey and Fuller (1976, 1979), Said and Dickey (1984), Phillips (1987), Phillips and Perron (1988), and others developed modified version of the Dickey-Fuller tests when the error term i.e ϵ_t is not *white noise*. These tests are called 'Augmented Dickey-Fuller test' (ADF). The results of the *Augmented Dickey-Fuller test* (ADF) help to determine the form in which the data should be applied in any econometric analysis. The alternative forms are as follows.

$$\Delta y_t = \gamma + \alpha y_{t-1} + \sum_{j=2}^k \theta_j \Delta y_{t-j+1} + e_t \quad (4.4)$$

$$\Delta y_t = \gamma + \delta_t + \alpha y_{t-1} + \sum_{j=2}^k \theta_j \Delta y_{t-j+1} + e_t \quad (4.5)$$

$$\Delta y_t = \alpha y_{t-1} + \sum_{j=2}^k \theta_j \Delta y_{t-j+1} + e_t \quad (4.6)$$

where, y_t = Model Variables (TD_t and BD_t) of Maldives;

Δy_t = First differenced series of y_t .

Δy_{t-j+1} = First differenced series of y_t at $(t-j+1)^{\text{th}}$ lags. ($j = 2 \dots k$)

The equation (4.4) contains a constant as exogenous, while equation (4.5) bears a constant along with a linear trend. However, equation (4.11) presents an auto-regressive process with no constant and linear trend.

4.7 The Dickey-Fuller- GLS Unit Root Test

The DF-GLS test has been developed by Elliott, Rothenberg and Stock (1996), and the test possesses greater precision than ADF test in identifying *non-stationarity*. The DF-GLS t-test is performed by testing the hypothesis $a_0 = 0$ in the regression equation.

$$\Delta y_t^d = a_0 y_t^d + a_1 \Delta y_{t-1}^d + \dots + a_p \Delta y_{t-p}^d + \text{error} \quad (4.7)$$

where y_t^d is the locally de-trended series y_t . The local de-trending depends on whether we consider a model with drift only or a linear trend.

- (i) DF-GLS unit root test without time trends (a model with drift only):

$$y_t^\mu = \alpha y_{t-1}^\mu + \sum_{i=1}^k \Psi_i \Delta y_{t-i}^\mu + u_t \quad (4.8)$$

(ii) DF-GLS unit root test with time trends (a model with linear trend):

$$y_t^\tau = \alpha y_{t-1}^\tau + \sum_{i=1}^k \Psi_i \Delta y_{t-i}^\tau + u_t \quad (4.9)$$

4.8 Phillips –Perron Unit Root Test

Phillips (1987), Phillips and Perron (1988) have generalized the Dickey –Fuller (DF) tests to the situations where disturbance processes, ϵ_t are serially correlated. The PP is intended to add a ‘correction factor’ to the DF test statistic.

Let the AR (1) model be,

$$Y_t = \mu + \Phi_1 Y_{t-1} + \epsilon_t, [t = 1 \text{ --- } T] \quad (4.10)$$

with $\text{Var}(\epsilon_t) \equiv \sigma_\epsilon^2$

If ϵ_t is serially correlated , the ADF approach is to add lagged ΔY_t to ‘whiten’ the residuals . The test statistic $T(\Phi_1-1)$ has been considered which is distributed as ρ_μ in the maintained regression with an intercept but with no time trend. The PP modified version is

$$Z_{\rho_\mu} = T(\Phi_1-1)\text{-CF} \quad (4.11)$$

where the correction factor CF is

$$CF = 0.5 (s_{T1}^2 - s_\epsilon^2) / (\sum_{t=2}^T (Y_{t-1} - \bar{Y}_{-1})^2 / T^2) \quad (4.12)$$

$$\text{and, } s_\epsilon^2 = T^{-1} \sum_{t=1}^T \epsilon_t^2 \quad (4.13)$$

$$s_{T1}^2 = s_{2\epsilon}^2 + 2 \sum_{s=1}^l W_{sl} \sum_{t=s+1}^T \epsilon_t \epsilon_{t-s} / T \quad (4.14)$$

where $W_{sl} = 1-s/(1+l)$ and $\epsilon_t = Y_{t-\mu} - \Phi_1 Y_{t-1}$

$$\bar{Y}_{-1} = (\sum_{t=2}^T Y_t) / (T-1) \quad (4.15)$$

4.9 KPSS Unit Root Test

Another frequently used method for the test of stationarity is the KPSS test developed by Kwiatkowski et al. the KPSS test is an analog of Phillips –Perron test. The model for KPSS test is:

$$\Phi(L)y_t = \alpha_t + \beta t + \epsilon_t \quad (4.16)$$

where $\alpha_t = \alpha_{t-1} + \eta_t, \alpha_0 = \alpha \quad (t = 1, 2, \dots, T)$

and $\epsilon_t \sim \text{IID}(0, \sigma_\epsilon^2), \eta_t \sim \text{IID}(0, \sigma_\eta^2)$

ϵ_t and η_t are independent and $\Phi(L)$ is a p^{th} order lag polynomial. The relevant hypothesis for the test of stationarity in this model are,

$$H_0 : \sigma_{\eta}^2 = 0$$

Against,

$$H_1 > 0$$

Under H_1 the process defines an ARIMA (p,d,q) model structure.

It has been argued that tests with stationarity as null can be used to confirm the results of the usual unit root tests. The relevant two tests are

Test 1 (usual test)	Test 2 (KPSS test)
$H_0 : y_t$ is nonstationary (unit root)	$H_0 : y_t$ is stationary
$H_1 : y_t$ is stationary	$H_1 : y_t$ is nonstationary (unit root)

If both tests reject their nulls, there will be no confirmation. But if test 1 rejects the null but test 2 does not (or vice versa) the confirmation can be drawn (G.S.Maddala 2001:553).

4.10 ERS Point Optimal Test

The *ERS point Optimal* test is largely based on the following quasi-differencing regression equation:

$$d(y_t / \alpha) = d(x_t / \alpha)' \delta(\alpha) + \eta_t \quad (4.17)$$

where x_t stands for either a constant or a constant along with trend and $\delta(\alpha)$ be the ordinary least square (OLS) estimates. The residual from this equation is

$$\eta_t(\alpha) = d(y_t / \alpha) - d(x_t / \alpha)' \delta(\alpha) \quad (4.18)$$

let, $SSR(\alpha) = \sum \eta_t^2(\alpha)$ be the sum of squared residuals function. The ERS point optimal test statistic of the “null that $\alpha=1$ ” against “the alternative that $\alpha = \bar{\alpha}$ ”, is then defined as

$$P_T = SSR(\bar{\alpha}) - \bar{\alpha} SSR(1) / f_0 \quad (4.19)$$

where, f_0 is an estimator of the residual spectrum at frequency zero. In order to compute the ERS test, it is necessary to specify the underlying set of exogenous regressors x_t and a technique for estimating f_0 .

4.11 NG and Perron (NP) Test

Ng and Perron (2001) estimate four test statistics that are based upon the GLS de-trended data y_t^d . These test statistics are modified forms of Phillips and Perron Z_{α} and Z_t statistics, the Bhargava (1986) R_1 statistics and the ERS Point Optimal statistic. The terms are defined as

$$k = \sum_{t=2}^T (y_{t-1}^d)^2 / T^2 \quad (4.20)$$

The modified statistics can be represented as

$$\begin{aligned}
 MZ_{\alpha}^d &= [T^{-1}(y_t^d)^2 - f_0] / 2k \\
 MZ_{\alpha}^d &= MZ_{\alpha} \times MSB \\
 MSB^d &= (k / f_0)^{1/2} \\
 MP_t^d &= \{\bar{c}^2 k - \bar{c} T^{-1}(y_t^d)^2 / f_0 \quad \text{if } x_t = \{1\}\} \\
 &= \{\bar{c}^2 k - (1 - \bar{c}) T^{-1}(y_t^d)^2 / f_0 \quad \text{if } x_t = \{1, t\}\} \\
 &\text{where, } \bar{c} = -7 \quad \text{if } x_t = \{1\} \\
 &= -13.5 \quad \text{if } x_t = \{1\}
 \end{aligned} \tag{4.21}$$

The NP tests require a specification for x_t and a choice of method for estimating f_0 .

4.12 Correlogram

One of the simple, intuitive and interesting methods of testing ‘stationarity’ is running a correlogram. Correlogram is nothing but simply a graphical representation of Auto-correlation Function (ACF) and Partial Auto-correlation Function (PACF). The nature of stationarity can also be found almost accurately in most of the cases with the help of Correlogram.

4.13 Cointegration Test

The *cointegration* test represents the gesticulation of long- run equilibrium relationship between two variables say y_t and x_t . Let both be integrated at one, that is $y_t \sim I(1)$ and $x_t \sim I(1)$. Then y_t and x_t are said to be cointegrated if there exist some β such that $y_t - \beta x_t$ is $I(0)$. This is denoted by saying y_t and x_t are CI (1,1) i.e y_t and x_t are *cointegrated*. Different types of *cointegration* techniques are available for the time series analysis. These tests include the *Engle and Granger Test* (1987), *Stock and Watson Procedure* (1988) and *Johansen’s Method* (1988, 1990, 1991)

4.14 Engle Granger Cointegration Test

This approach is also known as a residual test. If the variables included in an equation are integrated of the same order, say (1), the error term should be stationary, i.e. $I(0)$. Let us consider M time horizon time series ($Y_{t1} \dots Y_{tM}$), each of which is $I(1)$, Let one of the two regression model include a drift and no trend while the second one contains a drift and trend such that

$$Y_{t1} = \beta_0 + \sum_{j=2}^M \beta_j y_{t-j+1} + \varepsilon_t \tag{4.22}$$

$$Y_{it} = \beta_0 + \beta_1 t + \sum_{j=2}^M \beta_j y_{t-j+1} + \varepsilon_t \quad (4.23)$$

A test for no co-integration is assigned by a test for a unit root in the estimated error e_t . This can be achieved by applying ADF test to the residuals using the following equation:

$$\Delta e_t = \alpha e_{t-1} + \sum_{j=1}^p \Phi_j e_{t-j} + v_t \quad (4.24)$$

The null hypothesis $\alpha = 0$ is tested using the τ statistic.

4.15 Johansen Maximum Likelihood Cointegration Test

The *Johansen Maximum Likelihood Procedure* examines the relationship among stationary or non-stationary variables using the following equation:

$$x_t = \sum_{i=1}^p \Pi_i x_{t-i} + \varepsilon_t \quad (4.25)$$

The function can be presented according to the following VAR system:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^p \Phi_i \Delta x_{t-i} + \mu + \varepsilon_t \quad (4.26)$$

In which x_t is an $n \times 1$ random vector, ε_t is NIID $(0, \sigma^2_\varepsilon)$, and μ is deterministic terms. The long run relationships are derived through the coefficient Matrix of Π , denoted by r , which is between 0 and n . Then there are r linear combinations of the variables in the system that are $I(0)$ or *cointegrated*. Under Johansen (1991), and Johansen and Juselius (1990) procedures, two tests are available for the determination of *cointegrating vectors* and for the estimation of their values. These tests are the *Trace Test* and the *Eigen Value Test*. In this method a two-stage testing procedure has been followed. In the first stage, the '*null hypothesis of no Cointegration*' hypothesis is tested against the alternative hypothesis that the data are cointegrated with an unknown cointegrating vector. If the null hypothesis is rejected, a second stage test is implemented with Cointegration maintained under both the null and alternative. Gonzalo (1994) has suggested that Johansen's procedure has certain properties which are superior to alternative cointegrating testing methods.

4.16 Vector Error Correction Modeling

Vector Error Correction Modeling provides important information on the short run relationship (short run dynamics) between any two cointegrated variables. *Vector Error Correction* test has provided empirical evidence on the short-run causality between the trade deficits and budget deficits in Maldives.

In the present study the vector error correction estimates have been specified using the following models.

$$\Delta TD_t = \gamma_1 + \rho_1 z_{t-1} + \alpha_1 \Delta TD_{t-1} + \alpha_2 \Delta TD_{t-2} + \alpha_3 \Delta BD_{t-1} + \alpha_4 \Delta BD_{t-2} + \epsilon_{1t} \quad (4.27)$$

$$\Delta BD_t = \gamma_2 + \rho_2 z_{t-1} + \beta_1 \Delta TD_{t-1} + \beta_2 \Delta TD_{t-2} + \beta_3 \Delta BD_{t-1} + \beta_4 \Delta BD_{t-2} + \epsilon_{2t} \quad (4.28)$$

where, ΔTD_t = first difference of the trade deficit (real); ΔBD_t = first difference of Budget deficit (real); z_{t-1} = first lag of error term of cointegrating equation, ϵ_{1t} and ϵ_{2t} are white noise errors; $\alpha_1, \alpha_2, \alpha_3,$ and α_4 and $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients of lagged variables in the above models.

The focus of the vector error correction analysis is on the lagged z_t terms. These lagged terms are the residuals from the previously estimated cointegrating equations. In the present case the residuals from two lag specifications of the cointegrating equations have been used in the vector error correction estimates. Lagged z_t terms provide an explanation of short-run deviations from the long-run equilibrium for the test equations above. Lagging these terms means that disturbance of the last period impacts upon the current time period. Statistical significance tests are conducted on each of the lagged z_t terms in equations (4.27) and (4.28) in general. Finding a statistically insignificant coefficient of the z_t term implies that the system under investigation is in the short-run equilibrium as there are no disturbances present. If the coefficient of the z_t term is found to be statistically significant, and then the system is in the state of the short-run disequilibrium. In such a case the sign of z_t term gives an indication of the causal direction between the two test variables.

4.17 Granger Causality Tests:

For the sake of simplicity, let γ_{1t} and γ_{2t} be the two variables under study. Along this course, the focal idea that γ_{1t} is not *Granger-Caused* by γ_{2t} if the optimal predictor of γ_{1t} does not use information from γ_{2t} . While applying this idea, the predictor is usually restricted to be an optimal '*linear*' predictor and optimality is defined as minimizing the *mean squared error* (MSE) of the h-step predictor of γ_{1t} .

In order to be more specific, let γ_{1t} and γ_{2t} have vector autoregressive representation (VAR) in which γ_{1t} depends upon its own lags and lags of γ_{2t} and symmetrically γ_{2t} depends upon its own lags and lags of γ_{1t} .

Let us put the aforesaid VAR specification as following.

$$\gamma_{1t} = \mu_{10} + \pi_{11.1} \gamma_{1t-1} + \dots + \pi_{11.p} \gamma_{1t-p} + \pi_{12.1} \gamma_{2t-1} + \dots + \pi_{12.p} \gamma_{2t-p} + \epsilon_{1t} \quad (4.29)$$

$$\gamma_{2t} = \mu_{20} + \pi_{21.1} \gamma_{1t-1} + \dots + \pi_{21.p} \gamma_{1t-p} + \pi_{22.1} \gamma_{2t-1} + \dots + \pi_{22.p} \gamma_{2t-p} + \epsilon_{2t} \quad (4.30)$$

Here, first subscript denotes the variable and the second subscript denotes the observation index. This representation is bi-variate p^{th} order VAR. Since we are having a system of two

equations (4.29) and (4.30), the errors may be contemporaneously correlated. Therefore, any shock to one of the equations would have ‘*ripple effect*’ on the other equation.

For the specification of the ϵ_{it} , $i=1,2$; is assumed to have an innovation with zero mean, constant variance and no serial correlation that is $E\{\epsilon_{it}\} = 0$; $E\{\epsilon_{it}^2\} = \sigma^2_i$ for $i=1,2$ and $E\{\epsilon_{it}, \epsilon_{is}\} = 0$ for $t \neq s$ and $i=1,2$. Again it is also assumed that $E\{\epsilon_{1t}, \epsilon_{2s}\} = 0$ for $t \neq s$ for no serial correlation. Additionally, the ‘*ripple effect*’ is captured by the covariance between ϵ_{1t} and ϵ_{2t} , denoted by σ_{12} , which is assumed to be constant.

Along this course, for all t , the *error variance matrix* for the VAR with p lags will be

$$\Omega(\epsilon, p) \equiv E \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} (\epsilon_{1t} \epsilon_{2t}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad (4.31)$$

The equations (4.29) and (4.30) may be presented in a matrix form;

$$\begin{pmatrix} \gamma_t^1 \\ \gamma_t^2 \end{pmatrix} = \begin{pmatrix} \mu_{10} \\ \mu_{20} \end{pmatrix} + \begin{bmatrix} \pi_{11.1} & \pi_{12.1} \\ \pi_{21.1} & \pi_{22.1} \end{bmatrix} \begin{pmatrix} \gamma_{t-1}^1 \\ \gamma_{t-1}^2 \end{pmatrix} + \dots + \begin{bmatrix} \pi_{11.p} & \pi_{12.p} \\ \pi_{21.p} & \pi_{22.p} \end{bmatrix} \begin{pmatrix} \gamma_{t-p}^1 \\ \gamma_{t-p}^2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad (4.32)$$

Now again, a third subscript is necessary on the coefficients to make a distinction of lag length which runs from 1 lag through p lags. In that case, it may be written as

$$\gamma_t = \mu + \Pi_1 \gamma_{t-1} + \dots + \Pi_p \gamma_{t-p} + \epsilon_t \quad (4.33)$$

where, $\gamma_t = (\gamma_{1t}, \gamma_{2t})$, $\mu = (\mu_{10}, \mu_{20})$ and Π_i are 2×2 matrices defined in response to equation (4.32). If any of the coefficients of γ_{2t-1} ($i=1 \dots \beta$) significant, that γ_{2t} *Granger Causes* to γ_{1t} . Similarly, if any of the coefficients of γ_{1t-1} ($i=1 \dots \beta$) is significant, then γ_{2t} is *Granger Caused* by γ_{1t} .

4.18 Vector Auto-regression (VAR) Modeling

While testing the long-run dynamic relationship between model variables concerned, we may not make any *a priori* assumption of endogeneity and exogeneity of variables concerned. In such situation, *Vector Auto-regressive Model (VAR)* can be applied. This model treats all variables systematically without making reference to the issue of dependence or independence. A *VAR* additionally offers a scope for *Intervention Analysis* through the study of ‘*Impulse Response Functions*’ for the endogenous variables in the model. Moreover, a *VAR* model allows us to study the ‘*Variance Decompositions*’ for these variables and this help us understand the inter-relationships among the variables concerned. We, therefore, seek to develop following models for the Twin Deficit Relationship for the economy of Maldives.

4.18.1 The VAR Model

The *Vector Autoregression (VAR)* model for trade deficit and budget deficit for the economy of Maldives consists of the equation (4.34) and (4.35)

$$\Delta TD_t = \alpha_1 + \sum_{i=1}^k \beta_{1i} \Delta TD_{t-i} + \sum_{i=1}^k \gamma_{1i} \Delta BD_{t-i} + u_{1t} \quad (4.34)$$

$$\Delta BD_t = \alpha_2 + \sum_{i=1}^k \beta_{2i} \Delta BD_{t-i} + \sum_{i=1}^k \gamma_{2i} \Delta TD_{t-i} + u_{2t} \quad (4.35)$$

where, α_i , $i=1,2$ intercepts,

u_{1t} and u_{2t} = Stochastic error terms (alternatively called as impulses or innovations or shocks

in VAR Modeling), $\sum_{i=1}^k \beta_{1i} \Delta TD_{t-i}$ and $\sum_{i=1}^k \gamma_{2i} \Delta TD_{t-i}$ represent all summation values of

lagged terms of trade deficit in the model.

$\sum_{i=1}^k \gamma_{1i} \Delta BD_{t-i}$ and $\sum_{i=1}^k \beta_{2i} \Delta BD_{t-i}$ represent all summation values of lagged variables of

budget deficit in the model.

Furthermore, the VAR model consisting of equations (4.34) and (4.35) which requires that

- (i) ΔTD_t and ΔBD_t be stationary, and
- (ii) u_{1t} and u_{2t} be white noise terms such that

$$u_{1t} \sim \text{iid } N(0, \sigma^2 u_1), \text{ and}$$

$$u_{2t} \sim \text{iid } N(0, \sigma^2 u_2)$$

4.19 Impulse Response Function

Any shocks to any variables not only directly affects the respective variable only, but also it would be transmitted to all of the endogenous variables in the model through dynamic (lag) structure of VAR. An Impulse Response Function tries to find out the effect of one time shock to one of the innovations on current and future values of the endogenous variables. Due to this feature, Impulse Response Function (IRS) in VAR System is widely used in describing the dynamic behavior variables in the system concerned to shocks in the residual of the time series under study.

Innovations are normally correlated and may be viewed as having common properties which cannot be associated only to a specific variable. In order to explain the impulses, it is widely applied for a transformation (P) to the innovations so that they become uncorrelated such that

$$V_t = P\varepsilon_t \sim (0, D)$$

where, $D = \text{Diagonal co-variance matrix}$.

4.20 Variance Decomposition

Specifically, *Impulse Response Function* discovers the effects of a shock to one and thereby transmitted to other endogenous variables in the VAR System. However it cannot tell us the magnitude of shocks in the system. To overcome this problem, Variance Decomposition mechanism is applied to separate out the variation in an endogenous variable into the constituent shocks to the VAR System. So Variance Decomposition is applied in the models to find out the information about relative importance of every random innovation in the question of its effects to the variables concerned in the VAR system.
