CHAPTER FOUR

STATISTICAL AND ECONOMETRIC MODELS

A proper methodology for obtaining robust findings from available data is a basic principle for good analysis in any discipline. Since the present study is an empirical one, it is very important to choose models which can fit the data and meet the terms of the research without compromising the theoretical and technical demands. This chapter presents the econometric and statistical models that have been used in the present study.

The chapter is presented as follows. Section 4.1 describes the different methods used in the literature for calculating growth rate. Here an attempt has been made to specify the importance and necessity of these different methods of calculations according to purpose and availability of data set. Section 4.2 discusses the methodology in relation to stability test. The relative merits and demerits of different stability test procedures have also been discussed here. Section 4.3 presents the method of decomposing growth rate of an aggregate variable into its constituents. The literature of this section is highly inspired by Balakrishan and Parameswaran (2007). Section 4.4 has been devoted to develop the methodology in order to test convergence hypothesis. Methodology regarding both the concepts of convergence, namely, beta convergence and sigma convergence have been detailed here. Section 4.5 deals with statistical methods used to calculate measures of dispersion. Section 4.6 describes the econometric methodology for testing inter-linkage among variables. Models for testing stationarity property of time series and panel data series have been discussed in section 4.7. Finally, section 4.8 discusses the issues in relation to co-integration.

4.1: Measurement of Growth Rate

In the discipline of economics, undoubtedly the concept of ‘computation of growth rates’ has been used maximum number of times during the past four decades or so in research studies (see, for example, Panse, 1964; Dey, 1975; Reddy, 1978; Narain et al., 1985; Kumar and Rosegrant, 1994; Joshi and Saxena, 2002; Singh and Srivastava, 2003). In fact, one full issue of Indian Journal of Agricultural Economics...
(Vol. 35, No. 2) in 1980 was devoted exclusively to presentations of its various aspects. Growth rate shows the percentage change in the value of a variable from one period to another/the next, or the average change over a number of periods. For the calculation of the growth rates in the present study two methods have been applied to suit the specific nature of the data. These methods are: Simple One Period Growth Rate and Least-Square Growth Rate.

4.1a: Simple One Period Growth Rate

In order to calculate growth rate of a variable in period \( t \) over period \( t-1 \), following simple method has been applied:

\[
g_{t-1,t} = \left( \frac{Y_t - Y_{t-1}}{Y_{t-1}} \right) \times 100 = \left( \left( \frac{Y_t}{Y_{t-1}} \right) - 1 \right) \times 100 ............. (4.1.1)
\]

Here \( g_{t-1,t} \) is percentage growth/change in variable \( Y \) from period \( t-1 \) to \( t \). The one period growth rate has been used to measure growth of a variable from one period to the next.

4.1b: Least-Square Growth Rate

Least-squares growth rate can be used whenever there is a sufficiently long time series to permit a reliable calculation. The least-square growth rate is estimated by fitting a linear regression trend line to the logarithmic (annual) values of the variable in the relevant period. To illustrate, the formula for the one-period growth rate can be re-arranged as:

\[
Y_t = Y_{t-1} (1 + g_{t-1,t}) ................. ......................... (4.1.2)
\]

Since:

\[
Y_1 = Y_0 (1 + g) \\
Y_2 = Y_1 (1 + g) \\
Y_3 = Y_2 (1 + g) \\
\cdots \\
Y_t = Y_{t-1} (1 + g)
\]

It follows that:
Taking natural logarithm of the above equation, we can write

$$\ln Y_i = \ln Y_0 + t \ln(1 + g) \quad \text{.................................... (4.1.4)}$$

Now letting

$$\alpha = \ln Y_0$$
$$\beta = \ln(1 + g)$$

We can write (4.1.4) as

$$\ln Y_i = \alpha + \beta t \quad \text{.................................... (4.1.5)}$$

Adding the disturbance term to (4.1.5), we obtain

$$\ln Y_i = \alpha + \beta t + u, \quad \text{.................................... (4.1.6)}$$

This model is like any other linear regression model in that the parameters $\alpha$ and $\beta$ is linear. If $\hat{\beta}$ is the estimate of $\beta$, then the average annual compound growth rate is obtained as:

$$g = e^{\hat{\beta}} - 1 \quad \text{.................................... (4.1.7)}$$

Similarly, given the time serious data, we can also calculate exponential growth rate with the help of least square technique. The exponential growth rate is calculated by following formula

$$Y_i = Y_0 e^{\hat{\theta}} \quad \text{.................................... (4.1.8)}$$

Logarithmic transformation of this function gives

$$\ln Y_i = \ln Y_0 + gt \ln e \quad \text{.................................... (4.1.9)}$$

Or,

$$\ln Y_i = \ln Y_0 + gt \quad \{\text{Since } \ln e = 1\} \quad \text{.................................... (4.1.10)}$$

Now letting

$$\alpha = \ln Y_0$$
$$\beta = g$$

we can write equation (4.1.10) as

$$\ln Y_i = \alpha + \beta t \quad \text{.................................... (4.1.11)}$$

Adding the disturbance term to (4.1.11), we obtain
\[ \ln Y_t = \alpha + \beta t + u_t \]  \hspace{1cm} (4.1.12)

This model is also linear regression model in that the parameters \( \alpha \) and \( \beta \) are linear. If \( \hat{\beta} \) is the estimate of \( \beta \), then the average annual exponential growth rate is given by the value of \( \hat{\beta} \).

4.2: Test for Structural Shift

In estimating growth rates by least square technique in the last section, it is implicitly assumed that the parameter \( \hat{\beta} \) is constant for the entire sample, both for the data period used to estimate the model, and for any subsequent period used in the construction of forecasts. This implicit assumption can be tested using parameter stability tests.

The classical test for parameter stability is attributed to Chow (1960). This test procedure involves splitting the sample into two sub-periods, estimating the parameters for each of the sub-periods and testing for the equality of the two sets of parameters using the F-statistic. If the value of the test statistic is greater than the critical value from the F-distribution, then reject the null hypothesis that the parameters are stable over time. To illustrate suppose we want to test whether there has been a structural change at time \( t = n_1 \). The procedure is to divide the sample of \( n \) observations into two groups- group 1 consisting of the first \( n_1 \) observations and group 2 consisting of the remaining \( n_2 = n - n_1 \) observations. Then the model is estimated separately for each of the two sample groups and the sum of squared residuals \( ESS_1 \) and \( ESS_2 \) are computed. The unrestricted sum of squares is therefore given by \( ESS_u = ESS_1 + ESS_2 \). If we divide this by \( \sigma^2 \), this will have a chi-square distribution with degree of freedom (d.f) \( n_1 - k + n_2 - k = n - 2k \). Now assuming that the regression coefficient is the same before and after period \( n_1 \), we estimate the model again but with the pooled sample, and obtain \( ESS_p \). The appropriate test statistic is then

\[
F_c = \frac{(ESS_p - ESS_1 - ESS_2) \times k}{(ESS_1 + ESS_2) + (n - 2k)} \hspace{1cm} (4.2.1)
\]
The test procedure is to reject the null hypothesis that there is no structural change if \( F_C \) exceeds \( F^*_{k,n-2k} \), the point on the \( F \)-distribution with \( k \) and \( n-2k \) degrees of freedom (d.f.) such that the area to the right is equal to the level of significance.

But, as noted in Hansen (2001), the above procedure is based on the critical assumption that the break date is known a priori. If that is not the case, the Chow test involves two problems. First, the test turns out to be uninformative because one needs to pick up an arbitrary break date to perform the test and in the process, might miss the true break date. The other way is to select a date based on some feature of the data. But in this case, the test might be misleading, as the candidate break date is endogenous, i.e., it is correlated with the data and the test is likely to falsely indicate a break when no such break exists. Further, since the results can be highly sensitive to these arbitrary choices, different researchers can reach quite different conclusions. The solution to this problem, attributed to Quandt (1960), is to treat the break date as unknown, carry out the procedure for all the possible years and then select the largest Chow statistic over all possible break dates. The question then is about the critical values to be used because the Chi-square critical values are inappropriate, if the break date is a priori unknown. The solution to the problem was provided by many authors, but the appropriate critical values for assessing statistical significance were not developed until several decades later. Recent work by Andrews (1993), Andrews and Ploberger (1994), and Vogelsang (1997) have made it possible to implement the test by proposing critical values.

An alternative to the Quandt Likelihood Ratio (QLR) test for use in the situation where a researcher believes that a series may contain a structural break but is unsure of the date is to perform a recursive estimation. This is sometimes known as recursive least squares (RLS). Recursive estimation simply involves starting with a sub-sample of the data, estimating the regression; then sequentially adding one observation at a time and re-running the regression until the end of the sample is reached. It is common to begin the initial estimation with the very minimum number of observations possible, which will be \( k + 1 \). So at the first step, the model is estimated using observations 1 to \( k + 1 \); at the second step, observations 1 to \( k + 2 \) are used and so on; at the final step, observations 1 to \( T \) are used. The final result
will be the production of T - k separate estimates of every parameter in the regression model. It is to be expected that the parameter estimates produced near the start of the recursive procedure will appear rather unstable since these estimates are being produced using so few observations, but the key question is whether they then gradually settle down or whether the volatility continues through the whole sample. Seeing the latter would be an indication of parameter instability.

Given the above literature on denitrifying break dates, in the present study I will use Quandt Likelihood Ratio Test to find unknown break dates in time series data.

4.3: Decomposition of change in Aggregate Growth Rate

Decomposition of change in aggregate growth rate is important if the variable under study is composed of many other variables. Decomposition helps us to identify the variables among those sub-variables which are responsible for acceleration/deceleration in aggregate values of the variable. To illustrate the methodology of decomposition which is used in the present study, let us suppose we have a variable Y which is the sum of other three variables P, S and T. Then we can denote Y as:

\[ Y_t = P_t + S_t + T_t \] \hspace{1cm} (4.3.1)

where \( t \) stands for current period.

From (4.3.1), the growth rate of Y can be written as:

\[ g_Y = w_{P_t} \times g_{P_t} + w_{S_t} \times g_{S_t} + w_{T_t} \times g_{T_t} \] \hspace{1cm} (4.3.2)

where \( g_Y = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \) is the growth rate of Y in t and \( g_{P_t}, g_{S_t}, \) and \( g_{T_t} \) are defined similarly. Using (4.3.2), the average annual growth rate for the period 1 (\( g_{Y_{11}} \)) can be written as follows. Where a bar over notations has been used to denote average value.
\[ g_{y_t} = \bar{w}_i \times \bar{g}_{\bar{r}_i} + \bar{w}_s \times \bar{g}_{s_i} + \bar{w}_r \times \bar{g}_{r_i} \] ........................ (4.3.3)

where,
\[ \bar{g}_{\bar{y}_i} = N_1^{-1} \left( \sum_{t=1}^{N_1} g_{y_t} \right), \] and \[ \bar{w}_i \times \bar{g}_{\bar{y}_i} = N_1^{-1} \left( w_i \times g_{y_t} \right), \] N1 stands for number of observation in the first period.

Similarly, \( g_{y_t} \) can be written as below,
\[ g_{y_t} = \bar{w}_i \times \bar{g}_{\bar{r}_i} + \bar{w}_s \times \bar{g}_{s_i} + \bar{w}_r \times \bar{g}_{r_i} \] ........................ (4.3.4)

In (4), averages are taken over N2 observations in period two.

Subtracting (3) from (4),
\[ g_{y_2} - g_{y_1} = \left( \bar{w}_i \times \bar{g}_{\bar{r}_2} - \bar{w}_i \times \bar{g}_{\bar{r}_1} \right) + \left( \bar{w}_s \times \bar{g}_{s_2} - \bar{w}_s \times \bar{g}_{s_1} \right) + \left( \bar{w}_r \times \bar{g}_{r_2} - \bar{w}_r \times \bar{g}_{r_1} \right) \] ........... (4.3.5)

Dividing (4.3.5) throughout by the left-hand side value and multiplying it by 100 gives:
\[ 100 = \left( \frac{w_{p2} g_{r2} - w_{p1} g_{r1}}{g_{r2} - g_{r1}} \right) \times 100 + \left( \frac{w_{s2} g_{s2} - w_{s1} g_{s1}}{g_{s2} - g_{s1}} \right) \times 100 + \left( \frac{w_{r2} g_{r2} - w_{r1} g_{r1}}{g_{r2} - g_{r1}} \right) \times 100 \] ........... (4.3.6)

This equation gives us percentage contribution of each variable to the change in the average growth rate of Y in period two over that in previous period.

### 4.4 Economic Growth and Disparity: Test of Convergence

The convergence hypothesis basically postulates that economies with lower per capita incomes will grow faster than the leader with the highest income per capita in
a group. The literature on convergence has generated two tests of convergence that produce implications for convergence in per capita incomes across nations or regions. These are beta convergence and sigma convergence. Both these tests are used in the present study:

4.4.a Beta Convergence

The beta (\(\beta\))-convergence approach moves from the neoclassical Solow-Swan exogenous growth model (Solow, 1956; Swan, 1956), assuming a closed economic system, exogenous saving rates and a production function based on decreasing productivity of capital and constant returns to scale. On this basis authors like Mankiw et al. (1992) and Barro and Sala-i-Martin (1992) suggested the following statistical model:

\[
\ln \left( \frac{y_T}{y_0} \right) = a + (1 - e^{-\beta k}) \ln y_{0,j} + u_j \quad \text{............... (4.4.1)}
\]

where \(\ln(y_T / y_0)\) is the growth rate of the per-capita income over the entire period, as \(y_T\) is the value of per-capita income in the last period of time considered, and \(y_0\) is the value in the first period; \(\varepsilon_i\) is the error term. The parameter \(\beta\) measures the “speed of convergence”, i.e. how fast economies converge towards the steady state. The assumption on the probability model implicitly made in this context is that \(\varepsilon_i\) is normally distributed \((0,\sigma^2)\) independently of \(\ln y_{0,j}\). Finally, concerning the sampling model, it is assumed that \(\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}\) are independent observations of the probability model. Equation 4.4.1 is usually directly estimated through non-linear least-squares (Barro and Sala-i-Martin, 1995) or by re-parameterizing the statistical model setting \(b = (1 - e^{-\beta k})\) and estimating \(b\) by ordinary least squares. Absolute convergence is said to be present if the estimate of \(b\) negative and statistically significant. If the null hypothesis \((b = 0)\) is rejected, we would conclude that poor regions grow faster than rich ones, and that they all converge to the same level of per-capita income.
In the above specification, we considered the case of conditional convergence. However, the pace of growth depends also on the distance from the region-specific steady state, i.e. the further a region finds itself from its own steady state, the faster its growth rate will be. Assuming a kind of reversed gravity law, specific factors must be considered that could potentially affect the convergence process (conditional convergence). Accordingly, the following model has been implemented for β conditional convergence:

\[
\ln \left( \frac{y_{t,i}}{y_{0,i}} \right) = \alpha + (1 - e^{-\beta t}) \ln y_{0,i} + \gamma X' + \varepsilon_i, \quad \ldots \ldots \ldots (4.4.2)
\]

Note that this equation includes the matrix \( X' \) containing additional explanatory steady-state variables (physical or human capital, shares of production sectors to SDP, degree of political instability, ratio of public expenditures to SDP and other environmental variables) and the respective vector of associated coefficients \( \gamma \).

Though one can use the above specifications to test presence or absence of convergence, but one serious limitation of classical cross-section regression is that this approach does not give equal weight to each and every observation in the series under consideration. This problem can be solved by using panel data technique. The advantage of the panel data approach to convergence is that it can be helpful for the correction of the bias generated by omitted variables and heterogeneity in the classical cross-sectional regression (Islam, 2003). Panel data, in fact, allow for technological differences across regions, (or at least the unobservable and unmeasurable part of these differences), by modelling the regional specific effect. More formally, the panel version of the growth equation can be expressed in the following way:

\[
\ln \left( \frac{y_{t,i+k}}{y_{t,i}} \right) = a + b \ln y_{t,i} + u_{it} \quad \ldots \ldots \ldots (4.4.3)
\]

\[
\ln \left( \frac{y_{t,i+k}}{y_{t,i}} \right) = a + b \ln y_{t,i} + \gamma X' + u_{it} \quad \ldots \ldots (4.4.4)
\]
With \( i = 1, \ldots, N \) denoting regions, and \( t = 1, \ldots, T \) denoting time periods. The dependent variable \( \ln(y_{i,t}/y_{i,t-1}) \) is the annual growth rate of the per-capita income of the \( i \)th economy, \( \ln y_{i,t} \) is the value of the per-capita income at the beginning of each period over which the growth rate is calculated.

In this simple panel the coefficients \( a \), \( b \) and \( \gamma \) do not have any subscripts, suggesting that they will be the same for all units and for all years. We can introduce some degree of heterogeneity in this panel by relaxing the fact that the constant \( a \) should be identical for all cross-sections. That is

\[
\ln \left( \frac{y_{i,t+k}}{y_{i,t}} \right) = a + b \ln y_{i,t} + \epsilon_{i,t} \quad \ldots \ldots \ldots \quad (4.4.5)
\]

\[
\ln \left( \frac{y_{i,t+k}}{y_{i,t}} \right) = a + b \ln y_{i,t} + \gamma \lambda' + \epsilon_{i,t} \quad \ldots \ldots \ldots \quad (4.4.6)
\]

where, \( a \) can now differ for each cross-section unit in the sample.

In general, simple linear panel data models can be estimated using three different methods: (a) with a common constant as in equations (4.4.3 and 4.4.4), (b) allowing for fixed effects, and (c) allowing for random effects (Asteriou, 2006). If the \( a_i \) is assumed to be fixed parameters, equations (4.4.5 and 4.4.6) become fixed-effect panel data model. Conversely, if \( a_i \) are assumed to be random, equations (4.4.5 and 4.4.6) express the random effect panel data model. Generally, fixed-effect model is particularly indicated when the regression analysis is limited to a precise set of individuals (firms or regions). On the contrast, random effect is a more appropriate specification if we are drawing a certain number of individuals randomly from a larger population. Technically, however, the Hausman test is formulated to assist in making a choice between the fixed effects and random effects approaches.

4.4b : Sigma Convergence

Sigma \( (\sigma) \) convergence is a simpler concept. Data on per capita income are collected as a time series for each of the regions under analysis. Then the dispersion of the log of per capita income is computed for each year across the regions. This is a simple measure of income inequality for the sample data. If this dispersion
declines over time, per capita incomes are less dispersed and $\sigma$ convergence is implied. This concept, sigma-convergence, provides a measure of the extent of income inequality and how such inequality changes over time. Algebraically one can test presence of sigma convergence by estimating the following equation:

$$\sigma = a + bT$$  \hspace{1cm} (4.4.7)

Where, $\sigma$ stands for measure of dispersion, and $T$ is the time trend. The existence of sigma convergence is shown as $b < 0$.

Generally $\beta$ convergence implies $\sigma$ convergence, but the process may be offset by shocks that increase income dispersion (Barro and Sala-i-Martin, 2008). Put differently, $\beta$ convergence is a necessary, but not sufficient, condition for $\sigma$ convergence. To probe this mathematically (due to Barro & Sala-i-Martin, 2008), let us denote $\sigma_t^2$ as the cross-economy variance of $\ln(y_t)$ at time $t$. Equation (4.4.1) and the assumed properties of $u_t$ imply that $\sigma_t^2$ evolves over time in accordance with the first-order difference equation

$$\sigma_t^2 = e^{-2\beta} \sigma_{t-1}^2 + \sigma_u^2$$  \hspace{1cm} (4.4.8)

where it is assumed that the cross section is large enough so that the sample variance of $\ln(y_u)$ corresponds to the population variance.

If the variance of the disturbance, $\sigma_u^2$, is constant over time ($\sigma_u^2 = \sigma_u^2$ for all $t$), the solution of the first-order difference equation (4.4.4) is

$$\sigma_t^2 = \frac{\sigma_u^2}{1-e^{-2\beta}} + \left(\sigma_0^2 - \frac{\sigma_u^2}{1-e^{-2\beta}}\right)e^{-2\beta t}$$  \hspace{1cm} (4.4.9)

Where $\sigma_0^2$ is the variance of $\ln(y_0)$. Equation (4.3.9) implies that $\sigma_t^2$ monotonically approaches its steady-state value, $\sigma^2 = \frac{\sigma_u^2}{1-e^{-2\beta}}$, which rises with $\sigma_u^2$ but declines with the convergence coefficient, $\beta$. Over time, $\sigma_t^2$ rises (or falls) if the initial value $\sigma_0^2$ is less than (or greater than) the steady-state value, $\sigma^2$. Thus a positive
coefficient $\beta$ (convergence) does not imply a falling $\sigma^2$ (convergence). In other words, $\beta$ convergence is a necessary but not sufficient condition for $\sigma$ convergence.

4.5. Measures of Dispersion:

Dispersion is defined as the amount by which a set of observations deviate from their mean. When the values of a set of observations are close to their mean, the dispersion is less than when they are spread out widely from the mean (Everitt, 2006). There are two types of measures of dispersion which are: (a) Absolute Measure of Dispersion, and (b) Relative Measure of Dispersion. The absolute measures which have been used in the present study are:

1. The Range
2. Mean Deviation
3. The Standard deviation and Variance

Each absolute measure of dispersion can be converted into its relative measure.

1. Coefficient of Range or Coefficient of Dispersion.
2. Coefficient of Mean Deviation or Mean Deviation of Dispersion.
3. Coefficient of Standard Deviation or Standard Coefficient of Dispersion.
4. Coefficient of Variation

Another relative measure of dispersion which got very high popularity in recent years is the Gini coefficient as developed by the Italian statistician Corrado Gini (1912). The Gini approach starts from a base that is fundamentally different from measure such as mean deviation and coefficient of variation. Instead of taking mean value, it takes the difference between all pairs of values and simply totals the (absolute) differences. The Gini coefficient is normalised by dividing by population squared as well as mean income. In symbols, the Gini coefficient $G$ is given by (Ray, 2008)

$$G = \frac{1}{2n^2\mu} \sum_{j=1}^{m} \sum_{k=1}^{m} n_j n_k |Y_j - Y_k| \quad \text{...............} \quad (4.5.1)$$
The double summation sign implies that we first sum over all ks, holding each j constant, and then sum over all the js. This is like summing all pairs of differences weighted by the number of each pairs, $n_j n_k$. The Gini coefficient can range from 0 to 1; it is sometimes multiplied by 100 to range between 0 and 100. A low Gini coefficient indicates a more equal distribution, with 0 corresponding to complete equality, while higher Gini coefficients indicate more unequal distribution, with 1 corresponding to complete inequality.

4.6: Test for Causality among variables

The concept of causation is essentially a philosophical rather than an empirical issue, with theories ranging from an extreme "everything causes everything", to denying the existence of any causation whatsoever. The concept of causation or causality was first introduced into the econometric literature by Granger (1969) and as a consequence is often referred to as Granger causality\(^3\). This has been used in the present study. It can be described in the following way:

*If past values of a variable \(y\) significantly contribute to forecast the future value of another variable \(x\) then \(y\) is said to cause \(x\) (Granger, 1969).*

More specifically consider the following models:

\[
y_t = \gamma_0 + \sum_{k=1}^{M} \delta_k y_{t-k} + \sum_{l=1}^{N} \gamma_l x_{t-l} + e_t \quad \quad \quad (4.6.1)
\]

\[
x_t = \beta_0 + \sum_{k=1}^{M} \beta_k y_{t-k} + \sum_{l=1}^{N} \alpha_l x_{t-l} + u_t \quad \quad \quad (4.6.2)
\]

where, \(x_t\) and \(y_t\) are two stationary time series with zero mean, \(e_t\) and \(u_t\) are two uncorrelated error terms, \(t\) denotes the time period and 'k' and 'l' are the number of lags. In equation 1, the null-hypothesis (which is as $H_0 : \sum_{l=1}^{N} \gamma_l = 0$) is tested.
against the alternative hypothesis (which is as $H_1: \gamma_i's$ are jointly significant). If we reject $H_0$, we would conclude that $x$ Granger causes $y$. Similarly, in equation 2, the null hypothesis ($H_0: \sum_{k=1}^{N} \beta_k = 0$) is tested against the alternative one (which is as $H_1: \beta_k's$ are jointly significant). If we reject $H_0$, then we would conclude that $y$ granger causes $x$. If $x$ Granger causes $y$ but not the other way around, then causality is said to be uni-directional. But if both $\gamma_i$ and $\beta_k$ are significant, then causality is said to be bi-directional.

4.7 Stationarity of Time Series data:

Like most time series econometric models, an important requirement of testing Granger Causality is that data series have to be stationary. But, most economic time series data are collected as discrete data over time, and every individual observation in the time series are viewed as “just happen to be” or random or stochastic. Thus, a time series is a collection of “the representative of” or “the realisation of” random variables arranged in time. Although all the individual observations are good representatives of their populations at the particular time, to be able to generalise this representative to another time period, the order of “the representatives” in time are required to be well behaved – that is, the mean and the variance of the time series are required to be constant over time. More formally, a time-series variable, $Y_t$, is stationary if:

1. the mean of $Y_t$ is constant over time,
2. the variance of $Y_t$ is constant over time, and
3. the simple correlation coefficient between $Y_t$ and $Y_{t-k}$ depends on the length of the lag (k) but on no other variable (for all k).

If one or more of these three properties is not met, then $Y_t$ is non-stationary. If a series is non-stationary, that problem is often referred to as non-stationarity (Studenmund, 2005)
4.7.1: Unit Root Tests for Individual Time Series

One of the most popular methods of testing for nonstationarity is the unit root test. However, a number of unit root test methods can be employed for individual time series data, such as the Dickey-Fuller (DF) test, the Augmented Dickey-Fuller (ADF) test, the Phillips-Perron (PP) test, the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test, the Elliot, Rothenberg and Stock (ERS) point optimal test, and the Ng-Perron (NP) test. Some of the tests are specific to a particular time series scenario and the application is limited.

Though unit root test has been devised by different econometricians in different ways, in the present study only three among them have been considered. These are: Dickey-Fuller Unit Root Test The Phillips-Perron Unit Root Test, and Kwiatkowski-Phillips-Schmidt-Shin Unit Root Test.

Dickey-Fuller (1979, 1981) devised a procedure to test for non-stationarity. The key insight of their test is that testing for non-stationarity is equivalent to resting for the existence of a unit root. This test is thus based on the simple AR (1) model of the form:

\[ y_t = \psi y_{t-1} + e_t \]  

(4.7.1)

Here we need to examine whether \( \psi \) is equal to 1. Obviously, the null hypothesis is \( H_0 : \psi = 1 \), and the alternative hypothesis is \( H_a : \psi < 1 \).

Subtracting \( y_{t-1} \) from both sides of (4.7.1):

\[ y_t - y_{t-1} = \psi y_{t-1} - y_{t-1} + e_t \]
\[ \Delta y_{t-1} = (\psi - 1) y_{t-1} + e_t \]
\[ \Delta y_{t-1} = \delta y_{t-1} + e_t \]  

(4.7.2)

where \( \delta = (\psi - 1) \). Then, the null hypothesis is \( H_0 : \delta = 0 \) and the alternative hypothesis is \( H_a : \delta < 0 \), where if \( \delta = 0 \) then \( y_t \) follows a pure random-walk model.
Dickey-Fuller (1979) also proposed two alternative regression equations that can be used for testing for the presence of a unit root. The first contains a drift in the random-walk process as in the following equation:

\[ \Delta y_{t-1} = \alpha_0 + \delta y_{t-1} + e_t \]  

This is an extremely important case, because such processes exhibit a definite trend in the series, which is often the case for macroeconomic variables.

The second case is to allow, apart from a drift, also a non-stochastic time trend in the model, so as to have:

\[ \Delta y_{t-1} = \alpha_0 + \alpha_t + \delta y_{t-1} + e_t \]  

In all cases the test concerns whether \( \delta = 0 \). The DF-test statistic is the t statistic for the lagged dependent variable. If the DF statistic value is smaller than the MacKinnon (1991) critical value then we reject the null hypothesis of a unit root and conclude that \( y_t \) is a stationary process.

In conducting the DF test as in (4.7.2), (4.7.3), or (4.7.4), it was assumed that the error term \( e_t \) was uncorrelated. But in case the \( e_t \) are correlated, Dickey and Fuller have developed a test, known as the Augmented Dickey-Fuller (ADF) test. This test is conducted by “augmenting” the preceding three equations by adding the lagged values of the dependent variable \( \Delta y_t \) (Gujarati, 2007). To be specific, suppose we use (4.7.4), the ADF test here consists of estimating the following regression:

\[ \Delta y_{t-1} = \alpha_0 + \alpha_t + \delta y_{t-1} + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + e_t \]  

where \( e_t \) is a pure white noise error term. The number of lagged difference terms to include is determined by the Akaike Information Criterion (AIC) or Schwartz Bayesian Criterion (SBC), or more usefully by the lag length necessary to whiten the residual. In ADF we still test whether \( \delta = 0 \) and the ADF test follows the same asymptotic distribution as the DF statistic, so the same critical values can be used.

The distribution theory supporting the DF tests is based on the assumption that error terms are statistically independent and have a constant variance. So, when
using the ADF methodology we have to make sure that the error terms are uncorrelated and that they really have a constant variance. Phillips and Perron (1988) developed a generalization of the ADF test procedure that allows for fairly mild assumptions concerning the distribution of errors. The test regression for the Phillips and Perron (PP) test is the AR(1) process:

\[ \Delta y_{t-1} = \alpha_0 + \rho y_{t-1} + e_t \]  

while the ADF test corrects for higher order serial correlation by adding lagged difference terms on the right-hand side, the PP test makes a correction to the t statistic of the coefficient \( \delta \) from the AR(1) regression to account for the serial correlation in \( e_t \). The t statistic is calculated as:

\[ t_{pp}^{\alpha} = t_{\alpha} \left( \frac{\gamma_0}{f_0} \right)^{1/2} - \frac{T \left( f_0 - \gamma_0 \right) (S_\hat{\alpha})}{2 f_0^2 s} \]  

where \( \hat{\alpha} \) is the estimated \( \alpha_0 \) in (4.7.6), \( t_{\alpha} \) is the t-statistic of \( \alpha_0 \) in (4.7.6), \( S_\hat{\alpha} \) is the standard error coefficient for the t-statistic and \( s \) is the standard error of the test regression (4.7.6). \( \gamma_0 \) is a consistent estimate of the error variance in (4.7.6) and at frequency of zero, \( f_0 \) is an estimator of the residual spectrum. So, the PP statistics are just modification of the ADF t-statistics that takes into account the less restrictive nature of the error process\(^4\). The asymptotic distribution of the PP t-statistic is the same as ADF t-statistic and therefore the MacKinnon (1991) critical values are still applicable. As with the ADF test, the PP test can be performed with the inclusion of a constant, a constant and a linear trend, or neither in the regression.

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (1992) reversed the null hypothesis by assuming that the series is stationary (in contrast to the ADF test with the null hypothesis of nonstationary). It is an LM test based on the OLS residual from the regression of:

\[ y_t = x_t \delta + u_t \]  

84
Since \( x_i' \delta \) is the exogenous term of either constant or constant and trend specification, The KPSS test will be in the following two forms:

\[
y_i = c_i + u_i \quad \text{...................................................... (4.7.9)}
\]

and

\[
y_i = c_i + \beta t + u_i \quad \text{...................................................... (4.7.10)}
\]

The test statistic is:

\[
LM = \frac{\sum S_i^2}{T^2 f_0} \quad \text{...................................................... (4.7.11)}
\]

Where \( f_0 \) is same estimator as the PP test and \( S_i \) is the cumulative residual function of \( S_i = \sum_{c=1}^{t} \hat{u}_i \). To specify the KPSS test, similar to the PP test, we need to specify whether an intercept or a trend and intercept are present in the test regression. We also need to select the method of estimating \( f_0 \).

4.7.2: Unit Root Tests for Panel Data Series

The term Panel Data in this study refers to the pooling of time series data for all of the state economies in the study. Panel unit root tests are mostly similar to the unit root tests on single time series in Equation (4.7.3.) or (4.7.4), but modified as:

\[
y_{it} = \rho y_{i,t-1} + X_i' \delta_i + u_{it} \quad -1 \leq \rho \leq 1 \quad \text{......... (4.7.12)}
\]

where \( X_i' \delta_i \) is an exogenous variable, such as constant and/or constant and trend, and \( u_{it} \) is pure random term. And \( i=1,2,3, \ldots, N \) cross section series. Within the range of \(-1 \leq \rho \leq 1\), if \( \rho < 1 \), the series of \( y_i \) is said to be stationary, if \( \rho = 1 \), \( y_i \) is said to be unit root nonstationary series.

A number of panel data stationary tests have been developed in recent years. Those tests can be classified into two broad categories, (i) those cross-section data that can be assumed independent, and (ii) those cross-section data that cannot be assumed independent. For the category of cross-sectional independent panel unit

In the present study, we assume that the cross-sections of data in our sample are independent. The following subsections focus on the panel data unit root tests, which assume cross-section independence of the panel data. This category can be further classified into two subcategories: a more restrictive assumption of a common unit roots for all cross-sections or less restrictive assumption of individual unit roots for each cross-section units.

Levin, Lin and Chu (2002) assume a common unit root across the cross-sections. LLC started with a basic ADF specification for panel data:

\[
\Delta y_{it} = \alpha_i + \rho y_{i,t-1} + \sum_{L=1}^{n} \beta_{L} \Delta y_{i,t-L} + \delta_{i} + \theta_{t} + \epsilon_{it} \quad \ldots \ldots (4.7.13)
\]

This model allows for two-ways fixed effects, one from \( \alpha_i \) and the second from the \( \theta_t \). So we have both unit-specific fixed effects and unit-specific time trends. The unit-specific fixed effects are a very important component because they allow for heterogeneity since the coefficient of the lagged \( y_i \) is restricted to be homogeneous across all units of the panel.

The null and alternative hypotheses of this test are that:

\[
H_0: \rho = 0 \\
H_0: \rho < 0
\]

If the null hypothesis is rejected in favour of the alternative, then the panel data has no unit root. Under the assumption that the individual processes are cross-sectionally independent, the test derives conditions for which the pooled OLS estimator of \( \rho \) will follow a standard normal distribution under the null hypothesis.
Thus, the LL test may be viewed as a pooled DF or ADF test, potentially with different lag length across the different sections in the panel.

The major drawback of the LL test is that it restricts \( \rho \) to be homogeneous across all \( i \). Im, Pesaran and Shin (1997) extended the LL test allowing heterogeneity on the coefficient of the \( y_{it-1} \) variable and proposing as a basic testing procedure one based on the average of the individual unit-root test statistics.

The Im, Pesaran and Shin (IPS) test provides separate estimations for each \( I \) section, allowing different specifications of the parametric values, the residual variance and the lag length. Their model is given by:

\[
\Delta y_{it} = a_i + \rho_i y_{it-1} + \sum_{L=1}^{\infty} \beta_i \Delta y_{it-L} + \delta_i + \theta_i + u_t \quad \text{............ (4.7.14)}
\]

The null and alternative hypotheses are formulated as:

- \( H_0: \rho_i = 0 \) for all \( i \)
- \( H_0: \rho < 0 \) for at least one \( i \)

Thus, the null hypothesis of this test is that all series are non-stationary process under the alternative that a fraction of the series in the panel are assumed to be stationary. This is in sharp contrast with the LL test, which presumes that all series are stationary under the alternative hypothesis.

Im, Pesaran and Shin (1997) formulated their model under the restrictive assumption that \( T \) should be the same for all cross-sections, requiring a balanced panel to compute the \( \tilde{t} \) test statistic. Their \( \tilde{t} \) statistic is nothing else than the average of the individual ADF t-statistics for testing that \( \rho_i = 0 \) for all \( i \) denoted by \( t_{pi} \):

\[
\tilde{t} = \frac{1}{N} \sum_{i=1}^{N} t_{pi} \quad \text{................................. (4.7.15)}
\]

Im, Pesaran and Shin (1997) also showed that under specific assumptions \( t_{pi} \) converges to a statistic denoted as \( t_i \) which they assume that is independent and
identically-distributed (iid) and that also has finite mean and variance. They then computed values for the mean \(E[t_{it} \mid \rho_t = 1]\) and for the variance \(Var[t_{it} \mid \rho_t = 1]\) of the \(t_{it}\) statistic for different values of \(N\) and lag included in the augmentation term of equation (4.7.13). Based on those values, they then constructed the IPS statistic for testing for unit roots in panels given by:

\[
t_{IPS} = \frac{\sqrt{N} \left( t - 1/N \sum_{i=1}^{N} E[t_{it} \mid \rho_t = 0] \right)}{Var[t_{it} \mid \rho_t = 0]} \tag{4.7.16}
\]

which they have proved follows the standard normal distribution as \(T \to \infty\) followed by \(T \to \infty\) sequentially. The values of \(E[t_{it} \mid \rho_t = 0]\) and \(Var[t_{it} \mid \rho_t = 0]\) are given in their paper. Finally, they also suggested a group mean Lagrange multiplier test for testing for panel unit roots. Performing Monte Carlo simulations they proved that both their LM and \(t\) statistics have better finite sample properties than the LL test.

Breitung (2000) suggests a test statistic that does not employ a bias adjustment whose power is substantially higher than that of LLC or the IPS tests using Monte Carlo experiments. To have a brief review of Breitung’s test, consider the following AR (1) process for panel data:

\[
y_{it} = \rho_i y_{i,t-1} + \delta_t x_{i,t} + \epsilon_{i,t} \tag{4.7.17}
\]

where \(y_{i,t}\) represents the panel of original series of interest, \(x_{i,t}\) represents the deterministic variables in the model including any fixed cross effects or individual trends, \(\rho_i\) is the autoregressive coefficients, and the errors \(\epsilon_{i,t}\) are assumed to be mutually independent idiosyncratic errors. If \(|\rho_t| < 1\) then \(y_{i,t}\) is said to be weakly stationary. On the other hand, if \(|\rho_t| = 1\) then \(y_{i,t}\) contains a unit root. In Breitung’s test, it is assumed that the persistence parameters are common across cross-sections, i.e. \(\rho_i = \rho\) for all \(i\). Based on this assumption, the maintained hypothesis is that
\[ \Delta y_{i,t} = \alpha y_{i,t-1} + \sum_{j=1}^{p} \beta_j \Delta y_{i,t-j} + \delta_t x_{i,t} + \varepsilon_{i,t} \] ........................ (4.7.18)

Here we assume a common \( \alpha = \rho - 1 \) for all \( i \) but allow the lag order for the difference terms to vary across the cross-sections. The null and alternative hypotheses for the tests are:

\[ H_0 : \alpha = 0 \quad \quad H_1 : \alpha < 0 \]

4.8: Long Run Relationship between Variables: Test for Cointegration

Most macroeconomic variables follow a particular pattern of trend and therefore the spurious regression problem is highly likely to be present in most macroeconomic models. One way of resolving this is to difference the series successively until stationarity is achieved and then use stationary series for regression analysis. However, applying first differences of the variables leads to the loss of long-run properties, since the model has no long-run solution. The desire to have models which combine both short-run and long-run properties, and which at the same time maintain stationarity in all of the variables, has led to a reconsideration of the problem of regression using variables that are measured in their levels (Asteriou, 2006). The idea is that if there are economic time series that are integrated and on the same order, then try to check whether we can find a way to combine them together into a single series which itself is non-stationary. If this happens, then the series that exhibits this property is called cointegrated. To illustrate, suppose \( X_t \) and \( Y_t \) are two random walks and hence are not stationary. In general, we would expect that a linear combination of \( X_t \) and \( Y_t \) would also be a random walk. Yet, the two series may have the property that a particular linear combination of them \( (X_t - \alpha Y_t) \) is stationary. Thus, \( X_t \) and \( Y_t \) may each be integrated of order 1 \( \{ \text{that is, } I (1) \} \), but there may exist an \( \alpha \) such that \( X_t - \alpha Y_t \) is stationary \{that is, \( I (0) \) with a finite variance\}. If such a property holds, then we say that \( X_t \) and \( Y_t \) are
cointegrated. Two cointegrated series will thus not drift too far apart over the long run (Ramanathan, 2008)

Engle and Granger (1987) have developed a very simple test for the existence of cointegration relationship. In order to understand this approach, let us suppose we have two series \( X_t \) and \( Y_t \) which are I(1), then in general any linear combination of the two series, say

\[
\theta X_t - \theta Y_t
\]

will also be I(1). However, although this is the more likely case, there are exceptions to this rule, and we might find in rare cases that there is a unique combination of the series as in (4.6.1) that is I(0). If this is the case then we say that \( X_t \) and \( Y_t \) are cointegrated of order (1,1).

Technically, if the two series \( X_t \) and \( Y_t \) are I(1), then we estimate the long-run relationship of the form:

\[
Y_t = \alpha + \beta X_t + e_t \quad \text{......................... (4.8.2)}
\]

and then obtain the residuals of this equation. Now, in order to determine if the variables are actually cointegrated, denote the estimated residual sequence from this equation by \( \hat{e}_t \). Then a DF test is performed on the residual series to determine their order of integration. The form of the DF test is the following:

\[
\Delta \hat{e}_t = a \hat{e}_{t-1} + \sum_{i=1}^{n} b_i \hat{e}_{t-i} + \nu_t \quad \text{......................... (4.8.3)}
\]

If we find that \( \hat{e}_t \sim I(0) \) then we can reject the null hypothesis that the variables \( X_t \) and \( Y_t \) are not cointegrated.
One important problem with Engle and Granger (1987) procedure is that it does not refer anything about the order of the variables. When estimating the long-run relationship, one has to place one variable in the left-hand side and use the other as regressors. The test does not say anything about which of the variables can be used as regressor. A second problem is that when there are more than two variables there may be more than one cointegrating relationship, and the Engle and Granger (1987) procedure using residuals from a single relationship can not treat this possibility. A third problem is that it relies on two-step estimator. The first step is to generate the error series and the second step is to estimate a regression for this series in order to observe whether the series is stationary or not. Hence, any error introduced in the first step is carried into the second step (Asteriou, 2006). All these problems are resolved with the use of the Johansen (1988, 1991, 1995) approach. In order to present this approach, let us assume that we have three variables, $X_t$, $Y_t$, and $W_t$ which can all be endogenous, i.e. we have that (using matrix notation for $Z_t = [X_t, Y_t, W_t]$)

\[
Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \ldots + A_k Z_{t-k} + u_t \ldots (4.8.4)
\]

which is comparable to the single-equation dynamic model for two variables. Thus, it can be reformulated in a vector error correction model (VECM) as follows:

\[
\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \Gamma_2 \Delta Z_{t-2} + \ldots + \Gamma_{k-1} \Delta Z_{t-k-1} + \Pi Z_{t-1} + u_t \ldots (4.8.5)
\]

where:

\[
\Gamma_i = (I - A_i - A_2 - \ldots - A_k), \ i = 1, 2, \ldots, k-1.
\]

And, \( \Pi = -(I - A_i - A_2 - \ldots - A_k) \)

The matrix \( \Pi \) contains the information regarding the long-run relationship. In fact \( \Pi = \alpha \beta' \) where \( \alpha \) will include the speed of adjustment to equilibrium coefficient while \( \beta' \) will be the long-run matrix of coefficients.
Therefore, the $\beta'Z_{t-1}$ term is equivalent to the error-correction term 

\[(Y_{t-1} - \alpha - \beta X_{t-1})\] 

in the single equation case, except that now $\beta'Z_{t-1}$ contains up to 

(n-1) vectors in a multivariate framework.

For simplicity let’s assume $k=2$, so that we have only two lagged terms, and the 
model is then the following:

\[
\begin{align*}
\begin{bmatrix}
\Delta Y_t \\
\Delta X_t \\
\Delta W_t
\end{bmatrix}
&= \Gamma_1 
\begin{bmatrix}
\Delta Y_{t-1} \\
\Delta X_{t-1} \\
\Delta W_{t-1}
\end{bmatrix}
+ \Pi 
\begin{bmatrix}
\Delta Y_{t-1} \\
\Delta X_{t-1} \\
\Delta W_{t-1}
\end{bmatrix}
+ e_t 
\end{align*}
\]

\[\text{(4.8.6)}\]

Or

\[
\begin{align*}
\begin{bmatrix}
\Delta Y_t \\
\Delta X_t \\
\Delta W_t
\end{bmatrix}
&= \Gamma_1 
\begin{bmatrix}
\Delta Y_{t-1} \\
\Delta X_{t-1} \\
\Delta W_{t-1}
\end{bmatrix}
+ \begin{bmatrix}
   a_{11} & a_{12} \\
   a_{21} & a_{22} \\
   a_{31} & a_{32}
\end{bmatrix}
\begin{bmatrix}
   \beta_{11} & \beta_{21} & \beta_{31} \\
   \beta_{12} & \beta_{22} & \beta_{32}
\end{bmatrix}
\begin{bmatrix}
   Y_{t-1} \\
   X_{t-1} \\
   W_{t-1}
\end{bmatrix}
+ e_t
\end{align*}
\]

\[\text{(4.8.7)}\]

Let us now analyse only the error-correction part of the first equation which gives:

\[
\Pi_t Z_{t-1} = (a_{11} \beta_{11} + a_{12} \beta_{12}) \begin{bmatrix}
   a_{11} \beta_{21} + a_{12} \beta_{22} \\
   a_{11} \beta_{31} + a_{12} \beta_{32}
\end{bmatrix}
\begin{bmatrix}
   Y_{t-1} \\
   X_{t-1} \\
   W_{t-1}
\end{bmatrix}
\]

\[\text{(4.8.8)}\]

where $\Pi_t$ is the first row of the $\Pi$ matrix.

Equation 4.8.8 can be re-written as:

\[
\Pi_t Z_{t-1} = a_{11} (\beta_{11}Y_{t-1} + \beta_{21}X_{t-1} + \beta_{31}W_{t-1}) + a_{12} (\beta_{32}Y_{t-1} + \beta_{22}X_{t-1} + \beta_{22}W_{t-1}) 
\]

\[\text{(4.8.9)}\]

which shows clearly the two cointegrating vectors with their respective speed of 

adjustment terms $a_{11}$ and $a_{12}$.
Let us now examine the behaviour of the Π matrix under different circumstances. Given that \( Z_t \) is a vector of non-stationary I(1) variables, then \( \Delta Z_{t-1} \) are I(0) and \( \Pi Z_{t-1} \) must also be I(0) in order to have that \( u_t \perp I(0) \) and therefore to have ‘white nose’.

In general there can be three cases for \( \Pi Z_{t-1} \) to be I(0):

Case 1. When all the variables in \( Z_t \) are stationary. There is no problem of spurious regression and the simple VAR in levels model can be used.

Case 2. When there is no cointegration. In this case the VAR model should be used in first difference with no long-run elements.

Case 3. When there exists up to (n-1) cointegrating relationships of the form \( \beta'Z_{t-1} \perp I(0) \). In this case, \( r \leq (n-1) \) cointegrating vectors exist in \( \beta \). This simply means that \( r \) columns of \( \beta \) from \( r \) linearly independent combinations of the variables in \( Z_t \), each of which is stationary. Of course there will also be \( (n-r) \) non-stationary vectors as well.

In the third case, only the cointegrating vector in \( \beta \) should be included in the VECM, otherwise \( \Pi Z_{t-1} \) will not be I(0). This means that the last (n-r) columns of \( \alpha \) will be statistically significant. Therefore, in determining how many \( r \leq (n-1) \) cointegrating vectors exist in \( \beta \) is equivalent to testing which columns of \( \alpha \) are equal to zero. In other words, testing for cointegration is equivalent to determining the rank of the matrix \( \Pi \), which is given by finding the number of \( r \) linearly independent columns in \( \Pi \). Johansen (1988) developed a methodology that tests for the rank of \( \Pi \) and provides estimates of \( \alpha \) and \( \beta \) through a procedure known as reduced-rank regression.

Given the above explanation about the advantage of Johansen Cointegration test over Engle and Granger procedure, I used the former to get robust results.
References:


----------------------------------------
NOTES:

1. This section is inspired by Balakrishan and Parameswaran (2007).
2. For more details on the use of these two models for panel data see Asteriou (2006).
3. In what follows, we will take the terms "causality" and "Granger causality" to mean the same thing.
5. For detail on the derivation of t-statistics see Levin et al. (2002).
6. If some or all of the variables in the regression are I(1) then the usual statistical results may or may not hold. One important case in which the usual statistical results do not hold is spurious regression when all the regressors are I(1) and not cointegrated. For detail on estimating procedure see Johansen (1988).