CHAPTER III

RESEARCH METHODOLOGY

A. Method of Sampling, Type of Data and Data Gathering

1. Sample

Sampling is a subgroup of subset of the population. It comprises some members selected from it. In order words, some, but not all, elements of the population form the sample (Sekaran and Bougie, 2009: 263). The basic idea of sampling is that by selecting some elements in a population, we may draw conclusion about the entire population. Sample is worth doing because it includes lower cost, greater accurate of results, greater speed of data collection, and availability of population elements (Cooper and Shrindler, 2014: 338). So by studying sample, the research study should be able to draw conclusions that are generalizable to the population of interest. The validity of sample depends on two considerations: accuracy and precision (Cooper and Shrindler, 2014: 339). Accuracy is the degree to which bias is absent from the sample; and the precision is the degree at which the interpreted findings of research is closely represents the real state of whole population.

There are two types of sampling methods: probabilistic and non-probabilistic methods. This research adopts a nonprobability sampling. Non-probabilistic method is opted because it is more practical for our considerations in sampling, it makes more satisfactorily to meet our sampling objectives; the choice of sampling will be at our convenience, because we have a purposive samples, and so that the selected samples follow our subjective judgment.

2. Type of Data:

There are many ways of classifying data, but mainly, there are only two types of data: primary and secondary data. Primary data refer to information first-hand by the researcher on the variables of interest for the specific purpose of the study. Secondary data refer to information gathered from sources that already exist. Secondary data can obtained from company records or archives, government publications, industry analyses offered by media, websites, the internet (Sekaran and Bougie, 2009: 180).

The primary data are those which are collected afresh and for the first time, and thus happen to be original in character. The secondary data are those which are already been collected by someone else and which have already been passed through the statistical process.

The purpose of this research is to analyze the interrelation between capital market index and economic growth among United-States, England and Japan. Thus, secondary data are employed for doing the analysis. Secondary data means data that are already prepared by someone else. They refer to the data which have already collected and analyzed by other parties for another purpose.

Collecting secondary data is easy because secondary data are already available, but it is risky to trust available data, so before using secondary data, we ought to take notice of the reliability, the suitability, and the adequacy of the data. The reliability of the secondary data can be measured by asking those following question: who provide the data? What were the sources of the data? How were the processes of collecting data? The suitability of the secondary data to our research is also important to consider, so we must check carefully if the collection of primary data respected the terms, the norms and the time of collecting that are should applied, and studied if they are suited with what we expect to obtain. The adequacy of the data refers if the secondary data is enough or satisfactory to our purpose. The data is considered inadequate if they are related to an area which may be either narrower or wider that the area of the present enquiry.

For the research study of the interrelation between capital market index and economic growth in USA, England and Japan, all we need to do is choosing one of the capital market index of each countries cited above, be sided by the variation of their growth domestic product quarterly. Investopedia gives the definition of market index as an aggregate value produced by combining several stocks or other investment vehicles together and expressing their total values against a base value from a specific date. Market indexes are intended to represent an entire stock market and thus track the market's changes over time. Capital market index is a weight of all capital market proxies. The research is quantitative research because we attempt to measure the mutual causality between capital market and economic growth.

3. Data Gathering

Data can be collected in a variety ways, in different settings. Secondary data can be obtained from different sources: government publication, technical and trade journals, books, magazines, newspapers, reports, publications, reports prepared by research scholars or universities, and economists.

The secondary data that are going to be used in this research are expected to obtain from official websites. The growth of gross domestic product of United-states, England, and Japan are supposing to be obtained from official organizational governmental websites. Similarly, we predict to acquire the full data of capital market indexes of United-States, England and Japan from trusted websites. Here are some examples of website that might be the source of our secondary data:

a. Yahoo finance
b. Bloomberg
c. Cia.org
d. Imf.org
e. Worldbank.org
f. Indexmundi.com
g. World-statistics.org
h. Morgan Stanley Capital Indices
i. Reuters
j. Morningstar.

B. Research Variables and Period of the study

1. Variables

This research analyzes the mutual relationship between capital market index and economic growth using three great countries. Thus the variables of our study are the quarterly variation of GDP of each country concerned, capital market indexes such as S&P 100, FTSE 100, Nikkei 500 which respectively represents capital market index of United-States, of England, of Japan. The table below resumes the list of variables and their symbols.

Table 2
Variables and Symbol of Variables

Countries	**Variables**	**Symbols**
United-States	S&P 100	^OEX
	Gross Domestic Product USA	GDP_{USA}
England	FTSE 100	^FTSE
	Gross Domestic Product England	GDP_{Eng}
Japan	NIKKEI 225	^N225
	Gross Domestic Product Japan	GDP_{Jap}

Source: Author

<u>Brief explanation of each variable:</u>

S&P 100 (Standard & Poor’s 100) is the index used to measure the performance of large cap companies in the United-States. It is made up of 100 major blue chip companies that span multiple industry groups. In order to be included in the index, a company must make individual stock options available for each constituent, individual stocks option are listed for each index constituent (S&P Dow Jones Indices, 2017).

FTSE 100 or Financial Times Stock Exchange 100 stock index is an index based on the prices of hundred leading companies in London. This is the representative of approximately 80% of market capitalization of the London Stock Exchange in its entirely. Larger companies comprise a greater portion of the index because it is weighted by market capitalization (Investopedia, 2017).

Nikkei 225 is the short of *Japan Nikkei 225 Stock Average*, it is the leading and most respected index of Japanese stocks. It is price weighted index comprised of Japan's top 225 blue-chip companies traded on the Tokyo Stock Exchange (Investopedia, 2017).

2. Period of the Study

To assess the causal interrelationship between capital market index and economic growth of united-states, England, and Japan, the study uses time series quarterly data for a period from 1987:1 to 2016:4 of three countries mentioned. The study uses a range of 720 observations. This period is using because it is quite enough and satisfactorily to obtain a good result of analysis.

Table 3
A Sketch Form of the Future Data Gathered

	United-States		England		Japan	
	S&P 100	Δ Economic Growth	FTSE 100	Δ Economic Growth	NIKKEI 225	Δ Economic Growth
1987Q1	$\frac{OEX_{1987q1} - OEX_{1986q4}}{OEX_{1986q4}}$	$\frac{GDP_{1987q1} - GDP_{1986q4}}{GDP_{1986q4}}$	$\frac{FTSE100_{1987q1} - FTSE100_{1986q4}}{FTSE100_{1986q4}}$	$\frac{GDP_{1987q1} - GDP_{1986q4}}{GDP_{1986q4}}$	$\frac{N225_{1987q1} - N225_{1986q4}}{N225_{1986q4}}$	$\frac{GDP_{1987q1} - GDP_{1986q4}}{GDP_{1986q4}}$
1987Q2	$\frac{OEX_{1987q2} - OEX_{1987q1}}{OEX_{1987q1}}$	$\frac{GDP_{1987q2} - GDP_{1987q1}}{GDP_{1987q1}}$	$\frac{FTSE100_{1987q2} - FTSE100_{1987q1}}{FTSE100_{1987q1}}$	$\frac{GDP_{1987q2} - GDP_{1987q1}}{GDP_{1987q1}}$	$\frac{N225_{1987q2} - N225_{1987q1}}{N225_{1987q1}}$	$\frac{GDP_{1987q2} - GDP_{1987q1}}{GDP_{1987q1}}$
…	…	…	…	…	…	…
2016Q3	$\frac{OEX_{2016q3} - OEX_{2016q2}}{OEX_{2016q2}}$	$\frac{GDP_{2016q3} - GDP_{2016q2}}{GDP_{2016q2}}$	$\frac{FTSE100_{2016q3} - FTSE100_{2016q2}}{FTSE100_{2016q2}}$	$\frac{GDP_{2016q3} - GDP_{2016q2}}{GDP_{2016q2}}$	$\frac{N225_{2016q3} - N225_{2016q2}}{N225_{2016q2}}$	$\frac{GDP_{2016q3} - GDP_{2016q2}}{GDP_{2016q2}}$
2016Q4	$\frac{OEX_{2016q4} - OEX_{2016q3}}{OEX_{2016q3}}$	$\frac{GDP_{2016q4} - GDP_{2016q3}}{GDP_{2016q3}}$	$\frac{FTSE100_{2016q4} - FTSE100_{2016q3}}{FTSE100_{2016q3}}$	$\frac{GDP_{2016q4} - GDP_{2016q3}}{GDP_{2016q3}}$	$\frac{N225_{2016q4} - N225_{2016q3}}{N225_{2016q3}}$	$\frac{GDP_{2016q4} - GDP_{2016q3}}{GDP_{2016q3}}$

Source: Author

C. Method of Analysis

1. Time Series Stationary:

A random or stochastic process is a collection of random variables ordered in time (Gujarati, 2003:797). A stationary stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two times periods depends only on the distance or gab or lag between the two time periods and not the actual time at which the covariance is computed. A stochastic process can be weakly stationary. To explain weak stationary, let Y_t be a stochastic time series with these properties:

Mean: $E(Y_t) = \mu$ (1)

Variance: $var(Y_t) = E(Y_t - \mu)2 = \sigma 2$ (2)

Covariance: $\gamma k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$ (3)

where γk, the covariance (or autocovariance) at lag k, is the covariance between the values of Y_t and Y_{t+k}, that is, between two Y values k periods apart. If $k = 0$, we obtain $\gamma 0$, which is simply the variance of Y (= $\sigma 2$); if $k = 1$, $\gamma 1$ is the covariance between two adjacent values of Y. Suppose we shift from Y_t to Y_{t+m}, if Y_t is to be stationary, the mean, variance, and autocovariances of Y_{t+m} must be the same as those of Y_t.

If a time series is stationary, its mean, variance, and autocovariance (at various lag) remain the same no matter at what point we measure them; that is, they are time invariant. Such a time series will tend to return to its mean (called mean reversion) and fluctuations around this mean (measured by its variance) will have a broad constant amplitude. Oppositely, a nonstationary time series will have a time-varying mean or time-varying variance or both.

Stationary time series are so important because if a time series is nonstationary, we can study its behavior only for the time period under consideration. Each set of times series data will therefore be used for a particular episode. As a consequence, it is not possible to generalize it to other time periods. Therefore, for the purpose of forecasting, such (nonstationary) time series may be of little practical value.

a. Dickey-Fuller Test

DF test, developed by statisticians David Dickey and Wayne Fuller (1979), tests the null hypothesis of whether unit root is present in autoregressive model (Gujarati, 2003:814).

Let Y_t a stochastic process: $Y_t = pY_{t-1} + u_t$ $\quad -1 \leq p \leq 1 \quad (4)$

where u_t is a white noise error term. If p=1, (4) becomes a random walk model without drift, that means equation (4) becomes $Y_t = Y_{t-1} + u_t$ which is a nonstationary stochastic process.

The general idea behind the unit root test of stationary is to regress Y_t on its lagged value Y_{t-1} and find out if the estimated p is statistically equal to 1. If it is, them Yt is nonstationary (Gujarati, 2003:814).

$$Y_t – Y_{t-1} = pY_{t-1}- Y_{t-1} + u_t$$

$$\Delta Y_t = (p-1)Y_{t-1} +u_t$$

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (5)$$

Where $\delta = (p-1)$ is a coefficient and Δ the first difference operator.

In practice, we estimate and test the null hypothesis that $\delta = 0$. If $\delta = 0$, then $p = 1$, that is we have a unit root, meaning the time series under consideration is nonstationary.

Dickey and Fuller have shown that under the null hypothesis that $\delta = 0$, the estimated t value of the coefficient of Y_{t-1} in equation (5) follows the tau statistic. These authors have computed the critical value of the tau statistic on the basis of Monte Carlo simulations. The tau statistic or test is known as the DF test. If the hypothesis that $\delta = 0$ is rejected i.e. the time series is stationary, we can use the usual t-test.

To allow for the various possibilities, the DF test is estimated in three different forms, that is, under three different null hypotheses.

Y_t is random walk $\quad \Delta Y_t = \delta Y_{t-1} + u_t \quad$ (6)

Y_t is random walk with drift $\quad \Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \quad$ (7)

Yt is a random walk with drift around stochastic trend $\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$ (8)

where t is the time or trend variable.

In each case, the null hypothesis is that $\delta = 0$; that is, there is a unit root—the time series is nonstationary. The alternative hypothesis is that δ is less than zero ($\delta<0$); that is, the time series is stationary. If the null hypothesis is not supported, it means that Yt is a stationary time series with zero mean in the case of equation (6), that Yt is stationary with a nonzero mean [$=\beta_1/(1-\rho)$] in the case of equation (7), and that Yt is stationary around a deterministic trend in equation (8).

b. Augmented Dickey-Fuller Test

The error terms u_t in equations (6), (7), and (8) were assumed to be uncorrelated in order to conduct the Dickey-Fuller test. But if these error terms were correlated, a test known as augmented Dickey-Fuller ADF test has been developed by Dickey and Fuller in purpose to check the unit root test stationary. The ADF test is conducted by augmenting the preceding three equations ((6), (7) and (8)) by adding the lagged values of the dependent variable ΔYt. To be specific, suppose we use equation (8). The ADF test here consists of estimating the following regression:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \tag{9}$$

where ε_t is a pure white noise error term and where $\Delta Y_{t-1}=(Y_{t-1} - Y_{t-2})$, $\Delta Y_{t-2}=(Y_{t-2} - Y_{t-3})$, etc. The number of lagged difference terms to include is often determined empirically, the idea being to include enough terms so that the error term in equation (9) is serially uncorrelated. In ADF we still test whether δ = 0 and the ADF test follows the same asymptotic distribution as the DF statistic, so the same critical values can be used (Gujarati, 2004:617).

2. Engle and Granger Causality Test

After testing the stationary of the time series variables i.e. of capital market index and variation of gross domestic product data, the next step of our methodology is to test the bivariate correlation of the time series. To find out whether the two variables were interrelated between each other, the research employs Granger causality Test.

To explicit the Granger causality Test, consider our main statement problem: is the economic growth causes the variation of capital market index (growth of gross domestic product Granger causes the variation Capital markets index)? Or is the variation of capital market index causes the economic growth (variation of capital markets index Granger causes growth gross of domestic product)? The Granger causality test assumes that the information relevant to the prediction of the respective variables, GDP and Capital market index, is contained solely in the time series data on these variables. The test involves estimating the following regressions (Gujarati, 2003:697):

$$GDP_t = \sum_{i=0}^{n} \alpha_i \, CM_{t-i} + \sum_{j=0}^{n} \beta_j \, GDP_{t-j} + u_{1t} \quad (10)$$

$$CM_t = \sum_{i=0}^{n} \lambda_i \, CM_{t-i} + \sum_{j=0}^{n} \delta_j \, GDP_{t-j} + u_{2t} \quad (11)$$

Assume that the disturbances u_{1t} and u_{2t} are uncorrelated.

Equation (10) presumes that current gross domestic product (GDP_t) relies on the past values of GDP itself (GDP_{t-j}) and on the past values of capital market index

(CM_{t-i}). And equation (11) affirms that the current value of capital market index (CM_t) depends on the past value of capital market index (CM_{t-i}) and gross domestic product (GDP_{t-j}).

The causality between capital market index and economic growth depends on the value of slop parameters α_i in equation (2) and δ_j in equation (3). If $\sum_{i=0}^{n} \alpha_i \neq 0$, the test concludes that gross domestic product is granger caused by capital market index, and vice-versa. If $\sum_{j=0}^{n} \delta_j \neq 0$, the test concludes that capital market index is granger caused by gross domestic product, and inversely.

Based on Gujarati (2003), four cases may appear:

a. **Unidirectional causality**: Granger causality test reveals a one direction causality from capital market index to economic growth if the coefficient on the historical lagged of the capital market index in equation (10) is significantly different from zero (i.e. $\sum \alpha_i \neq 0$) and the coefficient of gross domestic product on equation (11) is not statistically different from zero (i.e. $\sum \delta_j \neq 0$).

b. Conversely, **unidirectional** causality from economic growth to capital market index reveals if the coefficient of capital market index in equation (10) is not statistically different from zero (i.e., $\sum\alpha_i=0$) and the coefficient in (11) is statistically different from zero (i.e., $\sum\delta_j\neq0$).

c. **Feedback**, or **bilateral causality** between capital market index and economic growth, is observed when the coefficients of capital market index and GDP are statistically significantly different from zero in the two equations (10) and (11), i.e. ($\sum \alpha_i \neq 0$ and $\sum \delta_j \neq 0$).

d. Finally, **independence** between capital market index and economic growth is concluded when the coefficient of capital market index and are not statistically significant in two equations (10) and (11), i.e. ($\sum \alpha_i = 0$ and $\sum \delta_j = 0$).