# Estimation, Analysis and Projection of India's GDP

Daga, Ugam Raj and Das, Rituparna and Maheshwari, Bhishma

2004

# Estimation, Analysis and Projection of India's GDP
# A Time Series Model

U. R Daga, Rituparna Das and Bhishma Maheshwari

## 1. Introduction

Gross domestic product or GDP, tells us the country's current aggregate production of goods and services. It is often considered the best measure of how well the economy is performing. GDP summarizes the aggregate of all economic activities in a given period of time. In any economy, however, goods and services produced are not homogenous. It is not possible to add, for example, 10 barrels of petroleum with 10 million matric tons of wheat. So, as a trick, quantities and volumes of all respective goods and services are multiplied by their prices and then summed up. This gives the money value of GDP. Prices however include indirect business taxes (IBT) i.e. sales taxes and excise duties. So this GDP is not a true measure of the productive activities in the economy. In order to get a true measure of GDP we deduct IBT from GDP. This is called GDP at factor cost. For all practical purposes the government uses data on GDP at factor cost. The government of India has started Economic Reform program following the guidelines of IMF and World Bank with a number of ends keeping in view, one of which is that this program would boost up the annual growth of GDP through liberalizing trade. The philosophy of comparative advantage tells that free trade can increase the GDP of the trading countries.

## 2. Research Questions

1. Whether GDP data of India is stationary/non stationary?
2. What is the form of GDP trend equation?
3. Which sector(s) of the economy is(are) responsible for such shape?
4. Did economic reforms have any impact on GDP?
5. Is the estimated GDP trend able to predict future value of GDP?

## 3. Objectives of the study

Following are the quadruple objectives of the study

1. Examine the trend of GDP in order to see whether there is any difference in the trend between the periods from 1981-82 to 1990-91 and from 1991-92 to 2001-2002, i.e. two decades before and after the commencement of liberalization program.
2. Examine the nature of such difference, if it exists
3. Find the causes of such difference
4. Produce a model of forecasting GDP.

## 4. What is trend?

Trend is the general tendency of the variable under consideration to take increasing or decreasing values over a long period of time. Trend is also called 'secular trend'. It exists in time series data on any economic or business variable if there is a smooth, long run and general tendency of the variable take increasing or decreasing values over the given period. Trend is the long run component of time series data. If the variable does not show the tendency to take increasing or decreasing values then it is deemed to have no trend.

Trend analysis is a time series analysis. Time series analysis is used to detect patterns of change in statistical information over regular intervals of time. These patterns are used to project the future and help to arrive at an estimate for the future. Thus, time series analysis helps projection of the future value of variable through curve fitting.

## 5. What is Stationarity of Time Series Data and What are Its Consequences?

Any time series data has an underlying stochastic process. A stochastic process is called stationary if its mean and variance are constant over time and the covariance between the values at two different periods depends only on the lag between those periods and not on the time of calculation of covariance.

There are two key concepts in time series analysis:

1. Trend stationary process (TSP): If in the regression $Y_t = a + bt + u_t$, error term $u_t$ is stationary then $Y_t = a + bt + u_t$ represents a TSP.
2. Difference stationary process (DSP): If $Y_t$ is generated as $Y_t - Y_{t-1} = c + u_t$, where c is a constant and $u_t$ is stationary then the process is called a DSP.

The consequence of a non-stationary time series data is that it makes least square estimators inconsistent and diagnostic statistics like *t* and *F* statistics do not have their standard limiting distributions. As a consequence the regression of an explanatory variable may appear significantly different from zero though it is not truly a determinant of the dependent variable.

## 6.What is the link between stationarity and trend estimation?

The practical significance of TSP lies in long run forecasting. Forecasting made from TSP is reliable. Forecasting made from a DSP is unlike a TSP, not reliable (Gujarati 1995).

## 7. Examining and Ensuring Stationarity

Nowadays test of stationarity is done by EViews software, which draws the correlogram and also runs the unit root tests. With help of EViews software we check whether GDP of India is a TSP or a DSP.

## 8.Choice of Method to Estimate/Measure Trend

There are four methods to estimate trend – a. Graphic Method, b. Semi Average Method, c. Method of Curve Fitting and d. Method of Moving Average. Out of these four methods we have chosen the third method because of three reasons.  Firstly it is characterized by the maximum degree of objectivity compared to fist, secondly the data on GDP fulfills all preconditions for a good curve fitting and thirdly the nature of data on GDP during the aforesaid periods is such that there is no need for the fourth method, because it is annual data and does not seem to contain cyclical component. The data takes the form of a smooth exponential curve when plotted on a two dimensional graph against time.

### 9. Collection of Sample

Government of India publishes every year the data on economic variables in Economic Survey. We have taken the time series data on GDP and individual GDP components of the five sectors of the Indian economy (a) Agriculture, Forestry and Logging, Fishing, Mining and Quarrying; (b) Manufacturing, Construction, Electricity, Gas and Water Supply; (c) Trade, Transport, Storage and Communication; (d) Financing, Insurance, Real Estate and Business Services; and (e) Public Administration and Defence and other services at 1993-94 prices from Economic Survey 2002-03. The data are secondary and the sample period from 1981-82 to 2001-2002 is divided into two sub-sample periods as per the first objective of the study.

### 10. Nature of Study

We use these data and feed them into the Microsoft Excel's 'Data Analysis' package. Any computer loaded with Microsoft Excel Software is useful for this purpose. So this is a doctrinaire study. It does not involve field survey. It includes an empirical element in the sense that it studies real life data on GDP.

### 11. Dependent and Independent Variables

Trend analysis of any economic variable involves one independent variable, time and one dependent variable, the economic variable under consideration. After finding the trend, if one further wants to analyze the causes of the trend then such analysis calls for involvement of a number of independent variable. Precisely the same is this study. First we estimate trend of GDP over the two sample periods. We find that GDP trend has taken a turn right after the onset of the liberalization program. We try to find the cause of the turn and examine the impact of every individual sector on the trend. So here we include four independent variables – the GDP components or contributions to GDP of four sectors of the economy. Finally we have included an independent dummy variable in order to segregate the individual trends of pre and post liberalization-commencement periods. So there are one dependent variable GDP and six independent variables - four sectoral GDPs, a time variable and a dummy variable

### 12. Steps of Data Analysis

Following are the steps in estimation and analysis of GDP trends during the two sample periods:

1. Plotting the annual GNP value against the corresponding year and draw a line diagram by connecting all the plotted points. Ascertain the nature of temporal movement of the value of the GDP variable over the sub-sample periods.

2. We have to check whether GDP is a non-stationary process. So we have to run unit root test. If GDP is found non-stationary, then we have to see whether GDP is a TSP after estimating trend.

3. Propose an exponential trend relationship between GDP as the dependent variable and time and a dummy variable for trade liberalization as independent variables and check that the data satisfies the prescribed criteria for running regression of GDP on time and the dummy variable for the entire sample period. We run above multiple variable regression using 'Regression' package of the Microsoft Excel Software. We examine the goodness of fit in terms of adjusted $R^2$ and 't' values at 95% level of significance. We drop the variable whose coefficient has a calculated absolute 't' value less than the table value.

4. On satisfying the conditions for a regression exercise we accept the model for regression purpose.

5. After estimating the model we calculate the residuals and run unit root test for $1^{st}$ difference. Thus we check whether GDP is a TSP.

6. We make the estimation of GDP within the sample for all the years and plot them on graph to find the difference between the estimated line and the actual line with an intension of improvement if there is any wide difference. We estimate a 95% confidence level interval for forecasting GDP.

7. In order to detect which sector of the economy is responsible for the turn of GDP right after liberalization we examine the correlations between GDP and each sector's contribution to GDP separately for pre and post liberalization-commencement periods. We ascertain which sector contributes maximum and which sector contributes minimum to the movement of GDP over the sample period, for, this information would be of use to economic planners.

All the above steps are performed in the appendix.

## 13. CONCLUSION

GDP of India is a stationary process. The trend equation proves a good fit after we drop the dummy variable. It gives a result contrary to the belief that economic reform causes a boost in the GDP. It gives however an adjusted $R^2$ as high as 99.7%. All the 't' values are found highly significant. While plotted on graph, the estimated GDP line just coincides with the actual line. So this estimation can be used for the purpose of GDP forecasting. This model has tracked well the path of past movements in the value of the variable. The sector comprising Trade, Transport, Storage and Communication is found to contribute the maximum and the sector comprising Financing, Insurance, Real Estate and Business Services is found to contribute the minimum to the GDP trend under study.

## 14. Appendix

### Step 1

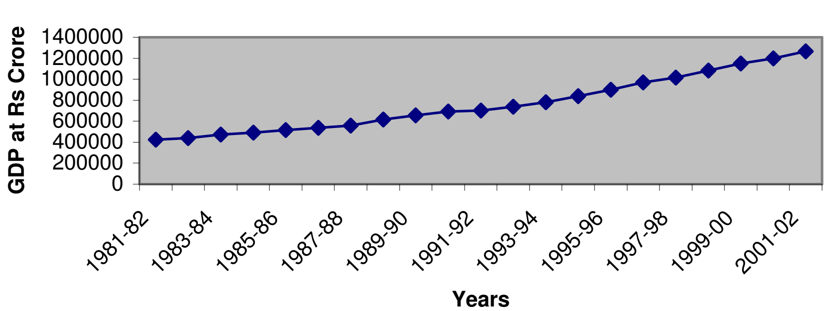
We plot Table 1 in Figure 1

**Table 1**

GDP from1981-81 to 2001-02 at 1993-94 prices

| Year | GDP at Factor cost (Rs Crores) |
|---|---|
| 1981-82 | 425073 |
| 1982-83 | 438079 |
| 1983-84 | 471742 |
| 1984-84 | 492077 |
| 1985-86 | 513990 |
| 1986-87 | 536257 |
| 1987-88 | 556778 |
| 1988-89 | 615098 |
| 1989-90 | 656331 |
| 1990-91 | 692871 |
| 1991-92 | 701863 |
| 1992-93 | 737792 |
| 1993-94 | 781345 |
| 1994-95 | 838031 |
| 1995-96 | 899563 |
| 1996-97 | 970083 |
| 1997-98 | 1016594 |
| 1998-99 | 1082748 |
| 1999-00 | 1148442 |
| 2000-01 | 1198685 |
| 2001-02 | 1265429 |

**Figure 1**




Source: *Economic Survey 2002-03*, Government of India, Delhi

The data given above is plotted on the graph. The 'X' axis has been taken as Years from 1981-82 to 2001-02, whereas, 'Y' axis has been taken as GDP. By seeing the line we come to the conclusion that GDP of India is showing an increasing trend with time. From the graph we can also infer that, as there are no fluctuations in the line, so there is no cyclical variation in the data. The data taken is annual, so there is no element of seasonal variation in the data. This proves that the given data is fit for regression analysis, provided, the following preconditions are satisfied: a. stationarity, b. appropriateness of nonlinear relationship, b. absence of autocorrelation in error terms, c. homoscadasticity of error terms and d. absence of multicollinearity between independent variables. If the series is not stationary we have to see whether it is trend stationary or difference stationary.

**Step 2**

We have to see whether GDP is a stationary series. After running alternative Dicky-Fuller unit root test for GDP at level without any lagged difference and without any intercept, we find that computed $|\tau|$ value is above all Mackinnon critical values. So we reject the hypothesis that GDP data is non-stationary. The results are given in the end of the appendix. Now we have to see whether GDP is a TSP after estimating trend.

**Step 3**

The first proposed equation is

$\log Y = a + bX + cD + u$,

where log Y = log natural of GDP, a = intercept coefficient, b = coefficient of time variable and c = coefficient of dummy variable, for the period 1981 – 82 the dummy variable will be '0' and for the rest of the period it will be '1' in order to capture the effect of liberalization on GDP, u = error

Assumptions:

1. No multicollinearity: There is no significant correlation between time variable and dummy variable.
2. No autocorrelation: There does not exist any correlation between $u_t$ and $u_{t-1}$.
3. Homoscadasticity: Every u is independently normally distributed with zero mean and uniform variance. This is checked by running regression of estimated u on X. A high adjusted $R^2$ and significant t values ascertain heteroscadasticity and viec versa.

In order to check condition (i) we calculate the correlation coefficient between the dummy variable and time. We get the value of the correlation coefficient as high as 86.6%. So we drop the dummy variable and decide that GDP grows by its own in the post liberalization period. Uchikawa1999 supports this conclusion.

**Step 4**

The modified data is in Table 2 followed by the multiple regression output

**Table 2**

| log Y | Year | X |
|---|---|---|
| 5.6285 | 1981-82 | 1 |
| 5.6416 | 1982-83 | 2 |

| | | |
|---|---|---|
| 5.6737 | 1983-84 | 3 |
| 5.692 | 1984-85 | 4 |
| 5.711 | 1985-86 | 5 |
| 5.7294 | 1986-87 | 6 |
| 5.7457 | 1987-88 | 7 |
| 5.7889 | 1988-89 | 8 |
| 5.8171 | 1989-90 | 9 |
| 5.8407 | 1990-91 | 10 |
| 5.8463 | 1991-92 | 11 |
| 5.8679 | 1992-93 | 12 |
| 5.8928 | 1993-94 | 13 |
| 5.9233 | 1994-95 | 14 |
| 5.954 | 1995-96 | 15 |
| 5.9868 | 1996-97 | 16 |
| 6.0071 | 1997-98 | 17 |
| 6.0345 | 1998-99 | 18 |
| 6.0601 | 1999-2000 | 19 |
| 6.0787 | 2000-01 | 20 |
| 6.1022 | 2001-02 | 21 |

We run regression of log Y on X only. The summary output is as follows:

SUMMARY OUTPUT

| Regression Statistics | |
|---|---|
| Multiple R | 0.99838 |
| R Square | 0.99676 |
| Adjusted R Square | 0.99659 |
| Standard Error | 0.00878 |
| Observations | 21 |

ANOVA

| | df | SS | MS | F | Significance F |
|---|---|---|---|---|---|
| Regression | 1 | 0.45115 | 0.45115 | 5851.25 | 4E-25 |
| Residual | 19 | 0.00146 | 7.7E-05 | | |
| Total | 20 | 0.45261 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|---|---|---|---|---|---|---|---|
| Intercept | 5.59194 | 0.00397 | 1407.36 | 3.8E-49 | 5.58363 | 5.60026 | 5.58363 | 5.60026 |
| X Variable 1 | 0.02421 | 0.00032 | 76.4935 | 4E-25 | 0.02354 | 0.02487 | 0.02354 | 0.02487 |

We find all t values are highly significant.
Before going to accept the model for forecasting we check conditions (ii) and (iii).
For condition (ii) we see Correlation coefficient between $e_t$ and $e_{t-1}$ = -0.01943 and correlation coefficient between $e_t$ and $e_{t-2}$ = 0.054776, which are negligible. Hence we decide that the problem of autocorrelation does not exist.
And for condition (iii) we run regression of estimated u on X and get the following

SUMMARY OUTPUT

| Regression Statistics | |
|---|---|
| Multiple R | 0.160100108 |
| R Square | 0.025632045 |
| Adjusted R Square | -0.025650479 |
| Standard Error | 14466.03088 |
| Observations | 21 |

| | Coefficients | Standard Error | t Stat |
|---|---|---|---|
| Intercept | -3550.42381 | 6545.969363 | -0.5423832 |
| X Variable 1 | 368.5623377 | 521.3195663 | 0.706979675 |

From the above results, we can see that Adjusted R square is coming as 0.025650479, which is very small. It shows that there is no relation between the time and residuals and hence the problem of heteroscadasticity does not exist.

**Step 5**
We have to check whether estimated residuals from above constitute a stationary series. The estimated residuals are $u_t$* given in
**Table 3**

$u_t$*

11906
1228
9849
3706
-2376
-9709
-20485
4744
10989
10535
-19587
-25015
-25189
-14736
-2088
16745
8607
16979
21579
7226
5671

We run the unit root test of $u_t$* for 1st difference and find that computed |τ| exceed Mackinnon critical values of 1%, 5% and 10%. So we conclude that GDP of India is a trend stationary process. The result of the test is given in the end of the appendix.

**Step 6**

Now, without hesitation we accept the model log Y = a + bX + u for the purpose of forecasting. The model comes out to be: log Y = 5.591917834 + 0.024208073 X

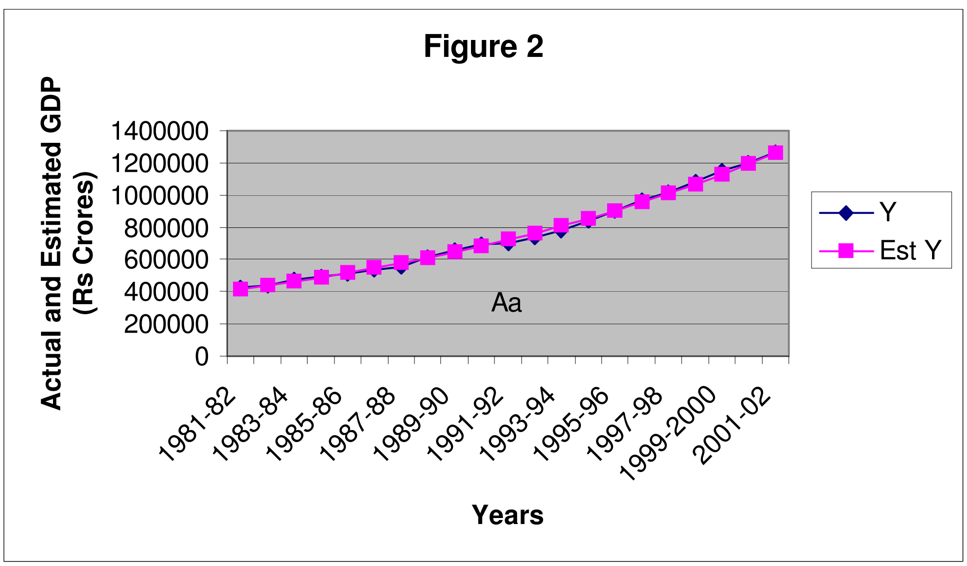
Estimated $Y = 10^{(5.591917834 + 0.024208073\,X)}$

We get the table 4 using the above model:

**Table 4**

| Year | Y | Estimated Y |
|---|---|---|
| 1981-82 | 425073 | 413167 |
| 1982-83 | 438079 | 436851 |
| 1983-84 | 471742 | 461893 |
| 1984-85 | 492077 | 488371 |
| 1985-86 | 513990 | 516366 |
| 1986-87 | 536257 | 545966 |
| 1987-88 | 556778 | 577263 |

| | | |
|---|---|---|
| 1988-89 | 615098 | 610354 |
| 1989-90 | 656331 | 645342 |
| 1990-91 | 692871 | 682336 |
| 1991-92 | 701863 | 721450 |
| 1992-93 | 737792 | 762807 |
| 1993-94 | 781345 | 806534 |
| 1994-95 | 838031 | 852767 |
| 1995-96 | 899563 | 901651 |
| 1996-97 | 970083 | 953338 |
| 1997-98 | 1016594 | 1007987 |
| 1998-99 | 1082748 | 1065769 |
| 1999-2000 | 1148442 | 1126863 |
| 2000-01 | 1198685 | 1191459 |
| 2001-02 | 1265429 | 1259758 |




Plotting table 2 in figure 2 shows that actual and estimated trends almost coincide.

So, the point estimation of GDP for 2002-03 is 1259759. The estimated interval is 1259759 ± 2(14664.67) or (1230429.4, 1289088) at 95% confidence interval. The actual figure lies in the estimated interval.

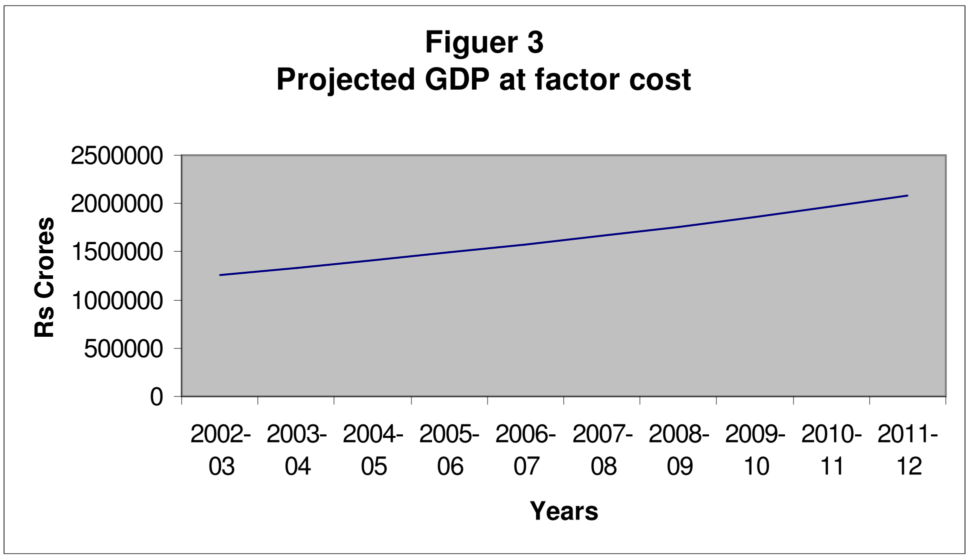
Nest we project the GDP figures for next years beyond the sample period on the basis of our model. We find the following table 5 and figure 3:

**Table 5**

| **Year** | **Projected GDP** |
|---|---|
| 2002-03 | 1259759 |
| 2003-04 | 1331973 |
| 2004-05 | 1408327 |
| 2005-06 | 1489058 |

| | |
|---|---|
| 2006-07 | 1574417 |
| 2007-08 | 1664668 |
| 2008-09 | 1760094 |
| 2009-10 | 1860989 |
| 2010-11 | 1967669 |
| 2011-12 | 2080463 |

This graph shows that our model has well captured the historical behavior of the variable.




**Step 7**

We check which sector contributes the maximum/minimum to the above shape of GDP line. We check correlation between GDP and each sector's GDP separately for pre and post liberalization periods. We get following two matrices:

Pre liberalization Period Correlation Coefficient Matrix

| Sectors | a | b | c | d | e | GDP |
|---|---|---|---|---|---|---|
| A | 1 | | | | | |
| B | 0.963554 | 1 | | | | |
| C | 0.944304 | 0.993814 | 1 | | | |
| D | 0.954695 | 0.997821 | 0.998476 | 1 | | |
| E | 0.943014 | 0.993354 | 0.999083 | 0.998394 | 1 | |
| GDP | 0.97374 | 0.998143 | 0.99376 | 0.997277 | 0.993249 | 1 |

(a. Agriculture, Forestry and Logging, Fishing, Mining and Quarrying; b. Manufacturing, Construction, Electricity, Gas and Water Supply; c. Trade, Transport, Storage and Communication; and d. Financing, Insurance, Real Estate and Business Services; e. Public Administration and Defence and other services)

Post liberalization Period Correlation Coefficient Matrix

| | a | b | c | d | E | GDP |
|---|---|---|---|---|---|---|
| A | 1 | | | | | |
| B | 0.975037 | 1 | | | | |
| C | 0.965765 | 0.991862 | 1 | | | |
| D | -0.29099 | -0.30126 | -0.2663 | 1 | | |
| E | 0.928737 | 0.961251 | 0.986338 | -0.20592 | 1 | |
| GDP | 0.975981 | 0.994378 | 0.998573 | -0.2739 | 0.982284 | 1 |

Comparison of above matrices shows that sector c has contributed maximum and sector d has contributed minimum in terms of change in the magnitude of correlation coefficient of the respective sectors with GDP from pre to post liberalization period.

The actual data of GDP and it's sectoral components taken from Economic Survey 2002-03 is

| Year | a | b | c | d | e | GDP |
|---|---|---|---|---|---|---|
| 1981-82 | 177341 | 93029 | 78387 | 28336 | 47979 | 425073 |
| 1982-83 | 177300 | 95695 | 82001 | 31272 | 51811 | 438079 |
| 1983-84 | 193508 | 103992 | 86013 | 34391 | 53838 | 471742 |
| 1984-85 | 196353 | 110474 | 90426 | 37320 | 57504 | 492077 |
| 1985-86 | 198353 | 115689 | 97555 | 41126 | 61267 | 513990 |
| 1986-87 | 198740 | 122847 | 103327 | 45768 | 65575 | 536257 |
| 1987-88 | 196735 | 131417 | 108742 | 49598 | 70286 | 556778 |
| 1988-89 | 227095 | 142738 | 115229 | 55251 | 74785 | 615098 |
| 1989-90 | 231389 | 157979 | 123740 | 62204 | 81020 | 656331 |
| 1990-91 | 242012 | 169703 | 129786 | 66990 | 84380 | 692871 |
| 1991-92 | 239253 | 167967 | 133080 | 75027 | 86536 | 701863 |
| 1992-93 | 252205 | 175175 | 140487 | 794030 | 90494 | 737792 |
| 1993-94 | 262059 | 185070 | 150500 | 90084 | 93632 | 781345 |
| 1994-95 | 276049 | 204092 | 166131 | 95085 | 96674 | 838031 |
| 1995-96 | 275153 | 229098 | 188167 | 102847 | 104298 | 899563 |
| 1996-97 | 299461 | 246848 | 202936 | 109995 | 110843 | 970083 |
| 1997-98 | 295050 | 256121 | 218822 | 122784 | 123817 | 1016594 |
| 1998-99 | 312485 | 265956 | 235757 | 131892 | 136658 | 1082748 |
| 1999-2000 | 314253 | 279130 | 255817 | 145863 | 153379 | 1148442 |
| 2000-01 | 313806 | 298689 | 273380 | 150910 | 161900 | 1198685 |
| 2001-02 | 330272 | 309291 | 297213 | 157701 | 170952 | 1265429 |

**Result of alternative Dicky-Fuller unit root test for GDP at level without any lagged difference and without any intercept**

| | | | | |
|---|---|---|---|---|
| Level test | 15.22617 | 1% | Critical Value* | -2.6889 |
| ADF Test Statistic | | 5% | Critical Value | -1.9592 |
| | | 10% | Critical Value | -1.6246 |

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GDP)
Method: Least Squares
Date: 02/13/04 Time: 07:44
Sample(adjusted): 1981-82 to 2001-02, Annual figures
Included observations: 20 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| GDP(-1) | 0.057179 | 0.003755 | 15.22617 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.558632 | Mean dependent var | 42017.80 |
| Adjusted R-squared | 0.558632 | S.D. dependent var | 19621.74 |
| S.E. of regression | 13035.80 | Akaike info criterion | 21.83749 |
| Sum squared resid | 3.23E+09 | Schwarz criterion | 21.88728 |
| Log likelihood | -217.3749 | Durbin-Watson stat | 1.811311 |

This shows that the variable is stationary.

**References**


Branson W. H., (2002), *Macroeconomic Theory and Policy*, AITBS Publishers, New Delhi, 2nd edition, Chapter 2

Delurgio S. A., (1998), *Forecasting Principles and Applications*, Irwin McGraw-Hill, Boston, 1st edition, Chapter 3

Gujarati D. N., (1995), *Basic Econometrics*, McGraw-Hill, Inc., New York, 3rd Edition, Chapter 21

Krugman P. R., (2000), International Economics, Theory and Policy, Addision Wesley Longman Singapore, Delhi, 5th edition, Chapter 2

Ministry of Finance and Company Affairs, (2003), *Economic Survey 2002-2003*, Government of India Press, New Delhi

Uchikawa S., (1999), Economic Reforms and Foreign Trade Policies Case Study of Apparel and Machine Tools Industries, *Economic and Political Weekly*, November 27