THE DATABASE

This Data Appendix explains the data along with their transformation used in the study as also their sources. The period of analysis is from 1950/1 to 1995/6 except for the cases where the data is not available. We shall use the following set of abbreviations for various publications for our purpose:

CPG  

CSO  
Central Statistical Organisation, Government of India.

ES  

MAS  

NAS  
*National Accounts Statistics*, published by the CSO (Annual).

RCF  

WIND  
*Wholesale Price Index*, published by the Ministry of Industry.

IPFS  
Indian Public Finance Statistics or Indian Economic Statistics (Public Finance) published by the Ministry of Finance

A.1 The data series and their transformations

The variables used in the analysis are as follows:

- $Y_{agr}$ refers to the GDPfc\(^1\) from the agricultural sector at current prices.
- $Q_{agr}$ refers to the GDPfc from the agricultural sector at constant (1980-81) prices.
- $Y_{ind}$ refers to the GDPfc from the industrial sector\(^2\) at current prices.

1. Refers to Gross Domestic Product at factor cost.

2. This consists of three broad sub-sectors: (i) Manufacturing; (ii) Mining and Quarrying, or Mining sub-sector; and (iii) Electricity, Gas and Water Supply, or the Electricity sub-sector. The manufacturing sub-sector with a value-added share of about 80 percent in the industrial sector has two broad sub-divisions: the Factory Sector and the Non-Factory Sector. It also
\( Q_{\text{ind}} \) refers to the GDPfc from the industrial sector at constant (1980-81) prices.

\( C_{\text{gov}} \) refers to the *Government final consumption expenditure* (or public consumption) at constant (1980-81) prices.

\( I_{r} \) refers to total real investment and is proxied by *Gross Capital Formation* (GCF) at constant (1980-81) prices. This includes GCF by *public sector, private corporate sector and household sector*.

\( I_{\text{pcs}} \) refers to *Gross Capital Formation (private corporate sector)* at current prices.

\( I_{\text{kh}} \) refers to *Gross Capital Formation (household sector)* at current prices.

\( I_{\text{gov}} \) refers to *public investment* and is proxied by *Gross Capital Formation (public sector)* at current prices.

\( Q_{\text{food}} \) is the *net production of foodgrains* (million tonnes).

\( \text{CPI}_{\text{AL}} \) is the *consumer price index (food) for agricultural labourers* (at 1970-71 base).

\( \text{CPI}_{\text{IW}} \) is the *consumer price index (food) for industrial workers* (at 1970-71 base).

\( P_{\text{agr}} \) refers to the GDP deflator for the agricultural sector. This is computed using the \( Y_{\text{agr}} \) and \( Q_{\text{agr}} \).

\( P_{\text{ind}} \) refers to the GDP deflator for the industrial sector. This is computed using the \( Y_{\text{ind}} \) and \( Q_{\text{ind}} \).

\( \text{WPI} \) *Wholesale Price Index for all commodities.*

\( M_{1} \) is the money supply with the public, is the narrow measure of money stock. It consists of currency and coins with the public, demand deposits of banks and salary earners' societies held by the public and 'other deposits' with the Reserve Bank. The most liquid and most generally accepted means of payment available as a medium of exchange and for final settlement of claims are included in \( M_{1} \), thus emphasizing the medium of exchange characteristic of money.\(^3\)

---

\(^3\)RBI(1965) p 97.
M₃ refers to the broad measure of money stock. It includes time deposits (excluding inter-bank time deposits) of banks in addition to M₁.⁴

H is the reserve money representing those liabilities of the central bank and the government that are deemed to be eligible for the purpose of deposit money creation in the system where fractional reserve ratio governs the creation of deposit money. Generally currency liabilities of the central bank and the government are considered as eligible for being held as bank reserves supporting deposit money creation. Accordingly reserve money in India is the sum total of currency with the public and bankers' deposits with the Reserve Bank, cash with the banks, and 'other deposits' with the RBI which are liabilities of the Reserve Bank to the non-bank sector, and hence equivalent to currency with the public in so far as their relevance to deposit money creation of the banks is concerned.⁵

Hg refers to the reserve money created by the Government of India.

FD refers to the Fiscal Deficit of the government, and is computed as difference between Total Expenditure and Total Revenue Receipts plus Non-debt creating Capital Receipts.

Dfin is used for Deficit Financing and is computed as the first difference of Hg.

RoIs is the rate of interest on one-year fixed deposit rate was taken to represent the short term rate.

RolI is the flat yield on non-terminable 3 percent Rupee was taken to represent the long term rate of interest.⁶

RRWages refers to the Real Rural Wages of Agricultural Workers in rupees at 1979-80 prices.⁷

---

⁴ Ibid, p 97.

⁵ In India reserve money consists of predominantly of currency with the public (Ibid, p 99-100).

⁶ See also Balakrishnan(1991).

⁷ Data for this series is taken from Papanek, 1989. p 140.
A.2 Data Sources

The source of all, except the RRWages, WPI, $Q_{food}$, $CPI_{AL}$ and $CPI_{IW}$ is Chandok and Policy Group (1990) for the period 1950-80 period. The data for $Q_{food}$, $CPI_{AL}$ and $CPI_{IW}$ series are taken from the ES. Since the $CPI_{IW}$ series is available only from 1965 onwards so we use the annual rate of growth of CPI (working class) to complete the series of $CPI_{IW}$ for the 1950-65 period for the purpose of our analysis in chapter 4. The $CPI_{AL}$ series is not available before the year 1966 so the use of this series is restricted to 1966 to 1995 period only as indicated in the Chapter 3.

The data source for the WPI series is the WIND.

The monetary variables, $H$, $Hg$, $M$, MoIs, RoII, used in the analysis for the 1980-95 period are taken from RCF. The data source for the rest of the series for the 1980-1995 period is the NAS. The FD is computed using data from the IPFS.
This Appendix discusses the econometric techniques used in the empirical analysis of the present study. First we elucidate the concept of stationarity and unit root test which is useful in determining the stationarity of a data series. After that, the technique of cointegration is explained.

B.1 Stationarity and Unit Root test

In the analysis we use the natural logarithm of the variables because of their variance stabilization property.

Since the variables with same order of integration can only be co-integrated, so as a first step we needed to determine the order of integration of the variables under consideration. The order of integration of a series refers to the number of times a variable needs to be differenced for it to become stationary.

A 'stationary' series tends to return to its mean value and fluctuate around it within a more-or-less constant range (i.e. it has a finite variance), while non-stationary series has a different mean at different points in time (and thus, the concept of mean is not necessarily applicable) and its variance increases with the sample size (Enders, 1995, p 358-64, Gujarati, 1995, p 713). The analysis with non-stationary variables can lead to the "problem of spurious regression whereby the results suggest that there are statistically significant long-run relationships between the variables in the regression model when, in fact, all that is there is evidence of contemporaneous correlations rather than meaningful casual relations".

For the purposes of establishing the order of integration (or the stationary properties) of the series used augmented Dickey-Fuller test (or the ADF test) were conducted on the time series of all the variables. This involved running following regression

$$\Delta X_t = \alpha_0 X_{t-1} + \sum_{i=2}^{4} \alpha_{1,i} \Delta X_{t-i} + e_t$$

If the coefficient of $X_{t-1}$, i.e. $\alpha_0$, was found to be negative and statistically significant according to the ADF-test statistic, the series is referred to as stationary series. It was also confirmed that all the variables, except Rols and $I$, were non-stationary when expressed in log levels. TABLE B.1 presents these results.

---

2 These three variables were found to be trend-stationary.
B.2 Cointegration and Error Correction Model

'Cointegration tests' were conducted, as the second step, to check whether these variables are cointegrated or not. We know that if two time series are both integrated of order \( d \), \( I(d) \), then in general any linear combination of the two series will also be \( I(d) \). If, however, there exists a vector \( b \), such that the disturbance term from the regression \( e_i = y_i - b.x_i \) is of lower order of integration, \( I(d-b) \), where \( b>0 \), then \( y \) and \( x \) are referred to as cointegrated of order \( (d,b) \), in short \( CI(d,b) \). Thus, if \( y \) and \( x \) were both \( I(1) \), and the least square residual of this regression equation \( e_i \) is found to be an \( I(0) \) variable, then the two series would be \( CI(1,1) \). This implies that if we wish to estimate the long-run relationship between \( y \) and \( x \) it is only necessary to estimate the static model. The concept of cointegration is very relevant to the problem of determination of long-run relationship in economics (Harris, (1995), p 8).

The next step was to check for the long-run relationship among these variables using the cointegration technique. For this, we first regress \( Y_t \) on \( X_t \) using the ordinary least squares (OLS) estimation technique. All the variables, as we have seen, are \( I(1) \) in log-levels, except \( L_t \) and RoIs. As these could give rise to the problem of spurious regression. This problem has been well documented in the literature. Such possibilities arise because most of the economic series exhibit non-stationary tendencies. High \( R^2 \) may arise due to the uncorrelated trends but not through economic relationships. Low DW statistic may be due to the non-stationary residuals. The standard method of overcoming this problem is to see whether the relationships discovered in the levels persists after first differencing the underlying series. The problem with such an approach is that it involves the loss of low frequency (long run) information. Cointegration and error correction modeling allows the identification of non-spurious relationships without forcing the loss of long run information (Engle and Granger, 1987). Moreover, error correction models allow for suitable economic interpretation since they incorporate equilibrium relationship as suggested by economic theory along with the possibility of variables responding to disequilibrium as favoured by time series econometricians (Enders, 1995, p 365-7, Gujarati, 1995, 723).

A common test for cointegration is the Engle & Granger (1987), or the EG, two step procedure. The following steps are involved in this procedure:

---

3 Engle and Granger(1987)

4 For example Granger and Newbold (1974)

5 This could be particularly likely when the \( R^2 \) is greater than the Durbin Watson (DW) statistic.
Before testing for cointegration one needs to ensure that the order of integration of the variables under consideration is same, as the variables of same order of integration can only be cointegrated. To do this standard Dickey Fuller (DF) and Augmented Dickey Fuller (ADF) tests are used. Following this, the first step in the EG two-step procedure involves running a least squares regression on the levels of the variables in question and testing the hypothesis of cointegration by determining the order of integration of the residuals. If the hypothesis of cointegration is to be accepted then the residuals must be integrated of order less than the order of integration of the series used in the regression. For instance, if dependent and independent variables are $I(1)$ then the LS residuals must be $I(0)$ for cointegration or long-run relationship to exist.

The first step of the EG procedure provides an intuitively compelling method of testing for stability. If, in fact, the relationship between the variables is stable then there must be cointegration among the variables.

The equation estimated in the first step above is generally referred to as the 'static regression'. The coefficients of this regression are referred to as the cointegrating parameters. The parameters can be interpreted as defining the long run (or equilibrium) relationship between the variables under consideration. After the presence of cointegration has been established the second step requires the construction of an error correction model (Gujarati, 1995, p 726, Enders, 1995, p 365-7).

Evidence of cointegration implies the existence of an error correction mechanism (ECM) and, therefore, we can set-up an ECM for better understanding of the short-run dynamics among the variables. This involves regressing the first difference of each variable in the cointegration onto the values of the first difference of all the variables plus the lagged value of the error correction term (which refers to LS residual from the cointegrating regression). The computed disturbance term (the LS residual) is used to capture the short-run dynamics for the explaining of the dependent variable if it is found to be integrated of lower order than the variables under consideration. The error correction term represents the extent of deviation from equilibrium explicitly in the short run regression equation and which, therefore, immediately displays information about the adjustment that a process makes to a deviation from the long-run relationship. The coefficient of the error correction term can be taken as a measure of the speed of adjustment of dependent variable ($Y_t$) to a discrepancy between $Y_t$ and $X_t$ (the independent variable) in the previous period. It is also referred to as short-run adjustment to a discrepancy. The role of differencing, here, is as a transform, which preserves co-

---

6 Error correction terms were used by Sargan (1964), Hendry and Anderson (1974) and Davidson et al (1978) as a way of capturing adjustment in a dependent variable which depends not only on the levels of explanatory variables, but also on the extent to which an explanatory variable deviated from an equilibrium relationship (Banerjee et al, 1993, p 501).
integration, and not as filter, which eliminates level variables and hence loses co-integration. The results are reported in the respective chapters along-with the critical values for the tests of presence of cointegration among the variables. The critical values are at always at 5 percent level of significance unless otherwise specified.

B.3 Causality Tests

a) bivariate case

Granger devised tests for causality starting from the premise that future cannot cause the present or the past. If event A occurs before event B, we know that B cannot cause A. At the same time, if A occurs before B, it does not necessarily imply that A causes B. In practice, we observe A and B as time series and we would like to know whether A precedes B or B precedes A, or are they contemporaneous. This is what is the purpose of tests of Granger causality. The test proceeds as follows:

Consider two time series, \{y_t\} and \{x_t\}. The series \(x_t\) *fails to Granger cause* \(y_t\) if in a regression of \(y_t\) on \(y_t\)'s and lagged \(x_t\)'s, the coefficient of latter are zero. That is, consider

\[
y_t = \sum_{i=1}^{k} \alpha_i y_{t-i} + \sum_{i=1}^{d} \beta_i x_{t-i} + e_t
\]

Then if \(\beta_i = 0\) (for \(i=1,2,...,k\)), then \(x_t\) fails to cause \(y_t\). The reverse regression tells us about the opposite Granger causality\(^7\). The detailed results are reported in the Endnotes of the respective chapters.

b) multivariate case

A dynamic specification of the equation was estimated in first difference of the variables with three set of lags, thought reasonable, of the variables. The model, as in the estimated equation, was progressively reduced to the most parsimonious representation possible. The Wald Coefficient test was, then, used to determine the importance of each series in its influence on the dependent variable. For the specific results refer to the Endnotes of the respective chapters.

\(^7\)The determination of lag length could be arbitrary (or subjective) or depending on the level of significance.
TABLE B.1: Test of Stationarity (Unit Root Tests)
(Using Augmented Dickey Fuller (ADF) Tests)

<table>
<thead>
<tr>
<th>Variables</th>
<th>log-levels (with trend)</th>
<th>log-difference (without trend)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample 1951-1995</td>
<td></td>
</tr>
<tr>
<td>Yagr</td>
<td>-2.26 (2)</td>
<td>-6.75* (0)</td>
</tr>
<tr>
<td>Yind</td>
<td>-1.78 (0)</td>
<td>-4.98* (0)</td>
</tr>
<tr>
<td>Qagr</td>
<td>-0.86 (5)</td>
<td>-5.12* (3)</td>
</tr>
<tr>
<td>Qind</td>
<td>-1.98 (1)</td>
<td>-4.85* (0)</td>
</tr>
<tr>
<td>Pagr</td>
<td>-2.92 (0)</td>
<td>-5.07* (0)</td>
</tr>
<tr>
<td>Pind</td>
<td>-2.74 (0)</td>
<td>-4.44* (0)</td>
</tr>
<tr>
<td>I_r</td>
<td>-3.74 (0)</td>
<td>-5.68* (3)</td>
</tr>
<tr>
<td>Cgov</td>
<td>-2.89 (1)</td>
<td>-4.40* (0)</td>
</tr>
<tr>
<td>I_gov</td>
<td>-2.82 (0)</td>
<td>-6.46* (1)</td>
</tr>
<tr>
<td>Ipcs</td>
<td>-4.84 (0)</td>
<td>-5.26* (3)</td>
</tr>
<tr>
<td>I_hh</td>
<td>-3.27 (0)</td>
<td>-8.49* (0)</td>
</tr>
<tr>
<td>M1</td>
<td>-2.24 (0)</td>
<td>-3.90* (1)</td>
</tr>
<tr>
<td>M3</td>
<td>-2.23 (1)</td>
<td>-4.46* (0)</td>
</tr>
<tr>
<td>H</td>
<td>-2.11 (0)</td>
<td>-3.32® (1)</td>
</tr>
<tr>
<td>Q</td>
<td>-2.68 (0)</td>
<td>-5.04* (3)</td>
</tr>
<tr>
<td>P</td>
<td>-3.29 (0)</td>
<td>-5.03* (0)</td>
</tr>
<tr>
<td>RoIs</td>
<td>-5.49 (1)</td>
<td>-5.57* (0)</td>
</tr>
</tbody>
</table>

NOTE: Figures in the parenthesis are the no. of significant lags (Ng and Perron, 1995). Rate of interest and Real investment has been found to be trend-stationary variables so they were detrended for achieving stationary transformations.

® indicates significance at 5 percent, * indicates significance at 1 percent level.