Chapter-3

OBJECTIVES, DATA BASE AND METHODOLOGY

3.0 **Objects of the study:**

After studying large literature, we have presented in this chapter the objects of the study, hypothesis to be tested, variables under study, the data sources then the methodologies to be used to test the hypothesis relating impact of corporate investment in the rural wholesale market and the strategic behaviour in the retail market of agricultural commodities. Before going to state object in detail, we should understand the nature of traders including small capital and large capital operating in the trading market. The types of traders are, first, $C_I$ i.e. large capital or corporate firm of input collector who use to collect from the farmers and intervene in the rural wholesale markets in collecting agricultural commodities as inputs in making final produce, for example, food processing firms, second one is $C_R$, i.e. large capital or corporate firm of wholesalers & retailers who use to collect the agriculture commodities from the farmers to resale i.e. they act like a wholesalers and/or retailers (multi brand and/or single brand) and the third one is the $C_L$ i.e. local traders of small capital in nature who use to collect agricultural commodities from the farmer and resale them. To show the empirical evidence of price behaviour and riskiness we have taken the data of vegetables as large capital or corporate firms are collecting vegetables throughout the year and little bit of food grains directly from the farmer. It was difficult to get price data of vegetables, which are producing and trading every week and/or month by all the 19 districts in West Bengal. So under vegetables we have used the data of wholesale price of potato of weekly basis of all the (based on data availability) markets of 18 districts (except district Darjeeling as data was not
available) of West Bengal then derived the average of wholesale price according to district and then applied the methodologies. To investigate district wise we have chosen vegetable potato, as potato is a regular diet and used as an input by the food processing firms. The behaviour of distribution of weekly data for wholesale price of potato in different districts in West Bengal we have collected data from January, 2006 to October, 2013 of weekly wholesale price of potato as 100% Foreign Direct Investment (FDI) in cash and carry has been opened science 2006 with automatic route. First, we have collected the available traded time series data on weekly wholesale price of potato for all the wholesale markets in West Bengal. After that, we have segregated according to district wise and derived average to get district wise average wholesale price of all the markets and after that took natural logarithm. We did not able to analyze district Darjeeling, as the data was not sufficient. Regarding the analysis of strategic behaviour of large capital retailer in setting prices of agricultural produce in the retail market, we surveyed the retail market in Kolkata and collected the data relating to the strategies of large capital retail stores. There after we relate that data and strategies with the standard industrial organization theory and analyze accordingly.

The study has been divided into two parts, one, impact of corporate investment on trading of agricultural commodities in the rural wholesale markets and second, strategic behaviour of large capital traders (mainly single and multi-brand retailers) in the retail market in setting prices of agricultural commodities. The main objectives related with our study are given below:

(1) To see the changes in the price behaviour of agricultural commodities in the rural wholesale market by dividing the districts from where the large capital or corporate
traders are collecting agricultural commodities through their collection centre with other districts.

(2) To see the inter district behaviour of wholesale price change of potato from 2006 to 2013.

(3) To see whether small capital traders (SCTs) and large capital traders (LCTs) are forming a cartel or operating independently in the short run and possibility of cartel between them in the long run.

(4) To know the strategic behaviour of large capital traders (mainly single and multi-brand retailers) in setting prices of agricultural commodities in the retail market and possible interaction with the small capital traders (small capital retailers).

(5) To examine whether large capital traders (retailers) will be able to offer agricultural commodities at the lower price or not in the retail market.

3.1 Hypothesis:

Based on the above objectives the hypotheses that are to be tested are set as given below:

**Hypothesis 1:**

The districts from where the corporate traders are collecting the agricultural commodities directly from the farmer through their collection centre, there the rural wholesale market price of the agriculture produce will concentrate with higher likelihood on the moderate price than the other districts.

**Hypothesis 2:**

Participation at the rural wholesale market by the large capital or corporate traders in collecting agricultural commodities raises the likelihood of higher price of agricultural commodities on all the districts.

**Hypothesis 3:**

Large capital or corporate traders (mainly retailers of single brand and multi-brand) will not be able to offer lower price of agriculture commodities in the retail market and in presence of them small capital traders (retailers) must not exit from the market who are engaged with the retailing of agricultural commodities.
3.2 Sources of data under study:

We have divided our study into two parts, one impact in the rural wholesale market of agricultural commodities and second one is the study of strategic behaviour of large capital or corporate firms (including single brand and multi brand retailer) in the retail market. For the study of first one, the data sources are given below:

Regarding field survey I surveyed, Hooghly district and North-24th pargana district, West Bengal, India for getting farm level information and evidences. Our work is based on the in-depth study and information of North 24th pargana district where three corporate firms including one 100% FDI in cash and carry trade are operating. These are Reliance Fresh, Metro cash and carry, Keventer Agro. For empirical analysis, I have collected published data from the sources given below:

1. Data on production, cost of cultivation, farm harvest price and export of potato have collected from the website of Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India (13).

2. We have used the data of wholesale price of potato of weekly basis of all the (based on data availability) markets of 18 districts (except district Darjeeling as data was not available) of West Bengal then derived the average of wholesale price according to district and then applied the methodologies. These data of weekly whole sale price of potato have collected from the website of Agriculture Marketing Information Network-AGMARKNET (14),

(13) Website: http://eands.dacnet.nic.in/
(14) Website: http://agmarknet.nic.in/
3. Data on Production of Principal Crops in West Bengal, Contribution of West Bengal to All India Production of Certain Selected Crops, Yield Rates of Some Selected Crops in West Bengal and India have collected from Economic Review, Statistical Appendix, and Govt. of West Bengal series from 1990 to 2011.


5. Data on Production of Fruits and Vegetables in West Bengal have collected from Directorate of Food Processing Industries and Horticulture, Govt. of West Bengal.

6. For the study of strategic behaviour of large capital or corporate firms (including single brand and multi brand retailer) in the retail market, we surveyed Kolkata retail market and collected data from the large capital or corporate firms and little bit from the small capital retail firms.

3.3 Methodologies:

We have shown the entire study into two ways, one we have shown some theoretical implications based on some standard theoretical base of economic theories secondly with data analysis and empirically based on statistical methods and econometrics. These are discussed below:

The theoretical methodologies are given below:

1. To present the matter theoretically we have used Nash bargaining solution where the Nash bargaining solution between buyer and seller is ½ to each of them over total gain
fulfilling all the axioms given by Nash are, INV (Invariance to equivalent Utility Representations), SYM (Symmetry), IIA (Independence of irrelevant alternatives), PAR (Pareto efficiency), (Nash, 1950a) and the role of risk aversion (Osborne & Rubinstein, Bargaining and Markets, 2005).

2. Theory of industrial organization, which has been used to study the competitive behaviour of small capital and large capital traders. The relevant reference, which we have found fruitful here, is Theory of Industrial Organization (Tirole, 2007).

The other references are (Arrow, 1970), (Pratt, 1964), (Sen A., 2007).

Methodologies for empirical analysis to test some theoretical implications and to check the matters in detail are given below:

3. **Summary statistics**

Under summary statistics, include mean, variance, coefficient of variations, maximum value and minimum value of natural logarithm of average wholesale price of potato for 18 districts individually in West Bengal.

4. **Skewness/Kurtosis tests for Normality**

Skewness/Kurtosis tests implements the test described by (D'Agostino, Belanger, & D'agostino jr, 1990) with the empirical correction developed by (Royston, 1991). Let $g_1$ denote the coefficient of skewness and $b_2$ denote the coefficient of kurtosis as calculated by summarize, and let $n$ denote the sample size. If weights are specified, then $g_1$, $b_2$, and $n$ denote the weighted coefficients of skewness and kurtosis and weighted sample size, respectively. To perform the test of skewness, we compute
\[ Y = g_1 \left\{ \frac{(n+1)(n+3)}{6(n-2)} \right\}^{\frac{1}{2}} \]

\[ \beta_2(g_1) = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} \]

\[ W^2 = -1 + \left[ 2 \beta_2(g_1) - 1 \right]^{\frac{1}{2}} \]

and, \( \alpha = \left( \frac{2}{w^2 - 1} \right)^{\frac{1}{2}} \)

Then the distribution of the test statistic

\[ Z_1 = \frac{1}{\sqrt{\ln w}} \ln \left[ \frac{Y}{\alpha} + \left( \frac{Y}{\alpha} \right)^2 + 1 \right]^{\frac{1}{2}} \]

is approximately standard normal under the null hypothesis that the data are distributed normally.

To perform the test of kurtosis, we compute

\[ E(b_2) = \frac{3(n-1)}{n+1} \]

\[ \text{var}(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \]

\[ X = \frac{\{b_2 - E(b_2)\}}{\sqrt{\text{var}(b_2)}} \]

\[ \sqrt{\beta_1(b_2)} = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \left\{ \frac{6(n+3)(n+5)}{n(n-2)(n-3)} \right\}^{\frac{1}{2}} \]

and, \( A = 6 + \frac{8}{\sqrt{\beta_1(b_2)}} \left[ \left\{ \frac{2}{\beta_1(b_2)} + 1 + \frac{4}{\beta_1(b_2)} \right\}^{\frac{1}{2}} \right] \]

Then the distribution of the test statistic
\[ z^2 = \frac{1}{2} \left[ (1 - \frac{2}{9A}) - \left\{ \frac{1 - \frac{2}{A}}{1 + X \sqrt{\frac{2}{A - 4}}} \right\} \frac{1}{3} \right] \]

is approximately standard normal under the null hypothesis that the data are distributed normally.

D’Agostino, Balanger, and D’Agostino Jr.’s omnibus test of normality uses the statistic

\[ k^2 = z_1^2 + z_2^2, \]

which has approximately a \( \chi^2 \) distribution with 2 degrees of freedom under the null of normality. The following adjustment proposed (Royston, 1991) to the test of normality, which sktest uses by default. Let \( \phi(x) \) denote the cumulative standard normal distribution function for \( x \), and let \( \phi^{-1}(p) \) denote the inverse cumulative standard normal function [that is, \( x = \phi^{-1}[\phi(x)] \)]. Define the following terms:

\[ z_c = -\phi^{-1}\{\exp(-\frac{1}{2}k^2)\} \]

\[ z_t = 0.55n^{0.2} - 0.21 \]

\[ a_1 = (-5 + 3.46 \ln n)\exp(-1.37 \ln n) \]

\[ b_1 = 1 + (0.854 - 0.148 \ln n)\exp(-0.55 \ln n) \]

\[ a_2 = a_1 - \frac{2.13}{1 - 2.37 \ln n} \cdot z_t \]

\[ and, \ b_2 = 2.13/(1 - 2.37 \ln n) + b_1 \]

If \( Zc < -1 \) set \( Z = Zc \); else if \( Zc < Zt \) set \( Z = a_1 + b_1 Zc \); else set \( Z = a_2 + b_2 Zc \). Define \( p = 1 - \phi(z) \). Then \( k^2 = -2 \ln p \) is approximately distributed \( \chi^2 \) with 2 degrees of freedom.

We have performed Skewness/Kurtosis tests for Normality of natural logarithm of average wholesale price of potato for 18 districts individually in West Bengal.
5. Test for unit root in time series

Dickey-Fuller (DF) test (Dickey & Fuller, 1979), Phillips-Perron test (Phillips & Perron, 1988) for unit root to test stationary in time series data. The ADF test adjusts the DF test to take care of possible serial correlation in the error terms by adding the lagged difference terms of the regressand. Phillips and Perron use nonparametric statistical methods to take care of the serial correlation in the error terms without adding difference terms. Therefore, to test stationary in time series data if ADF test model becomes insignificant then Phillips-Perron test for unit root will be appropriate.

a. Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) test for unit root

In the OLS estimation of an AR (1) process with Gaussian errors

\[ Y_t = \rho Y_{t-1} + \varepsilon_t, \]

Where \( \varepsilon_t \) are independently and identically distributed as \( N(0; \sigma^2) \) and \( y_0 = 0 \), the OLS estimate (based on an \( n \)-observation time series) of the autocorrelation parameter \( \rho \) is given by

\[ \hat{\rho}_n = \frac{\sum_{t=1}^{n} Y_{t-1} Y_t}{\sum_{t=1}^{n} Y_t^2} \]

If \(|\rho| < 1\), then

\[ \sqrt{n}(\hat{\rho}_n - \rho) \rightarrow N(0, \rho^2) \]
If this result were valid when $\rho = 1$, the resulting distribution would have a variance of zero.

When $\rho = 1$, the OLS estimate $\hat{\rho}$ still converges in probability to one, though we need to find a suitable non degenerate distribution so that we can perform hypothesis tests of

$$H_0 : \rho = 1 \quad (\text{Hamilto, 1994})$$

To compute the test statistics, we fit the augmented Dickey–Fuller regression

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \delta t + \sum_{j=1}^{k} \xi_j \Delta Y_{t-j} + e_t$$

Via OLS where, depending on the options specified, the constant term $\alpha$ or time trend $\delta t$ is omitted and $k$ is the number of lags specified in the lags option. The test statistic for

$$H_0 = \beta = 0$$

is

$$Z_i = \frac{\hat{\beta}}{\sigma_{\beta}}$$

Where $\sigma_{\beta}$ is the standard error of $\hat{\beta}$.

The critical values included in the output are linearly interpolated from the table of values that appears in (Fuller, 1996), and the MacKinnon approximate p-values use the regression surface published in (Mackinnon, 1994).

b. Phillips-Perron test for unit root

The Phillips–Perron test (Phillips & Perron, 1988) involves fitting the regression

$$Y_i = \alpha + \rho Y_{i-1} + \epsilon_i$$
Where we may exclude the constant or include a trend term. There are two statistics \( Z_{\rho} \) and \( Z_\tau \), calculated as

\[
Z_{\rho} = n(\hat{\rho}_n - 1) - \frac{1}{2} n^2 \frac{\hat{\sigma}^2}{s_n^2} (\hat{\lambda}_n - \gamma o, n)
\]

\[
Z_\tau = \sqrt{\frac{\hat{\gamma} o, n}{\hat{\lambda}_n}} \frac{\hat{\rho}_n - 1}{\hat{\sigma}} - \frac{1}{2} (\hat{\lambda}_n - \gamma o, n) \frac{1}{\hat{\lambda}_n} \frac{n \hat{\sigma}}{s_n}
\]

\[
\hat{\gamma}_{j,n} = \frac{1}{n} \sum_{i=j+1}^{n} u_i \hat{u}_{i-j}
\]

\[
\hat{\lambda}_n = \gamma o, n + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{q + 1}\right) \hat{\gamma}_{j,n}
\]

\[
s_n^2 = \frac{1}{n - k} \sum_{i=1}^{n} u_i^2
\]

Where \( u_i \) is the OLS residual, \( k \) is the number of covariates in the regression, \( q \) is the number of Newey–West lags to use in calculating \( \hat{\lambda}_n \), and \( \hat{\sigma} \) is the OLS standard error of \( \hat{\rho} \). The critical values, which have the same distribution as the Dickey–Fuller statistic (Dickey & Fuller, 1979) included in the output, are linearly interpolated from the table of values that appear in (Fuller, 1996), and the MacKinnon approximate p-values use the regression surface published in (Mackinnon, 1994).

We have tested unit root based on above two methods of average wholesale price of potato for 18 districts individually in West Bengal.
6. Nonparametric density estimation

Nonparametric density estimation (Pagan & Ullah, 1999) (Vailar & Goergen, 2009) is most frequently used for exploratory data analysis. It can also be useful for more sophisticated data analysis. There are different types of Kernel density functions. A kernel density estimate is formed by summing the weighted values calculated with the kernel function \( K \), as in

\[
\hat{f}_K = \frac{1}{qh} \sum_{i=1}^{n} w_i K \left( \frac{x - X_i}{h} \right)
\]

Where \( q = \sum_i W_i \). If weights are frequency weights or analytic weights, and \( q = 1 \) if weights are importance weights. Analytic weights are rescaled so that \( \sum_i W_i = n \). If weights are not used, then \( w_i = 1 \), for \( i = 1, \ldots, n \). Kernel density includes seven different kernel functions. The Epanechnikov is the default function if no other kernel is specified and is the most efficient in minimizing the mean integrated squared error.
### Table 3.1

Name of some kernel functions and their formulas:

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Formula</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biweight</td>
<td>$K[z] = \begin{cases} \frac{15}{16}(1-z^2)^2 &amp; \text{If }</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Cosine</td>
<td>$K[z] = \begin{cases} 1 + \cos(2\pi z) &amp; \text{If }</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Epanechnikov</td>
<td>$K[z] = \begin{cases} \frac{3}{4}(1 - \frac{1}{5}z^2) &amp; \text{If }</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Epan2</td>
<td>$K[z] = \begin{cases} \frac{3}{4}(1 - z^2) &amp; \text{If }</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$K[z] = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parzen</td>
<td>$K[z] = \begin{cases} 4/3 - 8z^2 + 8</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If $1/2 &lt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Rectangular</td>
<td>$K[z] = \begin{cases} 1 &amp; \text{If }</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
<tr>
<td>Triangular</td>
<td>$K[z] = \begin{cases} 1 -</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Otherwise</td>
</tr>
</tbody>
</table>
From the definitions given in the table, we can see that the choice of \( h \) will drive how many values are included in estimating the density at each point. This value is called the window width or bandwidth. If the window width is not specified, it is determined as

\[
m = \min \left( \sqrt{\text{Variance}_x}, \text{Interquartile range}_x/1.349 \right)
\]

\[
h = \frac{0.9m}{n^{1/5}}
\]

Where \( x \) is the variable for which we wish to estimate the kernel and \( n \) is the number of observations. The choice of kernel is not as important as the choice of bandwidth.

Here I shall use Gaussian Kernel and Epanechnikov Kernel density functions. The kernel can work on the variable on different length intervals, i.e. the weighting effect around \( x \) is more or less great and depends on the construction of each kernel. The kernels are represented on the graph below and these can be observed that the length of the compact support on which each kernel exits. Except the Gaussian kernel (support is \([-4, 4]\)), these can be noted that the supports of these kernels are \([-1, 1]\): the window of these kernels is written: \( 1_{[-1,1]}(x) \).

Figure-3.1 Epanechnikov kernel and Gaussian kernel
Although all the examples we included had densities of less than 1, the density may exceed 1. The probability density $f(x)$ of a continuous variable, $x$, has the units and dimensions of the reciprocal of $x$. If $x$ is measured in meters, $f(x)$ has units $1$=meter. Thus, the density is not measured on a probability scale, so it is possible for $f(x)$ to exceed 1.

We have derived Gaussian Kernel and Epanechnikov Kernel density functions of natural logarithm of average wholesale price of potato for 18 districts individually in West Bengal.