CHAPTER 5

STUDY DESIGN, METHODOLOGY AND DATA

This study aims to measure the change in productivity at the national level and the division level for Bangladesh in an attempt to identify the sources of growth of the economy. While there can be several approaches towards measuring change in the level of productivity for an economy, the selection of method depends largely on availability of relevant data. This study uses non-parametric approach to growth accounting to measure the Total Factor Productivity Growth (TFPG) at the aggregate national level and for the major sectors. One specialty of this study is that, it also takes into account the task of calculating productivity growth for six divisions of Bangladesh in the form of a regional study. Since relevant price information is not available at the division level in Bangladesh, a distance function approach for measuring a Malmquist index for each of the six divisions of Bangladesh is undertaken in this study. This chapter presents the models for applying growth accounting technique in productivity analysis at the national aggregate level and for the major sectors. The model for the division level analysis, however, stems from a distance function approach.

The models used in this study for calculating total factor productivity growth at the aggregate national level, for the major economic sectors and for the administrative divisions require only two reference points in time. But since it is difficult to find such points in time for which data on all of the key variables are available from reliable sources, some of the data have to be estimated. Besides, in case of inconsistency of definition from different surveys of the same variable, some adjustments have to be made. This chapter describes the limitations of the data set to be used in this study and the way it determined some aspects of this study, defines the variables one by one mentioning specific limitations of some of the key variables, also describes measures taken to adjust the data to make them compliant with this study.

5.1 Measuring Total Factor Productivity Growth at National Level

5.1.1 Selection of Model for Analysis at the National Level

While measuring change in Total Factor Productivity (TFP), either parametric approach or the non-parametric approach towards growth accounting can be undertaken. This study considers the non-parametric approach and measures TFPG at the national level not only for the whole economy at the aggregate level but also for the major sectors of the economy. The study considers two sectors of the economy - the agricultural sector, and the non-agricultural sector. The distance function approach cannot be used for the analysis at the national level, since this approach requires panel data.

5.1.2 Model for Growth Accounting

The model for non-parametric approach to growth accounting is as follows. Equation 5.1 will be applied in the analysis at the national aggregate level and in the agricultural sector analysis. The same model can be used for the analysis at the division level too, with the assumption that the factor income shares for the whole economy is applicable for the divisions too.

Equation 5.2, where expansion of land is not considered as a factor effecting growth, will be applied in the non-agricultural sector analysis. Here, it is necessarily assumed that, growth rate of land factor in non-agricultural sector is zero. In fact, acquisition of land is considered as an investment in the national accounts and this specific investment in turn is included in the capital stock.

$$TFPG^i = \frac{Y^i_2 - Y^i_1}{\Delta t \, . \, Y^i_1} - s^i_K . \frac{K^i_2 - K^i_1}{\Delta t \, . \, K^i_1} - s^i_N . \frac{N^i_2 - N^i_1}{\Delta t \, . N^i_1} - s^i_L . \frac{L^i_2 - L^i_1}{\Delta t \, . \, L^i_1} \qquad (5.1)$$

$$TFPG^i = \frac{Y^i_2 - Y^i_1}{\Delta t \, . \, Y^i_1} - s^i_K . \frac{K^i_2 - K^i_2}{\Delta t \, . \, K^i_1} - s^i_N . \frac{N^i_2 - N^i_1}{\Delta t \, . \, N^i_1} \qquad (5.2)$$

Here, $TFPG^i$ is the total factor productivity growth of unit i between two points in time, Y^i_t is the gross national product for unit i in year t in million taka, K^i_t is the stock of capital for unit i in year t in million taka, N^i_t is the size of labor force for unit i in year t thousand persons, L^i_t is the amount of cultivable land for unit i in year t in acres. s^i_K is the share of income going to capital, s^i_N is the share of income going to labor, and s^i_L is rental share of income going to land owners for unit i during the interval. i is 0 for the analysis at the national aggregate level, 1 for the analysis of agricultural sector, 2 for the analysis of non-agricultural sector, and n for the analysis of the n-th division. j is K for the input capital, N for the input labor, and L for the input land. t is 1 for the initial year and 2 for the final year of two points in time, whereas Δt is the time difference in years between period 1 and period 2

The equations 5.1 and 5.2, however, are the discrete approximations of the models 4.25 and 4.27 (see pages 26-27 of chapter 4) respectively. The study also assumes a constant returns to scale (CRS) technology, so that

$$s^i_N + s^i_K + s^i_L = 1 \tag{5.3}$$

The income shares used are the average of the income shares of the two points in time. However, as this study obtains this specific information from Social Accounting Matrix (SAM) constructed for Bangladesh economy for the financial year 1993-94 and assumes it to be constant over the entire period of study, there will be only one set of values anyway. If it is assumed that the same shares of factor income are applicable for the divisions as well, the same model can be applied for measuring TFPG for the division.

5.2 Measuring Productivity Growth at Division Level

5.2.1 Selection of Model for Analysis at the Division Level

Productivity assessment at the division level is a useful tool for identifying parts of the country exhibiting sluggishness in development. But such assessment faces the challenge of limitations of data, specially price data. In least developed countries like Bangladesh, reliable and detailed information on prices at the administrative division level is not usually available. In such cases, the distance function method may be a powerful tool for productivity analysis, since it does not need price information for its calculations. This study, thus, uses Data Envelopment Analysis (DEA) to construct a Malmquist Productivity Index (MPI) for every division of Bangladesh.

Here it should be noted that, DEA can be either input-orientated or output-orientated. In the input-orientated case, the DEA method defines the frontier by seeking the maximum possible proportional reduction in input usage, with output levels held constant, for each division. On the other hand, in the output-orientated case, the DEA method seeks the maximum proportional increase in output production, with input levels held fixed. The two measures provide the same technical efficiency scores when a CRS technology applies, but are unequal when variable returns to scale (VRS) is assumed. As a CRS technology is assumed in this study, the choice of orientation should not be a serious issue of concern. However, this study considers output distance functions with a rationale that, a division will in general be looking for maximizing its output from a given set of inputs, rather than the converse.

It is obvious that, the returns to scale properties of the technology are very important in productivity measurement. A CRS technology is used in this study for two reasons. Firstly, given that this study is using aggregate division-level data, it does not appear to be sensible to consider a VRS technology, as it may not be possible for a division to achieve scale economies (Coelli and Rao, 2003). The use of a VRS technology is understandable in a situation when the summary data is expressed on an average per farm basis, because one can then discuss the scale economies of the average farm. But when one is dealing with aggregate data, the use of a CRS technology seems to be the only sensible option (Coelli and Rao, 2003). Secondly, for both firm-level and aggregate data, a Malmquist index may not correctly measure TFP changes when VRS is assumed for the technology (Grifell-Tatjé and Lovell, 1995). Hence it is important that CRS be imposed upon any technology that is used to estimate distance functions for the calculation of a Malmquist index. Otherwise the resulting measures may not properly reflect the TFP gains or losses resulting from scale effects.

5.2.2 Model for Distance Functions Approach to Productivity Measurement

The use of distance functions to construct productivity index allowed this study to describe a multi-input multi-output production technology without the need to specify a behavioral objective such as cost minimization or profit maximization. The output distance function is defined on the output set *P(x)* as

$$d\,(x,y) = \min\{\delta : (y\,/\,\delta)\ \epsilon\ P(x)\}$$

The distance function $d\,(x,y)$ will take a value which is less than or equal to one if the output vector y is an element of the feasible production set $P(x)$. Furthermore, the distance function will take a value of 1 if y is located on the outer boundary of the feasible production set, and will take a value greater than 1 if y is located outside the feasible production set.

Following Färe et al (1994), the output-orientated Malmquist index for every division between period s (the base period) and period t (the recent period) is given by

$$M\,(y_s, x_s, y_t, x_t) = \left[\frac{d^s\,(y_t, x_t)}{d^s\,(y_s, x_s)} \times \frac{d^t\,(y_t, x_t)}{d^t\,(y_s, x_s)}\right]^{1/2} \qquad (5.4)$$

where $d^s\,(y_t, x_t)$ is the distance from the period t observation to the period s technology, $d^s\,(y_s, x_s)$ is the distance from the period s observation to the period s technology, $d^t\,(y_s, x_s)$ is the distance from the period s observation to the period t technology, and $d^t\,(y_t, x_t)$ is the distance from the period t observation to the period t technology.

A value of M greater than 1 indicates positive productivity growth from period s to period t while a value less than 1 indicates a productivity decline. Equation 5.4 is the geometric mean of two TFP indices, the first is evaluated with respect to period s technology and the second with respect to period t technology.

Equation 5.4 can be expressed allowing for a decomposition as equation 5.5.

$$M(y_s, x_s, y_t, x_t) = \frac{d^t(y_t, x_t)}{d_o^s(y_s, x_s)} \left[\frac{d^s(y_t, x_t)}{d^t(y_t, x_t)} \times \frac{d^s(y_s, x_s)}{d^t(y_s, x_s)} \right]^{1/2} \quad (5.5)$$

Here, the first term of the product at the right hand side measures the change in the output-orientated measure of Farrell technical efficiency between periods s and t. That is, the efficiency change is equivalent to the ratio of the technical efficiency in period t to the technical efficiency in period s. The remaining part of the index in equation 5.5 is a measure of technical change. It is the geometric mean of the shift in technology between the two periods, evaluated at x_t and also at x_s.

With available panel data, one can calculate the required distance measures for the MPI using DEA-like linear programs as described below. For the i-th division, one must calculate four distance functions to measure the TFP change between two periods, s and t. This requires the solving of four linear programming (LP) problems with assumption of CRS. The required LPs are shown below. Each LP maximizes technical efficiency (φ), subject to the constraint that output of unit i multiplied by technical efficiency term (φ) will be less than or equal to the weighted sum of outputs, and also subject to the constraint that input for unit i will be greater than or equal to the weighted sum of input factors.

$$\begin{aligned} &\text{maximize}_{\varphi,\lambda}\, \varphi && = [d_o^t(y_t, x_t)]^{-1} \\ &\text{subject to} && -\varphi y_{i\,t} + Y_t\lambda \geq 0 \\ & && x_{i\,t} - X_t\lambda \geq 0 \\ & && \lambda \geq 0 \end{aligned} \quad (5.6)$$

$$\begin{aligned} &\text{maximize}_{\varphi,\lambda}\, \varphi && = [d_o^s(y_s, x_s)]^{-1} \\ &\text{subject to} && -\varphi y_{i\,s} + Y_s\lambda \geq 0 \\ & && x_{i\,s} - X_s\lambda \geq 0 \\ & && \lambda \geq 0 \end{aligned} \quad (5.7)$$

$$\begin{aligned} &\text{maximize}_{\varphi,\lambda}\ \varphi && = [d_o^{\,t}(y_s,\ x_s)]^{-1} \\ &\text{subject to} && -\varphi y_{i\,s} + Y_t\lambda \geq 0 \\ & && x_{i\,s} - X_t\lambda \geq 0 \\ & && \lambda \geq 0 \end{aligned} \tag{5.8}$$

$$\begin{aligned} &\text{maximize}_{\varphi,\lambda}\ \varphi && = [d_o^{\,s}(y_t,\ x_t)]^{-1} \\ &\text{subject to} && -\varphi y_{i\,t} + Y_s\lambda \geq 0 \\ & && x_{i\,t} - X_s\lambda \geq 0 \\ & && \lambda \geq 0 \end{aligned} \tag{5.9}$$

Here, $y_{i\,j}$ is a $M\times1$ vector of output quantities for the i-th division at time j, $x_{i\,j}$ is a $K\times1$ vector of input quantities for the i-th division at time j, Y_j is a $M\times6$ matrix of output quantities for all 6 divisions at time j, X_j is a $K\times6$ matrix of input quantities for all *6* divisions at time j, λ is a 6×1 vector of weights, and φ is a scalar. For 6 divisions, i is from division1 to division6; for 2 periods j is s and t; for 1 output, M is GDP; and for 3 inputs, K is cultivable land area, size of labor force and stock of capital.

It can be noted here that in LPs 5.8 and 5.9, where production points are compared to technologies from different time periods, the φ parameter need not be greater than or equal to one, as it must be when calculating standard output-orientated technical efficiencies. The data point could lie above the production frontier. This will most likely occur in LP 5.9 where a production point from period t is compared to technology in an earlier period, s. If technical progress has occurred, then a value of $\varphi<1$ is possible. It can also be noted that it could also possibly occur in LP 5.8 if technical regress has occurred, but this is less likely.

The calculation of productivity for all the divisions by solving the LPs is a huge task. Fortunately, special software available for such analysis makes the task much easier. This study uses Warwick DEA Software for the division level analysis. The analyses using the software is shown in brief in Appendix F.

5.3 Sources of Productivity Growth

Theoretically, there are many factors that can influence TFPG for any country such as higher level of education of the workers, optimal allocation of resources, trade liberalization by the government, increase of foreign direct investment, extended research and development, etc. This study investigates the effects of human capital in terms of level of education of the workers and health condition of the population, growth of export and import volume, degree of openness, growth of infrastructure in terms of transport, communication, education, health and energy. However, since there are very few numbers of observations, the indicators of sources of productivity growth cannot be regressed against the computed value of TFPG, and only a qualitative analysis is conducted for the purpose of this study.

5.4 Limitations of the Analyses due to Limitations of the Data Set

5.4.1 Major Sectors under Study

Number of sectors under study is limited due to characteristic of the available capital stock data. This study uses capital stock data from Alam et al. (2005), which lumps together both manufacturing and service sector capitals under one heading. Hence, this study has to consider these two sectors under one heading of 'non-agricultural sector' and focus on the economy divided into two major sectors – agricultural sector and non-agricultural sector.

5.4.2 Period of the Study

Period of this study is limited due to unavailability of a long series of constant price series for Gross Domestic Product (GDP). Although data on GDP is a comprehensive information available from Bangladesh Bureau of Statistics (BBS), the data is available in three constant price series':

i) the constant 1972-73 price series available from 1972-73 to 1984-85,
ii) the constant 1984-85 price series available from 1984-85 to 1999-00,
iii) the constant 1995-96 price series available from 1995-96 to 2003-04.

As the longest series of GDP data at constant prices is available at constant 1984-85 prices, this study considers the corresponding time range, i.e. from 1984-85 to 1999-2000, as the period for analysis. It is possible, however, to join the second and the third series using data of the overlapping years. But this approach is constrained by the unavailability of capital stock data after 1999-2000.

5.4.3 Sub-Periods of the Study

Although data on only two points in time is enough to measure Total Factor Productivity Growth (TFPG) with the models proposed in this study, data for all the variables should be available at the same point in time. Unfortunately, there are very few points in time – for which data on all the variables is available simultaneously. This will be evident from Table 5.1 and the discussion afterwards.

As can be seen in Table 5.1, data on all key variables are available for the years 1984-85 and 1999-2000, although some adjustments are required for the national labor force in the year 1984-85.

No labor force survey (LFS) was conducted in the financial year 1989-90. But one LFS is available for the calendar year 1989. If LFS data for the calendar year 1989 is assumed applicable for the financial year 1989-90, then 1989-90 can be a common reference point in time for which data on all other key variables are available.

LFS was conducted for both of the years 1990-91 and 1995-95, but data on non-agricultural capital stock is missing for both of these years too. However, capital stock for these years can be estimated by adding previous year's national investment (deflated by wholesale price indices at 1984-85 prices) with previous year's capital stock data (net of current year investment at 1984-85 price) and allowing for depreciation.