Dual Measure of Technological Progress, Returns to Scale and Capacity Utilization

Hiung-Joon Kim

Department of Economics, College of Social Science,
Soochunhyang University, Asan 336-745, Korea

Abstract

The total-factor-productivity growth (TFPG) is interpreted as the rate of change in an index of outputs not explained by the change in an index of inputs or the rate of shift of an production function over time. Typically the rate of growth in an input aggregate is computed as a weighted average of the growth rates in the individual inputs using input cost-shares as weights. This measure of TFPG can be decomposed into two components: the rate of technical progress (RTP) and a returns to scale effect (RTS). The former captures the shift in the production function over time due to technical progress; the latter, variations in productivity along a specific production function due to non-constant returns to scale where relevant. This decomposition assumes competitive profit-maximizing behavior and full static optimization. In other words, all inputs are treated as variable and prices of individual inputs are assumed to be equal to their respective values of marginal products at the observed quantities.

However, a firm or industry can not achieve full economic efficiency in production process all the time and measured annual rates of TFPG, then, should be distorted. Similarly, when sub-optimal quantities of variable inputs are used, actual cost shares are inappropriate weights and distort the returns to scale effects. In general, most studies of total factor productivity growth ignore the effects of technical and/or allocative inefficiencies and considers all inputs to be variable and does not separate effects of changes in allocative
inefficiency from changes in capacity utilization.

The purpose of this paper is to take explicit account of the effects of variations in allocative efficiency and capacity utilization level on the rate of TFPG using the dual approach. We treat capital as a quasi-fixed input and examine how changes in capacity utilization level over time affect the conventional measure of productivity. This paper is organized as follows. Section 2 compares dual and primal approaches. Section 3 presents the standard approach which assumes competitive profit-maximizing behavior and full static optimization. Section 4 provides a theoretical exposition of the decomposition of the rate of TFPG into the different components using dual cost function. Section 5 is the conclusion.

2. Dual approach versus primal approach

The sources of output growth were identified and their contributions were measured from the production model. Same analysis can be implemented using the neoclassical duality theory. Duality theory states that all the economically relevant informations on a production technology can be recovered from a dual cost function. This duality between production function and cost function also hold along the respective frontier functions. Therefore, a dual frontier cost function can also be estimated to analyze the productivity growth and decomposition of the TFPG.

The dual cost approach is different from primal one in some respects. First of all, the dual approach enable one to analyze the effect of allocative inefficiency. In most studies about productivity, technical inefficiency is measured from the production model and it is identified as an important source of bias of the TFPG. However, there was no discussion of inefficiency resulting from non-optimal input mix, i.e., allocative inefficiency, since the primal model based only on the production relationship. In the primal model, all the variable inputs are assumed to be allocatively efficient. In other words, every variable inputs are employed in such a way that price of each input is equal to its marginal revenue product. In this paper, that assumption will be relaxed to include the impact of allocative efficiency on the TFPG.

Allocative efficiency is also directly related to the TFPG because the measure uses cost share as the weights in the index of inputs and allocative efficiency affects cost. For example, as a firm becomes more efficient allocatively, the production cost decrease and the weights of the measure would increase and thus raise the value of the measure, ceteris paribus.

Another different aspect of the dual approach is that the measure of technical efficiency from cost frontier is an ‘input-saving’ measure. Technical inefficiency is measured by finding the minimum possible cost to produce a given level of output, providing that a firm is allocatively efficient. Comparing actual cost and the minimum cost yield the measure of technical efficiency. Thus this measure indicates how much production cost can be reduced for the given output level. In the primal approach, technical efficiency was measured as a ratio of actual output to the frontier output representing the maximum output producible with a given input bundle. Thus it was an ‘output-augmenting’ measure of technical efficiency.

The input-saving and output-augmenting measures share the same value if production technology exhibits constant returns to scale. Otherwise, there exists a discrepancy between them, which depends on degree of scale economies. Furthermore, a full efficiency of the primal measure does not necessarily imply that the dual measure also shows full efficiency and vice versa. Suppose that an observation is proved to

---

1) Bauer (1990) tries to decompose multifactor productivity growth into technical progress, scale effect, and effects of changes in allocative and technical efficiencies. He is not able, however, to separate the two kinds of efficiencies in his empirical application and ignores the effects of capacity utilization.
be technically efficient in the primal model. It implies that output larger than the observed one can not be produced with the given input quantities. But it does not necessarily mean that the observed output can not be produced with smaller input quantities. In other words, input slacks may be present in production process. Therefore, it would be desirable to implement both primal and dual models to analyze efficiencies in production.

3. Dual cost framework – a standard approach

Consider a dual cost function for n-input, single-output production technologies:

\[ C(t) = h(w_1(t),..., w_n(t), y(t), t) \quad (1) \]

where \( y \) is output producible from a vector of \( n \) inputs \( x \), \( w_i \) is the price of \( i \)th input and \( t \) is an index of time capturing shifts in the cost function over time due to changes in technology.

Logarithmically differentiating (1) with respect to \( t \) yields:

\[
\frac{d \ln C(t)}{dt} = \sum_{i=1}^{n} \frac{\partial \ln h(t)}{\partial \ln w_i(t)} \frac{d \ln w_i(t)}{dt} + \frac{\partial \ln h(t)}{\partial \ln y(t)} \frac{d \ln y(t)}{dt} + \frac{\partial \ln h(t)}{\partial \ln t} \frac{d \ln t}{dt} \quad (2)
\]

Employing Shephard’s lemma (\( \frac{\partial C}{\partial w_i} = x_i \)),

\[
\frac{\dot{C}}{C} = \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{w}_i}{w_i} + \epsilon_{\sigma_y} \frac{\dot{y}}{y} + \frac{\dot{h}}{h} \quad (3)
\]

where \( \epsilon_{\sigma_y} \) is the cost elasticity of output and \( \frac{\dot{h}}{h} \) denotes the rate of shift in the cost function representing technological change. The rate of change in actual cost, \( \frac{\dot{C}}{C} \), is decomposed into three components: changes in aggregate input price, output and technology.

Ohta(1974) showed that the dual measure of technical change \( \frac{\dot{h}}{h} \) is related to the primal measure \( \frac{\dot{f}}{f} \).

Totally differentiating the cost identity, \( C = \sum_{i=1}^{n} w_i x_i \), and dividing it by \( dt \) and \( C \) yields:

\[
\frac{\dot{C}}{C} = \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{w}_i}{w_i} + \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{x}_i}{x_i} \quad (4)
\]

Or,

\[
\frac{\dot{C}}{C} - \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{w}_i}{w_i} = \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{x}_i}{x_i} \quad (5)
\]

Substituting (5) into (3) yields:

\[
-\frac{\dot{h}}{h} = \epsilon_{\sigma_y} \frac{\dot{y}}{y} - \sum_{i=1}^{n} \frac{w_i x_i}{C} \frac{\dot{x}_i}{x_i} \quad (6)
\]
or,

$$-\varepsilon_{cy}^{-1} \frac{h}{h} = \frac{\dot{y}}{y} - \varepsilon_{cy}^{-1} \sum_{i=1}^{n} \frac{w_i}{C} \frac{\dot{x}_i}{x_i}$$

(7)

Under the assumption of perfectly competitive input and output market, \( w_i = p \left[ -\frac{\partial f}{\partial x_i} \right] \) and \( p = p \left[ -\frac{\partial C}{\partial y} \right] \).

Therefore, (7) can be rewritten as:

$$-\varepsilon_{cy}^{-1} \frac{h}{h} = \frac{\dot{y}}{y} - \sum_{i=1}^{n} \varepsilon_{fi} \frac{\dot{x}_i}{x_i}$$

(8)

where \( \varepsilon_{cy} = \sum_{i=1}^{n} \varepsilon_{fi} \) is the dual measure of returns to scale.

(8) states that the rate of technological progress can also be obtained from dual cost function by taking the \( \varepsilon_{cy} \) and rate of change in cost.

4. Dual measure in the short-run equilibrium model

A short-run total cost function can be written as:

$$C(t) = g(w_1(t), \ldots, w_{n-1}(t), k(t), y(t), t) + w_k k(t)$$

(9)

Differentiating (9) with respect to time and rewriting in a growth rate form yields:

$$\frac{\dot{C}}{C} = \sum_{i=1}^{n} \frac{w_i}{C} \frac{\dot{w}_i}{w_i} + \varepsilon_{gy} \frac{g}{C} \frac{\dot{y}}{y} + \frac{-z_k k}{C k} + \frac{\dot{g}}{g} \frac{g}{C} \frac{k}{k} + \frac{w_k}{C} \frac{\dot{k}}{k}$$

(10)

where \( v_i \) denotes ith variable input, \( \varepsilon_{gy} = \frac{\partial g}{\partial y} \), and \( z_k = -\frac{\partial g}{\partial k} \). \( \varepsilon_{gy} \) measures short-run returns to scale and \( -z_k \) is the shadow price of capital.

Since \( \frac{\partial g}{\partial y} = \frac{\partial C}{\partial y} \), (10) can be rewritten as:

$$\frac{\dot{g}}{g} \frac{g}{C} = \frac{\dot{C}}{C} = \sum_{i=1}^{n} \frac{w_i}{C} \frac{\dot{w}_i}{w_i} - \frac{w_k}{C} \frac{\dot{k}}{k} - \varepsilon_{cy} \frac{\dot{y}}{y} + \frac{(-z_k + w_k) k}{C k}$$

(11)

This short-run cost diminution can be related to the primal measure of technological progress by differentiating the cost identity as shown in previous section:

$$\frac{\dot{C}}{C} = \sum_{i=1}^{n} \frac{w_i}{C} \frac{\dot{w}_i}{w_i} + \sum_{i=1}^{n} \frac{w_i v_i}{C} \frac{\dot{v}_i}{v_i} + \frac{w_k}{C} \frac{\dot{k}}{k} + \frac{w_k}{C} \frac{\dot{k}}{k}$$

(12)

Incorporating (12) into (11) yields:

$$\frac{\dot{g}}{g} \frac{g}{C} = \varepsilon_{cy} \frac{\dot{y}}{y} - \sum_{i=1}^{n} \frac{w_i v_i v_i}{C} \frac{\dot{v}_i}{v_i} - \frac{w_k}{C} \frac{\dot{k}}{k} + \frac{(-z_k + w_k) k}{C k} = \varepsilon_{cy} \frac{\dot{y}}{y} - \sum_{i=1}^{n} \frac{w_i v_i v_i}{C} \frac{\dot{v}_i}{v_i} - \frac{z_k}{C}$$

(13)
or,
\[-\varepsilon_y^{-1} \frac{\dot{g}}{g} \frac{C}{k} = \frac{\dot{y}}{y} - \varepsilon_y^{-1} \frac{\sum_{i=1}^{n} w_i v_i}{C} \frac{\dot{v}_i}{v_i} - \varepsilon_y^{-1} \frac{z_k}{C} \frac{k}{k} \]  \hspace{1cm} (14)

Again, the second term in the RHS is the sum of input growth rate weighted by their marginal productivities since
\[\varepsilon_y^{-1} \frac{w_i v_i}{C} = \frac{C}{p_y} \frac{\partial f_i}{\partial v_i} = \varepsilon_{h_i} \quad \text{for all } v_i, i = 1, 2, 3, \ldots, n-1 \]  \hspace{1cm} (15)

and \(z_k\) is the shadow price of the capital input which represents the reduction in the variable cost due to the increase in the stock of capital. This would be equal to the market price of capital in the long-run equilibrium.

From the short-run isocost function, we can derive the marginal product of capital:
\[\hat{g} = g(w, y, k)\]
\[0 = \frac{\partial g}{\partial y} dy + \frac{\partial g}{\partial k} dk\]  \hspace{1cm} (16)
\[z_k = -\frac{\partial g}{\partial y} \left( \frac{dy}{dk} \right)\]

Substituting (16) into the third term in the RHS of (14) yields
\[\varepsilon_y^{-1} \frac{z_k}{C} \frac{k}{k} = \frac{\dot{y}}{y} \frac{k}{k} = f_k \frac{k}{k} \]  \hspace{1cm} (17)

Therefore, (14) can be rewritten as:
\[-\varepsilon_y^{-1} \frac{\dot{g}}{g} \frac{C}{k} = \frac{\dot{y}}{y} - \sum_{i=1}^{n} \varepsilon_h \frac{\dot{v}_i}{v_i} - \varepsilon_{h_k} \frac{k}{k} \]  \hspace{1cm} (18)

which represents the pure rate of technical progress.

In the standard approach, it was shown that \(\varepsilon_y\) is the measure of returns to scale. However, \(\varepsilon_y\) no longer represents returns to scale if there is fixed input(s) since fixed input cannot be employed optimally in the short run. Instead it is interpreted as short-run returns to scale because it consists of two components, returns to scale and short-run disequilibrium component resulting from the fixity of the capital input. It is, therefore, necessary to separate the effect of non-optimality of fixed input from that of scale economies.

Morrison (1985) decomposed the short-run returns to scale into two components— the inverse of long-run returns to scale \((\rho)\) and disequilibrium component— by expressing long-run cost elasticity of output in terms of short-run cost elasticities:
\[\varepsilon_y = \frac{\partial C}{\partial y} \frac{y}{C} + \frac{\partial C}{\partial k} \frac{y}{C} \frac{dk}{dy} = \varepsilon_y + \varepsilon_{yk} \rho \]  \hspace{1cm} (19)

Or,
\[\varepsilon_y = \rho (1 - \varepsilon_{yk}) \]  \hspace{1cm} (20)
where \( \varepsilon_{Cy} \) is long-run cost elasticity of output, \( \rho = \left( \frac{1}{\frac{d\varepsilon}{dk} \left( \frac{k}{y} \right)} \right) \) is the inverse of long-run returns to scale. \( \varepsilon_{Cy} \) represents the percentage change in the long-run total cost required to increase output by one percent. If it is less than one, production exhibits long-run returns to scale. Therefore, \( \varepsilon_{Cy} \) is dual measure of long-run returns to scale and equals to \( \rho \).

The term, \( (1 - \varepsilon_{Cy}) \), measures the impact of short-run fixity of the capital input. To see why \( (1 - \varepsilon_{Cy}) \) measures the impact of short-run disequilibrium,

\[
(1 - \varepsilon_{Cy}) = 1 - \frac{(-z_k + w_k)k}{C} = \frac{C - w_k k + z_k k}{C} = \frac{C^*}{C}
\]

(21)

where \( C^* = g(w, y, k, t) + z_k k \), is the shadow cost.

\( C^* \) is different from \( C \) because the shadow valuation of capital deviates from the market price in the short-run. Therefore, \( C^*/C \) measures the rate of capacity utilization (cu) and \( (1 - \varepsilon_{Cy}) \) measures the impact of short-run disequilibrium.

The short-run cost elasticity of output can be expressed as

\[
\varepsilon_{Cy} = \frac{\partial C}{\partial y} \cdot \frac{y}{C^*} \cdot \frac{C^*}{C} = \varepsilon_{Cy} \cdot cu.
\]

(22)

which indicates that the short-run measure differs from the long-run measure by level of capacity utilization.

5. Conclusions

The traditional index measure of TFPG, which is expressed as the rate of change in an index of aggregate output not explained by the change in an index of aggregate input, has been very popular in productivity studies. However, most studies ignore the fact that the growth in total factor productivity indicates technological progress only under the restrictive assumptions such as perfectly competitive input-output markets, full economic efficiency, and full static equilibrium.

In this paper, we relax the full static equilibrium assumption. Producers are no longer in long-run equilibrium and so some inputs are not fully adjusted to their static cost-minimizing levels within one period. Then the dual measures of technical progress, returns to scale, and capacity utilization are derived from dual short-run cost function. We also examine how short-run measures are related to long-run measures and the level of capacity utilization.

Reference


