Tanzanian Economic Review, Vol. 12 No. 2, December, 2022: 102–119

Decomposition of Total Factor Productivity Growth in Referral Hospitals in Kenya: 2012-2016

Samuel Ojwang' Oyieke & Innocent John Karamagi§*

Abstract

In Kenya, health provision faces challenges of high poverty levels, high HIV-AIDS, malaria prevalence and poor road infrastructure. Using data from 14 county referral hospitals for the period 2012-2016, this study decomposed the DEA output-oriented Multi-factor Productivity Index (MPI) to identify the causes of productivity growth in Kenya's health sector. The findings show the mean MPI growth for the period was 2.69%, which is driven by a technical change of 3.19%, but dampened by a decline in technical efficiency change of 0.18%, scale efficiency change of 0.07% and pure technological change of 0.15%; with the technical change being scale-augmenting. The study finds RTS to be greater than STC, with both being less than one. Thus, hospitals could enhance productivity by adjusting their scales towards technological optimal scale size (TOPS), and addressing management challenges that debilitate the synergy between technology and human resource capacity.

Keywords: *decomposition, output-oriented, Malmquist total factor productivity, hospitals, Kenya.*

1. Introduction

Total factor productivity grew slowly over time across different regions of the world in the last two decades in contrast to the rapid technological growth experienced in the same period (Foster & Verspagan, 2017). Asia experienced a relatively higher total factor productivity growth compared to Europe, America and Sub-Saharan Africa (SSA) (ibid.). America and SSA have experienced a decline in total factor productivity over time, with a slight growth in the last decade (ibid.). In Kenya, total factor productivity growth has been low, rising immediately after independence, and peaking at around 1.7% per annum in the mid-seventies. It has, however, experienced a steady decline to less than 1% in 2000 (Onjala, 2002). The overall trend has been a slow growth of below 1% despite the rapid technology uptake as demonstrated by the Internet and mobile telephone penetrations.

The concept of total factor productivity was formally introduced into economic analysis by the neoclassical economists in the writings of Solow (1956, 1957). Solow emphasized that residual growth was not captured by changes in factor inputs (labour and capital), but rather by the different intensity of the use of capital, which he referred to as total factor productivity (TFP), or multifactor productivity (MFP). Total factor productivity measures the ratio of total output to the aggregate

* Jaramogi Oginga Odinga University of Science and Technology: samueloyieke@yahoo.com
§ School of Economics, University of Dar es Salaam: innocent.karamagi@gmail.com

 ♦ https://doi.org/10.56279/ter.v12i2.124

measure of all inputs used (Coelli et al., 2005). The term TFP and MFP have been used interchangeably in the literature, though TFP is a misnomer as it refers to all factors of production being combined as inputs simultaneously (ibid.). Such a state of production is rarely attained. TFP is, therefore, a function of: (i) production technology, (ii) efficiency of the production process (efficiency change), (iii) scale of operation, and (iv) managerial skill-set that maximizes the synergy between factor inputs (Sickles & Zelenyuk, 2018).

Total factor productivity change (growth/regress) identifies the change in TFP over time due to the usage of factor inputs to produce certain levels of output. This change occurs through an evolutionary process, in which processes with poor performance are replaced with those with better performance (Hulton, 1986). These process improvements are achieved through organisational structural change, management system upgrades, technological changes, works management improvements, manufacturing technique upgrades, and changing competitive structures (ibid. 1986).

This paper focused on the Lake Region Economic Block (LREB), which comprises 14 counties around Lake Victoria, in Kenya. The region has a population of 13.7m people, with a mean hospital of 1.2 per 100,000; and an average healthcare worker of 1.5 per 100,000 persons. The region is characterized by poor access to healthcare services, with the average distance to healthcare facilities being greater than the recommended 5km from where a population resides. The road infrastructure and other transport logistics are poor, especially during the two rainy seasons (KDH, 2014). This region has a high poverty index with over 65% of the population being classified as earning less than US$1 a day, with the majority mainly engaged in traditional subsistence farming, petty trading and animal husbandry. Landholdings average less than one acre per household. Households are largely female-headed due to proportionately high male mortality rates. The women are predominantly engaged in the low-paying informal sector and unpaid labour as caregivers. There is a high prevalence of HIV-AIDS and malaria, hence a low labour force participation rate in income-generating activities (KDH, 2014). Also, the area is food-insecure due to climate change and variability that is caused by environmental degradation as settlements and light industrial activities proliferate around the Lake Region (World Bank, 2016). Thus, the resulting health challenges and vulnerability exert extra pressure on the few hospitals in the region. Without efficient and productive use of the available health resources, these health challenges could potentially overwhelm the health system in the region (World Bank, 2016).

Thus, this study aims to identify the components of such changes in TFP to underscore their interactions and possible policy imperatives. This objective helps explain the incongruity between rapid technological progress, advancement in skills, and the slow growth in total factor productivity in Africa, and Kenya in particular. Such a decomposition is particularly of interest in the resource-scarce health sector in developing countries experiencing high poverty, unemployment and disease burdens, with low insurance uptakes. Under such a scenario, the disease burden rests heavily on public health providers (governments).

Therefore, the study has decomposed the total factor productivity changes and provided empirical evidence for areas of intervention to enhance the productive use of the scarce health resources, geared at enhancing the provision of health care. In addressing these issues, the paper contributes to the empirical evidence on how technology and productive use of scarce resources can improve the overall health provision in Kenya for better productivity and economic wellbeing.

2. Theoretical Framework

Among the earliest studies to analyse productivity growth was Solow (1957). Solow's (1957) model is summarized as:

$$Q_t = A_t F(K_t, L_t) \qquad (1)$$

Where A_t= total factor productivity, which measures the shift in the production function at given levels of labour (*L*), capital (*K*) and technology set.

This total factor productivity (A_t) is measured using a non-parametric index number (Hulton, 1986). Thus, the approach does not impose a specific form on the production function. Hence, equation (1) could be converted to a (logarithmic) differential of the production function as:

$$\left(\frac{Q^*}{Q_t}\right) = \left(\frac{\partial Q}{\partial K}\right)\left(\frac{K_t}{Q_t}\right)\left(\frac{K_t^*}{Q_t}\right) + \left(\frac{\partial Q}{\partial}\right) + \left(\frac{L_t^*}{L_t}\right) + \left(\frac{A^*}{A_t}\right) \qquad (2) \quad \text{(Hulton, 1986).}$$

That is, the growth rate of real output can be factored out into the growth rate of capital and the growth rate of labour, weighted by their output elasticities and the growth rate of the Hicksian efficiency index (Hulton, 1986). The growth rates of capital and labour, weighted by their respective elasticities, represent movements along with the production function (movement towards technological optimal production scale); while the Hicksian efficiency index measures the shift of the production function (Hulton, 1986).

By total differentiation of equation (2), Solow (1957) showed that the Hicksian efficiency index is a residual growth rate of output that is not accounted for by the growth in inputs (Hulton, 1986), which is given as:

$$\Re_t = \left(\frac{Q^*}{Q_t} - S_t^k\right) = \left(\frac{K_t^*}{K_t} - S_t^l\right) + \left(\frac{L_t^*}{L_t}\right) = \left(\frac{A_t^*}{A_t}\right) \qquad (3)$$

Thus the Solow residual ($\Re_t$) = the Hicksian index. Solow concluded that, theoretically, this growth rate was equal to the growth rate of the Hicksian efficiency parameter ($\frac{A_t^*}{A_t}$) (ibid.).

Abramovitz (1956) referred to this residual as a measure of the degree of our 'ignorance.' "This ignorance could be wanted (technical, scale and organizational innovation) or unwanted like (measurement errors, omitted variables, aggregation bias, and model misspecification)" (ibid: 10–11). In the hospital case, we assume that the unwanted ignorance is minimal, and hence attribute the Solow residual to technical, scale and organizational innovation (Zofio, 2007). Therefore, this residual is captured by technical change (TC), technical efficiency change (TEC), pure technological change (PTC), and the scale change (SC) components in the decomposition of total factor productivity growth (Zofio, 2007). The Solow model was criticized on grounds that it assumed constant returns to scale while production takes place under variable returns to scale (Hulton, 1986). Secondly, the residual is closely linked with the assumption of marginal cost pricing. Finally, the formulation is only valid if innovation improves the marginal productivity of all the inputs proportionally (Sickles & Zelenyuk, 2018).

3. Technology and Distance Functions

The study analyses the output-oriented Malmquist productivity index (MPI) for a panel of 14 hospitals over five years. In terms of technology and distance function, this can be summarized as a case of a panel of I =1… 14 DMUs analyzed for a time period t =1… 5 years. These DMUs transform input vector $x_i^t = (X_i^t, \dots, XN_i^t) \in \Re_+^N$ into output vector$y_i^t = (Y_i^t, \dots, YM_i^t) \in \Re_+^M$. The feasible technology set may therefore be presented as a combination of feasible input-output, as: $S^t = \{(x^t, y^t): x^t$ can produce $y^t\}$ (Coelli et al., 2005). From this framework, a valid representation of the technology from the i^{th} DMU is given by the Shephard's output distance function $D_0^t(x_i^t, y_i^t) = Inf_\theta\{\theta > 0: (x_i^t, y_i^t/\theta) \in S^t$, which is linearly homogeneous of degree +1 in y, and non-increasing in x. If $D_0^t(x_i^t, y_i^t) = 1$, then the focal DMU is said to be efficient, belonging to the best practice technology frontier; otherwise inefficient and outside the best practice technology frontier (ibid.).

This technology frontier is given by the subset:

$$Isoq.S^t(x, y) = \{(x, y): D_0^t(x_i^t, y_i^t) = 1\} \qquad (4)$$

If $D_0^t(x_i^t, y_i^t) < 1$, then a radial expansion of the output vector y_i^t is feasible within the production technology for the observed input level x_i^t, and the evaluated firm is said to be inefficient (ibid.). If period t technology were to exhibit global returns to scale, then the technology S^t implies a mapping $x \to y$ that is homogeneous of degree +1, i.e., $(x, y) \in$; and implies $(\lambda x, \lambda y) \in S^t$ for all $\lambda > 0$: Such technology is represented by:

$$\check{S}^t = \{(\lambda x^t, \lambda y^t): (x^t, y^t) \in S^t, \lambda > 0 \qquad (5)$$

This implies that the output distance function is defined on a linearly homogeneous technology, and is homogeneous of degree -1 in input (Sickles & Zelenyuk, 2018).

4. Decomposition of MPI

Productivity growth results from improved utilization of factor inputs due to input-specific, organization, environmental or market-related features. It could be the result of the synergy between parts, or all of these features, or their components. In this decomposition of productivity growth, the paper focused on a firm's behaviour around the efficiency frontier, and what happens to this frontier over time. This frontier could shift as a result of technology changes (TC). In this case, technology could be scale-augmenting or reduced. As a firm adjusts its scale of operation in response to both internal and external undercurrents, there is a need to adopt appropriate technology, which would then be a firm's response to exogenous factors. This is possible when a firm is operating under variable returns to scale. A firm, in its interactive environment, would strive to catch up with those operating on the efficiency frontier, thus improving its productivity growth (TEC). This catch-up-caused growth would be compounded by the frontier shift due to technology growth. Firms on the efficiency frontier may not necessarily be operating at the technological optimal production scale (TOPS). As such, such firms would be adjusting their operations in response to competitive (peer) forces to operate at TOPS. This adjustment would result in scale adjustment (SEC), thereby generating further growth. The role of management in the whole process is to maximize the synergy between these factor inputs. They could achieve this through their innovativeness (PTC). Thus, productivity growth is a product of all these forces, and may be expressed as: MPI = TC × TEC × SEC × PTC.

The various decompositions discussed are, therefore, an attempt to capture these forces as far as possible in a manner that is close to the reality of a firms' operating environment. Caves, Christensen and Diewert (1982) (CCD) introduced the Malmquist total factor productivity index (MPI) as a tool for efficiency analysis, where they compared the performance of a firm in period 2 using period 1 as the base under constant returns to scale. The index was therefore decomposed as:

$$m_0^1(x_i^1, y_i^1, x_1^2, y_i^2) = \frac{d_0^1(x_i^2, y_i^2)}{d_0^1(x_i^1, y_i^1)} = TC_0^{1,2}(x_i^2, y_i^2) \times TEC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \qquad (6)$$

Where m_o^1 refers to the output-oriented Malmquist index for period 2 output to the base period, while d_0^1 is the output distance function for period 2 to base period technology (Coelli et al., 2005).

This index decomposed productivity change into two components: technical change (TC), and technical efficiency change (TEC). Whereas $d_0^1(x_i^2, y_i^2)$ represents a mixed period distance function that compares period 2 output to the base period technology (Coelli et al., 2005; Fare, et al., 1989), $TC_0^{1,2}(x_i^2, y_i^2)$ captures the shift in technology between the two periods concerning the actual best-practice frontier. The second part of the equation $TEC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2)$ (technical efficiency change), measures the change in relative efficiency, i.e., how far actual observed production deviates from the maximum potential production (Fare et al., 1989).

This conceptualization did not take into consideration the proportionality property of being homogeneous of degree -1 in inputs and +1 in output (equation 5); and also it ignored the effect of returns to scale (RTS) on productivity changes (Sickles & Zelenyuk, 2018). Hence, the Caves, Christensen and Diewert's (1982) decomposition was criticized for giving an imprecise measure of productivity change as it ignored the scale factor and considered the actual best technology set (Sickles & Zelenyuk, 2018).

Fare et al. (1989) sought to address the weaknesses of the Caves, Christensen and Diewert's (1982) decomposition by proposing a decomposition of MPI that took into account the benchmark technology as opposed to the actual best-practice technology set. By indirectly defining the MPI concerning constant returns to scale cone-technology, the index imposed a technology representation that allowed for a comparison of a firm's productive performance to a technology optimal productive scale (Sickle & Zelenyuk, 2018).

According to Forsund, Lovell and Schmidt (1980), the proportionality property was critical to determining whether an index qualified as MPI. This property stated that, if outputs were to be increased in the same proportion between periods 1 and 2, while inputs remained constant in the same periods, then the productivity index ought to increase by the same proportion (Forsund et al., 1980). Conversely, a reduction in inputs in the same proportion—holding outputs constant—should lead to an equal increase in the productivity index. This property made it essential that the distance function be linearly homogeneous of degree +1 in output and -1 in input (equation 5) (Sickle & Zelenyuk, 2018). Thus, Fare, Grosskopf and Lindgren's (1989) decomposition is given as:

$$\breve{M}_0^1(x_t^1, y_t^1, x_t^2, y_t^2) = PTC_0^{1,2}(x_i^2, y_i^2) \times TEC_0^{1,2}(x_i^1, y_i^1, x_i^1, y_i^2) \qquad (7)$$

The difference between Fare, Grosskopf and Lindergren's (1989) decomposition and the Caves, Christensen and Diewert's (1982) decomposition is that in the former, the technical change term produced potential productivity change between DMUs operating at the technology optimal productive scale-size (Ray & Desli, 1997). Hence, $PTC_0^{1,2}(x_t^2, y_t^2)$ measures technical change concerning the virtual supporting cone-technology, which implies that it can only measure technical change when constant returns to scale are assumed (Ray & Desli, 1997). The efficiency change $TEC_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2)$ term now measures how far a DMU is from the benchmark cone productivity, and therefore relates to both technical and scale change terms (Ray & Desli, 1997).

In a later revision, Fare, Grosskopf and Lovell (1994) sought to incorporate the effect of returns to scale, $RTS_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2)$. They asserted that productivity change is determined by scale changes that can be captured by returns to scale (Fare et al., 1994). Therefore, they redefined the MPI by using the virtual cone-technology and scale component as:

$$\breve{M}_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2) = PTC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \times TEC_0^{1,2}(x_i^1, y_i^1, x_i^1, y_i^2)$$
$$\times SEC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \qquad (8)$$

However, this decomposition was criticized by Lovell (2003) on account that it ignored the shift in the best practice frontier. Hence, it could lead to exaggeration or underestimation of the scale efficiency value.

Ray and Desli (1997) criticized both the Caves, Christensen and Diewert's (1982) and the Fare, Grosskopf and Lindgren's (1989) decompositions on grounds that, in the case of scale efficiency change, a true production technology ought to exhibit variable returns to scale (VRS). Therefore, they proposed an alternative decomposition that had technical change measured relative to the VRS frontier, and a modified scale change component (return-to-scale) that was not exactly equivalent to the scale efficiency change of Fare, Grosskopf and Lindgren (1989). They did this by introducing a variant that measured how far a firm is from the benchmark cone productivity (Ray & Desli, 1997). This decomposition, therefore, comprised both technical and scale efficiency, such that:

$$\breve{m}_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) = TC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \times .TEC_0{}^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \times .RTS_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \qquad (9)$$

Fare, Grosskopf and Lovell (1994) had earlier argued that $SEC_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2) = RTS_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) / STC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2)$. Hence, treating returns to scale change to represent the scale efficiency may lead to incorrect identification of the scale properties of the essential technology. Therefore, if one accepts a decomposition that includes the acceptable notion of effective technical change at a firms' input scale, and not the potential productivity change at the optimal level, then one loses efficiency change term for a returns to scale component (Fare, Grosskopf & Lovell, 1994).

Simar and Wilson (1998) offered an alternative decomposition of the Malmquist productivity index, in which they added the scale efficiency change term that considers a firms' optimal scale (benchmark) technology as presented in Fare, Grosskopf and Lovell (1994). This addition requires a term in the index to reflect the scale-bias of the technical change $STC_0^{1.2}(x_i^1, y_i^1, x_i^2, y_i^2)$ (Simar & Wilson, 1998). They, therefore, proposed a recasting of the pure technical change as:

$$PTC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) = TC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \times STC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) \qquad (10)$$

This decomposition took into account the effect of scale-bias in the pure technical change by indicating that it is a product of technical change, and the scale-bias of technical change. Thus, the MPI is presented as:

$$\breve{M}_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) = TC_0^{1,2} \times TEC_0{}^{1,2} \times SEC_0^{1,2} \times STC_0^{1,2} \qquad (11)$$ (Zofio, 2007).

Thus, if the DMU experiences efficiency growth between the base and the focal period, then $SEC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2)$ >1; while if the DMUs scale-bias technical change works against its inputs, then $STC_0^{1.2}(x_i^1, y_i^1, x_i^2, y_i^2)$ >1 (Zofio, 2007). Such a situation only occurs if returns to scale make a positive contribution to productivity change, that is, $RTS_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2)$ >1; which is larger than the negative change in the scale-bias of the technical change (ibid.). Conversely, if a scale efficiency gain is accompanied by a favourable change of the scale-bias technical change $STC_0^{1.2}(x_i^1, y_i^1, x_i^2, y_i^2)$ <1, then the presence of increasing returns to scale $RTS_0^{1,2}(x_t^1, y_t^1, x_t^2, y_t^2)$ <1 dampens the effects of such efficiency gains (ibid.).

This decomposition is preferred in cases where scale is expected to play a role in explaining productivity growth (Sickles & Zelenyuk, 2018). This is because, this decomposition includes a scale-bias of technical change; thereby allowing a comprehensive analysis of the general framework of productivity change, efficiency change, and technological change: both from technical and scale perspectives (Sickles & Zelenyuk, 2018). It, therefore, affords a realistic coalescing structure for the classification of technical and efficiency changes (Sickles & Zelenyuk, 2018).

Scale efficiency compares the highest productivity attained by a DMU at the actual scale to the highest productivity observed at the optimal scale (Sickles & Zelenyuk, 2018). Thus, a scale efficiency change would compare scale efficiency in both the base and the comparison periods. Therefore, it is possible that in moving from the base to the comparison period, a DMU may improve its productivity by the use of returns to scale offered by the benchmark technology, or a scale change due to expansion of operations (Balk, 2001). The benchmark technology may also change from the base period to the comparison period during the same time to accommodate the expanded operations (Balk, 2001).

Taking the period 2 technology to measure returns to scale, and the base period (1) to measure the scale-bias of the technical change, then: $SEC_0^{1,2} = RTS_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2)/ STC_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2)$ (Zofio, 2007). Therefore, if $RTS_0^{1,2} > 1$, then the DMU's performance improves on a scale concerning the base period productivity benchmark by exploiting increasing returns to scale, and getting closer to the maximum potential scale size (ibid.). On the contrary, if $RTS_0^{1,2} < 1$, then the DMUs move away from the optimal scale. If, however, $RTS_0^{1,2} = 1$, then the DMU performance change is scale-neutral (ibid.).

However, these changes need to be considered alongside changes in the scale-bias of technical change $\left(STC_0^{1,2}\right)$. If $STC_0^{1,2} > 1$, then the positive effect on productivity is only possible if $RTS_0^{1,2} > 1$ and significant, to counterbalance the negative effect exerted, $STC_0^{1,2} > 1$ (ibid.). On the contrary, if a scale gain is accompanied by a positive scale-bias-change of technical change ($STC_0^{1,2} < 1$), then the existence of increasing returns to scale dampens such scale gains (ibid.).

If, however, $RTS_0^{1,2} < 1$, then scale gains are still possible so long as the positive scale-bias of technical change is not counterbalanced by those lowering returns to scale, i.e., $RTS_0^{1,2} > STC_0^{1,2}$; and both terms are less than one. Hence, the issue is whether technical change is scale-augmenting or reducing, or whether it results in increasing returns to scale, or decreasing returns to scale (ibid.).

In health, the role of technical change, technological innovation and scale changes are critical to delivering quality and affordable healthcare, since modern health services are technology-intensive, and the modes of health care delivery are ever-changing as a result of technical changes and technological innovations. This technological innovation would involve both products (namely, self-diagnostic kits and service delivery using modern technological platforms), and process innovations (namely, modern technology). Technological changes are endogenous, that is to say, they are a health sector's response to healthcare challenges in a sector characterized by rapid technological changes (Sloan & Hsieh, 2017). Thus, given the same input resources, the performance difference would be a result of an innovative use of this resource. Healthcare providers, therefore, need to leverage technology to improve service provision.

5. Empirical Literature on Productivity Change

Table 1 selectively summarizes the empirical literature. These studies were purposively selected to incorporate only those that used the DEA Malmquist total factor productivity index. The reviewed empirical studies demonstrate that productivity growth has elicited considerable interest across the world, as shown by the spectrum of the studies.

These studies are limited to the Caves, Christensen and Diewert's (1982) (CCD) decomposition, thus failing to capture adequately the effects of returns to scale and the relationship between scale and technology changes. In Kenya, there has been only one documented study of this nature (Kirigia et al., 2007), although decomposition is critical for a resource-scarce economy such as that of Kenya, and particularly on the health sector. Issues of low factor productivity growth and rapid technological innovation have been particularly evident in the last decade, yet no study has been documented in Kenya with this focus.

6. The Empirical Model

The current study adopts the output-oriented Malmquist total factor productivity index since hospitals in Kenya have little control over inputs because they receive them from the government, based on the government's allocation policy. The study adopted Simar and Wilson's (1998) decomposition, with the modifications proposed by Zofio (2007) and Sickles and Zelanyuk (2018), which defined pure technical change as (STC × TC) to account for managerial discretion. This model would be more representative of the operations of the counties' referral hospitals in Kenya during this transition phase (2012–2016). Thus, the model is presented as:

$$\breve{M}_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) = TC_0^{1,2} \times TEC_0{}^{1,2} \times SEC_0^{1,2} \times STC_0^{1,2} \qquad (12)$$ (Zofio, 2007).

Since the MTFP index measures change for either period 1 or period 2 technology, the Index is therefore the geometric mean of the change for the two periods' technology. It is given as:

$$m_0^{1,2}(x_i^1, y_i^1, x_i^2, y_i^2) = [\{m_0^1(x_i^1, y_i^1, x_i^2, y_i^2)\} \times \{m_0^2(x_i^1, y_i^1, x_i^2, y_i^2)\}]^{0.5} \qquad (13)$$ (Zofio, 2007).

The above Malmquist index requires the computation of four distance functions:

$d_0^2(y_i^2, x_i^2)$, $d_0^1(y_i^1, x_i^1)$, $d_0^1(y_i^2, x_i^2)$, and $d_0^2(y_i^1, x_i^1)$ (Coelli et al., 2005).

Table 1: Summary of Empirical Studies

Author (s)	Country	Period	Unit of Analysis	Percentage Changes
Nghiem et al. (2011)	Australia	1996-2004	Public Hospitals	TFP=1.6, PTE=1,SEC=0.4, TC=1.4
Quellete &Viestraete, 2004)	Canada	1997-98	15 hospitals	TFP=-0.08,TEC=-0.06, TC=5
Ganon (2008)	Ireland	1995-98	Hospitals (6 regional, 8 general) 22 Country	TFP=2.8, TEC=-0.06, TC=3.4. TFP=1.2, TEC=-0.01,TC=1.3 TFP=-0.03,TEC=0.5,TC=-0.08
Masri &Asbu (2018)	Eastern Mediterranean	2003-2014	20 Country hospitals	TFP=-3.8,TC=-3.8, PTE=-9.1,TEC=5.8,
Cheng et al. (2015)	China	2010-2013	114 Hospitals	TFP=7.8,TC=6.8, PTE=0.9,SEC=-1.9
Mogha et.al. (2015)	India	2001-2011	27 government hospitals	TFP=4.9, TC=2.2, TEC=2.6, PTE=2.3
Singh et al. (2017)	India	2013-2014	18 government hospitals	TFP=9.4, TC=5.9, TEC=3.3, SEC=3.2
Kirigia et al. (2012)	Benin	2003-2007	Zonal hospitals	TFP=-5.3, TC=-24, TEC=26, PTE=9.9, SEC=15.8.
Ali et al. (2017)	Ethiopia	2007/8-2012/13	12 hospitals	TFP=-4,TC=-5,TEC=1, PTE=2,SEC=-1
Kirigia et al. (2008)	Angola	2000-2002	28 municipal hospitals	TFP=4.5,TC=-7.3, TEC=12.7, PTE=5, SEC=7.3
Tlotlego et al.(2010)	Botswana	2006-2008	21 non-teaching hospitals	TFP=-1.5,TEC=-4.5,PTE=4.2,SEC=-1
Babola & Moodley (2020)	South Africa	2015-2017	38 district hospital	TFP=4.8,TC=6.9,PTE=-1.9
Mujasi & Kirigia (2016)	Uganda	2009/10-2013/14	17 referral hospitals	TFP=4.9,TC=3.1,TEC=2.3,PTE=1.2,SEC=0.8

Source: Summarized by authors.

Table 2: Variables Description and Data Sources

Variable	Description and Measurement	Data Sources
Medical staff	Consists of doctors, nurses both registered and community health nurses, clinical officers medical officers of health and laboratory technicians.	Individual hospitals' published reports
Number of hospital beds	Hospital beds and cots as at the end of every financial year.	Individual hospitals' published reports.
Outpatient department visits	All outpatient cases including repeat visits recorded for a given year.	Individual hospitals' published reports.
Deliveries	A total number of maternity cases in the hospitals per year.	Individual hospitals' reports.
Total inpatient admissions	A total number of admissions in a given year.	Individual hospitals' published reports.
Bed occupancy rate	(Inpatient days/ bed days)X100. Where inpatient days=admissions X ALOS, and bed days=number of beds X 365(6)	Individual hospitals' published reports
The average length of stay	ALOS=Discharge days for all services/total discharges + deaths	Individual hospitals' published reports
Teaching status	This a dummy variable taking a value of 1 if a teaching hospital and 0 if not	Individual hospitals' published reports
Catchment population	The population of a given county	KNBS (GOK).
Health inefficiency	$(1 - VRS_{TE})$[1]	DEA technical efficiency, estimated from research data.

Source: Compilation by Authors

7. Presentations and Discussion

7.1 Descriptive Statistics

Tables 3 and 4 provide the descriptive statistics of the data to show the distribution and dispersion around the central tendencies and measures of dispersion. Table 3 shows how the efficiency variables are distributed within a defined range. It, therefore, intends to give the reader a birds' eye view of the sample data.

Table 3: Descriptive Statistics

Descriptive Statistics	N Stat	Min Stat	Max Stat	Mean Stat	Std. Error	Std. Dev Stat	Skew Stat	Skew Std. Error	Kurt Stat	Kurt Std. Error
Beds	70	124	365	218.37	9.257	77.448	.716	.287	-.746	.566
Outpatients	70	9600	46375	17788.29	1096.545	9174.357	1.968	.287	2.883	.566
Medical Staff	70	53	350	139.86	9.897	82.804	1.359	.287	0.399	.566
Deliveries	70	1108	4068	2399.87	108.076	904.230	.191	.287	-0.405	.566

Source: Authors' Computations from the Research Data.

[1] The VRS_TE scores are derived from the DEA output oriented technical efficiency scores reported in the Thesis submitted to the University of Dar es Salaam titled "*Technical Efficiency and Total Productivity of County Referral Hospitals in Kenya: 2012-2016*" from where this paper is extracted.

Table 4 : Interquartile Distribution of the Vital Statistics

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Maximum
BED	124	158	192	218.4	250	365
OPD	9,600	12,533	14,530	16,559	18,222	35,640
MSTAFF	58	86.25	102.50	142.20	172.25	350
DELIVERIES	1,108	1,480	2,262	2,400	3,315	4,068

Source: Authors' computations from Research Data

7.2 Estimations and Discussions

The distance function takes a value less than—or equal to—one (1) if the output vector is an element of the feasible production set. Otherwise, the function takes a value greater than one (1) if the output vector is located outside the feasible production set. In this analysis, the average output distance function is less than one (1), implying that the output vector is an element of a feasible production set.

Table 5 presents a summary of the geometric mean of the decomposition of MPI into its parts, showing how these parts have affected the MPI.

Table 5: Geometric Means of the MPI and its Components per Referral Hospital (2012-2016):

Hospital	MPI	TEC	TC	PTC	STC	SEC	RTS
1	1.0561	1.0839	0.9796	0.8245	0.8417	1.0048	0.8457
2	1.0255	1.0000	1.0255	1.0000	0.9751	1.0000	0.9751
3	0.9589	1.0000	0.9596	1.0000	1.0421	1.0000	1.0421
4	0.9797	0.9510	1.0317	0.9714	0.9416	0.9788	0.9216
5	1.0728	0.9840	1.0897	1.0000	0.9177	0.9788	0.8982
6	0.7503	0.9648	1.0515	0.9949	0.9462	0.9840	0.9311
7	1.0094	0.9517	1.0676	0.9838	0.9215	1.0078	0.9287
8	1.0614	0.9781	1.0859	1.0000	0.9210	0.9671	0.8907
9	1.0458	1.0208	1.0226	1.0134	0.9910	0.9782	0.9694
10	1.0839	1.0012	1.0825	1.0000	0.9238	1.0070	0.9303
11	1.0048	1.0123	0.9965	0.9831	0.9866	1.0012	1.0156
12	1.0069	1.0114	0.9980	0.9926	0.9946	1.0294	1.0241
13	1.0167	1.0152	1.0063	1.0011	0.9948	1.0127	1.0074
14	1.0463	1.0000	1.0463	1.0000	0.9558	1.0000	0.9558

Note: 2012 is the base year. Hospitals have been assigned codes to conceal their identity as part of the ethical consideration.

Source*:* Authors' Computations from Research Data using R statistical software

7.2.1 Productivity Changes (MPI)

The study reports productivity gains for the period 2012–2016 for most of the county referral hospitals (78.6%). This productivity gain was driven by technical change and technical efficiency change. However, it was dampened by pure efficiency change and scale efficiency changes as discussed in the relevant sections.

7.2.2 Technical Efficiency Change (TEC)

There were six hospitals (42.86%) with efficiency change greater than one (1). This implies that there was a catching-up of these hospitals (ID 1, ID 9, ID 10, ID 11, ID

12, and ID 13) with their peers on the efficiency frontier. There were, however, 5 hospitals (35.71%) with an efficiency change of less than one (1), implying that these hospitals (ID 4, ID 5, ID 6, ID 7 & ID 8) were moving away from the efficiency frontier. There was an efficiency decline from 1.0254 in 2013 to 1.0122 in 2016 (1.32%), implying that the rate at which inefficient hospitals were catching up with their efficient peers declined throughout this period. This decline was larger for this sample in 2013–2014 (8.76%).

7.2.3 Technical Change (TC)

In terms of individual hospitals, there were 10 hospitals (71.43%) (ID 2, ID 4, ID 5, ID 6, ID 7, ID 8, ID 9, ID 10, ID 13, and ID 14), where there was technical progress (outward shift in the frontier). There were, however, 4 hospitals (28.57%) where there was technical regress (inward shift in the frontier). This result shows that the main driver of productivity growth in this period was technical progress, following the implementation of the devolved system, which necessitated the upgrading of district hospitals to county referral hospitals, and considerable technology upgrades. The technical progress reported by the majority of the hospitals (71.43%) could have theoretically arisen from the application of modern technology, enhanced skill-set, improved hospital equipment, and infrastructure. The source of this technical progress was, however, not investigated in this study.

7.2.4 Scale-bias Technical Change (STC)

In 92.9% of the hospitals, the scale-biased technical change was less than one (1) (STC<1), implying that the scale-bias of the technical progress had a positive effect (the technical progress is input augmenting). However, this positive effect was counterbalanced by the negative scale change effect resulting in six of the county referral hospitals (42.86%) experiencing a positive scale change effect.

7.2.5 Pure Technical Efficiency Change (PTC)

Many DMUs employ trained managers to create synergy in resource usage (surplus value). Managerial inefficiency was reported in 6 hospitals (42.9%) where pure technical efficiency change was negative. The role of the management was only input augmenting in 2 hospitals (14.3%). In the remaining six hospitals (42.9%), the management's role was productivity neutral. There was an overall decline in the pure technical efficiency of 1.45% in the period 2013–2016. This meant that the contributions of management and operational practices to productivity declined. In 2014, PTC regress was 0.9676, which means that management and operational practices contributed negatively to hospital productivity. The mean PTC for the period was 0.9985, with a maximum of 1.3227 and a minimum of 0.8334, which confirmed the management challenges facing the health sector as illustrated by the number of financial and human resource management challenges.

7.2.6 Scale Efficiency Change (SEC)

In 42.86% of the referral hospitals, scale efficiency was greater than one (1), which means that in these hospitals the scale of operations contributed positively to productivity growth, and these hospitals were moving towards the optimal scale size. In 21.43% of the county referral hospitals (ID2, ID3 & ID14), the scale change was one (1); that is to say, there was no scale change in these hospitals, thus no effect on productivity change and no movement towards or away from the optimal scale size. In 35.7% of the hospitals, scale contributed negatively to productivity growth, with these hospitals moving away from the optimal scale size since the negative RTS was smaller than the positive influence of the STC.

7.2.7 Returns to Scale Changes (RTS)

In 71.4% of the hospitals, RTS was less than one (1), implying that the performance of these hospitals declined on a scale concerning the base period; and the hospitals were moving away from the optimal scale size. This period witnessed scale inefficiency of hospitals partly explained by the euphoria of devolution characterized by the upgrading of the district level hospitals to county referral hospitals, resulting in improved access; while other filter facilities may have witnessed a reduced pressure. In 28.57% of the hospitals, the RTS >1, meaning that the performance of these hospitals improved the scale relative to the base period; and they moved towards the optimal scale size.

Table 6 gives a summary of the geometric mean of the decomposition of annual total factor productivity growth for all the hospitals decomposed into the various components for the period 2012–2016, with 2012 as the base period.

Table 6: Decomposition Of Annual Total Factor Productivity Change (Geometric Mean)

Year	MPI	TEC	TC	PTC	STC	SEC	RTS
2013	1.0547	1.0254	1.0311	1.0172	0.9865	1.0071	0.9935
2014	1.0454	0.9378	1.1164	0.9676	0.8667	0.9702	0.8409
2015	1.0039	1.0172	0.9885	1.0067	1.0184	1.0101	1.0287
2016	1.0038	1.0122	0.9918	1.0027	1.0199	1.0097	1.0298
Mean	1.0269	0.9982	1.0319	0.9985	0.9729	0.9993	0.9732

Source: Authors' Computations from Research Data using R statistical software

NB: 2012 is the base year.

As summarized in Table 6, the study found an annual total factor productivity growth (MPI) of 2.69%. This growth is attributable to an average technical change of 3.19%, which outweighed the marginal regress in technical efficiency change of 0.18%; a 0.15% regress in pure technical change; and a 0.07% regress in scale efficiency change. Thus, this productivity gain is dampened by moving away from the optimal scale of the scale-inefficient hospitals. This is due to the healthcare pressure and the perceived better quality care in these referral hospitals, and a weak referral system. There have been managerial challenges, as witnessed by various industrial disputes, and the fact that the full transition period was shortened due to the pressure from county governments, thereby denying these governments the necessary preparedness to handle the fully devolved healthcare. There has also been the occasional pull-and-push at the county budgetary process, and the central governments' delayed release of funds for the devolved functions. The confluence of these factors has compounded the healthcare delivery challenges.

The positive growth depicted in Table 6 is consistent with several of the reviewed studies, albeit with differing growth rates (Singh et al., 2015 (9.4%); Babola & Moodley, 2020 (4.8%); Cheng, Tao et al., 2015 (7.8%); Mijasi & Kirigia, 2016 (4.9%); Mogha et al., 2015 (4.9%); Ganon, 2008 (2.8%) and Nghiem et al., 2011 (1.6%)). In many of the studies, the growth was a result of technical change. Pure technical change played a minimal role in many of these studies, which points to the need for enhancing managerial skills in hospitals. Scale efficiency seems to be a general problem in many of these studies, as well as for Kenya. Some studies reported productivity regress, which was driven by efficiency decline or technical regress (Quellete & Viestraete, 2004 (-0.08%); Mastri & Asbu, 2018 (-3.8%); Kirigia et al. 2012 (-5.3%); Ali et al., 2017 (-4%) and Tlotlego et al., 2010 (-1.5%)).

Table 7 presents a summary statistical distribution of the components of the Malmquist productivity index showing their spread around the measures of central tendencies and dispersion.

Table 7: Summary of Estimation Results

Variable	Min.	1st Qu	Median	Mean	3rdQu.	Max
MPI	0.8613	0.9959	1.0177	1.0269	1.0476	1.2992
TEC	0.8241	0.9789	1.0000	0.9982	1.0105	1.3889
TC	0.8613	0.9818	1.0193	1.0319	1.0710	1.2376
PTC	0.8334	0.9882	1.0000	0.9985	1.0000	1.3227
SEC	0.8445	0.9927	1.0000	0.9993	1.0083	1.0737

Source: Authors' computation from research data using the R program

The results in Table 7 describe the distribution of the decomposition of productivity into TEC, TC, PTC and SEC. These variables show general skewness in their distribution. Technical efficiency change (TEC), pure technical change (PTC) and scale efficiency change (SEC) are negatively skewed, implying that the mean is influenced by a few low scores. The Malmquist productivity index (MPI) and the technical change are skewed to the right, implying that the mean is influenced by a few high scores.

Figure 2 depicts the trends in the components of the MPI over the period 2012–2016, with 2012 as the base period.

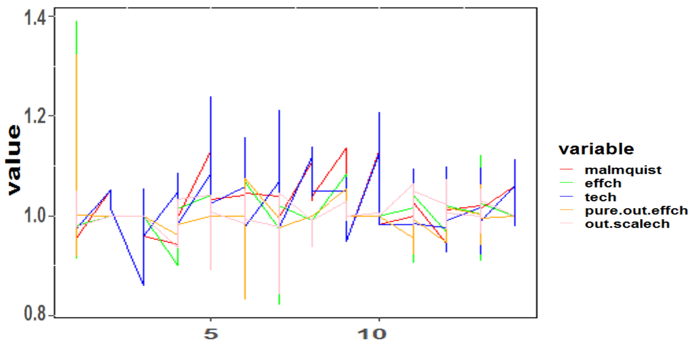

Figure 2: Trends in Productivity Changes
Source. Generated by the authors from the research data

Figure 2 shows that growths in MPI and technical changes were largest in 2012–2013. These trends, however, declined over time before reaching their lowest in 2015. Pure technical change (PTC) and scale changes had little variability, with PTC remaining below one (1), while the latter was around one (1). These long-term trends show that productivity growth was largely driven by technical change and, to a limited extent, efficiency change.

8. Conclusions

This study used secondary data for the period 2012–2016 from 14 county referral hospitals in Kenya to decompose the DEA output-oriented multi-factor productivity index (MPI) to identify the causes of productivity growth in Kenya's health sector. The decomposition of the total factor productivity changes has provided empirical evidence for areas of intervention to enhance the productive use of the scarce health resources, geared at enhancing the provision of healthcare.

The findings have shown that technical progress is the major driver of productivity growth. In addition, scale inefficiency is relatively high, which implies that the majority of the hospitals are operating at less than the technology optimally productive scale sizes (TOPS). The evidence shows that 60% of the hospitals recorded decreasing returns to scale, whereas only 30% recorded increasing returns to scale. These hospitals were operating beyond their capacity because access had been enhanced as the referral hospitals that upgraded their status increased from 2 to 14 within the studied region. This upgraded status exerted pressure on the equipment, facilities and staff. However, due to technical progress and modernization of equipment, productivity grew—on average—at 2.69% for the period.

Hospital managements faced more challenges as these facilities experienced an upsurge in the number of patients and resource constraints. This was reflected in the empirical results of minimal contributions to productivity growth of pure technical efficiency change. The growth in technical efficiency (catch-up effect) slowed down due to the majority of the hospitals becoming less efficient towards the end of this period.

In general, the findings indicate that the productivity growth would have been further enhanced if hospital managements were adequately prepared to handle matters of the devolution of health services. This would have required facility and infrastructure upgrades, and a well-motivated and skilled staff to leverage technical progress (TC).

References

Abramovitz, M. (1956). Resource and output trends in the United States since 1870. *American Economic Review,* 46: 5–23.

Ali, M., Debela, M. & Bamud, T. (2017). Technical efficiency of selected hospitals in Eastern Ethiopia. *Health Economics Review,* 7(24): 1–13. https: //doi.org/10.1186/s13561–017–0161–7.

Babola, T.K.& Moodley, I. (2020). Technical efficiency and productivity of public district hospitals in Kwa Zulu Natal Province, South Africa, *Journal of Public Health Research* 9(1). DOI: 10.4081/jpr.2020.1741.

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis,* 15: 159–183.

Caves, D.W., Christensen, L.R. & Diewert, W.E. (1982). The Economic theory of index numbers and the measurement of inputs, outputs and productivity. *Econometrica,* 50 *(*6*):* 1393–1414.

Cheng, Z., Tao, H., Cai, M. et al. (2015). Technical efficiency and productivity of Chinese county hospitals: An exploratory study in Henan Province, China. *BMJ Open,* 5(9): 1–10 5: e007267. DOI: 10.1136/bmjopen-2014–007267.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis* (2 ed.). Springer Science, USA.

Fare, R., Grosskopf, S., Lindgren, B. & Roos, P. (1989). Productivity changes in Swedish hospitals: A Malmquist output approach. In A. Charnes, W. Cooper, A. Lewin & L. Seiford (1994). *Data Envelopment Analysis: Theory, Methodology and Applications.* Boston MA. Kluwer Academic Publishers. Dordrecht.

Fare, R., Grosskopf, S. & Lovell, C. (1994). *The measurement of efficiency of production.* Boston MA. Kluwer- Nijhoff.

Fare, R., Grosskopf, M., Norris, M. & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries: Reply. *American Economic Review*, 87*(*5*):* 1040–1043.

Forsund,F.R., Lovell, C.A.K. & Schmidt, P. (1980). A Survey of frontier production functions and their relationship to efficiency measurement. *Journal of Econometrics,* 13*:* 5–25.

Foster, N.M. & Verspagen, B. (2017). Decomposing total factor productivity growth in manufacturing and services. *Asian Development Review,* 34*(*1*):* 88–115.

Ganon, B. (2008). Total factor productivity growth in hospitals in Ireland 1995–1998: A non-parametric approach. *Applied Economic Letters,* 15(2): 131–135.

Hulton, C.R. (1986). Productivity Change, capacity utilization, and the source of efficiency growth. *Journal of Econometrics,* 33*:* 31–50.

KDH (2014). *Demographic health survey*. Nairobi. Government Printers.

Kirigia, J.M., Emrouznejad, A., Cassoma, B., Asbu, E. & Barry, S. (2008). A performance assessment method for hospitals in Angola. *Journal of Medical Systems,* 32*(*6*):* 509–519.

Kirigia, J.M., Sambo, L.G., Mensah, O., Mwikisa, C., Asbu, E.Z., Makoudoda, P. & Hounnankan, A. (2012). Productivity changes in Benin zone hospitals: A non-parametric Malmquist approach. *African Journal of Health Economics*, 1–14.

Kirigia, J.M., Zere, E. A., Greene, A.W. & Emrouznejad, A. (2007). Technical efficiency, efficiency change, technical progress and productivity growth in national health systems of continental Africa. *Social Science Review,* 23*(*2*):* 19–40.

Lovell, C.A.K. (2003). The decomposition of Malmquist productivity indices. *Journal of Productivity Analysis,* 20: 437–438. *System Human Resources,* 30(6): 473–481.

Masri, M.D. & Asbu, E.Z. (2018). Productivity change of national health system in WHO Eastern Mediterranean region: Application of DEA base Malmquist productivity index. *Global Health Research Policy,* 3(22): 1–13. Doi: 10.1186/s41256–018–0077–8.

Mogha, S.K., Yadav, S.P. & Singh, S.P. (2015). Technical efficiency and productivity growth in public sector hospitals of Uttarakhand (India). *International Journal of System Assurance Engineering and Management,* 6*(*4*):* 390–406.

Mujasi, P. & Kirigia, J. (2016).Productivity and efficiency changes in referral hospitals in Uganda: An application of Malmquist total factor productivity index. *Health Systems and Policy Research,* 3(1)9. Doi: 10.21767/2254–9137.100028.

Nghiem, S., Coelli, T. & Barber, S. (2011). Sources of productivity growth in health services: A case study of Queensland public hospitals. *Economic Analysis and Policy,* 41(1): 37–48.

Onjala, J. (2002). *Total factor productivity in Kenya: The link with trade policy*. Nairobi Kenya, AERC.

Quelleta, P. & Vierstraete, V. (2004). Technical change and efficiency in the presence of quasi-fixed inputs; DEA application to the hospital sector. *European Journal of Operation Research,* 154: 755–768.

Ray, S.C. & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in Industrialized countries: Comment. *American Economic Review* 87*(*5*):* 1033–1039.

Sickle, R.C. & Zelenyuk, V. (2018). *Measurement of productivity and efficiency, theory and practice.* New York NY: Cambridge University Press.

Simar, L. & Wilson, P.W. (1998). Productivity *growth in industrial countries*. Discussion paper no. 9810, Universite Catholique de Louvain.

Singh, D., Singh, S. & Singh, K. (2017). Efficiency and total factor productivity of public hospitals in Haryana (India): A data envelopment analysis. *International Journal of Advances in Management and Economics*, 6*(*2*):* 71–84.

Sloan, F.A. & Hsieh, C.R. (2017). *Health economics*. UK: The MIT Press,.

Solow, R.M. (1957). Technical change and aggregate production function. *Review of Economics and Statistics,* 39*:* 312–320.

Tlotlego, N., Nonvignon, J., Sambo, L.G., Asbu, E.Z. & Kirigia, J.M. (2010). Assessment of productivity of hospitals in Botswana: A DEA application. *International Archives of Medicine,* 3(27): 1–14.

World Bank, (2016). *Reviving Lake Victoria by restoring livelihood.* The World Bank. Washington D.C.

Zere, E., McIntyre, D. & Addison, T. (2001). technical efficiency and productivity of public sector hospitals in three South African provinces. *South African Journal of Economics,* 69(2): 336–358.

Zofio, J.L. (2007). Malmquist productivity index decomposition: A unifying framework. *Applied Economics*, 39(18): 2371–2387. Taylor Francis (Routledge).