TECHNICAL EFFICIENCY AND TOTAL FACTOR PRODUCTIVITY GROWTH IN UGANDA'S DISTRICT REFERRAL HOSPITALS

By

Bruno Lule Yawe

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Economics) of the University of Dar es Salaam

University of Dar es Salaam
September, 2006

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance by the University of Dar es Salaam a thesis entitled: ***Technical Efficiency and Total Factor Productivity Growth in Uganda's District Referral Hospitals***, in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Economics) of the University of Dar es Salaam.

Prof. Nehemiah E Osoro
(Supervisor)

Date: ------------------------

Dr Rowena Jacobs
(Supervisor)

Date: ------------------------

DECLARATION

AND

COPYRIGHT

I, **Bruno Lule Yawe**, declare that this thesis is my own original work and that it has not been presented and will not be presented to any other University for a similar or any other degree award.

Signature: …………………………..

ACKNOWLEDGEMENTS

Many institutions and persons have contributed in myriad ways toward the completion of this thesis specifically, and the PhD program, in general. These institutions and persons are too many to be acknowledged individually in the limited space. As a consequence, from the outset I register my heartfelt appreciation to all those persons and institutions who contributed to the entire PhD program becoming a reality. To all those institutions and persons, I say a big thank you. Nevertheless, there are individuals who deserve special mention without implication.

First and foremost, I would like to take this opportunity to thank the African Economic Research Consortium for funding my PhD studies and for partial funding of this research project. I feel deeply honored to be a recipient of this scholarship award. Additional financial support from Makerere University under project No: Carnegie/Mak/2005/0036, is gratefully acknowledged.

My thesis supervisors Prof. Nehemiah E Osoro and Dr Rowena Jacobs have been the best supervisors every doctoral candidate could ever wish for. Besides being very supportive and helpful whenever things seemed difficult, Nehemiah and Rowena have always been there for me and their parental advice and support have been a necessary and sufficient condition for this dissertation to exist. I thank them most heartily.

Special thanks are due to my employer – Makerere University, for granting me a study leave which ensured uninterrupted progress on the PhD program. More specifically, encouragement, support, inspiration and advice, by the Dean, Faculty of Economics and Management - Prof. J Ddumba-Ssentamu, is gratefully acknowledged.

My lecturers at the University of Dar es Salaam, Department of Economics and the Collaborative PhD Program (CPP) Joint Facility for Electives (2003), for a job well done in shaping my thoughts that ultimately led to this study. In addition, my appreciation goes to the staff of the Department of Economics and the Economic Research Bureau, University of Dar es Salaam, for providing me with a cordial working environment.

Let me also say 'thank you' to the following people: The then Director General of Medical Services, Ministry of Health, Prof. Francis Omaswa for clearing this study; The Medical Superintendents, Hospital Administrators, staff of Records Departments, in the sample hospitals for maximum cooperation in gathering the necessary data; the patients on whom data were collected; colleagues at the Faculty of Economics and Management, Makerere University, for the encouragement and support; Dr Rowena Jacobs, Prof. Peter C Smith and Dr Andrew Street, The University *of York*, for training in the measurement of efficiency in public sector organizations; Prof. Paul Wilson, University of Texas for providing me with the 'FEAR' software package; Prof. Bruce Hollingsworth, Faculty of Business and Economics, Monash University; Dr Kirigia, WHO Regional Office for Africa, Congo Brazaville for providing me with African healthcare facility efficiency studies; Dr. EA Zere, WHO Namibia; Pantaleo Kessy, University of Colorado; thanks also go to Geraldine G Duwe and Clara G Karungi for research assistance.

Support from colleagues in the CPP class of 2002 at the University of Ibadan (Nigeria), Yaounde II (Cameroun) and the University of Dar es Salaam (Tanzania) is gratefully acknowledged. Specifically, collaboration from the 'Dar Group' – Tchaka Ndlovu, Asmah E Ekow, Eria Hisali, Nassor S Nassor, Felician Mutasa, Maureen

Were and Jehovaness U Aikaeli, was invaluable for the completion of the program. I thank you guys most sincerely for being there for us fulltime.

Last but not the very least, I would like to thank my family: my parents, my wife Grace, children: Alice-Martha, Pamela-Ann, Patricia-Ann and Pantaleo. Your support in terms of patience, understanding, love and encouragement, has been invaluable and without it, this thesis would not exist. Heartfelt thanks go to my brothers: Emmanuel, Frank, Kizito, sisters: Dorothy, Zoe, Norah, Rose, Immaculate, Resty, Teddy, Agnes, and Carol, for all the support.

Finally, notwithstanding the support, encouragement and inspiration from the above persons and institutions, this study and the PhD program in general, would be impossible to complete without the wisdom, good health, and resilience from the Almighty God. I will thank you and praise you for ever.

The usual disclaimer applies. None of the acknowledged institutions and persons are responsible for any errors, inadequacies and omissions in this study. They remain my responsibility.

DEDICATION

I dedicate this work to my parents, who brought me to this world and have guided me; To the Almighty God, who has comforted me at all times; To my loving wife Grace, who has stood by me at all times; To our lovely children: Alice-Martha, Pamela-Ann, Patricia-Ann, and Pantaleo, who have been and are the joy of my life.

ABSTRACT

The study measures the technical efficiency and total factor productivity growth of 25 district referral hospitals from three regions of Uganda over the 1999-2003 period. This study is motivated by a desire to evaluate the ongoing health sector reforms in Uganda which in part are seeking to improve the efficiency of health services.

Nonparametric Data Envelopment Analysis (DEA) is used in the measurement of hospital technical efficiency whilst the DEA-Malmquist index is used in the measurement of hospital total factor productivity change. The Hospital Management Information System launched in 1997 is the source of the data for this study.

The results indicate the existence of different degrees of technical and scale inefficiency in Uganda's district referral hospitals over the sample period. There were productivity losses for the sample hospitals which are largely due to technological regress rather than technical inefficiency. Thus, changes in technology are needed if the hospitals are to become more productive, for instance through improved diagnosis tests, hospital information management.

The findings illustrate one of the advantages of the frontier efficiency technique, namely the ability to identify the degree of emphasis that should be placed on improving technical efficiency vis-à-vis technological change. The study adds to the existing literature on health facility efficiency but additionally incorporates patient deaths in the measurement of hospital technical efficiency. Additionally, heterogeneity in the patient load is controlled for via a length of stay-based case-mix index. Quality of care was incorporated into the analysis by means of patient deaths. Super-efficiency was conducted to further distinguish between the technically efficient hospitals. To construct confidence intervals for individual hospitals technical efficiency scores, nonparametric bootstrapping was conducted. The efficiency vectors yielded have ready uses by policymakers in the hospital sector. Indicators of the relative efficiency of hospitals are needed to gauge whether hospital cost-containment efforts are succeeding, amongst other uses.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background to the Problem

Health care costs are causing an increasing concern to many governments in the world. Governments and the general public are concerned that patients receive the appropriate level of care and that the care be delivered as efficiently as possible. A key component of health sector efforts to improve operating efficiency has to do with making the best use of existing resources (Parker and Newbrander, 1994).

Uganda's health care system increasingly faces critical resource constraints in its efforts to extend health services of acceptable quality to the vast majority of people. The shortage of health care resources is attributed to among other factors, rapid growth of the population, an upsurge in diseases such as malaria, poor macroeconomic performance, the HIV/AIDS epidemic and cutbacks in public spending.

Public hospitals are an important part of health systems in developing countries, and depending on their capacity, act as first referral, secondary or last referral facilities. These hospitals are generally responsible for 50 to 80 percent of recurrent government health sector expenditure in most developing countries (Barnum and Kutzin, 1993), and utilize nearly half of the total national health expenditure in many of these countries (Mills et al., 1990). Under the prevailing conditions of increasing health care costs, it is thus no surprise that "hospitals, as the main spenders within the healthcare system, are in the limelight" (Montoya-Aguilar and Marchant-Cavieres, 1994). In a bid to find new resources to fund the high cost activities of the hospitals as well as to utilize existing resources more efficiently, governments in some developing countries have started giving varying degrees of autonomy to public hospitals in the hope that this would both reduce the financial burden of hospitals on governments and strengthen the efficiency and effectiveness of public hospitals.

The overall objective of Uganda's National Health Policy (1999) is to reduce mortality, morbidity and fertility, and the disparities therein by ensuring access to the Uganda National Minimum Health Care Package. To achieve the aforementioned policy objectives, the government has implemented a package of reforms to support this policy by providing additional resources, subsidies, guidelines, reviews and undertaking some restructuring and decentralization.

The overall expected outcome would be an effective, efficient, responsive and accountable National Health Care System. The goals and objectives and indicators are related to amongst other things improving health status, decreasing morbidity and mortality, changing health risk behaviors, implementing PHC, improving quality in health services, hospital services, as well as equity in health resources. Each objective may have one or more input, process, output, outcome and/or impact indicators.

The hospital sub-sector is a large consumer of scarce health care resources. Over the period 1999-2003, hospitals in Uganda used on average 70 percent of the total public sector expenditure on health (see Appendix Table 1). Thus, the efficiency of hospitals merits close attention and scrutiny due to their enormous consumption of resources. It is against the aforementioned background that the present study sought to examine the extent to which district referral hospitals being major components of the national health care system, are technically efficient.

Although efficiency is accorded a central place in the health policies of most countries, in practice much remains to be done. The literature on hospital efficiency studies (for instance, Zere, 2000 and Kirigia et al., 2004) in sub-Saharan Africa indicates that in practice not much attention is given to efficiency by health care administrators. A lot of the attention of policymakers, donors and health care systems researchers appears to be focused on health sector reforms, prominent of which is the mobilization of additional resources for health care via user fees and other means of financing (Zere, 2000).

1.2 Statement of the Problem

The success of policy in guiding the hospital sector towards best-practice depends upon among other things the ability to distinguish efficient from inefficient healthcare providers. Over the 1999-2003 period, the allocation of resources to Uganda's hospital sub-sector averaged 70 percent of the total public expenditure on health (see Appendix Table 1). Out of the 70 percent of the total public expenditure which is consumed by the hospital sub-sector, approximately 45 percent was allocated to district referral hospitals. However, the extent to which district referral hospitals in Uganda efficiently turn their inputs into desirable health outputs and outcomes is not yet known. This therefore points to a knowledge gap with regard to hospital efficiency in Uganda. Moreover, with improved efficiency, district referral hospitals as well as other hospitals can provide more quality healthcare services even without the need to increase the resources devoted to health. This can be attained if hospitals seek to realize efficient quality improvement.

1.3 Objectives of the Study

The ongoing health sector reforms in Uganda amongst others sought to improve: efficiency; equity; health of the citizenry; and quality of healthcare and sustainability of health services. This study focuses on the broad objective of improving efficiency. Thus, the main purpose of this study is to measure the technical efficiency of district referral hospitals located in three[1] out of four administrative regions of Uganda. Specifically, the study sought to evaluate the technical and scale efficiency as well as assess total factor productivity changes of district referral hospitals in Uganda over the 1999-2003 period. This study focuses on technical efficiency because it is crucial for the realization of allocative and economic efficiency. Besides, it has not hitherto been investigated and can readily be evaluated using routinely collected hospital data.

1.4 Significance of the Study

Health care costs in most developed and developing economies have grown dramatically over the last few decades and it is widely believed to be mainly due to the inefficiency of health care institutions (Montoya-Aguilar and Marchant-Cavieres, 1994). The study is

[1] District referral hospitals in the Central, Western and Eastern regions were covered.

significant in a number of respects. The study investigates the notion of technical efficiency thereby providing insight into the optimal allocation of hospital resources. The measurement of hospital efficiency is of considerable significance to policymakers. First, it draws attention to the fact that the hospital sector may be able to achieve a higher level of overall goal attainment without increasing its resource inputs. Second, with the measurement of efficiency, it is then feasible to investigate exogenous determinants of inefficiency; it may be possible to identify for example, if low efficiencies are related to factors such as inappropriate human resource management (e.g. low remuneration for medical staff) or particular ways that hospitals are organized and financed. Third, the regular measurement of efficiency over time is important for monitoring the impact of policy reform aimed at increasing technical and allocative efficiency of hospitals. Fifth, this study controls for patient load heterogeneity by means of a length of stay-based case-mix index; provides for quality via in hospital mortality rates; along with the examination of super-efficiency and bootstrapping. Finally, for health insurers and government authorities, it is important to know whether a hospital is efficient in the sense of producing a certain bundle of services at minimum cost.

1.5 Scope

The study places a particular focus on district referral hospitals for two reasons. First, the decentralized nature of health care delivery in Uganda means that district authorities determine the priorities of their districts while the Ministry of Health dispenses supervisory and monitoring roles. Second, district hospitals consume a bigger proportion of the resources used by the hospital sub-sector compared to national and regional hospitals. The study covers the 1999-2003 period due to amongst others, the following ongoing policy changes which in part stress the efficient delivery of healthcare services:

(i) The decentralized delivery of social services since 1997;

(ii) The 10-year National Health Policy - (2000-2010);

(iii) The 5-year Health Sector Strategic Plan: 2000/01-2004/05; and

(iv) The Poverty Eradication Action Plan (PEAP) – launched in 1997 but revised in 2000 and 2004 – which is the overall national planning framework for Uganda.

1.6 Organization of the Study

The study comprises eight chapters. The next chapter briefly covers Uganda's health care delivery system. It examines the country's administration and economy; population, poverty and health status; institutional setting in health care delivery; performance of the country's health care system; and concludes by briefly looking at the policies of the health sector. Chapter three reviews the study's relevant literature while chapter four presents the conceptual framework. The methodology used by the study is presented in chapter five. There are two empirical chapters. Chapters six and seven respectively present empirical results on technical efficiency and total factor productivity growth. The last chapter provides concluding remarks on policy implications, caveats and suggestions for future research.

CHAPTER TWO

UGANDA'S HEALTH CARE DELIVERY SYSTEM

2.1 Administration and Economy of Uganda

Administratively, as of 2006, Uganda is divided into four regions namely western, eastern, central and northern which are further partitioned into 69 districts. Uganda has a decentralized system of governance and several functions have been ceded to the local governments. However, the central government retains the role of making policy, setting standards, and supervising. National security is also the role of the central government.

The economy is predominantly agricultural with the majority of the population dependent on subsistence farming and light agro-based industries. The country is self-sufficient in food, although the distribution is uneven over all areas. Coffee, tea, and cotton are the major earners of Uganda's foreign exchange. During the period immediately following independence, from 1962 to 1970, Uganda had a flourishing economy with a gross domestic product (GDP) growth rate of 5 percent per annum, compared with a population growth rate of 2.6 percent per annum. However, in the 1970s through the early 1980s, Uganda faced a period of civil and military unrest, resulting in the destruction of the economic and social infrastructure. This seriously affected the growth of the economy and the provision of social services such as education and health care.

Since 1986, however, the government has introduced and implemented several reform programs that have steadily reversed the setbacks and aimed the country towards economic prosperity. Consequently, between 1996 and 2000, the country's real GDP grew at an average rate of 6.2 percent per annum. This is far higher than the population growth rate, which was estimated at 2.9 percent. The GDP per capita grew at a rate of 2.6 percent per annum.

2.2 Population, Poverty and Health Status

In the past, most demographic statistics in Uganda were derived from population censuses, which started in 1948. Subsequent censuses have been held in 1959, 1969, 1980, 1991 and 2002. In addition, Demographic and Health Surveys have been conducted in 1988, 1995, and 2000. Additional demographic data have been obtained from small-scale surveys devoted to specific subjects.

Civil registration was made compulsory in Uganda in 1973. However, its coverage is incomplete and is therefore unsatisfactory as a source of demographic statistics. Efforts to streamline the system were made between 1974 and 1978, but the achievements that were realized were later frustrated by the economic and civil instability. Since 1995, an attempt has been made to revive the civil registration system in the country, but thus far, it has not reached a satisfactory level.

Table 2.1 presents several demographic indices compiled from the population censuses of 1948 through 2002. The table shows that over the period, the population increased more than fourfold. As of 2002, Uganda's population was 24.4 million with an annual growth rate of 3.3 percent. This places the country in the group of countries with the fastest growing populations in the world. Slightly over 12 percent of the total population was living in urban areas (2002 Uganda Population and Housing Census). The high growth rate is brought about by amongst other things high fertility and declining mortality levels (Table 2.1). The level of urbanization is still low but has been increasing over time.

Up to the late 1960s, there were more males than females in Uganda. This was mainly due to large numbers of male immigrants who came to the country to work at factories and plantations. In the mid-1970s these migrants left because of the deteriorating economic situation. Since then, the number of females exceeds that of males.

Table 2.1 Uganda: Selected Demographic Indicators; 1948-2002

	Census Years					
Indicator	**1948**	**1959**	**1969**	**1980**	**1991**	**2002**
Population (million)	5.0	6.5	9.5	12.6	16.7	24.4
Population increase (million)		1.5	3.0	3.1	4.1	8.0
Average annual increase (thousands)		143.0	300.0	282.0	367.0	686.0
Crude birth rate	42.0	44.0	50.0	50.0	52.0	49.0
Total fertility	5.9	5.9	7.1	7.2	7.1	6.9
Crude death rate	25.0	20.0	19.0	20.0	17.0	18.0
Urbanization (%)	3.1	4.8	7.8	8.7	11.3	12.2
Population density	25.2	33.2	48.4	64.4	85.0	126.0

Source: Uganda Demographic and Health Survey 2001; 2002 Uganda Population and Housing Census

Health status and economic growth are key determinants of human welfare. They are so interrelated that it is impossible to generate economic growth in the developing world without solving the central health problems faced by these countries. It can be conjectured that it is not possible to improve health status of communities without generating economic growth (Sachs, 2002); unhealthy people are poor people and vice versa.

High population growth tends to overburden the development process. Despite the decline in income poverty from 56 percent in 1992 to 38 percent in 2002, the total population below the poverty line dropped by about 6 percent from 9.8 million to 9.2 million over the same period. The per capita GDP in real terms, increased by 3.7 percent between 1991 and 2002 (2002 Uganda Population and Housing Census). The poverty situation is examined by looking at the results of the 1999/00 and 2000/03 Uganda National Household Surveys. Poverty increased nationwide between 1999/00 and 2000/03 in the Uganda National Household Surveys irrespective of the poverty indicator used. The percentage of people living below the poverty line rose from 34 percent to 38 percent (Uganda Bureau of Statistics, 2004).

Poverty rose in both urban and rural areas between the 1999/00 and 2002/03 Uganda National Household Surveys. In rural areas, there was no growth in consumption and the proportion of people in poverty increased from 37 percent to 42 percent, which corresponds to between 7.0 million and 8.5 million persons in poverty. In urban areas, the corresponding increase was from 10 to 12 percent, recording an increase in absolute numbers of the poor from 0.3 million to 0.4 million (Uganda Bureau of Statistics, 2004). Although rural areas remain markedly poorer in living standards than urban areas, the proportionate rise in poverty is actually higher in urban areas. This is partly due to the increase in internally displaced persons due to insecurity in the northern region of the country.

2.3 Health Sector Policy Environment

The overall policy environment for health has been affected by policies related to poverty reduction, civil service reform, and political decentralization. These policies are around the topics of poverty reduction, health sector development, financing, and organizational changes and are interlinked to each other. Nevertheless, different policies often arise from the same pressures and address shared concerns; their effects overlap and sometimes conflict; and the relationships among them are synergistic.

Uganda started implementing health sector reforms in 1987 in the form of broad decentralization; broadening health financing and the introduction of user charges and later community pre-payment schemes; working with private Not-for-profit and private health care providers and also encouraging the autonomy of public hospitals; planning and resource allocation systems (bottom-up versus top-down practice); and finally human resource management systems under which there was retrenchment, pay reform, transparent remuneration structures and decentralized human resource management.

The Primary Health Care (PHC) concept which was introduced at the 1978 Alma Ata conference was later adopted by Uganda as the focus of health system development. However, its implementation was hampered in the early 1980s by continued bad governance and civil strife. As a result, the health system was in shambles by 1986. Due to the failure of the public system to provide for the health care needs of the populace, the private providers had entered the health care market with the associated inequities and inequalities of myriad forms with a resultant lack of recognizable PHC activities. Therefore, Uganda did not perform as expected in implementing the PHC objectives, goals and strategies agreed on in Alma Ata in 1978. In 1986, the Expanded Programme on Immunization (EPI) was re-launched, the Maternal and Child Health (MCH) program, Family Planning and the AIDS control programs were introduced. The early 1990s were characterized by the implementation of the health sector reforms.

The health sector has undergone rebuilding, reorientation, and reform since 1986. In the 1980s, health sector development focused on rebuilding the collapsed health system. This was followed in the early 1990s by a reorientation of health services from an overemphasis on specialized curative care towards preventive and primary health care services. Since the late 1990s, the reforms have focused on improving service coverage, especially for preventive and primary health care services, under a decentralized health system. The current reform is guided by the 10-year National Health Policy and the National Health Sector Strategic Plan II and the Poverty Eradication Action Plan (PEAP).

The national health policy further decentralized health service delivery from the district to the Health Sub-District (HSD). The HSD - an innovative creation by the National Health Policy are functional sub-divisions of the district health care system. They seek to: increase equity of access, improve the quality and management of routine health care service delivery and foster community involvement in the planning, management and delivery of basic health care services.

Administratively, a HSD is equivalent to a county with a population of over 100, 000 persons. The HSD strategy redefined the roles and responsibilities of the districts in the delivery of health care services. HSDs are charged with the responsibility of operational planning and management of health care delivery; implementation (health care service delivery); monitoring and supervision of all basic health care services in the HSD; offering emergency surgery (particularly emergency obstetric surgery) and treating referred patients from lower health care units (Republic of Uganda, 2003).

However, in order to get the full benefits from HSDs, measures must be taken to ensure that HSDs have the management capacity to carry out their responsibilities for Primary Health Care. Along with growing autonomy and responsibility for HSDs is the need for the identification of constraints which need to be resolved at a higher level (e.g. human resources) and capacity building.

Although decentralization is intended to promote efficiency, accountability, transparency, community participation and sustainability, these intended goals in the health sector are yet to be seen. The lack of effective supervision is an indication that mechanisms for technical accountability for service provision need to be strengthened. Ministries of health in decentralized systems often face constraints in carrying out their oversight role. Uganda has not yet shown that it has effective mechanisms in place to carry out its oversight role on a regular basis.

However, despite the progress on health systems development and positive trends in economic growth, health status indicators reveal a mixed picture. The country has been able to reverse the HIV/AIDS epidemic, reducing HIV prevalence from 18 percent in the early 1990s to six percent in 2003. On the other hand, while maternal mortality fell from 523 to 505 per 100,000 live births over the 1995-2004 period, there was virtually no improvement in infant and child mortality, which stand at around 100 and 150 per 1,000 live births, respectively. Similarly, there has been little improvement in children's nutritional status since 1995 while the total fertility rate remains high at an average of 6.9.

The financing of health care is still at a very low level, particularly with the largest share from private out-of-pocket expenditure. The government needs to further expand its current financing strategy to encourage the development of health insurance schemes and explore other options for risk pooling. Additionally, given that some healthcare services are provided by non-public sectors, the government needs to increase its role in regulatory and oversight to ensure quality of services and access by the poor and vulnerable groups.

2.4 Structure of Uganda's Health System

A health system, defined to include "all the activities whose primary purpose is to promote, restore or maintain health" (WHO, 2000) has three essential objectives: (i) improving the health of the population it serves; (ii) responding to people's expectations; and (iii) providing financial protection against the costs of ill-health. The first objective deals with ensuring better health, while the other two are concerned with allocative efficiency and equity. Thus, the main functions of the health systems are concerned with investing, delivering and financing health services. Since much of the health sector reform aims to change these aspects of the health system, it is imperative that any changes made by decentralization reforms ought to improve health system performance, or at the very least not have detrimental effects.

The financing and delivery of health care services in Uganda is undertaken by the government, the private sector and the nongovernmental sector. Government intervention in Uganda's health sector can be categorized into both healthcare financing and delivery. Government intervention in health does not imply that the government in all cases both finances and delivers the services itself. The National Health Policy (1999) and the Health Sector Strategic Plan (HSSP) emphasize the central role of Public-Private Partnership in Health (PPPH) in the financing as well as delivery of health care services. Table 2.2 highlights the different possible combinations of financing and provision, indicating where in Uganda these combinations can be found.

Table 2.2 Uganda's Public-Private Partnerships in Health

		Providers	
		Public	***Private***
Financing	***Public***	Public health facilities	Facility-based PNFP* Non-facility based PNFP (e.g. medical bureau) Private practitioners (subsidies, microfinance, commercial marketing strategies)
	Private	Public health facilities (user charges, private beds)	Traditional & complementary medical providers Private health practitioners Facility-based PNFP Non-facility based PNFP

*PNFP = Private-Not-For-Profit
Source: Republic of Uganda (2003); The Health Sector PEAP Review Paper, Ministry of Health.

Health services in Uganda are mainly financed by three sources: private out-of-pocket spending, government allocations, and external aid, with the private funds contributing almost 60 percent of the total health expenditure. In the fiscal year 2002/03, total per capita expenditure on health was estimated to be about US$ 17.1 out of which US$ 9.9 was out-of-pocket spending, US$ 3.9 (23 percent) from the government and the remaining US$ 3.3 (19 percent) funded by donors. Expenditure on health in Uganda has increased in recent years. However, its level remains very low, particularly in terms of public spending (World Bank, 2005).

Health care in Uganda is financed and delivered by various providers through a public-private partnership. The Public-Private Partnership in Health (PPPH) comprises of public health care providers, Private-Not-For-Profit and the Private Health Practitioners (PHPs); as well as traditional and complementary medicine practitioners. Strengthening the collaboration and partnership between the public and private sector in health is an important guiding principle of the National Health Policy. These are discussed in what follows.

(a) Public Health Care Providers

Government hospitals are divided into three categories: national referral; regional referral; and district/rural/general referral hospitals. District/rural/general referral hospitals are staffed with general doctors. Regional referral hospitals have specialists in a limited number of fields and are also teaching hospitals. Finally, there are two national referral hospitals, namely Mulago Hospital complex and Butabika Hospital located in

Kampala district. These are also teaching/research hospitals which provide comprehensive specialist services.

The health centers throughout the country are graded as Health Center II, Health Center III, and Health Center IV. The grading depends on the administrative zone served by the facility. They provide different types of services; however, a unit can work as HC II and III or IV. If a facility has more than one grade, the highest is considered.

An HC II serves a population of approximately 5,000 people. It provides outpatient care, antenatal care, immunization, and outreach. An HC II is supposed to be staffed by one enrolled nurse, one enrolled midwife, and two nursing assistants. According to the HSSP, all HC IIs provide community-based preventive and promotive health service.

An HC III serves a population of approximately 20,000 people. It provides all the services of an HC II plus inpatient care and environmental health. It is usually staffed by one clinical officer, one enrolled nurse, two enrolled midwives and one nursing assistant, one health assistant, one laboratory assistant, and a records officer. According to the HSSP, all HC IIIs provide the services offered in an HC II plus maternity services, inpatient health services, and laboratory services.

An HC IV or the district referral hospital serves a population of approximately 100,000 people and is the focus of this study. Additionally, it is the headquarters unit of the health sub-district. With a Ministry of Health prescribed staff mix,[2] it provides all the

---

[2] Each HC IV should have at least one medical officer, two clinical officers, one registered midwife, one enrolled nurse, one enrolled midwife, one registered comprehensive nurse, two nursing assistants, one laboratory technician, one laboratory assistant, one health inspector, one dispenser, one public health

services of an HC III, plus surgery, supervision of the lower-level units, collection and analysis of data on health, and development of plans for the health sub-district. The HSSP requires all HC IVs to provide emergency surgery and blood transfusion services as well as all the services offered at HC IIIs. Table 2.3 presents the availability of health care facilities in Uganda as at 2002.

Table 2.3 Availability of Health Facilities in Uganda, 2002

Region	Population	Hospitals	HCIV	HCIII	HCII	Total	Beds	Population per bed
Central	6683887	36	47	212	1141	1436	9006	742
Eastern	6301677	20	45	223	306	594	5763	1093
Northern	5345964	25	25	137	257	444	6156	868
Western	6417449	23	42	215	321	601	5847	1098
Total	**24748977**	**104**	**159**	**787**	**2025**	**3075**	**26772**	**Av. Pop. per bed: 950**

Source: Uganda Bureau of Statistics (2005): 2004 Statistical Abstract

(b) Private-Not-For-Profit (PNFP) Health Care Providers

The private not for profit category of health care providers is motivated by concern for the welfare of the population. The PNFP comprise agencies that provide health services from an established static health unit/facility to the population and those that work with communities and other counterparts to provide non facility-based health services.

Conversely, the non facility-based PNFP comprise the majority of local and international organisations working in the health sector commonly referred to as NGOs. They work with counterparts such as government, facility-based PNFP providers, private practitioners and communities. Their contribution is in areas ranging from social

dental assistant, one aesthetic officer, one assistant health educator, one records assistant, one accounts assistant, and two support staff.

awareness and advocacy to more specific aspects of service delivery. The area of emphasis tends to conform to agency expertise such as special disease programmes, technical assistance, training, capacity building, emergency and relief services and mainstream service delivery.

(c) Private Healthcare Practitioners (PHP)

This sub-sector encompasses all cadres of health professionals in the Clinical, Dental, Diagnostics, Medical, Midwifery, Nursing, Pharmacy and Public Health categories who provide private health services outside the Public, PNFP and the Traditional and Complementary Medicine establishment. Private health care practitioners provide mainly primary level services and limited secondary level services. In addition, some urban units offer tertiary and specialist care.

(d) Traditional and Complementary Medicine Practitioners

Traditional healers include herbalists, spiritual healers, bone-setters, traditional birth attendants, hydro therapists, traditional dentists amongst others. There are several associations with registered members at the sub-county and district levels, coordinated by cultural officers even though they remain unaffiliated to any association.

The low ratio of traditional practitioners and university-trained doctors in relation to the entire population in Uganda is revealing as is true for many African countries. There is at least one traditional healer for nearly 290 persons compared to one western-trained medical practitioner for every 10,000 people in the urban areas and 50,000 people in the rural areas, respectively (World Bank, 2003). This gross difference in numbers implies that traditional healers are usually the first line of contact for most patients with the health care system. Thus, traditional healers are an integral part of the local culture and sustainable sources of health care and knowledge on illness.

2.5 Performance of Uganda's Health Care Sector

The performance of Uganda's health care system is briefly analyzed by looking at various indicators. These include causes of morbidity, immunization rates for antigens, and access to health care facilities. These are healthcare system proxy outputs which ultimately seek to improve upon the health status of the population.

(a) Causes of Morbidity in Uganda

Health is defined in the World Health Organization's (WHO) constitution as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity," (United Nations, 1946). However, through time, this statement has been modified to include the ability to lead a "socially and economically productive life." The WHO definition is not without criticism, as some argue that health cannot be defined as a state at all, but must be seen as a process of continuous adjustment to the changing demands of living and of the changing meanings one gives to life. The WHO definition is therefore considered by many as an idealistic goal rather than a realistic proposition. The level of morbidity is one indicator that is used to analyze the health status of a population. The Uganda National Household Surveys collect information that monitors this indicator. Table 2.4 indicates the major causes of morbidity between the 1999/00 and 2002/03 National Household Surveys. The table shows that malaria accounts for more than 50 percent of all persons who fell sick. Respiratory infections ranked second as a cause of illness and its share rose by two percentage points (from 12 to 14 percent) between the two surveys. Intestinal infections and diarrhoea follow respiratory infections and their respective shares fell between the two surveys.

Table 2.4 Major Causes of Morbidity in Uganda (Percentage)

Disease	*1999/00*	*2002/03*
Fever/Malaria	56.0	56.0
Respiratory Infections	12.0	14.0
Intestinal infections	6.0	4.0
Diarrhoea diseases	5.0	4.0
Skin diseases	3.0	3.0
Other illnesses	18.0	19.0

Source: Uganda Bureau of Statistics (2005); 2004 Statistical Abstract

Apart from the heavy burden of infectious disease, Uganda is also simultaneously experiencing a marked upsurge in the occurrence of non-communicable diseases such as hypertension, cancer, diabetes, mental illness and chronic heart disease. Uganda has therefore, already entered the early phase of the epidemiological transition. While infectious diseases must be given priority, selective attention needs to be given to all the key determinants of ill health in Uganda, including unhealthy lifestyles and the rising toll of accidents (Uganda Bureau of Statistics, 2004).

(b) Immunization Rates for Antigens

The countrywide immunization rates for such antigens as Bacille Calmette Guérin (BCG), measles, polio and Diphtheria, Pertussis, and Tetanus (DPT) over the period 1998-2002 are shown in Table 2.5. As indicated, BCG recorded the highest immunization rate. Generally, there has been an upward trend for all antigens save for tetanus for pregnant women, which showed a downward trend over the period under review.

Table 2.5 Uganda's Reported Routine Immunization Coverage (%); 1998-2002

Vaccine	1998	1999	2000	2001	2002	2003
Bacille Calmette Guérin	85	93	83	90	102	96
Diphtheria, Pertussis, and Tetanus	56	60	57	60	72	81
Polio	57	62	57	61	73	82
Measles	59	63	61	62	77	83
Tetanus Toxoid (Pregnant)	48	50	45	45	50	48
Tetanus Toxoid (non pregnant)	7	9	7	5	4	10

Source: Uganda Bureau of Statistics (2005); 2004 Statistical Abstract

(c) Access to Health Care Services

The 2004 National Service Delivery Survey indicates that health care services were accessed from conventional static health units, mobile/outreach arrangements and sometimes from other non-traditional health institutions like schools. The facilities were largely owned by the government, the private for profit sector and private not-for-profit sector. The survey revealed that static government facilities remain the major source of general health care services, immunization services and birth-related services compared to other sources. For all illnesses reported, government health care facilities provided the first source of treatment as indicated in Table 2.6.

The efficiency of public health care system is reflected in the utilization of the services by the people for whom they are intended. The absence of a proper referral system leads to the paradoxical co-existence of under-utilization and crowding. The purpose of a referral system is simple and straightforward. The primary considerations are costs, efficiency and quality at all levels (primary and secondary) of medical care. Gate keeping is one way of ensuring proper referral system in the public sector. As per this system, the doctors at the Primary Health Centers (PHCs) must act as gate keepers for more specialized and more expensive healthcare services. Patients are required to have a referral from the PHC doctor to be able to use specialized healthcare services. In order to improve patient compliance, fast tracking can be employed as a supplementary strategy. Fast tracking would enable the referred patients to get preference over others and the gain to them will be in the form of reduction in waiting time. Referral system exists in European countries such as Denmark, Norway, Spain, the Netherlands and UK (Kulu-Glascow et al., 1998).

Table 2.6 Distribution of Persons who Fell Sick by First Source of Treatment, 2004

Type of Health Facility	Percentage using facility as first source of treatment
Government health facility	33.0
Private health facility	28.6
Pharmacy/drug shop	17.8
Home/self medication	10.6
None	3.8
Religious mission facility	2.7
Traditional healer	1.1
NGO health facility	1.0
Other	0.9
Community health workers	0.4
Total	**100.0**

Source: Uganda Bureau of Statistics (2005); 2004 National Service Delivery Survey

In Uganda, the referral system exists only for surgery and uncommon diseases. However, significant bypassing of the system is present with regard to common diseases for instance malaria due in part to the lack of user fees. The situation is compounded by the general lack of user fees at public healthcare facilities. By sending ‘price signals’, (Varatharajan, 1999), user financing can make the referral system work better by enhancing the cost effectiveness because highly expensive technology and personnel are reserved for more complex and referred illnesses. When the monetary prices are close to zero uniformly across the whole health care system, there will be a tendency among healthcare consumers to converge on the most expensive health service even for minor illness because there is no incentive for not doing so. The 2004 National Service Delivery Survey further notes that the average distance to the nearest government health care facilities where people sought first treatment was 5.2 km in the rural areas while in the urban areas it was 2.9 km as indicated in Table 2.7.

Table 2.7 Average Distance (km) to the Nearest Health Facility by Region, 2004

Region	*Rural*	*Urban*
Central	6.0	4.1
Eastern	5.1	2.1
Northern	4.9	2.4
Western	5.1	2.9
Average[3]	***5.2***	***2.9***

Source: Uganda Bureau of Statistics (2005); 2004 National Service Delivery Survey

The survey further analyzed the distance to health facilities where people sought first treatment by range of distance to establish persons who traveled distances longer than the target of 5 km set by the Poverty Eradication Action Plan. Table 2.8 presents the percentage distribution of people who traveled distances equal to or longer than the target of 5 km by region.

Table 2.8 Percentage Distribution of Population by Distance to Health Facility where First Treatment was Sought by Region, 2004

Region	**5 km or less**	**More than 5 km**
Central (minus Kampala)	80.8	19.2
Eastern	82.7	17.3
Western	78.1	21.9
Northern	78.9	21.1
Kampala	89.6	10.4
Average	***80.5***	***19.5***

Source: Uganda Bureau of Statistics (2005); 2004 National Service Delivery Survey

It is apparent from Table 2.8 that approximately 20 percent of the population sought first treatment from health facilities which were located more than 5 km away.

[3] Weighted by the number of health facilities in each region.

2.6 Conclusion

Uganda's healthcare indicators are far from perfect in light of the increasing challenges of rising disease burden and caring for a rapidly growing population coupled with resource constraints. This, therefore, underscores the need for efficiently utilizing the available resources to maximize healthcare outcomes and ultimately improved health status for the population.

CHAPTER THREE

PRODUCTION THEORY FOR HOSPITALS

3.1 Introduction

This chapter presents the conceptual framework within which the measurement of hospital technical efficiency takes place. The economic theories of hospital behavior are considered, followed by the performance framework of hospitals, along with the definition and measurement of hospital inputs and outputs. Risk adjustment of hospital outputs via case mix adjustment is also examined along with adjusting hospital production for quality of care.

3.2 Economic Theories of Hospital Behavior

In the classic, competitive model of the firm, the concept of profit maximization provides an effective postulation about the behavior of business firms. Most hospitals, however, are not profit-oriented enterprises and thus the assumption of profit maximization does not shed much light on hospital behavior (Lee, 1971). As a consequence several attempts have been made to develop a theory of behavior for nonprofit hospitals.

Several theories seek to explain the behavior of nonprofit hospitals. While all of these theories assume that some constituency controls the nonprofit hospital and uses this control to pursue some objective, the theories differ in the identity of the controlling constituency and the objective pursued. One theory argues that independent, nonprofit hospitals behave as consumer cooperatives (Lynk, 1994). According to this theory, community representation on a nonprofit hospital's board of directors ensures that the nonprofit hospital will set competitive prices even when the nonprofit hospital possesses market power.

Several other theories assume that hospital administrators control nonprofit hospitals. The first of these theories assumes that hospital administrators seek to maximize the hospital's output. Thus, this theory predicts that a nonprofit hospital would not exercise market power since doing so would only reduce the number of patients served (Baumol, 1967). The second of these theories assumes that hospital administrators obtain utility from both the quantity and the quality of a hospital's output (Newhouse, 1970). The desire to increase the quality of the output may prompt a hospital to offer a higher quality of service than consumers want. Thus, in this theory, nonprofit hospitals may exploit market power in order to subsidize services that are not economically viable.

One final theory argues that nonprofit hospitals are controlled by their medical staffs (Pauly and Redisch, 1973). In this theory, whether the hospital exploits market power depends on the hospital's policy toward granting staff privileges. If physicians can freely enter and obtain staff privileges, then both hospital care and physician care will be provided competitively. A hospital's medical staff generally has an interest in keeping prices low because an increase in the price of hospital care in an area reduces the demand for physician services in that area.

Yet other authors such as Rice (1966) have proposed a sales maximization model for the hospital which is similar to Baumol's sales revenue maximization hypothesis. Conversely, Newhouse (1970) suggests that hospital decision-makers have both quantity of output and prestige in their maximand. Moreover, Lee (1971) proposes a conspicuous production theory of hospital behavior. Lee's theory is based on the propositions that the preferences of hospitals are interdependent and certain inputs are acquired for purposes other than meeting the requirement of ordinary production (this represents the type of production function that will be adopted by a firm that is motivated by profit maximization).

Most of Uganda's public hospitals do not sell their output on a pro-rata basis. Most revenue is grant, funded out general taxation. Healthcare providers in Uganda also enjoy a quasi-monopoly status. They may not hold property rights to any cost savings they generate. All this means that providers in Uganda's public healthcare have more in common with government 'bureau' (Niskanen, 1971) than with either neoclassical firms of the economic textbook or the models of healthcare institutions developed for the American private healthcare sector. The 'bureaucratic' nature of Uganda's healthcare providers raises a range of efficiency-related issues for the funding agency. The issues are whether healthcare providers are (i) technically efficient; (ii) allocatively efficient; (iii) vertically and horizontally equitable; as well as (iv) producing the mix of outputs which society values. This study thus contributes to knowledge by investigating the technical efficiency and total factor productivity growth of district referral hospitals in Uganda.

The application of microeconomic principles to understanding the economic behavior of hospitals is not a straight forward exercise due to the complex nature of healthcare as a service or product. Health is uncertain and is dependent on a myriad of factors, many of which are beyond the control of the healthcare provider or the purchaser (Scott et al., 2001).

Nevertheless, economic thinking provides one common goal: efficiency, or getting the most from available resources. For instance, a hospital administrator is faced with the challenge of organizing resources in order to meet the organization's goals and objectives. The relationship between the range of productive inputs used and outputs produced can be characterized by a production function, which depicts the maximum amount of product that is obtainable from any specific combination of resources or inputs used in the production of a product or output. By identifying the relationship between outputs and inputs, one can find the combination of inputs and outputs which maximizes economic return (Scott et al., 2001).

However, the production process of hospitals is complicated both conceptually and empirically. Measuring the productive process in hospital care is complicated by the fact that the patient is both an input as well as an output in the process (that is, the patient's health status is a function of factors determined outside the hospital, for instance, genetics and lifestyle). Efficiency measurement in the health care sector is complicated by the nature of the production process. The measurement of the ideal output, improved health status, is difficult both conceptually and empirically (Grosskopf and Valdmanis, 1987). The complications are due to the fact that health status is a function of various factors, many of which are exogenous to the health sector – for instance education, the income of the household, and decisions within the household.

The hospital is unique conceptually, as its 'product' is not defined clearly and measurably. There have been six basic approaches to define hospital output in the literature (Berki, 1972): (i) patient days or number of discharges; (ii) hospital services; (iii) episodes of illness; (iv) health status level; (v) intermediate outputs; and (vi) composites of one or more of the previous measures. These approaches are deficient for they are not accepted universally or clear in their crystallization of the uniqueness of the health care process. Currently, treated case (patient discharge) has been used as one measure of hospital output. In accordance with the needs of output dimension in a meaningful way to enable inter-and intra-hospital comparison, the treated case (patient discharge) has been emerged as an accepted measure.

Hospitals may not fit into the economist's standard notion of a firm which gives rise to a host of challenges in properly estimating the cost function of hospitals. It is arguable to presume a definition and recognition of hospital output. If it is allowed (extending Grossman, 1972) that people utilize hospital services because their health stocks have fallen below some critical levels, then, the restoration of the health stock of its patients ought to be regarded as the outputs of hospitalization (Breyer, 1987 and Ellis, 1992). Ellis (1992) elucidates that measuring the improvement in health status is unrealistic, as health status is a multifaceted concept. Consequently, it is not easy to define and measure, in an operationally feasible manner, much less to compare and aggregate across patients.

Notwithstanding the difficulties of conceptualizing hospitals as firms, it is clear that the same economic processes which go on in firms also occur in and around hospitals. Decisions are made as to what services to produce, and how, what inputs to use, and how much to pay for the inputs/charge for the services. Are the "right" sorts of hospital outputs being produced, or are people being over-hospitalized, provided with unnecessary care? Are there too many surgical procedures, unnecessary diagnostic tests, overly prolonged hospital stays? And do hospitals use the least-cost mixes of labour and other inputs in producing their services? All these are the standard questions of economic performance except that in health care, the issues of appropriate quantities and mixes of servicing relate back to health care needs rather than to the conventional criterion of consumer willingness-to-pay.

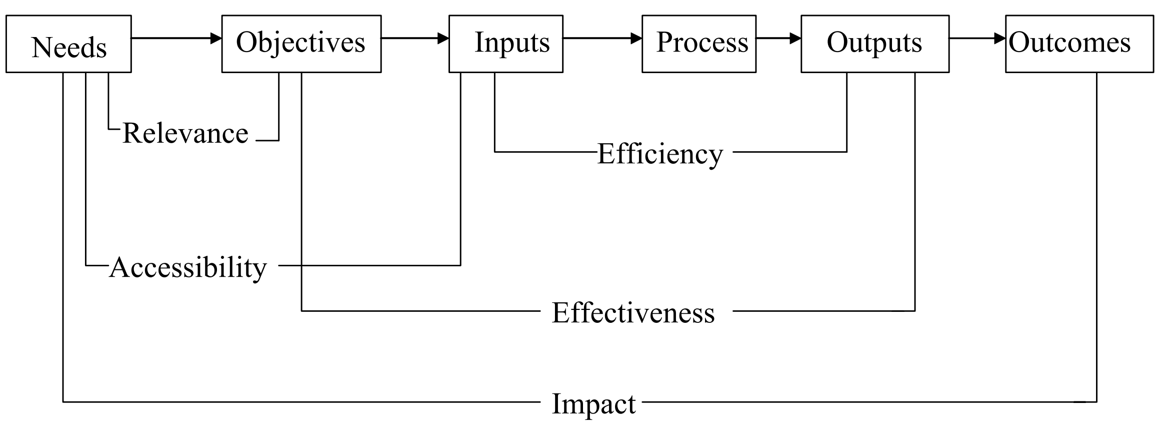
Empirical researchers have used measures of throughputs or intermediate outputs, in the number of cases treated, of patient-days served per hospital department, and of outpatient visitors. This strategy creates a new set of problems related to the homogeneity of hospital output. Two aspects that have received widespread attention are, (i) case-mix; and (ii) quality of care. These issues are covered in sections 3.4 and 3.5, respectively.

3.3 Performance Framework for Hospitals

Systems theory can be employed in the analysis of hospital production. A hospital can be viewed as a system from various perspectives. It is a physical system of buildings; it is a system of many interacting staff; it comprises a complex logistics system; it is a system for treating patients as well being an information system (Aldred et al., 1971). In approaching any given hospital any one of these systems perspectives might be most appropriate. However, the study chose to emphasize the hospital as a system for treating patients. A hospital is a system for making people well following illness or accident. It consists of a number of sub systems to achieve that purpose which include but are not limited to: the surgery system; clinical system; nutrition system; and care system.

Systems theory has had a significant effect on management science and the understanding of organizations. A system is a collection of parts unified to accomplish an overall goal. A system can be seen as having inputs, processes, outputs, outcomes, as well as impact. Systems share feedback among each of these aspects. The components of a simple system and the relationships between them are shown in Figure 3.1.

Figure 3.1 Performance Framework for Hospitals



By means of the schematic flow in Figure 3.1, one can look at a health care service organization. Inputs would include resources such as raw materials, money, technologies and people. These inputs go through a process where they are planned, organized, motivated and controlled, ultimately to meet the organization's goals of improving upon the health status of its patients.

3.3.1 Inputs

In a hospital setting, inputs are the resources needed to carry out a process or provide a health care service. Inputs required in health care are usually financial, physical structures such as buildings, supplies and equipment, and personnel (doctors, dentists, nurses, and other hospital staff).

The health care system in general and hospitals in particular utilize a variety of factor inputs in the provision of health care services. The inputs can be broadly categorized into two groups, capital and labor. By its very nature health care provision is a labor-intensive undertaking which involves the physical and mental skills of doctors, nurses, technicians, and administrators among other personnel.

Physicians play a significant role in the provision of hospital care. Cowing, Holtmann and Powers (1983) stress that although doctors supply what may be considered indispensable inputs in the treatment of patients in a hospital setting; they are often paid separately, either by patients or health insurance companies. Costs of their services are usually not reflective of production factors in hospital cost figures.

Physicians enjoy privileged relationships with their patients, which allow a wide degree of latitude in treatment procedure choice. The inputs used in this study included: beds; doctors; nurses and other employees.

3.3.2 Process

Process refers to a series of actions (or activities) that transform inputs (or resources) into a desired product, service, or outcome. In a health care delivery setting, these would include amongst others: training in Voluntary Counseling and Testing for HIV/AIDS, retention of revenue by hospitals, training of Primary Health Care nurses, and stock control of drugs.

3.3.3 Outputs

Output relates to the direct result of the interaction of inputs and processes in the system; the types and quantities of goods and services produced by an activity, project, or program. Where it is difficult to provide a direct measure of how changes in the hospital affect the problems being addressed, 'intermediate output' measures are often used instead. These can be measured more easily, but only give an indirect picture of how the health status of the population is affected. For a hospital, output includes but is not limited to antenatal care, counseling, and immunization coverage of children under 1 year as well as facilities offering Primary Health Care packages.

Defining and measuring hospital output is problematic. This is because the typical hospital provides a wide range of services to patients with a variety of different conditions. Additionally, teaching hospitals provide both teaching and research activities. Therefore, the hospital is actually a complex multi-product firm. One approach to this problem in hospital production function studies has been to use a single index of the services provided. The other approach has been to estimate a cost function instead, and treat output as a multidimensional vector.

In the measurement of hospital output a crucial distinction is whether the output is the actual provision of the medical treatment itself or the resulting improvement in patients' health status assessed by Quality Adjusted Life Years (QALYs) for example, as discussed by Butler (1995). It could be misleading to use intermediate outputs, such as patient days, laboratory procedures or surgical operations, as measures of output. Possibly, the most successful approach for determining the final products of hospital inpatient care is the Diagnostic Related Groups (DRGs) classification system, which considers a given hospital's case mix. The DRGs are a patient classification scheme which was originally developed as a means of relating the type of patients a hospital treats (i.e. its case-mix) to the costs incurred by hospitals. The design and development of the DRGs began in the late 1960s at Yale University (Fetter et al., 1980). Although the DRG system has been evolving for several years, DRGs still suffer from certain weaknesses; it is deemed a serious shortcoming that some of the DRGs manifest significant heterogeneity in resource use because they do not explicitly consider the severity of illness of the patient (Horn et al., 1985; Averill et al., 1992). Nevertheless, the effect of this bias is difficult to clarify when the DRG system is used as a method for aggregating hospital inpatient treatment into a single output measure.

A unit of output measurement commonly adopted is the patient day, with total output over a given period, then being taken as the total number of patient days provided over that time period. This is considered as an alternative to measure output, by using the number of cases treated. Feldstein (1967) states that "we must choose between two basic units of output: the case and the patient week". Lave and Lave (1970) employ the patient day as the unit of output measurement, arguing that although “a more relevant measure is the number of cases treated. The question of whether patients or patient-days is the better measure cannot be settled a priori.”

Some authors argue that the treated case is a priori, a more defensible unit of output measurement than the patient day. Indeed, the latter is more like an input measure relating to the time dimension of the production of a treated case. The number of patient days, which a hospital uses to produce a treated case, indicates the time period over which production of one unit of output takes place, for it does not measure the output itself. However, Feldstein (1969) disagrees and shows that there is no clear line to divide inputs and outputs, in the context of efforts to improve the level of community’s health.

The treatment conception of hospital output accords with the view of hospital output, as an intermediate product used as an input into the production function for health. This view is summarized by Berki (1972): “If we consider that the final product of the medical care process is the provision of the highest attainable health, given the state of the art, it is clear that the output of hospitals is more precisely an intermediate input into this process." Nevertheless, whatever position is adopted on these conceptual issues, empirical reality usually dictates the choice of hospital output.

It should be noted that the treated-case has been argued to be a more defensible unit of output than patient-days for treatment provided in hospitals. The argument does not necessarily apply to other institutions which may be providing a different type of output. In-patients in nursing homes do not generally receive treatment for a specific illness or illnesses, but are assisted or cared for in matters of everyday living (cooking, bathing, and so on). For such institutions a strong case can be made, that a day of care is the unit of analysis, not a treatment. This can give rise to problems in measuring hospital output, if nursing home patients are being cared for in hospitals (along with the usual patients requiring acute care). In discussing the concept and measurement of hospital output, this study briefly alludes to the fact that treatments may differ for different illnesses. Under these circumstances, a treated case may not be a homogeneous unit of output even within a hospital. This is another important dimension of the multi-product nature of the hospital. This study used the following hospital outputs: admissions; outpatient department attendances; surgical operations; deliveries and patient deaths (undesirable output).

3.3.4 Outcomes or Impact

Outcomes or impact try to answer the question: what are the final products of the organization? This area of performance is always difficult to measure with respect to hospitals as there is often little agreement on the ways in which health outcomes (that is, the change in health status of a person having been to a hospital) can be measured. There are a few direct measures, but proxy measures are also employed.

Outcome refers to the consequence of a process, including outputs, effects, and impacts. This can for instance be malaria prevalence, HIV/AIDS infection rates, prevalence of underweight children, safer lifestyles and behavior, reduced mental disorders and Tuberculosis cure rates.

Alternatively outcome may relate to impact which implies a change in the status (e.g., health, standard of living) of individuals, families, or communities as a result of a program, project, or activity. For example, the impact of an immunization program might be improved health status indicators which take the form of reduced infant mortality rate, maternal mortality rate and under five mortality rates, by a given percentage.

Health care is an intermediate input in the context of health production, but an output in the context of service provision. Some measures of efficiency in health services are formulated to compare resource use against provision for instance number of patient cases or consultations with service providers, instead of the resulting health benefits to service recipients. (Peacock et al., 2001). Theoretically, improved health status is the ultimate output of hospitals or the health care system generally. Nevertheless, the measurement of health status poses difficulties because health is multi-dimensional and involves assessing the quality of a patients' life (Clewer and Perkins, 1998). Because of the difficulty of accurately measuring improvement in health status, hospital output is measured in the study as an array of intermediate outputs (health services) that seek to improve health status (Grosskopf and Valdmanis, 1987). Therefore, the study views health care as an output in the service provision perspective and relates the inputs employed to the outputs produced.

3.4 Adjusting Hospital Production for Case-mix

Risk adjustment is an important tool in health services research. Briefly, case-mix or risk adjustment is a method to statistically control for the effects of patient characteristics that may influence outcome status. Controlling for these prognostic characteristics removes their influence on the outcome, which makes it possible to evaluate the relationship of treatment on outcome more directly. Adjusting for patient case-mix has often been referred to as “leveling the playing field” (MDRC, 1997) and is particularly important when there are systematic differences between patients across the units of analysis. The case-mix index is typically an arithmetic index that measures the relative costliness of inpatient or outpatient cases treated in a hospital compared to the country-wide average.

Conceptually, case-mix adjustment is similar to statistically controlling for pre-existing group differences in naturalistic or non-experimental designs, or in experimental designs in which randomization has not produced equivalent groups. In some treatment evaluations, researchers may want to adjust for patient characteristics that vary across treatment conditions and predict outcome. In contrast, the development of case-mix adjustment models typically focuses only on patient characteristics that have consistent prognostic significance, rather than on those that vary between programs or regions. When programs/facilities vary on the prognostic patient characteristics included in a case-mix model, however, case-mix adjustment will affect program/facility performance standings.

Hospital case-mix refers to the variety of illness treated in a hospital setting. The case-mix of cost function estimation presents two problems, (i) if hospitals do not treat the same kinds of ailments (or if they follow radically different treatment protocols), then their production and cost structures are bound to be different, and they ought not to be regarded as belonging to the same class of firms; and (ii) the correct specification of the cost function requires the inclusion of all outputs of hospitals in the set of regressors. Otherwise, the regression equation runs the risk of being mis-specified. Given the sheer number of diseases and conditions for which patients seek treatment in a hospital, aggregation of throughputs is necessary to avoid degrees of freedom problem (i.e., where the number of parameters to be estimated is greater than or equal to, the number of observations in the data set) in the estimation of cost function. Unfortunately, the appropriate method of aggregation is still unsettled in the literature, although there is no shortage of proposals.

Schemes employed in actual hospital cost function studies, range from simple breakdowns of medical cases into the number of outpatient visits and in-patient admissions (Wouters, 1993), to elaborate stratification. This includes number of in-patient days by hospital department and emergency room visiting (Cowing and

Holtmann, 1983) and frequencies of in-patient days by age group and mode of payment (Conrad and Strauss, 1983).

Breyer (1987) suggests that the case-mix issue be handled by grouping patients, according to an arbitrary (manageable) number of diagnostic categories and by specifying that each diagnostic group raises total costs only by a constant. Given N diagnostic groups and y_1, y_2, ..., y_N cases per group, the effect on total costs of these groups is given by

$$\sum_{n=1}^{N} \delta_n y_n$$

Wagstaff and Barnum (1992) noted, however, that this type of specification assumes away the possibility of economies of scope. That is, the cost of jointly producing various classes of outputs cannot be lower than the costs of producing each output category separately, if total costs are merely the sum of all outputs.

After recognizing that the approach used for traditional markets[4] is inappropriate in the case of hospitals because of price and non-price distortions in the market, Ellis (1992) argues for aggregating hospital throughputs into a case-mix index in the specification of the hospital cost function. This case-mix index is generated by dividing the severity-weighted sum of hospital admissions, in which diagnostic resource group (DRG) costs are used as measures of illness severity weighted by the total number of admissions. Ellis cautions, though, that a drawback in using such a case-mix index is

[4] Which is to deflate the price weighted sum of a subset of products by some price index in order to generate quantity indices of outputs.

that the index infects output variables with measures of inputs. Studies concerned with issues of hospital performance and cost, using diagnostic-specific case groups or other forms of hospitals, services and institutional surrogate measures, typically aggregate the hospital case-mix. The approach recognizes hospitals as multiple product firms. Researchers (Thompson, Fetter, Mross, 1975 and Lave & Lave, 1970) identified three approaches of guidance on how hospital case-mix can be measured: (i) Define groups of patients on amount and type of resources expended; (ii) Aggregate patients by clinical services (pediatrics, obstetrics, etc) and define case- mix in terms of the proportion of patients in these groups; and (iii) Classify patients by various International Classification of Diseases (ICD) groupings and define case-mix by the percentage of patients within these diagnostic groups.

In contrast to case-mix measures, Phillip and Hai (1976) use product related variables, which include facility availability and accreditation dummy variables, direct output measures of the ratio of births to discharges and of outpatient activity to in-patient activity. Controlling input factors (e.g., staffing patterns) and external factors (e.g., socioeconomic and physician availability), Phillip and Hai (1976) were able to explain 34 percent of the variation in cost per case.

Feldstein (1967) adopted surrogate variables as the output mix measure and examined the relationship between operating expenses per medical case and the proportion of cases in each of 28 diagnostic categories of case-mix. Various specialties were grouped into eight mutually exclusive categories, (i) general medicine; (ii) pediatrics; (iii) general surgery; (iv) Ear Nose and Throat (ENT); (v) trauma and orthopedic surgery; (vi) other surgery; (vii) gynecology and obstetrics.

Lave et al. (1974) adopted 17 broad ICD categories, hospital characteristics, patient population characteristics and severity measures, against which cost per day is regressed. Evans (1971) used 41 broad ICD diagnostic categories including age-sex variables, bed size, and the case flow rate, reduced subsequently to 10 categories through factor analysis, losing only 7 percent of the explanatory power.

Watts and Klastorin (1980) used a cross-section sample of 315 short term general hospitals in the US, to compare the ability of various measures of case-mix to explain the variation in average cost per admission per hospital. The study uses 10 case-mix variables and proxy variables, including the number of beds in the institution. Four-element variables counted the number of facilities and services in each of the four service categories (basic service, quality-enhancing services and complex services) and the weighted sum of a number of facilities and services reported by the hospital. The study was able to explain 70 percent of the inter-hospital variation, due to four service categories. Given dimensionality issues, direct case-mix variables performed better than the proxy variables.

Given the lack of a patient classification system in Uganda's healthcare system according to the complexity of illness as well as resource consumption in the provision of treatment; and the lack of data on the costing/billing of hospital services by type of disease, this study made use of length of stay by ward to compute the case-mix index. This is because of the expected direct relationship between the length of stay and hospital costs. Previous studies for instance, Winslow et al. (2002) found the number of respiratory complications experienced during the initial acute-care hospitalization for cervical spine injury to be a more important determinant of length of stay and hospital costs than level of injury.

3.5 Adjusting Hospital Production for Quality of Care

Typically, obtaining accurate insight into healthcare quality is difficult, yet important. Beginning with the seminal work of Donabedian (1988)[5] healthcare quality has been separated into three components: structure, process, and outcomes. Structure refers to the components of the healthcare system: personnel training and skills, adequacy of equipment resources (both diagnostic and therapeutic), and organizational systems to efficiently mobilize these resources for optimal patient care. Process refers to the use of appropriate diagnostic and therapeutic modalities for individual patients. To facilitate the interpretability of process assessments, “ideal” patient subsets—those without contraindications for therapy—are often used as the denominator, and those who received appropriate treatments are reported as the numerator. The term “outcomes” refers to the consequences of treatment and can represent markers of disease progression (mortality, readmission, and so on), health status (symptoms, functioning, and quality of life), and/or cost.

[5] Donabedian A (1988); “The Quality of Care: How can it be assessed? *JAMA,* 260:1743–1748.

Appropriate measures of quality of care predicate, (i) the inherent uncertainty about the outcomes of medical treatments, which makes mortality and re-admission rates indicative but noisy measures of quality at best; (ii) the bundling of (medical and non medical or hostel) services in a hospital stay, each of which may have a qualitative aspect; and (iii) the perception of patients regarding quality of care. Measures of quality proposed or used in literature highlight, (i) the teaching status of hospitals; (ii) the number or proportion of specialists on the medical staff; (iii) the location and accessibility of the hospital; (iv) the attributes of amenities (e.g., cleanliness of facilities, hospitality of the staff, quality of food); and (v) the occupancy rate of hospitals. The occupancy rate of hospitals was used by Friedman and Pauly (1981) as a measure of quality. This is based upon the argument that as admissions approach hospital capacity, lower overall quality of services may be provided.

In-hospital mortality for a surgical procedure determined from discharge data is clearly important information, interpreted cautiously; such mortality can sometimes be attributed to differences related to quality of care (McKee, 1993). However, correction for differences for case-mix must be made (Robin and Wu, 1992) but even then; evidence that the residual variation is related to quality of care is conflicting (Jencks and Dobson, 1987). This study used hospital mortality data an indicator of the quality of healthcare provided. It should be noted that mortality figures are a crude approximation of the quality of care because a small proportion of patients undergo treatment in which they die. Thus, it does not tell us anything about patient experiences by the majority who survive, for instance their satisfaction with the care received. Mortality rates are an incomplete indicator of the quality of care, in part, because death is an infrequent event. In addition, risk adjustment is an inherently subjective process.

3.6 Conclusion

Hospitals seek to improve upon the health status of their patients. Nevertheless, health improvement is bedeviled with conceptual and empirical difficulties. Consequently, this study uses proxy outputs that are geared toward improved health status. Heterogeneity exists in the patient loads of hospitals. Within the limitations of available data, this was crudely provided for via a length of stay-based case-mix index. The quality of care provided by the hospital is also approximated for by means of in-hospital mortality. The next chapter reviews methods for the measurement of hospital efficiency and productivity.

CHAPTER FOUR

REVIEW OF METHODS TO EXAMINE HOSPITAL EFFICIENCY AND PRODUCTIVITY

4.1 Introduction

This chapter reviews the methodological as well as empirical literature on efficiency measurement in healthcare systems. In particular it looks at parametric and nonparametric techniques that can be used to measure health care service efficiency and highlights their respective strengths and limitations. Hospital efficiency studies have been carried out in both developed and developing countries. This review focuses on technical efficiency and does not concern itself with other issues of health system performance measurement such as equity, effectiveness, quality and outcomes. It looks at the interplay between resource inputs and the resulting outputs of healthcare institutions. The lessons learnt from the empirical studies are also highlighted.

4.2 Efficiency and Productivity Measurement Techniques

Extraordinary progress has been made in the theory and empirical estimation of models of productivity and efficiency since 1978. Advances in statistical methodology, an explosion in data provision, and the ready availability of high quality software have among other things contributed to the prodigious growth of interest in the topic. Moreover – and unusually for such complex analytic technology – the methods are increasingly being considered by policy makers as a tool for influencing the real economic behavior of public service organizations.

Policy makers are therefore increasingly attracted to the development of overall measures of organizational performance. The arguments in favor of such endeavors according to Hibbard et al. (2003), are: (i) Unlike targets based on individual performance measures, global efficiency measures can offer local managers the freedom to set their own priorities, and to seek out improvements along dimensions of performance where they believe gains are most readily secured; (ii) Global measures of efficiency can be used to support other objectives, such as allocating finance or identifying the priority organizations for inspection and performance improvement; (iii) In contrast to piecemeal examination of single performance indicators, global indices of efficiency can offer a rounded assessment of system performance. This is particularly important when inputs (in the form of expenditure) cannot readily be attributable to specific activities, given limitations in data or accounting methods; and, (iv) Global measures of efficiency facilitate the publication of 'league tables' or rankings of entire organizations. Some analysts conjecture strongly that such rankings nurture public interest in the public services, promote accountability, as well as stimulate a search for improvement.

Askin and Standbridge (1993) define effectiveness as doing the right task, efficiency as doing a task right, and performance as accomplishing the right task efficiently. Sink and Tuttle (1989) maintain that system performance is a function of the complex interaction among seven criteria. These criteria are efficiency, effectiveness, quality, productivity, quality of work life, innovation, and profitability. Efficiency is concerned with measuring how inputs are converted into valued outputs. Productivity is the ratio of outputs that an organization produces to inputs that are used in the production process. Thus productivity may include the notion of efficiency but is not confined to it.

Since such authors as Debreu (1951), Koopmans (1951) and Farrell (1957) introduced the analysis of efficiency in the economic literature, there are numerous and wide ranging collections of papers and articles devoted to the measurement of productive efficiency. There has always been a close link between the measurement of efficiency and the use of frontier functions. Different techniques have been utilized to either calculate or estimate these frontier functions.

Ratio analysis and frontier techniques are used in the measurement of productivity and/or efficiency, but the latter have been widely employed in the analysis of the efficiency of healthcare services. Most of the studies related to the measurement of productive efficiency in healthcare services have based their analysis either on parametric or on non-parametric methods (see for instance, Jacobs 2001, Hollingsworth 2003, and Worthington 2004 amongst others). The choice of estimation method has been an issue of debate, with some researchers preferring the parametric approach (for example, Berger, 1993) and others the non-parametric approach (for instance, Seiford and Thrall 1990).

The literature on productivity and efficiency measurement has grown dramatically since the publication of seminal papers by Solow (1956), Aigner & Chu

(1968) and Aigner, Lovell & Schmidt (1977), Charnes, Cooper & Rhodes (1978). Increasingly sophisticated methodologies have been developed to tackle noisy data; reduce the restrictions imposed by functional forms and behavioral assumptions; identify the components and determinants of productivity growth; and address many other theoretical and econometric issues. The techniques have been applied to a large variety of contexts and sectors, in particular healthcare institutions, where hundreds of published papers on productivity and efficiency analysis exist.

Productivity as well as efficiency research has developed two broad schools of analytic thought intended to inform the development of global efficiency measures (Stone, 2002). Parametric methods, such as Stochastic Frontier Analysis (SFA), use multivariate statistical models to explore why output or costs differ across organizations. Conversely, non-parametric methods, pre-eminently Data Envelopment Analysis (DEA), attempt to measure efficiency by estimating the optimal level of output conditional upon the amount and mix of inputs.

Policy makers are increasingly showing keen interest in these techniques. Internationally, one of the highest profile examples is a study conducted by the World Health Organization that produced a ranking of the efficiency of national health systems (World Health Organization, 2000). This initiative stimulated a wave of academic and policy response (Scientific Peer Review Group, 2002).

The difficulty in identifying a straightforward process of efficiency measurement is not surprising in view of the nature of the phenomenon being studied. Inefficiency is inherently unobservable. Estimates of inefficiency must therefore be derived indirectly, after taking account of observable phenomena. In crude terms, this entails the following process: (i) measuring observable phenomena (for instance, inputs, outputs, prices, costs); (ii) specifying some form of relationship between these phenomena; (iii) defining 'efficient' behavior; (iv) calculating the difference between each organization's observed data and the maximum achievable as defined by the specified relationship; and judging how much of the difference is attributable to inefficiency. More detail on the theoretical underpinnings of the approaches appears elsewhere (Greene, 1993; Coelli et al., 1998).

Empirical estimates of efficiency measures entail two steps, namely the estimation of the frontier as well as the calculation of the individual decision making unit deviations from the frontier. There are two approaches used in estimating frontiers (Seiford and Thrall, 1990; Coelli et al., 1998): the parametric approach which employs econometric methods, and non-parametric approach, that entails linear programming techniques. From an implementation viewpoint, methods of measuring efficiency can be broadly classified into parametric and nonparametric approaches.

The main assumptions and estimation issues underlying ratio analysis, parametric, and non-parametric techniques of efficiency measurement are described briefly in the following section. Although both are consistent with the principles of the efficiency index, they are based upon slightly different methodological foundations. Moreover, within each empirical framework, a series of estimation decisions must be made, and there is no generally accepted methodology for guiding such decisions.

4.2.1 Ratio Analysis

The measurement of productive efficiency by means of ratio analysis generally entails computing and comparing one or both of the following two types of ratios, namely, input to output ratios as well as cost of inputs to output ratios (Bitran, 1992). However, a major shortcoming of this method is its inability to handle multiple input versus multiple output production. The input to output ratio approximates technical efficiency, whereas the cost of inputs to output ratio approximates economic efficiency.

There is a vast amount of literature on the measurement of health services efficiency by means of ratio analysis. Barnum and Kutzin (1990) provide an excellent review of that literature for developing countries. Ratio analysis has several advantages, including its conceptual simplicity, ease of computation, low cost, and being amenable to small samples. However, if standard statistical confidence is sought for differences in the ratios, the sample must be large enough. Nevertheless, ratio analysis has several limitations which include its inability to deal with multiple-output production as noted earlier. Moreover, in some situations, the approach requires that certain costs of multi-product providers be allocated among services, which calls for arbitrary assumptions about rules of allocation. More often than not, the results are very sensitive to the allocation rule employed. Contingent upon the circumstances, the advantages of ratio analysis may prevail over its demerits and its use, therefore, warranted. In addition, ratio analysis is a powerful tool in the assessment of efficiency within a particular health care institution.

4.2.2 The Parametric Technique

The parametric method to the analysis of efficiency entails the estimation of an econometric (multiple regression) model. On the one hand, if the purpose is to explore differences in output, a production function is estimated. Conversely, if differences in the cost are the focus of the empirical exercise, a cost function is estimated. In industries where firms produce multiple outputs (for instance hospitals), it is usually more convenient to work with a cost function which allows a single dependent variable to be estimated. If cost minimizing behavior can be hypothesized, the cost function is the dual of the production function, which makes the two approaches equivalent (Varian, 1978). However, this assumption may not hold in all industries.

An alternative taxonomy is deterministic versus stochastic models. Deterministic models identify the distance between observed and predicted frontier as inefficiency, whereas stochastic models relate the distance as a combination of inefficiency and random noise.

Parametric methods such as Stochastic Frontier Approach (SFA) which allow for this decomposition of the error term sometimes also referred to as the econometric frontier approach - specify a functional form for the cost, profit, or production relationship among inputs, outputs, and environmental factors, and allow for random error. SFA posits a composed error model where inefficiencies are assumed to follow an asymmetric distribution, usually the half-normal, while random errors follow a symmetric distribution, usually the standard normal. The other two parametric methods (Berger and Humphrey, 1997) include the distribution-free approach and the thick frontier approach. The Distribution-Free Approach (DFA) also specifies a functional form for the frontier, but separates the inefficiencies from the random error in a different way. The Thick Frontier Approach (TFA) specifies a functional form and assumes that deviations from predicted performance values within the highest and lowest performance quartiles of observations (stratified by size class) represent random error, while deviations in predicted performance between the highest and lowest quartiles represent inefficiencies.

The parametric approach to modeling efficiency closely mirrors the traditional statistical approach. It entails the following general process: (i) identification of a dependent variable – either output or cost (*y*); (ii) specification of a set of explanatory variables (**x**) that are unrelated to efficiency but are thought to explain or predict differences in output or cost; and (iii) interpretation of residual differences between observed and predicted output or cost (ε) as arising from either measurement error or inefficiency. The dependent and independent variables are related by specifying a statistical model of the general form:

$$y_i = \alpha + \beta X + \varepsilon_i \tag{4.1}$$

where *y* indicates either output or cost, ***x*** is a vector of independent variables, and the residual ε_i represents the deviation between the observed data and the relationship predicted by the independent variables in the model.

The error term in equation (4.1) consists of two components:

$$\varepsilon_i = v_i - u_i \tag{4.2}$$

The first component, v_i is a two-sided conventional random error term that is independent of u_i, and is assumed to be distributed as $N(0,\sigma_v^2)$.This component is supposed to capture statistical noise (i.e. measurement error) and random exogenous shocks that disrupt production. The second component u_i is also a random variable, but unlike v_i, it is only a one-sided variable taking non-negative values. This term captures technical inefficiency of a DMU in producing output. One of the disadvantages of the parametric method is that its estimation requires explicit specification of the distribution of the inefficiency term u_i. There is no consensus among econometricians as to what specific distribution u_i should have. In previous empirical studies a variety of distributions, ranging from the single-parameter half-normal, exponential and truncated normal distributions to the two-parameter Gamma distribution, have been used (see Bravo-Ureta and Rieger, 1990; and Battese, 1992; Jaforullah and Devlin, 1996; and Sharma et al., 1999).

However, a number of questions must be answered in order to estimate a parametric model of the specified form. These include but are not limited to: Should output or cost be the dependent variable? What functional form should be employed? Should the explanatory variables be transformed (for example to logarithmic form, amongst others)? Which explanatory variables should be included? And how should the residual be modeled and interpreted? Answers to the above questions are influenced by the scope and nature of data availability. However, the residual is not afforded special attention in traditional statistical methods, other than its being distributed according to the modeling assumptions. In contrast, the distinguishing feature of the policy use of productivity models is that the residuals for individual observations are often deemed the only phenomena of direct interest.

4.2.3 Merits and Demerits of the Parametric Technique

Statistical cost (and production) functions attempt to estimate an underlying cost (production) function from cross-sectional or time-series data. Whilst it is accepted that the estimated functions derived from such studies do not represent "efficient" production, this approach has particular value in terms of identifying the behavior of marginal costs at various output levels, and of drawing conclusions regarding the existence and importance of returns to scale (Hensher, 2001). The general technique of multiple regression analysis which is employed in these types of studies also lends itself to the inclusion of large numbers of independent variables, whose potential impact on cost can thus be estimated. This ability has been prized for allowing the incorporation of complex adjustment for case-mix factors in the healthcare setting.

However, this approach is not without its shortcomings. Crucially, as the true functional form is unknown a priori, there is always the inherent risk of mis-specification which yields misleading results. Nevertheless, attempts to use more flexible functional forms (Breyer, 1987) sacrifice capability to adjust for case-mix or other explanatory variables. There is also some debate as to whether such studies yield short-run or long-run functions, and the implications of this for their interpretation

(Aletras, 1999). Most important for the consideration of efficiency and its improvement is the criticism that the use of central tendency techniques inherently produce an analysis of average performance (Rosko and Chilingerian, 1999).

A further drawback of the econometric approach applied to the estimation of production functions is the difficulty of considering multiple outputs. The obvious solution to this is the estimation of a cost function. However, an alternative approach is to apply canonical regression methods (Hotelling, 1936). The aim is to find a linear combination, or weighted sum, of two sets of variables (for instance outputs and inputs) such that the correlation between the two is maximized (Giuffrida and Gravelle, 2001). Nevertheless, as yet, there have been limited applications of the technique (Ruggiero, 1998; Tofallis, 2001).

4.2.4 The Nonparametric Technique

Building upon the work of Farrell (1957), Charnes et al. (1978) proposed a formal linear programming approach known as Data Envelopment Analysis (DEA) for studying technical efficiency of “Decision making Units”, or DMUs, allowing for multiple output production which is a common attribute of most health care service providers. They proposed an iterative algorithm where within each iteration; the technical efficiency of a DMU is computed relative to all the other DMUs in the set. The efficiency measure is obtained as “the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that similar ratios for every DMU be less than or equal to unity.”

Charnes et al. (1978) DEA model assumes constant returns to scale (CRS) and is considered a sensitive model for measuring technical efficiency. Following the work of Banker, Charnes, and Cooper (1984), a second DEA model, which assumes variable returns to scale (VRS), was developed to separate pure technical efficiency from scale efficiency.

The precise formulation of the Charnes et al. (1978) approach is

$$Max \quad h_o = \frac{\sum_{r=1}^{s} u_r \cdot y_{ro}}{\sum_{i=1}^{m} v_i \cdot x_{io}} \tag{4.3}$$

subject to:

$$Max \quad h_o = \frac{\sum_{r=1}^{s} u_r \cdot y_{rj}}{\sum_{i=1}^{m} v_i \cdot x_{ij}} \leq 1;$$

$$for \quad j = 1,...n; \quad u_r, v_i \geq 0; \quad r = 1,...,s; \quad i = 1,...,m.$$

where the sub index 0 is used to represent the facility for which technical efficiency is being calculated and where "the y_{rj}, x_{ij} (all positive) are respectively the [observed] outputs and inputs of the j^{th} DMU and the u_r, $v_i \geq 0$ are the variable weights to be determined by the solution to this problem, for instance, by the data on all of the DMUs which are being used as a reference set. The efficiency of one member of this reference set of j = 1,…,n DMUs is to be rated to the others. According to this approach, the solution to the technical efficiency calculation problem is obtained by solving, in any order, n linear programming constrained maximization problems.

The CRS assumption is only appropriate if all DMUs are operating at an optimal scale. Amongst other things, imperfect competition as well as, constraints on finance, may cause a DMU not to operate at the optimal scale. When DMUs are not operating at an optimal scale, the technical efficiency can be decomposed into pure technical efficiency and scale efficiency. Therefore, in situations where the CRS does not hold, the technical efficiency measure is mixed with scale efficiency. To disentangle the effect of scale efficiency it is necessary to use a DEA model with a Variable Returns to Scale (VRS) assumption. To this end Banker et al. (1984) developed an extension of the original CRS model. Banker et al. (1984) suggested an extension of the Constant Returns to Scale (CRS) DEA model to account for Variable Returns to Scale (VRS) situations. The use of the CRS specification when not all DMUs are operating at the optimal scale, results in measures of technical efficiency (TE) which are confounded by Scale Efficiencies (SE). The use of VRS DEA specification permits the calculation of TE free of SE effects.

The precise formulation of the Banker et al. (1984) approach is

$$Max \quad \sum_{r=1}^{s} u_r y_{r0} - u_0 \tag{4.4}$$

subject to

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - u_0 \leq 0, \qquad j = 1,...,n,$$

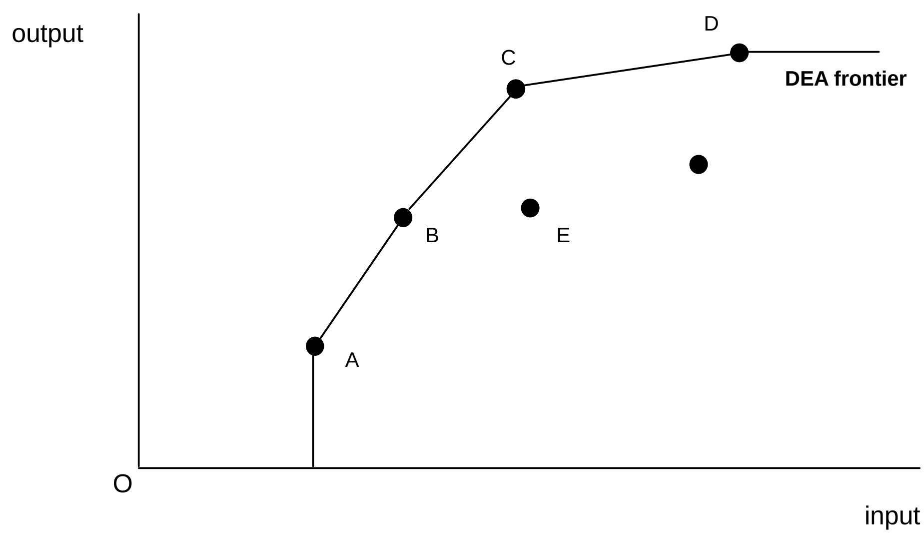
$$\sum_{i=1}^{m} v_i x_{i0} = 1, \qquad u_r, v_i \geq 0$$

and u_0 is unconstrained in sign and captures the returns to scale. From the VRS model, it is possible to analyze whether a DMU's production indicates increasing returns to scale, constant returns to scale, or decreasing returns to scale contingent upon the sign of the term u_0. Increasing returns to scale exists if the value of u_0 is less than zero (u_0<0), constant returns to scale if the value of u_0 is equal to zero (u_0=0), and decreasing returns to scale if the value of u_0 is greater than zero (u_0>0). Thus, one can examine the existence of economies of scale, confirm the most productive scale size (minimum efficient scale) of a DMU and estimate the number of DMUs operating at the efficient scale.

The assumption of variable returns to scale in a DEA framework results in higher estimates of efficiency than constant returns to scale. This is because the frontier 'envelopes' the data more tightly. The approach known as Free Disposal Hull (FDH) analysis allows an even closer fit of the frontier to the data, by fitting a piece-wise linear function that is permitted to display non-increasing segments (in the case of the production frontier) (Tulkens, 1993). FDH generates a frontier that increases as a series of step functions. Nevertheless, it is difficult to conjecture the economic rationale for why a frontier would display such characteristics.

Whilst the parametric approach is guided by economic theory, data envelopment analysis (DEA) is a data-driven approach. The location (and to a lesser extent, the shape) of the efficiency frontier is determined by the data. DEA is based on the notion that an organisation that employs less input than another to produce the same amount of output can be considered more efficient. The efficiency frontier is constructed of linear segments that join up those observations with the highest ratios of output to input. The resulting frontier "envelopes" all the observations. In Figure 4.1, observations A, B, C and D are considered efficient, given the scale of their operations. The inefficiency of observation E can be measured by either its vertical or horizontal distance from the frontier – it uses more input to produce a similar level of output to observation B and, despite employing a similar amount of input to observation C, it produces considerably less output.

Figure 4.1 Illustration of the Data Envelopment Analysis Technique



4.2.5 Merits and Demerits of the Nonparametric Technique

DEA has grown in popularity recently such that it has become the dominant approach to performance measurement in many sectors of the economy (see Chang et al., 1992; Kittelsen and Forsund, 1992; Kooreman, 1994; Hollingsworth et al., 1999, amongst others). The main attraction of DEA is that it can handle multiple inputs and multiple outputs easily. Additionally, being a non-parametric technique, gives it the added merit of requiring no assumptions with regard to the functional form of the production, cost or profit frontier. This reduces the need for a theoretical exposition of model specification, with the main points of contention focusing on what variables should be classified and included as inputs to and/or outputs of the production process. Banker et al. (1984) notes that DEA can also be employed in the evaluation of management and program efficiencies of DMUs of not-for-profit organizations such as schools and hospitals.

DEA employs flexible, nonparametric methods to construct the best-practice frontier and so allows the data to ‘speak for themselves’ (Bates et al., 1996). The other advantage of DEA over other techniques is that each input and output can be measured in its natural physical unit without the need to apply a weighting system to collapse the different units in money or other single unit measure. The instrument is useful in the identification of differences in efficiency over time and between hospitals.

Input and output slacks represent another problem. Input slacks occur in DEA analysis when the projection of an inefficient observation onto the efficient plane occurs in such a way that the further reduction of one or more inputs is possible. Ali and Seiford (1993) describe a method to deal with slacks using the output of the CRS DEA linear program (LP) as the input into a second-stage LP. Solving this LP from each observation yields the maximum sum of input and output slacks required to move an inefficient frontier point to an efficient frontier point. However, Coelli et al. (1998) argue that there two major problems associated with this second-stage LP. The first problem is that the sum of slacks is maximized rather than minimized. The second is that it is not invariant to units of measurement. The alteration of the units of measurement while leaving other units of measurement unchanged could result in the identification of different efficient boundary points and hence different slack values.

Unlike in Management Science where DEA became virtually an instant success, in economics, its welcome has been far less enthusiastic. According to Ray (2004), there are three principal reasons for skepticism about DEA from the economist's perspective. First, DEA being a nonparametric method, no production, cost or profit function is estimated from the data. This precludes evaluating marginal products, partial elasticities, marginal costs, or elasticities of substitution from a fitted model. Consequently, one cannot derive the usual conclusions about the technology, which are possible from a parametric functional form. Second, DEA employs Linear Programming (LP) instead of the familiar least squares regression analysis. In economics, constraints in standard optimization problems are typically assumed to be binding and Lagrange multipliers are almost always positive. An average economist feels uncomfortable with shadow prices that become zero at the slightest of perturbation of the parameters. Finally, being non-statistical in nature, the LP solution of a DEA problem produces no standard errors and leaves no room for hypothesis testing. In DEA, any deviation from the frontier is treated as inefficiency and there is no provision for random shocks. By contrast, the far more popular stochastic frontier model explicitly allows the frontier to either move up or down due to random shocks. Moreover, a parametric frontier yields elasticities and other measures regarding the technology that are useful for marginal analysis.

(a) Sensitivity to Variable Selection and Model Specification

The sensitivity of DEA efficiency estimates to input variable selection and model specification has been another concern for practitioners. Pedraja-Chaparro et al. (1999) have shown that DEA relative efficiency is influenced by the distribution of true efficiencies, the sample size, the number of inputs and outputs included in the analysis and the degree of correlation between inputs**.** The potential for model misspecification in DEA is huge, and it can be attributed to the absence of a credible model-building methodology. The main causes of model misspecification in DEA are known to be omission of relevant variables, inclusion of irrelevant variables and incorrect assumption on returns-to scale and the sample size, resulting in misleading conclusions. When important variables are omitted the results may be far from reality. On the other hand, increasing the number of variables decreases the ability of the model to differentiate individual production units in terms of efficiency.

(b) Lack of Confidence Intervals

Data Envelopment Analysis (DEA) is a deterministic method, meaning that DEA distance functions do not have any statistical properties such as standard deviations, confidence intervals, among others. To address this weakness associated with nonparametric methods (DEA), two methods can be used to generate confidence intervals around the efficiency score for each DMU namely the Jackknife and the Bootstrap. The Jackknife predates the Bootstrap but is a similar method. The confidence intervals are generated around two statistics; the estimate of the efficiency score for each DMU, *è,* as well as the mean level of the efficiency of the whole sample, $\overline{\theta}$.

The bootstrap (Efron, 1979) provides an alternative method of producing confidence intervals. The bootstrap is implemented by sampling with replacement from the data to produce a new data set each time, and from this new data, new estimates are calculated. The procedure is replicated a number of times, the larger the number of replications the more accurate the statistic, although between 50 and 200 replications are generally sufficient (Efron and Tibshirani 1993). Bootstrapping (Efron and Tibshirani, 1993) is a general method for estimating statistical properties of deterministic parameters.

(c) Weight Flexibility

In traditional DEA, the actual input-output data values are multiplied with the calculated weights to determine the efficiency scores. Recent variants of the DEA model impose upper and lower bounds on the weights to eliminate certain drawbacks associated with unrestricted weights. These variants are called weight restriction DEA models. Most weight restriction DEA models suffer from a drawback that the weight bound values are uncertain because they are determined based on either incomplete information or the subjective opinion of the decision-makers. Since the efficiency scores calculated by the DEA model are sensitive to the values of the bounds, the uncertainty of the bounds gets passed onto the efficiency scores. The uncertainty in the efficiency scores becomes unacceptable when one considers the fact that the DEA results are used for making important decisions like allocating funds and taking action against inefficient units.

(d) Missing and/or Badly 'Behaved' Data

Missing data are a chronic problem in applications of DEA (Charnes et al., 1978). Very often, potentially important input and/or output variables have insufficient coverage, or DMUs fail to report all required statistics. The art of composing dense data matrices required by DEA from the sparse statistics available is a critical step of the analysis in a wide variety of applications. Due to its nonparametric and multi-dimensional nature, DEA generally requires large numbers of DMUs to produce statistically meaningful results (see for instance, Simar and Wilson, 2000; for a discussion of statistical properties of DEA efficiency estimators). Thus, DEA is highly vulnerable to data problems. The question of blank output entries is closely related to the treatment of zeros in the data matrices (Thompson et al., 1993). Charnes et al. (1978) requires all input-output data of DMUs to be strictly positive, and hence leaves no room for blank entries. Subsequent basic as well as empirical research has further elaborated on the minimal data requirements, and considerably relaxed the positivity condition. The minimal conditions are known to be: (i) at least one DMU consumes/produces every input and output, (ii) each DMU consumes at least one input and produces at least one output (Fare and Grosskopf, 2002, following Shepard, 1970).

4.2.6 Productivity Measurement

In general terms productivity refers to an economy's ability to convert inputs into outputs. It is a relative concept with comparisons either being made across time or between different production units.[6] For example, if it is possible to produce more output in period $t+1$, when using the same amount of inputs that were used in period t, then

[6] For example, a comparison of productivity between two firms in an industry, between two industries within or between countries, or between countries.

productivity is said to have improved. In other words, productivity is higher in the second period compared to the first.

Different types of input measures give rise to different productivity measures. For example, labor productivity measures involve dividing total output by some measure that reflects the amount of labor used in production. The total number of worker hours is one such measure, although some studies have used total numbers employed. Capital productivity is measured by dividing total output by a measure reflecting the total amount of physical capital used in the production process. Productivity measures, such as labor productivity and capital productivity that only relate to one class of inputs are known as partial productivity measures. Caution needs to be applied when using partial productivity measures as changes in input proportions can influence these measures. A simple substitution of capital for labor within the input mix of a firm or industry can also raise average labor productivity. This means that movements in the average labor productivity statistics do not always represent true changes in the underlying productivity of labor (Dixon, 1990).

The level of Total Factor Productivity (TFP) can be measured by dividing total output by total inputs. Total inputs are often an aggregation of only physical capital and labor, and may overlook inputs such as land.[7] When all inputs in the production process are accounted for, TFP growth can be thought of as the amount of growth in real output that is not explained by the growth in inputs. This is why Abramovitz (1956) described

[7] Consequently some authors prefer to define the resulting productivity measure as Multi-Factor Productivity (MFP) due to it including multiple inputs but not all possible (i.e., total) inputs.

the TFP residual as a 'measure of our ignorance'. As TFP levels are sensitive to the units of measurement of inputs and outputs, they are rarely of primary interest. Rather, the measurement of TFP growth is of primary interest. Hence, it is common to use the notation "TFP" to refer to growth rather than levels and this is the convention adopted in this study.

The four main approaches to productivity measurement include the growth accounting approach; the index number approach; a distance function approach; and the econometric approach (see Mawson et al., 2003, for a review). This study employs the Malmquist TFP index - a distance function-based approach which is reviewed shortly.

Besides measuring technical efficiency, it is crucial to assess the evolution of Total factor Productivity (TFP) and efficiency through time in order to examine whether a change in efficiency has occurred. The concept of TFP is very closely related to the concept of efficiency. The growth of TFP is defined as the change in output due to technical change and technical efficiency change over time. Technical change is represented by a shift in the production frontier of period t and t+1 whereas efficiency change is represented by the movement of a Decision Making Unit (DMU) closer or further from the present and past frontiers. Technical change and technical efficiency change cannot be measured accurately using trends in annual average efficiency scores because the average scores are based on separate frontiers estimated for each year over the study period.

Malmquist (1953) proposed a quantity index for use in consumption analysis. The index scales consumption bundles up or down, in a radial fashion, to some arbitrarily selected indifference surface. In this context Malmquist's scaling factor turns out to be Shephard's (1953) input distance function, and Malmquist quantity indexes for pairs of consumption bundles can be constructed from ratios of corresponding pairs of input distance functions. Although it was developed in a consumer context, the Malmquist quantity index has recently enjoyed widespread use in a production context, in which multiple but cardinally measurable outputs replace scalar-valued but ordinally measurable utility. In producer analysis Malmquist indexes can be used to construct indexes of input, output or productivity, as ratios of input or output distance functions.

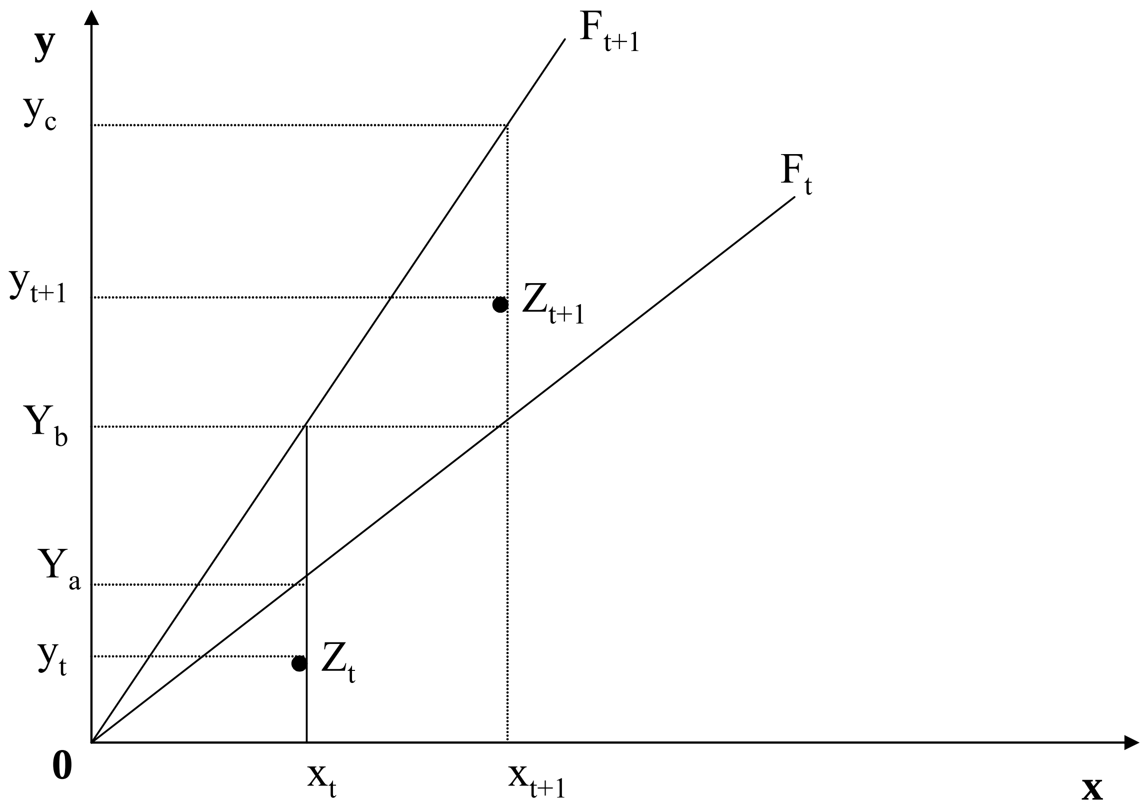
The Malmquist productivity index uses the idea of the distance function so that a preceding estimation of the corresponding frontier is required. The estimations are carried out using DEA, as it not only permits a consistent approach within the study, but also uses the same methodology as the majority of existing papers (Cummins and Weiss, 1999).

Malmquist indexes have a number of desirable features. They do not require input prices or output prices in their construction, which makes them particularly useful in situations in which prices are distorted or non-existent and it is usually the case in healthcare that we do not have price data. They do not require a behavioral assumption such as cost minimization or profit maximization, which makes them useful in situations in which producers' objectives differ, or are unknown. They are easy to compute, as Färe et al. (1995) have demonstrated. Under certain conditions they can be related to the superlative Törnqvist and Fisher ideal quantity indexes, as Caves et al. (1982) as well as Färe and Grosskopf (1992) have shown.

An attractive feature of the Malmquist productivity index is that it decomposes total factor productivity growth into technical efficiency change and technical or technological change. This was first demonstrated by Färe et al. (1995) using the geometric mean formulation of the Malmquist index. Färe et al. (1994) showed that the technical efficiency change index of the geometric mean of adjacent-period Malmquist productivity indexes, derived under the assumption of constant returns to scale, can be expressed as the product of an index of pure technical efficiency change, an index of scale efficiency change, and an index of congestion change. The value of each of these decompositions is that they provide insight into the sources of productivity change.

Figure 4.2 illustrates the framework under constant returns to scale following the methodology by Coelli et al. (1998). In this figure, a production frontier representing the efficient level of output (y) that can be produced from a given level of input (x) is constructed, premised on the assumption that this frontier can shift over time. The frontiers (F) thus obtained in the current (t) and future $(t+1)$ time periods are labeled accordingly. When inefficiency is assumed to exist, the relative movement of any given hospital over time will therefore depend upon both its position relative to the best practice frontier (technical efficiency) and the position of the frontier itself (technical or technological change). If inefficiency is ignored, then total factor productivity growth over time will be unable to distinguish between improvements that derive from a hospital's 'catching up' to the best practice frontier or those that result from the frontier itself shifting up over time.

Figure 4.2 Technical Efficiency, Technology and Productivity Changes



For a given hospital in period t depicted by the output/input bundle Z_t, the inputs are x_t and the output is y_t. But this is technically inefficient since the hospital lies below the production frontier. With the available technology and the same level of inputs the hospital should be able to produce output y_a. In the next period, there is a technology increase such that more outputs can be produced for a given level of inputs: the frontier moves upward to F_{t+1}. Suppose that the hospital's output/input bundle is now represented by Z_{t+1} with input x_{t+1} and output y_{t+1}. Once again, the hospital is inefficient, but in reference to the new technology, and should be producing output y_c if it were efficient. The challenge for productivity assessment is to decompose these increases in output relative to the level of inputs into that associated with the change in technical efficiency and that connected with the change in technology.

Following Färe et al. (1994) the input-oriented Malmquist total factor productivity index between periods t and t+1 is defined as:

$$M_i^{t+1}(x^t, y^t, x^{t+1}, y^{t+1}) = \left[\frac{D_i^t(x^{t+1}, y^{t+1})}{D_i^t(x^t, y^t)} \times \frac{D_i^{t+1}(x^{t+1}, y^{t+1})}{D_i^{t+1}(x^t, y^t)} \right]^{1/2} \tag{4.5}$$

where D_i is the input distance function and $M_i^{t+1}(x^t, y^t, x^{t+1}, y^{t+1})$ is the productivity of the most recent production unit, i.e. A(t+1), using period t+1 technology relative to the earlier production unit i.e. A(t), with respect to t technology. A score for M of less than unity indicates productivity progress in the sense that the hospital delivers a unit of output in period $t+1$ using fewer inputs. In other words, the hospital in period $t+1$ is more efficient relative to itself in period t. Similarly, a score greater than unity implies productivity regress and a unit score indicates constant productivity (Hollingsworth et al., 1999).

Following Färe et al. (1995) an equivalent way of writing this index is:

$$M = \Delta TECH \times \Delta EFF \tag{4.6}$$

where:

$$\Delta EF = \frac{D_i^{t+1}(x^{t+1}, y^{t+1})}{D_i^t(x^t, y^t)}$$

$$\Delta TECH = \left[\frac{D_i^t(x^{t+1}, y^{t+1})}{D_i^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D_i^t(x^t, y^t)}{D_i^{t+1}(x^t, y^t)} \right]^{1/2}$$

In this view, M – the Malmquist total factor productivity index, is the product of a measure of technical progress, ΔTECH,[8] and a change in efficiency, ΔEFF, over the same period. In order to compute these indices, it is necessary to solve several sets of linear programming problems. Suppose that there are n DMUs (e.g. hospitals) and that each DMU consumes varying amounts of m different inputs to produce s outputs in each period t. The j[th] DMU, in period t, is therefore represented by the vectors (x_j^t, y_j^t). It should be noted that either parametric or nonparametric methods can be used to compute M, but most empirical studies have used DEA methods. An input-oriented Malmquist index was specified because hospital managers can only influence the resource inputs into health care services but not the type and extent of the health care services demanded by their clients.

DEA is amenable to panel data, by calculating the Malmquist index. Changes in total factor productivity over time can be attributed to three separate explanations (Fare et al., 1994 and Giuffrida, 1999). First, the technical efficiency of an individual organization may change, at a given scale of operation. Second, the efficiency of the organization may change in response to a change in the scale of operation. Finally, the underlying technology may change, thereby inducing a shift in the production frontier, which will affect the efficiency of all organizations. The Malmquist index provides estimates of each of these effects by calculating separate distance functions in each period and by varying the assumption about the available technology.

---

[8] As measured by shifts in a frontier at period t+1 from period t (geometric average).

Simar and Wilson (1999) propose a bootstrap approach for estimating confidence intervals for Malmquist indices as well as their decomposition, which makes it possible to test hypotheses of whether the indices are significantly different from unity, i.e. whether significant TFP changes have taken place. The basic idea of bootstrapping is to simulate the original case study X times, each time recalculating the parameters of interest. This yields X estimates of parameters, which makes it possible to estimate the distributional properties of these. Seeing that it is often difficult or simply impossible to physically repeat the original case study, each new simulation is performed through re-sampling with replacement from the original dataset. The bootstrap procedure makes it possible to evaluate whether TFP changes are statistically significant, or whether they are caused by sampling noise.

The procedure for bootstrapping Malmquist indices involves re-sampling that is performed from the two sets of DEA distance parameters[9] for the time periods, as used to scale the observations in each time period, before evaluating the new set of Malmquist indices and their decompositions. The resulting X estimates of the indices make it possible to construct confidence intervals for the indices (Simar and Wilson, 1999).

4.2.7 Comparison of Techniques

Table 4.1 provides a summary of efficiency measurement techniques along with type of analysis as well as the kind of data each respective techniques employs. It compares

[9] Using bivariate smoothing that takes into account that the distance parameters in the two time periods may be correlated, (Simar and Wilson, 1999).

statistical techniques and mathematical programming. Broadly, ordinary least squares, stochastic frontier analysis and data envelopment analysis are amenable to both cross-sectional, time series as well as panel data although they differ in their assumption with regard to the measurement error.

Table 4.1 Summary of Efficiency Measurement Techniques

Type of Analysis	**Statistical Techniques** Ordinary Least Squares (OLS) and Stochastic Frontier Analysis (SFA)	**Mathematical Programming** Data Envelopment Analysis
Measurement of productivity change through time *Example* Measuring TFP growth/decline for a single entity for a period of two or more years.	***OLS*** • Can be applied to the measurement of productivity change ***SFA combined with Malmquist Index*** • Typically uses panel data; • TFP change can be decomposed into changes in technical efficiency, scale and technological; • Allows for measurement error.	***DEA combined with Malmquist Index*** • Typically uses panel data; • TFP change can be decomposed into changes in technical efficiency, scale and technological; • Assumes no measurement error.
Measurement of relative *technical efficiency* levels at a point in time ***Example*** Benchmarking the technical efficiency of a group of service delivery units of an entity for a given year	***OLS*** • Uses cross-sectional data; • Entities compared with average industry/sector performance; • Assumes no measurement error – residual is attributed to inefficiency. ***SFA*** • Uses cross-sectional data; • Comparison against best performing entity; • Residual decomposed into random error (measurement error) and inefficiency parts.	***DEA*** • Uses cross-sectional data; • Entities compared with best performers in the sample.

Source: Adapted from Hughes (2003).

4.3 Review of Selected Studies on Health Care Services Efficiency

Within the broad scope of healthcare services, frontier efficiency measurement techniques have been applied to a variety of healthcare institutions. These include but are not limited to hospitals (Banker et al., 1986; Fare et al., 1993; Chirikos, 1998), physician practices (Chillingerian, 1993; Defelice and Bradford, 1997), nursing homes (Hofler and Rungeling, 1994; Chattopadhyay and Ray, 1996) and substance abuse clinics (Alexander et al., 1998). And while the empirical literature has been predominantly concerned with the efficiency of North American healthcare institutions, applications in countries such as Kenya (Kirigia et al., 2004); Nigeria (Wouters, 1990); South Africa (Zere et al., 2000); Jordan (Al-Shammari, 1999); Spain (Wagstaff, 1989; Ley, 1991); Scandinavia (Fare et al., 1993; Luoma et al., 1996; Mobley and Magnussen, 1998), Taiwan (Lo et al., 1996) and the United Kingdom (Thanassoulis et al., 1996; Parkin and Hollingsworth, 1997; Jacobs, 2001; Hollingsworth, 2003; Worthington, 2004) have also been carried out. The main frontier technique employed in examining the efficiency of healthcare services has been Data Envelopment Analysis (Fizel and Nunnikhoven, 1992; Valdmanis, 1992; Kooreman, 1994; Thanassoulis et al., 1996; Parkin and Hollingsworth, 1997).

Comparisons between frontier efficiency measurement techniques have been made. For example, Jacobs (2001) compares the efficiency rankings from cost indices with those obtained using Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) in a sample of UK's National Health Service (NHS) hospitals. The paper recommends that several specifications should be used to develop ranges of inefficiency to act as signaling devices rather than point estimates. It is argued that differences in efficiency scores across different methods may be due to random "noise" and reflect data deficiencies. Linna (1998) examines DEA measures and stochastic frontier estimates of cost efficiency in Finnish acute care hospitals and concludes that the choice of approach does not significantly influence the results. Additionally, efforts have been made in healthcare services to compare ratio analysis and frontier techniques as alternative tools for performance assessment. For example, Thanassoulis et al. (1996) compares UK's NHS performance indicators for perinatal care units with DEA measures of productive performance. They conclude that there was no rationale for performance indicator values to be routinely accompanied by DEA measures of performance. Moreover, the multiple-input, multiple-output nature of DEA can be used to set performance targets. Nunamaker (1983) also compares univariate ratios and DEA, although in the form of cost per patient day.

Data Envelopment Analysis has also been compared to the conventional translog cost function. Banker et al. (1986) attempt to compare the results of the conventional translog cost function with DEA. Of special interest in this particular study was the level of similarities or differences between the two approaches in ascertaining increasing, constant or decreasing returns-to-scale, and estimating marginal rates of output transformation and technical inefficiencies of individual hospitals. Measuring inputs in terms of nursing, ancillary, administrative and general services and outputs in terms of patient days, Banker et al. (1986) using a sample of North Carolina hospitals finds that DEA was “able to examine the possibility of increasing or decreasing returns to scale prevailing in specific segments of the production possibility set”. More particularly, whereas the translog cost function indicated constant returns-to-scale across the sample, DEA found that the most productive scale size varied dramatically with different output mixes and capacity.

Banker et al. (1986) is vitally important for subsequent empirical research because it compares alternative techniques for efficiency measurement just like Wagstaff (1989). In addition, it sets a precedent for the specification of healthcare inputs as well as outputs. Most subsequent studies (for example, Kooreman, 1994; and Parkin and Hollingsworth, 1997) conceptualize healthcare as combining the inputs of labour (normally the number of healthcare staff) and capital (often proxied by bed capacity) to produce some easily-observable unit of output, for instance discharges or inpatient days. For example, Valdmanis (1992) conceptualizes Michigan hospitals as managing the inputs of physicians, house staff and nurses to maximize adult, pediatric and intensive care inpatient days and emergency and ambulatory visits. Similarly, Byrnes and Valdmanis (1989) study the technical efficiency for a group of 123 non-profit and teaching Californian community hospitals. They also use linear programming and information about input prices and technology to compute economic efficiency for individual hospitals in the sample. Their measures of output include acute medical surgical discharges, surgical intensive care unit discharges, and maternity discharges. The variable inputs they consider include registered nurse hours, management and administrative personnel hours; technical services personnel hours; aide and orderly hours; and licensed practical nurse hours. The fixed inputs chosen are the number of staff physicians and physicians with admitting privileges and the average staffed beds.

However, the emphasis that is placed on the production of inpatient care, because it usually comprises the largest component of hospital costs which can be readily measured, is questionable on a variety of grounds.

Magnussen (1996) notes that the selection of outputs and inputs for a DEA study requires careful thought as the distribution of efficiency is likely to be affected by the definition of outputs and the number of outputs and inputs included. Clewer and Perkins, (1998) note that conceptually, improved health status is the expected final output of the health care system in general or hospitals in particular. However, the measurement of health poses difficulties because health is multi-dimensional and there is subjectivity involved in assessing the quality of patients' life. Given the difficulty of accurately measuring improvements in health status, Grosskopf and Valdmanis (1987) argue that an array of intermediate outputs (health services) geared toward improving health status can be used to measure hospital output. Butler (1995) classifies hospital output into four broad categories which include inpatient treatment, outpatient treatment, teaching as well as research.

Chillingerian (1993) also argues that defining healthcare output by patient days, or discharges, or even cases, is acceptable so long as adjustment is made first for the case mix, or complexity of cases, and second for the intra-diagnostic severity of cases. Chillingerian (1993) using a sample of U.S. physicians, incorporates these concepts by classifying discharges on the basis of either a satisfactory (i.e. healthier state) or unsatisfactory outcome (i.e. the presence of morbidity or mortality).

The usual case is to engage in some form of aggregation in order to ensure homogeneous health care outcomes. For example, Zere et al. (2000) stratify hospitals according to their level to account for case-mix as well as staffing pattern and medical technology likely to impact on the quality of care delivered. Similarly, Grosskopf and Valdmanis (1987) disaggregate outputs by type of treatment which includes acute in-patient days, intensive care inpatients days and the number of surgeries. In the same vein, Banker et al. (1986) categorize outputs in terms of patient's age, namely, adult patients, Medicare patients and paediatric patients. Despite the foreseen attempts, Newhouse (1994) argues that case-mix controls by Diagnosis-related Groups (DRGs) usually encompass non-random variation. The problem of defining healthcare output is further highlighted when it is realized that even DRG outputs, which in turn are aggregated measures, are likely to involve several hundred separate categories. Citing earlier studies, Newhouse (1993) gives the example where patients may be disproportionately admitted to hospitals that are equipped to undertake specific treatments, and accordingly is not the result of variation in efficiency, rather variation in a healthcare institution's patients. This has implications for the validity of efficiency measures. Zere et al. (2000) observes that even though the use of Diagnostic-related groups (DRGs) may address the problem of hospital case-mix, the absence of such data in most developing countries makes its use limited.

4.4 Conclusion

The foregoing methodological and empirical review points to the fact that each broad category of efficiency measurement techniques has particular merits and demerits besides potentially measuring different aspects of efficiency. Given the relative strengths and weaknesses of ratio analysis, parametric and nonparametric techniques as well as the data limitations on inputs and outputs in the health care system, this study employs nonparametric techniques to examine the technical efficiency and total factor productivity growth of Uganda's district referral hospitals. The methodology employed is described in detail in the next chapter.

CHAPTER FIVE

METHODS FOR MEASURING HOSPITAL EFFICIENCY AND PRODUCTIVITY

5.1 Introduction

This chapter presents the methods used in analyzing technical efficiency and total factor productivity growth of district referral hospitals located in three out of four regions of Uganda. It looks at the microeconomics of efficiency and productivity; the input versus output orientation; the data envelopment analysis model; modeling production with undesirable outputs; incorporating undesirable outputs in DEA models; the unit of analysis, sample size, data collection, sample selection and modeling choices; choice of inputs and outputs; providing for case-mix; super-efficiency; and boostrapping.

5.2 Microeconomics of Efficiency and Productivity

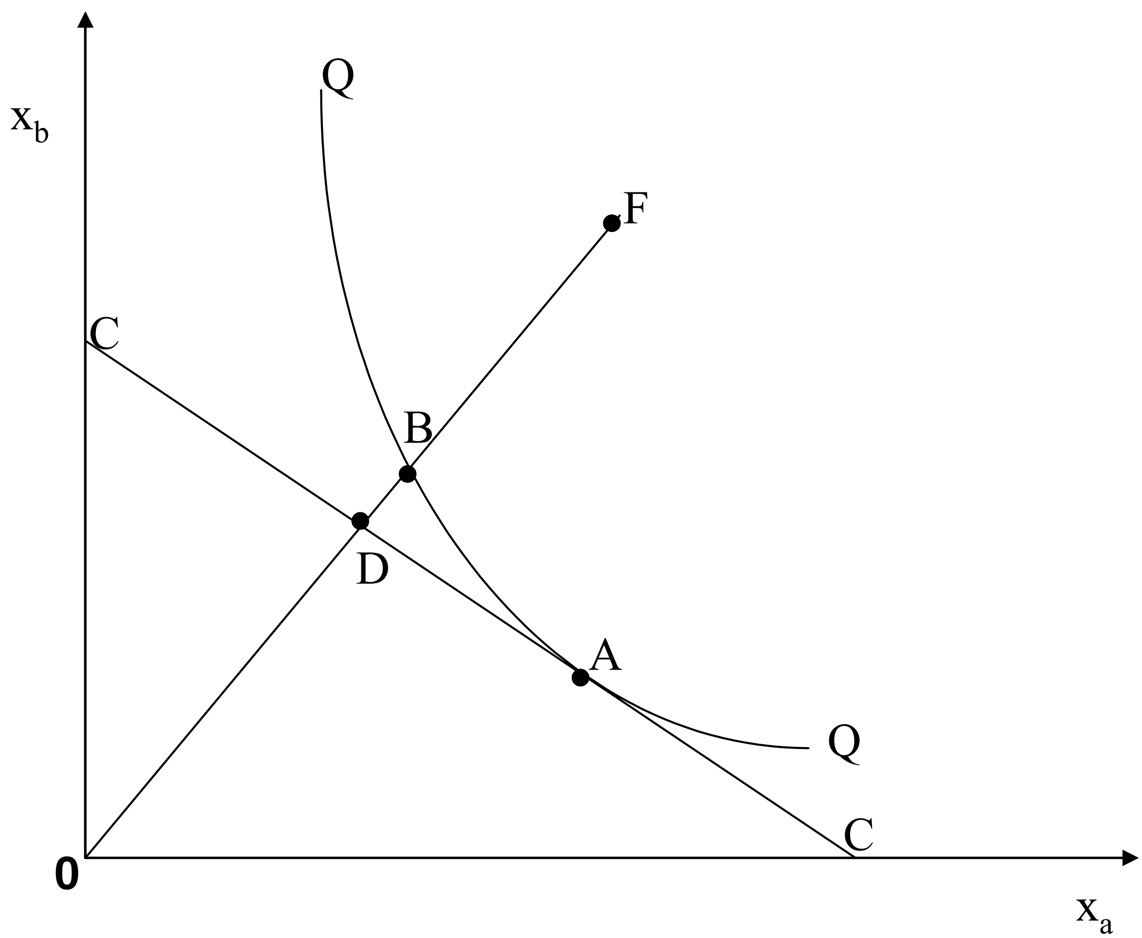
The concepts of productivity and efficiency mean different things to different people and in some instances are interpreted as being synonymous. Though the concepts are related, in general, productivity can be thought of as being a broader concept than efficiency. Both concepts can be related to a production function representing the transformation of inputs to output. The concept of productivity is widely accepted as a key performance benchmark for economic entities. Rising productivity is related to lower costs, and sustained competitiveness.

Productivity which is defined as the ratio of outputs to inputs can be analyzed at various levels – economy-wide, industry, firm/agency and operational unit. The focus of this study is at a more disaggregated level; how well district hospitals in Uganda convert their inputs of amongst others labor, materials and capital into outputs of health care services.

Drawing upon the work of Debreu (1951) and Koopmans (1951), Farrell (1957) introduced a measure of productive efficiency that avoids the impediments associated with traditional average productivity measures (ratios). Farrell refuted the idea of an absolute measure of efficiency and proposed that efficiency be measured relative to a best-performance frontier determined by a representative peer group. He further provided the definitions and computational framework for technical and allocative (in)efficiency.

Farrell (1957) developed a framework, susceptible to empirical estimation, to distinguish between technical and economic efficiency. His approach can best be explained graphically, by considering a process that produces a single output, Q, with two inputs, x_a and x_b (see Figure 5.1). All points along the isoquant, such as A and B, are technically efficient but only one point (point A, at the intersection of the isoquant and isocost lines) on the isoquant is economically efficient.

Figure 5.1 Graphical Interpretation of Technical and Economic Inefficiency



Consider the quantity of output produced at point F. At this point, there is technical and economic inefficiency. Farrell measures technical efficiency as the ratio $\frac{OB}{OF}$. Because at F more inputs are used to produce Q than is technically necessary, OF is greater than OB, $\frac{OB}{OF}$ is less than unity, and thus F is technically inefficient. But at point B (as well as point F) there is also economic inefficiency, because the least costly combination of inputs to produce Q is not chosen. Economic efficiency of point B is measured as $\frac{OD}{OB}$, where D is a point on the isocost line if it were to move to a point such as A. Since OB is greater than OD, $\frac{OD}{OB}$ is less than unity, and therefore there is economic inefficiency. Farrell thus states that a healthcare provider's deviation from minimum cost (or from maximum economic efficiency) can be attributed to, or decomposed into, technical and economic inefficiencies. Therefore:

$$Overall \quad Economic \quad Efficiency = Technical \quad Efficiency \times Allocative \quad Efficiency$$

Using the example from the figure, overall economic efficiency of production at point F would be

$$OE_F = \frac{OD}{OF} = \frac{OB}{OF} \times \frac{OD}{OB}$$

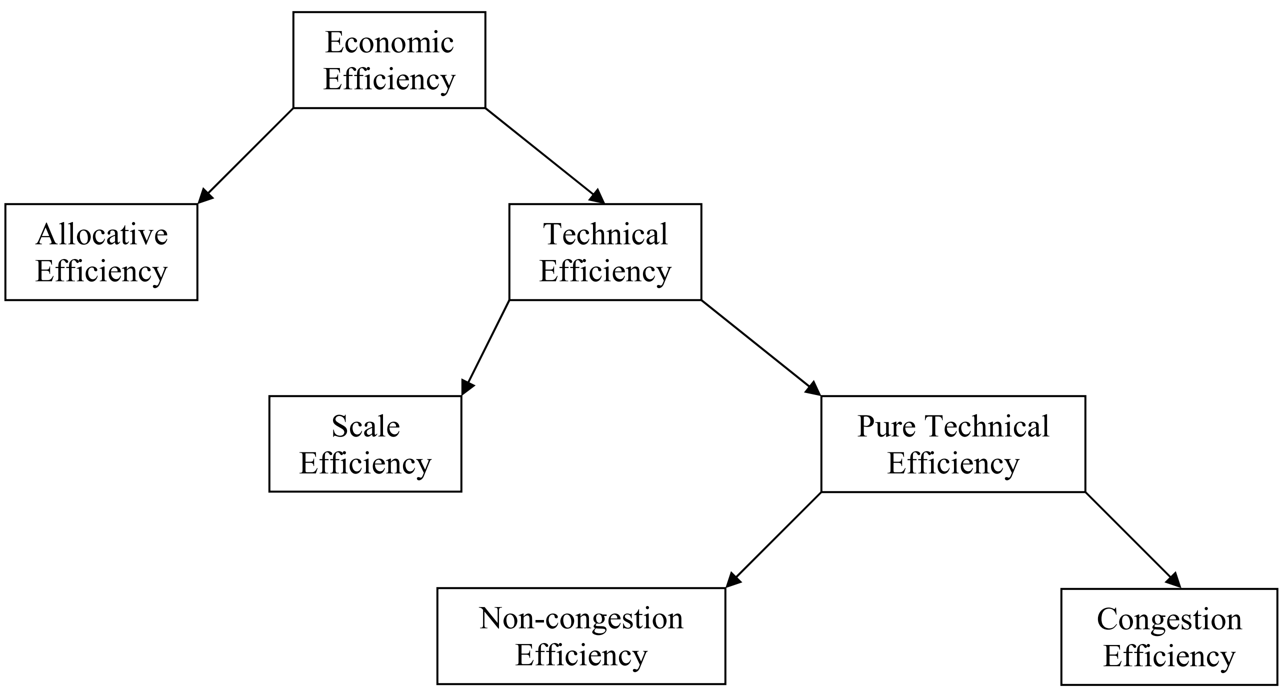
Technical efficiency refers to the conversion of inputs such as employees and machines into outputs relative to best practice. Technical efficiency is affected by managerial practices and the scale or size of operations. It relates to obtaining the greatest possible production of goods and services from available resources. In other words, resources are not wasted in the production process. This is also considered as engineering efficiency and should be contrasted with economic or allocative efficiency. Technical efficiency means that natural resources are transformed into goods and services without waste, that producers are doing the best job possible of combining resources to produce goods and services. There is no waste of material inputs. The maximum amount of physical production is obtained from the given resource inputs. In essence, production is achieved at the lowest possible opportunity cost.

Allocative efficiency refers to whether, given input prices, inputs are chosen to minimize the cost of production. Technical efficiency is a prerequisite for allocative and economic efficiency. Economic efficiency is achieved if the highest possible level of satisfaction is obtained from given resources. Because satisfaction is derived from consuming goods and services, economic efficiency requires the greatest possible level of production, that is, technical efficiency. However, while technical efficiency is necessary for economic efficiency, it does not guarantee economic efficiency.

The measurement of micro level efficiency involves the comparison between the observed and the optimal usage of inputs to produce an amount of output for each observation in the sample. Optimal input or output values are determined by the potential production possibilities, that is, the best observed practice in the sample. In this context, efficiency and productivity are defined using the values and ratios of "useful" inputs to outputs (Webster et al., 1998).

Nunamaker (1985) defines technical efficiency as a measure of the capacity of a micro level unit (referred to as a firm, observation or Decision Making Unit (DMU)) to avoid waste by producing as much output as input usage will allow, or using as little input as output level will allow. Allocative efficiency measures the ability of a DMU to avoid waste by producing a level of output at the minimal possible cost. Another decomposition occurs at the level of technical efficiency, which can be considered to be comprised of scale and non-scale effects, the latter being referred to as pure technical efficiency. Scale efficiency is the measure of the ability of a DMU to avoid waste by operating at the most productive scale. Finally, pure technical efficiency can be considered to be composed of congestion efficiency and other effects. Input congestion efficiency is the measure of the component of pure technical efficiency because of the existence of negative marginal returns to input, and the inability of a firm to dispose of unwanted inputs costlessly. The inability to costlessly dispose of unwanted inputs is referred to as weak disposability of inputs. Figure 5.2 sets out the progression of efficiency measures outlined above.

Figure 5.2 Decomposition of Economic Efficiency



Source: Adapted from Webster et al. (1998).

The level of technical efficiency of a particular firm is characterized by the relationship between observed production and some ideal or potential production

(Greene, 1993). The measurement of firm specific technical efficiency is based upon deviations of observed output from the best production or efficient production frontier. If a firm's actual production point lies on the frontier it is perfectly efficient. If it lies below the frontier then it is technically inefficient, with the ratio of the actual to potential production defining the level of efficiency of the individual firm.

The definition of technical efficiency by Farrell in 1957 led to the development of methods for estimating the relative technical efficiencies of firms. The common feature of these estimation techniques is that information is extracted from extreme observations from a body of data to determine the best practice production frontier (Lewin and Lovell, 1990). From this the relative measure of technical efficiency for the individual firm can be derived. Despite this similarity the approaches for estimating technical efficiency can be generally categorized under the distinctly opposing techniques of parametric and non-parametric methods (Seiford and Thrall, 1990).

5.3 Input versus Output Orientation

Output-oriented technical efficiency refers to a firm’s ability to obtain maximum output from a given amount of inputs. Formally, the level of technical efficiency is measured by the distance a particular firm is from the production frontier. Thus, a firm that sits on the production frontier is said to be technically efficient. This concept is important to firms because their profits depend highly upon their value of technical efficiency. Two firms with identical technologies and inputs but different levels of technical efficiency;

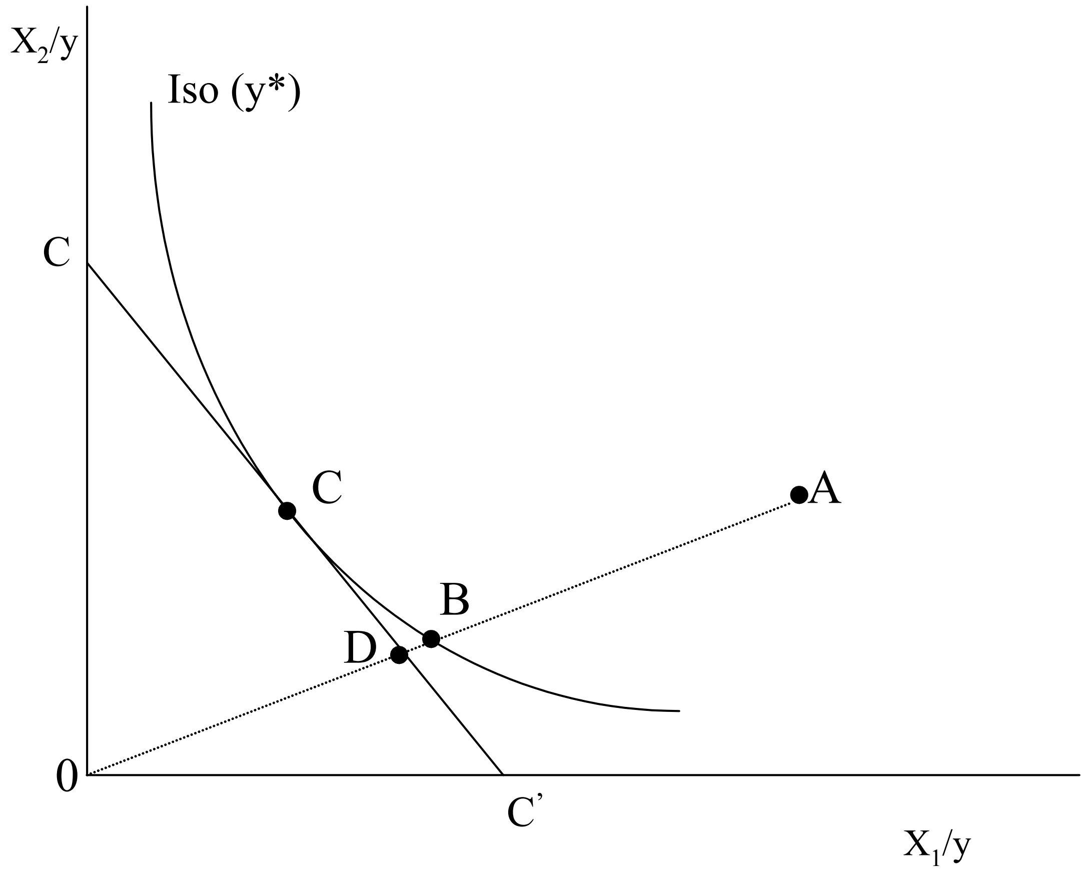
will have different levels of output. This will create higher revenue for one firm although both have the same costs, obviously generating a larger surplus for the more efficient firm. Conversely, Input-oriented technical efficiency refers to a firm's ability to minimize inputs from a given amount of output. It is noteworthy that Atkinson and Cornwell (1993) show that these two values are the same under the assumption of constant returns to scale.

The main reasons for examining technical efficiency as opposed to another type of efficiency are expressed by Kumbhakar and Lovell (2000). They state that technical efficiency is a purely physical notion that can be measured without recourse to price information and having to impose a behavioral objective on producers. It is well known that price data is often difficult to find and/or flawed. For this reason alone, one might decide to focus on technical efficiency. On the other hand, cost, revenue and profit efficiency are economic concepts whose measurement requires both price information and the imposition of an appropriate behavioral objective on producers. In addition, measuring output based technical efficiency seems to be more relevant in real life scenarios. A firm could more easily attempt to increase output with a given amount of inputs rather than decrease inputs to produce a given amount of output. In many cases, inputs lack liquidity or are costly to eliminate (e.g. unemployment benefits).

In Figure 5.3, the firm is producing a given level of output y^* using an input combination defined by point A. The same level of output could have been produced by radially contracting the use of both inputs back to point B, which lies on the isoquant associated with the minimum level of inputs required to produce y^* (i.e. Iso(y^*). The input-oriented level of technical efficiency (TE_I (y, x)) is defined by 0B/0A. However, the least-cost combination of inputs that produces (y^*) is given by point C (i.e. the point where the marginal rate of technical substitution is equal to the input price ratio w_2/w_1). To achieve the same level of cost (i.e. expenditure on inputs), the inputs would need to be further contracted to point D. The cost efficiency (CE(y, x, w)) is therefore defined by 0D/0A. The input allocative efficiency (AE_I (y, w, w)) is subsequently given by CE(y, x, w)/TE_I (y, x), or 0D/0B in Figure 5.3 (Kumbhaker and Lovell, 2000).

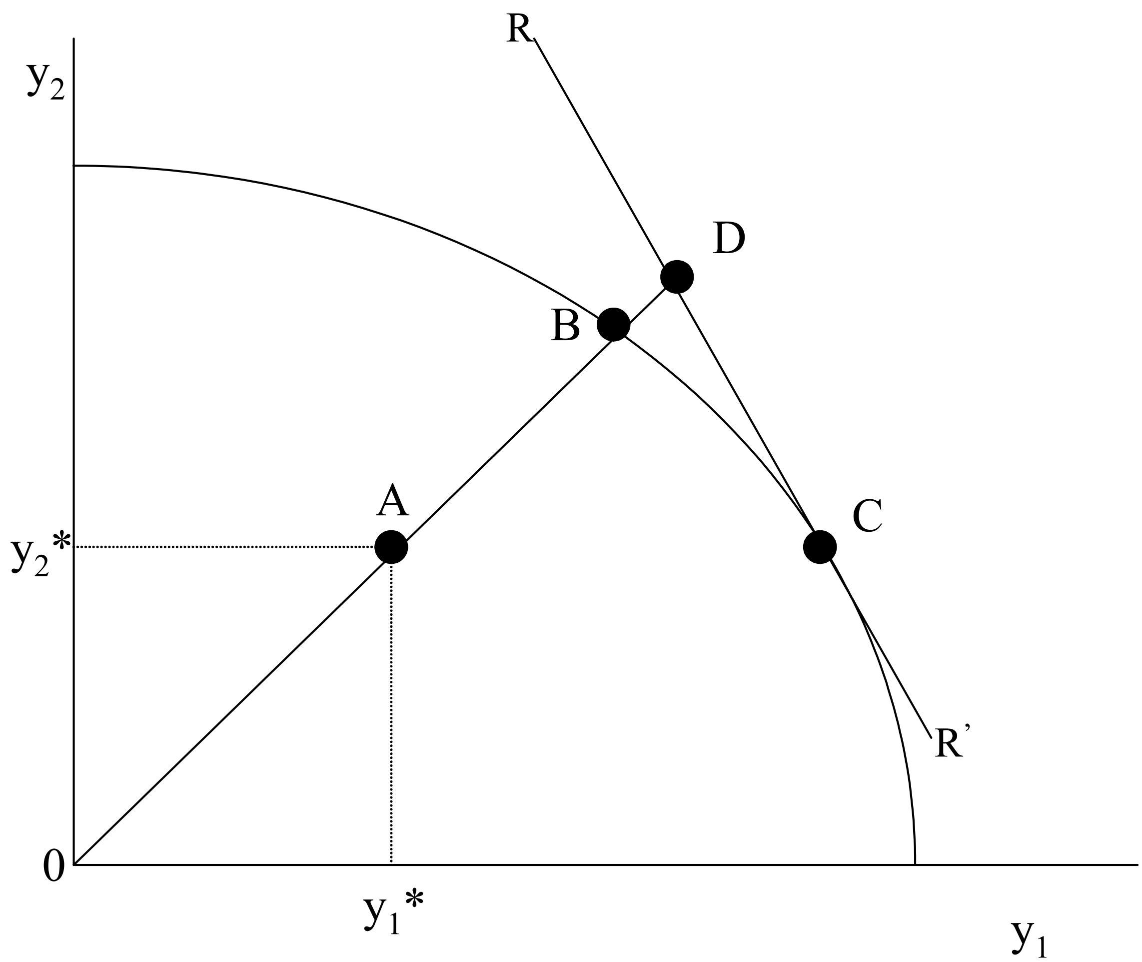
The production possibility frontier for a given set of inputs is illustrated in Figure 5.4 (i.e. an output-orientation). If the inputs employed by the firm were used efficiently, the output of the firm, producing at point A, can be expanded radially to point B. Hence, the output oriented measure of technical efficiency (TE_O (y, x)), can be given by 0A/0B. This is only equivalent to the input-oriented measure of technical efficiency under conditions of constant returns to scale. While point B is technically efficient, in the sense that it lies on the production possibility frontier, a higher revenue could be achieved by producing at point C (the point where the marginal rate of transformation is equal to the price ratio p_2/p_1, given by the line RR'). In this case, more of y_1 should be produced and less of y_2 in order to maximize revenue. To achieve the same level of revenue as at point C while maintaining the same input and output combination, output of the firm would need to be expanded to point D. Hence, the revenue efficiency (RE (y, x, p)) is given by 0A/0D. Output allocative efficiency (AE_O (y, w, w)) is given by RE(y, x, w)/TE_I (y, x), or 0B/0D in Figure 5.4 (Kumbhaker and Lovell, 2000).

Figure 5.3 Input-Oriented Efficiency Measures



Most studies use input-oriented specifications, whereby the focus is on the minimum input usage for given output levels. Any hospital utilizing more inputs to produce the same amount of outputs as compared to its peers would be deemed inefficient. Alternatively, an output-based model is used to demonstrate possible increases in outputs given fixed levels of inputs. The choice of model depends on the objective in question. The present study uses input-oriented DEA models due to the fact that hospital managers and administrators cannot influence the demand for healthcare[10] services they provide but rather the supply of healthcare services.

Figure 5.4 Output-Oriented Efficiency Measures



[10] This is dictated by the healthcare seeking behavior of patients.

5.4 Data Envelopment Analysis Model

Charnes et al. (1978) propose a Data Envelopment Analysis model which had an input-orientation and assumed Constant Returns to Scale (CRS). They specify a fractional linear program that computes the relative efficiency of each Decision Making Unit (DMU) by comparing it to all the other observations in the sample. Their exposition proceeds as follows.

This study adapts the methodology by Coelli (1996) and suppose that there is data on K inputs and M outputs on each of N firms or Decision Making Units (DMUs) as they are referred to in the DEA literature (for the present study the hospitals are the DMUs). For the i^{th} hospital these are represented by the vectors x_i and y_i, respectively. The KxN input matrix, X, and the MxN output matrix, Y, represent the data of all the hospitals. DEA constructs a non-parametric envelopment frontier over the data points such that all observed points lie on or below the production frontier.

By means of duality in linear programming, the input-oriented CRS DEA model can be specified as:

$$
\begin{aligned}
& Min_{\theta,\lambda}\theta \\
& subject \quad to \\
& -y_i + Y\lambda \geq 0, \\
& \theta x_i - X\lambda \geq 0, \\
& \lambda \geq 0
\end{aligned}
\qquad (5.1)
$$

where θ is a scalar and λ is an Nx1 vector of constants. This envelopment form entails fewer constraints than the multiplier form (K+M < N+1), and therefore is the generally preferable form to solve. The value of θ obtained will be the efficiency score for the i^{th}

hospital. It will satisfy $0 \le \theta \le 1$, with a value of 1 showing a point on the production frontier and therefore a technically efficient hospital according to Farrell's (1957) definition. It is worth noting that the linear programming problem must be solved N times, once for each hospital in the sample to yield a value of θ.

The CRS assumption is only appropriate when all hospitals operate at an optimal scale. Constraints in the operating environment for instance imperfect competition, financial and human resource constraints amongst other factors may cause a hospital to operate at non-optimal scale. Banker et al. (1984) suggest an extension of the CRS DEA model to provide for Variable Returns to Scale (VRS) situations. The use of the CRS specification when not all hospitals are operating at the optimal scale will result in a measure of technical efficiency which is confounded by scale efficiency. Therefore VRS DEA specification which permits the calculation of scale inefficiency was used.

The CRS linear programming problem can be modified to account for VRS by adding the convexity constraint: $N1'\lambda = 1$ to equation (5.1) where N1 is an Nx1 vector of ones, (Coelli 1996). This approach forms a convex hull of intersecting planes which envelope the data points more tightly than the CRS canonical hull and thus provides technical efficiency scores which are equal to or greater than those obtainable by means of the CRS model.

5.5 Modeling Production with Undesirable Output

Not all of the output from any production process might be desirable. Thus, the inclusion of only desirable output might not provide a true picture of the technical efficiency of a decision making unit. Many production processes have a negative side effect, namely the production of undesirable byproducts. A typical example in health is the death of a patient in the course of administering treatment. Traditional production theory does not lend itself to modeling joint production of good and bad outputs. To address this issue, Shephard and Färe (1974), introduce the notion of null-joint production which explicitly allows for the joint production of desirable and undesirable outputs/products. They also introduce the idea that disposal of undesirable outputs may not be 'free', which is the standard assumption made in traditional production theory. As an alternative, they propose the notion of 'weak' disposability; that is, disposability of some outputs may be costly, at least in the sense that reducing them may require diversion of resources away from the production of good outputs.

Another feature of undesirable outputs is that they are often not readily marketable, and therefore have no observable market prices. Using the notions of null-joint production and weak disposability, the shadow prices of these outputs may be retrieved using duality theory. Armed with a formal model of joint production and costly disposability, one can proceed to develop performance measures which take these features into account. As it turns out, the directional distance function is extremely useful in this case since it allows for simultaneous expansion of desirable outputs and contraction of undesirable outputs in assessing performance.

The production of desirable output is often accompanied by simultaneous or joint production of undesirable outputs. In our case, the primary desirable output is the treatment of the patient through medical attention/care and improved health status, whereas, the undesirable output is the death of the patient. Obviously, there are patients who leave the healthcare institution worse off than they arrived. However, for ease of modeling, we assume that there are two outcomes after receiving medical attention, namely being discharged alive or dead.

The problem facing the hospital's medical staff which makes decisions regarding the utilization of resources in order to improve the treatment of patients and thus the patient’s probability of survival, is that in the effort to save lives there is a risk of incurring patient deaths. Patient deaths may occur for any number of reasons: diagnostic or treatment failure, errors in judgment of medical staff, or as a random[11] occurrence beyond the medical staff's control. This simultaneous production of desirable (alive patients) and undesirable output (patient deaths) implies that reducing the bad output is costly in terms of increased technological capability or increased diagnostic and treatment capability of the medical staff.

Modeling undesirable outputs for instance, patient deaths, pollution[12] and other detrimental side-effects of production activities (such as noise) has attracted considerable attention among production economists. A common approach is to treat detrimental variables as inputs, based on the economic argument that both inputs and detrimental side-effects incur costs for a firm and one is thus usually interested in

[11] It is should be noted that DEA does not account for random error.
[12] For instance, emissions of harmful substances in air, water and ground.

decreasing both types of variables as much as possible (Cropper and Oates, 1992). Borrowing from Cropper and Oates (1992), one can conjecture that healthcare providing institutions implicitly seek to minimize their usage of factor inputs as well as undesirable outputs (for instance, the death of their clients). This is because, although it is the case that some patients die in the process of getting medical care, the overriding motive of a healthcare providing institution is to improve the health status of the patients or save their lives. However, Färe and Grosskopf (2003, 2004), argue that the treatment of undesirable outputs as inputs is inconsistent with the physical laws and the standard axioms of production theory. These authors advocate an alternative approach that models undesirable outputs by imposing an assumption that these are weakly disposable.

The measurement of hospital technical efficiency when both desirable and undesirable outputs are produced requires the explicit provision for their joint production. This study adapts the methodology by Dismuke and Sena (2001) to model deaths in hospital efficiency measurement. Denoting desirable outputs (if a patient is discharged alive) by $y \in \Re_{+}^{M}$, undesirable output (if a patient is discharged dead) by $b \in \Re_{+}^{I}$, and inputs by $x \in \Re_{+}^{N}$, then the technology may be written as

$$T = \{(x, y, b) : x \quad can \quad produce \quad (y, b)\} \tag{5.2}$$

The technology consists of all feasible input and output quantities, i.e., it consists of all desirable and undesirable outputs that can be produced by the given input vectors.

It is convenient to model the technology of the joint production of the good and bad outputs in terms of the output sets, i.e.,

$$P(x) = \{(x,b):(x,y,b) \in T\}\ . \tag{5.3}$$

The technology T can be recovered from $P(x)$ as follows:

$$T = \{(x,y,b):(y,b) \in P(x), x \in \Re_{+}^{N}\ . \tag{5.4}$$

Therefore, the technology is equivalently represented by either its technology set T or its output set $P(x), x \in \Re_{+}^{N}$.

One important feature from the viewpoint of healthcare organizations/institutions (hospitals in the context of this study) is that it is costly to reduce patient deaths or other undesirable outputs. This idea is modeled by imposing the weak disposability of outputs assumption, i.e.,

$$(y,b) \in P(x) \quad and \quad 0 \le \theta \le 1 \quad imply \quad (\theta y, \theta b) \in P(x) \tag{5.5}$$

Equation (5.5) states that reduction of patient deaths (undesirable outputs) is feasible if good outputs are also reduced, given fixed input levels. With respect to the good outputs, it is assumed that they are freely or strongly disposable, i.e., $(y,b) \in P(x) \quad and \quad y' \le y \quad imply \quad (y',b) \in P(x)$. Therefore, it is not feasible to reduce the undesirable outputs only, i.e., if (y,b) is feasible and $b' \le b$ then it may not be possible to produce (y,b') using x, i.e., $(y,b) \in P(x)$ and $(y,b') \notin P(x)$. This problem would not arise providing that undesirable outputs could be disposed of freely or costlessly. Desirable and undesirable outputs must be distinguished in terms of their disposability because on the one hand, desirable outputs have positive prices, while undesirable outputs are non marketable and thus do not have readily observable prices.

The notion that desirable and undesirable outputs are jointly produced is modeled by what Shephard and Färe (1974) refer to as null-jointness. This means that if no bad outputs are produced, then there can be no production of good outputs. If a hospital wishes to produce some live discharges, then there will be byproducts of undesirable outputs (patient deaths). Formally, we have

$$(y,b) \in P(x) \quad and \quad b=0 \quad then \quad y=0 \tag{5.6}$$

That is, if (y,b) is a feasible output vector composed of both desirable outputs (live discharges) y and undesirable outputs (patient deaths) b, then if no undesirable outputs are produced $(b=0)$ then by null-jointness, the production of positive desirable outputs is not feasible, thus $y=0$. Equation (5.6), states that the desirable outputs are "null-joint" with the undesirable outputs if the only way not to produce undesirable outputs is by not producing desirable output. In other words, the hospital must risk having some patient deaths in the effort to produce live patients (live discharges).

5.6 Incorporating Undesirable Outputs in DEA Models

Data Envelopment Analysis assumes non-negativity of all inputs and outputs. This assumption is not, however, always satisfied as it was the case in our application to hospitals, which led to the development of alternative models aiming at assessing efficiency in the presence of negative data. For DEA efficiency valuations it is crucial to choose appropriate inputs and outputs and make some general assumptions about the technology structure with regard to convexity, disposability and returns to scale. In the presence of negative data the use of radial measures of efficiency traditionally used in DEA is problematic.

Classical DEA models as described in Charnes et al. (1994) rely on the assumption that inputs have to be minimized and outputs have to be maximized. DEA assumes that inputs and outputs are isotonic, that is, increased input use reduces efficiency, while increased output increases efficiency. However, this is not always the case for some input and output variables. However it was mentioned already in the seminal work of Koopmans (1951) that the production process may also generate undesirable outputs like smoke pollution or waste.

The treatment of undesirable outputs has similarities with the treatment of negative outputs since both should be contracted rather than expanded[13]. Several approaches exist to deal with undesirable outputs as can be seen in the review by Allen (1999) as well as Dyckhof and Allen (2001). The approaches for incorporating undesirable outputs in DEA may be categorized into direct and indirect approaches (Scheel, 2001).

Direct Approaches. On the one hand, direct approaches use the original output data but modify the assumptions about the structure of the technology set in order to treat the undesirable outputs appropriately.

The direct approach suggested by Färe et al. (1989) replaces strong disposability of outputs by the assumption that outputs are weakly disposable while only the sub-

[13] Note that this wording is only valid when it refers to multiplying factors associated to negative data. Indeed, in the case of negative outputs one wants to increase good outputs, while at the same time decreasing undesirable outputs.

vector of good outputs is strongly disposable. Other direct approaches use the directional distance function, first proposed by Chung et al. (1997) to deal with negative output data. The main advantage of the directional distance approach over the existing approaches is that it is able to provide efficiency scores, similar in meaning to radial efficiencies traditionally used in DEA, while at the same time negative data are used without the need to subjectively transform them. Finally the approach yields targets that are, in general, easier to achieve than those resulting from the additive model.

Indirect Approaches. Conversely, indirect approaches transform the values of the undesirable outputs by a monotone decreasing function f such that the transformed data can be included as “normal” (desirable) outputs in the technology set T (since after retransformation increasing these values means decreasing the undesirable outputs).

The indirect approaches suppose that their transformed data are meaningful (for instance, consider the bad output “mortality rate” and its translated additive inverse “survival rate”), the direct approach uses the original output data by assuming that it is not possible to decrease bad outputs without simultaneously decreasing good outputs. Outputs are *strongly disposable* if $(x, y) \in T$ implies that $(x, y') \in T$ for every $y' \leq y$ and *weakly disposable* if $(x, y) \in T$ implies that $(x, \mu y) \in T$ for $0 \leq \mu < 1$.

Suppose that B represents the matrix of bad output data such that each row represents a given DMU whilst each column captures one bad output. In the presence of bad outputs, DMU k is efficient if there is no vector (x', b', y') in the technology set such that $x \leq x^k, b' \leq b^k \quad and \quad y' \geq y^k$ with at least one strict inequality.

Data Transformation. Traditionally negative data is handled in efficiency applications through some data transformation so that all negative data are converted into positive data (see for example, Pastor, 1994; Lovell, 1995). Such transformation of the data may have implications for the solution, classification, or ordering of the DEA results (Seiford and Zhu, 2002). There are some models whose solutions are invariant to data transformations, usually referred to as translation invariant. In the presence of negative data the most often used model is the variable returns to scale (VRS) additive model of Charnes et al. (1985), which is translation invariant as demonstrated by Ali and Seiford (1990). A translation invariant model is such that 'an affine displacement of data does not alter the efficient frontier' (Ali and Seiford, 1990). The BCC model has been found translation invariant (Ali and Seiford, 1990). However, if the efficiency scores should in addition not be affected, then the BCC output oriented model allows a translation of inputs and the input oriented model of outputs (Lovell and Pastor, 1995, Pastor, 1996).

The additive DEA model of Cooper, Thompson and Thrall (1996) is translation invariant. However, this model does not produce the usual efficiency scores (i.e., between 0 and 100 percent) and may not be as easily interpretable. An alternative is to recognize that the Banker, Charnes, Cooper (BCC) VRS input oriented DEA model is in fact invariant to any output translations.

The additive model is not however, in its original form, units invariant (independent of scales of measurement). Due to this limitation Lovell and Pastor (1995) put forward a units invariant version of the additive model that uses a weighted sum of slacks where the weights are the inverse of the standard deviations of each input and output (see also Pastor, 1996; Thrall, 1996) corresponding to the slack. The main drawback associated with the additive model is the fact that it yields in respect of an inefficient unit the `furthest' targets on the production frontier, besides not yielding an efficiency measure. Thus the model does not yield very practical guidance as to how a unit might improve its performance nor does it make it possible to readily rank units on performance.

The translation invariance of the additive model is subject to it being specified under VRS. Constant returns to scale (CRS) models are not translation invariant because they do not impose the sum of the intensity variables to equal unity (Banker et al., 1984). If data cannot be translated in a CRS model without changing the model's solution, then the issue is whether CRS models can be used in the presence of negative data.

The other way of incorporating bad outputs suggested by Koopmans (1951) and applied by Berg et al. (1992) is based upon a transformation known as the additive inverse. The bad outputs are incorporated as good outputs with values $f(B) = -B$. Apart from the sign of the bad outputs it generates the same technology set as incorporating bad outputs B as inputs. The classification of DMUs as efficient or inefficient is the same when the bad outputs are transformed via f or incorporated as inputs into the technology set. The classification is preserved if the values of bad outputs are "translated" in the sense of Ali and Seiford (1990). According to the methodology by Ali and Seiford (1990), one adds to the additive inverse of the bad output i a sufficiently large scalar λ_i such that the resulting output values $f_i^k(B) = -b_i^k + \lambda_i$ are positive for each DMU k.

For the present study, suppose that y_{rj}^g and y_{rj}^b represent the good (desirable) and negative (bad or undesirable) outputs, respectively[14]. It is the case that we wish to reduce y_{rj}^b while at the same time increasing y_{rj}^g in order to improve the hospital's performance. In the output-based BCC envelopment model, however, both y_{rj}^g and y_{rj}^b are supposed to improve performance. To increase the good outputs and to reduce the bad outputs, Zhu (2003) proceeds as follows. Each undesirable output is multiplied by "-1" and then a proper value v_r is found to let all negative undesirable outputs be positive, i.e. $\overline{y}_{rj}^b = -y_{rj}^b + v \rangle 0$. This can be attained for instance by $v_r = \max\{y_{rj}^b\} + 1$.[15] The use of a linear transformation preserves the convexity and as such is a good choice for a DEA model. This study adapts the methodology by Zhu (2003) and the technical efficiency scores when patient deaths are incorporated into the analysis are estimated by means of DEA Excel Solver[16].

---

[14] These can also be termed 'less-is-better outputs'.

[15] Alternatively, a monotone decreasing transformation (for instance, $1/y_{rj}^b$) may be applied to the undesirable outputs then using the adapted variables as outputs.

[16] This is included in Zhu (2003).

5.7 Unit of Analysis, Sample Size, Data Collection, Sample Selection and Modeling Choices

The study seeks to investigate the technical efficiency of district referral hospitals. As such, the study's unit of analysis is the district referral hospital. Twenty five (25) district referral hospitals drawn as follows, seven (7) from the Eastern, eight (8) from Western[17] and ten (10) from the Central regions of Uganda constitute the study sample. The Northern region was left out due to security concerns during data collection and because the operating environment of hospitals in this region is not comparable to that of their counterparts in the remaining regions. The reason for this being that the region has been insecure for the last 18 years or so and therefore including it would bias the sample. Twenty five (25) out of the 38 district referral hospitals in Uganda have been covered (see Table 5.1).

Table 5.1 Sample District Referral Hospitals by Region

Region	Hospital
Central	Entebbe, Gombe, Kalisizo, Kawolo, Kayunga, Kiboga, Nakaseke, Mityana, Mubende and Rakai.
Eastern	Bududa, Bugiri, Busolwe, Iganga, Kapchorwa, Pallisa and Tororo.
Western	Bwera, Itojo, Kagadi, Kambuga, Kiryandongo, Kisoro, Kitagata and Masindi.

There is a conscious attempt to account for the heterogeneity of the hospital environment. The sample of hospitals is limited to district referral hospitals indicating that the 'care mix' can be assumed to be fairly comparable. The assumption is that

[17] Bundibugyo district referral hospital was also left out due in part to insecurity and poor accessibility during data collection.

hospitals of similar organizational form produce similar types of health care. Because the sample hospitals have the same scope of service, it is reasonable to assume homogeneity in the range of health care services they provide.

The choice of the sample size, number of inputs as well as the number of outputs was guided by the rule of thumb proposed by Banker and Morey (1989), that $n \geq 3(m+s)$, where: n is the number of DMUs included in the sample; m is the number of inputs; and s is the number of outputs included in the analysis. The rule captures two issues, sample size and number of factors $[(m+s)]$. However, Pedraja-Chaparro et al., (1999) note that the rule ignores two other issues, distribution of efficiencies as well as the covariance structure of factors. Nevertheless, we still use the rule of thumb as a guide in the absence of any a priori view on the number of factors.
A schedule containing the data needed for the study (hospital inputs as well as outputs) was constructed. The schedule was piloted on three district referral hospitals which included Nakaseke, Kayunga and Entebbe. There was a discrepancy between the initial research instrument and the Health Management Information System databases. After the pilot study, the schedule was adjusted to the Health Management Information System (HMIS) databases.

A panel data set was assembled and a common set of input and output indicators was constructed to support the estimation of Data Envelopment Analysis (DEA) models. Input as well as output data were gathered for the twenty five hospitals over the 1999-2003 period. The potential gains from using panel data to measure technical efficiency appear to be quite large. A panel obviously contains more information about a particular Decision Making Unit than does a cross-section of the data.

The HMIS launched in 1997 is the source of the data for the present study. However, the study concentrates on the period 1999-2003 because this period yielded a balanced panel. Data on the hospital inputs and the outputs were sought from the HMIS databases of each hospital. Twenty five out of 38 district referral hospitals were selected in the regions of Western, Eastern and Central Uganda due to the decentralized delivery of healthcare services and their being conducive for data collection compared to the Northern region. Comparability of data across hospitals was ensured by a common database that all public district referral hospitals are required to submit to the District Director of Health Services on a monthly and annual basis. The HMIS captures data on a calendar year basis. Administrative data and annual reports were collected at each hospital to generate the dataset. Unfortunately, financial data on a majority of hospitals were not readily available and as a consequence, the variable total operating costs has been left out.

Specifically, data on admissions, deaths, inpatient days by ward[18] as well as surgical operations, outpatient department attendances was collected from the Hospital Annual Reports. In-hospital mortality was used to account for quality of care whilst a length of stay-based case-mix index was computed to provide for the heterogeneity of admissions.

[18] Inpatient days and admissions by ward were employed in the computation of the casemix index for each hospital.

DEA models to measure technical efficiency as well as DEA-Malmquist models to measure productivity are estimated by means of DEAP version 2.1; a Data Envelopment Analysis (DEA) Program developed by Coelli.[19] In order to check the stability and sensitivity of DEA results, a multi-pronged approach is adopted in the analysis of DEA results. This includes assessment of the efficiency of the sample hospitals, inclusion/exclusion of inputs/outputs, providing for case-mix in each hospital's patient load, analyzing the correlation between different models over time, running the models both on the cross-sectional and pooled datasets and assessing the performance of hospitals across all models based on their efficiency scores and rankings.

5.8 Choice of Inputs and Outputs

A typical healthcare institution like a hospital embraces a variety of resources (human, material and knowledge amongst others), which are used in a series of processes that ultimately aim to improve upon the medical condition of the patient and contribute to healthier communities.

The estimation of technical efficiency requires the careful choice of the sample size as well as the number of factors (number of inputs plus the number of outputs). A Data Envelopment Analysis (DEA) study requires the careful selection of inputs and outputs. This is due to the fact that the distribution of efficiency is likely to be affected by the definition of outputs and the number of inputs and outputs included (Magnussen, 1996).

[19] Centre for Efficiency and Productivity Analysis, University of New England, Australia.

Theoretically, improved health status is the ultimate output of hospitals or the health care system generally. Nevertheless, the measurement of health status poses difficulties because health is multi-dimensional and there is subjectivity involved in assessing the quality of life of patients (Clewer and Perkins, 1998). Because of the difficulty of accurately measuring improvement in health status, hospital output is measured as an array of intermediate outputs (health services) that improve health status (Grosskopf and Valdmanis, 1987).

The measures used in the study represent the general areas of direct services which hospitals provide to patients. Attempts are made to incorporate a fairly comprehensive list of inputs and outputs which reflect the general scope of hospital activities in order to obtain informative and robust results. However, the fact that DEA operates more powerfully when the number of DMUs exceeds the number of the combined total of inputs and outputs by at least twice (Drake and Howcroft, 1994) restricts the input and output measures chosen for the study.

Input Variables: Four inputs are constructed and include doctors, nurses, other staff, and beds. Due to lack of information on Full Time Equivalent staff, the study uses absolute numbers of human resources providing health care services to approximate the labor resources employed. Because there is some variation in how the hospitals record their staff in the registers, the study minimizes this by combining labor categories into three variables: 'doctors', 'nurses' and 'other employees'. The variable "doctors" includes all senior medical officers, medical officers as well as dental surgeons. The variable 'nurses' includes senior nursing officers, nursing officers, Uganda registered nurses, midwives, enrolled midwives, enrolled nurses, nursing assistants, and nursing aids. Finally, the variable "other staff" includes clinical officers, dispensers, anesthetics officers, radiographers, orthopedic officers, laboratory technologists and technicians, laboratory assistants, hospital administrator, accountant, clerical officers, supplies officers, stores assistants, telephone operators, stenographers, copy typists, records assistants, dark room attendants, mortuary attendants, drivers, kitchen attendants, security guards, artisans (carpenters), electrical technicians and plumbers. All the three staffing measures include only salaried hospital staff. It should be noted that the inclusion of only salaried staff might understate the hospitals' human resource complement.

There were no data for capital inputs for instance buildings and equipment. As a consequence, capital is approximated by the number of beds per hospital. Beds are often used to proxy for capital stock in hospital studies usually because a reliable measure of the value of assets is rarely available. District referral hospitals are distinguished from other public hospitals as being 100-bed hospitals. Nevertheless, the bed stock has been on the increase in some hospitals as they try to cope with increasing numbers of admissions. Moreover, in most hospitals due to limited bed capacity, there are what can be termed "floor admissions".[20] In the ideal world no hospital would admit when its bed stock is exhausted. However, being the only hospital with relatively 'free' healthcare in the district, admissions beyond available bed capacity are admissible given that patients

[20] Hospital records do not clearly distinguish "bed admissions" from "floor admissions" which complicates its tracking across hospitals and through time for a given hospital. They are all lumped together as admissions.

may have lack alternatives due partly to the high levels of poverty. These will unfortunately make some hospitals appear more efficient than others with respect to bed capacity as some of the hospitals' inpatients have no beds. This will also have implications for total factor productivity measures and in particular technology change.

Output Variables: The output measures focus on the process type or production volume style estimates of hospital output. The study examines a number of measures of district referral hospitals' output. These include admissions, deliveries, operations, and outpatient department attendances.

Inpatient Care: Inpatient care output for each district hospital was measured in two ways: first as annual cases treated, specifically annual admissions, and then as "case-mix adjusted" admissions. Case-mix adjusted admissions are defined as annual admissions times the case-mix index. The index is the (normalized) weighted sum of the proportions of the hospital's inpatients in different wards where the weights reflect the length of stay of its patient load. Case-mix adjusted admissions transform admissions into ward homogeneous patient loads. For a given level of admissions, the adjusted measure captures output differences due solely to case-mix variation. In particular, it controls for the fact that hospitals whose wards exhibit relatively longer average length of stay may be due to a more complex mix of patients compared to wards with relatively short average length of stay. The adjusted measure captures output differences due to variations in average length of stay, and by proxy, case-mix. While the data prohibits more detailed estimation of case-mix differences, this approach attempts to adjust output into more homogeneous and comparable groupings.

Deliveries: Deliveries include all deliveries in the hospital without adjusting for neonatal deaths because resources are expended irrespective of the status of the birth.

Surgical Operations: Surgical operations include major operations, minor operations as well as Caesarian sections.

Outpatient Department Attendances: Outpatient department attendances include new cases as well as re-attendances.

Patient Deaths: Patient deaths denote the total of dead discharges across the four wards (male, female, maternity and pediatrics).

A summary of variable definitions is provided in Table 5.2 while Table 5.3 contains descriptive statistics for the input and output variables for each sample year. The means and standard deviations reported in the table suggest that there are substantial variations across the sample with respect to the input and output variables.

Table 5.2 Definitions and Measurement of Input and Output Variables

Variables	***Definition and Measurement***
Inputs	
Beds	Total Number of beds
Doctors	Total Number of medical doctors (physicians, pharmacists, dentists, etc., including residents and interns)
Nurses	Total Number of nurses, including professional, enrolled, registered, community nurses, and nursing aids.
Other Employees	Total Number of paramedics and assistants, technicians and assistants; administrative staff; and other general staff.
Outputs	
Admissions	Total Annual Admissions
Outpatient Dept. Attendances	Annual Total Number of outpatient department attendances
Surgical Operations	Annual Total Number of surgical operations
Deliveries	Annual Total Number of deliveries in the hospital
Patient Deaths[21]	Annual sum of dead discharges from all wards

[21] Patient deaths represent undesirable or bad output of a hospital in the sense that they are of the form of "less is better."

The mean and standard deviation of inputs and outputs analyzed by the study are shown in Table 5.3 whereas Table 5.4 presents the Pearson correlation matrix of input and output variables. The mean and standard deviation vary marginally by year across the study period and for the pooled dataset. This implies that on average the variables display some degree of stability on a year to year basis across the study period and for the pooled dataset. In Table 5.4, supply-side factors are correlated, as are some measures of outputs (as expected) and where possible we will try to maintain parsimonious specifications and reduce double counting.

Table 5.3 Mean [and Standard Deviation] of Input and Output Variables

Variable/Year	1999 (n=25)	2000 (n=25)	2001 (n=25)	2002 (n=25)	2003 (n=25)	1999-2003 (n=125)
Inputs						
Beds	113.1	114.7	115.1	115.8	117.7	115.3
	[19.6]	[23.5]	[23.1]	[22.9]	[23.1]	[22.2]
Doctors	4.5	4.6	4.8	4.8	4.8	4.7
	[1.7]	[1.7]	[1.7]	[1.8]	[1.8]	[1.7]
Nurses	58.9	57.6	58.2	55.9	57.2	57.6
	[21.5]	[21.9]	[20.7]	[19.4]	[19.4]	[20.3]
Other Staff	64.4	66.0	67.4	65.7	65.6	65.8
	[28.0]	[28.6]	[27.3]	[24.9]	[25.2]	[25.4]
Outputs						
Admissions (unweighted)	7049.5	7063.3	7850.9	8185.4	8541.4	7738.1
	[4314.2]	[4738.1]	[5981.6]	[6363.1]	[6664.6]	[5627.3]
Casemix Adjusted	7052.6	7058.7	7845.4	8238.8	8571.1	7753.3
Admissions	[4298.0]	[4725.4]	[6400.6]	[6413.2]	[6284.0]	[5640.1]
Outpatient Attendances	29467.9	30482.0	35467.9	37373.4	36243.4	33806.9
	[14179.2]	[14033.7]	[14981.3]	[15046.6]	[17079.3]	[15201.7]
Surgical Operations	775.8	826.9	886.8	1046.5	1040.8	915.3
	[472.7]	[433.4]	[437.6]	[459.3]	[466.1]	[460.3]
Deliveries	1192.9	1148.1	1358.6	1474.5	1495.5	1333.9
	[475.8]	[506.2]	[529.8]	[612.3]	[666.9]	[571.6]
Patient Deaths	268.1	280.1	269.2	277.3	270.8	273.1
	[24.6]	[21.5]	[22.0]	[25.5]	[24.0]	[116.0]

Note:
Beds = Total number of available beds
Doctors = Total Number of medical doctors (physicians, pharmacists, dentists, etc., including residents and interns)
Nurses = Total Number of nurses, including professional, enrolled, registered, community nurses, and nursing aids.
Other Staff = Total Number of paramedics and assistants, technicians and assistants; administrative staff; and other general staff.
Admissions = Total annual admissions (Maternity, Male, Female and Pediatrics wards)
Case-mix Adjusted Admissions = Length-of-stay-based case-mix index multiplied by total annual admissions
Outpatient Attendances = Annual Total Number of outpatient department attendances.
Surgical Operations = Annual Total Number of surgical operations (major & minor operations and Caesarian sections)
Deliveries = Annual Total Number of deliveries in the hospital
Patient Deaths = Annual sum of dead discharges from all wards

Table 5.4 Pearson Correlation Matrix of Input and Output Variables (n=125), 1999-2003

	Admissions	Deaths	CAA	OPD	S/Operations	Deliveries	Beds	Doctors	Nurses	Other Staff
Admissions	1									
Patient Deaths	0.3719*	1								
CAA	0.9444*	0.2759*	1							
OPD	0.2527*	-0.0112	0.2481*	1						
S/Operations	0.2220*	0.5197*	0.1212	0.0581	1					
Deliveries	0.1483	0.4622*	0.0465	0.2469*	0.3941*	1				
Beds	0.1169	0.3844*	0.0474	-0.1106	0.4062*	0.4455*	1			
Doctors	-0.1238	0.2098*	-0.0689	0.2637*	0.0954	0.2859*	0.3487*	1		
Nurses	-0.0234	0.1141	-0.0222	0.3836*	0.1685	0.1577	0.2794*	0.4636*	1	
Other Staff	0.0245	0.0539	0.0178	0.2207*	0.1272	0.3335*	-0.0065	0.2457*	0.3364*	1

Note: *Significant at 5 percent level
CAA = Case-mix Adjusted Admissions
OPD = Outpatient Department Attendances
S/Operations = Surgical Operations (Minor, major and Caesarian sections)

Table 5.5 presents the six models estimated in the measurement of efficiency and productivity change. Modeling input-oriented DEA technical efficiency scores and DEA-Malmquist TFP index, model 1 includes four inputs namely beds, doctors, nurses and other staff, and four outputs admissions (un-weighted), outpatient department attendances, surgical operations and deliveries. Model 2 keeps the same inputs and outputs as Model 1 but replaces admissions (un-weighted) with case-mix adjusted admissions. Model 3 includes the same inputs as Models 1 and 2, as well as two outputs, case-mix adjusted admissions and outpatient department attendances. Models 4 and 5 have two inputs, beds and all staff grouped together; Model 4 includes the same outputs as Model 3 while Model 5 includes the same outputs as Model 2. The five models were run for individual years and the pooled dataset over the 1999-2003 period in the estimation of technical efficiency; and in the estimation of TFP change. Model 6 differs from model 1 in only one respect that it replaces outpatient department attendances with patient deaths. This is because patient deaths are more correlated to the other outputs (un-weighted admissions; surgical operations and deliveries) than to outpatient department attendances. However, it should be noted productivity was not estimated using model 6 for lack of an appropriate software package.

Table 5.5 Six DEA and Malmquist Model Specifications

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Inputs	*Inputs*	***Inputs***	***Inputs***	***Inputs***	***Inputs***
Beds	Beds	Beds	Beds	Beds	Beds
Doctors	Doctors	Doctors	All staff grouped together	All staff grouped together	Doctors
Nurses	Nurses	Nurses			Nurses
Other staff	Other staff	Other staff			Other staff
↓	↓	↓	↓	↓	↓
Outputs	*Outputs*	***Outputs***	***Outputs***	***Outputs***	***Outputs***
Admissions (un-weighted)	Admissions (case-mix adjusted)	Admissions (case-mix adjusted)	Admissions (case-mix adjusted)	Admissions (case-mix adjusted)	Admissions (un-weighted)
Outpatient attendances	Outpatient attendances	Outpatient attendances	Outpatient attendances	Outpatient attendances	Patient Deaths
Surgical operations	Surgical operations			Surgical operations	Surgical operations
Deliveries	Deliveries			Deliveries	Deliveries

In order to check the stability and sensitivity of DEA results, a multi-pronged approach is adopted in the analysis of DEA results. This includes simultaneous assessment of the efficiency of the sample hospitals and the inclusion/exclusion of inputs/outputs. In order to capture the variations in efficiency over time, Boussofiane et al., (1991) describe the following method. According to them, given n units with data on their input/output measures in k periods, then a total of nk units are assessed simultaneously. The study utilizes this method in its analysis. Following the methodology by Boussofiane et al., (1991), given 25 hospitals and data on their input/output measures over a 5-year period, a total of 125 hospitals are assessed simultaneously. This data pooling allows for a greater sample size and a comparison of efficiency estimates.

5.9 Providing for Case-mix

If the analysis to follow used inpatient days, deliveries, operations, as proxies for hospital output, a serious shortcoming in the analysis would exist: the failure to control for case-mix differences between hospitals. Specifically, while it might be the case that one hospital produces more outputs (e.g. inpatient days, operations, deliveries) for a given combination of inputs than another hospital, the first might be no more efficient if it consistently treats a relatively less sophisticated mix of cases, that is, a mix of cases requiring relatively fewer inputs per unit of output. Any study of hospital technical efficiency must then attempt to control for differences in the case mix between different hospitals.

Lacking data on individual hospital case mix as well as billing or cost data; the study adapted the English Department of Health's Casemix (Hernandez, 2002). The case-mix index (HI_i) for hospital i is approximated by means of the average length of stay to control for the case-mix among different hospitals as follows:

$$HI_i = \frac{\sum_j NALOS_j * Ad_{ji}}{TALOS * \sum_j Ad_{ji}} , \tag{5.7}$$

where:

HI_i	=	case mix index for hospital i;
$NALOS_j$	=	national weighted average length of stay for ward j;
Ad_{ji}	=	number of admissions in ward j (in hospital i);
$TALOS$	=	average weighted length of stay of wards;
$\sum_j Ad_{ji}$	=	total number of admissions treated by hospital i.

And

$$NALOS_j = \frac{\sum_i LOSAd_{ji} * Ad_{ji}}{\sum_i Ad_{ji}}, \qquad (5.8)$$

where:

$LOSAd_{ji}$ = unit length of stay of ward j 's admissions in hospital i;

Ad_{ji} = number of admissions in ward j (in hospital i);

$\sum_i Ad_{ji}$ = sum of ward j 's admissions for all hospitals.

And

$$TALOS = \frac{\sum_j \sum_i LOSAd_{ji} * Ad_{ji}}{\sum_j \sum_i Ad_{ji}}, \qquad (5.9)$$

where:

$LOSAd_{ji}$ = unit length of stay of ward j 's admissions in hospital i;

Ad_{ji} = number of admissions in ward j (in hospital i);

$\sum_j \sum_i Ad_{ji}$ = sum of all admissions for all hospitals.

The above approach to approximating the case mix index for a given hospital is premised on the assumption that the wards produce very similar types of output across hospitals. However the length of stay-based case-mix index has a number of shortcomings which include but are not limited to: (i) it is not based on individual level patient data (it does not account for age, gender, complexity); (ii) hospital wards may not use homogeneous definitions across hospitals; (iii) there is a likelihood of different length of stay policies across hospitals; (iv) length of stay is susceptible to outlier data (hospitals provide more than curative care for instance palliative care, social care, etc.,); and (v) discharges might be linked to the degree of integration with community care in which case hospitals might keep patients longer if there are weak community health service links. In order to address some of the weaknesses of nonparametric DEA, further analysis was conducted which included, super-efficiency DEA models and bootstrap resampling of the hospital-specific technical efficiency scores.

5.10 Super-efficiency

Two interrelated problems have been widely recognized with standard DEA models; namely, having weak discriminating power and unrealistic weight distribution (Li et al., 1999). The problem of weak discriminating power occurs when the number of DMUs under evaluation is not large enough compared to the total number of inputs and outputs. In this situation, classical DEA models (e.g., CCR model by Charnes et al., 1978; and BCC model by Banker et al., 1984) often yield solutions that identify too many DMUs as efficient. In order to improve DEA's discriminating power we adopt a super-efficiency DEA model.

Super efficiency indicates the extent to which the efficient DMUs exceed the efficient frontier formed by other DMUs. Zhu (2003) notes that when a DMU under evaluation is excluded from the reference set of the envelopment models, the resulting DEA models are termed super-efficiency DEA models. Charnes et al. (1992) use a super-efficiency model to study the sensitivity of the efficiency classifications. Zhu (1996b) and Seiford and Zhu (1998d) develop a variety of new super-efficiency models to determine the efficiency stability regions (see chapter 11 of Zhu 2003). The CRS super-efficiency model has been proposed by Andersen and Petersen (1993) in ranking the efficient DMUs. Additionally, the super-efficiency DEA models can be used in detecting influential observations (Wilson 1995) as well as in identifying the extreme efficient DMUs (Thrall 1996). The discriminatory power of the SE model provides insights that cannot be gained with standard DEA models.

However, there are three problematic areas with this methodology (Adler et al., 2002). First, Andersen and Petersen (1993) refer to the DEA objective function value as a rank score for all units, despite the fact each unit is evaluated according to different weights. Second, the super-efficient methodology can give "specialized" DMUs an excessively high ranking. To avert this problem, Sueyoshi (1999) introduces specific bounds on the weights in a super-efficient ranking model. Furthermore, in order to limit the super-efficient scores to a scale with a maximum of two, Sueyoshi develops an Adjusted Index Number formulation. The final problem lies with an infeasibility issue which if it occurs, means that the super-efficient technique cannot provide a complete ranking of all DMUs.

Infeasibility is addressed in the literature in a variety of ways. Lovell and Rouse (2003) propose a modification of the standard DEA model that overcomes the infeasibility problem often encountered in computing super-efficiency. In the Lovell-Rouse procedure one appropriately scales up the observed input vector (scale down the output vector) of the relevant super-efficient firm thereby usually creating its inefficient surrogate.

An alternative procedure proposed by Ray (2004) uses the directional distance function introduced by Chambers et al. (1996) to generate the Nerlove-Luenberger measure of super-efficiency. The fact that the directional distance function combines features of both an input-oriented and an output-oriented model, generally leads to a more complete ranking of the observations than either of the oriented models. An added advantage of this approach is that the Nerlove-Luenberger super-efficiency measure is unique and does not depend on any arbitrary choice of a scaling parameter.

Zhu (2003) presents the basic super-efficiency DEA models based upon the envelopment DEA models. The difference between the super-efficiency and envelopment models is that DMU_k under evaluation is not included in the reference set in the super-efficiency models. That is, the super-efficiency DEA models are based upon a reference technology constructed from the remaining DMUs.

Having specified input-based DEA models, we likewise specify input-oriented super-efficiency DEA models with variable returns to scale. Following Zhu (2003), the input-oriented super-efficiency DEA model is specified as follows:

$$
\begin{aligned}
& Min\theta^{\text{super}} \\
& subject \quad to \\
& \sum_{\substack{j=1 \\ j \neq o}}^{n} \lambda_j x_{ij} \leq \theta^{\text{super}} x_{io} \qquad i = 1,2,...,m; \\
& \sum_{\substack{j=1 \\ j \neq o}}^{n} \lambda_j y_{rj} \geq y_{ro} \qquad\qquad r = 1,2,...,s; \\
& \lambda_j \geq 0 \qquad\qquad\qquad j \neq o \\
& \sum\nolimits_{j \neq 0} \lambda_j = 1
\end{aligned}
\tag{5.10}
$$

Although super-efficiency models can differentiate the performance of the efficient DMUs, the efficient DMUs are not compared to the same “standard’. This is due to the fact that the frontier constructed from the remaining DMUs changes for every efficient DMU under evaluation which makes the ranking problematic. Chen (2002) notes that super-efficiency should be regarded as the potential input savings or output surpluses.

Charnes et al. (1991) observe that DMUs can be portioned into four classes $E,\quad E',\quad F\quad and\quad N$ described as in Table 5.6.

Table 5.6 Portioning of DMUs by Means of Super-efficiency Scores

Class	Attributes
E	Set of extreme efficient DMUs
E'	Set of efficient DMUs that are not extreme points. DMUs in this set can be expressed as linear combinations of DMUs in set E
F	Set of frontier points (DMUs) with non-zero slack(s). These are termed weakly efficient
N	Set of inefficient DMUs

Source: Adapted from Zhu (2003).

5.11 Boostrapping

By far the most serious impediment to a wider acceptance of DEA as a valid analytical method in economics lies in the fact that it is seen as non-statistical thereby not distinguishing random shocks from inefficiency. Although a satisfactory resolution of the problem is not at hand, efforts to add a stochastic dimension to DEA have been along various lines. The major directions of research in this area include but are not limited to: Banker’s *F* tests; chance-constrained programming; Varian’s statistical test of cost minimization; and bootstrap resampling (bootstrapping) for

DEA. Bootstrap resampling stands out as the most promising and is increasingly becoming popular. The bootstrap is a resampling method for statistical inference. It is commonly used to estimate confidence intervals, but it can also be used to estimate bias and variance of an estimator or calibrate hypothesis tests.

In 1979, bootstrapping was introduced as a computationally intensive statistical technique that permits the researcher to make inferences from data without making strong distributional assumptions (Money and Duval (1993); Efron and Tibshirani (1998)). Two distributions are considered. First, the underlying distribution of the data themselves, which is frequently described as a probability function (for instance, normal, binomial or Poisson) that shows all the values that the variables can take on and the probability or likelihood, that each will occur. Second, the distribution of the statistic (for example, the mean and median, amongst others) computed from the data.

The data as well as the computed statistic will vary in ways that can mathematically be described under the assumption that new sets of data were obtained or "sampled" and, for each respective dataset, a new statistic is computed. The statistic's sampling distribution is the probability of all possible values of the estimated statistic computed from a sample of size n drawn from a given population (Levy and Lemeshow, 1999). Bootstrapping employs resampling with replacement (or Monte Carlo resampling), to estimate the statistic's sampling distribution. The statistic's sampling distribution, if it can be determined, can then be used to estimate the standard errors and confidence intervals for the statistic under consideration.

The bootstrap method is an established statistical resampling method used to perform inference in complex problems. Lothgren and Tombour (1999), note that the crucial step in any application of the bootstrap is a clear specification of the Data Generating Process (DGP) underlying the observed data. The basic idea of the bootstrap method is to approximate the sampling distributions of the estimator by using the empirical distribution of resampled estimates obtained from a Monte Carlo resampling simulation of the estimation procedure whereby repeated resamples obtained from an estimate of the DGP produce repeated estimates. The performance of the bootstrap in terms of the validity of the conducted statistical inference ultimately depends on how well the DGP characterizes the true data generation and how well the DGP is mimicked in the resampling simulation.

The estimation of the confidence intervals using the bootstrap proceeds as follows. First, one uses resampling with replacement to construct m resampled datasets (or bootstrap samples), that comprise of the same number of data points (n) just like the original dataset. Conducting resampling with replacement, a data point or an observation is randomly selected from the original dataset and copied into the resampled dataset being created. While a given observation has been "used," it is not deleted from the original dataset or, precisely, it is "replaced." A different observation is then randomly selected, and the process is repeated until a resampled dataset of size n is created. Consequently, the same data point may be included in the resampled dataset once, twice, or more times or not at all. Second, the descriptive statistic of choice is calculated for each resampled dataset.

Finally, a confidence interval for the statistic is computed from the set of values obtained for the statistic. At this point in the analysis, there exist various options for calculating the confidence intervals. These include the normal approximation method, the percentile method, the Bias-Corrected (BC) method, the

Bias-Corrected and Accelerated (BCA) method, and the Approximate Bootstrap Confidence (ABC) method (Efron and Tibshirani, 1998). This study used the bias corrected and accelerated method.

The normal approximation method calculates an approximate standard error by means of the sampling distribution that results from all the bootstrap resamples. The confidence interval is then computed by means of the z-distribution (original statistic $\pm$ $1.96 \times$ standard error; for a 95 percent confidence interval). The percentile method uses the frequency histogram of the m statistics computed from the bootstrap samples. The 2.5 and 97.5 percentiles constitute the limits of the 95 percent confidence interval. The BCA method adjusts for bias in the bootstrapped sampling distribution, and is therefore deemed a substantial improvement over the percentile method (Efron and Tibshirani, 1998). The BCA confidence interval is an adjustment of the percentiles used in the percentiles method based upon the computation of two coefficients called "acceleration" and "bias correction." The acceleration coefficient adjusts for nonconstant variances within the resampled datasets. The ABC method is an approximation of the BCA method that requires fewer resampled datasets than the BCA method. Conversely, the bias correction coefficient adjusts for the skewness in the bootstrap sampling distribution. If the bootstrap sampling distribution is perfectly symmetric, then the bias correction will be zero (Efron and Tibshirani, 1998).

Mooney and Duval (1993) note that as a general guideline, at least $1{,}000$ resampled datasets should be used when computing a BCA confidence interval. When using the percentile method, Efron and Tibshirani (1986) suggest that a value of at least 250 can be used when estimating a confidence interval as a consequence of not having to compute bias correction. The variability of the confidence interval is inversely related to the number of resampled datasets, which implies that as the number of resampled datasets decreases, more variability is introduced into the confidence interval estimation (Efron and Tibshirani, 1986 & 1998).

The bootstrap can be implemented in a variety of ways and some of the previous studies have considered different questions. For instance, firm-specific efficiency scores (or distance functions) are considered in Gstach (1995), Ferrier and Hirschberg (1997) and Simar and Wilson (1998), whilst Atkinson and Wilson (1995) bootstrap sample-averages of efficiency and Malmquist indices. Following the methodology by Simar and Wilson (1998), bootstrap resampling was conducted on the hospital-specific technical efficiency scores in order to provide confidence intervals.

Nonparametric bootstrap resampling was conducted by means of FEAR (Frontier Efficiency Analysis with R) developed by Wilson (2005). Because the pooled dataset (1999-2003) yields approximately the same results as the 1999 dataset, data for 1999 were used in the bootstrapping and because the software package at our disposal could not handle panel data.

5.12 Conclusion

The methodology used by this study has been highlighted in the foregoing chapter. The unit of analysis, as well as the input and output variables employed in various data envelopment models have been examined. In addition, the following have been discussed: (a) dealing with undesirable outputs; (b) adjusting for case-mix; (c) using super-efficiency to distinguish between efficient DMUs, and (d) bootstrap resampling to obtain confidence intervals for the efficiency estimates.

CHAPTER SIX

TECHNICAL EFFICIENCY EMPIRICAL FINDINGS

6.1 Introduction

This chapter presents the empirical findings on technical efficiency of the sample hospitals. The next section commences with DEA technical efficiency results without undesirable output (patient deaths). The sampling distribution of the DEA technical efficiency scores generated by means of Bootstrap resampling, super-efficiency as well as DEA technical efficiency results with undesirable outputs follow. The chapter concludes with a discussion of the empirical results presented.

6.2 DEA Technical Efficiency Results without Undesirable Outputs

We present the empirical results obtained from applying the DEA technique to a number of model specifications. These specifications, based on the variable definitions in Chapter five, are presented in Table 6.1, which shows the input and output combinations used in six model specifications. Models 2 to 6 are based on slight modifications of model 1.

Table 6.1 DEA Model Specifications

Variables/Model	1	2	3	4	5	6
Inputs						
Beds	X	X	X	X	X	X
All staff grouped together				X	X	
Doctors	X	X	X			X
Nurses	X	X	X			X
Other staff	X	X	X			X
Outputs						
Admissions (un-weighted)	X					X
Admissions (case-mix adjusted)		X	X	X	X	
Outpatient Attendances	X	X	X	X	X	
Surgical Operations	X	X			X	X
Deliveries	X	X			X	X
Patient Deaths						X

Each of models 2 to 6 contains a minor definitional change (such as the inclusion or exclusion of a variable from a model) to the specification contained in model 1. For instance, model 2 uses the same inputs and outputs as model 1 with the exception of a different definition of the admissions (which have been adjusted by means of the case-mix index generated in Chapter five). Model 2 was chosen as the preferred model because it was decided that case-mix adjusted admissions was conceptually a better measure of output than admissions (un-weighted). The model also gives a sensible spread of efficiency scores for the whole sample and contains a plausible number of variables or factors (input plus outputs) when compared with the size of the overall sample.

The DEA method provides relative technical efficiency scores for the sample in question. One vital important consideration in this analysis is the sensitivity of efficiency estimates to the change in model specifications. Due to the non-parametric nature of

DEA, it is not possible to test model specifications or goodness-of-fit in the usual parametric manner associated with regression analysis. Because of this, the study employed a number of models to analyze the sensitivity of DEA results to changes in the choice of input or output variables.

The principal technical efficiency results reported were derived by imposing the assumption of variable returns to scale on each of the models outlined and input-oriented DEA. The results reported below do not take into account the undesirable output of patient deaths. We introduce this later in section 6.3.

Technical efficiency scores only refer to the relative performance within the sample. Hospitals with an efficiency score of one are efficient relative to all other hospitals in the sample, but may not be efficient by some absolute or standard necessarily. This is important – inefficiency is inherently unobservable – all we can do is benchmark DMUs against each other, not against some absolute standard.

The five models result in different measures of technical efficiency and Table 6.2 reports the efficiency scores from the five DEA models using annual data as well as the pooled dataset (1999-2003). The mean efficiency scores differ depending upon the model specification.

Table 6.2 Efficiency Scores from Five DEA Models, 1999-2003

	1999	2000	2001	2002	2003	Pooled 1999-2003
Model 1						
Mean	0.972	0.943	0.975	0.982	0.968	0.972
Standard Deviation	0.058	0.107	0.063	0.060	0.070	0.058
Minimum	0.786	0.606	0.757	0.728	0.698	0.786
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
Number on Frontier	19	17	20	22	17	19
Model 2						
Mean	0.973	0.946	0.975	0.983	0.971	0.973
Standard Deviation	0.054	0.104	0.061	0.060	0.069	0.055
Minimum	0.804	0.630	0.770	0.730	0.698	0.804
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
Number on Frontier	19	18	20	22	18	19
Model 3						
Mean	0.921	0.922	0.944	0.923	0.917	0.921
Standard Deviation	0.103	0.105	0.100	0.118	0.131	0.103
Minimum	0.594	0.642	0.602	0.602	0.591	0.594
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
Number on Frontier	9	13	16	14	12	9
Model 4						
Mean	0.902	0.905	0.938	0.916	0.888	0.902
Standard Deviation	0.101	0.112	0.109	0.126	0.127	0.101
Minimum	0.594	0.660	0.602	0.602	0.591	0.594
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
Number on Frontier	5	12	16	13	6	5
Model 5						
Mean	0.961	0.932	0.957	0.961	0.935	0.960
Standard Deviation	0.077	0.120	0.104	0.091	0.102	0.077
Minimum	0.743	0.580	0.630	0.693	0.657	0.743
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
Number on Frontier	15	15	20	20	11	15

To check for the robustness of the models to changes in the measurement of admissions, models 1 and 2 were run. Comparing models 1 and 2 indicates that in general the efficiency scores of hospitals rise when the admissions are adjusted by means of the case-mix index. For instance, the mean efficiency score marginally rises from 0.972 (97.2 percent) for Model 1 to 0.974 (97.4 percent) for Model 2 in 1999. This, therefore, implies that not adjusting admissions to the structure of the patient load, understates the efficiency scores of hospitals. Although the difference is small, model 2 is preferred to model 1 because this is better than no case-mix adjustment at all despite its being rather crude. Thus, in 1999, relative to the frontier of the sample hospitals, Uganda's district referral hospitals realized approximately 97 percent of their potential output. The same potential output is produced (97 percent) even when efficiency is estimated from the pooled dataset. On average 19 out of the 25 hospitals operated on the production frontier over the sample period when Models 1 and 2 are estimated.

Comparing models 3 and 4 generally shows that lumping human resources into one variable reduces the efficiency scores by an average 1.6 percent and reduces the number of hospitals on the frontier from 19 to 9 (for Model 3) and to 5 (for Model 4). When models 4 and 5 are compared, it is revealed that incorporation of more output variables increases the efficiency scores by an average of 4 percent and increases the number of hospitals on the production frontier by 6 hospitals (for Model 3) and 9 hospitals (for Model 4). These results are driven by the choice of variables in the modeling process. Also in line with expectations (Smith, 1997), the models with larger numbers of inputs and outputs yield higher average efficiencies.

As expected, the inclusion of additional variables or the disaggregation of existing variables (while holding the number of observations constant), has the effect of increasing efficiency scores for observations which were not previously efficient. This effect is seen by the difference in average efficiency scores between models 1 and 2 and models 3 or 4. Models 1 and 2 have the most factors, thus most hospitals end up on the frontier (Nunamaker, 1985).[22] The only shortcoming of these two models is that they are less discriminating. It is noteworthy that models 1 and 2 perform as well as the corresponding pooled dataset both in terms of efficiency scores and hospitals on the frontier. The similarity between the results for models 1 and 2 (n=25) vis-à-vis those for the pooled dataset (n=125), shows that DEA models perform better with large samples (Pedraja-Chaparro et al., 1999).

The technical and scale efficiency scores for individual hospitals, estimated with the preferred model (model 2)[23] using the pooled dataset (1999-2003) are presented in Table 6.3. On average, the sample hospitals had a technical efficiency (TE) score of 97.3 percent, while the scale efficiency (SE) score stood at 97.5 percent. Of the 25[24] hospitals, 19 (76 percent) were technically efficient since they had a relative technical efficiency score of 100 percent. The remaining 6 (24 percent) had a technical efficiency score of less than 100 percent, implying that they were technically inefficient. The TE score among the technically inefficient hospitals ranged from 80.4 percent in Gombe hospital to 97.3 percent in Masindi hospital. This empirical finding implies that Gombe and Masindi hospitals could potentially reduce their factor inputs by 19.6 percent and 2.7 percent, respectively while leaving their output levels fixed.

Sixteen (64 percent) of the hospitals had a scale efficiency score of 100 percent implying that they had the most productive size for that particular input-output mix. The

[22] Nunamaker (1985) shows that no firm can become 'less' efficient by the addition of a variable, so that firms which were previously fully efficient will remain fully efficient with the addition of extra variables.
[23] Model 2 is the preferred model because the admissions have been adjusted for the patient characteristics by means of the case-mix index generated via the average length of stay.
[24] The pooled dataset has 125 hospitals (25 hospitals over a five-year period). However, DEAP requires the number of DMUs as well as time periods to be specified and reports results on the specified number of DMUs.

remaining nine (36 percent) hospitals have a scale efficiency score of less than 100 percent and as such they were not scale efficient.

Constant returns to scale were exhibited by all the sixteen scale efficient hospitals. This implies that they were operating at their most productive scale sizes. Eight of the nine scale inefficient hospitals displayed increasing returns to scale (IRS) whereas one of the scale inefficient hospitals had decreasing returns to scale (DRS). To operate at the most productive scale size, a hospital with DRS should scale down its inputs as well as its outputs. In the same vein, a hospital exhibiting IRS should expand both its outputs and inputs.

Table 6.3 Technical and Scale Efficiency Scores: Model 2, Pooled Dataset (1999-2003

Hospital	*CRS Technical Efficiency*	*VRS Technical Efficiency*	*Scale Efficiency*
Bududa	1.000	1.000	1.000
Bugiri	1.000	1.000	1.000
Busolwe	1.000	1.000	1.000
Bwera	0.953	1.000	0.953
Entebbe	1.000	1.000	1.000
Gombe	0.791	0.864	0.916
Iganga	1.000	1.000	1.000
Itojo	0.805	0.859	0.938
Kagadi	0.800	0.947	0.845
Kalisizo	0.963	1.000	0.963
Kambuga	1.000	1.000	1.000
Kapchorwa	0.975	1.000	0.975
Kawolo	0.904	1.000	0.904
Kayunga	0.826	0.940	0.879
Kiboga	0.804	0.804	1.000
Kiryandongo	1.000	1.000	1.000
Kisoro	1.000	1.000	1.000
Kitagata	1.000	1.000	1.000
Masindi	0.911	0.912	0.998
Mityana	1.000	1.000	1.000
Mubende	1.000	1.000	1.000
Nakaseke	1.000	1.000	1.000
Pallisa	1.000	1.000	1.000
Rakai	1.000	1.000	1.000
Tororo	1.000	1.000	1.000
Mean	**0.949**	**0.973**	**0.975**
=1	15	19	16
<1	10	6	9

Note: CRS Technical Efficiency = technical efficiency from CRS DEA
VRS Technical Efficiency = technical efficiency from VRS DEA

$$\textit{Scale Efficiency} = \frac{\textit{CRS Technical Efficiency}}{\textit{VRS Technical Efficiency}}$$

Table 6.4 shows the Pearson correlation matrix of efficiency scores across the five DEA models for individual years and for the pooled dataset. This was done to check model stability over time. The year 1999 has the same number of significant Pearson correlation coefficients as 2000. Likewise, 2002 and 2003 have same number of significant coefficients. This, therefore, implies that the five models are reasonably stable for the years 1999, 2000, 2002 and 2003 but not for 2001 and for the pooled data (1999-2003). Caution has to be exercised in the interpretation of the results across models (i.e., consistency across model specifications) because some models are nested in other models.

Table 6.4 Pearson Correlation Matrix of Efficiency Scores Across Five DEA Models

1999	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9787*	1.0000			
Model 3	0.1639	0.2416	1.0000		
Model 4	0.2102	0.2524	0.8962*	1.0000	
Model 5	0.9261*	0.9473*	0.2451	0.2941	1.0000
2000	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9965*	1.0000			
Model 3	0.0156	0.0334	1.0000		
Model 4	0.0612	0.0808	0.9348*	1.0000	
Model 5	0.9725*	0.9658*	0.0383	0.1401	1.0000
2001	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9996*	1.0000			
Model 3	0.2978	0.3004	1.0000		
Model 4	0.3727	0.3745	0.9910*	1.0000	
Model 5	0.8390*	0.8364*	0.4629*	0.5660*	1.0000
2002	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9999*	1.0000			
Model 3	0.5087*	0.5063*	1.0000		
Model 4	0.5528*	0.5501*	0.9940*	1.0000	
Model 5	0.7201*	0.7165*	0.5535*	0.6208*	1.0000
2003	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9670*	1.0000			
Model 3	0.5042*	0.5445*	1.0000		
Model 4	0.4773*	0.4761*	0.9065*	1.0000	
Model 5	0.7333*	0.7748*	0.5727*	0.6256*	1.0000
Pooled 1999-2003	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9787*	1.0000			
Model 3	0.1639	0.2416*	1.0000		
Model 4	0.2102*	0.2524*	0.8962*	1.0000	
Model 5	0.926104*	0.9473*	0.2451*	0.2941*	1.0000

Note: *Significant at 5 percent level

Table 6.5 captures the change in efficiency scores and ranks for individual hospitals across the five DEA models using the pooled dataset. The changes in the rankings (and efficiency scores) are analyzed by looking at the top five, middle five and bottom five hospitals across the five models. It should be noted that the ranking of all hospitals is consistent across models 1 and 2. Models 4 and 5 are utilized to check the change in the efficiency scores and rankings of hospitals. The movement from Model 4 to Model 5 makes hospitals gain three places for the top five and five places for the middle five, on average. Conversely, the bottom five hospitals lose on average, three places as we move from Model 4 to Model 5. Some stability is exhibited although one cannot really distinguish between hospitals whose technical efficiency score equals unity.

To investigate whether there are any regional differences in hospital technical efficiency, hospitals were grouped into three categories by region, namely, Eastern, Western and Central. The null hypothesis tested was that there is no difference in the hospital scores across the three regions. The pooled dataset was used to estimate model 2 for the three regions. However, the results lacked discriminatory power to show any clear regional differences in hospital technical efficiency.

Table 6.5 Change in Efficiency Scores and Ranks for Individual Hospitals across Five DEA Models using the Pooled Dataset (1999-2003)

Model 1			**Model 2**			**Model 3**			**Model 4**			**Model 5**		
Hospital	**Score**	**Rank**	**Hospital**	**Score**	**Rank**	**Hospital**	**Score**	**Rank**	**Hospital**	**Score**	**Rank**	**Hospital**	**Score**	**Rank**
Bududa	1.000	1	Bududa	1.000	1	Bududa	1.000	1	Bududa	0.957	9	Bududa	1.000	1
Bugiri	1.000	2	Bugiri	1.000	2	Bugiri	1.000	2	Bugiri	1.000	1	Bugiri	1.000	2
Busolwe	1.000	3	Busolwe	1.000	3	Busolwe	0.800	3	Busolwe	0.800	23	Busolwe	1.000	3
Bwera	1.000	4	Bwera	1.000	4	Bwera	1.000	23	Bwera	0.960	8	Bwera	0.984	18
Entebbe	1.000	5	Entebbe	1.000	5	Entebbe	0.827	21	Entebbe	0.827	20	Entebbe	1.000	4
Iganga	1.000	6	Iganga	1.000	6	Iganga	0.594	25	Iganga	0.594	25	Iganga	1.000	5
Kalisizo	1.000	7	Kalisizo	1.000	7	Kalisizo	0.952	15	Kalisizo	0.952	12	Kalisizo	0.995	16
Kambuga	1.000	8	Kambuga	1.000	8	Kambuga	0.974	12	Kambuga	0.951	13	Kambuga	1.000	6
Kapchorwa	1.000	9	Kapchorwa	1.000	9	Kapchorwa	0.953	13	Kapchorwa	0.953	10	Kapchorwa	0.993	17
Kawolo	1.000	10	Kawolo	1.000	10	Kawolo	0.953	14	Kawolo	0.953	11	Kawolo	0.973	19
Kiryandongo	1.000	11	Kiryandongo	1.000	11	Kiryandongo	1.000	5	Kiryandongo	1.000	3	Kiryandongo	1.000	7
Kisoro	1.000	12	Kisoro	1.000	12	Kisoro	0.909	18	Kisoro	0.879	17	Kisoro	1.000	8
Kitagata	1.000	13	Kitagata	1.000	13	Kitagata	1.000	6	Kitagata	0.802	21	Kitagata	1.000	9
Mityana	1.000	14	Mityana	1.000	14	Mityana	0.975	11	Mityana	0.975	14	Mityana	1.000	10
Mubende	1.000	15	Mubende	1.000	15	Mubende	0.981	10	Mubende	0.981	6	Mubende	1.000	11
Nakaseke	1.000	16	Nakaseke	1.000	16	Nakaseke	0.929	16	Nakaseke	0.924	14	Nakaseke	1.000	12
Pallisa	1.000	17	Pallisa	1.000	17	Pallisa	0.801	22	Pallisa	0.801	22	Pallisa	1.000	13
Rakai	1.000	18	Rakai	1.000	18	Rakai	1.000	8	Rakai	1.000	4	Rakai	1.000	14
Tororo	1.000	19	Tororo	1.000	19	Tororo	1.000	9	Tororo	1.000	5	Tororo	1.000	15
Kagadi	0.943	20	Kagadi	0.947	20	Kagadi	1.000	4	Kagadi	1.000	2	Kagadi	0.947	20
Kayunga	0.940	21	Kayunga	0.940	21	Kayunga	0.915	17	Kayunga	0.915	15	Kayunga	0.937	21
Gombe	0.897	22	Gombe	0.864	23	Gombe	0.735	24	Gombe	0.735	24	Gombe	0.743	25
Masindi	0.866	23	Masindi	0.912	22	Masindi	1.000	7	Masindi	0.870	18	Masindi	0.814	23
Itojo	0.859	24	Itojo	0.859	24	Itojo	0.846	20	Itojo	0.830	19	Itojo	0.855	22
Kiboga	0.786	25	Kiboga	0.804	25	Kiboga	0.881	19	Kiboga	0.881	16	Kiboga	0.771	24

6.3 Sampling Distribution of the DEA Technical Efficiency Scores

The original DEA methodology assumes that the input-output vectors give a good representation of the complete production technology. For some applications, large, statistically representative data sets are available (Bartelsman and Doms, 2000). However, many applications including the present study involve small samples. Small samples generally do not give a full representation of the technology. Hence, inefficient DMUs can be wrongly classified as efficient, or 'true inefficiency' can be substantially underestimated (Cherchye and Post, 2003).

A recent theoretical development of DEA is to take explicitly into account the statistical properties of efficiency scores as estimators of unknown true scores by applying the technique of bootstrapping (Simar and Wilson, 2000). This provides bias correction of the efficiency scores and confidence intervals, thus signaling the quality of the estimates, and especially avoiding drawing wrong conclusions as to which DMUs should be used as role models for improvement when the density of observation is disregarded.

The confidence intervals via the homogeneous bootstrap method (Simar and Wilson (1998) for the DEA efficiency scores for Model 2, estimated by means of 1999 data using Wilson's (2005) FEAR version 0.913, software, are reported in Table 6.7. The number of bootstrap iterations is 2000.

As indicated in Table 6.6, on average, the bootstrapped DEA efficiency scores have a lower bound of 0.8327 and an upper bound of 0.8771. The mean VRS efficiency score stood at 0.973, while the mean bias-corrected efficiency score was 0.8404, which yielded a mean estimated bias of 0.1326.

Table 6.6 Original and Bias-Corrected DEA Efficiency Scores: Model 2, 1999 Data

Hospital	*Eff. Scores (VRS)*	*Eff. Bias-Corrected*	*Est. Bias*	*Lower Bound*	*Upper-Bound*
Bududa	1.0000	0.8237	0.1763	0.8063	0.9070
Bugiri	1.0000	0.8230	0.1770	0.8063	0.9089
Busolwe	1.0000	0.8234	0.1766	0.8064	0.9088
Bwera	1.0000	0.8053	0.1947	0.8064	0.8029
Entebbe	1.0000	0.8226	0.1774	0.8063	0.9091
Gombe	0.8640	0.9751	-0.1111	0.9750	0.9762
Iganga	1.0000	0.8230	0.1770	0.8063	0.9113
Itojo	0.8590	0.9362	-0.0772	0.9387	0.9274
Kagadi	0.9470	0.8084	0.1386	0.8061	0.8133
Kalisizo	1.0000	0.8061	0.1939	0.8096	0.7969
Kambuga	1.0000	0.8127	0.1873	0.8063	0.8202
Kapchorwa	1.0000	0.8076	0.1924	0.8081	0.8059
Kawolo	1.0000	0.8141	0.1859	0.8063	0.8274
Kayunga	0.9400	0.8468	0.0932	0.8487	0.8371
Kiboga	0.8040	1.0000	-0.1960	1.0000	1.0000
Kiryandongo	1.0000	0.8226	0.1774	0.8063	0.9112
Kisoro	1.0000	0.8220	0.1780	0.8063	0.9023
Kitagata	1.0000	0.8574	0.1426	0.8586	0.8530
Masindi	0.9120	0.8399	0.0721	0.8415	0.8359
Mityana	1.0000	0.8242	0.1758	0.8063	0.9096
Mubende	1.0000	0.8230	0.1770	0.8062	0.9070
Nakaseke	1.0000	0.8373	0.1627	0.8364	0.8443
Pallisa	1.0000	0.8223	0.1777	0.8063	0.9054
Rakai	1.0000	0.8094	0.1906	0.8063	0.8105
Tororo	1.0000	0.8230	0.1770	0.8063	0.8959
Mean	*0.9730*	*0.8404*	0.1326	*0.8327*	*0.8771*
Std. Dev.	*0.0551*	*0.0514*	*0.1040*	*0.0551*	*0.0554*

Out of the 25 hospitals in the original sample 19 hospitals were found to operate on the best practice frontier (VRS efficiency score = 1). This result does not provide much information to decision makers as it is not possible to distinguish between the performances of the majority of the hospitals. In these circumstances, the bootstrap procedure turns out to be a very useful tool.

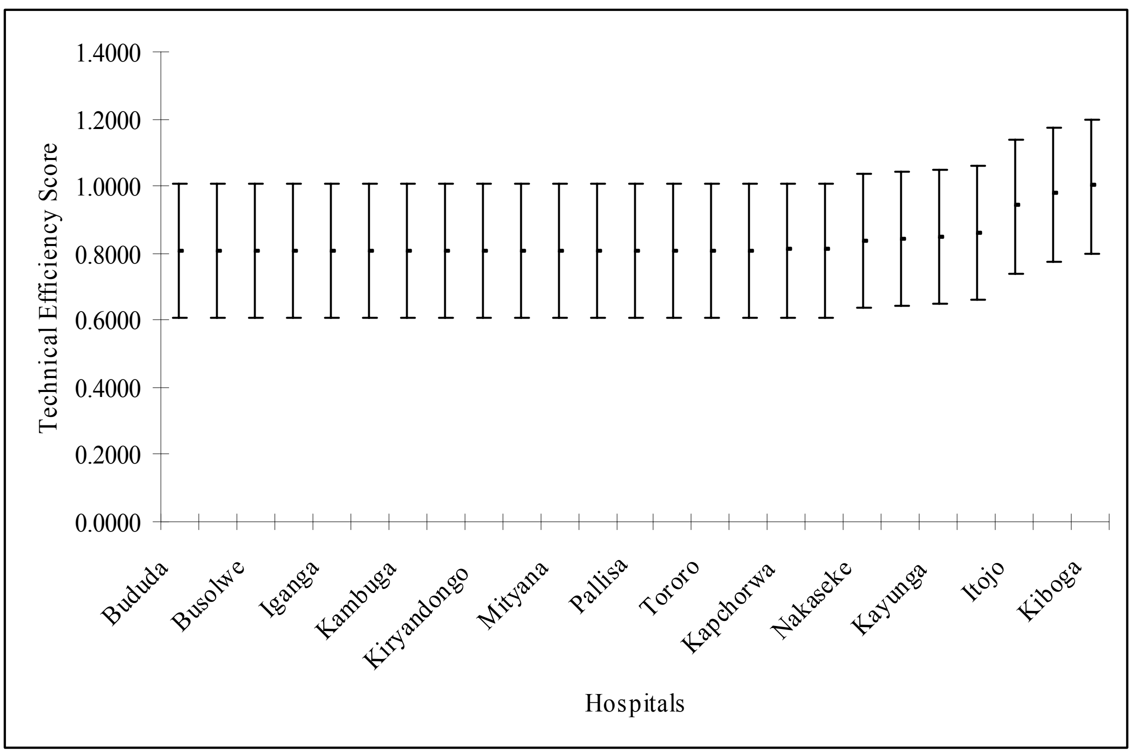
Columns 2–6 in Table 6.6 provide the original DEA efficiency score, the bias-corrected estimate, the bootstrap bias estimate (difference between the original efficiency score and the bias-corrected efficiency score), and 95 percent confidence intervals (lower bound and upper bound), respectively, for the bias-corrected efficiency scores.

These results reveal the sensitivity of the efficiency measures with respect to sampling variation. The bias-corrected estimates in column 3 reveal that differences in measured efficiency are of a different magnitude than when original efficiency scores are considered. Specifically, we observe that technical efficiency scores for all hospitals fall except for Gombe, Itojo and Kiboga. It should be noted that Kiboga hospital is the only hospital that is efficient by means of the bias-corrected efficiency scores.

However, to more properly interpret the bias-corrected estimators we must use information on confidence intervals, which define the statistical location of the true efficiency. Broadly, they show that efficiency scores, when adding statistical precision, overlap to a notable extent (Figure 6.1). It should be noted that the wider the interval, the higher the possibility of overlapping and vice-versa. Consequently, for 90 percent confidence intervals, overlapping would occur in fewer instances, whereas for 99 percent confidence intervals it would occur more often. Nevertheless, these findings cannot be compared due to lack of previous hospital efficiency studies which have implemented bootstrap resampling.

If it is assumed that technical inefficiency is invariant in time it is possible to construct confidence intervals for individual hospital scores. The confidence intervals around bias-corrected estimates of the efficiency scores for all observations in 1999 are shown in Figure 6.1. Hospitals are ranked in ascending order of the bias-corrected technical efficiency scores. The results show that we cannot really discern any differences between the hospitals which makes DEA in small samples unable to discriminate. Thus, we cannot use this exercise to rank hospitals. Nevertheless, it is very valuable as an improvement over straight DEA and the graph clearly indicates why.

Figure 6.1 Confidence Intervals for Individual Hospital Technical Efficiency Scores



6.4 Super-Efficiency

Data Envelopment Analysis provides an objective basis for ranking firms in an industry in order of their measured technical efficiency scores. This, however, is not possible for the sub-group of firms that lie on the graph of the technology and are all rated at 100 percent technical efficiency. A procedure first proposed by Andersen and Petersen (1993) uses the *super-efficiency* measures of these efficient firms to resolve this problem. A firm is regarded as super-efficient if its DEA efficiency score exceeds 100 percent when measured against a production possibility set constructed from the input-output data of *all other firms* in the sample.

Having specified input-based DEA models, we likewise specified input-oriented VRS super-efficiency DEA models with variable returns to scale and the results of model 2 using 1999 data are reported in Table 6.7. A comparison of the results for Model 2 using 1999 data with those of Table 6.7 reveals that the mean technical efficiency score rises from 0.974 to 1.0917 while the standard deviation also rises from 0.0784 to 0.6310. The minimum and maximum super-efficiency scores, stand at 0.8061 and 2.9211, respectively, whilst 12 hospitals attain super-efficiency scores of at least unity and 10 have super-efficiency scores below unity. Super-efficiency was infeasible for three hospitals, namely, Bugiri, Entebbe and Iganga.

Zhu (2003) explains the infeasibility of input-oriented VRS super-efficiency model as follows: for a specific extreme efficient $DMU_o = (x_o, y_o)$, the input-oriented VRS super-efficiency model is infeasible *if and only if* $(\chi x_o, y_o)$ is efficient under the VRS envelopment model for any $1 \leq \chi < +\infty$.

The results underscore the discriminatory power of the super-efficiency model which standard DEA models lack. For instance, the number of hospitals on the frontier falls from 19 with standard DEA models to 12 when the super-efficiency model is estimated. The ranking of hospitals although incomplete, is generally unchanged across the original VRS technical efficiency scores and the input-oriented VRS super-efficiency scores which points to some reasonable degree of stability.

Table 6.7 Original VRS Technical Efficiency Scores versus Input-Oriented VRS Super-efficiency Scores: Model 2, 1999 Data

Hospital	*Original VRS Technical Efficiency Score*	*Input-Oriented VRS Super Efficiency Score*
Bududa	1.0000	2.1441
Bugiri	1.0000	infeasible
Busolwe	1.0000	1.7690
Bwera	0.9530	0.9996
Entebbe	1.0000	infeasible
Gombe	0.7910	0.8271
Iganga	1.0000	infeasible
Itojo	0.8050	0.8587
Kagadi	0.8000	1.0052
Kalisizo	0.9630	0.9959
Kambuga	1.0000	1.0332
Kapchorwa	0.9750	0.9975
Kawolo	0.9040	1.0462
Kayunga	0.8260	0.9499
Kiboga	0.8040	0.8061
Kiryandongo	1.0000	1.7778
Kisoro	1.0000	1.4807
Kitagata	1.0000	0.9390
Masindi	0.9110	0.9583
Mityana	1.0000	1.3789
Mubende	1.0000	1.2575
Nakaseke	1.0000	0.9639
Pallisa	1.0000	2.9211
Rakai	1.0000	1.0161
Tororo	1.0000	1.1662
Mean	0.9493	*1.0917*
Standard Deviation	0.0784	*0.6310*
Minimum	0.7910	*0.8061*
Maximum	1.0000	*2.9211*
=>1	15	*12*
<1	10	*10*
Infeasible	Not applicable	*3*

6.5 DEA Technical Efficiency Results with Undesirable Outputs

The production of desirable output is often accompanied by simultaneous or joint production of undesirable outputs. In a healthcare setting, the primary desirable output is the patient recuperating after receiving medical attention/care, whereas the undesirable output is the death of the patient. Obviously, there are patients who leave the healthcare institution worse off than they came. Therefore, it makes economic sense to evaluate the technical efficiency of hospitals by crediting live discharges (desirable output) while at the same time penalizing patient deaths (undesirable output). The study tries to evaluate the technical efficiency of hospitals by treating patient deaths as undesirable outputs while the desirable outputs comprise admissions, operations, deliveries and outpatient department attendances.

Table 6.8, presents the descriptive statistics of live and dead discharges, for individual years and for the 1999-2003 period. The proportion of dead discharges to total admissions is also indicated.

Table 6.8 Descriptive Statistics: Live Discharges and Patient Deaths

	Live Discharges					
Year	***1999***	***2000***	***2001***	***2002***	***2003***	***1999-2003***
Mean	6781	6783	7582	7908	8271	7465
Standard Deviation	4282	4693	5944	6307	6621	5585
Minimum	2956	3093	1273	3734	2622	1273
Maximum	25086	27824	32962	36226	37936	37936
	Dead Discharges					
Year	***1999***	***2000***	***2001***	***2002***	***2003***	***1999-2003***
Mean	268	280	269	277	271	273
Standard Deviation	123	107	110	128	120	116
Minimum	50	107	129	69	23	23
Maximum	707	652	679	645	610	707
Total Dead Discharges/Total Admissions	*3.80*	*3.96*	*3.43*	*3.39*	*3.17*	*3.53*

It is apparent that the mean of live discharges lies between 6,781 and 7,908 with a standard deviation in the range 4,282-6,621. The dead discharges on the other hand, have a mean which lies between 268 and 280, with a standard deviation in the range 107-128. The proportion of dead discharges in the total admissions is in the range of 3.17 – 3.80. The deaths vary by ward or specialty.

The output-oriented technical efficiency scores when the undesirable output (the bad output (patient deaths) are not translated are presented in Table 6.9 by means of model 1[25] with 1999 data. Patient deaths are incorporated into the technical efficiency analysis and model 6 is estimated by means of Zhu's (2003) DEA Excel Solver.

Comparing Table 6.2 with Table 6.9 reveals that when all hospitals are credited for producing desirable outputs and penalized for producing an undesirable output by way of patient deaths, their mean technical efficiency score falls from 0.972 to 0.838 while the standard deviation also falls from 0.058 to 0.052. The environmentally-

[25] This is the model in which the admissions have not been adjusted by means of the case-mix index.

adjusted technical efficiencies are generally lower than the original ones (except for Itojo and Kayunga, for which the score rose), because the score is reduced for the use of inputs in generating the undesirable output which is costly to dispose of.

Table 6.9 Technical Efficiency Scores; with and without Bad Output: Model 1, 1999 Data

Hospital	*Technical Efficiency Score without Patient Deaths*	*Technical Efficiency Score with Patient Deaths*
Bududa	1.0000	0.8117
Bugiri	1.0000	0.8117
Busolwe	1.0000	0.8117
Bwera	1.0000	0.8117
Entebbe	1.0000	0.8117
Gombe	0.8970	0.8117
Iganga	1.0000	0.8117
Itojo	0.8590	1.0000
Kagadi	0.9430	0.8548
Kalisizo	1.0000	0.8117
Kambuga	1.0000	0.8117
Kapchorwa	1.0000	0.8117
Kawolo	1.0000	0.8117
Kayunga	0.9400	0.9669
Kiboga	0.7860	0.9219
Kiryandongo	1.0000	0.8117
Kisoro	1.0000	0.8117
Kitagata	1.0000	0.8864
Masindi	0.8660	0.8615
Mityana	1.0000	0.8117
Mubende	1.0000	0.8117
Nakaseke	1.0000	0.8397
Pallisa	1.0000	0.8117
Rakai	1.0000	0.8117
Tororo	1.0000	0.8117
Mean	0.9716	0.8377
Standard deviation	0.0581	0.0523
Minimum	*0.7860*	0.8117
Maximum	*1.0000*	1.0000
=>1	*19*	1
<1	*6*	24

The minimum technical efficiency score rises from 0.786 to 0.812. In addition, all hospitals attain an efficiency score of at least 81 percent. Therefore, the incorporation of patient deaths in the analysis impacts on the hospitals' efficiency scores by generally lowering them. The results thus indicate that excluding undesirable outputs generally overstates the technical efficiency of hospitals. Thus, it is important to account for quality of care provided.

6.6 Discussion

The technical efficiency of some hospitals in the sample is less than 100 percent and this should be of some concern to Ministry of Health policymakers and planners interested in good value for money. Given the existing levels of both technical and scale inefficiency, the attainment of the national Health Policy objectives as well as health-related global and regional targets such as Abuja targets and the Millennium Development Goals (MDGs) will be compromised. Therefore, the efficient use of existing resources should be the center piece of the national health policy.

Nevertheless, the degree of inefficiency and policy response should be contingent upon the hospital's operating environment and appropriate action ought to be taken only after a thorough investigation. While DEA is a useful diagnostic tool, it might not be appropriate to base funding and resource decisions or efficiency targets on the basis of the resultant efficiency estimates.

Technical as well as scale inefficiency is present in varying degrees in a majority of hospitals in both developing and developed countries (see for instance, Wouters 1993, McMurchy 1996, Ersoy et al. 1997, Ferrier and Valdmanis 1996, Hao and Pegles 1994, Ozcan et al. 1996, Rosko and Chilingerian 1999). Nevertheless, in Sub-Saharan Africa, few hospital efficiency studies have been carried out using frontier models (see for example, Kirigia et al. 2000, Zere et al. 2000, Kirigia et al. 2001, Kirigia et al. 2002, Kirigia et al. 2004, Osei et al. 2005, and Renner et al. (2005)). However, most of the African healthcare efficiency studies have mainly looked at public hospitals. Thus, there is no clear and quantifiable evidence on the type and degree of inefficiency given that the not-for-profit as well as private healthcare providers play a major role in healthcare delivery in Africa's healthcare systems.

Twenty-four percent of the sample hospitals were technically inefficient while 36 percent were scale inefficient. Similar hospital efficiency studies in Africa have found hospitals to be both technically and scale inefficient. For instance, Osei et al. (2005) found 47 percent of the public hospitals in Ghana to be technically inefficient while 59 percent were scale inefficient. While investigating the technical efficiency of peripheral health units in Pujehun district of Sierra Leone, Renner et al. (2005) found 59 percent of the 37 health units be technically inefficient, while 65 percent were scale inefficient. A similar study by Zere et al. (2000) and Kirigia et al. (2000), among 55 public hospitals in Kwazulu-Natal, South Africa, found 42 percent of the hospitals to be scale inefficient while 40 percent were technically inefficient.

CHAPTER SEVEN

TOTAL FACTOR PRODUCTIVITY GROWTH RESULTS

7.1 Introduction

Empirical findings with regard to the total factor productivity growth are the subject of this chapter. It looks at the total factor productivity growth indices from the five models estimated. In addition, results with regard to regional differences in hospital total factor productivity growth have been presented. The chapter concludes with a discussion of the empirical findings.

7.2 Total Factor Productivity Growth

The empirical results were obtained by applying the DEA technique to a number of model specifications in the estimation of the Malmquist total factor productivity index. These specifications, based on the variable definitions in Chapter V, are presented in Table 7.1, which shows the input and output combinations used in five model specifications. Models 2 to 5 are based on slight modifications of model 1.

Table 7.1 DEA Malmquist Model Specifications

Variables/Model	1	2	3	4	5
Inputs					
Beds	X	X	X	X	X
All staff grouped together				X	X
Doctors	X	X	X		
Nurses	X	X	X		
Other staff	X	X	X		
Outputs					
Admissions (un-weighted)	X				
Admissions (case-mix adjusted)		X	X	X	X
Outpatient Attendances	X	X	X	X	X
Surgical Operations	X	X			X
Deliveries	X	X			X

The principal total factor productivity change results reported were derived by imposing the assumption of constant returns to scale on each of the models outlined in Table 7.1 and input-oriented DEA. An input-oriented Malmquist index is specified because hospital managers can only influence the resource inputs into health care services but not the type and extent of the health care services demanded by their clients. The Malmquist TFP index represents the productivity of the production point (x_{t+1}, y_{t+1}) relative to the production point (x_t, y_t). The mean TFP indices differ depending upon the model specification. The mean TFP indices, range from 0.945 to 1.024.

The five models yield different TFP indices. Having specified an input-oriented Malmquist productivity change index, the estimated indices are interpreted as follows: a score of less than unity indicates productivity progress in the sense that the hospital delivers a unit of output in period $t+1$ using fewer inputs. In other words, the hospital in period $t+1$ is more efficient relative to itself in period t. Similarly, a score greater than unity implies productivity regress and a unit score indicates constant productivity (Hollingsworth et al., 1999).

Table 7.2 shows the mean, standard deviation, minimum, maximum of the Malmquist productivity index for each of the five models for the 1999-2003 period, as well as the number of hospitals that registered productivity progress and those that saw productivity regress. On average the sampled hospitals registered productivity regress over the sample period across all models except Model 3 where 23 out of 25 hospitals registered a 5.5 percent mean productivity gain. This may be due to the fact that model 3 has four inputs coupled with only two outputs when compared to other models. In general, this indicates that Uganda's district referral hospitals registered productivity losses.

Table 7.2 TFP Results from Five Malmquist Models, Pooled (1999-2003) Dataset

	Model 1	Model 2	Model 3	Model 4	Model 5
Mean	1.022	1.023	0.945	1.002	1.024
Std. Dev.	0.058	0.058	0.098	0.112	0.056
Min.	0.911	0.898	0.782	0.799	0.905
Max	1.120	1.121	1.251	1.352	1.139
>=1	15	11	2	9	9
<1	10	14	23	16	16

Note: Using Input-orientation and Constant Returns to Scale

Table 7.3 shows the Pearson correlation matrix between Malmquist TFP scores from the five models. Four out of ten correlation coefficients are significant at the 5 percent level. The TFP scores between models 1 and 2 are correlated and significant at the 5 percent level, which may be due to the fact that the two models only differ in the treatment of admissions. Whereas model 1 uses unweighted admissions, model 2 employs case-mix adjusted admissions. Models 1and 5 have TFP scores which are significantly correlated at the 5 percent level, possibly because while model 1 disaggregates labor (doctors, nurses and other staff), model 5 lumps together labor into a single factor input. Models 2 and 5 are also significantly correlated at the 5 percent level, perhaps because the former disaggregates labor while the latter aggregates labor. Finally, models 3 and 4 are correlated significantly at the 5 percent level due to how each treats the labor inputs, model 3 disaggregates labor whilst model 4 aggregates it. What accounts for the registered productivity losses in Model 1 is considered next.

Table 7.3 Pearson Correlation Matrix between Malmquist TFP Scores from Five Models

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Model 1	1.0000				
Model 2	0.9938*	1.0000			
Model 3	-0.1804	-0.1209	1.0000		
Model 4	-0.1311	-0.0779	0.8021*	1.0000	
Model 5	0.8532*	0.8568*	-0.1572	-0.0868	1.0000

Note: *Significant at 5% level

Table 7.4 reports the Malmquist TFP index summary of annual means for Model 1. It shows the annual means of Technical Efficiency Change, Technical Change, Pure Efficiency Change, Scale Efficiency Change and Total Factor productivity change over the study period (1999-2003) by means of Model 1.

Given an input-oriented Malmquist TFP index, the mean TFP change of 1.021 indicates that on average, over the sample period, there was a 2.1 percent productivity regress. Looking at the mean technical efficiency change (0.99) and the mean technical or technological change (1.034), productivity losses were largely the result of technological regress. This is because the mean technical efficiency change (0.99) is less than the mean technical or technological change of (1.034). Additionally, since the overall technical efficiency change $(TE\Delta)$ is the product of pure technical efficiency $(PE\Delta)$ and scale efficiency $(SE\Delta)$, such that $TE\Delta = (PE\Delta)(SE\Delta)$, pure efficiency change was 0.999 (99.9 percent) whereas scale efficiency change stood at 0.991 (99.1 percent), this implies that the relative source of technical inefficiency was scale inefficiency.

The result on technical regress as well as that of scale inefficiency has implications for the proper functioning of hospitals. Technical regress may imply that hospitals are not experiencing any technological progress (that they are operating on production possibility frontiers closer to the origin, than further away from it). However, the result of technical regress has to be interpreted with caution because the variable “beds” used to capture a hospital’s capital input is not a comprehensive representation of the capital employed by a hospital. Conversely, scale inefficiency may reflect the fact that hospitals are combining their factor inputs in less than optimal proportions. For instance, looking at labor as a factor input, all hospitals in the sample had fewer staff than they are supposed to have. Across the different labor types, there was fewer staff on the ground than the establishment stipulates. All sample hospitals had unfilled positions. The staffing levels used in the analysis were for the actual staff on the ground and do not include unfilled positions.

Table 7.4 Malmquist TFP Index Summary of Annual Means for Model 1, Pooled (1999-2003) Dataset

Year*	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
2000	0.937	1.106	0.965	0.971	1.036
2001	1.064	0.947	1.039	1.024	1.008
2002	1.001	1.066	1.008	0.993	1.067
2003	0.958	1.016	0.984	0.974	0.974
Mean	*0.990*	*1.034*	*0.999*	*0.991*	*1.021*

**Note that 2000 refers to the change between 1999 and 2000, and so on*
Total Factor Productivity Change = (Technical Efficiency Change)(Technical Change)
Technical Efficiency Change = (Pure Efficiency Change)(Scale Efficiency Change)

The Malmquist TFP index summary of annual means for Model 2 are reported in Table 7.5. Model 2 differs from Model 1, in only one respect that the admissions are adjusted by means of the case-mix index and it is our preferred model. This is because case-mix adjustment tries to account for the complexity of the treated patients.

Table 7.5 Malmquist TFP Index Summary of Annual Means; Model 2, 1999-2003

Year	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
2000*	0.934	1.107	0.967	0.966	1.034
2001	1.058	0.954	1.036	1.021	1.010
2002	1.004	1.063	1.007	0.996	1.067
2003	0.961	1.015	0.988	0.973	0.976
Mean	*0.988*	*1.033*	*0.999*	*0.989*	*1.021*

**Note that 2000 refers to the change between 1999 and 2000, and so on*
Total Factor Productivity Change = (Technical Efficiency Change)(Technical Change)
Technical Efficiency Change = (Pure Efficiency Change)(Scale Efficiency Change)

A comparison of the results of models 1 and 2 reveals that the annual Malmquist TFP index over the sample period remains unchanged even after adjusting the admissions for the hospital's patient load. Technical change falls by 0.001 in comparison to model 1 whereas efficiency change falls by 0.002. Therefore, even after adjusting for the patient load, the sample hospitals registered a total factor productivity decline whose causes are similar to those of model 1.

By means of model 2, figures 7.1 and 7.2 respectively, present the behavior of the TFP index and the efficiency change. Figure 7.1 shows trends of the TFP index vis-à-vis the efficiency change and technical change over the sample period.

Figure 7.1 Malmquist TFP Index versus Efficiency Change and Technical Change, 1999-2003

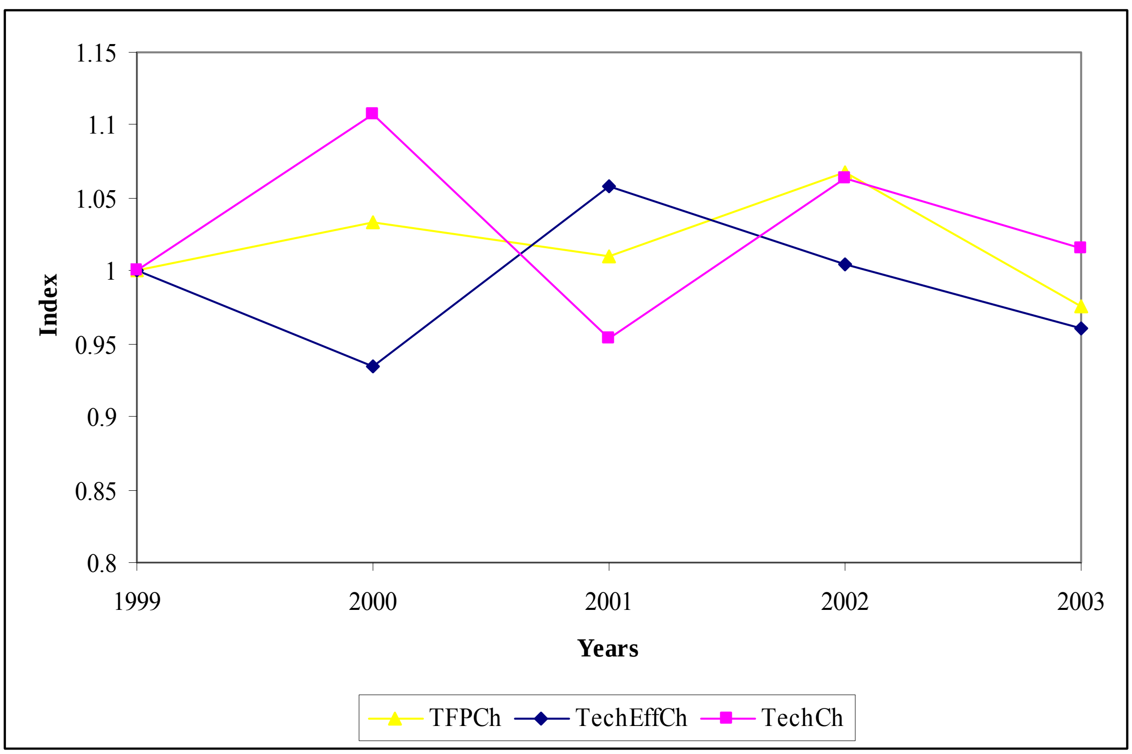


Figure 7.1 shows that overall, over the study period, the TFP decline was accounted for more by technical regress than technical inefficiency. Note that 1999 represents the base year and equals the value of unity. The graphs indicate that total factor productivity regress is driven more by technological declines than technical inefficiency. There might other drivers of productivity growth which have not been incorporated in the analysis. This is because technological change may not necessarily translate into TFP growth. Empirically, however, TFP growth is not necessarily caused by technological change.

Figure 7.2 **Technical Efficiency Change versus Pure Efficiency Change and Scale Efficiency Change, 1999-2003**

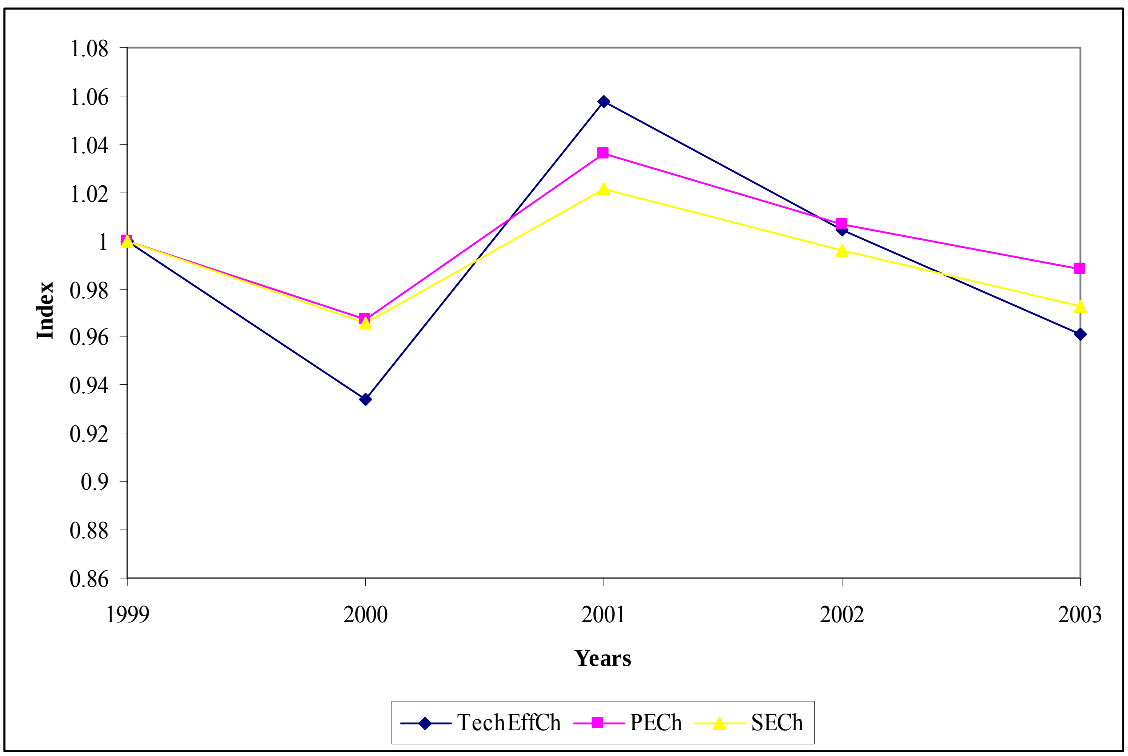


Figure 7.2 depicts the trends of the technical efficiency change vis-à-vis pure efficiency change and scale efficiency change. There was a general decline in the three efficiency changes over the 1999-2003 period save for 2000-2001 when there was a rise in the three efficiency changes. This rise might be attributed to the changes in the management of district hospitals brought about by the launch of the health sector strategic plan I. Over the sample period as shown in Figure 7.2, technical inefficiency was largely due to scale inefficiency. The graphs indicate that technical inefficiency is driven more by scale inefficiency than pure inefficiency.

Malmquist total factor productivity indices using Model 2 over the sample period (1999-2003) for every hospital are presented in Table 7.6. Having specified an input-based Malmquist total factor productivity index, a TFP change greater than unity implies total factor productivity loss, less than unity implies a gain while unity implies no change. On average, over the study period all the sample hospitals registered a total factor productivity loss of 2.1 percent. A closer look at Table 7.6 shows that eleven[26] out of 25 hospitals registered marginal total factor productivity gains that averaged 3.2 percent over the study period.

[26] Bududa, Entebbe, Itojo, Kagadi, Kalisizo, Kapchorwa, Kitagata, Mityana, Nakaseke, Pallisa, and Tororo.

Table 7.6 Malmquist Index Summary of Hospital Means using Model 2, 1999-2003

Hospital	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
Bududa	0.929	1.019	1.000	0.929	0.947
Bugiri	1.000	1.108	1.000	1.000	1.108
Busolwe	1.000	1.017	1.000	1.000	1.017
Bwera	0.998	1.070	1.000	0.998	1.068
Entebbe	0.979	1.019	1.000	0.979	0.998
Gombe	1.037	1.072	1.017	1.020	1.112
Iganga	1.000	1.068	1.000	1.000	1.068
Itojo	0.952	1.035	0.998	0.954	0.985
Kagadi	0.965	1.032	0.926	1.042	0.996
Kalisizo	0.941	1.049	0.995	0.946	0.987
Kambuga	1.000	1.025	1.000	1.000	1.025
Kapchorwa	0.956	1.026	0.998	0.958	0.981
Kawolo	1.025	1.094	1.000	1.025	1.121
Kayunga	1.028	1.044	1.005	1.023	1.073
Kiboga	1.023	1.045	1.023	1.000	1.069
Kiryandongo	1.000	1.044	1.000	1.000	1.044
Kisoro	1.000	1.040	1.000	1.000	1.040
Kitagata	0.994	0.973	1.000	0.994	0.967
Masindi	1.023	1.008	1.023	1.000	1.031
Mityana	0.983	0.981	1.000	0.983	0.964
Mubende	1.000	1.054	1.000	1.000	1.054
Nakaseke	0.962	1.014	1.000	0.962	0.975
Pallisa	1.000	0.950	1.000	1.000	0.950
Rakai	1.000	1.089	1.000	1.000	1.089
Tororo	0.924	0.972	1.000	0.924	0.898
Mean	**0.988**	**1.033**	**0.999**	**0.989**	**1.021**

A comparison of the mean technical efficiency change (0.988) and the mean technical (technological) change (1.033) reveals that total factor productivity losses over the study period were the result of technological regress. However, looking at individual hospitals there are a few hospitals where this general picture does not hold. The total factor productivity losses for Kitagata, Mityana and Pallisa hospitals were due to technical inefficiency rather than technical regress. On average, the sample hospitals were less than fully technically efficient and this was largely due to scale inefficiency (0.989) rather than pure inefficiency (0.999). This may imply that the sample hospitals were not combining their factor inputs in optimal combinations.

Regional Differences in Hospital Total Factor Productivity Growth

To investigate whether there are any regional differences in hospital total factor productivity growth, hospitals were grouped into three categories by region, namely, Eastern, Western and Central. The null hypothesis tested was that there is no difference in the hospitals' Malmquist TFP indices across the three regions. The pooled dataset (1999-2003) was used to estimate Model 2 for the three regions and the results are reported in Table 7.7.

Table 7.7 shows that there are regional differences in the hospitals' Malmquist total factor productivity growth indices and calls for an investigation to find out the possible causes of the differences. The Central and Western regions registered productivity losses while there was a marginal productivity gain in the Eastern region.

On average, the Eastern region recorded a productivity gain of 1 percent, while the Central and Western regions registered productivity losses of, 5.2 percent and 3.1 percent, respectively. It is also apparent that across the three regions, the productivity losses were largely due to technical regress rather than technical inefficiency. Looking at the pure efficiency and scale efficiency changes, it is clear that scale inefficiency rather than pure inefficiency largely explains technical inefficiency in the Eastern and

Western regions while pure inefficiency explains the technical inefficiency in the Central region marginally than scale inefficiency.

Table 7.7 Regional Differences in Hospital Malmquist TFP Indices: Model 2, Pooled (1999-22003) Dataset

Year	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
Central					
2000*	0.982	1.054	1.000	0.982	1.034
2001	1.023	1.005	1.000	1.023	1.029
2002	1.000	1.148	1.000	1.000	1.148
2003	1.000	1.002	1.000	1.000	1.002
Mean	**1.001**	**1.050**	**1.000**	**1.001**	**1.052**
Eastern					
Year	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
2000*	0.984	1.056	1.000	0.984	1.040
2001	0.995	0.951	1.000	0.995	0.946
2002	0.959	1.047	1.000	0.959	1.004
2003	1.010	0.963	1.000	1.010	0.973
Mean	**0.987**	**1.003**	**1.000**	**0.987**	**0.990**
Western					
Year	*Technical Efficiency Change*	*Technical Change*	*Pure Efficiency Change*	*Scale Efficiency Change*	*Total Factor Productivity Change*
2000*	1.000	1.076	1.000	1.000	1.076
2001	0.990	1.034	0.995	0.995	1.024
2002	1.010	0.992	1.005	1.005	1.002
2003	0.982	1.040	1.000	0.982	1.021
Mean	**0.996**	**1.035**	**1.000**	**0.996**	**1.031**

**Note that 2000 refers to the change between 1999 and 2000, and so on*
Total Factor Productivity Change = (Technical Efficiency Change)(Technical Change)
Technical Efficiency Change = (Pure Efficiency Change)(Scale Efficiency Change)

7.4 Discussion

There is a mixed picture with regard to total factor productivity change in healthcare facilities around the world. Whereas it has been found to rise in some studies, it has been found to fall in others. In addition, the causes of either the decline or the rise are mixed. In most of the cases where total factor productivity has declined, technical regress has explained more of the decline than technical inefficiency. The present study found total factor productivity decline mainly due to technical regress than technical inefficiency.

With the exception of Zere et al. (2000), there is no African study that we are aware of, which has addressed the issue of total factor productivity growth in healthcare facilities. Zere and others find that total factor productivity dropped by 12.1 percent over their sample period. This was largely due to technical regress as this present study has found out with regard to Uganda's district referral hospitals in the three regions studied.

In their analysis of the productivity growth in Norwegian psychiatric outpatient clinics, Halsteinli1et al. (2003), found increased overall productivity (2.5 per cent), with an important contribution from increased technical efficiency. Personnel growth had a negative influence on productivity growth, while growth in the share of university educated personnel improved productivity.

The reforms of the National Health System in the UK introduced in 1990 led to substantial changes in the organization of primary health care. Giuffrida (1999) analyzed the efficiency of primary care provision in the English Family Health Service

Authorities (FHSAs) over the period 1990/91–1994/95. Data Envelopment Analysis was used to measure Malmquist indices of productivity change. The analysis indicated a small improvement (between 1.26 and 1.56 percent) in the productivity over the period considered. This increase was attributed to pure technical efficiency improvement and positive change in scale efficiency, while the technology did not show significant change. The analysis suggests that there is very limited scope for productivity gains in this sector.

Sola and Prior (2001) found a 30 percent decline in the Malmquist total factor productivity index for Catalan hospitals which was largely due to technical regress (36 percent). The fundamental cause explaining the drop in the Malmquist index was technical regress which indicated that the best-practice frontier of more efficient hospitals had worsened over the study period. They however, note that their results should be interpreted with caution. They note, on the one hand, that given technological improvement, healthcare has the possibility of improving diagnosis and cure by making extensive use of technology. Conversely, while new technology improves levels of technical quality, it is costly and does not allow for wholesale substitution of older technology as it assumes a complementary role.

Dismuke and Sena (1998) used a two-stage procedure to assess the impact of actual Diagnostic Related Groups (DRG) payment on the productivity of diagnostic technology in Portuguese hospitals over the 1992-1994 time period using both parametric and non-parametric frontier models. The DRG payment system was found to have positively impacted on total factor productivity and technical efficiency of some commonly employed diagnostic technologies in Portugal.

The Malmquist quantity index is based on the concept of a distance function. In this general formulation, a distance function is very much an engineering-type relationship. In its most general form, it does not require assumptions about efficient producer behavior and about constant returns to scale technology. This property makes it a very versatile tool that is also suited for the measurement of non-market input, output and productivity in healthcare.

The DEA model has a number of merits which have enhanced its popularity among productivity analysts. DEA-based Malmquist index is a non-parametric local index, which means that productivity growth and its components are allowed to be producer-specific, time-varying, and there are no restrictions on the temporal patterns. Additionally, it is not necessary to impose any behavioral assumptions such as cost minimization, which is an important advantage in the analysis of public hospitals. Although the literature typically refers to DEA methods as being deterministic, it is possible to incorporate sensitivity analysis through bootstrapping (Simar and Wilson, 1998). However, the TFP estimates do not account for measurement error which limits the use of its findings for policymaking.

CHAPTER EIGHT

CONCLUDING REMARKS

8.1 Introduction

This chapter discusses the study's main conclusions, policy implications, limitations and offers suggestions for further research.

8.2 Discussion

The overall objective of Uganda's National Health Policy (1999) is to reduce mortality, morbidity and fertility, and the disparities therein by ensuring access to the Uganda National Minimum Health Care Package. The overall expected outcome is an effective, efficient, responsive and accountable national health care system. Uganda's hospital sub-sector is a large consumer of scarce health care resources. Over the 1999-2003 period, hospitals used on average 70 percent of the total public sector expenditure on health. Out of the 70 percent of the total public expenditure which is consumed by the hospital sub-sector, on average about 45 percent was allocated to district referral hospitals. Thus, the efficiency of district referral hospitals merits close attention and scrutiny due to their enormous consumption of resources.

The ongoing health sector reforms in Uganda amongst others seek to improve: efficiency; equity; health of the citizenry; and quality of healthcare and sustainability of health services. It is against the aforementioned background that the present study sought to examine the extent to which district referral hospitals being major components of the national health care system, are technically efficient. Thus, the main purpose of this study is to measure the technical efficiency as well as assess total factor productivity changes of district referral hospitals located in the Central, Western and Eastern regions over the 1999-2003 period.

Twenty five out of 38 district referral hospitals are drawn as follows; seven from the Eastern, eight from Western and ten from the Central regions of Uganda constitute the study sample. The Northern region is deliberately left out due to security concerns and because the operating environment of hospitals in this region is not comparable to that of their counterparts in the remaining three regions.

The Hospital Management Information System launched in 1997 is the source of the data for the study. However, the study concentrates upon the 1999-2003 period due to the fact that this timeframe provides a balanced panel dataset. A longitudinal data set was assembled and a common set of input and output indicators is constructed to support the estimation of Data Envelopment Analysis and Malmquist models. Specifically, data on beds, admissions, patient deaths, inpatient days by ward as well as surgical operations, deliveries and outpatient department attendances were collected from the Hospital Annual Reports. Data on the staffing captured in the annual reports were verified from the Hospital Administrator's records.

Nonparametric Data Envelopment Analysis (DEA) is used in the measurement of hospital technical efficiency whilst the DEA-Malmquist index is used in the measurement of hospital total factor productivity change. DEA and DEA-Malmquist are estimated by means of DEAP version 2.1; a Data Envelopment Analysis (DEA) Program developed by Coelli (1996). DEA super-efficiency models are estimated by means of DEA Excel Solver developed by Zhu (2003). Nonparametric bootstrap resampling is conducted by means of FEAR (Frontier Efficiency Analysis with R) developed by Wilson (2005). Technical efficiency scores when patient deaths are incorporated into the analysis are estimated by means of Zhu's (2003) DEA Excel Solver.

The DEA method provides relative technical efficiency scores for the sample hospitals. We test whether efficiency estimates change when we change model specifications. Due to the non-parametric nature of DEA, it is not possible to test this in the usual parametric manner associated with regression analysis. Because of this reason, the study employed a number of models to analyze the sensitivity of DEA results to the inclusion and exclusion of certain input or output variables and testing for the robustness of the results, as has been done in previous studies (Nunamaker, 1985; Valdmanis, 1992). Although the models proved to be robust in this regard, there was some inconsistency across the different model specifications. Therefore, caution should be exercised not to literally interpret the hospital efficiency scores, productivity indices and rankings. Reasonable correlations might be suggestive of convergent validity. However, the correlations were moderate across the different specifications.

The existence of a large amount of random "noise" in the study might be a potential reason for the lack of concurrence across the different model specifications which could be mistaken for inefficiency. The failure of DEA to provide for random "noise" affects the quality of its efficiency and productivity estimates.

The mean efficiency scores differ depending upon the model specification. Technical efficiency scores only refer to the relative performance within the sample.

Hospitals with an efficiency score of unity are efficient relative to all other hospitals in the sample, but may not be efficient by some absolute standard. This is important because inefficiency is inherently unobservable but all that we can do is benchmark hospitals against each other and not against some absolute standard. The different efficiency scores should thus not be interpreted as accurate point estimates of efficiency, but may be interpreted as indicative of the overall trends in inefficiency of certain hospitals. The point estimates of inefficiency are sensitive to specification and measurement error.

The degree of inefficiency and policy response is contingent upon the hospital's operating environment and appropriate action should be taken only after a thorough investigation. While DEA is a useful diagnostic tool, it might not be appropriate to base funding and resource decisions or efficiency targets on the basis of the resultant efficiency estimates (Batavia et al., 1993; Newhouse, 1994; Hadley and Zuckerman, 1994).

While case-mix was provided for by means of length of stay, this inadequately provides for the severity of treated illnesses. Proxies like transfers of patients between hospitals give some indication of severity of illness, as such; improved measures with regard to severity would improve upon the modeling. Nevertheless, Grosskopf and Valdmanis (1993) found that the inclusion of case-mix (as a weighting tool and as a separate output) made no statistical difference to hospital performance measurement in DEA. However, this study argues that case-mix (and severity) ought to be included in healthcare facility technical efficiency and productivity analysis for plausible theoretical rationale and for credibility purposes. To the extent that hospital output is heterogeneous, providing for case-mix is currently the best known avenue to consider this lack of homogeneity. A variety of effective ways exist to measure severity (Thomas and Ashcraft, 1991) and their proper application in health facilities in Uganda should be considered.

Case-mix affects the technical efficiency and productivity of healthcare facilities. Just like differences in the technical quality of care obscure comparisons of technical efficiency among providers, so do differences in case-mix. Case-mix is an important concept although hard to define. As a consequence it has been defined differently by various authors (Hornbrook, 1982a; Hornbrook, 1982b and Tatchell, 1983) for the various ways through which researchers attempt to define and measure case-mix. The available definitions involve some or all of the following information: facilities (or services) available; intermediate and final services actually provided; complexity of the cases treated; and patient characteristics (for instance, gender, age, education status). Ceteris paribus, one would expect efficient healthcare providers with different case-mix to use different levels of factor inputs. For instance, a health facility with a greater proportion of complex cases should be expected to use more resources in production than an otherwise identical facility treating a set of patients with less severe illnesses.

Relative efficiency assessment and target setting based upon only one method may send inappropriate incentives to hospital managers. Under DEA, the Pareto efficiency criterion has the merit of regarding each input and output as being equal in value which allows hospitals to be rated along their best dimensions. Nevertheless, this same merit may create the perverse incentive for hospital managers to act in unethical manner to improve their efficiency rating by engaging in alteration of the input-output mix, political lobbying and creative accounting

(Nunamaker, 1985) if DEA performance measures are indeed incorporated into an incentive scheme.

The merit of employing DEA to measure technical efficiency and productivity lies in the richness of details available from the model solutions and concrete connections to individual hospitals. However, this may be a problem because it is not always easy to find the reasons for specific findings such as why some hospitals are efficient while others are not. Additionally, it is does not provide insights on the drivers of productivity besides technological and technical efficiency changes.

Given that DEA is becoming more widely accepted as a valid efficiency and productivity measurement method, it should be noted that DEA itself is still under development, especially in terms of the choice of which DEA model is appropriate in different potential policy applications. The choice of models specified in terms of inputs and outputs, and the analytical method used (CRS or VRS, weight restrictions or not, output or input oriented among others) will affect the results. Unless the model used is founded upon a robust model of production and is tested for validity, it may produce results which do not in reality reflect the production process.

Inappropriately constructed output measures in any technique could lead hospitals to devote more resources to attain low priority outputs merely to improve their perceived efficiency. Priority outputs should be arrived at through mutual consensus between regulators, hospital managers, and healthcare personnel. However, hospitals should be allowed some time to adjust their activities to agreed-upon priority outputs before their relative efficiency is assessed. Moreover, regulators and hospitals managers should cultivate a culture of mutual trust to ensure cost-effective health services are delivered to the population.

Ultimately, data accuracy is invaluable to any efficiency and productivity analysis as inaccurate data in the DEA methodology does affect the hospital's efficiency and productivity rating but also those of other hospitals. The quality of the data mirrors the random "noise" level and will affect the efficiency and productivity measures arrived at. Therefore, improving upon data quality would perhaps contribute immensely to better efficiency and productivity estimates. In addition to the data currently being collected, data on treatment outcomes, quality of care (re-admission rates, in-hospital mortality, medical errors, hospital infection rates), hospital capital, teaching and research as well as recurrent and development expenditure would be critical. As expected, there may be a lack of self-interest in the accurate reporting of data by hospital managers which underscores the need for commensurate incentives to ensure this. Health facility efficiency and productivity research in the future should consider ways to improve upon the modeling through the inclusion of alternative variables. Omitted variables may have biased the technical efficiency and productivity estimates (Vitaliano and Toren, 1994).

A research agenda which accounts for some or all these concerns will greatly add to our knowledge of hospital efficiency and productivity and better interpretation of this study's findings.

8.3 Conclusions

Data Envelopment Analysis has been employed in Africa (Zere et al., 2000; Renner et al., 2005; Osei et al., 2005); Asia (Lo et al., 1996); Europe (Kooreman, 1994; Jacobs, 2001; Giuffrida and Gravelle, 2001); and the United States (Conrad and Strauss, 1983; Averill et al., 1992; Chattopadhyay and Ray, 1996), to measure the technical efficiency of healthcare providers. The current study adds to this literature.

The following conclusions can be made from this study. It is important to account for case-mix. The method used in this study is rather crude due to data limitations. This calls for more investment in better data management for proper policymaking. Quality of care should be provided for in any efficiency and productivity investigation. It would be crucial to supplement mortality data by routinely collecting information on other measures for instance access, responsiveness as well as patient experiences. There is not much variation in small samples and super-efficiency helps to discern the differences. When bootstrapping is used we cannot discern the differences which could be due to sampling error. This lack of variation may be due to: small sample coupled with relatively large number of factors and/or errors in reporting. This study has unpicked this variation further and has illustrated the value of bootstrapping and super-efficiency. Therefore, it is important that policymakers *cannot* rank hospitals, but rather use DEA as a can opener to explore good practice versus poor performance further by going into individual hospitals to assess what dictates and constrains their proper functioning.

The study has unearthed the prevalence of combined pure technical and scale inefficiencies and low total factor productivity growth. In a poor country with low levels of expenditure on healthcare, growing challenges in healthcare provision and limited access to healthcare services, the existing levels of inefficiency seriously impede the government's efforts to increase the population's access to healthcare services. Additionally, progress toward the attainment of the treasured national health policy objectives, regional (for instance the Abuja Declaration) and global health targets (e.g., the health-related Millennium Development Goals), would be gravely hampered.

8.4 Policy Implications

Although not amenable to neoclassical economics of the firm, there is a need for improved efficiency in Uganda's district referral hospitals to attain more value for money. This is partly due to the rising health expenditures, growing challenges to financing and sustainability, increasing demand to improve performance (e.g., responsiveness, quality, patient satisfaction); which add up to the motive for reform with a view to improving health system performance or value for money. Improved efficiency is one the corner stones of the decentralized healthcare delivery as well as the national health policy. Hence, there is need for evidence-based policy making with regard to improving efficiency and continuously measuring, monitoring, and taking note of the challenges to its realization.

Increasing health facility technical efficiency and productivity may be the only way of reconciling rising demands for healthcare with public financing constraints. There is scope for improvement in the cost-effectiveness of healthcare systems. This is because the health sector is typically characterized by market failures and heavy public intervention, both of which can generate excess or misallocated spending. Consequently, resources are wasted and opportunities missed to improve health status of the populace. Changing how health funding is spent, rather than mere cost-cutting, is crucially vital in achieving better value for money.

Labor is at the hub of the input spokes for any country's healthcare system because it determines how efficiently and effectively inputs and capital investments – such as infrastructure, equipment, consumables and drugs – are utilized, and also determines the consequent impact on care provision. There is need for appropriate training, remuneration and workforce balance given the rising challenges and risks of providing healthcare.

Information systems direct the flow of information, which is the "lifeblood" of a functional health system. The state of the health information management in Uganda is generally poor, which adversely affects the assessment of the country's overall health system performance. The bulk of the existing information systems consist of paper-based form filling, often duplicative, that are rarely ever used to formulate health policy and guide regulatory activities. Thus, the Health Information Management System (HIMS) should be consolidated. This is because better information systems and performance data are needed to support assessment and improvement of health system efficiency. Nevertheless, increasing efficiency may call for some additional, targeted investments (e.g., information systems, records management, and management improvements). There is need to computerize the HIMS for quick data capture and analysis. This should be coupled with proper accommodation for the records department in hospitals given the crucial role it plays in tracking the entity's day-to-day operations. In the majority of cases, the records department is housed in a small back office without proper lighting and hardly any computerization.

The overall expected outcome of Uganda's national health policy is an effective, efficient, responsive and accountable national healthcare system. Nevertheless, to the best of our knowledge there is hardly any healthcare system efficiency study that has been conducted apart from the present one. Yet, efficiency measures are a useful tool for health planning and policy evaluation. There is thus an urgent need to impart expertise in efficiency evaluation techniques to health facility managers along with improving information systems to improve the management of healthcare resources with a proper economic perspective.

Given the rising pressure on healthcare resources in Uganda, health policymakers, health facility administrators and clinicians should search for more efficient ways to deliver health services. Efficiency improvements in the health sector, even in small amounts, can yield savings of resources or expansion of services for the community.

Health policymakers should be cognizant of the efficiency-equity trade-off in their initiatives to promote health system performance. Although hospital technical efficiency and productivity are paramount, equity considerations might necessitate some degree of inefficiency to ensure equitable access to healthcare.

Finally, health policy-making should involve a careful balance of trade-offs, reflecting the weights assigned to a range of important goals and a great deal of uncertainty. The ultimate goal, certainly, is robust population health, but promoting health is not the only consideration. Health policy decisions also have considerable economic goals for instance, efficiency and productivity improvement.

8.5 Limitations of the Study

The study has both conceptual and methodological limitations like any empirical investigation. The study's findings are limited in their generalizability by sample selection bias. Because the study focuses on district referral hospitals, the results from the study may not be generalizable to the entire hospital subsector of Uganda's healthcare system. Although the findings of the study provide useful information on the technical efficiency of selected district referral hospitals, the results should be interpreted cautiously. Data Envelopment Analysis (DEA) yields relative efficiency scores for the sample hospitals and not with respect to any universal or standard benchmark.

DEA makes use of linear programming algorithm to construct an 'efficiency frontier', with the most efficient organizations within a group being studied to define the standard against which the performance of other DMUs is evaluated. The concept of efficiency is thus relative rather than absolute. It is noteworthy that DEA considers any deviation from the efficiency frontier to be the result of technical inefficiency. Thus, measurement errors and random influences on a hospital's output are ignored.

DEA is neither a measure of absolute efficiency nor maximum possible efficiency. It measures only relative efficiency, which may not suffice if the sample of hospitals as a whole is inefficient. Additionally, DEA does not measure effectiveness. The quality of the health care being provided is not being measured in the DEA model. It would also be interesting to investigate the impact of health care provided on the patient's health status.

The principal strength of DEA lies in its ability to combine multiple inputs and outputs into a single summary measure of efficiency without requiring the specification of any a priori weights. Nevertheless, a shortcoming of DEA is that the distribution of efficiency scores is typically highly skewed, with an unknown theoretical distribution, which creates problems when attempting to test hypotheses concerning the relative efficiency of different groups or the changes in efficiency over time.

Input and output variables are arbitrarily selected in DEA. An important practical issue is the selection of the input-output variables. There is need to reflect upon which inputs and outputs can be deemed exogenous and which inputs and outputs are controllable. This is because a meaningful efficiency measure should include controllable dimensions only. Similarly, there is the issue of distinguishing between inputs and outputs. Typically, labelling a variable as an input when it is actually an output (or vice versa) seriously distorts the analysis. The study included inputs and outputs on which data were readily available.

There are currently no reliable empirical specification tests in DEA. Although various specification tests have been proposed based on formal statistical tests and bootstrap approaches, in the nonparametric setting, these tests use minimal assumptions for the sampling distribution, and they are asymptotic by nature. These tests can be very useful for hypothesis testing in large samples. However, the issue of specification testing is relevant mostly for small samples. Unfortunately, none of these tests has been demonstrated to possess acceptable size and power in small samples.

On the conceptual front, the study has the following shortfalls. The study did not capture the explicit process or treatment outcome measures for quality. Differences found in efficiency across units may not necessarily indicate that efficient hospitals produced their output mix by using minimum inputs, but rather may show an unobserved quality variable bias. Therefore, how efficiency is achieved in a nonprice-competitive market remains a prominent issue. Measures such as hospitalization rates, quality of hospial care, waiting time, and the quality of life of people who utilize hospital services should be included in an efficiency analysis if differences in the quality care in the production of hospital outputs are to be taken into account.

There were no data for capital inputs for instance buildings and equipment. As a consequence, capital is approximated by the number of beds per hospital. Beds are often used to proxy for capital stock in hospital studies usually because a reliable measure of the value of assets is rarely available. Efforts should be directed at correctly and regularly establishing the capital input in hospital operations for a proper assessment of their efficiency.

The study has not incorporated patient deaths in the estimation of total factor productivity growth due to lack of an approapriate software package. It would be interesting to find out how total factor productivity behaves through time when this undesirable output is included in the analysis.

Healthcare is an intermediate input from the health production perspective, but an output from the viewpoint of service provision. The service provision perspective was adopted in the measurement of hospital technical efficiency and productivity change in this study. A fundamental concern for adopting the service provision perspective in efficiency analysis is whether efficiency is measured in a manner that reflects the objective of improving upon the health of the population. Thus, measures of efficiency in healthcare services need to be formulated to ascertain the resulting health benefits to the patients instead of resource-use vis-à-vis service provision (for instance, number of outpatient department attendances, deliveries, surgical operations, community outreach programs).

8.6 Suggestions for Further Research

Whilst the study provides valuable insights, it has highlighted limitations that future research can address. The study covers 25 out of 38 district referral hospitals in Uganda. Thus, there is need for a comprehensive study to gauge the performance of the country's entire healthcare system by comparing different levels of healthcare providers.

Given that a big proportion of the health sector's budget is currently being directed toward Primary Health Care, there is also a need to examine its performance. Additionally, the performance of health sub-districts merits an investigation given their position in Uganda's healthcare delivery system. The causes of inefficiency should be investigated and necessary efficiency measures be instituted to augment the government's initiatives to address healthcare access issues in Uganda. There is also a need to estimate the level of efficiency savings in the entire healthcare system.

Future efficiency analysis should seek to measure economic efficiency. Thus, there is need to comprehensively, consistently and regularly gather data on quantities of all inputs, outputs, outcomes, and average or median prices per unit of each input, from private, public and not-for-profit healthcare providers. In order to aid monitoring and evaluation of the impacts of various healthcare reforms on the efficiency of individual healthcare providers through time by means of the Malmquist productivity index analysis, it is necessary to gather data for at least one year prior to the introduction of specific reforms and for subsequent years. The study has established the existence of regional differences in the hospital total factor productivity growth; this necessitates an investigation into the possible causes of the observed differences. Additionally, it is crucial to investigate how patient deaths affect total factor productivity change.

New public management has been implemented in Uganda's hospitals with a view to strengthening them. Hospital strengthening is a broad term that implies all activities that seek to promote the efficiency, effectiveness, and responsiveness of the hospital within its environment. The creation of autonomous hospitals has been one way of providing more direct incentives. Hospital autonomy describes organizational arrangements where the managers have a greater degree of authority than in a traditional, directly managed public service. Thus, it is crucial to evaluate this policy to find out whether it has attained its intended goals, challenges and constraints to its implementation.

This study has used a length of stay-based case-mix index in the adjustment of admissions on the understanding that inpatients with complex illnesses have longer stays and thus cost the hospital more. However, this was due to the lack of cost data for patients in the different wards. Consequently, there is need for a comprehensive case-mix which incorporates some or all of the following information: operating environment of the healthcare provider; healthcare services available; intermediate and final services actually provided; complexity of the cases treated; and patient characteristics (for instance, gender, age, education status).

New health facility efficiency research in Uganda should not only be directed at DEA- and Stochastic Frontier Regression-specific studies, but also on comparative analyses of the results yielded by each type of technique. Comparative studies are needed because DEA and Stochastic Frontier Regression appear to be treated in some parts of the literature as sufficiently close substitutes to afford decision makers a choice and/or a means for cross-validating results, yet in other parts they are treated as complementary methods yielding triangulated results.

The provision of healthcare is a labor-intensive undertaking. Uganda's education sector has over the years produced health personnel of various categories. However, some healthcare personnel have been lost through brain drain to the developed countries. It is crucial to find out how the production of healthcare personnel when adjusted for brain drain has affected the health status of the population. Additionally, it would be insightful to investigate ways of stemming the brain drain of healthcare personnel given the growing challenges of providing healthcare.

BIBLIOGRAPHY

Abramovitz M (1956); “Resources and Output Trends in the U.S. since 1870,” *American Economic Review,* Vol. 46, pp. 5-23.

Aigner D, K Lovell, K and P Schmidt (1977); “Formulation and Estimation of Stochastic Frontier Models,” *Journal of Econometrics* Vol. 6, pp. 21-37.

Aigner DJ and SF Chu (1968); On Estimating the Industry Production Function,” *American Economic Review,* Vol. 58, pp, 826-39.

Aigner DJ, CAK Lovell and P Schmidt (1977); “Formulation and Estimation of Stochastic Frontier Production Function Models,” *Journal of Econometrics*, Vol. 6, pp. 21-37.

Aldred K, A Barr, A Hindle, M Luck, P Shipp and D Rigby (1971); “The Systems Approach to Hospitals,” *Operations Research Quarterly* (1970-1977), Vol. 22, Special Conference Issue, pp. 39-55.

Aletras VH (1999); “A Comparison of Hospital Scale Effects in Short-run and Long-run Cost Functions,” *Health Economics*, Vol. 8, pp. 521-530.

Alexander JA, JRC Wheeler, TA Nahra, and CH Lemack (1998); “Managed Care and Technical Efficiency in Outpatient Substance Abuse Treatment Units,” *Journal of Behavioural Health Services and Research*, Vol. 25, No. 4, pp. 377-396.

Ali AI and LM Seiford (1990); “Translation Invariance in Data Envelopment Analysis,” *Operations Research Letters,* Vol. 9, pp. 403-405.

Ali AI and LM Seiford (1993); “The Mathematical Programming Approach to Efficiency Analysis,” In *The Measurement of Productive Efficiency: Techniques and Applications*, Edited by HO Fried, CAK Lovell and SS Schmidt, (New York, Oxford University Press).

Allen K (1999); “DEA in the Ecological Context - An Overview.” In *Data Envelopment Analysis in the Service Sector*, edited by Westermann G, Gabler Edition, Wissenschaft.

Allen R, AD Athanassopoulos, RG Dyson and E Thanassoulis (1997); “Weight Restrictions and Value Judgements in DEA: Evolution, Development and Future Directions,” *Annals of Operations Research*, Vol. 73, pp. 13-34.

Al-Shammari M (1999); “A Multi-criteria Data Envelopment Analysis Model for Measuring the Productive Efficiency of Hospitals, *International Journal of Operations and Production Management*, Vol. 19, No. 9 pp. 879-890.

Andersen P and NC Petersen (1993); “A Procedure for Ranking Efficient Units in Data Envelopment Analysis,” *Management Science*, Vol. 39, pp. 1261-1264.

Askin RG and CR Standbridge (1993); *Modeling and Analysis of Manufacturing Systems*, New York, NY: John Wiley and Sons, Inc.

Atkinson SE and C Cornwell (1993); “Measuring Technical Efficiency with Panel Data,” *Journal of Econometrics*, Vol. 6, pp. 257-61.

Atkinson SE and PW Wilson (1995); “Comparing Mean Efficiency and Productivity Scores from Small Samples: A Bootstrap Methodology”, *Journal of Productivity Analysis*, Vol. 6, pp. 137-152.

Averill A, T McGuire, B Manning, D Fowler, S Horn, P Dickson, M Coye, D Knowlton and J Bender (1992); “A Study of the Relationship Between Severity of Illness and Hospital Cost in New Jersey Hospitals,” *Health Services Research*, Vol. 5, pp. 587-604.

Banker RD and RC Morey (1989); "Incorporating Value Judgments in Efficiency Analysis,” *Research in Government and Nonprofit Accounting*, Vol. 5, pp. 245-267.

Banker RD, A Charnes and WW Cooper (1984); “Models for Estimating Technical and Scale Efficiencies in Data Envelopment Analysis,” *Management Science* Vol. 30, No. 9, pp. 1078-1092.

Banker RD, RF Conrad and RP Strauss (1986); “A Comparative Application of Data Envelopment Analysis and Translog Methods: An Illustrative Study of Hospital Production,” *Management Science*, Vol. 32, No. 1, pp. 30-44.

Bannerman RH, J Burton and C Wen-Chieh (1993); Traditional Medicine and Health Care Coverage, World Health Organization, Geneva.

Barnum H and J Kutzin (1990); “Public Hospitals in Developing Countries: Resource Use, Costs and Financing.” Population and Human Resources Department, The World Bank: Washington, D.C.

Barnum H and J Kutzin (1993); "Public Hospitals in Developing Countries: Resource Use, Cost, Financing," John Hopkins University Press: Baltimore, Maryland.

Bartelsman EJ and M Doms (2000); “Understanding Productivity: Lessons from Longitudinal Microdata,” *Journal of Economic Letters*, Vol. 38, pp. 569-594.

Batavia AI, RJ Ozminkowski, G Gaumer and M Gabay (1993); “Lessons for States in Inpatient Rate Setting Under the Boren Amendment,” *Health Care Financing Review*, Vol. 15, No. 2, pp. 137-154.

Bates JM, D Baines and DK Whynes (1996); "Measuring the efficiency of prescribing by General Practitioners," Journal *of Operational Research,* Vol. 47, No.12, pp. 1443-1451.

Battese GE (1992); "Frontier Production Functions and Technical Efficiency: A Survey of Empirical Applications in Agricultural Economics," *Agricultural Economics,* Vol. 7, pp. 185-208.

Berg SA, FR Forsund and ES Jansen (1992); "Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking, 1980-89," *Scandinavian Journal of Economics, Supplement*, Vol. 94, pp. 211-228.

Berger AN (1993); "Distribution-free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis*, Vol. 4, pp. 261-292.

Berger AN and DB Humphrey (1997); "Efficiency of Financial Institutions: International Survey and Directions for Future Research", the Wharton Financial Institutions Center Working Paper No. 97-05.

Berki S (1972); *Hospital Economics.* Lexington Books, Mass.

Bickel PJ and DA Freedman (1981); Some Asymptotic Theory for the Bootstrap, *Annals of Statistics*, Vol. 9, pp. 196-217.

Bitran R (1992); "Technical and Economic efficiency in the Production of Health Services; Phase 1: Review of Concepts and Literature and Preliminary Field Work Design," Health Financing and Sustainability Major Applied Research Paper No. 2.

Boussofiane A, RG Dyson and E Thannassoulis (1991); "Applied Data Envelopment Analysis," *European Journal of Operational Research*, Vol. 52, pp. 1-15.

Bravo-Ureta BE and L Rieger (1990); "Alternative Production Frontier Methodologies and Dairy Farm Efficiency," *Journal of Agricultural Economics,* Vol. 41, pp. 215-226.

Breyer F (1987); "The Specification of a Hospital Cost Function, a Comment on the Recent Literature," *Journal of Health Economics*, Vol. 6, pp. 147-157.

Butler JR (1995); *Hospital Cost Analysis*, (Dordrecht: Kluwer Academic Publishers).

Byrnes P and V Valdmanis (1989); "Variable Cost Frontiers and Efficiency: An Investigation of Labor costs in Hospitals. *Lecture Notes in Economics and Mathematical Systems Series*. Berlin: Springer-Verlag, Vol. 332, pp. 958-976.

Caves DW, LR Christensen and WE Diewert (1982b); “The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity”, *Econometrica*, Vol. 50, pp. 1393-1414.

Chambers RG, Y Chung and R Färe (1996); “Benefit and Distance Functions,” *Journal of Economic Theory,* Vol. 70, pp. 407-419.

Chang PL, SN Hwang and WY Cheng (1995); “Using Data Envelopment Analysis to Measure the Achievement and Change of Regional Development in Taiwan,” *Journal of Environment Management*, Vol. 43, No. 1, pp. 49-66.

Charnes A, S Haag, P Jaska, and J Semple (1992); “Sensitivity of Efficiency Classifications in the Additive Model of Data Envelopment Analysis,” *International Journal of Systems Science*, 23, pp. 789-798.

Charnes A, WW Cooper and E Rhodes (1978); “Measuring Efficiency of Decision Making Units,” *European Journal of Operational Research*, Vol. 2, pp. 429-444.

Charnes A, WW Cooper, AY Lewin and LM Seiford (1994); *Data Envelopment Analysis: Theory, Methodology, and Application.* (Dordrecht: Kluwer Academic Publishers).

Charnes A, WW Cooper, B Golany, L Seiford, and J Stutz (1985); “Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions,” *Journal of Econometrics*, Vol. 30, pp. 91-107.

Chattopadhyay S and SC Ray (1996); “Technical, Scale and Size Efficiency in Nursing Home Care: A Nonparametric Analysis of Connecticut Homes,” *Health Economics*, Vol. 5, No. 4, pp. 363-373.

Chen Y (2002); “On Infeasibility and Super-efficiency DEA,” Working Paper.

Cherchye L and T Post (2003); “Methodological Advances in DEA: A Survey and an Application for the Dutch Electricity Sector,” *Statistica Neerlandica*, Vol. 57, No. 4, pp. 410-438.

Chilingerian JA (1993); “Exploring Why Some Physicians’ Hospital Practices are More Efficient: Taking DEA inside the Hospital, in A Charnes, WW Cooper, AY Lewin and LM Seiford (eds.) *Data Envelopment Analysis: Theory, Methodology and Applications*, (Boston: Kluwer Academic Publishers).

Chirikos TN (1998); “Identifying Efficient and Economically Operated Hospitals: The Prospects and Pitfalls of Applying Frontier Regression Techniques,” *Journal of Health Politics*, Vol. 23, No. 6, pp. 879-904.

Chung Y, R Färe and S Grosskopf (1997); "Productivity and Undesirable Outputs: A Directional Distance Function Approach," *Journal of Environmental Management*, Vol. 51, No. 3, pp. 229-240.

Clewer A and D Perkins (1998); *Economics for Health Care Management*, London: FT Prentice Hall.

Coelli T, DSP Rao and GE Battese (1998); *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers, Boston.

Coelli TJ (1996); *A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program*. Centre for Efficiency and Productivity Analysis (CEPA) Working Paper 96/8. Department of Econometrics, University of New England, Australia.

Conrad RF and RP Strauss (1983); "A Multiple-output Multiple-Input Model of the Hospital Industry in North Carolina," *Applied Economics* Vol. 15, No. 3, pp.341-352.

Cooper WW, RG Thompson, and RM Thrall (1996); "Introduction: Extensions and new Developments in DEA", *Annals of Operational Research,* Vol. 66, pp. 3-46.

Cowing TG and AG Holtmann (1983); "Multiproduct Short-run Hospital Cost Function: Empirical Evidence and Policy Implication from Cross Section Data," *Southern Economic Journal*, Vol. 49, pp. 637-653.

Cowing TG, AG Holtmann and S Powers (1983); "Hospital Cost Analysis: A Survey and Evaluation of Recent Studies," *Advance in Health Economics and Health Services Research*, Vol. 4, pp. 257-303.

Cropper ML and WE Oates (1992); "Environmental Economics: A Survey," *Journal of Economic Literature*, Vol. 30, pp. 675-740.

Cummins JD and MA Weiss (1999); "Analyzing Firm Performance in the Insurance Industry Using Frontier Efficiency and Productivity Methods. In *Handbook of Insurance,* Edited by Dionne G) (Boston: Kluwer Academic Publishers).

Debreu G (1951); "The Coefficient of Resource Utilization," *Econometrica,* Vol. 19, pp. 273-292.

Defelice LC and WD Bradford (1997); "Relative Inefficiencies in Production between Solo and Group Practice Physicians," *Health Economics*, Vol. 6, No. 5, pp. 455-465.

Dismuke C and V Sena (1998); "Has DRG Payment Influenced The Technical Efficiency and Productivity Of Diagnostic Technologies In Portuguese Public Hospitals?" An Empirical Analysis Using Parametric And Non-Parametric

Methods," The University of York Discussion Papers in Economics, Number 1998/06.

Dismuke CE and V Sena (2001); "Is there a Trade-off between Quality and Productivity? The Case of Diagnostic Technologies in Portugal," *Annals of Operations Research*, Vol. 107, pp. 101-116.

Dixon S (1990); "The Measurement of Productivity in New Zealand: An introduction," New Zealand Department of Labor Occasional Paper Series (1990/3).

Drake L and B Howcroft (1994); "Relative Efficiency in the Branch Network of a UK Bank: An Empirical Study," *OMEGA*, Vol. 22, No. 6, pp. 429-444.

Dyckhof H and Allen K (2001); "Measuring Ecological Efficiency with Data Envelopment Analysis," *European Journal of Operational Research*, Vol. 132, No. 2, pp. 312-325.

Efron B (1979); "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, Vol. 7, pp. 1-26.

Efron B and G Gong (1983); "A Leisurely Look at the Bootstrap, the Jackknife, and Cross Validation," *American Statistician*, 37, pp. 36-48.

Efron B and R Tibshirani (1986); "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy, *Statistical Science* Vol. 1, pp. 54-77.

Efron B and RJ Tibshirani (1998); *An Introduction to the Bootstrap*, New York: Chapman & Hall/CRC.

Efron B and Tibshirani RJ (1993); *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Ellis R (1992); "Hospital Cost Function Estimation When Firms May not Try to Minimize Total Costs." Manuscript. Boston University, Boston.

Ersoy K, S Kavuncubasi, YA Ozcan and JM Harris (1997); "Technical Efficiencies of Turkish Hospitals: DEA Approach," *Journal of Medical Systems*, Vol. 21, No. 2, pp. 67-74.

Evans RG (1971); "Behavioural Cost Functions for Hospitals," *Canadian Journal of Economics* Vol. 4, No. 2, pp. 198-215.

Färe R and S Grosskopf (1992); "Malmquist Indexes and Fisher Ideal Indexes," *The Economic Journal* Vol. 102, No. 410, pp. 158-160.

Färe R and S Grosskopf (2002); "Two Perspectives on DEA: Unveiling the Link between CCR and Shephard," *Journal of Productivity Analysis,* Vol. 17, No. (1-2), pp. 41-47.

Färe R and S Grosskopf (2003); "Nonparametric Productivity Analysis with Undesirable Outputs: Comment," *American Journal of Agricultural Economics*, Vol. 85, pp. 1070-1074.

Färe R and S Grosskopf (2004); *New Directions: Efficiency and Productivity*, Boston: Kluwer Academic Publishers.

Färe R, S Grosskopf, and CAK Lovell (1994); *Production Frontiers;* (Cambridge: Cambridge University Press).

Färe R, S Grosskopf and P Roos (1995); "Productivity and Quality Changes in Swedish Pharmacies," *International Journal of Production Economics,* Vol. 39, No. (1,2), pp. 137-147.

Färe R, S Grosskopf, B Lindgren and P Roos (1994); "Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach." In *Data Envelopment Analysis: Theory, Methodology and Applications,* edited by A Charnes, WW Cooper, AY Lewin and LM Seiford, (Boston: Kluwer Academic Publishers).

Fare R, S Grosskopf, M Norris and Z Zhang (1994); "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *American Economic Review*, Vol. 84, pp. 66-83.

Farnham PG, SP Ackerman and AC Haddix (1996); "Study Design," In *Prevention Effectiveness: a Guide to Decision Analysis and Economic Evaluation,* edited by Haddix AC, SM Teutsch, PA Shaffer, and DO Dunet.. New York: Oxford University Press.

Farrell MJ (1957); "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A*, CXX, Part 3, pp. 253-290.

Feldstein MS (1967); *Economic Analysis for Health Service Efficiency*. North-Holland, Amsterdam.

Feldstein MS (1968); *Economic Analysis for Health Service Efficiency* (Chicago: Markham Publishing Company).

Feldstein MS (1969); "Discussion" In VR Fuchs (eds), *Production and Productivity in the Service Industries*. National Bureau of Economic Research, New York.

Ferrier GD and JG Hirschberg (1997); "Bootstrapping Confidence Intervals in Linear Programming Efficiency Scores with an Illustration using Italian Banking Data," *Journal of Productivity Analysis*, Vol. 8, pp. 161-176.

Ferrier GD and V Valdmanis (1996); “Rural Hospital Performance and its Correlates,” *The Journal of Productivity Analysis*, Vol. 7, pp. 63-80.

Fetter R, Y Shin, J Freeman, R Averill and J Thompson (1980); “Case Mix Definition by Diagnostic-Related Groups,” *Medical Care*, Vol. 18, No. 2, pp. 1-53.

Försund FR, CAK Lovell and P Schmidt (1980); “A Survey of Frontier Production Functions and their Relationship to Efficiency Measurement,” *Journal of Econometrics,* Vol. 13, pp. 5-25.

Friedman B and M Pauly (1981); “Cost Functions for a Service Firm with Variable Quality and Stochastic Demand: The Case of Hospitals,” *Review of Economics and Statistics*, Vol. 63, pp. 620-624.

Gijbels I, E Mammen, BU Park and L Simar (1999); “On Estimation of Monotone and Concave Frontier Functions,” *Journal of the American Statistical Association*, 94, pp. 220-228.

Giuffrida A (1999); “Productivity and Efficiency Changes in Primary Care: a Malmquist Index Approach,” *Health Care Management Science*, Vol. 2, pp. 11-26.

Giuffrida A and H Gravelle (2001); “Measuring Performance in Primary Care: Econometric Analysis and DEA,” *Applied Economics*, Vol. 33, pp. 163-175.

Greene WH (1993); "Frontier Production Functions," EC-93-20. Stern School of Business, New York University.

Grosskopf S (1996); “Statistical Inference and Non-parametric Efficiency: A Selective Survey”, *Journal of Productivity Analysis*, 7, pp. 161-176.

Grosskopf S and V Valdmanis (1987); "Measuring Hospital Performance: A Non-Parametric Approach." *Journal of Health Economics*, Vol. 6, pp. 89-107.

Grossman M (1972); *The Demand for Health: A Theoretical and Empirical Investigation*. National Bureau of Economic Research, New York.

Gstach D (1995); “Comparing Structural Efficiency of Unbalanced Sub-samples: a Resampling Adaptation of Data Envelopment Analysis,” *Empirical Economics*, Vol. 20, pp. 531-542.

Hadley J and S Zuckerman (1994); “The Role of Efficiency Measurement in Hospital Rate Setting,” *Journal of Health Economics*, Vol. 13, No. 3, pp. 335-340.

Halsteinli V, J Magnussen and SAC Kittelsen (2003); “Productivity Growth in Norwegian Psychiatric Outpatient Clinics: A Panel Data Analysis of the Period 1996-2001,” Health Economics Research Programme at the University of Oslo, HERO 2003.

Hao S and CC Pegles (1994); “Evaluating Relative Efficiencies of Veterans Affairs Medical Centers using Data Envelopment Analysis,” *Journal of Medical Systems*, Vol. 18, No. 2, pp. 55-67.

Hellinger FJ (1975); “Specification of Hospital Production Function,” *Applied Economics*, Vol. 7, pp. 149-160.

Hensher M (2001); “Financing Health Systems through Efficiency Gains,” Commission on Macroeconomics and Health Working Paper Series Paper No. WG3: 2.

Hernandez RA (2002); “Casemix Indices for Acute and Specialized Hospitals in England,” Unpublished MSc Health Economics Dissertation, University of *York*, UK.

Hibbard J, J Stockard and M Tusler (2003); "Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?," *Health Affairs*, Vol. 22, No. 3, pp. 84-94.

Hofler RA and B Rungeling (1994); “US Nursing Homes: Are they Cost Efficient?” *Economic Letters*, Vol. 44, No. 3, pp. 301-305.

Hollingsworth B (2003); “Non-parametric and Parametric Applications Measuring Efficiency in health care,” *Health Care Management Science*, Vol. 6, No. 4, pp. 203-218.

Hollingsworth B, PJ Dawson and N Maniadakis (1999); “Efficiency Measurement of Health Care: a Review of Nonparametric Methods and Applications,” *Health Care Management Science*, Vol. 2, pp. 161-172.

Horn D, R Horn, P Sharkey and A Chambers (1985); “Interhospital Differences in Severity of Illness,” *The New England Journal of Medicine*, July 4, pp. 20-24.

Hornbrook MC (1982a); “Hospital Case Mix: Its Definition, Measurement and Use; Part I. The Conceptual Framework,” *Medical Care Review*, Vol. 39, pp. 1-43.

Hornbrook MC (1982b); “Hospital Case Mix: Its Definition, Measurement and Use; Part II. Review of Alternative Measures,” *Medical Care Review*, Vol. 39, pp. 73-123.

Hotelling H (1936); “Relations between Two sets of Variates,” *Biometrika*, Vol. 28, pp. 321-377.

Hughes A (2003); *Guide to the Economic Performance Measurement*; TPP 01-03; New South Wales Treasury, Australia.

Jacobs R (2001); "Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis," *Health Care Management Science,* Vol. 4, pp. 103-115.

Jaforullah M (1999); "A Comparison of Production Frontier Methods Using Bangladesh Handloom Textile Industry Data," *Economics Discussion Papers*, No. 9911, University of Otago, Dunedin.

Jencks S and A Dobson (1987); "Refining Case-mix Adjustment; The Research Evidence," New English Journal of Medicine, 317:679-686.

Kirigia JM, A Emrouznejad and LG Sambo (2002); "Measurement of Technical Efficiency of Public Hospitals in Kenya: using Data Envelopment Analysis," *Journal of Medical Systems*, Vol. 26, No. 1, pp. 39-45.

Kirigia JM, A Emrouznejad and LG Sambo, N Munguti and W Liambila (2004); "Using Data Envelopment Analysis to Measure the Technical Efficiency of Public Health Centers in Kenya," *Journal of Medical Systems*, Vol. 28, No. 2, pp. 155-166.

Kirigia JM, E Lambo and LG Sambo (2000); "Are Public Hospitals in Kwazulu-Natal Province of South Africa Technically Efficient?" *African Journal of Health Sciences*, Vol. 7, No. 3-4, pp. 25-32.

Kirigia JM, LG Sambo and H Scheel (2001); "Technical Efficiency of Public Clinics in Kwazulu-Natal Province of South Africa," *East African Medical Journal*, Vol. 78, No. 3, pp. S1-S13.

Kittelsen SAC and FR Forsund (1992); "Efficiency Analysis of Norwegian District Courts," *Journal of Productivity Analysis*, Vol. 3, No. 3, pp. 277-306.

Koopmans TC (1951); "An Analysis of Production as Efficient Combination of Activities." In *Activity Analysis of Production and Allocation,* edited by TC Koopmans, (Cowles Commission for Research in Economics, New York).

Kooreman P (1994); "Nursing Home Care in the Netherlands: a Nonparametric Efficiency Analysis," *Journal of Health Economics*, Vol. 13, pp. 301-316.

Kulu-Glasgow I, D Delnoij and D de Bakker (1998); "Self-referral in a Gate Keeping System: Patients' Reasons for Skipping the General Practitioner," *Health Policy*; 45: 221-238.

Kumbhakar SC and CAK Lovell (2000); *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

LaFond A (1995); Sustaining Primary Health Care, Save the Children Fund: New York.

Lave JR and LB Lave (1970); “Hospital Cost Functions,” *American Economic Review* Vol. 60, No. 3, pp. 379-395.

Lave JR, LB Lave and LP Silverman (1974); “Hospital Cost Estimation Controlling for Case-mix,” *Applied Economics*, Vol. 4, pp. 165-180.

Lee K (1991); “The Search for Value-for-money in Health Care: Some UK Observation,” *Economics and Health* Proceeding of the 13th Australian Conference of Health Economists, 22-57. School of Health Management, University of New South Wales.

Lee ML (1971); “A Conspicuous Production Theory of Hospital Behavior," *Southern Economic Journal*, Vol. 38, No. 1, pp. 48-58.

Levy PS and S Lemeshow (1999); *Sampling of Populations: Methods and Applications*, Third Edition, (New York: John Wiley & Sons).

Lewin AY and CAK Lovell (1990); Editors Introduction, *Journal of Econometrics*, Vol. 46, pp. 3-5.

Li X, B Reeves and R Gary (1999); “A multiple Criteria Approach to Data Envelopment Analysis,” *European Journal of Operational Research,* Vol. 115, pp. 507-517.

Linna M (1998); “Measuring Hospital Cost Efficiency with Panel Data Models,” *Health Economics*, Vol. 7, No. 5, pp. 415-427.

Lo JC, KS Shih and KL Chen (1996); “Technical Efficiency of the General Hospitals in Taiwan: An Application of DEA, *Academia Economic Papers*, Vol. 24, No. 3, pp. 275-296.

Lothgren M and M Tambour (1999); “Testing Scale Efficiency in DEA Models: A Bootstrapping Approach,” *Applied Economics*, Vol. 31, pp. 1231-1237.

Lovell CAK (1995); “Measuring the Macroeconomic Performance of the Taiwanese Economy,” *International Journal of Production Economics*, Vol. 39, pp. 165-178.

Lovell CAK and APB Rouse (2003); “Equivalent Standard DEA Models to Provide Super-Efficiency Scores,” *Journal of the Operational Research Society,* Vol. 54, No. 1, pp. 101-108.

Lovell CAK and JT Pastor (1995); “Units Invariant and Translation Invariant DEA Models,” *Operations Research Letters*, Vol. 18, pp. 147-151.

Lynk WJ (1994); “Property Rights and the Presumptions of Merger Analysis,” *Antitrust Bulletin*, Vol.39, pp. 363-383.

Macrae J, A Zwi and L Gilson (1996); “A Triple Burden for Health Sector Reform: ‘Post’-conflict rehabilitation in Uganda,” *Social Science and Medicine*, Vol. 42, pp. 1095-1108.

Magnussen J (1996); “Efficiency Measurement and the Operationalization of Hospital Production”, *Health Services Research*, Vol. 31, pp. 21-37.

Malmquist S (1953); "Index Numbers and Indifference Surfaces," *Trabajos de Estadistica*, Vol. 4, pp. 209-42.

Management Decision and Research Center (1997); “Risk Adjustment: A Tool for Leveling the Playing Field,” Boston, MA: VA Health Services Research and Development Service in Collaboration with the Association for Health Service Research.

Mann JK and DE Yett (1968); “The Analysis of Hospital Costs: A Review Article,” *Journal of Business*, Vol. 41, pp. 191-202.

Mawson P, KI Carlaw and N McLellan (2003); “Productivity Measurement: Alternatives Approaches and Estimates,” New Zealand Treasury Working Paper 03/12.

McKee M (1993); “Routine Data: A Resource for Clinical Audit,” *Quality of Health Care*, 2:104-111.

McMurchy D (1996); *Efficiency Studies of Public Sector Hospitals in the Eastern, Northern and Western Cape Provinces*. HEU Working Paper No.6. Cape Town: Health Economics Unit, University of Cape Town.

Meeusen W and J van den Broeck (1977); “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error,” *International Economic Review*, Vol. 18, pp. 435-44.

Mills A, JP Vaughan and D Smith and Iraj Tabibzadeh (1990): *Health System Decentralization*, World Health Organization, Geneva.

Mobley LR and J Magnussen (1998); “An International Comparison of Hospital Efficiency: Does Institutional Environment Matter?” *Applied Economics*, Vol. 30, No. 8, pp. 1089-1100.

Montoya-Aguilar C and L Marchant-Cavieres (1994); “The effect of economic changes on health care and health in Chile,” *International Journal of Health Planning and Management*, Vol. 9, No. 4, pp. 279-294.

Mooney CZ and RD Duval (1993); *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Beverly Hills, CA: Sage Publications.

Newhouse J (1970); “Toward a Theory of Nonprofit institutions: An Economic Model of a Hospital,” *American Economic Review*, Vol. 60, No. 1, pp. 64-74.

Newhouse JP (1994); “Frontier Estimation: How useful a Tool for Health Economics?” *Journal of Health Economics*, Vol. 13, No. 3, pp. 317-322.

Niskanen WA (1971); *Bureaucracy and Representative Government*. Aldine-Atherton, Chicago.

Nunamaker TR (1983); “Measuring Routine Nursing Service Efficiency: A Comparison of Cost per Patient day and Data Envelopment Analysis Models,” *Health Services Research*, Vol. 18, No. 2, pp. 183-205.

Nunamaker TR (1985); “Using Data Envelopment Analysis to Measure the Efficiency of Non-profit Organizations: A Critical Evaluation,” *Journal of Managerial and Decision Economics*, Vol. 6, pp. 50-58.

Osei D, Kirigia JM, S D’Almeida, MO George, AO Mensah and LH Kainyu (2005); “Technical Efficiency of Public District Hospitals and Health Centers in Ghana: a Pilot Study,” *Cost Effectiveness and Resource Allocation* (Online).

Ozcan YA, MJ McCue and AA Okasha (1996); “Measuring the Technical Efficiency of Psychiatric Hospitals,” *Journal of Medical Systems*, Vol. 20, No. 3, pp. 141-150.

Parker D and W Newbrander (1994); ``Tackling Wastage and Inefficiency in the Health Sector", *World Health Forum*, Vol. 15, pp. 107-13.

Parkin D and B Hollingsworth (1997); “Measuring Production Efficiency of Acute Hospitals in Scotland, 1991-94: Validity Issues in Data Envelopment Analysis,” *Applied Economics*, Vol. 29, No. 11, pp. 1425-1433.

Parliament of The Republic of Uganda Webpage; http://www.parliament.go.ug/social%20rpt7_session3.htm, accessed 27th May 2004.

Pastor JT (1994); “How to Discount Environmental Effects in DEA: an Application to Bank Branches,” Working Paper N. 011/94, Depto. De Estadistica e Investigación Operativa, Universidad de Alicante, Spain.

Pastor JT (1996); “Translation Invariance in Data Envelopment Analysis: a Generalization,” *Annals of Operations Research*, Vol. 66, pp. 93-102.

Pauly MV (1980); *Doctors and their Workshops*, (Chicago: University of Chicago Press).

Pauly MV and M Redisch (1973); “The Not-for-profit Hospital as a Physicians’ Cooperative,” *American Economic Review*, Vol. 63, pp. 87-99.

Peacock S, C Chan, M Mangolini and D Johansen (2001); "*Techniques for Measuring Efficiency in Health Services*," Productivity Commission Staff Working Paper, July.

Pedraja-Chaparro F, J Salinas-Jimenez and PC Smith (1999); "On the Quality of the Data Envelopment Analysis Model," *Journal of the Operational Research Society*, Vol. 50, No. 6, pp. 636-644.

Phillip J and A Hai (1976); *Hospital Cost: An Investigation of Causality in The Nature of Hospital Costs: Three studies*. Hospital Research and Education Trust, Chicago.

Pitt M and LF Lee (1981); "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry," *Journal of Development Economics*, Vol. 9, pp. 43-64.

Portela MCAS, E Thanassoulis and GPM Simpson (2004); "Negative Data in DEA: A Directional Distance Approach Applied to Bank Branches," *Journal of the Operational Research Society*, Vol. 55, pp. 1111-1121.

Ray SC (2004); "The Directional Distance Function and Measurement of Super-Efficiency: An Application to Airlines Data," University of Connecticut, *Department of Economics Working Paper Series,* Working Paper 2004-16.

Ray SC (2004); *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, (Cambridge, Cambrigde University Press).

Renner A, JM Kirigia, EA Zere, SP Barry, DG Kirigia, C Kamara, and LHK Muthuri (2005); "Technical Efficiency of Peripheral Health Units in Pujehun District of Sierra Leone: a DEA Application," BMC *Health Services Research*, (Online).

Republic of Uganda (1997); The Local Government Act (1997).

Republic of Uganda (2003); The Health Sector PEAP Review Paper, Ministry of Health.

Rice RG (1966); "Analysis of the Hospital as an Economic Organism," *Modern Hospital*, Vol. 106, pp. 87-91.

Rosko MD and JA Chilingerian (1999); "Estimating Hospital Inefficiency: Does Case-mix Matter?" *Journal of Medical Systems*, Vol. 24, No. 2, pp. 57-71.

Rubin HR and AW Wu (1992); "The Risk of Adjustment," *Medical Care*, 30:973-975.

Ruggiero J (1998); “A New Approach for Technical Efficiency Estimation in Multiple Output Production,” *European Journal of Operational Research*, Vol. 111, pp. 369-380.

Sachs J (2002), “Macroeconomics and Health: Investing in Health for Economic Development,” Report of the Commission on Macroeconomics and Health work at the Institute for Global Health.

Scheel H (2001); “Undesirable Outputs in Efficiency Valuations,” *European Journal of Operational Research*, Vol. 132, pp. 400-410.

Schmidt P and RC Sickles (1984); “Production Frontiers and Panel Data,” *Journal of Business and Economic Statistics*, Vol. 2, pp. 367-74.

Scientific Peer Review Group (2002); *Report of the Scientific Peer Review Group on Health Systems Performance Assessment*, Geneva: World Health Organization.

Scott RD, SL Solomon, and JE McGowan (2001); “Applying Economic Principles to Health Care,” *Emerging Infectious Diseases* Vol. 7, No. 2, pp. 282-285.

Seiford LM and J Zhu (1998d); “Stability Regions for Maintaining Efficiency in Data Envelopment Analysis,” *European Journal of Operations Research*, Vol. 108, No. 1, pp. 127-139.

Seiford LM and RM Thrall (1990); “Recent Development in DEA: the Mathematical Programming Approach to Frontier Analysis,” *Journal of Econometrics*, Vol. 46, pp. 7-38.

Seiford ML and J Zhu (2002); “Modeling Undesirable Factors in Efficiency Evaluation,” *European Journal of Operational Research*, Vol. 142, No. 1, pp. 16-20**.**

Sharma KR, P Leung and HM Zaleski (1999); “Technical, Allocative and Economic Efficiencies in Swine Production in Hawaii: A Comparison of Parametric and Non-Parametric Approaches,” *Agricultural Economics*, Vol. 20, pp. 23-35.

Shephard RW (1953); *Cost and Production Functions*. (Princeton: Princeton University Press).

Shephard RW and R Färe (1974); "The Law of Diminishing Returns," *Zeitschrift für Nationalökonomie*, pp. 69-90.

Simar L (1992); “Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-parametric and Semi-parametric Methods with Bootstrapping”, *Journal of Productivity Analysis*, 3, pp. 171-203.

Simar L and P Wilson (2000); "Statistical Inference in Nonparametric Frontier Models: The State of the Art," *Journal of Productivity Analysis*, Vol. 13, No. 1, pp. 49-78.

Simar L and PW Wilson (1998); "Sensitivity Analysis of Efficiency Scores: how to Bootstrap in Nonparametric Frontier Models," *Management Science*, Vol. 44, pp. 49-61.

Simar L and PW Wilson (1999); "Estimating and Bootstapping Malmquist Indices," *European Journal of Operational Research*, Vol. 115, pp. 459-471.

Sink DS and TC Tuttle (1989); *Planning and Measurement in Your Organization of the Future. Institute of Industrial Engineers,* Noncross, GA: Industrial Engineering and Management Press.

Smith P (1997); "Model Misspecification in Data Envelopment Analysis," *Annals of Operations Research*, Vol. 73, pp. 233-252.

Sola M and D Prior (2001); "Measuring Productivity and Quality Changes using Data Envelopment Analysis: An Application to Catalan Hospitals," *Financial Accountability & Management*, Vol. 17, No. 3, pp. 219-245.

Solow R (1956); "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, Vol. 70, No. 1, pp. 65-94.

Stone M (2002); "How to Measure the Efficiency of Public Services (And How One Might)," *Journal of the Royal Statistical Society, Series A*, Vol. 165, pp. 405-434.

Tatchell M (1983); "Measuring Hospital Output: A Review of the Service Mix and Case Mix Approaches," *Social Science Medicine*, Vol. 17, No. 13, pp. 871-883.

Thanassoulis E, A Boussofiane and RG Dyson (1996); "A Comparison of Data Envelopment Analysis and Ratio Analysis as Tools for Performance Measurement," *OMEGA, International Journal of Management Science*, Vol. 24, No. 3, pp. 229-244.

Thomas JW and MLF Ashcraft (1991); "Measuring Severity of Illness: Six Severity Systems and their Applicability to Explain Cost Variations," *Inquiry*, Vol. 28, pp. 39-55.

Thompson JD, FB Fetter and CD Mross (1975); "Casemix and Resource Use," *Inquiry* XII, December, pp. 300-312.

Thompson RG, PS Dharmapala and RM Thrall (1993); "Importance for DEA of Zeros in Data, Multipliers, and Solutions," *Journal of Productivity Analysis*, Vol. 4, pp. 379-390.

Thrall RM (1996); “Duality, Classification and Slacks in DEA,” *Annals of Operations Research*, Vol. 66, pp. 109-138.

Tofallis C (2001); “Combining Two Approaches to Efficiency Assessment,” *Journal of the Operations Research Society*, Vol. 52, pp. 1225-1231.

Tulkens H (1993); On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit,” *Journal of Productivity Analysis*, Vol. 4, No.1/2, pp. 183-210.

Uganda Bureau of Statistics (2004); 2003 Statistical Abstract.

Uganda Bureau of Statistics (2005): 2004 Statistical Abstract.

Uganda Bureau of Statistics (2005); 2002 Uganda Population and Housing Census, Main Report.

Uganda Bureau of Statistics (2005); 2004 National Service Delivery Survey.

United Nations (1946); *Constitution of the World Health Organization*.

Valdmanis V (1992); “Sensitivity Analysis for DEA Models: An Empirical Example using Public vs NFP Hospitals,” *Journal of Public Economics*, Vol. 48, pp. 185-205.

Varatharajan D (1999); “Improving the Efficiency of Public Health Care Units in Tamil Nadu, India: Organizational and Financial Choices,” Takemi Program in International Health Harvard School of Public Health, Research Paper No. 165.

Varian HR (1978); *Microeconomic Analysis*. New York: WW Norton Co Inc.

Vitaliaono DF and M Toren (1994); “Frontier Analysis: A Reply to Skinner, Dor and Newhouse,” *Journal of Health Economics*, Vol. 13, pp. 341-343.

Wagstaff A (1989); “Estimating Efficiency in the Hospital Sector: A Comparison of Three Statistical Cost Frontier Models,” *Applied Economics*, Vol. 21, No. 5, pp. 659-672.

Wagstaff A and H Barnum (1992); “Hospital Cost Function for Developing Countries.” *Policy Research Working Paper Series 1044,* Washington, D.C.: World Bank Population and Human Resource Department.

Watts CA and TD Klastorin (1980): “The Impact of Case mix on Hospital Cost: A Comparative Analysis,” *Inquiry*, Vol. 17, pp. 357-367.

Webster R, S Kennedy and L Johnson (1998); “Comparing Techniques for Measuring the Efficiency and Productivity of Australian Private Hospitals,” Working Papers in Econometrics and Applied statistics No. 98/3.

Wilson PW (1995); "Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs," *Journal of Business and Economic Statistics*, Vol. 11, pp. 319-323.

Wilson PW (2005); "FEAR: A Package for Frontier Efficiency Analysis with R," unpublished Working Paper, Department of Economics, University of Texas at Austin, Austin, Texas.

Winslow C, Bode RK, D Felton, D Chen and PR Meyer (2002); "Impact of Respiratory Complications on Length of Stay and Hospital Costs in Acute Cervical Spine Injury," *Chest*, Vol. 121, pp. 1548-1554.

World Bank (2003); "Traditional Medicine Practice in Contemporary Uganda," Indigenous Knowledge Notes No. 54.

World Bank (2005); *Improving Health Outcomes for the Poor in Uganda; Current Status and Implications for Health Sector Development, Human Development Sector*, (Africa Region: The World Bank).

World Health Organization (2000); *The World Health Report 2000. Health Systems: Improving Performance,* (Geneva: World Health Organisation).

Worthington A (2004); "Frontier Efficiency Measurement in Health Care: A Review of Empirical Techniques and Selected Applications," *Medical Care Research and Review*, Vol. 61, No. 2, pp. 135-170.

Wouters A (1993); "The Cost and Efficiency of Public and Private Healthcare Facilities in Ogun State, Nigeria," *Health Economics*, Vol. 2, No. 1, pp. 31-42.

Zere EA, T Addison and D McIntyre (2000); "Hospital Efficiency in Sub-Saharan Africa: Evidence from South Africa," *South African Journal of Economics*, Vol. 69, No. 2, pp. 336-358.

Zhu J (1996b); "Robustness of the Efficient DMUs in Data Envelopment Analysis," *European Journal of Operations Research*, Vol. 90, No. 3, pp. 451-460.

Zhu J (2003); *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Boston: Kluwer Academic Publishers.

APPENDIX

Appendix Table 1 Uganda's Health Sub-sector Allocations (%): 1999-2003

Budget Area	***1999***	***2000***	***2001***	***2002***	***2003***
District Referral Hospitals	32	44	47	49	54
National Referral Hospitals	22	13	14	12	12
Regional Referral Hospitals	14	10	11	8	8
Sub-total	*68*	*67*	*72*	*69*	*74*
Ministry of Health Headquarters	30	31	26	29	24
Other Agencies	2	2	2	2	2
Grand Total	**100**	**100**	**100**	**100**	**100**

Source: Parliament – The Republic of Uganda Webpage:
http://www.parliament.go.ug/social%20rpt7_session3.htm, accessed 27th May 2004.

Appendix Table 2 Hospitals in Uganda as at 2001

National Referral Hospitals

District	Hospital	Number of Beds
Kampala	Mulago	1500
Kampala	Butabika	850

Regional Referral Hospitals

District	Hospital	Number of Beds
Arua	Arua	278
Gulu	Gulu	250
Hoima	Hoima	139
Jinja	Jinja	431
Kabale	Kabale	200
Kabarole	Fort Portal	300
Masaka	Masaka	330
Mbale	Mbale	332
Mbarara	Mbarara	250
Soroti	Soroti	140

Government District Hospitals

District	Hospital	Number of Beds
Kotido	Abim	100
Adjumani	Adjumani	100
Gulu	Anaka	100
Apac	Apac	100
Kumi	Atutur	100
Sironko	Bududa	100
Bugiri	Bugiri	100
Bundibugyo	Bundibugyo	100
Tororo	Busolwe	100
Kasese	Bwera	100
Mpigi	Entebbe	147
Mpigi	Gombe	100
Iganga	Iganga	100
Mbarara	Itojo	100
Kotido	Kaabong	100
Kibale	Kagadi	100
Rakai	Kalisizo	70
Rukungiri	Kambuga	100
Kapchorwa	Kapchorwa	100
Mukono	Kawolo	100
Mukono	Kayunga	100
Kiboga	Kiboga	100
Masindi	Kiryadongo	100
Kisoro	Kisoro	109
Bushenyi	Kitagata	100
Kitgum	Kitgum	185
Lira	Lira	254
Masindi	Masindi	100
Mubende	Mityana	100

Moroto	Moroto	90
Moyo	Moyo	100
Mubende	Mubende	100
Luwero	Nakaseke	100
Nebbi	Nebbi	100
Pallisa	Pallisa	100
Rakai	Rakai	60
Tororo	Tororo	217
Yumbe	Yumbe	100

Uganda Catholic Medical Bureau Non-governmental Hospitals

District	Hospital	Number of Beds
Apac	Pope John, Aber	205
Arua	Maracha	180
Bushenyi	Comboni	100
Gulu	Lacor	457
Iganga	St. Francis, Buluba	164
Kabarole	Virika	105
Kampala	St. Francis, Nsambya	361
Kampala	Rubaga	300
Kamuli	Kamuli	153
Kasese	Kilembe	110
Kisoro	St. Francis, Mutolere	203
Kitgum	Kalongo	360
Kitgum	St. Joseph, Kitgum	250
Kotido	Morulem	40
Masaka	Kitovu	220
Masaka	Villa Maria	120
Mbarara	Ibanda	152
Moroto	St. Kizito, Matany	244
Mpigi	Kisubi	74
Mpigi	Nkozi	90
Mukono	Nagalama	150
Mukono	Nkokonjeru	100
Mukono	St. Francis, Nyenga	100
Nebbi	Angal	234
Nebbi	Nyapea	121
Rukungiri	Nyakibale	165
Soroti	Iwala	135
Tororo	St. Anthony, Tororo	110

Uganda Protestant Medical Bureau Non-governmental Hospitals

District	Hospital	Number of Beds
Arua	Kuluva	160
Bushenyi	Ishaka	80
Kabarole	Kabarole	70
Kampala	Mengo	285
Kampala	Namungona	20
Kampala	St. Stephen's, Mpererwe	56
Kasese	Kagando	220
Kumi	Kumi	200
Kumi	Ngora	180
Mbarara	Rushere	50
Moroto	Amudat	75
Lira	Amai	80
Luwero	Kiwoko	114
Rukungiri	Kisiizi	200

Uganda Muslim Medical Bureau Non-governmental Hospitals

District	Hospital	Number of Beds
Arua	Oriajini	50
Kampala	Kibuli	70
Kampala	Old Kamapala	30

Private Hospitals

District	Hospital
Jinja	Kakira
Kampala	International Hospital
Kampala	Kololo
Kampala	Bugolobi
Kampala	Mckenzie Valley
Kampala	Makerere University
Gulu	Gulu Independent

Hospitals Belonging to Other Ministries

District	Hospital
Kampala	Mbuya Military Hospital
Kampala	Murchison Bay Hospital
Luwero	Bombo Military Hospital
Kampala	Makerere University Hospital