Discussion Paper Series

The KENYA INSTITUTE for PUBLIC
POLICY RESEARCH and ANALYSIS

Cost Efficiency and Scale Economies of Kenya's Water Service Providers

Hellen Kalunde Musyoki

DP/146/2013

THE KENYA INSTITUTE FOR PUBLIC POLICY
RESEARCH AND ANALYSIS (KIPPRA)

YOUNG PROFESSIONALS (YPs) TRAINING
PROGRAMME

Cost Efficiency and Scale Economies of Kenya's Water Service Providers

Hellen Kalunde Musyoki

Infrastructure and Economic Services Division
Kenya Institute for Public Policy
Research and Analysis

KIPPRA Discussion Paper No. 146
2013

KIPPRA in Brief

The Kenya Institute for Public Policy Research and Analysis (KIPPRA) is an autonomous institute whose primary mission is to conduct public policy research leading to policy advice. KIPPRA's mission is to produce consistently high-quality analysis of key issues of public policy and to contribute to the achievement of national long-term development objectives by positively influencing the decision-making process. These goals are met through effective dissemination of recommendations resulting from analysis and by training policy analysts in the public sector. KIPPRA therefore produces a body of well-researched and documented information on public policy, and in the process assists in formulating long-term strategic perspectives. KIPPRA serves as a centralized source from which the Government and the private sector may obtain information and advice on public policy issues.

Published 2013

Bishops Garden Towers, Bishops Road
PO Box 56445-00200 Nairobi, Kenya
tel: +254 20 2719933/4; fax: +254 20 2719951
email: admin@kippra.or.ke
website: http://www.kippra.org

ISBN 9966 058 12 6

The Discussion Paper Series disseminates results and reflections from ongoing research activities of the Institute's programmes. The papers are internally refereed and are disseminated to inform and invoke debate on policy issues. Opinions expressed in the papers are entirely those of the authors and do not necessarily reflect the views of the Institute.

This paper is produced under the KIPPRA Young Professionals (YPs) programme. The programme targets young scholars from the public and private sector, who undertake an intensive one-year course on public policy research and analysis, and during which they write a research paper on a selected public policy issue, with supervision from senior researchers at the Institute.

KIPPRA acknowledges generous support from the Government of Kenya, the African Capacity Building Foundation (ACBF), and the Think Tank Initiative of IDRC.

Abstract

Kenya has adopted Integrated Water Resources Management and instituted water sector reforms, which have led to commercialization of water supply and sanitation (WSS) services to enhance sustainability in the management of water resources and improve water supply efficiency. Inefficiencies in water utilities are a major cause of poor access to WSS. This study seeks to estimate which size of water service providers is operationally efficient and to establish an optimal size after determining the impact of key sector variables on the cost structure of the Kenya's WSPs. A variable cost function with outputs, input prices and network characteristics is formulated as a transcendental logarithmic model and estimated using panel stochastic frontier analysis. Results show that the volume of water produced and treatment level, prices of inputs (materials and administration), the number of connections, unaccounted for water, staff productivity and also the population density per connection influence the running costs and affect the cost efficiency of WSPs. The cost efficiency across WSPs regardless of the size is nearly nonexistent at around 1.1 percent and economies of scale, output and customer density are consistently present only in the medium and large WSPs. The very large and small firms exhibit negative economies of scale and should not be expanded further. Instead, the small WSPs may be merged to form medium or large firms. These exhibit positive economies and serve an average of 800,000 people each year. The very large ones may be retained at their current size.

Abbreviations and Acronyms

CAACs	Catchment Area Advisory Committees
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
ECD	Economies of Customer Density
EOD	Economies of Output Density
EOS	Economies of Scale
GoK	Government of Kenya
MDGs	Millennium Development Goals
ML	Maximum Likelihood
MSL	Minimum Service Levels
MWI	Ministry of Water and Irrigation
MWMP	National Water Master Plan
O&M	Operations and Maintenance
SFA	Stochastic Frontier Analysis
SPAs	Service Provision Agreements
UFW	Unaccounted for Water
WARIS	Water Regulation Information System
WASREB	Water Services Regulatory Board
WRMA	Water Resources Management Authority
WSBs	Water Services Boards
WSP	Water Service Provider
WSPs	Water Service Providers
WSS	Water Supply and Sanitation

Table of Contents

1. Introduction

1.1 Background

The Dublin statement on Water and Sustainable Development of 1992 noted that efficient management of Water Supply and Sanitation (WSS) services is a key component of the United Nations' agenda for sustainable development. The International Water Association (IWA) also uses the language of sustainable development in describing water systems to include those for both the delivery and use of water. As such, the performance of water sectors remains a high international priority because global expectations have not been met by existing water initiatives (Water Operators' Partnership-WOP, 2009)). Kenya's blueprint for development (Kenya Vision 2030) under the social pillar proposes to ensure access to adequate water and improved sanitation to all by year 2030. Similarly, target 10 of the Millennium Development Goal (MDG) 7 aims to "halve the proportion of people without sustainable access to safe drinking water and adequate sanitation services" by year 2015. These priorities signify the need for concerted efforts to improve service quality, expand networks and optimize operations of water utilities.

Kenya has adopted Integrated Water Resources Management (IWRM), of which sustainable water supply is a major component. The government also enacted the Water Act of 2002 (Government of Kenya, 2007) to address similar issues. The ensuing water sector reforms were anchored on incorporating all stakeholders to enhance efficiency and sustainability in the management of water resources.

Inefficiencies in African water utilities are a major cause of poor access to WSS services. Both the water services strategy and water resources management strategies currently being implemented in Kenya prioritize sustainability if the long-term associated human rights are to be met (Ministry of Water and Irrigation, 2008). This calls for efforts to increase efficiency in existing systems to reduce wastage, improve service quality and secure cash flows since, in the face of scarce resources, revenues are insufficient to cover operating costs or expand service coverage (WOP, 2009).

1.2 Water Supply and Sanitation Sector in Kenya

During the period before 2002 when the role of water management in Kenya was performed by the public sector, it was associated with poor management, under-funding and inadequate budgetary provisions for operation and maintenance (Government of Kenya, 2007). The Water Act enacted in 2003 targeted efficiency, sustainability and affordability. It provided for the creation of new institutions at both national and local levels to facilitate separation of policy from service delivery.

For oversight on conservation and management of the water resource, the Water Resources Management Authority (WRMA) was established with regional offices and Catchment Area Advisory Committees (CAACs) in all Kenyan localities. The Water Services Regulatory Board (WASREB) was formed as the independent WSS industry regulator to oversee provision of water and sanitation. Eight regional Water Services Boards (WSBs) located in each catchment block were created, mandated with owning and developing WSS infrastructure. The WSBs contract Water Services Providers (WSPs) through Service Provision Agreements (SPAs) as service provision agents. The WSPs maintain their utilities' infrastructure and hence strive to operate at full cost recovery to meet operating costs and attain sustainability. They also pay a regulatory levy to WASREB, abstraction fees to WRMA and administration charges to the respective WSB (WASREB, 2011b). These layers of authorities also impose a cost burden on the WSPs.

WSPs differ in size (as indicated in Table 1.1), and operate in differing environments since they are located in distinct areas within which they are prescribed to provide formal water and sanitation through service agreements.

Table 1.1 is based on information submitted by WSPs for annual performance assessment carried out by WASREB and, therefore, reflects only the portion of the WSS sector whose performance is evaluated by the regulator. The total population living within the service areas targeted by these WSPs is about 20.6 million compared to the 8.6 million people who have actually been served (WASREB, 2012), implying that the demand for WSS services is yet to be met.

Table 1.1: Number of WSPs by size and population served annually

Size of WSP (Category)	2005/06	2006/07	2007/08	2008/09	2009/10	2010/11
Very Large ≥35,000 connections	2	2	3	4(3.5)	4(3.55)	5(4.21)
Large – 10,000-34,999 connections	3	15	18	19 (2.1)	17(5.53)	17(2.39)
Medium - 5,000-9,999 connections	8	12	17	21(0.73)	14(1.09)	15(1.29)
Small <5,000 connections	12	26	34	33(0.57)	27(0.96)	28(0.71)
	25	55	72	77 (6.9)	62(8.13)	65(8.61)

NB: The numbers indicate total WSPs in each category and brackets show millions of people served

Source: WASREB (various)

WASREB assesses WSPs' performance based on nine key sector indicators, which influence overall sector performance. Their annual reviews indicate that coverage levels for WSS are still low. There is high unaccounted for water (ufw) equivalent to about Ksh 8.6 billion per year in revenue terms. The reliability and quality of WSS is unpredictable, staff numbers are unbalanced (in some cases overstaffed), and only about 50 per cent of the WSPs manage to achieve varying levels of cost recovery (WASREB, 2011a).

Further, most capital investments are financed from public funds (WASREB, 2011b) where about 60 per cent is from development partners. Efforts to guarantee cost recovery have seen at least 37 Regular Tariff Adjustments (RTAs) approved by the regulator for WSPs. However, by 2011, only 56 per cent of WSPs could recover O&M costs. This percentage falls to 25 per cent among the small WSPs, which constitute 44 per cent of the total number considered, and incidentally also have the most expensive tariffs that impact negatively on low income populations (WASREB, 2011a). According to Government of Kenya (2007), WSS coverage must improve over time to justify the utilization of public resources. This means that water utilities must demonstrate value for money in addition to recovering costs for sustainability.

The 2009 Census revealed that overall access to piped water declined over the last two decades from 32 per cent in 1989 to 31 per cent in 1999 and to 30 per cent in 2009. The IMPACT report (a performance report of the water services sector available also at the WASREB website) of 2011 confirmed this trend where urban populations served by WSPs declined to 39 per cent in 2009/10 from 46 per cent in 2008/09, with only marginal improvement in sanitation access from 46 per cent to 56 per cent in the same period. WASREB (2011a) attributes the poor performance of utilities to lack of proper investment planning and high expenditures on administration. Based on the performance evaluations, small WSPs are inefficient and thus need to be clustered to gain scale economies.

The perception of inefficiency of small WSPs will therefore be assessed in this study by applying a frontier framework for econometrically analyzing the inefficiencies of WSS to identify basis and the extent to which such clustering can be done. This is done by exploring the cost structure of WSPs to evaluate the causes of cost efficiencies and determine existence of economies of scale (EOS). There is currently no study in Kenya that addresses the efficiency question among water utilities. Identification of the key potential areas for improved operational efficiency will inform policy on the strategic areas of intervention, optimal structure of WSS sector, and choice of the best strategies to adopt (including clustering).

1.3 Problem Statement

Kenya is generally characterized by low access levels to WSS, posing a challenge for the sector institutions. Registered WSPs dotted across the country target to provide WSS to an average of 60 per cent of the country's population, but achieve only about 35 per cent (WASREB, 2011a). Small WSPs constitute 45 per cent of all WSPs but serve only 8 per cent of the targeted population, while the large ones comprising about 30 per cent supply more than 80 per cent of the total population served annually. The substantial variation in performance by WSPs of different sizes is attributed to their capacity to realize economies of scale (Worthington and Higgs, 2011), and this determines their desirable structure. For the last five years (2005/06 to 2010/11), WASREB has been evaluating WSPs based on their independent performance, with prevailing recommendations to cluster small WSPs since they are considered inefficient, unsustainable and are generally outperformed by larger ones.

The performance evaluations did not assess the existence of economies of scale in any size of WSPs, and also did not establish the optimal size in which EOS may be expected to materialize. The proposal to cluster small WSPs presents a paradox, since larger WSPs may not necessarily effectively minimize costs or benefit from economies of scale, as any gains realized may be offset by the increased expenditure for extending supply to distant lower density areas. Therefore, determination of the reasons for (un)sustainability is necessary through identification of the important cost factors and the degree to which they cause inefficiency for the WSP. Further, the presence or absence of EOS needs to be established to justify the chosen structure for the water sector, where suggested consolidations can be supported by establishing optimally sized WSPs.

1.4 Objectives

The main objective of this study is to establish whether small WSPs are indeed inefficient as compared to the larger ones, in order to justify clustering to create larger WSPs where economies of scale offer cost containment benefits.

The specific objectives are to:

(i) Determine the impact of various key sector indicators on the costs of Kenyan WSPs

(ii) Estimate the cost efficiency score for WSPs

(iii) Establish the optimal size of WSP for Kenya

Merging of water utilities is a strategic management tool applied to enhance their performance efficiency, through cost minimization accruing from economies of scale. Empirical evidence will inform strategic policy interventions in operations. An inefficiency measure will indicate the value for money of the resources used by each WSP, while presence of scale economies will point to potential cost reductions associated with increases in size. It will also determine strategic investments, if necessary, to create larger WSPs where demand is high, to increase production capacity and serve more customers.

This study contributes to the water economics literature in Kenya, and the findings will be instrumental for county governments in fulfilling their constitutional mandate of providing water and sanitation. County administrators will therefore need to base their decisions on whether to merge existing WSPs or choose an appropriate size of utilities that best suits their given circumstances. Finally, for optimal inter-county cooperation, institutions established under the devolved governance structure must take cognizant of efficiency and sustainability of water utilities as they tend to have widespread externalities cutting across borders.

1.5 Scope of the Study

While recognizing social and ecological considerations that are key to efficiency, this study focuses only on economic considerations in terms of cost efficiency. It is an investigation of the cost structure of WSPs in Kenya, by using secondary data on water produced as the main output. Its focus is limited to the period following enactment of the Water Act 2002, when performance data is assumed to be consistently available for comparisons across WSPs.

The study covers the six year period for which data is captured in the Water Regulation Information System (WARIS) hosted by the regulator from the year 2005/06 to date. However, consistent data was only available for 2008/09 to 2010/11, further limiting the study scope to these three years. The number of WSPs covered was thus based on an average of 77 that submitted annual data to WASREB for evaluation and benchmarking through WARIS. This data is used to generate the annual IMPACT report and, as may be expected, it covers only variables required by WASREB for performance evaluation. It does not allow for differentiation of the default environments in which the WSPs operate for all the years to suit the purposes of this study.

2. Literature Review

2.1 Background

A general indicator of potential efficiency benefits from mergers has been found to be economies of scale, which refers to cost advantages obtained by a production unit as it increases scale of operation. This is normally assessed by analysis of the cost structure of utilities to compare returns to scale across different sizes, thus determining an optimal size. Various studies of this nature have been undertaken in different countries from various parts of the world. The findings, though mostly consistent, may however not be transferable since utilities in each country or even region operate under different environments/conditions (WOP, 2009, Antonioli and Fillipini, 2001 and Farsi *et al.*, 2003).

WASREB (2010) recognizes that cost-reflective tariffs may cushion and allow sustainability of most of the WSPs. However, poor governance, dilapidated infrastructure leading to inadequate capacity to meet demand, and small size of WSPs where economies of scale are ignored constitute other factors that must be addressed. According to WASREB (2011a), utilities must take advantage of economies of scale to enhance cost-effective and affordable WSS services and at commercially viable levels. This would also create possibilities for cross-subsidization to benefit vulnerable and marginalized groups, avoid unjustified costs to consumers, and thus lead to progressive realization of the human right to water and sanitation (Government of Kenya, 2007).

2.2 Theoretical Literature

2.2.1 Cost structure of water utilities

The overall costs (or expenditure) required to operate a water utility are categorized either as operating costs or capital costs. Operating expenditure (O&M) is the day-to-day expenditure incurred by a water utility in running its business, including administration charges, costs on personnel, chemical and electricity plus other miscellaneous costs. Capital expenditure, on the other hand, relates to those amounts typically invested in long-term assets, which can be depreciated over time.

Worthington and Higgs (2011) suggest that one of the acceptable long-run cost objectives for water utilities is the production of desired output as stipulated by regulation and/or as required by customers at minimum cost. Cost efficiency thus entails operation at the frontier with cost minimization in the production process.

When technical efficiency is achieved, maximum output is produced using the given set of inputs. Minimum costs are incurred to produce that given output at existing input prices and production technology. Sjodin (2006) split technical efficiency into pure technical (cost) efficiency to describe the success in converting inputs to outputs and scale efficiency to describe operation at the most productive size for decision making units (DMUs).

The basic key ingredients for provision of WSS services are water quantity and quality, both of which are affected by environmental attributes. WSPs need to generate revenues to pay for environmental services, which can adversely affect the quality of water used, hence affecting the condition of infrastructure (Madrigal *et al.*, 2010). Poor quality of raw water – requiring extra purification or treatment before entering the system - and scarcity or water losses can cause utilities to be inefficient as they impose high costs-to-production value (Martins *et al.*, 2012; Guerrini *et al.*, 2011 and Follmi and Miester, 2011).

WSS services entail social considerations, including quality of services to consumers in addition to environmental obligations. While the increasing and competing water uses impose a greater economic value to water than its social value, the government's mandate of equitable access to water moderates the valuation. Still, it is clear that the social water demand need not supersede economic water needs, as the competing uses form the basis for mobilization of resources necessary for sustainability in water resources management. Therefore, in addition to consistency in customer satisfaction, WSS services should satisfy the technical efficiency criteria, as well as managerial effectiveness (Correia and Marques, 2011).

The outsourcing arrangement adopted for the provision of WSS services can also affect the efficiency of utilities. For instance, Kenya adopted commercialization involving the creation of accountable and financially autonomous wholly owned enterprises. These institutions deliver water and sanitation services, while the principal retains asset ownership, regulates tariffs and controls service. The enterprises are also publicly managed by boards of directors drawn from the principal (government), consumer representatives and other stakeholders (Government of Kenya, 2002). These layers of authority institutions add to the costs of WSPs. Stone and Webster (2004) find that consolidation of common support services or asset management functions can lead to significant cost savings, reduced financing costs and pooled water resources management risk.

Regulation safeguards citizens against extraction of monopoly rents, ensures water quality is adequate, and guarantees investors return on long term assets. According to Abrate *et al.* (2008), a regulation system is based on local regulatory authorities' entitlement to determine final customer tariffs, and plan and monitor the capital investment programmes as well as quality levels. They establish long-term economic and financial plans of the WSS business, which becomes the basis for tariff setting for WSPs. In Italy, for instance, local authorities are supposed to provide incentives for efficiency based on the comparison of their budget plans with a given benchmarking formula defined at the national level. The supply agents are assumed to incorporate adequate efficiency improvements by using better knowledge of specific operating environments to separate the effect of heterogeneity from cost inefficiencies.

Water utilities tend to be natural monopolies localized in specific areas of the country. According to Correia and Marques (2011), natural monopolists can undercut costs by exploiting economies of scale. Water supply and sanitation is a core infrastructure service typified by high sunk costs, a sub-additive cost function and economies of scale (Stone and Webster, 2004; Guerrini *et al.*, 2011; Sjodin, 2006). This begs the conclusion that smaller utilities operate at higher average total costs, wasting scarce resources and transferring costs to consumers through higher prices. This necessitates nationalization of water supply and sanitation provision with regulated price controls requiring monitoring of minimum service levels. Economies of scale are, however, not sufficient prerequisites for efficiency, and thus may not be used to prescribe small scale production where there are diseconomies of scope (Baumol, 1977).

Regulation is also a necessity for water supply and sanitation services as it controls the market by rigorously comparing the producing utilities through efficiency rankings. Regulators also assess the potential efficiency gains for individual WSPs from better joint use of inputs, and the effects of exogenous factors to advise management of critical areas of focus (Estache and Kouassi, 2002).

2.2.2 Efficiency estimation

Environmental engineers use economic and financial tools to analyze water systems. These include engineering economics, where the time value of money is compared between discrete alternatives; microeconomics defines theoretical frameworks, optimization selects the optimal choice (least cost or maximum benefits), game theory evaluates the winners or losers and finally risk analysis evaluates the reliability (safety) of the selected solution (Heaney, 2009). The analyses carried out in this research and arguments arrived at thereof will borrow from variants of these tools.

Worthington and Higgs (2011) recognize that assessment of the efficiency of water utilities is necessary to highlight deficiencies in their management, and recognize and quantify the effect of structural factors and barriers to effective outcomes of production. Three key measures used are the technical, allocative and productive (total economic) efficiency. Technical efficiency, which is defined as production of "maximum (minimum) possible output (input) from (for) a given set of inputs (outputs)" (Worthington and Higgs, 2011), refers to the physical relationship between resources used and the outputs; allocative efficiency refers to the optimal combinations of inputs to maximize outputs; and finally economic efficiency subsumes both technical and allocative efficiency in operations.

Empirical measurement of the productive performance of water utilities is usually based on frontier efficiency measurement techniques. In the frontier methods, a set of most efficient solutions is determined, and then observed performance measured against them. Most studies use production frontiers where output is specified as a function of the inputs. However, since WSPs produce multiple outputs, the cost frontier typical for water utility efficiency analyses is preferred. With the cost frontier, efficiency cannot be decomposed into allocative and technical, hence all deviations from the frontier are taken to reflect cost inefficiency (Worthington and Higgs, 2011).

Cost efficiency measures indicate the possible reductions in cost for an efficient WSP. The efficiency measure is usually assessed against an efficient frontier, since cost functions are not directly observable and estimation is either non-parametric or parametric, with deterministic or stochastic specification of models and equations.

Parametric estimation assumes a functional form of an efficient frontier defined a priori, while the non-parametric approach calculates the frontier empirically from a sample of observations (Shih *et al.*, 2004, M-Zamorano, 2004). Parametric approaches include deterministic ('full frontier') and stochastic models. Deterministic models are estimated by either mathematical programming or econometric techniques, and assume away the role of exogenous variables and specification errors on the measured efficiency, thus the disturbance term represents the inefficiency level. Conversely, stochastic frontier procedures apply econometric techniques to model technical inefficiency independently by introducing a decomposed error term, which separates out the effects of data errors and uncontrollable factors from the inefficiency term (Murillo-Zamorano, 2004). Shumilkina (2010) shows that performance can also be measured using indexing methods, where ratios compare performance between peer utilities or by econometric models where equations describe relationships between specific indicators and other factors.

Various authors propose common outputs, inputs and environmental factors applicable in water utilities' efficiency studies. Common outputs include amount of water produced, treated or delivered with characteristics of water quality, reliability of supply and number of connections to network as well as input variables such as customer density (Worthington and Higgs, 2011). Guerrini *et al.* (2011) proposes the production inputs as staff productivity measured by staff per 1000 connections, the quality of pipe networks measured by unaccounted-for water, quality of service measured by number of distribution interruptions, and quality of customer service and quality of water measured by compliance with chemical and microbiological standards.

2.2.3 Scale economies in water utilities

According to the United States Environmental Protection Agency (EPA, 2012), effective utilities are operationally resilient, thus sustainable. They optimize, have improved quality and adequacy, customer satisfaction and stakeholder support. This sets an important standard for WSPs in Kenya, as they operate to enhance adequate access to water services to all. Their mandate forms the basis of WSPs' performance as assessed against minimum service levels (MSLs) and internationally recognized benchmarks for water sector objectives (WASREB, 2008). These attributes can only be achieved where the Decision Making Unit (DMU) operates at the most efficient scale, minimizing costs and producing maximum outputs.

Returns to scale can be defined as the minimization of costs at all levels of outputs, such that there is a less than proportionate increase in costs as outputs and scale of operation are increased. According to Kim (1987), the existence and nature of economies of scale in an industry determines the required public policy or industry practice (Worthington and Higgs, 2011; Fraquelli and Moiso, 2005). If economies of scale exist, this provides an argument for creation of large firms to maximize on these benefits, and since consolidation in such an industry would benefit the small firms, correct assessment of the magnitude of economies of scale can avert biased policy decisions. This raises the question of what size or range of sizes of utilities will realize economies of scale.

Economies of scale and density, which can be estimated from the cost function, are important determinants of water policies and guide production planning and tariff setting (Baranzini and Faust, 2009). In addition, according to Antoniolli and Filippini (2001), even when two water companies produce similar output, the cost level could be different because of differences in the network size. As observed by Triebs *et al.* (2012), economies of scale and scope are modeled and estimated using cost functions that are common to all firms in an industry.

2.3 Empirical Literature Review

Kim (1987) estimated a multi-product translog cost function (TCF) while analyzing the scale and scope economies to determine the cost minimizing number of water utilities. The variables of study included the amount of water delivered to residential and non-residential customers as outputs, service distance and capacity utilization rate as network characteristics and input prices of labour, capital and energy. The main findings were that water utilities generally exhibit constant returns to scale, but did not rule out economies of scale for small firms and slight diseconomies for the larger ones.

Ashton (1999) estimated a variable cost model of the Unite Kingdom water industry using data of 20 English and Welsh water companies. Using a restricted variable cost function, he modeled the production structure of the water supply industry and found that slight but significant diseconomies of scale with substantial diseconomies of capital utilization existed in the industry.

With a panel data set of 32 water distribution firms operating at provincial levels over the period 1991-1995 Antoniolli and Fillipini (2001) estimated a variable cost function using three inputs: labour, capital and energy plus a single output, volume of water distributed. The number of customers and size of water system - defined by length of pipe network - were included to allow for the distinction of economies of output density, customer density and scale. Using a Cobb-Douglas formulation, they found that the price of labour, water loss and service area characteristics were significant in explaining the costs of WSPs.

In an application of frontier analysis by Estache and Kouassi (2002), an unbalanced panel of 21 utilities was used to assess the potential for efficiency improvements and importance of scale of operation for African utilities. Their results indicated that institutional problems of corruption and governance significantly drive performance of water firms. In another study, Shih *et al.* (2004) examined production costs of water supply systems to estimate their economies of scale. They estimated the total unit cost and individual component cost elasticities of production, finding positive economies of scale. They concluded that the observed economies may indeed reflect production economies, or may imply that larger systems buy at a bargain unlike smaller systems whose inputs have higher unit costs. In addition, physical mergers of systems are not necessary for the object to realize such economies.

Fraquelli and Moiso (2005) attempted to verify both scale economies and inefficiency using a thirty-year unbalanced panel data from business plans of 18 water utilities distributed throughout Italy. They estimated a TCF frontier in order to assess the behaviour of returns to scale, the inefficiency score and the impact of network characteristics. Their model was based on total costs, the amount of delivered water, and the inputs consisted of the price of labour, electricity and the price of materials, services and capital. The network length and level of losses as output characteristics were also included. In their findings, network characteristics were significant in explaining the inefficiency scores. The presence of scale economies suggested that the supply structure could improve if fragmentation was reduced at the local level.

Filippini, Hrovatin and Zoric (2008) applied the stochastic TCF to estimate inefficiency and EOS for an unbalanced panel of water distribution utilities in Slovenia. They found that inefficiency scores obtained using different model specifications were not robust, and the specifications applied did not separate unobserved heterogeneity of utilities from inefficiency. They concluded, like Brown (2005), that inclusion of environmental and exogenous factors and using either generalized least squares (GLS) or maximum likelihood (ML) estimation with random effects, heterogeneity could be controlled for. This would enable deliberation on substantive differences existing across geographic regions, and hence enable identification of the specific determinants of a utility's performance. This way, the accuracy and precision of performance measurement would be expected to inform and improve decision making.

Baranzini and Faust (2009) also estimated a TCF for Swiss utilities to explore their cost structure. Their model was based on a variable cost function with one output - water delivered; with three inputs – the price of labour, energy and materials; and the other variables were capital stock, number of customers, proportion of water to different types of customers and network losses. Their key finding was that economies of production density decreased with the size of utility. Similarly, Zschille and Walter (2011) estimated a cost function to calculate the economies of scale and density for 72 utilities in Germany for the period 1998-2007. They found that this varied significantly depending on the size, where the largest utilities showed diseconomies and a few utilities operated below the optimal size. They concluded that there existed high cost reduction potentials and, with improved efficiency, consumers could enjoy price reduction.

Horn and Saito (2011) estimated a stochastic cost frontier using a true fixed-effect model for 831 Japanese water utilities from 1999 to 2008 to assess their cost efficiency and economies of scale. Results showed that smaller water utilities enjoyed increasing returns to water delivery volume as well as positive economies of scale compared to larger utilities. They concluded that merging water utilities into larger scale of operations is not always suitable, and water utilities should instead be of an optimal size. Urakami and Parker (2009) also assessed the effects of consolidation among Japanese water utilities by analyzing a hedonic cost function. They found that the cost savings realized were generally offset by extra expenditure for expanding supply.

Worthington and Higgs (2011) used stochastic functions of operating and capital costs to calculate product-specific economies of scale and scope for 55 major urban water utilities over a four-year period of 2005/6-2008/9. They found that strong economies of scale existed at relatively low levels of output. This implied that horizontal merger of utilities in close proximity would realize efficiency benefits as long as no significant network cost investments were needed. They also found product-specific economies of scale in chemical compliance, water quality and service complaints, and the number of connected properties. Capital costs for water losses and water main breaks realize economies at relatively high levels of output only attainable by the large WSPs.

Zschille (2012) used Data Envelopment Analysis (DEA) to analyze the potential efficiency gains from mergers between water utilities at the county level in Germany. The key finding is that reducing individual inefficiencies leads to the greatest efficiency gains, but mergers also have gains in terms of technical and scale efficiency.

2.4 Overview of Literature

Empirical literature has shown that an optimal scale for water utilities exists and is dependent on the operating environment including utility-specific circumstances. There is no universal conclusion to be drawn for any country or region (Zschille, 2012; Triebs *et al.*, 2012). The sample used in this study was characterized by great variations in the outputs of small and large WSPs, hence outliers were expected to pose serious challenges in over/understatement of efficiency scores. SFA, which enables separation of statistical errors from the inefficiency term (Horn and Saito, 2011) was selected for estimation as the alternative DEA, which is deterministic and includes all residuals in the inefficiency term (M-Zamorano, 2004).

The empirical works reviewed show that to calculate the economies of scale in production, one must explore the cost structure of water utilities. This has been done by specifying either a total cost function (Shih *et al.*, 2004; Fraquelli and Moiso, 2005; Horn and Saito, 2011) or variable cost function (Ashton, 1999; Antonioli and Fillipini, 2001; Baranzini and Faust, 2009) with functional forms such as Cobb-Douglas, log-linear and the more common flexible translog formulation. The estimations are based on either cross sectional or panel datasets, with the latter preferred as it allows consideration of multiple observations for individual firms over time. Other reasons for estimating the cost function are that it accommodates multiple outputs as explanatory variables. It also caters for cases where output prices are determined in discretionary policies, and profit maximization is not the overriding objective as in the case of WSS. This enables calculation of alternative cost indices for policy analysis, in order to suggest choice areas of potential cost reduction benefits. One may also identify opportunities for efficiency improvements (Stone and Webster, 2004) as well as provide measures to reduce marginal costs as determinants of planning and tariff setting (Baranzini and Faust, 2009).

The key variables that affect costs of water utilities (including WSPs in Kenya) are identified in literature as output with key characteristics of water quality, reliability, chemical treatment and number of metered connections as well as inputs and prices of inputs. Incorporation of structural factors of the network includes environmental variables such as length of network, customer density, staff productivity, water losses, proportion of customer types, improved environmental quality standards and service coverage which differentiate the heterogeneous WSPs (Nauges & Berg, 2008; Von Hirschhausen *et al.*, 2008; Fillipini *et al.*, 2008).

Tsegai (2009) lists the choices of models available as the Cobb-Douglas, Quadratic, Translog and generalized Leontief cost functions but suggests that the translog function is the most flexible functional form. It does not require the assumptions of homotheticity and separability *a priori* and it enables a second order approximation to the unknown cost function (Nauges and Berg, 2008).

Since the details of the production process are not within the scope of the current study, translog flexible functional form of the cost function is chosen to describe the relationship between costs and output as influenced by input prices and environment variables. The cost function is appropriate to take account of the linear, non-linear and cross-product terms found in the specifications for more than one output, and allows the economies of scale to vary with different levels of input.

3. Methodology

3.1 Theoretical Framework

For a WSP to effectively increase coverage for WSS, it must be technically efficient to minimize costs and maximize desired outputs by internalizing environmental effects without compromising its production process. Efficiency improves productivity (the productivity ratio is used as an important measure of performance) and should be a desirable operational requirement for WSPs. Coelli *et al.* (2005) posit that the estimation of cost functions and the study of cost efficiency applies stochastic frontier analysis (SFA) to model these functional relationships as proposed by theory.

The cost function of an efficient WSP represents the minimum costs that are applicable for similar utilities. A cost frontier (representing the borderline minimum cost) therefore needs to be defined and estimated for the i^th water utility. The models used in this analysis also allow for the estimation of economies of scale as done by Kim (1987) and Triebs *et al.* (2012), among others.

Based on the textbook definition of economies of scale, it is assumed that WSPs utilize a vector X of N exogenous inputs to produce a vector Y of M exogenous outputs. Thus, the cost function is expressed as:

$$C(Y/X) = Min(PX) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

where C is the cost of producing Y, P is a vector of the input prices and $C=\{X/Y(P,X)\geq 0\}$ the convex input requirement set. Thus, the minimum cost of production is now determined by the output and input prices given by $C(Y,P)$. This function contains information on both input prices and output quantities, thus satisfies the conditions for a dual multi-product cost function. The ratio of the observed costs of WSPs to the estimated frontier cost then gives the cost inefficiency score.

Economies of scale, also referred to as returns to scale, are important basis to examine the potential for economic benefits of running larger or smaller entities. Kim (1987) indicates that economies of scale are usually defined in terms of the relative increase in outputs resulting from a proportionate increase in inputs. The formal measure, he suggests, is given by the relationship of average cost to marginal cost. Thus, quoting Baumol (1976) and Panzar and Willig (1977) for any m number of outputs, a local measure of economies of scale or the overall returns to scale for a multi-product firm can be given by equation (3.2).

$$EOS(Y,P) = \frac{C(Y,P)}{\sum_i^m MC_i} = \frac{1}{\sum_i^m \varepsilon CY_i} = \frac{1}{\frac{\delta \ln C}{\delta \ln Y_i}} \ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

Where

i=1,2,.........n are the DMUs under consideration,

$MC_i = \frac{C}{Y_i} * \varepsilon CY_i$ is the marginal cost of the i^{th} output, and

$\varepsilon CY_i = \frac{\partial \ln C}{\partial \ln Y_i}$ is the cost elasticity of the i^{th} output.

This implies that the overall degree of economies of scale in the short run is given by the ratio of production cost to the revenues that would accrue to the firm by pricing the outputs at marginal costs. Variable costs are used in this study, as the WSPs only lease the infrastructure, thus their capital costs are assumed to be fixed in the short term.

If $SL>1$, economies of scale exist; if $SL=1$, there are constant returns to scale; and if $SL<1$ there are diseconomies of scale. This also implies that the revenues generated from pricing outputs at their MC is below, equal to or exceeds cost of production, respectively. Thus, by operating at the minimum efficient scale producing output at minimum costs, if the firm experiences economies of scale, its must price its product at above the MC to recover costs (Kim, 1987).

The underlying assumption in the above measure of overall economies of scale in equation 3.2 is that the output is homogenous and increases by the same proportion. For network industries such as WSPs, however, Fraquelli and Moiso (2005) observe that the output (W, volume of water produced) varies together with the change in output characteristics, thus a distinction between the economies of output density and scale is required. Economies of output density (EOD) is the relative increase in operating costs brought about by a relative increase in output, keeping all other variables constant. When the number of customers is allowed to vary as production increases, then the economics of customer density (ECD) can show the behaviour of production costs.

The presence of density or scale economies shows that the average cost falls when the output or the size of firm increases, thus providing useful information about the optimum WSP size from an operating point of view (Fraquelli and Moiso, 2005).

3.2 Conceptual Framework

A WSP represents a DMU that utilizes resources as inputs in production to generate required outputs. In the Kenyan case, WSPs may not minimize costs in the short to medium term with respect to capital, since investment in infrastructure is the mandate of respective WSBs, thus capital stock can be assumed to be fixed in their individual perspectives. Since WSP outputs are determined by demand, SPA conditions and MSLs, the WSP is only at liberty to minimize input costs, while internalizing the effects of their environment to achieve optimization.

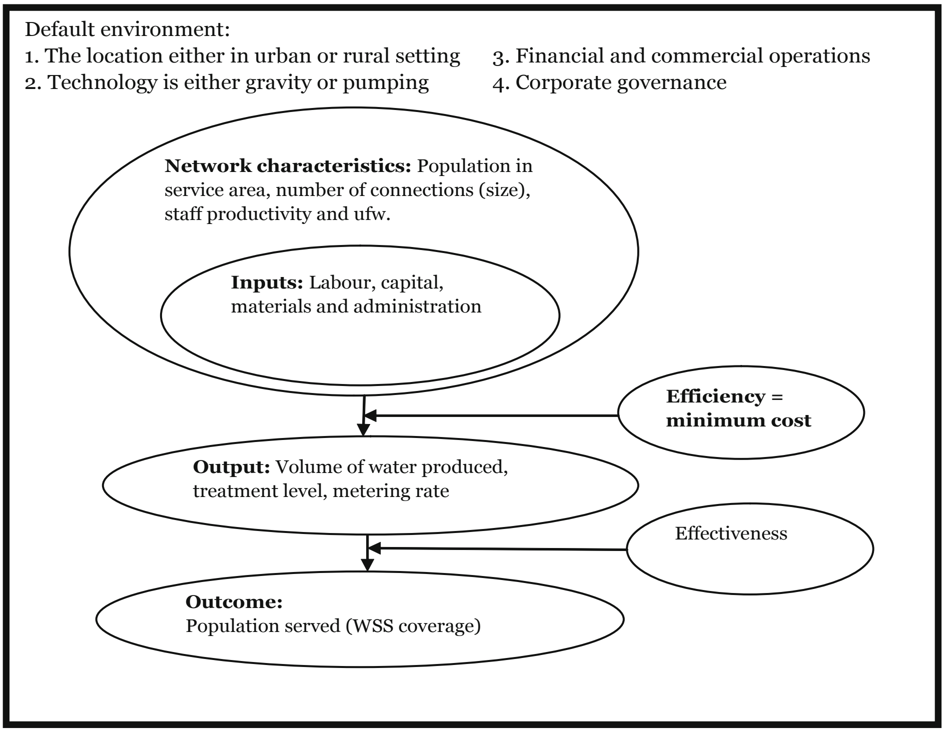
All functional areas involved in converting inputs to outputs for a specified outcome, including the operating environment, determine efficiency of water utilities. Corporate governance, human resources, accountability towards customers, financial, commercial and technical operations are all potential loopholes for inefficiency. However, as observed by Worthington and Higgs (2011), the tendency of larger WSPs in Kenya to outperform smaller ones may suggest that they enjoy cost benefits from economies of scale.

The objectives of WSPs broadly include targets for providing water to populations living within the areas of jurisdiction as prescribed by MSLs in their respective SPAs. Thus, WSPs utilize inputs usually in terms of labour and various forms of capital. In addition, they must internalize the effects of their operating environment to minimize costs, while achieving required outputs. Cost recovery of expended costs is necessary for sustainability. Costs are either in capital investment for infrastructure and/or as O&M costs for production processes, treatment, storage and distribution to water users. Antoniolli and Fillipini (2008) suggest that network characteristics such as number of connections and production capacity also distinguish between the returns to output and scale effects.

Conventionally, the functional ability of a water utility is assessed through analysis of efficiency/productivity, effectiveness and impact in relation to broad objectives. Efficiency assessment has been identified as a prerequisite to ensuring value for money, where a utility gets to understand their sources of inefficiencies, learn about potential solutions from peers, and thus prioritize their activities. Following Kim (1987), the most important objectives for WPSs are fulfilled by increasing the number of people served with acceptable quality of reliable potable water. Since they supply different classes of spatially distributed consumers, the actual volume produced and delivered cannot be used in isolation to measure how well they perform. Therefore, efficiency, defined as the capacity to minimize the consumption of inputs given a certain amount of output (Romano and Guerrini, 2011), becomes a useful consideration. This notion on how WSPs meet their objectives is represented in Figure 3.1.

The default environments in which the WSPs operate in influence the attempts of cost minimization (Estache and Kouassi, 2002). However, since they are not differentiated in this research due to data unavailability, their effect is assumed to be more or less the same for the attainability of efficiency by each WSP. Therefore, the individual network characteristics will be used to determine how efficiently inputs can be turned into necessary outputs to achieve the desired outcome effectively.

Figure 3.1: Conception of the production path for WSPs in Kenya



The following research questions sought to be answered in this study:

(a) What is the impact on WSP operating cost of the various sector indicators?

(b) What are the efficiency levels of the WSPs?

(c) What is the optimal size of WSPs given their cost structures and outputs?

3.3 Model Specification

To satisfy the three objectives of the study, the following variable cost model used by Antoniolli and Fillipini (2001) and Botasso and Conti (2007) is borrowed to analyze the cost structure of WSPs in Kenya.

$$C = F\left((Outputs(Q);WTLWCSRMR)(Inputprices(P);PL,PM,PA\right)quasi-fixedcapital(K)NC;(environment(Z);PD,SP,LO))\cdots\cdots\cdots\cdots(3.3)$$

where C represents the variable *O&M* cost, which is the sum of all expenditures on personnel (P_L), energy and chemicals (materials) P_M, and other costs, taken as the price of administration P_A. The price of electricity is weighted by the supply reliability in hours of supply per day. Information on capital expenditure is unavailable from the data sources used and, therefore, the number of connections (a most consistent form of capital available for the sampled WSPs) is used as a proxy for capital stock, which is fixed in the medium term. *W* is the annual volume of water produced in cubic meters; *TL* represents a quality adjustment measure (Bottasso and Conti, 2003) given as the level of water treatment to comply with standards; *MR* represents the metering rate of connections attained by WSP, and *WC* and *SC* represent the proportion of the population in service area that are served with water and sanitation services, respectively. The environment/network characteristics are represented by the population density (*PD*) obtained as the number of people living in the service area divided by the connections (Stone and Webster, 2004); SP which is the staff productivity given by the average number of staff per 1000 connections a standard measure of efficiency of staff (Worthington and Higgs, 2011); and the system water losses/ufw *LO*, according to Fraquelli and Moiso (2005), as a structural phenomenon in the water sector. The number of connections *NC* indicates the quasi fixed capital of the WSPs, as these are not at their long term equilibrium as they also vary with scheduled investments and time.

The differences in network and environmental characteristics influence the production process and, therefore, the costs (Filippini and Farsi, 2008). Hence, the model selection is shaped by the assumptions on these variables and their expected influences on the running costs of WSPs, and whether they vary over time. The generic TCF is as in equation 3.4:

$$\begin{aligned} InC_i = {} & \beta_o + \sum_a \beta a InQ_a + \sum_i \beta_i InP_i + \sum_k \beta_k InK_k + \sum_z \beta_z InZ_z + 0.5\sum_q\sum_q \beta_{qq} I_nQ_a I_nQ_b + 0.5\sum_p\sum_p \beta_{pp} I_nP_i I_nP_j + 0.5\beta_{kk} In(K_k)^2 \\ & +0.5\sum_z\sum_z \beta_{zz} I_nZ_1 I_nZ_2 + \sum_q\sum_k \beta_{qk} Q_a I_nK_k + \sum_q\sum_z \beta_{qz} I_nQ_a I_nZ_z + \sum_i\sum_k \beta_{ik} I_nP_i I_nK_k + \sum_p\sum_z \beta_{pz} I_nP_i I_nZ_z + \sum_k\sum_z \beta_{kz} I_nK_k I_nZ_z \end{aligned} \qquad \cdots\cdots (3.4)$$

Where:

C_i is variable costs

i is a firm index

Q is an index for the various output characteristics

P is an index of various input prices

K represents the quasi-fixed capital

Z is an index of the environmental/network characteristics

The translog is preferred, but loss of degrees of freedom due to a large number of parameters to be estimated can be problematic if the observations are insufficient. Some literature proposes estimation of the system of equations involving the TCF and input shares to improve efficiency of the output and factor input parameter estimates. However, there are contradicting views. Filippini and Farsi (2005) suggest that allocative inefficiencies, which enter the demand function, can complicate the error system and thus over or underestimate the efficiency term. The following parameters from equation 3.4 representing the TCF will be estimated to obtain statistical estimates for the impact of the explanatory variables on costs $\beta_o + \beta_a + \beta_i + \beta_k + \beta_z + \beta_{qq} + \beta_{pp} + \beta_{kk} + \beta_{zz} + \beta_{qk} + \beta_{qz} + \beta_{ik} + \beta_{pz} and \beta_{kz}$. The deviation from the estimated frontier will be interpreted as the cost inefficiencies of individual WSPs (Aikaeli, 2002).

For the cost function to be used meaningfully for empirical inference, it must meet various regularity conditions to conform to the theoretical U-shaped average cost curve of a cost minimizing firm. First, the cost function must be homogenous of degree one in input prices so that a proportional increase in all input prices shifts cost by the same amount when output is held constant. Homogeneity is satisfied if $\sum_p \beta_{ij} = 1, \sum_q \beta_q = 0, \sum_k \beta_k = 0, \sum_z \beta_z = 0$ and $\sum \varepsilon_i = 0$

and it will be imposed by normalizing the costs and input prices by the price of labour (by dividing input prices and the cost by the price of labour). In addition, the hessian of the cost function must also be symmetric in input prices, and this may be ensured by symmetry imposed *a priori* by Young's theory, which dictates that $\beta_{ab} = \beta_{ba} and \beta_{ij} = \beta_{ji}$.

The cost efficiency scores will be estimated from the non-negative component of the error term u_{it}, calculated as:

$$Eff_{it} = \left(\frac{C_{it}}{C_{it}^F} \right) = \exp(-u_{it}) \quad \text{.........(3.5)}$$

where C_{it} is the observed total cost and C_{it}^F is the frontier cost of the i-th WSP in time *t*. A score of one indicates an efficient WSP, which operates on the frontier, but any score above one indicates inefficiency.

The output cost elasticity will provide an estimate of the economies of increasing output levels

$$EOD = \frac{1}{\partial \ln c / \partial \ln Q} \quad \text{.........(3.6)}$$

Similarly, the elasticity of increasing the percentage of the population covered to output can indicate economies of customer density.

$$ECD = \frac{1}{\frac{\partial \ln C}{\partial \ln C} + \frac{\partial \ln C}{\partial 1WC}} \quad \text{.........(3.7)}$$

The degree of economies of scale will be obtained by the following expression also defined by Fillipini *et al.* (2008):

$$ECD = \frac{1}{\frac{\partial \ln C}{\partial \ln C} + \frac{\partial \ln C}{\partial 1WC} + \frac{\partial \ln C}{\partial \ln Z}} \quad \text{.........(3.8)}$$

$E_S>1$ implying that economies of scale exist and it would be beneficial to increase the size of service area. Economies of scale measure the reaction of production costs as outputs increase or decrease, and is thus a useful indicator for the optimal size of utility.

3.4 Data Description

The WSPs operating in Kenya provide water and sanitation services to customers in distinct urban or rural[1] areas. Focus is limited to water production as the kind of data maintained by WASREB in their water regulation information system (WARIS) since 2005/06 only records the overall coverage rate for sewerage/ sanitation information. The WSPs also vary in terms of size defined by the number of connections to network and are differentiated mainly by their network characteristics, which form their respective operating environment, such as population in service area, staff productivity (staff per 1000 connections) and system water losses (ufw).

The WARIS-based data used for this study is the same data used to generate the IMPACT report, which publishes WSPs' performance in selected nine key sector indicators. Due to limitations in accessing the information system, a big portion of the data used in this study has actually been extracted from the published IMPACT reports, with necessary transformations and conversions to fit model requirements. Additional information on selected variables was obtained from WASREB through face to face interviews, thus more interrogation of the WARIS.

Even though the information system has been in place for about six years, data for the first three years contains numerous gaps due to infrequent recording of observations. The key variables captured also changed in the latter three years to include general data that were actually found to be more precise and useful for the current study. These included figures for population in the service area and actual number of people served, volume of water produced, as well as actual number of connections and the number of staff in each WSP. Due to this mismatch, collation of initial data resulted in a balanced panel of 72 WSPs, with observations for only three years of 2008/09, 2009/10 and 2010/11.

The cost variable (C) is used in reference to the sum of O&M costs, including the WSPs' expenditure on labour, electricity, chemicals and materials and also the miscellaneous costs for administration per year. The key output is the volume of water produced in cubic meters and since WSPs attain varying percentages of treatment and water quality, the volume of treated water by each utility was also

---

[1] Described according to where a WSP derives the bigger share of revenue from WSS services

identified as an output characteristic in addition to metering rate and coverage of both water and sanitation services. The prices of materials and energy are obtained as the amount each WSP spends in chemicals and electricity, respectively, per one unit of water produced. However, since different WSPs have different hours of supply per day, the price of energy was weighted by the respective average hours per day to give the prorated cost of electricity to each WSP. Miscellaneous expenditures were assumed to constitute the price paid for administration in order to obtain production of one unit of water. The price of labour is calculated as the personnel expenditure per staff of WSP. The capital stock, which is quasi-fixed (Bottasso and Conti, 2003) is proxied by the number of connections. Environmental characteristics considered to affect the costs of a WSP include the network population density, calculated as population in service area divided by number of connections size of utility, the existing staff productivity rate, as well as ufw, which reduces the output (Horn and Saito, 2011).

Table 3.2: Variables description

Type	Variable	Label	Expected sign
Goal	O&M Expenditure (variable cost)	C	Dependent Variable
Input	Capital = Capital Expenditure (proxy by no. connections)	K	-
Input	Price of Labour = staff expenditure/ numbers of staff	P_L	+
Input	Price of materials: P_M=C-P_L	P_M	+
Input	Price of energy	P_E	+
Input	Price of administration	P_A	+
Output	Water Produced in M_3	W	+
Output characteristics	Water treated in M_3	T_W	+
Output characteristics	Metering rate	M_R	-
Output characteristics	Water coverage	W_C	+
Output characteristics	Sewerage coverage	S_C	+
Exogenous	Staff productivity	S_P	-
Exogenous	System water losses (ufw)	L_O	+
Exogenous	Population density	P_D	-

3.5 Estimation

The cost efficiency model is based on the stochastic frontier analysis as applied by Battesse and Coelli (1995). The translog is a flexible functional form, which provides a second order approximation to the unknown cost function. Since it is a generalization of the Cobb-Douglas function and it is linear in parameters, it can be estimated using least squares method with homogeneity conditions imposed as restrictions on the parameters.

Focus of the SFA is not to estimate the cost function but rather the inefficiency component represented in the error term. Equation 10 is expressed with a decomposed error term ε_i (Correia and Marques, 2011):

$$\ln C_{it} = \ln C(Y_{it}W_{it};\beta) + (v_{it} + u_i)\text{...}(3.9)$$

Where i=1,2,......N; t=1,2,......T and $(v_{it}+u_{it}) = \varepsilon_{it}$

$u_{it} \geq 0$ is independently distributed of v_{it} and the other regressors and represents the non-negative cost inefficiency term with iid $N(\mu,\sigma^2 v)$.

v_{it}- Symmetric statistical noise and is assumed to be iid $N(\mu,\sigma^2 v)$.

The term $C(.)$ is the functional form in equation 3.9; $C(Y_{it}, W_{it})$ is the stochastic cost frontier curve, where Y_{it} and W_{it} represent the vector of output and input prices, respectively, plus other exogenous factors and β is the vector of unknown parameters to be estimated.

Borrowing from Aigner *et al.* (1977), the inefficiency effects $u_{it}s$ are assumed to be a function of a set of specific regressors that affect efficiency (Z_{it}) and the respective unknown coefficients (δ). The cost inefficiency is therefore expressed as:

$$u_{it} = Z_{it}\delta + W_{it}\text{..}(3.10)$$

This follows specifications introduced by Aigner *et al.* (1977) where W_{it} is a random variable which makes the inefficiency stochastic, defined by truncation of the normal distribution at $-Z_{it}\delta$ such that $W_{it} \geq -Z_{it}\delta$ (Fraquelli and Moiso, 2005).

The likelihood function proposed is therefore in terms of the variances where $\sigma^2=\sigma_v^2+\sigma_u^2$ and $\gamma=\sigma_u^2/\sigma^2$, therefore cost inefficiency *(CE) = exp* $(-u_{it})$.

4. Empirical Findings

4.1 Descriptive Statistics

The descriptive statistics in Table 4.1 summarize the key characteristics of the variables used to assess the cost structure of WSPs in Kenya.

Table 4.1 shows that the standard deviations of C,W,P_M,P_A,NC,PD and SP are larger than the mean, indicating that they are widely spread out from their mean values. As anticipated at the model selection stage, the effect of outliers cannot be ignored in the analysis. The actual status of cost elements (as per available data) is also shown in Appendix Table 2. The large WSPs generally have higher water production, serve more people, manage longer hours of supply per day and have consistently retained operating surpluses throughout the three years. The smallest WSPs, on the other hand, have performed poorly, in comparison, and have been running operating deficits over the years, implying that they are unsustainable without financial subsidies.

Table 4.1: Summary statistics

Variable	Units	Abbr.	Mean	Std. Dev.	Min	Max
Operating cost	Ksh (M)	C	106	325	5.25	2,720
Water production	M3 (000)	W	5,358	19,300	92	168,000
Water treatment level	%	TL	82	15	19	100
Water coverage (population served)	%	WC	42	21	5	95
Sanitation Coverage	%	SC	67	30	0	100
Metering rate	%	MR	71	27	0	100
Price of labour (personnel expenditure/ No. staff)	Kshs	P_L	321,103	198,953	42,653	1,057,168
Unit price of materials (Expenditure on energy+ chemicals)/W	Kshs	P_M	12.3	16.4	0.03	110
Unit price of administration (Expenditure on other expenses)/W	Kshs	P_A	8.83	8.87	0.08	64.11
Capital stock (Number of connection)	Numbers	NC	17,167	50,736	162	415,229
Population density (Population in service area/number of connections)	Ratio	PD	28.06	32.14	2.71	278.6
UFW (Network water losses)	%	LO	51	16	0	98
Number of staff	Numbers	S	105	254	5	2,112
Staff productivity (# staff/1000connections)	Ratio	SP	9.16	9.22	2.27	98.77

4.2 Empirical Results

4.2.1 Impact of cost elements

Correlation analysis shows that expansion of sewerage coverage (SC), high populations per connection (PD), high number of staff per 1,000 connections (SP) and unaccounted for water (LO), negatively affect operating costs of WSPs. However, only LO is statistically significant for r (187) = 0.13, $p < 0.05$.

As the differences across WSPs are expected to influence their operating costs, the inefficiency term is estimated and compared for both the Cob Douglas (C-D) and translog models. The Hausman test was not conclusive as the variance across panels was inconsistent. Thus, to check the model's suitability, the Wald test was applied. This test checks the linear hypotheses and joint significance of all the independent variables. The null that parameters of the regressors are zero was rejected with a p-value of 0.00 for eleven degrees of freedom. Therefore, all the selected regressors actually have an influence on costs of WSPs.

Table 4.2 shows that the estimated C-D function was well fit with a positive log likelihood ratio and chi-square p=0.00. The coefficients showed that the volume of water produced (W), price of administration (P_A) and the number of connections (NC) as well as the staff per 1,000 connections (SP) significantly affect the operational costs. For instance, when the water produced increases by one unit, that is one cubic meter, the costs are likely to increase by about 31 per cent. The effect is similar for the PA, NC and SP where costs may rise by 32 per cent, 66 per cent and 64 per cent, respectively. On the other hand, and even though the negative parameters for SC and MR were not statistically significant, WSPs could marginally reduce operating costs by increasing sewerage coverage and metering rate.

To meet theoretic requirements, regularity conditions which require that a cost function is homogeneous and non-increasing in input prices had to be imposed. First, the functional form where all variables were expressed in terms of their natural logarithms ensured that the costs are positive, since no output can be produced with zero or negative costs. Incidentally, these coefficients of the cost function can also be read as cost elasticities to provide a measure of the effect of various factors on costs. Secondly, homogeneity in input prices was imposed by normalizing the operating cost and prices of materials and administration by the numeraire (in Filippini and Farsi, 2005) price of labour.

From the TCF estimation, all coefficients were jointly significantly different from zero, with a positive log likelihood ratio and chi-square p=0.00. The parameters that were independently significant at 95 per cent confidence level for both the C-D and translog models are as shown in Table 4.2. For the insignificant variables, only (+) and (-) signs are indicated to show the direction of the variable relationship with the independent variable (operating costs).

The default for the stochastic frontier model in (Stata software) applies ML as the optimization method and assumes that the white noise is *iid* $N(0,\sigma^{v}{}_{2})$ and the inefficiency term is asymmetrical *iid* $N(\mu,\sigma^{u}{}_{2})$. The hypothesis of no cost inefficiency in the model as indicated by the Log likelihood of 32.7(11) in the C-D and 77.9 (53) in the translog model, both with a chi-square p = 0.0000, is rejected.

Table 4.2: Estimated parameter coefficients

Variable	**Parameter**	**C-D**	**Translog**
Vol. water produced (W)	β_{w}	0.3151125 (0.0573055)***	-
Metering rate (MR)	β_{mr}	-	-
Water treatment level (TL)	β_{tl}	+	-
Water coverage (WC)	β_{wc}	+	-
Sewerage coverage (SC)	β_{sc}	-	-
Price of materials (PM)	β_{pm}	+	0.2414959 (0.1443631)*
Price of administration (PA)	β_{pa}	0.3243255 (0.0247298)***	0.8169599 (0.2455433)***
Number of connections (NC)	β_{nc}	0.6670716 (0.0666679)***	+
Population density (PD)	β_{pd}	+	-
Staff productivity (SP)	β_{sp}	0.645421 (0.066474)***	+
Unaccounted for water (LO)	β_{lo}	+	-
W*TL	β_{wtl}		-0.5295244 (0.2198646)**
TL*TL	β_{tl2}		0.3469499 (0.1922813)*
TL*NC	β_{tlnc}		0.8864784 (0.2696274)***
TL*PD	β_{tlpd}		0.4707086 (0.1814462)***
TL*SP	β_{tlsp}		0.5023521 (0.2449532)**
SC*SC	β_{sc2}		-0.0654182 (0.0323435)*
SC*NC	β_{scnc}		0.1716992 (0.101626)*
LO*LO	β_{lo2}		-0.139121 (0.0706681)**
	const	-3.675635 (0 .526322)***	+
Degrees of freedom		11	53
Model power	*Chi²*	p = 0.00	p = 0.00

Notes: Standard errors in brackets;

***statistically significant at 1%, ** significance at 5%, * significant at 10% for the 95% confidence interval

The first differential coefficients of the prices of administration (PA) and materials (PM) are positive as required by theory and also statistically significant at p=0.00 and p=1. This indicates that, on average, a one percent increase in the cost of administration services may lead to over 81 per cent increase in the costs of running a WSP. It is noted that the second differentials, however, remain positive and thus concavity of the cost function in input prices is not satisfied.

The first differentials for two outputs (sewerage coverage and metering rate) have unexpected negative signs, whereas all the other output variables analyzed have the expected negative second differential coefficients albeit insignificant. There is a slight change when their cross products are considered. The square of treatment level (β_{tl}^2) is positive and significant, but its product with volume of water produced (β_{wtl}) has a statistically significant negative coefficient. Treating water is a costly business and, therefore, to garner returns, the WSP increases treatment needs to raise the volume produced so as to provide more clients. This can be reinforced by further findings on interactions of water treatment level and network characteristics, where β_{tlnc}, β_{tlpd} and β_{tlsp} are also positive and statistically significant. Increasing the treatment level while also increasing the size of a WSP by adding number of connections will proportionately increase the operating costs by more than 88 per cent. Similarly, if the number of people per connection is also rising and the staff managing every one thousand connections increases, then the costs is bound to increase by 47 per cent and 50 per cent respectively, for the WSP.

The coefficients of the capital proxy (*NC*), population density (*PD*), staff productivity *SP* and water losses (*LO*) are not independently statistically significant. The square terms β_{sc}^2 and β_{lo}^2 for sewerage coverage and water losses, respectively, are negative and statistically significant at the 5 per cent level.

4.2.2 Cost efficiency (CE)

The CE is estimated using equation 3.5, and the scores calculated by size of WSP are indicated in Table 4.3.

From Table 4.3, when the respective outputs are considered alongside costs of inputs and network characteristics, the average cost efficiency attained in the various categories of the WSPs is estimated at only 1.2 and 1.1 per cent under the C-D and the translog models, respectively.

4.2.3 Returns to scale

Overall, scale economies are a measure of total cost elasticity as the DMU simultaneously expands its outputs (EOD), customers (ECD) and scale of

Table 4.3: The estimated cost efficiency scores

Cost efficiency(CE) at 95% confidence interval		
WSP by size	C-D	translog
Small WSP	0.0140518 (0.0019816)	0.0121305 (0.0014102)
Medium	0.0124994 (0.0013883)	0.0119527 (0.0013419)
Large	0.0108726 (0.0010958)	0.0100528 (0.0009803)
Very Large	0.0082063 (0.0010856)	0.0081603 (0.0009558)
Average	0.011407525	0.01057408

operations (ES) and these are estimated according to equations 3.6, 3.7 and 3.8, respectively. Table 4.4 presents the results that have been calculated according to the size of WSPs as per the existing categorizations by number of connections.

Table 4.4 shows that the measure of EOD for the very large WSPs is actually a diseconomy at -6.611, positive ECD of 1.7 and high negative EOS at -23.7. The large and medium sized utilities have positive EOD, ECD and EOS, while the small category of WSPs has positive economies only in ECD but negative EOD and EOS.

All the four categories of WSPs exhibit positive ECD. Thus, by increasing their customer base in 2010 by 1 per cent, they actually could reduce their operational costs more than proportionately. Since none of the WSPs had a 100 per cent coverage rate within their service areas, they would all benefit from serving more customers without altering their output and with the existing conditions of network and input costs.

4.3 Discussion of Results

The model is estimated using both a C-D and translog formulation in which the RE model was extended to include time-variant inefficiency in Battese and Coelli (1995), with efficiency varying across firms. The efficiency of the model would have been improved by estimation of a system of equations, constituting the TCF and share equations in two out of three inputs (administration and materials)

Table 4.4: Returns to scale

ESC	EOD			ECD			EOS		
	2008	2009	2010	2008	2009	2010	2008	2009	2010
Very large	-2.08333	.	-6.6111	-0.81548	-2.00287	1.696972	0.982457	-0.90968	-23.7104
Large	-3.66905	-3.20992	5.93518	-0.39963	0.310171	4.965269	0.36231	0.898758	8.669655
Medium	9.1143	0.969579	23.33669	0.37098	32.01074	16055.7	-1.59623	10.89987	7.804787
Small	2.228566	15.68753	-8.79961	-7.55752	-4.1049	4.772299	-6.39967	-1.37051	-21.0289

EOD- economies of density; ECD-economies of customer density; EOS- economies of scale to avoid singularity problems. The model, however, failed to converge and this could be due to lack of complete data on input prices, which could have caused this inconsistency.

Various attempts to estimate a TCF, including the entire cross effects of outputs, inputs and fixed factors, resulted in unexpected signs of the parameter estimates, particularly the factor prices of labour and administration. This could be attributed to the multicollinearity of operating cost with outputs and environmental variables. Thus, in order to reduce the parameters, the cross products between input prices and the outputs were dropped and a homothetic cost function considered instead (Filippini and Farsi, 2005). This is supported by the assumption that economic returns are actually influenced by factor prices only through the fixed costs of production technology.

4.3.1 Factors that affect the costs of WSPs

Conceptualizing the operating conditions of the WSPs as localized for each unit leads to the assumption that various unobserved variables affect their cost efficiency. These represent heterogeneity, because the unmeasured effects may be different between WSPs or similar but change or remain unchanging over time. Since the WSPs also exhibit individual variations in output characteristics such as treatment level of water produced, metering rate and unaccounted for water, these variables are included as independent regressors in the cost function to differentiate them.

Theory proposes that a cost function is non-decreasing in outputs, output characteristics and input prices. In this analysis, the condition is only met for two outputs but is not consistent for metering rate and sewerage coverage, as well as the prices of inputs whose second derivatives are negative. This may be because these outputs are not stand alone, but are simply characteristics complementing the volume of water delivered, and their increases imply value addition which may require more expenditures to achieve. Higher metering rate in the context of water utilities has potential gains realized from reduced water losses, increased revenue generation and increased accountability for produced water. Sewerage coverage also requires high investment for its specialized infrastructure and may also necessitate hiring of additional staff. However, expansion of SC may not cause the costs to increase due to economies of scope (not examined in this study), which may arise from sharing of some overheads as well.

Therefore, even though the cost function is found to be only partly concave in input prices, cost functions are not known to satisfy all required properties except under cost minimizing behaviour. Therefore, according to Filippini *et al.*

(2008), in the presence of model inefficiencies, a cost function may not accurately represent the economic reality of a given production possibility set.

The results indicate that the cost of administration services leads to large increases in the costs of running a WSP, suggesting that it is an important area of concern for WSPs. A similar issue was also flagged by the regulator during their annual performance evaluations, where they found that WSPs' expenditure on administration issues including board meetings was more than proportionately over-budget and had to be reduced (WASREB, 2010).

The interactions between water treatment level with the number of connections, population density and even staff productivity implies that treating more water impacts on efficiency of WSP operations. Benefits may arise to the utility, including but not limited to reducing wastage or misuse, improving revenues or even preserving the infrastructure. It is also a costly business and, therefore to garner returns, the WSP that increases treatment needs to also increase the volume produced to provide more clients. These variables, therefore, portend useful considerations when mergers are considered to avoid increasing operating costs unnecessarily. For instance, when population per connection increases as treatment level is increased, the WSP does not save on costs because these require more complex connections (Horn and Saito, 2011) or may be characterized by frequent bursts and prevalent illegal connections, which add to costs.

The increase in costs due to increasing ufw implies that there is an opportunity for WSPs to reduce their operating costs by reducing the level of water lost from their systems. Other studies have indicated that ufw is significant in efficiency (Fraquelli and Moiso, 2005; Martins *et al.*, 2012) and effectiveness of water utilities and this study corroborates those findings. Sewerage coverage also increases the operating costs as its infrastructure is elaborate and requires high capital as well as maintenance, which are expected to increase the spending budget for WSPs. However, as the current research does not establish economies of scope, this area could be explored to show how a utility may benefit from adding sewerage to its water business without changing some of the overhead costs.

4.3.2 Cost efficiency scores

The cost efficiency scores presented in Table 4.3 show that WSPs operate with cost structures that are far flung from the cost frontier, given their expenditure on production and supply of water and sanitation services. From the results obtained, it is clear that WSPs are not cost efficient, thus this is not a distinguished criteria to evaluate or rank them. This could be because their business involves the provision of a human right that cannot be measured by the amount of costs incurred, but are measured instead by the impact of their outputs.

Also, it is possible that the results were influenced by the kind of data used, as prices of inputs were not available, but were derived from expenditures of WSPs. In addition, important variables that determine the operating cost, such as management structure, including ownership (Estache & Kouassi, 2002; Triebs *et al.*, 2012), nature of provider (whether private or public) and level of capital stock (Stone & Webster, 2004) or investments were noticeably missing from the analysis. Furthermore, researchers in the topic of inefficiency agree that estimation of inefficiency scores using stochastic frontier analyses depend on the different assumptions regarding cost inefficiency under the different models used. Therefore, their interpretation and use remain debatable, since time invariant effects resulting from unobserved and therefore un-captured heterogeneity and or fixed characteristics may lead to underestimation or overestimation, respectively (Filippini *et al.*, 2008).

4.3.3 Economic returns

For the three years considered, very large utilities have experienced no economies of output density (EOD). It appears their output level has reached its maximum given the existing structures. Increasing the amounts of water currently produced without further expanding infrastructure can only represent losses. They also have very negative economics of scale (EOS), meaning the size of the networks has already gone beyond the optimal point, and additions would not realize positive returns.

The large firms seem to be on a growth path as their EOD turned from negative in the first two years to positive in 2010. This can be attributed to improved capacity (as connections and or meters among other infrastructure increase) and technology or due to increased customer numbers. In comparison, the medium firms have consistently benefited from positive EOD, ECD as well as EOS in the last two years. Thus, they can still further increase their level of outputs, serve more customers, and expand the necessary infrastructure and still realize positive returns on investments.

Finally, the small WSPs that show increasingly strong diseconomies of scale should not attempt to increase their size. Instead, they could increase the customers served using existing infrastructure and output. This can be done by improving accountability of the water produced and network surveillance to reduce ufw as they have already overextended their production capacity. Increasing the volume of water produced could be causing their operational costs to sky rocket, hence aggravate their sustainability.

5. Conclusion and Policy Recommendations

5.1 Conclusions

This study attempted to analyze the cost structure of water utilities in Kenya. The key objectives were to identify cost elements with the greatest impact on operating costs of WSPs, estimate their cost efficiency and suggest an optimal scale of operation for WSPs as well as the extent to which small WSPs may be clustered. The key inspiration is the need to justify the recommendations from performance evaluations by WASREB to cluster small 'inefficient' WSPs so as to improve their sustainability and reduce their numbers for better management.

Focus was on the registered WSPs for which data was available in the WARIS system for the years 2005 to 2010. The scope of the study was, however, constrained by lack of consistent data for the first three years for all WSPs and even in the latter three years for some variables selected for analyses. In spite of this, a representative sample of 72 WSPs out of the 100 evaluated in 2010/11 were analyzed, and the results are expected to represent the actual sector situation.

Based on the data analyzed, the key factors that affect the running costs and the cost efficiency of WSPs in this study are therefore the volume of water produced and treatment level, the prices of inputs (materials and administration), the number of connections, unaccounted for water, the level of staff productivity, and the population density per connection. The population density per connection may be greatly influenced by the nature of physical planning, where compact cities or planned settlements would greatly reduce the need for extensive networks. More customers would be served in an area using the same basic infrastructure, increasing coverage, saving costs and thus improving efficiency. These findings also present a challenge for WSPs to improve treatment levels of produced water, increase connections and reduce the number of staff per connection.

Owing to the contributing impacts of these elements, overall cost efficiency of WSPs operating in Kenya is extremely low at around 1.1 per cent. None out of the four categories of small, medium, large and very large WSPs is efficient. In fact, from this analysis, the very large WSPs that are efficient and have the lowest CE scores at 0.8 per cent. This implies that efficiency cannot be based merely on size without considering the interaction of the other key variables or consideration of other overriding objectives being met by the WSPs.

Finally, in terms of the overall economies of scale of operation, the very large and small categories show substantially negative levels. This may imply that small WSPs have no capacity to increase their size (number of connections) given their current conditions. On the other hand, it appears that the very large utilities have already exceeded their minimum efficient scale and cannot expand further and attain positive returns. The large and medium categories exhibit the greatest potential for economic returns, implying that they have sufficient room for extra production, more customers and increased capacity.

The large and medium WSPs spend, on average, between 40 million and 80 million shillings per year in operating costs. They produce an average of between 2-4 million cubic meters of water and treat about 73-86 per cent of the produced water; serve up to 40 per cent of their targeted population with about 5 per cent connection rate. When compared to the very large WSPs, which do not show any evidence of CE or economies of output, density and scale, their service is not better in per capita terms. The large WSPs that spend on average at least one billion Kenya shillings treat 90 per cent of the water they produce, serve 70 per cent of their targeted population, and manage more than 10 per cent of network connections.

Since this research has shown that all categories of WSPs are inefficient and positive economic returns are only realizable with the medium to large WSPs, then the clustering solution proposed by WASREB could become handy, provided the optimal size is not exceeded. Otherwise, the realized economies may disappear immediately the scale of operation gets to a very large size. For instance, in Kenya, single WSPs cannot effectively and efficiently serve customers numbering above the average of 800,000 people currently being served by the very large WSPs.

5.2 Policy Recommendations

This study provides the following recommendations to guide the structuring of WSPs in order to attain efficiency in operations, hence effectiveness in services provision.

First, WSPs should incorporate geographical information systems (GIS) where platforms for documentation of network elements can be availed to create spatially enabled databases for planning, monitoring and real-time identification of network problems. This would also enhance the availability of regular data on important indicators for the sector, which have been unavailable due to lack of documentation. GIS would also assist in efforts to reduce unaccounted for water by enabling more accurate surveillance as well as timely maintenance and rehabilitation of infrastructure.

Secondly, clustering as a strategy to improve efficiency and effective service delivery should be adopted, albeit with caution. This research suggests that the very large WSPs must not have their size increased, but they may be maintained as they still have potential to serve more customers with the current capacity. It has also been established that they have been performing well in the key sector indicators, achieving high service coverage and treatment levels. The small WSPs, on the other hand, may be merged to form only medium or large utilities since production at their existing capacity may not be economical. It is, however, important to note that the mere size of a utility is not related to its efficiency as research has shown that CE does not improve nor worsen with the utilities' size.

The third recommendation is for the enhancement of planned settlements and designated network routes for utility services provision. The county governments now have an opportunity to make amends as they implement their development and spatial plans in which they may incorporate adequate provisions for the establishment of utility service networks. In addition, each county needs to prioritize and undertake research to establish cost efficiency in order to effectively manage their respective WSPs and improve WSS supply conditions.

Finally, the water sector players must populate and avail consistent data to allow rigorous investigations in the performance of the sector institutions. It is noted that this research did not incorporate important information on the type of consumers served, which would indicate the importance of treatment; the size and morphology of the service area, which can determine the different costs of energy for pumping; and the water sources whether underground or surface, which would determine the required investment in treatment chemicals and wear of the infrastructure. Data on the soft infrastructure was also unavailable for management structures such as qualification of staff, ownership (whether private or public) and nature of business.

5.3 Areas for Further Research

This research could be furthered by using a cost benefit analysis approach to establish the size of WSPs that is most appropriate for each area. This would help incorporate considerations of the human right to water as well as technical challenges in each region in terms of availability of water sources and socio economic qualities.

After establishing that the price of administration significantly increases the operating costs of WSPs, an indepth investigation is required to identify the individual components. These could include managerial capacity and their significance to the operations of WSPs in order to suggest necessary policy and management decisions.

The research also points to the need to investigate the causes of system water losses, including illegal connections and lack of accountability. Despite high expenditures on rehabilitation and maintenance of water system networks to reduce the ufw, the problem has persisted, mainly due to old and dilapidated infrastructure (WASREB, 2010).

In conclusion, it is observed that data was a key challenge for this analysis. This posits an urgent need to identify and establish which indicators are critical for the sector and use them to launch a universal database that can facilitate required investigations aimed at improving the sector's performance. The exercise should also suggest the format of such a database, and the most appropriate custodians to host them for the sector.

References

Abrate, G., Erbetta, F. and Fraquelli, G (2008), "Cost Inefficiency or Just Heterogeneity? An Application of Stochastic Frontier Models to the Italian Water Industry," HERMES Working Paper 3.

Aigner D., Lovell K.C.A. and Schmidt, P.G. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, 6, pp. 21-37.

Aikaeli, J. (2002), "Commercial Banks Efficiency in Tanzania", AERC's CPP Graduates, Class of 2002, A Selection of Research Papers Drawn from Doctoral Theses.

Antonioli B. and Filippini, M. (2001), "The Use of a Variable Cost Function in the Regulation of the Italian Water Industry", *Decanato della Facoltà di Scienze Economiche*.

Ashton J. (1999), "Economies of Scale, Economies of Capital Utilization and Capital Utilization in the English and Welsh Water Industry", Bournemouth University School of Finance and Law Working Paper Series No. 17.

Baranzini A. and Faust A. (2009), "The Cost Structure of Water Utilities in Switzerland", Hes.So Geneve, Working Paper.

Battese, G.E. and Coelli T. (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Econometrics*, 20: 325-332.

Baumol W. J. (1977), "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry", *American Economic Review*, American Economic Association, Vol. 67(5), pp 809-22.

Botasso, A. and Conti, M. (2003), "Cost Inefficiency in the England and Welsh Water Industry: A Heteroskedastic Stochastic Cost Frontier Approach," Economic Discussion Papers, University of Essex Department of Economics.

Brown, T. L. (2005), "Evaluating Geographic Programme Performance Analysis: Weighing the Benefits and Costs of Geographically Disaggregating Performance Data", *Public Performance and Management Review*, Vol. 29, No. 2, pp. 164-190, Dec.

Coelli, T., Rao D. S. and Battese G. E. (2005), "An Introduction to Efficiency and Productivity Analysis", 2nd edition, Springer Verlag.

Correia, T. and Marques, R.C. (2011), "Performance of Portuguese Water Utilities: How Do Ownership, Size, Diversification and Vertical Integration Relate to Efficiency?", Water Policy 13 343–361, IWA Publishing.

EPA (Online), "Effective Water Utility Management", Available at www.http://water.epa.gov/infrastructure/sustain/watereum.cfm, Accessed on 10th September, 2012.

Estache, A. and Kouassi, E. (2002), "Sector Organization, Governance, and the Influence of African Water Utilities", World Bank Policy Research Working Paper 2890.

Filippini M., Hrovatin N. and Zoric´ J. (2008), "Cost Efficiency and Economies of Scale of Slovenian Water Distribution Utilities: An Application of Stochastic Frontier Methods", *J Prod Anal* 29:169-182.

Filippini M. and Farsi M. (2008), Cost Efficiency and Scope Economies in Multi-Output Utilities in Switzerland", Federal Department of Economic Affairs, Effingerstrasse 31, 3003, Berne.

Filippini M., Farsi M. (2005), "Cost Efficiency of Slovenian Water Distribution Utilities: An Application of Stochastic Frontier Methods", *J Prod Anal* 29:169-182.

Farsi M., Filippini M. and Kuenzle M. (2003), "Unobserved Heterogeneity in Stochastic Cost Frontier Models: a Comparative Analysis", Quaderno N. 03-11.

Follmi, R and Miester, U. (2011), "Enhancing the Efficiency of Water Supply: Product Market Competition versus Trade", Discussion Paper No. 2011-21, School of Economics and Political Science, Department of Economics, University of St. Gallen.

Fraquelli G and Moiso V. (2005), "Cost Efficiency and Economies of Scale in the Italian Water Industry", *Pubblicazione Internet Realizzata Con Contributo Della Società Italiana*.

Government of Kenya (2006) and Ministry of Water and Irrigation (2006) in Collaboration with Ministry of Planning and National Development, "MDGs Needs Assessment Report; Requirements for Goal No. 7, Target 10: Providing Sustainable Water and Sanitation Services.

Government of Kenya (2007) and Ministry of Water and Irrigation, "The National Water Resources Management Strategy" (NWRMS) 2007-2009, First Edition-January.

Government of Kenya (2007), "The National Water Services Strategy (NWSS): 2007-2015", Ministry of Water and Irrigation.

Government of Kenya (2010), "Water Sector Reform Secretariat, Information Brief", Communication Issue 4–Role of NGOs: Ministry of Water and Irrigation.

Guerrini, A., Romano, G. and Campedelli, B. (2011), "Factors Affecting the Performance of Water utility Companies", *Intl Jnl of Public Sector Management*, Vol. 24 Issue: 6 pp. 543-566.

Heaney J, P. (2009), *Economic Efficiency and Equity Aspects of Urban Water Systems*, Presentation at Grand Challenges Conference, Association of Environmental Engineering Professors, Iowa City, IA.

Horn, T. and Saito, H. (2011), "Cost Efficiency and Scale Economies of Japanese Water Utilities", *International Conference on Applied Economics*– ICOAE

Kim, Y. (1987), "Economies of Scale in Multi-Product Firms: An Empirical Analysis", *Economica,* 54, 185-206.

Madrigal, R. Alpízar, F. and Schlüter, A. (2010), "Determinants of Performance of Drinking-Water Community Organizations: A Comparative Analysis of Case Studies in Rural Costa Rica", Environment for Development Discussion Paper Series.

Martins R., Coelho F. and Fortunato A. (2012), "Water Losses and Hydrographical Regions Influence on the Cost Structure of the Portuguese Water Industry", *J Prod Anal* 38:81–94.

Ministry of Water and Irrigation (2008), 'Implementation Plan of the National Water Services Strategy, Water Sector Reform Programme of the GTZ.

Murillo-Zamorano, L.R. (2004), "Economic Efficiency and Frontier Techniques," *Journal of Economic Surveys*, 18(1), 33-77

Nauges, C. and Berg, C. (2008), "Economies of Density, Scale and Scope in the Water Supply and Sewerage Sector: A Study of Four Developing and Transition Economies", *J Regul Econ* 34:144–163.

Panzar, J. C. and Willig, R. D. (1977) "Economies of Scale in Multi-Output Production", *Quarterly Journal of Economics*, 91, 481-494.

Romano, G. and Guerrini, A. (2011), "Measuring and Comparing the Efficiency of Water Utility Companies: A Data Envelopment Analysis Approach," *Utilities Policy*, Elsevier, Vol. 19(3), pp 202-209.

Shih, J.S., Harrington, W., Pizer, W.A. and Gillingham, K. (2004), "Economies of Scale and Technical Efficiency in Community Water Systems", Discussion Paper 04–15, Resources for the Future.

Shumilkina, E. (2010), "Utility Performance: How Can State Commissions Evaluate it Using Indexing, Econometrics and Data Envelopment Analysis?" National Regulatory Research Institute, available at at www.nrri.org, Accessed on 15 August 2012.

Stone and Webster (2004), *Investigation into Evidence for Economies of Scale in the Water and Sewerage Industry in England and Wales,* Final Report Stone and Webster Consultants, Ltd, January.

Sjodin, J. (2006), "Determinants of the Performance of Public Water Services in Chile 1977-1999," Internship Paper, Economic Commission for Latin America and the Caribbean (ECLAC) Santiago, Chile.

Triebs T P., Saal D S., Arocena P. and Kumbhakar S. C. (2012), "Estimating Economies of Scale and Scope with Flexible Technology", Ifo Working Paper No. 142.

Tsegai D.W., Linz T. and Kloos, J. (2009), "Economic Analysis of Water Supply Cost Structure in the Middle Olifants Sub-Basin of South Africa," Discussion Papers, University of Bonn, Centre for Development Research (ZEF).

Urakami, T. and Parker, D. (2009), "The Effects of Consolidation amongst Japanese Water Utilities: A Hedonic Cost Function Analysis", JEL Classification, L95, L11.

WASREB (2008), "IMPACT: A Performance Report of Kenya's Water Services Sub-Sector", Issue No. 1.

WASREB (2010), "IMPACT: A Performance Report of Kenya's Water Services Sub-Sector.", Issue No. 3.

WASREB (2011a), "IMPACT: A Performance Report of Kenya's Water Services Sub-Sector", Issue No. 4.

WASREB (2012), "IMPACT: A Performance Report of Kenya's Water Services Sub-Sector.", Issue No. 5.

WASREB (2011b), "Financing Urban Water Services in Kenya Utility Shadow Credit Rating", A Publication of the Water Services Regulatory Board in Collaboration with the Water and Sanitation Programme.

Water Operator Partnership (2009), Africa Utility Performance Assessment, WSP, UN-Habitat, ESAR, AIWA.

Worthington, A.C. and Higgs, H. (2011), “Economies of Scale and Scope in Australian Urban Water Utilities”, JEL Classifications, C21, D24, L95, Griffith Business School.

Von Hirschhausen C., Astrid C., Walter M., Wand R. and Zschille M. (2008), “Quo Vadis Efficiency Analysis of Water Distribution? A Comparative Literature Review”, Efficiency Analysis Working Papers, Utilities Policy, Dresden University of Technology (WP-EA-18).

Zschille, M. and Walter, M. (2011), “The Performance of German Water Utilities: A (Semi) -Parametric Analysis”, JEL, L95.

Zschille M. (2012), “Consolidating the Water Industry: An Analysis of the Potential Gains from Horizontal Integration in a Conditional Efficiency Framework”, Discussion Papers, DIW Berlin German Institute for Economic Research.

Appendix

Appendix Table 1: Water sector institutions and their responsibilities

SCOPE	INSTITUTION	ROLES AND RESPONSIBILITIES	ROLE
NATIONAL LEVEL	Ministry of Water and Irrigation (MWI)	Developing legislation, national policies and strategies Sector coordination and guidance Monitoring and evaluation of progress Planning and mobilization of sector finances	POLICY FORMULATION
	National Water Conservation and Pipeline Corporation (NWCPC)	Construction of dams and drilling of boreholes	POLICY, REGULATION AND SERVICE PROVISION
	National Irrigation Board (NIB)	Development of irrigation infrastructure	
	Kenya Water Institute (KEWI)	Training and research	
	Water Appeals Board (WAB)	Arbitration of water related disputes and conflicts	
	Water Services Trust Fund (WSTF)	Finance pro-poor investments	
NATIONAL LEVEL	Water Services Regulatory Board (WASREB)	Regulation and performance monitoring of Water Supply and Sanitation (WSS) services Licensing of Water Services Boards Setting standards for water services provision Development of sector guidelines, e.g. tariffs.	REGULATION
	Water Resources Management Authority (WRMA)	Planning and management of water resources Allocation and Monitoring of water resources Catchment protection and conservation Regulation and control of water usage	
REGIONAL LEVEL	Water Services Boards (WSBs) Athi Tanathi Tana Coast Northern Rift Valley Lake Victoria North Lake Victoria South	Provide efficient and economical WSS services Develop water facilities Plan and implement investments Rehabilitate and where possible replace infrastructure Enforce regulations on water services and tariffs Procure/lease water and sewerage facilities, where necessary contract WSPs	SERVICE PROVISION
	Catchment Area Advisory Committees (CAACs)	Advising WRMA on water resources issues at the catchment area level	
LOCAL LEVEL	Water Service Providers (WSPs)	Agents in the provision of WSS services Utilize acceptable business principles Assets maintenance Reach performance standards set by WASREB	
	Water Resources Users Associations (WRUAs)	Involved in decision making to identify and register water users Collaborate in water allocation and catchments management Assist in water monitoring and information gathering Conflict resolution and co-operative management of water resources	USE

Source: NWSS 2007-2015

Appendix Table 2: Descriptive characteristics of WSPs in Kenya by size from 2008-2010

	Very large			**Large**			**Medium**			**Small**		
	2008/09	2009/10	2010/11	2008/09	2009/10	2010/11	2008/09	2009/10	2010/11	2008/09	2009/10	2010/11
Number of utilities	4	4	4	19	21	23	20	20	20	28	28	25
Average volume of water produced (millions)	49	45.9	51.4	3.8	4.1	4.2	3.7	1.3	1.8	0.714	0.758	0.672
Average number of connections	139,243	143,933	142,692	15,965	16,183	16,977	7,380	7,064	7,546	2,279	2,282	2,269
Population served	877,321	887,806	998,626	112,844	128,993	125,975	90,635	69,888	80,249	20,280	24,237	25,109
Number of individuals per connection	12	11	11	27	24	22	77	46	42	43	47	71
Average duration of supply	14.75	14.25	14.50	17.53	17.95	17.65	14.61	14.35	13.05	13.08	11.86	12.50
Average operating deficit (Ksh millions)	429.9	814.7	2,147.8	(198.4)	158.1	504.75	(134.2)	(42.1)	104.3	(225.4)	(205.6)	(87.5)
Staff per 1000 connections	4	4	4	19	21	23	19	19	20	21	21	17

ISBN 9966 058 12 6

Kenya Institute for Public Policy Research and Analysis
Bishops Garden Towers, Bishops Road
PO Box 56445, Nairobi, Kenya
tel: +254 20 2719933/4, 2714714/5, 2721654, 2721110
fax: +254 20 2719951
email: admin@kippra.or.ke
website: http://www.kippra.org