Water utility efficiency and stated choice responses: status quo effects, effects of presentation format and response time

By

Genius Murwirapachena

Thesis in fulfilment of the requirements for the degree of:

Doctor of Philosophy (Economics)

In the

Faculty of Business and Economic Sciences

Department of Economics and Economic History

Nelson Mandela University

Supervisor: **Associate Professor Johane Dikgang**

15 March 2019

Declaration by candidate

Name: Genius Murwirapachena

Student number: 214069850

Qualification: Doctor of Philosophy (PhD) in Economics

Title of thesis: Water utility efficiency and stated choice responses: status quo effects, effects of presentation format and response time

Declaration:

In accordance with Rule G5.6.3, I hereby declare that the above-mentioned thesis is my own work and that it has not previously been submitted for assessment to another University or for another qualification.

Signature:

Date: 15 March 2019

Abstract

Water regulators and policymakers around the world are increasingly influencing water systems towards efficiency and sustainable consumption. In pursuit of these, most regulators mainly use traditional economic-analysis methods to benchmark water utilities and elicit water-service preferences. There have been discussions of several other techniques that extend the commonly used traditional economic analysis tools in the literature. Regardless of these discussions, the practical application of new economic analysis tools in the water sector remains relatively low. This study intends to extend the existing literature by providing more robust methods that could be useful to water regulators. The study asks four research questions to shed light on whether more robust methods are the way forward in water regulation. More precisely, the study investigates the consistency of efficiency scores obtained from the data envelopment analysis (DEA), stochastic frontier analysis (SFA) and stochastic non-parametric envelopment of data (StoNED) techniques on a sample of South African water utilities. Additionally, the study examines the impact of status quo bias, presentation format and response time on results from discrete choice experiments conducted using a case of the South African water sector.

The study reports four main findings. First, we find that the StoNED method (based on the methods of moments estimator) outperformed both SFA and DEA. However, SFA outperformed StoNED, when the latter was based on the pseudolikelihood estimator. Second, we find that including a partially relevant status quo reduced status quo bias but did not significantly affect empirical estimates. Major differences are noted in the marginal willingness to pay (MWTP) estimates reported for one of the sub-samples. Third, we find that presenting attributes and levels using the visuals format generated more statistically significant coefficients than presenting them as text or text-and-visuals. Generally, we find that the presentation format significantly affects choice. Finally, we find that removing fast or slow responses from the sample did not significantly affect both utility function and MWTP results.

Based on these findings, the study makes four main recommendations. Firstly, the study argues that StoNED (method of moments estimator) and SFA are more appropriate for estimating efficiency in heterogenous water sectors. The study makes recommendations for future studies that seek to do a methodological cross-checking of the three efficiency analysis techniques in the water sector. Secondly, the study argues that a text-and-visuals experiment improves choice task clarity and yields more robust estimates. Thus, more research on the effects of presentation formats is required in environmental economics so that guidelines on developing valid presentation formats for choice tasks can be established. Finally, the study argues against the exclusion of fast and slow responses from the dataset; and recommends approaches for future studies that investigate the impact of response time on choice.

Keywords: efficiency, choice experiments, status quo bias, presentation format, response time.

Acknowledgements

This thesis could not have been successful had it not been for the help I received from various people and institutions. I hereby acknowledge my indebtedness to several people whose support was valuable during my journey as a student.

I am grateful to my supervisor, Associate Professor Johane Dikgang for his continued guidance, patience, constructive criticism, support and all the opportunities he gave me. It was not an easy road, but he made it possible. He was an 'ideal' supervisor, an inspiration and a role model. My most sincere word of gratitude is the hardest to convey.

I also want to express gratitude to Professor Ronney Ncwadi, the Head of the Department of Economics and Economic History at the Nelson Mandela University (NMU), for all the support and his ability to make difficult processes less complex. Many thanks also go to Professor Stephen Hosking for introducing me to the field of environmental economics.

Furthermore, I am thankful to the staff and postgraduate students in the Public and Environmental Economics Research Centre (PEERC) at the University of Johannesburg (UJ). The support, collegiality and constructive criticism I received from them were helpful. I am so thankful to the PEERC's PhD brown-bag seminars and various other opportunities that improved my work. Special thanks go to Associate Professor Beatrice Simo-Kengne, Dr Hiywot Menker, Dr Dambala Gelo, Jugal Mahabir, Shavon Serfontein, Phindile Nkosi, Amanda Musandiwa, Akhona Mgwele, Isaiah Magambo, Celiwe Samkange, Jean-Luc Mubenga-Tshitaka, Nouran Zen El Abden, Mashekwa Maboshe, and Frank Bannor. Many thanks also go to Dave Buchanan for proof-reading my work.

I also want to thank my colleagues from the Department of Public Management and Economics at the Durban University of Technology (DUT) for their valuable support. Many thanks go Dr Jason Davis, Dr Kudayja Parker, Moonsamy Pillay, Nqobile Mpala and Christopher Ifeacho.

Special thanks also go to Dr Richard Mulwa for his guidance and support. This thesis was only possible because of his unstinted help. I also want to thank Dr Edwin Simiwu, Dr Leeward Jeke, Dr Tafadzwa Chitenderu, Samson Mukanjari, Zvikomborero Nyamazunzu, Taurayi Sihamba and David Damiyano. Many thanks for the times we spent sharing challenges and supporting each other in our respective journeys as PhD candidates. Furthermore, I want to thank Thelma Zindoga, Nothando Sithole, Slindile Mdletshe, Jerome Heugh, Simone Moswete and Moses Seleka for their help during the various phases of data collection.

I am also thankful to the Water Research Commission (WRC) for funding this research. Special thanks go to Jay Bhagwan for believing in us and for his team of water practitioners whose criticisms improved this research. Special thanks also go to the Research and Capacity Development (RCD) department at NMU for the bursary they gave me during my studies. The funds went a long way in making this dream come true. I would also want to thank the staff in the Office for International Education at NMU Special thanks go to Mrs Natasha September for her patience, kindness and unwavering support.

Furthermore, I want to express my sincere gratitude to the Swedish International Development Cooperation Agency (SIDA) and the Environment for Development Initiative (EfD) of the Department of Economics, University of Gothenburg for the opportunity to attend the PhD specialisation course in environmental valuation. Additionally, I would also like to thank my teachers in the discrete choice analysis course, in the Institute of Transport and Logistics Studies at the University of Sydney. Many thanks go to Professor David Hensher, Professor William (Bill) Greene, Professor Michiel Bliemer and Dr Andrew Collins.

Finally, I would like to thank my family for all the support they gave me during this journey. Special thanks go to my wife, my friend and lover, Tariro. She made so many sacrifices and had to go through a lot of hardships to make my dream come true. This journey would not have been successful without her love, care, support and understanding. For that, I am so thankful. Lastly and most importantly, I would like to thank God for making this dream come true.

Dedication

I dedicate this thesis to my wife Tariro, daughter Tanatswa Nicole and grandmother Gladys.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AIC	Akaike Information Criterion
ASC	Alternative Specific Constants
AVC	Asymptotic Variance-Covariance
BIC	Bayesian Information Criterion
CE	Choice Experiment
CLM	Conditional Logit Model
CNLS	Convex Nonparametric Least Squares
CoGTA	Cooperative Governance and Traditional Affairs
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Units
DWS	Department of Water and Sanitation
GMXL	Generalized Mixed Logit
IBT	Increasing Block Tariff
IIA	Independence from Irrelevant Alternatives
IID	Independently and Identically Distributed
IWA	International Water Association
KPI	Key Performance Indicator
KZN	Kwazulu-Natal
LL	Log Likelihood
MM	Methods of Moments
MNL	Multinomial Logit
MRS	Marginal Rate of Substitution
MWTP	Marginal Willingness to Pay
MXL	Mixed Logit
NIDS	National Income Dynamics Study
NL	Nested Logit
NRPL	Nonlinear Random Parameters Logit
OLS	Ordinary Least Squares
PPP	Public Private Partnership

PSL	Pseudo-Likelihood
RPL	Random Parameter
SALGA	South African Local Government Association
SFA	Stochastic Frontier Analysis
SQ	Status Quo
StoNED	Stochastic Non-Parametric Envelopment of Data
TE	Technical Efficiency
UK	United Kingdom
USA	United States of America
VRS	Variable Returns to Scale
WSA	Water Services Authority
WSP	Water Services Provider
WTP	Willingness to Pay

Chapter 1: Introduction

1.1. Introduction

The question of how to influence water systems towards more efficiency and more sustainable consumption has become a major focus in the water sector around the world and is gaining importance among policymakers and the general public. The water sector is highly regulated globally, despite a wave of reforms. According to Cabrera et al. (2018), the water sector is characterised by natural monopolies in many parts of the world; hence, it is highly regulated by governments. This high level of regulation is driven by the desire to protect the interests of consumers. Among the roles of water regulators, the Lisbon Charter (IWA, 2015) lists provisions to supervise water tariffs, oversee and promote suitable quality of service, protect consumer rights and conserve water.

Policymakers have been addressing these roles – at least in part – by collecting key data. This data allows regulators to compare the performance of water utilities by benchmarking them, with the aim of nudging underperforming utilities into improving their performance. The policymakers' focus in this instance is on managing water supply. Regulators also use their regulatory tools to nudge consumers towards water conservation. Stated-preference techniques such as choice experiments (CE) are useful for shedding light on water users' preferences. This implies that policymakers are also involved in water demand management. Effective water regulation therefore requires policymakers to be involved in both supply and demand management.

The literature contains a wealth of efficiency and stated-preference studies in the water sector. Stochastic Frontier Analysis (SFA), a parametric method, and Data Envelopment Analysis (DEA), a non-parametric method, are the two frontier techniques employed most often to evaluate and compare the performance of water utilities (see Byrnes et al., 2010; Romano and Guerrini, 2011; Worthington, 2014). Unlike parametric methods, DEA does not require definition of the functional form of the production/cost frontier (De Witte and Marques, 2010a). Nonetheless, DEA is a deterministic technique; and thus, cannot deal with imprecise data, or provide information about uncertainty (Kao and Liu, 2014). This suggests that traditional DEA models (Charnes et al., 1978; Banker et al., 1984) do not allow stochastic variations and uncertainty in data and require that the exact values of all inputs and outputs are known. However, these assumptions may be incorrect, since some data cannot be measured accurately enough in practice (Eslami et al., 2012). In fact, uncertainty exists naturally in data that are collected, monitored or recorded in the water sector.

To overcome the problems associated with traditional DEA and SFA and to take uncertainty into account, several extensions have been proposed; such as Monte Carlo simulation, the α-level-based approach, the chance constraint, bootstrapping, fuzzy ranking, DEA tolerance (Cabrera, 2018) and the stochastic non-parametric envelopment of data (StoNED). In the case of water utilities, these methodologies have barely been used – with the exception of De Witte and Marques (2010a), who applied the order-*m* method to incorporate environmental variables into water utility efficiency assessments; and De Witte and Marques (2010b), Ananda (2014), See (2015) and Molinos-Senante et al. (2016), who employed a double-bootstrap DEA approach to compute bias-corrected efficiency scores. Each of these methodological approaches has its advantages and shortcomings.

However, Bonilla et al. (2004) showed that the DEA tolerance method is simpler and quicker than the bootstrapping technique and leads to similar results. Moreover, Dong et al. (2017) found that the DEA tolerance approach is less subjective than the fuzzy approach, since it does not need the fuzzy sets of variables to be defined for units. Moreover, the DEA tolerance approach can be combined with the system of indicators proposed by Boscá et al. (2011), which allows units to be benchmarked in an uncertain context. StoNED, which combines the strength of both DEA and SFA – and hence is categorised as a semi-parametric technique – has not been used at all in the water sector.

Choice experiments are also commonly used in water policy formulation to elicit preferences for water products and services. According to Lancsar and Louviere (2008), there is growing acknowledgement that CE can provide more than information on preferences, especially given its potential to contribute more directly to outcome measurement for applications in economic valuation. Almost uniquely, CE could potentially contribute to outcome measurement for use in both cost-benefit analysis and cost-utility analysis.

Regulators predominantly still use traditional economic-analysis methods for benchmarking water utilities and eliciting water-service preferences. Techniques that extend these tools are increasingly being discussed, but applications remain relatively scarce in the water literature.

In this thesis, we progress the existing literature on these techniques by providing more robust methods that may prove very promising for water regulators. This research asks four research questions, posed to shed light on whether more robust methods are the way forward in water regulation.

1.2. Research objectives

The need for effective regulation for efficient water resources management policies and sustainable water use is well documented in the economics literature. This need is also emphasised in this thesis; hence, the objectives of this study are as follows:

1. Chapter 2 compares empirical efficiency results from DEA, SFA and StoNED approaches, using the South African water sector as a case study;

2. Chapter 3 tests for the effects of introducing a partially relevant status quo option aimed at reducing status quo bias in choice experiments;

3. Chapter 4 investigates whether presenting attributes and levels as text, visuals, or text-and-visuals generates significantly different results with respect to attribute interpretation, relative importance, and willingness-to-pay estimates; and

4. Chapter 5 tests the effects of response time on respondents' choices under a self-administered face-to-face survey environment.

1.3. Relevance of the study

The relevance of this study is fourfold. Firstly, the methodological cross-checking process using three tools adopted to analyse efficiency in South African water utilities provides more robust, reliable and useful information for regulatory analysis and for policymakers. The study is a novel application that contributes towards bridging the gap between methodology and empirical practice, making it relevant to policymakers, practitioners and scholars. It uses an innovative approach, and to the best of our knowledge is the first cross-checking process using three methods to be applied in the water sector. Furthermore, this is the first application of the StoNED efficiency-analysis approach in the water sector. The approach has previously been applied in the electricity sector, the banking sector and the agricultural sector (Eskelinen and Kuosmanen, 2013; Kuosmanen, 2012; Vidoli and Ferrara, 2015).

Secondly, the study adopts a unique approach to test for the effects of introducing a partially relevant status quo option aimed at reducing status quo bias in choice experiments. In doing this, the study is one of few studies (if any) that test for the effects of reducing status quo bias by dividing a population based on economic segmentation, into 'wealth' and 'poverty' groups, and presenting each sub-population with two different choice experiments. The first treatment gives participants in each sub-population a series of choice sets, each with a status quo option that resonates with them. In another treatment, each sub-population is presented with a series of choice sets, each with a status quo option that does not fully reflect their current situation.

Thirdly, the study evaluates more presentation formats than are evaluated in similar studies. In most cases, studies in the literature compare text to visual presentations, or text to text-and-visuals. However, this study compares three presentation formats (i.e. written text, visuals, and text-and-visuals). To the best of our knowledge, this is one of the very few studies in environmental economics to test the impact of the way in which attribute profiles are presented. Similar studies are found in other disciplines – for example, housing, urban planning and consumer studies. Therefore, this study bridges the gap in the choice experiment literature regarding environmental and resource economics studies that investigate the impact of presentation formats and establish guidelines on developing valid presentation formats.

Finally, in testing for response time, this study uses a self-administered face-to-face survey, with data captured by electronic devices. The use of electronic devices allows for the accurate capturing of response time. This makes our study unique, because most similar studies in the literature are predominantly web-based. Although convenient, web-based surveys are difficult to monitor, whereas researchers in face-to-face surveys can monitor the respondents' involvement. This study also differs from studies that give participants time to think before responding. Participants completed the survey in the presence of enumerators, allowing the enumerators to monitor activity and ensure it was done correctly, without necessarily pacing the participants.

1.4. Proposed structure of the study

This first chapter has established the nature of the study, introducing the objectives of and the relevance of the study. The research issues raised in this chapter are addressed in the following four chapters.

Chapter 2 compares efficiency results from DEA, SFA and StoNED, using the South African water sector as a case study. The chapter presents a methodological process for cross-checking the use of these three efficiency-analysis tools in the South African water sector.

Chapter 3 is an attempt to test for the effects of introducing a partially relevant status quo option aimed at reducing status quo bias in choice experiments. It examines the impact of the relevance or partial relevance of the status quo on the utility functions and marginal willingness-to-pay estimates.

Chapter 4 attempts to investigate whether presentation formats matter in environmental economics discrete-choice experiments. The chapter tests whether presenting attributes and levels in text, visuals, or text-and-visuals generates significantly different empirical results.

Chapter 5 is an attempt to test for the impact of response times on empirical estimates. It tests the hypothesis that fast responses reflect random decision-making and affect empirical results if they are not accounted for.

Chapter 6 presents the conclusion and recommendations of the study. The chapter also identifies some limitations of the study and suggests areas for future research.

List of references

Ananda, J., 2013. Evaluating the performance of urban water utilities: Robust nonparametric approach. *Journal of Water Resources Planning and Management,* 140**,** 04014021

Banker, R.D., Charnes, A. and Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30***,** 1078-1092.

Bonilla, M., Casasús, T., Medal, A. and Sala, R., 2004. An efficiency analysis with tolerance of the Spanish port system. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti,* 31**,** 379-400.

Bosca, J.E., Liern, V., Sala, R. and Martinez, A., 2011. Ranking decision making units by means of soft computing DEA models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 19**,** 115-134.

Byrnes, J., Crase, L., Dollery, B. and Villano, R., 2010. The relative economic efficiency of urban water utilities in regional New South Wales and Victoria. *Resource and Energy Economics*, 32**,** 439-455.

Cabrera, E., Estruch-Juan, E. and Molinos-Senante, M., 2018. Adequacy of DEA as a regulatory tool in the water sector. The impact of data uncertainty. *Environmental Science and Policy*, 85, 155-162.

Charnes, A., Cooper, W. W. and Rhodes, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2**,** 429-444.

De Witte, K. and Marques, R.C., 2010a. Designing performance incentives, an international benchmark study in the water sector. *Central European Journal of Operations Research*, 18**,** 189-220.

De Witte, K. and Marques, R.C., 2010b. Influential observations in frontier models, a robust non-oriented approach to the water sector. *Annals of Operations Research*, 181**,** 377-392.

Eslami, R., Khodabakhshi, M., Jahanshahloo, G.R., Lotfi, F.H. and Khoveyni, M., 2012. Estimating most productive scale size with imprecise-chance constrained input–output orientation model in data envelopment analysis. *Computers and Industrial Engineering*, 63**,** 254-261.

Eskelinen, J. and Kuosmanen, T. 2013. Intertemporal efficiency analysis of sales teams of a bank: stochastic semi-nonparametric approach. *Journal of Banking and Finance*, 37**,** 5163-5175.

IWA - International Water Association, 2015. The Lisbon Charter. [Online] IWA. Available: http://www.iwa-network.org/wp-content/uploads/2015/04/Lisbon_Regulators_Charter_SCREEN_EN_errata.pdf [Accessed 5 December 2018]

Kao, C. and Liu, S.T., 2014. Measuring performance improvement of Taiwanese commercial banks under uncertainty. *European Journal of Operational Research*, 235**,** 755-764.

Kuosmanen, T. 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. *Energy Economics*, 34**,** 2189-2199.

Kuosmanen, T. and Kortelainen, M. 2012. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38**,** 11-28.

Lancsar, E. and Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*, 26**,** 661-677.

Molinos-Senante, M., Donoso, G. and Sala-Garrido, R., 2016. Assessing the efficiency of Chilean water and sewerage companies accounting for uncertainty. *Environmental Science and Policy*, 61**,** 116-123.

Romano, G. and Guerrini, A., 2011. Measuring and comparing the efficiency of water utility companies: a data envelopment analysis approach. *Utilities Policy*, 19**,** 202-209.

See, K.F., 2015. Exploring and analysing sources of technical efficiency in water supply services: Some evidence from Southeast Asian public water utilities. *Water Resources and Economics*, 9**,** 23-44.

Vidoli, F. and Ferrara, G. 2015. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. *Empirical Economics*, 49**,** 641-658.

Worthington, A.C., 2014. A review of frontier approaches to efficiency and productivity measurement in urban water utilities. *Urban Water Journal*, 11**,** 55-73.

Chapter 2: Efficiency in South African water utilities: a comparison of estimates from DEA, SFA and StoNED

Abstract

For efficiency analysis to be useful to policymakers, the various approaches used should produce estimates that are consistent in identifying the best and worst firms, as well as overall rankings of firms in terms of their efficiency levels. This paper investigates the consistency of efficiency scores obtained from the data envelopment analysis (DEA), stochastic frontier analysis (SFA), and stochastic non-parametric envelopment of data (StoNED) methods. We estimate cost efficiency based on cross-sectional data from 102 South African water utilities in the period 2013/14. The results suggest that the StoNED method (based on the methods of moments estimator) outperforms SFA and DEA. However, based on the pseudo-likelihood estimator, SFA outperformed StoNED. Overall, the results suggest moderate consistency across the three methods. Based on the findings, we conclude that our results are robust.

Keywords: water utilities, StoNED, DEA, SFA, frontier efficiency.

2.1. Introduction

Climate change has brought renewed and increasing attention to the productivity and efficiency of the water sector. This has stimulated interest, which has manifested itself in the increased application of statistical tools to measure productivity and efficiency in the water sector. Policymakers in developed countries are already making use of statistical analyses of water systems for determining productivity and efficiency. The best-known examples are Switzerland, the United Kingdom (UK), the United States of America (USA), Germany, the Netherlands and Italy (see Baranzini et al., 2010). One of the most commonly used statistical tools for determining productivity and efficiency is the efficiency frontier model.

There are generally two types of techniques used in frontier analysis, namely, non-parametric approaches (i.e. mathematical programming) such as data envelopment analysis or DEA (see Charnes et al., 1978; Farrell, 1957); and parametric (i.e. econometric) approaches, such as the stochastic frontier analysis or SFA (see Aigner et al., 1977; Meeusen and Vandenbroeck, 1977). DEA and SFA are commonly used by researchers, practitioners and policymakers when carrying out frontier efficiency analysis. In estimating efficiency, DEA places less emphasis on the shape of the efficiency frontier and is credited for its axiomatic properties that can accommodate a multiplicity of inputs and outputs, as well as its ability to consider returns to scale. On the other hand, the strength of SFA is in its ability to decompose deviations from the frontier into random noise and inefficiency terms. These tools are not direct competitors, but complement each other, due to their respective advantages.

Since DEA and SFA use different assumptions, efficiency scores from the two methods may be inconsistent. Where these two techniques are used concurrently, it creates complexity as to which scores to adopt. In some cases, regulators use the arithmetic average of the firm-specific DEA and SFA efficiency estimates, whereas other regulators choose the highest out of the DEA and SFA estimates (see Kuosmanen, 2012; von Hirschhausen et al., 2006). However, both the arithmetic average and taking the highest score violate the assumptions of DEA and SFA. For this reason, Kuosmanen and Kortelainen (2012) developed the Stochastic Non-parametric Envelopment of Data (StoNED) method, which combines the axiomatic, non-parametric, piecewise linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise. StoNED is more robust to both model misspecification and noise, because its less restrictive assumptions imply a wider range of applicability.

Policymakers in developing and emerging economies including Asia, Africa and South America are now beginning to collect data that can serve as a basis for performance comparison, which can assist decision-makers to identify under-performing water utilities (Corton and Berg, 2009). An increasing number of countries are adopting performance appraisals to promote efficiency improvement in water provision. Performance appraisals almost invariably involve some form of benchmarking, or the comparison of actual performance versus some reference performance. Although there are various steps that can be used to undertake benchmarking, the process generally entails identifying relevant performance indicators; determining where performance should be, versus where it is at the time the evaluation is done (i.e. identifying performance gaps); determining the performance gap drivers; and designing an action plan to deal with the gaps. This process is then repeated continuously, as the organisation continuously improves its products and services.

Although some developing countries have tried benchmarking, most are still lagging. Those that do perform benchmarking often use performance indicators, which provide the ratio of an input to an output and vice versa (for example, total debt to total assets, or workers to number of connections). Benchmarking using the ratio of one input to a single output lacks scientific rigour and does not accurately portray the overall performance of the utility (Greenberg and Nunamaker, 1987). But in developing countries, the application of rigorous and more robust tools such as DEA and SFA is often limited to academics.

The main objective of this paper is to compare efficiency results between parametric, non-parametric and semi-non-parametric approaches, using the South African water sector as a case study. We extend existing studies comparing frontier analysis methods by introducing a robust semi-non-parametric methodological tool. This methodological cross-checking process using three methodological tools provides more robust, reliable and useful information and diagnostics for regulatory analysis and policymakers. This is an innovative approach, and to the best of our knowledge this is one of few such studies – and, the first cross-checking process using three methods – to be applied to the water sector. Furthermore, this is the first application of the StoNED approach to the water sector. Given the recent drought in South Africa which resulted in water crises, the efficiency analyses for the water sector in South Africa is very important.

The rest of the paper is organised into seven sections. Section 2.2 discusses the benchmarking methods used in the study. Section 2.3 gives an overview of the South African water sector.

Section 2.4 reviews some empirical literature, while Section 2.5 presents the empirical approach. Section 2.6 discusses the data used in the study. Section 2.7 presents and discusses the results. Section 2.8 concludes the study.

2.2. The three frontier efficiency tools

Benchmarking is a process of comparing the performance of one decision-making unit (DMU) with best practice among all peer DMUs (Zhu, 2014). Predominantly, the rationale for benchmarking is to promote competition, encourage information-sharing and transparency, give performance trends, and provide accurate information to decision-makers. Parametric and non-parametric efficiency analysis techniques have become strategic tools for benchmarking activities. Parametric techniques include corrected ordinary least squares or COLS (Winsten, 1957), parametric programming (Timmer, 1971), and SFA. Non-parametric methods include convex non-parametric least squares or CNLS (Hildreth, 1954), corrected concave non-parametric least squares or C^2NLS (Kuosmanen and Johnson, 2010), DEA, and StoNED. Of these methods, DEA and SFA (and recently StoNED) are the most commonly used approaches for benchmarking utilities. In this section, we discuss these three efficiency analysis techniques.

2.2.1. Data envelopment analysis (DEA)

DEA (Charnes et al., 1978, Farrell, 1957) constructs a non-parametric envelopment frontier over given data points, such that all observed points are on or below the frontier. If there are data on K inputs and M outputs on each of N decision-making units (DMUs), for the i^{th} DMU these variables are represented by the vectors x_i and y_i respectively. The KxN input matrix (x) and the MxN output matrix (y) represent the data for all N DMUs. For each DMU, the idea is to obtain a measure of the ratio of all outputs over all inputs, such as $u'y_i/v'x_i$, where u is an $Mx1$ vector of output weight and v is a $Kx1$ vector of input weights. Optimal weights are selected by specifying the following mathematical problem:

$$\max_{u,v}(u'y_i/v'x_i),$$

$$\text{s.t.} \quad u'y_j/v'x_j \leq 1, j = 1,2,\dots,N, \quad u,v \geq 0 \tag{2.1}$$

The process involves obtaining values for u and v such that the efficiency measure of the i^{th} DMU is maximised (subject to the constraint that all efficiency measures are equal to or less than one). To avoid the problem of an infinite number of solutions, the constraint $v'x_i = 1$ is imposed. The imposed constraint provides that:

$$\max_{\mu,v}(\mu'y_i),$$

$$\text{s.t.} \quad v'x_i = 1, \quad \mu'y_j - v'x_j \leq 0, j = 1,2,\dots,N, \quad \mu,v \geq 0 \tag{2.2}$$

This form is called the multiplier form of the linear programming problem, where the change of notation from u and v to μ and v reflects the transformation. When duality in linear programming is used, an equivalent envelopment form of the programming problem is derived. The equivalent envelopment form is presented as:

$$\min_{\theta,\lambda} \theta,$$

$$\text{s.t.} \quad -y_i + Y\lambda \geq 0, \quad \theta x_i - X\lambda \geq 0, \quad \lambda \geq 0 \tag{2.3}$$

where θ is the scalar and λ is an $Nx1$ vector of constants. An envelopment form of this nature includes lesser constraints than the multiplier form ($K + M < N + 1$). The obtained value of θ will be the efficiency score of the i^{th} DMU. This score satisfies $\theta \leq 1$, with a value of 1 indicating a point on the frontier, that is, a technically efficient DMU (Farrell, 1957). Best-practice utilities are relatively efficient, shown by a DEA efficiency rating of θ = 1, while inefficient utilities are shown by an efficiency rating of less than 1 (i.e. θ < l). DEA provides an efficiency rating that is generally between zero and 1, interchangeably referred to as an efficiency percentage between the range of 0 and 100%. The DEA model proposed by Charnes et al. (1978) had an input orientation that assumed constant returns to scale (CRS).

To estimate efficiency in South African water utilities, our study adopts the input-oriented assumption proposed by Charnes et al. (1978). However, since utilities in South Africa are quite diverse in terms of size, type and operating environment, we assume that they are at different stages of the production process. As such, we adopt the variable returns to scale (VRS) assumption in our DEA estimation. More precisely, we estimate an input-oriented DEA that assumes VRS.

2.2.2. Stochastic frontier analysis (SFA)

SFA is a parametric efficiency analysis technique that assumes a Cobb-Douglas, a log-linear or a translog functional form (Aigner et al., 1977; Meeusen and Van Den Broeck, 1977). The efficiency of DMUs is determined based on the specified functional form. The original formulation that is the foundation of SFA is:

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, \tag{2.4}$$

where y is the observed outcome (goal attainment), $\boldsymbol{\beta}'\mathbf{x} + v$ is the optimal frontier goal pursued by the DMU (e.g. minimum cost), $\boldsymbol{\beta}'\mathbf{x}$ is the deterministic part of the frontier, and $v \sim \mathrm{N}[0, \sigma_v^2]$ is the stochastic part. The two parts together constitute the stochastic frontier. The amount by which the observed DMU fails to reach the optimum (i.e. the frontier) is u, where $u = |U|$ and $U \sim \mathrm{N}[0, \sigma_u{}^2]$. The stochastic cost frontier then changes to $v + u$, where u represents inefficiency.

Different specifications of the terms u and v distinguish stochastic frontier models. According to Aigner et al. (1977), the normal-half normal model is the basic form of the stochastic frontier model. It assumes u to be independently half-normally $[\mathrm{N}+(0,\sigma_u{}^2)]$ distributed, with the idiosyncratic component v independently normally $[\mathrm{N}(0,\sigma_v)]$ distributed over the observation. Other SFA model specifications are the normal-exponential model (where u is independently exponentially distributed with variance $(\sigma_u{}^2)$), and the truncated-normal model (where u is independently $[\mathrm{N}+(\mu,\sigma_u{}^2)]$ distributed with truncation point at 0). For the sake of simplicity, our study uses the normal-half normal SFA model specification to estimate efficiency in South African water utilities. The study estimates an SFA cost function where total operation cost of providing water services (TC_i) is a function of the volume of water supplied (Q_i), the length of the water pipes $(MAINS_i)$, and the number of customers connected (CON_i). Therefore, the SFA functional form assumes:

$$lnTC_i = \alpha_i + lnQ_i + \ lnMAINS_i + lnCON_i - u_i + \ v_i \tag{2.5}$$

where v_i is the noise term assumed to be in normal distribution $v_i \sim \mathrm{N}[0,\sigma_\mathrm{v}^2]$. The u_i notation is the non-negative inefficient term (which is the distance from the observed cost to the cost on the frontier). The assumptions on the error term require both u_i and v_i to be homoscedastic. However, South African water utilities are diverse in size and operating environment. Such differences are likely to be captured in v_i, resulting in heteroscedasticity. But heteroscedasticity in v_i could lead to biased estimates, while heteroscedasticity in u_i leads to deceptive efficiency scores (Kumbhakar and Lovell, 2000). To address heteroscedasticity in an SFA function, one can account for the key drivers of the variation when estimating the efficiency term. This is done by estimating a simultaneous regression on the cost function, the inefficiency term, and the random noise term. This is specified as follows:

$$u_i = \alpha_1 + \alpha_2 lnCON_i + \delta_i \tag{2.6}$$

In this study, the variation in the inefficiency term u_i driven by heteroscedasticity is controlled for by regressing u_i on the total number of metered and unmetered water connections. This will ensure that the size of each utility is accounted for, minimising the impact of size on the efficiency estimates. Likewise, we also control for heteroscedasticity in the noise term v_i by regressing v_i on the total population served by each water utility (i.e. $v_i = \alpha_1 + \alpha_2 \mathrm{lnPOP}_i + \delta_i$). In doing this, we control for bias estimates that could have emanated from heteroscedasticity in the noise term, and biased efficiency scores because of heteroscedasticity in the inefficiency term.

2.2.3. Stochastic non-envelopment of data (StoNED)

Developed by Kuosmanen and Kortelainen (2012), StoNED combines the axiomatic, non-parametric, piecewise linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise. Combining a linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise makes StoNED more robust to both model misspecification and noise. The method has two main stages. The first stage estimates the shape of the total cost function using the convex non-parametric least squares (CNLS) regression, which belongs to the set of continuous, monotonic increasing and globally concave functions whose disturbances satisfy the Gauss-Markov assumptions. The second stage (which will be our focus in this study) estimates the expected inefficiency (μ), variance parameters (σ_u^2, σ_v^2), and DMU-specific inefficiencies. Kuosmanen (2012) suggests we introduce a composite error term ($\varepsilon_i = u_i + v_i$), and linearise the cost frontier function by taking the natural logs of both sides, to obtain:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \tag{2.7}$$

The main challenge in the least squares estimation of equation 2.7 is that the expected value of the composite error term is negative, due to the inefficiency term $u > 0$; that is, $E(u_i) = \mu > 0$. Kuosmanen (2008) reiterates that the composite error term in the model violates the Gauss-Markov properties, which can be restored by rewriting equation 2.7 as:

$$lnTC_i = (lnf(\boldsymbol{y}_i) - \mu)(\varepsilon_i + \mu) = lng(\boldsymbol{y}_i) + v_i \tag{2.8}$$

$$\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u \sqrt{2/\pi}$$

where $lng(\boldsymbol{y}_i) = lnf(\boldsymbol{y}_i) - \mu$ is the average practice cost function which can be contrasted with the best practice cost frontier $lnf(\boldsymbol{y}_i)$, while $v_i = \varepsilon_i + \mu$ is the modified composite error term. Since μ is a constant, the average practice function $lng(\boldsymbol{y}_i)$ inherits concavity and monotonicity properties from the best practice function $lnf(\boldsymbol{y}_i)$. The modified error term v_i satisfies the Gauss-Markov assumptions. The average practice frontier function can be estimated by a non-parametric regression technique such as StoNED. In the StoNED model, the assumption is that the cost of providing water (TC) by WSPs depends on a vector of outputs $\boldsymbol{y}$. Therefore, for each water utility, the CNLS problem is to find $g \in F_2$ that minimises the sum of square deviations of the average practice function, given as:

$$
\begin{aligned}
&\underset{f,v}{Min} \sum_{i=1}^{n} v_i^2 \left| lnTC_i = lng(\boldsymbol{y}_i) + v_i \right. \qquad \forall_{\,i} = 1, \dots, n \\
&\qquad \text{s.t.} \qquad g \in F_2
\end{aligned}
\tag{2.9}
$$

The CNLS estimator for the water utilities cost function is obtained as the optimal solution to the following least squares problem, which can be solved by convex programming algorithms and solvers:

$$
\begin{aligned}
&\underset{y,\boldsymbol{\beta},v}{Min} \sum_{i=1}^{n} \left(v_i^{CNLS}\right)^2 \\
&s.t. \quad lnTC_i = \alpha_i + \boldsymbol{\beta}_i' ln\boldsymbol{y}_i + v_i^{CNLS} \qquad \forall_i = 1, \dots, n \\
&\qquad \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{y}_i \le \alpha_h + \boldsymbol{\beta}_h'\boldsymbol{y}_h \qquad \forall_i, \forall_h = 1, \dots, n \\
&\qquad \boldsymbol{\beta}_i \ge 0 \qquad \forall_i = 1, \dots, n \\
&\qquad\qquad g \in F_2
\end{aligned}
\tag{2.10}
$$

where α_i is the intercept and $\boldsymbol{\beta_i}$ represents the coefficient of the tangent hyperplanes, which can also be interpreted as the marginal costs of output variables. These coefficients are analogous to the multiplier weights in DEA; and in contrast to the linear regression model, they are specific to each DMU (Kuosmanen, 2012). Parameter v_i^{CNLS} is the CNLS residual, the

CNLS estimator of g is monotonic increasing and concave, while ε_i^{CNLS} does not need to be identically and independently distributed but is uncorrelated with outputs $\boldsymbol{y}$ (see Kuosmanen and Johnson, 2010).

After the estimation of the CNLS residuals ($\hat{v}_i^{CNLS}$), the next step – which is the basis of our study – disentangles inefficiency from noise by imposing more specific distributional assumptions. Following the basic SFA developed by Aigner et al. (1977), we assume the half-normal distribution for the inefficiency term, and a normally distributed noise term. Usually the noise term is symmetrically distributed, and any skewing in the CNLS residual estimates can be attributed to inefficiency. According to Kuosmanen and Fosgerau (2009) it is essential to test if the skewing is statistically significant, in which case one can use the method of moments (MM) or the pseudo-likelihood (PSL) functions to estimate the variance parameters of the inefficiency and noise terms (σ_u^2, σ_v^2). When MM is used, assuming a half-normal inefficiency term and a normally distributed noise term, the second and third central moments of the composite error are given by:

$$M_2 = \left[\frac{\pi-2}{\pi}\right]\sigma_u^2 + \sigma_v^2, \qquad M_3 = \left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]\sigma_u^3. \tag{2.11}$$

Based on the distribution of CNLS residuals, these moments can be expressed as;

$$\widehat{M}_2 = \sum_{i=1}^{n} \frac{(\hat{v}_i - \hat{E}(v_i))^2}{n}, \quad \widehat{M}_3 = \sum_{i=1}^{n} \frac{(\hat{v}_i - \hat{E}(v_i))^3}{n}. \tag{2.12}$$

The third moment M_3, which is the skewness of the distribution, depends on the standard deviation of the parameter σ_u. This implies that the estimated $\widehat{M}_3$ should be positive in the case of a cost frontier. The σ_u parameter can be estimated as:

$$\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\left(\sqrt{2/\pi}\right)\left[1-\frac{4}{\pi}\right]}}. \tag{2.13}$$

Additionally, the standard deviation of the error term σ_v can also be estimated as follows:

$$\hat{\sigma}_v = \sqrt[2]{\hat{M}_2 - \left[\frac{\pi-2}{\pi}\right]\hat{\sigma}_u^2}. \tag{2.14}$$

Citing Aigner et al. (1977) and Greene (2008), Kuosmanen (2012) suggests that these MM estimators are unbiased and consistent, but not as efficient as maximum likelihood estimators. Using the estimator $\hat{\sigma}_u$ from MM, the best cost frontier function $lnf(\boldsymbol{y}_i)$ can be presented as:

$$ln\hat{f}(\boldsymbol{y}_i) = ln\hat{g}_{min}(\boldsymbol{y}_i) + \hat{\sigma}_u\sqrt{2/\pi} \tag{2.15}$$

According to Kuosmanen and Johnson (2010), this is like shifting the average practice frontier obtained from the CNLS by the expected value of the inefficiency term. The firm-specific inefficiency component u_i can be inferred indirectly from Jondrow et al's (1982) conditional distribution of inefficiency u_i given ε_i, irrespective of how the estimators of σ_u and σ_v are obtained. Under the assumption of a normally distributed error term and half-normally distributed inefficiency term, Jondrow et al. (1982) derive the conditional distribution of u_i given ε_i, and propose the conditional mean of the point estimate of u_i (i.e. $E(u_i|\varepsilon_i)$) as:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \mu_* + \sigma_*\left[\frac{f(-\mu_*/\sigma_*)}{1-F(-\mu_*/\sigma_*)}\right] \tag{2.16}$$

where f represents the standard normal density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ and F is the cumulative density function. Note that $(-\mu_*/\sigma_* = \varepsilon\lambda/\sigma)$ where $(\lambda = \sigma_u/\sigma_v)$. Given the parameter estimates $\hat{\sigma}_u$ and $\hat{\sigma}_v$ of the conditional inefficiency obtained from the method of moments, the conditional mean of u (assuming a truncated normal distribution) is given as:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \frac{\hat{\sigma}_u\hat{\sigma}_v}{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\left[\frac{f\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)}{1 - \mathrm{F}\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)} - \frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right] \tag{2.17}$$

where $\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u\sqrt{2/\pi}$ is the estimate of the composite error term, and not the CNLS residual. The obtained conditional expected value from equation 2.17 is an unbiased but inconsistent estimator of u_i; and irrespective of the sample size, each DMU will have a unique value of u_i. Technical efficiency estimates are given by $TE_i = e^{\hat{E}(u_i|\hat{\varepsilon}_i)}$ (see Dios-Palomares et al., 2002). The TE estimates can also be estimated by defining the ratio $(y_i/\hat{g}(x_i)$, where $\hat{g}(x_i) = \hat{E}(y_i|x_i) + \hat{\mu}$ is the estimated non-parametric frontier, which can be expressed as $\hat{g}(x_i) = \hat{E}(y_i|x_i) + \hat{\sigma}_u\sqrt{2/\pi}$ for the method of moments. Jondrow et al's (1982) estimates for $\hat{u}_i$ can be converted to cost efficiency measures (CE), expressed in the percentage scale by using CE = 100% x exp $(\hat{u}_i)$. The range of the cost efficiency scores CE is [0%, 100%], where CE=100% corresponds to the cost-efficient activity level (see Kuosmanen, 2012).

2.3. Benchmarking efforts in the South African water sector

Water sectors across the world are characterised by natural monopolies. In developing countries, the provision of water is usually the responsibility of public entities. This is because water provision is not a lucrative business in the developing world, where most citizens are poor and access to water is considered a basic human right; hence, water in these countries is a public good. In developed countries, the private sector takes part in water provision. Nevertheless, government regulations in most of these countries still make water service providers monopolies (Aubert and Reynaud, 2005; De Witte and Marques, 2010). Private participation in the water sector is also increasing in emerging economies (see Carvalho et al., 2015; Estache and Rossi, 2002; Souza et al., 2007).

Water services provision is a process involving the movement of water from source to final user. The process is comprised of water treatment works, storage and distribution. The delivery of water services is dependent on a sequential process along a value chain. Key players in the South African water sector value chain are the Department of Water and Sanitation (DWS), Water Services Providers (WSPs), Water Services Authorities (WSAs) and the final water users. The sequential interrelation of key players in the South African water sector value chain is shown in Figure 2.1.

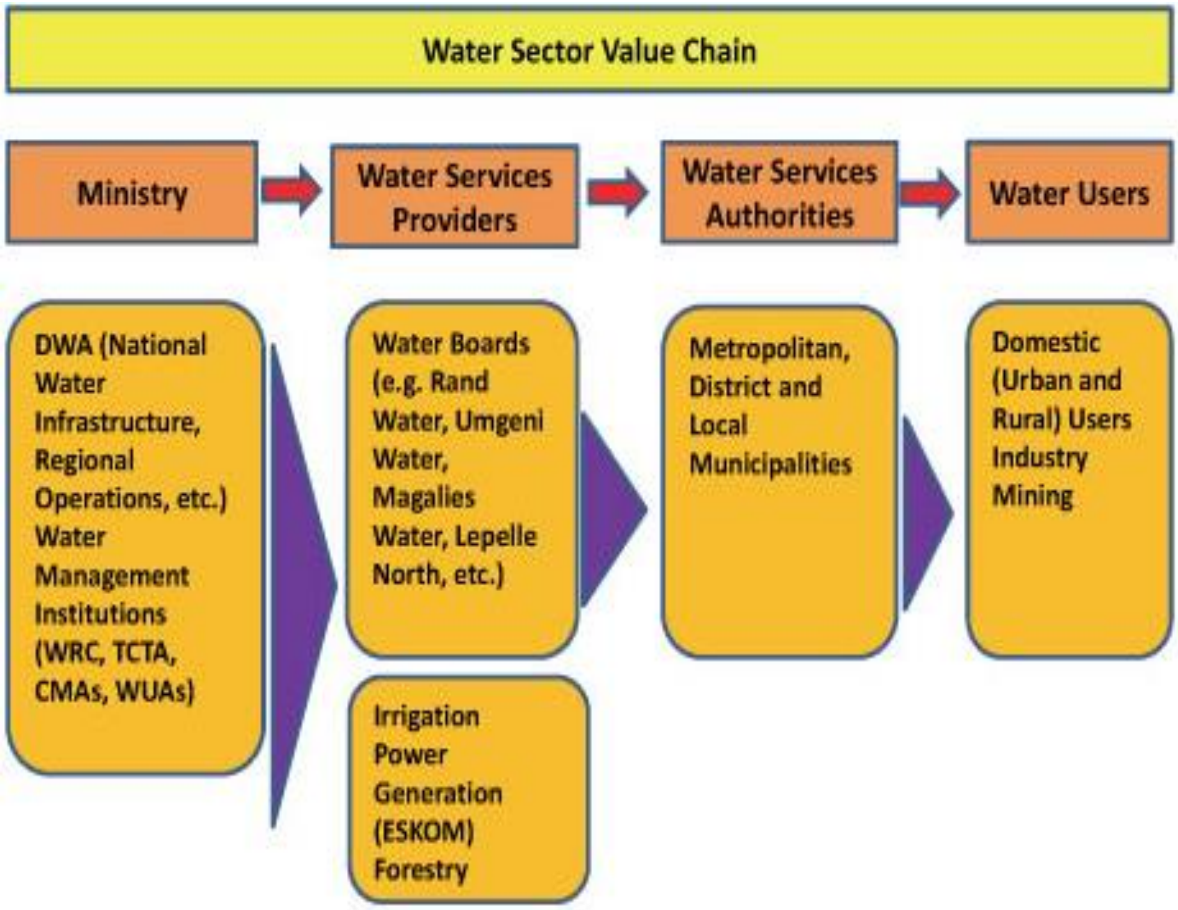


Figure 2.1: The water sector value chain in South Africa

Source: Ruiters (2013)

In the figure, Ministry refers to the DWS, which is the custodian of the country's water resources. The DWS is primarily responsible for the formulation and implementation of the policies that govern the water sector. The legislative mandate of the DWS is to ensure that the country's water resources are protected, managed, used, developed, conserved and controlled in a sustainable manner that benefits all people and the environment. In pursuing this mandate, the DWS plays a regulatory role through setting national norms and standards for water services, monitoring the performance of WSAs, providing support to WSAs, and intervening in cases of water service delivery failure. As a regulator, the DWS develops a knowledge base and implements policies, procedures and integrated planning strategies for both water resources and services. The DWS' regulatory role also includes supporting institutions in complying with existing regulations and institutional reforms, as well as enforcing set standards by way of incentivising performance and sanctioning non-performance.

The authority to supply potable water is a competence of municipalities, which act as water utilities. A municipality accorded the responsibility of providing water services is called a WSA. Although South Africa has 278 municipalities, only 152 are WSAs. The Minister of Cooperative Governance and Traditional Affairs (CoGTA) is responsible for determining which municipalities qualify to be WSAs. The 152 WSAs encompass district municipalities that deliver within the jurisdiction of their local municipalities, and local municipalities that deliver within their own jurisdictions. In most cases, where a district municipality is authorised to provide water, the local municipalities in the area do not have such authority; and in instances where the local municipalities within a district are authorised, the related district municipality is not authorised. If the local municipality is deemed to have a large enough budget, then it is authorised, as opposed to the district municipality. This usually occurs when the local municipality is considered a 'secondary city' (i.e. a local municipality with a large town or city as its urban core). The asymmetric delivery of water services across municipalities is due to the incapacity of many local municipalities to deliver water services, particularly those in the former homeland areas[1].

Although these arrangements have some merit, the ultimate delivery of water services in the country can be complicated. This is apparent where WSAs have the legal option to appoint a third party to provide all or part of the water services on their behalf. Section 76 of the

[1] Under the pre-1994 apartheid government, these were areas that were designated for specific black ethnic groups, with a high degree of political autonomy and even so-called 'independence' from South Africa.

Municipal Systems Act (Act 32 of 2000) differentiates between internal and external service-delivery mechanisms. The former is the delivery of water services by a department, administrative unit or business unit within a municipality. The latter includes the partial or complete outsourcing or commercialising of the delivery of water services. External delivery of a service by an authorised WSP would include outsourcing the service to another municipality, municipal entity, organ of state or the private sector, through commercialising the delivery of the service or by public-private partnerships (PPPs).

The need for standardised information, transparency and accountability has recently intensified, resulting in benchmarking efforts gaining momentum in South Africa. The primary goal of benchmarking is to provide key performance indicators (KPIs) that will enable utilities to compare their performance with the performance of other utilities and identify areas of improvement. However, in South Africa as in other developing countries, the conventional benchmarking approaches used in developed countries are not applicable in cases where water supply is intermittent, accessed by non-piped means, unmetered, and/or has a significant number of poorer customers on shared public connections (Mehta et al., 2013). Although water services provision is widespread in South Africa, there is a lack of data regarding the quality and level of service. Very little is known about how South African municipalities compare, in their capacity as WSAs. This is mainly due to a lack of standardised data, gaps in existing data, and lack of data verification.

Earlier performance benchmarking, initiated by the South African Local Government Association (SALGA) in 2001, was a failure. In 2006, government made further efforts; and since then, much has been achieved in the monitoring of municipal service performance through the Blue and Green Drop Certifications. The former proactively measures aspects contributing to sustainable safe drinking water, while the latter identifies and develops the core competencies required to sustainably improve the level of wastewater management. Although these programmes are plausible efforts towards performance benchmarking, water sectors across the world use scientific benchmarking tools such as DEA and SFA (see Baranzini et al., 2010; da Cruz, et al., 2013; Guerrini et al., 2015). These tools estimate technical and cost efficiency scores that are essential to performance benchmarking. In South Africa, DEA and SFA are only mentioned in the academic literature and have not been implemented by authorities for benchmarking (see Brettenny and Sharp, 2016; Tsegai et al., 2009).

Although WSAs are the legal custodians of water services within their jurisdictions, the methods used to deliver water services vary. Using scientific efficiency-analysis tools to benchmark municipal performance can deal with the inconsistences in the way water is provided across municipalities in the country. Efficiency-analysis methods can cater for instances where there are differences in treatment methods, or instances in which both a district and a local municipality are providing water, even though only the former is authorised. These discrepancies are accounted for in the inputs used and outputs produced, thus making technical and cost efficiency analysis a more accurate benchmark than other one-dimensional or ordinal methods.

2.4. Literature review

Most water-efficiency studies use either DEA or SFA. Over the years, studies in the literature have relied on DEA to estimate the efficiency of water utilities (see Brettenny and Sharp, 2016; Carvalho et al., 2015; Cruz, Carvalho et al, 2013; De Witte and Marques 2010; Guerrini, et al., 2015). Although DEA is a useful efficiency-analysis tool, it is often criticised for not allowing random error by assuming that any deviation from the frontier is inefficiency; an assumption that exaggerates inefficiency if noise is present (see Coelli et al., 2005; Leleu, 2006; Simar and Wilson, 2008). Assuming away the noise term makes DEA biased in small samples, and sensitive to outliers. Because of these criticisms, several studies that estimate the efficiency of water utilities opt for SFA as a better tool (see Aubert and Reynaud, 2005; Baranzini et al, 2010; Filippini et al., 2007; Horn and Saito 2011; Souza et al., 2007; Vishwakarma and Kulshrestha, 2010). SFA is credited for its ability to control for heterogeneity in the sample. However, it is often criticised for its functional form assumption, which is arbitrary and difficult to justify. Many commonly used functional forms fail to capture the economies of scope in joint production (Kuosmanen and Kortelainen, 2012).

In order to benefit from the advantages of both DEA and SFA, there is a large body of studies in the literature that use both tools to estimate the efficiency of utilities (see Dong et al., 2014; Herwartz and Strumann, 2012; Lannier and Porcher, 2014; Zschill and Walter, 2012). Apart from the scholarly literature, some regulators (mostly in Europe) use both DEA and SFA to measure efficiency and benchmark utilities for regulatory purposes. For example, from 2008 to 2011 the Finnish electricity regulator estimated both DEA and SFA and determined efficiency improvement targets using the arithmetic average of the firm-specific DEA and SFA scores (Kuosmanen et al., 2013). In Germany, the electricity regulator also estimated both DEA and SFA scores, but chose the maximum (von Hirschhausen et al., 2006). However, both taking the arithmetic average and taking the highest score violate the assumptions of DEA and SFA. Efficiency scores from DEA and SFA are estimated based on different assumptions; using the scores interchangeably violates the theories underpinning the models.

In a study that applied both DEA and SFA, Dong et al. (2014) used a panel data set of Chinese banks and found efficiency scores generated by SFA to be slightly higher than scores from DEA. The study also revealed that DEA and SFA were moderately consistent in identifying the best and worst quartile decision-making units regarding cost efficiency. In a different study, Herwartz and Strumann (2012) used DEA and SFA to examine whether hospital efficiency had emerged after the financial reform on spatial interdependence in Germany. Results showed that the SFA efficiency scores were higher than the DEA scores, reflecting that DEA identifies all deviations from the frontier as inefficiencies, while SFA separates inefficiency from noise.

The complexity associated with having to choose between DEA and SFA led to the development of StoNED. Introduced as a replacement to DEA and SFA in the regulation of electricity distribution utilities in Finland, StoNED is increasingly garnering attention in the literature. Kuosmanen et al. (2013) compared DEA, SFA and StoNED in the context of regulating electricity distribution, using data from Finland. The study compared the impacts of methodological choices on cost efficiency estimates and acceptable cost. In the results, the efficiency estimates were highly correlated, while the cost targets revealed major differences. StoNED yielded a root mean squared error of 4%, and its precision improved as the sample size increased. DEA yielded a root mean squared error of 10%, but performance deteriorated as the sample size increased. SFA had a root mean squared error of 144%, its poor performance explained to be due to the wrong functional form and multicollinearity. These comparisons demonstrate that the choice of method has significant effects on the regulatory outcomes.

Following the work of Kuosmanen (2012) and Kuosmanen et al. (2013), StoNED is gaining momentum in its use for efficiency analysis in the electricity sector. Cheng et al. (2015) examined the productivity development of Norwegian electricity distribution companies for the period 2004 to 2013. The study used DEA, SFA, and StoNED to examine productivity change, with the usual decompositions into efficiency change, technical change, and scale efficiency change. Based on the hypothesis that increasing investment and use of accounting-based capital costs leads to a negative bias in the productivity change estimates, analysis in the study was performed with and without capital costs, and results indicated a negative productivity development. In a different study, Li et al. (2016) applied SFA and StoNED to estimate efficiency for 23 Chinese power-grid companies, using data for the period 2005 to 2009. Among other findings, the study revealed that StoNED efficiency estimates were no different from those estimated by the various functional forms of SFA. StoNED has also been applied in various other studies in the energy literature (see Dai and Kuosmanen, 2014; Johnson and Kuosmanen, 2015; Mekaroonreung and Johnson, 2012; Sabouhi Sabouni and Kenari, 2014).

The application of StoNED to estimate efficiency is also found in the literature in areas such as banking, agriculture, and manufacturing. Eskelinen and Kuosmanen (2013) used StoNED to examine the efficiency and performance of sales teams over time in a bank branch network. The study estimates the intertemporal sales frontier from a panel of monthly data for the years 2007 to 2010. Using StoNED to assess the efficiency and performance development of the sales teams of a bank is one major contribution to the banking sector, where efficiency is central to sustainability. In a different study conducted in the agriculture sector, Vidoli and Ferrara (2015) use StoNED to estimate efficiency on Italian citrus firms. Using agricultural micro-data, the study maps out the overall level of efficiency, focusing on the evaluation of the differences observed due to the presence of contextual variables. Following a different method, Andor and Hesse (2014) used Monte Carlo simulations to evaluate the performance of StoNED relative to DEA and SFA, and found that in scenarios without noise, the rivalry is between DEA and SFA; while in noisy scenarios, StoNED pseudo-likelihood was a promising alternative to SFA.

Depite StoNED being applied across various fields, a gap exists in the literature on studies that use the approach to estimate efficiency in the water sector. To the best of our knowledge, the performance of StoNED has not been tested in benchmarking water utilities. The approaches common in the water literature are the conventional efficiency analysis techniques (i.e. DEA and SFA); they have also been extensively adopted by water regulators to benchmark utilities. A study closer to ours is Kuosmanen et al. (2013), which compared DEA, SFA and StoNED in the electricity distribution sector, and found that StoNED yielded the most precise efficiency scores in both heterogeneous and noisy samples.

By comparing efficiency estimates from DEA, SFA and StoNED in the South African water sector, our study becomes one of the first to use these three methods in the context of the water sector, especially in developing countries. Developing countries are mostly associated with inconsistent and inaccurate data, which make it difficult to use efficiency-analysis tools such as DEA, SFA and StoNED that require consistent data. However, some attempts have been made (see Brettenny and Sharp, 2016; Carvalho et al., 2015; Souza et al., 2007; Vishwakarma and Kulshrestha, 2010). These studies use either DEA or SFA, and report inefficiencies in the water utilities of developing countries. Considering that no study has used StoNED to compare the consistency of water efficiency levels, our study extends the current literature in that arena.

2.5. Empirical approach

Water provision is a process involving several operational costs (Filippini et al., 2007). To characterise the process, it is essential to assume the existence of a mathematical relationship between water supply inputs and outputs. Water provision costs (in the case of South African water utilities) include bulk water purchases, labour, interest on capital, depreciation of fixed assets, and other general expenditure, such as fuel and oil, printing and stationery, and hiring of plant equipment. We aggregate these cost components into one total operation cost variable denoted by TC_i. We then use cost frontier models to relate TC_i to output variables that influence each utility's cost structure. Output variables in this study are the volume of water supplied (Q_i), length of water pipes ($MAINS_i$), the number of connections (CON_i), and the population (POP_i), which is an exogenous variable. Therefore, our study assumes the cost frontier model to be:

$$TC_i = f\left(Q_i^{\alpha} MAINS_i^{\beta} CON_i^{\gamma} POP_i^{\Omega}\right). exp(v_i + u_i) \tag{2.18}$$

where u_i is a random variable representing the cost inefficiency of the water utility i, v_i is a stochastic noise term that captures the effects of measurement errors, omitted variables and other random disturbances, and $\alpha, \beta, \gamma, \Omega$ are parameters to be estimated. When the vector $\boldsymbol{y}$ is used to represent the output variables $Q, MAINS$, CON and exogenous variable POP, equation 2.18 can be written as:

$$TC_i = f(\boldsymbol{y}_i).exp(v_i - u_i) \tag{2.19}$$

If a composite error term $\varepsilon_i = v_i - u_i$, which consists of an inefficiency term $u > 0$, and a random parameter term $v = 0$ is introduced, and the cost function is linearised in logs, then equation 2.19 will be rewritten as:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \tag{2.20}$$

The cost function presented in equation 2.20 is estimated using DEA, SFA and StoNED. For DEA, we estimate an input-oriented DEA that assumes VRS. In doing this, we first estimate efficiency scores for the whole sample (consisting of all water utilities in the sample). Then we group utilities based on their sizes and estimate efficiency scores for each category. Grouping utilities is essential, since DEA is very susceptible to the influence of outliers (Banker, 1993). The South African water sector has variations in the sizes and operating environments of utilities; as such, categorisation is essential. Efficiency estimates from pooled data are compared to those derived from grouped utilities. Eventually, DEA efficiency scores will be compared to scores from the other two methods.

SFA is used for two main purposes. Firstly, to present the stochastic cost frontier to show the impact of output variables on the input variable. When doing this, we also control for heteroscedasticity in both the noise and the inefficiency terms. This is essential, because heteroscedasticity in the noise term can result in biased estimates; while heteroscedasticity in the inefficiency term can lead to misleading efficiency scores (Kumbhakar and Lovell, 2000). Secondly, we report on the efficiency scores for each water utility from the SFA model. SFA efficiency scores are then compared to those estimated using the other two methods.

StoNED has two main process stages: the first stage estimates the shape of the cost function, while the second stage estimates inefficiencies (Kuosmanen et al., 2013). This study does not report on the shape of the cost frontier but presents results on the utility-specific efficiency scores, which are then compared to estimates from the other two models. This approach is in line with the objective of the study. We present results from both the MM and PSL estimators of StoNED; but for further analysis, we adopt the estimator with efficiency scores that have less variance around the mean. Kuosmanen (2012) suggests that the MM is unbiased and consistent, while Andor and Hesse (2014) suggest that PSL gives more robust estimates. Therefore, for the StoNED model we present scores from both MM and PSL, and subsequently compare their variance around the mean. A complete list of all DMU-specific efficiency scores will be presented for DEA, SFA and StoNED – that is, in addition to the separate comparison of efficiency scores for big and small water utilities.

2.6. Raw data

The sample in our study comprises cross-sectional data for the 2013/4 period for 102 water utilities. This implies that we have just one (i.e. annual) observation per utility. We could not include all 152 water utilities, due to missing data; nor could we use panel data for other periods, due to too many gaps in the dataset. The period we are using for our analysis had the most complete data. The sample is representative of city, big-town, small-town and rural South African water utilities. As is the case in Dong et al. (2014), which follows the intermediation approach (defining input and output variables), treating Chinese banks as multi-product firms that employ inputs X_i at given prices W_i that minimise total costs TC to produce outputs Q_i, this study treats South African water utilities in exactly the same manner. Our study uses a single input (TC_i) with three outputs $(Q_i, MAINS_i, CON_i)$ and an environmental variable (POP_i), used to control for heterogeneity in the operating environments of the utilities[2]. (Similar variables are also used in Kuosmanen (2012), in the context of electricity distribution utilities).

Total cost (TC) is the total water-related operating cost[3] for each water utility. The total cost data is expressed in South African Rands[4], and comprises both direct and indirect costs resulting from bulk water purchases, labour, interest on capital, depreciation of fixed assets, and other general expenditure. In the context of this study, total cost is used as the only input

[2] In the context of this study, a water utility refers to a WSA that provides water services to final users.

[3] A water utility typically distributes water and other related services, such as wastewater treatment. However, in the case of South Africa, this is the responsibility of a local municipality. A local municipality provides various services such as refuse removal, electricity, roads and storm water drainage, and street lighting. Considering local municipalities incur costs for all services provided, we are interested in this study only in "water-related" costs, hence the use of that term. Non-water-related operating costs therefore includes costs incurred towards provision of street lights, municipal roads and electricity.

[4] The Rand is the South African currency. As at 24 October 2018, US$1 = ZAR14.30.

variable, and water utilities are expected to minimise cost, given output variables. The rationale for using operating cost is motivated by the reality that operating cost gives a true reflection of the actual costs of running a water services department each year.

Water output (Q) is the total quantity of water supplied by each water utility. The authorised consumption expressed in kilolitres (kl) per annum is used to account for water output. Authorised consumption is defined by the DWS as the total volume of metered and/or non-metered water taken by registered customers, the water supplier itself, or others who are implicitly or explicitly authorised to do so by the water supplier. Water output is used in this study as one of the three output variables. Water utilities buy bulk raw water from water boards and are charged per kilolitre for the quantity bought. Hypothetically, the higher the quantity of water supplied by a utility, the higher will be the total costs, *ceteris paribus*. Therefore, utilities are expected to minimise the cost of providing water services, given a certain quantity supplied.

Total connections (CON) is the total number of metred and non-metered water connections for each utility. The connections variable shows the number of water consumer units for a water utility and is used as one of the three output variables. Hypothetically, more connections for a utility implies more consumer units, which may result in higher costs of providing water services, *ceteris paribus*. Therefore, the rationale for each water utility is to minimise its cost given a certain number of water connections.

Length of mains (MAINS) is the total length in kilometres of the water pipes owned by each water utility during the year. This shows the distance that the water moves, from point of extraction to the last consumer for each water utility. The variable is used as one of the three output variables. Hypothetically, utilities with longer pipe networks incur more costs from water losses and from transporting water over long distances. As such, water utilities are expected to minimise the cost of providing water according to the length of the pipe network.

Population (POP) is the number of people served by each utility. The hypothesis is that the total cost of providing water services is likely to be higher for utilities with higher population figures. Arbitrary higher total cost figures for utilities with lower population numbers may be attributed to inefficiency, *ceteris paribus*. The population figures express the size of each utility's distribution network. As such, this variable is used as an exogenous variable necessary to control for heterogeneity. In the literature, population is used extensively to control for heterogeneity (see Baranzini et al., 2010; Filippini et al., 2007; Tsegai et al., 2009). The figure below shows the map of where the utilities (i.e. municipalities) are located.

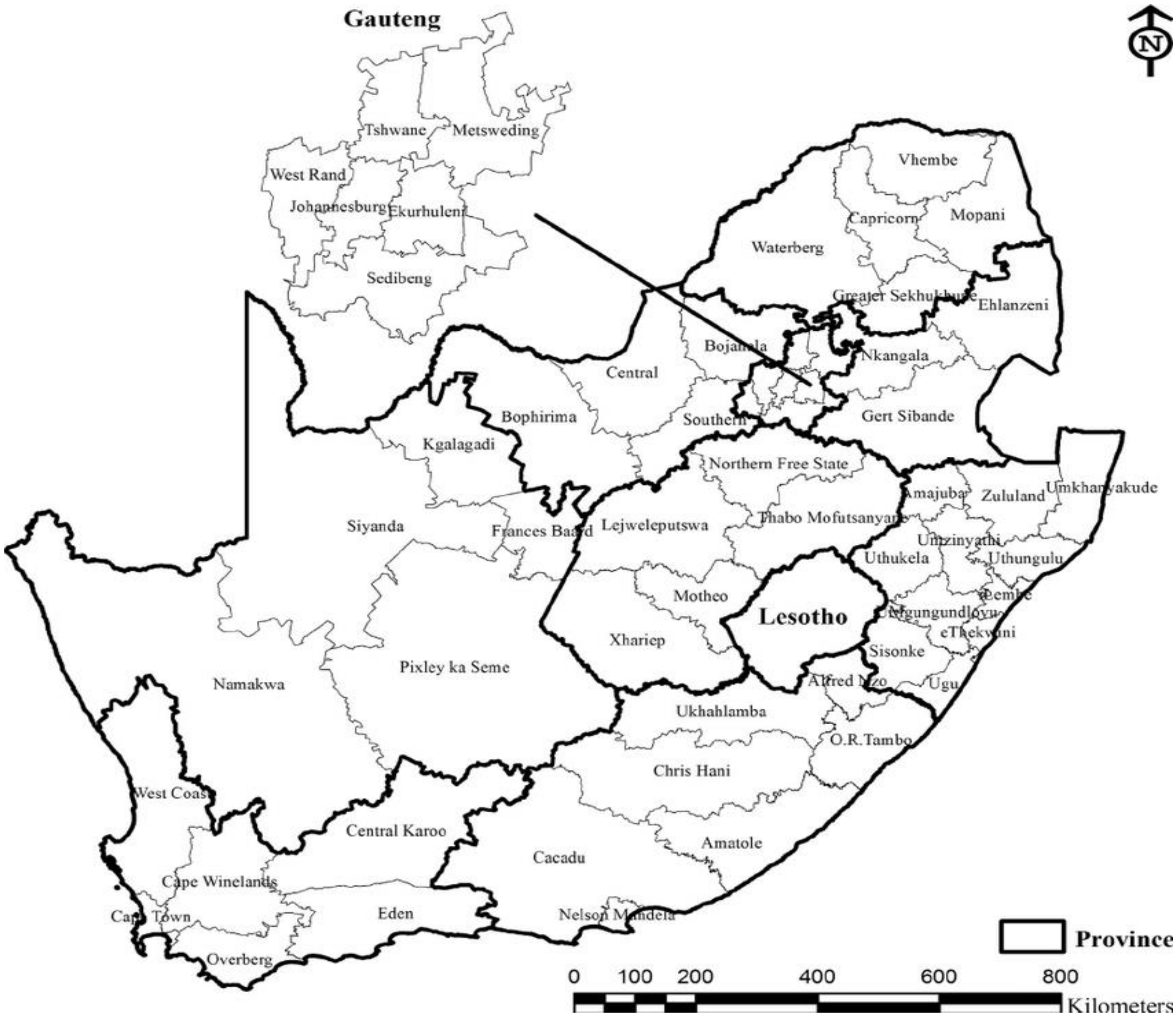


Figure 2.2: Map of where South African district municipalities are located

Source: Wabiri et al. (2016)

The analyses in the study are based on a sample of South African water utilities depicted in the figure above. The district municipalities depicted in the figure are further divided into various local municipalities which are also included in our analyses. Summary statistics of all the variables used in the analysis are presented in Table 2.1.

Table 2.1: Descriptive statistics

Variable	Description	Mean	Std. Dev	Min	Max
TC	Total cost in thousands of Rands	286,000	813,000	990	5,380,000
Q	Quantity of water in thousands of kilolitres	22,700	56,700	350.4	352,000
CON	Total number of connections	73,543	130,269	2,306	713,143
MAINS	Length of mains in kilometres	1,485	2,593	46	12,479
POP	Population served in thousands	413.3	801.2	10.6	4,500

Notes: Observations (N) =102.

The table above reveals that water distribution in South Africa is heterogeneous. The distinctions in size are evident from the statistics provided. Our sample contains all the categories of water utilities (i.e. city, big-town, small-town and rural water utilities). These categories vary in terms of size, operational environment, and resources. City and big-town water utilities serve huge populations, because they have urban cores that are highly populated due to urbanisation, which is prevalent in South Africa. On the other hand, utilities serving small towns and rural areas are relatively poor and have low densities in terms of population distribution. For such utilities, population levels may be relatively less and the number of water connections relatively few. However, the length of the mains and the total cost of providing water services may be relatively larger, because water is distributed across widely spaced household units.

There are statistical variations in the sample suggesting heterogeneity. For example, the total cost of providing water varies from R990,000 to R5.38 billion, while the quantity of water supplied by utilities varies from 350,400 kilolitres to 352 million kilolitres. The number of water connections within the sample varies from 2,306 to 713,143, while the total length of water pipes varies from 46 kilometres to 12,479 kilometres. The population statistics also show evidence of heterogeneity, varying from 10,578 people to 4.5 million people. Total costs are higher in large water utilities, due to larger population sizes and many connections.

The implication of these variations is that they affect the selection of efficiency-analysis tools, and how the selected tools are used. The most applicable estimation tools in such heterogeneous samples are SFA and StoNED, because of their ability to separate and control for noise (Andor and Hesse, 2014; Kuosmanen et al., 2013). If one decides to use DEA, as Brettenny and Sharp (2016) did, one needs to carefully separate water utilities according to their sizes and operating environments. This is because DEA is more susceptible to the influence of outliers (Banker, 1993), and is likely to make utilities serving lower population numbers appear relatively more efficient, while utilities serving higher population numbers are deemed relatively less efficient. This is probably due to smaller utilities with smaller inputs being compared to larger utilities. Such results highlight the importance of accounting for utility size, given the heterogeneous nature of South African water utilities.

Sample size has been identified as one important factor influencing the performance of efficiency estimation methods (Andor and Hess, 2014). Where smaller samples are used, more variables should be included for each DMU, as a measure to increase the number of observations. One major weakness of few observations is that it becomes hard to get coefficient estimates with small standard errors in parametric modelling like SFA. On the other hand, DEA is affected by a variation in sample size, but the direction of the effect depends on the underlying scenario. Andor and Hess (2014) explain that in scenarios without noise, the performance of DEA improves with an increasing number of DMUs, while it deteriorates with a growing number of DMUs in noise scenarios. For StoNED, computations with more than 300 observations can take several days (see Kuosmanen, 2012; Andor and Hess, 2014). As such, the number of observations for this study may be deemed fewer for DEA but are relevant for real-world application of SFA and StoNED.

2.7. Results and discussion

This section presents the estimated results using the three models defined previously: the SFA, DEA and StoNED. The results are presented in four main steps. Firstly, we report on the summary statistics of the efficiency scores generated by each method. It is imperative to note that here, we present summary statistics for scores from an SFA model, two StoNED models (MM and PSL), and three DEA models (whole sample, big water utilities, and small water utilities). Secondly, we show the distribution of utility-specific scores around the mean for all methods.

After the first and second steps, we compare the standard deviations of the two StoNED methods (MM and PSL) and adopt the method with less variation for further analysis. The standard deviation is essential for showing how well each model controls for heterogeneity, where less variation around the mean efficiency score implies the model’s ability to control heterogeneity. Thirdly, we present the frequency distribution of scores for each model by analysing the frequency of utilities below the mean efficiency score of each method, and the frequency of those above the mean. Finally, we group utilities into big and small water utilities and then compare the scores of each group for all models.

We extracted efficiency scores for each water utility from the SFA model and compared the SFA efficiency scores to those estimated using DEA and StoNED. Our study estimated an input-oriented DEA, which assumes VRS (because water utilities are at different levels of production). Since DEA is susceptible to the influence of outliers (Banker, 1993), we group utilities using their categories into big and small utilities and estimate efficiency scores. This allows for a comparison of water utilities of similar sizes. However, for consistency with the other estimation models used in this study, we also pooled all the utilities together and used DEA to estimate efficiency scores, which were then compared to those from big and small utilities as well as SFA and StoNED. For the StoNED analysis, we used both the MM and the PSL techniques. Summary statistics of the efficiency scores based on each method are given in Table 2.2.

Table 2.2: Summary statistics of efficiency scores based on method

	StoNED (MM)	**StoNED (PSL)**	**SFA**	**DEA (All utilities)**	**DEA (Big utilities)**	**DEA (Small utilities)**
Mean	0.681	0.529	0.662	0.447	0.587	0.461
Minimum	0.396	0.018	0.223	0.095	0.154	0.104
Maximum	0.762	0.724	0.896	1.000	1.000	1.000
Std. Dev	0.079	0.146	0.139	0.280	0.256	0.289

The table shows average efficiency scores for the 102 water utilities ranging from 0.447 in DEA to 0.681 in StoNED MM. Using the StoNED MM estimate, the average efficiency score is interpreted to mean that on average, water utilities in the sample are 68% efficient (i.e. 32% inefficient). This implies that utilities could reduce their operating costs by 32% and still afford to supply the same quantity of water, serve both the same population, and number of connections. When utilities are grouped together, the DEA average efficiency estimate is 45%, which is almost the same as the average estimate for small utilities (i.e. 46%), but less than the average estimate for big utilities, which is 59%.

In terms of standard deviation statistics, which express how much utility-specific efficiency scores vary from the mean score, the StoNED MM function reported the lowest standard deviation, of 0.079; followed by SFA, with a standard deviation of 0.139. In this regard, SFA performed better than the more sophisticated StoNED PSL. This confirms the findings in Andor and Hess (2014) that in noise samples, there is competition between StoNED PSL and SFA. Furthermore, it justifies why pioneering studies on StoNED, such as Kuosmanen et al. (2013) and Cheng et al. (2015), used the MM function. DEA reported the most variations which are similar across all three DEA categories. Based on the variations observed, the StoNED MM method performed better than the other models.

Although the summary statistics presented in Table 2.3 above give a clear snapshot of the performance of the three models, it is equally essential to examine how the efficiency scores for all the 102 water utilities in our sample were distributed around the mean. To do this, we graphically illustrate all utility-specific efficiency scores extrapolated from the three analysis models. For the DEA scores, we use estimates from the combined (pooled) sample. The distribution of the efficiency scores for all water utilities around the mean is presented in Figure 2.3.

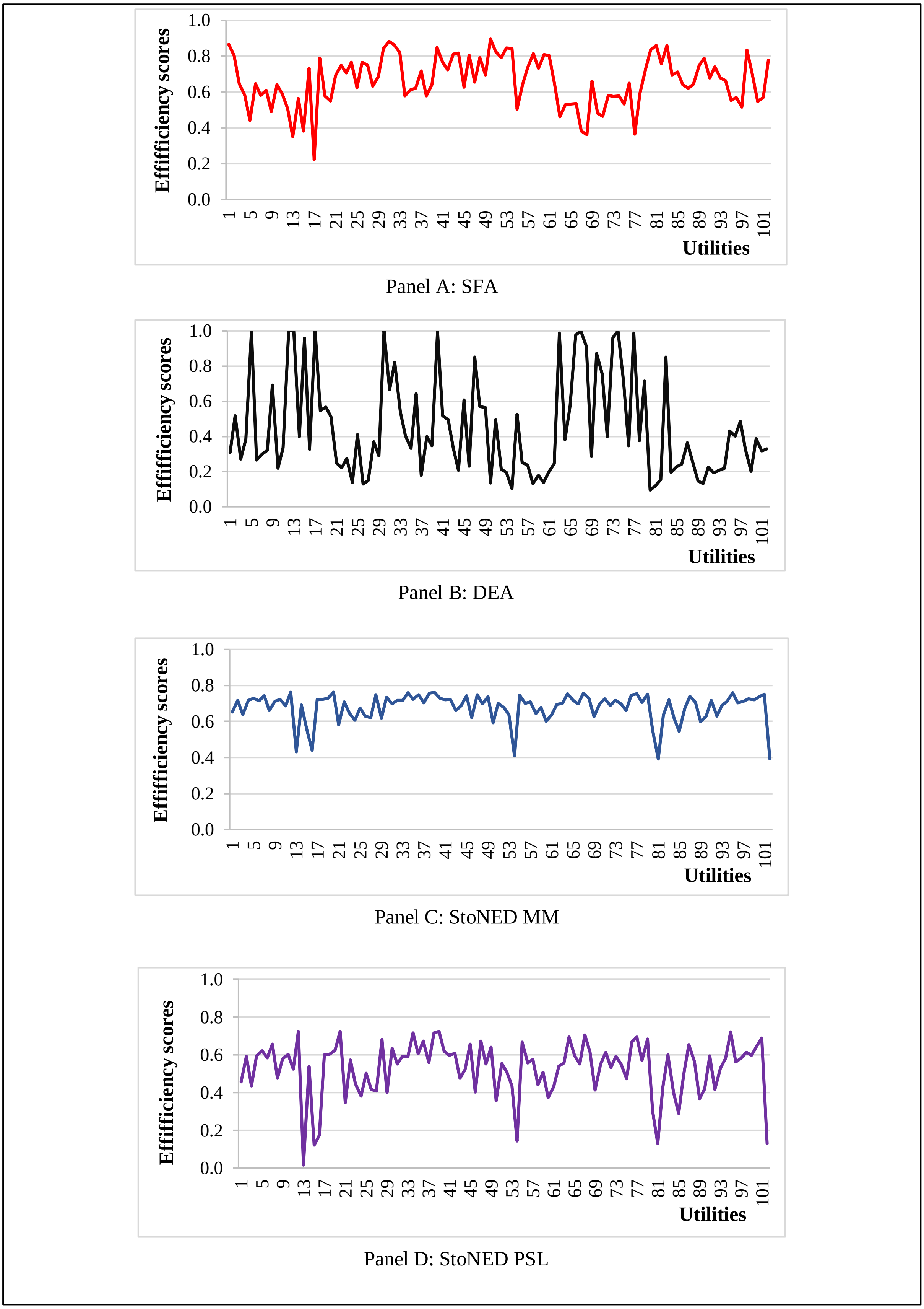


Figure 2.3: Distribution of efficiency scores around the mean

The figure shows less variation around the mean in StoNED MM efficiency scores, followed by SFA. This is coherent with the summary statistics presented earlier in Table 2.2. The variation in scores generated by DEA was expected, because we took estimates for the whole sample; yet the method does not account for heterogeneity. Rather, DEA identifies the best-performing utilities in the sample, then compares the other utilities to the best-performing ones (see Charnes, et al., 1978; Farrell, 1957). This would then make smaller utilities with smaller populations and fewer connections appear more efficient compared to bigger utilities. Regarding StoNED, the theory assumes the PSL function to be more efficient than the MM function in estimating efficiency scores (see Fan et al., 1996). However, our estimates in both Figure 2.3 and Table 2.2 above show MM performing better than PSL in terms of variation from the mean efficiency score. Since the StoNED MM scores show less standard deviation relative to PSL scores, our subsequent analysis will be based on StoNED MM scores.

To gain more insight into each estimation technique, we present the frequency distribution of utilities with efficiency scores below the model's mean score, as well as the frequency of utilities with scores above the model's mean. This implies that scores from each model are compared to the mean for the model. The results are presented in Figure 2.4 below.

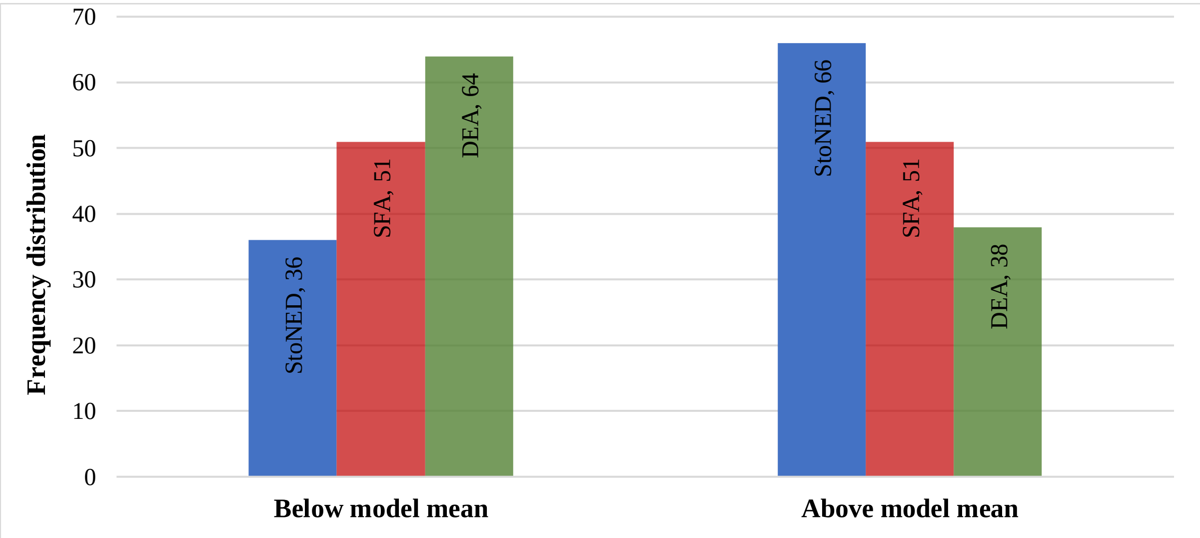


Figure 2.4: Frequency distribution of utilities below and above model average score

The results in Figure 2.4 show that using StoNED, 66 of the 102 water utilities (i.e. 65% of the utilities) scored efficiency estimates above the method's mean of 0.681. This implies that 65%

of the utilities are more than 68.1% efficient under StoNED. On the other hand, SFA results show an even distribution of utilities across the two categories. Each of the two categories had 51 utilities. That is, 50% of the utilities reported efficiency scores below SFA's mean of 0.662. SFA results were somewhat coherent with those reported in StoNED. In fact, DEA reported the most inconsistent results relative to the other models. DEA scores show 64 utilities below the model's mean of 0.447, implying that 63% of the utilities reported efficiency estimates below 44.7% under DEA[5].

Since our main aim is to estimate the efficiency scores for each utility given the three methods, it is essential to present the estimated efficiency scores for each utility, using each method. Since we have a relatively large sample of 102 utilities, we therefore present efficiency scores for selected utilities from the sample. To give an intuitive snapshot of our diverse sample, we present a range of scores, from both big utilities (in cities and big towns) and small (in small towns and rural areas). The chosen utilities were randomly selected from the total of 102, and they cover the nine South African provinces[6]. Table 2.3 presents efficiency scores for these selected utilities.

[5] Further analysis revealed that 5 utilities in StoNED, 3 utilities in SFA and 67 utilities in DEA recorded efficiency scores below 50%. Further analysis of the distribution of DEA scores reported for big and small utilities show that 56% of the former had scores below the mean, while 67% of the latter had scores below the mean.

[6] A complete list of the efficiency scores for all the water utilities of the sample is given in Appendix 2.1.

Table 2.3: Efficiency scores for selected municipalities using the three estimation methods

Big water utilities				Small water utilities			
Utility	**StoNED**	**SFA**	**DEA**	**Utility**	**StoNED**	**SFA**	**DEA**
Nelson Mandela Bay	0.653	0.867	0.309	Camdeboo	0.644	0.647	0.272
Buffalo City	0.717	0.803	0.517	Ikwezi	0.737	0.443	1.000
Amathole	0.762	0.508	1.000	Sunday's River	0.663	0.609	0.320
Chris Hani	0.557	0.382	0.960	Baviaans	0.718	0.489	0.692
Joe Gqabi	0.724	0.223	1.000	Kouga	0.722	0.642	0.219
Mangaung	0.722	0.789	0.546	Kou-kamma	0.688	0.593	0.335
City of Johannesburg	0.735	0.843	1.000	Tsolwana	0.696	0.564	0.398
City of Tshwane	0.700	0.883	0.666	Tswelopele	0.588	0.691	0.249
Ekurhuleni Metro	0.718	0.863	0.823	Setsoto	0.649	0.706	0.275
eThekwini Metro	0.728	0.850	0.996	Mantsopa	0.680	0.623	0.411
Ugu	0.720	0.765	0.518	Richtersveld	0.712	0.461	0.988
Umgungundlovu	0.663	0.812	0.331	Karoo Hoogland	0.628	0.535	0.976
Uthukela	0.686	0.819	0.207	Umsobomvu	0.729	0.361	0.911
Amajuba	0.623	0.805	0.230	Thembelihle	0.700	0.482	0.872
Uthungulu	0.700	0.791	0.569	Siyathemba	0.729	0.464	0.756
iLembe	0.737	0.696	0.563	!Kai! Garib	0.693	0.582	0.398
Mopani	0.594	0.896	0.133	!Kheis	0.700	0.578	1.000
Vhembe	0.701	0.828	0.494	Tsantsabane	0.664	0.532	0.708
Capricorn	0.641	0.845	0.196	Dikgatlong	0.755	0.365	0.987
City of Cape Town	0.720	0.860	0.852	Gamagara	0.715	0.593	0.375
Average score	**0.690**	**0.756**	**0.596**	**Average score**	**0.691**	**0.548**	**0.607**

The table shows that the average efficiency scores for the selected big utilities lie in the range 0.596 to 0.756, while those for small utilities lie in the range 0.548 to 0.691. StoNED reported no major differences in the average scores for both big and small utilities. However, SFA reported big utilities to be more efficient (with an average efficiency score of 0.756, i.e. 75.6%) than small utilities (with an average efficiency score of 0.548, i.e. 54.8%). On the other hand, DEA reported small utilities to be relatively more efficient than big utilities. However, in the case of DEA the margin is not as large – small utilities are on average 60.7% efficient, while big utilities are on average 59.6% efficient. This revelation shows increasing returns to scale among water utilities.

Regarding the actual utility-specific efficiency scores, in many instances DEA reported very high scores (such as 100% efficiency) or very low scores, relative to the other methods. By contrast, StoNED estimates lay mostly in between the DEA and SFA scores. For example, for

Nelson Mandela Bay, DEA reported an efficiency score of 0.309 while SFA reported a score of 0.867; StoNED was in between, with 0.653. The same trend is observed in utilities that were 100% efficient according to DEA. One example is Ikwezi, with a DEA score of 1, an SFA score of 0.443 and a moderating StoNED score of 0.737. In addition to the Nelson Mandela Bay and Ikwezi, the same trend is generally observed in other water utilities including Chris Hani, Joe Gqabi, Mopani, Capricorn, Richtersveld, and Tswelopele. StoNED proved to be more robust in estimating efficiency scores for our sample. This finding agrees with Kuosmanen et al. (2013), in which DEA, SFA and StoNED were compared and StoNED yielded the most precise results.

The efficiency scores reported in this study are consistent with those reported in the literature where DEA and SFA are used to estimate the efficiency of water utilities (see Brettenny and Sharp, 2016; Estache and Rossi, 2002; Horn and Saito, 2011). Brettenny and Sharp (2016) used an input-oriented DEA and estimated efficiencies for 88 South African water utilities and revealed average efficiency scores of 0.636 for urban water utilities and 0.526 for rural water utilities. As for scores from international water utilities, South African scores are comparable to those presented in Horn and Saito (2011), in which efficiencies were estimated for 831 Japanese utilities using SFA and the average scores were between 0.596 and 0.621. Similarly, Estache and Rossi (2002) SFA and estimated efficiencies for water utilities in Asia and the Pacific region. The study found the average efficiency to be within the range 0.72 to 0.78 in Bangkok, 0.66 to 0.69 in Beijing, 0.70 to 0.77 in Delhi, 0.66 to 0.77 in Hong Kong, 0.24 to 0.35 in Jakarta, 0.83 to 0.87 in Kuala Lumpur, and 0.74 to 0.75 in Singapore, among many others. These results indicate that South African water utilities compare well to international utilities.

After estimating efficiency scores, several efficiency analysis studies in the literature proceed to regress the estimated efficiency scores against the output variables. This practice is common in studies that use a two-stage DEA approach, as researchers try to ascertain the drivers of efficiency scores (see Dunghana et al., 2004; Lannier and Porcher, 2014; Ren, Li and Guo, 2017; Sharma et al., 1999; Zschill and Walter, 2012). According to Battese and Coelli (1995), one way of doing this is by including the variables in the stochastic frontier model, as presented earlier in the study. The advantages of this approach are explained in Kumbhakar and Lovell (2000). We followed this approach; the impact of the output variables on total cost were presented earlier, in the SFA model results.

2.8. Conclusion

In this paper, we employ parametric (SFA), non-parametric frontier (DEA) and semi-non-parametric (StoNED) approaches on a sample of South African water utilities, for methodological cross-checking purposes. In many developing countries there is a need to introduce rigorous benchmarking of the water sector, due to the low operational efficiency of existing public water utilities. As climate change intensifies, and competition increases between different needs for water, inefficiencies in the water sector in developing countries such as South Africa are bound to rise significantly. Efficiency gains is a potential adaptation strategy that the water sector could use to address several emerging trends driven by climate change. Using the three efficiency analysis techniques, our study reports four key findings.

Firstly, a comparison of standard deviations revealed that StoNED MM had the least standard deviation, followed by SFA, which outperformed StoNED PSL. DEA reported the most variations, which were similar across all three DEA categories employed in the study. Standard deviation is a key measure of robustness in the context of this study, as it expresses how much utility-specific efficiency scores within the given sample vary from the mean score. A technique that produces the least variation is deemed more robust, as it manages to control for heterogeneity in the sample. In this study DEA remained susceptible to outliers, even when we separated utilities and categorised them according to size.

Secondly, we observed that while SFA efficiency scores were distributed evenly, most utilities recorded scores above the model's mean under StoNED while most utilities reported scores below the model's mean under DEA (this was consistent even when utilities were grouped into big and small). Precisely, 65% of utilities had efficiency scores above the method's mean of 0.681 under StoNED while in SFA, 50% of the utilities reported efficiency scores above model's mean of 0.662. For DEA, only 37% of the utilities reported efficiency scores above the model's mean of 0.447. Another key observation regarding DEA is on its mean efficiency score which was below 50% efficiency. Further analysis revealed that 67 utilities under DEA had efficiency scores below 50%. This is consistent with the theory that DEA does not account for noise but treats any deviation from the frontier as inefficiency.

Thirdly, for most of the utilities, efficiency scores estimated using the StoNED model moderated those from DEA and SFA. Where DEA gave a higher efficiency score and SFA gave a lower efficiency score (and vice versa), StoNED usually gave a median score for the two. This trend was observed in most of the utilities in the sample. Fourthly, we observed some key empirical observations from the results. We noted that inefficiencies exist in the provision of water in South Africa. Average efficiency scores reported for the sample were 0.447 (DEA), 0.662 (SFA), 0.681 (StoNED MM) and 0.529 (StoNED PSL). After grouping utilities into big and small categories, we observed little variation between the average scores from big and small utilities using StoNED. However, SFA reported that big utilities were more efficient (75.6% efficient on average) than small utilities (54.8% efficient on average), whereas DEA reported that small utilities were more efficient than big utilities.

Based on the performance of StoNED MM relative to the other techniques, we join other studies in the literature in arguing that this method is more appropriate for heterogeneous samples. StoNED MM controls heterogeneity well and leads to efficiency estimates with low standard deviations even in noisy scenarios. Where StoNED MM cannot be used, we argue that SFA is the next-best efficiency-analysis tool for noisy samples. Our study shows that the use of DEA – even when water utilities are grouped according to size – is not ideal in heterogeneous samples. Utilities in developing countries operate in very distinct environments with hugely distinct budgets. In consequence, benchmarking such utilities requires using techniques that can control for heterogeneity.

One of the weaknesses of our study is that it focused only on the second stage of StoNED, which estimates utility-specific efficiency scores. We recommend that future studies wishing to compare DEA, SFA and StoNED in the water sector should include other forms of distribution for SFA, e.g. exponential and truncated distributions. Our study only used the half-normal distribution. Regarding DEA, we recommend that future studies also test other forms of returns to scale. Our study focused only on an input-oriented VRS, because South African municipalities are different and are expected to be at different levels of production.

List of references

Agrell, P., Bogetoft, P. and Tind, J. 2005. DEA and dynamic yardstick competition in Scandinavian electricity distribution. *Journal of Productivity Analysis* 23**,** 173-201.

Aigner, D., Lovell, C. K. and Schmidt, P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics,* 6**,** 21-37.

Andor, M. and Hesse, F. 2014. The StoNED age: the departure into a new era of efficiency analysis? A monte carlo comparison of StoNED and the "oldies" (SFA and DEA). *Journal of Productivity Analysis,* 41**,** 85-109.

Aubert, C. and Reynaud, A. 2005. The impact of regulation on cost efficiency: an empirical analysis of Wisconsin water utilities. *Journal of Productivity Analysis,* 23**,** 383-409.

Banker, R. D. 1993. Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science,* 39**,** 1265–1273.

Baranzini, A., Faust, A.-K. and Maradan, D. 2011. Water supply: Costs and performance of water utilities: Evidence from Switzerland. 13th International Water Resources Association World Water Congress, 1-4 September 2008 Montpellier.

Battese, G. E. and Coelli, T. J. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics,* 20**,** 325-332.

Brettenny, W. and Sharp, G. 2016. Efficiency evaluation of urban and rural municipal water service authorities in South Africa: A data envelopment analysis approach. *Water SA,* 42**,** 11-19.

Byrnes, J., Crase, L., Dollery, B. and Villano, R. 2009. An analysis of the relative efficiency of wastewater utilities in non-metropolitan New South Wales and Victoria. *Australasian Journal of Regional Studies, The,* 15**,** 153.

Carvalho, P., Pedro, I. and Marques, R. C. 2015. The most efficient clusters of Brazilian water companies. *Water Policy,* 17**,** 902-917.

Charnes, A., Cooper, W. W. and Rhodes, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2**,** 429-444.

Cheng, X., Bjorndal, E., Lien, G. and Bjorndal, M. H. 2015. Productivity development for Norwegian electricity distribution companies 2004-2013. Discussion Paper. Norwegian School of Economics.

Coelli, T., Rao, D. S. P. and Battese, G. E. 2005. *An introduction to efficiency and productivity analysis*, Boston/Dordrecht/London, Kluwer Academic Publishers.

Coelli, T. J. 1995. Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis. *Journal or Productivity Analysis,* 6**,** 247-268.

Corton, M. L. and Berg, S. V. 2009. Benchmarking central American water utilities. *Utilities Policy*, 17, 267-275.

Da Cruz, N. F., Carvalho, P. and Marques, R. C. 2013. Disentangling the cost efficiency of jointly provided water and wastewater services. *Utilities Policy,* 24**,** 70-77.

Dai, X. and Kuosmanen, T. 2014. Best-practice benchmarking using clustering methods: Application to energy regulation. *Omega,* 42**,** 179-188.

De Witte, K. and Marques, R. C. 2012. Gaming in a benchmarking environment. A non-parametric analysis of benchmarking in the water sector. *Water Policy,* 14**,** 45 - 66.

Dios-Palomares, R., Ramos, A. and Roldán, J. 2002. A Monte Carlo study on the technical efficiency estimation in the stochastic frontier model. *Questiió,* 26**,** 443-459.

Dong, Y., Hamilton, R. and Tippett, M. 2014. Cost efficiency of the Chinese banking sector: A comparison of stochastic frontier analysis and data envelopment analysis. *Economic Modelling,* 36**,** 298-308.

Dunghana, B. R., Nuthall, P. L. and Nartea, G. V. 2004. Measuring the Economic Efficiency of Nepalese Rice Farms Using Data Envelopment Analysis. *The Australian Journal of Agricultural and Resource Economics,* 48**,** 347-369.

Eskelinen, J. and Kuosmanen, T. 2013. Intertemporal efficiency analysis of sales teams of a bank: Stochastic semi-nonparametric approach. *Journal of Banking and Finance,* 37**,** 5163-5175.

Estache, A. and Rossi, M. A. 2002. How different is the efficiency of public and private water companies in Asia? *The World Bank Economic Review,* 16**,** 139-148.

Estache, A., Rossi, M. A. and Ruzzier, C. A. 2004. The case for international coordination of electricity regulation: evidence from the measurement of efficiency in South America. *Journal of Regulatory Economics,* 25**,** 271-295.

Farrell, M. J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General),* 120**,** 253-290.

Filippini, M., Hrovatin, N. and Zorić, J. 2008. Cost efficiency of Slovenian water distribution utilities: an application of stochastic frontier methods. *Journal of Productivity Analysis,* 29**,** 169-182.

Forsund, F. R. and Kittelsen, S. A. C. 1998. Productivity development of Norwegian electricity distribution utilities. *Resource Economics,* 20**,** 207-224.

Garcia-Valiñas, M. A. and Muñiz, M. A. 2007. Is DEA useful in the regulation of water utilities? A dynamic efficiency evaluation (a dynamic efficiency evaluation of water utilities). *Applied Economics,* 39**,** 245-252.

Greenberg, R. and Nunamaker, T. 1987. A generalized multiple criteria model for control and evaluation of nonprofit organizations. *Financial Accountability and Management,* 3**,** 331-342.

Greene, W. H. 2008. *The econometric approach to efficiency analysis*, Oxford University Press.

Guerrini, A., Romano, G., Leardini, C. and Martini, M. 2015. Measuring the efficiency of wastewater services through data envelopment analysis. *Water Science and Technology,* 71**,** 1845-1851.

Herwartz, H. and Strumann, C. 2012. On the effect of prospective payment on local hospital competition in Germany. *Health Care Management Science,* 15**,** 48-62.

Hildreth, C. 1954. Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, 49, 598-619.

Hjalmarsson, L. and Veiderpass, A. 1992. Productivity in Swedish electricity retail distribution. *Scandinavian Journal of Economics,* 94**,** 193-205.

Johnson, A. L. and Kuosmanen, T. 2012. One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research,* 220**,** 559–570.

Johnson, A. L. and Kuosmanen, T. 2015. An introduction to CNLS and StoNED methods for efficiency analysis: Economic insights and computational aspects. *Benchmarking for Performance Evaluation.* Springer.

Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Economics,* 19**,** 233–238.

Ke, W., Wei, H., Jie, W. and Ying-Nan, L. 2014. Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. *Omega,* 44**,** 5-20.

Kirkpatrick, C., Parker, D. and Zhang, Y.-F. 2006. An empirical analysis of state and private-sector provision of water services in Africa. *The World Bank Economic Review,* 20**,** 143-163.

Korhonen, P. and Syrjänen, M. 2003. Evaluation of cost efficiency in Finnish electricity distribution. *Annals of Operations Research,* 121**,** 105-122.

Kumbhakar, S. C. and Lovell, C. A. K. 2000. *Stochastic Frontier Analysis,* Oxford University Press.

Kuosmanen, T. 2008. Representation theorem for convex nonparametric least squares. *Economic Journal,* 11**,** 308–325.

Kuosmanen, T. 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. *Energy Economics,* 34**,** 2189-2199.

Kuosmanen, T. and Fosgerau, M. 2009. Neoclassical versus frontier production models? Testing for the presence of inefficiencies in the regression residuals. *Scandinavian Journal of Economics,* 111**,** 317–333.

Kuosmanen, T. and Johnson, A. L. 2010. Data envelopment analysis as nonparametric least squares regression. *Operations Research* 58**,** 149–160.

Kuosmanen, T. and Kortelainen, M. 2012. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis,* 38**,** 11-28.

Kuosmanen, T., Saastamoinen, A. and Sipiläinen, T. 2013. What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods. *Energy Policy,* 61**,** 740 -750.

Lannier, A. L. and Porcher, S. 2014. Efficiency in the public and private French water utilities: prospects for benchmarking. *Applied Economics,* 46**,** 556-572.

Leleu, H. 2006. A linear programming framework for free disposal hull technologies and cost functions: primal and dual models. *European Journal of Operational Research,* 168**,** 340–344.

Lewis, H. F., Mallikarjun, S. and Sexton, T. R. 2013. Unoriented two-stage DEA: The case of the oscillating intermediate products. *European Journal of Operational Research,* 559**,** 529–539.

Li, H.-Z., Kopsakangas-Savolainen, M., Xiao, X.-Z., Tian, Z.-Z., Yang, X.-Y. and Wang, J.-L. 2016. Cost efficiency of electric grid utilities in China: A comparison of estimates from SFA–MLE, SFA–Bayes and StoNED–CNLS. *Energy Economics,* 55**,** 272-283.

Meeusen, W. and Van Den Broeck, J. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review***,** 435-444.

Mehta, D., Meera, M. and Anitha, I. 2013. A Review of Performance Benchmarking Urban Water Supply and Sanitation. CEPT University.

Mekaroonreung, M. and Johnson, A. L. 2012. Estimating the shadow prices of SO2 and NOx for US coal power plants: a convex nonparametric least squares approach. *Energy Economics,* 34**,** 723-732.

Olson, J. A., Schmidt, P. and Waldman, D. M. 1980. A Monte Carlo study of estimators of stochastic frontier production functions. *Journal of Economics,* 13**,** 67–82.

Ren, C., Li, R. and Guo, P. 2017. Two-stage DEA analysis of water resource use efficiency. *Sustainability* 9.

Ruiters, C. 2013. Funding models for financing water infrastructure in South Africa: framework and critical analysis of alternatives. *Water SA*, 39**,** 313-326.

Sabouhi Sabouni, M. and Kenari, R. E. 2014. Measuring technological gap ratio of wheat production using StoNED approach to metafrontier. *International Journal of Applied Operational Research,* 4.

Sharma, K. R., Leung, P. and Zaleski, H. M. 1999. Technical, allocative and economic efficiencies in Swine production in Hawaii: a comparison of parametric and nonparametric approaches. *Agricultural Economics* 20**,** 23-35.

Simar, L. and Wilson, P. W. 2008. Statistical inference in nonparametric frontier models: recent developments and perspectives. *The Measurement of Productive Efficiency and Productivity Growth***,** 421-521.

Souza, G. S., Faria, R. C. and Moreira, T. B. S. 2007. Estimating the relative efficiency of Brazilian publicly and privately-owned water utilities: a stochastic cost frontier approach. *Journal of the American Water Resources Association* 45**,** 1237-1244.

Timmer, C. P. 1971. Using a probabilistic frontier production function to measure technical efficiency. *Journal of Political Economy*, 79, 776-794.

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. Working Paper. Social Science Research Network (SSRN).

Vidoli, F. and Ferrara, G. 2015. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. *Empirical Economics,* 49**,** 641-658.

Vishwakarma, A. and Kulshrestha, M. 2010. Stochastic production frontier analysis of water supply utility of urban cities in the State of Madhya Pradesh, India. *International Journal of Environmental Sciences,* 1**,** 357.

Von Hirschhausen, C., Cullmann, A. and Kappeler, A. 2006. Efficiency analysis of German electricity distribution utilities–non-parametric and parametric tests. *Applied Economics,* 38**,** 2553-2566.

Wabiri, N., Shisana, O., Zuma, K. and Freeman, J. 2016. Assessing the spatial nonstationarity in relationship between local patterns of HIV infections and the covariates in South Africa: a geographically weighted regression analysis. *Spatial and Spatio-temporal Epidemiology,* 16**,** 88-99.

Westhuizen, G. and Dollery, B. 2009. Efficiency measurement of basic service delivery at South African district and local municipalities. *Journal for Transdisciplinary Research in Southern Africa,* 5**,** 162-174.

Winsten, C. 1957. Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society,* 120, 282-284.

Yang, Z. J. 2006. A two-stage DEA model to evaluate the overall performance of Canadian life and health insurance companies. *Mathematical and Computer Modelling,* 43**,** 910-919.

Zhu, J. 2014. *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets*, Springer.

Zschill, M. and Walter, M. 2012. The performance of German water utilities: a (semi)-parametric analysis. *Applied Economics,* 44**,** 3749-3764.

Appendix 2.1: Efficiency scores for all water utilities

Table A1: Efficiency scores for all utilities using the three methods

	Category[7]	StoNED MM	StoNED PSL	SFA	DEA
Nelson Mandela Bay	A	0.653	0.455	0.867	0.309
Buffalo City	A	0.717	0.589	0.803	0.517
Camdeboo	B3	0.644	0.440	0.647	0.272
Blue Crane Route	B3	0.723	0.604	0.580	0.385
Ikwezi	B3	0.737	0.641	0.443	1.000
Makana	B2	0.714	0.583	0.647	0.265
Ndlambe	B3	0.742	0.655	0.582	0.300
Sunday's River Valley	B3	0.663	0.474	0.609	0.320
Baviaans	B3	0.718	0.591	0.489	0.692
Kouga	B3	0.722	0.602	0.642	0.219
Kou-kamma	B3	0.688	0.523	0.593	0.335
Amathole	C	0.762	0.724	0.508	1.000
Inxuba Yethemba	B3	0.443	0.018	0.350	1.000
Tsolwana	B3	0.696	0.541	0.564	0.398
Chris Hani	C	0.557	0.126	0.382	0.960
Gariep	B3	0.447	0.176	0.733	0.325
Joe Gqabi	C	0.724	0.599	0.223	1.000
Mangaung	A	0.722	0.602	0.789	0.546
Letsemeng	B3	0.730	0.622	0.578	0.568
Kopanong	B3	0.762	0.724	0.550	0.511
Tswelopele	B3	0.588	0.349	0.691	0.249
Matjhabeng	B1	0.710	0.572	0.751	0.222
Setsoto	B3	0.649	0.449	0.706	0.275
Dihlabeng	B2	0.608	0.379	0.768	0.137
Mantsopa	B3	0.680	0.507	0.623	0.411
Moqhaka	B2	0.630	0.416	0.767	0.129
Ngwathe	B3	0.624	0.406	0.749	0.151
Metsimaholo	B2	0.750	0.679	0.634	0.369
Mafube	B3	0.620	0.398	0.686	0.289
City of Johannesburg	A	0.735	0.634	0.843	1.000
City of Tshwane	A	0.700	0.550	0.883	0.666
Ekurhuleni Metro	A	0.718	0.591	0.863	0.823
Emfuleni	B1	0.717	0.589	0.821	0.542
Midvaal	B2	0.761	0.716	0.577	0.405
Lesedi	B3	0.723	0.604	0.612	0.331
Mogale City	B1	0.748	0.671	0.622	0.644
Randfontein	B2	0.704	0.558	0.720	0.178
Westonaria	B2	0.759	0.716	0.580	0.398
Merafong City	B2	0.762	0.723	0.643	0.347

[7] Category A is metropolitan municipalities, category C is district municipalities, and category B is local municipalities. Local municipalities are further classified into category B1 (municipality with a large town or city as its urban core), category B2 (municipality with a medium town as its urban core), category B3 (municipality with a small town as its urban core), and category B4 (municipality with no urban core). In the context of this study, categories A, C, B1 and B2 are large water utilities, while categories B3 and B4 are small water utilities.

eThekwini Metro	A	0.728	0.618	0.850	0.996
Ugu	C	0.720	0.596	0.765	0.518
The Msunduzi	B1	0.724	0.607	0.724	0.494
Umgungundlovu	C	0.663	0.474	0.812	0.331
Uthukela	C	0.686	0.521	0.819	0.207
Newcastle	B1	0.743	0.657	0.626	0.607
Amajuba	C	0.623	0.404	0.805	0.230
City of uMhlathuze	B1	0.747	0.670	0.656	0.851
Uthungulu	C	0.700	0.550	0.791	0.569
iLembe	C	0.737	0.640	0.696	0.563
Mopani	C	0.594	0.357	0.896	0.133
Vhembe	C	0.701	0.553	0.828	0.494
Polokwane	B1	0.680	0.507	0.792	0.212
Capricorn	C	0.641	0.435	0.845	0.196
Lephalale	B3	0.412	0.144	0.843	0.104
Mogalakwena	B2	0.746	0.667	0.504	0.526
Msukaligwa	B2	0.703	0.556	0.643	0.252
Govan Mbeki	B1	0.710	0.573	0.739	0.238
Emalahleni (MP)	B1	0.644	0.441	0.815	0.133
Steve Tshwete	B1	0.679	0.507	0.733	0.178
Dr J S Moroka	B4	0.604	0.373	0.810	0.137
Mbombela	B1	0.639	0.431	0.804	0.201
Umjindi	B3	0.695	0.540	0.638	0.245
Richtersveld	B3	0.712	0.577	0.461	0.988
Nama Khoi	B3	0.755	0.693	0.531	0.382
Hantam	B3	0.722	0.602	0.533	0.574
Karoo Hoogland	B3	0.628	0.412	0.535	0.976
Khai-Ma	B3	0.761	0.717	0.382	1.000
Umsobomvu	B3	0.729	0.616	0.361	0.911
Emthanjeni	B3	0.632	0.419	0.663	0.285
Thembelihle	B3	0.700	0.550	0.482	0.872
Siyathemba	B3	0.729	0.620	0.464	0.756
!Kai! Garib	B3	0.693	0.535	0.582	0.398
//Khara Hais	B2	0.721	0.599	0.577	0.961
!Kheis	B3	0.700	0.550	0.578	1.000
Tsantsabane	B3	0.664	0.476	0.532	0.708
Sol Plaatje	B1	0.746	0.666	0.650	0.346
Dikgatlong	B3	0.755	0.694	0.365	0.987
Gamagara	B3	0.715	0.585	0.593	0.375
Rustenburg	B1	0.751	0.681	0.726	0.715
Mafikeng	B2	0.554	0.301	0.834	0.095
Ramotshere Moiloa	B3	0.396	0.131	0.861	0.116
Tlokwe	B1	0.638	0.429	0.758	0.154
City of Cape Town	A	0.720	0.597	0.860	0.852
Matzikama	B3	0.621	0.401	0.696	0.195
Cederberg	B3	0.547	0.290	0.712	0.226
Bergrivier	B3	0.675	0.497	0.642	0.241
Saldanha Bay	B2	0.743	0.656	0.621	0.363
Swartland	B3	0.706	0.563	0.643	0.261
Witzenberg	B3	0.599	0.366	0.746	0.147
Drakenstein	B1	0.631	0.418	0.789	0.133
Stellenbosch	B1	0.718	0.593	0.678	0.224
Breede Valley	B2	0.632	0.418	0.741	0.192
Langeberg	B3	0.690	0.528	0.679	0.206

Theewaterskloof	B3	0.713	0.579	0.664	0.218
Overstrand	B2	0.761	0.722	0.552	0.432
Cape Agulhas	B3	0.709	0.571	0.571	0.400
Swellendam	B3	0.713	0.581	0.516	0.485
Mossel Bay	B2	0.725	0.610	0.835	0.327
George	B1	0.720	0.597	0.693	0.201
Bitou	B3	0.739	0.645	0.548	0.388
Knysna	B2	0.753	0.687	0.569	0.316
Beaufort West	B3	0.396	0.131	0.777	0.330

Chapter 3: An empirical examination of reducing status quo bias in heterogeneous populations: evidence from the South African water sector

Abstract

Choice experiments typically include a status quo option, which often describes the current scenario. This is to secure the validity and applicability of choice experiments. People have a propensity to choose what they are familiar with, despite being presented with alternatives that seem better (i.e. the ‘status quo effect’). Various experiments have reliably demonstrated this effect. The tendency to prefer the current scenario disproportionally does not mimic real-life preferences; therefore, status quo bias is undesirable. In a split sample framework, we test for the effects of reducing status quo bias by considering a heterogeneous sample. We use generalised mixed logit models to carry out the tests.

Keywords: choice experiments, heterogeneous, generalised mixed logit, status quo bias.

3.1. Introduction

Choice experiments typically include a 'status quo' (SQ) option that describes the current situation. Respondents who participate in a choice experiment (CE) are generally asked to choose several times between hypothetical options and an SQ option. Thus, the SQ option is an 'opt-out' option to the offered alternatives in the choice sets. To secure the validity and applicability of CE studies, the SQ option should mimic real-life status quo choices. The SQ option essentially avoids the undesired effects linked to forced choices or ascertains whether respondents are satisfied with current packages. In certain circumstances, when respondents choose the SQ option it signifies their level of satisfaction with the current packages (Lanz and Provins, 2015).

However, a large body of literature argues that the SQ option leads to the problem of SQ bias. This issue was first identified by Samuelson and Zeckhauser (1988), who described it as the respondents' disproportionate tendency to choose a default option. In the literature, it is often argued that respondents usually choose the SQ option to the extent that no real trade-offs are made between the given attributes. If this happens, there is a risk of biased empirical results, which may lead to incorrect practical inferences. SQ bias has been argued to emerge from issues such as task complexity, and the reality that respondents normally prefer the SQ they are currently experiencing, compared to designed options whose utility is hypothetical (Meyerhoff and Liebe, 2009; Scarpa et al., 2007). The existence of such factors makes SQ bias inevitable in CE surveys.

An SQ bias problem is undesirable because it leads to an underestimation of the welfare changes of a proposed policy change. In addition, as the SQ alternative is preferred too frequently, information is reduced regarding the relative values of different types of attributes associated with a policy change. This consequently reduces the effectiveness of stated preference surveys to elicit and identify preferences (Bonnichsen and Ladenburg, 2015). The status quo bias can have serious effects on empirical results as well as on policy choices.

Several attempts have been made in the literature to address SQ bias and the effects of imposing SQs on a heterogeneous sample. A significant number of CE studies impose hypothetical baselines as SQs. Some completely exclude the SQ option in the choice sets, while others replace it with a 'none' choice as an alternative opt-out option. Although excluding the SQ option may be a solution, in some instances including the SQ option is essential, because SQ choices may reflect genuine preferences for current packages (see Lanz and Provins, 2015). Another possible solution involves the use of individual-specific SQ options; in such cases, the SQ option is left blank, and respondents determine their own perceived SQs (Campbell et al., 2008; Hess and Rose, 2009).

Status quo bias studies tend to assume that treatment choice is on homogenous populations. Most of the studies test for bias without considering whether there might be a heterogeneous population. The implication of a heterogeneous population may be that the alternative representing the status quo situation does not resonate with some subpopulations. To address this, some studies divide the population into subgroups, and then present different status quos accordingly. In developing countries with high levels of inequality, such as South Africa and Brazil, the optimal treatment rule may be to divide the population into subpopulations, each of whose members share the same current situation; and then to conduct an experiment in which each subpopulation faces a status quo alternative with which they are familiar. This is an optimal method to secure the validity and applicability of the experiments. However, empirical studies show that when presented with alternatives that seem superior, CE participants tend to prefer what they already have. Considering that splitting the sample initially is meant primarily to ensure that each sub-sample is presented with an SQ that they already know, it is plausible that this trend would remain even in a case in which the population is divided, meaning that the status quo bias problem will persist. In this study, therefore, we are interested in assessing whether presenting a partially relevant SQ[8] could reduce the SQ bias problem.

Our paper tests for the effects of introducing a partially relevant status quo aimed at reducing SQ bias in a CE eliciting households' preference for water service packages. We test these effects by comparing the utility functions and marginal willingness to pay estimates of two subpopulations (i.e. suburbs and townships). Each subpopulation is presented with two experiments: one containing a relevant SQ[9], and another containing a partially relevant SQ. The focus of the paper is on testing whether participants' likelihood of choosing the SQ is driven by the relevance or partial relevance of the SQ option. We make use of data from

[8] A 'partially relevant status quo' in this study implies an SQ option that does not completely capture the current scenario for the participants in each subsample. Ideally, it is an SQ option that does not resonate with the water service package currently received by most participants in each subpopulation.
[9] A 'relevant status quo' in this study implies an SQ option that captures the current scenario for most participants in each subsample. It shows the water service package currently received by most participants in each subpopulation.

experiments on households' satisfaction with their municipal water service packages in Durban, a city in South Africa.

To the best of our knowledge, this is the first paper to test for the effects of reducing status quo bias by dividing population (based on economic segmentation) into two subpopulations (i.e. wealth and poverty) and presenting each subpopulation with two different choice experiments. In the first treatment, respondents in each subpopulation are presented with a series of choice sets, each with a status quo choice that resonates with them (i.e. it is relevant). In another treatment, each subpopulation is presented with series of choice sets, each with a status quo choice that does not fully reflect their current situation. In both cases, participants are presented with a choice between two alternative hypothetical water service packages: an SQ alternative, representing their current or perceived current water service package, and an 'opt out' option.

The study uses a novel approach in that the sample is split into two strata, with each stratum presented with two choice experiments. The fundamental difference between the two experiments is the SQ option. Results from the two experiments in each stratum are then compared, against each other and against results from the other stratum. By doing this, we can detect the impact and magnitude of the SQ bias on the utility functions and estimated MWTP figures, and the magnitude of this impact. It is common practice for studies that compare estimates across experiments to make comparisons based on the statistical significance, sign and absolute value of the coefficient (see Bateman et al., 2009; Orzechowski et al., 2005; Patterson et al., 2017; Vriens et al., 1998). In these studies, the numbers of statistically significant coefficients in each experiment are compared. The absolute value of the coefficient is used to measure the magnitude of impact a change in an attribute has on the respondents' utility, while the sign of the coefficient gives the direction of the impact. Our study follows this approach in comparing experiments.

The rest of the paper is organised into eight sections. Section 3.2 reviews the literature on SQ bias. Section 3.3 gives an overview of the South African water sector. Section 3.4 discusses the experimental design of the study. Section 3.5 discusses modelling approaches. Section 3.6 presents the experimental data. Section 3.7 presents and discusses the empirical findings. Section 3.8 concludes the study.

3.2. SQ bias theories

Since the pioneering work of Samuelson and Zeckhauser (1988), SQ bias has received much attention in the literature. Various theories were developed to explain the main causes of SQ bias. Most of these theories fall within the domain of psychology. The most notable theories to explain the origin and drivers of SQ bias include the loss aversion theory, the inertia theory, the decision avoidance theory, and the incomplete preferences theory. All these theories explain the psychological reasons for SQ bias. After the establishment of the SQ bias theories, empirical studies emerged in the literature validating the drivers of SQ bias, as identified earlier in the theoretical literature. This section discusses some of the theories on SQ bias. The empirical literature on SQ bias is also used to explain and discuss the SQ bias theories.

Proposed by Tversky and Kahneman (1991), the loss aversion theory suggests that the SQ option serves as a reference point, and the losses relative to this reference point have greater impact on preferences than gains do. The theory argues that individuals keen to avoid losses have a strong tendency to remain in the SQ, because the disadvantages of leaving it appear larger than the advantages. This argument also holds true when respondents are not sure of the good, or when they face complexity in understanding the given choices. In such cases, respondents choose the SQ, whose utility they currently experience as minimising losses linked to hypothetical utilities from experimentally designed options. The empirical literature identifies various instances in which respondents choose the SQ to avoid loss. The most notable instances include when respondents are faced with complex choice sets, and when too many choice sets, attributes and levels are included (see Meyerhoff and Liebe, 2009; Moon, 2000; Oehlmann et al., 2017; Scarpa et al., 2007). In such cases, respondents tend to choose the default option whose utility they currently experience.

The inertia theory (Ritov and Baron, 1992; Schweitzer, 1994) argues that keeping the SQ requires only inaction, and respondents are known to have some preference for inaction. In this theory, respondents are believed to have an attachment to and persistence in the use of the status quo, even in the presence of better alternatives and/or incentives to change. This is normally the case where the perceived value of change is low, which means respondents may show strong resistance to change. From a rational decision perspective, inertia would be due to loyalty to respondents’ current status, or as a result of the respondents trying to minimise their losses. Inertia due to loyalty to the current status is explained in several empirical studies, among them Scarpa et al. (2007) and Dubé et al. (2010). The argument put forward in these studies is that when respondents are loyal to the current status, they are not likely to choose hypothetical alternatives provided in choice sets. In such cases, they tend to choose the SQ option ahead of any other available option.

In the decision-avoidance theory, Anderson (2003) argues that when expected to decide between many options, respondents usually choose not to decide. This arises when respondents avoid making choices by postponing, or through choosing an easy way that may involve no action or no change (Anderson, 2003). The decision avoidance theory builds on earlier postulates in Beattie et al. (1994), which state that respondents desire to make or avoid decisions independent of any consequence that this may cause. Depending on the context, respondents are therefore assumed to be either decision seeking or decision averse. In explaining the decision avoidance theory, Anderson (2003) acknowledges that marked preferences for avoidant options have been discovered in diverse areas of the literature. The more specific instances include when respondents generally prefer no change (status quo bias, Samuelson and Zeckhauser, 1988), no action (omission bias, Ritov and Baron, 1992; inaction inertia, Tykocinski et al., 1995), or delay (choice deferral, Dhar, 1996).

In the incomplete preference theory, Mandler (2004) argues that respondents who have an unchanging but incomplete preference usually prefer the SQ. The theory suggests that respondents with incomplete preferences choose to maintain the SQ when they follow the simple rule of refusing to trade their endowment for unranked bundles. Such respondents do this while waiting to be offered an alternative that is ranked as superior. This process of refusing to trade their endowment for unranked bundles respects respondents' interests, and respondents will not be led to outcomes they judge to be inferior (Mandler, 2004). Eventually, respondents persistently maintain their SQ. The incomplete preference theory is built on findings from Diamond and Hausman (1994) that respondents with incomplete preferences do not make preference judgements between certain pairs of bundles. This is more prevalent when intangible goods are involved; respondents may not form a definitive view of the monetary value of an incremental unit of a good.

Subsequent to the establishment of these SQ theories, several empirical studies have been conducted on the drivers of SQ bias. In the empirical literature, loyalty to the SQ is commonly identified as a key determinant of SQ bias (see Ren, 2014; Scarpa et al., 2007; Dubé et al., 2010). When respondents are loyal to the SQ, they tend to stick with the SQ option rather than choosing one of the hypothetical alternatives provided. Marsh et al. (2011) identify respondents' knowledge of the good as another key driver of SQ bias. Respondents who are not completely aware of the good try to avoid the risk associated with choosing hypothetical alternatives. In such situations, respondents maintain the SQ option, as argued in the risk aversion theory (Tversky and Kahneman, 1991) and in the incomplete preference theory (Mandler, 2004). Other notable drivers of SQ bias identified in the empirical literature include protest attitude, attitude towards the good, perceived choice task complexity, number of attributes and levels, and number of choice profiles (see Boxall et al., 2009; Meyerhoff and Liebe, 2009; Oehlmann et al., 2017; Moon, 2000; Ren, 2014; Zhang and Adamowicz, 2011).

Notably – because of the various reasons discussed earlier in this section – SQ bias is inevitable in CEs, though efforts are made to address the problem. Such efforts include omitting the SQ option (see Hensher et al., 2005; Saldías et al., 2016), and using individual-specific SQ options (see Campbell et al., 2008; Hess and Rose, 2009; Marsh et al., 2011). Our study examines whether all these efforts are necessary. It does so by examining whether the presence of SQ bias affects empirical results and welfare measures. One study that has some similarities to our study is Boxall et al. (2009), which examines the impact of SQ bias on welfare measures. In the next section, we discuss the experimental design used in this study.

3.3. The South African water sector

South Africa has a complex water governance environment, with both considerable successes in and significant ongoing challenges to achieving sustainable, adequate and equitable water access (Beck et al., 2016), both of which impact water service levels. Historically, water supply and distribution schemes in South Africa were created and managed during the colonial and apartheid eras to serve predominantly white populations. Investment in aspects such as pipes, dams and other water-related infrastructure were differentially applied in different areas during apartheid, with homelands, townships and informal settlements receiving significantly less funding, and generally a lower quality of water service (Goldin, 2010). This led to access to water services in South Africa being highly differentiated by race and income, as well as an extremely fragmented water management system (Herrfahrdt-Pähle, 2010) and undemocratic participatory engagement, resulting in challenges that are still experienced today.

The South African water sector is an ideal subject for a case study, due to the fragmentation that has resulted in a differentiated water service. We chose to conduct our experiments in

Durban because of its unique characteristics. Although Durban is South Africa's third-biggest city, it is unique in that it is the only city that has some township/informal settlement, some suburban, and some rural components. This last is uncommon in cities and towns in South Africa. The profile of the city implies that the level of water service is not uniform, which makes it ideal for our experiment. Water provision in each area of South Africa is the responsibility of the area municipality, which also acts as a water utility. They are commonly referred to as Water Service Authorities (WSAs). WSAs are responsible for the provision of water services within their area of jurisdiction. A WSA may carry out the functions of a water services provider (WSP) itself, or it may sub-contract the delivery to a third party. For Durban, the eThekwini Metropolitan municipality is the WSA, as well as being a WSP (see the map of eThekwini below).

Figure 3.1: Map of the eThekwini Metropolitan Municipality

Source: Local Government Handbook (2012)

Durban is in the eastern part of South Africa, in the province of KwaZulu-Natal (KZN). The municipality has a population of about 3.6 million people (eThekwini Municipality, 2015). Figure 3.1 shows suburban, township and rural areas in the municipality. Due to the apartheid history of segregation, there are townships for black South Africans and townships for Indian South Africans. The former includes areas such as KwaMashu, Inanda, Clermont and Umlazi, while the latter includes Chatsworth and Phoenix. Rural areas in the municipality include Umbumbulu, and areas such as Umhlanga, Verulam and Westville are suburbs. Township areas are densely populated relative to suburban areas. For example, the 2011 National Census shows that the total population in KwaMashu was 175 663 people (50 683 households), Inanda had 178 418 people (44 736 households), Umlazi had 404 811 people (104 914 households) and Chatsworth had 196 580 people (54 497 households). These numbers are much larger than those reported for suburban areas such as Umhlanga (24 238 people and 9 256 households), Verulam (37 273 people and 10 896 households) and Westville (30 508 people and 8 814 households)[10].

The minimum standard of water service provided by the municipality is a community tap designed to serve a community where the maximum distance from the furthest dwelling should not be greater than 200 metres (eThekwini Municipality, 2014). However, such facilities are mostly found in informal settlements (e.g. Bhambayi, close to Inanda township, has community taps). The 2011 National Census revealed that while 60.2% of households in the municipality had access to piped water services inside their dwellings, about 17% obtained potable water from community taps. Where households do not access potable water from inside the dwelling or from a community tap, they receive it through a tap in the yard, a phenomenon common mostly in townships. Based on statistics from the 2011 National Census, the percentage of suburban households accessing piped water inside the dwelling was 99% for Umhlanga, 94% for Westville and 81% for Verulam. The same statistics in the townships were 50% in Umlazi, 35% in Inanda and 44% in KwaMashu.

The municipality uses an increasing block tariff (IBT) pricing structure to charge for water services. IBT is a volumetric tariff system proportional to consumption (i.e. it increases with consumption). IBT is used in South African municipalities as a measure to address the problems of unequal income distribution, by providing fair access to water in a country that

[10] These statistics are based on the 2011 National Census data published by Statistics South Africa. Our sample is a true representation of the population. Descriptive statistics for our sample are presented later in this study in section 3.6.2.

has huge income disparities (Banerjee et al., 2010; Jansen and Schulz, 2006; Muller, 2008). In line with the country's Free Basic Water Policy of 2002, which states that indigent households should receive at least 6 000 litres of free water per month, the eThekwini municipality provides 9 000 litres free to each of its indigent households. To determine who is indigent, the municipality uses a property value-based targeting approach in which households occupying properties valued at less than R250 000 (i.e. $17 483) are considered indigent and qualify for free basic water services. Such property values are predominantly found in townships, informal settlements and rural areas; hence, most recipients of free basic water services are from these areas.

Due to budget constraints, we could not collect data from all areas of the municipality as illustrated in Figure 3.1. Various areas were selected based on their population size, demographic representation, geographical location and economic status. The suburban areas surveyed were Morningside, Musgrave and Overport in central Durban, as well as La Lucia, Umhlanga, and Verulam in northern Durban. For townships, data was collected from Inanda, Ntuzuma and Phoenix in northern Durban, as well as Chesterville, Chatsworth and Umlazi in southern Durban. Respondents from informal settlements in Bhambayi and Umlazi were also surveyed. Rural households were surveyed in Umbumbulu.

The sampled areas were selected for three main reasons. First, they represent the suburbs, townships, informal settlements and rural areas that are the true spatial segmentations of the study area. Second, our sampled areas represent areas where different racial groups reside. For example, Phoenix and Chatsworth represent townships where Indian South Africans reside predominantly, while the other townships represent areas that are predominantly occupied by black South Africans. These two racial groups are the most dominant in townships and other low-income areas. Third, the sampled areas represent the northern, central and southern parts that form the geographic demarcations of Durban. An exploration of these diverse areas gave us a clear picture of the range of water service packages in the municipality. Since access to water in townships is very similar to access in informal settlements and rural areas, we grouped these segments together into one sub-sample.

3.4. Experimental design

CEs are conducted to determine the independent influence of different attributes on the choices that are observed to be made by sampled respondents. The first step in CE modelling is selecting relevant and realistic attributes. Attributes are translated features and characteristics that show the objective properties of a commodity. These can be deduced from literature reviews, focus group discussions, pilot studies and expert consultations. After attributes are selected, feasible, realistic, and non-linearly spaced levels that span the range of respondents' preference maps are assigned to each attribute (Hanley et al., 2001; Louviere et al., 2000). The attributes and levels are then experimentally designed into various choice profiles. Hensher et al. (2015) describe experimental design as the effect on a response variable following the specialised manipulation of the levels of one or more other variables. This section discusses the attributes and levels used in the study, as well as how they are designed into choice profiles.

3.4.1. Attributes and levels

This study tests the impact of SQ bias on empirical results using a case of household preference for water service packages in one of South Africa's metropolitan municipalities. Water provision is deemed an ideal case study, because the diversity and complexity found within South African municipalities makes it difficult to come up with an SQ that applies to the whole population. Although the quality of water received by different households in different areas of the same municipality may be similar, the total package of the water service may differ in terms of location of tap, reliability of supply, water pressure, and monthly household water bill. To determine the attributes and levels for the study, two focus groups were established. One focus group contained residents from the suburbs, while the other had residents from townships. Through focus group discussions and a review of the literature, it emerged that households are concerned about the position of the tap, the reliability of the supply, and the pressure, quality and cost of water services. Collectively these features make up a typical water service package; and are adopted as attributes in this study.

Subsequently, the literature review and the focus groups were used to come up with levels for each attribute. Furthermore, a panel of experts assembled from the water stakeholders was used to discuss and refine the attributes and levels. The final attributes and levels[11] that emerged from the panel discussions are presented in Table 3.1.

Table 3.1: Attributes and levels used in the study

Attribute	Description	Attribute Levels
Piped water 	Access to piped or tap water in the dwelling, on-site or off-site. This shows how piped water is delivered to households.	**Level 1:** Inside dwelling **Level 2:** In yard **Level 3:** Community tap: less than 200m from dwelling **Level 4:** Community tap: more than 200m from dwelling **Level 5:** No access to piped water
Reliability of supply 	Whether the household had any interruption in piped water supply in the last month.	**Level 1:** Yes **Level 2:** No
Water pressure 	Water pressure is a measure of the force that gets the water through our mains and into your pipes.	**Level 1:** High water pressure **Level 2:** Low water pressure
Water quality 	The chemical, physical, and biological characteristics of water, usually in respect to its suitability for an intended purpose. Colour in water is a concern of water quality for aesthetic reason. Taste and odour are human perceptions of water quality.	**Level 1:** Safe to drink **Level 2:** Has colour **Level 3:** Has a taste **Level 4:** Has a smell
Cost 	Cost per month.	**Level 1:** R120 **Level 2:** R220 **Level 3:** R400 **Level 4:** R680 **Level 5:** R980

[11] To develop levels for the cost attribute, the current domestic water tariff structure published by the municipality was used. The water tariff structure has five successive and increasing blocks, and the average costs in each block were used as levels for the cost attribute. For the position of the tap (indicated here as the piped water attribute), the literature shows that the main access points for piped water in the municipality are inside the dwelling, in the yard, or community taps.

The attributes and levels presented in Table 3.1 were used to generate the choice profiles used in the study. As mentioned earlier, various classes of experimental design exist in the literature. Each of these classes has its own merits and demerits. An analysis of each of the classes of experimental design is essential before one adopts a class for designing choice profiles. In the next sub-section, the study presents a brief discussion of the classes of experimental design that are commonly used in the CE literature. The sub-section also explains the class of design adopted in this study, its advantages over the other classes of designs, and how the attributes and levels presented in Table 3.1 were designed into choice profiles used to collect stated preference data for the study.

3.4.2. Choice experiment design

The most common choice experiment designs are the full factorial, orthogonal, and efficient designs. A full factorial design contains choice situations spanning all possible attribute level combinations, and will result in the maximum number of choice situations that can be produced without allowing repeated choice situations (Rose and Bliemer, 2009). Full factorial designs were used predominantly in the early studies (see Addelman, 1962; Hahn and Shapiro, 1966; Hanley et al., 2001; Holland and Cravens, 1973; Street et al., 2005). They are simple, quick and easy to construct, because they are mostly general and do not require sophisticated software coding to design (Rose and Bliemer, 2009; Street et al., 2005).

However, because all the possible treatment combinations in full factorial designs are not designed using a statistical package, orthogonal designs became preferred to full factorial designs (see Kanyoka et al., 2008; Rose et al., 2008; Snowball et al., 2008; Street et al., 2005). An orthogonal design relates to the correlation structure between the attributes of the experiment (Louviere et al., 2000; Rose and Bliemer, 2009). When the experimental design is orthogonal, the attributes in the experiment are statistically independent of each other. This theoretically allows for an independent determination of each attribute's influence on the observed choices. Orthogonal designs produce fewer confounding estimates of the population parameters, due to the enforced statistical independence between the attributes contained within the design (Ferrini and Scarpa, 2007; Rose and Bliemer, 2009). Nevertheless, orthogonal designs are criticised for their inapplicability to non-linear models such as discrete choice models, and because data loses orthogonality when researchers attempt to maintain orthogonality (see Bliemer and Rose, 2006; Bliemer et al., 2008; Bliemer et al., 2017).

Efficient designs address most of the shortcomings encountered in other designs. Efficient designs produce more robust data, which leads to more reliable parameter estimates with even lower sample sizes and smaller widths in the confidence intervals. Using efficient designs requires some knowledge of prior parameters. Where prior parameters are not known, designers can draw them using the Bayesian parameter distributions (Bliemer et al., 2008). Bayesian parameter distributions are less sensitive to misspecification of priors, because they assume prior parameter values to be approximately known and randomly distributed. When Bayesian parameter distributions are used to draw prior parameters for a D-*error* statistic, the experimental design becomes a Bayesian D-*error* design (D_b-*efficient*). This is represented as:

$$D_b - error = \int_{\tilde{\beta}} \det\left(\Omega_1(X, \tilde{\beta})\right)^{1/K} \phi(\tilde{\beta} \,|\, \theta) d\tilde{\beta}. \tag{3.1}$$

In equation 3.1, D_b is Bayesian design, Ω_1 is the asymptotic variance-covariance (AVC) matrix of the design, X is the experimental design, $\tilde{\beta}$ represents prior parameters, and K is the number of parameters to be estimated. The Bayesian D-*error* design is commonly used in efficient designs where the true population parameters are not known with certainty. Rose and Bliemer (2009) identify this as an improvement to most early studies, which assumed all prior parameters to be zero (i.e. $D_z - error = \det(\Omega_1(X, 0))^{1/K}$). Furthermore, it is also an improvement to studies which assumed non-zero parameters that were known with certainty (i.e. $D_p - error = \det(\Omega_1(X, \beta))^{1/K}$).

Due to the several advantages of efficient designs relative to other designs, this study uses an efficient design to create the hypothetical choice sets used to test for status quo bias. Following Rose and Bliemer (2009), the parameter estimates are drawn using a normally distributed Bayesian D-*efficiency* parameter. To determine the number of draws for Bayesian priors, we use the Gaussian method. The rule of thumb for the absolute minimum Gaussian quadrature is 2^K, where K is the number of Bayesian priors.

Our study adopted an efficient design and used the maximum possible Gaussian draws (i.e. 32 draws). The maximum possible draws allow for a more efficient design. Using the attributes and levels presented earlier in Table 3.1, six choice sets of two profiles each were generated by means of a normally distributed Bayesian D-*efficiency* method. Two experiments were designed for each of the two sub-samples. The first experiment contains an SQ that resonates with each sub-sample, while the second experiment contains an SQ that is partially relevant to each sub-sample. Each experiment has six choice sets, each of which consists of four options: an SQ option, Options 1 and 2, and a 'None' option. The 'None' option was included as a way of giving respondents room to opt out. The "none" option was presented to respondents as a protest bid. Including an opt-out option is essential, as it gives respondents the chance not to select any of the given alternatives. If the opt-out option is not given, respondents are forced to choose between their SQ and the hypothetical, experimentally designed options. If a respondent prefers neither the SQ nor the experimentally designed alternatives, incorrect inferences will be deduced if an opt-out option has not been included. An example of a choice set with an SQ relevant to the township stratum is shown in Table 3.2.

Table 3.2: Example of a choice set with an SQ relevant to the township sub-sample

	STATUS QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	In yard	Inside dwelling	
Reliability	No	Yes	No	
Water pressure	Low pressure	Low pressure	High pressure	
Water quality	Safe to Drink	Has colour	Has a smell	
Monthly cost	R0	R120	R400	
I WOULD CHOOSE:	☐	☐	☐	☐

The SQ presented in Table 3.2 was the baseline that resonated with most of the township dwellers. As argued first in Samuelson and Zeckhauser (1988) and in many subsequent studies, we expected most respondents to choose the SQ option in the experiment that presented them with their own SQ. However, since the main purpose of this study is to test whether SQ bias affects empirical results, we presented the township respondents with a second experiment. The second experiment contained an SQ option perceived to be less relevant to the township sub-sample.

By presenting the township stratum with a second experiment containing a partially relevant SQ, we can test whether utility functions and MWTP estimates differ across experiments. The difference between the first and second experiments is mainly in the SQ options. Therefore, we expect less SQ bias in the experiment in which the SQ was less relevant. This is because respondents may view the SQ option as one of the experimentally designed options. Psychological literature implies that if respondents are made to choose between their current SQ and other hypothetical options, they tend to disproportionately choose the SQ option. Reasons for this bias selection have been given in Samuelson and Zeckhauser (1988) and subsequent theories (see Tversky and Kahneman, 1991; Ritov and Baron, 1992; Anderson, 2003; Mandler, 2004). An example of a choice set with an SQ perceived to be less relevant to the township sub-sample is given in Table 3.3.

Table 3.3: Example of a choice set with an SQ partially relevant to the township stratum

	STATUS QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	No access to piped water	Inside dwelling	In yard	
Reliability	Yes	No	Yes	
Water pressure	High pressure	High pressure	Low pressure	
Water quality	Bad Taste	Has a smell	Safe to drink	
Monthly cost	R0	R680	R220	
I WOULD CHOOSE:	☐	☐	☐	☐

The same approach explained above for the township sub-sample was also applied to the suburban sub-sample. Respondents in the suburban sub-sample were also presented with two experiments. The first experiment contained a baseline SQ option perceived to be relevant to the sub-sample, while the second experiment presented respondents with an SQ option perceived as less relevant to them. As with the township sub-sample, we anticipated SQ bias in the first experiment and real trade-offs in the second. Utility functions and MWTP estimates from each of these two experiments will be compared against each other, as well as against those from the township stratum.

3.5. Modelling

Choice experiments are stated preference surveys that give respondents a series of alternatives that differ in attribute levels (Hanley et al., 2001). Respondents compare the available alternatives and choose the one that maximises their utility. The theoretical foundation of choice experiments arose from the random utility theory, which hypothesises that an individual makes choices based on the characteristics of the good, along with a random component (McFadden, 1974). According to Ben-Akiva and Lerman (1985), the random component could emerge from the uniqueness of the individual's preferences, or due to researchers having incomplete information about the individual observed. Given this, the random utility theory hypothesises that the utility U_{ij} of individual i obtained from alternative j is not known but can be decomposed into a deterministic component V_{ij} and an unobserved random component ε_{ij}. Therefore, the individual utility function will be presented as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{3.2}$$

Equation 3.2 is the basic utility function, which could alternatively be expressed by decomposing the indirect utility function for individuals into the deterministic component V_{ij}, which is normally specified as a linear index of the attributes in a choice set, and a stochastic component ε_{ij}, representing the error term. Therefore, the function assumes the form:

$$U_{ij} = V_{ij}\left(X_{ij}, C_{ij}, \beta\right) + \varepsilon_{ij} \tag{3.3}$$

Parameter U_{ij} is the true but unobservable utility of individual i associated with alternative j, while X_{ij} is a vector of the attributes associated with alternative j, parameter C_{ij} is the cost of alternative j, parameter β is a vector of preference parameters for the population in the sample, and ε_{ij} is the stochastic component (random term) with a zero mean. The utility function expressed in equation 3.3 can simply be expressed as linear in parameters, as follows:

$$U_{ij} = \sum_{k=1}^{K} \beta_x X_{ij} + \beta_c C_{ij} + \varepsilon_{ij} \quad (3.4)$$

The random utility theory assumes that any rational individual i will choose alternative j over alternative k if $U_{ij} > U_{ik}$. Each alternative consists of a bundle of attributes. When one alternative is selected over the other, it suggests that the hypothetical utility derived by an individual from the chosen alternative is greater than the utility of the other alternative not chosen (Greene, 2003, Hensher et al., 2015)[12]. This is expressed as follows:

$$P_i(j) = \text{Prob}\left(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}\right) \forall\, k \in C, k \neq j. \quad (3.5)$$

If the error terms are independently and identically distributed (IID) with an extreme value type I distribution, the variance of which is $var(\varepsilon) = \pi^2\tau^2/5$, where τ is a scale parameter used to normalise the model, then the choice probability of an alternative is expressed as:

$$P_{ij} = \exp\left(\frac{v_{ij}}{\tau}\right) / \sum_{k=1}^{K} \exp\left(\frac{v_{ik}}{\tau}\right) \quad (3.6)$$

The basic conditional logit model (CLM) – also known as the multinomial logit (MNL), in cases where there are no choice varying attributes – assumes the choice probability illustrated in equation 3.6. Historically, for many years the MNL model was the primary basis for the analysis of multinomial choices (Keane and Wasi, 2012), and has been labelled the 'workhorse' for discrete choice experiments (Hensher et al., 2015). However, the model is only good as a basic model because it relies on two unrealistic assumptions (Meyerhoff and Liebe, 2009).

[12] In deriving the probability of choosing an alternative within the random utility model, the choice of alternative with higher utility is not certain. The expectation has always been that there is a high chance that a respondent will choose the alternatives with higher utility.

Firstly, it assumes respondents to have homogeneous tastes for observed attributes. Secondly, it also assumes that the random part of utility obeys the independence from irrelevant alternatives (IIA) as well as the independence and identical distribution (IID) properties. According to Hensher et al. (2015), these assumptions rule out persistent heterogeneity in taste for both observed and unobserved product attributes.

To address the problems associated with the IIA and IID assumptions of the basic MNL model, the discrete choice analysis literature suggests various other models. These suggested models are more advanced, and include the nested logit (NL), the mixed logit (MXL – also called the Random Parameter (RPL)), the generalised mixed logit (GMXL), and the non-linear random parameters logit (NRPL) models (see Greene, 2012; Hensher et al., 2015). To assess the impact of the SQ options on choice experiment studies, this study uses the MXL and GMXL models.

The MXL model allows coefficients to vary randomly across individuals, reflecting the fact that different respondents have different tastes and preferences for attributes in each choice set (Hensher and Greene, 2003; Hensher et al., 2005). MXL explicitly accounts for both observed and unobserved heterogeneity in the preference parameters. It can be estimated using single cross-sectional data as well as panel data (Hensher et al., 2015). According to Greene (2012), the MXL model formulation is a one-level MNL model for individuals $i = 1, \dots, N$ in choosing alternative j. The model breaks down coefficients into a population mean and an unobserved individual's deviation from that mean. This is shown as:

$$U_{ij} = \beta_1 X_{ij} + \varepsilon_{ij} = \beta X_{ij} + \eta_{ij} + \varepsilon_{ij} \tag{3.7}$$

Parameter β_1 is the population mean and η_{ij} is the individual deviation from the population mean (i.e. the individual specific heterogeneity, with mean zero and standard deviation one, according to Greene (2012)). If θ is then used to represent the distribution of the parameters of β, the probability of individual i choosing alternative j can therefore be represented as:

$$\overline{P}_{ij} = \int P_{ij}\, f(\beta \mid \theta) d\, \beta \tag{3.8}$$

P_{ij} is as given in equation 3.5 and $f(\beta \mid \theta)$ is the probability density function for the coefficient β over the vector of parameter θ. Due to $\overline{P}_{ij}$ not having a closed form and its value being approximated numerically by way of simulation, estimation becomes somewhat complicated.

MXL models are good at identifying taste heterogeneity only. As such, Hensher et al. (2015) explain that there is growing interest from researchers in establishing a mechanism that also accounts for scale heterogeneity across individuals – that is, the variance of the variance term, also explained as the standard deviation of utility over different choice sets. The GMXL model is identified as a better tool that recognises the relationship between scale and taste heterogeneity (Fiebig et al., 2010; Keane and Wasi, 2012). Therefore, we also use the GMXL model to estimate households' utility functions.

According to Hensher et al. (2015), the GMXL model builds on the specifications of the MXL model and the generalised multinomial logit (GMNL) model suggested by Fiebig et al. (2010). Both observed and unobserved scale heterogeneity across choices are accommodated in the model by random alternative-specific constants (Hensher et al., 2015). The essential format of the GMXL model suggested by Fiebig et al. (2010) is a mixed logit model, illustrated as:

$$U_{ij} = \boldsymbol{\beta}_i'\mathbf{x}_{ij} + \varepsilon_{ij} \tag{3.9}$$

$$\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\Gamma}\mathbf{w}_i,\ \mathbf{w_1} \sim \mathrm{N}[\mathbf{0},\mathbf{I}], 0 \le \gamma \le 1 \tag{3.10}$$

$$\sigma_i = \exp\left(-\frac{\tau^2}{2} + \tau v_i\right), v_i \sim \mathrm{N}[0,1] \tag{3.11}$$

Equation 3.10 is an MNL model based on the extreme value distribution of the error component ε_{ij}. The general form of the GMXL model combines the scaled MNL model with the random parameter model. A random scaling factor σ_i with mean 1 and variance $\exp(\tau^2 - 1)$ is included in the model. Greene (2012) and Hensher et al. (2015) suggest that Gamma γ is central to the GMXL model, as it controls the relative importance of the overall scaling of the utility function. When γ = 0, it implies a scaled MXL model, while when γ = 1 it means a hybrid model. The other important element of the GMXL model is the Tau scale τ. When τ is equal to zero ($\tau = 0$), it implies that γ is not identified (that is, γ is not estimable).

The study will estimate utility functions using both MXL and GMXL models. However, these two models will be compared in terms of their goodness-of-fit statistics. The most fit model between MXL and GMXL will be adopted and its results presented. The three main goodness-of-fit parameters commonly used to compare models in the discrete choice experiment literature are the log likelihood function (LL), Akaike information criterion (AIC), and Bayesian information criterion (BIC). LL expresses how many times more likely the data are under one model than the other. A model with a larger LL estimate is deemed more robust relative to the other. AIC is an information-based measure of the relative quality of statistical models for a given set of data. Given a set of models, the model to be preferred is the one with the lowest AIC value (see Aho et al., 2014; Akaike, 1998; Burnham and Anderson, 2004). BIC is also an information-based measure, for which the decision rule is that the model with the lowest BIC is preferred (Burnham and Anderson, 2004).

Additionally, the study estimates the marginal willingness to pay (MWTP) for the water service attributes. MWTP estimates are welfare measures that show the marginal rate of substitution (MRS) for attributes. It is deemed essential to also examine the impact of SQ bias on welfare measures. This is because it could be possible for SQ bias to have an impact on the utility functions, but not have any impact on welfare measures, and vice versa. Hensher et al. (2015) suggest that one important output from choice models is the MRS between specific attributes of interest, with a financial variable typically being in the trade-off so that MRS is expressed in monetary terms. MRS is commonly referred to as the MWTP and gives the average estimates of what households are prepared to pay for or against improvements in each attribute. Assuming a linear utility function with attribute X and a cost C:

$$U_{ij} = \beta X_j + \mu\left(C_i - p_j\right) + \varepsilon_{ij} \tag{3.12}$$

U_{ij} is the utility of the respondent i for alternative j; while β_j and μ are the marginal (dis)utilities of attribute X (attribute of interest) and cost, respectively. Dikgang and

Muchapondwa (2014) point out that it is possible for researchers to use a set of observed discrete choices to determine different marginal values for each attribute used in explaining the policy alternatives, instead of a single value for the whole policy scenario. As such, this current study follows that route of determining the marginal values for each attribute.

We appreciate that there have been developments in the choice experiment literature. Researchers using advanced estimation models (such as the GMXL model) use measures such as the WTP space (Greene, 2012). According to Hensher et al. (2015), the GMXL model provides a straightforward method of re-parameterising the model to estimate taste parameters in WTP space, which continues to enjoy attention as an alternative way of directly obtaining WTP estimates. Since the primary objective of this current study is to compare estimates across two strata and establish whether SQ options bias results (as opposed to simply establishing empirical WTP estimates), we stick to the traditional way – which establishes WTP utility, and not WTP space. MWTP estimates will be compared across the different strata identified earlier in the study.

3.6. Experimental data

3.6.1. Data collection

The study is based on experimental data collected from 999 household heads from Durban during the period September to November 2016. Survey instruments[13] were prepared in English, and four enumerators fluent in both English and isiZulu (i.e. the local language) were recruited from a group of postgraduate students. These enumerators were trained and supervised during the data collection process. A total of 500 responses was collected in the suburbs, while 499 were collected in townships. Responses from the suburbs were further divided into 249 collected using a questionnaire with a relevant SQ (block 1), and 251 collected using a questionnaire with a partially relevant SQ (block 2)[14]. For the township sub-sample, 250 complete responses were collected in block 1 and 249 in block 2. Choice experiments

[13] An example of the questionnaire used to collect information is given in Appendix 3.1. The questionnaire presented is for block 1 of the township sub-sample.

[14] Henceforth, the experiment involving a relevant SQ will be identified as block 1, while the experiment with the partially relevant SQ will be identified as block 2. It is important to note that each sub-sample has both blocks.

allow even smaller samples to produce many observations. This is because each respondent is asked repetitively, which then increases the number of observations.

When collecting data, households were conveniently selected until the required data points from each area were obtained. To avoid fatigue and receive maximum cooperation, each respondent took part in only one experiment. For example, if the first respondent in the township sub-sample took part in an experiment using the SQ perceived to be relevant, the next respondent would then take part in the experiment with the SQ perceived as less relevant. In addition to the choice experiments, each of the four questionnaires contained two other sections. The second section collected general information, while the third section collected biographic details. Except for the question on household monthly income, which had higher values in the suburbs sub-sample, all the questions in the second and third sections were similar across the four questionnaires (i.e. respondents answered the same questions in these sections).

3.6.2. Descriptive statistics

In addition to stated preference data, the collected data included detailed information on the demographic characteristics of the respondents, as well as some general information on how the respondents received water services at the time. The demographic characteristics collected included household size, level of education, age and income, as well as source of income. In addition to these demographic data, some general information was collected on how each household accessed piped water services, whether they received free basic water, how often they experienced water interruptions, and how they perceived the quality of water they received. Such information is deemed to be essential, as it determines households' preferences for water service packages. Table 3.4 provides an overview of the descriptive statistics of the main variables of interest for all blocks in the two sub-samples.

Table 3.4: Descriptive statistics of respondents

	Suburbs		Townships	
	Block 1 Mean	Block 2 Mean	Block 1 Mean	Block 2 Mean
No. of respondents (N)	***249***	***251***	***249***	***250***
Female respondents (%)	41	41	44	45
Household head (%)	39	45	49	55
Average household size	4	4	5	5
Married respondents (%)	51	45	29	34
Race (%): *African*	42	4	87	79
Indian	45	43	12	20
Coloured	5	6	1	1
White	7	10	0	1
Age (%): *16-24 years*	12	14	15	20
25-34 years	32	37	24	23
35-44 years	24	25	21	22
45-54 years	14	14	14	15
55-64 years	12	6	12	11
65+ years	6	4	12	10
Secondary education and above (%)	76	74	30	32
Monthly income (%):				
<R2500	-	-	75	67
R2500 < R5000	-	-	14	18
R5000 < R10000	-	-	9	10
>R10000 < 15000	52	46	2	4
R15000 < R30000	31	30	-	-
R30000 < 50000	12	14	-	-
>R50000	6	10	-	-
Recipients of free basic water (%)	-	-	39	66
Access to piped water (%):				
Inside dwelling	100	100	71	77
In yard	0	0	25	19
Community tap	0	0	4	4
Water supply interruptions (%):				
Very often	8	6	21	24
Once in a while	53	51	69	62
Not at all	38	43	10	14
Water quality experiences (%):				
Not clear	35	24	29	34
Bad taste	3	4	1	2
Bad smell	0	1	0	0
Has colour	12	7	2	2
Good quality	50	65	68	62

Table 3.4 shows various systematic differences between the suburbs and township sub-samples. It is noted that the average household size is slightly lower in the suburbs sub-sample with a mean of 4 family members, compared to the townships sub-sample where the average household size is 5 family members. This dynamic reflects the actual trends in South African communities, where township families usually include extended family members. It is also noted that most township dwellers are black South Africans, at around 80%, reflecting the

legacy of a segregated past in which non-white South Africans were confined to township areas. However, it is equally important to note that although South Africans of Indian origin constitute the greatest number of suburban dwellers, many black South Africans also live in the suburbs. Another important dynamic noted in the table is that most of the respondents from the suburbs sub-sample have at least secondary-school education. Statistics for this variable sit at above 70% in the suburbs sub-sample, while in the township sub-sample the percentage is around 30%.

In terms of access to piped water, all households in the suburbs sub-sample access piped water inside their dwellings, while the percentage of households with such a facility in the township sub-sample is around 70%. Households without piped water inside their dwellings access it mostly from the yard, while about 4% of the township households surveyed still access piped water from community taps. These statistics reflect that some township households receive inferior water service packages compared to suburban households. However, it is imperative to note that overall, most respondents in the sample indicated that they receive water of good quality, even though some respondents suggested that the water they receive is not clear. The satisfaction with water quality across all sub-samples clearly indicates that even though households may access water differently, the quality of the water is good. Finally, although households from all sub-samples experience water interruptions “once in a while”, more interruptions are experienced by the township sub-sample than by the suburbs sub-sample.

3.6.3. Frequency distribution of stated preference choices

An understanding of the stated preference data collected using the different experiments is essential. Prior to the survey, we hypothesised that if respondents were presented with their relevant SQ, most of them would choose the SQ ahead of hypothetical designed options, as suggested in Samuelson and Zeckhauser (1988). Probable reasons for this behaviour have been discussed in the previous sections of this study, and include loyalty to the SQ, choice task complexity, loss aversion, inertia, and many others (see Anderson, 2003; Dubé et al., 2010; Lanz and Provins 2015; Mandler, 2004; Meyerhoff and Liebe, 2009; Moon, 2000; Oehlmann et al., 2017; Ritov and Baron, 1992; Scarpa et al., 2007; Tversky and Kahneman, 1991). Therefore, we expected to observe SQ bias in block 1 experiments, where respondents were presented with their realistic SQ option. Furthermore, we expected less SQ bias in block 2 experiments, where respondents were presented with unrealistic SQ options. To explain this phenomenon, Figure 3.2 presents the frequency distribution of choices across the different options in each experiment.

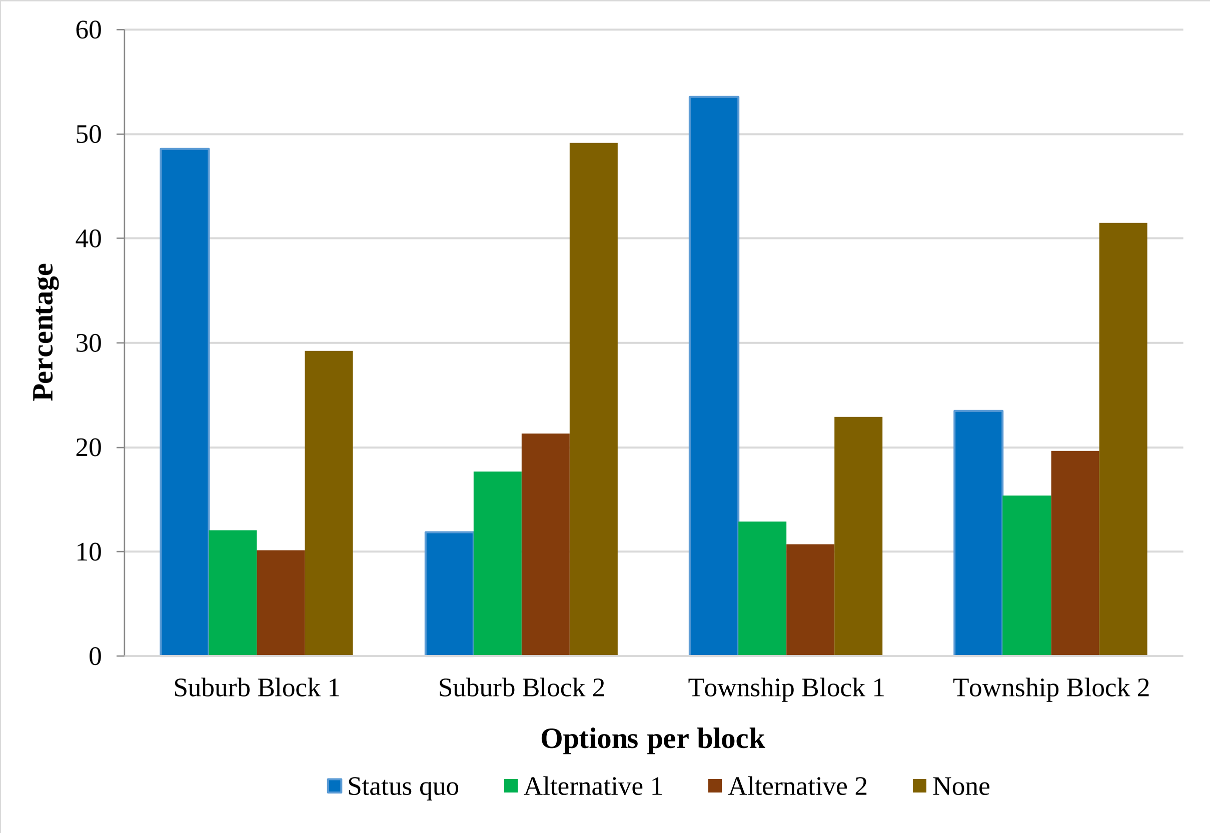


Figure 3.2: Frequency distribution of choices made by respondents

The frequency distribution statistics presented in Figure 3.2 confirm our hypothesis that most respondents would choose the SQ option if presented with a relevant SQ. This is shown in the number of respondents who chose the SQ options in the first blocks of the two sub-samples. In suburbs block 1, close to 50% of respondents chose the SQ option; while in the township sub-sample, more than 50% of respondents chose the SQ option in block 1. However, in the second blocks of each sub-sample, where respondents were presented with partially relevant SQ options, the frequency of SQ choices was lower than that observed in the block 1 experiments. For the suburbs sub-sample, about 11% of respondents chose the SQ option in block 2; while in the townships sub-sample, around 23% of respondents chose the SQ option in block 2. Interestingly, it is also noted that when presented with a partially relevant SQ option, relatively large numbers of respondents chose to opt out, by selecting the 'none' option. About 49% of the respondents opted out in suburbs block 2, while around 41% of township respondents opted out in block 2. However, since all options in the choice sets were selected, it is still possible to achieve real trade-offs. As such, the next section presents estimations results from the econometric analysis of stated preference data.

3.7. Empirical findings

To examine the impact of SQ bias on empirical estimates, we use the mixed logit (MXL) and the generalised mixed logit (GMXL) models as estimation tools. The robustness of each of these estimation tools in so far as our data is concerned is compared, and the most fit tool is adopted. Although GMXL is more advanced and is expected to perform better than MXL, Hensher et al. (2015) suggest that the latter may perform better, depending on the dataset. The model that performs best is subsequently used to estimate utility functions. Since our sample is stratified into two sub-samples, and each sub-sample responds to two survey instruments, we estimate utility functions for each sub-sample. MWTP will also be estimated for each sub-sample, using the model that fits best. The ultimate rationale for these two important analyses is to compare estimates across sub-samples and see if there are variations in terms of statistical significance, sign and magnitude of parameter estimates.

To estimate utility functions, the study adopts unconstrained MXL and GMXL models that do not control for heterogeneity in the means of normally distributed random parameters. The five attributes of the study are modelled as normally distributed random parameters while alternative specific constants (ASCs) and socioeconomic characteristics are modelled as fixed parameters. Results are obtained using the Halton sequence for simulation based on 200 draws. We adopted 200 draws because estimation was collapsing in one of the four models if more than 200 draws were included. For uniformity across models, we adopted 200 draws. However, we are cognisant that more draws improve the quality of the estimates. Additionally, this section also presents MWTP estimates for each sub-section. The MWTP estimates are then compared across blocks and sub-samples.

3.7.1. Goodness of fit: MXL versus GMXL models

Since each of the two sub-samples has two blocks, presenting estimation results for both the MXL and GMXL models may lead to too many columns that might render the work clumsy. Consequently, we assess the goodness-of-fit parameters of the two estimation tools. The better-performing tool is adopted, and its results presented. The three main goodness-of-fit parameters commonly used in the literature are the LL, AIC, and BIC. A model with a larger LL estimate is deemed more robust. For AIC and BIC, the model to be preferred is one with the lowest value (see Aho et al., 2014; Akaike, 1998; Burnham and Anderson, 2004). Goodness-of-fit results for the MXL and GMXL models used in this study are presented in Table 3.5.

Table 3.5: Goodness of fit statistics

	MXL Model				**GMXL Model**			
	Suburb		**Township**		**Suburb**		**Township**	
	Block 1	**Block 2**	**Block 1**	**Block 2**	**Block 1**	**Block 2**	**Block 1**	**Block 2**
LL	-1117.0	-1528.7	-1640.5	-1786.9	-525.6	-1132.0	-965.9	-1195.5
AIC	2270.0	3093.5	3316.9	3609.8	1111.2	2324.1	1991.8	2451.1
BIC	2365.4	3189.2	3412.4	3705.4	1270.3	2483.6	2150.9	2610.4
N	1485	1505	1484	1498	1485	1505	1484	1498

As hypothesised, the estimates for goodness-of-fit parameters presented in Table 3.5 show that all GMXL models performed better than the MXL models. Under the GMXL models, the LL estimates for both suburbs and townships are larger than those presented under the MXL models. The GMXL estimates also show lower AIC and BIC statistics than those recorded in the MXL models, implying that the GMXL models outperformed the MXL models. Therefore, we present empirical results from the GMXL models only.

3.7.2. GMXL model estimates

Using unconstrained GMXL models, this study estimates utility as a function of normally distributed attributes. Utility functions for all blocks in each sub-sample assume the form:

$$U_{ij} = ASC + \beta_1 PIPE_{ij} + \beta_2 RELIABILITY_{ij} + \beta_3 PRESSURE_{ij} + \beta_4 QUALITY_{ij} + \beta_5 COST_{ij} + \varepsilon_{ij} \quad (3.14)$$

ASC in equation 3.14 is the alternative specific constant, which in the choice experiment literature is often used to capture SQ bias (see Boxall et al., 2009; Lanz and Provins, 2015; Meyerhoff and Liebe, 2009; Oehlmann et al., 2017; Ortoleva, 2010). In studies that test for the existence of SQ bias, the common approach is to include at least two ASCs; one for the status quo option, (ASC_{SQ}) and the other for experimentally designed options. Such studies report a positive and statistically significant ASC_{SQ} as evidence of SQ bias (see Kahneman et al., 1991; Korobkin, 1997; Maltz and Romagnoli, 2015; Meyerhoff and Liebe, 2009). However, our study does not primarily seek to capture the existence of SQ bias.

This study notes and acknowledges the existence of SQ bias, and tests whether respondents' preferences change when they are presented with experiments containing different SQ options. This is done by examining the sign, significance and magnitude of the attribute parameters in the utility functions from each block of the two sub-samples. Therefore, although this study also reports on the sign and significance of the ASC estimates across sub-samples, the main emphasis will not be on the existence of SQ bias, but on its impact on the utility function. A comparison of attribute parameter estimates in each utility function informs us whether preferences differ with changes in the SQ. Table 3.6 presents the estimation results.

Pipe water and water quality are qualitative variables with three categories and as in the norm in econometric estimations, a norm is to introduce two dummy variables for a qualitative variable with three categories. In contrast, we introduce only one variable for each of these qualitative variables. We have performed some transformations on these qualitative variables, hence our use of one variable.

Table 3.6: Estimation results based on GMXL models[15]

	Suburbs				Townships			
	Block 1		Block 2		Block 1		Block 2	
	Par. Est.	Std. Err	Par. Est.	Std. Err	Par. Est.	Std. Err	Par. Est.	Std. Err
Random parameters in utility functions								
COST	0.010***	0.003	-0.005*	0.002	-0.018***	0.003	-0.013**	0.005
PIPE	-4.229	3.180	-2.726***	1.045	0.376	0.436	0.206	0.259
RELIABILITY	-3.570	2.690	1.166	1.129	0.826	0.814	-0.003	0.732
PRESSURE	5.255*	2.373	5.645**	2.555	0.606	0.997	-0.615	0.914
QUALITY	-10.190*	5.954	-11.111**	4.790	-3.799***	0.691	-3.419**	1.470
Nonrandom parameters in utility functions								
ASC_{SQ}	11.707***	1.435	2.566**	0.869	3.674***	1.590	2.516**	0.942
$ASC_{1\text{ and }2}$	6.905*	1.922	3.463***	0.302	5.815**	1.603	3.822***	0.912
AGE	-0.924	0.563	-0.084	0.227	-0.264	0.300	-0.287*	0.163
EDUCATION	-0.284	0.338	-0.084	0.103	-0.447**	0.216	0.064	0.141
GENDER	-1.176	1.413	-0.596	0.433	0.978	0.682	-1.452***	0.531
INCOME	-1.163	0.773	0.762***	0.261	-0.255	0.385	-0.326	0.268
RACE	0.104	0.778	-0.541**	0.234	-0.009	0.430	0.144	0.271
STATUS	2.414	1.552	-0.539	0.525	-0.389	0.858	0.067	0.507
Diagonal values in Cholesky matrix, L.								
NsCOST	0.003	0.003	0.012**	0.005	0.017***	0.003	0.011**	0.004
NsPIPE	0.792	2.231	4.301**	1.749	1.972***	0.529	2.186**	0.883
NsRELIABILITY	2.405**	1.142	1.300	1.210	2.392***	0.840	0.759	0.742
NsPRESSURE	0.373	1.334	4.527**	2.166	0.923**	0.372	4.026**	1.627
NsQUALITY	3.768**	1.557	5.077***	1.898	0.306	0.456	1.742*	0.903
Below diagonal values in L matrix. V = L*Lt								
PIPE:COST	-1.419	1.840	-0.913	0.620	-2.265***	0.465	-1.877***	0.715
RELIA:COST	-2.918	2.045	5.218**	2.482	-1.189	0.821	0.797	0.674
RELIA:PIPE	-2.568	2.410	-1.388	1.305	-3.228***	1.163	-3.250**	1.488
PRESS:COST	0.444	2.594	-0.876	1.082	0.663	0.805	-0.384	0.766
PRESS:PIPE	-0.954	1.809	0.787	0.863	0.086	0.907	-3.161**	1.555
PRESS:RELIA	-1.637	1.121	0.461	1.196	2.717***	0.847	0.110	0.523
QUAL:COST	4.312	4.205	-4.454**	1.814	2.189***	0.614	0.011	0.482
QUAL:PIPE	-3.775	4.629	-3.252**	1.507	2.925***	0.775	3.153**	1.310
QUAL:RELIA	-3.563	2.554	-2.318	1.652	-0.865	0.630	-1.108	0.744
QUAL:PRESS	-1.314	2.878	-2.372*	1.422	0.808*	0.477	1.761**	0.818
Variance parameter tau in GMX scale parameter								
TauScale	1.336***	0.168	1.297***	0.248	0.800***	0.116	1.243***	0.242
Weighting parameter gamma in GMX model								
GammaMXL	0.0	0.124	0.0	0.031	0.00	0.062	0.0	0.033
Sample Mean and Sample Std. Dev.								
Sigma(i)	0.906	1.480	0.922	1.443	0.966	0.836	0.926	1.376
Standard deviations of parameter distributions								
sdCOST	0.003	0.002	0.012**	0.005	0.017***	0.003	0.011**	0.004
sdPIPE	1.625	2.327	4.401***	1.622	3.004***	0.598	2.882***	0.326
sdRELIABILITY	4.571***	0.997	5.554**	2.295	4.190***	0.770	3.431**	1.421
sdPRESSURE	1.981	1.446	4.700**	2.064	2.946***	0.737	5.134***	0.942
sdQUALITY	7.840***	2.780	8.197***	1.170	3.853***	0.711	4.160***	1.522

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Par Est. = parameter estimates. Std. Err = standard errors

[15] In additional to these results, GMXL also reports on the correlation between the random parameters in the utility functions. The correlation matrices for the four models are presented in Appendix 3.2 as Table A1. Most of the values have low correlation indices, suggesting minimum levels of multi-collinearity.

The ASC_{SQ} estimates for the first blocks of each sub-sample are both statistically significant at the 1% significance level. Positive and statistically significant ASC_{SQ} estimates in the first blocks of the two sub-samples indicate the existence of SQ bias. These findings are consistent with revelations from Samuelson and Zeckhauser (1988) and several other previous studies that revealed that many respondents disproportionately choose the SQ option if offered their relevant SQ. Regarding the second blocks of each sub-sample, we expected the ASC_{SQ} estimates to be statistically insignificant. However, contrary to our prior hypothesis, the ASC_{SQ} estimates for the second blocks were also positive and statistically significant[16]. As explained earlier, the aim of this study is not to show evidence of SQ bias, but to test whether utility functions differ across blocks in the given sub-samples. Therefore, the estimation results presented in Table 3.6 will be interpreted based on the sign, significance and magnitude of the random parameter coefficients that make up the utility functions. More precisely, the study compares the coefficients of the random parameter estimates in each stratum, and across the sub-samples. A comparison of these coefficients gives information on whether utility functions vary across blocks. Before interpreting the results presented in Table 3.6, we substitute coefficients for each parameter estimate into the utility function given earlier in equation 3.14. Four equations are then applied, as follows:

$$U_{SUBURBS\,B1} = 11.71 - 4.23xPIPE_{ij} - 3.57xRELIABILITY_{ij} + 5.26xPRESSURE_{ij} - 10.19xQUALITY_{ij} + 0.01xCOST_{ij} + \varepsilon_{ij} \tag{3.15}$$

$$U_{SUBURBS\,B2} = 2.57 - 2.73xPIPE_{ij} + 1.17xRELIABILITY_{ij} + 5.65xPRESSURE_{ij} - 11.11xQUALITY_{ij} - 0.01xCOST_{ij} + \varepsilon_{ij} \tag{3.16}$$

$$U_{TOWNSHIP\,B1} = 3.67 + 0.38xPIPE_{ij} + 0.83xRELIABILITY_{ij} + 0.61xPRESSURE_{ij} - 3.80xQUALITY_{ij} - 0.02xCOST_{ij} + \varepsilon_{ij} \tag{3.17}$$

[16] This could be because of various other reasons. We suspect that one of those reasons could be that most respondents chose neither the SQ option nor the hypothetical options, but rather the 'none' option, as revealed earlier in Figure 3.2. Such choices may be reflected by the positive and significant ASC_{SQ}.

$$U_{TOWNSHIP\ B2} = 2.52 + 0.21xPIPE_{ij} - 0.003xRELIABILITY_{ij} - 0.62xPRESSURE_{ij} - 3.42xQUALITY_{ij} - 0.01xCOST_{ij} + \varepsilon_{ij} \quad (3.18)$$

Equations 3.15 to 3.18 show the utility functions of the four blocks of the two strata. Positive attribute coefficients indicate household preference for changes in the attribute. For instance, the positive coefficient of PRESSURE in equations 3.15 to 3.17 indicates that households prefer changes in the water pressure. To be specific, a unit change in PRESSURE increases households' utility by about 5.23 units in suburbs block 1, and by 5.65 units in suburbs block 2. On the other hand, negative coefficients suggest that households do not prefer changes in the attribute. For example, the negative coefficients of PIPE in equations 3.15 and 3.16 indicate that suburban households do not prefer changes in the way they access piped water services. More precisely, the PIPE results in equations 3.15 and 3.16 suggest that a unit change in access to piped water reduces utility by about 4.23 units in suburbs block 1, and 2.72 units in suburbs block 2.

Illustrations in equations 3.15 to 3.18 show the signs and magnitudes of the coefficients of parameters. These are essential for defining the utility function, as they indicate the impact of changes in an attribute and the extent of that impact on utility. The first step in our bid to examine the impact of SQ bias will test whether variations exist in the utility functions across blocks. Therefore, we compare utility functions in each sub-sample in terms of the signs and magnitudes of the coefficients. In the suburbs sub-sample, two variations are noted in the signs of COST and RELIABILITY across the two blocks. In the township sub-sample also, two variations are noted across blocks, in the signs of RELIABILITY and PRESSURE. Regarding the magnitudes of parameter estimates, no huge discrepancies are noted across blocks in either sub-sample. Magnitudes of coefficients are consistent across the blocks of each sub-sample.

Although the signs and magnitudes of parameter estimates are essential when examining utility functions, the statistical significance of the parameters is more important in determining preferences. Statistical significance shows the attributes that are important to households, as well as those that are not important. Therefore, a comparison of the statistical significance of the attribute parameter coefficients across the blocks of each sub-sample is essential. In the suburbs sub-sample, only the coefficients for PIPE are not consistent across the two blocks; the coefficient is insignificant in block 1, but significant at 1% in block 2. The rest of the coefficients are consistent in terms of statistical significance across the two suburban blocks. Notable variations are seen only in the level of significance where, for example, the significance level for COST is 1% in block 1 but 10% in block 2. Overall, there is consistency in the statistical significance of parameter estimates across the suburban blocks. Regarding statistical significance in the township sub-sample, coefficients for all the five attributes are consistent across blocks. Where estimates are statistically significant in the first block, they are also significant in the second block, and vice versa.

In addition to the five attributes modelled as random parameters in the utility functions, we also controlled for six selected socio-economic characteristics (AGE, EDUCATION, GENDER, INCOME, RACE and STATUS). We considered these variables because we hypothesise that they are essential in determining how individuals make choices. For example, elderly respondents are expected to make different choices to younger respondents. Equally, a respondent's level of education, income level, race and marital status are also expected to be key determinants of choice. The empirical results in Table 3.6 show that none of these socio-economic characteristics were significant determinants of respondents' choices in suburbs block 1. However, in the second block of the suburbs sub-sample, INCOME and RACE were important determinants of choice. In the township sub-sample, EDUCATION was an important determinant in block 1, while AGE and GENDER were important determinants in block 2.

3.7.3. MWTP estimates

We also examined whether the different utility functions given in the different blocks would also yield the same welfare measures. To do this we estimated the households' MWTP for the given attributes. MWTP is a welfare measure that shows average estimates of what households are prepared to pay for or against improvements in each attribute. Positive and significant estimates show the average amount that households are willing to pay, while negative and significant estimates show how much they are willing to accept as compensation. The MWTP estimates are presented in Table 3.7.

Table 3.7: Marginal willingness to pay estimates (in US Dollars)[17]

	Suburbs				**Townships**			
	Block 1		Block 2		Block 1		Block 2	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
PIPE	31.06	26.09	-42.23***	14.06	1.46	1.71	1.10	1.22
RELIABILITY	26.21	24.11	18.06	14.23	3.22	3.26	-0.01	3.91
PRESSURE	-31.24*	16.63	87.45***	24.91	2.36	3.79	-3.28	5.08
QUALITY	74.82**	32.67	-172.12***	58.30	-14.79***	2.63	-18.26***	3.25
Wald Statistic	**1.85**		**0.96**		**2.42**		**2.41**	
Prob. from Chi^2	**0.000**		**0.008**		**0.000**		**0.000**	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Std. Err are standard errors.

The MWTP estimates presented in Table 3.7 are consistent in terms of statistical significance in the township sub-sample, where parameter estimates for all attributes are insignificant except for the parameter estimates of the QUALITY attribute. The MWTP estimates for QUALITY are both negative and statistically significant at 1% significance level. The negative sign suggests that households are not willing to pay for any improvements in the quality of the water they receive. More precisely, township households in block 1 are willing to accept $14.79 as compensation for improvements in the quality of water, while in block 2 they are willing to accept $18.26 as compensation for improvements in the quality of water. In a nutshell, different SQs did not affect MWTP estimates across the blocks of the township sub-sample, as the estimates are consistent across the two blocks.

In the suburbs sub-sample, there are inconsistencies across the two blocks. The only attribute with consistent parameter estimates in terms of sign and significance is RELIABILITY, which is statistically insignificant across the two suburban blocks. The rest of the attributes recorded inconsistent estimates in terms of sign and/or significance. Firstly, the MWTP estimate for PIPE is positive and statistically insignificant in block 1, but positive with a statistical significance of 1% in block 2. Secondly, the MWTP estimate for PRESSURE is negative with a statistical significance of 10% in block 1, but positive with a statistical significance of 1% in block 2. Finally, although the MWTP estimates for the QUALITY attribute are both statistically significant, in block 1, households from the suburbs are willing to pay $74.82 for improvements in the quality of their water; yet in block 2, they are willing to accept $172.12

[17] As at 24 October 2018, US$1 = ZAR14.30

as compensation if the water quality improvements. Inconsistencies in the suburbs sub-sample suggest that SQs may affect the MWTP estimates.

3.8. Conclusion

This paper tests for the effects of reducing status quo bias considering a heterogeneous sample. We test this by introducing a partially relevant status quo aimed at reducing SQ bias in a choice experiment that elicits household preferences for water service packages in Durban, South Africa. To achieve this, we stratify our sample into two sub-samples (i.e. suburbs and townships). Each sub-sample is presented with two experiments, one containing a relevant SQ (block 1) and another containing a partially relevant SQ (block 2). We test whether the likelihood of a participant choosing the SQ is driven by the relevance or partial relevance of the SQ option. Subsequently, we test whether this affects empirical results by comparing the significance, sign and absolute values of attribute parameters as well as MWTP estimates across the two blocks in each sub-sample. We use the GMXL and MXL models as estimation tools. However, we present empirical estimates from GMXL based on goodness-of-fit tests, which revealed that GMXL outperformed MXL.

Results from our tests revealed that parameter estimates across the two blocks of the township sub-sample were largely similar in terms of sign, statistical significance and the absolute value of the parameters' magnitude. Only COST and QUALITY emerged statistically significant in both blocks. These two attributes contained the same signs and had coefficients of the same magnitude in absolute terms. Similarities in the two blocks of the township sub-sample were also observed in the MWTP estimates. QUALITY was the only attribute with statistically significant MWTP estimates that also had the same sign and a similar coefficient magnitude across the two blocks.

In the suburban sub-sample, we found that all attributes except one reported the same statistical significance across the two blocks. The statistically significant parameters had the same signs, except for COST, which had a positive coefficient in block 1 and a negative coefficient in block 2. A positive coefficient in block 1 suggests that respondents preferred higher monthly water bills. Such a revelation is not consistent with either our prior expectations or the common findings in the literature, where the cost attribute generally has a negative coefficient (see Anand, 2001; Bhaduri and Kloos, 2013; Brouwer et al., 2015; Hensher et al., 2005). An analysis of the size of the coefficients showed no major differences across the two suburban blocks. We found that the attribute parameters were of the same magnitude, in absolute terms.

Estimation results in the suburban sub-sample imply that to a large extent, the relevance or partial relevance of the SQ did not affect the utility functions. However, the MWTP estimates in the two blocks of the suburban sub-sample reported disagreeing results in terms of sign and significance. Block 1 had two attributes that were statistically significant, while block 2 reported statistical significance on three attributes. Most importantly, we observed that all the statistically significant MWTP estimates had different signs across the two suburban blocks. Overall, we argue that the inclusion of a partially relevant SQ reduced SQ bias, but did not affect estimates of attribute parameters, in both suburbs and townships.

List of references

Addelman, S. 1962. Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics,* 4, 21-46.

Aho, K., Derryberry, D. and Peterson, T. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology,* 95, 631-636.

Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. *Selected Papers of Hirotugu Akaike.* Springer.

Alpizar, F., Carlsson, F. and Martinsson, P. 2001. Using choice experiments for non-market valuation. *Economic Issues 8*, 83–110.

Anderson, C. J. 2003. The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological Bulletin,* 129**,** 139.

Banerjee, S. G., Foster, V., Ying, Y., Skilling, H. and Wodon, Q. T. 2010. *Cost recovery, equity, and efficiency in water tariffs: evidence from African utilities*. Africa Infrastructure Country Diagnostic (AICD).

Bateman, I. J., Day, B. H., Jones, A. P. and Jude, S. 2009. Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management,* 58**,** 106-118.

Beattie, J., Baron, J., Hershey, J. C. and Spranca, M. D. 1994. Psychological determinants of decision attitude. *Journal of Behavioral Decision Making,* 7**,** 129-144.

Beck, T., Rodina, L., Luker, E. and Harris, L., 2016. Institutional and policy mapping of the water sector in South Africa. Vancouver, Canada: The University of British Columbia, Institute for Resources, Environment and Sustainability.

Ben-Akiva, M. E. and Lerman, S. R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand,* MIT Press.

Bliemer, M. C. and Rose, J. M. 2006. *Designing Stated Choice Experiments: State of the Art.* Emerald.

Bliemer, M. C., Rose, J. M. and Chorus, C. G. 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. *Transportation Research Part B: Methodological,* 102**,** 83-104.

Bliemer, M. C., Rose, J. M. and Hess, S. 2008. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling,* 1**,** 98-126.

Bonnichsen, O. and Ladenburg, J., 2015. Reducing status quo bias in choice experiments. *Nordic Journal of Health Economics,* 3, 47-67.

Boxall, P., Adamowicz, W. L. and Moon, A. 2009. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Australian Journal of Agricultural and Resource Economics,* 53**,** 503-519.

Brouwer, R., Job, F. C., Van der Kroon, B. and Johnston, R. 2015. Comparing willingness to pay for improved drinking-water quality using stated preference methods in rural and urban Kenya. *Applied Health Economics and Health Policy,* 13**,** 81-94.

Burnham, K. P. and Anderson, D. R. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research,* 33**,** 261-304.

Campbell, D., Hutchinson, W. G. and Scarpa, R. 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics,* 41**,** 401-417.

Dhar, R. and Simonson, I. 2003. The effect of forced choice on choice. *Journal of Marketing Research,* 40**,** 146-160.

Diamond, P. A. and Hausman, J. A. 1994. Contingent valuation: is some number better than no number? *Journal of Economic Perspectives,* 8**,** 45-64.

Dikgang, J. and Muchapondwa, E. 2014. The economic valuation of nature-based tourism in the South African Kgalagadi area and implications for the Khomani San 'bushmen'community. *Journal of Environmental Economics and Policy,* 3**,** 306-322.

Dubé, J. P., Hitsch, G. J. and Rossi, P. E. 2010. State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics,* 41**,** 417-445.

Ethekwini Municipality. 2014. *Provision of Potable Water, eThekwini Municipality* [Online]. eThekwini Municipality. Available: http://www.durban.gov.za/City_Services/water_sanitation/Services/Pages/Provision.aspx [Accessed 15 May 2015].

Ethekwini Municipality. 2015. *eThekwini's Official Unemployment Rate Lowest in the Country* [Online]. eThekwini Municipality. Available: http://www.durban.gov.za/Resource_Centre/Press_Releases/Pages/EThekwini's-Official-Unemployement-Rate-Lowest-In-The-Country.aspx [Accessed 2 May 2016].

Ferrini, S. and Scarpa, R. 2007. Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, 53**,** 342-363.

Fiebig, D. G., Keane, M. P., Louviere, J. and Wasi, N. 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, 29**,** 393-421.

Greene, W. H. 2003. *Econometric Analysis*, Pearson Education India.

Greene, W. H. 2012. NLOGIT Version 5 Reference Guide. Econometric Software, Inc.: Econometric Software, Inc. Plainview, NY.

Goldin, J. A., 2010. Water policy in South Africa: trust and knowledge as obstacles to reform. *Review of Radical Political Economics*, *42*(2), 195-212.

Hahn, G. J. and Shapiro, S. S. 1966. *A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments*, General Electric, Research and Development Center.

Hanley, N., Mourato, S. and Wright, R. E. 2001. Choice modelling approaches: a superior alternative for environmental valuatioin? *Journal of Economic Surveys*, 15**,** 435-462.

Hensher, D., Shore, N. and Train, K. 2005a. Households' willingness to pay for water service attributes. *Environmental and Resource Economics*, 32**,** 509-531.

Hensher, D. A. and Greene, W. H. 2003. The mixed logit model: the state of practice. *Transportation*, 30**,** 133-176.

Hensher, D. A., Rose, J. M. and Greene, W. H. 2005b. *Applied Choice Analysis: A Primer*, Cambridge University Press.

Hensher, D. A., Rose, J. M. and Greene, W. H. 2015. *Applied Choice Analysis*, Cambridge University Press.

Herrfahrdt-Pähle, E., 2010. South African water governance between administrative and hydrological boundaries. *Climate and Development*, *2*, 111-127.

Hess, S. and Rose, J. M. 2009. Should reference alternatives in pivot design SC surveys be treated differently? *Environmental and Resource Economics,* 42**,** 297-317.

Holland, C. W. and Cravens, D. W. 1973. Fractional factorial experimental designs in marketing research. *Journal of Marketing Research***,** 270-276.

Jansen, A. and Schulz, C.E., 2006. Water demand and the urban poor: A study of the factors influencing water consumption among housholds in Cape Town, South Africa. *South African Journal of Economics*, *74*(3), 593-609.

Kahneman, D., Knetsch, J. L. and Thaler, R. J. 1991. The endowment effect, loss aversion, and status quo bias. *Journal of Economics Perspective,* 5**,** 193-206.

Kanyoka, P., Farolfi, S. and Morardet, S. 2008. Households' preferences and willingness to pay for multiple use water services in rural areas of South Africa: an analysis based on choice modelling. *Water SA,* 34**,** 715-723.

Keane, M. P. and Wasi, N. 2012. Estimation of discrete choice models with many alternatives using random subsets of the full choice set: with an application to demand for frozen pizza.

Korobkin, R. 1997. Status quo bias and contract default rules. *Cornell L. Rev.,* 83**,** 608.

Lanz, B. and Provins, A. 2015. Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics,* 47**,** 300-324.

Louviere, J. J. 2001. Choice experiments: an overview of concepts and issues. *The Choice Modelling Approach to Environmental Valuation***,** 13-36.

Louviere, J. J., Hensher, D. A. and Swait, J. D. 2000. *Stated choice methods: analysis and applications*, Cambridge University Press.

Maltz, A. and Romagnoli, G. 2015. The effect of ambiguity on status quo bias: an experimental study. Working Paper.

Mandler, J. M. 2004. *The Foundations of Mind: Origins of Conceptual Thought*, Oxford University Press.

Marsh, D., Mkwara, L. and Scarpa, R. 2011. Do respondents' perceptions of the status quo matter in non-market valuation with choice experiments? An application to New Zealand freshwater streams. *Sustainability,* 3**,** 1593-1615.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour. *in Frontiers in Econometrics, ed.* P. Zarembka. New York: Academic Press, 105–142.

Meyerhoff, J. and Liebe, U. 2009. Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity. *Land Economics,* 85**,** 515-528.

Moon, A. 2004. Assessing the impacts of complexity in stated preference methods. Unpublished MSc thesis, Department of Rural Economy, University of Alberta, Edmonton, Alberta Canada, 126.

Muller, M., 2008. Free basic water - a sustainable instrument for a sustainable future in South Africa. *Environment and Urbanization*, 20, 67-87.

Oehlmann, M., Meyerhoff, J., Mariel, P. and Weller, P. 2017. Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management,* 81**,** 59-73.

Orzechowski, M., Arentze, T., Borgers, A. and Timmermans, H. 2005. Alternate methods of conjoint analysis for estimating housing preference functions: Effects of presentation style. *Journal of Housing and the Built Environment,* 20**,** 349-362.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J. and Yazdizadeh, A. 2017. Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning,* 157**,** 63-74.

Ritov, I. and Baron, J. 1992. Status-quo and omission biases. *Journal of Risk and Uncertainty,* 5**,** 49-61.

Rose, J. M. and Bliemer, M. C. 2009. Constructing efficient stated choice experimental designs. *Transport Reviews,* 29**,** 587-617.

Saldías, C., Speelman, S., Van Huylenbroeck, G. and Vink, N. 2016. Understanding farmers' preferences for wastewater reuse frameworks in agricultural irrigation: lessons from a choice experiment in the Western Cape, South Africa. *Water SA,* 42**,** 26-37.

Samuelson, W. and Zeckhauser, R. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty,* 1**,** 7-59.

Scarpa, R., Willis, K. G. and Acutt, M. 2007. Valuing externalities from water supply: status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management,* 50**,** 449-466.

Schweitzer, M. 1994. Disentangling status quo and omission effects: an experimental analysis. *Organizational Behavior and Human Decision Process* 58**,** 457-476.

Snowball, J., Willis, K. and Jeurissen, C. 2008. Willingness to pay for water service improvements in middle-income urban households in South Africa: a stated choice analysis. *South African Journal of Economics,* 76**,** 705-720.

Statistics South Africa 2011. National Census 2011. Pretoria, South Africa.

Street, D. J., Burgess, L. and Louviere, J. J. 2005. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing,* 22**,** 459-470.

Tversky, A. and Kahneman, D. 1991. Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics,* 106**,** 1039-1061.

Tykocinski, O. E., Pittman, T. S. and Tuttle, E. E. 1995. Inaction inertia: foregoing future benefits as a result of an initial failure to act. *Journal of Personality and Social Psychology,* 68**,** 793.

Vriens, M., Loosschilder, G. H., Rosbergen, E. and Wittink, D. R. 1998. Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *Journal of Product Innovation Management,* 15**,** 455-467.

Zhang, J. and Adamowicz, W. L. 2011. Unraveling the choice format effect: A context-dependent random utility model. *Land Economics*, 87**,** 730-743.

Appendix 3.1: Example of the questionnaire used for townships

NELSON MANDELA UNIVERSITY

HOUSEHOLD WATER PROVISION IN DURBAN

Name of interviewer: ______________________________________

Location: ____

Date of the interview: ____/___/2016

To improve on the provision of water services to households in the eThekwini Metropolitan Municipality, households' preferences for water services should be established. This questionnaire seeks to elicit households' willingness to pay for water services packages, given varying levels of attributes such as access to piped water, reliability of water supply, water pressure, water quality, and cost per month. The questionnaire has three sections. Section A of the questionnaire provides 6 choice-sets. Section B of the questionnaire provides some general questions on water provision in the municipality. Section C of the questionnaire gathers personal information of the respondents, to help us understand factors that affect the way people feel about water provision in the municipality.

SECTION A: CHOICE EXPERIMENTS

Several attributes may be noted about how portable water is provided to households. These attributes include the way households receive piped water, reliability of water supply, water pressure, water quality, and monthly cost. The table below shows these attributes and their levels which are used in the choice experiments.

Attribute	Description	Attribute Levels
Piped water	Access to piped or tap-water in the dwelling, off-site or on-site. This shows how piped water is delivered to households.	**Level 1:** Inside dwelling **Level 2:** In yard **Level 3:** Community tap: less than 200m from dwelling **Level 4:** Community tap: greater than 200m from dwelling **Level 5:** No access to piped water
Reliability of water supply	Whether the household had any interruption in piped water supply in the last one month.	**Level 1:** Yes **Level 2:** No
Water pressure	Pressure is the force that pushes water through pipes. Water pressure determines the flow of water from the tap.	**Level 1:** High water pressure **Level 2:** Low water pressure
Water quality	A measure of the suitability of water for a particular use based on selected physical, chemical and biological characterises.	**Level 1:** Safe to drink **Level 2:** Has colour **Level 3:** Has a taste **Level 4:** Has a smell
Cost	Cost per month.	**Level 1:** R120 **Level 2:** R220 **Level 3:** R400 **Level 4:** R680 **Level 5:** R980

Six (6) choice-sets generated from the table above are shown below. In each choice-set, assume that the four alternatives (Status-Quo, Alternative 1, Alternative 2 and 'None') were the only water provision options available in the eThekwini municipality. If you do not like any of Alternative 1, Alternative 2 or the Status Quo, then please choose the box marked 'None'. You may notice that while the levels of the attributes of the status quo always stay the same, the levels in the columns of alternatives 1 and 2 change in each choice-set. It is very important to consider each choice-set based on its own outcomes, regardless of whether the prior and subsequent choice-sets provide better packages.

We would like to know which option you prefer the most in each choice-set. Some options may seem unrealistic; please note that we do not describe how each option would be brought about.

Choice-Set 1

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	Inside dwelling	Community tap: greater than 200m from dwelling	
Reliability of water supply	No	No	Yes	
Water pressure	Low pressure	Low pressure	High pressure	
Water quality	Safe to Drink	Safe to drink	Has a smell	
Cost per month	Free	R220	R400	
I WOULD CHOOSE:	☐	☐	☐	☐

Choice-Set 2

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	Inside dwelling	Community tap: less than 200m from dwelling	
Reliability of water supply	No	Yes	No	
Water pressure	Low pressure	High pressure	Low pressure	
Water quality	Safe to Drink	Safe to drink	Has colour	
Cost per month	Free	R980	R120	
I WOULD CHOOSE:	☐	☐	☐	☐

Choice-Set 3

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	No access to piped water	Inside dwelling	
Reliability of water supply	No	Yes	No	
Water pressure	Low pressure	High pressure	Low pressure	
Water quality	Safe to Drink	Has colour	Has a taste	
Cost per month	Free	R400	R120	
I WOULD CHOOSE:	☐	☐	☐	☐

Choice-Set 4

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	In yard	Inside dwelling	
Reliability of water supply	No	Yes	No	
Water pressure	Low pressure	Low pressure	High pressure	
Water quality	Safe to Drink	Has colour	Has colour	
Cost per month	Free	R120	R680	
I WOULD CHOOSE:	☐	☐	☐	☐

Choice-Set 5

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	Community tap: less than 200m from dwelling	No access to piped water	
Reliability of water supply	No	No	Yes	
Water pressure	Low pressure	High pressure	High pressure	
Water quality	Safe to Drink	Has a smell	Safe to drink	
Cost per month	Free	R120	R980	
I WOULD CHOOSE:	☐	☐	☐	☐

Choice-Set 6

	STATUS-QUO	ALTERNATIVE 1	ALTERNATIVE 2	NONE
Piped water	In yard	Community tap: greater than 200m from dwelling	In yard	
Reliability of water supply	No	No	Yes	
Water pressure	Low pressure	Low pressure	Low pressure	
Water quality	Safe to Drink	Has a taste	Safe to drink	
Cost per month	Free	R680	R220	
I WOULD CHOOSE:	☐	☐	☐	☐

SECTION B: GENERAL QUESTIONS

1. **When you made your choices, which attributes mostly influenced your decision?** *(Tick the appropriate box/boxes).*

☐ Piped water ☐ Water pressure
☐ Reliability of water supply ☐ Cost per month
☐ Water quality

2. **How does your household receive water?** *(Tick the appropriate box).*

☐ Piped water in dwelling ☐ Piped water in yard
☐ Piped water on community tap ☐ Water tanker
☐ Borehole ☐ Spring, river or stream

3. **Do you receive free basic water?** *(Tick the appropriate box):*

☐ Yes ☐ No

4. **How often do you experience water interruptions?** *(Tick the appropriate box).*

☐ Very often ☐ Occasionally ☐ Not at all

5. **Describe the quality of the water you receive** *(Tick the appropriate box).*

☐ Not clear ☐ Bad Taste ☐ Bad smell ☐ Has colour
☐ Other (Please specify): ___________________________________

SECTION C: PERSONAL INFORMATION

6. **How many people are members of your household?** *(Write the number in the box):*

☐

7. **Are you the head of the household?** *(Tick the appropriate box):*

☐ Yes ☐ No

8. **What is your gender?** *(Tick the appropriate box):*

☐ Male ☐ Female

9. **What is your marital status?** *(Tick the appropriate box).*

☐ Single ☐ Married ☐ Other (*Please specify*): __________________

10. **Do you have any infants in your family?** *(Tick the appropriate box):*

☐ Yes ☐ No

11. **Do you have any children under the age of 16 years?** *(Tick the appropriate box):*

☐ Yes ☐ No

12. Which racial group do you belong to? *(Tick the appropriate box. You may choose not to answer this question):*

☐ Black/African ☐ White ☐ Indian/Asian ☐ Coloured

13. What is your highest education level? *(Tick the appropriate box):*

☐ Never attended school ☐ Primary school

☐ High school ☐ Certificate

☐ Diploma ☐ Degree

☐ Postgraduate ☐ Other *(Please specify)*: ______________________

14. Which of the following is your age category?

☐ 16-24 years ☐ 25-34 years

☐ 35-44 years ☐ 45-54 years

☐ 55-64 years ☐ 65^{+} years

15. What is your household's main source of income? *(Tick the appropriate box):*

☐ Salary/Wages ☐ Income from investments

☐ Government grant ☐ Allowances/hand-outs

☐ Pension ☐ Other *(Please specify)*: ______________________

16. What is your household's monthly average income? *(Tick the appropriate box):*

☐ Less than R2 500 per month

☐ Greater than R2 500 but less than R5 000 per month

☐ Greater than R5 000 but less than R10 000 per month

☐ Greater than R10 000 per month

Thank you for taking the time to answer this questionnaire.

Appendix 3.2: Correlation matrices for random parameters

Table A1: Correlation matrices for random parameters in all blocks and sub-samples

	Cost	Pipe	Reliability	Pressure	Quality
Suburbs Block 1					
Cost	1.000	-0.623	-0.638	0.224	0.550
Pipe	-0.623	1.000	0.284	-0.430	-0.515
Reliability	-0.638	0.284	1.000	0.562	-0.320
Pressure	0.224	-0.430	0.562	1.000	-0.052
Quality	0.550	-0.515	-0.320	-0.052	1.000
Suburbs Block 2					
Cost	1.000	-0.212	0.640	-0.186	-0.543
Pipe	-0.212	1.000	-0.443	0.203	-0.273
Reliability	0.640	-0.443	1.000	-0.240	-0.478
Pressure	-0.186	-0.203	-0.240	1.000	-0.216
Quality	-0.543	-0.273	-0.478	-0.216	1.000
Townships Block 1					
Cost	1.000	-0.754	-0.284	0.225	0.568
Pipe	-0.754	1.000	-0.292	-0.150	0.070
Reliability	-0.284	-0.292	1.000	0.440	-0.574
Pressure	0.225	-0.150	0.440	1.000	0.009
Quality	0.568	0.070	-0.574	0.009	1.000
Townships Block 2					
Cost	1.000	-0.652	0.232	-0.075	0.003
Pipe	-0.652	1.000	-0.670	-0.418	0.573
Reliability	0.232	-0.670	1.000	0.570	-0.776
Pressure	-0.075	-0.418	0.570	1.000	-0.141
Quality	0.003	0.573	0.776	-0.141	1.000

Chapter 4: The effects of presentation formats in choice experiments

Abstract

Although stated-preference surveys take various forms, the use of either text or visuals to represent attributes is uncontroversial, and they remain the commonly used formats. While prior research has investigated the impact of these commonly used formats in other disciplines, little is known about their effects on results in terms of relative importance in environmental economics literature. We compare three presentation formats, namely text, visuals, and both text and visuals. We test whether these three formats generate differences in estimated utilities and willingness to pay. This research sheds light on how to develop a valid presentation method for attribute levels in choice experiments, which is critical considering most environmental economics goods and services are not traded in the market. Our results show that the visuals format generates more statistically significant coefficients than the other formats. This suggests that the presentation format has significant impacts on choice. The choice between the three elicitation formats may imply a trade-off in choice precision. Our findings suggest that more research on presentation formats in environmental economics is warranted.

Keywords: choice experiments, format, text, presentation, visuals

4.1. Introduction

In most choice experiment (CE) studies in environmental economics, the attributes of non-traded environmental goods are communicated to respondents in the form of a table consisting of verbal descriptions (written text). The table normally consists of attributes, their detailed descriptions, and levels. Thereafter, respondents are presented with a series of choice sets, which are often in the form of written text. It is assumed at this stage that the respondents can fully comprehend the attributes and the attribute levels. Respondents are expected to form their preferences in response to the information provided to them in choice sets pertaining to the environmental good in question (Bateman et al., 2009).

Respondents form their preferences by cognitively combining the utilities they derive from the attribute levels that make up choice alternatives according to some function. Information plays an important role in the formation of preferences, particularly for the estimation of value for non-traded environmental goods/services, where experience of the good/service and the hypothetical market may be limited (Munro and Hanley, 2001). According to Green and Tunstall (1999), the accuracy and 'face-value' comprehension of information provided to respondents with the non-traded valuation studies should not be taken for granted.

We argue that in addition to the information presented to respondents, the presentation format in which the information is conveyed to respondents may also influence how preferences are formed. Presentation format pertains to the way the attribute alternatives are presented. This aspect is taken for granted in the environmental economics literature. Evidence on the influence of format on how preferences are formed has been observed in the literature of other fields, such as housing (see Timmermans and van Noortwijk, 1995; Wang and Li, 2004), urban planning (see Jansen et al., 2009) and consumer studies (see Townsend and Kahn, 2013). This literature attempts to address concerns about whether respondents can truly articulate their preferences, if their responses are an artefact of the experimental task, and if they can fully comprehend the typical presentation format often used to convey attribute levels, which can be complex – particularly for unfamiliar goods.

It has been argued that the presentation of attribute levels may be captured better graphically or visually. On the other hand, respondents may pay more attention to certain features of the visuals in the experiment. Moreover, some attribute levels (such as cost or other monetary attributes) may not lend themselves to visual representation (Orzechowski et al., 2005). Our study is designed to contribute to the limited but growing literature pertaining to whether presentation formats matter in choice experiments. To be specific, our study reports on the findings of a test on whether written text, visual representations or a combination of written text and visuals for attribute profiles in choice experiments generate differences in estimated empirical results.

This paper attempts to investigate whether the text, visuals or text-and-visuals presentation formats matter for discrete choice experiments in environmental and resource economics. The objective of the study is to test whether presenting attribute levels in these three presentation styles generates significantly different results with respect to attribute interpretation, relative importance, probability of adopting water conservation technologies, and willingness to pay estimates. Households completed a CE questionnaire that contained three versions of the same six choice tasks with three alternatives (status quo, alternatives 1 and 2), in which the attribute levels were presented in text, visuals or text-and-visuals. The status quo was undefined, as only the households knew their current situation. Five attributes relating to the decision of households to adopt water-saving technologies were included. Mixed logit models were used to estimate the relative importance of the attribute levels.

The contribution of this study is twofold. Firstly, more presentation formats are evaluated than in most studies in the literature; most studies compare either written text to visuals, or written text to a combination of text and visuals (see Jansen et al., 2009; Muller et al., 2010; Patterson et al., 2017; Townsend and Kahn, 2013). Our study compares three formats. Secondly, to the best of our knowledge, this is the first study in environmental economics to examine the impact of the way in which attribute profiles are presented. Evidence from other disciplines suggests that presentation formats matter (see Arentze et al, 2003; Bateman et al., 2009; Orzechowski et al., 2005). It is not clear in the environmental economics literature how to develop valid presentation methods for attribute levels in choice experiments; the investigation carried out in this study therefore sheds light on this, and so doing, contributes to the establishment of guidelines to developing valid presentation formats.

Presentation formats such as visuals improve respondents' understanding of the goods/services involved. However, considering the nature of environmental economics goods and services, coming up with the most appropriate visuals requires a lot of effort and resources. If less appropriate visuals are used to depict attribute profiles, there is a good chance that such visuals could contain various distracting effects that might bias respondents' choices (Scarpa et al., 2009). In such instances, a text presentation might be better, as it would provide clearer and more precise descriptions of the environmental good/services involved. A combination of both efforts may yield even better results, as it would combine the strengths of both approaches. Considering the effort and precision needed to come up with the most appropriate visuals, it is important to examine whether such effort does in fact improve the quality of data collected.

The rest of the paper is organised into seven sections. Section 4.2 reviews the literature. Section 4.3 presents the experimental design. Section 4.4 discusses the case study. Section 4.5 discusses the modelling approaches. Section 4.6 presents the experimental data. Section 4.7 presents and discusses the empirical findings, and Section 4.8 concludes the study.

4.2. Literature on presentation formats

The issue of presentation formats is not new. Many studies in the neuropsychology literature discuss the various merits and demerits of presenting survey instruments as text, visuals, or a combination of both. Early contributions on presentation formats are found in Holbrook and Moore (1981), Childers and Houston (1984) and MacInnis and Price (1987). These studies explain how respondents process information presented either as text or visuals. A common conclusion is that information presented as text and information presented as visuals are processed differently, and by different areas of the brain. Over the years, the discussion regarding presentation formats has continued in the literature; however, only in recent years have studies emerged in the literature that examine the role of presentation formats in the choice experiment domain.

Several advantages of text presentations are discussed in the literature. The commonly identified advantage is that text presentations provide clear and appropriate descriptions of attributes and levels, as they do not have the problem of attribute interaction associated with visual presentations. Typically, text presentations do not result in the distracting effects intrinsic to visual presentations (Scarpa et al., 2009). According to Vriens et al. (1998), text presentations facilitate judgment, making it possible for respondents to make real trade-offs between given attributes and levels. These advantages are consistent with the psychological literature, which suggests that visuals dominate attributes in terms of colour and form, and may distort responses (see Holbrook and Moore, 1981; Wittink et al., 1994). As such, Patterson et al. (2017) and many other studies recommend only using visuals ahead of text when it is absolutely necessary.

However, the literature also demonstrates the advantages of visual representation. Predominantly, the use of visuals is supported in psychology literature by studies that argue that respondents are inclined to process images more readily than written text (Berlyne, 1971; Childers and Houston, 1984; Hetherington et al., 1993; Wohlwill, 1976). It is argued that visuals improve respondents' understanding and comprehension of the survey instrument. This is because it is a relatively less irritating cognitive process to perceive cues depicted in visuals than to perceive those in text (Fitzsimons et al., 2002). This assertion is consistent with the argument in Childers and Houston (1984) that advertisements presented using visual representations are remembered easily compared to those provided by verbal representations.

There is no consensus in the literature on which of the two formats to adopt when designing survey instruments. In the choice experiments literature, written text is the commonly used format for presenting choice sets (see Abdullah and Mariel, 2010; Arentze et al., 2003; Bhaduri and Kloos, 2013; Lanz and Provins, 2015; Vásquez et al., 2012). However, some choice experiment studies combine both text and visuals, to capitalise on the individual benefits of each presentation format (see Kanyoka et al., 2008; Snowball et al., 2008; Saldías et al., 2016). Importantly, most of the studies that combine text and visuals only include visuals in the table when explaining the attributes and levels. Very few include visuals in the actual choice sets. Where visuals are included in the choice sets, they are normally limited to attributes, with very few studies including them in the profiles of each choice set. Our study includes visuals to represent both the attributes and the attribute levels in choice profiles.

A clear link between presentation format and the preferences of respondents does not exist in the choice experiments literature. This has prompted an emerging interest in examining the impact of presentation formats; but the studies that test the impact of presentation formats in choice experiments are mostly in domains other than environmental economics. Most of these studies report inconsistent results (see Bateman et al., 2009; Jansen et al., 2009; Lovett et al., 2015; Muller et al., 2010; Patterson et al., 2017). While some report that text presentations and visual representations affect empirical results, others show that using either of the two formats has no meaningful effect on empirical estimations. Our study joins this debate, assessing the impact of text presentations, visual representations, and text and visual representations together on empirical estimates. By including an experiment in which information is presented only in visuals, our study is a step ahead of most similar studies in the literature.

Jansen et al. (2009) test the impact of including visuals in choice experiments on housing preferences. The study reveals that including visuals in the choice sets led to several differences in the results compared to those from text presentations. These differences are explained as emanating from accidental details in the images. Coming from a different angle, Bateman et al. (2009) test the impact of text presentations and visual representations in a choice experiment on coastal land use. The study found text presentations generated higher gain/loss asymmetry than visual representations. Differences between results from text presentations and visual representations are further confirmed in Syrengelas (2017), in which different presentation formats yielded different welfare estimates. Other choice experiment studies to find that presentation formats affect empirical results include Muller et al. (2010) and Orzechowski et al. (2005). Most of these studies find that visual representations tend to produce more parameter estimates that are statistically significant and with larger absolute coefficients, compared to text presentations.

On the other hand, Patterson et al. (2017) use choice experiments on preferences for landscape and urban planning to test the impact of presentation formats, and find no evidence of major differences between the results from text presentations and those from visual representations. The study reports that respondents' preferences in the text survey were based on their mental images; whereas in the visuals survey, preferences were based on the displayed images. Despite this, similar results were reported from the two separate experiments. Findings from Patterson et al. (2017) are consistent with earlier work by Arentze et al. (2003), which used choice experiments on choice of transport mode to test the impact of presentation formats on empirical results. The study revealed that including visuals to text for attributes affected neither the error variance nor the measurement of attribute weights; as such, the effort it takes to develop pictorial material is not compensated for by better-quality data.

It is evident from this section that the existing choice experiment literature on presentation formats is predominantly in domains other than environmental economics. Our current study attempts to address this by examining concerns around the way choice experiment profiles are presented in environmental economics. As revealed in this section, the few emerging choice experiment studies in the literature compare text-only experiments to those employing text and visuals. Our study is a step ahead of these studies because in addition to comparing text experiments to text-and-visual experiments, it also compares a visuals-only experiment to the other two. More precisely, our current study compares three presentation formats, as opposed to the more common procedure of comparing only two formats. In comparing three presentation styles, hopefully the study will shed light on the extent of the bias found in the empirical estimates produced by the environmental economics literature.

4.3. Case study: households' willingness to adopt water-saving technologies

One of the biggest criticisms of environmental and resource economics choice experiments is that the goods and services being evaluated are not traded in the market, so respondents are not familiar with them; hence they may find making trade-offs very difficult. This suggests that the behaviour underlying CE results is not well understood. It is therefore possible that by default, respondents may resort to a simplified decision rule – particularly in instances where choices are too difficult. There are ongoing debates about the complexity of choice tasks in environmental-related choice experiments, and the extent to which respondents can comprehend choice tasks as intended in the experiment. Despite this, there is increasing use of CE experiments in environmental economics. Since some of these experiments may be used for policymaking, the accuracy and validity of the measured preferences are key to avoiding incorrect policy choices based on invalid experiments.

Because of global warming and growing water scarcity, policymakers are increasingly exploring ways to conserve water. As households are among the biggest water users, they are often targeted by decision-makers using a variety of tools, some of which are intended to change user consumption behaviour through the adoption of conservation technologies. The case study in this study therefore has serious policy implications; hence, it is essential that participants in the experiment understand the choice tasks as fully as possible, to reveal their true preferences. It is critical that participants in CEs better understand the included attribute-related information, in order to make choices that reflect preferences more accurately.

A failure on the part of the households in our experiment to understand the adoption of water-saving technologies will put the experiment at risk. In this area, little is known about how much of the attribute-related information respondents understand. This is worrisome, considering the increasing use of this technique in environmental economics. One of the reasons for failure to fully comprehend the experiment is lack of understanding of the attribute levels in the experiment. There is great diversity in the way environmental economics information is translated into attribute levels, in how they are explained to participants, and in how choice tasks are presented in CEs on environmental or environmentally related topics.

It is argued that the presentation format used in an experiment may impact the respondents' understanding of the information on attribute level contained in the experiment. Considering that in environmental economics most studies use text only, it is plausible that visual attribute-level communication might help to make choice tasks easier, and therefore improve the quality of respondent feedback. It is also plausible that attribute-level information that contains both text and visuals may yield more consistent results. The question is whether any presentation format gives more consistent responses, and whether the participants would prefer one format over another. There is no empirical evidence on this score in the environmental economics literature. The adoption of water-conservation technologies can be presented easily using either format, so it is deemed suitable for the purposes of this study. We therefore test for differences in CE results using the three presentation styles.

4.4. Experimental design

The first step in choice experiments involves selecting relevant attributes and assigning realistic levels to each attribute. Selected levels assigned to each attribute should be feasible, realistic, non-linearly spaced, and should span the range of respondents' preference maps (Hanley et al., 2001). Both attributes and levels can be deduced from a literature review, focus groups, pilot studies, and expert consultations. After attributes are identified and relevant levels are assigned to each attribute, experimental design commences. Experimental design is explained as the specialised and scientific manipulation of the levels of one or more attributes to generate choice profiles (Hensher et al., 2015). The most common classes of experimental design in the literature are full factorial, orthogonal, and efficient designs. This section discusses the attributes and levels used in the study, as well as how they are experimentally designed into choice profiles.

4.4.1. Attributes and levels

Choice experiments are primarily conducted to determine the independent influence of different attributes on the choices that are observed to be made by surveyed respondents. In determining the attributes for our experiment, we used a combination of both literature review and expert consultation. The literature shows that household water-efficient technologies can be categorised based on areas in a home. A typical South African middle-income household of four spends 25% of their water use in flushing the toilet, 25% on garden and outdoor activities, 24% on bathing or showering, 13% on laundry, 11% in the kitchen, and 2% on other activities (Price, 2009). We use these areas as attributes and adopt the various technologies that may be fitted into each of these areas as levels. From a series of expert consultations, four key areas were adopted as attributes, namely kitchen, shower, toilet and garden/outdoor. Choice experiments also include a monetary attribute, which is essential for measuring social welfare (Hensher et al., 2015). We include the monthly water bill as our monetary attribute.

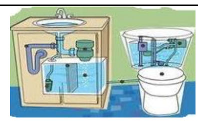
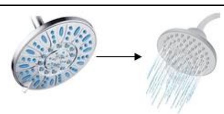
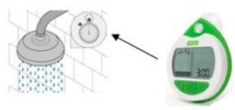
Several water-efficient technologies that can be installed to save water in a homestead are identified in the literature (see Hering and Ingold, 2012; Jones and Hunt, 2010; Makki et al., 2013; Mini et al., 2015; Still and Bhagwan, 2008; Willis et al., 2013). After consulting with experts, our study has adopted water-efficient technologies that are deemed necessary in the South African context. For the kitchen devices, as levels we use efficient dishwashers, efficient taps, and a system for collecting used water. In terms of shower devices, we use efficient showerheads and shower timers as levels, while for toilet devices we use dual-flush cisterns, interruptible (multi) flush cisterns, and cistern displacement devices (hippo bags) as levels. For garden/outdoor devices, the levels are time-based irrigation controllers, micro-drip irrigation systems, and water tanks for harvesting rainwater[18].

Investing in water-efficient technologies essentially reduces a household's monthly water bill. Therefore, the various possibilities for reduced monthly water bills are used as levels for the monetary attribute. To determine these possibilities, we consider the average monthly water bill for households in the three selected municipalities. Using data collected by the National

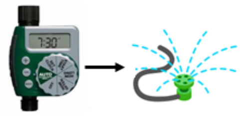
[18] We agree that the technologies chosen as levels may also be used as attributes in other studies. However, in the context of our study the emphasis is on the areas in a home where households can save water by installing efficient technologies (i.e. kitchen, shower, toilet and outdoor). As such, water-efficient devices that can be fitted in these areas are used as levels in our choice experiment.

Income Dynamics Study (NIDS), an average of R450 (around $31.47)[19] per month was determined to be the current average water bill. If households were to adopt water-saving technology, their bill would be reduced by 75%, 50% or 30%, that is, from R450 to one of R110, R225 or R315 ($7.69, $15.73 or $22) per month, respectively. The final list of attributes and levels is presented in Table 4.1.

Table 4.1: Attributes and levels used in the study

Attribute	Description	Attribute Levels	
 **Kitchen devices**	A typical household uses about 11% of total water use in the kitchen. A standard tap flows at about 8l per minute. Installing water-flow regulators or tap-head aerators saves water and makes a standard tap more efficient by 60%. An efficient dishwasher uses 15l per cycle, using 50% less water than is used in a conventional dishwasher.	**Level 1:** Efficient dishwasher	
		Level 2: Efficient tap	
		Level 3: System collecting used water	
Shower devices 	A typical household uses about 24% of total water in the shower. Shower timers result in shorter showers. Efficient showerheads save 65% of water used in the shower.	**Level 1:** Efficient shower head	
		Level 2: Shower timer	
Toilet devices 	A typical household uses about 25% of total water use in the toilet. Replacing a 12l cistern with a 3l dual cistern saves about 75% of water. An interruptible flush cistern allows users to control how long the toilet flushes. Hippo bags displace water in the cistern and save about 1.2l per flush.	**Level 1:** Dual flush cistern sized 3-6L	
		Level 2: Interruptible flush cistern	
		Level 3: Cistern displacement (hippo bag)	

[19] As at 24 October 2018, US$1 = ZAR14.30.

Garden & outdoor devices 	A typical household uses about 25% of total water use on garden/outdoor activities. Efficient gardening technologies reduce water use by 30%. These include time-based irrigation control and micro-drip systems. Irrigating gardens using water collected with water tanks also saves water.	**Level 1:** Time-based irrigation controller	
		Level 2: Micro-drip systems	
		Level 3: Use harvested rain water	
Monthly water bill 	The average water bill for a household is R450 per month. Installing water-efficient technologies will reduce the monthly water bill by 30%, 50% or 75%.	**Level 1:** R110 **Level 2:** R225 **Level 3:** R315	

The attributes and levels presented in Table 4.1 are experimentally designed into choice set profiles. In addition to the designed profiles, we also include a status quo (SQ) profile. An SQ profile essentially avoids the undesirable effects associated with forced choices (Dhar and Simonson, 2003; Ferrini and Scarpa, 2007). Our study uses an undefined SQ in each choice set. An undefined SQ is an individual-specific SQ in which where each respondent envisages their own current status and compares it to the experimentally designed hypothetical options (Hess and Rose, 2009). Undefined SQs and similar approaches are common in the literature (see Campbell et al., 2008; Hess and Rose, 2009; Marsh et al., 2011; Scarpa et al., 2007; Train and Wilson, 2008; Willis et al., 2005). They are commonly used when it is difficult to ascertain the current situation for the sample. This is the case in our experiment, because we cannot determine the current use of water-efficient technologies with certainty. In South Africa, household water-conservation practices are not clearly documented. The use of an individual-specific SQ essentially avoids the problems associated with the risk of imposing an inapplicable SQ, which could have been the case in our study.

4.4.2. Description of design

This study uses an efficient design to generate the choice profiles presented to respondents. Efficient designs are praised in the literature for producing robust data that give more reliable parameter estimates with an even lower sample size than designs such as full factorial and orthogonal. According to Rose and Bliemer (2009), efficient designs give smaller widths of confidence intervals observed around the parameter estimates, and maximised asymptotic t-ratios for each parameter, thereby improving the reliability results. However, efficient designs are only efficient if prior parameters are known. If incorrect prior parameters are used, efficient designs become inefficient (Bliemer et al., 2008). To address this problem, the literature recommends drawing parameter estimates using the Bayesian parameter distributions. Bayesian parameter estimates are sensitive to misspecification of priors, because they assume prior parameter values to be approximately known and randomly distributed. Using a D-*error* efficient measure with prior parameters drawn by means of Bayesian parameter distribution, the design becomes a Bayesian D-*error* design (i.e. D_b-*efficient*) defined as:

$$D_b - error = \int_{\tilde{\beta}} \det\left(\Omega_1(X, \tilde{\beta})\right)^{1/K} \phi(\tilde{\beta} \mid \theta) d\tilde{\beta}. \tag{4.1}$$

D_b is the Bayesian design (where $'b'$ is from 'Bayesian'), Ω_1 is the asymptotic variance-covariance (AVC) matrix of the design, X is the experimental design, $\tilde{\beta}$ represents prior parameters, and K is the number of parameters to be estimated. This Bayesian D-*error* design is commonly used to examine efficient designs where the true population parameters are not known with certainty. According to Hensher et al. (2015), earlier studies assumed all population parameters to be zero (i.e. D_z–*error* = $\det(\Omega_1(X, 0))^{1/K}$); later on, studies assumed non-zero population parameters to be known with certainty (i.e. D_p–*error* = $\det(\Omega_1(X, \beta))^{1/K}$). In both assumptions, D-*errors* are functions of the experimental design X and the prior values $\tilde{\beta}$.

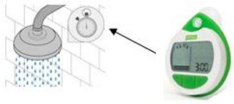
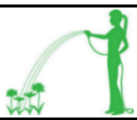
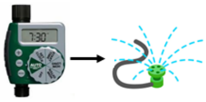
Using a normally distributed Bayesian D-*efficiency* design, this study experimentally designs six choice sets of two profiles each. Following suggestions by Bliemer et al. (2008) that the Gaussian method is the best approximation method for Bayesian efficient designs, we adopt the Gaussian method to come up with the number of draws for Bayesian priors. The rule of thumb for determining the absolute minimum Gaussian quadrature is 2^K, where K is the number of Bayesian priors. Given the number of attributes and levels in our experiment, we use the maximum possible Gaussian draws (i.e. 32 draws). These Gaussian draws are used in the normally distributed Bayesian D-efficient design adopted to populate the six choice sets of two profiles each that are experimentally designed in this study. The generated choice profiles are then presented to respondents using the three presentation formats being tested in this study. An example of a choice set used in the text experiment is given in Table 4.2.

Table 4.2: Example of a text-only choice set

	Status quo	**Option 1**	**Option 2**
Kitchen devices		Efficient dishwasher	System collecting used water
Shower devices		Shower timer	Efficient shower head
Toilet devices		Hippo bag	Dual-flush cistern
Garden/outdoor devices		Time-based irrigation controller	Use harvested rain water
Monthly water bill	R450	R225	R225
YOUR CHOICE			

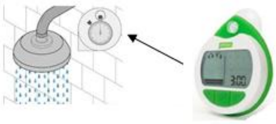
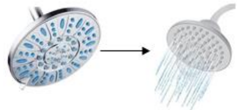
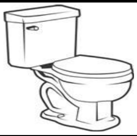
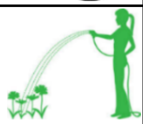
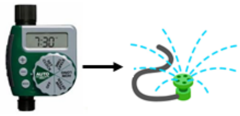
Respondents were asked to select their preferred profile from options 1 and 2, or they could opt out by choosing the status quo option. The same was done in the second experiment, in which attributes and levels were presented as visuals (except for the levels of the monetary attribute). Studies that use only visuals are rare in the literature, where most studies add text to visuals. The scarcity of such studies could be attributed to the disadvantages of visuals that are highlighted in the literature (see Holbrook and Moore, 1981; Vriens et al., 1998; Wittink et al., 1994). However, there are also several advantages of visuals over text representations, making it imperative to compare empirical results derived from the two formats. An example of the visual presentation choice set used in this study is given in Table 4.3.

Table 4.3: Example of a visual presentation choice set

	Status quo	Option 1	Option 2
			
			
			
			
	R450	R225	R225
YOUR CHOICE			

The visuals presented in each profile of Table 4.3 represents the same information that was presented as text in Table 4.2. Additionally, the designed choice profiles were also presented as both text and visuals. Our text-and-visuals experiment is a step ahead of most choice experiment studies, which only include images in the attributes (see Kanyoka et al., 2008; Snowball et al., 2008). Our choice sets include images in the profiles. Saldías et al. (2016) used this style to present choice sets in a study that elicited farmers' preferences for wastewater re-use frameworks in agricultural irrigation. Table 4 gives an example of the choice sets used in the text-and-visuals experiment.

Table 4.4: Example of a text-and-visuals choice set

	Status quo	Option 1	Option 2
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Shower timer 	Efficient shower head 
Toilet devices 		Hippo bag 	Dual-flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rain water 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

In addition to the choice experiment, our questionnaire also included other sections. The second section collected general information on households' water conservation behaviour and technology. Such information is essential, because a relationship between water-efficient technology and water-use behaviour is identified in the literature (see Davis, 2008; Freire-Gonzalez, 2011; Ghosh and Blackhurst, 2014; Smeets et al., 2014). The literature argues that households adopt non-efficient habits when they install efficient technologies.

The third section collects the socio-economic characteristics of the respondents, which are essential when establishing the drivers of respondents' choices. Various socio-economic characteristics of respondents are identified as key determinants of choices in the literature (see Millock and Nauges, 2010; Martinez-Espineira and García-Valiñas, 2013; Pérez-Urdiales and García-Valiñas, 2016)[20].

[20] It is important to note that the three questionnaires used in this study only differed in the formats used to present the choice experiment section. The information presented was similar across all three questionnaires. An example of the questionnaire used to collect information is given in Appendix 4.3.

4.5. Modelling

Developed from the random utility theory, choice experiments assume that individuals are rational decision-makers who choose the most preferred (utility-maximising) option when faced with a possible set of options (Abelson and Levy, 1985; Howard, 1977; McFadden, 1973). According to McFadden (1973), these rational individuals make choices based on the characteristics of the good, along with a random component. The random component could emerge from the uniqueness in the individual's preferences, or due to researchers having incomplete information about the individual observed (Ben-Akiva and Lerman, 1985). The literature proposes that the utility derived from an option by an individual is not known, but can be decomposed into a deterministic component and an unobserved random component, as follows:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{4.2}$$

Parameter U_{ij} represents the utility of individual i obtained from option j, parameter V_{ij} is the deterministic component which is normally specified as a linear index of the attributes in a choice set, and ε_{ij} is the unobserved random component of latent utility which captures the consequence for choice of uncertainty due to incomplete information. Equation 4.2 represents the basic utility function and may be expressed by decomposing the indirect utility function for individual U_{ij} into two main components (see Ben-Akiva and Lerman, 1985; Hensher et al., 2015; McFadden, 1973). If that occurs, the utility function then assumes the form:

$$U_{ij} = V_{ij}\left(X_{ij}, C_{ij}, \beta\right) + \varepsilon_{ij} \tag{4.3}$$

Equation 3 decomposes V_{ij} into attributes X_{ij} and C_{ij}. Parameter X_{ij} represents the vector of non-monetary attributes associated with option j, while parameter C_{ij} is the monetary attribute of option j, parameter β is the vector of preference parameters for the population in the sample, and ε_{ij} is the stochastic component (random term) with a zero mean. The utility function expressed in equation 4.3 can be expressed as linear in parameters as:

$$U_{ij} = \sum_{k=1}^{K} \beta_x X_{ij} + \beta_c C_{ij} + \varepsilon_{ij} \tag{4.4}$$

Random utility posits that any rational individual i chooses option j over option k if $U_{ij} > U_{ik}$. Each option consists of a bundle of attributes. When an individual selects one option over the other, it suggests that the hypothetical utility derived by the individual from the chosen option is greater than the utility of the other option not chosen (Greene, 2003; Louviere, 2001)[21]. Therefore, the probability P_i of selecting option j because $U_{ij} > U_{ik}$ is illustrated as:

$$P_i(j) = \text{Prob}\left(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}\right) \;\; \forall \;\; k \in C, k \neq j. \tag{4.5}$$

If the error terms are independently and identically distributed (IID) with an extreme value type I distribution, the variance of which is $var(\varepsilon)=\pi^2\tau^2/5$, where τ is a scale parameter that is used to normalise the model; then the choice probability of an option is expressed as:

$$P_{ij} = \exp\left(\frac{v_{ij}}{\tau}\right) / \sum_{k=1}^{K} \exp\left(\frac{v_{ik}}{\tau}\right) \tag{4.6}$$

[21] It is imperative to note that in deriving the probability of choosing an alternative within the random utility model, the choice of alternative with higher utility is not certain. The expectation has always been that there is a high chance that a respondent will choose the alternatives with higher utility.

Several logistic models are then used to estimate the probability defined in equation 4.6. The most basic of these logistic models is the conditional logit model (CLM). Also known as the multinomial logit (MNL) model if there are no choice-varying attributes, the model uses the maximum likelihood estimation approach (Hensher et al., 2015). MNL has enjoyed extensive use in the literature, and Hensher et al. (2015) identify it as the ‘workhorse’ for discrete choice experiments. However, MNL is criticised for assuming that respondents have homogenous tastes for observed attributes, and that the random part of the utility obeys the independence from irrelevant alternatives (IIA) as well as the independence and identical distribution (IID) properties. These assumptions are unrealistic, as they rule out persistent heterogeneity in taste for observed and unobserved product attributes (Greene, 2012; Hensher et al., 2015; Keane and Wasi, 2012).

Models that address the criticisms of MNL include the mixed logit (MXL – also known as the Random Parameter Logit, or RPL) model. Our current study uses the MXL model to examine the impact of presentation formats on empirical results. The MXL model is also used in many other studies that examine presentation formats (see Caussade et al., 2005; Syrengelas, 2017; Patterson et al., 2017). It allows coefficients to vary randomly across individuals, reflecting the reality that different respondents have different tastes and preferences for attributes in each choice set. The many advantages of MXL include its ability to account for both observed and unobserved heterogeneity in the preference parameters; and that it is versatile, with both single cross-sectional and panel data (Hensher et al., 2015). MXL breaks down coefficients into a population mean, and an unobserved individual’s deviation from that mean (Greene, 2012), as follows:

$$U_{ij} = \beta X_{ij} + \eta_{ij} + \varepsilon_{ij} \tag{4.7}$$

Parameter β in equation 4.7 is the population mean, while η_{ij} is the individual deviation from the population mean which shows the individual specific heterogeneity, with mean zero and standard deviation one (Greene, 2012). If θ is used to represent the distribution of the parameters of β, the probability of individual i choosing option j can therefore be represented as:

$$\overline{P}_{ij} = \int P_{ij}\, f(\beta \mid \theta) d\, \beta \tag{4.8}$$

Parameter P_{ij} represents the choice probability of an option as given in equation 4.6, while $f(\beta \mid \theta)$ is the probability density function for the coefficient β over the vector of parameter θ.

As is common in the literature, we also test for variations in welfare measures by examining the marginal willingness to pay (MWTP) estimates across the three presentation formats. MWTP estimates show the marginal rate of substitution (MRS) between each attribute and the monetary attribute; this is an important output of choice models, as it gives average estimates of what respondents are prepared to pay for or against each attribute (Hensher et al., 2015). Assuming a linear utility function with attribute X and a monetary attribute C , then:

$$U_{ij} = \beta X_j + \mu\left(C_i - p_j\right) + \varepsilon_{ij} \tag{4.9}$$

In the context of our study, it is essential to assess the impact of the presentation formats on welfare measures. This is because the literature reports inconsistent results on variations in the actual preferences and welfare measures (see Fitzsimons et al., 2002; Jansen et al., 2009; Lovett et al., 2015; Orzechowski et al., 2005; Patterson et al., 2017; Ro et al., 2009; Scarpa et al., 2009; Syrengelas, 2017; Vriens et al., 1998). Therefore, we examine whether variations in the MWTP would be observed across the three different presentation formats.

4.6. Experimental data

4.6.1. Data collection

The study is based on experimental data collected from 894 heads of households in the Gauteng Province, during the period November to mid-December 2017 and mid-January to February 2018. Survey instruments were prepared in English, and enumerators conversant in both English and other local languages were recruited from residents in the study area. These enumerators were trained and supervised during the data collection process. The survey, the main aim of which was to elicit the impact of presentation formats, collected stated-preference data on household preferences for water-efficient technologies. A split-sample survey was adopted, in which the first sub-sample was presented with a text experiment and data was collected from 232 respondents. The second sub-sample was presented with the same information, but using visual representations, and 257 complete responses were collected. The third sub-sample was presented with a questionnaire that combined both text and visual representations, and 405 complete responses were collected.

Respondents in each sub-sample were from the same residential area, hence their socio-economic characteristics could be expected to be similar. This is typical of South Africa, where historically, residential areas are clustered, mainly for socio-economic and historical reasons. Enumerators spent a week in a residential area collecting data using the first survey instrument. Once certain expected data points were achieved, enumerators would then move to another area, still using the first instrument. After collecting enough data from our targeted residential areas using the first instrument, enumerators went back to the same areas with the second questionnaire. However, on the second visit, different households from those interviewed in the first survey were interviewed. This process was followed until enough data points had been collected using the three questionnaires.

4.6.2. Descriptive statistics

The questionnaires used to collect information were made up of three sections. In the first section, respondents were presented with the choice experiment. The second section collected some general information on households' current water-use behaviour, as well as their current use of water-efficient technology. Such information is essential in determining people's choices. For example, households without water-efficient technologies installed – hypothetically – would prefer changes to their current water appliances compared to those with efficient technologies currently installed in their homes. The third section collected the biographical information of the respondent. The biographic characteristics collected include the respondent's gender, household size, education, age, marital status, race, income and source of income. The literature identifies these variables as key drivers in how respondents process information.

The information collected in the second and third sections of the questionnaire is essential in our study, which uses a split-sample approach and compares empirical results across sub-samples. For us to be able to compare empirical results across sub-samples, there should be some similarity and consistency in the socio-economic characteristics of the respondents across the sub-samples. The descriptive statistics of these biographical characteristics are given in Table 4.5. In addition to the three sub-samples, the table also presents statistics on pooled data[22].

Table 4.5: Descriptive statistics of respondents

	Text-only Mean	**Visuals-only Mean**	**Text and Visuals Mean**	**Pooled data Mean**
Number of respondents (N)	***232***	***257***	***405***	***894***
Male respondents (%)	52	64	51	55
Average household size	4	4	4	4
Average age	41	41	45	43
Married respondents (%)	57	45	57	53
Race (%):				
Black	79	89	82	84
White	8	3	13	9
Indian/Asian	9	4	4	5
Coloured	3	4	1	2
Education (%):				
Never attended school	0	0	1	1
Primary	3	1	3	2
High school	68	65	69	67
Certificate	14	21	13	15
Diploma	12	11	10	11
Degree	2	2	4	3
Postgraduate	1	1	1	1
Source of income (%):				
Salaries/wages	57	54	56	56
Business	22	26	20	22
Pension	13	6	17	13
Grants/allowances	3	3	3	3
Other	5	11	3	6
Monthly household income (%):				
<R5 000	34	34	43	38
R5 000 to R10 000	47	50	37	43
R10 000 to R20 000	17	16	19	17
R20 000 to R40 000	3	0	1	1
R40 000 to R60 000	0	0	0	0
>R60 000	0	0	0	0

[22] Pooled data is a combination of data from the three presentation formats. It is possible to combine these datasets because the information in all the three questionnaires is similar, i.e. the three questionnaires collected the same information.

Except for the number of respondents, where the text-and-visuals experiment had more respondents than the other two experiments, the descriptive statistics presented in Table 4.5 above show some consistency in socio-economic characteristics across the three experiments. Across all the experiments, there were slightly more male respondents than female respondents. Equally, in all the experiments there were more respondents belonging to the 'black' racial group than to the other groups. Most of the respondents in all three experiments had high school education and receive salaries or wages as their main source of income. The consistency of the socio-economic characteristics across the three experiments makes it possible for us to compare the empirical results estimated in each experiment.

4.6.3. Frequency distribution of efficient technologies and water-consumption habits

The current use of water-efficient technologies by households may have an impact on their choices. In South Africa, little is known about households' water-consumption behaviour, and the extent to which they make use of efficient technologies. Therefore, eliciting such information is essential for creating new knowledge useful for policymaking. Equally, the literature suggests a link between water-use behaviour and the installation of water-efficient technology. We asked respondents to indicate whether they currently have water-efficient technologies installed. Eight questions on water-efficient technologies were asked, using a four-point Likert scale with the options 'Yes', 'No', 'Not applicable' and 'Not sure'. To be specific, respondents were asked to indicate whether they currently have water-collection tanks, cistern displacement devices, water-flow regulators, efficient showerheads, efficient toilet cisterns, multi-flush toilet cisterns, dishwashers, and/or efficient garden devices. Except for the efficient toilet devices, the modal response for all technologies was 'No', indicating that households in our sample did not currently use water-efficient technologies. This result was observed consistently across all three experiments[23].

Furthermore, we elicited the possible reasons for not installing efficient technologies. Although there could be various reasons, respondents were asked to choose between 'I cannot afford', 'I did not know about them', 'I have no infrastructure to connect them', 'They are not important to me', and 'Other'. We assume that the reason for not installing water-efficient devices has an

[23] Full responses on the use of water-efficient technology are presented in Appendix 4.1, Table A1.

impact on both the respondents' choices and the format used to present the attribute levels. For example, respondents who 'did not know about water-efficient devices' are more likely to make informed decisions when the technologies are presented both textually and visually because presenting the technology as text only may not give enough information. Equally, respondents who 'do not have the infrastructure' to install certain technologies are likely to ignore choice profiles that contain such technologies. Therefore, it is imperative to elicit such information. Figure 4.1 presents the frequency distribution of the reasons for respondents' not installing efficient water technologies. The frequency distribution is reported for each of the three experiments.

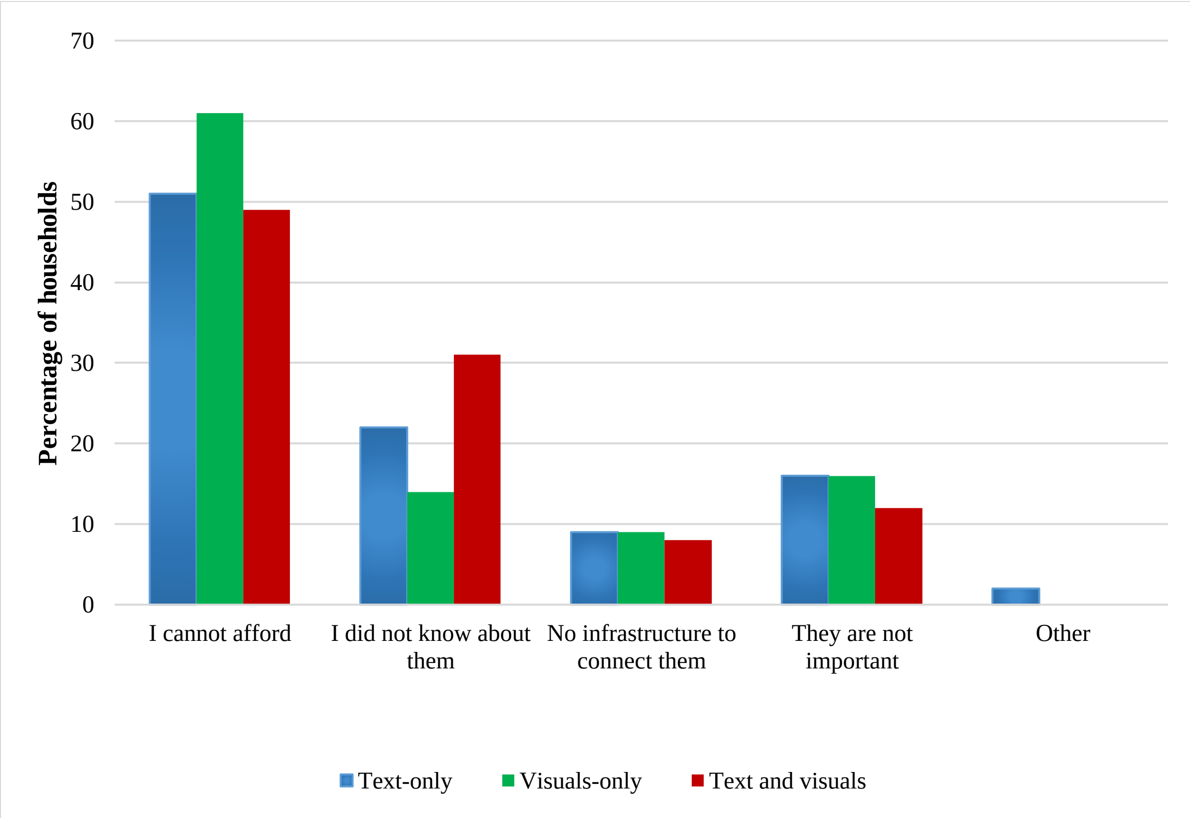


Figure 4.1: Reasons for not having water-efficient technology

Across all three experiments, the main reason for not adopting water-efficient technologies is that households 'cannot afford the technology'. Interestingly, respondents also indicated that they 'did not know about water-efficient technologies. This justifies the assertion by

Vloerbergh et al. (2007) that the nature of water makes it a low-involvement product, such that people do not think about it as long as it is available and does not have colour or smell or taste odd. Figure 1 above shows that the main reasons for not currently installing water-efficient technologies are consistent across the three experiments. This clearly shows some similarity in the respondents sampled in each experiment, which makes it possible to compare estimation results across the experiments.

Additionally, respondents were asked eleven behavioural questions, using a four-point Likert scale with the options 'Never', 'Once in a while', 'Always' and 'Not applicable'. The eleven questions asked were on inefficient water-use behaviour; if respondents indicated 'Never', it showed that they were practising efficient water-use behaviour, while if they indicated 'Always', it showed they were practising inefficient water-use behaviour. Generally, we observed that households in the sample practised efficient water-use behaviour. Responses were mostly consistent across the three experiments, which indicates that our sampled respondents possessed almost similar characteristics across sub-samples[24].

Finally, we present the frequency distribution of the stated preference choices. A presentation of the frequency distribution of how each alternative was chosen in each experiment is important when checking if choices were consistent across presentation formats. Where consistent choices are observed across formats, it makes the comparison of empirical estimates possible. However, if inconsistences are observed across formats, it implies that the presentation format affected respondents' choices. This information is shown in Figure 4.2.

[24] Complete frequency distributions of responses for each question are presented in Appendix 4.2, Table A2.

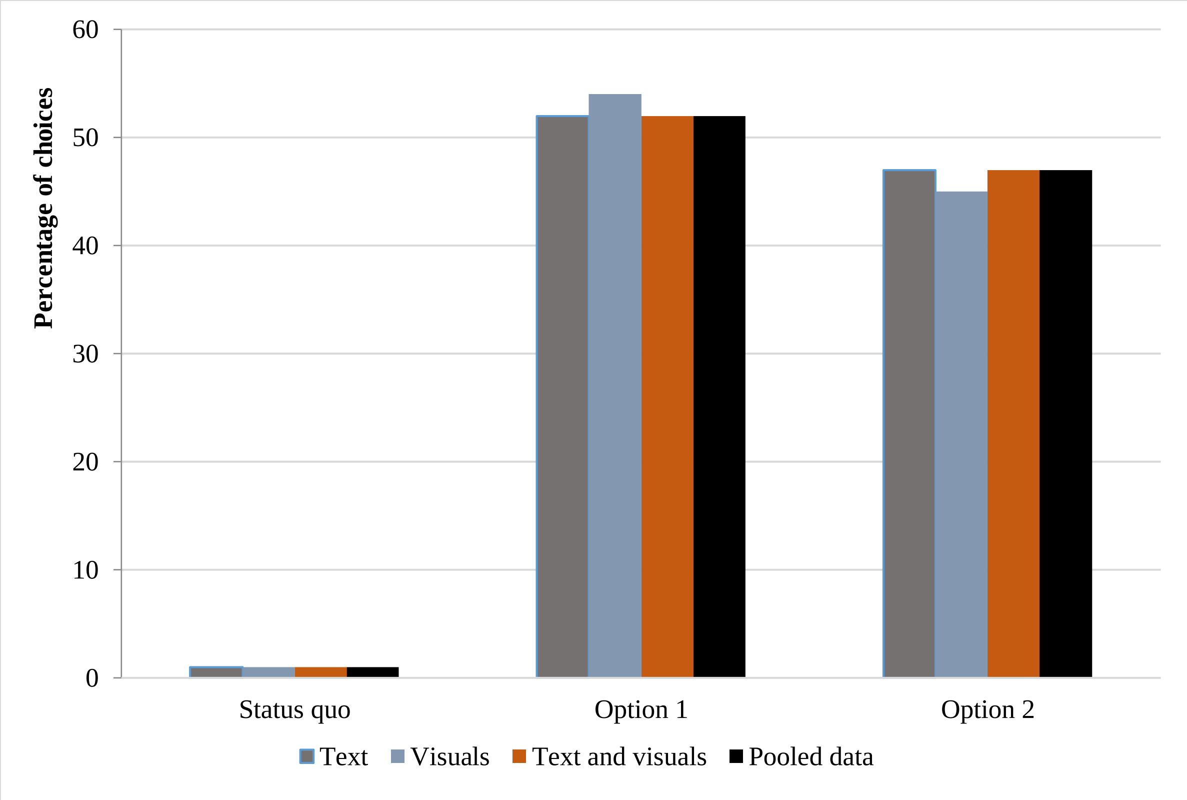


Figure 4.2: Frequency distribution of stated preference choices made by respondents

Figure 4.2 above shows that options 1 and 2 had an almost equal chance of being selected by respondents, implying that there were real trade-offs between the two options. In most choice experiments, the problem of status quo bias is reported, where respondents resort to choosing the status quo option they know as opposed to hypothetically designed options (Lanz and Provins, 2015; Marsh et al., 2011; Samuelson and Zeckhauser, 1988). The problem of status quo bias makes econometric analyses complex, and is alleged to bias estimation results (Meyerhoff and Liebe, 2009). In our current study, we avoided this problem by using an individual specific status quo (see Campbell et al., 2008; Hess and Rose, 2009; Marsh et al., 2011). The distribution of choices shows trade-offs between options 1 and 2, which will make results from our econometric analyses more robust. Inconsistency of choice distributions across the experiments implies that the treatment effect is not negligible. The next section presents and discusses the estimation results based on the stated preference data.

4.7. Empirical findings and discussion

To examine the impact of presentation formats on empirical estimates, we use the mixed logit (MXL) model as an estimation tool. Since our study conducted three experiments, we estimate utility functions for each of the three experiments. The MWTP estimates are also estimated for each experiment. The rationale of these two important analyses is to compare estimates across experiments and see if there are variations in terms of the statistical significance, sign and magnitude of the estimates. This section presents and discusses the estimation results. To estimate utility functions, the study adopts unconstrained MXL models where the five attributes of the study are modelled as normally distributed random parameters while alternative specific constants (ASCs) are modelled as fixed parameters. Results are obtained using the Halton sequence for simulation, based on 1000 draws. Utility models estimated in this study are defined as:

$$U_{ij} = \beta_0 + \beta_1 KITCHEN_{ij} + \beta_2 SHOWER_{ij} + \beta_3 TOILET_{ij} + \beta_4 GARDEN_{ij} + \beta_5 BILL_{ij} + \varepsilon_{ij} \tag{4.11}$$

Parameter β_0 represents the ASCs, while parameters β_1 to β_5 are coefficients of attributes and ε_{ij} is the random error component. The utility function presented in Equation 4.11 is estimated for the three experiments. The estimation results are presented in Table 4.6.

Table 4.6: Estimation results on household preferences for water-efficient technology

	Text		Visuals		Text and visuals	
	Par. Est.	Std. Err	Par. Est.	Std. Err	Par. Est.	Std. Err
Random parameters in utility functions						
KITCHEN	0.008	0.083	0.145**	0.068	0.332***	0.062
SHOWER	0.483***	0.114	-0.065	0.109	-0.204*	0.123
TOILET	-0.075	0.060	0.197***	0.052	-0.052	0.054
GARDEN	0.041	0.074	0.144**	0.073	-0.051	0.065
BILL	-0.012***	0.001	-0.011***	0.001	-0.001***	0.0004
Non-random parameters in utility functions						
ASC	0.0	0.206	0.0	0.427	0.0	0.613
Diagonal values in Cholesky matrix, L.						
NsKITCHEN	0.448***	0.091	0.306***	0.093	0.387***	0.151
NsSHOWER	0.736***	0.159	0.558***	0.151	0.204	0.380
NsTOILET	0.341***	0.087	0.150	0.125	0.001	0.397
NsGARDEN	0.015	0.134	0.346	0.289	0.170	1.234
NsBILL	0.620	0.002	0.211	0.289	0.0004	0.010
Below diagonal values in L matrix. V = L*Lt						
SHOWER:KITCHEN	-0.160	0.183	-0.082	0.184	0.733***	0.151
TOILWT:KITCHEN	-0.122	0.113	0.108	0.103	0.084	0.088
TOILET:SHOWER	0.237*	0.122	0.141	0.122	0.046	0.208
GARDEN:KITCHEN	-0.220**	0.092	-0.238*	0.139	0.026	0.101
GARDEN:SHOWER	-0.065	0.111	-0.723	0.151	0.023	0.264
GARDEN:TOILET	-0.223***	0.082	-0.247	0.319	0.051	4.169
BILL:KITCHEN	0.002	0.002	0.004**	0.002	-0.003***	0.001
BILL:SHOWER	-0.004**	0.002	-0.004*	0.002	0.002	0.003
BILL:TOILET	-0.007***	0.001	-0.004*	0.003	-0.0004	0.033
BILL:GARDEN	-0.0003	0.002	0.003	0.003	-0.001	0.009
Standard deviations of parameter distributions						
sdKITCHEN	0.448***	0.091	0.306***	0.093	0.387***	0.092
sdSHOWER	0.753***	0.155	0.564***	0.152	0.761***	0.217
sdTOILET	0.433***	0.094	0.233**	0.115	0.096	0.155
sdGARDEN	0.321***	0.094	0.493	0.396	0.181	0.111
sdBILL	0.008***	0.001	0.008***	0.003	0.003	0.002
LL Function	-811.4		-897.0		-2459.4	
McFadden Pseudo R^2	0.4		0.4		0.1	
AIC	1664.8		1836.1		4960.9	
BIC	1774.8		1947.8		5082.5	
Number of observations	1392		1515		2324	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Par Est. = parameter estimates. Std. Err = standard errors

We interpret the estimation results presented in Table 4.6 above based on the sign, magnitude and statistical significance of the random parameters. The parameter estimate of each attribute indicates the utility derived by respondents. To be specific, the sign of the parameter estimate shows the direction of the relationship between an attribute and the respondents' utility derived, while the magnitude of the parameter estimate shows the extent of the impact. The statistical significance of the parameter estimate shows the importance of an attribute to respondents. If the parameter estimates reported in Table 4.6 are substituted into equation 4.11 above, the following utility functions emerge:

$$U_{Text} = 0.01xKITCHEN_{ij} + 0.48xSHOWER_{ij} - 0.08xTOILET_{ij} + 0.04xGARDEN_{ij} - 0.01xBILL_{ij} + \varepsilon_{ij} \quad (4.12)$$

$$U_{Visuals} = 0.15xKITCHEN_{ij} - 0.07xSHOWER_{ij} + 0.20xTOILET_{ij} + 0.14xGARDEN_{ij} - 0.01xBILL_{ij} + \varepsilon_{ij} \quad (4.13)$$

$$U_{Text\ and\ visuals} = 0.33xKITCHEN_{ij} - 0.20xSHOWER_{ij} - 0.05xTOILET_{ij} - 0.05xGARDEN_{ij} - 0.01xBILL_{ij} + \varepsilon_{ij} \quad (4.14)$$

Equations 4.12 to 4.14 show the utility functions of the three experiments. Positive parameter estimates show that respondents prefer improvements in the attribute, whereas negative estimates show that respondents do not prefer improvements. Using the attribute parameter estimates reported for KITCHEN devices in all three equations, the results are interpreted to mean that households prefer improvements in the kitchen devices. In equation 4.12 above, for example, a unit improvement in the KITCHEN devices will increase respondents' utility by approximately 0.01 units, that is, a 10% improvement in KITCHEN devices increases the respondents' utility by about 0.1%. Regarding the negative attribute parameters, a unit increase in the BILL, for example, reduces the respondents' utility by approximately 0.01 across all the utility functions. This implies that when making choices, respondents did not prefer alternatives with higher water bills.

The utility functions presented in equations 4.12 to 4.14 above give information on the sign and magnitude of parameter estimates. Variations in these signs and magnitudes across experiments are interpreted to mean that presentation formats affect empirical results. The equations show significant differences in the magnitude of the parameter estimates for each attribute across the three experiments, especially for shower which shows opposite signs in two regressions. The magnitudes of the parameter estimates are well within the same range, in absolute terms. This result is consistent with findings from Arentze et al. (2003) and Patterson et al. (2017), where the size of the coefficients in absolute terms showed little difference across different experiments. However, our estimates are not consistent with results in similar studies that show large coefficients for visuals experiments compared to text experiments (see Bateman et al., 2009; Orzechowski et al., 2005; Vriens et al., 1998). This could be because those experiments use visuals to convey different information about the choice options than we have conveyed here.

Only KITCHEN and BILL reported parameters with the same signs across the three experiments. Nevertheless, some similarities are observed when comparisons are made between any two of the three experiments. For example, SHOWER has the same sign in the visuals and the text-and-visuals experiments, TOILET has the same sign in the text and the text-and-visuals experiments, and GARDEN has the same sign in the text and the visuals experiments. However, if the signs of the attribute parameters are compared across all three experiments, there are noticeable differences in most of the parameter estimates. This is in line with the results in some studies in the literature, which show that presenting information as visuals is likely to give different results compared to scenarios where the same information is presented as text (see Molin, 2011; Rizzie et al., 2012; Wittink et al., 1994). The argument usually put forward is that visuals present greater evaluability, which reduces the respondents' judgement error (Bateman et al., 2009).

An analysis of the significance of the parameter estimates presented in Table 4.6 above shows only two attributes that are statistically significant in the text experiment, while four attributes are statistically significant in the visuals experiment and three attributes are statistically significant in the text-and-visuals experiment. BILL is the only attribute that is statistically significant across all three experiments. KITCHEN is statistically significant in the visuals experiment and the text-and-visuals experiment but is statistically insignificant in the text experiment. SHOWER is statistically significant in the text and the text-and-visuals experiments. Except for the parameter estimates for BILL, there are no other consistent estimates between the text and the visuals experiments. These results are consistent with findings in the literature that visuals always have more statistically significant coefficients than the other presentation formats. Example of studies whose findings are consistent with ours include Jansen et al. (2009), Orzechowski et al. (2005) and Vriens et al. (1998). According to Patterson et al. (2017), most studies in the literature show visually presented variables taking on more importance than variables presented through text.

While the sign, significance and magnitude of random parameter estimates are essential when comparing empirical results, random parameter estimates themselves show the population mean. Therefore, it is also important to compare the dispersion that exists around the sample population in each format. This information is given by the standard deviations of the parameter distributions. Insignificant parameter estimates for derived standard deviations indicate that the dispersion around the mean is statistically equal to zero, suggesting that all information in the distribution is captured within the mean (Hensher et al., 2015). On the other hand, statistically significant parameter estimates for derived standard deviations of a random parameter suggest the existence of heterogeneity in the parameter estimates over the sampled population around the mean parameter estimate. According to Hensher et al. (2015), this implies that different individuals possess individual-specific parameter estimates that may be different from the sample population mean parameter estimate.

In terms of the standard deviations of random parameters, our results show that the text-and-visuals model had more estimates that were statistically insignificant than the other two models. Only two attribute parameters in the text-and-visuals model (KITCHEN and SHOWER) had statistically significant standard deviations. In the visuals model, all estimates except for GARDEN were statistically significant; while in the text model, all estimates were statistically significant. This suggests that in the text and the visuals models, different respondents possessed individual-specific parameter estimates that may be different from the sample population mean parameter estimate. However, in the text-and-visuals model the dispersion around the mean of most estimates is statistically equal to zero, suggesting that all information in the distribution is captured within the mean. This implies that the text-and-visuals experiment was able to capture the true preferences of respondents better than the other experiments.

It is also common practice in the literature to compare empirical estimates on the measures of welfare across presentation formats (see Bateman et al., 2009; Patterson et al., 2017). This section presents MWTP estimates, which are commonly used as welfare measures in the literature. MWTP estimates show the average estimates of what respondents are prepared to pay for or against improvements in each attribute. Positive and significant figures show the average amount that households are willing to pay for improvements in the attribute, whereas negative and significant figures show how much households are willing to accept as compensation for changes in the attribute. Empirical estimates for MWTP for the study are presented in Table 4.7.

Table 4.7: Estimates on MWTP for changes in water-efficient devices (in US Dollars)[25]

	Text		Visuals		Text and visuals	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
KITCHEN	0.05	0.49	0.90**	0.43	15.59***	5.58
SHOWER	2.84***	0.78	-0.40	0.66	-9.55	6.69
TOILET	-0.44	0.34	1.22***	0.34	-2.45	2.46
GARDEN	0.24	0.44	0.89*	0.48	-2.38	2.94
Wald Statistic	**1.06**		**1.38**		**0.61**	
Prob. from Chi^2	**0.005**		**0.001**		**0.070**	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Std. Err = standard errors.

MWTP estimates presented in Table 4.7 are interpreted to mean that in the text experiment, respondents are willing to pay $2.84 for improvements in SHOWER devices. In the visuals experiment, respondents are willing to pay $0.90, $1.22 and $0.89 for improvements in KITCHEN, TOILET and GARDEN devices respectively. In the text-and-visuals experiment, respondents are willing to pay $15.59 for improvements in SHOWER devices. Two main observations are made from a comparison of the statistical significance of the MWTP estimates. Firstly, the visuals experiment has more MWTP estimates that are statistically significant than the other two experiments, which have one statistically significant MWTP estimate each. This observation is consistent with earlier results on utility functions, where the visuals experiment also emerged as having more attribute parameter estimates that were statistically significant than the other experiments.

Secondly, we observe that the MWTP estimates reported in the text-and-visuals experiment are larger in absolute terms than those from both the text and the visuals experiments. When the sizes of MWTP estimates for the text and the visuals experiments are compared, it can be observed that the latter has more estimates that are bigger than the former in absolute terms. This agrees with findings in the literature that images tend to produce estimates that are mostly bigger than those from text experiments, in absolute terms (Bateman et al., 2009; Orzechowski et al., 2005; Syrengelas, 2017; Vriens et al., 1998). Overall, we observe that MWTP estimates were largely different in terms of sign, significance and magnitude across the three presentation formats. Based on these results, we argue that the presentation format also affects MWTP estimates. However, we note that for most of the attributes, the MWTP figures are very low

[25] As at 24 October 2018, US$1 = ZAR14.30

considering the water and monetary saving benefits associated with adopting the suggested technologies.

4.8. Conclusion

This paper uses choice experiments to examine the impact of presentation formats on empirical results. The focus of the paper was to establish whether utility functions and MWTP estimates are affected by the format used to present choice experiments. To achieve this, we used data from experiments on household preferences for water-efficient technologies in the Gauteng province of South Africa. The study compares three experiments, namely a text experiment, a visuals experiment and a text-and-visuals experiment. In the text experiment, respondents answered choice questions with alternatives presented as text, while in the visuals experiment, the same information was presented to respondents in the form of images. The third experiment provided the same information again but using both text and visual representations. By presenting an experiment that is entirely visual, our study is a step ahead of many similar studies in the choice experiment literature, which mainly compare text presentations with text and visual presentations (see Jansen et al., 2009; Orzechowski et al., 2005; Patterson et al., 2017). Our study uses the MXL model for empirical estimation, and four main findings can be reported.

Firstly, we found that while only two attributes emerged as important in the text experiment, four attributes were important in the visuals experiment and three were important in the text-and-visuals experiment. The literature explains the importance of attribute parameters as based on the size and statistical significance of the coefficients. Although there was not much difference in the size of the coefficients (in absolute terms) across the three experiments, we found that the visuals experiment had more statistically significant coefficients than both the text and the text-and-visuals experiment. This result is consistent with those of similar studies in the literature, which also find visual experiments to have more statistically significant coefficients than text experiments (see Jansen et al., 2009; Orzechowski et al., 2005; Vriens et al., 1998). Since the visuals and the text-and-visuals experiments both had more coefficients that were statistically significant than the text experiment, we argue that including visuals in the choice profiles increased the number of attributes that were important to respondents.

Secondly, a comparison of attribute parameters across all three experiments showed some differences in the signs of each parameter, with only two attributes having the same sign across all three experiments. However, a few similarities in the signs were observed when comparisons were made between any two of the three experiments. Prior to the tests, we hypothesised that although the magnitude and statistical significance of each attribute parameter may differ across experiments, the sign of each parameter should be the same. This is because descriptive statistics in our study showed similarities in the socio-economic characteristics of respondents across the experiments. Our hypothesis was shaped by Patterson et al. (2017), where no meaningful differences were observed in results across experiments. However, the results in this study confirm reports in the literature that visuals and text experiments give different results (see Molin, 2011; Rizzie et al., 2012; Wittink et al., 1994).

Thirdly, we observed that the text-and-visuals experiment reported fewer attribute parameters with dispersion around the sample population than the text and the visuals experiments. Only two parameters in the text-and-visuals experiment had statistically significant standard deviations. This indicates that the random parameter estimates reported in the text-and-visuals experiment correctly reflect respondents' choices, except for two attributes. In the visuals experiment, only one parameter was statistically insignificant; while in the text model, all estimates were statistically significant. This suggests that in the text and the visuals experiments, different respondents possessed individual-specific parameter estimates that may be different from the sample population mean. Considering this, we argue that the text-and-visuals experiment was able to capture the true preferences of respondents better than the other experiments.

Finally, the MWTP estimates showed that households were willing to pay for more attributes in the visuals experiment than in the other two experiments. In the visuals experiment, respondents were willing to pay for three attributes, whereas they were only willing to pay for one in the text experiment and one in the text-and-visuals experiment. Again, this confirms reports in the literature that visual experiments tend to have more significant parameters than text experiments (see Bateman et al., 2009; Orzechowski et al., 2005; Syrengelas, 2017; Vriens et al., 1998). A comparison of the magnitude of the MWTP estimates across the three experiments showed that the text-and-visuals experiment had larger estimates than the other two experiments, in absolute terms. Overall, the MWTP estimates were largely different in terms of sign, significance and magnitude across the three presentation formats.

Based on the results presented in this study, we join other studies in the literature in arguing that visually-presented attributes tend to take on more importance than attributes presented through text. However, we advise caution when presenting experiments as visuals, since other less important aspects such as colour and form may distort preferences. On the other hand, the text-and-visuals experiment showed some consistency with both the text and the visuals experiments in terms of the sign and significance of parameters. By combining both text and visuals, the experiment was able to clarify attributes to respondents, thereby yielding more robust stated preference data and empirical estimates. Overall, we argue that the format of presenting information matters in choice experiments conducted in environmental economics.

List of references

Abdullah, S. and Mariel, P. 2010. Choice experiment study on the willingness to pay to improve electricity services. *Energy Policy,* 38**,** 4570-4581.

Abelson, R. P. and Levy, A. 1985. *Decision Making and Decision Theory. (In Lindzey, G. and Aronson, E. Eds). Handbook of Social Psychology 1.,* New York: Random House.

Arentze, T., Borgers, A., Timmermans, H. and Delmistro, R. 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review,* 39**,** 229-244.

Bateman, I. J., Day, B. H., Jones, A. P. and Jude, S. 2009. Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management,* 58**,** 106-118.

Ben-Akiva, M. E. and Lerman, S. R. 1985. *Discrete choice analysis: theory and application to travel demand,* MIT press.

Berlyne, D. E. 1971. *Aesthetics and psychobiology,* JSTOR.

Bhaduri, A. and Kloos, J. 2013. Getting the water prices right using an incentive-based approach: an application of a choice experiment in khorezm, uzbekistan. *The European Journal of Development Research,* 25**,** 680-694.

Bliemer, M. C., Rose, J. M. and Hess, S. 2008. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling,* 1**,** 98-126.

Campbell, D., Hutchinson, W. G. and Scarpa, R. 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and resource economics,* 41**,** 401-417.

Caussade, S., De Dios Ortúzar, J., Rizzi, L. I. and Hensher, D. A. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological,* 39**,** 621-640.

Childers, T. L. and Houston, M. J. 1984. Conditions for a picture-superiority effect on consumer memory. *Journal of consumer research,* 11**,** 643-654.

Dhar, R. and Simonson, I. 2003. The effect of forced choice on choice. *Journal of Marketing Research,* 40**,** 146-160.

Ferrini, S. and Scarpa, R. 2007. Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management,* 53**,** 342-363.

Fitzsimons, G. J., Hutchinson, J. W., Williams, P., Alba, J. W., Chartrand, T. L., Huber, J., Kardes, F. R., Menon, G., Raghubir, P. and Russo, J. E. 2002. Non-conscious influences on consumer choice. *Marketing Letters,* 13**,** 269-279.

Green, C. and Tunstall, S. 1999. *A psychological perspective*, Oxford University Press.

Greene, W. H. 2003. *Econometric analysis*, Pearson Education India.

Greene, W. H. 2012. NLOGIT Version 5 Reference Guide. Econometric Software, Inc.: Econometric Software, Inc. Plainview, NY.

Hanley, N., Mourato, S. and Wright, R. E. 2001. Choice modelling approaches: a superior alternative for environmental valuatioin? *Journal of economic surveys,* 15**,** 435-462.

Hensher, D., Shore, N. and Train, K. 2005. Households' willingness to pay for water service attributes. *Environmental and Resource Economics,* 32**,** 509-531.

Hensher, D. A., Rose, J. M. and Greene, W. H. 2015. *Applied Choice Analysis*, Cambridge University Press.

Hering, J. G. and Ingold, K. M. 2012. Water resources management: what should be integrated? *Science,* 336**,** 1234-1235.

Hess, S. and Rose, J. M. 2009. Should reference alternatives in pivot design SC surveys be treated differently? *Environmental and Resource Economics,* 42**,** 297-317.

Hetherington, J., Daniel, T. C. and Brown, T. C. 1993. Is motion more important than it sounds?: The medium of presentation in environment perception research. *Journal of Environmental Psychology,* 13**,** 283-291.

Holbrook, M. B. and Moore, W. L. 1981. Feature interactions in consumer judgments of verbal versus pictorial presentations. *Journal of Consumer Research,* 8**,** 103-113.

Howard, J. A. 1977. *Consumer Behaviour in Marketing Strategy.*, New Jersey: Prentice Hall.

Jansen, S., Boumeester, H., Coolen, H., Goetgeluk, R. and Molin, E. 2009. The impact of including images in a conjoint measurement task: evidence from two small-scale studies. *Journal of housing and the built environment,* 24**,** 271-297.

Jones, M. P. and Hunt, W. F. 2010. Performance of rainwater harvesting systems in the southeastern United States. *Resources, Conservation and Recycling,* 54**,** 623-629.

Kanyoka, P., Farolfi, S. and Morardet, S. 2008. Households' preferences and willingness to pay for multiple use water services in rural areas of South Africa: an analysis based on choice modelling. *Water SA,* 34**,** 715-723.

Keane, M. P. and Wasi, N. 2012. Estimation of discrete choice models with many alternatives using random subsets of the full choice set: With an application to demand for frozen pizza.

Landon, A. C., Kyle, G. T. and Kaiser, R. A. 2016. Predicting compliance with an information-based residential outdoor water conservation program. *Journal of Hydrology,* 536**,** 26-36.

Lanz, B. and Provins, A. 2015. Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics,* 47**,** 300-324.

Liu, A., Giurco, D. and Mukheibir, P. 2015. Motivating metrics for household water-use feedback. *Resources, Conservation and Recycling,* 103**,** 29-46.

Louviere, J. J. 2001. Choice experiments: an overview of concepts and issues. *The choice modelling approach to environmental valuation***,** 13-36.

Lovett, A., Appleton, K., Warren-Kretzschmar, B. and Von Haaren, C. 2015. Using 3D visualization methods in landscape planning: An evaluation of options and practical issues. *Landscape and Urban Planning,* 142**,** 85-94.

MacInnis, D. J. and Price, L. L. 1987. The role of imagery in information processing: Review and extensions. *Journal of consumer research,* 13**,** 473-491.

Makki, A. A., Stewart, R. A., Panuwatwanich, K. and Beal, C. 2013. Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies. *Journal of Cleaner Production,* 60**,** 129-146.

Marsh, D., Mkwara, L. and Scarpa, R. 2011. Do respondents' perceptions of the status quo matter in non-market valuation with choice experiments? An application to New Zealand freshwater streams. *Sustainability,* 3**,** 1593-1615.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior.

Meyerhoff, J. and Liebe, U. 2009. Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity. *Land Economics,* 85**,** 515-528.

Mini, C., Hogue, T. and Pincetl, S. 2015. The effectiveness of water conservation measures on summer residential water use in Los Angeles, California. *Resources, Conservation and Recycling,* 94**,** 136-145.

Molin, E. J. 2011. Conjoint analysis. *The measurement and analysis of housing preference and choice.* Springer, Dordrecht.

Munro, A. and Hanley, N. D. 2001. Information, uncertainty, and contingent valuation. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries. Oxford University Press, Oxford***,** 258-274.

Orzechowski, M., Arentze, T., Borgers, A. and Timmermans, H. 2005. Alternate methods of conjoint analysis for estimating housing preference functions: Effects of presentation style. *Journal of Housing and the Built Environment,* 20**,** 349-362.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J. and Yazdizadeh, A. 2017. Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning,* 157**,** 63-74.

Price, G. 2009. Water Conservation Guideline. eThekwini Municipality: eThekwini Municipality

Rizzi, L. I., Limonado, J. P. and Steimetz, S. S. 2012. The impact of traffic images on travel time valuation in stated-preference choice experiments. *Transportmetrica,* 8**,** 427-442.

Ro, T., Singhal, N. S., Breitmeyer, B. G. and Garcia, J. O. 2009. Unconscious processing of color and form in metacontrast masking. *Attention, Perception, and Psychophysics,* 71**,** 95-103.

Rose, J. M. and Bliemer, M. C. 2009. Constructing efficient stated choice experimental designs. *Transport Reviews,* 29**,** 587-617.

Rose, J. M., Bliemer, M. C., Hensher, D. A. and Collins, A. T. 2008. Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological,* 42**,** 395-406.

Saldías, C., Speelman, S., Van Huylenbroeck, G. and Vink, N. 2016. Understanding farmers' preferences for wastewater reuse frameworks in agricultural irrigation: lessons from a choice experiment in the Western Cape, South Africa. *Water SA,* 42**,** 26-37.

Samuelson, W. and Zeckhauser, R. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty,* 1**,** 7-59.

Scarpa, R., Campbell, D. and Hutchinson, W. G. 2007. Benefit estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment. *Land Economics,* 83**,** 617-634.

Scarpa, R., Gilbride, T. J., Campbell, D. and Hensher, D. A. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics,* 36**,** 151-174.

Snowball, J., Willis, K. and Jeurissen, C. 2008. Willingness To Pay For Water Service Improvements In Middle-Income Urban Households In South Africa: A Stated Choice Analysis. *South African Journal of Economics,* 76**,** 705-720.

Still, D. and Bhagwan, J. 2008. The Status and Use of Potable Water Conservation and Savings Devices in the Domestic and Commercial Environments in South Africa. *Water Distribution Systems Analysis 2008.*

Syrengelas, K. 2017. Examining text versus visual presentation of choice experiments: Does the presentation method affect consumer preferences and willingness to pay? Master's Thesis, University of Tennessee.

Timmermans, H. and Van Noortwijk, L. 1995. Context dependencies in housing choice behavior. *Environment and Planning A,* 27**,** 181-192.

Train, K. and Wilson, W. W. 2008. Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B: Methodological,* 42**,** 191-203.

Townsend, C. and Kahn, B. E. 2013. The "visual preference heuristic": The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *Journal of Consumer Research,* 40**,** 993-1015.

Vásquez, W. F., Franceschi, D. and Van Hecken, G. T. 2012. Household preferences for municipal water services in Nicaragua. *Environment and Development Economics,* 17**,** 105-126.

Vloerbergh, I., Fife-Schaw, C., Kelay, T., Chenoweth, J., Morrison, G. and Lundéhn, C. 2007. Assessing consumer preferences for drinking water services-Methods for water utilities. *Techneau Project Report D,* 6.

Vriens, M., Loosschilder, G. H., Rosbergen, E. and Wittink, D. R. 1998. Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *Journal of Product Innovation Management,* 15**,** 455-467.

Wang, D. and Li, S.-M. 2004. Housing preferences in a transitional housing system: the case of Beijing, China. *Environment and Planning A,* 36**,** 69-87.

Willis, K. G., Scarpa, R. and Acutt, M. 2005. Assessing water company customer preferences and willingness to pay for service improvements: A stated choice analysis. *Water Resources Research,* 41.

Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R. and Mousavinejad, A. 2013. End use water consumption in households: impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production,* 60**,** 107-115.

Wittink, D. R., Vriens, M. and Burhenne, W. 1994. Commercial use of conjoint analysis in Europe: Results and critical reflections. *International journal of Research in Marketing,* 11**,** 41-52.

Wohlwill, J. F. 1976. Environmental aesthetics: The environment as a source of affect. *Human behavior and environment: Advances in theory and research,* 1**,** 37-86.

Appendix 4.1: Households' use of water efficient technology

Table A1: Frequency distribution of households' responses to having water-efficient technology

			Text	Visuals	Text-and-visuals	Pooled data	Modal response
		Respondents (N)	***232***	***257***	***405***	***894***	
1.	**Water-collection tank (Jojo tank) (%)**	*Yes*	5	5	6	6	No
		No	89	93	88	90	
		Not applicable	5	1	5	3	
		Not Sure	1	1	1	1	
2.	**Cistern displacement device (Hippo bag) (%)**	*Yes*	3	4	3	3	No
		No	77	93	79	82	
		Not applicable	19	2	16	13	
		Not Sure	2	1	1	1	
3.	**Water-flow regulators (%)**	*Yes*	14	8	13	12	No
		No	83	91	85	86	
		Not applicable	2	-	1	1	
		Not Sure	1	1	1	1	
4.	**Efficient showerheads (%)**	*Yes*	40	23	28	30	No
		No	56	75	68	67	
		Not applicable	3	1	3	2	
		Not Sure	1	1	1	1	
5.	**Efficient toilet cistern, sized 3-6 litres (%)**	*Yes*	57	47	42	48	No
		No	41	52	58	52	
		Not applicable	1	-	-	-	
		Not Sure	-	1	-	-	
6.	**Interruptible/multi-flush cistern (%)**	*Yes*	91	94	90	91	Yes
		No	9	6	9	9	
		Not applicable	-	-	-	-	
		Not Sure	-	-	1	-	
7.	**Dishwasher (%)**	*Yes*	17	11	14	14	No
		No	82	89	86	86	
		Not applicable	1	-	-	-	
		Not Sure	-	-	-	-	
8.	**Efficient garden devices (%)**	*Yes*	14	5	9	9	No
		No	84	94	90	89	
		Not applicable	1	1	-	1	
		Not Sure	1	-	1	1	

Appendix 4.2: Households' daily water-use behaviour

Table A2: Households' daily water-use behaviour

			Text	Visuals	Text and visuals	Pooled data	Modal response
		Respondents (N)	*232*	*257*	*405*	*894*	
1.	**Take bath instead of shower (%)**						
		Never	23	29	29	27	
		Occasionally	15	7	15	12	Always
		Always	40	44	46	44	
		Not applicable	22	21	10	16	
2.	**Take shower for more than 5 minutes (%)**						
		Never	19	20	33	25	Not applicable
		Occasionally	31	12	27	24	
		Always	22	15	15	17	
		Not applicable	28	53	25	34	
3.	**Run shower for some time, waiting for hot water (%)**						
		Never	27	21	45	33	Never
		Occasionally	13	7	9	10	
		Always	33	22	23	25	
		Not applicable	28	50	23	32	
4.	**Keep the tap running when brushing teeth (%)**						
		Never	84	86	83	84	
		Occasionally	8	4	10	8	Never
		Always	8	9	7	8	
		Not applicable	-	1	-	-	
5.	**Ignore water leaks from the toilet tank (%)**						
		Never	96	93	95	95	
		Occasionally	4	5	4	4	Never
		Always	-	1	1	1	
		Not applicable	-	1	-	-	
6.	**Keep tap running when washing dishes (%)**						
		Never	77	92	85	85	
		Occasionally	7	6	4	7	Never
		Always	16	2	1	8	
		Not applicable	-	-	-	-	
7.	**Rinse cutlery and glasses under running water (%)**						
		Never	66	80	75	74	
		Occasionally	14	11	11	12	Never
		Always	20	9	15	14	
		Not applicable	-	-	-	-	
8.	**Use running water to defrost frozen food (%)**						
		Never	91	90	90	90	
		Occasionally	5	5	5	5	Never
		Always	3	6	5	5	
		Not applicable	-	-	-	-	
9.	**Ignore a dripping tap (%)**						
		Never	96	97	97	97	
		Occasionally	4	2	1	3	Never
		Always	-	1	1	-	
		Not applicable	-	-	1	-	

10. Ignore kids wasting water (%)						
	Never	97	98	95	96	
	Occasionally	3	2	4	4	Never
	Always	-	-	1	-	
	Not applicable	-	-	-	-	
11. Keep water running while washing face or hair (%)						
	Never	97	92	92	93	
	Occasionally	2	4	5	4	Never
	Always	1	4	3	3	
	Not applicable	-	-	-	-	

Appendix 4.3: Example of the questionnaire used in the experiment (text and visuals)

NELSON MANDELA UNIVERSITY

HOUSEHOLDS' INTENTIONS TO ADOPT WATER-SAVING TECHNOLOGY IN JOHANNESBURG

Time interview began ____:____

Date of the interview: ____/___/_____

Name of interviewer: ____________________________________

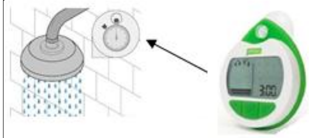
Area study is taking place: _______________________________

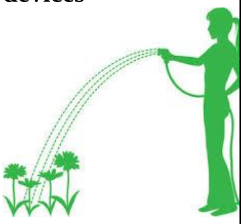
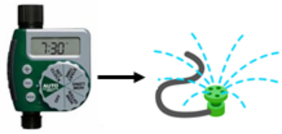
As is the case in the rest of South Africa, Johannesburg is facing water shortages; yet little is known about households' water-conservation efforts. Households are encouraged to install water-saving devices as part of addressing the water shortage. This can only be achieved if households are aware of water-saving options and the cost-savings benefits. We employ choice experiments to evaluate the intention of households to adopt water-saving technologies.

The survey has three sections. Section A provides choice experiments by which households' intention to adopt water-saving devices is evaluated. Section B provides general questions on households' water-consumption behaviour. Section C collects the biographical information of respondents.

SECTION A: CHOICE EXPERIMENT

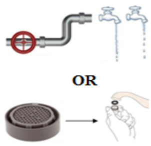
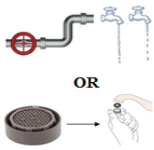
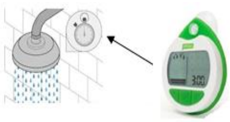
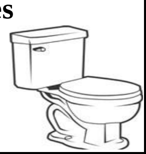
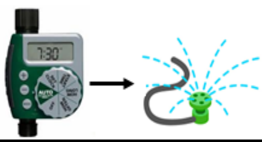
Table 1: Attributes and levels used in the study

Attribute	Description	Attribute Levels	
Kitchen devices 	A typical household uses 11% of its water in the kitchen. A standard tap flows at about 8l per minute. Installing water-flow regulators or tap-head aerators makes a standard tap more efficient and saves water by 60%. An efficient dishwasher uses 15l per cycle, using 50% less water than is used in a conventional dishwasher.	**Level 1:** Efficient dishwasher	
		Level 2: Efficient tap	
		Level 3: System collecting used water	
Shower devices 	A typical household uses 24% of its water in the shower. Shower timers result in shorter showers. Efficient showerheads save 65% of water used in the shower.	**Level 1:** Efficient showerhead	
		Level 2: Shower timer	
Toilet devices 	A typical household uses 25% of its water for flushing the toilet. Replacing a 12l cistern with a 3l dual cistern uses about 75% less water. An interruptible-flush cistern allows users to control how long the toilet flushes. Hippo bags displace water in the cistern and save about 1.2l per flush.	**Level 1:** Dual-flush cistern sized 3-6l	
		Level 2: Interruptible-flush cistern	
		Level 3: Cistern displacement (hippo bag)	

Garden & Outdoor devices 	A typical household uses 25% of its water in the garden or for outdoor activities. Efficient gardening technologies reduce water use by 30%. These include time-based irrigation control systems, and micro-drip systems. Irrigating gardens using water collected with water tanks also saves water.	**Level 1:** Time-based irrigation controller	
		Level 2: Micro-drip systems	
		Level 3: Use harvested rain water	
Monthly water bill 	The average water bill for a household is R450 per month. Installing water-efficient technologies will reduce the monthly water bill by 30%, 50% or 75%.	**Level 1:** R110 **Level 2:** R225 **Level 3:** R315	

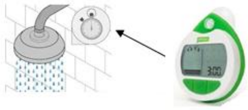
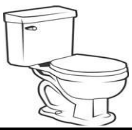
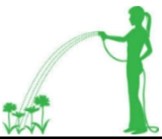
Six choice sets with three alternatives (Status Quo, Option 1 and Option 2) are generated. The Status Quo is undefined, as only you know your current situation. We would like to know which option you prefer the most. Please treat each choice set independently.

CHOICE SET 1

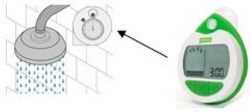
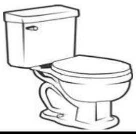
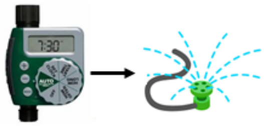
	Status quo	Option 1	Option 2
Kitchen devices 		Efficient tap 	Efficient tap 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Dual-flush cistern 	Hippo bag 
Garden & outdoor devices 		Use harvested rain water 	Time-based irrigation controller 

Monthly water bill 	R450	R315	R110
YOUR CHOICE			

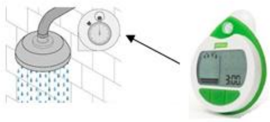
CHOICE SET 2

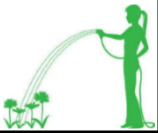
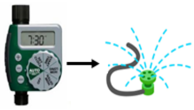
	Status quo	**Option 1**	**Option 2**
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Efficient showerhead 	Shower timer 
Toilet devices 		Dual-flush cistern 	Hippo bag 
Garden & outdoor devices 		Micro-drip irrigation system 	Micro-drip irrigation system 
Monthly water bill 	R450	R110	R315
YOUR CHOICE			

CHOICE SET 3

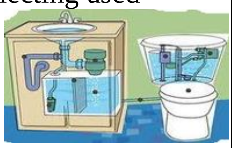
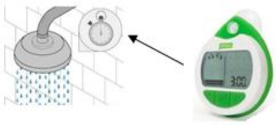
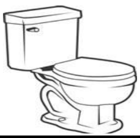
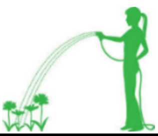
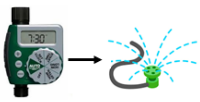
	Status quo	Option 1	Option 2
Kitchen devices 		Efficient tap 	Efficient tap 
Shower devices 		Efficient showerhead 	Shower timer 
Toilet devices 		Hippo bag 	Dual-flush cistern 
Garden & outdoor devices 		Use harvested rain water 	Time-based irrigation controller 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

CHOICE SET 4

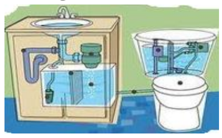
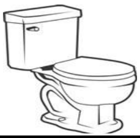
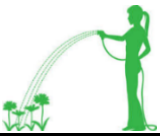
	Status quo	Option 1	Option 2
Kitchen devices 		System collecting used water 	Efficient dishwasher 
Shower devices 		Efficient showerhead 	Shower timer 

Toilet devices 		Interruptible-flush cistern 	Interruptible-flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rain water 
Monthly water bill 	R450	R315	R110
YOUR CHOICE			

CHOICE SET 5

	Status quo	**Option 1**	**Option 2**
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Hippo bag 	Dual flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rain water 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

CHOICE SET 6

	Status quo	Option 1	Option 2
Kitchen devices 		System collecting used water 	Efficient dishwasher 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Interruptible-flush cistern 	Interruptible-flush cistern 
Garden & outdoor devices 		Micro-drip irrigation system 	Micro-drip irrigation system 
Monthly water bill 	R450	R110	R315
YOUR CHOICE			

SECTION B: WATER CONSUMPTION BEHAVIOUR AND TECHNOLOGY

1. When you made your choices, which attribute most influenced your decision?
Please tick in the box next to the attribute

Kitchen devices	
Shower devices	
Toilet devices	
Garden devices	
Water bill	

2. Do you have the following water technology at home?
Please select one answer per row

	Yes	No	Not applicable	Not sure
Water-collection tank (Jojo tank)				
Cistern displacement device (‘hippo bag’)				
Water-flow regulators				
Efficient showerheads				
Efficient bathtub				

Efficient toilet cistern sized 3-6 litres				
Interruptible-flush (multi-flush) cistern				
Dishwasher				
Efficient garden devices				

3. **If any of your answers in QUESTION 2 above was NO, what is your main reason?**
Please select one reason you think is the main reason

I cannot afford them	
I did not know about them	
I have no infrastructure to connect them	
They are not important	
Other *(Please specify):*	

4. **How often do you do the following in your daily life?**
Please select one answer per row

	Never	Occasionally	Always	Not applicable
Take bath instead of shower				
Take showers longer than 5 minutes				
Run shower for some time, waiting for hot water				
Keep the tap running when brushing teeth				
Ignore water leaks from the toilet tank				
Keep the tap running when washing dishes				
Rinse cutlery and glasses under running water				
Use running water to defrost frozen food				
Ignore a dripping tap				
Ignore kids wasting water				
Keep water running while washing face or hair				

SECTION C: PERSONAL INFORMATION

1. **How many people are in your household?**

☐

2. **Do you have the following in your household?**
Please select one answer per row

	Yes	No
Infant (0-2 years)		
Child (3-15 years)		

3. **If YES to QUESTION 2 above, how many infants/children do you have?**

☐

4. **What is your gender?**

☐ Male ☐ Female

5. **Which racial group do you belong to?** *(Optional)*:

☐ Black/African ☐ White ☐ Indian/Asian ☐ Coloured

6. What is your marital status?

☐ Single ☐ Married ☐ Other (*Please specify*): _____________

7. What is your highest education level?

☐ Never attended school ☐ Primary school ☐ High school

☐ Certificate ☐ Diploma/Degree ☐ Postgraduate

8. What is your year of birth?

☐

9. What is your household's main source of income?

☐ Salary/Wages ☐ Business ☐ Investments ☐ Grant/Pension/Allowance

10. What is your household's monthly average income?

< R5 000	
R5 000 – R10 000	
R10 000 – R20 000	
R20 000 – R40 000	
R40 000 – R60 000	
> R60 000	

11. Did you answer this questionnaire on your own? *(If you answered on your own without the interviewer ticking boxes for you, select YES):*

☐ Yes ☐ No

Time interview ended ______:______

Chapter 5: The link between response time and choices in choice experiments

Abstract

Response time is a possible indicator of the cognitive processes employed by choice experiment participants when making choices. The decision-making literature suggests a positive correlation between slower response time and rational thinking, which is consistent with standard theories of decision-making. The aim of the paper is to investigate the relationship between response time and respondents' choices. We disentangle preference and willingness-to-pay estimates and explore whether response time sheds light on these aspects. Our approach entails dividing the data (ordered by response time) into three subsets. While the effects of response time have been investigated previously, this paper's emphasis is on assessing the time respondents require to answer self-administered face-to-face stated-preference surveys. We make use of electronic devices for data collection instead of traditional paper-based methods to accurately capture response time. We use data on water-efficient technologies to test the impact of response time. Using generalised mixed logit models, we compare results from an analysis of average responses, fast responses, slow responses and whole sample data. Overall, we find that response times did not affect results.

Keywords: choice experiments, response time, generalised mixed logit models.

5.1. Introduction

Although very costly, face-to-face surveys remain the main mode for collecting stated-preference data in developing and poorer countries. The main reasons for this are low internet penetration, slow internet connectivity and low literacy levels. This trend is in contrast to what is emerging in developed countries, where web-based online surveys have gained popularity. The advantages of web-based surveys are well documented in the literature (see Fleming and Bowden, 2009; Linhjem and Navrud, 2011), hence their growing popularity. However, for the reasons I have mentioned, developing countries are still using traditional survey modes such as face-to-face interviews, despite their many shortcomings. These shortcomings include the fact that they are very costly, inflexible (questionnaires cannot be adjusted), prone to human error and have a slow data collection rate.

To mitigate the shortcomings associated with face-to-face surveys, our experiment makes use of electronic devices. According to Nkosi and Dikgang (2018), the use of electronic devices has gained popularity over the orthodox paper method, because of its efficiency. This method minimises human error, since the coding of the survey into the gadget occurs in advance, which makes it easier and less time-consuming for the enumerator when collecting data. It is argued that this systematic method reduces the number of questions that might otherwise be mistakenly skipped when rushing to complete the survey, and the entering of incorrect information when capturing data, since data capture occurs automatically when the survey is completed.

The use of electronic devices allows for accurate capturing of response time. The aim of the study is to investigate the effects of response time on respondents' choices in a self-administered face-to-face survey environment. Most of the studies that investigate the relationship between response time and choices use online stated-preference surveys. Few studies that use face-to-face interviews tend to leave questionnaires with respondents overnight or give respondents 'time to think'. 'Time to think' is also used in web-based surveys. However, because our interest is to isolate response time and assess how it links to respondents' choices, the presence of 'time to think' may contaminate response time.

Our study differs from other studies in that experiment participants are not given 'time to think', so contamination should be minimised. Moreover, the use of electronic devices means human error is minimised and response time is captured more accurately. As in Campbell et al. (2018), we disentangle preference, variance and processing heterogeneity, and investigate whether response time sheds light on these kinds of heterogeneity. To achieve this, we divide our data (ordered by response time) into three sub-groups (i.e. fast, average and slow responses) and assess linkages to cognitive processing. According to Stupple et al (2017), response-time patterns may be considered an indicator to the cognitive processes employed by the experiment respondents when making choices.

Response-time data is an example of non-choice data that may help economists to understand the experimental process. The impact of response time on perceptual and cognitive processes in decision-making has received serious and considerable attention in experimental psychology, consumer research and marketing research (Campbell et al., 2017). However, it is not given serious and adequate attention in the economics literature, most importantly environmental economics. Although interest is beginning to emerge in the economics literature, very few choice experiment studies investigate the impact of response times. There is considerable scope for more research on the impact of response times in choice experiments (Bonsall and Lythgoe, 2009).

In every survey, some respondents answer quickly while others take longer. The literature suggests a contentious nexus between response time and quality of data. Some studies in the literature argue that slow respondents apply deliberative thought in considering all information provided (Chen and Fischbacher, 2015; Haaijer et al., 2000; Recalde et al., 2014). Such a view assumes that longer response times generate good-quality data, while quick responses signify insufficient effort and are presumed to result in data of suspicious validity and reliability. In that context, quick responses are argued to be a source of random error when utility functions are estimated and are referred to as ‘quick and dirty’ in certain studies (see Conrad et al., 2017; Huang et al., 2015; Schwappach and Strasmann, 2006; Wood et al., 2017). The implication of these studies is that any failure to account for response times would result in incorrect estimates and inferences.

However, it is imperative to appreciate that personal characteristics such as age, level of education, cognitive ability and self-reported decision-making style can contribute to the time that each respondent takes to complete a survey. For example, some respondents will understand a question at their first attempt, while others will only understand the same question after several attempts at reading it. In such cases, there is a risk that fast responses which may have quality data will be thrown out. Börger (2016) and Campbell et al. (2017) argue that fast responses are inevitable, and recommend that researchers do not discard them. Although it might seem appropriate to believe that fast responses reflect random decision-making, this assumption could be misguided, because there is no clear link in the literature between response time and cognitive effort.

This paper uses stated-preference data from households' preferences for water-efficient technologies to test the impact of response times on empirical estimates. The study is designed to contribute to the growing but still limited literature on whether response times matter in choice experiments. We test the hypothesis that fast responses reflect random decision-making and affect empirical results if they are not accounted for. In testing the impact of response time, we join the topical debate on the need for economists to take note of non-choice data when conducting choice experiments. Essentially, choice-experiment researchers put substantial effort into the design of experiments, without considering how involved respondents will be in reading and processing the content before they make choices (Vista et al., 2009). Different levels of involvement are believed to result in systematic differences in estimated empirical results, because answering a survey requires respondents to invest a great deal of cognitive effort (Krosnick, 1991; Lenzner et al., 2010).

Our study makes two main contributions. Firstly, we use a face-to-face experiment to collect stated-preference data. This makes our study unique, because similar studies in the literature are predominantly web-based (see Campbell et al., 2013; Downes-Le Guin et al., 2012; Savage and Waldman, 2008; Wood et al., 2017). Although the internet is increasingly becoming a preferred and convenient survey mode, it is difficult for researchers to monitor the survey. For example, some respondents may take longer because they are performing other tasks during the time that they answer survey questions. This problem is taken care of in face-to-face surveys, as researchers can take note of the respondents' activities during the survey. Secondly, our study generally differs from similar studies in that respondents are not given time to think (typically a night) before responding (see Cook et al., 2012; Cook et al., 2007; Svedsater, 2007; Whittington et al., 1992).

The rest of the paper is organised into seven sections. Section 5.2 reviews the literature on response time. Section 5.3 presents the case study. Section 5.4 discusses the experimental design. Section 5.5 presents the modelling approach. Section 5.6 discusses the experimental data. Section 5.7 presents and discusses the empirical findings. Section 5.8 concludes the study.

5.2. Literature on response time

Spiliopoulos and Ortmann (2014) identify three waves of response-time studies in the literature, classified according to the types of tasks investigated. The first wave investigated judgment tasks such as perceptual acuity or memory retrieval (Anderson et al., 1998; Ratcliff, 1978; Schooler and Anderson, 1997). A second wave emerged first in cognitive psychology and later in economics, investigating individual decision-making choice tasks that required valuation processing rather than judgments. Studies in this wave investigated decision-making under risk and lottery choices, and multi-alternative and multi-attribute choice (Dror and Hartman, 1999; Rieskamp and Hoffrage, 2008; Rieskamp and Otto, 2006; Wilcox, 1993). The third wave, which is the most recent, involves the analysis of response times in strategic decision-making or games (Arad and Rubinstein, 2012; Di Guida and Devetag, 2013; Gill and Prowse, 2017; Kuo et al., 2009; Rubinstein, 2007; Rubinstein, 2016).

The use of response time to study decision-making in economics started with the work of Wilcox (1993), in which response time was viewed as a proxy for decision cost. Subsequently, literature on response time investigated the decision processes employed by individuals to make inferences about preferences, and to predict choices across various domains (Recalde et al., 2014). Opportunities for expanding the knowledge and understanding of preferences for public and environmental goods have been made possible through advances in behavioural economics. Rand et al. (2012) were among the first to use response time with the intention of identifying intuitive and deliberate actions in public-good games. There has also been increasing interest in the cognitive aspects of preference revelation. Studies on this element of the literature highlight that choices in stated-preference surveys reflect respondents' true preferences as well as their cognitive efforts and errors (Beshears et al., 2008; Kessler and Meier, 2014; Kocher et al., 2017; Lohse et al., 2017; Nielsen et al., 2014).

In the choice-experiment literature, Campbell et al. (2017) examine the impact of response time on preferences, variance heterogeneity and class membership. Using the scale-adjusted latent class model, the study reports that preference heterogeneity and the variance of observed factors are sensitive to response times. Error variance was found generally to decrease with increasing response time. However, response time was reported not to affect MWTP estimates. These results are consistent with findings in Börger (2016), in which the generalised multinomial logit model was used to examine the impact of response time on choices and scale. Longer response times were found to increase scale but had no impact on welfare estimates.

Using mixed logit and multinomial logit models, Rose and Black (2006) also found that response time affected both the mean and the variance of random parameter distributions, and suggested that any failure to account for response times could result in incorrect model inferences being drawn.

Konovalov and Krajbich (2017) investigated several ways in which the relationship between response time and choices could be used to infer preferences when choice outcomes are uninformative or unavailable. The results revealed that response times from a single two-alternative choice problem would be enough to rank respondents usefully according to their degree of loss aversion. Using long response times, the study predicts choices that are inconsistent with a respondent's utility function, and likely to be reversed later. These results agree with those from the work of Dellaert et al. (2012), which examined choice complexity and individual differences in response time as determinants of choice-experiment-based models. That study reported that complexity and individual response time affect error scale in the utility function – findings that were later confirmed, in Börger (2016), Campbell et al. (2017) and Konovalov and Krajbich (2017).

Haaijer et al. (2000) and Holmes et al. (1998) suggested that response time should be incorporated as a parameter in estimation. The argument in these studies is that if response time is not considered, there is a greater chance of drawing incorrect conclusions. Using a multinomial probit model, Haaijer et al. (2000) showed that including response times as a parameter in the estimation model significantly improves the fitness of the model, provides narrower confidence intervals, reduces heterogeneity, and generates better holdout predictions. These results are also suggested in Rose and Black (2004), where the mixed logit model is used. After incorporating response time as a parameter in the estimation, Haaijer et al. (2000) found it to have a negative coefficient, implying that choice heterogeneity decreases as response time increases.

In a study that examined the determinants of response time, Bonsall and Lythgoe (2009) found that response times varied with the characteristics of the choice set, the order of presentation, and the personal characteristics of respondents. Personal characteristics that emerged as important were age, education and self-reported decision-making style. However, Vista et al. (2009) tested the impact of respondents' demographics and attitudes to the product on response time, and found that demographic characteristics were not significant determinants of response time. When the sample was divided into high and low response times, Vista et al. (2009)

reported no significant results in terms of the amount of time respondents spent on attribute information, experiment description, choice questions, or completion time for the entire survey.

The literature suggests a relationship between response time and the random-error component of the utility function. However, there is no consensus on the direction of this relationship. In the literature reviewed in this section, three main issues are observed. Firstly, the literature on response time is predominantly in domains other than the field of environmental economics. Secondly, most of the existing studies are web-based, with very few, if any, using face-to-face interviews. Thirdly, studies on response time are predominantly conducted in developed nations, with less attention given to developing nations. Our current study addresses these issues by conducting a face-to-face survey in the Gauteng province of South Africa. We examine the impact of response time using data on household preferences for water-conservation technologies. Response time for each respondent is captured and categorised as either fast or slow. In so doing, our study addresses the current lack of studies in environmental economics that use face-to-face surveys to test for the impact of response time in the context of a developing country.

5.3. Case study: household preferences for water-efficient technology

South Africa is one of the 30 driest countries in the world. It is among those that will have renewable water resources below the threshold of 1,500 cubic metres per capita per year by the year 2030 (Rijsberman, 2006; Yang et al., 2003). In addition to being naturally water-scarce, South Africa continues to experience water supply problems due to various issues including population growth, economic growth, increased urbanisation and changing climatic patterns. With an average per capita water consumption of 235 litres per day, South Africa’s water consumption far exceeds the international benchmark of approximately 180 litres per day (Department of Water and Sanitation, 2017). For instance, in Gauteng province the average consumption is 305 litres per person per day. Considering this mismatch between available water resources and demand for water, there is a need for the country to prioritise efficiency in the management of water resources.

Households can play a part in promoting efficiency in the management of water resources, through installing efficient water devices. Common devices that save water include efficient showerheads, shower timers, dual-flush cisterns, multi-flush cisterns, cistern displacement devices, time-based irrigation controllers, and micro-drip irrigation systems (see Jones and Hunt, 2010; Makki et al., 2013; Mini et al., 2015; Still and Bhagwan, 2008; Willis et al., 2013). Because of the high cost of purchasing and installing some of these devices, most households in developing countries do not have them installed. In South Africa, very little is known about households' use of efficient water technologies. Therefore, an elicitation of household preferences for these technologies is a step in the right direction for both conservation and the generation of knowledge.

A case concerning preferences for water-efficient technologies is ideal for testing the impact of response times on choice experiments. By using the setting of the water sector, our study bridges the existing gap in choice-experiment studies in environmental economics that test the impact of response time in the context of a developing country. We conducted our experiments in Gauteng province. Gauteng was chosen because it is the largest water user in the country. Gauteng, KwaZulu-Natal and the Western Cape together account for 66% of the country's water demand. Gauteng has the highest per capita consumption due to the high number of wet industries in its supply area and has been unable to reduce its water demand over the years. In 2015/16 alone, water balances for all municipalities in Gauteng showed water losses (i.e. system leaks) of about 27.4% (Department of Water and Sanitation, 2017).

Located in the northern part of South Africa, Gauteng is the smallest province in terms of land area; but it is the most urbanised and populous province, with approximately 14.7 million people, about 27% of the country's total population (Statistics South Africa, 2018). The province is the only one in South Africa with three metropolitan municipalities. These are municipalities with urban cores that are highly populated due to urbanisation. We conducted our experiment in three municipalities, namely City of Johannesburg Metropolitan, Ekurhuleni Metropolitan and Mogale City. The total population for these three municipalities is around 8.9 million people; 5.1 million in the City of Johannesburg, 3.4 million in Ekurhuleni, and 383,864 in Mogale City (Statistics South Africa, 2017). Our sample included both suburbs and townships. The former are affluent areas characterised by better water infrastructure and lower population densities, while the latter are low-income areas that were designed for non-white South Africans under the apartheid segregation laws. Municipalities in Gauteng are shown in the map in Figure 5.1.

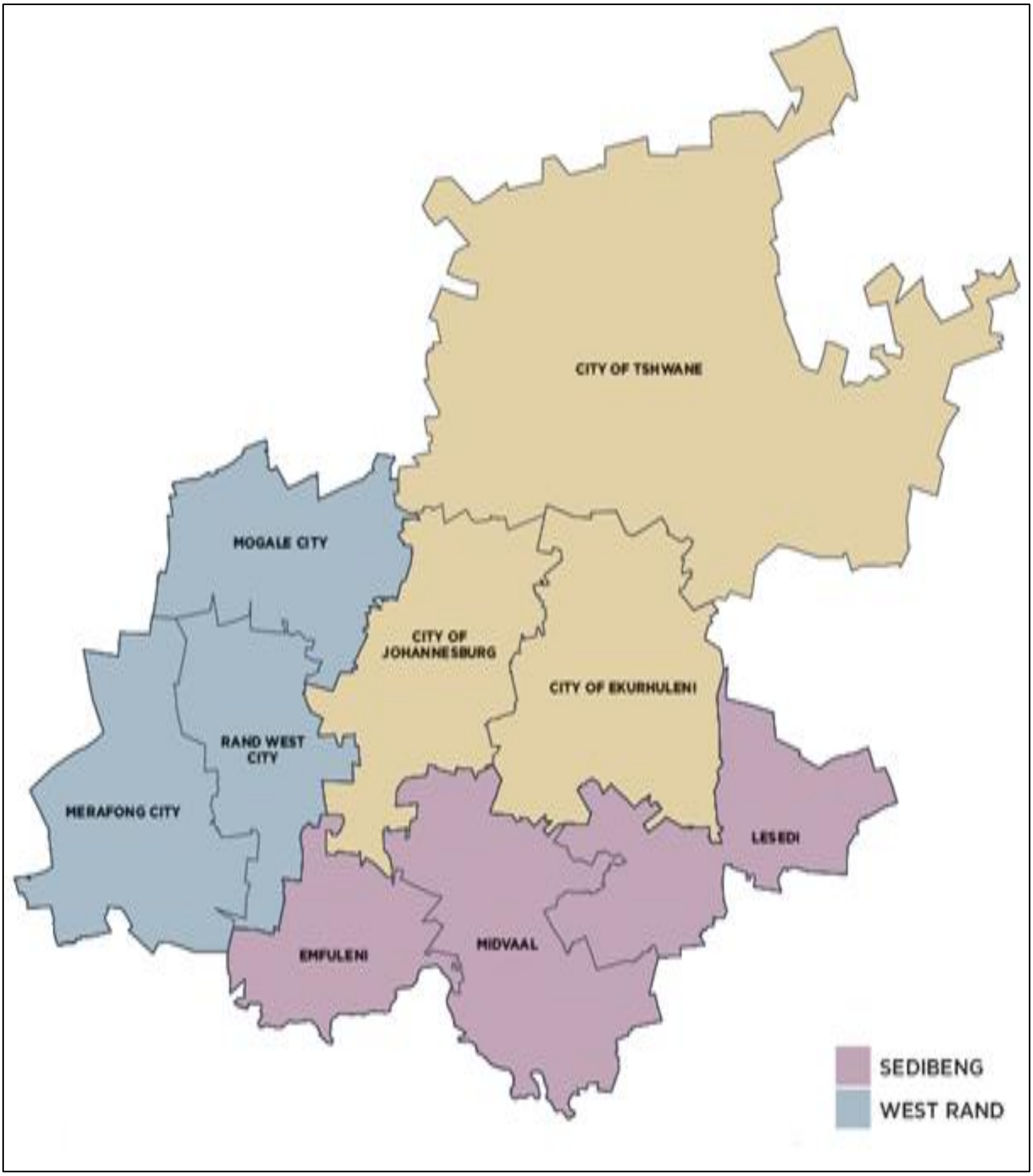


Figure 5.1: Map of Gauteng province

Source: Municipalities of South Africa (2018)

Although each of the three municipalities selected is itself comprised of several areas, divided into suburbs and townships, we surveyed only a few areas in order to represent the geographic and demographic characteristics of each municipality. This decision to sample a few areas in each municipality was due to budget constraints. Areas were selected based on their population statistics, socio-economic characteristics and geographical locations. In the City of Johannesburg, we conducted the survey in Soweto, South Africa's most populous township, with a population of more than 1.5 million people. Suburban households surveyed in the City of Johannesburg were from Ennerdale, Lenasia, Midrand, Randburg, Roodepoort and Sandton. These areas represent the eastern, western, northern and southern parts of the municipality. In Ekurhuleni, the survey was conducted in three townships, namely Duduza, Tembisa and Tsakane; and three suburban areas, namely Benoni, Kempton Park and Springs. In Mogale City, we conducted the survey in two townships, namely Kagiso and Munsievile, while Krugersdorp was the only suburban area surveyed. These areas are truly representative of areas in the province.

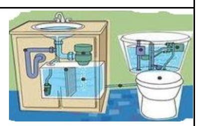
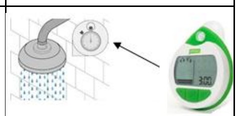
5.4. Experimental design

Experimental design is the specialised and scientific manipulation of the levels of one or more attributes in order to generate choice profiles (Hensher et al., 2015). Attributes are translated features and characteristics that show the objective properties of a commodity. Each attribute consists of various sub-features called levels. When conducting choice experiments, the initial step is the selection of relevant attributes and the assigning of feasible, realistic, and non-linearly-spaced levels to each attribute. Normally, a literature review, focus groups, pilot studies and expert consultations are used to determine the most relevant attributes and levels. Once these have been determined, experiment design commences. This section presents the attributes and levels used in the study, and how these are designed into choice profiles.

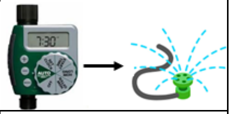
5.4.1. Attributes and levels used in the study

In determining the attributes for our experiment on household preferences for water-efficient technologies, we used a combination of both literature reviews and expert consultation. The literature shows that water-efficient devices can be grouped based on the areas of a home, namely the kitchen, shower, toilet and garden/outdoors. This is because a typical South African middle-income household uses 25% of water in the toilet, 25% in garden/outdoor activities, 24% in the bath/shower, 13% in the laundry, 11% in the kitchen, and 2% in other activities (Price, 2009). We use these areas as attributes and adopt technologies that can be fitted in these areas as levels. Additionally, we include the monthly water bill as a monetary attribute. Noting that installing water-efficient devices reduces the monthly water bill, the possibly reduced monthly water bills are used as levels. Using data from the National Income Dynamics Study (NIDS), an average of R450 (around $31.47) per month was the current average water bill[26]. If households were to adopt water-efficient devices, their monthly water bill would decrease by 75%, 50% or 30% – that is, from R450 to one of R110, R225 or R315 ($7.69, $15.73 or $22 respectively) per month. Table 5.1 below presents the refined list of attributes and levels used in the study.

Table 5.1: Attributes and levels used in the study

Attribute	Description	Attribute Levels	
Kitchen devices 	A typical household uses about 11% of its water in the kitchen. A standard tap flows at about 8l per minute. Installing water-flow regulators or tap-head aerators make a standard tap more efficient and saves water by 60%. An efficient dishwasher uses 15l per cycle, using 50% less water than is used in a conventional dishwasher.	**Level 1:** Efficient dishwasher	
		Level 2: Efficient tap	
		Level 3: System collecting used water	
Shower devices 	A typical household uses about 24% of its water in the shower. Shower timers result in shorter showers. Efficient showerheads save 65% of water used in the shower.	**Level 1:** Efficient showerhead	
		Level 2: Shower timer	

[26] As at 24 October 2018, US$1 = ZAR14.30. This exchange rate will be used throughout this chapter in instances when South African Rands (ZAR) must be converted to United States Dollars (US$).

Toilet devices 	A typical household uses about 25% of its water in the toilet. Replacing a 12l cistern with a 3l dual cistern saves about 75% of water. An interruptible flush cistern allows users to control how long the toilet flushes. Hippo bags displace water in the cistern and save about 1.2l per flush.	**Level 1:** Dual-flush cistern sized 3-6l	
		Level 2: Interruptible-flush cistern	
		Level 3: Cistern displacement (hippo bag)	
Garden & Outdoor devices 	A typical household uses about 25% of its water in the garden or in outdoor activities. Efficient gardening technologies reduce water use by 30%. These include time-based irrigation control systems, and micro-drip systems. Irrigating gardens using water collected with water tanks also saves water.	**Level 1:** Time-based irrigation controller	
		Level 2: Micro-drip systems	
		Level 3: Use harvested rain water	
Monthly water bill 	The average water bill for a household is R450 per month. Installing water-efficient technologies will reduce the monthly water bill by 30%, 50% or 75%.	**Level 1:** R110 **Level 2:** R225 **Level 3:** R315	

We use the attributes and levels given in Table 5.1 above to design the choice profiles presented to respondents in our choice experiment. We agree that the technologies used as levels may also be used as attributes in other studies. However, in the context of our study the emphasis is on the areas in a home where households can save water by installing efficient technologies. As such, water-efficient devices that can be fitted in these areas are used as levels in our experiment.

5.4.2. Choice experiment design

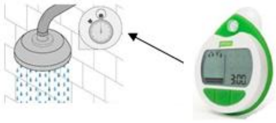
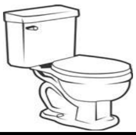
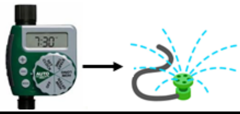
Various classes of experimental designs exist in the literature. The most common are full factorial, orthogonal and efficient designs (Bliemer et al., 2017; Hensher et al., 2015; Rose and Bliemer, 2009). This study uses an efficient design to generate choice profiles. Efficient designs address most of the shortcomings encountered in the other designs. They produce more robust data that lead to more reliable parameter estimates with even lower sample sizes and smaller confidence-interval widths. Using efficient designs requires some knowledge of prior parameters, because if incorrect prior parameters are used, the design becomes inefficient. If prior parameters are not known, they can be drawn using Bayesian parameter distributions. These are less sensitive to misspecification because they assume parameter values to be approximately known and randomly distributed (Bliemer et al., 2008). If a D-*error* statistic is used and prior parameters are drawn using Bayesian distributions, the design is called a Bayesian D-*error* design (i.e. D_b-*efficient*):

$$D_b - error = \int_{\tilde{\beta}} \det\left(\Omega_1\left(X, \tilde{\beta}\right)\right)^{1/K} \phi\left(\tilde{\beta} \,\middle|\, \theta\right) d\tilde{\beta}. \tag{5.1}$$

where D_b is Bayesian design, Ω_1 is the asymptotic variance-covariance matrix of the design, $\tilde{\beta}$ represents prior parameters, X is the experimental design, and K is the number of parameters to be estimated. A normally distributed Bayesian D-efficiency criterion was employed to design choice sets experimentally. The maximum possible Gaussian draws (i.e. 32) were used to determine the number of draws for Bayesian priors.

The final design consisted of six choice sets of two profiles each. In addition to the two designed profiles, each choice set included an undefined status quo (SQ). This is an individual-specific SQ, for which each respondent envisages their own current status and compares it to the experimentally designed hypothetical options (Hess and Rose, 2009). Undefined SQs are commonly used when it is difficult to ascertain the current situation for the sample (see Campbell et al., 2008; Hess and Rose, 2009; Scarpa et al., 2007). This was the case in our experiment, where we could not determine the current use of water-efficient devices with certainty. An example of the choice sets used in the experiment is given in Table 5.2 below.

Table 5.2: Example of the choice sets used in the experiment

	Status quo	Option 1	Option 2
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Hippo bag 	Dual-flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rainwater 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

In addition to the choice experiment (the first section), our questionnaire contained two other sections. The second section collected general information on water-conservation technology and behaviour, while the third section collected the biographical details of the respondents. Furthermore, we captured each respondent's response time by taking note of the time when the interview commenced and the time it ended. The duration of the interview was determined to be the difference between the start and finish times of the interview. Our approach is different from other studies, which consider response time to be the time taken by respondents to complete each choice task (see Börger, 2016; Campbell et al., 2017). Such an approach is convenient in web-based surveys where timers are incorporated but may be complex in face-to-face surveys such as ours. Our current study measures response time as the duration for completing the entire questionnaire[27].

[27] The questionnaire used to collect information is given in Appendix 5.1.

5.5. Modelling

The theoretical foundation of choice experiments is based on the random utility theory, which assumes that individuals make choices based on the characteristics of a good, along with an error component (Ben-Akiva and Lerman, 1985; McFadden, 1974). The error component is due to the uniqueness of each participant's preferences, or because researchers do not have enough information on the observed participant. Therefore, the utility of a participant obtained from an option is not known with certainty, but can be decomposed into a deterministic and an unobserved error component, as follows:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{5.2}$$

where U_{ij} is the utility of participant i obtained from option j, V_{ij} is the deterministic component, and ε_{ij} is the unobserved error term. Equation 5.2 is the basic utility function, which can alternatively be written by decomposing the deterministic component as follows:

$$U_{ij} = V_{ij}\left(X_{ij}, C_{ij}, \beta\right) + \varepsilon_{ij} \tag{5.3}$$

where X_{ij} is a vector of the attributes associated with option j, C_{ij} is a vector of the monetary attribute of option j, and β is a vector of preference parameters for the population in the sample. In the random utility theory, participants are assumed to be rational. Therefore, participant i is expected to choose option j over option k if $U_{ij} > U_{ik}$. The selection of one option over the other signifies that a participant's hypothetical utility from the chosen option is greater than the utility of the option not chosen (Greene, 2003)[28].

[28] Important to not is that in deriving the probability of choosing an alternative within the random utility model, the choice of alternative with higher utility is not certain. The expectation has always been that there is a high chance that a respondent will choose the alternatives with higher utility.

To estimate utility, studies such as ours use the multinomial logit (MNL) model (Campbell et al., 2017; Haaijer et al. 2000) and the mixed logit (MXL) model (Dellaert et al., 2012; Rose and Black, 2004; Rose and Black, 2006). The former has the problem that it assumes participants have homogeneous tastes for observed attributes, and that the random part of the utility obeys the properties of independence from irrelevant alternatives (IIA) and independence and identical distribution (IID). These are unrealistic assumptions, as they rule out persistent heterogeneity in taste for observed and unobserved product attributes (Greene, 2012; Hensher et al., 2015). On the other hand, MXL can identify taste heterogeneity, but fails to account for scale heterogeneity across participants. Scale heterogeneity is the variance of the variance term, also explained as the standard deviation of utility over different choice sets (Hensher et al., 2015). This problem is addressed in generalised mixed logit (GMXL) models[29].

Therefore, we follow a similar study to Czajkowski et al. (2014) in adopting the GMXL model as a tool to estimate household preferences for water-efficient technologies. The ability to take care of scale heterogeneity makes the model favourable to our data set, which we hypothesise to have such a problem because it contains respondents from different socio-economic statuses. Developed by Fiebig et al. (2010), GMXL builds on the specifications of the mixed logit model and the generalised multinomial logit model. The essential format of the GMXL model is:

$$U_{ij} = \boldsymbol{\beta}_i' \mathbf{x}_{ij} + \varepsilon_{ij} \tag{5.4}$$

$$\boldsymbol{\beta}_i = \sigma_i \boldsymbol{\beta} + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\Gamma}\mathbf{w}_i,\ \mathbf{w_1} \sim \mathrm{N}[\mathbf{0}, \mathbf{I}], 0 \leq \gamma \leq 1 \tag{5.5}$$

$$\sigma_i = \exp\left(-\frac{\tau^2}{2} + \tau \mathrm{v}_i\right), \mathrm{v}_i \sim \mathrm{N}[0,1] \tag{5.6}$$

Equation 5.5 is an MNL model based on the extreme value distribution of the error component ε_{ij}. The general form of the GMXL model combines the scaled MNL model with

[29] Several advantages of GMXL are given in Greene (2012) and Hensher et al. (2015).

the random parameter model. A random scaling factor σ_i with mean 1 and variance $\exp(\tau^2 - 1)$ is included in the model. Hensher et al. (2015) suggest that Gamma γ is central to the GMXL model, as it controls the relative importance of the overall scaling of the utility function. The other important element of GMXL is the Tau scale τ. These parameters are interpreted as:

$\tau = 0$ suggests the random parameters model, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_i$
$\gamma = 0$ suggests a scaled random parameter logit model, $\boldsymbol{\beta}_i = \sigma_i[\boldsymbol{\beta} + \sigma_i\boldsymbol{\Gamma}\mathbf{w}_i]$
$\gamma = 1$ suggests a hybrid model, $\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_i$

We also estimate the MWTP for the non-monetary attributes. MWTP shows the marginal rate of substitution between each attribute and the monetary attribute. Such estimates are important in choice modelling, because they show what respondents are prepared to pay for or against changes in each attribute. Some studies in the literature show that response time has no impact on willingness-to-pay estimates (Börger, 2016; Campbell et al., 2017; Rose and Black, 2006). However, it is essential for us to test the impact of response time on MWTP estimates in the context of environmental economics.

The literature provides very little a priori guidance on what characterises fast and slow response times. We follow Campbell et al. (2017) to determine average, fast and slow response times, by calculating the median and mean response times. Subsequently, we use all data points around the sample median and mean as average responses. We then benchmark individual response times against the calculated median and mean response times. Where response time is less than the median, it is deemed fast; and when response time is greater than the mean, it is considered slow. Estimations of utility functions and marginal willingness-to-pay estimates are then based on these sub-datasets.

5.6. Experimental Data

5.6.1. Data collection and descriptive statistics

The study is based on experimental data for 307 household heads, collected in Gauteng during the periods November to mid-December 2017 and mid-January to February 2018. The number of respondents interviewed was 405; however, the data for 98 respondents did not have response time captured correctly and were excluded from analysis[30]. Our questionnaire was prepared in English, and enumerators conversant in both English and the local languages were recruited, trained and supervised during the data-collection process.

Our questionnaire had three sections. The first section was the choice experiment, while the second section collected general information on households' current use of water-efficient technology. The rationale for including this section was to gather information on the number of households in our sample that currently have water-efficient devices installed. We assume that respondents who currently have water-efficient devices will have lower response times because they are familiar with the devices presented in the experiment. The third section of the questionnaire collected the biographical information of the respondents. Such information is necessary because the literature suggests a link between response time and biographical characteristics (see Bonsall and Lythgoe, 2009). Response time was captured by noting the time the survey began and the time it ended. Therefore, response time is defined as the duration each respondent took to answer the questionnaire[31]. Descriptive statistics of the data are given in Table 5.3.

[30] Small samples in choice experiments produce many observations, because each respondent makes multiple choices. Additionally, the experimental design used in this study produces more robust data with lower sample sizes.

[31] We acknowledge that it was essential to capture response times for the choice experiments section only as opposed to the whole questionnaire. However, it would have been a complex exercise to use that approach in our face-to-face survey. Therefore, we opted for capturing the time taken to complete the survey.

Table 5.3: Descriptive statistics of respondents

	Whole sample[32] Mean	Average responses Mean	Fast responses Mean	Slow responses Mean
Response time (minutes):				
Minimum	1	5	1	11
Median	6	7	3	16
Mean	7	7	3	17
Maximum	28	10	4	28
Male respondents (%)	51	56	47	41
Average household size	4	4	4	4
Average age	45	44	38	42
Married respondents (%)	57	46	64	48
Race (%):				
Black	82	79	85	76
White	13	15	11	21
Indian/Asian	4	5	2	3
Coloured	1	1	1	0
Education (%):				
Never attended school	1	1	0	3
Primary	3	3	2	3
High school	69	74	63	69
Certificate	13	11	20	7
Diploma	10	8	10	7
Degree	4	2	4	10
Postgraduate	1	1	0	
Source of income (%):				
Salaries/wages	56	52	64	41
Business	20	21	19	24
Pension	17	20	12	24
Grants/allowances	3	4	0	3
Other	3	4	4	7
Monthly household income (%):				
<R5,000	43	53	34	41
R5,000 to R10,000	37	28	41	45
R10,000 to R20,000	19	17	23	14
R20,000 to R40,000	1	2	1	0
R40,000 to R60,000	0	0	0	0
>R60,000	0	0	0	0
Number of respondents (N)	***307***	***188***	***90***	***29***

Out of the whole sample, the fastest respondent took 1 minute to complete the survey, while the slowest respondent took 28 minutes. The average response time for the sample was 7 minutes, with a median of 6 minutes. Average respondents constituted approximately 61% of the total sample, while fast respondents and slow respondents constituted about 29% and 10% respectively. The proportion of female respondents who took longer to complete the survey was greater than that of male respondents (only 41% of slow respondents were males). It is also observed that the average age of the respondents was higher for slow responses than for fast responses, implying that the older a respondent was, the more time they would take to

[32] The "whole sample" in this case implies all data, consisting of both fast and slow responses. It is imperative to also estimate results for the combined dataset and see how they compare to those from fast and slow responses.

complete the survey. We also note that education played a role in the time respondents took to complete the survey. The number of respondents with post-high-school certificates and diplomas was greater in the fast responses than in the slow responses. This implies that the higher the education level, the less time a respondent would take to complete the survey.

The summary statistics presented in Table 5.3 above show that response times varied from 1 to 28 minutes, with a median response time of 6 and a mean of 7 minutes. Since response time is an important variable in our analysis, it is imperative to show its distribution across all 307 respondents. We use the diagram in Figure 5.2 to show the response time for each respondent in our sample.

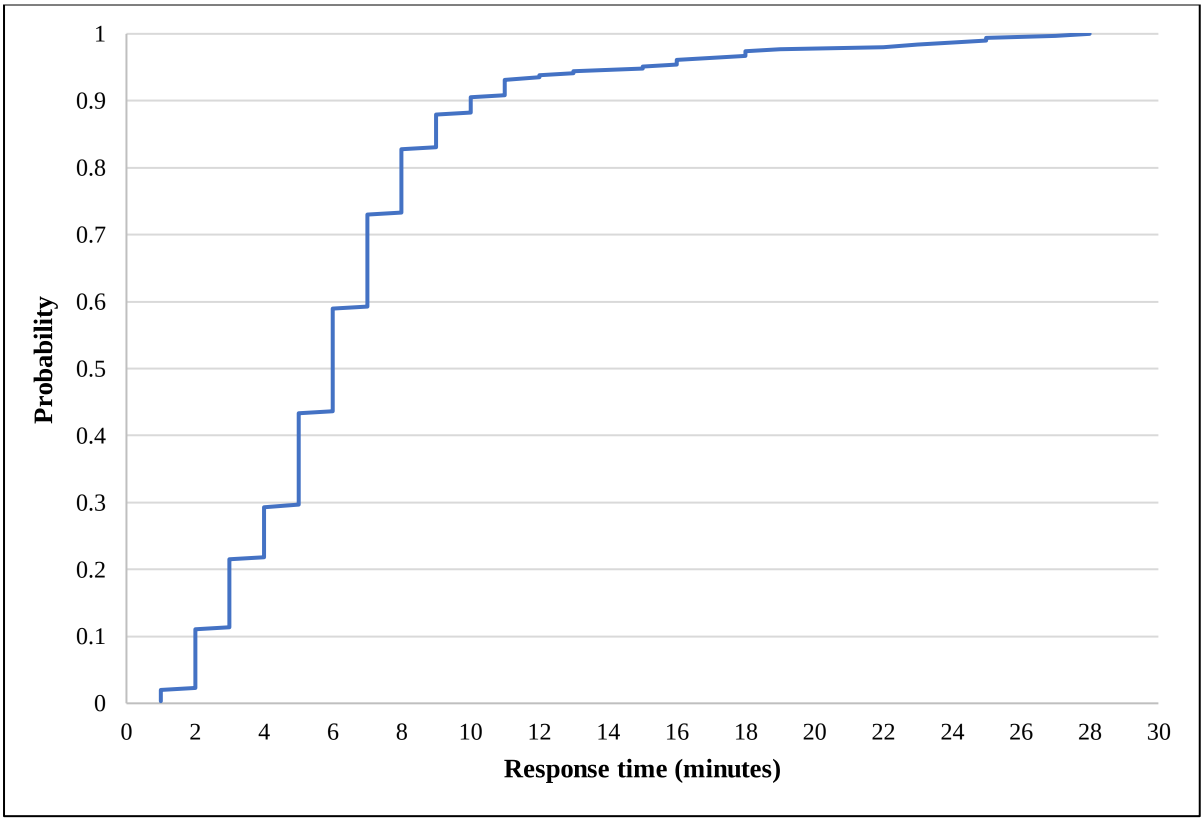


Figure 5.2: Distribution of response times

The figure shows that most of the respondents took between 5 and 10 minutes to complete the survey. However, variations around the mean are observed, as some respondents took more than 20 minutes, while many others took less than 4 minutes to complete the survey. The distribution in the figure is in line with the descriptive statistics presented earlier, and warrants testing for the impact of response time on empirical estimates. However, before we adopt econometric tools to test whether response time affects empirical estimates, we present statistics on households' current use of water-efficient devices.

Regarding the current use of water-efficient technologies, we used a 4-point Likert scale to collect information on whether households have water-efficient devices currently installed. Respondents were asked eight questions, using a scale with the options 'Yes', 'No', 'Not applicable' and 'Not sure'. Overall, the modal response was 'No', indicating that respondents did not have water-efficient devices at the time[33]. As a follow-up, we collected information on the possible reason's households had not installed efficient devices. We asked respondents to choose between 'I cannot afford them', 'I did not know about them', 'I have no infrastructure to connect them', and 'They are not important to me'. Summary statistics[34] show that most households cannot afford the technology. A few other respondents indicated that they did not know about water-efficient technologies or that they are not important.

Finally, we present the frequency distribution of the stated-preference choices. A presentation of the distribution of choices made by respondents is important when checking if respondents made real trade-offs between given choice profiles. Where consistent choices are observed across the response-time categories, it makes the comparison of empirical estimates possible. This information is given in Figure 5.3 below.

[33] Summary statistics on the use of water-efficient technology are presented in Table A1 in Appendix 5.2.

[34] The frequency distribution of reasons for not installing efficient technologies is given in the Figure in Appendix 5.3.

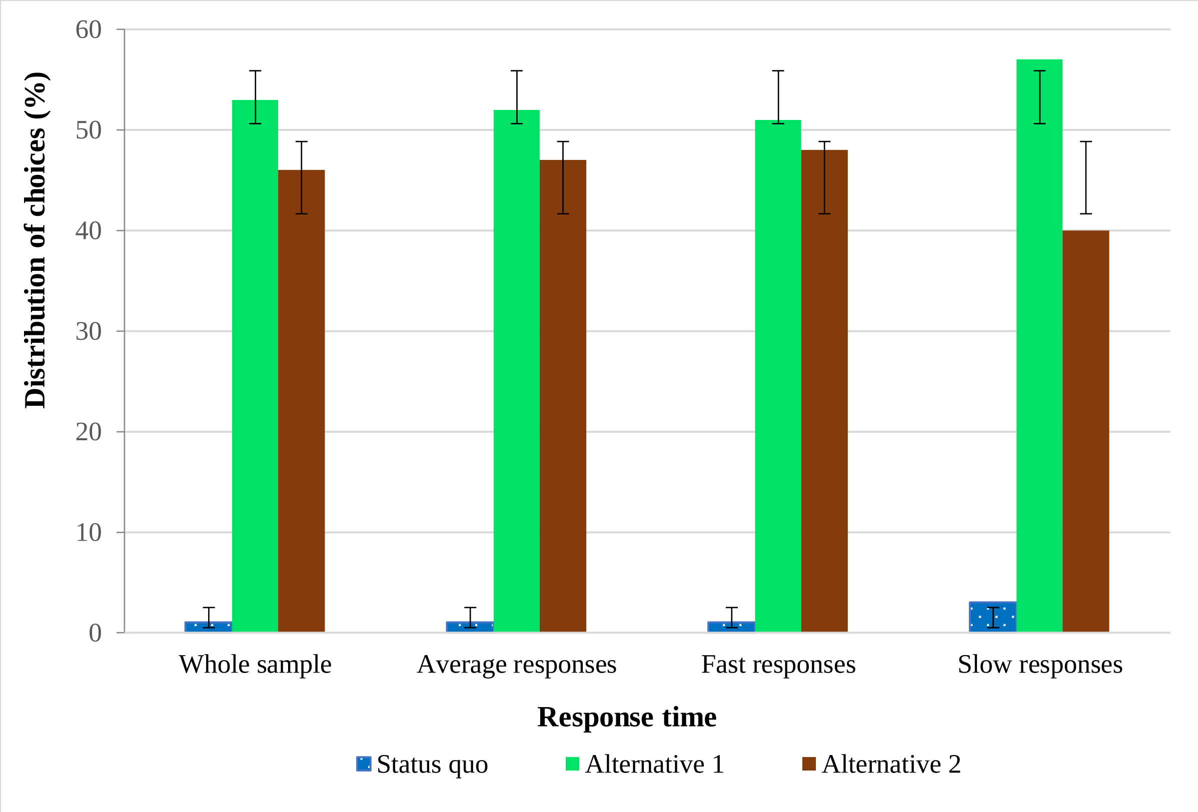


Figure 5.3: Frequency distribution of choices based on response times

For all categories (average, fast and slow response times), comparable trade-offs are noted between the two experimentally-designed alternatives. It would be of interest to find out how the empirical results from these categories compare with each other. It is also interesting to note that the introduction of an individual-specific status quo addressed the problem of status quo bias, which is commonly reported in the choice-experiment literature (see Anderson, 2003; Dubé et al., 2010; Lanz and Provins 2015; Mandler, 2004; Meyerhoff and Liebe, 2009). We also observe that the distribution of choices made in all three categories is consistent with the distribution shown in the whole sample. This makes comparisons of empirical estimates necessary. In the next section, we present and discuss the empirical results from the utility functions and marginal willingness-to-pay estimates for the response-time categories.

5.7. Empirical findings

This section presents the results of the study. The section is divided into three main parts. The first part presents the results of the determinants of response time, by estimating response time as a function of selected biographic characteristics of respondents. In doing this, we join various other studies in the literature that test the impact of biographical characteristics on response time (see Bonsall and Lythgoe, 2009; Börger, 2016; Recalde et al., 2014; Vista et al., 2009). The second part presents results on the estimated utility functions. These results are the basis of our study, as they indicate the impact of response time on utility estimates. Finally, we present results on the impact of response time on MWTP estimates.

5.7.1. The determinants of response time

To examine the determinants of response time, we use the ordinary least squares (OLS) as an estimation tool. OLS is used in many other similar studies to test for the determinants of response time (see Börger, 2016; Recalde et al., 2014). In this study, we estimate response time as a function of selected biographical characteristics of respondents. Additionally, we control for interviewer bias and for the way the interview was administered. Regarding the former, three enumerators conducted the survey, and enumerator bias is hypothesised to affect the time taken by each respondent to complete the survey. In terms of the way the interviews were conducted, we assume that respondents who completed the survey on their own would take more time than the time taken by respondents who were assisted by the enumerator. This information was captured in our questionnaires through the inclusion of a question in which respondents were asked to indicate whether they answered the questions on their own with very minimal help from the enumerator. The functional form of the OLS model estimated in this study is:

$$RT = \beta_0 + \beta_1 GENDER + \beta_2 RACE + \beta_3 STATUS + \beta_4 EDU + \beta_5 INCOME + \beta_6 ANSWER + \beta_7 NUMERAT + \varepsilon \quad (5.8)$$

where RT is response time, $GENDER$ is the gender of the respondent, $RACE$ is the racial group of the respondent, $STATUS$ is the marital status of the respondent, EDU is the education level of the respondent, $INCOME$ is the income level of the respondent, $ANSWER$ captures whether the respondent answered the survey on their own, $NUMERAT$ is the identity of the enumerator, Parameter β_0 is the constant and parameters β_1 to β_7 are the coefficients of the determinants, while parameter ε is the error term. The estimation results are presented in Table 5.4 below.

Table 5.4: OLS results on the determinants of response time

	Coefficients	**Standard errors**
GENDER	0.107	0.148
RACE	-0.313	0.203
STATUS	-1.117***	0.144
EDUCATION	-0.148	0.104
AGE	0.050***	0.006
INCOME	-0.128	0.124
ANSWER	1.454***	0.373
NUMERAT	1.743***	0.177
_CONS	0.273	0.838
Model parameters		
Number of observations	2160	
Probability > F	0.000	
R^2	0.84	
Adjusted R^2	0.84	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively.

The OLS results presented in Table 5.4 show that STATUS, AGE, ANSWER and NUMERAT were statistically significant determinants of response time. The negative coefficient of STATUS suggests that married respondents took less time to complete the survey relative to respondents whose marital status was single. On the other hand, the positive coefficient of AGE implies that older respondents took more time to complete the survey than younger respondents. The statistical insignificance of some demographic variables agrees with studies such as Vista et al. (2009), which found that biographical characteristics do not affect response time. Since most of the biographical characteristics were statistically insignificant, we argue that although marital status and age were statistically significant, biographical variables in general did not affect response time.

The results also show that response time was determined by the way the survey was conducted, as well as by the enumerator conducting the survey (i.e. ANSWER and NUMERAT), which confirms the existence of interviewer bias. ANSWER is a binary variable where a value of 1 was assigned to a "yes" response, meaning that the respondent completed the survey on their own and 0 otherwise. The positive coefficient for ANSWER suggests that respondents who completed the survey on their own took longer than those who asked the enumerator to tick the boxes for them. On the other hand, the positive coefficient of NUMERAT suggests that response time was also determined by the enumerator. This implies that it is imperative to control for interviewer bias when estimating models based on data collected by several enumerators. GENDER is another binary variable were a value of 1 was assigned to male respondents while a value of 0 assigned to female respondents.

5.7.2. Utility function estimates

To examine the impact of response time on the utility functions, we use the GMXL model as an estimation tool. We estimate two sets of GMXL models. The first set of models estimates utility functions for average responses, fast responses, and slow responses. Results from these three sub-categories are compared across sub-categories, as well as to results for the whole sample. In the second set of models, we estimate utility functions using data that exclude fast responses and data that exclude slow responses. Results from these two estimations are then compared across each dataset, and to the results generated by the whole sample data. The five attributes of the study are modelled as normally distributed random parameters, while the alternative specific constant (ASC) is modelled as a fixed parameter. Results are obtained using the Halton sequence for simulation based on 1,000 draws. Utility functions for the first set of estimates are given in Table 5.5.

Table 5.5: Comparison of estimates from whole sample, average, fast and slow responses

	Whole sample		Average responses		Fast responses		Slow responses	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Random parameters in utility functions								
B_KITCHEN	0.164***	0.062	0.227***	0.068	0.124	0.196	-0.935	17.554
B_SHOWER	0.095	0.121	0.131	0.153	-0.124	0.337	2.318	33.765
B_TOILET	0.018	0.060	0.055	0.067	-0.085	0.208	0.293	4.249
B_GARDEN	-0.002	0.071	-0.008	0.076	-0.080	0.204	0.714	14.192
B_BILL	-0.002***	0.001	-0.003***	0.001	-0.001	0.002	-0.009	0.158
Non-random parameters in utility functions								
ASC	0.265	0.724	0.0	0.954	0.0	0.145	0.0	0.223
Diagonal values in Cholesky matrix, L.								
NsB_KITCHEN	0.259**	0.131	0.287**	0.135	0.261	0.386	1.490	11.623
NsB_SHOWER	0.386	0.507	0.438	0.317	0.396	1.445	2.088	11.093
NsB_TOILET	0.017	0.703	0.011	0.156	0.041	0.491	0.954	6.132
NsB_GARDEN	0.254	0.290	0.086	5.287	0.012	15.220	0.904	17.373
NsB_BILL	0.001	0.008	0.002	0.005	0.002	0.043	0.005	0.427
Below diagonal values in L matrix. V = L*Lt								
B_SHO:B_KIT	0.630**	0.291	-0.574**	0.250	-0.640	0.775	-0.2.529	14.897
B_TOI:B_KIT	0.031	0.116	-0.062	0.126	0.042	0.596	0.777	3.925
B_TOI:B_SHO	0.056	0.181	0.030	0.145	0.168	0.530	0.868	7.479
B_GAR:B_KIT	0.028	0.151	-0.041	0.164	0.043	0.478	-0.178	14.461
B_GAR:B_SHO	0.092	0.229	0.033	0.204	0.199	0.588	0.973	6.504
B_GAR:B_TOI	-0.038	1.442	-0.254	1.851	-0.115	1.180	1.802	11.076
B_BIL:B_KIT	-0.002	0.003	0.002	0.002	0.003	0.004	0.003	0.150
B_BIL:B_SHO	0.002	0.004	0.003	0.002	0.001	0.009	0.021	0.086
B_BIL:B_TOI	-0.495	0.012	0.001	0.011	0.003	0.018	0.0003	0.062
B_BIL:B_GAR	-0.002	0.004	0.001	0.037	0.0003	0.135	-0.012	0.127
Variance parameter tau in GMX scale parameter								
TauScale	0.397	1.374	0.048	2.248	0.00	8.184	2.160	8.972
Weighting parameter gamma in GMX model								
GammaMXL	0.707	3.126	0.100	Fix. Par.	0.100	Fix. Par.	0.100	Fix. Par.
Sample Mean	**Sample Std.Dev.**							
Sigma(i)	0.992**	0.404	0.998***	0.064	0.998***	0.043	0.761	2.579
Standard deviations of parameter distributions								
sdB_KITCHEN	0.259**	0.131	0.287**	0.135	0.261	0.386	1.490	11.623
sdB_SHOWER	0.739	0.490	0.722***	0.170	0.753*	0.429	3.279	18.225
sdB_TOILET	0.066	0.260	0.070	0.099	0.178	0.476	1.506	7.940
sdB_GARDEN	0.274	0.444	0.273**	0.121	0.234	0.835	2.247	10.187
sdB_BILL	0.004	0.006	0.004	0.001	0.005***	0.001	0.0025	0.128
N	1839		1122		539		174	
LL	-1790.9		-1058.4		548.4		-157.5	
AIC	3627.7		2160.8		1140.9		358.9	
BIC	3754.6		2271.3		1235.3		428.4	
Pseudo R^2	0.1		0.2		0.1		0.2	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Std. Err are standard errors.

Results in the table are compared based on the statistical significance, sign and magnitude of attribute parameter estimates. The statistical significance of attribute parameters is the most important aspect to consider, as it shows the attributes that are important to respondents and those that are not. When an attribute has a statistically significant coefficient, it implies that the attribute is important to the respondents, while a statistically insignificant coefficient suggests that the attribute is not important to them (Hensher et al., 2015). The sign and magnitude of the attribute parameters respectively show the direction of the impact an attribute has, and the extent of the impact to respondents' utility (see Campbell et al., 2017; Patterson et al., 2017; Rose and Black, 2006).

Positive and statistically significant attribute parameters suggest that households prefer changes in the attribute, while negative and statistically significant attribute parameters suggest otherwise. Using this interpretation, positive attribute parameters imply that improvements in the attribute would increase household utility, whereas negative attribute parameters suggest that changes decrease household utility. For example, in the 'whole sample' model, household utility increases by about 0.164 following an improvement in KITCHEN technologies, while an increase in the monthly water BILL reduces utility by about 0.002. These results are consistent with both prior expectations and the literature, and imply that when making choices, respondents were likely to choose profiles with improvements in KITCHEN technologies and were likely to avoid profiles with a higher monthly water BILL.

The table shows similarities in the estimates reported in the 'whole sample' model and the 'average responses' model. In these two models, all parameters are largely similar in terms of statistical significance, sign and magnitude. On the other hand, we observe that all attributes in the 'fast responses' model and the 'slow responses' model reported insignificant parameter estimates[35]. It is important to note that the results from fast and slow responses could be due to the sample sizes of these sub-categories, which were relatively low compared to the average responses sub-category. This could have driven the similarities observed between the estimates from the 'average responses' and 'whole sample' models, which implies that results for the whole sample were largely determined by data from respondents who took 'average time' to complete the survey.

[35] It was also observed that both fast and slow responses had larger confidence intervals than average responses. Confidence intervals as well as standard errors for slow responses were very large. These observations are consistent with various similar studies in the literature, which argue that quick responses are a source of random error (see Conrad et al., 2017; Huang et al., 2015; Wood et al., 2017).

In the literature, fast responses are sometimes argued to be 'quick and dirty', implying that failure to account for them can lead to incorrect model inferences being drawn (see Conrad et al., 2017; Huang et al., 2015; Rose and Black, 2006; Schwappach and Strasmann, 2006; Wood et al., 2017). Our results show that fast and slow response datasets both produced statistically insignificant attribute parameter estimates. Therefore, it is imperative to test whether fast and slow responses distort estimates when included in the whole-sample dataset. To do this, we estimate two utility functions, one without fast responses and the other without slow responses. Results from these two models are compared to the estimation results based on the whole-sample data. Utility function results for these three models are presented in Table 5.6 below.

Table 5.6: Utility functions for the whole sample, samples without fast and slow responses

	Whole sample		**Without fast responses**		**Without slow responses**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Random parameters in utility functions						
B_KITCHEN	0.164***	0.062	0.187***	0.054	0.199***	0.067
B_SHOWER	0.095	0.121	0.183**	0.090	0.052	0.123
B_TOILET	0.018	0.060	0.052	0.046	0.009	0.067
B_GARDEN	-0.002	0.071	0.021	0.052	-0.027	0.072
B_BILL	-0.002***	0.001	-0.003***	0.001	-0.002***	0.001
Nonrandom parameters in utility functions						
ASC	0.265	0.724	0.0	0.208	0.0	0.753
Diagonal values in Cholesky matrix, L.						
NsB_KITCHEN	0.259**	0.131	0.272**	0.109	0.234*	0.123
NsB_SHOWER	0.386	0.507	0.345	0.216	0.411	0.347
NsB_TOILET	0.017	0.703	0.003	0.067	0.042	0.133
NsB_GARDEN	0.254	0.290	0.157	0.169	0.044	2.196
NsB_BILL	0.001	0.008	0.001	0.001	0.0001	0.009
Below diagonal values in L matrix. V = L*Lt						
B_SHO:B_KIT	0.630**	0.291	0.675***	0.160	0.584**	0.256
B_TOI:B_KIT	0.031	0.116	0.029	0.079	0.022	0.111
B_TOI:B_SHO	0.056	0.181	0.035	0.076	0.041	0.138
B_GAR:B_KIT	0.028	0.151	0.055	0.132	0.032	0.164
B_GAR:B_SHO	0.092	0.229	0.096	0.170	0.096	0.199
B_GAR:B_TOI	-0.038	1.442	-0.238*	0.142	-0.266	0.434
B_BIL:B_KIT	-0.002	0.003	0.003***	0.001	-0.003	0.002
B_BIL:B_SHO	0.002	0.004	0.003**	0.001	0.002	0.003
B_BIL:B_TOI	-0.495	0.012	0.001	0.001	0.002	0.017
B_BIL:B_GAR	-0.002	0.004	-0.001	0.001	-0.001	0.017
Variance parameter tau in GMX scale parameter						
TauScale	0.397	1.374	0.021	0.116	0.004	3.599
Weighting parameter gamma in GMX model						
GammaMXL	0.707	3.126	0.100	Fix. Par	0.100	Fix. Par
Sample Mean Sample Std.Dev.						
Sigma(i)	0.992**	0.404	0.998***	0.045	0.998***	0.043
Standard deviations of parameter distributions						
sdB_KITCHEN	0.259**	0.131	0.272**	0.109	0.234*	0.123
sdB_SHOWER	0.739	0.490	0.758***	0.212	0.715*	0.383
sdB_TOILET	0.066	0.260	0.045	0.085	0.063	0.144
sdB_GARDEN	0.274	0.444	0.306	0.197	0.251	0.753
sdB_BILL	0.004	0.006	0.004***	0.001	0.004	0.005
N	1839		1300		1665	
LL	-1790.9		-1222.6		-1625.2	

AIC	3627.7	2489	3413.7
BIC	3754.6	2602.9	2107.6
Pseudo R^2	0.1	0.2	0.1

The table shows that estimations based on the ‘whole sample’ dataset resulted in two statistically significant attribute parameters (KITCHEN and BILL). The same results are noted in the model where ‘slow responses’ were removed from the dataset. However, the model based on data that excludes ‘fast responses’ reported three statistically significant parameter estimates (KITCHEN, SHOWER and BILL). The only difference noted between the model ‘without fast responses’ and the other models is in the statistical significance of the SHOWER parameter. All the other estimates reported the same signs (except for GARDEN in the model ‘without fast responses’), statistical significance, and sizes of coefficients. Despite the minor differences observed, we argue that removing either the fast responses or the slow responses from the sample did not significantly affect the estimation results in terms of the statistical significance, sign and magnitude of attribute parameter estimates[36].

In addition to the utility functions, we also compare the standard deviations of the parameter distributions reported in each model. The standard deviations of parameter distributions show the dispersion that exists around the sample population. While random parameter estimates show a preference for the population mean, it is important to note that there might be some dispersion around the estimated population mean. This is shown by the standard deviations of random parameter distributions. Where estimates are significant, it implies the existence of heterogeneity in preferences, suggesting that different individuals have individual-specific parameter estimates that may be different from the population mean parameter estimate (Hensher et al., 2015).

Results from the table show that in the ‘whole sample’ model, only KITCHEN had a statistically significant estimate for the standard deviations of parameter distributions. However, the model ‘without fast responses’ had three significant random parameters (KITCHEN, SHOWER and BILL), while the model ‘without slow responses’ had two (KITCHEN and SHOWER), each with a 10% level of significance. These results suggest that

[36] However, we noted that the model without fast responses had lower confidence intervals for each attribute parameter than the model without slow responses. Confidence intervals reported for the ‘whole sample’ were larger than those reported when fast responses were removed from the dataset, but smaller than those reported when slow responses were removed from the dataset. This implies that including fast responses in the dataset increased the confidence intervals.

the ‘whole sample’ model performed better than the other two models in capturing the preferences of the population in the mean parameter estimates. It is evident that estimation based on data ‘without fast responses’ gave population mean estimates that did not capture the preferences of most respondents in the sample. This implies that the model had individual-specific estimates that were different to the population mean. Based on this, we argue that removing fast and slow responses from the dataset did not improve our results. This argument is consistent with Börger (2016) and Campbell et al. (2017), where it is also suggested that data generated from fast responses should not be thrown away.

5.7.3. Estimation of MWTP

We also test for variations in the measures of social welfare in each response-time model. To do this, we estimate MWTP figures, which show the average estimates that households are prepared to pay if they adopt each of the given technological devices. Positive MWTP estimates show the average amount that households are willing to pay, whereas negative estimates show what households are willing to accept as compensation for changes in the attribute. As we did in the sub-section on utility functions, we present two sets of MWTP estimates. The first set compares MWTP estimates generated from average responses, fast responses, and slow responses. These three are also compared to estimates for the whole sample. Subsequently, we compare MWTP estimates for the whole sample, the sample without fast responses, and the sample without slow responses. In presenting the MWTP estimates, we follow Börger (2016) and include the confidence interval for each estimate. Table 5.7 presents the first set of MWTP estimates (figures are in US dollars).

Table 5.7: MWTP estimates for the whole sample, average, fast and slow responses

	Whole sample		Average responses		Fast responses		Slow responses	
	Estimate	95% Conf. Int	Estimate	95% Conf. Int	Estimate	95% Conf. Int	Estimate	95% Conf. Int
KITCHEN	4.78** (2.07)	0.72 to 8.84	5.36*** (2.05)	1.34 to 9.38	7.70 (17.14)	-25.89 to 41.29	7.09 (51.99)	-108.99 to 94.80
SHOWER	2.76 (3.35)	-3.81 to 9.33	3.09 (3.43)	-3.64 to 9.81	-7.70 (27.10)	-60.82 to 45.42	17.59 (134.00)	-205.83 to 241.02
TOILET	0.53 (1.77)	-2.93 to 3.99	1.30 (1.77)	-1.96 to 4.55	-5.25 (11.26)	-27.31 to 16.81	2.22 (18.79)	-34.60 to 39.05
GARDEN	-0.07 (2.08)	-4.14 to 4.00	-0.18 (1.81)	-3.72 to 3.36	-4.99 (11.92)	-28.35 to 18.37	5.42 (48.18)	-89.00 to 99.85
Wald Stat	**0.72**		**0.84**		**0.04**		**0.001**	
Prob. from Chi^2	**0.036**		**0.016**		**0.954**		**0.999**	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Standard errors in parentheses.

The table shows that only the MWTP estimates for KITCHEN devices in the 'whole sample' model and the 'average responses' model are statistically significant. These are interpreted to mean that respondents are willing to pay $4.78 in the 'whole sample' model and $5.36 in the 'average responses' model for improvements in KITCHEN devices. Estimates for all attributes are statistically insignificant for fast and slow responses. Overall, except for KITCHEN devices in the whole sample and average responses models, all estimates are insignificant for all models. However, we note that the confidence intervals are larger in the fast- and slow-responses models than in the whole sample and average response models. We also observe that the slow-responses model reported very large standard errors. These observations may be due to the sample sizes of the fast and slow responses, which were relatively small compared to the average responses.

Therefore, it is imperative to examine whether fast and slow responses affect utility estimates when included in the sample. To do this, we estimate and compare MWTP figures using data without fast responses as well as data without slow responses. Results are also compared to MWTP figures estimated using the whole-sample data. These estimation results are presented in Table 5.8 (figures are in US dollars).

Table 5.8: MWTP for the whole sample, sample without fast and slow responses

	Whole sample		Without fast responses		Without slow responses	
	Estimate	95% Conf. Int	Estimate	95% Conf. Int	Estimate	95% Conf. Int
KITCHEN	4.78** (2.07)	0.72 to 8.84	4.55*** (1.50)	1.60 to 7.50	5.90** (2.52)	0.97 to 10.84
SHOWER	2.76 (3.35)	-3.81 to 9.33	4.43** (2.11)	0.30 to 8.56	1.53 (3.63)	-5.58 to 8.64
TOILET	0.53 (1.77)	-2.93 to 3.99	1.27 (1.20)	-1.07 to 3.61	0.26 (2.00)	-3.66 to 4.18
GARDEN	-0.07 (2.08)	-4.14 to 4.00	0.50 (1.27)	-2.00 to 2.99	-0.79 (2.11)	-4.94 to 3.35
Wald Statistic	**0.72**		**1.71**		**0.55**	
Prob. from Chi^2	**0.036**		**0.000**		**0.095**	

Note: ***, ** and * = significance at 1%, 5%, 10% level, respectively. Standard errors in parentheses.

The table shows that except for the estimate for SHOWER devices, which is statistically significant in the 'without fast responses' model, all MWTP estimates are largely similar across the three models in terms of significance, sign and magnitude. These results agree with findings in the literature, where responses times are found to have no meaningful impact on the willingness-to-pay estimates (see Börger, 2016; Campbell et al., 2017; Dellaert et al., 2012; Konovalov and Krajbich, 2017). However, although no major differences are noted across the three models, we note that excluding fast responses from the dataset gives more statistically significant estimates than when they are included. To an extent, this is consistent with the suggestion in Haaijer et al. (2000) that fast responses should be accounted for in estimation, because otherwise wrong inferences will be deduced.

Furthermore, we observe from the table that the model that excludes fast responses reported lower confidence intervals than the other two models. The model that excludes slow responses reported the largest confidence intervals. This implies that including fast responses in the dataset has an implication on the confidence intervals (i.e. increases the intervals). This observation is consistent with findings in Börger (2016) and Campbell et al. (2017), which also reported that confidence intervals decrease as response time increases.

5.8. Conclusion

This paper investigates the effects of response time on respondent choices in a self-administered face-to-face survey environment. The focus of the paper was to establish whether utility functions and MWTP estimates are affected by the time respondents take to complete a survey. To achieve this, we used data for 307 household heads from an experiment on household preferences for water-efficient technologies in Gauteng province, South Africa. The study follows Campbell et al. (2017), and divides the dataset into average responses, fast responses and slow responses.

Prior to the analysis of stated-preference data, we join studies such as Bonsall and Lythgoe (2009), Börger (2016) and Recalde et al. (2014) in examining the determinants of response time. We find that only marital status and age were significant determinants of response time, while gender, race, education, and income were statistically insignificant. This result is consistent with Vista et al. (2009), in which biographical characteristics were found to have no effect on response time. We also found that respondents who completed the survey on their own took longer than those who asked the enumerator to tick boxes for them. Although several interpretations can be deduced, this implies that enumerator involvement made the experiment less complex. This addresses the problem of choice task complexity, which is a common determinant of response time (see Dellaert et al., 2012; DeShazo and Fermo, 2002; Konovalov and Krajbich, 2017). We argue that where several enumerators are used, there is a need to control for interviewer bias.

To test the impact of response time on utility functions, we use the GMXL model as an estimation tool. Estimation is done in two stages. The first stage estimates utility functions using data for average responses, fast responses, and slow responses. The second stage estimates utility functions using data for the whole sample, data without fast responses, and data without slow responses. Findings in each stage were compared, mainly in terms of attribute parameter estimates.

In the first stage, we found that results reported for 'average responses' were similar to those reported for the 'whole sample' in terms of statistical significance, sign and magnitude. However, we found that 'fast responses' and 'slow responses' generated statistically insignificant parameters for all attributes. Thus, we argue, results for the whole sample were mostly determined by 'average response' data, implying that fast and slow responses did not affect estimates for attribute parameters. This could have been prompted by the sample sizes of the fast and slow responses, which were relatively small compared to the dataset for average responses. Therefore, it was possible for average responses to drive results in the whole sample. Nevertheless, we found that parameter estimates for the whole sample were not affected by the inclusion of fast and slow responses; thus, we join Börger (2016) in arguing that fast responses should not be thrown away.

Results from the second stage show that the whole sample reported two statistically significant attribute parameters. The same results were observed when slow respondents were removed from the dataset. However, when fast responses were removed from the dataset, three attribute parameters emerged as statistically significant. Two of the three statistically significant parameter estimates were for the same attributes that also reported statistical significance in the 'whole sample' model, as well as the model 'without slow responses'. Overall, we found that parameter estimates for all three sets of data in this stage of our analysis generated similar results in terms of sign, significance and magnitude of coefficients. Thus, we argue that removing fast responses or slow responses from the sample did not significantly affect results.

In addition to the utility functions, we test whether response times affect MWTP estimates. We follow the same approach used in the estimation of utility functions and estimate MWTP figures in two stages.

The first stage estimated MWTP figures for average responses, fast responses, and slow responses, while the second stage estimated MWTP figures for the whole sample, the sample excluding fast responses, and the sample excluding slow responses. Results from the first stage showed that all MWTP figures in the fast responses model and the slow responses model were statistically insignificant. On the other hand, only one attribute had statistically significant estimates in the whole sample and average response models. Regardless of this difference, there were no other major differences in the MWTP for all models in terms of statistical significance. Findings from our tests agree with those in studies such as Börger (2016), Campbell et al. (2017), Dellaert et al. (2012), and Konovalov and Krajbich (2017), which also found response time to have no meaningful impact on willingness-to-pay estimates.

Results from the second stage showed that in general, all MWTP figures were largely similar across the models in terms of statistical significance, sign and magnitude. The only difference noted across the models was that while other models had one statistically significant attribute each, the estimation based on data excluding fast responses had two significant attributes. Furthermore, we found that the model that excluded fast responses reported lower confidence intervals than the other two models.

Overall, we argue that including fast and slow responses in stated-preference data did not affect results in terms of sign, significance and magnitude of parameter estimates. The first limitation of our study is that it uses the time taken by each respondent to complete the whole survey as proxy for response time. For future environmental economics studies using self-administered face-to-face surveys, we recommend that response times be captured only for the choice experiment section, or for each choice task. This is common in online studies and can give the exact time respondents took in answering the choice experiment section of the questionnaire. The main challenge associated with our approach is that there is a chance that some respondents may take less time in the choice experiment section but take longer in other sections of the questionnaire. Such respondents will be considered slow, yet they are fast based on the time taken to complete the section on choice experiment, which is key in stated-choice analysis. The second limitation of our study is that sample sizes were relatively low for fast and slow responses. This implies that caution should be employed when using our results. We recommend that future studies use larger samples and ensure that the sub-sample sizes for fast and slow responses are relatively large, for more robust results.

List of references

Anderson, J. R., Bothell, D., Lebiere, C. and Matessa, M. 1998. An integrated theory of list memory. *Journal of Memory and Language,* 38**,** 341-380.

Arad, A. and Rubinstein, A. 2012. Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game. *Journal of Economic Behavior and Organization,* 84**,** 571-585.

Ben-Akiva, M. E. and Lerman, S. R. 1985. *Discrete choice analysis: theory and application to travel demand*, MIT press.

Beshears, J., Choi, J. J., Laibson, D. and Madrian, B. C. 2008. How are preferences revealed? *Journal of Public Economics,* 92**,** 1787-1794.

Bliemer, M. C., Rose, J. M. and Chorus, C. G. 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. *Transportation Research Part B: Methodological,* 102**,** 83-104.

Bliemer, M. C., Rose, J. M. and Hess, S. 2008. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling,* 1**,** 98-126.

Bonsall, P. and Lythgoe, B. 2009. Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments. *Journal of Choice Modelling,* 2**,** 216-236.

Börger, T. 2016. Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. *Environmental and Resource Economics,* 65**,** 389-413.

Campbell, D., Morkbak, M. and Olsen, S. 2013. How quick can you click? The role of response time in online stated choice experiments. *Bioecon Conference 2013.* Cambridge.

Campbell, D., Mørkbak, M. R. and Olsen, S. B. 2017. Response time in online stated choice experiments: the non-triviality of identifying fast and slow respondents. *Journal of Environmental Economics and Policy,* 6**,** 17-35.

Campbell, D., Mørkbak, M. R. and Olsen, S. B. 2018. The link between response time and preference, variance and processing heterogeneity in stated choice experiments. *Journal of Environmental Economics and Management,* 88**,** 18-34.

Chen, F. and Fischbacher, U. 2015. Cognitive processes of distributional preferences: A response time study. Working Paper.

Conrad, F., Tourangeau, R., Couper, M. and Zhang, C. Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, 2017. 45-61.

Cook, J., Jeuland, M., Maskery, B. and Whittington, D. 2012. Giving stated preference respondents “time to think”: results from four countries. *Environmental and Resource Economics,* 51**,** 473-496.

Cook, J., Whittington, D., Canh, D. G., Johnson, F. and Nyamete, A. 2007. Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. *Economic Inquiry,* 45**,** 100-114.

Czajkowski, M., Hanley, N. and Lariviere, J. 2014. The effects of experience on preferences: theory and empirics for environmental public goods. *American Journal of Agricultural Economics,* 97**,** 333-351.

Dellaert, B. G., Donkers, B. and Soest, A. V. 2012. Complexity effects in choice experiment–based models. *Journal of Marketing Research,* 49**,** 424-434.

Department of Water and Sanitation 2017. Benchmarking of Water Loss, Water Use Efficiency and Non-Revenue Water in South African Municipalities (2004/05 to 2015/16).

Deshazo, J. and Fermo, G. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and management,* 44**,** 123-143.

Di Guida, S. and Devetag, G. 2013. Feature-based choice and similarity perception in normal-form games: An experimental study. *Games,* 4**,** 776-794.

Downes-Le Guin, T., Baker, R., Mechling, J. and Ruyle, E. 2012. Myths and realities of respondent engagement in online surveys. *International Journal of Market Research,* 54**,** 1-21.

Dror, M. and Hartman, B. C. 1999. Stopping rules for utility functions and the St. Petersburg gamble. *Applied Mathematics and Computation,* 98**,** 279-291.

Fiebig, D. G., Keane, M. P., Louviere, J. and Wasi, N. 2010. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science,* 29**,** 393-421.

Fleming, C. M. and Bowden, M. 2009. Web-based surveys as an alternative to traditional mail methods. *Journal of Environmental Management,* 90**,** 284-292.

Gill, D. and Prowse, V. L. 2017. Using response times to measure strategic complexity and the value of thinking in games. [Online] SSRN. Available: https://dx.doi.org/10.2139/ssrn.2902411 [Accessed 2 October 2018].

Greene, W. H. 2003. *Econometric analysis*, Pearson Education India.

Haaijer, R., Kamakura, W. and Wedel, M. 2000. Response latencies in the analysis of conjoint choice experiments. *Journal of Marketing Research,* 37**,** 376-382.

Hensher, D. A., Rose, J. M. and Greene, W. H. 2015. *Applied Choice Analysis,* Cambridge University Press.

Holmes, T., Alger, K., Zinkhan, C. and Mercer, D. E. 1998. The effect of response time on conjoint analysis estimates of rainforest protection values. *Journal of Forestry Economics, 4*.

Huang, J. L., Liu, M. and Bowling, N. A. 2015. Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology,* 100**,** 828.

Jansen, A. and Schulz, C. E. 2006. Water demand and the urban poor: A study of the factors influencing water consumption among housholds in Cape town, South Africa. *South African Journal of Economics,* 74**,** 593-609.

Jones, M. P. and Hunt, W. F. 2010. Performance of rainwater harvesting systems in the southeastern United States. *Resources, Conservation and Recycling,* 54**,** 623-629.

Kessler, J. B. and Meier, S. 2014. Learning from (failed) replications: Cognitive load manipulations and charitable giving. *Journal of Economic Behavior and Organization,* 102**,** 10-13.

Kocher, M. G., Martinsson, P., Myrseth, K. O. R. and Wollbrant, C. E. 2017. Strong, bold, and kind: Self-control and cooperation in social dilemmas. *Experimental Economics,* 20**,** 44-69.

Konovalov, A. and Krajbich, I. 2017. Revealed indifference: Using response times to infer preferences. Working Paper.

Krosnick, J. A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology,* 5**,** 213-236.

Kuo, W.-J., Sjöström, T., Chen, Y.-P., Wang, Y.-H. and Huang, C.-Y. 2009. Intuition and deliberation: two systems for strategizing in the brain. *Science,* 324**,** 519-522.

Lanz, B. and Provins, A. 2015. Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics,* 47**,** 300-324.

Lenzner, T., Kaczmirek, L. and Lenzner, A. 2010. Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology,* 24**,** 1003-1020.

Lindhjem, H. and Navrud, S. 2011. Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics,* 70**,** 1628-1637.

Lohse, J., Goeschl, T. and Diederich, J. H. 2017. Giving is a Question of Time: Response Times and Contributions to an Environmental Public Good. *Environmental and Resource Economics,* 67**,** 455-477.

Makki, A. A., Stewart, R. A., Panuwatwanich, K. and Beal, C. 2013. Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies. *Journal of Cleaner Production,* 60**,** 129-146.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour. *in Frontiers in Econometrics, ed.* P. Zarembka. New York: Academic Press, 105-142.

Meyerhoff, J. and Liebe, U. 2009. Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity. *Land Economics,* 85**,** 515-528.

Mini, C., Hogue, T. and Pincetl, S. 2015. The effectiveness of water conservation measures on summer residential water use in Los Angeles, California. *Resources, Conservation and Recycling,* 94**,** 136-145.

Nielsen, U. H., Tyran, J.-R. and Wengström, E. 2014. Second thoughts on free riding. *Economics Letters,* 122**,** 136-139.

Nkosi, N. P. and Dikgang, J. 2018. Pricing electricity blackouts among South African households. *Journal of Commodity Markets,* 11**,** 37-47.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J. and Yazdizadeh, A. 2017. Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning,* 157**,** 63-74.

Price, G. 2009. *Water Conservation Guideline. eThekwini Municipality: eThekwini Municipality*. [Online]. Ethekwini Municipality. Available: http://www.durban.gov.za/City_Services/development_planning_management/environmental_planning_climate_protection/Publications/Documents/GG_Water_Guide.pdf [Accessed 5 September 2018].

Rand, D. G., Greene, J. D. and Nowak, M. A. 2012. Spontaneous giving and calculated greed. *Nature,* 489**,** 427.

Ratcliff, R. 1978. A theory of memory retrieval. *Psychological Review,* 85**,** 59.

Recalde, M. P., Riedl, A. and Vesterlund, L. 2014. Error prone inference from response time: The case of intuitive generosity.Working Paper.

Rieskamp, J. and Hoffrage, U. 2008. Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica,* 127**,** 258-276.

Rieskamp, J. and Otto, P. E. 2006. SSL: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General,* 135**,** 207.

Rijsberman, F. R. 2006. Water scarcity: fact or fiction? *Agricultural Water Management,* 80**,** 5-22.

Rose, J. M. and Black, I. R. 2004. Response time influences on stated preference choice experiments. *33rd European Marketing Academy Conference*, Mucia, Spain.

Rose, J. M. and Black, I. R. 2006. Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters,* 17**,** 295-310.

Rubenstein, A. 2013. Response time and decision making: An experimental study. *Judgment and Decision Making,* 8**,** 540.

Rubinstein, A. 2007. Instinctive and cognitive reasoning: A study of response times. *The Economic Journal,* 117**,** 1243-1259.

Rubinstein, A. 2016. A typology of players: Between instinctive and contemplative. *The Quarterly Journal of Economics,* 131**,** 859-890.

Savage, S. J. and Waldman, D. M. 2008. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics,* 23**,** 351-371.

Schooler, L. J. and Anderson, J. R. 1997. The role of process in the rational analysis of memory. *Cognitive Psychology,* 32**,** 219-250.

Schwappach, D. and Strasmann, T. 2006. 'Quick and Dirty Numbers'? The Reliability of a Stated-Preference Technique for the Measurement of Preferences for Resource Allocation. *Journal of Health Economics* 25**,** 432-448.

Spiliopoulos, L. and Ortmann, A. 2014. The BCD of response time analysis in experimental economics. *Experimental Economics***,** 1-51.

Statistics South Africa 2017. The state of basic service delivery in South Africa: In-depth analysis of the Community Survey 2016 data. Pretoria: Statistics South Africa.

Statistics South Africa 2018. Mid-year population estimates. Statistical release PO302. Pretoria: Statistics SA.

Still, D. and Bhagwan, J. 2008. The Status and Use of Potable Water Conservation and Savings Devices in the Domestic and Commercial Environments in South Africa. *Water Distribution Systems Analysis 2008.*

Stupple, E. J., Pitchford, M., Ball, L. J., Hunt, T. E. and Steel, R. 2017. Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test. *PloS One,* 12**,** e0186404.

Svedsater, H. 2007. Ambivalent statements in contingent valuation studies: inclusive response formats and giving respondents time to think. *Australian Journal of Agricultural and Resource Economics,* 51**,** 91-107.

Vista, A. B., Rosenberger, R. S. and Collins, A. R. 2009. If you provide it, will they read it? Response time effects in a choice experiment. *Canadian Journal of Agricultural Economics/Revue Canadienne D'agroeconomie,* 57**,** 365-377.

Vloerbergh, I., Fife-Schaw, C., Kelay, T., Chenoweth, J., Morrison, G. and Lundéhn, C. 2007. Assessing consumer preferences for drinking water services-Methods for water utilities. *Techneau Project Report D,* 6.

Whittington, D., Smith, V. K., Okorafor, A., Okore, A., Liu, J. L. and Mcphail, A. 1992. Giving respondents time to think in contingent valuation studies: a developing country application. *Journal of Environmental Economics and Management,* 22**,** 205-225.

Wilcox, N. T. 1993. Lottery choice: Incentives, complexity and decision time. *The Economic Journal,* 103**,** 1397-1417.

Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R. and Mousavinejad, A. 2013. End use water consumption in households: impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production,* 60**,** 107-115.

Wood, D., Harms, P., Lowman, G. H. and Desimone, J. A. 2017. Response Speed and Response Consistency as Mutually Validating Indicators of Data Quality in Online Samples. *Social Psychological and Personality Science***,** 1948550617703168.

Yang, H., Reichert, P., Abbaspour, K. C. and Zehnder, A. J. 2003. A water resources threshold and its implications for food security. ACS Publications.

Appendix 5.1: The questionnaire used in the survey

NELSON MANDELA
UNIVERSITY

HOUSEHOLDS' INTENTIONS TO ADOPT WATER-SAVING TECHNOLOGY IN JOHANNESBURG

Time interview began ____:____

Date of the interview: ____/___/_____

Name of interviewer: ____________________________________

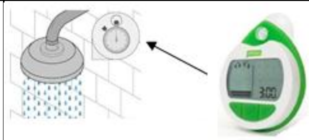
Area study is taking place: _______________________________

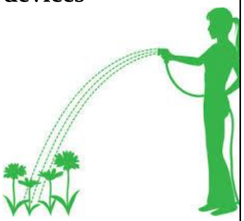
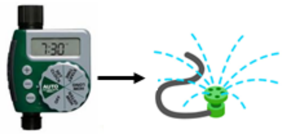
As is the case in the rest of South Africa, Johannesburg is facing water shortages; yet little is known about households' water-conservation efforts. Households are encouraged to install water-saving devices as part of addressing the water shortage. This can only be achieved if households are aware of water-saving options and the cost-savings benefits. We employ choice experiments to evaluate the intention of households to adopt water-saving technologies.

The survey has three sections. Section A provides choice experiments by which households' intention to adopt water-saving devices is evaluated. Section B provides general questions on households' water-consumption behaviour. Section C collects the biographical information of the respondents.

SECTION A: CHOICE EXPERIMENT

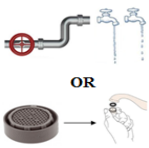
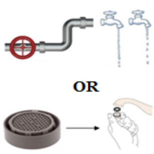
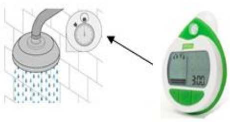
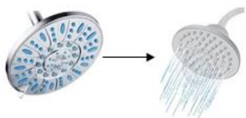
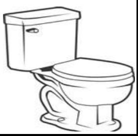
Table 1: Attributes and levels used in the study

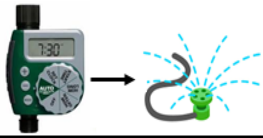
Attribute	Description	Attribute Levels	
Kitchen devices 	A typical household uses 11% of its water in the kitchen. A standard tap flows at about 8l per minute. Installing water-flow regulators or tap-head aerators makes a standard tap more efficient and saves water by 60%. An efficient dishwasher uses 15l per cycle, using 50% less water than is used in a conventional dishwasher.	**Level 1:** Efficient dishwasher	
		Level 2: Efficient tap	
		Level 3: System collecting used water	
Shower devices 	A typical household uses 24% of its water in the shower. Shower timers result in shorter showers. Efficient showerheads save 65% of water used in the shower.	**Level 1:** Efficient showerhead	
		Level 2: Shower timer	
Toilet devices 	A typical household uses 25% of its water for flushing the toilet. Replacing a 12l cistern with a 3l dual cistern uses about 75% less water. An interruptible-flush cistern allows users to control how long the toilet flushes. Hippo bags displace water in the cistern and save about 1.2l per flush.	**Level 1:** Dual-flush cistern sized 3-6l	
		Level 2: Interruptible-flush cistern	
		Level 3: Cistern displacement (hippo bag)	

Garden & Outdoor devices 	A typical household uses 25% of its water in the garden or for outdoor activities. Efficient gardening technologies reduce water use by 30%. These include time-based irrigation control systems, and micro-drip systems. Irrigating gardens using water collected with water tanks also saves water.	**Level 1:** Time-based irrigation controller	
		Level 2: Micro-drip systems	
		Level 3: Use harvested rain water	
Monthly water bill 	The average water bill for a household is R450 per month. Installing water-efficient technologies will reduce the monthly water bill by 30%, 50% or 75%.	**Level 1:** R110 **Level 2:** R225 **Level 3:** R315	

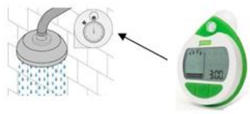
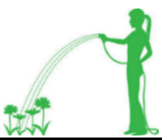
Six choice sets with three alternatives (Status Quo, Option 1 and Option 2) are generated. The Status Quo is undefined, as only you know your current situation. We would like to know which option you prefer the most. Please treat each choice set independently.

CHOICE SET 1

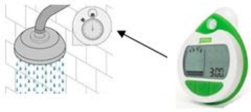
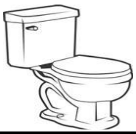
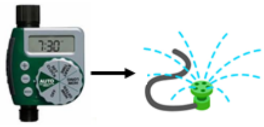
	Status quo	Option 1	Option 2
Kitchen devices 		Efficient tap 	Efficient tap 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Dual-flush cistern 	Hippo bag 

Garden & outdoor devices 		Use harvested rain water 	Time-based irrigation controller 
Monthly water bill 	R450	R315	R110
YOUR CHOICE			

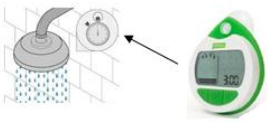
CHOICE SET 2

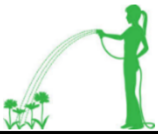
	Status quo	**Option 1**	**Option 2**
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Efficient showerhead 	Shower timer 
Toilet devices 		Dual-flush cistern 	Hippo bag 
Garden & outdoor devices 		Micro-drip irrigation system 	Micro-drip irrigation system 
Monthly water bill 	R450	R110	R315
YOUR CHOICE			

CHOICE SET 3

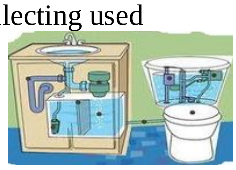
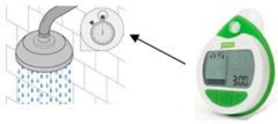
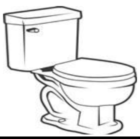
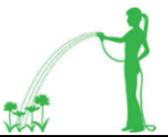
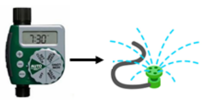
	Status quo	Option 1	Option 2
Kitchen devices 		Efficient tap 	Efficient tap 
Shower devices 		Efficient showerhead 	Shower timer 
Toilet devices 		Hippo bag 	Dual-flush cistern 
Garden & outdoor devices 		Use harvested rain water 	Time-based irrigation controller 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

CHOICE SET 4

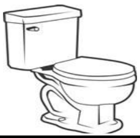
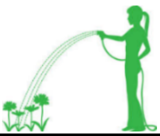
	Status quo	Option 1	Option 2
Kitchen devices 		System collecting used water 	Efficient dishwasher 
Shower devices 		Efficient showerhead 	Shower timer 

Toilet devices 		Interruptible-flush cistern 	Interruptible-flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rainwater 
Monthly water bill 	R450	R315	R110
YOUR CHOICE			

CHOICE SET 5

	Status quo	**Option 1**	**Option 2**
Kitchen devices 		Efficient dishwasher 	System collecting used water 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Hippo bag 	Dual flush cistern 
Garden & outdoor devices 		Time-based irrigation controller 	Use harvested rainwater 
Monthly water bill 	R450	R225	R225
YOUR CHOICE			

CHOICE SET 6

	Status quo	Option 1	Option 2
Kitchen devices 		System collecting used water 	Efficient dishwasher 
Shower devices 		Shower timer 	Efficient showerhead 
Toilet devices 		Interruptible-flush cistern 	Interruptible-flush cistern 
Garden & outdoor devices 		Micro-drip irrigation system 	Micro-drip irrigation system 
Monthly water bill 	R450	R110	R315
YOUR CHOICE			

SECTION B: WATER CONSUMPTION BEHAVIOUR AND TECHNOLOGY

1. When you made your choices, which attribute most influenced your decision?
Please tick in the box next to the attribute

Kitchen devices	
Shower devices	
Toilet devices	
Garden devices	
Water bill	

2. Do you have the following water technology at home?
Please select one answer per row

	Yes	No	Not applicable	Not sure
Water-collection tank (Jojo tank)				
Cistern displacement device (‘hippo bag’)				
Water-flow regulators				

Efficient showerheads				
Efficient bathtub				
Efficient toilet cistern sized 3-6 litres				
Interruptible-flush (multi-flush) cistern				
Dishwasher				
Efficient garden devices				

3. **If any of your answers in QUESTION 2 above was NO, what is your main reason?**
   Please select one reason you think is the main reason

I cannot afford them	
I did not know about them	
I have no infrastructure to connect them	
They are not important	
Other *(Please specify):*	

4. **How often do you do the following in your daily life?**
   Please select one answer per row

	Never	Occasionally	Always	Not applicable
Take bath instead of shower				
Take showers longer than 5 minutes				
Run shower for some time, waiting for hot water				
Keep the tap running when brushing teeth				
Ignore water leaks from the toilet tank				
Keep the tap running when washing dishes				
Rinse cutlery and glasses under running water				
Use running water to defrost frozen food				
Ignore a dripping tap				
Ignore kids wasting water				
Keep water running while washing face or hair				

SECTION C: PERSONAL INFORMATION

1. **How many people are in your household?**

2. **Do you have the following in your household?**
   Please select one answer per row

	Yes	No
Infant (0-2 years)		
Child (3-15 years)		

3. **If YES to QUESTION 2 above, how many infants/children do you have?**

4. **What is your gender?**

☐ Male ☐ Female

5. **Which racial group do you belong to?** *(Optional)*:

☐ Black/African ☐ White ☐ Indian/Asian ☐ Coloured

6. **What is your marital status?**

☐ Single ☐ Married ☐ Other (*Please specify*): _____________

7. **What is your highest education level?**

☐ Never attended school ☐ Primary school ☐ High school

☐ Certificate ☐ Diploma/Degree ☐ Postgraduate

8. **What is your year of birth?**

☐

9. **What is your household's main source of income?**

☐ Salary/Wages ☐ Business ☐ Investments ☐ Grant/Pension/Allowance

10. **What is your household's monthly average income?**

< R5 000	
R5 000 – R10 000	
R10 000 – R20 000	
R20 000 – R40 000	
R40 000 – R60 000	
> R60 000	

11. **Did you answer this questionnaire on your own?** *(If you answered on your own without the interviewer ticking boxes for you, select YES):*

☐ Yes ☐ No

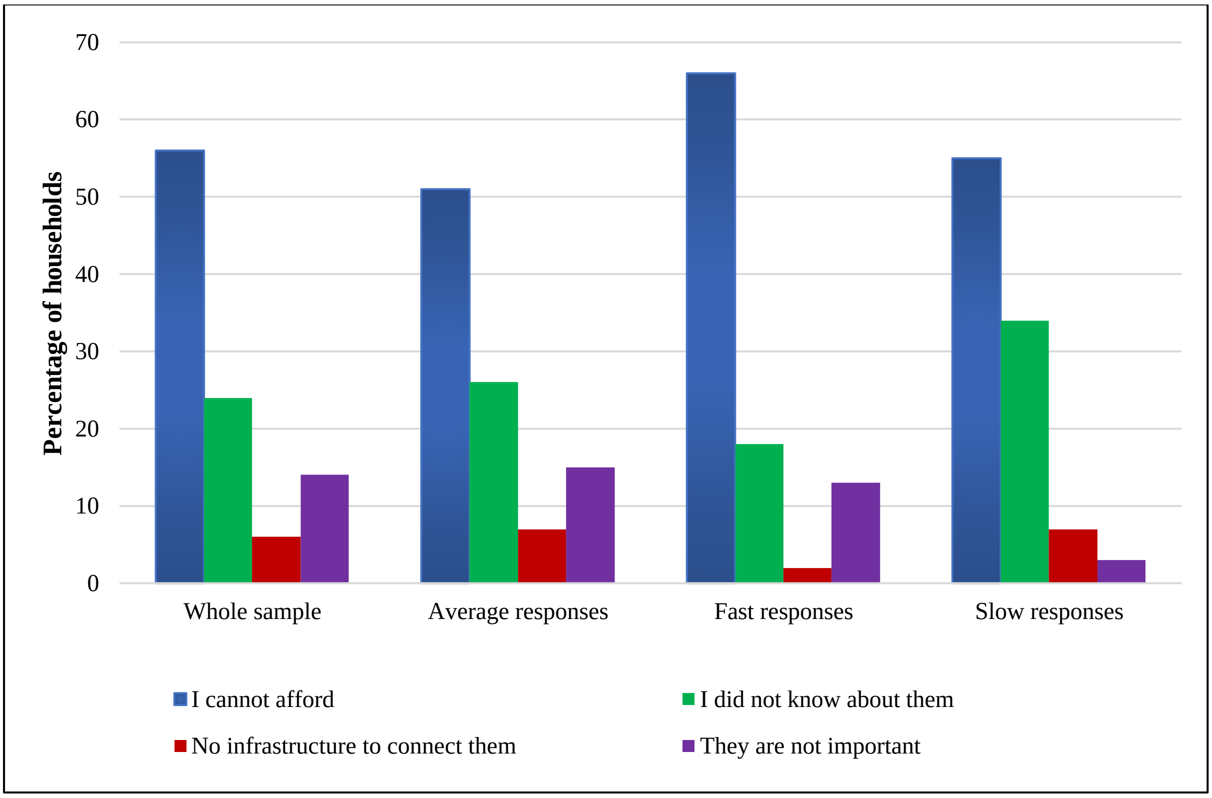
Time interview ended ______:______

Appendix 5.2: Current use of water-efficient devices

Table A1: Summary statistics of households' responses to having water-efficient technology

		Whole sample	Modal answer
	Respondents (N)	***307***	
9. Water collection tank (Jojo tank) (%)			
	Yes	8	No
	No	87	
	Not applicable	4	
	Not sure	1	
10. Cistern displacement device (Hippo bag) (%)			
	Yes	5	No
	No	75	
	Not applicable	19	
	Not sure	2	
11. Water-flow regulators (%)			
	Yes	17	No
	No	80	
	Not applicable	2	
	Not sure	1	
12. Efficient showerheads (%)			
	Yes	32	No
	No	64	
	Not applicable	3	
	Not sure	1	
13. Efficient toilet cistern sized 3-6 litres (%)			
	Yes	44	No
	No	55	
	Not applicable	1	
	Not sure	-	
14. Interruptible/multi-flush cistern (%)			
	Yes	88	Yes
	No	11	
	Not applicable	-	
	Not sure	1	
15. Dishwasher (%)			
	Yes	16	No
	No	83	
	Not applicable	-	
	Not sure	1	
16. Efficient garden devices (%)			
	Yes	12	No
	No	87	
	Not applicable	-	
	Not sure	1	

Appendix 5.3: Reasons for not having water-efficient technology



Chapter 6: Conclusion

6.1. Summary

Influencing water systems towards efficiency and sustainable consumption is increasingly gaining importance among policymakers and the general public. In most water sectors across the world, regulators predominantly use traditional economic-analysis methods for benchmarking water utilities and eliciting water-service preferences. Several techniques that extend these commonly used tools are increasingly being discussed in the literature. However, their practical application in the water sector remains relatively low. This study is intended to extend the existing literature by providing more robust methods that could be useful to water regulators. The study asked four research questions, to shed light on whether more robust methods are the way forward in water regulation. More precisely, the study investigated the consistency of efficiency scores obtained from DEA, SFA and StoNED on a sample of South African water utilities; as well as the impact of status quo bias, presentation format and response time on results from choice experiments conducted using a case of the South African water sector. These issues constitute the four objectives of this study.

To investigate the consistency of efficiency scores obtained from DEA, SFA and StoNED, the study uses cross-sectional data from 102 South African water utilities for the period 2013/14. Results showed that StoNED (based on the methods of moments estimator) controlled heterogeneity better than DEA and SFA, as it produced efficiency scores with lower standard deviations. However, we also found that while DEA reported the most variations, SFA performed better than StoNED in terms of the variation of scores around the sample mean, when the latter was based on the pseudo-likelihood estimator. Furthermore, we found that under StoNED, most utilities reported efficiency scores above the model's mean; while under DEA, most utilities reported scores below the model's mean. More precisely, 65% of utilities had efficiency scores above the model's mean of 0.681 under StoNED, while in SFA, 50% of the utilities reported efficiency scores above the model's mean of 0.662. Under DEA, only 37% of the utilities reported efficiency scores above the model's mean of 0.447. Additionally, for most of the utilities, efficiency scores estimated using StoNED moderated those resulting from DEA and SFA. Where DEA gave a higher efficiency score and SFA gave a lower efficiency score (or vice versa), StoNED usually gave a median score for the three.

The second objective of the study tested for the effects of reducing status quo bias in discrete choice experiments conducted in environmental economics. In pursuit of this objective, the study used a case of household preferences for water-service packages in South Africa. The study divided the sample into suburbs and townships, and presented each sub-sample with two different choice experiments. In each sub-sample, the first treatment presented respondents with a series of choice sets, each with a status quo option that resonated with them. The second treatment presented respondents with a series of choice sets, each with a status quo option that did not fully reflect their current situation. Estimation was done using generalised mixed logit models and we found that both attribute parameter and MWTP estimates across the two treatments in the township sub-sample were largely similar in terms of sign, statistical significance and the absolute value of their magnitude. This was also true in the suburban sub-sample, except that the MWTP estimates in the two treatments of this sub-sample reported disagreeing results. Thus, we argued that overall, including a partially relevant status quo reduced status quo bias, but did not affect empirical estimates (except for the MWTP estimates in the suburbs).

Furthermore, the study tested whether presenting attributes and levels as text, visuals, or text-and-visuals would generate differences in estimated utilities and willingness to pay. This was tested using discrete-choice experiments on South African households' preferences for water-efficient technologies. Estimation of stated-preference data was done using mixed logit models, and three main findings were reported. Firstly, we found that the visuals experiment had more statistically significant coefficients than both the text and the text-and-visuals experiments. We argued that including visuals in the choice profiles increased the number of attributes that were important to respondents. Secondly, we found that the text-and-visuals experiment was able to capture the true preferences of respondents better than the other experiments. This was observed in the standard deviations of random parameters, where the text-and-visuals experiment had fewer statistically significant random parameters than the other experiments. Finally, we found that MWTP estimates were largely different in terms of sign, significance and magnitude across the three presentation formats. Respondents were willing to pay for three attributes in the visuals experiment, whereas they were willing to pay for only one attribute in each of the other experiments. The text-and-visuals experiment was also found to have larger MWTP estimates than the other two experiments.

Finally, we tested for the impact of response time on empirical results generated in discrete-choice experiments conducted in environmental economics. To achieve this, we used a case of household preferences for water-efficient technologies in South Africa. Generalised mixed logit models were adopted, and the study first estimated utility functions using data for the whole sample, average, fast, and slow responses. Thereafter, we estimated utility functions using data without fast responses and data without slow responses. Overall, we found from the first estimation that fast and slow responses generated statistically insignificant parameters for all attributes, while average responses and the whole sample reported similar results. Thus, we argued that results for the whole sample were mostly determined by 'average response' data, implying that fast and slow responses did not affect estimates. In the second estimation, we found that all three datasets generated mostly similar results in terms of the sign, significance and magnitude of coefficients. Therefore, we argued that removing fast or slow responses from the sample did not significantly affect results. Regarding MWTP, we found that there were no major differences in estimates for all models in each of the two stages of estimation. Thus, we concluded that response time had no meaningful impact on willingness-to-pay estimates.

6.2. Future work

Based on the findings in this study, we make four main recommendations for future studies. In conducting efficiency analyses, our study did not report on the shape of the total cost function, focusing only on estimating utility-specific efficiency scores. Our study only used the half-normal distribution to estimate SFA, and an input-oriented variable returns to scale for DEA. We recommend that future studies also include other forms of SFA distribution (e.g. exponential and truncated) as well as other forms of returns to scale for DEA (i.e. constant, decreasing, and increasing returns to scale).

In testing for the impact of status quo bias, this study used different respondents with similar socio-economic characteristics to answer the two experiments presented in each sub-sample. Additionally, the inclusion of the 'none' option in our experiment was problematic in the second blocks of each sub-sample, where many respondents opted for the 'none' option. Therefore, we recommend against the inclusion of a 'none' option when conducting experiments such as ours. In this regard, we argue that omitting the 'none' option would motivate respondents to make real trade-offs, by having to choose between the hypothetical designed options and the status quo.

Furthermore, results on the impact of presentation formats showed that attributes presented visually took on more importance than those presented through text. However, we argue that care should be taken when presenting choice profiles as visuals, because preferences may be distorted by less important features such as the colour and form of the image. Thus, we argue in favour of presenting choice profiles using a combination of written text and visuals. This is because our study showed that the text-and-visuals experiment had considerable consistency with both the text and the visuals experiments, in terms of sign and significance of parameters. Our argument is that combining text and visuals improved clarity of attributes and/or levels to respondents, thereby yielding more robust empirical estimates. Thus, we recommend that future research on the effects of various presentation formats is required in environmental economics, so that guidelines on how to develop valid presentation formats for attribute levels in the choice tasks can be established.

Finally, the results on the impact of response time showed that including fast and slow responses did not affect empirical results; thus, we argued against the exclusion of fast and slow responses from the dataset. Similarities were observed in the empirical estimates reported when fast and slow responses were included and excluded in the estimation. However, since this study uses the time taken by each respondent to complete the whole survey as a proxy for response time, we recommend future environmental economics studies only capture the time taken to complete the choice experiment section, or each choice task. Additionally, we emphasise that our results should be interpreted with caution, since sample sizes were relatively low for fast and slow responses compared to those for average responses. Given this situation, the results reported for the whole sample could largely be determined by the average responses. Therefore, we recommend that future studies use larger samples, and ensure that the sub-sample sizes for fast and slow responses are relatively large, for more robust results.