UNIVERSITY OF THE WITWATERSRAND
JOHANNESBURG

School of Mechanical, Industrial and Aeronautical Engineering

Dissertation Title:

A Comparison of Non-Competitive Tender Prices in the South African Public Procurement Sector to Identify Pricing Deviations

Authored by:
Trinity Nkosi

Student Number:
707312

Supervised by:
(Supervisor) Ms Bontle Tladi
(Co-Supervisor) Dr Ellsworth Jonathan

A Dissertation Submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, In Fulfilment of the Requirements for the Degree of Master of Science in Engineering

Date of Submission: Monday, 20 September 2021

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

SCHOOL OF MECHANICAL, INDUSTRIAL & AERONAUTICAL ENGINEERING

Plagiarism Declaration

As Stated in The University Senate Plagiarism Policy

I ***Trinity Pholosho Nkosi,*** (Student number: ***707312)*** am a student registered for the degree of ***Master of Science in Industrial Engineering by Research*** in the academic year **2020.**

- I hereby declare the following:
- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included in Appendix A; Section 1 is the report from "Turnitin" (*or other approved plagiarism detection*) software indicating the level of plagiarism in my research document.

Signature: ______________________ Date: *20 September 2021*

Abstract

The South African Public Procurement system is riddled with corruption issues, including above-market price inflations of goods and services acquired. The reviewed literature revealed there are inadequate tools/indices readily available to flag such procurements individually. As such, the study aimed to remedy this shortfall by creating a cost comparison model through benchmarking municipal procurement prices to identify specific cases of pricing deviations.

This was achieved by using multiple linear regression to create a municipal market-value price prediction model for low valued, non-competitive *fencing supply and installation* procurements. It was assumed clean audited municipalities had a true market value for the topic contract to which the model training data was collected from seventeen clean audited municipalities in the Republic of South Africa.

The developed model was used to assess how contracts from five Gauteng non-clean audited municipalities would have been priced should they have been issued through a clean audited municipality and by what margin, if any, was the pricing deviation. Participation in the study was voluntary for the approached governmental entities, and all samples were collected based on the *quota sampling method*.

The cost comparison model was developed successfully and contrary to the expected outcomes had depicted that non-clean audited municipalities procured the topic contract at a lower price than clean audited municipalities. No correlation was found between the clean audit status of a municipality and the pricing of their non-competitive contracts. Furthermore, the model revealed it is possible to individually compare public procurements from different governmental organisations to identify pricing deviations.

***Keywords**: Multiple Linear Regression, Pricing Deviations, Price Inflations, Public Procurement, Benchmarking*

Dedication

I dedicate this dissertation to both my parents. My father, the late Malutha Nkosi whose unofficial *mantra* to his children growing up was always "*Hambani niyofunda*", which translates to "*go learn".* He instilled the culture of education both formal and informal in myself and my siblings, who too, has been instrumental on the journey to my Masters. My mother, Naume Nkosi, has been a source of inspiration and strength during moments of anguish and disheartenment. Her prayers, care and support, have propelled me to do better continually.

***Thank You, Mama noBaba*.**

Author
Trinity Nkosi

Acknowledgements

I would first like to thank my dissertation supervisor *Ms Bontle Tladi* from the School of Mechanical, Industrial and Aeronautical Engineering at the University of Witwatersrand. She allowed this paper to be my work but corrected me whenever she saw necessary, aiding me in narrowing my scope.

Second, I would like to thank my dissertation co-supervisor *Dr Ellsworth Jonathan*, a public sector procurement specialist, who planted the idea of me pursuing my Masters. He equipped me with the relevant tools needed to ease the development of this paper and provided constructive criticism from a qualitative perspective.

Third, I would like to thank *Dr Raeesa Ganey* from the School of Statistics and Actuarial Science at the University of Witwatersrand. She conducted a preliminary peer review of the Methodology and Findings Chapters. Thereafter, recommended crucial corrective measures.

Finally, I must express my very profound gratitude to all the *procurement personnel* from the different municipalities, approached during this study, for taking the time to provide me with the required data amidst the pandemic. This accomplishment would not have been possible without them.

Thank You All

Author
Trinity Nkosi

Table of Contents

List of Figures

List of Tables

List of Equations

List of Symbols

Symbol	Definition
Ps	Points scored for the comparative price of tender or offered under consideration.
Pt	Comparative price of bid or offer under consideration
$Pmin$	Comparative price of the lowest acceptable bid or offer
ε	Error term: a measure of the *noisy elements* in a regression function
β	Coefficients acquired for a regression equation using dataset samples
X	Independent variables (covariant or predictor); the scope of work variables
Y	Dependent variables; the predicted output variable, observed as pricing for study
n	The number of datasets in a sample for regression analysis (sample size)
m	The exponential value for a polynomial regression value determined by the modeller
Q	The desired output for a data envelopment analysis based on the Cobb-Douglas Model
A	The derived constant for a data envelopment analysis based on the Cobb-Douglas Model
L	The labour input a data envelopment analysis based on the Cobb-Douglas Model
K	The capital input a data envelopment analysis based on the Cobb-Douglas Model
a and b	The elastic outputs data envelopment analysis based on the Cobb-Douglas Model (labour and capital, respectively).
r	Pearson's Correlation Coefficient, which is a statistical tool that measures the linear correlation of a data set. It ranges from $-1 \leq r \leq 1$.
k	The total number of regressor per data point
p	The total number of regressor per data point plus one
SSE	Error sum of squares, a mathematical expression of squaring the deviations between the modelled outcome and the true outcome followed by the summation of all the squared deviations
SST	The total sum of squares, a mathematical expression of squaring the deviations between the mean outcome and the true outcome followed by the summation of all the squared deviations.
SSR	Regression sum of squares, a mathematical expression of squaring the deviations between the mean outcome and the modelled outcome followed by the summation of all the squared deviations
$\mathbf{R^2}$	The coefficient of multiple determination used to assess the fit of the model to the sample data given.
Adjusted $\mathbf{R^2}$	A changed version of $\mathbf{R^2}$ that was adjusted for the number of predictors in the model. Used to assess model accuracy in this study.
F-statistic	A general regression test result used to assess a hypothesis

Symbol	Definition
Critical **F-**Value	Used to assess whether the calculated F-statistic proves we can reject the null hypothesis by F-statistic being greater than it.
P-value	A measure of the probability that an observed difference could have occurred just by random chance.
MSE	Mean square error, a measure of the quality of an estimator. It is always non-negative, and values closer to zero are better.
k-folds	The number of partitions in a segmented data set for a k-folds cross-validation
α	The significance level is the probability of rejecting the null hypothesis when it is true
t or t_{stat}	**t** distribution, any member of a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the sample size is small and the population standard deviation
ROF	The independent variable for the removal of previous fencing structures
FMCE	The independent variable for the erection of a fence (study anchor predictor)
HGMCE	The independent variable for the installation of a hinged gate
SGMCE	The independent variable for the installation of a sliding gate
MoE	The margin of Error, the percentage of error per predicted value.
P_{GTV}	The generalised tender value to the preferred month aided by CPI
$P_{awarded}$	The official awarded contract value with no adjustments
CF	Cost Factor, the factor used to transform the fencing material type to a numerical figure
EF	Escalation Factors, the factor used to transform the motorisation of a gate to a numerical figure
S_{price}	The supplier pricing score that was presumed to be 80 for this study
S_{BEE}	The supplier BEE score out of 20
P_{SGTV}	SGTV; Scaled generalised contract value of the awarded tender price by incorporating the relevant CPI and the supplier BEE score.
SEE	Standard estimation error, the error estimate for a model generated by IBM SPSS Statistics 26
RMSE	Rooted mean square error, a measure of the quality of an estimator. It is always non-negative, and values closer to zero are better.
P_{pred}	The model predicted price (outcome)
P_{LB}	The lower bound of the predation interval
P_{UB}	The upper bound of the predation interval

List of Definitions

Term	Abv.	Definition
Anchor Predictor	-	The primary input (independent) variable that is assumed to have the highest correlation with the theoretical outcome as it does in the actual world.
Analysis of Variance	ANOVA	A statistical analysis to test the degree of differences between two or more groups of an experiment. It is summarised in table form by statistical software.
Apartheid	-	A policy or system of segregation or discrimination on the grounds of race that was prominent in South Africa pre-1994
Auditor General of South Africa	AGSA	The Auditor-General produces audit reports on all government departments, public entities, municipalities, and public institutions every year – to help ensure accountability and good governance in the public sector
Benchmarking Data	-	Non-Clean audited municipal data that was assessed using the developed and tested benchmarking model to identify cases of pricing deviations
Bidding Price Fixing	-	When a collective group of suppliers control the outcome of the accepted price for a bidding process
Black Economic Empowerment	BEE	South Africa's policy of black economic empowerment is an initiative to redress the wrongs of the South African past
Clean Audit	-	An audit outcome given to an organisation that passes all the required auditing assessments
Competitive Tender	-	A Municipal South African procurement valued at more than R 200 000 (excluding vat)
Confidence Level	-	The percentage of all possible samples that can be expected to include the true population parameter
Convenience Sampling	-	A sampling method based on easily accessible data to acquire an estimate.
Consumer Price Index	CPI	The change in the cost of goods and services sold over certain periods.
Cross-Validation	CV	A method of validating the data used to build a model.
Generalised Tender Value	GTV	The awarded contract value by incorporating the relevant CPI transformations
Good Performance Index (GPI)	GPI	An index developed by Good Governance Africa to assess municipalities holistically based on three themes: administration, economic development and service delivery
Hierarchal Regression	-	A sequential building of several multiple linear regression models, each adding and removing independent variables
Homoscedastic	-	When the variance of error terms are similar across the values of the independent variables
Hypothesis Test	-	A statistical test to assess the validity of a premise.
Insignificant Variables	-	Variables that have minor to no significands in a modelling assessment

Term	Abv.	Definition
Irregular Expenditure	-	A financial transaction that did not follow the due process
Multiple Linear Regression	MLR	Attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data
Model Building Data	-	Clean audited municipal data that satisfies the 10-folds cross-validation technique and was used to build the model
Model Testing Data	-	Clean audited municipal data left out from the model building data to be used to test the benchmarking model
Model Out-of-Range Data	-	Clean audited municipal data that is above R230 000 (including vat) used to assess the predictive accuracy of the model for out of price range contracts
Multicollinearity	-	When the relationship between either of the independent variables is highly correlated
Municipal Finance Management Act	MFMA	Intended to secure sound and sustainable management of the financial affairs of municipalities and other institutions in the local sphere of government; to establish treasury norms and standards for the local sphere of government
Municipal Performance Index	MPI	An index developed by PWC to assess municipalities holistically based on three categories; socioeconomic outcomes, service delivery and municipal governance
Municipality	-	An urban district having corporate status and powers of self-government
Non-Clean Audit	-	An audit outcome given to an organisation that has not passed some, if not all the required auditing assessments
Non-Competitive Tender	-	Low-Cost Procurements; a Municipal South African procurement valued at less than R 200 000, sometimes R 230 000 when including vat
Pearson’s Correlation Coefficient	-	A statistical tool that measures the linear correlation of a data set
Polynomial Regression	-	Polynomial regression is like linear regression, however, can handle non-linear datasets
Producer Price Index	PPI	The change in the cost of goods and services produced over time
Prediction Interval	-	A range of figures where there can be a high certainty the true predicted value lies based on the confidence level
Predictors	-	Also referred to as independent variables, the variables used as inputs in a model built
Preferential Procurement Policy Framework	PPPFA	Provided for the implementation of a procurement policy for a category of preference in awarding contracts, and for the protection and/or advancement of persons or categories of persons disadvantaged by past unfair discrimination
Price Inflation	-	The above market payments of goods and services procured
Pricing Deviations	-	When the price of a procurement received deviates significantly from the expected amount, be it an over-price or an under-price

Term	Abv.	Definition
Promotion of Accesses to Information Act	PAIA	An act that mandates governmental bodies to disclose information that is or was available to the public
Public Procurement	-	When the public sector buys goods or services from suppliers in the local or international markets, usually employing an official contract referred to as a tender
Quota Sampling Method	QSM	A nonprobability sampling technique in where the researcher selects the population-based on certain desired features
Rand	R	The South African currency that was equated at a US $1.00 = R17.00 exchange rate for this study.
Scaled generalised contract value	SGTV	The awarded contract value by incorporating the relevant CPI and the supplier BEE score
Secondary Data		Data that was collected by another user. Common sources include government organisations, censuses and data that was originally collected for previous research.
Statistics South Africa	StatsSA	The national statistics agency of South Africa established under the Statistics Act (Act No.6 of 1999) to produce timely, accurate and accessible official statistics
Supply Chain Management	SCM	The department in an organisation that is in charge of all procurements for the company
Tender	-	An official contract for procurements in the South African public sector.
Tender Price Index	TPI	An index used by the construction industry as a guide to estimate project cost
The *Constitution*	-	The supreme law that governs the Republic of South Africa to protect all the rights of the citizens
Trend Analysis	-	A method that enables a modeller to look at quantitative data that has was collected over a duration of time
Value-for-Money	-	An industry terminology that supposes the quality matches the prices

***Keywords**: Multiple Linear Regression, Pricing Deviations, Price Inflations, Public Procurement, Benchmarking*

Chapter One: Introduction

1.1 Research Background

The opening declaration of the Preamble to the *Constitution* of South Africa (Republic of South Africa, 1996) reads "*We, the people of South Africa, Recognise the injustices of our past*". This acknowledgement has had a profound and transformative effect in all sectors of the country. The effect is more prevalent in the public procurement system as it is being used as a policy tool to rectify the prejudices experienced in the country's apartheid past (Bolton, 2006; Warf, 2019).

To deal with the apartheid legacy, direct mediation known as Black Economic Empowerment (BEE) came about (Acemogluy *et al,* 2007). BEE is the economic enablement of previously disadvantaged ethnic groups, in particular black, Indian, Asian and mixed-race persons (Akwasi, 2020). Among the aims set for BEE, is to increase black-owned, managed and controlled companies (DTI, 2018).

The policy was revised to focus more on the distribution of the nation's wealth across all races and genders (Pillay, 2020). The revised policy is known as *the Broad-Based Black Economic Empowerment Act 53 of 2003* that led to the term "B-BBEE" (Horn and Raga, 2012). Though these terms are occasionally used interchangeably in the industry and for this reason, they are referred to as BEE for the duration of this study.

The BEE initiative, while it has been successful to an extent in integrating managerial roles and diversifying business ownerships (Da Silva, 2006); has failed most times for public procurement; a process in which the government acquires goods and services from the private sector (Ambe, 2016). This has led to an increase in costs to the economy without fully supporting the initial target for its development to redress the imbalances in the country caused by apartheid (Badenhorst-Weiss and Ambe, 2012).

These failures may be linked to systemic inefficiencies which include a possibly flawed BEE system (Gumede, 2020) or ineffective recruitment processes that lead to the appointment of ill-equipped individuals in key positions of governmental organisations, including procurement departments (Motubatse, 2016). However, corruption, considering that it is one of the biggest challenges in the South African public procurement sector, is the more likely cause of these failures (Thornhill, 2006).

In the scrutinising of government necessities, Munzhedzi (2016) found that most government institutions need more or less the same goods and services namely, laptops, photocopy machines, furniture and stationery. These outcomes are understood by the author to mean that the expenditure for equally scaled municipalities when observing the size of the population they cater for and the economic activity within their area of operation, should not differ extensively or have valid reasons for significant variations in overheads (e.g. having undertaken multiple construction projects).

On the analysis of an article detailing municipal finances (National Treasury, 2017), the author noted that cities at similar stages of economic development, catering to like population sizes with identical revenue collections, vary significantly when compared against their overall expenditures. In comparison with the work of Munzhedzi (2016), stated above, this irregularity was identified as an anomaly requiring further exploration. The result of which is this research focus on public procurement expenditure comparisons for municipalities in South Africa.

1.2 Problem Statement

Pooe *et al.* (2015) stated that compliance with guidelines, that are meant to contain procurement costs in order to prevent price inflations (pricing regulations), remains a serious challenge for South African municipalities. As a result, municipalities are unable to carry out their mandate of service delivery effectively. Non-compliance issues include pricing deviations in supply chain management (SCM) departments not abiding by regular procurement procedures. Reasons behind the practices could be as a result of an *emergency procurement* clause (Republic of South Africa, 2017) or, more likely, as a result of fraudulent behaviour in procurements (Ambe, 2012).

Furthermore, an independent institution known as the Auditor General of South Africa (AGSA) publishes annual audited financial reports of governmental organisations, including but not limited to municipalities. These results are classified as *clean* or *non-clean audited*. A clean audit of a municipality shows its working within the prescribed legislation, whereas a non-clean does not (the concept will be explored further in the next chapter).

The 2019 audit report revealed municipalities in the country had over R32.06 billion (~US. $2 billion) [9.9% of the overall municipal revenue] in irregularly awarded contracts in the 2018/2019 financial year; an increase of R7.68 billion [31.50%] (~US. $500 million) from the previous year's audit outcomes (AGSA, 2020a). However, according to the National Treasury's *Irregular Expenditure Framework* (2019), irregular expenditure does not translate directly to corruption because it is defined as a financial transaction that did not follow the due process. An increase in expenditure that does not follow due process may indicate a rise in corrupt activities (Lambert-Mogiliansky and Kosenok, 2009).

Additionally, the results of the Auditor General's report stated that approximately 17% of irregular expenditure for the 2018/2019 financial year was attributed to unfair procurements that led to over-priced goods and services (AGSA, 2020a). The result of which was an adverse impact on municipal finances and their ability to perform their *Constitutional* decree of delivering services to communities (Republic of South Africa, 1996).

These indiscretions are a common sight in South Africa, with 45% of all reported corruption accounts, by whistle-blowers emanating in the Province of Gauteng, including public procurement corruption (BusinessTech, 2019; Daniel, 2019). Furthermore, more than 90% of the municipalities in the province have had recurring non-clean audit outcomes in the past five financial years (AGSA, 2020a). For these reasons, the Gauteng Province is the focus of the study.

Contrary to the example of the Gauteng Province, multiple municipalities in South Africa have repeatedly received clean audits; this being a sign that due process was followed when procurement was taking place (AGSA, 2020a). Based on the work by Aadnesgaard and Willows (2016), there is a positive correlation between the clean audit status of a municipality and their capability to deliver on services. As there is a direct relationship between service delivery and the ethical conduct of a municipality (Bizana, 2013); it can be assumed the probability of price inflations through corruption in clean audited municipalities is fairly low.

Therefore, an index or tool is required to compare these two classes of municipalities to identify cases of price inflations to understand which departments may be susceptible to corruption. While similar indices already exist in South Africa (as explored in the following chapter), they employ a holistic approach when comparing municipal performance. As a result, there is potentially a dilution of individual cases of price inflations that may have taken place. The research gap was thus identified; a tool/index/method is required to compare municipal procurements individually.

The above establishes a further-narrowed research focus for the study; a method is required to compare the clean audited municipalities in South Africa against non-clean audited municipalities in the Gauteng Province. The comparison would have to be for procurements at an individual level of assessment. As stated by Bizana (2013), "there are various types of public procurements including but not limited to *non-competitive, low valued tenders* (*less than R200 000 excl. vat [~US $11 700]*) and *competitive tenders, high valued* (*more than R200 000 excl. vat*)."

However, high-valued, competitive contracts are vetted more vigorously in South Africa compared to non-competitive contracts (Mantzaris, 2014). Following from this, it is interpreted that low valued, non-competitive contracts would be more vulnerable regarding public procurement corruption and in particular, the inflation of contract prices. This practice results in bloated municipal budgets which harm local service delivery (Manyathi, 2019). Based on the above evidence and trends; the purpose statement of the study is summarised as follows:

A comparative assessment of low-valued tenders between clean audited municipalities in South Africa and non-clean audited municipalities in Gauteng to identify price inflations.

1.3 Significance of the Study

According to a study by Hawkins *et al.* (2011), most public office-bearers are motivated by self-interest to amass power for personal gain. These claims are substantiated by the numerous corruption cases against senior government office bearers (Pooe, Makhubele and Mafini, 2015), mostly associated with public procurement fraudulent cases. Consequently, there has been a detrimental effect on local service deliveries by municipalities (AGSA, 2019).

Therefore, this research aimed to develop a benchmarking model to introduce a method for evaluating contracts on a discrete basis. "*Discrete basis*" refers to the benchmarking of a contract to compare its pricing to what the true market offers. This may provide a tool for the South African government to monitor and flag pricing deviations in the past, present and future procurements. This means that it shall present a proactive measure of control against public procurement corruption as compared to the reactive measures currently in place (e.g. auditing).

The various hindrances to assessing public procurements are explored carefully in chapter three. These include a disparity in procurement requests between municipalities (e.g. requesting different fencing types, a drastic variation in required BEE levels or different volumes of the procured item) which introduces complications when equating contracts without the aid of the proposed comparison method.

This practice makes it easier for cost inflations to take place unnoticed. The study aimed to provide a tool to bypass this drawback and compare contracts despite disparities that may be present. Essentially providing a simple method of identifying the relationship between the suppliers' controlled variables and the awarded pricing.

This was done through the aid of computerised modelling techniques that are further explored in chapters two and three. The study outcomes are an outlined framework for developing a procurement comparative tool for any low-cost public procurement to identify pricing deviations in the past, present and future contracts and answering the critical research question (CRQ) through the help of the below-mentioned objectives.

1.4 Critical Research Question and Objectives

1.4.1 Critical Research Question

What are the pricing deviations in low-valued procurements when comparing clean audited municipalities in South Africa against non-clean audited municipalities in the Gauteng Province?

1.4.2 Research Objectives

The following objectives were used as milestones when conducting the study. Not only were these research objectives satisfied but also aided in developing a structure used for chapters Three, Four, Five and Seven throughout the research:

1. Develop a framework for distinguishing between significant and non-significant variables in the collected municipal tender archive data; defined as those that did not have a noteworthy impact on the price of a tender. Thereafter, develop the comparative model based on the significant variable identified.

2 Using the generated comparative model, identify which contracts in the Province of Gauteng that may have incurred pricing deviations and by what margin. Subsequently, identify the reasons behind the deviations.

3 Based on the methods used in the study, generate a framework outline that can be used as a pricing comparison guideline for future procurements beyond the assessed tender category and identify drawbacks in this study for future researchers to avoid.

1.5 Research Hypothesis

The study aimed to develop a benchmarking model to identify possible cases of pricing deviations that may have occurred in recurrent non-clean audited municipalities in the Province of Gauteng. The *null* hypothesis (H_0) and the hypothesis (H_1) were derived from the critical research question and the previously explored literature. Thereafter, tested using relevant computerised modelling techniques explored further in the study:

H_0: Clean Audited Municipalities in South Africa Do Not Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng

H_1: Clean Audited Municipalities in South Africa Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng

1.6 Scope for the Study

The research is confined to the comparison of procurement costs of non-competitive contracts between clean audited municipalities in South Africa against non-clean audited municipalities in the Gauteng Province. The comparison explores which factors may have a direct impact on the pricing of these contracts by the removal of insignificant variables.

Thereafter, construct a cost comparative model with the significant variables through the aid of existing computerised modelling techniques so to assess each procurement individually. This will flag which of the procurements may have had price inflations. Then the findings will be compared with the existing works as described in the literature review.

Based on the audited outcomes for the thematic focus, the scope of this research study can be categorised into five groups. These groups are derived from the populations the study wishes to compare and the key attributes required for the modelling stages. They are a culmination of all the desired key features of each municipal audit type. These are as listed below followed by a summary of the scope of the study in Table 1.

1. The population of interest (clean and non-clean audited municipalities).
2. Focus location (provincially and nationally).
3. The purpose of the population (model building and benchmarking).
4. Sample sizes (based on minimum statistically allowable and data availability).
5. The applicable duration (period to collect samples).

Table 1. Research Scope of the Study

Group	Justification	Clean Audited Municipalities	Non-Clean Audited Municipalities
Population of Interest	These are the two selected populations, including clean audited South African and non-clean audited Gauteng municipalities required for the benchmarking. The ±five year assessment period was so to have ample data on the municipalities.	Municipalities under consideration were ones with three or more clean audits in the past five years to have a sufficient samples size for model building. The clean audit standard of this study was based on frequency.	Municipalities under consideration were ones with four or more non-clean audits in the past five years to have high certainty in the municipality's non-clean audited status while having a sufficient sample size.
Focus Location	The focus locations so as to limit the number of contracts to assess.	South Africa (including Gauteng) to have a sufficient sample size for model building purposes.	Gauteng province as 45% of all reported corruption cases stemming from this province.
Purpose of the Population	This is the purpose of each population for the different modelling stages. The descriptive function of each dataset is available in Table 27 in chapter three.	Used to build, test the comparison model and teach it how *clean audited* municipalities would price the specific procurement.	Used to assess how the pricing of the non-clean audited municipalities would change should the tender been issued by a clean audited municipality.
Sample Size	The sample sizes are the minimum requirements to have a reputable benchmarking exercise. These vary according to the sample data purpose (e.g. for training or testing).	This data required a minimum sample size that was explored in the methodology, as was the size required for a credible model build.	The minimum sample size was determined by data availability and willing participation by the municipalities but the ideal sample size was explored in the methodology.
Applicable Duration	The study had a set time limit therefore one also had to be set for the data collection duration.	Collected and processed over twelve months because of having a greater sample size	Collected and processed over six months because of having a smaller sample size and expected delays
Final Comments	Because of resource constraints, the research was restricted to one frequently requested low-cost procurement in South African municipalities. The procedure used to select the study contract focus is explored in chapter three.		

1.7 Ethical Clearance for the Study

Ethical clearance was required for the study as it dealt with secondary data that may be viewed as confidential by the parties involved (including municipalities and the University). Though it is governmental data and by law is supposed to be readily available to the public, due process was required by applying for an official University ethical clearance certificate.

The ethical clearance also covered instances, where specific municipalities would be contacted for details and official governmental requisition documents known as promotion of access to information forms (PAIA), were used. School protocols for obtaining the certificate were followed which led to the research being viewed as a “Minimal Risk” by the University of Witwatersrand *Non-Medical Human Research Ethics Committee*. The university issued ethical clearance certificate is available in Appendix A, Section 2.

1.8 Brief Chapters Overview

Below are the chapter synopses for the research paper. It should be noted, all extensive chapters, sections and subsections have a summary/conclusion of their respective contents at the end of each segment. These can additionally be used as side notes to clear up any doubts when requiring clarity on a subsection that required pre-reading of the previous segments. The list of symbols and definitions can be referenced to understand any non-familiar concepts in the summaries without having the review the entire segment.

1.8.1 Chapter Two Synopsis

A literature review presents a blend of empirical works according to relevant variables and justifies how the study addresses a problem (Bloomberg and Volpe, 2018). This review is segmented into a thematic and theoretical review. The thematic review explores all the new concepts introduced in chapter one and explored the viability of this study being integrated into governmental procurement systems. The theoretical review assessed which benchmarking tool would be ideal to base the study on. Thereafter, reviewed a past study to identify likely errors associated with the selected benchmarking technique.

1.8.2 Chapter Three Synopsis

The methodology chapter provides a detailed description of all aspects of the design and procedures of the study (Bloomberg and Volpe, 2018). This chapter is a detailed account of how to construct the study regression model with the aid of the relevant quality controls. It explored the probability of clean audited municipal procurements that may have been corrupted.

1.8.3 Chapter Four Synopsis

The study's main results are arranged and reported in the findings chapter, including the presentation of relevant quantitative and qualitative results (Bloomberg and Volpe, 2018). The findings depicted all the study results that followed the structure developed in the methodology. Each subsection has a brief data analysis so the reader understands why the subsequent steps were taken.

1.8.4 Chapter Five Synopsis

The analysis and discussion chapter offers an opportunity to reflect thoroughly on the study's findings, and the practical and theoretical implications (Bloomberg and Volpe, 2018). This chapter is a detailed analysis of all the results (intermediate and final) to test the research hypothesis. It also critiques the modelling assumptions made and the study limitations then partly connect the results to the reviewed literature.

1.8.5 Chapter Six Synopsis

The transferable findings chapter for this study is used to create a framework outline to fulfil the third objective and to draw real-world applications of the developed model. It further explored the economic impact of the 2020 COVID-19 global pandemic and provides guidelines as to how the research outcomes could be used to remedy some of the financial damage incurred by the country because of price inflations.

1.8.6 Chapter Seven Synopsis

The conclusion presents an integration of the study findings, analysis, interpretation, and synthesis (Bloomberg and Volpe, 2018). The conclusion summarised the validity of the results. These included the assessment of the statistical significance of the model and its alignment to existing works. Thereafter, attempting to link the research findings to the reviewed literature.

Chapter Two: Literature Review

The topic explored in this study required a qualitative examination of existing literature and government procurement practices as the justification of the selected populations of interest. Thereafter, quantitative analysis to select and assess the ideal tool for the benchmarking process. This type of method is known as *mixed-method research*, which suits the literature review as a two-part evaluation. Figure 1 is a chronologic depiction of the overall literature review layout and the thought process implemented.

Thematic Review: Review of qualitative works to explore the concepts introduced in Chapter One

- *Municipal Comparison Tools*: To identify and qualitatively substantiate the research gap based on the drawbacks introduced in the opening chapter on current existing indices/tools
- *Public vs Private Sector Procurement*: The contrasting of public and private sector procurement practises is to better understand why the research could not use private sector prices as the true market value of the benchmarking. Included in the section is the introduction of regulations from the public sector that act as gatekeepers to make cross-sector pricing comparisons not viable which lead to the assumption of clean audited municipalities having the true market value.
- *Clean Audited vs Non Clean Audited Municipalities*: The section is a breakdown of what a clean audit entails and its significance to the research. This is to better understand the assumption made in the opening chapter of clean audited municipalities having the true market value of public procurements.
- *Review of Price Inflation Cases in Gauteng*: The segment is a brief study of price inflation cases in the Gauteng Province to illustrate the reasons behind the selection of it as the study focus.
- *Benchmarking Municipalities Variables*: This section is a scrutiny of the key variables to consider when conducting the benchmarking exercise to aid in selecting the ideal comparison tool.
- *Municipal Procurement Process*: This division is a depiction of generic municipal procurement processes in South Africa to identify where the presented study solution can be implemented.

Theoretical Review: Reviewed quantitative works to select & evaluate an ideal benchmarking tool

- *Methods of Benchmarking*: The section assesses the existing benchmarking tools available and selects the ideal one based on simplicity and compatibility with the study objectives
- *Past Work into Understanding MLR Modelling Errors*: This section evaluates the likely errors to be incurred with the selected benchmarking tool to better interpret the study outcomes.

Figure 1. Outline of Literature Review

2.1 Thematic Review

In South Africa, during the transition to the new democratic state we now know, a supreme law known as *the Constitution* was drawn up. Its primary mandate is to protect the rights of all citizens and to ensure the injustices experienced in its apartheid past are not carried into the new era (Republic of South Africa, 1996).

Amongst the mandates set in the *Constitution* for local government, is the provisioning of services to surrounding communities sustainably (Republic of South Africa, 1996). However, various studies have revealed this is not the case for most municipalities in the country. One study, in particular, is by Badenhorst-Weiss and Ambe (2012). Their research revealed that there are multiple challenges experienced by local municipalities in delivering goods and services to their respective communities. According to Pooe *et al.* (2015), this problem frequently stems from procurement corruption.

The research made use of the purposive sampling method to evaluate 14 governmental institutions, six of which were municipalities (Badenhorst-Weiss and Ambe, 2012). Their findings revealed that inadequacies and inconsistencies in risk management leading to irregularities in supply chain management (SCM), were some of the many challenges faced by governmental organisations, together with municipalities (Badenhorst-Weiss and Ambe, 2012).

This implies that there is a “lack of internal control environment and implementing risk mitigation procedures through the effective utilisation of SCM policy and procedure” (Badenhorst-Weiss and Ambe, 2012). These findings are further substantiated by the research of Pooe *et al.* (2015), that explained the failure to comply with SCM regulations is one of the many challenges faced by governmental institutions, more so in the Province of Gauteng. The above occurrences lead to irregularities within municipal SCM departments in the country.

There are various reactive measures available to assess municipal irregularities in South Africa. However, only two are explored in this review. First, was an annual independent audit report of all municipalities in the country that has been researched further in the chapter. Second, two comparative models to assess municipal performances as shown below.

2.1.1 Municipal Comparison Tools

2.1.1.1 Good Performance Index (GGA, 2019)

The Good Performance Index (GPI) is developed by Good Governance Africa (GGA); a non-profit organisation that assesses 213 municipalities in South Africa. These only include metropolitan (Category A) and local municipalities (Category B) and ignores district municipalities (Category C) as their activities are inclusive in the local municipalities (GGA, 2019, p. 4).

The index is constructed by assessing three themes: administration, economic development and service delivery; Table 2 below is a breakdown of each theme:

Table 2. Breakdown of the Assessed Themes for GPI ***(source: GGA, 2019)***

Theme	**Indicator**		
Administration	Municipal capacity	Compliance	Financial soundness
Economic Development	Poverty	Work opportunities	Unemployment rate
	Individual income		
Service Delivery	Water	Police coverage	Sanitation
	Health facilities	Refuse removal	Education
	Electricity	Informal housing to formal housing	

All 213 municipalities are assessed using the indicators listed in Table 2. The data required is sourced from Statistics South Africa (StatsSA) and the Auditor General of South Africa (AGSA). The indicators use a five-point *Likert scale* from one (best performing) to five (worst performing). An example of the scaling method for access to electricity would be 80.1% to 100% as one and 0% to 20% as five. The overall municipal scores are added up, and the ranking is derived using the same concept so the lower overall scoring municipalities ranking best.

2.1.1.2 Municipal Performance Index (Williams and Rampai, 2020)

The Municipal Performance Index (MPI) was developed by PWC, a multinational corporation operating in South Africa. The index is used to compare municipalities to identify where limited resources should be targeted. It bases the analysis on three categories; socioeconomic outcomes, service delivery and municipal governance.

The performance categories are further divided into indicators similar to GPI as shown in Table 3 below. These indicators are over a thousand in total meaning more variations are accounted for. The MPI tool is a cloud base instrument that allows for the index to be recalculated to depict live scores as top quartile or median. The company updates the data quarter-yearly, as per the results published by municipalities.

Table 3. Breakdown of Each of the Assessed Subcategories ***(source: Williams & Rampai, 2020)***

Category	**Performance Category**		
Socio-Economic Outcomes	Employment & Economy	Living Standards	Health
	Safety	Education	Environment
Economic Development	Water	Amenities	Electricity
	Transport	Sewerage	Internet
	Refuse	House	
Service Delivery	Financial Performance	Operations	Financial Management
	Revenue Management	People	Working Capital
	Strategy	Ease of doing business	
	Governance, Risk Management and compliance		

2.1.1.3 Municipal Comparison Tools Summary

The above indices, though similar in assessed categories, are fairly different regarding validity. The Municipal Performance Index is more credible than the Good Performance Index in its analysis as it factors in more indicators that can explain more variance in the outcomes. Both these indices share an identifiable drawback of having a holistic approach to municipal assessments.

SCM contracts are not assessed as standalone items but integrated into the outcomes. This may dilute the results, consequently, some irregualties within SCM contracts may be overlooked. The challenges posed by municipal comparison tools informed the research gap for this study. A tool/index is required to scrutinise individual tenders within municipalities to identify contracts where procurement irregularities such as price inflations may have occurred.

However, as previously stated, high-valued, competitive contracts are vetted more vigorously compared to lower-valued procurements (Mantzaris, 2014). As a result, low-valued, less competitive contracts would be vulnerable when the inflation of procurement prices is at play. Therefore, the tools/index developed should be modelled around low-valued procurements within municipalities.

The developed comparative tool/index requires a baseline for what the “true market value” is for procurement. Therefore, a sample is necessary to teach the index/tool how the pricing should be structured. Ideally, the data would be sourced directly from the private sector, yet as explained in the following section, this is not the case.

2.1.2 Public vs Private Sector Procurement Processes

Supply Chain Management (SCM) is the system in which procurement takes place. SCM integrates both demand and supply management processes into a singular department in organisations and this process includes planning, implementing, and controlling supply (Oppelt, 2019). Ambe (2012) further defines SCM as a process to manage and coordinate all the supply chain activities necessary to support the organisation's strategy of delivering the right quantity of the product [or service] to the right place at the right time.

Procurement in both the private and public sectors can be further explained through subdivisions within the process; Figure 2 is a depiction of the segments. These may be termed as a general practice, but as explained further in the review, different sectors have altered approaches to implementing the process.

1. Identifying an item or service required to fulfil the needs of the organisation.
2. Identifying the most suitable service provider to satisfy that need.
3. Establishing a fair and reasonable price for the item or service to be purchased.
4. Enforcing the agreement for the purchase.
5. Managing the relationship to ensure there is delivery and at the agreed upon standard.

Figure 2. Five-Stages in the Supply Chain Management Process ***(source: Oppelt, 2019)***

Public and private sector procurement though similar in many aspects differs when assessing reporting structures, regulating bodies, funding sources and operating motives (Larson, 2009; Damas, 2013; Heller, 2013). Larson (2009) further states that the public procurement sector is governed by the relevant legislative bodies, laws, and regulations, whereas the private sector is overseen by boards of directors and business plans.

Table 4 is a generic contrast of the two sectors. The differences vary significantly when assessing a national governments' procurement policy framework and an individual private company's procurement policy. Preceding the table is an exploration of the key differences in the sectors when comparing the South African public procurement system to that of the private sector. The compared items are based on the fundamental differences between the sectors as identified by Green (2020), and Newman & Swift (2001). These primarily include the legislative and regulative practices.

Table 4. Public vs Private Sector Procurement ***(source: Ambe, 2012, p. 138)***

Feature	Private Sector Procurement	Public Sector Procurement
Goal	Profit-making from customers	Quality service delivery to all citizens in the country
View of SCM	Procurement is viewed as an element of SCM	SCM is viewed as a procurement tool
Sources of revenue	Sales of goods and services	Taxes and fees
Governance	Guided by a board of directors and business plans	Legislative bodies, laws and regulations
Skills	Tend to have highly skilled actors	Tend to have less skilled actors
Receptiveness	Emphasis on innovation and entrepreneurship	Emphasis on accountability and transparency
Organisational structures	Firms of many sizes with room for new entrants (less complex)	A highly complex system of organisations with various tasks

2.1.2.1 Regulations And Legislation

Every organisation, public or private is governed by a set of regulations and legislations on which their operational methods are based on. However, the public sector is largely constrained by its regulations when compared with the private sector (Lau, 2018). Though the private sector in South Africa is still inhibited by policies such as BEE requirements to trade with the state. Nevertheless, there is more freedom from legislation in the sector (Smith, 2017). Contrarily, the public sector is constrained by a centralised procurement policy framework which may delay the procurement processes through bureaucratic procedures (Lau, 2018).

2.1.2.1.1 Private Sector Regulations and Legislations

Green (2020) stated the private sector procurement policies are geared towards optimising their primarily profit-driven business goals. This allows for sourcing of suppliers as and when required with no official bidding process (NCPPC, 2019). As found by Heller (2013), the procurement cycle in the private being far lower than that in the public sector.

Though the private sector makes use of a centralised code of corporate governance. Particularly in South Africa, a document known as the King’s Report (King IV, 2016), publishes a set of principles and leading practices by which corporate governance may abide. These are voluntary and are not in any way imposed on the private sector.

2.1.2.1.2 Public Sector Regulations and Legislations

As previously mentioned, the transition to the new democratic state of South Africa led to the formation of a supreme law known as *the Constitution*. Included in the Preamble of *the Constitution* is the statement “Heal the divisions of the past; social justice and fundamental human right” (Republic of South Africa, 1996). This aim focuses on redressing the imbalances caused by the country’s oppressive past and can be interpreted as social, political, environmental (spatial) and most relevant to the study, economic inequities.

A method in which South Africa addresses this aim is the introduction of black economic empowerment (BEE) in supply chain management (SCM) (Krüger, 2011). BEE was introduced to deal with the country’s past economic imbalances brought upon by Apartheid (Acemogluy, Gelbz and Robinsonx, 2007). The BEE policy was integrated into a procurement structure known as the *Preferential Procurement Policy Framework Act 5 of 2000* [PPPFA] (Pooe, Makhubele and Mafini, 2015).

The policy led to a BEE scoring system based on the diversity of a company’s operation. It is ideal for a company to be BEE compliant as the key incentive is *preferred* trade with the government (Krüger, 2011). This being a favourable situation for any company as, according to Hawkins *et al.* (2011, p. 561), “The government purchasing market makes up the largest business sector in the world”.

The BEE scoring system varies according to company size. Three primary factors determine a BEE score as described in Table 5. Including, direct empowerment by company ownership and management; employment equity and upskilling employees; and indirect empowerment by procuring and encouraging local black-owned companies (Acemogluy, Gelbz and Robinsonx, 2007).

Table 5. Elements and Weightings of the Generic BEE Scorecard ***(source: Ambe 2016: 284)***

BEE Pillar		**Weight**	**Regulatory Objective**
Direct Empowerment	Ownership	20%	Encourages the sharing of ownership which will result in voting rights for black people in general and women in particular.

	Management	10%	Encourages senior black decision-making at the executive board and senior top management levels
Employment Equity	Employment equity	15%	Encourages companies to recruit black people in professional, middle and lower management positions
	Skills development	15%	Encourages companies to develop black talent through spending on skills development and learnerships
Indirect Empowerment	Preferential procurement	20%	Encourages the development or expansion of black small, and medium enterprises
	Enterprise development	15%	Measures the extent to which enterprises procure from BEE-compliant companies. Encourages spending on small and micro-enterprises.
Residual	Socio-economic	5%	Encourages initiatives intended to directly provide black people with a means of generating income for themselves.

There are two parts to the preferential point system for municipalities. The 80/20 rule is valid for contract values from R30 000 (~US $1 700) to R1 million (~US $58 800). Then the 90/10 rule for contract values above R1 million (Dobie and Xinwa, 2015). Where the 10 and the 20 fragments show the potential suppliers' BEE compliancy scores which are gained using Table 5.

The BEE score for the above points system totals 100. After which, is scaled down to either 10 or 20 depending on the contract value as shown in Table 6.

Table 6. Calculation of Points for BEE Rating ***(source: National Treasury, 2001)***

BEE level	Points for 90/10	Points for 80/20
1	10	20
2	9	18
3	8	16
4	5	10
5	4	8
6	3	6
7	2	4
8	1	2
Non-Compliant	0	0

Thereafter, the preferential score is calculated using the below formula for the 80 or 90 segments of the scorecard (Dobie and Xinwa, 2015). Where P_s represents the points scored by the supplier, P_t is the price offer under consideration from the supplier and last, P_{min} is the price of the lowest acceptable offer for the item being procured. The outcome would be a tally of both scores.

$$P_s = (80\ or\ 90\ ) \times (1 - \frac{P_t - P_{min}}{P_{min}})$$

The supplier with the highest score would be the favoured to win the contract. However, as stated by Lowman (2017) and Acemogluy *et al*. (2007), the above method could prioritise a costly supplier that is BEE compliant over an affordable seller that is none compliant; a direct consequence, increasing the cost of procurement.

As the PPPFA was completed, an act was required to ensure municipalities upheld their constitutional mandate. This is known as the *Local Government Municipal Finance Management Act 56 of 2003* (Ambe, 2016). From the Municipal Finance Management Act (MFMA), a regulatory procurement policy was derived for municipalities (Republic of South Africa, 2017, p. 91).

The procurement policy came as guidelines for municipalities to adhere to (Badenhorst-Weiss and Ambe, 2012); these measures slow down the acquisition process but are however supposed to limit the possibility of unethical practices taking place as mandated by *section 217(1) in the Constitution of South Africa* (Republic of South Africa, 1996). Though this has not been the case for some municipalities in the country particularly the Gauteng Region where 45% of all the corruption cases in the country were reported in 2018 (BusinessTech, 2019; Daniel, 2019).

2.1.2.2 Public vs Private Sector Procurement Processes Summary

Procurement in the private sector has many similarities to the public sector. However, the private sector follows a less socially sustainable but more cost-effective procurement practice as stated by Jenks (2017). In contrast, the public sector; more especially in South Africa, is people orientated and as a result, an increased cost of procurement because of the attempt to correct the country’s injustices of the past (Acemogluy, Gelbz and Robinsonx, 2007; Lowman, 2017).

Therefore, using private sector contract prices to find the “true market value” for the study would be inaccurate and bias toward the sector; for this reason, the benchmarking exercise is restricted to comparing governmental structures against one another. This is to ensure an *apples-to-apples* comparison, and that the focus of the assessment being pricing deviations between similar goods or services procured in past awarded contracts by municipalities.

The two groups being compared, as previously stated, are clean audited municipalities in South Africa and non-clean audited municipalities in the Gauteng Province. The following section elaborates on the differences and justifies the assumption made on clean audited municipalities having the actual market value for public procurements in South Africa, and follows a logical approach to explaining when and how the auditing body came about.

2.1.3 Clean Audited Municipalities Vs Non-Clean Audited Municipalities

2.1.3.1 Chapter 9 Institutions of South Africa (Auditor General of South Africa)

In South Africa, the supreme law known as *the Constitution* exists to uphold the rights of the citizens. Under the *Constitution*, independent organisations known as *Chapter 9 Institutions* were formed; and subject only to the law. As declared, when conducting their daily operations require impartiality and performing their functions without fear, favour or prejudice (Republic of South Africa, 1996).

The *Auditor General of South Africa* (AGSA) is one of the Chapter 9 Institutions and its mandate is to audit and advise government establishments on their financial management and accounts (Republic of South Africa, 1996; Tom, 2014; AGSA, 2020b). As a direct result, they are enabling oversight, accountability and good governance in the public sector (AGSA, 2020b). Under *the Constitution,* the AGSA conducts yearly audits on all municipalities in the country. Thereafter makes the results available to the public (Ngoepe and Ngulube, 2013; Mailovich and Phakathi, 2018).

2.1.3.1 Assessment Criteria to Acquire Status

Municipalities are assessed using pre-developed criteria that include financial statements, reporting on the previous year's recommendations and compliance with the legislation set up by the government (Republic of South Africa, 1996). Once an audit has been conducted, a set of the stated correctional objectives are given to the municipality which shall form part of the following year's assessment (AGSA, 2019).

The correctional objectives usually comprise of administrative corrections, vacancy filling and the enforcement of a *culture of consequence* (AGSA, 2019; Middleton, 2019). The outcome of the audit is summarised in Table 7 accompanied by the details behind the status (AGSA, 2020c).

Table 7. Contrasting Municipal Performances ***(source: AGSA, 2011)***

<table>
<tr><th colspan="2">Audit Status</th><th>Financial Statements</th><th>Reporting on Set Objectives</th><th>Compliance with the Legislation</th></tr>
<tr><td>Clean Audit</td><td>Clean Audit</td><td>Free from material misstatements.</td><td>No material findings</td><td>No material findings</td></tr>
<tr><td rowspan="4">Non-Clean Audited</td><td>Financially Unqualified</td><td>Free from material misstatements.</td><td>Possible material findings</td><td>Possible material findings</td></tr>
<tr><td>Qualified Audit</td><td colspan="3">Contain material misstatements in particular amounts, or there is insufficient evidence for AGSA to conclude that specific amounts included in the financial statements are not substantially misstated.</td></tr>
<tr><td>Adverse Audit</td><td colspan="3">Contain material misstatements that are not confined to specific amounts, or the misstatements represent a substantial portion of the financial statements</td></tr>
<tr><td>Disclaimer of Audit</td><td colspan="3">Provided inadequate evidence as documentation on which to base an audit opinion. The lack of adequate evidence is not limited to definite amounts or represents a considerable portion of the information in the financial statements.</td></tr>
</table>

2.1.3.2 Clean Audited Municipalities Vs Non-Clean Audited Municipalities Concluding

As shown in Table 7, clean audited municipalities conducted their daily operations as per the legislative framework setup. These regulations are summarised in the MFMA (2017) and include supply chain management guidelines that ensure fair, ethical and effective procurement practices are followed (Republic of South Africa, 2017; Hanks *et al.,* 2008). Therefore, it is acceptable to assume their collective contracting costs are to a certain degree, the true governmental market value for any required good or service.

The South African public procurement system is built on five core pillars under *section 65 (2) of the MFMA* (Republic of South Africa, 2017), these are explored further in the review. As stated by Munzhedzi (2016), if either of these pillars is disrupted during the procurement process, it compromises the integrity of the procurement structure. These interruptions lead to public procurement corruption and can be achieved through various methods as explored in the following subsection.

2.1.4 A Review of the Types of Public Procurement Corruption

Public procurement corruption as established in the opening chapter of this study is an issue currently plaguing the South African government (De Lange, 2011). As previously explored, this has led to an inflation of procurement costs in the public sector (Krappe and Kallayil, 2003; Munzhedzi, 2016; Child, 2018).

Table 8 depicts six types of public procurement corruption found. As observed, all six fraudulent procurement types have a direct impact on the pricing of procurement by inflated costs. Five out of the six methods, with the exclusion of "*collusion by bidders",* require the participation of a government official (buyer) for it to be achieved. This shows how an ethical supply chain management department would drastically decrease public procurement corruption.

Table 8. The Six Types of Procurement Frauds in the Public Sector ***(source: Nextenders, 2016)***

Type	Term	Definition
1	*Kickbacks & Corrupt Payments*	A promised bribe by a contractor to the buyer after they've received payment for the winning project. Since the contract is guaranteed the supplier can charge *an above market rate*.
2	*Corrupt Influence*	*Paying over market rates* by various means, including tailored specifications (the buyer creates specifications that only one supplier can meet, essentially monopolising the procurement).
3	*Collusion by Bidders*	A group of suppliers rotating available contracts to monopolise the field allowing them to dictate the market rate (price-fixing). Referred to as *cover quoting* in South Africa (Bizana *et al.,* 2015) and usually results in *inflated prices*.
4	*Billing Fraud*	Intentionally submitting false, *duplicate or inflated invoices* by a supplier, usually colluding with the buyer.
5	*Conflicts of Interest*	Where an employee purchases items through their private company and bills this to the state. This enables them to accept *above-market rates* for the procured items as they are to sign off.
6	*Delivery Fraud*	When a contractor submits a successful low bid (in collusion with a buyer) and subsequently submits further multiple variations to increase financial gain *ergo inflating the contract* (*scope creep*).

The various public procurement corruption methods lead to price inflation. This being the one common symptom that further substantiated the need for the study. To better understand how procurement price inflation has occurred in the real world, the below subsection is a survey of past cases focused on the Gauteng Province. This being the case as the non-clean audited municipalities in the Province is the focus of this research.

2.1.5 A Review of Price Inflation Cases in the Gauteng Region

The South African Public Procurement System has over the past years fallen victim to several corruption scandals (Managa, 2012; Pillay and Erasmus, 2015); amongst which are inflated contract costs. According to Brown (2016), the inflation of bidding prices have cost the South African taxpayer over R233 billion (*~US $16.3 billion*) annually. Municipalities in the Gauteng Province are a significant contributor to this amount (SABC News, 2014). This to the extent that it is viewed as the most corrupt region in

South Africa with over 45% of all the reported corruption cases in the country emanating from the province in 2018 (BusinessTech, 2019; Daniel, 2019).

To better explore this and how it relates to the study, some examples of price inflation cases published by investigative journalists are presented in Table 9. The table depicts the municipality in question, the article summary, the cost to state and the likely corruption type as reviewed in Table 8. These articles were cited to depict how price inflation, through tender corruption, takes place in the province.

The cases were selected based on their relation to the research topic as they are all price inflation cases from the Gauteng Province; their recent nature of publication as they were published within a year of the research being conducted and the financial *footprint* of these municipalities, as their combined expenditure accounts for above 40% of the national revenue for municipalities (AGSA, 2020b).

Table 9. Review of the Municipal Price Inflation Cases in the Gauteng Province

No.	Municipality	Brief Article Review	Amount	Type
1	*City of Joburg* by Modingoane & IAB (2019)	The City took advantage of an "Emergency Procurement" clause in which a municipality could procure certain goods and services without adhering to the correct procedures between 2014/2015 to 2017/2018 financial years.	Over R3.9 billion (~US $250 million)	**Type-1** and **Type-3**
2	City of Tshwane *by Moatshe* (2019)	The City had awarded a fuel contract to a supplier that allegedly inflated its prices. The supplier had links to senior officials in the city and the matter is currently still under investigation.	Undisclosed	**Type-1** and **Type-5**.
3	*City of Ekurhuleni By* Tabelo (2019)	The City ordered a tender for the leasing of portable toilets over three years. The project was said to be a "cash cow" and alleged that the awarded companies also had links to officials in the city who were hired as unofficial "*middlemen*" and rented on behalf of the municipality. This led to inflated prices.	R1.9 billion (~US. $133 million)	**Type-1** and **Type-5**.

2.1.5.1 Price Inflation Cases Concluding

There are various frameworks in place to deal with these types of cases. Among them is the *Irregular Expenditure Framework* to avail measures to be followed by accounting officers when dealing with irregular spending (National Treasury, 2019). Though, even with such frameworks in place, corruption by inflated bidding prices remains rampant in

South Africa, generally in the Gauteng Province as evident above. It should be noted, the explored municipalities in this subsection have had recurring non-clean audits over the past four financial years (AGSA, 2018, 2019, 2020a).

To have a credible and accurate comparison of the clean audited municipalities in South Africa and the non-clean audited municipalities in Gauteng, all the critical variables required to conduct a benchmarking exercise needed to be factored into the study. This is so the *ideal* benchmarking tool that will aid in realising the study objectives would be selected. This could only be attained by an in-depth assessment of the type of data the research would work with as explained in the next section of this review.

2.1.6 Benchmarking Municipalities Variables

There are various complexities involved in the pricing of procured goods and services (Pillay and Erasmus, 2015). Multiple factors need to be considered before a tender may be priced by a supplier among which include work specifications, investment risks and the number of competitors as identified by Stiti and Yape (2019). These factors can be broken down and grouped as follows as identified by the author in Table 10 and are elaborated in the following subsections:

Table 10. Variables Considered When Assessing the South African Public Procurement System

Term	Definition
Controlled Variables	These are the scope of work items disclosed in the procurement document and are controlled by the supplier such as the technical requirements and deliverables.
Uncontrolled Variables	These are the external factors outside the control of the supplier such as environmental issues.
Procurement Process Variables	These are the public procurement standards set up by the national government and are summarised in the five procurement pillars.

2.1.6.1 Controlled Variables (Scope of Work)

These variables are stated in the scope of work set by the municipality to the potential supplier. The variables differ from project to project. However, there are common elements that include the black economic empowerment (BEE) level, experience level, company size (SA Tenders, 2020). Though, according to Gumede (Gumede, 2020), it is a possibility that the BEE scoring system is inherently flawed. However, because of resource and time constraints, this concept has not been further explored in this study; only the supplier BEE scores impact on the pricing was assessed for this research.

2.1.6.2 Uncontrolled Variables (External Factors)

These variables are "*onsite issues*" that are frequently incurred once work has commenced and are unquantifiable and costly to incorporate into the study by the author. These include the geological factors, labour unions requirements, potential damages by criminal activities and logistical requirements (Davidson and Atkins, 2013). These external variables may have a significant impact on the final pricing (Achilles, 2020). However, based primarily on the lack of resources to analyse and quantify each of these individually, they were excluded from the study for the time being.

2.1.6.3 Procurement Processes Variables (Five Pillars of Procurement)

The South African public procurement system is built on five core pillars for procurement to be done effectively (Munzhedzi, 2016). Either of these pillars may have a significant impact on the final price obtained in the tendering processes. Not all these pillars were explored in the research. Table 11 depicts each of the pillars, their purpose, their impact on the pricing of a tender and the study purpose:

Table 11. The Five Pillars of Procurement and Their Impact on the Study

Pillars	Purpose	Impact of the Pillar On Pricing	Study Purpose
Value for Money	Ensures that the quality matches the prices (Horn and Raga, 2012).	Higher demand in the "value" of work usually increases the "money" to procure.	Irrelevant to the study because of data unavailability but briefly explored in the *conclusion*.
Open and Effective Competition	Ensures that the procurement process is transparent and is competitive (Horn and Raga, 2012).	Increase competitive bidding leads to a decrease in pricing and an increase in quality (Grega and Nemec, 2015).	Was comprised in the study as part of a *critical research limitation* that carries into *data modelling* and *analysis chapters*.
Ethics and Fair Dealings	Ensures the prevention of corrupt activities (Horn and Raga, 2012).	Failure to uphold this pillar leads to an increase in procurement costs (Grega and Nemec, 2015).	Was *used throughout the study* as this pillar assesses the likelihood of unethical conduct, this being price inflations for the study.
Account-ability and Reporting	Ensures that all actions are traceable to their source (Hawkins, Gravier and Powley, 2011).	No accountability leads to a decrease in operational efficiency and increased procurement costs (AGSA, 2019)	Was integrated into the study as part of the *data collection* as reporting is how the study archived data was accessed.
Equity	Assists previously disadvantaged persons (Econoserv, 2018).	Increase the cost of the contract as it encourages BEE compliance in all public acquisitions (Lowman, 2017).	Was incorporated in the study as part of the *data modelling, analysis and conclusion* chapters.

2.1.6.4 Benchmarking Municipalities Variables Summary

The study has focused on acquiring the relationship between the controlled variables and the price of the tender because of the resource and time constraints for this research. Four of the five procurement pillars are explored, Table 12 summarizes their relevance to the study. The value for money pillar was excluded from the research because of data unavailability and resource constraints. However, it was explored in the closing chapters to better appreciate its likely impact on the study outcomes.

Table 12. The Four of the Five Pillars of Procurement Relevant to the Study

Pillars	Relevance to the Study
Equity	Provides the independent variable of the Black Economic Empowerment (BEE) score from the suppliers.
Accountability and Reporting	All the data collected stemmed from archived data which form part of reporting in governmental entities.
Open and Effective Competition	Explored by using the findings by Stiti and Yape (2019) to ignore its effects for the duration of the study.
Ethics and Fair Dealings	The study purpose as it assesses the probability of unethical conduct from price inflations.

The key variables have been described for this study to develop a tool to identify or, at best, highlight procurements that likely contained pricing deviations. An assessment was conducted to determine where this tool would be implemented for the procurement process. The following section is a breakdown of the South African municipal procurement process.

2.1.7 Municipal Procurement Processes In the Republic of South Africa

The South African public procurement processes, though similar in many aspects, differs from one organisation body to another; as cited in Pooe *et al*. (2015, p. 68) *section 217(3) of the Constitution* (Republic of South Africa, 1996) states, each municipality has a right to develop its procurement policy and the national legislation must recommend a framework within which the policy may be implemented.

Municipalities, as previously mentioned, are governed by a framework known as the Municipal Finance Management Act (MFMA). This Act covers a range of municipal aspects among which is the Supply Chain Management (SCM) and how it should be undertaken. Below is a summary of how the standard procurement process of a municipal structure should be conducted.

2.1.7.1 Standard Municipal Procurement Process

All municipalities in South Africa have a generic prescribed framework for procurement as stated by Pooe *et al*. (2015). According to the structure, the processes of procuring differs according to the value of an item. Figure 3 is the model used for identifying which procurement process is relevant based on a particular item value. This research has focused on low valued procurements ranging from R30 000 to R200 000 that require a minimum of three bidders, advertised for seven days and applies the 80/20 BEE preferential point scoring system.

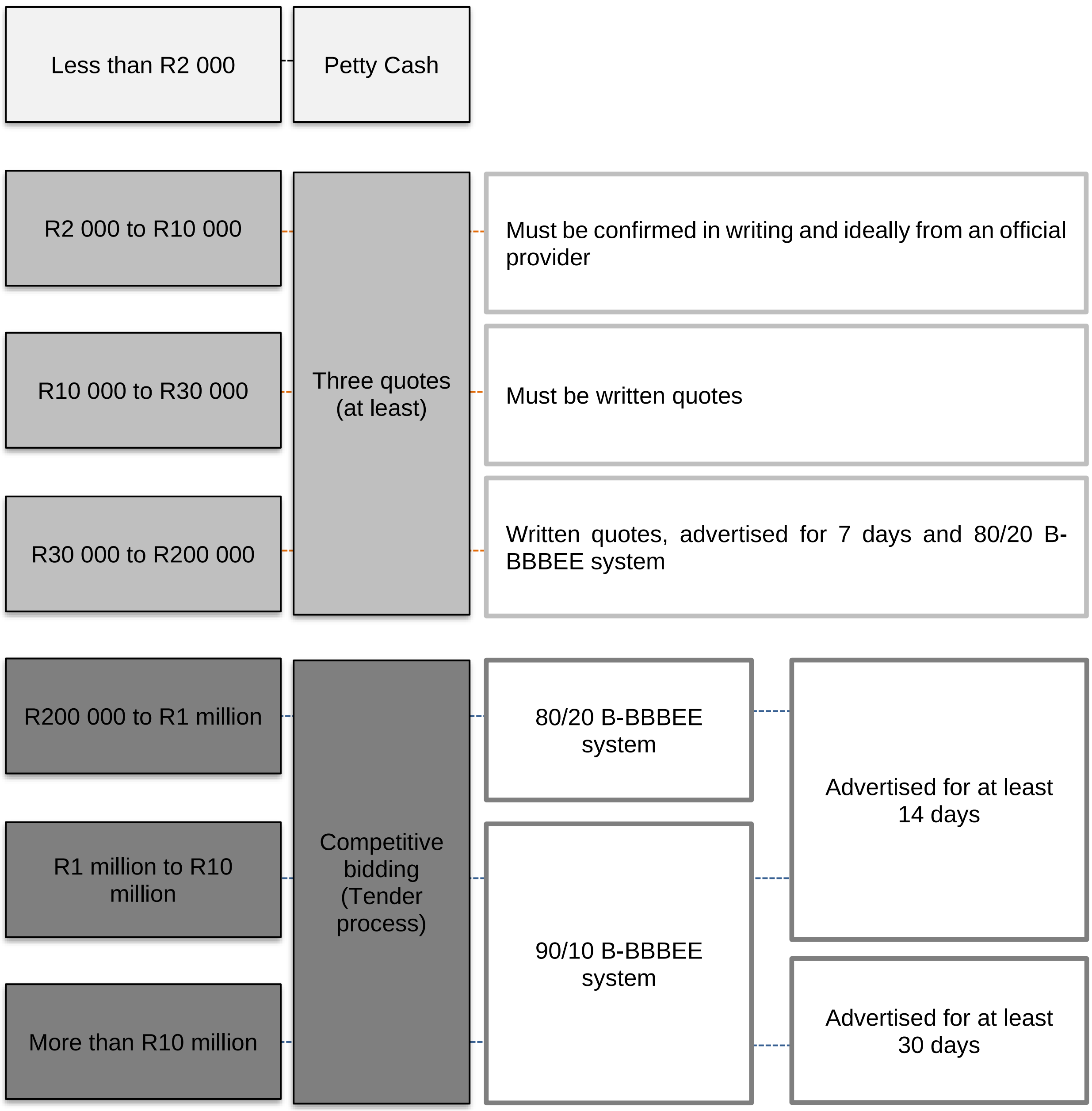


Figure 3. Procurement Method for Expected Contract Value ***(source: Dobie & Xinwa 2015)***

The below Figure 4 is a breakdown of all the procurement processes used by municipalities based on an example procurement for a printer by the town engineer followed by the elaboration of all steps involved in the process in Table 13 below:

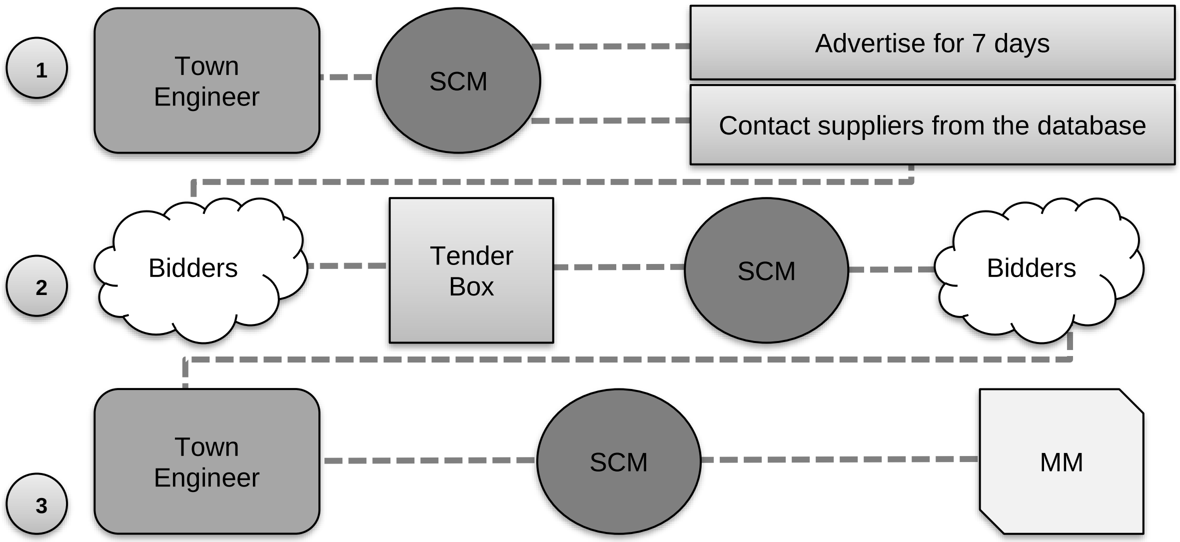


Figure 4. Example of the Municipal Procurement Process Value ***(source: Dobie & Xinwa, 2015)***

Table 13. Steps Involved in an Example Procurement ***(source: Dobie & Xinwa, 2015)***

Row	**Elaboration**
One	The town engineer gives the bid requirement to the Supply Chain Management (SCM) who will contact suppliers from the database and place the advertisement.
Two	The bidders must submit their bids in the tender box before the cut-off time.
	Official opening and a witnessing official must stamp and sign bids and logged in to a register.
	Bids must be opened in public directly after the cut off time and the supplier may be present.
	The names of the bidders and, if practical, the bid amounts must be read in public. This is an excellent opportunity to ensure transparency by attending.
Three	SCM does a check to see whether you have submitted all the right documents. You must comply otherwise your bid will be disqualified.
	The town engineer (or whoever needs the goods/services) does a technical evaluation (if there is a functionality criterion).
	SCM calculates scores based on functionality, price, and B-BBEE criteria.
	They give their written recommendation to the Municipal Manager, or another manager who has the delegated authority, to make the award.

2.1.7.2 Municipal Procurement Processes Concluding

As shown in Table 13 and Figure 4, the framework set up for the South African municipal procurement system factors in all five of the public procurement pillars explored earlier in the review. However, these are highly reliant on the integrity of all the personnel involved in the procurement process and can be evaded to benefit specific persons within the SCM process (Munzhedzi, 2016).

2.1.8 Thematic Review Summary

This Thematic Literature Review has qualitatively explored public procurement issues currently running rampant in South African municipalities (Managa, 2012; Pillay and Erasmus, 2015); mainly contract price inflations due to misconduct by elected officials, particularly in the Gauteng Province as shown in the articles reviewed in Section 2.1.4. The practice is going against the very *Constitution* the democratic state is built on.

There are various regulatory bodies in place to ensure that such unethical behaviour is identified and eradicated (Republic of South Africa, 1996). However, the practice of inflated procurement costs is still regularly observed in South African municipalities (Ntsaluba, 2019), and has led to a crippling of service delivery to local communities and bloated budgets. The Gauteng Province is the most challenging area in South Africa regarding corruption, more so public procurement (SABC News, 2014). Though, as explored, not all government municipalities have had critical reviews.

According to the AGSA (2019), some municipalities in the country have conducted their operations ethically, and by the set legislation. These are known as "*clean audited*". Non-clean audit refers to municipalities that have not performed the services as per the AGSA requirements. This established the basis of this study, a procurement comparison between clean audited municipalities in South Africa and non-clean audited municipalities in the Gauteng Province to identify possibly inflated procurement costs.

If successful, this research would have aided in adding a monitoring tool in the procurement process. It may be integrated into the procurement manager's office or during the SCM valuation to assess in actual time whether overcharging through price inflation is taking place. This was achieved by comparing the contracting costs with clean audited municipalities in South Africa against that of the particular item being required for non-clean audited municipalities in the Gauteng Province.

2.2 Theoretical Review

The theoretical review comprises two segments:

- First, is the selection of a benchmarking model which satisfies all the modelling requirements of a simplistic modelling approach explored in *chapter one,* *Significance of the Study,* Section 1.3 and aligns with the study objectives of having a quantitative outcome, Section 1.4.2.

- The second is an analysis of a past study into the likely errors that may have been encountered for the type of benchmarking methodology being undertaken. This is to better interpret the study outcomes as shown in section 3.4.2 when validating the model.

2.2.1 Methods of Benchmarking Municipalities

The items listed below are the *modelling requirements* for the benchmarking method:

1. Must be quantitative to express the modelling outcomes numerically for benchmarking.
2. Accommodating multiple input (independent) variables as procurement documents scope of works have many specifications unique to the contract type that impact the pricing.
3. Must have a simplistic approach to model building and be easily programmable because of limited computing resources.
4. Must have a *margin of error*; introduced to account for the *error element* presented by the previously mentioned unused variables or a quantitative method of measuring the modelling accuracy.

There are various quantitative benchmarking techniques, that satisfy the first modelling requirement, available to conduct a cost comparison study (Davidson and Atkins, 2013; Dittgen, 2020). However, only three were assessed for this study among them includes support vector machine (SVM), a popular predictive modelling technique used in financing (Henrique *et al.*, 2018).

This technique was disregarded beforehand as it is complex and mostly used in classification exercises that do not fit the quantitative approach for the study. The remaining two methods were found to be viable for a quantitative benchmarking assessment. The first being regression-based followed by data envelopment analysis.

2.2.1.1 Regression-Based Approach

The benchmarking study data had both independent (*specifications/scope of work*) and dependent (*pricing*) data as previously explored in the thematic exploration; this made a regression analysis possible (Montgomery and Runger, 2011). There are various regression techniques available, including simple/multiple linear, polynomial, ridge, elasticnet, lasso regression as explored by Dittgen (2020) and Seif (2018).

The author focused on two which were multiple linear regression (MLR) and polynomial regression. This is due to the fact that the remaining three were more complex and mainly used for exercises in which the size of the dataset is highly limited and non-conforming to basic regression assumptions such as non-multicollinearity and assumed homoscedasticity which was explored further in the study (Ismail *et al.,* 2009).

Polynomial was removed as a possible candidate because it is an immediate extension of MLR as it is meant for data that has a polynomial relationship rather than linear. However, according to Wilhelm (2016) and PennState (2019), nonlinear (polynomial) data can be transformed to be linear, ergo removing the need for the polynomial regressive technique. Therefore, MLR was selected as the regressive choice.

Multiple linear regression (MLR) is a model created for identifying the relationship between multiple independent (input) variables (X) and an output dependent variable (Y); this model assumes that the relationship between the variables is linear (Montgomery and Runger, 2011, p. 450). The formula is as follows (Ismail *et al*., 2009):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots\ldots + \beta_n X_n + \varepsilon$$

In the above equation, β_n are the coefficients acquired using the relevant formulas; β_0 is the y-axis intercept, ε is the random error term used to account for known and possibly unknown errors and last, n is the number of predictors.

2.2.1.2 Data Envelopment Analysis Approach

Data Envelopment Analysis (DEA) is a nonlinear variable analysis method that considers a set of weighted inputs and outputs (Klimberg and Lawrence, 2009); a mathematical programming-based method to assess the relative performance of businesses and used to assess technical efficiencies (AGPS, 1997; Santos, Ros and Cavique, 2011).

It is commonly used to identify the performance of governmental entities by grouping them into decision-making units (AGPS, 1997). According to Cooper *et al.* (2014), various models were developed in the past; though this study has focused on *the Cobb-Douglas Model.* The model is expressed in the function of a single output variable and two inputs; and is summarised below (Emrouznejad and Cabanda, 2015):

$$Q = ALaKb$$

Q is the desired output, and A is the derived constant. L is labour input and K is capital input. Last, a and b are the elastic respective outputs of labour and capital. DEA does not require any underlying assumptions for the inputs and output (Emrouznejad and Cabanda, 2015, p. 236); as compared against MLR, making it a more favourable modelling technique for this study. However, this is not the case as shown below.

2.2.1.3 Benchmarking Method Concluding

Table 14 has explored the reasons which have warranted the selection of the multiple linear regression modelling technique. The table is built around the benchmarking modelling aims of simplicity and compatibility with the multiple input data the study will work with. Additionally, the modelling accuracies had to be considered to know the level of confidence each technique has on the outcome.

Table 14. Data Envelopment Analysis Vs Multiple Linear Regression Summary

Modelling Requirement	**Data Envelopment Analysis (DEA)**	**Multiple Linear Regression (MLR)**
Compatibility with the Multiple Input Data	The *Cobb-Douglas Model* is not ideal for the research as it is restricted to two input variables. Cooper *et al*. (2014) did state that various other DEA models can be redesigned to fit modelling needs. Thus, DEA has no restrictions on the number of input variables.	MLR is only restricted by its underlying assumption of homogeneity, linearity and non-multicollinearity (Stiti and Yape, 2019) and no restriction on the number of input variables.
Simplicity in Modelling	According to Wojcik *et al*. (2019), a simple approach to DEA is problematic in non-production assessments. This makes it non-ideal for the modelling requirements of this study.	MLR was employed in various price prediction studies as explored by Gustafsson and Wogenius (2014). This was done without changes to the general equation and the simplistic solution aligns with the modelling aims.
Model Accuracy	According to Thanassoulis (1993), DEA has higher predictive accuracy than MLR. This being the desired outcome for the study as it would mean there is higher confidence in the results obtained.	Thanassoulis (1993) states that MLR provides greater stability in accuracy modelled than DEA. This is ideal for the research as a large instability in prediction accuracies would lead to faulty results when applying the model.

2.2.2 Past Work for Understanding Modelling Errors of Multiple Linear Regression

Modelling errors are a likelihood in most probability-based research and mitigation methods need to be explored beforehand (Frost, 2020b). This section is a brief survey of past research to understand modelling errors associated with MLR. This was achieved by analysing the past work of Klein and Rossin (1999).

The research by Klein and Rossin focused on understanding the effects of data quality of linear regression models used in the prediction of net asset values. In the study, it was found that the magnitude of errors ranging from 5% to 10% for the data used to train a model had a positive effect on the predictive accuracy (an error rate of 5% to 20%). Whereas, the magnitude of errors ranging from 5% to 10% for the data used to test a model harmed the predictive accuracy (an error rate of 25% to 100%). The findings are explored further when analysing error elements of both the model training and testing data samples in the study and applied when validating the model in chapter three, section 3.4.2.

2.2.3 Theoretical Review Summary

MLR better suited than DEA

This theoretical review was primarily based on selecting the ideal municipal procurement benchmarking model. As recognised, the multiple linear regression (MLR) modelling approach is best suited to data envelopment analysis (DEA) (Wojcik, Dyckhoff and Clermont, 2019) for benchmarking in the type of research being conducted, as it satisfies the multiple input variables and simplicity requirements explored earlier in chapter one, and it provides greater stability in accuracy modelled than data envelopment analysis (Thanassoulis, 1993).

Errors Associated with MLR

MLR on price prediction based on procurement is not without the risk of error as explored by Klein and Rossin (1999). Their research revealed errors in training data have a positive impact on predictive accuracy, whereas errors in testing data have a negative effect. Even so, there are various mitigation methods available to authenticate and validate such models (Montgomery and Runger, 2011; Stiti and Yape, 2019). These mitigation and validation techniques available for MLR models are explored as part of the methodology chapter in sections 3.1.4 and 3.1.5.

Chapter Three: Methodology

This study aimed to develop a benchmarking model to identify possible cases of price inflations that may have occurred in recurrent non-clean audited municipalities in the Gauteng Province. This was achieved by comparing their procurement costs to those of clean audited municipalities in South Africa, as stated by the critical research question (CRQ). A hypothesis (H_1) and null hypothesis (H_0) were derived from the CRQ and tested using a statistical regression function known as multiple linear regression:

H_0: *Clean Audited Municipalities in South Africa Do Not Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng*

H_1: *Clean Audited Municipalities in South Africa Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng*

The chapter has integrated the logical scholarly layout used by Stiti and Yape (2019) for building an MLR price prediction model:

- The first section is an empirical framework exploration of the basics of multiple linear regression and the model validation techniques available.
- Second, the data management plan included the selection of a commonly procured item that the research was modelled on. Afterwards the relevant data collection preparation and storage procedures.
- Third, is the preparation of the model design, which includes the removal of non-significant variables and testing for MLR model building assumptions.
- Fourth is the construction of the model and validation thereof. Thereafter exporting the derived equation for benchmarking to identifying municipalities that may have overpaid on past procurements and by what margin.

Figure 5 is a depiction of the overall research methodology layout. It describes all the relevant stages of the model building process and the target outcomes. These were used to map out the methodology chapter and will be used to signpost the remaining chapters as they all follow an identical structure mirroring this chapter. As explained below, each subchapter is indicated by the relevant heading number that may be followed to clear up any ambiguities and the detailed definitions of all the mentioned dataset is available in Table 27.

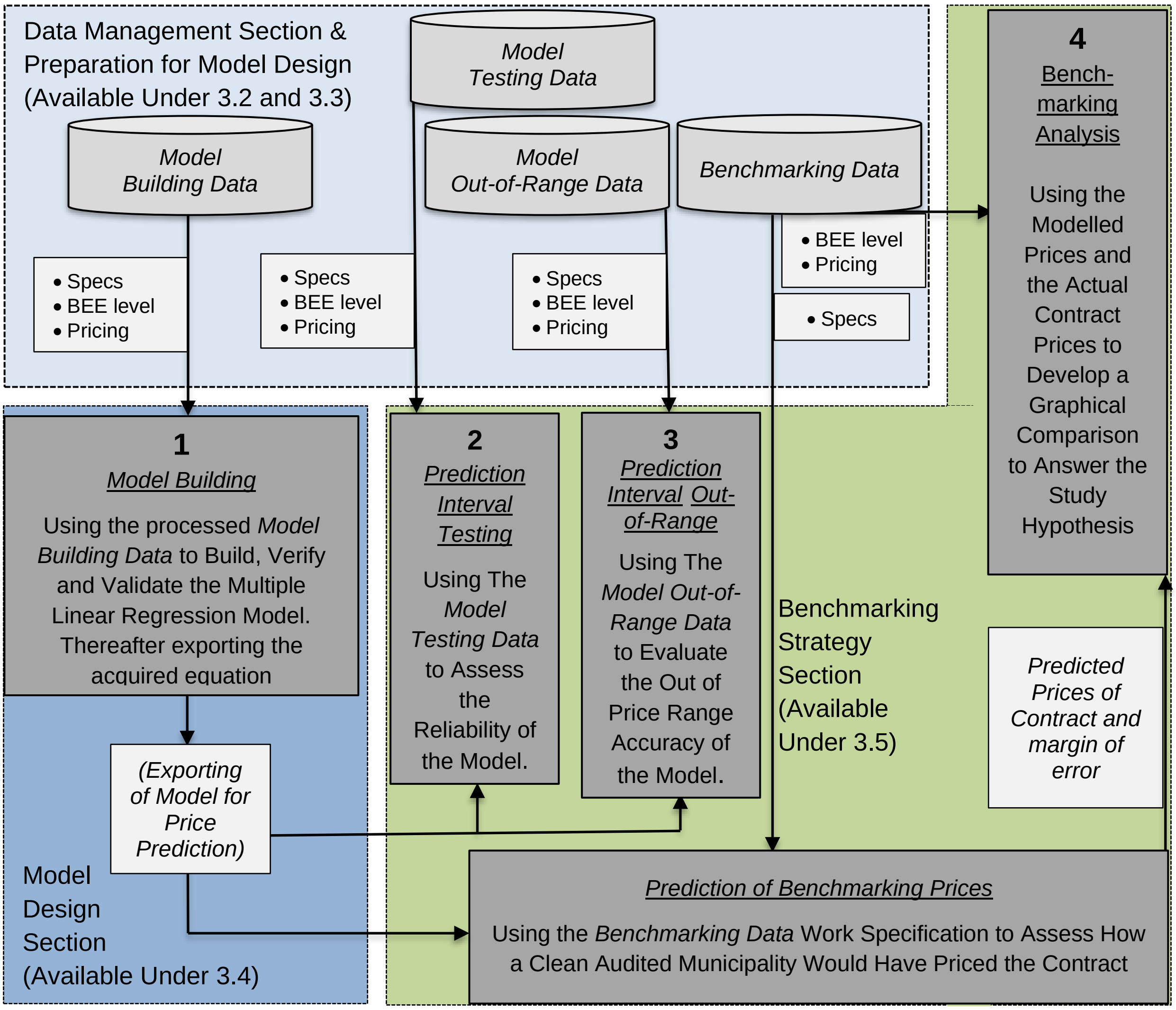


Figure 5. Research Model Design for Benchmarking Summary

All the collected and processed data is from clean audited municipalities, excluding the *benchmarking data*. All datasets contain specifications, BEE score and awarded contract value. The below list depicts the chronological sequence as numbered above:

1. *Model building data* is used to derive an equation (inputs as specs) that expresses how clean audited municipalities price (output) their goods and services procured.
2. *Model testing data* is used to confirm the equation's accuracy within the specified price range.
3. *Model out-of-range data* is then used to assess if the equation can be used for contract values above the study R 200 000 (excl. vat) cap.
4. Using the equation to predict what the non-clean audited municipalities (*benchmarking data*) pricing would be if it were issued by a clean audited municipality and comparing it to the actual awarded price.

3.1 Empirical Framework Exploration

3.1.1 Multiple Linear Regression Analysis

Linear Regression (LR) is a modelling technique used to acquire the linear relationship between a single independent (input) and dependent variable (output), and is summarised in the equation shown below (Montgomery and Runger, 2011, p. 407; Stiti and Yape, 2019). In the below equation: Y is the targeted outcome; β_0 is the intercept; β_1 are the regression coefficients; X_1 the required independent variable, also known as a predictor or explanatory variable, and ε the random error term.

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad \text{(Equation 1)}$$

However, LR would not work as it was established in the previous chapter; multiple linear regression is the ideal method to apply because this study is based on finding the relationship between numerous independent variables and a singular dependent output variable. As identified in Montgomery and Runger (2011, p. 451), the basic MLR formula is described by the following notation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots\ldots + \beta_k X_k + \varepsilon \quad \text{(Equation 2)}$$

Where Y is the outcome (dependent variable); β_0 is the y-axis intercept; β_k are the regression coefficients; X_k the required inputs (independent variables) and ε the random error term; which accounts for the known and unknown errors present in the model.

3.1.2 Limitations/Assumption Of Multiple Linear Regression

As with most computerised modelling techniques, MLR has its limitations. Table 15 is a summary of the limitations of the model, as identified by Ismail *et al.* (2009). Thereafter, an elaboration of each limitation and the remedial steps taken.

Table 15. MLR Limitations/Assumption ***(source: Statistical Solutions, 2020c)***

No.	Limitations/Assumption	Definition
1	*Linearity*	The relationship between either of the independent variables and the dependent variable must be close to linear.
2	*Multicollinearity*	The relationship between either of the independent variables themselves must not have a high correlation.
3	*Homoscedasticity*	The variance of error terms must be similar across the values of the independent variables, this is known as homoscedasticity.

First Limitation/Assumption of Linearity:

The relationship between either of the independent variables and the dependent variable must be close to linear (Stiti and Yape, 2019). A method of assessing the linearity between the variables is by using a scatter plot for a visual assessment

(Statistical Solutions, 2020b); or the use of *Pearson's Correlation Coefficient* (**r**) which is a statistical tool that measures the linear correlation of a data set; it ranges from $-1 \leq \mathbf{r} \leq 1$. Table 16 below shows the degree of correlation according to the values obtained:

Table 16. Ranking Scale for the Degree of Correlation ***(source: Statistical Solutions, 2020a)***

Degree of Correlation	**Explained**
Perfect	The value is near ± 1, (nearly 100% correlation)
High degree	The value lies between ± 0.50 and ± 1, (50% to 100% correlation)
Moderate degree:	The value lies between ± 0.30 and ± 0.49 (30% to 49% correlation)
Low degree	The value lies below ± 0.29 (0% to 29% correlation)
No correlation	The value is zero

If a moderate or above degree is not obtained, data transformation may take place to improve the linearity of the relationship (Wilhelm, 2016; PennState, 2019). This concept is explored further in the chapter. However, it may happen that a variable does not have a clearly defined relationship (linear or nonlinear) with the dependent variable. Though, from a logical perspective, the variable should have an effect (e.g. BEE score). This is owing to the fact that the variable has a collaborative effect with another variable and as a result, cannot stand alone; a phenomenon that is explored further in the chapter.

Second Limitation/Assumption Multicollinearity:

The relationship between either of the independent variables must not have a high correlation. Should this be the case, a statistical issue is known as multicollinearity is introduced to a model built (Stiti and Yape, 2019). This may lead to highly inaccurate modelling outcomes (Shah, 2009). There are various methods available to identify whether multicollinearity is present in a model. Table 17 depicts seven methods that were suitable to be used in this study. The first and seventh methods listed were selected as the multicollinearity assessment tools for this study because of their modelling simplicity and are widely preferred for statistical analytics and robustness, respectively (Grace-Martin, 2019).

The first method is the observation of the correlation between multiple variables aided by the *Pearson's Correlation Coefficient* (**r**) for every independent variable pairing, and the second is the calculation of the variance inflation factor (VIF) for each independent variable. Those found to have a high correlation [greater than± 0.50] or VIF [greater than 5] (Statistical Solutions, 2020c), one must be removed based on their significance.

Table 17. Methods to Identify Multicollinearity in a Model ***(sources: Grace-Martin, 2019).***

Type	Explanation
1	This method is based on the correlation between independent variables. If it's significantly high, the rule of thumb says you have multicollinearity.
2	If standard errors are magnitudes higher than their coefficients (**β**), multicollinearity is present.
3	The overall model is significant, but none of the coefficients are. Meaning if the **P**-values for the coefficients are above the set standard for the study, multicollinearity is present.
4	If the predictors are completely independent of each other, their coefficients won't change when you add or remove one. Should this not be the case, multicollinearity is present.
5	The coefficients have signs opposite is expected from theory, multicollinearity is present.
6	If the coefficients on different samples are not similar, multicollinearity may be present.
7	High variance inflation factor (VIF) is a direct measure of how much the variance of the coefficient is being inflated because of multicollinearity.

Third Limitation/Assumption of Homoscedasticity:

The variance of error terms must be similar across the values of the independent variables, this is known as homoscedasticity (Manyathi, 2019). A plot of standardised residuals (Statistical Solutions, 2020c) can show whether points are equally distributed across all values of the independent variables. As demonstrated in Figure 6 homoscedasticity has a more rectangular distributed plot whereas non-homoscedasticity (heteroscedasticity) has a cone-shaped plot. A solution may be nonlinear data transformation (Glen, 2015) or the use of weighted least squares regression, but this would require several assumptions (Statistics Solutions, 2013).

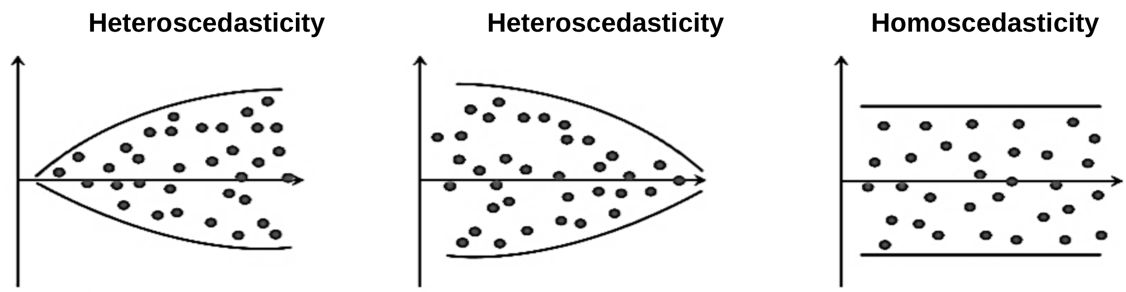


Figure 6. Heteroscedasticity vs Homoscedasticity ***(source: Statistical Solutions, 2020b)***

3.1.3 Calculation Of Coefficients

Squares is a statistical method of measuring a model's error by assessing the sum of squared errors, which is the vertical deviations from each sample data point to the modelled line (Ray, 2015), least squares is a method used to minimise this error. The method is further elaborated by Montgomery and Runger (2011) as used to estimate the regression coefficients ($\boldsymbol{\beta}_k$) in an MLR model. As cited in Stiti and Yape (2019), it is ideal for expressing the data available in the regression equation with the below-notated matrix form when calculating squares:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$ (Equation 3)

Where:

$$\boldsymbol{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$ (Equation 4)

In the above equation; n and k represent the observations (sample size of past bids received) and the regressor variables (identified number of relevant items in the scope of work that influence pricing) respectively (Montgomery and Runger, 2011). As stated in Montgomery and Runger (2011); $\boldsymbol{y}$ is an (*n x 1*) vector of the observations. $\boldsymbol{X}$ is an (*n x p*) matrix of the levels of the independent variables, **β** is a (p *x 1*) [where *p = k + 1*] vector of the regression coefficients and **ε** is a (*n x 1*) vector of random errors. The below equation is used in order to find the *Least Squares Estimate* of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$$ (Equation 5)

3.1.4 Regression Model Verification Tools

The likelihood of a multiple linear regression model being perfect is improbable (Montgomery and Runger, 2011). Below is the notation for calculating the error sum of squares (SS_E) which is a mathematical expression of squaring the deviations between the modelled outcome (*predicted value*) [$\hat{y}_i$] and the true outcome (actual value) [y_i] followed by the summation of all the squared deviations.

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$ (Equation 6)

The error equation is used to acquire the total sum of squares (SS_T). This is the maximum error an MLR model can have. The formula below is a mathematical expression of squaring the deviations between the mean outcome [$\bar{y}_i$] and the true outcome (*actual value*) [y_i] followed by the summation of all the squared deviations.

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$$ (Equation 7)

Lastly is the formula of regression sum of squares (SS_R), a mathematical expression of squaring the deviations between the mean outcome [$\bar{y}_i$] and the modelled outcome (predicted value) [$\hat{y}_i$] followed by the summation of all the squared deviations. Alternatively, can be obtained by subtracting SS_E from SS_T.

$$S_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$$ (Equation 8)

The resultant equation is the *analysis of variance identity* in the equation below. SS_T is constant for a model as it is the total sum of squares and the goal of any model is to minimise SS_E ergo maximise SS_R because of them being inversely proportionate.

$$SS_T = SS_R + SS_E \quad \text{(Equation 9)}$$

The above calculations were used to develop the two primary regression model verification tools in this study. These include the *Coefficient of Multiple Determination* and the *Test for Significance of Regression* and are explored further on in the chapter.

3.1.4.1 Alpha (α) Parameter Defined

Alpha (α) is a value (ranges from 0 to 1) that is chosen by a researcher thereafter used to assess whether a test statistic is statistically significant (Lavrakas, 2011). Alpha represents an acceptable probability of a Type I error in a statistical test; rejecting a true *null hypothesis* (Hughes and Fisher, 2020).

According to Montgomery and Runger (2011, p. 251); the ideal **α**-values are 0.10 (10%), 0.05 (5%) or 0.01 (1%). This study has chosen the highest α-value of 0.1 (10%) from the above-listed options. This as smaller alpha values increase the likelihood of *not* rejecting the null hypothesis (Everitt and Skrondal, 2010). As the study wished to find a distinction between procurements from clean and non-clean audited municipalities, the rejection of the null hypothesis was favourable at the α-value of 0.1.

3.1.4.2 Coefficient of Multiple Determination

The *coefficient of multiple determination* is a figure used to assess the fit of the model to the sample data given, by utilising the previously explored calculated errors and ranges between $0 \leq \mathbf{R^2} \leq 1$ where a higher value indicates the model's accuracy (Goeppner, 2008; Montgomery and Runger, 2011). However, the coefficient can be manipulated by adding more independent variables to the model being built to increase it regardless of the predictive accuracy of the model (Hughes and Fisher, 2020).

$$\boldsymbol{R^2} = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad \text{(Equation 10)}$$

A mitigation technique for the manipulation is by using adjusted $\mathbf{R^2}$. Adjusted $\mathbf{R^2}$ is an easy way to guard against including many independent variables by penalising the model builder for including excessive independent variables, it ranges from 0 ≤ adjusted $\mathbf{R^2} \leq 1$ (Moore, 2003). There is no ideal adjusted $\mathbf{R^2}$ value as this value changes according to the study aims. Yet a value of ±0.5 is viewed as *a high degree* ((Statistical Solutions, 2020b); therefore, a suitable standard for the study.

$$\boldsymbol{R^2_{adj}} = \frac{SS_R/(k)}{SS_T/(n-1)} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} \qquad \text{(Equation 11)}$$

3.1.4.3 Test for Significance of Regression

A model may give us an estimate to establish a close enough linear relationship between the dependent variable and the independent variables. However, it cannot prove the reliability of these relationships (Aggarwal, 2020). A method of testing a regression model's adequacy is by conducting a *hypothesis test* by using **F-**statistic and the **P-**value obtained by calculation or simply given by an ANOVA table (Shah, 2009; Montgomery and Runger, 2011).

$$S_E = \frac{\boldsymbol{SS_E}}{n-1} \qquad \text{(Equation 12)}$$

The above equation is used to calculate the mean square error (MSE) which is obtained through dividing the error sum of squares (*SSE*) by the total *degrees of freedom* and shall calculate **F-**statistic as explained in the equation below. **F-**statistic is a general regression test used to assess a hypothesis (Montgomery and Runger, 2011). The two notations below express the theorised scenario and the equation for **F-**statistic. This process can be automated by making use of the relevant software packages such as Microsoft Office Excel and IBM SPSS

$$H_1{:}\, \boldsymbol{\beta_1} \neq 0 \ (where\ \boldsymbol{\beta_1}\ are\ the\ coefficient\ from\ \boldsymbol{\beta_1}\ to\ \boldsymbol{\beta_k}) \qquad \text{(Equation 13)}$$

$$F_0 = \frac{SS_R(\beta_1|\beta_2)}{MS_E} \qquad \text{(Equation 14)}$$

To assess whether the calculated **F-**statistic proves we can reject the null hypothesis, an F-test must be undertaken by comparing it to the *critical **F-**value*. The critical **F-**value can be acquired from a relevant statistical F-table by having the *study alpha value*, *numerator degrees of freedom* (no. input variables) and the *denominator degrees of freedom* [sample size minus (no. input variables + 1)]. If the value of **F-**statistic is greater than the critical **F-**value, the null hypothesis is rejected, and it is proof that at least one of the coefficients of the independent variable ($\boldsymbol{\beta_1}$) is useful for predicting the outcome. This being one of the study *quality safeguards.*

3.1.5 K-Fold Cross Validation Method

K-Fold Cross-Validation is a statistical technique that involves partitioning the data into subsets, training the model on a portion of the subsets and using the rest to test the model's performance (Krishni, 2018; Brownlee, 2019). This validation approach is conducted to ensure the model has a robust build, and the data used did not give a low test error simply through *blind luck* (Quantcademy, 2020). The steps to conducting a *K-Fold Cross-Validation* as cited in Ray (2018) is:

1. Randomly split your dataset into a ***k*** number of folds.
2. For each k-fold in your dataset, build your model on k-1 folds of the dataset. Then, test the model using the k^{th} fold.
3. Record the *mean squared error* (MSE) you see on each of the forecasts.
4. Repeat this until each of the k-folds has served as the test set.
5. The average of your k recorded errors obtained using the below equation is called the cross-validation error and will serve as your performance metric for the model.

$$\frac{1}{k}\sum_{i=1}^{k} MS_{E_i} \qquad \text{(Equation 15)}$$

3.1.6 Prediction Interval Of Multiple Linear Regression

Intervals are a range of figures where there's a high certainty the real value lies to assess a regression model. The most popular of these include the prediction interval and the confidence interval. The confidence interval is where there is a high certainty the true regression mean lies. Whereas the prediction interval is where there's a high certainty the true individual future value is (Montgomery and Runger, 2011).

However, as reasoned by Hyndman (2014) because of the confidence interval only assessing the *true mean* value, it is much narrower than the prediction interval, and as a result likely to give an exaggerated sense of accuracy in predicting. For this reason, the prediction interval is the *interval method* for this research.

A prediction interval is a tool used to express the error in predicting a future observation from the distribution at point X_0 (Montgomery and Runger, 2011). It is a range of figures that likely contain the true value of the dependent variable for a single assessment given specific values of the independent variables (Frost, 2020a).

$$\hat{y}_0 - t_{\frac{\alpha}{2},n-p}\sqrt{\left(\frac{SS_E}{n-1}\right)^2\left(1 + \mathbf{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x_0}\right)} \le Y_0 \le \hat{y}_0 + t_{\frac{\alpha}{2},n-p}\sqrt{\left(\frac{SS_E}{n-1}\right)^2\left(1 + \mathbf{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x_0}\right)} \qquad \text{(Equation 16)}$$

The prediction interval accounts for the uncertainty in estimating the population mean and the random variants of the individual values (Graphpad, 2019). It is expressed by the above equation, where the *t*-distribution is based on the desired confidence level 100(1 - α)%.

3.2 Data Management

This portion of the methodology explores the entire research data system, consisting of data collection and preparation. The logical order followed in the section is based on chapter two of the *Data Science Process* by Manning Publications (2020), in which the topic of interest is broken down into three sections that include data collection, cleansing and transformation.

3.2.1 Selection of Contract Type for Benchmarking

This research is restricted to a single commonly procured low-valued, non-competitive item. As a result, preparation work was required to identify such a candidate to base the entire research on. This was carried out by hand using the *convenience sampling* method on clean audited municipalities over the past 5 years concluding on 1st September 2019 so to begin the official data collection for the study. The *convenience sampling* method is based on easily accessible data to acquire an estimate (Gaurav, 2017). It was conducted by examining clean audited municipal websites to assess all publicly available data of past awarded low-valued procurements. Thereafter, the most common contracts amongst them were tallied up, marked as Appendix B, Section 1.

The outcome showed the three most common low-cost procurements among the municipalities including Matzikama, Witzenberg, Cape Agulhas, Hessequa, Midvaal and Nkangala; were *fencing supply and installation* (27 contracts), *furniture supply* (9 contracts) and *catering services* (6 contracts). Further analysis was required of the above contract types based on the below-listed items in Table 18 developed with the help of *Data Science Process* by Manning Publications (2020) to identify the idyllic contract type that will meet the ideal data sample requirements.

Table 18. Requirements for Ideal Sample Data

No.	Requirement for Sample Data
1	It must be easy to distinguish between significant and insignificant variables.
2	It must not have many significant variables (e.g. bulky specifications).
3	It must have a probable big sample size (quantified further in the section).

The *catering services* contract type is an undesirable candidate because of detailed dietary requirements and the disparity in catering work specifications by different municipalities. As some may include decoration, cutlery and servers whereas other municipalities may not. This violates the second requirement as tabled above.

The *furniture supply* contract type was disregarded because of the likelihood of having many significant variables due to detailed specifications required in furniture contracts. This may include stipulations such as type of wood, general variation in furniture style and office sizes. This violates the first and second requirements as tabled above.

The *fencing supply and installation* contract type was the selected study focus low-cost procurement. This is due to it having easily identifiable significant variables (e.g. fence lengths), limited significant variables because of industry standards and probable big sample size based on the pre-feasibility work as the fencing procurements are 64% of the total commonly procured items. Subsequently, the factors influencing the final price of a fencing contract had to be explored to assess which scope of work variables will be likely significant as shown in Table 19.

Table 19. Variables To Factor Into Pricing Of Fencing ***(source: Tim, 2016; Gibson, 2019)***

No.	Variable
1	The removal of previous structures.
2	Type of material used, the height, the length and the finishing of the fencing.
3	The spacing between poles, the type of foundation required, and labour needed.
4	Whether items such as gating or fencing accessories (e.g. barbed wire) will be required.
5	In South Africa, the supplier Black Economic Empowerment (BEE) level (Bolton, 2006).

3.2.2 Fencing Supply and Installation Data Selection Criteria

The study focused on *fencing supply and installation* contracts from the municipalities of interest that is explored further in the chapter. Possible significant independent variables needed to be selected from the scope of work. The *convenience sampling* method was again used to assess past procurement documents to identify significant specifications and cross-reference with those recognised by industry specialists for fencing erection; Gibson (2019) and Tim (2016) as cited in Table 19. This was done to evaluate which specifications have most likely affected the final price of this type of contract. The outcome for this exercise is presented in Table 20 and was used as a checklist for the study to judge whether a contract can or cannot be used. It should be noted, the items listed below are *potential* input variables.

Table 20. Identified Compulsory Variables from the Fencing Scope of Work

Compulsory Items	Elaborated/Reason
Fence Removal [m²]	Only removal of old structures has been assessed and not the removal of rubble. The reason being a fence cannot be erected on non-cleared land so the element is assumed to be factored in by the fence length.
Fence Height [m]	Fence height must be specified as it can vary which will impact the price of the contract when assessing items such as the cost of the material.
Fence Length [m]	Similar reasons as stated for fence height.
Fence Material [No Unit]	This is the material used to erect the fence (e.g. steel mesh or concrete). This undoubtedly has played an important role in the final price when assessing items such as material cost.
Gate Length [m]	The gate length played an important role as some contracts require more gating structures erection than fencing and as a result, increase contracting cost disproportionately to other contracts.
Gate Height [m]	Some contracts may have varying gate heights for different areas which amount to additional material costs.
Gate Type [No Unit]	The type of gate, whether it is a (swinging) hinged or a sliding gate; as theory suggests a hinged gate may have fewer costs involved (Howmuch, 2020).
Gate Area Material [No Unit]	This is the material used to cover the surface area of the gate to have a rough estimate of the cost of the gate.
Supply and Erect Contract [No Unit]	The contractor must have both supplied and erected the structure if different suppliers were used the contracts shall then be merged where possible.
Gate Motor [Numerical]	This is an additional requirement to the gate which may impact the final price of the contract.
BEE level [numerical]	The BEE rating of the company impacts the final price of the awarded contract as a higher BEE score increases prices as explored in earlier chapters.
<R200 000 Contract Value	Contracts lower than R230 000 (incl. 15% VAT) before transforming to present value are regarded as non-competitive and low-cost.
Date of Procurement	This is the date the bidding process was closed so to transform the data accordingly using the consumer price index (CPI).

Table 20 shows the possible independent variables to be included in the samples collected. However, more variables were excluded by the author because it risks having many independent variables. As a result, these were neglected based on existing industry practice guidelines as shown in Table 21.

Table 21. Neglected Variables from the Fencing Scope of Work

Disregarded Items	Elaborated/Reason
Pole Type, Spacing & Number of Poles	There are industry standards available for pole spacing depending on the type of fence and height such as that of palisades by the South African Bureau of Standards [SABS] (2004). These help in assuming the pole spacing has been accounted for by the type of material and height of the fence.
Hole Depth and Concrete mixture	There are industry standards available for hole depth depending on the type of fence and concrete mixture strength (SABS, 2004).
Compulsory Add-ons	It is assumed that based on the type of material certain compulsory add-ons have been factored in (e.g. nuts and bolts for panels and tie wire for mesh).
Labour	It is assumed labour costs have been diluted into all the compulsory items such as the fence material and length.

3.2.3 Data Collection

The population under consideration is *secondary data* in government archives from clean audited municipalities in South Africa and non-clean audited municipalities in the Gauteng Province. Two sampling methods were relevant to the study. These include the *systematic sampling method*; a regular interval sampling from a list; and the *quota sampling method (QSM)*; a nonprobability sampling technique in where the researcher selects the population-based on certain desired features (Gaurav, 2017).

However, before a method could be selected, the three sampling parameters required acknowledgement as listed in Table 22. These parameters were noted for the critiquing of the modelling outcomes. The ideal sampling method based on the study objectives is the *QSM* as the population is based on the desired audit outcome.

Table 22. Three Parameters Required When Constructing a Sample ***(source: Gaurav, 2017)***

Parameter	Definition
Consistency	The sample should have a level of consistency over changes across time and space where any change in the data reflects real-world changes (e.g. regressing economy must be depicted by a lowering of contract values).
Diversity	The sample should have a high level of diversity to avoid biases in the data collected (e.g. data collected should be evenly distributed between provinces and years).
Transparency	The sample data collection constraints must be mentioned beforehand so any biases and assumptions are understood and the data collected can be viewed from the right perspective.

Although it should be noted QSM may sometimes yield highly unreliable results as stated by Aprameya (2016). This method relies on the researcher's judgement in choosing the right subgroups, yet their preferences may skew the sample data and give a misleading representation. However, this drawback is examined in chapter five by the analysis of the three sampling parameters.

The data obtained from some municipalities had not been completed (predominately from non-clean audited). Items such as BEE level, fencing heights and overall fence length were excluded. As a result, additional logical data sorting assumptions were taken. This is to ensure that the sample size to build the model and benchmarking thereof was adequate and good data had not been discarded. These are depicted in Table 23. Though some assumptions were unique to specific procurements, these are noted in the opening of chapter four.

Table 23. Generic Data Sorting Assumptions for the Study

Item	Assumption	Justification
Assumptions on Cost	All procurement costs are inclusive of vat if not specified.	As these are awarded amounts, it is assumed the supplier included VAT.
	Combining the prices for separate procurements is the accurate representation of a complete procurement (e.g. fencing supply plus erection gives total contract value).	Fencing erection is a two-part process that requires the purchasing of material and installation thereof.
	The closing date for the contract advert is used to convert the cost using CPI.	The suppliers pricing is based on the month the tender closes.
	If unavailable, it is assumed the BEE score is 20 to not affect the price and assumed BEE levels are stated in the Appendix by the use of “#”.	85% of contracts in this study had a BEE score of 20.
Assumptions on Fencing Structure	For barbed or electrified fences, the overall height includes these accessories.	All these were made in favour of simplifying the data collection and processing stages. There is no official justification for these.
	Partial and full fencing repair procurements are viewed as a *supply and installation.*	
	Pedestrian gates are all hinged unless stated otherwise (*swinging*).	
	All fence and gate heights are equal unless stated otherwise.	
	If the overall length is not specified, the summation of the total panel lengths for fencing material gives a good estimation.	

3.2.3.1 Clean Audited Data Collection

Clean audited municipalities in South Africa were selected for the regression model building as they were assumed to have the true market value for fencing procurements. The awarded municipal prices for *supply and installation of fencing contracts*, BEE level of supplier and the work specifications (SOW) for previously awarded low-valued procurements from 2014/2015 to 2019/2020 (cut-off date 1 March 2020) financial years were obtained from government websites and by formal request to the municipalities.

Table 24. Minimum Sample Size Guide for Model Building Data ***(source: Alshibly, 2018)***

Number of Independent Variables (k)	Calculated Minimum Sample Size (MSS)	Minimum Sample Size to satisfy 10-fold cross-validation
1	58	50 or 60
2	66	60 or 70
3	74	70 or 80
4	82	80 or 90
5	90	90

The minimum sample size (MSS) used in this research was obtained using the above Table 24 and only applied to the *model building data* from the clean audited municipalities. The MSS was maintained through to the data sorting and model building stages. Included in the table is the *minimum sample size to satisfy* the *10-fold cross-validation* technique, a concept that is explored further in the chapter.

The start-off year for data collection was selected because of an act that mandates municipalities to release the details of every contract awarded, effective 2016 (National Treasury, 2016). Some municipalities applied the regulation to their previous financial year's procurements. Therefore, the data collection started from the 2014/2015 financial year and conclude in the 2018/2019 financial year (5-year period).

The study assumed that municipalities that have had the majority clean audit outcomes in the past five years *(+60% clean audit*) were observed as *clean audited,* regardless of when the clean audits were obtained. This is based on there being a demand by the model for a large sample size as illustrated in Table 24 and a majority clean audit in the past financial years shows the dependability of the data.

The data collected ended on 1 March 2020, because of the national government fiscal year-end; and the audit results conclude in the municipal financial year of 30 June 2019. As a result, it is assumed the *clean audited* status of the municipalities has remained unchanged for the outstanding months from the last recorded clean audit to the closing date for data collection. This is to ensure that data collected sufficed to construct a credible multiple linear regression model.

Where information was not freely available, formal requests were sent to the relevant municipalities through the aid of a PAIA form (promotion of access to information). A PAIA is a form that mandates governmental bodies to disclose information that is or was available to the public (City of Joburg, 2000). Thereafter, procurement personnel were contacted after the relevant ethical clearances had been granted. All data collected was captured and categorised on the same *Microsoft Office Excel 2013 Workbook*.

Table 25 below displays all the municipalities that were approached for participation in the research by the author, including the province and the year in which the clean audit was conducted; "*clean*" shows a clean audit, "*Fin.U*" indicates a financial unqualified and "*Q.A*" shows a qualifying audit. All none "*clean*" results are observed as non-clean audited in the study.

Table 25. Clean Audited Municipalities ***(source: AGSA, 2016, 2017, 2018, 2019, 2020)***

Municipality	Class	Province	Audited Years Status				
			2014/2015	2015/2016	2016/2017	2017/2018	2018/2019
Municipality 1	B	WC	Clean	Clean	Clean	Clean	Clean
Municipality 2	B	WC	Clean	Clean	Clean	Clean	Clean
Municipality 3	B	WC	Clean	Clean	Clean	Clean	Fin.U
Municipality 4	B	GP	Clean	Clean	Clean	Clean	Clean
Municipality 5	B	EC	Clean	Clean	Clean	Clean	Clean
Municipality 6	B	WC	Clean	Clean	Clean	Clean	Clean
Municipality 7	B	WC	Fin.U	Fin.U	Clean	Clean	Clean
Municipality 8	B	WC	Fin.U	Clean	Clean	Clean	Clean
Municipality 9	C	WC	Clean	Clean	Clean	Clean	Clean
Municipality 10	B	WC	Clean	Clean	Clean	Clean	Fin.U
Municipality 11	B	WC	Clean	Clean	Clean	Clean	Clean
Municipality 12	C	WC	Clean	Clean	Clean	Clean	Fin.U
Municipality 13	C	WC	Clean	Clean	Clean	Clean	Clean
Municipality 14	B	WC	Clean	Clean	Clean	Fin.U	Clean
Municipality 15	B	WC	Clean	Clean	Clean	Q.A	Fin.U
Municipality 16	B	WC	Clean	Clean	Clean	Fin.U	Clean
Municipality 17	B	WC	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 18	B	WC	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 19	B	WC	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 20	B	WC	Clean	Clean	Clean	Fin.U	Clean
Municipality 21	C	NC	Clean	Clean	Clean	Clean	Fin.U
Municipality 22	B	KZN	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 23	B	KZN	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 24	B	KZN	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 25	B	KZN	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 26	B	KZN	Clean	Clean	Clean	Clean	Clean
Municipality 27	C	MP	Clean	Clean	Clean	Fin.U	Fin.U
Municipality 28	C	MP	Clean	Clean	Clean	Fin.U	Clean
Municipality 29	B	WC	Clean	Clean	Clean	Clean	Fin.U

3.2.3.2 Non-Clean Audited Data Collection

A similar data collection strategy for non-clean audited municipalities was followed. Municipal prices, supplier BEE level and the corresponding scope of work for previously awarded low-valued procurements in the financial year of 2014/2015 to 2019/2020 (cut-off date 1 March 2020) were obtained from government websites and requested if unavailable by using a requisite PAIA form.

Non-clean audited municipalities in Gauteng were assessed based on data availability. Therefore, several municipalities were approached, but because of data in-accessibility and refusal to take part, only a few contributed. It should be noted for this study it was assumed a municipality that has had only one or fewer clean audits *(-20% clean audit)* in the past five years were to be viewed as non-clean audited municipality. Table 26 are all the non-clean audited municipalities approached.

Table 26. Non-Clean Audited Municipalities ***(source: AGSA, 2016, 2017, 2018, 2019, 2020)***

Municipality	**Class**	**Province**	**Audited Years Status**				
			2014/2015	**2015/2016**	**2016/2017**	**2017/2018**	**2018/2019**
Municipality A	A	GP	Q.A	Fin.U	Fin.U	Fin.U	Fin.U
Municipality B	A	GP	Fin.U	Fin.U	Fin.U	Fin.U	Fin.U
Municipality C	A	GP	Clean	Fin.U	Fin.U	Fin.U	Fin.U
Municipality D	C	GP	Fin.U	Fin.U	Fin.U	Fin.U	Fin.U
Municipality E	C	GP	Clean	Fin.U	Fin.U	Fin.U	Fin.U
Municipality F	B	GP	Fin.U	Fin.U	Fin.U	Fin.U	Fin.U
Municipality G	B	GP	Fin.U	Fin.U	Fin.U	Fin.U	Incomplete
Municipality H	B	GP	Clean	Q.A	Q.A	Fin.U	Q.A
Municipality I	B	GP	Unavailable	Unavailable	Fin.U	Fin.U	Fin.U
Municipality J	B	GP	Fin.U	Fin.U	Fin.U	Fin.U	Fin.U

3.2.4 Data Preparation

The Manning Textbook (2020) states that most data retrieved is likely to be a diamond in the rough. Therefore, data requires *polishing* to be used in the modelling stages of an assessment. Included in the study is the data cleaning for limiting errors and transformation to fit the model assumptions. It should be noted this section applied to both data samples; clean and non-clean audited municipalities; and was carried out using *Microsoft Office Excel 2013* and *IBM SPSS Statistics 26*.

3.2.4.1 Data Cleansing

Data cleansing is a sub-process of the data science procedure that focuses on eliminating errors in your data so a true and consistent representation of the processes is created (Manning Publication, 2020). However, before data cleansing may begin, categorisation and sorting of the collected data were required. This was achieved by dividing it into four segments as shown in Table 27. The *ideal* minimum sample sizes from the other datasets are based on an industry assumption of testing data being 10 per cent of the model training data as explained further in the chapter when conducting k-fold cross-validation (Openm, 2020).

Table 27. Four Data Classification Segments for the Study

Data Name	Purpose	Sample Size
Model Building Data	Clean audited municipal data that satisfies the 10-folds cross-validation technique and has been used to build the model (cost less than R230 000 including VAT).	The *compulsory* minimum sample size is determined by Table 24.
Model Testing Data	Clean audited municipal data left out from the model-building to be used to test the benchmarking model (cost less than R230 000 including VAT).	The *ideal* minimum sample size ±10% of *model building data.*
Model Out-of-Range Data	Clean audited municipal data that is above R230 000 (incl. VAT) is used to assess the predictive accuracy of the model for out of price range procurements.	The *ideal* minimum sample size ±10% of *model building data*.
Benchmarking Data	Non-Clean audited municipal data that has been assessed using the developed and tested benchmarking model to identify cases of pricing deviations (cost less than R230 000 incl. VAT).	The *ideal* minimum sample size ±10% of *model building data.*

Errors may cause inconsistencies in Table 27. As the information collected depends on archived data (governmental *secondary data*) made available by the approached municipalities. Data capturing errors may have occurred because of human error or system faults because of data storage hardware malfunctioning (Manning Publication, 2020); this may lead to various issues including but not limited to outliers.

A partial remedial technique for outliers used were scatterplot graphs to identify data points isolated from the general trend. These were a plot of awarded contract price versus the *anchor predictor*; the primary input (independent) variable that is assumed to have the highest correlation with the theoretical outcome as it does in the actual world. This followed by an investigation of reasons behind the outlying and depending on the validity of the outcomes were kept or removed from the sample.

A majority of the clean audited municipalities are in the Western Cape. It is, for this reason, the consumer price index (CPI) was selected as the sole measure of pricing differences for similar procurements on different dates and locations. Essentially a control to satisfy the *consistency* sampling parameter. An alternative to the CPI is the Producer Price Index (PPI). However, most of the suppliers do not produce their raw materials but acquire them from a first-hand dealer, ergo the CPI is a more suitable index and as a result, employed for the study.

$$P_{GTV} = P_{awarded} \times \left(\frac{CPI\ for\ General\ Month\ at\ National\ Level}{CPI\ for\ Tender\ Month\ at\ Provincial\ Level}\right) \quad \text{(Equation 17)}$$

The above equation shows how the contract values were standardised by converting them to the end of the 2020 financial year at the national level; where P_{GTV} is the generalised tender value to the preferred month and $P_{awarded}$ is the real awarded contract value with no adjustments. This was achieved through the aid of CPI figures obtained from Statistics South Africa (available in Appendix B, Section 2).

3.2.4.2 Data Transformation

As stated previously, not all data collected will conform to the assumption made by the multiple linear regression model of all independent variables (scope of work) having a linear relationship to the dependent variable (pricing). Using *Tukey's ladder of transformation*; the data was transformed to acquire a close to linear relationship as expressed in Table 28 (Shah, 2009), aided by the *Pearson's Correlation Coefficient* (**r**) function available on *Microsoft Excel 2013* to detect the linearity of sample data.

Table 28. Tukey's Ladder of Transformation

Tukey's ladder of transformation			
Quartic	x^4	*Inverse Square-rooted*	$-1/\sqrt{x}$
Cubic	x^3	*Inverse Linear*	$-1/x$
Quadratic	x^2	*Inverse Quadratic*	$-1/x^2$
Linear	x	*Inverse Cubic*	$-1/x^3$
Square-rooted	$\sqrt{x}$	*Inverse Quartic*	$-1/x^4$
Logarithm	$\log x$		

The ideal data transformation was selected based on the highest *Pearson's Correlation Coefficient* obtained between the independent variable and the awarded pricing. Though it should be noted as the research is based on *fencing supply and installation* contracts for public procurements, some stand-alone variables were not highly correlated with the price and needed to be consolidated with others to depicted their true impact (e.g. BEE score)

These may have had a collaborative influence on the predictive accuracy of the final model. And so, some variables had to be combined with others to assess their collective significance and is elaborated on further. Numerical values had to be assigned to each of the variables that could not be conveyed numerically. Table 29 depicts the assigned values for the non-numerical data used in the study.

Over-fitting by including many predictors is an ever-present possibility when dealing with statistical modelling techniques (Shaikh, 2018), such as multiple linear regression

(MLR). It is for this reason the author combined variables where possible to reduce the number of independent variables (*predictors*). These are depicted in Table 30 which summarizes the refined list of compulsory independent variables to be used in the study for all four data segments. Included are the combined items, the reason and the methodology used to combine them.

Table 29. Assigned Values for Gate Automation and Fencing Material (Categorical Data Conversion)

Escalation Factors for Gates (EF)

Justification for Categorical Scale Used:

According to data found on *howmuch.net* (2020); the average price for the installation of a manual gate is ±US $4000, and the cost of motorising a sliding or hinged gate is ±US $2400 (including labour). This means the cost of motorising a gate is ±1.6 times greater than the original bill.

Assuming the same standard applies to South Africa, the below factors were developed:

If the Gate is automated	EF=1.6
If the Gate is manual	EF=1

Fence and Gate Materials Cost Factors (CF)

Justification for Categorical Scale Used:

The following items are the cost per metre for fences based on the type of material used for a height of 1.8 metres (BusinessNews, 2018).

- Wire mesh is ±R550 (+R80 for barb)
- Precast concrete is ±R820
- ClearVu ±R1250
- Palisade is ±R1 800
- Face brick is ±R1 850
- Steel palisade and wall pillars are ±R2 750

Using these figures, a relative material cost scale [rand/metres^2] was obtained by dividing the costs by using a unit-less value of 1.8 to scale down the cost to *per square metre*. Assumptions were made for materials not included in the article (*underlined items*) based on an internet search for the actual cost of the materials then grouping the unaccounted for materials with those that had similar pricing:

Item Name	Value (rand/metres^2)
Wire Mesh	306
Precast Concrete	458
ClearVu	703
Palisade Panels	1008
<u>Steel Slatted</u>	1008
Face Brick	1039
Palisade and Pillars	1528
<u>Concrete Pillars</u>	1528
Plus Barbed	+46
<u>Electric Accessory</u>	+46

The above Table 29 was used to refine and prepare that data accordingly. This resulted in five possible independent variables as explained in Table 30. As stated earlier the likelihood of all these highly correlating to the price is fairly low. This resulted in a hierarchal regression model building approach when short-listing independent variables. However, this concept is explored further in the chapter.

Table 30. Refined List of Compulsory Variables by Consolidation

Variables	**Grouping**	**Reason and Methodology**
0. Generalised Contract Value.	Rand [R]	The value of the winning contract after the CPI transformations has taken place. (*listed as zero because it does not form part of the predictors*).
1. Fence Removal [m^2]	Single [m^2]	Removal of old fences and is usually priced as 1m multiplied by the length of the existing fence.
2. Fence Material Cost Estimate [R] *(Anchor Predictor)*	Length [m] Height [m] CF [R/m^2]	The length, height and material will vary for fences according to the work specifications. Combining these variables will give the total estimated relative value of the total material used in the erecting of a fence. However, a contractor may be tasked with erecting multiple fences, for this reason, the preferred equation to merge these is $\sum length\ x\ width\ x\ Cost\ Factor$. Additionally, the fencing material cost estimate (FMCE) was selected as the anchor predictor as according to Lumber (2020) these collective variables have the highest influencenon the pricing of a fencing contract.
3. Sliding Gate Material Cost Estimate [R]	Length [m] Height [m] EF [Unit less] CF [R/m^2]	The length, height, material, motorising and number of sliding gates may vary according to the work specifications. Combining these variables will give a rough estimated relative value of the material used. The preferred equation to merge these: $\sum \boldsymbol{length\ x\ width\ x\ Cost\ Factor\ x\ Escalation\ Factor}$.
4. Hinged Gate Material Cost Estimate [R]	Length [m] Height [m] EF [Unit less] CF [R/m^2]	Same as Sliding Gate Material Cost Estimate equation: $\sum \boldsymbol{length\ x\ width\ x\ Cost\ Factor\ x\ Escalation\ Factor}$.
5. BEE Score	[Unit less]	The black economic empowerment (BEE) score has remained as is and ranged from 0 to 20.

The *BEE Score* variable has an impact on pricing as explored in the Literature Review. However, as a stand-alone variable, it will not highly correlate with the pricing. This may be due to contracts worth R200 000 and R45 000 having the same BEE score because of a variation in specifications. This leads to an unclear relationship between BEE Score and the pricing, yet existing theory already states it has an effect.

A remedial step taken by the author was to integrate it into the awarded pricing using the procedure preferential point system highlighted in the Literature Review. The procedure includes both the competitiveness of the pricing (out of 80) and the *BEE score* (out of 20). Thereafter scored to a maximum score of 100.

As all the contractors in this study were winning bidders and because of data unavailability, it was assumed all suppliers had scored a ±80 for the pricing assessment. This means they were the cheapest or one of the cheapest suppliers. The data for BEE scores were availed and therefore a *pricing scaling transformation* based primarily on a suppliers *BEE score* was possible.

The *pricing scaling transformation* method developed for this study is an adaptation of the preferential scorecard as explored in the previous chapter (Dobie and Xinwa, 2015). It is achieved by summing the pricing *score* out of 80 and the supplier *BEE score* out of 20. This proceeded with the division of the score by 100. Thereafter, dividing the CPI adapted winning bid price (generalised contract value) by the calculated amount.

The below equation summarizes the methodology used to transform the pricing by factoring in the supplier BEE score. P_{SGTV} is the calculated scaled generalised contract value (SGTV). S_{price} is the supplier pricing score that was presumed to be 80 for this study. S_{BEE} is the supplier BEE score out of 20 and last, P_{GTV} is the actual awarded procurement price after CPI correction (*generalised contract value*).

$$P_{SGTV} = {P_{GTV}} \Big/ \left(\frac{S_{price}+S_{BEE}}{100}\right) \quad \text{(Equation 18)}$$

In the equation, a supplier that has a low BEE score [anything significantly lower than 20] would have an increased P_{SGTV}, whereas a supplier with a compliant BEE score will remain unchanged. This formula was developed based on the theory explored in chapter two suggesting that a high BEE score is accompanied by an increase in procurement costs (Acemogluy, Gelbz and Robinsonx, 2007; Lowman, 2017). The *pricing scaling transformation* is essentially an attempt to *flatten the playing field* by removing the BEE element in the pricing.

However, the above scaling exercise assumes there is a linear relationship between a suppliers BEE score and the pricing given; this has yet to be proven. Therefore, a test was required to assess whether the *pricing scaling transformation* is effective by using the data from clean audited municipalities classified as *model building data*. This was realised by evaluating *Pearson's Correlation Coefficient* for two datasets. Both the P_{SGTV} and P_{GTV} were plotted against the anchor predictor to assess the correlation. The pricing dataset with the larger *Correlation* was selected for the model building and the other discarded. Below Figure 7 depicts the process followed to achieve this.

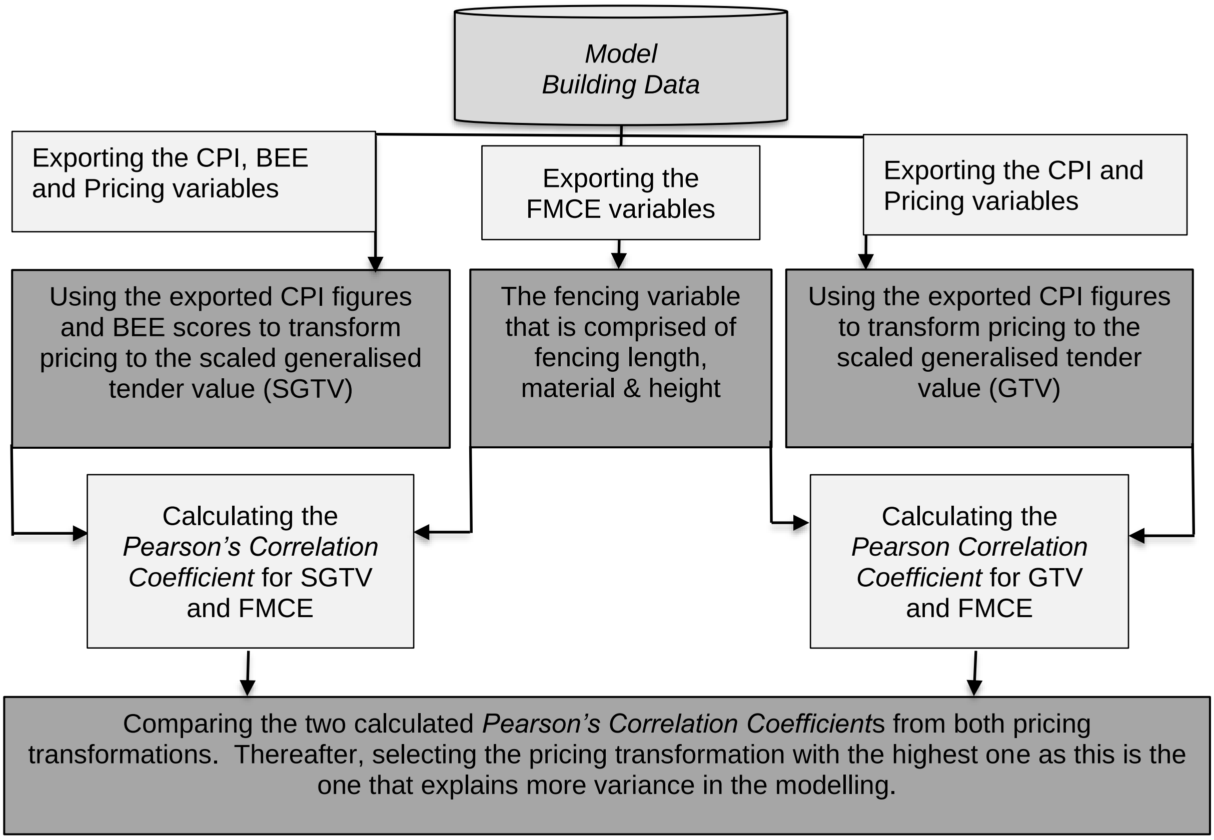


Figure 7. Method Used to Compare SGTV to GTV

Because of the above *pricing scaling transformation,* the BEE score would no longer be a primary variable. This means the shortlisted refined independent variables for the study would total four. Thereafter, all four variables from the *model-building data* were plotted against the winning pricing transformation to acquire their respective true study linear transformations using the previously outlined procedure. Subsequently, the identified transformations were applied to the remaining three data segments.

3.2.5 Data Storage

The refined data was then returned to their respective four data segments. All data segments were kept on the cloud storage facilities provided by the *University of Witwatersrand* and accessed remotely on any smart device or computer. Though the primary storage method used during the study was offline on an *HP 650 Notebook Laptop* and for confidentiality were erased after the study had concluded.

3.2.6 Data Management Summary

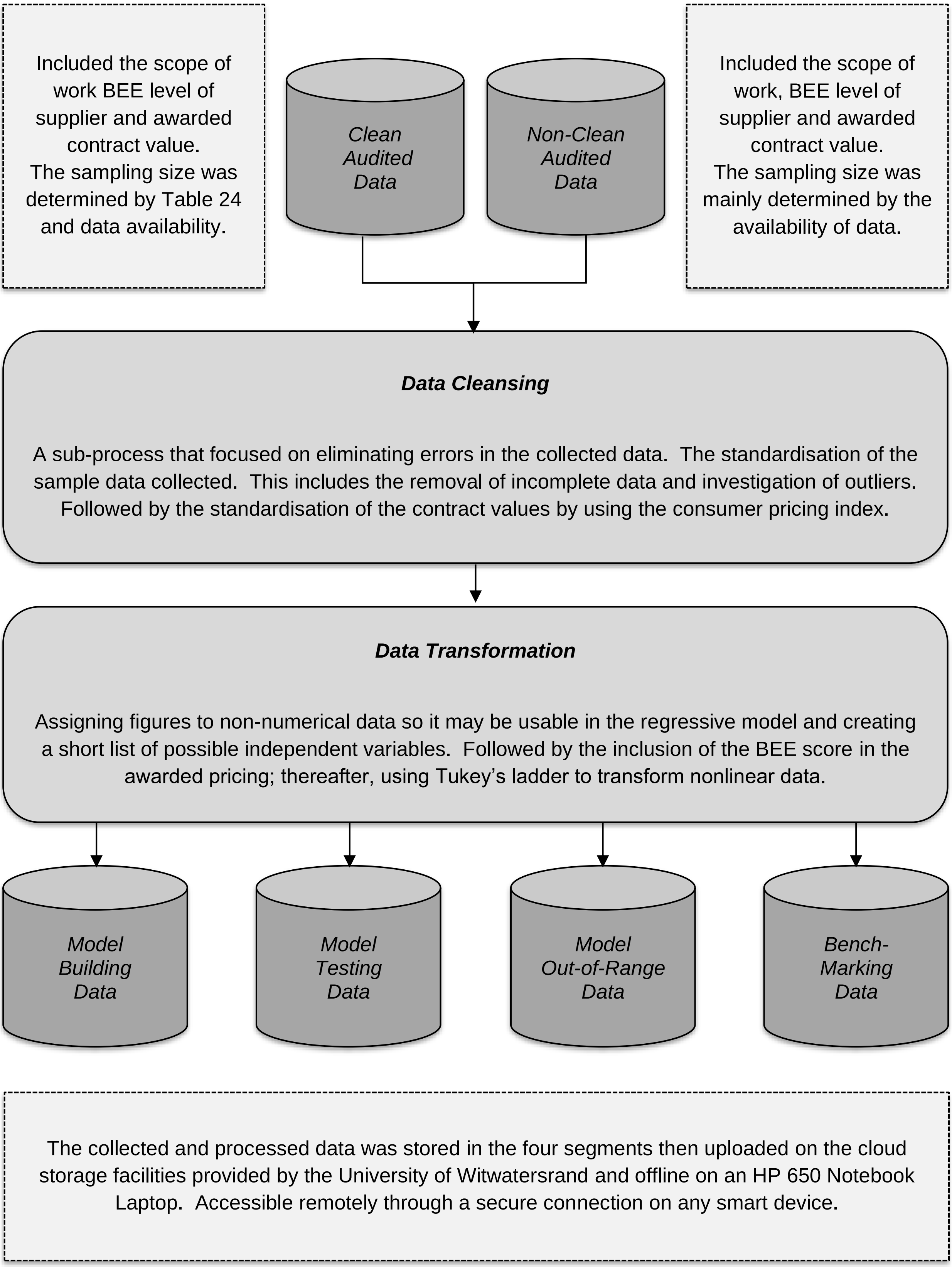


Figure 8. Research Data Management Method Summary

3.3 Preparation for Model Design

This subsection is a logical breakdown of the preparation that was required to develop a benchmarking model using multiple linear regression (MLR). The rational order followed is partly based on chapter two of the *Data Science Process* by Manning Publications (2020). It is a two-phased approach that includes identifying cases of multicollinearity and the removal of non-significant variables. The model structure is based on the *hierarchal regression* model building procedure and this section only makes use of the *model-building data*. It should be noted this subsection was used to fulfil the *first portion of the first objective* by identifying significant variables in the datasets to develop the model.

3.3.1 Identifying Multicollinearity Cases And Testing For Homoscedasticity

MLR models are restricted by three limitations/assumptions, the first of which was explored in the previous subsection by establishing a linear relationship between the independent variables and the pricing. The remaining two limitations/assumptions are explored in Table 31 that includes their requirement to pass the assessment, methodology used to address them and the target value set for the study.

Table 31. Methodology to Address Multicollinearity and Homoscedasticity Limitations/Assumptions

Limitation	Requisite	Method	Target
Multi-collinearity	No significant relationship existing between the independent variables (Stiti and Yape, 2019).	The *IBM SPSS Statistics 26* was used for every independent variable combination in the scope of work. which depicted the relationship between all the variables using the Pearson's Correlation Coefficient (**r**).	Non-high degree r coefficient value; between -0.5 and +0.5 (lower to moderate).
		The variance inflation factor (VIF) was calculated using *IBM SPSS Statistics 26.*	VIF is lower than 5 for all variables.
Homo-scedasticity	The variance of error terms is similar across the values of the variables (Ogee, 2013).	Automated using *IBM SPSS Statistics 26* Plot for standardised residuals function.	Have a "*squared*" like plot of the residuals.

First, if there had been a substantial linear relationship between the independent variables; the one with the higher collective significance was selected for the modelling application and the other discarded. This was done without removing the anchor predictor (*fence material cost estimate [FMCE])* identified in the hierarchal Figure 9. Second, if the residuals plot did not follow a *"squared*" like trend, the data transformations used in the previous sections would have been reviewed to obtain the desired outcome.

3.3.2 Removal of Non-Significant Variables

To conduct an MLR model build, the variables used were limited to the ones which had the strongest collaborative influence on the assessed outcome. This means a variable may have a low significance but increases collaboratively with other variables as explored in the previous section by combining the *BEE score* variable and the *GTV.*

However, known key independent variables may be viewed as irrelevant when testing for significance. For this reason, the study has followed the *hierarchal regression* model building procedure aided by the adjusted $\mathbf{R^2}$. Which is a sequential building of several multiple linear regression models, each adding and removing various independent variables (Petrocelli, 2003; Grace-Martin, 2020).

The independent variables are based on prior knowledge of the field and are ranked from highest to lowest importance by the modeller based on how influential each may be in the real world outcome. The highest-ranked is the primary variable (*anchor predictor*) and is assumed to have the greatest correlation with the theoretical result as it does in the real world.

One common practice of *hierarchal regression* model building is to start by only adding key control variable(s) to the model. In the next model, you can add variables of interest, to see if they predict the dependent variable above and beyond the effect of the controls (Grace-Martin, 2020). This exercise helped identify significant independent variables assessing their respective coefficients (**β**) and resultant adjusted $\mathbf{R^2}$ using the *IBM SPSS Statistic 26 Linear Regression Function*.

The significance of each independent variable's coefficient was depicted by both the **t**-value and **P**-value. If a coefficient had a **P**-value lower than 0.05 it meant it had more significance in the regression outcome. Should the **t**-value be above +2 or below -2, it meant there is confidence in the variable as a predictor (Barrons, 2020).

This assessment was used to test each independent variables level of importance to the regression model and cross-reference it with the below hierarchal diagram to judge whether removing the variable would be acceptable. Figure 9 shows the independent variable hierarchy used in this study based on existing knowledge of which item most likely had the greatest impact on the pricing (Howmuch, 2020).

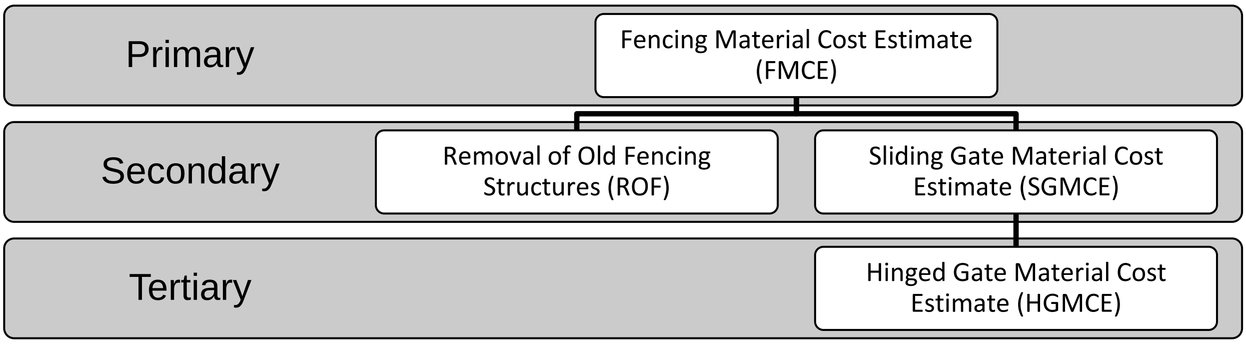


Figure 9. Hierarchal Ranking of Independent Variables

The FMCE is viewed as the primary variable (anchor predictor) as it is expected to have the highest influence on the final pricing (Tim, 2016). The ROF and SGMCE are presumed to be secondary variables as they are expected to have a noticeable impact on the final pricing. HGMCE was ranked as the tertiary variable as it cost less to install than SGTV and could not be ranked on the same level (Howmuch, 2020).

3.3.3 Preparation for Model Design Summary

This section of the methodology was primarily focused on the qualitative preparatory steps required to develop a credible multiple linear regression model build. These steps included the separation of the significant and insignificant independent variables to the study; so the noteworthy variables may be used in the model building process. The result for this section was the importing of the significant independent variables and the corresponding pricing variable from the *model-building data* for the next section.

3.4 Model Design

3.4.1 Developing Model

The development of the benchmarking model was conducted using the variables from the *model-building data* which made it through all the prior processes. As previously stated, the *hierarchal regressive* model building procedure was followed successively in the methods explored earlier. The model accuracy (adjusted $\mathbf{R^2}$) showed the successfulness of a build. This means the model accuracy should have been maintained at ±0.5 or the model would be rejected and a work review conducted. This subsection was used to fulfil the *second portion of the first objective* by using the identified significant variable to develop the comparative model.

3.4.1.1 Calculation of The Coefficients

To calculate the coefficients by using equation 5, the significant variables identified in the previous section were express in a matrix that has multiple independent variables (**X**) and single corresponding dependent variables (**Y)** form which satisfies the notation expressed by equation 4. This allowed for the automated calculation of all the coefficients (**β**) for the model by using *IBM SPSS Statistics 26*, *Regression Function* after which, testing the significance of the regression model by using the below tools.

3.4.1.2 Test for Significance of Regression

As stated by Montgomery and Runger (2011, p. 470); a method of testing a model's adequacy is by conducting a *hypothesis test* by using **F-**statistic and can be achieved by using equation 14. This study has made use of an automated method of acquiring **F-**statistic. Moreover, **F-**statistic must be greater than the critical **F-**value as a sign that at least one of the coefficient of the independent variables (**β**) is useful for predicting the outcome.

The critical **F-**value was calculated to be 2.01, based on an *alpha value* of 2.01, numerator *degrees of freedom* of 4 and *denominator degrees of freedom* of 75 using the methodology highlighted in the empirical framework section. Should this not have been the case, a work review was conducted.

3.4.2 Model Validation

Hughes and Fisher (2020) stated, a model may fail when presented with data not included in the building process, and for MLR, this may occur through over-fitting. Over-fitting is when a model can predict its data with high accuracy due to adding more independent variables to a model than is needed as demonstrated in Figure 10 (Montgomery and Runger, 2011). A model may be underfitting when insufficient variables are used an idyllic model is one the fits the data trend "just right".

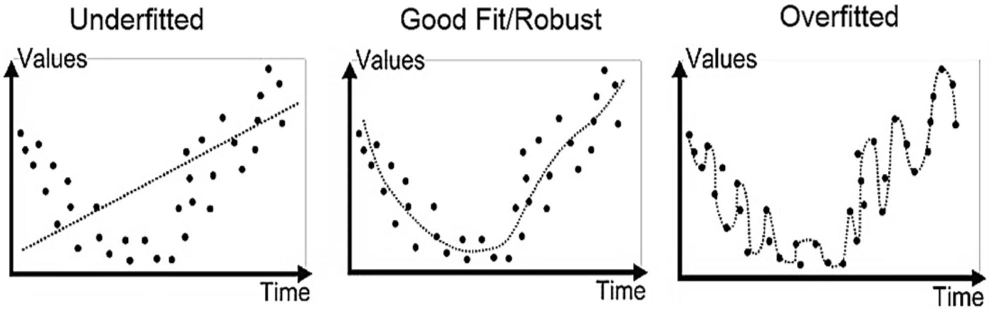


Figure 10. Under-Fitted vs Good-Fitted vs Over-Fitted ***(source: Bronshtein, 2017)***

A method of curbing the inflation of the predicted accuracy, as explored earlier, is by using adjusted $\mathbf{R^2}$. Adjusted $\mathbf{R^2}$ factors in the number of independent variables used to acquire the modelling accuracy and penalises the modeller for using many variables (Moore, 2003). However, additional safeguards can ensure that a model's precision is validated even with data samples not used in the construction process through the aid of the previously explored *k-fold cross-validation* technique.

According to Quantcademy (2020); for an MLR model, the preferred "k" value is 5 or 10. However, because of this research having a limited sample size; the 10-Fold cross-validation technique was utilised (k=10) as it resulted in higher sample sizes when training the models. The data was split into 90:10 ratios per fold, where 10 per cent of the data was used for testing and 90 per cent for training. The benefit of this method is that all datasets were used for both model building and validation (Openm, 2020).

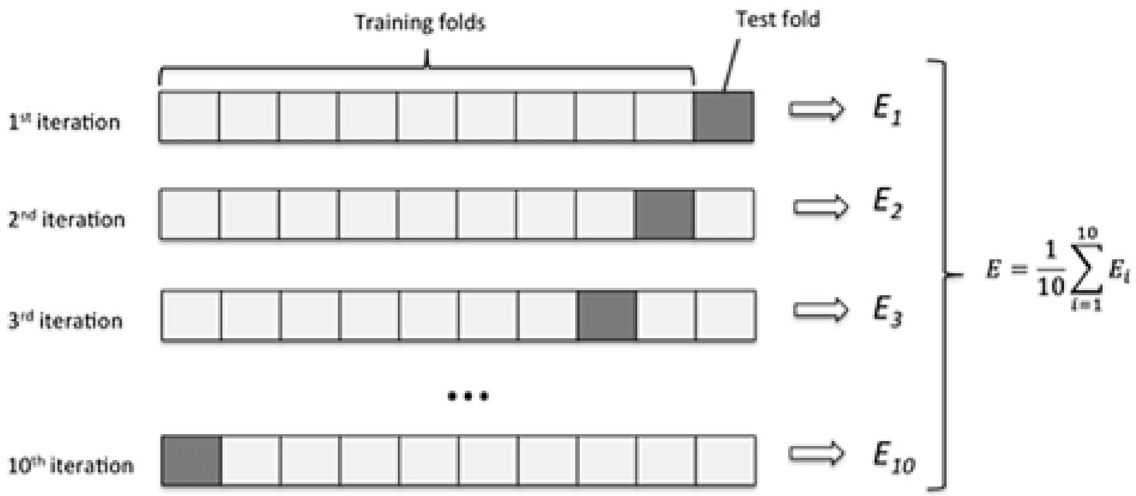


Figure 11. Breakdown of the 10-Fold Cross-Validation Method ***(source: ResearchGate, 2020)***

Because of the 10-Fold cross-validation technique, the model validation process was undertaken ten times and in every iteration, the data used for building and testing the model was changed as shown in Figure 11. The figure depicts how the divisions will appear for the validation method.

Coupled with this, the data sample used to build the model was a multiple of 10 so to fit in the validation as depicted in Table 24. The performance measure reported by 10-fold cross-validation is the average of the values computed in the loop (Scikit-Learn, 2020). Below is a depiction of the validation process.

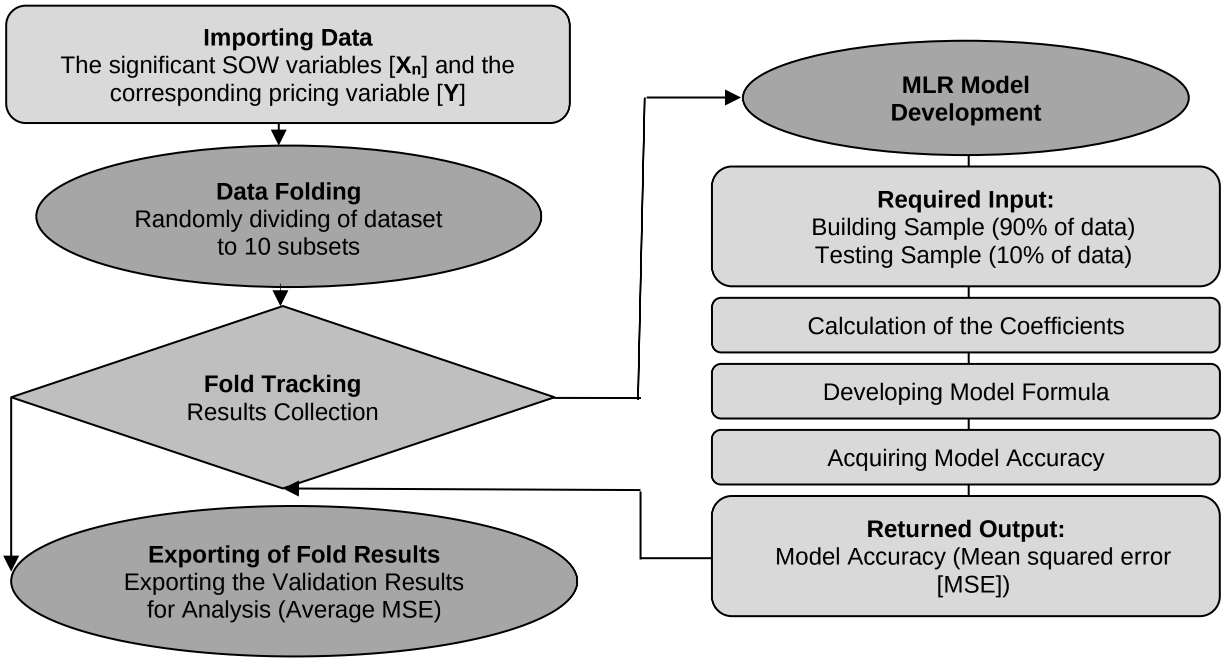


Figure 12. Breakdown of the Study Model Validation Process

3.4.3 Model Design Concluding

The result for this section was the final MLR model developed using the *model-building data*, followed by the perceived model reliability using the 10-fold cross-validation technique. The cross-validation indicated whether the model developed is credible by using the datasets provided. This was observed by rooting the average *mean squared error* (MSE) obtained to acquire the *root mean squared error* (RMSE).

The RMSE was then compared against the training model error that is given by the *standard estimation error* (SEE). SEE is calculated automatically through the aid of the *IBM SPSS Statistics 26* software package as part of the training outcome. The acquired RMSE should be close to the SEE (assumed ±1% deviation) for the model built to be viewed as robust and acceptable.

If RMSE is greater than SEE, it would have inconclusively correlated with the work reviewed by Klein and Rossin (1999) in the previous chapter, which stated training data errors (SEE) have a lesser impact on pricing predictive accuracy than testing data errors (RMSE). This is based on an assumption of homogeneity in the distribution of the error samples collected. However, this has been further reviewed in chapter five when analysing the research findings as the sample sizes used in the calculation of both SEE and RMSE may have played a crucial role.

3.5 Benchmarking Strategy

The research benchmarking was conducted in this section and achieved by using the model built to predict the pricing of fencing contracts awarded by past non-clean audited municipalities in the Gauteng region. This is to assess what the contract values may have been, should the procurement have been conducted by a clean audited municipality. The section was carried out using *IBM SPSS Statistic 26* and *Microsoft Office Excel 2013* and used to *fulfil the first portion of the second study objective* by identifying which contracts in the Gauteng Province may have incurred pricing deviations.

3.5.1 Prediction of Prices

The remaining three segments of the stored refined data were called on in this subsection. This included the *model testing data* for assessing the models' accuracy, the *model out-of-range data* to assess its out of price range capabilities, and the *benchmarking data* used to assess the research hypothesis. The independent variables for all the above datasets were used to calculate the modelled predicted contract values.

This was continued by comparing the modelled predicted contract values against the scaled generalised tender values for each of the procurements. However, the likelihood of the model being, perfect is low. Therefore, the benchmarking exercise had to consist of a tolerance for the inherent error. This was achieved by using the prediction interval.

3.5.2 Creation of Prediction Interval

A prediction interval is a range of values that likely contain the true figure of the dependent variable for a single assessment given values of the independent variables (Frost, 2020a). It is a range of figures in which there can be a high certainty the real value lies (Montgomery and Runger, 2011). This was achieved by using equation 16 as shown below:

$$\hat{y}_0 - t_{\frac{\alpha}{2},n-p}\sqrt{\left(\frac{SS_E}{n-1}\right)^2\left(1 + x_0'(X'X)^{-1}x_0\right)} \leq Y_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2},n-p}\sqrt{\left(\frac{SS_E}{n-1}\right)^2\left(1 + x_0'(X'X)^{-1}x_0\right)} \quad \text{(Equation 16)}$$

The notation is essentially the *margin of error* (MoE) for a prediction and is the basis on which the prediction interval is created. Adjusted $\mathbf{R^2}$ is inversely proportionate to the margin of error (Frost, 2020b). As a result, a greater adjusted $\mathbf{R^2}$ leads to a smaller *MoE,* consequently improves the modelling outcomes for flagging pricing deviations; henceforth the selected ±0.5 research standard.

To acquire the MoE, a confidence level was required. This is obtained as an expression of one minus the significance level (1- α [alpha value]). As previously stated, the assumed significance level for this study is 0.1 (10%). This translates to a 90% confidence level. Thereafter, the calculation of the individualised upper and lower bounds for each of the predicted variables using equation 16.

The developed model was tested using the remaining data samples from the clean audited municipalities which did not satisfy the multiple of ten assumptions (*the model testing data sample*). Using the independent variables, the prices from the procurements were calculated, after which the prediction range was obtained. Since the confidence level of this prediction interval was set at 90%, this meant at worst one out of 10 of the sample testing data can fall outside the prediction range.

An additional assessment was undertaken to evaluate whether the model built could assess procurements that fell outside the pricing scope (greater than R 230 000 incl. VAT). This was achieved by repeating the above steps with the *model out-of-range data* segment to evaluate whether the model could accurately predict *out-of-price-range data.* This assessment was used to identify the number of procurements from the clean audited municipalities that fell outside the prediction interval to obtain an accuracy depicted by a percentage.

Thereafter, the study purpose was conducted by using the built model to assess how clean audited municipalities would have priced their procurements should they have been given the same contracts from the non-clean audited municipalities. This was achieved using the above procedure, by using the scope of work from the *benchmarking data segment* to predict the probable real market pricing. Thereafter, comparing with the actual awarded values.

Besides the above, a supplementary test was conducted to identify the contracts for the clean audited municipalities that fell out of the pricing. This was achieved by using the prediction interval to assess the *model-building data* via the previously outlined procedures. This aided in the evaluation of how truthful the chosen confidence level of 90% was when compared against the sample used to build the model.

3.5.3 Benchmarking Strategy Summary

The results for the *benchmarking data segment* assessment were graphed and both the probable actual market price of the procurement and the scaled generalised tender values were plotted on the same diagram for a visual comparison. A *trend analysis* was used to examine the results, it is a method that enables a modeller to look at quantitative data that was collected over a duration of time (Surendran, 2018).

First, this was a visual depiction of whether pricing deviations may have taken place by the scaled generalised tender values being above or below the limits of the plotted graph. If overcharging (price inflations) likely had taken place, this was shown by the scaled generalised tender values being above the upper limit of the *prediction interval* and this also depicted by what margin the price was possibly inflated. Conversely, if undercharging may have taken place, this was shown by the scaled generalised tender values being below the lower limit of the *prediction interval.*

However, it should be noted that an optimistic evaluation favouring the procurements that deviated from the pricing was taken when analysing the data. This meant the justifications for the pricing deviations were explored; including but not limited to; faults within the sampling methods, model building process, and deviations in pricing due to study limitations and assumptions.

Second, this indicated whether indeed there was a correlation between a *clean audit* status of a municipality and the pricing of the topic low-valued contracts. Therefore, regardless of what the results may be, a conclusion was obtained on the two above points. The outcome has either shown there is a relationship between *clean audited* municipalities and pricing of tenders or the audit status of a municipality is no sign of probable pricing deviations for the assessed low-valued fencing contracts.

3.6 Limitations of The Study

3.6.1 First Limitation

This study was limited by the inability to assess more variables influencing the pricing of *supply and installation of fencing contracts* because of limited resources and being focused on acquiring a relationship between the scope of work (controlled variables) and the pricing of a procurement. This has presented an element of ignorance to the constructed model. The ignorance was conveyed as the "*error term"* (ε) in the multiple linear regression equation (Sarip *et al.,* 2016).

This has certainly led to a large prediction error in the model. However, various mitigation tools are explored in this chapter. Including the *10-fold cross-validation technique* to assess the reliability of the *model-building data*. Thereafter, the use of the *prediction interval* to give a range of figures where there is a high certainty the true contract value laid. Last, assessing how well the model can predict data not used in the building process using the *model testing data* aided by the prediction interval.

3.6.2 Second Limitation

The research was primarily focused on the pricing of *fencing supply and installation contracts* from different dates and locations. Perhaps there are various methods available to generalise these and to have all contracts presented in the same theoretical time and location. However, the allocated method for this study of assessing the changes in these was the consumer price index [CPI] (StatsSA, 2017).

3.6.3 Third Limitation

The study could not assess the impact of the competitiveness of a contract because of limited resources as explored in chapter two. The minimum required number of suppliers bidding for low-cost procurements in municipalities is three as stated by Dobie and Xinwa (2015). This may not be viewed as competitive, but based on the findings by Stiti and Yape (2019), companies tend to bid more evenly with lower-priced contracts irrespective of the competitiveness of the bidding process. Therefore, the total number of bids review per past contract assessed may be neglected for this study.

3.7 Methodology Summary

Figure 13 is the overall research methodology layout. It depicts the work summary of the tasks explored in the chapter to develop the benchmarking model. This included an empirical framework exploration; followed by an elaboration of the regression-based benchmarking model building procedures and the relevant validation and verification tools used in the study. Thereafter, details how the benchmarking procedure was undertaken for the study.

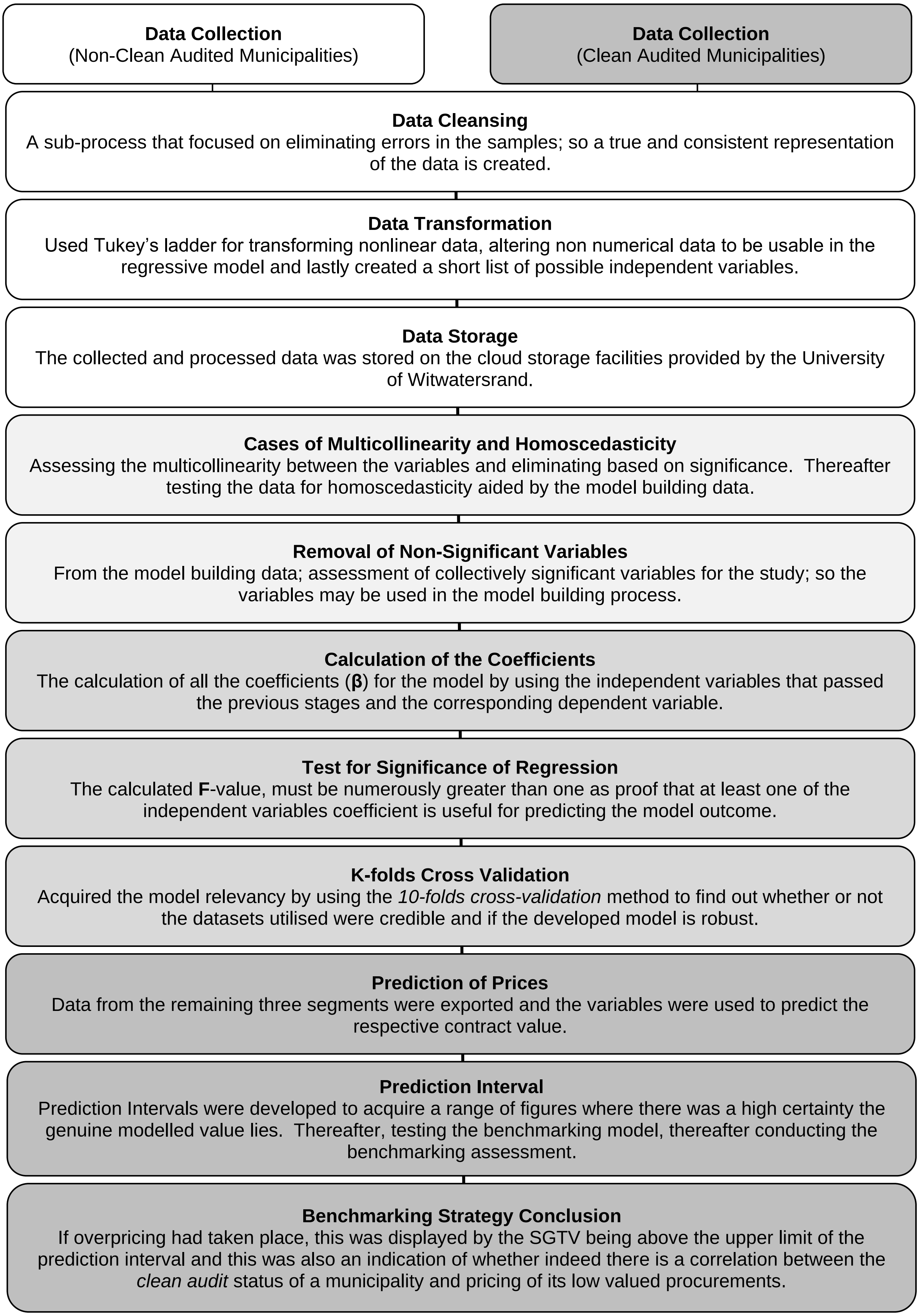


Figure 13. Overall Research Methodology Summary Diagram

Chapter Four: Findings

This chapter details the findings obtained in the research using the framework outlined in chapter three. The first section is the data management results obtained in the study. It includes the capturing and categorising of the data collected; the cleaning and standardisation thereof; and lastly the application of the relevant transformations and the consolidation of key independent variables.

The above is seconded by the modelling preparatory steps identified in chapter three. These identified multicollinearity and homoscedasticity cases that may have occurred in the model building stages and the remedial steps are taken. Thereafter, the removal of non-significant variables from the refined data for modelling based on their hierarchal level and significance to the study.

The third section is the multiple linear regression model building which included the calculation of the independent variable coefficients (**β**). Then assessing the models' significance by observing the key indicators including **F-**statistic, **P**-value and adjusted $\mathbf{R^2}$**.** Subsequently, validating the datasets used to build the model using the 10-Folds cross-validation technique on the data used to construct the model.

The last section is the benchmarking process in which past procurements from the non-clean audited municipalities in the Gauteng Province were assessed using the model. Thereafter, using the prediction interval to identify where pricing deviations may have occurred, but not without testing the reliability of the interval. Every section in this chapter has a clarification on how the results were found and what their significance is, to recognise why the subsequent step was taken.

4.1 Supply and Construction of Fence Benchmarking Model

4.1.1 Data Management

This section is the compilation of the data management results for the research. It comprises the total number of assessed procurements and the refinement procedures undertaken. Thereafter, transforming the datasets to meet the linearity assumption/limitation and the combining of key independent variables where applicable. Included in these outcomes was a brief explanation and justification for certain variables behaviour to the value of the contracts.

4.1.1.1 Key Assumptions for Section

The data obtained from some municipalities was incomplete. As a result, additional assumptions and data disclosures to specific procurements had to be taken, and are all listed in Table 32. This is to satisfy the sampling parameter of *transparency* for the decisions made by the author not to discard *good data* for the study. As depicted; *clean audited procurements are referred to numerically whereas non-clean audited are alphabetical*, and *none of these contracts was used for model building*.

Table 32. Data Sorting Assumptions Unique to Certain Procurements

Item	Tender	Assumption/Disclosures
Clean Audited	72	Included biometric scanners at the gates for controlled access to employees (*not part of modelling data*).
	109	Included twenty bollards and two boom gates at the entrances. The budgeted amount in the yearly financial plan was used as the awarded value was unavailable (*not part of modelling data*).
Non-Clean Audited	E & F	These contracts were procured separately but presented as a total on the municipal budget summary (R394 000). As a result, it was assumed both the procurements are worth R197 000 as the maximum allowable contract value for each was R200 000 (excluding vat) for each.
	J	Included the replacement of lights in various places and the supply and installation of three air-conditioning units.
	L	Additional items were required including a shed relocation and a 23m^3 parking pavement. It was assumed the cheapest seller won as only the prices given by all the bidders were available.

4.1.1.2 Data Collection

Various past *supply and installation of fencing* contracts were assessed from different clean audited municipalities in South Africa and non-clean audited in Gauteng. However, due to disproportionately divided contracts of supply and erection, incompletion of data in the scope of work, unavailability of final pricing, refusal to take part in the studied and non-responsiveness of some municipalities. The data sizes were limited to 109 for clean audited and 12 for non-clean audited.

To better understand the characteristics of the data collected the following tables were drawn up as a summary of the number of contracts refined to be used in the study. These are expressed by various distribution methods including date (year procured), provincial (location) and price distribution (range in costs). The detailed non-refined data tables' results for both clean and non-clean audited municipalities are available in Appendix C, Section 1. The appendix does not reveal the province/municipality of the individual contract data, only the municipal "*clean*" status was disclosed.

Table 33 is the distribution of contract values for the original awarded price (non-CPI adjusted) for both the clean audited and non-clean audited municipal data samples. On the assessment of the municipal data distribution; the data collected is highly grouped to the procurements ranging from R100 000 to R230 000 (~49.5% of the *overall clean audited data*, ~62.5% of the *model-building data*, ~50% of the *model testing* and ~75% of the *benchmarking data*). This violates the sampling parameter of *diversity* as the data is not evenly distributed between the price ranges.

Table 33. Fencing Contracts Data Sample Distribution by Price

Price Range	**Clean Audited Municipalities**		**Non-Clean Audited Municipalities**	
(Rands)	*Total*	*Note*	*Total*	*Note*
R0 to R30 000	0	*Model Building Data*	0	*Benchmarking Data (Used to Assess Cases of Pricing Deviations in Gauteng)*
R30 000 to R100 000	30		3	
R100 000 to R230 000	50		9	
R0 to R30 000	0	*Model Testing Data*		
R30 000 to R100 000	4			
R100 000 to R230 000	4			
R230 000 to R350 000	8	*Model Out-of-Range Data*		
R350 000 to R500 000	6			
R500 000 to R750 000	1			
Above R750 000	6			
Total	**109**		**12**	

Table 34 is the distribution of procurement contracts by date. As depicted, the contracts are heavily grouped to the latest three financial years; from the year 2017/2018 to partial 2019/2020 (~73.39% of the *overall clean audited data*, ~80% of the *model-building data,* ~100% of the *model testing data*, ~71.43% of the *model out-of-range data* and ~75% of the *benchmarking data*). As a result, the data was skewed toward the recent three financial years and, as a result, violating the sampling parameter of *diversity* and possibly *consistency.*

Table 34. Fencing Contracts Data Sample Distribution by Date

Year	**Model Building**	**Model Testing**	**Model Out-of-Range**	**Benchmarking**
2014/2015	3	0	1	3
2015/2016	7	0	3	0
2016/2017	13	0	2	0
2017/2018	16	2	4	1
2018/2019	28	5	6	7
Partial 2019/2020	13	1	5	1
Total	**80**	**8**	**21**	**12**

Table 35 is primarily a depiction of how the data used to create the model was distributed by location through the provinces. The Western Cape makes up the largest sample data for clean audited municipalities (~94.5% of the *overall clean audited data*, ~93.75% of the *model-building data,* ~75% of the *model testing data* and ~100% of the *model out-of-range data*). This violated the sampling parameter of *diversity.*

Table 35. Fencing Contracts Data Sample Distribution by Province

Province	Model Building	Model Testing	Model Out-of-Range	Benchmarking
Western Cape	75	6	21	N/A
Gauteng	2	0	0	12
Mpumalanga	0	1	0	N/A
Kwa-Zulu Natal	3	1	0	
Total	**80**	**8**	**21**	**12**

All the above violations have introduced elements of biases to the model constructed. These violations may present incorrect results which may lead to inconclusive findings. The implications for the above issues are further explored in the next chapter and their potential impact on the study outcomes. This is to satisfy the third and last sampling parameter, *transparency*.

4.1.1.3 Data Preparation

The data collected required cleaning and transformation; cleaning was conducted by converting the past procurement price to the present value through the aid of the consumer price index (CPI) using figures obtained from Statistics South Africa and equation 1. The equation shows how the procurement prices were standardised by bringing them to 1 March 2020 at the national level (CPI = 115.2). Where P_{GTV} is the *generalised tender value* (GTV) and $P_{awarded}$ is the awarded procurement price.

$$P_{GTV} = P_{awarded} \times \left(\frac{CPI\ for\ General\ Month\ at\ National\ Level}{CPI\ for\ Tender\ Month\ at\ Provincial\ Level}\right) \qquad \text{(Equation 17)}$$

Thereafter, creating scatterplot graphs for all the independent variables plotted against the pricing to assess the correlation. Though linear transformations were required for non-linear data through the aid of *Tukey's ladder of transformation,* as was explained in the previous chapter found in Table 28*.* Some variables could not stand alone and required paring with others to have an identifiable linear correlation to the pricing. The data results for the consolidation of the independent variables are available in Appendix C, Section 1.

However, after these variables were combined, the GTV could not be used without assessing the impact of the supplier black economic empowerment (BEE) score on the pricing outcome, as outlined in the previous chapter. This was achieved by assessing the *Pearson's Correlation Coefficient* for the GTV with and without the supplier BEE score because of temporarily using the anchor predictor as the dependent variable, the process is highlighted in Figure 7.

$$P_{SGTV} = {P_{GTV}}\Big/{\left(\frac{S_{price}+S_{BEE}}{100}\right)} \quad \text{(Equation 18)}$$

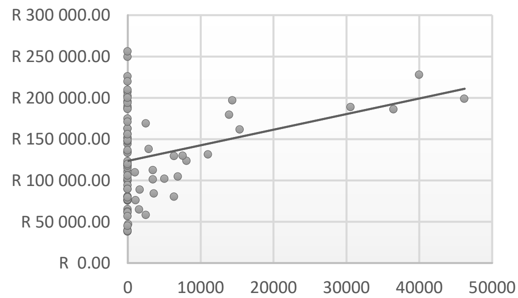
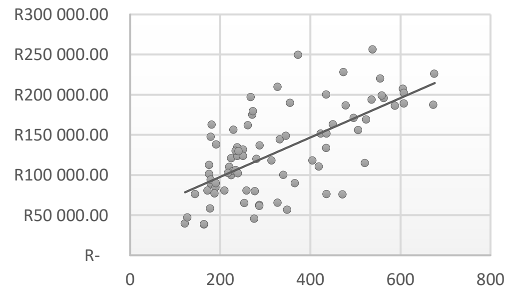
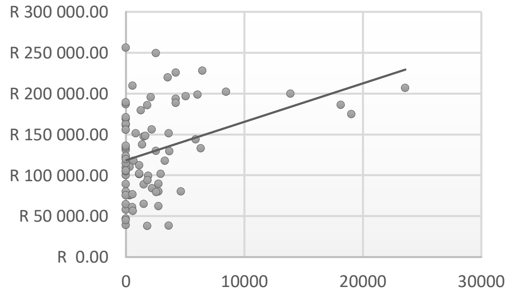
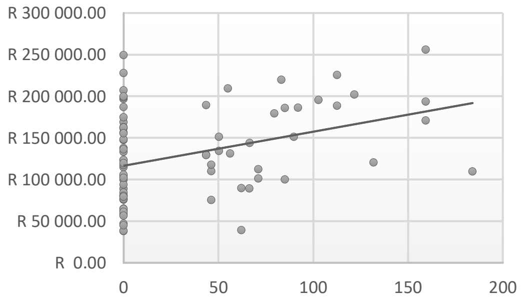
The GTV adjusted for BEE, this being SGTV, was acquired using the above equation. Where P_{SGTV} is the calculated scaled generalised tender value (SGTV). S_{price} is the supplier pricing score that was presumed to be 80 for this study. S_{BEE} is the supplier BEE score out of 20 and last, P_{GTV} is the actual awarded procurement price after CPI correction.

Using the *Microsoft Office Excel 2013 Pearson Function* and the *model-building data*. It was found that the correlation between the anchor predictor and the GTV (*without* the BEE score change) is *65.25%* (0.652515981). Whereas, the SGTV (*with* the BEE score change) is *66.12%* (0.661218742). This means the correlation has improved by a percentage (~1%) when factoring in a supplier's BEE score, ahead of the already three per cent (~3%) improvement by factoring CPI in the pricing calculations.

Therefore, the research target variable (dependent variable) is the SGTV. As a result, the *Pearson's Correlation Coefficient test* for all the refined independent variables were plotted against the SGTV. This was achieved using the *Microsoft Office Excel 2013 Pearson's Function* and the *model-building data* paired with a scatter plot diagram for each of the independent variables against the SGTV.

Table 36 is a visual and numerical depiction of each *model building data*'s independent variable's significance to the SGTV using *Pearson's Correlation Coefficient test*. It includes the graph generated through the aid of *Microsoft Office Excel 2013* (SGTV as the y-axis variable), and the commentary on the ideal transformation found followed by a description of the degree of significance. The transformed data table summaries can be found in Appendix C, Section 2.

Table 36. Depiction of Each Independent Variable's Significance to The SGTV and Commentary

<table>
<tr><th>Graph</th><th>Comments</th></tr>
<tr><td>
</td><td>The graph depicts the relationship between the removal of the old fence (ROF) and scaled generalised tender value (SGTV) [y-axis]. The ideal transformation was an exponent of two. This meant the data for ROF followed a square rooted trend when plotted against the SGTV. The Pearson's Correlation Coefficient result is 30.82% (0.308169991), a moderate degree of significance by this variable.</td></tr>
<tr><td>
</td><td>The graph depicts the relationship between fence material cost estimate (FMCE) and SGTV. The ideal transformation was square rooting. This means the data for FMCE followed a quadratic trend when plotted against the SGTV. The Pearson's Correlation Coefficient result is 66.12% (0.661218742), a high degree of significance by this variable which justifies its selection as the anchor predictor.</td></tr>
<tr><td>
</td><td>The graph depicts the relationship between hinged gate material cost estimate (HGMCE) and SGTV. There was no ideal transformation found. This means the data for HGMCE was already linear when plotted against the SGTV. The Pearson's Correlation Coefficient test result is 37.05% (0.370527885), a moderate degree of significance by this variable.</td></tr>
<tr><td>
</td><td>The graph depicts the relationship between sliding gate material cost estimate (SGMCE) and SGTV. The ideal transformation was square rooting. This means the data for SGMCE followed a quadratic trend when plotted against the SGTV. The Pearson's Correlation Coefficient test result is 36.56% (0.365644685), a moderate degree of significance by this variable.</td></tr>
</table>

The above results show how all the independent variables have an identifiable linear relationship with the dependent variable. This satisfies the first assumption/limitation of linearity for building a multiple linear regression model. However, either of these variables may still be discarded because of violating the remaining two assumptions.

These are, as explored in the previous chapter, assumed non-multicollinearity, neither one of the independent variables should have a high correlation with another; and assumed homoscedasticity, the variance of error terms should be similar across the values of the independent variables.

4.1.1.4 Data Management Summary

This section is a breakdown of the data collected and the processing of the results. The first section shows how the collected data was split into four segments; the *model-building data*, the model testing data, the *model out-of-range data* and the *benchmarking data*. Thereafter, how the data violated the sampling parameters of *consistency* and *diversity*. The implications for this are explored in chapter five.

On further processing of the data, it was found the inclusion of a suppliers BEE score to the CPI-adjusted pricing improved the linearity of the data by a percentage. This led to it being included in the dependent variable (awarded pricing) for the duration of the study. The implications for confirming the direct relationship between pricing received and a suppliers BEE score are profound and further analysed in chapter five.

Table 37. Summary of Pearson's Correlation Coefficient for Each Independent Variable

Independent Variable	**Correlation to SGTV**	**Transformation Used**
Removal of Old Fence (ROF)	(0.31) 30.82%	Squared
Fence Material Cost Estimate (FMCE)	*(0.66) 66.12%*	Square rooting.
Hinged Gate Material Cost Estimate (HGMCE)	(0.37) 37.05%	None
Sliding Gate Material Cost Estimate (SGMCE)	(0.37) 36.56%	Square rooting.

The theory of the *fencing material cost estimate* (FMCE) being the anchor predictor was confirmed through the obtained results. However, it was the only independent variable with a high correlation above +0.5 to the dependent variable (SGTV) as shown in Table 37. The table displays the correlation each independent variable had to the pricing, and the transformation used on the variable to find it. These results obtained in this subchapter are satisfactory and meet the minimum requirements explored in the previous chapter.

4.1.2 Model Preparation

The study has followed the *hierarchal regression* model building procedure aided by the adjusted $\mathbf{R^2}$. *Hierarchal regression* is a sequential building of several multiple linear regression models, each adding or removing independent variables (Petrocelli, 2003; Grace-Martin, 2020). The independent variables are based on prior knowledge of the field and are ranked from the highest to the lowest level of importance by the modeller, in which the highest rank is the anchor predictor, this being the fencing material cost estimate [FMCE].

4.1.2.1 Cases of Multicollinearity and Testing for Homoscedasticity

4.1.2.1.1 Multicollinearity Test Results

Two tests were conducted for multicollinearity, the first by acquiring the *Pearson's Correlation Coefficient* for every independent variable pairing. As previously stated, a correlation between -0.5 and +0.5 is idyllic for the study as this means there is a non-high degree of correlation between the independent variables. Second, the variance inflation factor (VIF) was observed for each of the variables and the ideal value should be a number below 5 for each independent variable.

Table 38. Multicollinearity Correlation Test and Variance of Inflation Results *(sample size = 80)*

Independent Variable	**ROF**	**FMCE**	**HGMCE**	**SGMCE**
Removal of Old Fence (ROF)	*n/a*	0.233	0.084	0.017
Fence Material Cost Estimate (FMCE)	0.233	*n/a*	0.304	0.364
Hinged Gate Material Cost Estimate (HGMCE)	0.084	0.304	*n/a*	0.001
Sliding Gate Material Cost Estimate (SGMCE)	0.017	0.364	0.001	*n/a*
Variance of Inflation Factor (VIF)	1.064	1.361	1.119	1.179

Table 38 are the results for the multicollinearity test. As has been shown, SGMCE and HGMCE have a moderate correlation with FMCE. Though the correlation is visible, the significance is low enough to be ignored as it has not introduced cases of multicollinearity as all VIFs are kept well below 5. The test results obtained from *IBM SPSS Statistics 26* are available in Appendix C, Section 3 and Section 4.

4.1.2.1.2 Homoscedasticity Test Results

The below Figure 14 is a plot of standardised residuals to test for homoscedasticity aided by *IBM SPSS Statistics 26 Residual for Standardised Plot*. As explained, the data does not follow a particular trend. However, it favours a more squared trend. As a result, it has passed the study assumption of homoscedasticity.

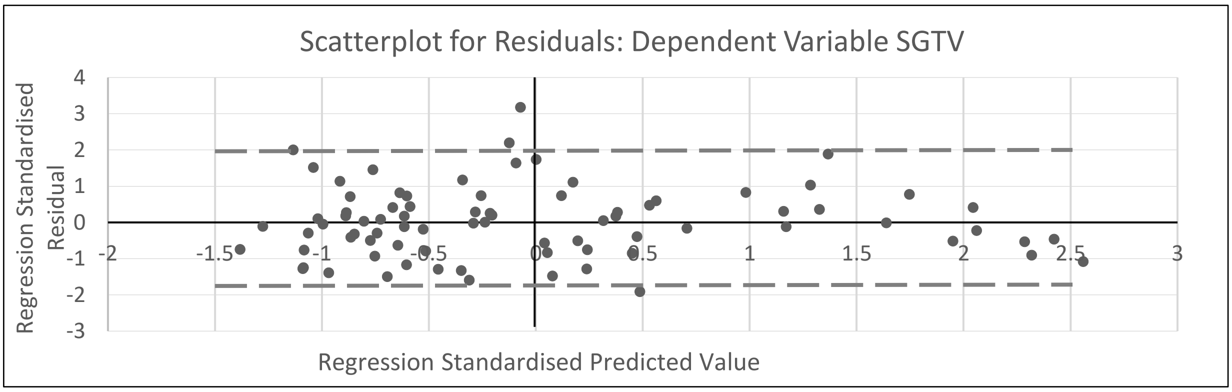


Figure 14. Residual for Standardised Scatterplot to Test for Homoscedasticity

4.1.2.2 Removal of Non-Significant Variables

All the above-mentioned independent variables have a noteworthy significance to the SGTV as explored earlier, and there are no valid cases of multicollinearity or heteroscedasticity. The subsequent step is to sequentially build several multiple linear regression models. This is to identify which independent variable combination has given the most stable and accurate outcome.

First, the model's stability was acquired by assessing the coefficient of each independent variable's (**β**) modelling significance to the study by calculating the **P**-value and **t**-value. If a coefficient displayed a **P**-value lower than 0.05 and a **t**-value above +2 or below -2; it meant the independent variable was significant to the model. Second, the model's accuracy was attained by observing the adjusted $\mathbf{R^2}$ and an adjusted $\mathbf{R^2}$ of ±0.5 is viewed as high degree ergo, ideal for the study.

Figure 9 was used with the results generated in Table 39 to identify which independent variables would be ideal for the model building stage*.* Eight test models were built as depicted in the table and the anchor predictor was not removed throughout the various model builds. Every model observed adjusted $\mathbf{R^2}$; **P-**value and **t-**value for each independent variable's coefficient (**β**).

The results show that the best model when assessing adjusted $\mathbf{R^2}$ is the model that includes all the independent variables. P-values are maintained below 0.05 and **t**-values above +2 for all coefficients. This has resulted in all independent variables being passed for the study model building process. The data summary results that are inclusive of the *beta* values for each iteration are available in Appendix C, Section 4.

Table 39. Results for Testing Several Model Builds (**P**-value and **t**-value)

	β_{ROF}		β_{FMCE}		β_{HGMCE}		β_{SGMCE}		Adjusted
	P	**t**	**P**	**t**	**P**	**t**	**P**	**t**	$\mathbf{R^2}$
Model One	0.039	2.102	0.000	5.273	0.017	2.446	0.037	2.121	0.496
Model Two	-	-	0.000	5.786	0.019	2.402	0.057	1.929	0.474
Model Three	0.44	2.047	0.000	6.174	-	-	0.079	1.778	0.463
Model Four	0.060	1.908	0.000	6.462	0.034	2.154	-	-	0.473
Model Five	-	-	0.000	6.742	-	-	0.114	1.597	0.441
Model Six	0.062	1.893	0.000	7.252	-	-	-	-	0.448
Model Seven	-	-	0.000	6.934	0.035	2.143	-	-	0.455
Model Eight	-	-	0.000	7.784	-	-	-	-	0.430

4.1.2.3 Model Preparation Summary

This section has explored the required model building preparatory steps to construct a credible regression model as identified in chapter two of the *Data Science Process* by Manning Publications (2020). These included both the identification of multicollinearity cases and the removal of insignificant variables from the *model-building data* through the aid of the variable hierarchal ranking in Figure 9.

The results showed there is a noticeable correlation between SGMCE and HGMCE with FMCE. However, the correlation did not have enough significance to introduce multicollinearity. This is because *Pearson's Correlation Coefficients* were maintained between -0.5 and +0.5 and the VIF values for all the independent variables were kept well below 5. As an effect, all independent variables have passed this assessment and the data satisfied the homoscedasticity assumption test.

No independent variables were removed from the *model-building data* as they all had significance to the regression model building. This was identified by building several regression models by removing and adding specific independent variables. The results showed the best model that generated the highest accuracy (*adjusted* R^2) was when all the independent variables were included. The acquired *adjusted* R^2 was 0.496 (49.6% accuracy) which meets the set study standard of ±0.5 as explored earlier in the previous chapter four.

These results are satisfactory as the *model building data* shows no signs of violating any of the multiple linear regressing limitations/assumptions as explored in chapter three. This means all the model-building data that was refined for the design portion of the study may be exported and used for building the low-valued "*supply and installation of fencing*" prediction model in the next section of the chapter.

4.1.3 Construction and Validation of Model

The section is a detailed breakdown of how the model was built and assessed using the refined variables that made it through the previous processes. The model coefficients were obtained through automated calculation on *IBM SPSS Statistics 26.* As stated in chapter three; **F**-statistic, adjusted R^2 and **P**-value were the quality assurance indicators selected for this study to test for model significance.

F-statistic had to be greater than 2.01 (critical **F-**value) as an indication that at least one of the coefficients of the independent variables is useful for predicting the outcome (Aggarwal, 2020). This is based on the condition that the modelled regression **P**-value would be lower than 0.05. The adjusted $\mathbf{R^2}$ was maintained at the set study standard of ±0.5 as explored in the previous section at the value of 0.496.

The last stage of the section is the 10-folds cross-validation technique used to validate the constructed model. The results for this process is a rooted average mean squared error (RMSE) that is compared against the models' standard estimation error (SEE) generated by the *IBM SPSS Statistics 26* software package. As stated in the previous chapter, ideally SEE and RMSE should not have a deviation that is greater than ±1%, as this is the set standard for the study to confirm the reliability of the model based on the sample used.

4.1.3.1 Calculation of the Coefficients

The calculation of the model coefficients as mentioned in the previous chapters was automated using *IBM SPSS Statistics 26*. Table 40 supplies the variables required to formulate the final equation for estimating "*Supply and Installation of Fencing"* procurements valued at less than R230 000 based on a scope of work that includes removal of old fence (ROF), fence material costs estimate (FMCE), hinged gate material costs estimate (HGMCE) and sliding gate material costs estimate (SGMCE). The results from *IBM SPSS Statistics 26* are available in Appendix C, Section 4.

Table 40. Calculated Regression Model Coefficients

Description	**Coefficient**	**Unit**	**Value**
Removal of Old Fence	(β_1)	*Rand/Metres*^{4}	1.061
Fence Material Costs Estimate	(β_2)	*Rand*$^{0.5}$	182.362
Hinged Gate Material Costs Estimate	(β_3)	*Unit less*	2.622
Sliding Gate Material Costs Estimate	(β_4)	*Rand*$^{0.5}$	205.450
Constant	(β_0)	*Rand*	52594.925

4.1.3.2 Test for Significance of Regression

Table 41 summarizes all the model quality assurance tools used for this study. It includes the set targets, the corresponding results and remarks on the results. The models acquired **P**-value is 0.000, the **F-**statistic is 20.463 and adjusted $\mathbf{R^2}$ is 0.496 for the model built; meaning it has passed the significance test. The summary of the results obtained is available in Appendix C, Section 4.

Table 41. Summary of Findings for Model Quality Assurance Tools

Item Name	Model Target	Model Value	Remark
***F**-statistic*	Greater than 2.01	20.463	Passed
***P**-value*	Lower than 0.05	0.000	Passed
adjusted R^2	±0.5 or greater	0.496	Passed

4.1.3.3 10-Folds Cross-Validation

The model validation was conducted using 10-fold cross-validation on the refined *model building data* available in Appendix C, Section 2. Various models were built using the prescribed methodology explored in chapter three. These were only used to calculate the mean squared error (*MSE*) for every iteration, thereafter discarded.

Table 42 depicts all the calculated *MSE* values*,* followed by the average *MSE* for all the models. This being the cross-validations result for every iteration. Included below, is the rooted average MSE value (RMSE) and the standard estimation error (SEE) for the research model built as shown in Appendix C, Section 4*:*

Table 42. Results for 10-Fold Cross-Validation

Iteration Number	Unit	Mean Squared Error Value
One	Rand Squared	2828242139.46
Two	Rand Squared	1636302324.57
Three	Rand Squared	691355385.66
Four	Rand Squared	1713138050.41
Five	Rand Squared	441487620.87
Six	Rand Squared	651591420.77
Seven	Rand Squared	308467797.12
Eight	Rand Squared	1623989295.08
Nine	Rand Squared	2210883763.23
Ten	Rand Squared	2943514329.25
Average MSE	1504897212.64	
RMSE	R 38 793.00	
SEE	R 38 469.41	
Difference	R323.59	
Deviation	0.841% (RMSE above SEE)	

As previously stated, the SEE and RMSE should not have a large deviation (assumed ±1% deviation) to show the dataset used to build the model was satisfactory. Based on the values depicted in the table, the deviation is 0.841% between the values. This shows the model built with the samples is robust and can be used for its intended purpose of benchmarking.

4.1.3.4 Construction and Validation of Model Summary

This section describes the model building steps and subsequent results. Included, is the calculation of the equation coefficients (**β**), and the applicable model quality assurance tools as advised by Montgomery and Runger (2011). These include the **F**-statistic, adjusted R^2 and **P**-value. Thereafter, the 10-fold cross-validation method to authenticate the reliability of the data set was used to build the model.

The results showed all the quality assurance tools had passed the set standards. This was depicted by **F**-statistic being greater than 2.01 (critical **F-**value), adjusted R^2 being maintained at ±0.5 and **P**-value being less than 0.05. Though these results were fitting, they were not enough to pass the model. The 10-fold cross-validation method was required to validate the constructed model by assessing the sample used to build it.

The success of the cross-validation was judged based on the deviation between the rooted average mean squared error (RMSE), and the built models' standard estimation error (SEE). The study target deviation was placed at an assumed ±1% for the validation to be viewed as a success. This was accomplished as the acquired deviation was 0.841%. All the above results are acceptable ergo the model built is robust and may be exported to the next section for the benchmarking assessment.

4.1.4 Construction and Validation of Prediction Interval

In this section, the construction of the benchmarking model has taken place. As explored in the previous chapter, this was achieved by predicting the prices of previously awarded procurements by non-clean audited municipalities and comparing them against the BEE scaled and CPI-adjusted awarded procurement cost (SGTV).

However, before this could be done. The benchmarking model required further authentication by using the *model testing data*. Thereafter, assess whether the model could evaluate procurements above the R230 000 limit by using the *model out-of-range data*. Subsequently, is the benchmarking assessment to identify pricing deviations that may have taken place by using the *benchmarking data.*

4.1.4.1 Prediction of Prices

The refined datasets from the above three mention data segments were imported into this section for price predictions. The price predictions were obtained using the coefficients calculated earlier in the chapter. Below is the notation of the equation used to acquire these predictions:

$$\boldsymbol{Y} = (52594.925) + (1.061) \times \boldsymbol{X_1} + (182.362) \times \boldsymbol{X_2} + (2.622) \times \boldsymbol{X_3} + (205.450) \times \boldsymbol{X_4}$$

(Equation 19)

Table 43 depicts the price prediction results for the *benchmarking data*. The *model testing data* and *model out-of-range data* results were excluded in the chapter and instead attached in Appendix C, Section 5 because of the quantity of the results. As shown below 66.6% of the sample has a minor deviation to the predicted value and the remaining have an above 40% deviation. Though significant, these results are irrelevant without the prediction interval. The next section in the chapter has given a better interpretation of what these results entail.

Table 43. Summary of the Price Predicted Results for Benchmarking Data

Identifier	Awarded Value (SGTV)	Predicted Value	Deviation Above Predicted
A	R 214 571.79	R 246 521.86	-12.96%
B	R 128 169.66	R 119 746.20	7.03%
C	R 206 809.92	R 144 731.16	42.89%
D	R 63 180.99	R 144 990.60	-56.42%
E	R 258 184.30	R 263 958.59	-2.19%
F	R 247 485.28	R 171 477.87	44.32%
G	R 156 342.86	R 131 046.55	19.30%
H	R 178 408.01	R 337 007.37	-47.06%
I	R 184 171.69	R 214 051.48	16.22%
J	R 208 351.62	R 185 681.53	-10.88%
K	R 35 430.03	R 83 558.32	135.84%
L	R 95 483.73	R 97 453.41	2.06%

4.1.4.2 Prediction Interval

This section has developed the prediction interval for *benchmarking datasets*. The study has followed a simplified approach for most of the required calculations. As a result, the below equations were derived from the previous chapter to calculate the prediction interval, they include both the lower and upper bound:

$$P_{LB} = P_{pred} - t_{stat} \times SEE \quad \text{(Equation 20)}$$

$$P_{UB} = P_{pred} + t_{stat} \times SEE \quad \text{(Equation 21)}$$

In the above equations, P_{LB} is the pricing lower bound; P_{pred} is the model predicted amount, and P_{UB} is the pricing upper bound. The *t* distribution is t_{stat} and calculated using an **α**-value of 0.10 and the model degrees of freedom for the residuals (df) available in the model summary results in Appendix C, Section 4. The SEE has already been acquired from the previous sections. The calculated values for both these are 1.6654 for t_{stat} based on a *df* of 75 and 38469.41 for the standard error estimate.

The next subsections depict the summaries of the prediction interval results for all three previously mentioned datasets. The range between the upper and lower bound of the predicted value are portrayed by the shaded grey area, and the corresponding awarded value with the relevant transformations displayed by the dark greyed diamond. The lower bound was restricted to a minimum figure of zero as procurement values cannot be priced negatively. Lastly, the x and y-axis show the contract identifier for the study and the value, respectively.

4.1.4.2.1 Prediction Interval Results for the Model Testing Data

Figure 15 depicts the results of the model testing process. This was achieved by randomly selecting clean audited procurements from the *model-building data* that did not satisfy the 10-folds cross-validation minimum sample. As this was the only reason for its exclusion, it is assumed one of the eight samples may fall outside the prediction range based on a 90% confidence level.

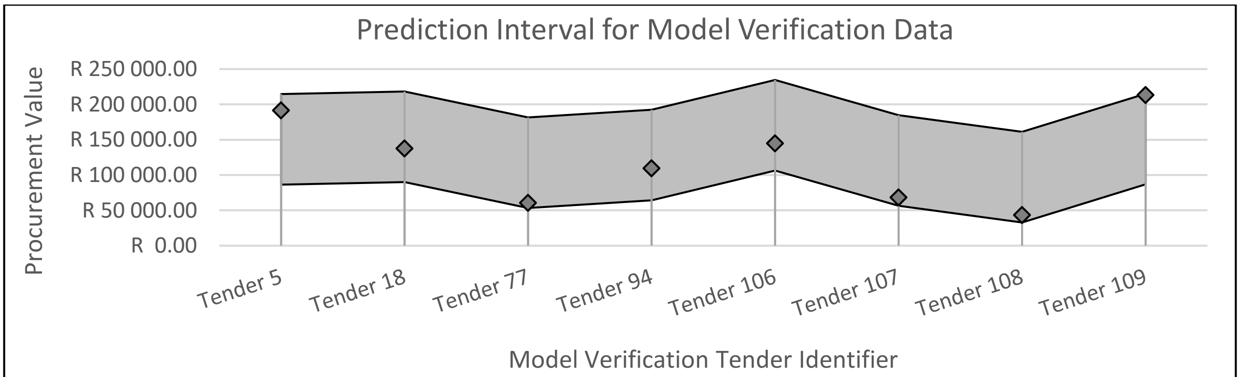


Figure 15. Prediction Interval and SGTV for Model Testing Data

All eight model testing samples passed the assessment, as they are within the prediction interval, ergo the model is passed. Tenders 77, 107, 108 and 109 are close to their respective edges. Based on the data in Appendix C, Section 5, they are the only ones of all the assessed procurements that have an above 40% deviation from the predicted value (-48.33%, -43.42%, -55.46% and +41.42% respectively).

4.1.4.2.2 Prediction Interval Results for the Model Out-of-Range Data

Figure 16 illustrates the results from the model out-of-range process to assess whether the model could evaluate open competitive contracts. This was achieved by collecting and sorting through clean audited contracts above the R230 000 limit (*model out-of-range dataset*). Thereafter, applying the relevant transformations.

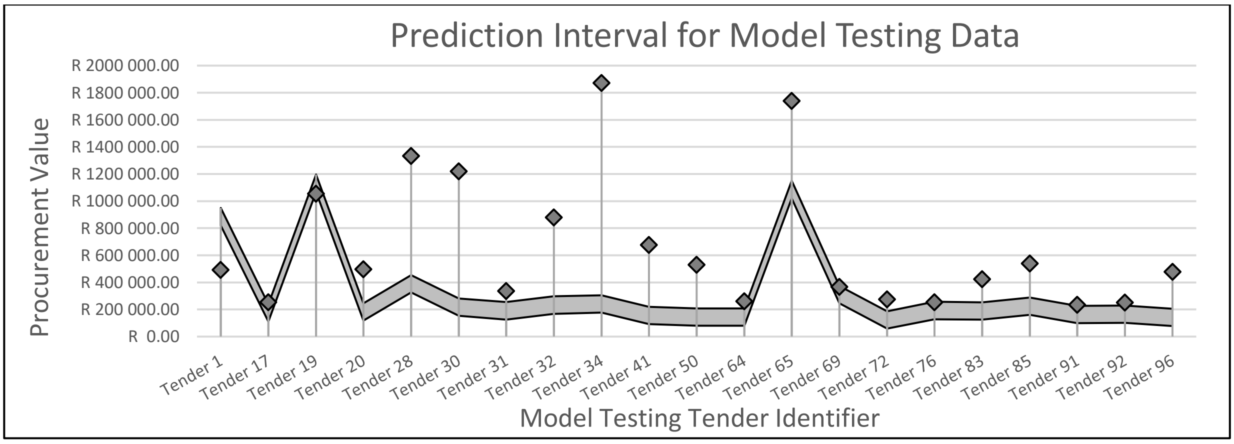


Figure 16. Prediction Interval and SGTV for Model Out-of-Range Data

Of all the 21 assessed procurements, only three fell within the prediction range. This shows the model has a success rate of 9.52% for out of price range data. For this reason, the model fails the out of price range assessment, and should therefore not be used to assess open and competitive "supply and installation of fencing" contracts. The data used to construct the figure is available in Appendix C, Section 5.

4.1.4.2.3 Prediction Interval Results for the Benchmarking Data

The last item for this subsection is the assessment of non-clean audited municipalities to answer the critical research question (CRQ). This was achieved by using the predicted contract amounts and the prediction interval to assess how clean audited municipalities would have priced the same procurement.

Table 44. Prediction Interval Results and Remarks for Benchmarking Data

Identifier	SGTV	Lower Bound	Upper Bound	Remark	Margin
A	R 214 571.79	R 182 453.92	R 310 589.80	None	N/A
B	R 128 169.66	R 55 678.26	R 183 814.14	None	N/A
C	R 206 809.92	R 80 663.22	R 208 799.09	None	N/A
D	R 63 180.99	R 80 922.66	R 209 058.54	Under Paid	21.92%
E	R 258 184.30	R 199 890.66	R 328 026.53	None	N/A
F	R 247 485.28	R 107 409.93	R 235 545.81	Over Paid	5.07%
G	R 156 342.86	R 66 978.61	R 195 114.48	None	N/A
H	R 178 408.01	R 272 939.43	R 401 075.31	Under Paid	34.63%
I	R 184 171.69	R 149 983.54	R 278 119.42	None	N/A
J	R 208 351.62	R 121 613.60	R 249 749.47	None	N/A
K	R 35 430.03	R 19 490.38	R 147 626.26	None	N/A
L	R 95 483.73	R 33 385.47	R 161 521.35	None	N/A

Table 44 shows the awarded procurement value (SGTV) and the calculated upper and lower bounds based on the predicted procurement amounts. Thereafter, remark on the results obtained and if applicable, the percentage above the upper bound or below the lower bound. Using these results the below Figure 17 was obtained.

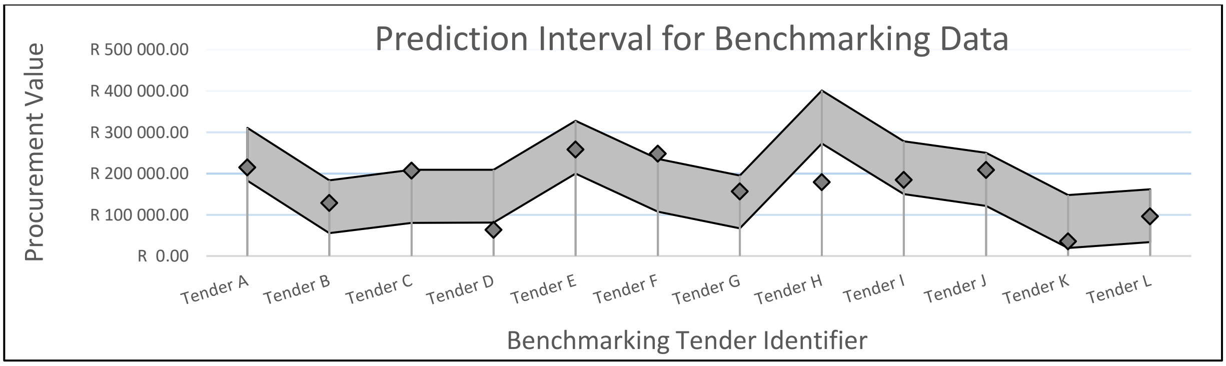


Figure 17. Prediction Interval and SGTV for Benchmarking Data

As shown above, out of the twelve samples assessed three fell out of range. This being Tender D, Tender F and Tender H (under-priced 21.92%, over-priced 5.07% and under-priced 34.63% respectively). In total 25% of the benchmarking sample data deviated from the model built. These results though complete are inconclusive at this stage as they required further analysis that is conducted in the following chapter five.

4.1.4.2.4 Supplementary Prediction Interval Results for the Model Building Data

A supplementary test was conducted to identify which of the clean audited *model building data* deviated at a 90% confidence level. This was achieved by using the prediction interval to assess the *model-building data*. Six contracts fell outside the prediction interval as depicted in Table 45.

Table 45. Prediction Interval Model Building Data Assessment

Identifier	SGTV	Lower Bound	Upper Bound	Remark	Margin
8	R 75 227.46	R 85 073.90	R 213 209.77	Under Paid	11.57%
36	R 249 197.40	R 63 256.60	R 191 392.48	Over Paid	30.20%
51	R 162 390.36	R 21 628.25	R 149 764.13	Over Paid	8.43%
67	R 196 648.45	R 66 098.17	R 194 234.05	Over Paid	1.24%
86	R 255 888.52	R 119 528.02	R 247 663.90	Over Paid	3.32%
101	R 209 359.18	R 61 202.50	R 189 338.37	Over Paid	10.57%

As the selected confidence level for the study is 90%, it was expected that ±90% of the sample would fall within the prediction interval. The results showed the 92.5% (74 procurements) of the sample fell within the interval and depicts which samples from the clean audited municipalities deviated in pricing.

4.1.4.3 Construction and Validation of Prediction Interval Summary

This section describes the prediction interval construction process and the sequential results that followed. Included is the construction and analysis of various prediction intervals using the segmented data samples identified earlier in the chapter which included the *model testing data*, *model out-of-range data* and lastly the *benchmarking data* to be used to answer the critical research question.

On assessment of the *model testing data,* it was found that the entire dataset fell within the prediction interval. This while four past procurements were fairly close to the boundary and had an above 40% deviation from their predicted values but the model was still viewed as reliable and used to analyse the remaining datasets.

The *model out-of-range data* to assess whether the model could evaluate open and competitive procurements that have a greater than R230 000 (incl. VAT) failed. This was shown by only 9.52% of the assessed data being within the prediction interval. Therefore, the fencing model may only be used for low priced procurements. To finish, the procurements in Table 46 are the ones that deviated from both model building and benchmarking data. However, these results are only analysed in chapter five.

Table 46. Summary of All Tenders That Are Viewed as Pricing Deviations by the Model.

Identifier	SGTV	Remark	Margin	Source	Location
8	R 75 227.46	Under Paid	11.57%	Clean Audited	Western Cape
36	R 249 197.40	Over Paid	30.20%	Clean Audited	Western Cape
51	R 162 390.36	Over Paid	8.43%	Clean Audited	Western Cape
67	R 196 648.45	Over Paid	1.24%	Clean Audited	Western Cape
86	R 255 888.52	Over Paid	3.32%	Clean Audited	Western Cape
101	R 209 359.18	Over Paid	10.57%	Clean Audited	Gauteng
D	R 63 180.99	Under Paid	21.92%	Non-Clean Audited	Gauteng
F	R 247 485.28	Over Paid	5.07%	Non-Clean Audited	Gauteng
H	R 178 408.01	Under Paid	34.63%	Non-Clean Audited	Gauteng

4.2 Results Summary

This chapter has presented all the findings for the study in the logical form laid out in the research methodology. It was divided into four subsections the first of which was the data collection and preparation step. Thereafter, the model preparatory steps included the short-listing of significant independent variables. Subsequently, the model building and validation stage through the 10-folds cross-validation method. Last, the construction and validation of the prediction interval to be used for benchmarking.

Table 47. Overall Research Findings Summary and Relevant Commentary

<table>
<tr><th>Section</th><th>Subsection</th><th>Results</th></tr>
<tr><td rowspan="2">Data Management</td><td>Data Collection</td><td>On analysis, the data revealed there are multiple sampling parameter violations including Diversity and Consistency in the data and the effects of these violations have been analysed in the next chapter.</td></tr>
<tr><td>Date Preparation</td><td>First, the subsection passed the first MLR assumption by all independent variables having an identifiable linear relationship with the pricing. Second, it confirmed the impact of a supplier's BEE level on the awarded pricing.</td></tr>
<tr><td rowspan="2">Model Preparation</td><td>Multicollinearity & Homoscedasticity Results</td><td>None of the independent variables had multicollinearity present by passing Pearson's Correlation Coefficient test and variance inflation factor assessment. As an effect, all the independent variables have passed this assessment and the data satisfied the homoscedasticity test.</td></tr>
<tr><td>Removal of Non-Significant Variables</td><td>The subsection showed the best model is the one that uses all the independent variables. This resulted in an adjusted $\mathbf{R}^2$ value of 0.496 (49.6% accuracy). It also validated the selection of FMCE as the anchor predictor by it being the only independent variable with a P-value below 0.000 and t-value above 5</td></tr>
<tr><td rowspan="3">Construction And Validation Of Model</td><td>Calculations of Coefficients</td><td>All coefficients were calculated as seen in the below equation: $\boldsymbol{Y} = (52594.925) + (1.061) \times \boldsymbol{X_1} + (182.362) \times \boldsymbol{X_2} + (2.622) \times \boldsymbol{X_3} + (205.450) \times \boldsymbol{X_4}$</td></tr>
<tr><td>Test for Significance</td><td>The results showed all the quality assurance tools had passed by F-statistic being greater 2.01, adjusted $\mathbf{R}^2$ being maintained at ±0.5 and P-value being less than 0.05.</td></tr>
<tr><td>10-Folds Cross Validations</td><td>The validation was a success by the generated rooted average validation generated error (RMSE) having a deviation of less than 1% when compared against the model generated error.</td></tr>
<tr><td rowspan="4">Construction And Validation Of Prediction Intervals</td><td>Model Testing Results</td><td>The reliability of the interval was confirmed by using the model testing data that had a 100% prediction success rate</td></tr>
<tr><td>Model Out-of-Range Results</td><td>The model failed to accurately predict the pricing of procurements that exceeded the value of R230 000; meaning it may not be used to assess open and competitive contracts.</td></tr>
<tr><td>Benchmarking Results</td><td>The assessment of the benchmarking data was limited to twelve samples of which three fell out of their respective prediction interval ranges.</td></tr>
<tr><td>Supplementary Test Results</td><td>Six contracts from the model-building data fell outside the prediction interval. It was expected that ±90% of the sample would fall within the prediction interval. The results showed the 92.5% of the sample fell within the range which meets the study expectations.</td></tr>
</table>

Table 47 summarizes the crucial findings in this chapter. Included are the section and subsection labels, the noted results and their impact on the study. To conclude the chapter, the results obtained have shown that the use of the multiple linear regression technique was ideal as it succeeded in all the verification and validation steps, and satisfied all the modelling assumptions.

Chapter Five: Analysis and Discussion

This chapter is a detailed analysis of the results obtained in chapter four using the layout developed in the methodology. The first section is an analysis and critiquing of the intermediate results, this includes the data management, preparatory model building requirements and model building results. The section has attempted to link the findings to the existing literature explored in chapter two.

Thereafter, an examination of the findings from the benchmarking, which includes the model testing and out-of-range assessment, supplementary test findings and benchmarking outcomes. Subsequently, critiquing the study assumptions and limitations to better understand the reasons behind the outcome. Last, answering the critical research question to pass or fail the below-listed research hypothesis.

H_0: *Clean Audited Municipalities in South Africa Do Not Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng*

H_1: *Clean Audited Municipalities in South Africa Procure Identical Low-Valued Goods and Services at a Lesser Price than Non-Clean Audited Municipalities in Gauteng*

This was achieved by analysing the benchmarking results obtained in chapter four; through assessing whether the model was built using clean audited municipal *supply and installation of fencing* as the sample procurements, had flagged any non-clean audited municipal procurements from Gauteng. Figure 18 is a recap of the methodology to better understand the structure of this chapter.

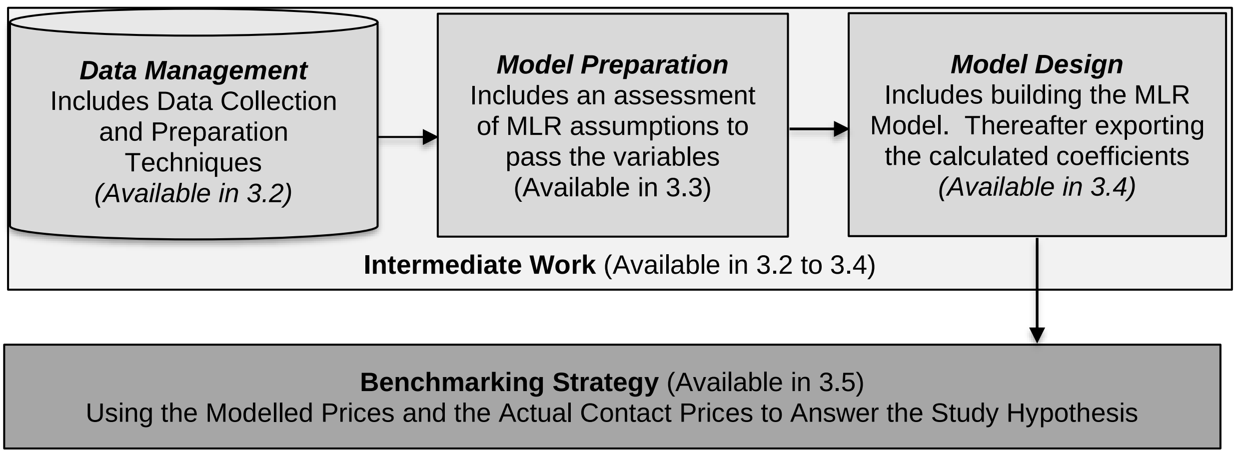


Figure 18. Research Methodology Recapitulation

5.1 Analysis of Intermediate Findings

5.1.1 Data Management

This subsection is an analysis of the complete research data organisation. Primarily broken down into two segments that include data collection and preparation. Two of the three data sampling parameters are used to critique the sampling biases introduced by using the *quota sampling method* (QSM). Thereafter, a partial linking of the reviewed literature in chapter two to the results obtained in chapter four.

5.1.1.1 Data Collection

A secondary data collection procedure from municipal archives was followed as outlined in the methodology. It was found the contracts from the clean audited municipalities to *build* and *testing* the model were skewed by awarded price, date, and location in various forms. Table 48 summarises all the noted violations followed by a detailed breakdown of the relevant observations and/or remedial steps to each violation.

Table 48. Summary of Incurred Sampling Violations and Effect on Outcomes

Item	Violation Description	Extent	Effect on Outcomes
Awarded Price	Skewed by contract values ranging from R100 000 to R230 000	61.36 %	Violated the *diversity* parameter, would lead to the model predicting some price ranges with high accuracy and others poorly.
Date	Skewed toward recent years ranging from 2018 to 2020	73.86%	Violated the *diversity* and *consistency* parameter, would “blind” the model to real-world changes over the years which would invalidate the results.
Location	Skewed toward the Western Cape Province	92.05%	Violated the *diversity* parameter, the model would effectively be a Gauteng to Western Cape comparison instead of Gauteng to South Africa.

The above data indiscretions violated the parameter of *diversity* and *consistency*. As explored by Gaurav (2017), violating diversity leads to time and location biases in favour of certain samples, and breaching the parameter of *consistency* leads to “blindness” of real-world changes in the collected sample. However, remedial steps were taken such as consumer price index (CPI) changes as explored below.

The data being marginally skewed toward upper valued procurements violated the *diversity* parameter because it favoured greater valued low-cost procurements; but the *benchmarking data* was highly skewed toward the same range. Furthermore, the data followed normality in distribution which shows a fairly robust sample when building a multiple linear regression (MLR) model (Glen, 2015). Therefore, it is fair to conclude that this violation had not played a crucial role in the research outcomes.

The data skewed by *date* violated two parameters; *diversity* because the data favoured specific years, and *consistency* as there were various changes through the years not reflected in the pricing. The remedial step taken was the inclusion of the CPI. As evident in the research by Biggeri *et al*. (2017) that found using CPI data for spatial comparisons is preferable. These findings are substantiated by an improved correlation between the anchor predictor and the awarded pricing of above three per cent (~3%). Therefore, these violations had too, not played a critical role in the research outcomes.

The data skewed by location violated *diversity* as this favoured procurements from the Western Cape. This may convey the image that the developed benchmarking model is a comparison of non-clean audited municipalities in Gauteng against clean audited municipalities in the Western Cape rather than South Africa. The model building and testing data comprised procurements from other provinces. Though small, this introduced an element of impartiality to the model. This impartiality is further immersed into the model by the use of the prediction interval (Frost, 2020a). Therefore, it is fair to conclude this violation had too, not played a vital role in the research outcomes.

5.1.1.2 Data Preparation

The data preparation included the already mentioned CPI transformations of the awarded contract price (incl. vat) to get the generalised tender value (GTV). However, as found by the literature revised in chapters one and two, a higher BEE score (e.g. a score of 20 as opposed to 12) of a supplier is likely to increase the pricing of a procurement (Lowman, 2017). Thus, an additional test was conducted, as explored in the methodology, to incorporate the BEE score in the GTV to calculate the SGTV.

The test results show that the statement derived from the literature is true. This as including a supplier's BEE score to the pricing improved the correlation between the anchor predictor and the pricing by a percentage as it explained more variance in the model. The above results served as a complimentary validation for the findings.

On further analysis of the findings, as expected, the data revealed the combined variable of fencing material cost estimate (FMCE) [anchor predictor], which consists of the type of material, the length and the height of a fence had the highest correlation to the pricing of a contract (Lumber, 2020). The results meeting the theoretical expectation shows that the method used to combine the variables was appropriate.

5.1.1.3 Data Management Analysis Summary

This section shows how the various sampling parameters had been violated during the model building stages. Potentially adding the author's biases to the research and this being an issue associated with the q*uota sampling method* (Aprameya, 2016). However, the remedial steps taken such as the use of prediction intervals and CPI changes have revealed that the biases were not neutralised but minimised for the study.

On further analysis of the results, the confirmation of BEE having a direct impact on pricing was made. This discovery aligning with the literature reviewed by Lowman (2017) and Ambe (2012). The theory on the type of material, the length and the height of a fence having the biggest influence in fencing procurements aligned with the study findings. This providing additional validations for the credibility of the model built.

5.1.2 Model Preparation

This portion of the paper was used to shortlist the significant variables from the scope of study that impacts the pricing. By doing so *fulfilling the first portion of the first objective* which reads, "develop a framework for distinguishing between significant and non-significant variables in the collected municipal tender archive data, which are defined as those that did not have a noteworthy impact on the price of a tender."

5.1.2.1 Case of Multicollinearity

On analysis of the correlations as shown in Table 49; FMCE is highly correlated to the cost of HGMCE and SGMCE. However, this is to be expected as the material used for the FMCE is used for the gating required in most contracts (Tim, 2016). This had resulted in a noticeable correlation with the two gating variables.

Table 49. Snippet Summary of Multicollinearity Pearson's Correlation Test Results

Independent Variable	**FMCE**	**HGMCE**	**SGMCE**
Fence Material Cost Estimate (FMCE)	n/a	0.304	0.364
Hinged Gate Material Cost Estimate (HGMCE)	0.304	n/a	0.001
Sliding Gate Material Cost Estimate (SGMCE)	0.364	0.001	n/a

Conversely, there was no correlation between HGMCE and SGMCE though they were composed of the same material in some contracts as shown in Appendix C, Section 1. It is understood that despite both these variables usually being procured as a collective, there is a disparity in dimensional requirements. This resulted in an unclear relationship between the two variables.

5.1.2.2 Removal of Non-Significant Variables

The central focus for this subsection was the repetitive model builds that included different independent variables. On analysis of these results, every independent variable added significance to the model. Although, on observation of the significance for the SGMCE and HGMCE, the model viewed HGMCE as more critical.

This goes against the research theory as the developed hierarchal table had viewed HGMCE as a tertiary variable below the SGMCE that was ranked as secondary; based on the knowledge that SGMCE requires more accessories (Howmuch, 2020). However, the result is explained in the variable counts. The HGMCE had 51 appearances from 80 samples in the modelling data, whereas the SGMCE only had 31.

Meaning more HGMCE variance was accounted for in the model than the SGMCE. When more variance is accounted for in an independent variable, the model may view it as more significant as stated by Lewis (2007). Leading to the model viewing HGMCE as more significant due to it having a more distinctive effect on the pricing. Table 50 was developed as a remedial rule to mitigate such an occurrence in future works partly based on the equation for minimum sample size (Alshibly, 2018).

Table 50. Reviewed Minimum Sample Size Count Rule per Variable

Sample Observations	**Minimum Occurrence of Each Variable**
For a sample size smaller than 100	Each independent variable must appear in 50 samples.
For a sample size greater than 100	Each independent variable must appear in 50% of the data samples.

5.1.2.3 Model Preparation Analysis Summary

The section was focused on the variables with a noticeable correlation. This included sliding (SGMCE) and hinged gate (HGMCE) with fencing erection (FMCE). The analysis showed the importance of sample size and more so the number of dependent variables individual sample sizes. A known fact that sliding gates would cost more than hinged was disregarded by the model as the sliding gate sample size was almost half that of the hinged gate.

Therefore, the model viewed the hinged gate as a more significant variable due to it having a clearer relationship with the awarded pricing than the sliding gate. It is assumed this issue is avoidable by ensuring all independent variables appear at least 50 times in the *model-building data* sample or at least 50% of a sample size greater than 100.

5.1.3 Construction and Validation of Model

This segment of the paper was used to develop the benchmarking model by deriving the multiple linear regression equation. By doing so *fulfilling the second portion of the first objective* which reads, “thereafter, using the identified significant variable to develop the comparative model on.”

5.1.3.1 Analysing the Coefficient of Multiple Determination

The model *coefficient of multiple determination* (adjusted $\mathbf{R}^2$) value is 0.496 (49.6%). As previously stated, it assesses the fit of the regression model to the sample data given and indicates the variance in the dependent variable that the independent variables explain. This was viewed as an acceptable value for the study. However, the 50.4% unexplained variance is yet to be accounted for.

As explored in the chapter two subsection titled “Benchmarking Municipalities Variables. External factors and procurement processes that may have had an impact on the pricing of the contracts were not included in the modelling. This undoubtedly led to an increase in the unexplained variance. However, the impact of each of these factors is assessed in the remainder of this chapter.

5.1.3.2 Analysing the Work of Klein and Rossin (1999)

As the model build passed all quality assurances set for this study, no analysis was required of the procedures used. However, the 10-fold cross-validation technique used to validate the model required analysis. The method involves splitting the data into subgroups, training the model on 90% of the subsets and using the rest to evaluate the model’s performance; thereafter repeating the procedure ten times (Krishni, 2018; Brownlee, 2019). The results were a rooted average mean squared error (RMSE), recorded as the *testing data error* for the research.

The outcome from the 10-folds cross-validation was used to test the work by Klein and Rossin (1999) that found the scale of errors ranging from 5% to 10% for the data used to train a model had a positive effect on the predictive accuracy for pricing based assessments. Whereas, the scale of errors ranging from 5% to 10% for the data used to test a model harmed the predictive accuracy.

This was validated by the acquired predictive accuracy for the testing data errors (RMSE) being greater than training data errors (SEE). This being the case under the assumption that the data had homogeneity in the distribution of the error samples collected. It should be noted that the testing data used to find the RMSE was always lower than the training data. Ergo, likely magnifying the existing errors, leading to an inconclusive confirmation of the work by Klein and Rossin (1999).

5.2 Analysis of Benchmarking Outcomes (Study Findings)

The closing segment of this subsection (Study Benchmarking Results Analysis) was used as an attempt to justify the pricing deviations in the identified contracts from the non-clean audited municipalities in the Gauteng Province. By doing so *fulfilling the first and second portion of the second objective* which reads, "Using the generated comparative model to identify which contracts in the Province of Gauteng may have incurred pricing deviations and by what margin. Subsequently, find the reasons behind the deviations."

5.2.1 Model Testing Results Analysis

As observed in the previous chapter, the model passed the authentication through all the past procurements used for testing being within the prediction interval. Though, four samples deviated significantly from their predicted values. Table 51 are the reasons found for the near-deviation post procurement document assessment.

Table 51. Analysis of the Near Deviated Procurements from the Model Testing Data

Tender	**Reasons Found**
77	The procurement was an "*evaluate and repair and/or replace*". However, the model viewed the procurement as a replacement contract that would inflate the predicted value, ergo flagging the contract as a near-outlier on the lower bound.
107	No reasons were found for the near outlying.
108	No reasons were found for the near outlying.
109	The contract had additional items including *20 concrete bollards* and *two boom gates*, which would inflate the pricing. Thus, flagged as a near-outlier on the upper bound.

5.2.2 Model Out-of-Range Results Analysis

The model had failed to predict 90.48% of the *out-of-price-range* data, this meant it could not be used for procurements above the R230 000 cap. Besides this, 80.95% of the assessed procurements were flagged as over-priced by the model. This may be the case because of the ignored factors, such as the competitiveness of open tenders resulting in inconsistencies in the pricing as found by Stiti and Yape (2019).

Table 52. Patterns Identified for Procurements above the R230 000 Cap

No.	Description	Remark
1	Contracts are usually stretched over two to three years and are in various locations.	Costs would escalate over the years which is not factored into the model. Hence flagged as over-priced.
2	Required staffing care items such as ablution facilities and round-the-clock security.	An additional cost that is not usually required in low-valued procurements that may inflate the pricing.
3	High-level CIDB compliant contractor; an experienced-based rating scale used in the South African construction industry.	This required specialising which would act as gatekeepers therefore, the procurement would not be as competitive as expected.
4	Required construction equipment is always available onsite.	Additional cost not required in low-valued procurements which would inflate the pricing.

To better appreciate the outcome from the *out-of-range* test, the author analysed the procurement documents for all the above tenders. A pattern emerged for most of the contracts mentioned above, as explained in Table 52. The table shows why the model would have issues predicting the pricing of open and competitive procurements. This is because it has built-in flaws that are based on the preferences of low-valued contracts.

These include low ranking construction industry development board (CIDB) levels, less than a year’s duration, and only needing the minimum equipment onsite for work execution and staffing care. These requirements for low-valued (non-competitive) contracts reduce the cost of the projects immensely and for that reason flawed for competitive procurements.

5.2.3 Supplementary Testing Results Analysis

The model showed that at a 90% confidence level, 92.5% (74 procurements) of the *model-building data* would fall within the prediction interval. This being an excellent result as it shows the model built meets the statistical expectations. However, the primary purpose of this assessment was to identify the six procurements (7.5%) from clean audited municipalities that fell out of range and were flagged as a pricing deviation.

Table 53 summarizes deviated pricing from the clean audited municipalities and the plausible reasons behind it. Followed by an elaboration of each procurement in the below subsections. Of the six flagged contracts, the first five procurements are from the Western Cape Province and the last is from Gauteng.

Table 53. Summary of Deviated Pricing from the Clean Audited Model Building Data

Tender	Result	Reasons Found
8	Under-priced	No reason was found for undercharging.
36	Over-priced	No reason was found for overcharging.
51	Over-priced	Geological may have inflated price, rejected the overpricing.
67	Over-priced	No valid reason was found for overcharging.
86	Over-priced	No reason was found for overcharging.
101	Over-priced	Local labour union interference but not enough to justify the overcharging.

5.2.3.1 Analysis of Tender 8

No reasons were found behind the outlying. However, based on the literature reviewed on public procurement corruption in chapter two, Table 8. This may be a case of *scope creep* in which the supplier purposely underbids to secure the contract. Thereafter, escalate the value by requesting additional funds as the project begins.

5.2.3.2 Analysis of Tender 36 and Tender 86

On the assessment of these contracts, no specifications deviated from the typical format of a fencing procurement. Both contracts requested the minimum requirements for fencing procurement. It is for this reason the study concludes this may have been an over-price because of corrupted price inflation (an above-market value procurement).

5.2.3.3 Analysis of Tender 51

This contract had a unique prerequisite for the supplier as the fence had been erected on a steep hillside and required additional reinforcement and “*stepping*” of the fence. The document had also flagged a hard rock environment that may require non-standard equipment. The above mentioned would have led to a rise in costs.

These identified additional expenses being an *environmental factor* that is classified as an uncontrollable variable. As explored in chapter two of the study, was ignored based on a lack of resources to analyse and quantify these individually. For this reason, it can be concluded that the over-price on this contract was justified.

5.2.3.4 Analysis of Tender 67

Tender 67 required additional staffing care items such as ablution facilities and round-the-clock security. This may have justified the increased cost if not for this being the particular standard for all fencing procurements issued by the municipality. Because this contract is the only one that deviated from the municipality, it can only be concluded this may have been an over-price because of corrupted price inflation.

5.2.3.5 Analysis of Tender 101

The above contract had a unique requirement that stipulated the supplier had to hire a community liaison officer at the total cost of R11 660.00. This may be a local labour union requirement which is another uncontrolled element that was excluded by the study and may have justified the inflated costs if not for the contract being an over-price of R20 020.81. Therefore, it was concluded that the procurement may have been an over-price because of corrupted price inflation.

As the above tender is the only one not in the Western Cape, additional analysis was given to it. The contract is from the only clean audited municipality in Gauteng that has had repetitive clean audits. However, inflated pricing can be justified in various ways. It may be that this was an untested procurement. This is based on the municipal budget summary as it favoured competitive tenders when requesting fencing procurements.

Another reason may be that fencing contracts in the province are slightly costlier because of less demand for the service. This is because the Gauteng Province is the smallest by landmass in the country. However, this presumption was tested and failed when analysing the benchmarking results in the next subsection as the focus is on municipalities in the Gauteng region.

5.2.4 Study Benchmarking Results Analysis

The section was used to entirely answer the critical research question (CRQ). The results showed 33.3% (4) of the municipal fencing contracts from Gauteng deviated highly from their predicted values. However, on the assessment of the benchmarking results, three of the four deviated procurements fell outside the prediction interval. Table 54 summarizes the pricing deviations from the benchmarking assessment and the probable reasons behind them. Thereafter, an explanation of each procurements deviation is in the below subsections.

Table 54. Summary of Deviated Pricing from the Non-Clean Municipal Data

Tender	Result	Reasons Found
D	Under-priced	No valid reason was found for the undercharging.
F	Over-priced	Contained ambiguities in the results that led to rejecting the flagged pricing deviation.
H	Under-priced	No valid reason was found for the undercharging.

5.2.4.1 Analysis of Tender D

The procurement was flagged as under-priced by the model. As explored earlier, these may be early symptoms of *scope creep*. However, before that can be concluded the tender document required evaluation. On assessment of the specifications, it revealed no additional data that would justify the under-price. Though, on analysis of the modelling data assumptions, the material type used for this procurement "concrete pillars" was assumed to have a similar rate to a "*palisades and pillars*" fencing erection.

On further assessment, this assumption was not adequately incorporated into the *model-building data* as only one clean audited municipal contract had been concrete pillars. This has once more introduced a level of ignorance by the model. However, on an evaluation of the non-clean audited *benchmarking data* it was found that three other contracts had "*concrete pillars*" yet not flagged as pricing deviations. Therefore, it would be fair to flag the above contract as an under-price that may be linked to scope creep.

5.2.4.2 Analysis of Tender F

The contract was flagged as an over-price by the model. However, as described in Table 32 of chapter four. Tender E and F were procured separately but presented as a merged final amount including vat (R394 000). As a result, it was assumed both procurements are worth R197 000 as the maximum allowable contract value for each was R200 000 (excluding vat). On the assessment of the contracts, Tender F was over-priced by R 11 939.47 whereas Tender E was R69 842.23 below the upper bound.

This would suggest the assumption of each procurement being half the total value might be incorrect. If it was further assumed the Tender E was more expensive than Tender F when including vat. That would decrease the Tender F value to below the upper bound, ergo not flagged for pricing deviations. However, unless the municipality had disclosed the individual amounts, there can be no certainty in the above results. This leads to ambiguity in the results. As the study has an optimistic evaluation favouring the suppliers that deviated from the pricing, it can only be concluded that this procurement was conducted ethically.

5.2.4.3 Analysis of Tender H

Tender H was flagged as having the highest under-price (*34.63% below the lower bound)* in the study. The procurement was for "palisade panels", a recurring contract that was well factored in for the designed model. On assessment of the procurement document, no additional specification had been made that would substantiate a price decrease (e.g. repairing some parts on the fence).

Because of the above evaluation, it can be said Tender H had been under-priced. This may have resulted from *scope creep*. However, as shown in the next subsection, various reasons may explain the model flagging Tender H and the previously mentioned procurements as either over-priced or under-priced.

5.2.5 Analysis of Benchmarking Outcomes Summary

The section was used to fulfil the second research objective. Thereafter, answer the critical research question. The results had shown how the excluded variables such as environmental factors and local labour union influence have a significant impact on procurement pricing. The model revealed that the assumption of repair contracts being equivalent to supply and installation was incorrect.

The outcomes further showed clean audited municipalities had 6.25% of their procurements flagged as unjustified pricing deviations that included one from the Gauteng Province and four from the Western Cape and of these five, four were flagged as over-priced. In contrast, the non-clean audited municipalities from Gauteng had two unjustified pricing deviations which were 16.7% of their sample. However, all over-priced procurements stemmed from clean audited municipalities.

This resulted in the *observational rejection of the study hypothesis* as clean audited municipalities seemed to have procured the topic contracts at rates above the Gauteng non-clean audited. This is further substantiated by Tender J and L, as shown in Table 32, which had additional items above that of fencing yet still not flagged as over-priced by the model. The implications for this are explored and summarised in chapter seven to conclude the study.

5.3 Analysis of Study Assumptions/Limitations

5.3.1 Generic Data Sorting Assumptions Analysis

The data obtained from some municipalities had not been completed, hence additional logical data sorting assumptions were taken to have an adequate sample size. Table 55 summarizes the study assumptions and their impact; followed by the subsections elaborating how some of these assumptions may have affected the research, and where relevant relating to the previous sections.

Table 55. Summary of Critical Study Assumptions and Their Impact of the Research Outcomes

No.	Assumption	Impact on the Study
1	All contract values include vat.	May present a well-priced contract as under-priced.
2	Combining separate supply and installation contracts.	No impact as no procurement used in the study had this occurrence.
3	Assumption of BEE score being 20.	Would decrease non-BEE compliant suppliers pricing leading to a false under-pricing.
4	Assumptions of fencing repairs are viewed as a supply and installation.	May flag some repair contracts as under-priced.

5.3.1.1 Assumption of Procurement Values Including VAT

The assumption may present a well-rated procurement as under-priced. This being the case as the final procurement value would be reduced by ~13% at a 15% vat. This may have been the case for Tender 8 from the clean audited municipalities as the procurement was flagged as an 11.57% under-price; as opposed to Tender H, which was flagged as an under-price of 34.63%.

However, it should be noted should the inclusion and exclusion of vat had played a significant role in the model building, more variables from the *supplementary* and *model testing* assessments would have been flagged as under-priced. Therefore, it is fair to conclude this assumption played little to no role in the modelling outcomes.

5.3.1.2 Assumption of Combining Separated Supply and Installation Contracts

This assumption expected the summation of two contracts that included the supply of material and installation thereof to be an equivalent of one contract. However, only one procurement had this occurrence and was removed from the data sample because of a disparity in items supplied (e.g. 100m of fencing material but 65m was installed). Therefore, this assumption unquestionably played no role in the modelling outcomes.

5.3.1.3 Assumption of BEE Score Being 20

This assumption was for the non-clean audited municipalities as these were the ones that suffered from none BEE score disclosures. The assumed BEE score of 20 would effectively decrease non-BEE compliant suppliers adjusted pricing. Essentially, flagging a fairly priced contract as under-priced or an inflated one as market-priced. However, based on the *modelling data* outcomes; 85% of suppliers are 100% BEE compliant.

Therefore, the likelihood of the assumed BEE score of 20 being incorrect is relatively low. Thus, this may have played a minimal role in the modelling outcomes.

5.3.1.4 Assumptions of Fencing Repairs Are Viewed As a Supply and Installation

This assumption may falsely flag contracts as under-priced or as observed with Tender 77, flags them as a near under-price. This is because of a fence repair contract being indisputably cheaper than an installation and resulting in the model viewing the actual procurement as under-priced. However, as observed with Tender 77, the procurement was not flagged as an outlier but a close outlier. This stems from the fact that some *model building data* contracts were also repairs; allowing the model to partially account for repair contracts. Therefore, this assumption played a minor role in the outcomes.

5.3.2 Study Limitations Analysis

These limitations were introduced to reduce the scope of the study. They include the *error term*, CPI assumption and non-competitive tender bids limitation. Table 56 is a summary of the limitations and their impact on the research outcomes followed by an elaboration of each in the below subsections.

Table 56. Summary of Critical Study Limitations and Their Impact of the Research Outcomes

No.	Limitations	Impact on the Study
1	Not assessing additional variables outside the controlled variables.	The ignored uncontrolled variables may lead to the justification of inflated procurement costs as observed with Tender 51 and partially Tender 10.
2	Not assessing other pricing generalisation factors beyond consumer price index (CPI).	Though it improved the model correlation by 3%, it had introduced ignorance to the model as a result decreasing its overall accuracy.
3	Not assessing the competitiveness of a procurement.	No impact on low valued contracts but led to the inability to assess tenders above the R230 000 contract value cap.

5.3.2.1 Limitation on Not Assessing Addition Variables outside *Controlled Variables*

This limitation had been introduced because of resource restrictions as explored in chapter two. It was assumed the prediction interval would account for most of the ignored variables including labour union interference and geological challenges. However, as observed in Tender 51 and partially Tender 101, the ignored uncontrolled factors had led to the justification of inflated procurement costs.

This would suggest there was further ignorance that the model may have had that was undetected in the other procurements which were flagged for pricing deviations.

Though as observed with the two contracts above, the factors that were unaccounted for by the model are usually stipulated in the tender documentation. Therefore, it would be safe to assume the pricing deviations with the data would stand as is.

The model did not assess the *value-for-money* of work executed. This means it assumed all work conducted was up to standard. However, should the variable have been accessible by the author, it could have been incorporated into the awarded pricing which would have led to more flagged pricing deviations. This was yet another limitation for the study which was explored as part of the recommendations.

5.3.2.2 Limitation on Not Assessing Other Pricing Generalisation Factors Beyond CPI

This limitation was introduced because of a lack of resources to further investigate the factors that influence the different pricing of fencing contracts from the provinces. Upon CPI adjustments, the contracts from different provinces and dates correlated highly with each other when assessing the anchor predictor and the pricing which led to a three per cent improved correlation. This would suggest that the assumption of CPI being the primary pricing change between contracts from different dates and locations are justified.

As shown with the 10-folds cross-validation, the data was dependable with the available CPI pricing transformation and 91.67% of the benchmarked contracts were within their respective intervals or had a good reason for deviating; this indicating the model's reliability. Ergo, the pricing deviation results with the current data would stand as is.

5.3.2.3 Limitation on the Competitiveness of a Procurement

This limitation was introduced by the literature explored in chapter two by Stiti and Yape (2019). The literature stated that suppliers bidding for low-valued contracts usually bid more evenly with each other. Therefore, the number of bidders and the competitiveness of the process did not influence the pricing. This statement was substantiated by the research findings, as all the low-valued procurements correlated highly with each other regardless of the number of bidders.

However, the contracts from the clean audited municipalities above the R230 000 cap (incl. vat), as observed in Figure 16, have a randomised pattern. This suggests that other factors not accounted for, such as the competitiveness of the tender may have played a crucial role. As this study was focused on non-competitive procurements, the pricing deviations flagged by the model with the information would stand as is.

5.3.3 Analysis of Study Assumptions/Limitations Summary

This subsection was used to critique the study assumptions/limitations. The analysis has shown how the generic data sorting assumptions have played no significant role in the modelling outcomes. However, the study limitations have played a more important role. This resulted in the flaw within the model that flagged market-valued contracts as either over-priced or under-priced.

On further analysis of the procurement documentation, the justification for the deviation in contract costs was found. Thus depicting how the flaw within the model can be remedied. Therefore it is safe to state that the study assumptions and limitations did not play a critical role that would skew the outcomes negatively. As a result, the conclusions drawn would stand as is.

5.4 Data Analysis Summary

This chapter has conducted a critical analysis of the research findings, including Table 57 which contains the intermediate results that were composed of data management, model preparation, construction and validation thereof. Then evaluating the benchmarking outcomes and critiquing of the study assumptions and limitations.

Table 57. Overall Research Data Analysis Summary

Item	Section	Analysis Summary
Intermediate Data Analysis	***Data Management***	The segment showed how the various sampling parameters had been violated by using the quota sampling method. Though, the remedial steps taken minimised the negative impact on the study's credibility. The suppliers' BEE scores were confirmed to having a direct impact on pricing and the theory on the type of material, length and height of a fence (FMCE) having the strongest influence in contract values aligned with the study findings.
	Model Preparation	The section revealed why the sliding and the hinged gate had a noticeable correlation with the FMCE. The hinged gate was viewed as more significant to the modelling than the sliding gate; although sliding gates cost more than hinged (Howmuch, 2020). This

		resulted from the sliding gate sample size being almost half that of the hinged gate. Therefore, the model viewed the hinged gate as more significant due to it having a clearer relationship with the awarded pricing. It is assumed this issue would be mitigated by each independent variable appearing in at least half the model-building data for samples above 100, and at least 50 occurrences for samples below 100.
	Construction and Validation of Model	The outcome from the 10-folds cross-validation aligned with the work by Klein and Rossin (1999) that found errors from training data had a positive effect on the predictive accuracy whereas, in testing, data harmed the accuracy. However, the testing data used to acquire the RMSE was always lower than the training data likely magnifying the existing errors, leading to an inconclusive confirmation of the work.
Benchmarking Outcomes		The results showed clean audited municipalities had 6.25% of their contracts flagged as groundless pricing deviations that included one from the Gauteng Province and four from the Western Cape. Of these five, four were flagged as over-priced. While the non-clean audited municipalities from Gauteng had two unwarranted pricing deviations which were 16.7% of their sample. All over-priced contracts stemmed from clean audited municipalities.
Assumptions and Limitations		The analysis of assumptions and limitations was used to further explore the factors which may have harmed the modelling outcomes. Though, on assessment, it was found that the generic data sorting assumptions have played no significant role in the modelling outcomes as opposed to the study limitations that resulted in the model flagging market-priced contracts as pricing deviations. Though the above was remedied by assessing the procurement document and finding explanations for the deviation.

The post data analysis, *at a confidence level of 90%,* revealed that clean audited municipalities were more likely to have inflated procurement costs for low-valued contracts as opposed to the non-clean audited. This resulted in an ***observational rejection of the study hypothesis*** and was further corroborated by Tender J and L that had additional items above that of fencing installations yet still not flagged as over-priced by the model. The implications for this have been explored in chapter seven to conclude the study.

Chapter Six: Transferability and Contribution of Findings

The developed benchmarking model is suitable for narrowing procurements that may require auditing because of pricing deviations and should in no way be used as an impartial indicator of corruption taking or have taken place. This solution helps in partly satisfying the third research objective by contributing a warning system to flag *possibly* corrupted contracts and as stated in chapter two, can be incorporated into the procurement manager's office as part of the existing governmental controls.

However, the developed model is based on low-valued, fencing procurements. As a result, an outlined framework was required for developing similar models for various low-valued contracts. Figure 19 results from fulfilling the third objective. The framework has intentionally incorporated the flaws discovered in the research. This is because the faults are known, and a procedure has been developed to mitigate their negative impact on the outcomes as explained in the following subsection.

6.1 Limitations of Framework and Outlining for Development

The framework is based on the study methodology and as a result, taken on its drawbacks. Table 58 is a breakdown of the primary limitations of the benchmarking framework including; *MLR, external variables (factors)* and *abidance with the limiting three of the five pillars of procurement* as the *Equity* (BEE variable) and *Ethics (*model outcome of flagging pricing deviated procurements*)* pillar have already been accounted for in the model. Then the remedial steps that can be taken:

Table 58. Primary Limitations of the Benchmarking Framework

Type	Limitation	Remedial Step
MLR	Linearity, multicollinearity and homoscedasticity data assumptions/limitations.	Relevant procedures are noted down in the fourth (4th) level of the framework.
External Factors	Neglecting external variables including environmental factors and local labour union requirements.	Usually disclosed in procurement documents and it would be the modeller's responsibility to examine all outliers to identify plausible reasons for outlying.
Abidance With the Limiting Three of the Five Pillars of Procurement	*(Open and Effective Competition)* The study could not assess the impact of this pillar on the findings because of limited time and resources.	It is remedied by a key assumption based on the work by Stiti and Yape (2019), that found companies tend to bid evenly with low-priced contracts regardless of competitiveness.
	(Accountability and Reporting) Study data was source through the aid of this pillar and sample sizes were decreased because of the non-receptiveness of some municipalities.	No remedial steps are available. This study has made use of a "*work with what you are given approach*".
	(Value for Money) Has been ignored throughout the study which may view market-valued contracts as inflated due to sub-par work.	Value-for-money evaluations can be conducted as part of the post modelling results analysis. However, further studies have to be conducted to find methods of integrating it into the model.

1 Selection of Benchmarking Interests

It is assumed the modeller has selected a procurement type to benchmark. The next step would be to find two sources you wish to compare. This could either be by audit status (clean vs. non-clean) or budget (over vs. under budget). The modeller must be certain the selected contract type has enough data to pass the minimum sample size requirement.

2 Collection of Data

Thereafter, is the collection of all past procurement data relating to the contract of interest (scope of works, BEE level of winning bidder and awarded procurement value). Then segmenting them into the model build, model testing and target *benchmarking data*; followed by the application of the data cleaning measures as outlined in the Methodology.

3 Selection of Independent Variables

Next assessing which scope of work variables are important based on experience using the *model-building data*. These will be the models' independent variables and must abide by the rules drafted in Table 50. Combining variables may be required to get a direct relationship to the pricing, then ranking them by their most likely significance.

4 Testing for the MLR Assumptions

The model has to conform to MLR assumptions. The first being the independent variables must have an identifiable *Linear* relationship to the pricing. This is achieved by data changes thereafter; the independent variables must be tested for *Multicollinearity* and *Homoscedasticity* using the procedures outlined in Chapter Three.

5 Coefficients and Significance Testing

Using the hierarchal regressive model building method to develop several MLR models, all without removing the primary variable as stipulated in the Methodology. Thereafter, selecting the most accurate model (±50%) that ideally does not have insignificant variable coefficients (**t**-value must be above +2 or below -2; **p**-value lower than 0.05).

6 Validation for Regression Model

This step validates the model built by using the 10-folds cross-validation method to find out whether the datasets utilised were credible and if the developed model is robust. This is achieved by comparing the standard estimation error to the rooted mean squared error as explored in the Methodology.

7 Creating Prediction Intervals

The prediction interval can then be built based on the outlined procedure in Chapter Three. However, a confidence level is required based on the reliability of the *model-building data*. A confidence level of 90% is a good assumption, the value can be changed to see which contracts deviated at different confidence levels. Thereafter, tested using the model *testing data*.

8 Identifying Outliers (Benchmarking)

The next step is to use the model to predict the *benchmarking data* then applying the relevant intervals. Thereafter, plotting the true awarded values on the same graph to identify where pricing deviations may have occurred. However, the same must be done for the *model-building data* to identify procurements that fell outside the confidence level.

9 Reduction of Outliers and Justification

The procurements flagged for pricing deviations from the *benchmarking data, model building data* and *model testing data* must be assessed individually. This is to identify likely reasons for the deviation. This is achieved by assessing out of scope variables such as those explored in the Literature Review by examining the procurement document.

10 Final Verdict

Thereafter, noting down all contracts that have been flagged as unjustified pricing deviations. Then assessing the individual procurements that would include value-for-money evaluations and procurement processes used.

Figure 19. Outlined Framework for Developing a Low-Value Contract Benchmarking Model

6.2 Relevance of Theoretical Literature Reviewed Discussion

The work by Stiti and Yape (2019) aided in developing a key assumption that led to a narrowed scope. However, this assumption was further substantiated by a high correlation in the data points collected when plotting the anchor predictor (FMCE) against the transformed awarded pricing (SGTV). This is a confirmation of the supposition that suppliers bid more evenly in low valued procurements regardless of the competitiveness in the bidding processes. Though, this assumption restricted the study scope as described in Table 58 but can be adapted accordingly in certain scenarios as explained in the following subsection.

As explored in the previous chapter, the work by Klein and Rossin (1999) stating that errors in training data have a positive effect on the predictive accuracy of a model, whereas errors in testing data have a negative impact, aligned with the study findings. This was validated by the acquired predictive accuracy for the testing data errors (RMSE) being greater [0.841% more] than training data errors (SEE), under the assumption that the data had consistency in the spread of erred samples. However, the multiple testing data samples used to acquire the RMSE were lower than the training data. This would have likely magnified the existing errors, leading to an inconclusive confirmation of the work by Klein and Rossin (1999).

6.3 Recommendation for Future Industrial Applications Framework

In the 2020 year, the COVID-19 virus pandemic swept across the globe (Villiers *et al.,* 2020). In response to deal with the pandemic, many countries employed procurement procedures to speed up the sourcing of required goods and services (Steingrüber *et al.,* 2020). This saw breaches in various countries of anti-corruption principles such as a person in power taking advantage of the crisis to increase their private benefits (Steingrüber *et al.*, 2020).

The above scenario is opportune to assess the industrial applications for the model. These being the many cases concerning the inflation costs of medical personal protective equipment (PPE) in South Africa during the pandemic (Villiers, Cerbone and Van Zijl, 2020). An emergency procurement clause used by the government that essentially removes the competitiveness in the procurement process to source goods and/or services promptly was abused, leading to inflated prices (Ramaphosa, 2020).

The government, as stated by His Excellency, President Cyril Ramaphosa (2020), began measures to clamp down on such cases. This is precisely the identified opportunity to test the model. It should be noted that these contracts are mostly above the R230 000 (incl. vat) study limit which would classify them as competitive. Because of the emergency procurement clause, they were sourced non-competitively and would lead to an increase in contracting costs (Moore and White, 2005). However, it would remove the competitive element (pillar) in the procurement process which would align with the explored study constraints.

On the assessment of a standard COVID-19 PPE contract scope of work document. It was typically comprising of standard quantifiable items such as the number of branded facial masks and shields, litres of sanitiser, temperature monitors and the area required for sterilisation in squared metres. These are all numerically quantifiable and it would be possible to find a relationship to the pricing without a need for categorical data transformations similar to those in Table 29.

Uncorrupted procurements would be required to teach the model what the true market value for a typical PPE contract is. This data could be sourced from previous year's contracts or, more accurately, governmental institutions that were well below budget when crossed referenced with the risk factor of COVID-19 outbreaks in the region and comparing the results with neighbouring cities. The following example is a depiction of how such a study could be led with the aid of Table 59.

Table 59. Example Data for COVID-19 Benchmarking Exercise

City	Population	COVID-19 Budget	Absolute Budget Used	Risk Rating
City One	1 million	R 40 million	R 20 million	Moderate Risk
City Two	4 million	R 100 million	R 12 million	High Risk
City Three	750 000	R 35 million	R 5 million	Low Risk
City Four	250 000	R 20 million	R 15 million	Moderate Risk

From the above, City Two has a high-risk rating in a population of 4 million inhabitants but has used an absolute lower budget for PPE than City One and Four with a moderate risk rating in a population of 1 million and 250 00 citizens, respectively. City One and Four exhibit signs of price inflation, whereas City Two, is an excellent candidate for the modelling data as it is likely *uncorrupted* and City One and Four would be the focus of the benchmarking exercise.

Chapter Seven: Conclusion

This chapter concludes the study by drawing significance from the post data analysis research outcomes. Included in the opening section of the chapter are the methodologies used in the study, the conclusions drawn from the findings, and the established methods and decisions to all research objectives. Thereafter, the statistical modelling summary followed by the relevance of the thematic literature was reviewed. Subsequently, close the study with the relevant future works recommendations.

7.1 Conclusions Drawn by Research

7.1.1 Summary of Methodology

As defined in Chapter One, the study aimed to compare past municipal procurements from clean and non-clean audited municipalities in South Africa and Gauteng, respectively. The *quota sampling method* was used to source data from past procurements obtained by accessing municipal records, ergo the study was based on secondary data (McCombes, 2019). This may have led to data inconsistencies introduced by human data capturing errors (Manning Publication, 2020). However, relevant sample data management techniques were explored in chapter three to remedy the issue.

Thereafter, the multiple linear regression (MLR) technique based on Montgomery and Runger (2011), aided by the hierarchal model building procedure as outlined in chapter Three was employed to develop the benchmarking model. To which all the required model building verification and validation tools, as explored in the methodology, were applied. The model built was used to assess all procurements including those from clean audited municipalities to identify all past contracts in the study that may have had pricing deviations.

7.1.2 Conclusion of Findings

The post data analysis findings, <u>*based on a confidence level of 90%,*</u> revealed three *"supply and installation of fencing"* contracts from Gauteng [Tender 101, D and H] and four from the Western Cape [Tender 8, 36, 86 and 67] deviated from their predicted pricing. This is 6.25% of the clean audited data and 16.7% of the non-clean audited data. All the flagged over-priced procurements stemmed from the clean audited municipal data sourced from the Province of Gauteng and Western Cape.

These results led to the *observational rejection of the study hypothesis* and by so doing favouring the *null* hypothesis. Therefore, based on the topic contract; clean audited municipalities in South Africa do not procure identical low-valued goods and services at a lower rate than non-clean audited municipalities in Gauteng. As procurement falls under the municipal finance departmental duties these findings align with the work by Kaklar *et al.* (2013) as cited in Motubatse (2016), which found there is a weak relationship between financial management and audit quality.

Motubatse (2016) attributed this weak relationship to shortages of appropriately qualified personnel in the accounting divisions of these entities. This essentially means a clean audited municipality may be unjustly viewed as non-clean audited simply because of accounting irregularities brought upon by the inabilities of accounting officers and not because of operational indiscretions.

The above is further substantiated in the 2018 municipal audit report by the late Auditor General, Mr Kiwi Makwetu (AGSA, 2019, p. 10), stating; “key officials lacking appropriate skills and competencies in financial reporting led to over-reliance on consultants and negatively impacted financial planning, record keeping and reporting”. These findings would explain the failed hypothesis as some non-clean audited Gauteng municipalities may truly be clean audited if not for accounting failures.

However, it should be noted the study did intentionally exclude external variables that may have further explained the rejection of the hypothesis. Though it was found that procurement documents usually disclose external factors that may affect the cost. This was used to eliminate some contracts that deviated from the pricing. As a result, the ignored external variables played a negligible role in changing the rejected hypothesis, though more research should be conducted on their direct effect on the pricing.

The *value-for-money* pillar of procurement which was overlooked throughout the study may have had an unknown impact on the pricing of a contract. The modelled assumed all the delivered work was up to par. Though, this is a flawed assumption as project reworks because of substandard services is a common sight in South Africa, as stated by Bowen *et al.* (2012). Therefore, to truly answer the critical research question holistically the effects of the pillar would have to be investigated as is described in the study recommendations.

7.1.3 Summary of Responses to Research Objectives

Table 60 is a depiction of the answers to the research objectives established in chapter one. Included are the detailed sub-objectives, their respective locations of the detailed methods and a summary of the conclusions drawn.

Table 60. Realisation of Study Objectives Summary

<table>
<tr><th>Objective</th><th>Sub-Objective</th><th>Methodology</th><th>Conclusion</th></tr>
<tr><td rowspan="2">One</td><td>Develop a framework for distinguishing between significant and non-significant variables in the collected municipal tender archive data, which are defined as those that did not have a noteworthy impact on the price of a tender.</td><td>Achieved by using the outlined procedure in the Preparation for Model Design section. Including testing for compliance with multiple linear regression modelling assumptions/limitations.</td><td>Successfully developed and implemented, resulting in all 4 variables being observed as significant by the model and fence material cost estimate (FMCE) meeting the theoretical expectations of having the highest significance as the anchor predictor</td></tr>
<tr><td>Thereafter, using the common factors in the various datasets to develop the comparative model.</td><td>Achieved using the Preparation for Model Design and Model Design subsection.</td><td>Realised by creating an MLR model using all four variables resulting in a predictive accuracy of 49.6%.</td></tr>
<tr><td rowspan="2">Two</td><td>Using the generated comparative model to identify which contracts in the Province of Gauteng may have incurred pricing deviations and by what margin.</td><td>Achieved using the results obtained using the Model Design and outlined method explored in the Benchmarking Strategy subsections.</td><td>Successfully identified three contracts from the Gauteng Province that had pricing deviations using the prediction interval and expressed numerically each contract deviation by a percentage and Rands.</td></tr>
<tr><td>Subsequently, finding the reasons behind the deviations.</td><td>Achieved as part of the Study Benchmarking Results Analysis.</td><td>Recognised by removing some unjustly flagged pricing deviated tenders.</td></tr>
<tr><td>Three</td><td colspan="2">Based on the methods used in the study, generate an outlined framework as a pricing comparison guideline for future procurements beyond the assessed research tender category and identify drawbacks incurred for this study for future researchers to avoid.</td><td>The framework was successfully constructed using the methodology followed in this study; therefore, taken on the flaws experienced in the research. This is because the faults are known, and a procedure has been developed to mitigate their negative impact on the outcomes. It is summarised in Figure 19.</td></tr>
</table>

7.2 Statistical Modelling Summary

The secondary data used to build the model satisfied the statistical rule of thumb for the minimum sample size for four independent variables as explored by Alshibly (2018) [80 contracts]. This resulted in the confirmation of the theoretical assumption that the length, the height and the type of material (FMCE) used to erect a fence have the highest significance in the pricing of *supply and installation of fencing* contracts.

The data sample used to construct the model succeeded in all MLR assumption tests as outlined by Ismail *et al.* (2009). All independent variables had an identifiable linear relationship with the dependent variable, satisfying the *linearity* assumption, and none had a significant linear relationship to each other, sustaining the *multicollinearity* assumption. The data was shown to satisfy the *homoscedasticity* assumption by the plot of standardised residuals having a square-like trend.

The 10-folds cross-validation technique was used to authenticate the reliability of the *model-building data* using the procedure outlined in the methodology. Subsequently, the dependability of the prediction interval was confirmed by all the *model testing data* being within their respective ranges. By meeting all the MLR limitations and assumptions, validation and verification requirements the model shows itself as statistically significant.

7.3 Relevance to Thematic Literature Reviewed Discussion

The reviewed works served as unofficial validations for the model. This being the case as the existing research states the black economic empowerment (BEE) level of a supplier is directly correlated to their pricing (Badenhorst-Weiss and Ambe, 2012). When factoring in the BEE level of a supplier, the correlation between the anchor predictor and pricing improved by a percentage ergo explaining more variance in the model. This matches the existing literature and confirms to an extent, the validity of the model.

The theory on the Gauteng Province suggesting it may have had inflated low-valued contracts did not align with the findings. Though 45% of reported corruption cases stemming from the province (Daniel, 2019), and its municipalities have had recurring non-clean audits (AGSA, 2020a). The study results have shown this cannot be taken as a sign of the likelihood for the inflation of non-competitive contract prices.

The two municipal benchmarking indices researched in the opening of chapter two (GGA, 2019; Williams and Rampai, 2020); were shown to have a flawed holistic approach to comparing municipal finances. This may have led to the neglecting of some pricing deviations within procurements. The developed benchmarking method has shown how this drawback can be remedied. It must be noted this does not discredit the indices but is an add-on to the existing tools.

7.4 Recommendations for Further Research

7.4.1 Recommendation for Critical Study Limitation

Though the model was proven to be conclusive by succeeding in all validation and verification tests, the outcomes have shown there are various elements unaccounted for which have had a significant impact on the study findings. These, as explored in chapter five include the *value-for-money* element and external factors such as geological layout and local labour union interference.

However, it was found the external factors that are likely to have a significant impact on the pricing were disclosed in the procurement documents work specifications. The *value-for-money* appraisal results conducted by municipalities are not easily availed. Therefore, these results must be accessed for future studies to better appreciate their impact on low-valued procurement costs.

7.4.2 Recommendation for Competitive Tendering Analysis Model Build

The model built for this study was successful in flagging pricing deviations as were described in the previous chapters. However, it has failed in assessing open and competitive contracts because of the simplified model building approach used. New variables would have to be included in the model for it to improve the 9.52% accuracy prediction rate for competitive contracts. Table 61 are some identified newly accounted for variables.

Table 61. New Variables to Consider When Assessing Open and Competitive Procurements

<table>
<tr><th>Variable</th><th>Impact on Modelling</th><th>Integration</th></tr>
<tr><td>Competitiveness</td><td>Decreased Pricing.</td><td rowspan="3">These would have to be integrated as part of the pricing element (dependent variable) similar to the BEE and CPI adjustments.</td></tr>
<tr><td>Additional Regulatory Requirements e.g. CIDB ranking</td><td>Act as unofficial gatekeepers which are supposed to uphold the quality of work but may increase pricing as it makes the process less competitive.</td></tr>
<tr><td>Project Duration</td><td>Roles over to next years and requires cost price adjustments which would lead to increased pricing.</td></tr>
</table>

The above would lead to a reviewing of one of the study limitations introduced by Stiti and Yape (2019). This means for open contracts; the competitiveness of the procurement process will undoubtedly play a part. However, these additional factors can be accounted for in the model. By so doing, creating a benchmarking model that transcends pricing restrictions which would lead to the realisation of the study purpose by adding a proactive monitory tool against the scourge of public procurement corruption currently plaguing the Republic of South Africa.

References

Aadnesgaard, V. and Willows, G. (2016) ‘Audit Outcomes And The Level Of Service Delivery Within Local Government Municipalities In South Africa’, *Corporate Ownership & Control*, 13(2), pp. 546–555.

Acemogluy, D., Gelbz, S. and Robinsonx, J. A. (2007) *Black Economic Empowerment and economic performance in South Africa*, *Massachusetts Institute of Technology*. Massachusetts Institute of Technology.

Achilles (2020) *How external factors can affect procurement*, *achilles.com*. Available at: www.achilles.com/industry-insights/how-external-factors-can-affect-procurement/ (Accessed: 30 March 2020).

Aggarwal, S. (2020) *Multiple Linear Regression*, *towardsdatascience.com*. Available at: https://towardsdatascience.com/multiple-linear-regression-8cf3bee21d8b?gi=b13ce526e910 (Accessed: 15 May 2020).

AGPS (1997) *A Technique For Measuring The Efficiency Of Government Service Delivery*, *Steering Committee for the Review of Commonwealth/State Service Provision*. Canberra: Commonwealth of Australia.

AGSA (2011) *Background to the Three Aspects We Audit*. Pretoria: MFMA.

AGSA (2016) *Consolidated General Report on the Outcomes of Local Government (2015)*, *MFMA*. Johannesburg.

AGSA (2017) *Consolidated General Report on the Outcomes of Local Government (2016)*, *MFMA*. Johannesburg.

AGSA (2018) *Consolidated General Report on the Outcomes of Local Government (2017)*, *MFMA*. Johannesburg.

AGSA (2019) *Auditor-general flags lack of accountability as the major cause of poor local government audit results*, *MFMA*. Johannesburg.

AGSA (2020a) *Consolidated General Report on The Local Government Audit Outcomes (2019)*, *MFMA*. Johannesburg.

AGSA (2020b) ‘The Role of The Auditor-General’, in AGSA (ed.) *What Does The Auditor-General Do?* Pretoria: Auditor General South Africa, pp. 1–7.

AGSA (2020c) *The Three Aspects We Audit.*, *Auditor-General of South Africa*. Available at: https://www.agsa.co.za/AuditInformation/AuditTerminology.aspx (Accessed: 9 April 2020).

Akwasi, T. (2020) *What is BEE? BBBEE levels explained 2020*, *Briefly*. Available at: www.google.com/amp/s/briefly/co/za/amp/28111/what_bee_bbbee/ (Accessed: 23 April 2020).

Alshibly, H. H. (2018) *Regression Sample Size*, *Al-Balqá*. Available at: https://www.researchgate.net/post/How_to_calculate_the_sample_size_in_regression_studies (Accessed: 31 August 2020).

Ambe, I. (2012) *The Perspectives Of Supply Chain Management In The Public Sector*. The University of

South Africa.

Ambe, I. M. (2016) ‘Public procurement trends and developments in South Africa’, *Research Journal of Business and Management*, 3(4). doi: 10.17261/Pressacademia.2016.351.

Aprameya, A. (2016) *Quota Sampling: When To Use It and How To Do It Correctly*, *humansofdata.atlan.com*. Available at: https://humansofdata.atlan.com/2016/04/quota-sampling-when-to-use-how-to-do-correctly/ (Accessed: 25 August 2020).

Badenhorst-Weiss, J. A. and Ambe, I. M. (2012) *Procurement Challenges In The South African Public Sector*, *University of South Africa (Unisa)*. The University of South Africa.

Barrons (2020) *T-value Dictionary*, *Allbusiness.com*. Available at: www.allbusiness.com/barrons/dictionary/dictionary-t-value-4942040-1 (Accessed: 22 August 2020).

Biggeri, L., Laureti, T. and Polidoro, F. (2017) ‘Computing Sub-national PPPs with CPI Data: An Empirical Analysis on Italian Data Using Country Product Dummy Models’, *Social Indicators Research*, 131(1), pp. 93–121. doi: 10.1007/s11205-015-1217-x.

Bizana, N. (2013) *Improving Service Delivery By Local Government: A Case For Improved Supply Chain Performance Management*. The University of Kwazulu-Natal.

Bizana, N., Naude, M. J. and Ambe, I. M. (2015) ‘Supply chain management as a contributing factor to local government service delivery in South Africa’, *Journal of Contemporary Management*, 12, pp. 664–683.

Bloomberg, L. and Volpe, M. (2018) *Completing Your Qualitative Dissertation A Road Map From Beginning to End*, *SAGE Publications*. The University of Columbia.

Bolton, P. (2006) ‘Government procurement as a policy tool in South Africa.’, *Journal of Public Procurement*, 6, pp. 193–217.

Bowen, P. A., Edwards, P. J. and Cattell, K. (2012) ‘Corruption in the South African construction industry: a thematic analysis of verbatim comments from survey participants’, *Construction Management and Economics*, 30(10), pp. 885–901. doi: 10.1080/01446193.2012.711909.

Bronshtein, A. (2017) *Train/Test Split and Cross-Validation in Python*, *towardsdatascience.com*. Available at: www.towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6 (Accessed: 20 May 2020).

Brownlee, J. (2019) *A Gentle Introduction to k-fold Cross-Validation*, *machinelearningmastery.com*. Available at: www.machinelearningmastery.com/k-fold-cross-validation_ (Accessed: 20 May 2020).

BusinessNews (2018) *Good fences*, *Economix*. Available at: www.grocotts.co.za-2018-04-10-good-fences (Accessed: 7 June 2020).

BusinessTech (2016) *Shocking levels of fraud and inflated prices cost South Africa R233 billion*, *businesstech.co.za*. Available at: www.businesstech.co.za/newsgovernment139193shocking-levels-of-fraud-and-inflated-prices-cost-south-africa-R233-billion (Accessed: 14 April 2020).

BusinessTech (2019) *New report shows where most corruption takes place in South Africa*, *businesstech.co.za*. Johannesburg. Available at: https://businesstech.co.za/news/government/310452/new-report-shows-where-most-corruption-takes-place-in-south-africa/ (Accessed: 12 July 2021).

Child, K. (2018) *How Gauteng Health Department was plundered*. Available at: www.timeslive.co.zanewssouth-africa2018-06-21-how-gauteng-health-department-was-plundered (Accessed: 14 April 2020).

City of Joburg (2000) 'What is PAIA ?', *Promotion of Access to Information Act 2000 (Act 2 of 2000)*, 2000(1), p. 1.

Damas, Y. I. (2013) *Adoption of Procurement Best Practices in the Private Sector: A Case Study of Banking Industry*, *Mzumbe Dar es Salaam Campus College*. Mzumbe Dar es Salaam Campus College.

Daniel, L. (2019) *Johannesburg bucks the national trend, SAPS bribery increases Bribery*, *The South African*. Johannesburg. Available at: https://www.thesouthafrican.com/news/johannesburg-most-corrupt-city-in-south-africa/ (Accessed: 12 July 2021).

Davidson, S. and Atkins, J. (2013) 'Cost analysis and benchmarking', *QS & Construction Standards GN*, 1(1), pp. 9–11. doi: 978 1 84219 921 3.

Dittgen, M. (2020) *Benchmarking simple models with feature extraction against modern black-box methods*, *towardsdatascience.com*. Available at: https://towardsdatascience.com/benchmarking-simple-machine-learning-models-with-feature-extraction-against-modern-black-box-80af734b31cc (Accessed: 14 January 2020).

Dobie, K. and Xinwa, N. (2015) 'Understanding the municipal procurement process: A guide for businesses', *ethicssa.org*, pp. 2–8. Available at: https://klcbt.co.za/understanding-municipal-procurement-a-guide-for-businesses/.

DTI (2018) 'Codes Of Good Practice On Broad-Based Black Economic Empowerment', 1(42021), pp. 123–125. Available at: https://www.gov.za/sites/default/files/gcis_document/201905/42496gen306.pdf.

Econoserv (2018) *Crash Course in BEE – B-BBEE in 5 minutes*, *econoserv*. Available at: https://www.econoserv.co.za/crash-course-in-bee-b-bbee-in-5-minutes/ (Accessed: 23 January 2020).

Emrouznejad, A. and Cabanda, E. (2015) 'Introduction to Data Envelopment Analysis and its applications', in *Handbook of Research on Strategic Performance Management and Measurement Using Data Envelopment Analysis*. Virginia: IGI Global, pp. 235–255. doi: 10.4018/978-1-4666-4474-8.ch004.

Everitt, B. S. and Skrondal, A. (2010) *Cambridge University Press*. 4th edn, *The Cambridge Dictionary of Statistics*. 4th edn. London.

Frost, J. (2020a) *Confidence Intervals vs Prediction Intervals vs Tolerance Intervals*, *statisticsbyjim.com*. Available at: statisticsbyjim.com_hypothesis-testing_confidence-prediction-tolerance-intervals_ (Accessed: 20 May 2020).

Frost, J. (2020b) *Standard Error of the Regression vs. R- squared*, *statisticsbyjim.com*. Available at: https://statisticsbyjim.com/regression/standard_error_regression/ (Accessed: 17 June 2020).

Gaurav, J. (2017) *6 Sampling Techniques: How to Choose a Representative Subset of the Population*, *humansofdata.atlan.com*. Available at: https://humansofdata.atlan.com/2017/07/6-sampling-techniques-choose-representative-subset/ (Accessed: 3 March 2020).

GGA (2019) *Good Performance Index*. Johannesburg. Available at: https://www.africaportal.org/documents/20129/Government_performance_index_HC7LSvx.pdf.

Gibson, C. (2019) *How much does fencing cost per metre?*, *Cost guide*. Available at: https://hipages.com.au/article/how_much_does_fencing_cost_per_metre (Accessed: 9 May 2020).

Glen, S. (2015) *Homoscedasticity/Homogeneity of Variance/ Assumption of Equal Variance*, *StatisticsHowTo.com: Elementary Statistics for the rest of us!* Available at: https://www.statisticshowto.com/homoscedasticity/#:~:text=The assumption of equal variances,they came from different populations.&text=Running a test without checking,may even invalidate them completely. (Accessed: 12 July 2021).

Goeppner, K. (2008) *Multiple Regression Analysis – A Case Study*, *American Society of Appraisers*. Virginia. Available at: http://www.appraisers.org/docs/default-source/discipline_rp/goeppner-multiple-regression-analysis_a-case-study.pdf?sfvrsn=2.

Grace-Martin, K. (2019) *Eight Ways to Detect Multicollinearity*, *theanalysisfactor.com*. Available at: https://www.theanalysisfactor.com/eight-ways-to- detect-multicollinearity (Accessed: 17 September 2020).

Grace-Martin, K. (2020) *Confusing Statistical Term #4: Hierarchical Regression vs. Hierarchical Model*, *The Analysis Factor*. Available at: www-theanalysisfactor-com-confusing-statistical-term-4-hierarchical-regression-vs-hierarchical-model- (Accessed: 5 June 2020).

Graphpad (2019) *The distinction between confidence intervals , prediction intervals and tolerance intervals.*, *graphpad.com*. Available at: graphpad.com_support_faq_the-distinction-between-confidence-intervals-prediction-intervals-and-tolerance-intervals_ (Accessed: 20 May 2020).

Green, A. (2020) *Difference in Public Vs. Private Procurement*, *Chron*. Available at: https://work.chron.com/difference-public-vs-private-procurement-28985.html (Accessed: 14 January 2020).

Grega, M. and Nemec, J. (2015) *Factors Influencing Final Price Procurement Evidence from Slovakia*, *16th Annual Conference on Finance and Accounting*. The University of Economics. doi: 10.1016/S2212-5671(15)00768-6.

Gumede, W. (2020) *How real BEE can help ordinary folk*, *University of the Witwatersrand*. Available at: https://www.wits.ac.za/news/latest-news/opinion/2020/2020-11/how-real-bee-can-help-ordinary-folk.html (Accessed: 12 July 2021).

Gustafsson, A. and Wogenius, S. (2014) *Modelling Apartment Prices with the Multiple Linear Regression Model*. Royal Institute of Technology. Available at: http://www.diva-portal.org/smash/get/diva2:725045/FULLTEXT01.pdfMultiple.

Hanks, J., Davies, H. and Perera, O. (2008) ‘Sustainable public procurement in South Africa’, *International Institute for Sustainable Development*, (May), pp. 44–45.

Hawkins, T. G., Gravier, M. J. and Powley, E. H. (2011) ‘Public Versus Private Sector Procurement

Ethics and Strategy: What Each Sector can Learn from the Other', *Journal of Business Ethics*, (November), pp. 568–579. doi: 10.1007/s10551-011-0881-2.

Heller, J. (2013) 'Supply Chain Management Perspectives in the Public and Private Sectors', *Advances in Business Research*, 4(1), pp. 107–113.

Henrique, B. M., Sobreiro, V. A. and Kimura, H. (2018) 'ScienceDirect Stock price prediction using support vector regression on daily and up to the minute prices', *The Journal of Finance and Data Science*, 4(3), pp. 183–201. doi: 10.1016/j.jfds.2018.04.003.

Horn, G. S. and Raga, K. (2012) *Municipal supply chain management challenges and solutions: An empirical study*, *Administratio Publica*. The University of Pretoria.

Howmuch (2020) *Driveway Gates Installation Costs*, *Howmuch.net*. Available at: https://howmuch.net/costs/driveway-gate-installation (Accessed: 4 September 2020).

Hughes, M. and Fisher, T. (2020) 'Introduction to Statistical Modeling', in *IntroStatModelling*. Miami: Miami University, pp. 12–30, 50–55.

Ismail, Z., Yahya, A. and Shabri, A. (2009) 'Forecasting Gold Prices Using Multiple Linear Regression Method', *University Technology Malaysia*, pp. 1509–1511.

Jenks, S. (2017) *Public vs. Private Procurement: What's the difference?*, *kodiakrating.com*. Kodiak Rating. Available at: https://medium.com/@KodiakRating/public-vs-private-procurement-whats-the-difference-d68aef7e0006 (Accessed: 11 November 2019).

Kaklar, H. M., Kangarlouei, S. . and Motavassel, M. (2013) 'Audit quality and financial reporting quality: Case of Tehran Stock Exchange', *Innovative Journal of Business and Management*, 1(3), pp. 43–47.

King IV (2016) *Report On Corporate Governance for South Africa*. Johannesburg. Available at: https://www.adams.africa/wp-content/uploads/2016/11/King-IV-Report.pdf.

Klein, B. and Rossin , D. (1999) 'Data Quality in Linear Regression Models: Effect of Errors in Test Data and Errors in Training Data on Predictive Accuracy', *University of Michigan-Dearborn USA*, 2(2), pp. 33–43.

Klimberg, R. K. and Lawrence, K. D. (2009) 'Using regression and Data Envelopment Analysis (DEA) to forecast bank performance over time', *Applications of Management Science*, 18(October), p. 3. doi: 10.1108/S0276-8976(2009)0000013010.

Krappe, K. and Kallayil, G. (2003) 'Contract management is more out of control than you think.', *Journal of Contract Management*, p. 3.

Krishni, K. (2018) *K-Fold Cross Validation*, *medium.com*. Available at: https://medium.datadriveninvestor.com/k-fold-cross-validation-6b8518070833?gi=4bb6e147c8f5 (Accessed: 12 July 2021).

Krüger, L. P. (2011) 'The impact of black economic empowerment (BEE) on South African businesses: Focusing on ten dimensions of business performance', *Southern African Business Review*, 15(3), pp. 207–233.

Lambert-Mogiliansky, A. and Kosenok, G. (2009) 'Fine-tailored for the cartel-favoritism in procurement', *Review of Industrial Organization*, 35(1–2), pp. 95–121. doi: 10.1007/s11151-009-9220-5.

De Lange, S. (2011) *Irregular State Expenditure Jumps 62%.*, *Businesslive*. Available at: www.businessday.co.za (accessed on 20 October 2011).

Larson, P. (2009) 'Public vs. Private sector perspectives on supply chain management', *Journal of Public Procurement*, 9(2), pp. 221–246. doi: 10.1108/JOPP-09-02-2009-B003.

Lau, T. (2018) *Contrasting Procurement Practices between the Private and Public Sector*, *Singapore Institute of Purchasing and Materials Management Academy*. Available at: http://keydifferences.com/difference-between-public-sector- and-private-sector.html, accessed 02/09/2017 (Accessed: 15 October 2019).

Lavrakas, P. J. (2011) 'Alpha, Significance Level of Test', *Encyclopedia of Survey Research Methods*, p. 1. doi: https://dx.doi.org/10.4135/9781412963947.n13 + More information Online Publication Date: January 1, 2011 Disciplines: Anthropology, Business and Management, Communication and Media Studies, Criminology and Criminal Justice, Economics, Education, Geography, Health, Marketing, Nursing, Political Science and International Relations, Psychology, Social Policy and Public Policy, Social Work, Sociology Print ISBN: 9781412918084 | Online ISBN: 9781412963947 Buy in print.

Lewis, M. (2007) 'Stepwise versus Hierarchical Regression: Pros and Cons', in *Southwest Educational Research Association 2007 Annual Meeting*. Denton, pp. 5–15. Available at: https://www.researchgate.net/publication/235464734_Stepwise_versus_hierarchical_regression_Pros_and_cons.

Lowman, S. (2017) *A bottomless pit? Increasing the BEE burden yet again*, *Biznews.com*. Available at: https://www.biznews.com/thought-leaders/2017/03/16/bee-burden-anthea-jeffery (Accessed: 28 August 2019).

Lumber, W. (2020) *Factors to Consider Before Building a Fence*, *WithersLumber.com*. Available at: https://witherslumber.com/blog/69082/factors-to-consider-before-building-a-fence (Accessed: 9 May 2020).

Mailovich, C. and Phakathi, B. (2018) *Treasury and Auditor-General Join Forces to Cut Wasteful Expenditure in Municipalities*, *bussinesslive*. Available at: https://www.businesslive.co.za/bd/national/2018-10-24-treasury-and-auditor-general-join-forces-to-cut-wasteful-expenditure-in-municipalities/ (Accessed: 11 December 2019).

Managa, A. (2012) 'Unfulfilled promises and their consequences : A reflection on local government performance and the critical issue of poor service delivery in South Africa', *Africa Institue of South Africa*, (76), pp. 1–9.

Manning Publication (2020) 'The data science process', in *Introducing Data Science*. Manning Publication, pp. 1–39. Available at: https://livebook.manning.com/book/introducing-data-science/chapter-2/.

Mantzaris, E. (2014) *Public procurement, tendering and corruption Realities, challenges and tangible solutions*, *Stellenbosch University*. Stellenbosch University. Available at: http://hdl.handle.net/2263/58120.

Manyathi, S. (2019) *The Influence Of Procurement Practices On Service Delivery: An Analysis Of Government Legislation – Service Delivery Relationships In South Africa*. Stellenbosch

University. Available at: https://scholar.sun.ac.za/handle/10019.1/107179.

McCombes, S. (2019) *Correlational Research*, *scribbr.com*. Available at: scribbr.com/methodology/correlation-research/ (Accessed: 23 August 2019).

Middleton, S. (2019) *S A's municipal audit 'not a pretty*, *University of Cape Town*. Available at: www.news.uct.ac.zaarticle/2019-08-01/sas-municipal-audit-not-a-pretty-picture (Accessed: 12 April 2020).

Moatshe, R. (2019) *Tshwane mayor Mokgalapa orders probe into fuel tender*, *IOL News*. Available at: www.iol.co.za/pretoria-news/tshwane-mayor-mokgalapa-orders-probe-into-fuel-tender-33840666 (Accessed: 10 June 2020).

Modingoane, N. and IAB (2019) *Emergency! City of Joburg's tender processes investigated*, *Mail & Guardian*. Available at: www.mg.co.za/article2019-07-11/emergency-city-of-joburgs-tender-processes-investigated (Accessed: 13 April 2020).

Montgomery, D. and Runger, G. (2011) 'Applied Statistics and Probability for Engineers', in *John Wiley & Sons, Inc.* Danvers: John Wiley & Sons, pp. 251, 400–502.

Moore, G. W. (2003) 'Estimating Procurement Cost Growth Using Logistic and Multiple Regression', pp. 55–64. Available at: http://www.dtic.mil/dtic/tr/fulltext/u2/a413830.pdf.

Moore, G. and White, E. (2005) 'A Regression Approach for Estimating Procurement Cost', *Journal of Public Procurement*, 5(2), pp. 187–209. doi: 10.1108/JOPP-05-02-2005-B003.

Motubatse, K. (2016) *An Evaluation Of Factors Affecting The Progression To Clean Audit Outcomes In South African*. University of Limpopo. Available at: http://ulspace.ul.ac.za/handle/10386/1701.

Munzhedzi, P. H. (2016) 'South African public sector procurement and corruption: Inseparable twins?', *Journal of Transport and Supply Chain Management*, 10(1), pp. 3–8. Available at: https://jtscm.co.za/index.php/jtscm/article/view/197/407.

National Treasury (2001) 'Preferential Procurement Regulations, 2001 Pertaining To the Preferential Procurement Policy Framework Act: No 5 Of 2000', (5), pp. 5–7. Available at: https://ppp.worldbank.org/public-private-partnership/library/preferential-procurement-policy-framework-act-5-no-5-2000#:~:text=The Preferential Procurement Policy Framework,historically disadvantaged persons%2C including females.

National Treasury (2016) 'Advertisement of Bids and The Publication of Awards on etendering', *National Treasury of South Africa*, p. 4. Available at: http://www.treasury.gov.za/legislation/pfma/TreasuryInstruction/Treasury Instruction No. 1 of 2015 2016 on eTender Publication Portal.pdf.

National Treasury (2017) 'Local Government Revenue And Expenditure: First Quarter Local Government Section 71 Report', *National Treasury of South Africa*, pp. 13–14. Available at: http://www.treasury.gov.za/comm_media/press/2021/2021030301 Media Statement - Local Government Revenue and Expenditure- Second Quarter Local Government Section 71 Report.pdf.

National Treasury (2019) 'Irregular Expenditure Framework', *PFMA Compliant Institutions*, p. 7. Available at: http://www.treasury.gov.za/legislation/pfma/TreasuryInstruction/Annexure A Irregular Expenditure Framework 20192020.pdf.

NCPPC (2019) 'Public Vs Private Private Sector Procurement', *Similarities and Differences between Public and Private Sector Procurement*, p. 1. Available at: http://www.goncppc.org/similarities_and_differences_between_public_and_private_sector_procurement.html.

Newman, D. and Swift, D. (2001) 'Comparing public- and private-sector procurement', *The Chatroom*, pp. 10–11.

Nextenders (2016) *6 types of procurement fraud and how to spot them*, *nextenders.co.uk*. Available at: http://www.nextenders.co.uk/6-types-of-procurement-fraud-and-how-to-spot-them/ (Accessed: 24 August 2020).

Ngoepe, M. and Ngulube, P. (2013) 'Contribution Of Record-Keeping To Audit Opinions: An Informetrics Analysis Of The General Reports On Audit Outcomes Of The Auditor-General Of South Africa', *ESARBICA Journal*, 32(1), p. 52. Available at: https://www.researchgate.net/publication/274953793_Contribution_of_record-keeping_to_audit_opinions_an_informetric_analysis_of_the_general_reports_on_audit_outcomes_of_the_Auditor-General_of_South_Africa.

Ntsaluba, G. (2019) *Gauteng, KZN the most corrupt provinces – Corruption Watch*, *The Citizen*. Johannesburg. Available at: https://citizen.co.za/news/south-africa/crime/2115709/gauteng-kzn-the-most-corrupt-provinces-corruption-watch/ (Accessed: 12 July 2021).

Ogee, A. (2013) *How to Interpret Regression Analysis Results: P-values and Coefficients How Do I Interpret the Regression Coefficients for Linear Relationships ?*, *blog.minitab.com*. Available at: blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients (Accessed: 22 August 2020).

Openm (2020) *10-fold Crossvalidation*, *openml.org*. Available at: www.openml.org/a_estimation_procedures/1 (Accessed: 13 May 2020).

Oppelt, C. (2019) 'Developing a Monitoring and Evaluation System in Supply Chain Management at Overstrand Municipality', *Stellenbosch University*, pp. 22, 152. Available at: https://scholar.sun.ac.za/handle/10019.1/105969.

PennState (2019) *Linear Regression Methods*, *newonlinecourses.science.psu.edu*. Available at: https://newonlinecourses.science.psu.edu/stat501/lesson/1 (Accessed: 24 August 2019).

Petrocelli, J. V (2003) 'Hierarchical Multiple Regression in Counseling Research: Common Problems and Possible Remedies', *Measurement and Evaluation in Counseling and Development*, 36, pp. 9–22. doi: 10.1080/07481756.2003.12069076.

Pillay, C. (2020) *Brush Up on Your BEE Knowledge*, *smesouthafrica.co.za*. Available at: https://smesouthafrica.co.za/Brush-up-on-your-BEE-knowledge/#:~:text=It is essentially a growth,African economy's weakest point%3A inequality.&text=As such%2C this strategy stresses,the redistribution of existing wealth. (Accessed: 7 October 2020).

Pillay, K. and Erasmus, J. (2015) 'Understanding Tender Corruption', *Corruption Watch*, 1(1), pp. 1–16. Available at: https://www.corruptionwatch.org.za/wp-content/uploads/2015/06/Corruption-Watch-Understanding-tender-corruption.pdf.

Pooe, D., Makhubele, D. T. and Mafini, C. (2015) 'Investigating Municipal Procurement Challenges in South Africa: A Qualitative Study', *International Business & Economics Research Journal*, 14(1), pp. 67–78. doi: https://doi.org/10.19030/iber.v14i1.9033.

Quantcademy (2020) *Using Cross-Validation to Optimise a Machine Learning Method - The Regression Setting*, *quantstart.com*. Available at: www.quantstart.com_articles_Using-Cross-Validation-to-Optimise-a-Machine-Learning-Method-The-Regression-Setting_ (Accessed: 20 May 2020).

Ramaphosa, C. (2020) *Ramaphosa says COVID-19 corruption will be put on lockdown*, *ENCA*. South Africa: ENCA. Available at: https://www.enca.com/news/ramaphosa-says-covid-19-corruption-will-be-put-on-lockdown (Accessed: 12 July 2021).

Ray, S. (2015) *7 Regression Techniques you should know*. Available at: https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/ (Accessed: 12 August 2019).

Ray, S. (2018) *Improve Your Model Performance using Cross-Validation (in Python and R)*, *analyticsvidhya.com*. Available at: https://www.analyticsvidhya.com/blog/2018/05/improve-model-performance-cross-validation-in-python-r/ (Accessed: 20 May 2020).

Republic of South Africa (1996) *The Constitution*, *The Republic of South Africa*. Pretoria: South Africa. Available at: https://www.justice.gov.za/legislation/constitution/saconstitution-web-eng.pdf.

Republic of South Africa (2017) *Local Government Municipal Finance Management Act 56 Of 2003*. Pretoria: National Treasury. Available at: https://www.gov.za/documents/local-government-municipal-finance-management-act-0.

ResearchGate (2020) *10-fold Cross Validation Diagram*, *researchgate.net*. Available at: www.researchgate.net_figure_Ten-fold-cross-validation-diagram-The-dataset-was-divided-into-ten-parts-and-nine-of_fig1_326465007 (Accessed: 15 May 2020).

Rob and Hyndman (2014) *The difference between prediction intervals and confidence intervals*, *robjhyndman.com*. Available at: robjhyndman.com/hyndsight_intervals_ (Accessed: 20 May 2020).

SA Tenders (2020) *Which Documents Do I need to Tender*, *SA Tenders*. Available at: www.sa-tenders.co.za/content/hints-tips-and-new/which-documents-do-i-need-tender (Accessed: 30 March 2020).

SABC News (2014) *Gauteng Among the Highest Corrupt Provinces*, *ngopulse.org*. Available at: http://www.ngopulse.org/newsflash/gauteng-among-highest-corrupt-provinces (Accessed: 12 July 2021).

SABS (2004) 'South African National Standard', (1), pp. 5–9. Available at: https://store.sabs.co.za/pdfpreview.php?hash=8cdc7ade1f886f0bfb9cf3258efc2685dc7615a8&preview=yes.

Santos, J., Ros, E. and Cavique, L. (2011) 'Introduction to Data Envelopment Analysis', in *International Series in Operations Research & Management Science*. 1st edn. Evora: ResearchGate, pp. 1–2. doi: 10.1007/978-94-007-5739-4.

Sarip, A. G., Hafez, M. B. and Lumpur, K. (2016) 'Application Of Fuzzy Regression Model For Real Estate Price Prediction', *Malaysian Journal of Computer Science*, pp. 1–13.

Scikit-Learn (2020) *Cross-validation: evaluating estimator performance*, *sci-kit learn*. Available at:

scikit_learn.org/stable_modules/cross_validation (Accessed: 13 May 2020).

Seif, G. (2018) *5 Types of Regression and their properties*, *towardsdatascience.com*. Available at: https://towardsdatascience.com/5-types-of-regression-and-their-properties-c5e1fa12d55e (Accessed: 10 August 2019).

Seiford, L., Zhu, J. and Cooper, W. (2014) ‘Data Envelopment Analysis: History, Models, and Interpretations’, in *Handbook on Data Envelopment Analysis. International Series in Operations Research & Management Science*. Austin: Springer, Boston, MA, pp. 3–7. doi: https://doi.org/10.1007/978-1-4419-6151-8_1.

Shah, M. (2009) *Multiple Regression model building*, *Study Trails*. Available at: https://www.studytrails.com/blog/multiple-regression-model-building/ (Accessed: 10 January 2020).

Shaikh, R. (2018) *Cross Validation Explained: Evaluating estimator*, *towardsdatascience.com*. Available at: www.towardsdatascience.com_cross-validation-explained-evaluating-estimator-performance-e51e5430ff85.

Da Silva, I. (2006) *ACA survey shows increased Black ownership*, *bizcommunity.com*. Available at: http://www.bizcommunity.com/Article/196/12/12606.html (Accessed: 4 December 2007).

Smith, P. (2017) *What Are the Differences Between Public and Private Sector Procurement?*, *publicspendforum.net*. Available at: https://www.publicspendforum.net/blogs/peter-smith/2017/07/27/differences-public-private-sector-procurement/%0A (Accessed: 23 April 2020).

Statistical Solutions (2020a) *Multiple Linear Regression Assumptions*, *StatisticalSolutions.com*. Available at: www.statisticssolutions.com/assumptions-of-multiple-linear-regression (Accessed: 8 May 2020).

Statistical Solutions (2020b) *Pearson’s Correlation Coefficient*, *StatisticalSolutions.com*. Available at: www.statisticssolutions.com-pearsons-correlation-coefficient (Accessed: 17 June 2020).

Statistical Solutions (2020c) *Testing Assumptions of Linear Regression in SPSS*, *StatisticalSolutions.com*. Available at: www.statisticssolutions.com-testing-assumptions-of-linear-regression-in-spss (Accessed: 15 June 2020).

Statistics Solutions (2013) *Homoscedasticity*, *statisticssolutions.com*. Available at: https://www.statisticssolutions.com/homoscedasticity/ (Accessed: 26 October 2020).

StatsSA (2017) ‘The South African CPI Sources and Methods Manual’, *Public Document*, p. 1. Available at: http://www.statssa.gov.za/cpi/documents/The_South_African_CPI_sources_and_methods_26Feb13.pdf.

Steingrüber, B. S. *et al.* (2020) ‘Corruption in the time of COVID-19: A double-threat for low-income countries’, in *Anti-Corruption Resource Centre*.

Stiti, K. and Yape, S. J. (2019) *Bid Forecasting in Public Procurement*. Kth Royal Institute of Technology. Available at: https://www.diva-portal.org/smash/get/diva2:1354986/FULLTEXT01.pdf.

Surendran, A. (2018) *Quantitative Data: Definition, Types, Analysis and Examples*, *questionpro.com*. Available at: https://www.questionpro.com/blog/quantitative-data/amp/ (Accessed: 9 August 2019).

Thanassoulis, E. (1993) 'A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments', *Journal of the Operational Research Society*, 44(11), pp. 1129–1144.

Thornhill, C. (2006) 'Effective municipal government and administration as preconditions for efficient service delivery', *Journal of Public Administration*, 1(1), p. 322. Available at: http://hdl.handle.net/2263/3914.

Tim (2016) *How Many Cost Factors Affect Fence Installation*, *Fence Doctor Home*. Available at: https://fencedoctorct.com/index.php/2016/08/26/how-many-cost-factors-affect-fence-installation/ (Accessed: 9 May 2020).

Timse, T. (2019) *The Ekurhuleni R1.9 billion toilet tender that stinks to high heaven*, *Daily Maverick*. Johannesburg. Available at: https://www.dailymaverick.co.za/article/2019-07-03-the-ekurhuleni-r1-9-billion-toilet-tender-that-stinks-to-high-heaven/ (Accessed: 13 July 2021).

Tom, X. (2014) 'Auditor-General South Africa on its role and mandate: workshop', in ANC (ed.) *Arts and Culture*. Pretoria: Parliamentary Monitoring Group, p. 1. Available at: pmg-org-za-committee-meeting-17587.

Villiers, C. De, Cerbone, D. and Van Zijl, W. (2020) 'The South African government's response to COVID-19', *Journal of Public Budgeting, Accounting & Financial Management*, 1(1), pp. 1–9. doi: 10.1108/JPBAFM-07-2020-0120.

Warf, B. (2019) *Global Corruption from a Geographic Perspective*. 1st edn. Kansas City: University of Kansas. doi: 10.1007/978-3-030-03478-8.

Wilhelm, J. (2016) *Why do we transform non-linear regression models to linear regression models?*, *researchgate.net*. Available at: https://www.researchgate.net/post/why_do_we_transform_non_linear_regression_models_to_linear_regression_models (Accessed: 18 August 2019).

Williams, J. and Rampai, K. (2020) *Municipal Performance Index (MPI)*, *pwc.co.za*. Available at: www.pwc.co.za/en/industries/public-sector/cities-and-urbanisation/municipal-performance-index.html (Accessed: 25 May 2020).

Wojcik, V., Dyckhoff, H. and Clermont, M. (2019) 'Is data envelopment analysis a suitable tool for performance measurement and benchmarking in non-production contexts?', *Business Research*, 12(2), pp. 559–595. doi: 10.1007/s40685-018-0077-z.

Appendix

Appendix A

Section 1

Figure A: Turn-it-in Results for the Dissertation

707312:Nkosi-Tladi-MScDis_(final_submission).pdf

ORIGINALITY REPORT

9%	5%	4%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	M Ambe Intaher, A Badenhorst Weiss Johanna. "Supply chain management challenges in the South African public sector", African Journal of Business Management, 2012 Publication	2%
2	hdl.handle.net Internet Source	1%
3	Submitted to University of Witwatersrand Student Paper	<1%
4	www.ethicsa.org Internet Source	<1%
5	Submitted to Mancosa Student Paper	<1%
6	www.src.uct.ac.za Internet Source	<1%

Section 2

Figure B: Ethic Clearance Certificate

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Research Office

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
R14/49 Nkosi

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER: H20/05/29**
PROJECT TITLE	A Comparison of Tender Prices in the South African Public Procurement Sector to Identify Pricing Deviations
INVESTIGATOR(S)	Mr T Nkosi
SCHOOL/DEPARTMENT	Mechanical, Industrial and Aeronautical Engineering/
DATE CONSIDERED	22 May 2020
DECISION OF THE COMMITTEE	Approved Risk Level: Minimal
EXPIRY DATE	23 June 2023

DATE 24 June 2020 **CHAIRPERSON** ______________________
(Professor J Knight)

cc: Supervisor : Dr E Jonathan and Ms B Tladi

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University. Unreported changes to the application may invalidate the clearance given by the HREC (Non-Medical)

I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the researcl procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to completion of a yearly progress report.**

____
Signature

Date 26 / 10 / 2020

PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

Appendix B

Section 1

Table A: Data for Exploration of Commonly Procured Items (Fencing Tenders)

Fencing Tenders					
Item No.	**Municipality**	**Date**	**Tender Number**	**Price**	**BEE Score**
1	Cape Agulhas	18/03/2019	Q13/2018/19	R 107 870.00	20
2	Cape Agulhas	17/04/2019	SCM35/2018/19	R 116 000.00	20
3	Cape Agulhas	17/11/2017	Q9/2017/18	R 158 980.00	20
4	Cape Agulhas	26/04/2018	Q16/2017/18	R 70 460.44	20
5	Hessequa	16/05/2016	INF 01/1516	R 141 220.00	20
6	Hessequa	30/10/2016	INF 06/1516	R 48 338.00	12
7	Hessequa	1/11/2016	Site 971985	R 165 000.00	20
8	Hessequa	26/02/2016	INF 06/1516	R 48 338.00	12
9	Hessequa	1/12/2016	RFQ 52840	R 124 945.00	20
10	Hessequa	04/03/2016	INF 02/1516	R 57 700.00	16
11	Hessequa	06/04/2016	HES-WRS 07/1516	R 411 814.70	20
12	Hessequa	07/09/2017	RFQ 55897	R 45 000.00	20
13	Hessequa	13/04/2017	RFQ 54460	R 68 699.12	20
14	Hessequa	02/06/2017	Not Available	R 30 000.00	20
15	Hessequa	22/05/2018	RFQ: 59608	R 115 100.00	20
16	Hessequa	24/06/2018	RFQ: 60455	R 39 200.00	20
17	Hessequa	13/09/2019	HES-TECH 30/1920	R 216 840.00	20
18	Hessequa	25/03/2019	RFQ 63401 & 63399	R 203 763.00	20
19	Hessequa	29/05/2019	RFQ: 63923	R 189 400.00	20
20	Hessequa	07/08/2019	RFQ: 65034	R 166 555.00	20
21	Hessequa	07/07/2016	RFQ: 959495	R 56 056.08	20
22	Matzikama	02/05/2017	Q6-2017	R 74 855.00	20
23	Matzikama	12/06/2018	FQ17-2018	R 114 934.30	20
24	Matzikama	11/05/2018	FQ13-2018	R 107 443.74	20
25	Midvaal	20/12/2018	AHT4533	R 199 182	20
26	Midvaal	18/07/2018	Not Available	R 89 355	20
27	Witzenberg	08/05/2017	08/2/13/77	R 1 188 290.40	0

Table B: Data for Exploration of Commonly Procured Items (Catering Tenders)

Catering Tenders					
Item No.	**Municipality**	**Date**	**Tender Number**	**Price**	**BEE Score**
1	Cape Agulhas	18/03/2019	Q17/2018/19	R 55 200.00	20
2	Midvaal	2019/04/24	AHT4544RE	R 87 175	20
3	Midvaal	18/07/2018	AHT4506	R 89 355	20
4	Midvaal	07/06/2019	AHT4560	R 64 687.50	20
5	Nkangala	14/01/2016	YU/002/01/16	R 42 500.00	18
6	Witzenberg	27/09/2016	08/2/13/80	R 199 815.00	20

Table C: Data for Exploration of Commonly Procured Items (Office Furniture)

Office Furniture Tenders					
Item No.	**Municipality**	**Date**	**Tender Number**	**Price**	**BEE Score**
1	Cape Agulhas	29/09/2017	Q1/2017/18	R 65 890.86	0
2	Cape Agulhas	06/09/2018	SCM11/2018/19	R 114 229.50	20
3	Hessequa	1/11/2016	RFQ 52416	R 110 032.80	18
4	Hessequa	22/12/2017	RFQ 57516	R 30 680.00	0
5	Hessequa	04/10/2018	RFQ 60873	R 21 282.58	20
6	Hessequa	08/02/2018	REQ: 58351	R 26 191.48	20
7	Hessequa	30/09/2019	RFQ: 65864	R 48 400.60	18
8	Midvaal	05/11/2018	AHT4520	R 129 860.30	20
9	Nkangala	31/10/2016	Not Available	R 827 500.00	20

Section 2

Table D: Consumer Price Index Data Transformation Figures.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
Geographic indices - CPI per province - Western Cape, Conversion factor: 0.7961783439														
2014	Index	84.3	85	86.1	86.4	86.7	86.9	87.7	88	88.3	88.5	88.7	88.5	87.1
2015	Index	88.5	88.8	90	90.8	91.1	91.4	92.3	92.3	92.6	92.9	93.1	93.3	91.4
2016	Index	93.9	95	95.8	96.4	96.4	97.1	97.9	97.7	98.1	98.6	99	100	97.2
2017	Index	100.5	101.9	102.5	102.5	102.7	102.9	103.2	103.3	104.3	104.8	104.9	105.3	103.2
2018	Index	105.7	106.7	107	107.9	108	108.5	109.3	109.3	110.1	110.5	110.9	110.8	108.7
2019	Index	110.6	111.7	112.9	113.5	113.8	114.3	114.6	114.7	115.2	115.2	115.4	115.7	114
2020	Index	116.2	.	.	.	.	.	.	.	.	.	.	.	.
Geographic indices - CPI per province - KwaZulu-Natal, Conversion factor: 0.8														
2014	Index	84.8	85.6	87	87.6	87.7	87.9	88.6	88.8	88.7	88.8	88.7	88.6	87.7
2015	Index	88.4	88.8	90	90.7	91	91.4	92.4	92.4	92.6	92.8	92.8	93	91.4
2016	Index	93.8	95	96	97	97.1	97.8	98.9	98.7	99	99.4	99.8	100	97.7
2017	Index	100.6	101.6	102	102.1	102.3	102.6	102.7	102.9	103.2	103.5	103.7	104.1	102.6
2018	Index	104.4	104.9	105.2	105.9	106.2	106.4	107.3	107.2	107.6	108.1	108.4	108.1	106.6
2019	Index	108.2	108.7	109.5	110	110.4	110.7	111.3	111.6	111.9	112	112.1	112.1	110.7
2020	Index	112.5	.	.	.	.	.	.	.	.	.	.	.	.
Geographic indices - CPI per province - Gauteng, Conversion factor: 0.8019246191														
2014	Index	85.2	86.4	87.5	87.9	88	88.3	88.9	89.3	89.2	89.3	89.3	89.1	88.2
2015	Index	88.9	89.7	90.9	91.7	92.1	92.4	93.3	93.3	93.3	93.5	93.6	93.9	92.2
2016	Index	94.6	96	96.6	97.4	97.7	98.2	99	98.8	99	99.4	99.8	100	98
2017	Index	100.7	101.8	102.4	102.5	102.8	103.2	103.5	103.5	103.9	104.2	104.3	104.8	103.1
2018	Index	105.1	105.9	106.4	107.2	107.4	107.9	108.9	108.9	109.2	109.7	109.9	109.6	108
2019	Index	109.4	110.3	111	111.7	112.1	112.5	113	113.4	113.6	113.6	113.6	113.9	112.3
2020	Index	114.2	.	.	.	.	.	.	.	.	.	.	.	.
Geographic indices - CPI per province - Mpumalanga, Conversion factor: 0.8110300081														
2014	Index	85.6	86.5	87.5	87.9	88.3	88.5	88.9	89.5	89.5	89.6	89.6	89.5	88.4
2015	Index	89.5	90	90.9	91.7	92.1	92.5	93.1	93.3	93	93.1	93.3	93.5	92.2
2016	Index	94.5	96.2	96.8	97.6	97.9	98.1	98.8	98.9	99	99.5	99.8	100	98.1
2017	Index	100.6	101.4	101.7	101.8	102.2	102.4	102.5	102.5	102.8	103	103.3	103.8	102.3
2018	Index	104.2	104.7	105	105.7	105.8	106.2	106.7	106.6	107.7	108.2	108.3	108	106.4
2019	Index	108.1	108.9	109.6	110.2	110.5	110.8	111.1	111.3	111.7	111.8	112.1	112.2	110.7
2020	Index	112.6	.	.	.	.	.	.	.	.	.	.	.	.

Appendix C

Section 1

Table A: Clean Audited Municipalities Data Summary and Combined Variables *(**CF**=cost factor, **EF**= escalation factor, **MC**= material cost)*

Tender Number	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
1	16/04/06	R 411 814.70	96.4	785	2.1	960	352	709632	4	2.1	1	352	2956.8	6	2.1	1.6	352	7096.32	20
2	16/05/16	R 141 220.00	96.4	50	1.8	500	306	275400	0	0	0	0	0	0	0	0	0	0	20
3	16/12/01	R 124 945.00	100	0	2.1	150	352	110880	8	2.1	1	352	5913.6	6	2.1	1	352	4435.2	20
4	20/02/12	R 152 492.00	116.2	0	2	89	1008	179424	1.8	2	1	1008	3628.8	4	2	1	1008	8064	20
5	19/05/29	R 189 400.00	113.8	0	1.8	300	458	247320	5	1.8	1	306	2754	0	0	0	0	0	20
6	17/06/09	R 122 946.50	102.9	54	1	120	306	36720	4.5	1	1	306	1377	0	0	0	0	0	20
7	17/11/17	R 158 980.00	104.9	0	2.1	35	1008	74088	9	2.1	1	1008	19051.2	0	0	0	0	0	20
8	18/04/26	R 70 460.44	107.9	0	1.2	608	306	223257.6	0.9	1.2	1	306	330.48	3.65	1.2	1.6	306	2144.448	20
9	19/03/18	R 107 870.00	112.9	0	1.2	480	306	176256	0.9	1.2	1	306	330.48	3.65	1.2	1.6	306	2144.448	20
10	19/04/17	R 116 000.00	113.5	0	1.2	448	306	164505.6	1.8	1.2	1	306	660.96	3.65	1.2	1.6	306	2144.448	20
11	19/09/16	R 39 150.00	115.2	0	1.8	27	306	14871.6	0	0	0	0	0	7	1.8	1	306	3855.6	20
12	19/10/28	R 76 000.00	115.4	0	1.2	520	306	190944	0.9	1.2	1	306	330.48	0	0	0	0	0	20
13	15/09/23	R 88 006.97	92.6	32	2.1	23	1008	48686.4	0	0	0	0	0	10	2.1	1.6	1008	33868.8	20
14	16/11/18	R 103 492.62	99	2	2.4	21	1008	50803.2	0	0	0	0	0	4.5	2.4	1.6	1008	17418.24	20
15	16/11/18	R 66 200.00	99	0	0.6	244	458	67051.2	0	0	0	0	0	0	0	0	0	0	16
16	17/11/17	R 137 615.00	104.9	0	1.8	301	352	190713.6	1	2.4	1	352	844.8	3	2.4	1	352	2534.4	20
17	18/07/27	R 240 880.00	109.3	0	2.4	185	1008	447552	0	0	0	0	0	0	0	0	0	0	20
18	18/08/01	R 130 640.00	109.3	55	2.4	48	1008	116121.6	0	0	0	0	0	8	2.4	1.6	1008	30965.76	20
19	18/11/27	R 1 014 121.75	110.9	850	2.2	850	1008	1884960	4.5	2.2	1	1008	9979.2	9	2.2	1.6	1008	31933.44	20
20	19/10/11	R 495 834.00	115.4	0	2.4	216	703	364435.2	0	0	0	703	0	6	2.1	1	703	8857.8	20
21	16/07/29	R 175 673.40	97.9	0	1.8	202.5	1008	367416	13	1.8	1	1008	23587.2	0	0	0	0	0	20
22	17/03/17	R 173 707.50	102.5	0	2.1	150	1008	317520	1	2.1	1	1008	2116.8	5	2.1	1	1008	10584	20
23	19/05/03	R 37 455.00	113.8	0	1.8	15	1008	27216	1	1.8	1	1008	1814.4	0	0	1	0	0	20

Tender Number	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
24	19/05/10	R 199 404.25	113.8	0	2.1	175	1008	370440	4	2.1	1	1008	8467.2	7	2.1	1	1008	14817.6	20
25	20/02/21	R 115 575.00	116.2	0	1.8	150	1008	272160	0	0	0	0	0	0	0	0	0	0	20
26	20/02/21	R 38 720.00	116.2	0	1.8	15	1008	27216	2	1.8	1	1008	3628.8	0	0	0	0	0	20
27	17/06/21	R 105 165.00	103.2	0	2	162	306	99144	6	1.8	1	306	3304.8	0	0	0	0	0	20
28	17/05/10	R 1 188 290.00	102.7	0	2.4	1200	1008	2903040	4.2	2.4	1	1008	10160.64	0	0	0	0	0	20
29	17/11/14	R 170 085.00	104.9	0	3	430	352	454080	0	0	0	0	0	0	0	0	0	0	20
30	18/03/06	R 1 133 177.00	107	0	2.4	551	352	465484.8	0	0	0	0	0	45	2.4	1	352	38016	20
31	18/05/08	R 314 433.00	108	0	1.8	170	1008	308448	1.2	1.8	1	1008	2177.28	12	1.8	1	1008	21772.8	20
32	18/05/16	R 823 285.00	108	0	2.4	867	352	732441.6	1.2	2.4	1	352	1013.76	13	2.4	1	352	10982.4	20
33	18/06/13	R 214 434.00	108.5	200	1.6	200	703	224960	3.6	1.6	1.6	703	6478.848	0	0	0	0	0	20
34	19/01/29	R 1 759 334.00	110.6	0	2.4	783	352	661478.4	0	0	0	0	0	45	2.4	1	352	38016	18
35	15/04/10	R 122 892.00	90.8	0	2.4	72	306	52876.8	3	2.4	1	306	2203.2	0	0	0	0	0	20
36	15/05/08	R 197 064.96	91.1	0	2.4	165	352	139392	3	2.4	1	352	2534.4	0	0	0	0	0	20
37	15/08/07	R 51 992.89	92.3	0	2.1	73	703	107769.9	1.2	1.8	1	703	1518.48	0	0	0	0	0	20
38	16/02/05	R 81 900.00	95	3	1.8	80	352	50688	3	1.8	1	352	1900.8	0	0	0	0	0	20
39	16/05/06	R 67 055.00	96.4	0	1.8	47	352	29779.2	4	3.3	1	352	4646.4	0	0	0	0	0	20
40	16/10/21	R 49 476.00	98.6	50	1.8	50	352	31680	0	0	0	0	0	0	0	0	0	0	20
41	16/11/11	R 581 085.00	99	0	2.4	306	352	258508.8	5	2.4	1	352	4224	0	0	0	0	0	20
42	17/02/03	R 142 831.54	101.9	124	1.6	122	352	68710.4	0	0	0	0	0	0	0	0	0	0	20
43	17/07/28	R 131 850.00	103.2	2.4	1.8	51	352	32313.6	2.4	1.8	1	352	1520.64	0	0	0	0	0	20
44	17/07/28	R 166 269.00	103.2	191	1.8	191	1008	346550.4	1	1.8	1	1008	1814.4	4	1.8	1	1008	7257.6	20
45	17/10/01	R 136 080.00	104.8	0	1.8	112	1008	203212.8	0	0	0	0	0	0	0	0	0	0	12
46	17/10/06	R 112 000.00	104.8	0	1.8	90	352	57024	0	0	0	0	0	0	0	0	0	0	20
47	17/10/13	R 83 562.00	104.8	0	1.8	64	1008	116121.6	0	0	0	0	0	4	1.8	1	1008	7257.6	12
48	17/12/08	R 119 780.00	105.3	105	1.8	100	352	63360	0	0	0	0	0	5	1.8	1	352	3168	20
49	17/12/08	R 112 750.00	105.3	90	2	90	352	63360	0	0	0	0	0	0	0	0	0	0	20
50	18/03/23	R 492 000.00	107	0	1.8	196	703	248018.4	0	0	0	0	0	0	0	0	0	0	20

Tender Number	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
51	18/04/06	R 152 100.00	107.9	0	2.4	39	352	32947.2	0	0	0	0	0	0	0	0	0	0	20
52	18/04/06	R 105 000.00	107.9	59	1.8	49	352	31046.4	1.8	1.8	1	352	1140.48	8	1.8	1	352	5068.8	20
53	18/05/11	R 97 970.00	108	83	1.8	83	352	52588.8	0	0	0	0	0	0	0	0	0	0	20
54	18/08/03	R 127 208.80	109.3	0	1.8	90	352	57024	0	0	0	0	0	4	1.8	1	352	2534.4	20
55	18/08/03	R 40 800.00	109.3	9	1.8	9	1008	16329.6	0	0	0	0	0	0	0	0	0	0	12
56	18/08/03	R 95 845.00	109.3	59	1.8	49	352	31046.4	1.8	1.8	1	352	1140.48	8	1.8	1	352	5068.8	20
57	18/08/03	R 84 147.50	109.3	41	1.8	51	352	32313.6	2.4	1.8	1	352	1520.64	0	0	0	0	0	20
58	18/08/03	R 96 820.00	109.3	0	1.8	75	352	47520	1.8	1.8	1	352	1140.48	0	0	0	0	0	20
59	18/09/07	R 42 940.00	110.1	0	1.8	42	1008	76204.8	0	0	0	0	0	0	0	0	0	0	20
60	19/02/01	R 67 405.89	111.7	33	1.8	33	352	20908.8	0	0	0	0	0	0	0	0	0	0	12
61	19/02/01	R 125 644.22	111.7	87	1.8	87	352	55123.2	4	1.8	1	352	2534.4	3	1.8	1	352	1900.8	20
62	19/02/01	R 81 462.00	111.7	60	1.8	57	352	36115.2	3	2.1	1	352	2217.6	0	0	0	0	0	20
63	19/02/01	R 125 310.21	111.7	80	2.1	79	352	58396.8	5	2.1	1	352	3696	3	1.8	1	352	1900.8	20
64	19/03/01	R 255 031.10	112.9	173	1.8	173	352	109612.8	0	0	0	0	0	0	0	0	0	0	20
65	19/04/18	R 1 713 902.07	113.5	900	2.4	900	352	760320	0	0	0	0	0	2.4	6	1	352	5068.8	20
66	19/05/17	R 92 704.00	113.8	0	1.8	51	352	32313.6	2.9	1.8	1	352	1837.44	0	0	0	0	0	20
67	19/05/17	R 194 258.63	113.8	120	1.8	114	352	72230.4	8	1.8	1	352	5068.8	0	0	0	0	0	20
68	19/05/17	R 104 209.00	113.8	0	1.8	87	352	55123.2	0	0	0	0	0	0	0	0	0	0	20
69	19/08/16	R 365 600.00	114.7	260	2.4	260	1528	953472	3.6	2.4	1	352	3041.28	0	0	0	0	0	20
70	19/09/06	R 76 570.00	115.2	0	1.8	56	352	35481.6	0.9	1.8	1	352	570.24	0	0	0	0	0	20
71	19/09/06	R 119 440.00	115.2	0	1.8	125	352	79200	0	0	0	0	0	0	0	0	0	0	20
72	19/10/04	R 273 786.60	115.2	19	1.8	172	352	108979.2	6	1.8	1	352	3801.6	0	0	0	0	0	20
73	19/10/25	R 189 243.00	115.2	0	1.8	200	352	126720	0	0	0	0	0	3	1.8	1	352	1900.8	20
74	19/11/01	R 179 486.00	115.4	118	1.8	118	352	74764.8	2	1.8	1	352	1267.2	10	1.8	1	352	6336	20
75	20/02/01	R 102 506.31	116.2	71	2.1	78	352	57657.6	4	2.1	1	352	2956.8	0	0	0	0	0	20
76	18/09/27	R 241 500.00	110.1	0	1.8	780	352	494208	0	0	0	0	0	5	1.8	1	352	3168	20

Tender Number	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
77	18/10/09	R 58 305.00	110.5	0	1.2	300	306	110160	4.7	1.2	1	306	1725.84	0	0	0	0	0	20
78	18/11/08	R 86 135.00	110.9	0	1.8	244	306	134395.2	5	1.8	1	306	2754	7	1.8	1	306	3855.6	20
79	18/12/12	R 178 945.75	110.8	0	2.4	95	1008	229824	7.5	2.4	1	1008	18144	3.5	2.4	1	1008	8467.2	20
80	19/04/24	R 61 237.50	113.5	0	1.8	150	306	82620	5	1.8	1	306	2754	0	0	0	0	0	20
81	19/11/12	R 198 858.00	115.4	215	1.8	380	458	313272	11	1.8	1	306	6058.8	0	0	0	0	0	20
82	14/07/01	R 48 790.00	87.7	80	1.8	80	306	44064	5	1.8	1	306	2754	0	0	0	0	0	0
83	14/08/01	R 323 075.54	88	0	1.8	203	1528	558331.2	0	0	0	0	0	0	0	0	0	0	20
84	16/01/01	R 136 400.00	93.9	0	1.8	136	1008	246758.4	0	0	0	0	0	14	1.8	1	1008	25401.6	18
85	16/05/01	R 451 642.89	96.4	0	1.8	199	1528	547329.6	0	0	0	0	0	17	1.8	1	1008	30844.8	20
86	16/07/01	R 213 111.60	97.9	0	1.8	160	1008	290304	0	0	0	0	0	14	1.8	1	1008	25401.6	18
87	17/02/01	R 136 800.00	101.9	0	2.1	136	1008	287884.8	2	2.1	1	1008	4233.6	12	2.1	1	1008	25401.6	0
88	17/02/15	R 199 488.60	101.9	0	2.1	216	1008	457228.8	2	2.1	1	1008	4233.6	6	2.1	1	1008	12700.8	20
89	17/04/12	R 167 648.40	102.5	175	2.1	175	1008	370440	2	2.1	1	1008	4233.6	6	2.1	1	1008	12700.8	20
90	19/11/29	R 220 000.00	115.4	0	2	220	703	309320	2.4	2.1	1	703	3543.12	4.7	2.1	1	703	6938.61	20
91	19/11/26	R 235 017.45	115.4	0	2	165	703	231990	6	2.1	1	703	8857.8	0	0	0	0	0	20
92	20/03/01	R 253 287.50	116.2	0	2	170	703	239020	6	2.1	1	703	8857.8	0	0	0	0	0	20
93	17/01/02	R 69 094.51	100.5	0	2.1	120	306	77112	4	2.1	1	306	2570.4	0	0	0	0	0	20
94	17/08/17	R 98 227.09	103.3	0	1.8	210	306	115668	0	0	0	0	0	8	1.8	1	306	4406.4	20
95	15/11/13	R117 294.60	93.1	0	1.8	218	306	120074.4	3	1.8	1	306	1652.4	0	0	0	0	0	18
96	15/06/02	R379 527.66	91.4	0	1.5	521	306	239139	0	0	0	0	0	0	0	0	0	0	20
97	19/02/17	R 61 327.00	111.7	40	2	46	703	64676	0	0	0	0	0	0	0	0	0	0	18
98	19/03/29	R 59 733.00	112.9	0	1.8	150	306	82620	1	1.8	1	306	550.8	0	0	0	0	0	20
99	19/05/10	R 131 327.00	113.8	0	1.8	150	703	189810	5	1.8	1	703	6327	0	0	0	0	0	20
100	19/06/14	R 198 202.50	114.3	0	1.8	150	703	189810	11	1.8	1	703	13919.4	0	0	0	0	0	20
101	18/12/20	R 199 182.00	109.6	0	1.8	170	352	107712	0.9	1.8	1	352	570.24	3	1.8	1.6	352	3041.28	20
102	16/04/21	R 112 800.79	98.00	97.40	1.8	151	306	83170.8	0	0	0	0	0	0	0	0	0	0	18

Tender Number	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
103	19/02/27	R 84 170.00	108.70	0	1.8	20	1011.3	36406	0	0	0	0	0	8	1.8	1	306	4406.4	20
104	19/05/29	R149 000.00	110.40	0	2.4	350	306	257040	0	0	0	0	0	0	0	0	0	0	20
105	19/05/31	R 54 000.00	110.40	0	2	40	1528	122240	1	2	1	306	612	0	0	0	0	0	20
106	18/10/01	R138 869.92	110.50	0	2	149	703	209494	0	0	0	0	0	20	2	1	703	28120	#20
107	18/11/16	R 65 600.00	110.90	0	2	120	306	73440	5	2	1	703	7030	0	0	0	0	0	#20
108	18/02/15	R 39 992.25	106.70	0	2.1	80	352	59136	0	0	0	0	0	0	0	0	0	0	20
109	19/12/03	R200 000.00	108	0	1.8	160	1008	290304	0	0	0	0	0	0	0	0	0	0	#20

Table B: Non-Clean Audited Municipalities Data Summary and Combined Variables

Tender Identity	Date	Price (as awarded)	CPI	Fencing Specifications															
				Fence	Fence				Hinged Gate					Sliding Gate					BEE
				Removal	H	L	C.F	M.C	L	H	E.F	C.F	M.C	L	H	E.F	C.F	M.C	Score
A	18/11/30	R 204 700.00	109.90	208	2	206	1008	415296	1.5	2	1	1008	3024	6	2	1	1008	12096	20
B	18/01/16	R 112 255.26	105.10	0	1.8	40	750	54000	7	1.8	1	750	9450	0	0	0	0	0	16
C	18/08/01	R 195 500.00	108.90	0	1.8	80	703	101232	1.2	1.8	1	703	1518.48	17	1.8	1	703	21511.8	20
D	19/01/11	R 60 000.00	109.40	0	2.4	70	1528	256704	0	0	0	0	0	0	0	0	0	0	#20
E	14/04/15	R 197 000.00	87.90	220	1.8	280	1528	770112	0	0	0	0	0	0	0	0	0	0	#20
F	15/04/08	R 197 000.00	91.70	0	1.8	150	1574	424980	0	0	0	0	0	0	0	0	0	0	#20
G	16/04/01	R 133 000.00	98.00	0	1.8	102	1008	185068.8	0	0	0	0	0	0	0	0	0	0	#20
H	18/11/17	R 170 200.00	109.90	325	2.1	325	1008	687960	0	0	0	0	0	5	2.1	1	1008	10584	20
I	19/11/15	R 181 613.75	113.60	0	3	171	1528	783864	0	0	0	0	0	0	0	0	0	0	20
J	18/11/19	R 198 766.00	109.90	0	1.8	204	1008	370137.6	0	0	0	0	0	4	1.8	1.6	1008	11612.16	20
K	18/11/28	R 33 800.00	109.90	0	2.2	13	1008	28828.8	0	0	0	0	0	0	0	0	0	0	#20
L	15/05/29	R 73 283.76	92.10	0	1.8	22	1528	60508.8	0	0	0	0	0	0	0	0	0	0	16

Section 2

Table C: Clean Audited Municipalities Transformed Variables Data Summary *(For Model Building)*

Tender Number	ROF^2	√FMCE	HGMCE	√SGMCE	Scaled GTV
2	2500	524.7856705	0	0	R 168 760.83
3	0	332.9864862	5913.6	66.59729724	R 143 936.64
4	0	423.5847023	3628.8	89.79977728	R 151 179.68
6	2916	191.6246331	1377	0	R 137 642.73
7	0	272.1911093	19051.2	0	R 174 590.05
8	0	472.5014286	330.48	46.30818502	R 75 227.46
9	0	419.8285364	330.48	46.30818502	R 110 067.53
10	0	405.5928993	660.96	46.30818502	R 117 737.44
11	0	121.9491697	0	62.09347792	R 39 150.00
12	0	436.9713949	330.48	0	R 75 868.28
13	1024	220.649949	0	184.0347793	R 109 485.99
14	4	225.3956521	0	131.97818	R 120 427.78
15	0	258.9424647	0	0	R 80 242.42
16	0	436.7076826	844.8	50.34282471	R 151 127.24
21	0	606.1484967	23587.2	0	R 206 716.81
22	0	563.4891303	2116.8	102.8785692	R 195 230.28
23	0	164.972725	1814.4	0	R 37 915.78
24	0	608.6378233	8467.2	121.7275647	R 201 857.38
25	0	521.6895629	0	0	R 114 580.38
26	0	164.972725	3628.8	0	R 38 386.78
27	0	314.8714023	3304.8	0	R 117 393.49
29	0	673.8545837	0	0	R 186 785.43
33	40000	474.2994834	6478.848	0	R 227 675.55
35	0	229.9495597	2203.2	0	R 155 915.84
36	0	373.3523805	2534.4	0	R 249 197.40
37	0	328.2832618	1518.48	0	R 64 892.53
38	9	225.1399565	1900.8	0	R 99 314.53
39	0	172.5665089	4646.4	0	R 80 132.12
40	2500	177.9887637	0	0	R 57 805.63
42	15376	262.1266869	0	0	R 161 473.93
43	5.76	179.7598398	1520.64	0	R 147 181.40
44	36481	588.6853149	1814.4	85.19154888	R 185 602.60
45	0	450.7913043	0	0	R 162 591.44
46	0	238.7969849	0	0	R 123 114.50
47	0	340.7661955	0	85.19154888	R 99 841.75
48	11025	251.7141236	0	56.28498912	R 131 041.37
49	8100	251.7141236	0	0	R 123 350.43
51	0	181.5136359	0	0	R 162 390.36
52	3481	176.1998865	1140.48	71.19550548	R 112 103.80
53	6889	229.3224804	0	0	R 104 501.33
54	0	238.7969849	0	50.34282471	R 134 075.51
55	81	127.7873233	0	0	R 46 741.72

Tender Number	ROF2	√FMCE	HGMCE	√SGMCE	Scaled GTV
56	3481	176.1998865	1140.48	71.19550548	R 101 018.70
57	1681	179.7598398	1520.64	0	R 88 689.77
58	0	217.9908255	1140.48	0	R 102 046.33
59	0	276.052169	0	0	R 44 929.05
60	1089	144.5987552	0	0	R 75 563.02
61	7569	234.7833043	2534.4	43.5981651	R 129 581.15
62	3600	190.0399958	2217.6	0	R 84 014.52
63	6400	241.6542985	3696	43.5981651	R 129 236.67
66	0	179.7598398	1837.44	0	R 93 844.47
67	14400	268.7571394	5068.8	0	R 196 648.45
68	0	234.7833043	0	0	R 105 491.01
70	0	188.365602	570.24	0	R 76 570.00
71	0	281.4249456	0	0	R 119 440.00
73	0	355.9775274	0	43.5981651	R 189 243.00
74	13924	273.4315271	1267.2	79.59899497	R 179 174.93
75	5041	240.11997	2956.8	0	R 101 624.16
78	0	366.599509	2754	62.09347792	R 89 474.77
79	0	479.3996245	18144	92.01738966	R 186 051.90
80	0	287.4369496	2754	0	R 62 154.71
81	46225	559.7070662	6058.8	0	R 198 513.36
82	6400	209.9142682	2754	0	R 80 111.29
84	0	496.7478233	0	159.3787941	R 170 755.69
86	0	538.7986637	0	159.3787941	R 255 888.52
87	0	536.5489726	4233.6	159.3787941	R 193 318.94
88	0	676.1869564	4233.6	112.6978261	R 225 525.88
89	30625	608.6378233	4233.6	112.6978261	R 188 420.45
90	0	556.165443	3543.12	83.29831931	R 219 618.72
93	0	277.6904752	2570.4	0	R 79 200.87
95	0	346.517532	1652.4	0	R 148 099.89
97	1600	254.3147656	0	0	R 64 539.40
98	0	287.4369496	550.8	0	R 60 949.88
99	0	435.6718949	6327	0	R 132 942.62
100	0	435.6718949	13919.4	0	R 199 763.15
101	0	328.1950639	570.24	55.14780141	R 209 359.18
102	0	288.3934812	0	0	R 136 138.07
103	0	190.8035639	0	66.38072009	R 89 203.16
104	0	506.9911242	0	0	R 155 478.26
105	0	349.6283741	612	0	R 56 347.83

Table D: Clean Audited Municipalities Transformed Variables Data Summary *(For Model Testing)*

Tender Number	ROF^2	√FMCE	HGMCE	√SGMCE	Scaled GTV
5	0	497.3127788	2754	0	R 191 730.05
18	3025	340.7661955	0	175.9709067	R 137 691.93
77	0	331.9036005	1725.84	0	R 60 784.94
94	0	340.0999853	0	66.38072009	R 109 542.70
106	0	457.7051453	0	167.6901905	R 144 776.60
107	0	270.998155	7030	0	R 68 143.55
108	0	243.1789465	0	0	R 43 178.14
109	0	538.7986637	0	0	R 213 333.33

Table E: Clean Audited Municipalities Transformed Variables Data Summary *(For out of scope Data)*

Tender Number	ROF^2	√FMCE	HGMCE	√SGMCE	Scaled GTV
1	616225	842.3965812	2956.8	84.23965812	R 492 127.11
17	0	668.9932735	0	0	R 253 882.67
19	722500	1372.938455	9979.2	178.6993005	R 1 053 442.97
20	0	603.6846859	0	94.11588601	R 494 974.67
28	0	1703.830978	10160.64	0	R 1 332 921.21
30	0	682.2644649	0	194.9769217	R 1 220 018.60
31	0	555.3809503	2177.28	147.556091	R 335 395.20
32	0	855.8280201	1013.76	104.7969465	R 878 170.67
34	0	813.3132238	0	194.9769217	R 1 869 905.13
41	0	508.4376068	4224	0	R 676 171.64
50	0	498.0144576	0	0	R 529 704.67
64	29929	331.0782385	0	0	R 260 226.60
65	810000	871.963302	0	71.19550548	R 1 739 572.85
69	67600	976.4589085	3041.28	0	R 367 193.72
72	361	330.1199782	3801.6	0	R 273 786.60
76	0	702.9992888	0	56.28498912	R 252 686.65
83	0	747.2156315	0	0	R 422 935.25
85	0	739.8172747	0	175.6268772	R 539 722.62
91	0	481.6534024	8857.8	0	R 234 610.14
92	0	488.8967171	8857.8	0	R 251 107.75
96	0	489.0184046	0	0	R 478 354.34

Table F: Non-Clean Audited Municipalities Transformed Variables *(For Benchmarking)*

Tender Identity	ROF^2	√FMCE	HGMCE	√SGMCE	Scaled GTV
A	43264	644.4346359	3024	109.9818167	214571.79
B	0	232.3790008	9450	0	128169.66
C	0	318.169766	1518.48	146.6690151	206809.92
D	0	506.6596491	0	0	63180.99
E	48400	877.5602543	0	0	258184.30
F	0	651.904901	0	0	247485.28
G	0	430.1962343	0	0	156342.86
H	105625	829.4335416	0	102.8785692	178408.01
I	0	885.3609433	0	0	184171.69
J	0	608.389349	0	107.7597327	208351.62
K	0	169.7904591	0	0	35430.03
L	0	245.9853654	0	0	95483.73

Section 3

Table G: IBM SPSS 2016 Model Building Data Pearson Correlation Assessment

Correlations					
		ROF^(2)	FMCE^(0.5)	HGMCE^(1)	SGMCE^(0.5)
ROF^(2)	Pearson Correlation	1	.233*	0.084	0.017
	Sig. (2-tailed)		0.037	0.461	0.878
	N	80	80	80	80
FMCE^(0.5)	Pearson Correlation	.233*	1	.304**	.364**
	Sig. (2-tailed)	0.037		0.006	0.001
	N	80	80	80	80
HGMCE^(1)	Pearson Correlation	0.084	.304**	1	0.001
	Sig. (2-tailed)	0.461	0.006		0.991
	N	80	80	80	80
SGMCE^(0.5)	Pearson Correlation	0.017	.364**	0.001	1
	Sig. (2-tailed)	0.878	0.001	0.991	
	N	80	80	80	80

Section 4

Table H: IBM SPSS Various Model Builds Assessment Outputs

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
1	.722a	0.522	0.496	38469.41464

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121134682133.83	4	30283670533	20.463	.000b
	Residual	110992189696.32	75	1479895863		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	52594.925	10800.886		4.87	0
	ROF^(2)	1.061	0.505	0.173	2.102	0.039
	FMCE^(0.5)	182.362	34.581	0.491	5.273	0
	HGMCE^(1)	2.622	1.072	0.207	2.446	0.017
	SGMCE^(0.5)	205.45	96.866	0.184	2.121	0.037

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
2	.703a	0.494	0.474	39324.59932

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
2	Regression	114598639330.72	3	38199546444	24.702	.000b
	Residual	117528232499.42	76	1546424112		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
2	(Constant)	51126.79	11017.874		4.64	0
	FMCE^(0.5)	199.07	34.403	0.536	5.786	0
	HGMCE^(1)	2.632	1.096	0.207	2.402	0.019
	SGMCE^(0.5)	190.51	98.752	0.17	1.929	0.057

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
3	.695a	0.484	0.463	39710.01554

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
3	Regression	112283586453.07	3	37427862151	23.735	.000b
	Residual	119843285377.07	76	1576885334		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
3	(Constant)	51149.275	11132.494		4.595	0
	ROF^(2)	1.066	0.521	0.174	2.047	0.044
	FMCE^(0.5)	209.083	33.868	0.563	6.174	0

	SGMCE^(0.5)	176.459	99.238	0.158	1.778	0.079

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
4	.702a	0.493	0.473	39344.89115

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
4	Regression	114477316887.02	3	38159105629	24.65	.000b
	Residual	117649554943.12	76	1548020460		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
4	(Constant)	50908.082	11016.701		4.621	0
	ROF^(2)	0.982	0.515	0.16	1.908	0.06
	FMCE^(0.5)	210.746	32.612	0.568	6.462	0
	HGMCE^(1)	2.343	1.088	0.185	2.154	0.034

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
5	.675a	0.455	0.441	40523.85149

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
5	Regression	105678816272.41	2	52839408136	32.176	.000b
	Residual	126448055557.74	77	1642182540		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
5	(Constant)	49667.81	11336.606		4.381	0
	FMCE^(0.5)	225.983	33.519	0.609	6.742	0
	SGMCE^(0.5)	161.327	100.991	0.144	1.597	0.114

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
6	.680a	0.462	0.448	40263.58175

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
6	Regression	107297858671.12	2	53648929336	33.093	.000b
	Residual	124829013159.03	77	1621156015		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
6	(Constant)	49812.228	11261.905		4.423	0
	ROF^(2)	0.997	0.527	0.163	1.893	0.062
	FMCE^(0.5)	231.36	31.904	0.623	7.252	0

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
7	.685a	0.469	0.455	40013.55926

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
7	Regression	108843332656.97	2	54421666328	33.99	.000b
	Residual	123283539173.18	77	1601084924		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
7	(Constant)	49655.53	11184.016		4.44	0
	FMCE^(0.5)	224.379	32.361	0.604	6.934	0
	HGMCE^(1)	2.372	1.107	0.187	2.143	0.035

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
8	.661a	0.437	0.430	40924.9829

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
8	Regression	101488242264.45	1	1.01488E+11	60.595	.000b
	Residual	130638629565.70	78	1674854225		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
8	(Constant)	48526.759	11426.075		4.247	0
	FMCE^(0.5)	245.457	31.532	0.661	7.784	0

Table I: Model Selected for Benchmarking Assessment

Model Summary

Model	r	R Square	Adjusted R Square	Std. Error of the Estimate
1	.722a	0.522	0.496	38469.41464

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121134682133.83	4	30283670533	20.463	.000b
	Residual	110992189696.32	75	1479895863		
	Total	232126871830.14	79			

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	VIF
1	(Constant)	52594.9245536835	10800.886		4.87	0	
	ROF^(2)	1.0605303664291	0.505	0.173	2.102	0.039	1.064
	FMCE^(0.5)	182.362407474448	34.581	0.491	5.273	0	1.361
	HGMCE^(1)	2.6215959163187	1.072	0.207	2.446	0.017	1.119
	SGMCE^(0.5)	205.450246629663	96.866	0.184	2.121	0.037	1.176

Section 5

Table J: Summary of the Price Predicted Results for Model Testing Data

Identifier	Awarded Value (SGTV)	Predicted Value	Deviation Above Predicted
5	R 191 730.05	R 150 505.96	27.39%
18	R 137 691.93	R 154 099.24	-10.65%
77	R 60 784.94	R 117 646.12	-48.33%
94	R 109 542.70	R 128 254.31	-14.59%
106	R 144 776.60	R 170 515.13	-15.09%
107	R 68 143.55	R 120 444.62	-43.42%
108	R 43 178.14	R 96 941.62	-55.46%
109	R 213 333.33	R 150 851.55	41.42%

Table K: Summary of the Price Predicted Results for Model Out-of-Range Data

Identifier	Awarded Value (SGTV)	Predicted Value	Deviation Above Predicted
1	R 492 127.11	R 884 800.31	-44.38%
17	R 253 882.67	R 174 594.15	45.41%
19	R 1 053 442.97	R 1 132 075.72	-6.95%
20	R 494 974.67	R 182 020.45	171.93%
28	R 1 332 921.21	R 389 946.74	241.82%
30	R 1 220 018.60	R 217 072.37	462.03%
31	R 335 395.20	R 189 898.92	76.62%
32	R 878 170.67	R 232 854.01	277.13%
34	R 1 869 905.13	R 240 970.74	675.99%
41	R 676 171.64	R 156 388.45	332.37%
50	R 529 704.67	R 143 414.04	269.35%
64	R 260 226.60	R 144 711.76	79.82%
65	R 1 739 572.85	R 1 085 264.98	60.29%
Identifier	Awarded Value (SGTV)	Predicted Value	Deviation Above Predicted
69	R 367 193.72	R 310 329.18	18.32%
72	R 273 786.60	R 123 145.51	122.33%
76	R 252 686.65	R 192 359.33	31.36%
83	R 422 935.25	R 188 858.97	123.94%
85	R 539 722.62	R 223 592.37	141.39%
91	R 234 610.14	R 163 651.97	43.36%
92	R 251 107.75	R 164 972.88	52.21%
96	R 478 354.34	R 141 773.50	237.41%

Table L: Prediction Interval Results and Remarks for Model Testing Data

Identifier	SGTV	Lower Bound	Upper Bound	Remark	Margin
5	R 191 730.05	R 86 438.02	R 214 573.89	None	N/A
18	R 137 691.93	R 90 031.30	R 218 167.18	None	N/A
77	R 60 784.94	R 53 578.18	R 181 714.06	None	N/A
94	R 109 542.70	R 64 186.37	R 192 322.25	None	N/A
106	R 144 776.60	R 106 447.19	R 234 583.07	None	N/A
107	R 68 143.55	R 56 376.68	R 184 512.56	None	N/A
108	R 43 178.14	R 32 873.68	R 161 009.56	None	N/A
109	R 213 333.33	R 86 783.61	R 214 919.49	None	N/A

Table M: Prediction Interval Results and Remarks for Model Out-of-Range Data

Identifier	SGTV	Lower Bound	Upper Bound	Remark	Margin
1	R 492 127.11	R 820 732.37	R 948 868.25	Under Paid	40.04%
17	R 253 882.67	R 110 526.21	R 238 662.09	Over Paid	6.38%
19	R 1 053 442.97	R 1 068 007.78	R 1 196 143.66	Under Paid	1.36%
20	R 494 974.67	R 117 952.51	R 246 088.39	Over Paid	101.14%
28	R 1 332 921.21	R 325 878.80	R 454 014.68	Over Paid	193.59%
30	R 1 220 018.60	R 153 004.43	R 281 140.31	Over Paid	333.95%
31	R 335 395.20	R 125 830.98	R 253 966.85	Over Paid	32.06%
32	R 878 170.67	R 168 786.07	R 296 921.95	Over Paid	195.76%
34	R 1 869 905.13	R 176 902.80	R 305 038.68	Over Paid	513.01%
41	R 676 171.64	R 92 320.51	R 220 456.39	Over Paid	206.71%
50	R 529 704.67	R 79 346.10	R 207 481.98	Over Paid	155.30%
64	R 260 226.60	R 80 643.82	R 208 779.70	Over Paid	24.64%
65	R 1 739 572.85	R 1 021 197.04	R 1 149 332.92	Over Paid	51.36%
69	R 367 193.72	R 246 261.24	R 374 397.12	None	N/A
72	R 273 786.60	R 59 077.57	R 187 213.45	Over Paid	46.24%
76	R 252 686.65	R 128 291.39	R 256 427.27	None	N/A
83	R 422 935.25	R 124 791.03	R 252 926.91	Over Paid	67.22%
85	R 539 722.62	R 159 524.43	R 287 660.31	Over Paid	87.62%
91	R 234 610.14	R 99 584.03	R 227 719.91	Over Paid	3.03%
92	R 251 107.75	R 100 904.94	R 229 040.82	Over Paid	9.63%
96	R 478 354.34	R 77 705.56	R 205 841.44	Over Paid	132.39%