DEVELOPMENT OF LIBYAN ENERGY GENERATION MODEL USING ROBUST PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODEL (RPLS-SEM) THROUGH WINSORIZATION

FADYA RAMADAN M. SHAKHIM

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2019

DEVELOPMENT OF LIBYAN ENERGY GENERATION MODEL USING ROBUST PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODEL (RPLS-SEM) THROUGH WINSORIZATION

FADYA RAMADAN M. SHAKHIM

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE AND MATHEMATICS
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2019

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (√):

Project Paper
Masters by Research
Master by Mixed Mode
PhD

INSTITUTE OF GRADUATE STUDIES
DECLARATION OF ORIGINAL WORK

This declaration is made on the **12** day of **NOVEMBER** 20**19**.

i. Student's Declaration:

I, **FADYA RAMADAN M. SHAKHIM, P20161000978, FACULTYSCIENCE AND MATHEMATICS** hereby declare that the work entitled **DEVELOPMENT OF LIBYAN ENERGY GENERATION MODEL USING ROBUST PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODEL (RPLS-SEM) THROUGH WINSORIZATION** is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Signature of the student

ii. Supervisor's Declaration:

I, **ZULKIFLEY MOHAMED** hereby certifies that the work entitled **DEVELOPMENT OF LIBYAN ENERGY GENERATION MODEL USING ROBUST PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODEL (RPLS-SEM) THROUGH WINSORIZATION** was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a full fulfillment for the conferment of **DOCTOR OF PHILOSOPHY**, and the aforementioned work, to the best of my knowledge, is the said student's work.

Date

Signature of the Supervisor

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

UPSI/IPS-3/BO 31
Pind : 00 m/s: 1/1

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / Title: DEVELOPMENT OF LIBYAN ENERGY GENERATION MODEL USING ROBUST PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODEL (RPLS-SEM) THROUGH WINSORIZATION

No. Matrik /*Matric's No.*: P20161000978

Saya / *I* : FADYA RAMADAN M. SHAKHIM

mengaku membenarkan [Tesis/Disertasi/Laporan-Kertas-Projek] (Doktor-Falsafah/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-
acknowledged that my [Thesis/Dissertation/Project-Paper] is kept at Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) and reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
The thesis is the property of Universiti Pendidikan Sultan Idris

2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan sahaja.
Tuanku Bainun Library has the right to make copies for the purpose of research only.

3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
The Library has the right to make copies of the thesis for academic exchange.

4. Perpustakaan tidak dibenarkan membuat penjualan salinan Tesis/Disertasi ini bagi kategori **TIDAK TERHAD**.
The Library is not allowed to make any profit for 'Open Access' Thesis/Dissertation.

5. Sila tandakan (√) bagi pilihan kategori di bawah / *Please tick (√) for category below:-*

☐	**SULIT/*CONFIDENTIA***	Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*
☐	**TERHAD/*RESTRICTE***	Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restircted information as specified by the organization where research was done.*
☐	**TIDAK TERHAD / *OPEN ACCESS***	

(Tandatangan Pelajar/ Signature)

Tarikh: _________________

(Tandatangan Penyelia / *Signature of Supervisor*)
(Nama & Cop Rasmi / *Name & Official Stamp*)

Catatan: Jika Tesis/Disertasi ini **SULIT** @ **TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

ACKNOWLEDGMENTS

First of all, I thank Allah the almighty without whose mercy this work would not have been achieved. I would like to thank my supervisor Associate Prof Dr.Zulkifley bin Mohamedfor all the valuable comments, guidance, advice and support that provided a central role in the production of this study. Also, I would like to thank the thesis examiners for their valuable suggestions and comments. My gratitude is to my family, my children and my friend Amal for their continuous prayers and encouragement. I am particularly indebted to my father and my mother for their veryhelpful prayers and support throughout the period of this study. Words fail me to express my appreciation to my husband, whose dedication, love and persistent confidence in me, has taken the load off my shoulder. I owe him for being unselfishly let his intelligence, passions, and ambitions collide with mine. My thanks to my country Libya for helping me to get financial support for my study, with thanks to the management and employees at the Al-Zawya steam power plant for their cooperation and their participation especially engineer NaserShakhim. I am also grateful to all members of staff at Universiti Pendidikan Sultan Idris, Tanjong Malim, Malaysia.

ABSTRACT

This research aimed at developing Libyan energy generation models. Specifically, the research focused on the problems of outliers, non-normal and mullticollinearity. The presence of multivariate outliers was determined using Mahalanobis distance (MD) method. The non-normal of the data distribution was tested using a Mardia's test of multivariate skewness and kurtosis. While the multicollinearity of the data were determined using two methods, namely (i) correlation coefficient, and (ii) Variance Inflation Factor (VIF). The robust partial least squares-structural equation modelling (RPLS-SEM) was utilized to develop Libyan energy generation model.The research proposed the use of an alternative robust measure of central tendency, namely the Adaptive Winsorized Mean (AWM), which used winsorization for determining the data distribution as symmetrical or asymmetrical. The results from RPLS-SEM analysis revealed that each block in the developed model consisted of strong latent variables. The results also showed that the standard measures of reliability, validity and model fit represented their respective latent constructs adequately. In addition, an overall evaluation of the structural model through Goodness of Fit (GoF) index indicated that the model fits the data well. The performance comparison between conventional Partial Least Squares-Structural Equation Model (PLS-SEM) and RPLS-SEM showed that the RPLS-SEM was more effective than the PLS-SEM. In conclusion, the research has successfully developed and evaluated the Libyan energy generation model using RPLS-SEM in the presence of outliers, non-normal and multicollinearity of data. In implication, the new proposed robust model can sustain effectively and also helps to achieve robust predictions.

PEMBINAAN MODEL PENJANAAN TENAGA LIBYA MENGGUNAKAN MODEL PERSAMAAN BERSTRUKTUR-KAEDAH KUASA DUA KECIL SEPARA TEGUH (MPB-KTST) MELALUI KAEDAH WINSORIZATION

ABSTRAK

Penyelidikan ini bertujuan membina model penjanaan tenaga Libya. Secara khusus, penyelidikan ini memberi tumpuan kepada masalah data pencilan, ketaknormalan dan multikolinearan. Kewujudan data pencilan multivariat ditentukan dengan kaedah Jarak Mahalanobis (JM). Taburan ketaknormalan data diuji menggunakan ujian kepecongan dan kurtosis Mardia. Manakala data multikolinearan ditentukan menggunakan dua kaedah, iaitu (i) pekali korelasi, dan (ii) Faktor Inflasi Varians (FIV). Model Persamaan Berstruktur-Kuasa dua Terkecil Separa Teguh (MPB-KTST) digunakan untuk membina model penjanaan tenaga Libya. Penyelidikan ini mencadangkan penggunaan ukuran kecenderungan memusat teguh alternatif, iaitu Min Penyuaian *Winsorized* (MPW), yangmenggunakan*winsorized* dalam menentukantaburan data secara simetrikal atau tak-simetrikal. Dapatan daripada analisis MPB-KTST menunjukkan bahawa setiap blok dalam model yang dibina mengandungi pemboleh ubah pendam yang kukuh. Dapatan juga menunjukkan bahawa ukuran piawai kebolehpercayaan, kesahan dan penyuaian model mewakili secukupnya konstruk pendam. Sebagai tambahan, penilaian keseluruhan model berstruktur melalui indek Kebagusan Penyuaian Model (KPM) menunjukan bahawa model disuaikan baik dengan data. Perbandingan prestasi antara Model Persamaan Berstruktur - Kuasa dua Terkecil Separa (MPB-KTS) konvensional dan MPB-KTST menunjukkan bahawa MPB-KTST adalah lebih efektif berbanding MPB-KTS. Kesimpulannya, penyelidikan berjaya membina dan menilai model penjanaan tenaga Libya menggunakan MPB-KTST dengan kehadiran data pencilan, ketaknormalan dan multikolinearan. Implikasinya, model teguh baharu yang dicadang dapat dikekalkan efektif dan membantu bagi memperoleh angggaran teguh.

TABLE OF CONTENTS

CHAPTER 2 LITERATURE REVIEW

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AVE	Average Variance Extracted
AWM	Adaptive Winsorized Mean
CA	Chemical Additive
CB-SEM	Covariance-Based SEM approach
CFA	Confirmatory Factor Analysis
DVs	Dependent Variables
DW	Desalination Water
GECOL	General Electricity Company of Libya
GLM	General Linear Model
GOF	Goodness of Fit
GUI	Graphical User Interface
JKW	Joreskog-Keesling-Wiley
LISREL	Linear Structural Relations
LVs	Latent Variables
MD	Mahalanobis Distance
MLE	Maximum Likelihood Estimator
MW	Megawatt
MSE	Mean-Squared Error
MVs	Manifest Variables
MVT	Multivariate Trimming
NOC	National Oil Company

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

OLS	Ordinary Least Squares
OP	Maintenance and Operation
Output	Electricity and Fresh Water
PLS	Partial Least Squares
PLS-SEM	Partial Least Square- Structural Equation Models
PM	Path Modelling
PROC	Procedure
R^2	Coefficient of Determination
RSIMPLS	Robust Straightforward Implementation Partial Least Squares
RPLS-SEM	Robust Partial Least Square- Structural Equation Models
SAS	Statistical Analysis System
SD	Standard Deviation
SEM	Structural Equation Modelling
SPP	Steam Power Plant
SPSS	Statistical Package for Social Sciences
SSE	Sum of Squared Errors
VB-SEM	Variance-BasedSEM
VIF	Variance Inflation Factor
WIN	Winsorization
WM	WinsorizedMean

CHAPTER 1

INTRODUCTION

1.1 Introduction

The Structural Equation Modelling (SEM) method consists of a family of various statistical techniques (a secondgeneration multivariate data analysis process) that can predict, describe, or test the correlation between the latent constructs and the measured variables (Makina 2016). SEM is a more effective statistical technique for the following (Hair et al.,2016):

i. Analysing the path analysis with multiple dependents.
ii. Analysing regressions with multicollinearity problem.
iii. Estimating the correlation and covariance in a model.
iv. Modelling the inter-relationships between variables in a model.

Two different processes can be applied for estimating the relationships in the SEM process. The most popular techniques include the Covariance-Based SEM (CB-SEM), which is used for approving (or rejecting) all theories (Hair et al., 2016), along with the Variance-Based SEM (VB-SEM) process which helps in the development of theories in the exploratory research.Table 1.1below illustrates the differences between CB-SEM and PLS-SEM (Roldán & Sánchez-Franco, 2012).

Table 1.1

Comparison of the Variance-based and Covariance-based SEM techniques

SEM Approach	**VB-SEM**	**CB-SEM** **AMOS**
Estimator Used	Partial Least Squares Estimator (PLS)	Maximum Likelihood Estimator (MLE)
Research Objective	• Exploratory in nature • Predicting • Theory development	• Confirmatory in nature • Theory testing by comparing all theories
Statistical Approach	Based on a non-parametric approach	Based on a parametric approach
Model Assessment	• Model predictiveness • Coefficient of determination, R^2	• Overall (absolute) fit measures

(continue)

Table 1.1 (*continued*)

	• Predictive relevance Q^2 and Average Variance Extracted, (AVE) • Stability of estimates, application of the resampling processes (Jack-knifing and Bootstrapping)	• Comparative (incremental) fit measures • Model parsimony
Software	SmartPLS	LISREL, AMOS and EQS
Type of fitting algorithm	Multi-stage iterative process using OLS.A subset of parameters estimated separately	simultaneous estimation of parameters by minimising the discrepancies between the observed and predicted Covariance/correlation matrix

The Partial Least Square (PLS) technique is also known as the component-based SEM, composite-based SEM or the variance-based SEM. PLS was seen to be a very effective process when the research aimed to make predictions or investigate a modelling method. On the other hand, the CB-SEM is applied when the research aim to develop a confirmatory modelling process (Garson, 2016).

The PLS technique or the PLS-path modelling process was developed by Herman Wold (Wold 1975; 1982; 1985). Thereafter, it became very popular and has been applied in various research areas like marketing (Albers, 2009), education (Campbell & Yates, 2011), and in the field of social science (Jacobs et al., 2011). Lohmöller (1989) for a mathematical presentation of the path modelling variant of PLS, which compares PLS with OLS regression, principal components factor analysis, canonical correlation, and structural equation modeling with LISREL.

PLS could be used as a regression model that can forecast multiple dependents from various sets of independent parameters. It can be applied as the path model, which includes causal paths that connect the predictors, along with the paths which connect the predictors and the response variable(s) (Garson, 2016). Furthermore, PLS is implemented as a regression model by SPSS and by SAS's PROC PLS. SmartPLS is the most prevalent implementation as a path model.

The path models consist of two elements, i.e., a structural model (called the inner model in the PLS-SEM), which describes the relationship between the latent variables and the measurement models (called the outer model in the PLS-SEM) that describe the relationship between the latent variables and their measures (or indicators). two types of measurement models exist: Type 1 is used for exogenous latent variables (or constructs that describe other model constructs) and Type 2 is used for endogenous latent variables (or constructs that are described in the model) (Joseph et al., 2014). When the causal arrow in the path diagram arises from the latent variables (factors) to the measured variables, the path model is known as reflective. However, if the arrow is present between the observed measures and latent variables, the path model is seen to be formative. The reflective models are often referred to as the “Mode A” models while the formative models are called the “Mode B” models (Garson, 2016).

The PLS-SEM technique is especially effective for a small sample size and for complex models (which include many constructs and indicators) and makes no assumptions regarding the underlying data, i.e., data need not be normally distributed. The PLS-SEM technique can handle the formative and the reflective measurement models, along with the single item constructs, without affecting their identification efficiency (Hair et al., 2016). Hence, it is used for various research applications.

Some advantages of using the PLS model include its capability to model multiple dependent/ independent variables and manage the multicollinearity existing among its independent variables; and ability to deal with the missing data or the noise in the data. The PLS method can also directly develop independent latent variables based on the cross-products related to all dependent variables (Garson, 2016), which helps it make accurate predictions.

Several robust methods can be used for decreasing or eliminating the effect of the outlying data points (Gonzaleza et al., 2009). In this research, a robust technique known as winsorization was proposed, where the outliers were penalised and retained within the data, without eliminating them or restricting their effect on the estimators’ variance. Winsorization was seen to be a simple, robust and effective process(Favre-Martinoz, Haziza, & Beaumont, 2015), which can minimise the outlier effect. Unlike the transformation, the winsorization method makes changes at the distribution tail. In the past few years, winsorization has become very popular as it decreases the effect of the outliers without altering the sample sizes. The research investigated the published literature and stated that none of the studies have solved the outlier-related issue using the winsorization with the PLS-SEM process.

PLS-SEM does not require normally-distributed data and can handle a small sample size (Hair et al., 2014). It effectively handles inadequate data, namely (i) a skewed distribution of all variables; (ii) multicollinearity within the LVs or blocks of variables; and (iii) in an incorrectly specified structural model like the exclusion of regressor(s) (Roldan & Sanchez-Franco, 2012). The PLS-SEM technique was considered to be the best process as the research assign a score for the LVs while analysing the predictive appropriateness. PLS-SEM was also considered to be an effective technique which could be used in studies that were based on the models with structural paths and novel measures and was seen to support the prediction-related study objectives (Chin, 2010; Hair et al., 2011).

Here, the research developed and assessed the energy generation model, which was based on the Robust PLS-SEM (RPLS-SEM), which integrated PLS-SEM and the winsorization (WIN) method. This model was used for solving three major modeling-related issues namely outliers, non-normal data distribution and multicollinearity. Each of these issues could be handled independently by applying a similar least squares technique for parameter estimation. Due to the limitations noted in the existing software which would be used for addressing the above-mentioned issues, the research developed a novel methodology that could be applied on the real data, derived from the Libyan oil and gas industries. Furthermore, the research compared the novel RPLS-SEM approach with the traditional PLS-SEM approach.

1.2 Problem of Statement

The OLS method is linked to a variety ofproblems. The least-squaresestimates, for example, are particularly sensitive to outliers, and this is particularly true in small samples (Wooldridge, 2009). In addition, multicollinearity is present in economic factors and could have a strong influence on parameter estimation. The degree of multicollinearity will determine the results, and in many instances it does so in a passive manner, for instance, by increasing the variance of OLS and decreasing the reliability of the model and jeopardizing its stationarity (ESalam, 2011). Since the rank of the parameter estimation is close to zero, the diagonals of the covariance matrices will be too large. This indicates that the variance inflation factor (VIF) is infinite. This would finally result in some of the crucial independent variables being omitted, thereby reducing reliability of the model (Shang & Zhang, 2009). The process of extracting different data from the estimates of sample parameters, multicollinearity will cause unstable estimation and the estimates obtained will lack in stationarity (D'Ambra & Sarnacchiaro, 2010).

The PLS-Pathmodelingtechnique assumes that structural models are linear. Hence regression techniques can be used to estimate structural coefficients. Despite of this, OLS regression modeling is the frequently used modeling technique since it does not set any requirement (Mateos-Aparicio, 2011). Therefore, the criterion employed to assess the correctness of the adjustment is the normal measure for this type of models, i.e. the coefficient of determination, R^2, which is the measure between variability explained by the regression model and the total variability. The coefficient can be estimated by utilizing the OLS method. However, PLS regression must be employed either in the structural model or the measurement model when multicollinearity exist among the explanatory variables.

In the outer model (measurement model), the presence of multicollinearity within the factor indicators will raise a problem of high correlations between the indictors for one factor and the indicators of another factor. The degree of multicollinearity determines the extent of a simple factor structure in PLS. Factor cross-loadings will make it difficult to distinguish, interpret, and label PLS factors. With regard to the multicollinearity of indicators within one factor, the research need to examine the contents of the items to ensure that the high correlation is not related to an artefact such as non-substantive word item variations (Garson, 2016). Since PLS factors are orthogonal by technical definition, mathematical multicollinearity will not present not a problem in the PLS regression. This is the main reason why the PLS models are sometimes preferred over SEM or OLS regression models (Garson,2016).

High levels of collinearity among formative indicators is a serious problem since they influence the estimation of weights and on their statistical significance. Specifically, high degree of collinearity influence the analysis results in two ways. Firstly, collinearity increases standard errors and, consequently, reduces the ability of the model to show that the estimated weights differ significantly from zero. This is of a particular problem in PLS-SEM analyses performed on small samples where generally the standard errors are large due to sampling errors. Secondly, high collinearity could result in erroneous estimation of the weights and reversal of their signs (Joseph et al., 2014).

In the analysis of real data, it is not uncommon for some observations to differ from the majority of the observations. These observations are referred to as outliers. They are sometimes the results of mistakes whencopying or recording the data, such as permutation of two digits or misplaced decimal point (Rousseeuw et al., 2006). Even though outliers differ from most of the data they are not necessarily erroneous (Liebmann, 2009). Observations of outliers are typically made under extraordinary conditions. On the contrary, they may belong to other statistical population(s). Classical methods often fail to identify outliers. The generated model could attempt to fit the outlying observations, hence conceal their inaccuracy (masking effect).Certain good data points, on the contrary, could appear as outliers (swamping effect).

Two methods can be employed to identify outlying points (Rousseeuw et al., 2006). The first method is a classical method, where the residuals are then used to compute a number of diagnostics. Outliers typically have a strong effect on classical methods which prevent the fitted model from identifying the observed deviation.

Thisis known as masking effect. In a condition known as swamping several valid data points might be identified as outliers (Chatterjee & Hadi, 2006). Another method of identifying outliers is by utilizing the robust method. Robust statistics seeks to find a comparable fit to one which would have been found without the outliers. The solution obtained through this method makes it possible to determine the outliers by their eresiduals from the robust fit (Huber & Ronchetti, 2009). Given that most classical methods are sensitive to outliers, robust statistics seek to establish methods that are robust against the possibility that one or more unanticipated outliers could be present anywhere in the data (Hubert et al., 2008).

Outliers could have a strong influence on the performance of the PLS method. Hence the characteristics of abnormal data should be taken into account in practical applications. From a statistical perspective, outliers are samples with excessive values that form a different population instead of the larger part of the data. Outliers can be classified into two categories: (i) high leverage points that are located far from the centre of the data within the space of the measurable variables; and (ii) high residual points that are very different from the observed values and their corresponding predictions within the space of the product quality variables (Yin et al.,2014).

Outliers could have adverse impacts on statistical analyses. They increase error variance and reduce the power of statistical tests. Outliers that are non-randomly distributed could reduce normality (and, in multivariate analyses, they violate the assumptions of sphericity and multivariate normality), thereby alter the likelihood of making both Type I and Type II errors. Outliers could also exert preference or influence estimates that that could be of fundamental interest (Osborne & Overbay, 2004).

There are however two conditions which enable PLS-SEM to provide the ultimate solution for models by using nonnormal data. The research should firstly recognize that highly skewed data might decrease the statistical power of an analysis. Specifically, the significance of the model parameters is contingent upon the standard errors from bootstrapping, which could be overinflated when handling data with highly skewed data (Hair et al., 2014a). Secondly, CB-SEM has several alternative estimation procedures, which makes it difficult to assume that PLS-SEM is the natural choice when dealing with data distribution (Hair et al., 2012). Even though PLS-SEM does not require for data to be normally distributed, the occurrence of excessive outliers will result in non-normality. Non-normality is usually linked to the presence of outliers.

This thesis demonstrates the development of Libyan energy generation model using RPLS-SEM for estimating parameters. It attempts to solve three fundamental problems in modeling: outliers, non-normal, and multicollinearity. Each problem is dealt with separately. Robustmethod such as winsorization is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers. Based on winsorization adaptive Winsorized mean achieves to handle non-normal distribution. In addition, Partial least squares structural equation modeling (PLS-SEM) is one of the effective ways of dealing with the multicollinearity problem Due to the specific constraints in the software accessible to the research in dealing with these problems, the development of the methodology will be demonstrated using real data pertaining to the generation of electricity derived from the oil and gas industry in Libya. The research also compares the performance of the

RPLS-SEM method with that of the most widely used method, the partial least squares method.

1.3 Research Questions

The research questions addressed in this research are as follows:

i. How does the Libyan energy generation model can be developed using PLS-SEM?
ii. Which effective methods can handle the general issues related to the PLS-SEM?
iii. What are the effective methods for solving the common problems associated with the PLS-SEM?

iv. What are the differences in the performance of the proposed model and the existing PLS-SEM on the development of the Libyan energy generation model?

1.4 Research Objectives

This research aims to:

i. propose a Libyan energy generation model with the help of the SEM.
ii. modify a robust PLS-SEM model using the winsorization method
iii. evaluate the ability of the proposed robust model in resolving the issues of outliers, non-normality, and multicollinearity.

iv. evaluate and compare the performances of the proposed method with the existing PLS-SEM process on the Libyan energy generation model.

1.5 Conceptual Framework

The research developed a conceptual framework (Figure 1.1) using different concepts, models, and source materials based on a conceptual framework theory (Ethridge, 2004). They used the PLS-SEM technique for specifying the model and determining the method for measuring the constructs, collecting all data and further processing the data using the SmartPLS3 software. The software could easily fit the data to a specific model and generate the results. The data processing output included the parameter estimates and the general model fit statistics. The research used actual data regarding the energy generation that was compiled by the employees of the Technical Department, Al-Zawiya Steam Power, and Libya. Data, extending to 2016 (sample size is 65 days), was used in this study.

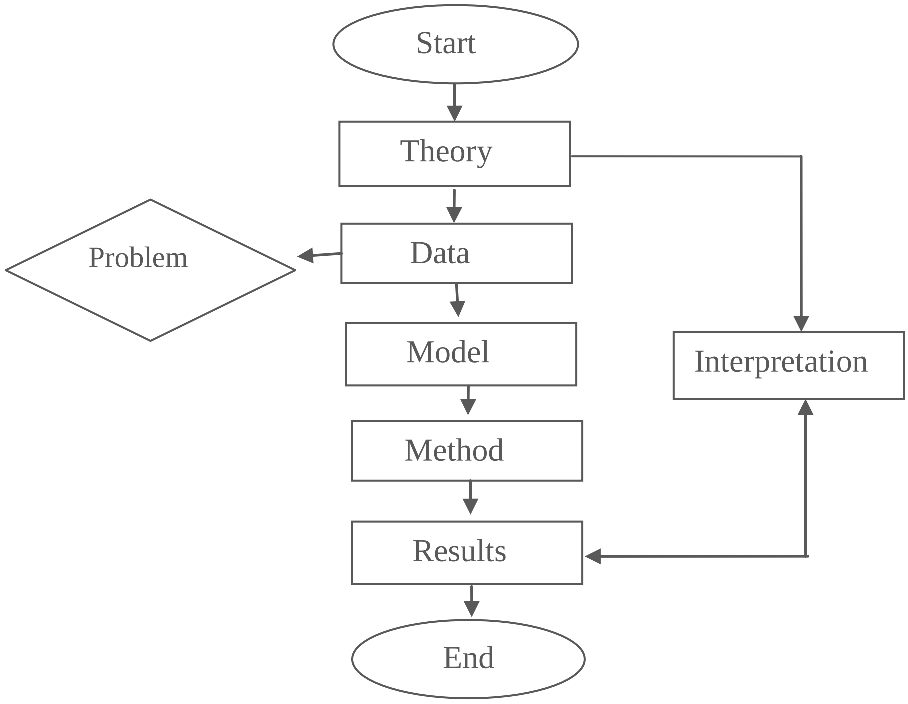

Figure 1.1 The Conceptual Framework of the Study

1.6 Libyan Energy Generation Model and Data Collection

In this research the data were collected from the Al-Zawiya Steam Power plant in Libya. Real data related to energy generation was collected and compiled by the Technical Department of the AL-Zawiya Oil Refining Company. This secondary data was compiled in year 2016. Some data was also collected from other sources like the government reports on the Oil Surveys in Libya, the Energy Research Centre and the General Directorate of Power Plant.

1.6.1 Libyan Oil and Gas Industry

Libya has a very large oil reserve, and the people are severely dependent on oil and gas industries, which generate higher than 90% of the total income. These industries offer facilities throughout the country, both off-shore and on-shore, but they require a large quantity of fresh water. These facilities include the oil and gas fields, gas processing plants, chemical complexes, pipelines and terminal ports for the oil and gas export. Many desalination plants are also constructed in the industrial region for ensuring an uninterrupted water supply, which can fulfil the industrial and domestic requirements of the people.

The oil industries in Libya are managed by the National Oil Company (NOC), which controls the local and international oil organisations and companies. The desalination plants in the country were developed and used by the oil and gas industries. One of the first desalination plants was constructed in Mersa Al-Brega in 1964, and it can generate 757 m^3/day of water. The desalination plants developed by the oil industries constitute 15% of the total number of plants in the country. Many desalination processes are applied in these plants (Al-Hengar et al., 2007).

The Libyan electric grid consisted of a high voltage network of 12,000 km, a medium voltage network of 12,500 km and a low voltage network of 7,000 km. This electric power sector was developed in the past couple of decades. The maximum load increased from 5,759 MW (2010) to 5,981 MW (2012), with an average growth rate of 3.85%. The electric energy generated also increased from 32,558 GWh (2010) to 33,980 GWh (2012), at an average growth rate of 4.37%. This value reached 48,497 GWh (2017). A majority of the existing Libyan power plants were converted from oil to natural gas, and many new power plants were developed which utilised natural gas.

These natural-gas based plants displayed a higher thermal efficiency, lower emission characteristics and were seen to be more economical than the oil-based power plants. Out of the total electric energy generated in Libya in 2012, 53% was due to standalone gas turbine units, 37% was due to the combined cycles while 10% were from the steam turbine units (Fellah & Noba, 2016).

1.6.2 Al-Zawiya Steam Power Plant

The Al-Zawiya Steam power plant was seen to be one of the largest power plants in Libya and was located near the Zawiya city, and 50 km in the west of Tripoli. This plant generated a total electric power of 1,440MW. The officials conducted an energy analysis for Unit three of the Al-Zawiya combined power plant. Unit three comprised of two units of gas turbines, two heat recovery steam generators and one steam turbine unit. It has a maximal capacity of 450MW. Unit one was commissioned in 2003, while the last unit was developed in 2007. This plant was operated by the General Electricity Company of Libya (GECOL).

1.6.3 Variables Included in the Libyan Energy Generation Model

In this research five latent constructs and twelve indicators (continuous variables), fourEndogenous latent variables (Output items, Input data, Steam Power Plant and Chemical Additives) and one exogenous latent variables (Maintenance and Operation) were included in the development of the model:

i. Output items: Electricity (megawatt) and freshwater (Cubic meters)

ii. Input data: The major input variables included - Desalination (DW) data; total steam generated (tons/day) and seawater (Cubic meters/ day) needed for producing freshwater.

iii. Steam Power Plant (SPP) requirements - steam turbine (tons/day) and boiler (Cubic meters/ day of distilled water)

iv. Chemical Additives (CA) -Morphine (Liter/day), Phosphate (Kilograms/day), anti-scale (Liter/day) and hydrazine (Liter/day).

v. Maintenance and OPeration (OP) – mean costs for the chemical treatment and fuel (Libyan dinar/day).

1.6.3.1Generation of Electricity at the Steam Power Plants

The Libyan electric grid consists of a 120,000 km high voltage, a 12,500 km medium voltage, and a 7,000 km low voltage network (Ramelli et al., 2006). This electric sector was developed in the past one-two decades ago. The total peak load increased from 5,759 MW (2010) to 5,981 MW (2012), which indicated a mean annual increase of 3.85%. The total electricity generated increased to 33,980 GWh (2012) from 32,558 GWh (2010), which represented an annual mean growth rate of 4.37%. This value is expected to increase to 48,497 GWh by 2020 (Fellah and Noba, 2016).

1.6.3.2Steam Power Plants

Electricity generation at the steam electric plants needs a steam turbine/electrical generator; steam generator (i.e., boilers) and condenser where the combustible fuel can be used as the primary energy source for steam generation. One of Libya's biggest electric plants is located at Al-Zawiya City, which was 50 km west of Tripoli. This was the type Gas Power Plant and had a 1440 MWe design capacity. This plant was GECOL-operated and consisted of nine units, where the Unit one began operating in 2003; while the final unit became operational in 2007. Unit three consisted of two gas turbine units; two heat recovery steam generators; and one steam turbine unit. The total capacity of this plant was 450 MW (Fellah and Ben Noba, 2016).

1.6.3.3Desalination Units

The water derived from the surface or groundwater sources contains many scale-forming elements, oxygen radicals, or acids. These chemicals must be removed since the efficient operation of a steam cycle needed a very high-water quality. A good water quality increases the efficiency of the plants by decreasing the scale deposition on the steam tubes, thereby reducing the general maintenance, and increasing the system availability. This can further decrease the total costs and increase revenue (Woodruff et al., 2005).

1.6.3.4Steam Generators

A steam generator or a boiler refers to a closed container where the pressurised water can be converted into steam after heat application. The open containers which generate steam at the atmospheric pressure cannot be called boilers. Chemical energy present in the fuel, contained in the furnace, gets converted to heat. This heat is then converted to water by the boiler very efficiently. Boilers can generate steam at pressure values higher than that at the atmospheric temperatures by absorbing all the heat which is generated during the fuel combustion (Woodruff et al., 2005).

1.6.3.5Chemical Additives to the Water

The waters in a boiler must be chemically treated regardless of the fact that the water was pre-treated. Thus, the external water treatment can be supplemented by the internal treatment, thereby ensuring that no impurities are passed through the water system into the boilers (like hard water, oxygen radicals, silica, iron, etc.). The chemical additives using for water treatment include morphine, hydrazine, sodium triphosphate, and anti-scale.

1.6.3.6Use of Steam plants

A majority of the industrial processes produce steam. It is considered one of the most significant power sources for electricity generation. Steam is very useful as it can be generated anywhere by using heat from various fuel sources available in the area. It has many properties which help in energy production. It can be recycled, converted to water or back to steam, without adversely affecting the environment. The existing steam plants are seen to be a complex arrangement of engineering system which can efficiently and economically produce steam.

1.6.3.7 Steam Turbines

Steam turbines refer to a turbo-machine, which consists of inter-connected blades and nozzles. It transforms the energy from the high-pressure and high-temperature steam to mechanical energy (or shaft work). They are mainly used for marine propulsion. Operation of a steam turbine is based on the dynamic activity of the steam expanding within the nozzles. The capacity of the steam turbines, used for generating electricity in the steam power plants, ranges between one and 1500 MW.

1.6.3.8Operational and Maintenance-Related Costs

The primary capital cost, along with the Operational and Maintenance costs are the major factors that must be considered from an economic viewpoint. The operational costs for the power plants include the fuel and the water treatment costs (including the chemical additive costs). As the fuel cost is very high, a primary fuel must be selected, along with the equipment that can ensure the overall efficiency of the plant and decrease the environmental effect. The industrial-scale thermal desalination process includes evaporation, which is an expensive and energy-intensive process (Breeze, 2008).

1.6.4 Libyan Energy Generation Model

In this section, the research described the process they applied for testing the relationship between the input and output energy generation model used in the Al-Zawiya Steam Power Plant. The research developed the estimation parameters with the help of the RPLS-SEM technique and the winsorization method for handling the non-normality, outliers and multicollinearity. Anderson and Gerbing (1988) suggested the application of a two-step model building technique which could analyse twotheoretically different models, i.e., the measurement or the structural models. The measurement model described the relationship between the measured (or observed) variables, occurring due to the LVs, while the structural model highlighted the relationship between the latent variables as stipulated by theory. The Smart PLS software was generally used for such models (Ringle et al., 2015).In this research, the measurement model consisted of five external models, i.e., the *DW* or desalination units (steam and seawater); Steam Power Plants (*SPP*) including the steam and boiler turbines; Chemical Additives (*CA*) like sodium triphosphate, hydrazine, morphine, and antiscale; Maintenance and Operation (*OP*) like chemical treatment or fuel-related costs; and freshwater or electricity was the final Output. On the other hand, the internal model consisted of five structural models, namely a *DW* structural model (η_1), *SPP* structural model (η_2) ,*CA* structural model (η_3), *OP* structural model (ξ_1), and an *Output* structural model (η_4).

Furthermore, the RPL-SEM was developed according to the recommendations suggested earlier (Piacentino 2004; Breeze 2008; Wu et al., 2013; Thomas et al., 2016). The research described the relationship between the *Output* and the *DW*, *SPP*, *CA* and *OP* parameters. The technique that was used for the proposed model was further described by applying the model to the output factors like fresh water and electricity in the Libyan oil and gas industry. The additive model described below was included in the theoretical study:

$$\eta_4 = \Lambda_0 + \Lambda_{14}\eta_1 + \Lambda_{24}\eta_2 + \Lambda_{34}\eta_3 + \Omega_{14}\xi_1 + \zeta_4$$

(1.1)

Where; Λ_0 refers to the general factor efficiency parameter that can be used for the composite primary factor inputs; while the Λ_{14}, Λ_{24}, Λ_{34}, Ω_{14}, parameters refer to the electricity generation elasticities; ξ_4 was an error term.

1.7 Scope of this Research

The Libyan oil and gas industry contributed to its economic development. The economic planners in the country have highlighted the development of the energy-related industries from an early economic planning stage. In this thesis, the research has proposed and developed a new model which could assess the energy generation in Libya and included the energy resource externalities.

Proposed model would be applied to the input secondary datawhich were important for the energy generation in Libya. This research proposed using a two-step strategy for formulating the model using two theoretically differing models, i.e., measurement and structural model. The measurement model described the latent correlation between the measured (or observed) variables; while the structural model describes the relationship between the LVs, as postulated by theory.

1.8 Significance of this Research

PLS-SEM is a very popular technique that is used for estimating the factors which were not noticeable. In this technique, the factors were called the LVs, and all indicators of the LV can be measured. The major objective of the PLS-SEM was to predict the variance between the endogenous constructs and the MVs of the constructs. The models consisting of some significant Jack-knife or Bootstrapparameter estimates were considered as null, based on a predictive perspective.

In this research, the used of a classification process, based on the predictions in the PLS-SEM, for estimating the SEM parameters were proposed. The researchstated that the PLS estimates could generate a very accurate measure for the SEM path coefficients compared to the OLS estimates. The PLS regression and Path Modelling (PM) approaches were seen to be novel techniques that have not been applied for aenergy generation model. Hence, no research used the PLS-SEM technique for making energy generation-related estimates at the Libyan oil and gas industries. This research developed a RPLS-SEM technique for the Libyan power generation plant, based on the data collected from the Al-Zawiya Steam Power Plant.

This researchoffered a better estimation of the OLS. It also addressed the issues related to the outliers, non-normal data and multicollinearity. To the best knowledge of the researcher, no solutions were offered for handling these issues simultaneously. This motivated the researcher to develop an optimal process for making reasonable predictions related to the oil and gas industries in Libya, by applying the RPLS-SEM technique. The results developed using this model could be effectively applied to the oil and gas industries in other countries. The real data consisted of many outliers. Hence, the research have summarised the various breakdown estimation methods, using the winsorization technique. These approaches can mask the outlier clusters and can identify the inliers and the outliers.

1.9 Organisation of the Thesis

This thesis consisted of six chapters. Chapter onepresentsthe background of the PLS-SEM method. It also addressed the issues related to outliers and non-normal data distribution, though the research focused on the multicollinearity in the PLS-SEM. Sections 1.2 and 1.3 described the problem statement and research objectives, respectively, whereas Section 1.4 represented the conceptual framework and describes the Libyan oil and gas industry.

Chapter twopresentsthe literature review of all relevant studies. Section 2 describes the issues of non-normal, outliers, and multicollinearity along with the PLS-SEM method. Section 2.1 presents the robust statistics and winsorization technique; while Section 2.2 describes the PLS-SEM and the OLS-SEM methods and compared the two. Section 2.3 presents the issue of multicollinearity and used the Variance Inflation Factor (VIF) as the major indicator of the multicollinearity. Section 2.5 reviews the PLS-SEM method, while Section 2.7 summarises the whole chapter.

Chapter three proposed the research methodology used in the thesis. This chapter presents the techniques used in this research and highlighted the research objectives and the nature of all problems and methods used (as mentioned in Chapter 1). Section 3.2 describes the sources used for collecting the data and presented the data collection techniques. Section 3.3 discusses all variables used in the research, namely steam electric plants, electric power generation, steam generator, desalination units, chemical-based water treatment, steam turbines and maintenance and operations. Section 3.4 describes the research framework and the overall parameter estimation. Sections 3.7-3.11 define the research strategies and modelling processes; while Sections 3.12-3.13 present the data analysis, model structures and model evaluation techniques.

Chapter four describes the results of data analysis. Section 4.1 focuses on the model development. This chapter also describes the inner and outer models, model development, and modelling results.

Chapter five outlines all results related to this research. Section 5.2 presents the methods used for identifying the outliers, non-normallity and the multicollinearity. Section 5.3 presents the results of the RPLS-SEM model, while Section 5.4 highlights the RPLS-SEM analysis, assessment of the measurement and structural models.

Section 5.5 compares the PLS-SEM and the RPLS-SEM models. Section 5.7 describes the novel RPLS-SEM.

Chapter six summarises the research conclusions. It highlights the contributions made by the research and offered suggestions for future work.