ESTIMATION OF COBB-DOUGLAS PRODUCTION FUNCTION PARAMETER THROUGH A ROBUST PARTIAL LEAST SQUARES

MARYOUMA ELAKDER ENAAMI

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2012

ESTIMATION OF COBB-DOUGLAS PRODUCTION FUNCTION PARAMETER THROUGH A ROBUST PARTIAL LEAST SQUARES

MARYOUMA ELAKDER ENAAMI

THESIS SUBMITTED TO FULFILL THE REQUIREMENTS TO OBTAIN Ph.D OF SCIENCE

FACULTY OF SCIENCE AND MATHEMATICS
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2012

DECLARATION

I hereby declare that the work in this dissertation is own except for quotation and summaries which have duly acknowledged.

Date

Signature

Name

Registration no.

Acknowledgement

There are numerous people to whom I am indebted for the completion of this work. Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Zulkifley Mohamed for his continuous support, patience, motivation, enthusiasm and immense knowledge which enabled me to successfully complete my Ph.D study and research. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would also like to take this opportunity to thank Dr. Sazelli Abdul Ghani, my thesis committee for his encouragement, insightful comments, and enlightening questions.

Last but not the least, I would like to thank my beloved husband, Salem Mohamed Alkhabuli, and my children, Salwa and Mohamed, for their continuous love, motivation and support as I worked on this thesis and towards the completion of my Ph.D. Their constant encouragement and quiet sacrifice were a blessing to me. In addition, I would like to thank my parents for their undying support as my life goes in this new and intriguing direction. Their unconditional love has always been my strength. I am also thankful to my brothers and sisters who continue to challenge me and keep the spirit of determination burning inside me.

Thank you.

ABSTRACT

This research aimed to develop a new model for estimation problems of Cobb-Douglas production function. Specifically, the researcher focused on the problems of outliers and multicollinearity, and on solutions that were proposed to deal with these issues simultaneously. The Cobb-Douglas production function was usually fitted by first linearizing the models through logarithmic transformation and then applying the method of least squares. However the log transformation process did not get rid of all the outliers and multicollinearity problems. These problems appeared to have wide repercussions. Yet, these problems had not been known for their written solutions, and no serious research had been conducted in this field. The researcher attempted to develop the best possible method for providing estimates in the context of the Libyan agricultural sector by using the robust partial least squares path modeling (RPLS-PM). The results from RPLS-PM analysis showed that each block consisted of strong latent variables and one or more indicator variables. Furthermore, the results revealed that the standard measures of reliability, validity and model fit represented their respective latent constructs adequately. In addition, an overall assessment of the structural model was carried out based on the goodness of fit (GoF) index. The GoF value obtained indicated that the model fit the data. In other words, the findings from the model correspond to very good results and provide important new insights. This supports the thesis that this approach has the potential for solving problems in the agricultural sector as well as in other sectors. Also, the researcher noted that the model performed better than the least squares method when the multicollinearity and outlier problems existed in the data.

ABSTRAK

Kajian ini bertujuan untuk membangunkan satu model baharu bagi masalah anggaran dalam fungsi pengeluaran Cobb-Douglas. Secara khusus, pengkaji memberi tumpuan kepada masalah data pencilan dan multikolinearan, dan penyelesaian yang dicadang untuk menangani isu-isu ini secara serentak. Fungsi pengeluaran Cobb-Douglas biasanya disuaikan dengan penglinearan awal melalui transformasi logaritma dan kemudian menggunakan kaedah kuasa dua terkecil. Walaubagaimanapun proses transformasi log tidak menyingkirkan semua data pencilan dan masalah multikolinearan. Munculnya masalah-masalah sebegini memberi kesan yang meluas. Namun, tidak terdapat penyelesaian bertulis dalam masalah sebegini, dan tiada penyelidikan yang bersungguh telah dijalankan dalam bidang ini. Penyelidik membangunkan kaedah terbaik bagi melakukan anggaran dalam konteks sektor pertanian Libya dengan menggunakan pemodelan kuasa dua terkecil separa teguh (RPLS-PM). Keputusan daripada analisis RPLS-PM menunjukkan bahawa setiap blok mengandungi pembolehubah-pembolehubah pendam dan satu atau lebih pembolehubah penunjuk yang kukuh. Tambahan, keputusan mendedahkan bahawa ukuran piawai kebolehpercayaan, kebersahan dan penyuaian model mewakili konstruk pendam yang berkaitan secukupnya. Seterusnya, penilaian keseluruhan model berstruktur dilakukan berdasarkan indeks kebagusan penyuaian (GoF). Nilai GoF yang diperoleh menunjukkan model disuaikan dengan data. Dengan kata lain, pemodelan menghasilkan keputusan yang baik dan memberi wawasan baharu yang penting. Ini menyokong tesis bahawa pendekatan yang dilakukan berpotensi bagi menyelesaikan masalah dalam sektor pertanian dan juga sektor-sektor lain. Begitu juga, penyelidik menjelaskan bahawa model adalah lebih baik berbanding kaedah kuasa dua terkecil apabila wujudnya data pencilan dan multkolinearan dalam data.

CONTENTS

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

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

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

CHAPTER 4 MODEL DEVELOPMENT

CHAPTER 5 RESEARCH RESULTS AND FINDINGS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	Agriculture inputs (seeds, chemical fertilizers, and pesticides)
AVE	The average variance extracted
D-D plot	The distance-distance plot
GoF	The Goodness of Fit
HW	Operating hours and wages.
LVs	Latent variable
LW	Land and water
MCD	The minimum covariance determinant
MD	The Mahalanobis Distance
MVs	Manifest variables
OLS	Ordinary least squares
OP	Operation and maintenance costs
Output	Wheat crop output
R^2	The coefficient of determination
PLS	Partial least squares
PLS-PM	Partial least squares path modeling
RPLS	Robust partial least squares
RPLS-PM-CD	Robust partial least squares-path modeling for the Cobb-Douglas production function
SEM	Structural equation models
VIF	The variance inflation factor

CHAPTER 1

INTRODUCTION

1.0 Introduction

Regression coefficients computed by the ordinary least squares (OLS) method may give a false impression of precision. The OLS regression analysis assumes linear relationships. It is meaningless when relationships are non-linear. If relationships are severely non-linear, then regression coefficients will tell a misleading story. The OLS method which is used to estimate regression coefficients may be inappropriate to estimate population regression coefficients when the data exhibit atypical properties such as outliers, or when the variables are characterized by heteroscedasticity, multicollinearity and autocorrelation. The OLS estimates will be biased, inefficient or non-robust under such circumstances (Pennings *et al*., 2006).

Sometimes a non-linear relationship between an independent variable and the dependent variable can be converted into a linear relationship by a numerical transformation of the variables. Exponential relationships, which are quite common in rational theories of economic, political and social conduct, such as the so-called Cobb-Douglas production function, can be turned into linear relationships by taking logs of the separate variables.

In the mathematical statistician's thought, in an experiment of an infinite number of samples from the same population, an estimate is said to be unbiased when on average it hits the mark precisely. Fortunately, the OLS estimates are indeed unbiased and are efficient when the variables involved in regression analysis are distributed normally and the relationships between them are linear (Pennings *et al*., 2006). There are some noteworthy exceptions, however. Unusual properties of the data which render OLS estimates dubious or even simply mistaken will be discussed in this subsection. These include outliers, multicollinearity, heteroscedasiticity and autocorrelation.

Outliers are single cases which have a disproportionate effect on the slope of the regression line. Generally these cases have extreme values on the dependent or on the independent variables in the regression model. In the presence of outliers a regression coefficient does not tell a story about the majority of cases, but a story about a few outliers. Multicollinearity, on the other hand, means that two or more independent variables are almost inextricable. Disentangling almost inextricable variables will result in regression estimates that depend on small residual variations in the variables which will be measurement errors. Meanwhile, heteroscedasticity expresses that the residual variance is much larger for some values of the independent values of the independent variable (e.g. for cases with high values on the independent variable) than for others (e.g. for cases with low values).

Autocorrelation often arises in the context of regression analysis with time periods as units. Autocorrelation implies that a failure to explain the state of the dependent variable at one point in time reflects on subsequent time periods. The residual from the regression equation at one point in time depends upon the residual at some previous point(s) in time. The number of independent observations is in fact smaller than the number of time periods. These are some important statistics and econometrics issues in the estimation of regression and production functions. This dissertation focuses on the problems of outliers and multicollinearity. These issues will be discussed next in the context of a Cobb-Douglas production function. However, the arguments and results can be extended to more general specifications of production functions.

In statistics and econometrics, attention is given to the techniques that can deal with data having a typical observation which may be traced back to outliers and/or typographic errors and/or miscoding, amongst other possible reasons. Knowledge of these atypical observations and of their roots is particularly important in non-linear regression models and in time series, least squares (LS) and maximum likelihood estimators (MLE) (Cizek, 2008). On the other side, the use of methods robust to atypical observations is infrequent and usually limited, even in recent applications, to detection of outliers (Woo, 2003), although exceptions exist (Preminger & Franck, 2007). Within this context, a statistical procedure is called robust if it is insensitive to the occurrence of gross errors in the data (Cizek, 2011).

The multicollinearity problem exists in linear modeling when the regressors exhibit strong pairwise and/or simultaneous correlations, causing the design matrix to become non-orthogonal or worse, ill-conditioned (Barrios & Vregorio, 2007). Once the design matrix is ill-conditioned, the least squares estimates are seriously affected, e.g., instability of parameter estimates, reversal of the expected signs of coefficients, masking of the true behavior of the linear model being explored, etc.

In addition to this, multicollinearity commonly exists among economic

indicators influenced by similar policies that lead to their simultaneous movement along similar directions. Whether joint integration does, or does not, exist among the predictors, simultaneous drifting away in some directions, especially among time series that exhibit non-stationary behavior, is common (Barrios & Vregorio, 2007). There are some solutions proposed in the literature to address this problem (Section 2.8).

In several linear regression and prediction problems, the independent variables may be numerous and many of them may be highly collinear. This phenomenon is called multicollinearity and it is known that in this case the OLS estimator for the regression coefficients or predictor(s) based on these estimates may give very poor results (Gunst & Mason, 1979). Even for finite or moderate samples, a collinearity problem may still exist. Plenty of methods have been developed to overcome this problem, such as principal component regression (PCR), ridge regression (RR) and partial least squares (PLS). Yeniay and Goktas (2002) found that PLS regression yields somewhat better results in terms of the predictive ability of models obtained when compared to the other prediction methods.

The PLS regression is one of the regression techniques commonly used in presence of multicollinearity (Branden & Hubert, 2003). It makes use of the OLS regression steps in the calculation of loadings, scores and regression coefficients. It is a linear regression style developed to relate many regressors to one response variable or more (Gonzalez *et al*., 2009). However, it is known that the popular algorithms for PLS regression (Nonlinear Iterative Partial Least Squares (NIPALS) and

Straightforward Implementation of a Statistically-Inspired Modification of the PLS method (SIMPLS)) are very sensitive to outliers in the data set (Gonzalez1 *et al*., 2009). Outliers (either in the independent or dependent variables) pose a serious threat to such least squares analysis (Rousseeuw & Leroy, 2003) and least squares methods are strongly influenced by the presence of outliers (Marco *et al*., 2009).

Robust methods are presented to reduce or remove the effects of outlying data points (Gonzalez *et al*., 2009). The target of PLS regression is to predict Y from X and to describe their common structure (Abdi, 2010). Whether the number of predictors is large compared to the number of observations or not, X is likely to be singular and the regression approach is no longer feasible because of multicollinearity (Byrnes, 2007). The PLS regression seeks components from X that are also relevant for *Y*. Specifically, the PLS regression searches for a set of components (called latent vectors) that performs a simultaneous decomposition of *X* and *Y* with the constraint that these components explain as much as possible of the covariance between *X* and *Y* (Abdi, 2010).

On the other hand, the Partial Least Squares-Path Modeling (PLS-PM) is a statistical approach for modeling complex multivariable relationships among observed and latent variables (Vinzi *at el.*, 2009). In the past few years, these approaches have enjoyed increasing popularity in several sciences. Structural Equation Models (SEM) include statistical methodologies that allow for estimation of a causal theoretical network of relationships linking latent complex concepts, each measured by means of a number of observable indicators. As an alternative to the classical covariance-based approach, the PLS-PM is claimed to seek for optimal linear predictive relationships rather than for causal mechanisms thus privileging a predictive relevance oriented discovery process to the statistical testing of causal hypotheses.

The PLS-PM is a component-based estimation method. It is an iterative algorithm that separately solves out the blocks of the measurement model and then, in a second step, estimates the path coefficients in the structural model (Tenenhaus, 2008). Therefore, the PLS-PM is an attempt to explain at best the residual variance of the latent variables and, potentially, also of the manifest variables (MVs) in any regression run in the model (Trinchera & Russolillo, 2010). That is why PLS-PM is considered more of an exploratory approach than of a confirmatory one.

In PLS-PM, the latent variables (LV) are estimated as linear combinations of the manifest variables and thus they are more naturally defined as emergent constructs rather than latent constructs (Vinzi *et al*., 2009). This mode is based on multiple OLS regressions between each latent variable and its own formative indicators. It is now well established that the OLS regression may yield unstable results in presence of important correlations between explanatory variables; it is not feasible when the number of statistical units is smaller than the number of variables or when missing data affect the dataset. The PLS regression can nicely replace OLS regression for estimating path coefficients whenever one or more of the following problems occur: missing latent variable scores, strongly correlated latent variables, a limited number of units as compared to the number of predictors in the most complex structural equation.

According to PLS-PM structure, each part of the model requires to be validated: the measurement, structural and overall models. A PLS-PM is described by two models:

(i) a *measurement model* relating the MVs to their own LVs and

(ii) a structural model relating some endogenous LVs to other LVs.

The structural, inner and measurement models are also called the outer model (Tenenhaus *et al.*, 2004a). In line with this, evaluation of a research model using PLS analysis consists of two distinct steps. The first step includes the assessment of the *measurement model* and deals with evaluation of the characteristics of the latent variables and measurement items that represent them. The second step involves the assessment of the *structural model* and evaluation of the relationships between the latent variables as specified by the research model.

Economical problems underlie many events or problems that seem to be hard to explain and solve (Armagan & Ozden, 2007). The efforts for economical development have increasingly become important. Production functions are a basic component of all economics domains. As such, estimation of production functions has a long history in applied economics, starting from as early as the early 1800's. Unfortunately, this history cannot guarantee unequivocal success as many of the econometric problems that hampered early estimation are still an issue today (Ackerberg *et al*., 2006).

In economics, the Cobb-Douglas form of the production functions is widely used to represent the relationship of an output to inputs. It was proposed by Wicksell (1851–1926) and tested against statistical evidence by Charles Cobb and Paul Douglas during the period 1900–1928 (Felipe *et al*., 2008). In applied work, most researchers often commence by estimating the Cobb-Douglas production function using OLS regression modeling approach, hoping to obtain estimates of the labor and capital output elasticities that look plausible and interpretable from a theoretical point of view (Armagan & Ozden, 2007). According to several issues discussed above the OLS is not the best method.

The efforts for economical development have become important and highlighted the necessity of finding the best estimation method (Okafor & Eiya, 2011), though due to the macroeconomic level of the agricultural sector, no enough data with steady trend is readily available, besides other data-related problems (Diao *et al*., 2007). The PLS method is adopted here in analyzing agriculture data to avoid the preceding limitations of OLS. The PLS is especially good when dealing with small sample data, a multitude of variables and multicollinearity. It can greatly improve reliability and precision of the created model (Shang & Zhang, 2009).

Based on that the OLS regression is therefore not the best method. This research proposes to use the most widely used robust methods, namely, the minimum covariance determinant (MCD) to detect outliers and to use a popular regression method, that's, the PLS method. The proposed method is more efficient in detecting outliers and solving multicollinearity problems. Each of the issues is handled separately but using the same method of least squares for parameter estimation. This research is more theoretical and should be seen as a new approach to estimation of the Cobb-Douglas production function. The researcher proposes the development of Cobb-Douglas production function parameter using structural equation modeling (SEM) implemented with a robust RPLS-PM method to be applied on data drawn from a Libyan agricultural production sector.

1.1 Statement of the Problem

The Cobb-Douglas production function (Cobb & Douglas, 1928) is still today the most ubiquitous form in theoretical and empirical analyses of growth and productivity. The estimation of the parameters of aggregate production functions fits many of today's work on growth, technological change, productivity, and labour (Felipe & Adams, 2005).

Empirical estimates of aggregate production functions are a tool of analysis that is essential in macroeconomics while important theoretical constructs such as potential output, technical change, or the demand for labour, are based on them. It is usually fitted by first linearizing the models through logarithmic transformation and then applying the method of least squares (Prajneshu, 2008). However, the ordinary least squares (OLS) method is not the best estimation method (Kahane, 2001) and the log-transformation is also not the best transformation way, because log transformation process does not get rid of all the outliers and multicollinearity problems.

In fact, many problems are attendant to the OLS method. For example, least squares estimates are very sensitive to outliers, particularly in small samples (Wooldridge, 2009). In addition, multicollinearity often exists between the economic factors and it can greatly affect parameter estimation. The seriousness of multicollinearity will affect the results, mostly negatively, for example by increasing the OLS variance, reducing reliability of the model and compromising its stationarity

(El-Salam, 2011). Since the rank of parameter estimation is close to zero, the diagonal of the covariance matrices will become too big. This in turn means that the variance inflation factor (VIF) will be infinite. The end result will be elimination of some important explaining variables and a reduction in the reliability of model (Shang & Zhang, 2009). At the same time, when extracting different data from sample parameter estimation, multicollinearity will lead to fluctuant estimation and lack of stationarity (D'Ambra & Sarnacchiaro, 2010).

All these implications decrease the accuracy of the OLS estimates to a great extent and mask model's significance. Moreover, in order to ascertain its statistical validity, the OLS method needs a great amount of sample data (Shang & Zhang, 2009) which is not a common case in the agricultural sector due to the

macroeconomic level of the economy of this sector where no much data with steady trends are available, besides other problems (Diao *et al*., 2007). The PLS method is adopted here in analyzing agricultural production data to avoid the preceding limitations of OLS. The PLS is especially good at dealing with small sample data, plenty of variables and multicollinearity. It can greatly improve reliability and precision of the model (Shang & Zhang, 2009).

Also when applying a statistical method, in practice it often occurs that some observations deviate from the usual assumptions (Hubert *et al*., 2008). When collecting and analyzing real data, it is quite common that some observations are different from the majority of the samples (Bryman & Hardy, 2009). More precisely, some observations deviate from the model that is suggested by the major part of the data or they do not satisfy the usual assumptions. Such observations are called outliers. Sometimes they are simply the result of transcription errors (e.g., a misplaced decimal point or the permutations of two digits). Often the outlying observations are not incorrect but are entries that were produced under exceptional circumstances. Another possibility is that these observations belong to another population and they consequently do not fit the model well (Rousseeuw & Hubert, 2011). It is very important to be able to detect these outliers. They can then be used for example to pinpoint a change in the production process or in the experimental conditions. To find the outlying observation two strategies can be followed (Rousseeuw *et al*., 2006).

The first approach is to apply a classical method, followed by the computation of several diagnostics that are based on the resulting residuals. In general, classical methods can be so strongly affected by outliers that the resulting fitting model will not allow for detection of the deviating observations. This is called the masking effect. Additionally, some good data points may even show up as outliers, which is a phenomenon known as swamping (Chatterjee & Hadi, 2006).

A second strategy to detect outliers is to apply robust methods. The goal of robust statistics is to find a fit that is similar to the fit we would have found without the outliers. That solution then allows us to identify the outliers by their residuals from that robust fit (Huber & Ronchetti, 2009). However, as many classical methods are sensitive to outliers, the robust statistics aim at developing methods that are robust against the possibility that one or several unannounced outliers may occur anywhere in the data (Hubert *et al*., 2008).

These methods then allow detecting outlying observations by their residuals from a robust fit (Hubert *et al*., 2008). The target of robust statistics is thus to provide tools not only for assessment of robustness properties of classical procedures but also for producing estimators and tests that are robust to model deviations. In the case of robust estimation of multivariate location and scatter, robust covariance had been investigated for the first time by Maronna (1976).

The classical estimators of the location and scatter of a multivariate data set with *n* observations and *p* variables are the sample mean and sample covariance matrix. However, these estimators are not renitent to the presence of outliers in the data set, so, robust alternatives that yield reliable estimates even in the presence of contamination are desirable (Roelant *et al*., 2008). Since covariance matrices form a cornerstone in multivariate statistical analysis, robust estimators of shape or scatter can be used to construct robust multivariate methods (Roelant *et al*., 2008).

In many fields of applied statistics, it is common practice to log-transform data to make them near symmetrical. A very small value becomes large negative when log-transformed and zeroes become negatively infinite. Therefore, this application is of practical interest (Maronna *et al*., 2006). Linear transformations are usually performed to change the mean and standard deviation of the data distribution to something that is easier to work with (Brown, 1993).

One problem when estimating a Cobb-Douglas production function with micro data is how to deal with the observations that show positive output but do not use some of the inputs (Soloaga, 1999). The standard approach for fitting a Cobb-Douglas production function to micro data with zero values is to replace those values with "sufficiently small" numbers to facilitate the logarithmic transformation. In general, the estimates obtained are extremely sensitive to the transformation chosen, generating doubts about the use of a specification that assumes that all inputs are essential (as the Cobb-Douglas does) when they are not (Soloaga, 1999).

To overcome the above-mentioned problems, the PLS has become an important statistical tool for modeling relations between sets of observed variables by means of latent variables, especially for statistical problems dealing with high dimensional data sets. Despite the fact that the PLS handles the multicollinearity problem, it fails to deal with data containing outliers since it is based on maximizing the sample covariance matrix between the responses and a set of explanatory variables, which is known to be sensitive to outliers (Turkmen & Billor, 2010). Multicollinearity and presence of outliers are not an exception in real data sets, thus highlighting the need for robust PLS methods in several application areas such as chemometrics (Turkmen & Billor, 2010).

Structural equation models (SEMs) (Kaplan & Haenlein, 2004) can be a powerful method for dealing with multicollinearity in sets of predictor variables. SEMs include a number of statistical methodologies meant to estimate a network of causal relationships that are defined according to a theoretical model linking two or more latent complex concepts, each measured through a number of observable indicators. The basic idea is that complexity inside a system can be studied taking into account a causality network among latent concepts, called latent variables (LVs), each measured by several observed indicators usually defined as manifest variables (MVs).

The PLS approach to SEMs, also known as PLS-PM has been proposed as a component-based estimation procedure different from the classical covariance-based LISREL-type approach. According to Vinzi *et al.*, 2009, the PLS objective, unlike that of covariance-based SEM, is latent variable prediction and the method is not Covariance-based but Variance-based. So, the PLS tried to maximize the variance explained of the dependent variables. PLS-PM can be applied in smaller samples than SEM, SEM is usually used with an objective of model validation and needs a large sample. Also standard errors are estimated in PLS-PM using adjunct methods, including bootstrapping. There are generally no model fit statistics of kind available in SEM (Kline, 2010).

In Wold's (1975) seminal paper, the main principles of *partial least squares* for *principal component analysis* (Wold, 1966) were extended to situations with more than one block of variables. Other presentations of PLS-PM given by Wold appeared in the same year (Wold, 1975). Wold (1980) provided a discussion on the theory and the application of PLS for path models in econometrics. The specific stages of the algorithm are well described in Wold (1982) and in Wold (1985).

Extensive reviews on the PLS approach to SEMs with further developments are given in Chin (1998) and in Tenenhaus *et al* (2004b).

This thesis highlighted the development of a new model based on Cobb-Douglas production function and the use of RPLS-PM for parameter estimation. It attempts to solve two main problems in modeling, namely, multicollinearity and outliers. Each of these issues is handled separately but using the same method of least squares for parameter estimation. Due to certain limitations in the current software available to deal with the above problems, this developed methodology will be illustrated by applying it on real data originating from the agricultural production sector in Libya. The researcher additionally compares the new method (RPLS-PM) with the most widely used least squares method.

1.2 Research Questions

A number of interesting and important questions can be asked about this research, and it is the purpose of this research to investigate the answers to these questions.

(i) what is the importance of robust method in Cobb-Douglas production function?

(ii) do the multicollinearity, and outliers problems affect the model in Cobb-Douglas production function?

(iii) what are the best methods to solve the problems in Cobb-Douglas production function?

(iv) what is the importance of robust partial least squares-path modeling in Cobb-Douglas production function?

1.3 Research Objectives

The objectives of research are as follows:

(i) to solve outliers and multicollinearity problems and evaluate the parameters in Cobb-Douglas production function by method of robust partial least squares.

(ii) to determine and evaluate the agricultural model and to fit the real data with the model by means of Cobb- Douglas method through robust partial least squares-path modeling.

(iii) to develop the Agricultural Production model for the Libyan agricultural sector based on the model suggested by the Libyan Government underlining the country policy in promoting this sector.

1.4.0 Conceptual Framework

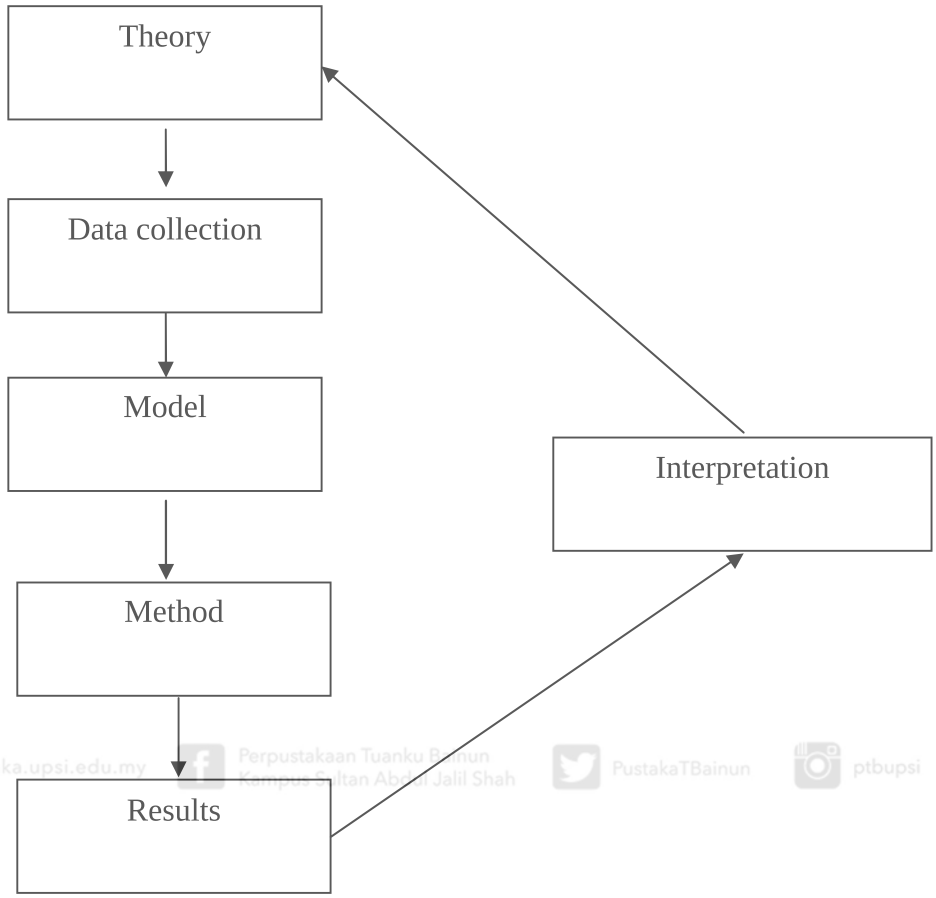
The conceptual framework of this research was formulated based on several models and concepts that have been applied and modified as well as on source material for the conceptual framework theory (Ethridge, 2004) shown in Figure 1.1.

In this research the researcher first specified a model based on Cobb-Douglas production function theory and then determined how to measure constructs, collect data, and then introduced the data to the vPLS1software package. The package fitted the data to the specified model and produces the results, which included overall model fit statistics and parameter estimates. This research was based on data pertaining to the 2004/05 production year of the Al-Kafrah Agricultural Production Project which is managed by the Raising National Capacities Project for Compilation and Analysis of Field Data under the patronage of the Public Administration of the Agricultural Production Projects. The input data and processing outputs in terms of wheat production per hectare are summarized in Table 1.1, and the items have been describing in Page 88.

Table 1.1. ***Input data and estimated wheat production per hectare in the 2004/2005 production year***

Item	**Unit**	**Irrigated Wheat (Al kofra Project)**		
		Value	**Price**	**Quantity**
I. Output Items				
Total Output	LYD	**1,381**		
II. Cost Items				
1. Agricultural inputs		**202**		
- Seeds	Kg	38	0.202	187.205
- Chemical fertilizers		132		
- DAP	Kg	86	0.400	216.114
- Urea 46%	Kg	25	0.052	482.525
- TSP	Kg	6	0.114	56.277
- Mixed	Kg	7	0.250	29.101
- Micro elements	Kg	7	1.000	6.558
- Pesticides		33		
- Herbicide (wide - prominal)	Liter	20	15.217	1.341
- Insecticide (aloxan)	Liter	11	19.773	0.573
- Insecticide (mylathion)	Liter	1	13.340	0.070
2. Operating & mainten. Cost		***425***		
- Electricity		226	226.422	1
- Fuel	Bale	14	14.469	1
- Spare parts		176	176.294	1
- Oil & lubrications	Bale	8	7.698	1
3. Wages, salaries & admin cost		***377***		
- Wages & salaries	LYD	246	245.627	1
- Social security	LYD	26	25.791	1
- Camping costs	LYD	9	8.534	1
- Administrative costs	m3	97	97.419	1
4. Depreciation	LYD	***47***	46.709	1
5. land	LYD			1
6. Water use	m3			
Total Cost	LYD	**1,051**		

Figure 1.1. The Conceptual Framework



1.4.1 Theory

In economics, the Cobb–Douglas functional form of the production function is widely used to represent the relationship of an output to input. It was proposed by Knut Wicksell (1851–1926), and tested against statistical evidence by Charles Cobb and Paul Douglas during 1900–1928. They considered a simplified view of the economy in which production output is determined by the amount of labor involved and the amount of capital invested. The function they used to model production was of the form:

$$P(L,K) = b\ L^{\alpha}\ K^{\beta} \tag{1.5}$$

where P is a total production (the monetary value of all goods produced in a year), L is a labour input (the total number of person-hours worked in a year), K is a capital input (the monetary worth of all machinery, equipment, and buildings), b is a total factor productivity, and α and β are the output elasticities of labour and capital, respectively.

These values are constants determined by available technology. Output elasticity measures the responsiveness of output to a change in levels of either labour or capital used in production. For example, if α = 0.15, a 1% increase in labour would lead to approximately 0.15% increase in output. Further, if $\alpha + \beta = 1$, the production function has constant returns to scale. That's, if L and K are each increased by 20%, then the total production (P) increases by 20%. Returns to scale refers to a technical property of production that examines changes in output subsequent to a proportional change in all inputs where all inputs increase by a constant factor.

If the output increases proportionally with this change, then there are constant returns to scale (CRTS). If the output increases by less than the proportional change, then there are decreasing returns to scale (DRS). If the output increases by more than proportion, then there are increasing returns to scale (IRS). However, if $\alpha + \beta < 1$, returns to scale are decreasing, and if $\alpha + \beta > 1$, returns to scale are increasing. Assuming perfect competition, then α and β can be shown to be the labour's and capital's shares of output.

1.4.1.1 Theories of the PLS-PM and SEM Approaches

The SEMs include a number of statistical methodologies meant to estimate, for example a network of causal relationships defined according to a theoretical model linking two or more latent complex concepts each measured through a number of observable indicators. The basic idea is that complexity inside a system can be studied as latent concepts called latent variables (LVs) each measured by several observed indicators usually defined as manifest variables (MVs). It is in this sense that SEM represents a joint-point between path analysis (Tukey, 1964; Alwin & Hauser, 1975) and confirmatory factor analysis (CFA) (Thurstone, 1931).

The SEMs essentially combine path models and confirmatory factor models, in other words, the SEMs incorporate both latent and observed variables. The early development of SEMs was due to Randall & Richard (2009). The approach was initially known as the JKW model but became known as the linear structural relations model (LISREL) with the development of the first software program, LISREL, in 1973 (Randall & Richard, 2009).

The PLS approach to SEM, also known as PLS-PM, has been proposed as a component-based estimation procedure different from the classical covariance-based LISREL-type approach. The presentations of the PLS-PM provided by Wold (1975, 1980) discuss the theory and application of PLS to path models in econometrics. The specific stages of the algorithm are well described in Wold (1982) and in Wold (1985). Extensive reviews on the PLS approach to SEM with further developments can be found in Chin (1998) and in Tenenhaus *et al*. (2004a).

The PLS-PM is a component-based estimation method (Tenenhaus, 2008). It is an iterative algorithm that separately solves out the blocks of the measurement model and then, in a second step, estimates the path coefficients in the structural model. Therefore, the PLS-PM is claimed to at best explain the residual variance of the latent variables and, potentially, of the manifest variables in any regression run of the model (Fornell & Bookstein, 1982). Unlike the classical covariance-based approach, the PLS-PM does not aim at reproducing the sample covariance matrix.

Finally, the PLS- PM is more oriented to optimizing predications (explained variances) than to the statistical accuracy of the estimates. The path coefficients are estimated afterwards by means of a regular regression between the estimated latent variable scores. According to Vinzi *et al* (2009), path coefficients are estimated via OLS multiple/simple regressions among the estimated latent variable scores. The PLS regression can nicely replace the OLS regression for estimating path coefficients wherever one or more of the following problems occur: missing latent variable scores, strongly correlated latent variables, a limited number of units as compared to the number of predictors in the most complex structural equation.

1.4.2 Data Collection

In this research, the data were collected from Al-Kufra Agricultural Production Project, Libya, pertaining to agricultural production of the wheat crop from the period 1960 to 2010. All reference data were collected from the Libyan Authorities; the main sources of data were government reports. Furthermore, some data were collected using the interview approach to data collection.

1.4.2.1 Libya Agricultural Sector

Libya, a North African country, lies along the southern coast of the Mediterranean, approximately between the latitudes 18^0 and 33^0 North and 9^0 and 25^0 East. Its total area is about 1,759,540 Km^2, of which more than 90% is desert. Most of the agricultural activities are scattered across oases in the desert. The population is about 6.1 millions (2008) and the average family size can range from six to seven persons per family.

Agriculture, fishing and forestry contributed 2% to Libya's gross domestic product (GDP) in 2007, a figure that has been steadily decreasing since 2002 when these three sectors contributed 4.3% to the GDP according to the 2008 international monetary fund (IMF) statistics. In the year 2008, the GDP per capita was US$ 8,900 with labor force strength of 1.5 million people. The Libyan currency is Dinar and its code is LYD. The prevailing climatic conditions are typical of the Mediterranean region which is characterized by variability and unpredictability; the rainfall fluctuates in quantity, frequency and distribution.

The major barriers to the growth of Libya are limited arable land (1.7% of Libya's total land area) and water resources, excessive-use of arable land and fertilizers, and a shortage of labour. The agricultural activities concentrate mainly along the coastline. Besides barley and wheat, the chief agricultural products are fruits and vegetables. These comprise dates, almonds, grapes, citrus fruits, watermelon, olives, and tomatoes, which constitute about 80% of the Libyan annual agricultural production.

Libya is investing a large share of the national revenues in agriculture on the hope that someday in the future the country becomes agriculturally self-sufficient where cultivation switches from subsistence farming to highly mechanized operations. Development plans aim at increasing the areas of irrigated lands and at extending the use of advanced farming techniques. In these plans, seeds and fertilizers have been subsidized. Areas singled out for development cover are Al-Jifara Plain in Tripolitania; Jabal Akhdar, east of Banghazi; part of Fezzan; and the oases of Kufrah and Sarir.

Wheat is mainly grown in the south-east and south-central plateaus. In the west near Tripoli, where the population is 1.2 million capita, there are about 18,000 ha of arable land used to mainly produce fruits and vegetables. On the other side, Benghazi (east), which has a population of 0.7 million capita, constitutes the centre of a region including Jebel Al Akhder where approximately 385,000 ha of rain-fed farmlands are situated. But even in the plateaus, rainfall is scarce and this is reflected in non-steady harvests. In the desert the choice of crops is limited to dates and olives, mostly, and all other crops are planted in and/or irrigated scattered oases, e.g., Al Kufra grain project.

1.4.2.2 AL-Kufra Agricultural Production Project

The main purpose of Kufra Agricultural Production Project was to support plant and animal production. Starting from 1960 it supported cultivation of wheat as the main crop, barley and legumes, as well as feed and Khertan in winter season. In addition, this project supported farming of cane, corn and Khertan in the summer season. On the other hand, in its beginning the project considered animal production as the second major target but in 1960 shifted concern to subsidiary target and based feeding sheep and camel flocks on the various plant products. The target area of the project amounted to 10,000 hectares of irrigated agriculture following the circular irrigation method comprising 100 circuits of 100 ha each. The project covered five farm units and the main source of irrigation water was well water drawn from 100 productive wells in the project-covered areas.

1.4.2.3 Agricultural Data Variables

The following variables have been used to estimate the production model:

Output items: Wheat production (ton/ha)

Input data:

(i) land and water (*LW*) data; *Harvested area* (thousand ha of land farmed with wheat) and *water use* (Million cubic meters of water/ha).

(ii) agricultural inputs (*AR*): Average per hectare (the value per hectare): *seeds* (seed ton/ha), *Chemical fertilizers* (ton/ha), and *Pesticides* (Liter/ha).

(iii) hours of operation and wages (*HW*): *Labour* use (hours/day) and *wages*, salaries and administrative cost (Average cost per season).

(iv) operation and maintenance (*OP*): *Electricity, fuel, spare parts, and oil and lubricants* (Average cost per season).

1.4.3 The Model

The researcher outlined a methodology that tests the relationship between output and input production function for Al-Kufra Agricultural Production Project, Libya. The production function was specified as a Cobb-Douglas function in the form of:

$$Y = \alpha_0 \prod_{i=1}^{11} x_i^{\alpha_i} \qquad (1.1)$$

where Y is the wheat crop output; the coefficient α_0 is the total factor efficiency parameter for composite primary factor input; the parameters α_i (i=1,…,11) are production elasticities; x_1 = water; x_2 = land; x_3 = seeds; x_4 = chemical fertilizers; x_5 = pesticides; x_6 = operation hours; x_7 = wages; x_8 = spare parts; x_9 = fuel; x_{10} = oil and lubricants; and x_{11} = electricity.

However, by taking the logs of both sides of Equation (1.1), the following log-linear relationship is obtained:

$$logY = \log\alpha_0 + \sum_{i=1}^{11} \alpha_i \log \; x_i + e \qquad (1.2)$$

This equation is described as log-linear because both the dependent variable and the regressors have been log-transformed. Where e is a error term. The coefficients in log-linear equations are elasticities since the Cobb-Douglas production function mentioned above is estimated only by least-squares estimators. However, to deal with outliers and multicollinearity problems, the development of an estimation parameter for the Cobb-Douglas production function is by the RPLS-PM. The model can be divided into two parts. The measurement model, which is the part which relates measured variables to latent variables, and the structural model, which is the part that relates latent variables to one another. Among the techniques available in the software used in this work is LVPLS-Lohmoller (Lohmoller, 1987, 1989).

The measurement model consists of five external models: *Harvested area* and *water use* (Land and Water (*LW*)); *seeds*, *chemical fertilizer* and *pesticide* (Agriculture inputs (*AR*)); *labour use* and *wages* (Hours of operation and Wages (*HW*)); *fuel, spare parts, oil and lubricants*, and *electricity* (Operation and maintenance (*OP*)); and *wheat crop* (*Output*). While the internal model consists of five structural models: the *LW* structural model (η_1), the *AR* structural model (η_2), the *HW* structural model (η_3), the *OP* structural model (η_4) and the Output structural model (η_5).

Creation of RPLS-PM-CD model was enlightened and guided by the recommendations of Kherallah (2000), Mahagayu *et al.* (2007) and Carver (2009) who expressed that there is a correlation between *Output* and each of *LW*, *AR*, *OP*, and *HW*. The suggested model methodology is illustrated by applying it to the data of an important output (wheat) of the Libyan agricultural production sector. In the theoretical research the following additive model was included:

$$\eta_3 = \Lambda_0 + \Lambda_{13}\eta_1 + \Lambda_{23}\eta_2 + \Omega_{13}\xi_1 + \Omega_{23}\xi_2 + \zeta_3 . \qquad (1.3)$$

where the coefficient Λ_0 is the total factor efficiency parameter for composite primary factor inputs. Parameters $\Lambda_{13}, \Lambda_{23}, \Omega_{13}, \Omega_{23}$ are production elasticities. ζ_3 is a error term.

1.4.4 Method

The Cobb-Douglas production function through the RPLS-PM-CD developed in this research describes the structural relationships between the *LW*, *AR*, *HW*, and *OP* besides their structural relationships with the *Output*. The researcher proposed that utilization of the MCD and PLS estimator first provides an estimate of the measurement model and describes the structural relationships between latent variables through the RPLS-path modeling. The strength of this research is manifested in the development of a new model based on Cobb-Douglas production function using RPLS-PM for parameter estimation. It attempts to solve outlier and multicollinearity problems.

1.5 Scope of Study

The agriculture sector is a critical component of, and contributor to, the process of economic development. With the recognition of this fact, the Libyan economic planners have emphasized on the development of agricultural and allied sectors right from the beginning of the economic planning process in Libya. In this research the researcher attempts to present a model for evaluating Libyan agricultural production that incorporates agriculturally-induced resource externalities.

The relationships between agricultural inputs and outputs are documented in various studies. This research used Cobb Douglas production function to determine the contribution of a particular input to the total production. Studies made earlier in Libya about the agricultural production function (Harvest Report 2008, the National Center for the Final Results of Improved Seeds to Harvest, and the Agricultural Research Center, Tripoli (Libya)) suggest that important inputs to wheat crop production encompass:

(i) land data requirements, which include : *Harvested area* and *water use (LW)*,

(ii) agricultural inputs: *Seeds*, *chemical fertilizers* and *pesticides (AR)*,

(iii) hours of operation *(Labour* use) and *Wages (HW)*,

(iv) operating and maintenance; e*lectricity, fuel, spare parts, and oil and lubricants (OP).*

In the Cobb-Douglas production function, this research used the following formula:

$$Y = \alpha_0 \prod_{i=1}^{11} x_i^{\alpha_i} \tag{1.1}$$

where Y is output of the wheat crop and α_0 is the total factor efficiency parameter for composite primary factor input. The parameters α_i (i=1, 2,…, 11) are production elasticities. And x_1 = Water, x_2 = Land, x_3 = Seeds, x_4 = Chemical fertilizer, x_5 = Pesticide, x_6 = Hours of operation, x_7 = Wages, x_8 = Fuel, x_9 = Spare parts, x_{10} = Oil and lubricants and x_{11} = Electricity.

Since Equation (1.1) demonstrates that the relationship between output and input is non-linear log-transformation the data was log-transformed before application of the OLS methods. However, this method is not the best method to estimate these parameters. The suggested model is to be applied to the input data is important for wheat production in Libya. The research proposed a two-step modelbuilding approach that emphasized on the analysis of two conceptually distinct models: a measurement model and a structural model.

The measurement model specifies the latent relationships among measured (observed) variables. The structural model specifies relationships among the latent variables as posited by theory. According to Enaami *et al* (2011b) the latent variables were *LW* which stands for *Land* and *Water*; *AR* (Agriculture inputs) which refers to input *Seeds*, *Chemical Fertilizers*, and *Pesticides*; *HW* which denotes *Operating Hours* and *Wages*; *OP* which expresses the operation and maintenance costs (such as *Fuel*, *Spare Parts*, *Oil and Lubricants*, and *Electricity*); and wheat as the *Output*.

To assess the deleterious effects of measurement error, a product indicator approach in conjunction with PLS-PM is proposed. The predictor and dependent variables are now considered as latent variables, which can not be measured directly. Instead, multiple indicators (i.e., measures) for these latent variables need to be obtained. Each indicator is modeled as being influenced by both the underlying latent variable and error. Product indicators reflecting the latent interaction variables are then created.

Each set of indicators reflecting their underlying latent variable are then introduced to the PLS for the estimation resulting in a more accurate assessment of the underlying latent variables and their relationships (Chin *et al.*, 2003). Therefore, there are two parts involve in this research: The first part involves testing the measurement model and it primarily deals with the validation of the latent constructs including the model. The second part involves the assessment of the structural model and testing of the hypothesized relationships between the latent constructs of the research model. The results of the assessment are based on the significance of the structural paths, which can be estimated by using PLS.

1.6 Significance of the Study

This research will benefit the estimation problems of Cobb-Douglas production function. Empirical searches of production functions, by and large, employ a function of the Cobb-Douglas type for reasons of computational economy. Typically, in this research; the researcher concentrates on the problems of outliers and multicollinearity, and on solutions that have been proposed to deal with these issues together.

A large number of research papers dealing with Cobb-Douglas production functions published in the area of agricultural economics are a testimony to the important role played by these models, by using the OLS regression methodology. These problems appear to have wide application for example; (e.g., Murthy (2002), Salvatore and Reagle (2002), Kennedy (2003), and Aguirregabiria (2009)). Yet, these problems have no known written solutions, and no serious research has been conducted in this field.

That is the main reason why the researcher attempted to develop the best possible method for providing estimates in the context of the Libyan agricultural sector by using the RPLS-PM method which may be applying equally to the agricultural production sectors of other countries. In addition, outliers are present in many real data. Hence, a brief overview of various high breakdown estimation methods based on MCD has been given. These methods not only masked the points in the outlier cluster, but also identified some inliers as outliers.

The PLS-PM is one of the most widespread approaches to estimation of unobservable factors. In the PLS-PM terminology, describe factors are called latent variables (LVs) and the indicator measured for each of them. The goal of PLS-PM is primarily to estimate the variance of endogenous constructs and in turn their respective manifest variables (MVs). Models with significant bootstrap or jackknife parameter estimates may still be considered invalid in a predictive sense. In this research, a prediction-oriented classification method in RPLS-PM is proposed to estimate Cobb-Douglas parameters. The PLS estimates are believed to produce more accurate estimates of SEM path coefficients than the OLS.

Finally, the RPLS approach and path-modeling (PM) are new, and have never been used to model Cobb-Douglas production function. Within this context, it seems that there is no published work using the Cobb-Douglas production function for estimates related to the Libyan agricultural sector. The main objective of this research is to develop a Cobb-Douglas production function based on wheat production data for Libya.

1.7 Organization of the Thesis

This work is organized in six chapters as shown in Figure 1.2. Chapter 1, where the researcher discusses the research framework, and introduces to this research. It begins with a briefing on the outliers problem then describes the robust statistics methods and the OLS problems. The researcher also concentrates on the problem of multicollinearity in Cobb-Douglas production function and on solutions proposed for resolving these problems. Then, a review of the problem of the present research is offered in Section 1.2. The research objectives of the research are presented in Section 1.3. In addition, the researcher outlines the conceptual framework in Section 1.4. In Section 1.5, a brief description of the Libyan Agriculture Sector is given. Finally, the researcher outlines the significance of the study in Section 1.6.

In Chapter 2 the researcher reviews the relevant literature. Section 2.0 starts by introducing the multicollinearity and outliers problems and Cobb-Douglas production function. In Section 2.1 the researcher discusses robust statistics, breakdown, MCD, and the MD. In Section 2.2 the researcher reviews the OLS, PLS and PLS-PM methods. While, Section 2.3 gives definition of the Cobb-Douglas production function. Section 2.4 elaborates on the problems of multicollinearity and variance inflation factor (VIF), also draws on current development in dealing with the multicollinearity and in the logarithmic transformation problems in the Cobb-Douglas production function. And lastly Section 2.5 summarizes the whole chapter.

Chapter 3 is concerned with the methodology followed by this research to realize its objectives. This chapter elaborates on the methods used in the current research. It begins with the research aims and a brief account of the nature of the problem and the methodology used. In Section 3.2 a brief overview of the data sources and methods of data collection are presented. The methods of collecting data are given. On the other hand, in Section 3.4 the variables considered by this research such as wheat production, farming seasons, the amounts of seeds used in the project, fertilization rates, the process of grass control in the project farms, irrigation in farms of the project, employment, quantity and value of losses of wheat, the harvesting process, productivity per hectare of wheat and barley, and operation and maintenance are discussed briefly.

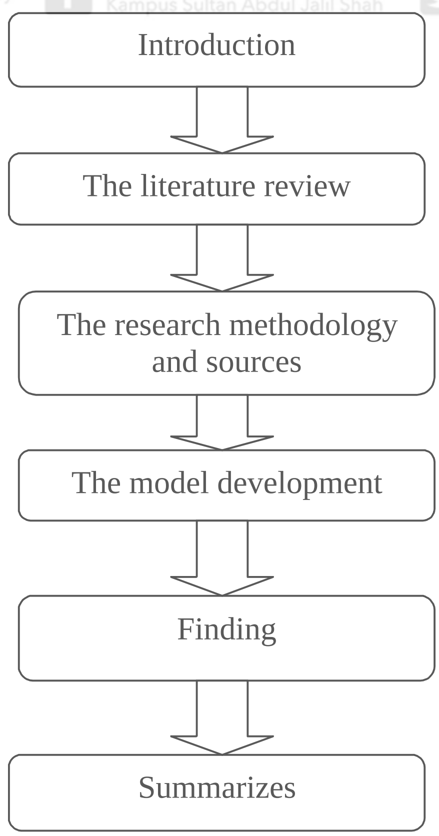
While, Section 3.5 gives illustrate research framework. The Sections 3.6, 3.7, 3.8 and 3.9 describes the overall research approach and modeling, and the overall estimation of a parameter is discussed. In Sections 3.10, 3.10, and 3.12 respectively show the data analysis, and the overall structure of model testing and evaluation. Finally, a summary of the chapter is presented.

Chapter 4 focuses on model development. Section 4.1 introduces and describes the model, the researcher discuses the measurement and the structured models, model development. And at the end of this chapter a section summarizes all foregoing sections.

Chapter 5 pertains to an overview of the research and its findings. In this chapter the researcher starts with an introduction. Results of the Cobb-Douglas estimates were shown in Section 5.1. In Section 5.2 the researcher presents outlier and multicollinearity detection methods. Section 5.3 provides details on the results of the RPLS-PM-CD model. The PLS analysis; assessment of the measurement and the structural models, is given in Section 5.4. And in the last section, Section 5.5, the researcher gives a summary of the chapter. And lastly, Chapter summarizes the whole research and includes some suggestions for future research as well as contribution of study and recommendations.

Figure 1.2: The outline of the research

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



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