

**AN EXPLORATORY STUDY ON GOAL PROGRAMMING AS AN  
ALTERNATIVE METHOD TO DEVELOP PREDICTION EQUATIONS**

**LAU CHIK KONG**

**A dissertation submitted in partial fulfillment of the requirements for the award of  
degree of Master of Science (Mathematics)**

**Faculty of Science  
Universiti Teknologi Malaysia**

**OCTOBER, 2004**

**Especially for my loving parents, dad Lau Heng Tiong and mum Loi Kiik Hee and  
my young brother, Lau Chik Muan.**

## ACKNOWLEDGEMENTS

A very great gratitude and appreciation expressed to all those who make a part to the successful cease of this dissertation with title 'An Exploratory Study As An Alternative Method to Develop Prediction Equations', either directly or indirectly.

In particular, I would like to thank my supervisor Dr. Maizah Hura Ahmad, for her dedication, support, useful guidance and advice. Her patience and helps really encourage me to finish this dissertation right on time. Also, I need to thank her for having the trust and confidence in me to handle this dissertation on myself.

Lots of gratitude and special thanks also to my family for their support, advice and encouragement throughout my study in Universiti Teknologi Malaysia (UTM). Last but not least, lost of thanks to my friends for their cooperation and help.

## ABSTRACT

One of the most promising techniques for multiple objective decision analysis is goal programming. Goal programming is a powerful tool which draws upon the highly developed and tested technique of linear programming, but provides a simultaneous solution to a complex system of competing objectives. Least squares method in regression analysis is also a popular technique used in decision making. It is an approach used in the study of relations between variables, particularly for the purpose of understanding how one variable depends on one or more other variables. However, one of the main problems is that the method of least squares is biased by extreme cases. This study proposes goal programming as an alternative to analyze such problems. The analysis were done by using QM for Windows and MINITAB software package.

## **ABSTRAK**

Pengaturcaraan gol adalah satu kaedah yang paling berkesan dalam penganalisan keputusan objektif berganda. Ia juga merupakan suatu teknik yang lebih baik berbanding pengaturcaraan linear dalam penyelesaian serentak untuk sistem kompleks. Kaedah kuasa dua terkecil dalam analisis regresi juga adalah satu teknik yang terkenal dalam membuat keputusan. Kaedah ini mengkaji hubungan antara pembolehubah terutama dalam memahami bagaimana satu pembolehubah bersandar kepada satu atau lebih pembolehubah yang lain. Bagaimanapun, masalah utama bagi kaedah kuasa dua terkecil ialah pengaruh kes ekstrim. Kajian ini mencadangkan pengaturcaraan gol sebagai kaedah alternatif untuk mengatasi masalah tersebut. Kajian ini menggunakan program QM for Window dan MINITAB dalam analisis.

## TABLE OF CONTENTS

<b>CHAPTER</b>	<b>SUBJECT</b>	<b>PAGE</b>
	<b>COVER</b>	i
	<b>DECLARATION</b>	ii
	<b>DEDICATION</b>	iii
	<b>ACKNOWLEDGEMENTS</b>	iv
	<b>ABSTRACT</b>	v
	<b>ABSTRAK</b>	vi
	<b>TABLE OF CONTENTS</b>	vii
	<b>LIST OF TABLES</b>	xiii
	<b>LIST OF FIGURES</b>	xv
	<b>LIST OF SYMBOLS</b>	xvi
	<b>LIST OF APPENDICES</b>	xix
 <b>CHAPTER 1</b>	 <b>RESEARCH FRAMEWORK</b>	 1
	1.0 Introduction	1
	1.1 Research Background	2
	1.2 Objectives of the Study	3
	1.3 Importance of the Study	3
	1.4 Scopes of the Study	3
	1.5 Thesis Organization	4

<b>CHAPTER 2</b>	<b>RESEARCH METHODOLOGY</b>	<b>5</b>
2.0	Introduction	5
2.1	Regression	5
2.1.1	Definition of Regression	5
2.1.2	The Purposes and Benefits of Regression Analysis	6
2.1.3	Simple Regression	7
2.1.4	Possible Criteria for Fitting a Line	10
2.1.5	Using Residuals to Test the Assumption of the Regression Model	12
2.2	Multiple Regression	13
2.2.1	The Mathematical Model	16
2.3	The Method of Least Squares	17
2.3.1	Polynomial Least Squares Fitting	18
2.4	The Least Squares Line	19
2.4.1	Residuals of Least Squares Line	21
2.5	Linear Multiple Regression Least Squares	22
2.5.1	Residuals of Multiple Regression Least Squares	25
2.6	Linear Goal Programming	25
2.6.1	History of Goal Programming	26
2.6.2	Advantages and Disadvantages of the Goal Programming	26
2.7	Goal Programming Model Formulation	27

2.7.1	Preemptive Goal Programming	29
2.7.2	Weighted Goal Programming	30
2.8	Solution Method of Goal Programming	36
2.8.1	The Graphical Method	36
2.8.2	The Modified Simplex (Multiphase) Method	40
2.9	Goal Programming Complications	53
2.9.1	Negative Right-hand Side Value	53
2.9.2	A tie in Selecting the Incoming Variable (Pivot Column)	53
2.9.3	A Tie in Selecting the Outgoing Variable (Pivot Row)	54
2.9.4	Alternative Optimal Solutions	54
2.9.5	An Infeasible Problem	54
2.9.6	An Unbounded Solutions	55
2.10	Sensitivity/Post Optimality Analysis	55
2.10.1	A Change in $v_s$	56
2.10.2	A Change in $u_i$	57
2.10.3	A Change in $b_i$	58
2.10.4	A Change in $y_{i,s}$	59
2.10.5	Additional of a New Goal	59
2.10.6	Additional of a New Decision Variable	60
2.11	Regression Analysis for Determining Relative Weighting	



	or Goal Constraint Parameter	
	Estimation	60
2.12	Summary	61
<b>CHAPTER 3</b>	<b>CASE STUDY ON USING LEAST SQUARES METHOD</b>	<b>62</b>
3.0	Introduction	62
3.1	Background of Data	62
3.2	Outliers	66
3.2.1	Box Plot (Box and Whisker Plots)	67
3.2.2	Existence of Outliers	69
3.3	Analysis Using Least Squares Method	74
3.3.1	Analysis on Data Set 1	74
3.3.2	Analysis on Data Set 2	77
3.3.3	Analysis on Data Set 3	79
3.4	Concluding Remarks	80
<b>CHAPTER 4</b>	<b>CASE STUDY ON USING GOAL PROGRAMMING</b>	<b>81</b>
4.0	Introduction	81
4.1	Converting Least Squares Problem into a Goal Programming	81
4.2	Analysis Using Goal Programming	82
4.2.1	Analysis on Data Set 1	82
4.2.2	Analysis on Data Set 2	87
4.2.3	Analysis on Data Set 3	93
4.3	Concluding Remarks	97

<b>CHAPTER 5</b>	<b>COMPARISON BETWEEN LEAST SQUARES METHOD AND GOAL PROGRAMMING</b>	<b>98</b>
5.0	Introduction	98
5.1	Mean Absolute Percentage Error (MAPE)	99
5.2	MAPE for Data Set 1 Analyzed Using Least Squares Method and Goal Programming	100
5.2.1	Discussion of the Results: Data Set 1	103
5.3	MAPE for Data Set 2 Analyzed Using Least Squares Method and Goal Programming	104
5.3.1	Discussion of the Results: Data Set 2	108
5.4	MAPE for Data Set 3 Analyzed Using Least Squares Method and Goal Programming	109
5.4.1	Discussion of the Results: Data Set 3	112
5.5	Discussion of the Results	112
5.6	Concluding Remarks	114
<b>CHAPTER 6</b>	<b>CONCLUSIONS AND SUGGESTIONS FOR FUTURE INVESTIGATION</b>	<b>115</b>
6.0	Introduction	115
6.1	Conclusions	115
6.2	Suggestions for Future Investigation	118

<b>REFERENCES</b>	120
<b>APPENDICES</b>	124
Appendix A – F	125 – 153

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Seven observations of annual starting salary and grade-point average	8
2.2	Procedure for achieving a goal	29
2.3	Data for single-goal model	32
2.4	The general initial modified simplex tableau	41
2.5	The general initial simplex tableau when the system constraints exist	44
2.6	The initial simplex tableau	48
2.7	The second tableau	50
2.8	The third tableau	51
2.9	The fourth tableau	51
2.10	The final tableau	52
3.1	The data of mass and elemental carbon	63
3.2	The data of enrollment, number of mailings and lead time	64
3.3	The data of gold, copper, silver and aluminium	65
5.1	The predicted values and errors for set 1	101
5.2	The predicted values and errors for set 1 with mild outlier removed	102
5.3	The predicted values and errors for set 1 with extreme outlier removed	102
5.4	The predicted values and errors for set 1 with both mild and extreme outliers removed	103
5.5	MAPE for data set 1	103
5.6	The predicted values and errors for set 2	105

5.7	The predicted values and errors for set 2 with the first mild outlier removed	106
5.8	The predicted values and errors for set 2 with the second mild outlier removed	106
5.9	The predicted values and errors for set 2 with the third mild outlier removed	107
5.10	The predicted values and errors for set 2 with all the mild outliers removed	107
5.11	MAPE for data set 2	108
5.12	The predicted values and errors for set 3	109
5.13	The predicted values and errors for set 3 with the first extreme outlier removed	110
5.14	The predicted values and errors for set 3 with the second extreme outlier removed	111
5.15	The predicted values and errors for set 3 with both extreme outliers removed	111
5.16	MAPE for data set 3	112
5.17	MAPE for data set containing outliers	113

## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	Scatter of the data on annual starting salary and grade-point average	8
2.2	Straight line fit based on inspection of the data	10
2.3	Typical error in fitting points with a line	11
2.4	The weakness of using $\sum(Y_i - \hat{Y}_i)$ to fit a line	12
2.5	Scatter of demand and price	14
2.6	Effect of price on demand, holding advertising constant	15
2.7	The multiple regression plane	24
2.8	Achievement of the system constraints	38
2.9	Achievement of the profit goal	39
2.10	Achievement of the profit and workforce goals	40
3.1	Box plots	69
3.2	Comparative box plot for mass ( $\mu\text{g}/\text{cm}^2$ )	70
3.3	Comparative box plot for number of mailings (x 1000)	71
3.4	Comparative box plot for silver (\$ per oz)	73
3.5	Comparative box plot for aluminium (cents per lb)	74
3.6	The scatter plot for mass ( $\mu\text{g}/\text{cm}^2$ ) depend on elemental carbon ( $\mu\text{g}/\text{cm}^2$ )	75

## LIST OF SYMBOLS

$a$	-	Intercept of response variable, $y$
$b$	-	Estimate regression coefficient
$\mathbf{b}$	-	Vector of $b$
$e$	-	Error / residual term
GP	-	Goal programming
$i$	-	$1, 2, \dots, m$
$iqr$	-	Interquartile range
$j$	-	$1, 2, \dots, n$
$k$	-	$1, 2, \dots, K$
$K$	-	The total number of preemptive priority factors
$L$	-	Minimize function
LS	-	Least squares
MAD	-	Mean absolute deviation
MAPE	-	Mean absolute percentage error
MPE	-	Mean percentage error
MSE	-	Mean square error
$n$	-	Number of prediction / forecasts
$p$	-	The total number of system constraints
$p(x)$	-	Polynomial function / fitting curve
QM	-	Quantitative method
RMSE	-	Root mean square error
$s$	-	$1, 2, \dots, S$
$S$	-	The total number of the decision and deviational variables where $S = n + 2m$
$X$	-	Independent (predictor) variable
$Y$	-	Dependent (response) variable

$\sum_{i=1}^n (Y_i - \hat{Y}_i)$	-	Sum of errors
$Z$	-	Objective function
$\beta_x$	-	Parameter of regression equation
$i'$	-	The pivot row
$s'$	-	The pivot column
$x^k$	-	Independent variable of $k$
$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	-	Sum of the squares of the error / least squares / ordinary least squares
$a_{ij}$	-	The coefficient associated with variable $j$ in the $i$ th goal
$A_i$	-	Artificial variable
$b_i$	-	The associated right hand side value
$C_b$	-	The column for assigning the preemptive priority factors and weights to the basic variables
$C_j$	-	The row for assigning the preemptive priority factors and weights to the basic and nonbasic variables
$E_i$	-	Excess or surplus variable
$g_k$	-	level of achievement of the goal in priority $k$ , where $\mathbf{g} = (g_1, g_2, \dots, g_k)$
$P_0$	-	Super priority factor/artificial objective function
$P_k$	-	The priority factor of the $k$ th goal
$r_{k,s}$	-	The index number for priority $k$ under $s$ th basic or nonbasic variable
$S_i$	-	Slack variable
$u_i$	-	The function of preemptive factors and weights associated with the $i$ th basic variable
$v_s$	-	The function of preemptive priority factors and weights associated with the $s$ th basic or nonbasic variable
$x_b$	-	Basic variable
$x_j$	-	The $j$ th decision variable
$y_{i,s}$	-	Element in the $i$ th row under the $s$ th basic or nonbasic variable. That is, the coefficient of the $s$ th basic or nonbasic variable in goal $i$
$Z_j - C_j$	-	The index row
$d_i^-$	-	Negative deviational variable from $i$ th goal (underachievement)



- $d_i^+$  - Positive deviational variable from  $i$ th goal (overachievement)
- $W_i^-$  - Positive numerical weight assigned to the negative deviational variable,  $d_i^-$  of the  $i$ th constant
- $W_i^+$  - Positive numerical weight assigned to the positive deviational variable,  $d_i^+$  of the  $i$ th constant
- $x_i^*$  - The optimal value of decision variables
- $\hat{b}_i$  - New value of  $b_i$
- $\hat{g}_k$  - New value of  $g_k$
- $\hat{r}_{k,s}$  - New value of  $r_{k,s}$
- $\hat{u}_i$  - New value of  $u_i$
- $\hat{v}_s$  - New value of  $v_s$
- $\hat{y}_i$  - Predicted / estimated value
- $\hat{y}_{i,s}$  - New value of  $y_{i,s}$

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
A1	The Output of MINITAB for Data Set 2 (remove first mild outlier)	125
A2	The Output of MINITAB for Data Set 2 (remove second mild outlier)	125
A3	The Output of MINITAB for Data Set 2 (remove third mild outlier)	126
A4	The Output of MINITAB for Data Set 2 (remove all mild outliers)	126
B1	The Output of MINITAB for Data Set 3 (contain outliers)	127
B2	The Output of MINITAB for Data Set 3 (remove first extreme outlier)	127
B3	The Output of MINITAB for Data Set 3 (remove second extreme outlier)	128
B4	The Output of MINITAB for Data Set 3 (remove both extreme outliers)	128
C1	Solution from QM for Window for Data Set 1 (contain outliers)	129
C2	Solution from QM for Window for Data Set 1 (remove mild outlier)	130
C3	Solution from QM for Window for Data Set 1 (remove extreme outlier)	132
C4	Solution from QM for Window for Data Set 1 (remove both mild and extreme outliers)	133

D1	Solution from QM for Window for Data Set 2 (contain outliers)	135
D2	Solution from QM for Window for Data Set 2 (remove first mild outlier)	137
D3	Solution from QM for Window for Data Set 2 (remove second mild outlier)	139
D4	Solution from QM for Window for Data Set 2 (remove third mild outlier)	141
D5	Solution from QM for Window for Data Set 2 (remove all mild outliers)	143
E1	Solution from QM for Window for Data Set 3 (contain outliers)	145
E2	Solution from QM for Window for Data Set 3 (remove first extreme outlier)	146
E3	Solution from QM for Window for Data Set 3 (remove second extreme outlier)	147
E4	Solution from QM for Window for Data Set 3 (remove both extreme outliers)	148
F	Terminology	150

## **CHAPTER 1**

### **RESEARCH FRAMEWORK**

#### **1.0 Introduction**

Many decision problems involve multiple objectives. Often, these objectives are conflicting with each other. A number of techniques have been proposed for multiple-objective decision making. One of the most promising techniques for multiple objective decision analysis is goal programming. Goal programming is a powerful tool which draws upon the highly developed and tested technique of linear programming, but provides a simultaneous solution to a complex system of competing objectives (Lee,1981). Goal programming can handle decision problems having a single goal with multiple subgoals.

Generally, many decision problems in organizations which involve multiple objectives are not easy to analyze by optimization techniques such as linear programming. Multiple-criteria decision making (MCDM) or multiple-objective decision making (MODM) has been a popular topic of management science during the past decade. A number of different approaches of MCDM or MODM have been proposed, such as multiattribute utility theory, multiple-objective linear programming, goal programming, compromise programming and various heuristics which are methods based on rules that are developed through experience. Goal programming is among the best widely accepted and applied technique. The primary reason for the wide popularity of goal programming appears to be its underlying philosophy of “satisficing” (Lee and Shim, 1986).

Nobel laureate Herbert A. Simon (1981) suggested that the satisficing approach, rather than optimizing is based on the concept of bounded rationality. This approach has emerged as a pragmatic methodology of decision making.

## 1.1 Research Background

A regression model is a mathematical equation that describes the relationship between two or more variables. The dependent variable is the one being explained, and the independent variables are the ones used to explain the variation in the dependent variable.

Regression techniques are associated with the fitting of straight lines, curves, or surfaces, to set of observations, where the fit is for one reason or another imperfect. The straight line is the simplest curve that can be fitted to a set of  $n$  paired observations  $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ . The least squares method is the most frequently used procedure for obtaining a linear function. A problem of fitting occurs only if the fit is for some reason imperfect. To be a statistical problem there must be some random element present in the data which leads to this inexactitude of fit. It is the nature of this random element that determines the appropriate method of fitting, that is of estimating the constants or parameters in the equation.

In simple linear regression analysis, the estimated regression model is  $\hat{Y} = a + bX$  ( $Y$  denotes the predicted dependent variable and  $X$  denotes the independent variable). In multiple regressions, the estimated regression model is

$$\hat{Y} = a + \sum_{i=1}^n b_i X_i$$

( $Y$  denotes the predicted dependent variable and  $X_i$  denotes the

independent variables). Although the method of least squares is one of the best known and probably widely utilized methods employed in the analyses of making predictions or forecasts of the future, most previous efforts in this area however, suffer from several disadvantages. According to Campbell (1972), one of the main

problems is that the method of least squares is biased by extreme cases. The current study proposes goal programming as an alternative to analyze such problems.

## **1.2 Objectives of the Study**

The following are objectives of this study:

- i. To develop prediction equations using least squares method.
- ii. To develop prediction equations using goal programming.
- iii. To compare the accuracy of goal programming and least squares method.

## **1.3 Importance of the Study**

Many prediction equations have been obtained using the least squares method. These equations have been used in various areas such as educational system planning, financial planning and economic policy analysis. This study explores goal programming as an alternative method to produce prediction equations. This is because goal programming is a widely accepted and applied technique in multiple objective decision analysis.

## **1.4 Scopes of the Study**

This study focuses on the use of the linear goal programming method to produce prediction equations in regression analysis problems. Only three data sets are considered. The first set consists of only one independent variable, the second set has two independent variables while the third set has three independent variables.

## **1.5 Thesis Organization**

This dissertation is organized into six chapters. Chapter 1 discusses the research framework. It begins with the introduction to goal programming and the least squares method. The objectives, importance and scope of this study are also presented.

Chapter 2 reviews the least squares method and goal programming. First, the least squares line and multiple regression least squares will be reviewed. Then, the modeling of the goal programming will be discussed. The discussion starts with the background of goal programming. Formulation and methodology of the goal programming model are also presented. Finally, some complications and post optimality analysis in goal programming are explained.

Chapter 3 begins with a discussion on outliers in a data set. In this chapter analysis of data sets using the least squares methods are carried out. Chapter 4 presents on the analysis of the same data sets using the goal programming model.

In chapter 5, comparison between the least squares and goal programming are made.

Chapter 6 summarizes and concludes the whole study and makes some suggestions for future investigation.