NEW APPROACHES IN ESTIMATING LINEAR REGRESSION MODEL PARAMETERS IN THE PRESENCE OF MULTICOLLINEARITY AND OUTLIERS

MOHAMMAD SABRY ABO AL-MASH

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

January 2017

This thesis is dedicated to my beloved father (Alhaj Sabry Abo Al-Mash)

ACKNOWLEDGEMENT

Praise is to the Almighty Allah the God of the Universe who gave me chances to live this beautiful life. This piece of work would not become possible without the contributions from many people and organizations. In this segment, I would like to acknowledge each and every person who has contributed their effort in this study by whatever means directly or indirectly. May the peace and blessings of Allah be upon The Final Messenger, Prophet Muhammad his family and combinations, and those who follow his path.

First and foremost, I would like to acknowledge my supervisor, **Assoc. Prof. Dr. Robiah Adnan** for her kind assistance and advice, beneficial criticisms, suggestions when I am hesitate and observations throughout this master thesis. Her invaluable help of constructive comments and useful advice throughout my research have contributed to the success of this research. Without their continued support and interest, this thesis would not have been the same as presented here. As appreciation also goes to the Universiti Teknologi Malaysia (UTM).

Many thanks go to my relatives back home especially to almarhum my beloved father, almarhum my beloved mother and my family. Not to forget, my loving wife, Ban who has been supported me throughout my study and to all my brothers, my sisters and my friends from whom I have received a great deal of support while conducting this research as well as studying at UTM. For the rest of the persons who had not been mention here, who have participated in various ways to ensure my research succeeded, thank you to all of you.Your kind and generous help will always be in my mind.

ABSTRACT

In multiple linear regression models, the ordinary least squares (OLS) method has been the most popular technique for estimating parameters of model due to its optimal properties and ease of calculation. OLS estimator may fail when the assumption of independence is violated. This assumption can be violated when there are correlations between the exploratory variables. In this situation, the data is said to contain multicollinearity and eventually will mislead the inferential statistics. However, the problem becomes more complicated when there are abnormal observational data known as outliers. It is now evident that presence of outliers has a serious threat on model with multicollinearity. In this research new procedures on how to improve the parameter estimation method in the presence of multicollinearity and outliers are put forward. The Principal Component Regression (PCR) and Ridge Regression (RR) individually are not resistant to outliers. The results of the research have showed that even if the PCR and RR produced good results with multicollinearity model, it may fail in the presence of outliers. The motive behind this research to find new procedures which are best with high break down point to estimate the model of regression with multicollinearity and outliers characteristics. The proposed methods are called Principal Component regression with Least Trimmed Squares (LTS) based on Tukey bisquare weighted (RWPCLTS) and Principal Component regression with Least Median Squares (LMS) based on Tukey bisquare weighted (RWPCLMS). Empirical applications of cigarette data according to its weight, tar, nicotine, and carbon monoxide contents for different brand of domestic cigarette were used to compare the performance between RWPCLTS and RWPCLMS with the existing methods of PCR and RR methods. A comprehensive simulation study evaluates the impact of multicollinearity and outliers on the proposed methods and existing methods. The considered percentages of outliers in the simulation are 0%, 5%, 10%, 15% and 20%. A selection criterion is proposed based on the best model with bias and root mean squares error for the simulated data and low standard error for real data. Results for both real data and simulation study suggest that the proposed criterion is effective for RWPCLTS and RWPCLMS in multicollinearity and outliers. Moreover, for both methods, the RWPCLTS tend to be the best followed by RWPCLMS when multicollinearity and outliers are present. This research shows the ability of the computationally intense method and viability of combining weighting procedures namely robust LTS-estimation or LMS-estimation and multicollinearity diagnostic methods of PC to achieve accurate regression model. In conclusion, the proposed methods are able to improve the parameter estimation of linear regression by enhancing the existing methods to handle the problem of multicollinearity and outliers in the data set. This improvement will help the analyst to choose the best estimation method in order to produce the most accurate regression model in the presence of multicollinearity and outliers.

ABSTRAK

Dalam pelbagai model regresi linear, Kaedah kuasa dua terkecil biasa (OLS) telah menjadi teknik yang paling popular untuk menganggar parameter yang ada pada model kerana sifat-sifat optimumnya dan cara pengiraan yang mudah. Penganggar OLS mungkin akan gagal apabila andaian kemerdekaan dilanggar. Andaian ini boleh dilanggar apabila terdapat korelasi antara pembolehubah penerokaan. Dalam situasi ini, data tersebut dikatakan mengandungi multikolinearan dan akhirnya akan memesongkan statistik inferensi. Walau bagaimanapun, masalah ini menjadi lebih rumit apabila terdapat ketaknormalan data pemerhatian yang dipanggil titik terpencil. Ia kini jelas bahawa kehadiran titik terpencil boleh menjadi satu ancaman yang serius kepada model dengan adanya multikolinearan. Dalam kajian ini, prosedur baharu untuk memperbaiki kaedah anggaran parameter dengan kehadiran multikolinearan dan titik terpencil dikemukakan. Regresi Komponen Prinsipal (PCR) dan Ridge Regresi (RR) secara individu tiada daya tahanan pada titik terpencil. Keputuson kajian telah menunjukkan bahawa walaupun PCR dan RR menghasilkan keputusan yang baik dengan model multikolinearan, ia mungkin gagal dengon kehadiron titik terpencil. Motif di sebalik kajian ini untuk mencari prosedur baharu yang terbaik dengan titik pecahan tinggi untuk menganggarkan model regresi dengan multikolinearan dan yang mempunyai ciri-ciri titik terpencil. Kaedah yang dicadangkan adalah dipanggil regresi Komponen Prinsipal yang LTS berdasarkan Tukey bisquare berwajaran (RWPCLTS) dan Principal Component regresi dengan LMS berdasarkan Tukey bisquare berwajaran (RWPCLMS). Aplikasi empirikal data rokok mengikut berat, tar, nikotin, dan kandungan karbon monoksida untuk pelbagai jenama rokok tempatan telah diguna untuk membandingkan prestasi antara RWPCLTS dan RWPCLMS dengan kaedah PCR yang sedia ada dan kaedah RR. Satu kajian simulasi menyeluruh menilai kesan multikolinearan dan titik terpencil pada kaedah yang dicadangkan dan juga pada kaedah yang sedia ada. Peratusan titik terpencil yang dipertimbangkan dalam simulasi adalah 0%, 5%, 10%, 15% dan 20%. Satu kriteria pemilihan adalah dicadangkan berdasarkan model terbaik dengan kecenderungan dan ralat punca kuasa dua min bagi data simulasi dan ralat piawai yang rendah untuk data sebenar. Keputusan untuk kedua-dua data sebenar dan kajian simulasi menunjukkan bahawa kriteria yang dicadangkan itu adalah berkesan untuk RWPCLTS dan RWPCLMS dalam multikolinearan dan titik terpencil. Lebih-lebih lagi, untuk kedua-dua kaedah, RWPCLTS cenderung untuk menjadi kaedah yang terbaik diikuti oleh RWPCLMS dengon kehodiron multikolinearan dan titik terpencil. Kajian ini menunjukkan keupayaan kaedah berkomputer yang amat rumit dan daya kebolehan menggabungkan prosedur-prosedur berpemberat iaitu teguh LTS-anggaran atau LMS-anggaran dan kaedah multikolinearan diagnostik PC untuk mencapai model regresi tepat. Kesimpulannya, kaedah yang dicadangkan dapat meningkatkan anggaran parameter regresi linear dengan meningkatkan kaedah sedia ada untuk menangani masalah multikolinearan dan titik terpencil dalam set data. Peningkatan ini akan membantu penganalisis untuk memilih kaedah anggaran yang terbaik untuk menghasilkan model regresi yang paling tepat dengan kehadiran multikolinearan dan titik terpencil.

TABLE OF CONTENTS

LIST OF TABLEs

CHAPTER 1

INTRODUCTION

1.1 Background of the Problem.

Regression analysis is a technique used in all fields of engineering, science, and management that required fitting a model to a set of data. It is a customary method used to mathematically model a response variable as a function of the explanatory variables. Explanatory variables can be defined as factors that can be different or manipulated in an experiment and normally denoted by x. Dependent variables are the response variables to the explanatory variables that are present in an experiment. We can have several independent variables which influence one or more dependent variables at the same time. This situation was known as multiple linear regressions. There are many methods available for estimating the model parameters, but ordinary least squares (OLS) method is the most popular method in statistics applications.

The ordinary least squares (OLS) is usually used to estimate the parameter coefficients of the linear regression model because of its optimal properties and straight forward computation. It is one of the oldest statistical methods, dating back to the age of slide rules until today. Computers are abundant, high-quality statistical software is free, and statisticians have developed several new estimation methods in making it easier to understand this model and thus, linear regression is still popular (Rao *et al*., 2008). The OLS method was discovered independently by Gauss in 1795 and Legendre in 1805

(Sorenson, 1970). OLS minimizes the sum of the squared distances for all points from the actual observation to the regression surface. The least squares estimator is attractive because of computational simplicity, availability of software, and statistical optimality properties. From the Gauss-Markov theorem, least squares are always the best linear unbiased estimator (BLUE). BLUE means that among all unbiased estimators, OLS has the minimum variance. If ε is assumed to be normally, independently distributed with mean 0 and variance σI, least squares is the uniformly minimum variance unbiased estimator. In multiple linear regression thus, BLUE property no longer exists in the presence of multicollinearity.

Under this assumption, inference procedures such as hypothesis tests, confidence intervals, and prediction intervals are powerful. However, if ε is not normally distributed, then the OLS parameter estimates and inferences can be flawed.

Violations of the independent assumption can results to multicollinearity in the data set. The inference procedures estimated based on the presence of multicollinearity will invalidate the model parameter. Multicollinearity or collinearity refers to the situation where there is either an exact or approximately exact linear relationship among the explanatory variables (Gujarati, 2003). When multicollinearity is present in a set of explanatory variables, the ordinary least squares (OLS) estimates of the multiple linear regression coefficients tend to be unstable. This will results in causes the ratios of one or more coefficients tend to be statistically insignificant (Chatterjee and Hadi, 2006). Because of its large variances and covariance matrix, the parameter estimate to be less precise (Adnan., 2006) and can result in the wrong inferences.

Therefore, the greater the multicollinearity, the less interpretable are the parameters. In such circumstances, there are many alternative estimation reduction regression methods that are used such as Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR). Although all three reduction regression models are biased (with big variances), they tend to have more precision when measured by Mean Square Error (Hoerl and

Kennard, 1976) and (Draper and Smith, 1998). OLS estimates are preferred because they are unbiased, consistent, and have smaller standard errors when there are no problems in model like multicollinearity and the model is robust.

Coefficient of Determination (R^2) is one of the most important tools in statistics which is widely used in data analysis in economics, physics, chemistry and many more fields. The coefficient of determination is equal to the regression sum of squares (that is, explained variation) divided by the total sum of squares (that is, total variation). Coefficient of determination allows us to forecast or predict the possible outcomes and possible variability in data. Coefficient of determination is denoted by R^2. The value of coefficient of determination lies between 0 and 1. The higher the value of R^2, the better the prediction becomes. That is, $0 \leq R^2 \leq 1$ in mathematical terms. An R^2=0, means that the dependent variable cannot be predicted from the independent variable. An R^2 of 1 means the dependent variable can be predicted without error from the independent variable.

However, the problem of multicollinearity is usually occurs in a multivariate situation, not bivariate variables. That means the bivariate correlation matrix is not sufficient to eliminate consideration of the problem of multicollinearity. The problem is not only that the two independent variables are highly correlated, but that one independent variable is highly correlated with at least one of the other independent variables. That means we need to examine the R^2's of each independent variable regressed on the other independent variables. Evidence of collinearity is provided by the correlation matrix among the regression coefficients. The weight/coefficient in regression model indicate the contribution of independent variable to the dependent variable.

Therefore, the existing of multicollinearity in regression model can be misleading of the effects or contribution of independent variables. Additionally, the standard errors of the coefficients are artificially inflated. Hence, there is a greater probability that we will incorrectly conclude that a variable is not statistically significant. Multicollinearity is likely to be present to some extent in most economic models. The issue is whether the multicollinearity has a significant effect on the regression results (Mela and Kopalle, 2002).

However, outliers are values in a data set that are far from the other values and far from the line implied by the rest of the data. An observation in which its standardized residual is large relative to other observations in the data set, it is considered an outlier that lies at a distance from the rest of the data set (Montgomery et al., 2015). Outliers, which occur in real data, are due to many reasons including interchanging of values, typing or computation errors, unintended observations from different populations and transient effects. Outliers can also be due to genuinely long-tailed distributions. Hampel *et al*. (2011) summarized the results of numerous studies of the frequency of outliers in real data and concluded that altogether 1-10% outliers in routine data are more the rule rather than the exception.

However, there are several methods proposed in the literature that handle the multicollinearity and outlier identification problems yet, there is little guidance for the practitioner on which methods perform well in representative under multicollinearity and outlier scenarios. Few methods are readily available on standard statistical packages for multicollinearity and outlier identification.

However, the use of the Variance Inflation Factors (VIF) is the most reliable way to examine multicollinearity. As a rule of thumb, if any of the VIF is greater than 10 (greater than 5 to be very conservative) there is a multicollinearity problem. Prior to estimating the regression equations, if we notice that any of the bivariate correlations among the independent variables are greater than 0.70, we may be facing the problem of multicollinearity (Ethington, 2013).

Mostly, analysts used method to detect outliers is visualization. For this thesis, we will use visualization method, like box plot to detection outlier values. Typically, for each of the independent variables (predictors) and dependent variable, the following plots are drawn to visualize the following behavior by box plot. Box plot is to spot any outlier observations in the variable. Having outliers in the predictor can drastically affect the predictions as they can easily affect the direction/slope of the line of best fit. By default, any value is higher than the 1.5* interquartile range' (1.5 * IQR) above the upper quartile (Q3), the value will be considered as outlier. Similarly, if any value is lower than the 1.5* interquartile range' (1.5 * IQR) below the lower quartile (Q1), the value will be considered as outlier.

Adnan *et al*., (2006) discussed several approaches for handling multicollinearity problem that have been developed such as Principal Component Regression, Partial Least Squares Regression and Ridge Regression. Principal Components Regression (PCR) is a combination of principal component analysis (PCA) and ordinary least squares (OLS) to handle multicollinearity. Partial Least Squares (PLS) is an approach similar to PCR because one needs to construct a component that can be used to reduce the number of variables. Ridge Regression is the modified least squares method that allows biased estimators of the regression coefficient.

The criterion is obtained by minimizing the ordered squared residual. Thus, the procedure leads to estimated regression coefficients that minimize the median of the squared residual. The ordinary least squares regression (OLS) can be duly effected by the presence of outliers and the multicollinearity measures.

However, it is now evident that the ordinary least squares regression (OLS) can be duly effected by the presence of outliers. Many robust regressions have been introduced to handle the problems of outliers, for example, Least Median Squares (LMS) regression. There is another robust regression which uses the Least Trimmed of Squares (LTS). LTS regression is obtained by minimizing the sum of squared residuals, where the squared residuals are ordered from smallest to largest. We might let h have the same value as for LMS, so that it will be a high breakdown point estimator. But sometimes 50% breakdown point produces poor results in this case when $h = \frac{n}{2}$. The results imply that it is better to use the larger value of n when explaining a trimming percentage α. Rousseeuw and Leroy (1987) proposed that h be selected as $h = [n(1-\alpha)] + 1$.

1.2 Statement of the Problem

In multicollinearity diagnostics methods, the methods used to estimate the regression model are based on OLS estimate which will be affected by the presence of outliers. Thus there is a requirement to find a suitable robust estimators that will not be much affected by outliers and multicollinearity problems. This prompted us to introduce a new method that is reliable in situations where the problems of outliers and multicollinearity occur simultaneously.

1.3 Objectives of the Study

The research objectives are:

(i) To develop an alternative robust estimation techniques for multiple linear regression model in the presence of multicollinearity and outliers by combining robust LTS with initial and scale estimate of LTS-estimator and principal component using Tukey bisquares weighting procedures.

(ii) To propose new approaches of robust estimation techniques for multiple linear regression model in the presence of multicollinearity and outliers by combining robust LMS with initial and scale estimate of LMS-estimator and principal component using Tukey bisquares weighting procedures.

(iii) To compare the performance of the proposed methods with RR,PCR and OLS estimation method for handling the multicollinearity problem in the presence of outliers.

1.4 Scope of the Study

This research will emphasize the problem of multicollinearity and outliers in linear regression models using real data and simulated data. The method of Ridge Regression (RR), Principal Component Regression (PCR) and ordinary least squares (OLS) are discussed in detail. The linear regression techniques based on multicollinearity diagnostic measures are used to remedy the problems of multicollinearity in the data. The method of Ridge Regression is obtained by adding a suitable small bias estimator to the diagonal elements which is considered as modified procedures of least squares estimator. On the other hand, the principal component analysis computes the linear combination of the independent variables. However, outliers have a great impact on the regression model, and the presence of outliers will invalidate the parameter estimate results in producing wrong inferential statistics. This work will compare the performance of robust estimators, Least Median Square (LMS) and Least Trimmed of Square (LTS) which are combined with Principal Component and weighting procedures of Tukey weighted function to handle multicollinearity in the presence of outliers.

However, the robust methods of LTS and LMS with the multicollinearity measures will be computed using the weighted least squares method procedures of Tukey bisquares weighted function introduced by Huber (1973). The performance of the proposed methods Robust Weighted Principal Component Regression Least Trimmed Squares (RWPCLTS) and Robust Weighted Principal Component Regression Least Median Squares (RWPCLMS) will be compared with the existing OLS and the multicollinearity measures of RR and PCR which are also obtained based on OLS-estimator using real data and simulated study.

Real data and simulation studies are the primary tool used to accomplish the objectives outlined in Section 1.3. In most cases, the simulation studies are set up as design instruments to gain the maximum performance of each estimation method.

In this thesis, there are enough replicates for the simulation procedures to get a clear indication of the performance of each estimator. The simulation data of multicollinearity and outliers problems in linear regression model will consider the number of parameters p to be significantly smaller than the number of cases of sample size (n). This study will be analyzed using R-programme software version 3.2.4.

1.5 Significance of Study

Presence of multicollinearity results in producing large variance and covariance for the least squares estimator of the regression coefficients causing biases in the variance of the covariance matrix that are used to estimate the standard error, confidence intervals and other coefficients of the regression model. However, the problem becomes more complicated when there are outliers in the data which will cause inaccurate parameter estimation of the regression model resulting in producing unreliable result. The existing methods deal with outliers and multicollinearity problems separately; therefore there is a need to introduce a new robust method that will handle the problems of multicollinearity in the presence of outliers at the same time.

The finding of this study will help us in modeling any complicated data where multicollinearity and outliers usually occur simultaneously. This study will also help us to promote the medical impact of the growing nation. The real dataset is useful for introducing the ideas of multiple regression and provides examples of multicollinearity and an outlier in variables. We have also modified a workable friendly computer coding method for the data with the situation of this kind using R software and Microsoft Excel.

1.6 Summary and Outline of Study

The aim of this study is to find the best method and procedure to handle multicollinearity and outlier problems by comparing the performances of the five methods to determine which method is superior to the others in terms of practicality. Practicality means how effective or convenient a method is in actual use. The algorithms for each method used in this study are shown in Chapter 3.

Chapter 2 reviews the relevant literature on published work done recently concerning the problems of multicollinearity and outliers. Discussion on methods for handling multicollinearity and outliers problems in linear regression analysis are presented in Chapter 3. Chapter 4 describes the simulation data set and real data and the analysis of the five methods. Chapter 5 discusses the performances of the five methods and makes comparisons among them and concludes the study and makes recommendations for further study.

REFERENCES

Abbott, C. A., Vileikyte, L., Williamson, S., Carrington, A. L., & Boulton, A. J. (1998). Multicenter study of the incidence of and predictive risk factors for diabetic neuropathic foot ulceration. *Diabetes care*, *21*(7), 1071-1075.

Abdi, H. (2007). Singular value decomposition (SVD) and generalized singular value decomposition. *Encyclopedia of measurement and statistics*, 907-912.

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(1), 97-106.

Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, *20*(10), 1263-1271.

Adnan, N. B. (2006). COMPARING THREE METHODS OF HANDLING MULTICOLLINEARITY USING SIMULATION APPROACH.

Adnan, N., Ahmad, M.H. and Adnan, R., 2006. A comparative study on some methods for handling multicollinearity problems. Matematika 22(2): 109-119.

Aguilera, A. M., Escabias, M., & Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, *50*(8), 1905-1924.

Alamgir, S., & Ali, A. (2013). Split Sample Bootstrap Method. *World Applied Sciences Journal*, *21*(7), 983-993.

Al-Hassan, Y. M. (2010). Performance of a new ridge regression estimator. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, *9*(1), 23-26.

Alkhamisi, M. A., & Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics—Theory and Methods*, *37*(4), 544-564.

Alkhamisi, M., Khalaf, G., & Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics—Theory and Methods*, *35*(11), 2005-2020.

Alma, Ö. G. (2011). Comparison of robust regression methods in linear regression. *Int. J. Contemp. Math. Sciences*, *6*(9), 409-421.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583-618.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 327-351.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics, 16*(4), 523-531.

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, *2*(2), 123-146.

Bagheri, A., & Midi, H. (2009). Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. *Journal of Mathematics and Statistics*, *5*(4), 311-321.

Bagheri, A., & Midi, H. (2011). On the performance of robust variance inflation factors. *International Journal of Agricultural and Statistical Sciences*, *7*(1), 31-45.

Bagheri, A., Habshah, M., & Imon, R. H. M. R. (2012). A novel collinearity-influential observation diagnostic measure based on a group deletion approach. *Communications in Statistics-Simulation and Computation*, *41*(8), 1379-1396.

Barnett, V., & Lewis, T. (1984). Discordancy tests for outliers in univariate samples. *Outliers in statistical data*, *3*, 120-121.

Barnett, V., & Lewis, T. (1994). Outliers in Statistical Data (Probability & Mathematical Statistics).

Beaton, A. E., & Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, *16*(2), 147-185.

Belsey, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity* (Vol. 571). John Wiley & Sons.

Ben-Gal, I. (2005). Outlier Detection in Maimon, O. and Rockach, L. (eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers.

Björkström, A. (2007). Ridge regression and inverse problems.

Björkström, A., & Sundberg, R. (1999). A generalized view on continuum regression. *Scandinavian Journal of Statistics*, *26*(1), 17-30.

Blatná, D. (2006). Outliers in regression. *Trutnov, 30*, 2006-03.

Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. J. R. Statist. Soc. B 37, 149-192.

Çamdevýren, H., Demýr, N., Kanik, A., & Keskýn, S. (2005). Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*, *181*(4), 581-589.

Chatterjee, S., & Hadi, A. S. (2006). Simple linear regression. *Regression Analysis by Example, Fourth Edition,* 21-51.

Chauvenet, W. (1863). A Manual of Theoretical and Practical Astronomy: Embracing the General Problems of Spherical Astronomy, the Special Applications to Nautical Astronomy, and the Theory and Use of Fixed and Portable Astronomical Instruments, with an Appendix on the Method of Least Squares.

Christensen, G., Saif, M., & Soliman, S. (2007). A new algorithm for finding the optimal solution of the least absolute value estimation problem. *Canadian Journal of Electrical and Computer Engineering*, *1*(32), 5-8.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15-18.

Cook, R. D., & Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, *22*(4), 495-508.

Costa, C., Menesatti, P., & Spinelli, R. (2012). Performance modelling in forest operations through partial least square regression. *Silva Fennica*, *46*(2), 241-252.

D'Ambra, A., & Sarnacchiaro, P. (2010). Some data reduction methods to analyze the dependence with highly collinear variables: A simulation study.*Asian J. Math. Stat*, *3*(2), 69-81.

Dempster, A. P., Schatzoff, M., & Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, *72*(357), 77-91.

Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information systems outsourcing: a survey and analysis of the literature. *ACM Sigmis Database*, *35*(4), 6-102.

Donoho, D. L., & Huber, P. J. (1983). The notion of breakdown point. *A festschrift for Erich L. Lehmann*, *157184*.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Dorugade, A. V., & Kashid, D. N. (2011). Parameter estimation method in Ridge Regression. *Statistical Simulations, 31*(4), 653-672.

Draper NR, Smith H. (1998). Applied Regression Analysis, 3rd edition. Wiley: New York.

Draper, N. R., & Smith, H. (1998). Wiley series in Probability and Statistics. *Applied Regression Analysis, Third Edition*, 707-713.

Duzan, H., & Shariff, N. S. B. M. (2015). Ridge Regression for Solving the Multicollinearity Problem: Review of Methods and Models. *Journal of Applied Sciences*, *15*(3), 392.

EDGEWORTH, F.Y. (1887), "On Observations Relating to Several Quantities," Hermathena, 6, 279-285.

El-Dereny, M., & Rashwan, N. I. (2011). Solving multicollinearity problem using ridge regression models. *Int. J. Contemp. Math. Sciences*, *6*(12), 585-600.

Engelen, S., Hubert, M., Branden, K. V., & Verboven, S. (2004). Robust PCR and robust PLSR: A comparative study. In *Theory and applications of recent robust methods* (pp. 105-117). Birkhäuser Basel.

Escabias, M., Aguilera, A. M., & Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, *16*(3-4), 365-384.

Ethington, Dr. Bunty, EDPR 7/8542, Univ of Memphis (2013). https://umdrive.memphis.edu/yxu/public/Multicollinearity.pdf 2013-07-02.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.

Ferguson, T. S. (1961). On the rejection of outliers. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 1, pp. 253-287). University of California Press Berkeley.

Fornell, C., & Robinson, W. T. (1983). Industrial organization and consumer satisfaction/dissatisfaction. *Journal of Consumer Research*, 403-412.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of statistical software*, *8*(15), 1-27.

Fox, J., & Weisberg, S. (2011). An R companion to applied regression 2 edition Sage. *Thousand Oaks*.

Freund, A. M., & Baltes, P. B. (1998). Selection, optimization, and compensation as strategies of life management: correlations with subjective indicators of successful aging. *Psychology and aging*, *13*(4), 531.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246-263.

Garibaldi, U., & Penco, M. A. (1985). Probability theory and physics between bernoulli and laplace: The contribution of jh lambert (1728-1777). In *Proceeding of the Fifth National Congress on the History of Physics, Rome* (Vol. 9, pp. 341-346).

Gentleman, J. F., & Wilk, M. B. (1975). Detecting outliers. II. Supplementing the direct analysis of residuals. *Biometrics*, 387-410.

Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, *76*(373), 131-139.

Giloni, A., & Padberg, M. (2002). Least trimmed squares regression, least median squares regression, and mathematical programming. *Mathematical and Computer Modelling*, *35*(9), 1043-1060.

Gladwell, M. (2008). *Outliers: The story of success*. Hachette UK.

Greene, W. H. (2000). Econometric analysis (International edition).

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, *23*(4), 519-529.

Gujarati, D.N. (2003).Basic Econometrics, MG Graw-Hill, New York.

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 761-771.

Hadi, A. S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, 393-396.

Hadi, A. S., & Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, *88*(424), 1264-1272.

Hadi, A. S., Imon, A. H. M., & Werner, M. (2009). Detection of outliers. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 57-70.

Hadi, S. (2006). Metodologi Penelitian Kuantitatif untuk Akuntansi dan Keuangan. *Yogyakarta: Ekonisia*.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), 383-393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). Linear Models: Robust Estimation. *Robust Statistics: The Approach Based on Influence Functions*, 307-341.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011).*Robust statistics: the approach based on influence functions* (Vol. 114). John Wiley & Sons.

Hampel, R. L., National Association of Independent Schools. Commission on Educational Issues, & National Association of Secondary School Principals (US). (1986).*The last little citadel: American high schools since 1940* (pp. 35-41). Boston: Houghton Mifflin.

Han, K. H., & Kim, J. H. (2004). Quantum-inspired evolutionary algorithms with a new termination criterion, H ε gate, and two-phase scheme. *Evolutionary Computation, IEEE Transactions on*, *8*(2), 156-169.

Harper H.L. (1974–1976). The method of least squares and some alternatives. Part I, II, II, IV, V, VI. International Statistical Review.

Hawkins, D. M. (1980). Identification of outliers (Vol. 11). London: Chapman and Hall.

Hawkins, S., He, H., Williams, G., & Baxter, R. (2002). Outlier detection using replicator neural networks. In *Data warehousing and knowledge discovery* (pp. 170-180). Springer Berlin Heidelberg.

Hekimoglu, S., Erenoglu, R. C., & Kalina, J. (2009). Outlier detection by means of robust regression estimators for use in engineering science. Journal of Zhejiang University SCIENCE A, 10(6), 909-921.

Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in statistics-Simulation and Computation*, *17*(2), 581-607.

Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Computer applications in the biosciences: CABIOS*, *8*(2), 189-191.

Ho, P., & Fung, D. (1991, June). Error performance of multiple symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels. In *Communications, 1991. ICC'91, Conference Record. IEEE International Conference on* (pp. 568-574). IEEE.

Hocking, R. R. (2013). *Methods and applications of linear models: regression and the analysis of variance*. John Wiley & Sons.

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, *58*(3), 54-59.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Hoerl, A. E., & Kennard, R. W. (1976). Ridge regression iterative estimation of the biasing parameter. *Communications in Statistics-Theory and Methods*, *5*(1), 77-88.

Hoerl, A. E., Kannard, R. W., & Baldwin, K. F. (1975). Ridge regression: some simulations. *Communications in Statistics-Theory and Methods*, *4*(2), 105-123.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Hotelling, H. (1957). The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, *10*(2), 69-79.

Huber, H., Lewis, S. M., & Szur, L. (1964). The Influence of Anaemia, Polycythaemia and Splenomegaly on the Relationship between Venous Haematocrit and Red-Cell Volume. *British journal of haematology*, *10*(4), 567-575.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73-101.

Huber, P. J. (1968). Robust confidence limits. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *10*(4), 269-278.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799-821.

Huber, P. J. (1981). Robust Statistics New York.

Huber, P. J. (2011). Robust statistics (pp. 1248-1251). Springer Berlin Heidelberg.

Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics, Hoboken. *NJ: Wiley. doi*, *10*(1002), 9780470434697.

Huber, S. C. (1981). Interspecific variation in activity and regulation of leaf sucrose phosphate synthetase. *Zeitschrift für Pflanzenphysiologie*, *102*(5), 443-450.

Hubert, J., Münzbergová, Z., Nesvorná, M., Poltronieri, P., & Santino, A. (2008). Acaricidal effects of natural six-carbon and nine-carbon aldehydes on stored-product mites. *Experimental and Applied Acarology*, *44*(4), 315-321.

Hulland, J. S. (1999). The effects of country-of-brand and brand name on product evaluation and consideration: A cross-country comparison. *Journal of International Consumer Marketing*, *11*(1), 23-40.

Hutcheson, G. D., & Moutinho, L. (2008). *Statistical modeling for management*. Sage.

Irving, M. R., Owen, R. C., & Sterling, M. J. H. (1978). Power-system state estimation using linear programming. *Electrical Engineers, Proceedings of the Institution of*, *125*(9), 879-885.

Jabr, R. A. (2006). Power system state estimation using an iteratively reweighted least squares method for sequential L 1-regression. *International Journal of Electrical Power & Energy Systems*, *28*(2), 86-92.

Jabr, R. A., & Pal, B. C. (2004). Iteratively reweighted least-squares implementation of the WLAV state-estimation method. In *Generation, Transmission and Distribution, IEE Proceedings-* (Vol. 151, No. 1, pp. 103-108). IET.

Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. *Applied Statistics*, 225-236.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (Vol. 4). Englewood Cliffs, NJ: Prentice hall.

Jolliffe, I. T. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics*, 300-303.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*.

Kaiser, H. F. (1991). Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychological reports*, *68*(3), 855-858.

Kendall, D. G. (1957). Some problems in the theory of dams. *Journal of the Royal Statistical Society. Series B (Methodological)*, *19*(2), 207-233.

Khalaf, G., & Shukur, G. (2005). Choosing ridge parameter for regression problems.

Khan, M. H., & Akter, S. (2009, December). Multiple-case outlier detection in least-squares regression model using quantum-inspired evolutionary algorithm. In *Computers and Information Technology, 2009. ICCIT'09. 12th International Conference on* (pp. 7-12). IEEE.

Krivulin, N. (1992). An analysis of the Least Median of Squares regression problem. In *Computational Statistics* (pp. 471-476). Physica-Verlag HD.

Laitinen, E. K. (2006). Partial least squares regression in payment default prediction. *Investment Management and Financial Innovations*, *3*(1), 64-77.

Lawless, J. F., Wang, P. (1976). A simulation study of ridge and other regression estimators. Communications in Statistics A5:307–323.

Lawrence, K & Arthur, JL. (1990), Robust Regression; Analysis and Applications; Marcel Dekker: INC.

Leroy, A. M., & Rousseeuw, P. J. (1987). Robust regression and outlier detection. *Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1987*, *1*.

Lewis-Beck, C., & Lewis-Beck, M. (2015). *Applied regression: An introduction* (Vol. 22). Sage publications.

Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*, *28*(9), 1635-1647.

Long, D. E. (1993). *Model checking, abstraction, and compositional verification* (Doctoral dissertation, The Technion).

Lorber, A., Wangen, L. E., & Kowalski, B. R. (1987). A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*, *1*(1), 19-31.

Mal, C.Z., and Y.L. Dul (2014) "Generalized shrunken type-GM estimator and its application "International Conference on Applied Sciences (ICAS2013)

Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*,*2*(1), 187-197.

Maronna, R. A. R. D., Martin, D., & Yohai, V. (2006). *Robust statistics* (pp. 978-0). John Wiley & Sons, Chichester. ISBN.

Maronna, R., Bustos, O., & Yohai, V. (1979). Bias-and efficiency-robustness of general M-estimators for regression with random carriers. In *smoothing techniques for curve estimation* (pp. 91-116). Springer Berlin Heidelberg.

Marquaridt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*(3), 591-612.

Martens, H., Naes, T., & Norris, K. H. (1987). Multivariate calibration by data compression.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, *60*(309), 234-256.

McDonald, G. C., Galarneau, D. I. (1975). A Monte Carlo evaluation of some Ridge-type estimators. Journal of the American Statistical Association 70:407–416.

McIntyre, L. (1994). Using cigarette data for an introduction to multiple regression. *Journal of Statistics Education*, *2*(1).

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics*, *34*(6), 667-677.

Mendenhall, W., and Sincich, T. (1992), *Statistics for Engineering and the Sciences* (3rd ed.), New York: Dellen Publishing Co.

Mesko, I. (1989): Discrete piecewise linear Ll approximation, L'analyse Numerique et la Theorie de L'approximation, 18, 139-145.

Midi, H., & MOHAMMED, M. A. A. (2014). Robust Latent Root Regression in the Presence of Multicollinearity and Outliers.

Midi, H., Bagheri, A., & Imon, A. H. M. (2010). The Application of Robust Multicollinearity Diagnostic Method Based on Robust Coefficient Determination to a Non-Collinear Data. *Journal of Applied Sciences*, *10*(8), 611-619.

Midi, H., Rana, M. S., & Imon, A. R. (2009). The performance of robust weighted least squares in the presence of outliers and heteroscedastic errors. *WSEAS Transactions on Mathematics*, *8*(7), 351-361.

Mili, L., Phaniraj, V., & Rousseeuw, P. J. (1991). Least median of squares estimation in power systems. *Power Systems, IEEE Transactions on*, *6*(2), 511-523.

Molina, I., Peña, D., & Pérez, B. (2009). Robust estimation in linear regression models with fixed effects.

Montgomery, D. C., & Friedman, D. J. (1993). Prediction using regression models with multicollinear predictor variables. *IIE transactions*, *25*(3), 73-85.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to linear regression analysis*. John Wiley & Sons.

Monyak, J. T. (1998). *Mean squared error properties of the ridge regression estimated linear probability model* (Doctoral dissertation, University of Delaware).

Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*.

Muniz, G., & Kibria, B. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation®*, *38*(3), 621-630.

Muthukrishnan, R. (2015). Contributions to a Study on Robust Statistics and Its Applications in Computer Vision.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (Vol. 1127). Homewood, IL: Irwin.

Noonan, R., & Wold, H. (1980). PLS Path Modelling with Latent Variables: Analysing School Survey Data Using Partial Least Squares-Part II 1, 2.*Scandinavian Journal of Educational Research*, *24*(1), 1-24.

Önder, M., & Gürer, M. A. (2001). The multiple faces of Behçet's disease and its aetiological factors. *Journal of the European Academy of Dermatology and Venereology*, *15*(2), 126-136.

Pardoe, I. (2012). *Applied regression modeling: a business approach*. John Wiley & Sons.

Pasha, G. R., & Shah, M. A. (2004). Application of ridge regression to multicollinear data. *Journal of research (Science)*, *15*(1), 97-106.

Pavlou, P. A., & Chai, L. (2002). What Drives Electronic Commerce across Cultures? Across-Cultural Empirical Investigation of the Theory of Planned Behavior. *J. Electron. Commerce Res.*, *3*(4), 240-253.

Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*(2), 559.

Pison, G., Rousseeuw, P. J., Filzmoser, P., & Croux, C. (2003). Robust factor analysis. *Journal of Multivariate Analysis*, *84*(1), 145-172.

Prescott, P. (1975). An approximate test for outliers in linear models. *Technometrics*, *17*(1), 129-132.

Rajab, J. M., MatJafri, M. Z., & Lim, H. S. (2013). Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. *Atmospheric Environment*, *71*, 36-43.

Rao, C. R., Toutenburg, H., Shalabh, H. C., & Schomaker, M. (2008). Linear models and generalizations. *Least Squares and Alternatives (3rd edition) Springer, Berlin Heidelberg New York*.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (2001). *Applied regression analysis: a research tool*. Springer Science & Business Media.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of marketing research*, *41*(3), 293-305.

Rey, G. (1983). A reason for doubting the existence of consciousness. In Consciousness and self-regulation (pp. 1-39). Springer US.

Ronald R. Hocking (2003). Methods and Application of Linear Models.

Rosa, A. J., & Horne, R. N. (1991). Automated well test analysis using robust (LAV) nonlinear parameter estimation, paper SPE 22679 presented at the 66th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers. *Dallas, Tex., Oct*, *6*(9).

Rosa, A. J., & Horne, R. N. (1995). Automated well test analysis using robust (LAV) nonlinear parameter estimation. *SPE Advanced Technology Series*, *3*(01), 95-102.

Rosen, D. H. (1999). The diagnosis of collinearity. A Monte Carlo simulation method. *Dissertation*.

Rosseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverate points. *Journal of the American Statistical Association*, *85*, 633-639.

Rosseeuw, P. J., & Yohai, V. (1984). Robust regression by means of S-estimates. *Lecture Notes in Statistics*, *26*, 256-272.

Rousseeuw, F. H. E. R. P., Hampel, F. R., Ronchetti, E. M., & Stahel, W. A. (1986). Robust statistics: the approach based on influence functions. *J. Wiley, New York*.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, *79*(388), 871-880.

Rousseeuw, P. J. (1994). Unconventional features of positive-breakdown estimators. *Statistics & Probability Letters*, *19*(5), 417-431.

Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*(3), 212-223.

Rousseeuw, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection* (Vol. 589). John Wiley & Sons.

Rousseeuw, P. J., & Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, *8*(3), 193-203.

Rousseeuw, P. J., & Van Driessen, K. (2006). Computing LTS regression for large data sets. *Data mining and knowledge discovery*, *12*(1), 29-45.

Rousseeuw, P. J., & van Zomeren, B. C. (1991). Robust distances: simulations and cutoff values. In *Directions in Robust Statistics and Diagnostics* (pp. 195-203). Springer New York.

Rousseeuw, P. J., & van Zomeren, B. C. (1992). A comparison of some quick algorithms for robust regression. *Computational statistics & data analysis*, *14*(1), 107-116.

Rousseeuw, P., & Leroy, A. (1987). Robust regression and outlier detection: Wiley Interscience. *New York*.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: a GLM approach*. Sage.

Ryan, T. P. (2008). *Modern regression methods* (Vol. 655). John Wiley & Sons.

Ryan, T. P. (2011). *Statistical methods for quality improvement*. John Wiley & Sons.

Schlossmacher, E. J. (1973). An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association*, *68*(344), 857-859.

Schumacker, R. E., Monahan, M. P., & Mount, R. E. (2002). A comparison of OLS and robust regression using S-PLUS. *Multiple Linear Regression Viewpoints*, *28*(2), 10-13.

Schumann, D. (2009). *Robust Variable Selection*. ProQuest.
Sharma, A., & Paliwal, K. K. (2007). Fast principal component analysis using fixed-point algorithm. *Pattern Recognition Letters*, *28*(10), 1151-1155.

Shihadeh, A., & Eissenberg, T. (2015). Electronic cigarette effectiveness and abuse liability: predicting and regulating nicotine flux. *Nicotine & Tobacco Research*, *17*(2), 158-162.

Simpson, J. A., Rholes, W. S., & Nelligan, J. S. (1992). Support seeking and support giving within couples in an anxiety-provoking situation: The role of attachment styles. *Journal of personality and social psychology*, *62*(3), 434.

Soliman, S. A., Christensen, G. S., & Rouhi, A. (1988). A new technique for curve fitting based on minimum absolute deviations. *Computational Statistics & Data Analysis*, *6*(4), 341-351.

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, *22*(1), 97-103.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Stigler, G. J. (1973). General economic conditions and national elections. *The American Economic Review*, *63*(2), 160-167.

Stone, M., & Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 237-269.

Strand, J. E., & Nybäck, H. (2005). Tobacco use in schizophrenia: a study of cotinine concentrations in the saliva of patients and controls. *European Psychiatry*, *20*(1), 50-54.

Sufian, A. J. M. (2005). Analyzing collinear data by principal component regression approach–An example from developing countries. *Journal of Data Science*, *3*(2), 221-232.

Sundberg, R. (1993). Continuum regression and ridge regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 653-659.

Susanti, Y., Sri Sulistijowati, H., & Pratiwi, H. (2014). Analysis of Rice Availability in Indonesia Using Multi-Dimensional Scaling. In *International Seminar on Innovation in Mathematics and Mathematics Education*. Department of Mathematics Education Faculty of Mathematics and Natural Science Yogyakarta State University.

Swallow, W. H., & Kianifard, F. (1996). Using robust scale estimates in detecting multiple outliers in linear regression. *Biometrics*, 545-556.

Tietjen, G. L., Moore, R. H., & Beckman, R. J. (1973). Testing for a single outlier in simple linear regression. *Technometrics*, *15*(4), 717-721.

Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing*, *8*(8), 527-533.

Tukey, J. W. (1953). SECTION OF MATHEMATICS AND ENGINEERING: SOME SELECTED QUICK AND EASY METHODS OF STATISTICAL ANALYSIS*. *Transactions of the New York Academy of Sciences*, *16*(2 Series II), 88-97.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to probability and statistics*, *2*, 448-485.

Tukey, J. W. (1977). Exploration Data Analysis. Wesley Canada: Addison.

Tutz, G., & Binder, H. (2007). Boosting ridge regression. *Computational Statistics & Data Analysis*, *51*(12), 6044-6059.

Wang, G., Hoffman, E. A., McLennan, G., Wang, L. V., Suter, M., & Meinel, J. (2003). Development of the first bioluminescent CT scanner. *Radiology*, *229*, 566.

Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.

Wichn, D. and Wahba (1988).Applied Regression Analysis: A Research Tool, Pacific Gross, California.

Wold, S., Ruhe, A., Wold, H., & Dunn, III, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*,*5*(3), 735-743.

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q. P., & Lillard Jr, J. W. (2014). A study of effects of multicollinearity in the multivariable analysis. *International journal of applied science and technology*, *4*(5), 9.

Yuliana, S., Hasih, P., & Sri, S. H. (2013). OPTIMASI MODEL REGRESI ROBUST UNTUK MEMPREDIKSI PRODUKSI KEDELAI DI INDONESIA. In*Prosiding Seminar Nasional Matematika dan Pendidikan Matematika (2013): Peningkatan Peran Matematika dan Pendidikan Matematika untuk Indonesia yang lebih Baik*. Jurusan Pendidikan Matematika FMIPA UNY.