ASPECTS OF MODEL DEVELOPMENT USING REGRESSION QUANTILES AND ELEMENTAL REGRESSIONS

Edmore Ranganai

Dissertation presented for the degree of Doctor of Philosophy
at Stellenbosch University

Promoter: Prof T. de Wet

Co-Promoter: Dr J.O. van Vuuren

December 2007

DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my own orignal work and I have not previously in its entirety or in part submitted it at any University for a degree.

Signature:_____

Date:_____

SUMMARY

It is well known that ordinary least squares (OLS) procedures are sensitive to deviations from the classical Gaussian assumptions (outliers) as well as data aberrations in the design space. The two major data aberrations in the design space are collinearity and high leverage. Leverage points can also induce or hide collinearity in the design space. Such leverage points are referred to as collinearity influential points. As a consequence, over the years, many diagnostic tools to detect these anomalies as well as alternative procedures to counter them were developed. To counter deviations from the classical Gaussian assumptions many robust procedures have been proposed. One such class of procedures is the Koenker and Bassett (1978) Regressions Quantiles (RQs), which are natural extensions of order statistics, to the linear model. RQs can be found as solutions to linear programming problems (LPs). The basic optimal solutions to these LPs (which are RQs) correspond to elemental subset (ES) regressions, which consist of subsets of minimum size to estimate the necessary parameters of the model.

On the one hand, some ESs correspond to RQs. On the other hand, in the literature it is shown that many OLS statistics (estimators) are related to ES regression statistics (estimators). Therefore there is an inherent relationship amongst the three sets of procedures. The relationship between the ES procedure and the RQ one, has been noted almost “casually” in the literature while the latter has been fairly widely explored. Using these existing relationships between the ES procedure and the OLS one as well as new ones, collinearity, leverage and outlier problems in the RQ scenario were investigated. Also, a lasso procedure was proposed as variable selection technique in the RQ scenario and some tentative results were given for it. These results are promising.

Single case diagnostics were considered as well as their relationships to multiple case ones. In particular, multiple cases of the minimum size to estimate the necessary parameters of the model, were considered, corresponding to a RQ (ES). In this way regression diagnostics were developed for both ESs and RQs. The main problems that affect RQs adversely are collinearity and leverage due to the nature of the computational procedures and the fact that RQs’ influence functions are unbounded in the design space but bounded in the response variable. As a consequence of this, RQs have a high affinity for leverage points and a high exclusion rate of outliers. The influential picture exhibited in the presence of both leverage

points and outliers is the net result of these two antagonistic forces. Although RQs are bounded in the response variable (and therefore fairly robust to outliers), outlier diagnostics were also considered in order to have a more holistic picture.

The investigations used comprised analytic means as well as simulation. Furthermore, applications were made to artificial computer generated data sets as well as standard data sets from the literature. These revealed that the ES based statistics can be used to address problems arising in the RQ scenario to some degree of success. However, due to the interdependence between the different aspects, *viz.* the one between leverage and collinearity and the one between leverage and outliers, “solutions” are often dependent on the particular situation. In spite of this complexity, the research did produce some fairly general guidelines that can be fruitfully used in practice.

OPSOMMING

Dit is bekend dat die gewone kleinste kwadraat (KK) prosedures sensitief is vir afwykings vanaf die klassieke Gaussiese aannames (uitskieters) asook vir data afwykings in die ontwerpruimte. Twee tipes afwykings van belang in laasgenoemde geval, is kollinearitiet en punte met hoë hefboom waarde. Laasgenoemde punte kan ook kollineariteit induseer of versteek in die ontwerp. Na sodanige punte word verwys as kollinêre hefboom punte. Oor die jare is baie diagnostiese hulpmiddels ontwikkel om hierdie afwykings te identifiseer en om alternatiewe prosedures daarteen te ontwikkel. Om afwykings vanaf die Gaussiese aanname teen te werk, is heelwat robuuste prosedures ontwikkel. Een sodanige klas van prosedures is die Koenker en Bassett (1978) Regressie Kwantiele (RKe), wat natuurlike uitbreidings is van rangorde statistieke na die lineêre model. RKe kan bepaal word as oplossings van lineêre programmeringsprobleme (LPs). Die basiese optimale oplossings van hierdie LPs (wat RKe is) kom ooreen met die elementale deelversameling (ED) regressies, wat bestaan uit deelversamelings van minimum grootte waarmee die parameters van die model beraam kan word.

Enersyds geld dat sekere EDs ooreenkom met RKe. Andersyds, uit die literatuur is dit bekend dat baie KK statistieke (beramers) verwant is aan ED regressie statistieke (beramers). Dit impliseer dat daar dus ‘n inherente verwantskap is tussen die drie klasse van prosedures. Die verwantskap tussen die ED en die ooreenkomstige RK prosedures is redelik “terloops” van melding gemaak in die literatuur, terwyl laasgenoemde prosedures redelik breedvoerig ondersoek is. Deur gebruik te maak van bestaande verwantskappe tussen ED en KK prosedures, sowel as nuwes wat ontwikkel is, is kolliniearitiet, punte met hoë hefboom waardes en uitskieter probleme in die RK omgewing ondersoek. Voorts is ‘n lasso prosedure as veranderlike seleksie tegniek voorgestel in die RK situasie en is enkele tentatiewe resultate daarvoor gegee. Hierdie resultate blyk belowend te wees, veral ook vir verdere navorsing.

Enkel geval diagnostiese tegnieke is beskou sowel as hul verwantskap met meervoudige geval tegnieke. In die besonder is veral meervoudige gevalle beskou wat van minimum grootte is om die parameters van die model te kan beraam, en wat ooreenkom met ‘n RK (ED). Met sodanige benadering is regressie diagnostiese tegnieke ontwikkel vir beide EDs en RKe. Die belangrikste probleme wat RKe negatief beinvloed, is kolliniearitiet en punte met hoë hefboom waardes agv die aard van die berekeningsprosedures en die feit dat RKe se

invloedfunksies begrensd is in die ruimte van die afhanklike veranderlike, maar onbegrensd is in die ontwerpruimte. Gevolglik het RKe ‘n hoë affiniteit vir punte met hoë hefboom waardes en poog gewoonlik om uitskieters uit te sluit. Die finale uitset wat verkry word wanneer beide punte met hoë hefboom waardes en uitskieters voorkom, is dan die netto resultaat van hierdie twee teenstrydige pogings. Alhoewel RKe begrensd is in die onafhanklike veranderlike (en dus redelik robuust is tov uitskieters), is uitskieter diagnostiese tegnieke ook beskou om ‘n meer holistiese beeld te verkry.

Die ondersoek het analitiese sowel as simulasie tegnieke gebruik. Voorts is ook gebruik gemaak van kunsmatige datastelle en standard datastelle uit die literatuur. Hierdie ondersoeke het getoon dat die ED gebaseerde statistieke met ‘n redelike mate van sukses gebruik kan word om probleme in die RK omgewing aan te spreek. Dit is egter belangrik om daarop te let dat as gevolg van die interafhanklikheid tussen kolliniariteit en punte met hoë hefboom waardes asook dié tussen punte met hoë hefboom waardes en uitskieters, “oplossings” dikwels afhanklik is van die bepaalde situasie. Ten spyte van hierdie kompleksiteit, is op grond van die navorsing wat gedoen is, tog redelike algemene riglyne verkry wat nuttig in die praktyk gebruik kan word.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the following people and institutions:

- To my promoter Prof T. de Wet (Stellenbosch University) and co-promoter Dr J.O. van Vuuren (Stellenbosch University).
- To members of the Department of Statistics and Actuarial Science (Stellenbosch University).
- The National Research Foundation for research funding.
- The Stellenbosch University Postgraduate Bursary Office for funding.
- My mother.

SOLI DEO GLORIA!

CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Background and motivation

Over the years the ordinary least squares (OLS) have become standard tools in building and analysing models. However, it is well known that OLS techniques are highly sensitive to deviations from the classical Gaussian assumptions (outliers) as well as to data aberrations in the design space. As a consequence, since the advent of OLS the list of diagnostic tools to identify these anomalies as well as procedures to develop alternative model building procedures, is ever growing.

One such class of procedures to counter deviations from the classical Gaussian assumptions is that of the Koenker and Bassett (1978) Regression Quantiles (RQs), which are natural extensions of order statistics to the linear model. Since Koenker and Bassett's pioneering 1978 paper, RQs have been further developed by them and other researchers as a powerful set of tools to deal with these problems (see *e.g.* the recent monograph by Koenker, 2005). RQs can be found as the solution to linear programming problems (LP's) and can therefore be obtained using the very efficient LP algorithms (see *e.g.*, Koenker and Park, 1996).

An elemental subset (ES) regression, which consists of a subset of observations of minimum size p, to estimate the necessary parameters of a given model, was introduced by Boscovich in 1755 and in recent years revived and further developed by *e.g.* Hawkins *et al.* (1984), Mayo and Gray (1997) as well as other researchers. However, due to ESs extreme computational demands, they have only become practical propositions during the last few years with the rapid development in computing power. ES regressions contain the set of all feasible solutions to the LP problem of which a RQ is an optimal solution. There is therefore an inherent relationship between these two sets of procedures. Also, many OLS estimators (statistics) can be expressed as weighted averages of ESs (RQs included) estimators (statistics) (see *e.g.* Hawkins, 1993). Consequently there is thus an inherent relationship between ES (RQ) estimators (statistics) and OLS estimators (statistics). It is this three-tier relationship that we aim to exploit in order to address the above mentioned issues of deviations from the classical Gaussian assumptions as well as aberrations in the design space. Deviations from the classical Gaussian assumptions which imply outliers, result in poor **prediction**. In the RQ scenario, **prediction** can only meaningfully be done using the middle RQ (L_1 estimator) which passes through the centre of the data since **prediction** becomes

poorer as the RQ hyperplane (by definition) moves away from the centre of the data. However, viewing a RQ as an ES, we will use the term **prediction** and devise cut-off values that mimic this **prediction** pattern so that extreme RQs are not classified as outlying or influential on the basis that they are extreme ones (see section 2.4.1).

Although various researchers have achieved success in the RQ arena in many fields (see Yu, Lu and Stander, 2003), there are still many unresolved issues remaining in using RQs in addressing the above issues, especially those that involve the design space. The major design space aberrations are collinearity and outliers in the predictor space. The former comprise exact linear dependencies and near exact linear dependencies amongst predictor variables (see *e.g.* Hocking and Pendleton, 1983). The latter are referred to as leverage points. RQs are fairly robust to outliers but susceptible to leverage points since their influence functions are bounded in the response space but unbounded in the predictor space. Since leverage points can also influence the eigenstructure of the regressor matrix, thereby inducing or hiding collinearity (see *e.g.*, Chatterjee and Hadi, 1986, 1988), they are also referred to as collinearity influential points. Whatever the causes of collinearity may be, it has undesirable manifestations on various regression statistics (see *e.g.*, Greene, 1990) which can be worse at the RQ level. A leverage point and/or outlier that has undesirable effects on various regression statistics (estimators) is referred to as an influential point.

A single observation that is a leverage point, an outlier or influential point, is referred to as a single case. On the other hand, observations that are leverage points, outliers or influential points jointly (in subsets of cases) are referred to as multiple cases. Multiple case diagnostics are important, since there may be situations where observations are jointly influential, but not individually. Not only is joint influence difficult to detect, it can also be more serious (see, *e.g.* Barrett and Gray, 1997a). Furthermore, single case diagnostics have been found to be ineffective in the presence of "masking" (which makes outliers appear inlying) and "swamping" (which makes inliers appear outlying) (see, *e.g.*, Rousseeuw and van Zomeren, 1990). Due to these phenomena an observation which is a single case may cease to be harmful at the multiple case level (and hence, RQ level) and vice versa. In order to deal with these problems, several procedures have been proposed to identify multiple cases (see, *e.g.*, Cook and Weisberg, 1982; Gray and Ling, 1984; Barrett and Gray, 1992, 1995; Hadi and Simonoff, 1993). However, all these multiple case procedures are not necessarily aimed at subsets of cases of size p corresponding to RQs.

In this thesis we focus on RQs which are the solutions to the LP problems corresponding to specific ESs (subsets of size p). As a consequence, if the p observations are jointly leverage points, outlying or influential they can therefore be viewed as multiple cases. Note that the RQs multiple cases are slightly different from other multiple cases in the sense that we are only concerned with specific multiple cases which are ESs of size p corresponding to RQs.

In order to have a holistic view of the diagnostics and model building in the RQ arena, analytical tools for collinearity, leverage, outlier and influential diagnosis and model selection still need to be addressed. One practical problem that usually arises is the determination of the size of the influential set. However, since RQs are specific ESs of size p, it is more convenient and natural to consider multiple cases of size $n-p$ since these remaining observations can be used to construct predictive validation statistics. As a consequence, the primary aim of this thesis is to contribute to RQs diagnostics by extending the usual OLS diagnostics to the RQ scenario using the ES regression procedures.

The main objectives are:

- To further explore the properties of the three-tier relationships amongst OLS statistics, RQs and ESs statistics.
- To investigate the properties of the elemental regression weight (ERW) (2.2.1) since it is the vehicle through which OLS statistics are related to ESs (RQs) statistics.
- To develop RQ based diagnostics by extending the existing OLS regression diagnostic techniques.
- To investigate the properties of the determined RQ statistics and procedures using analytic means as well as simulation and application to artificial and standard data sets from the literature.

We outline the major contributions of this thesis with regard to these objectives in the following section.

1.2 Thesis contributions

This thesis contributes to the understanding of the three-tier relationship amongst ESs, RQs and the OLS procedure and the use of ES procedures in addressing the problems that affect RQs. These problems comprise collinearity, leverage and "prediction". Specifically the following are addressed;

- The ERW can reveal vital information on the two major design space aberrations *viz.*, collinearity and leverage.
- Considering only a leverage view of the ERW, it is shown that the ERW is involved in many generalized OLS (multiple case) statistics.
- Using artificial data sets, it is shown that ERWs associated with RQs are often large, especially if leverage points are present in the design matrix.
- Further results which relate OLS single case leverage statistics to the RQ (ES) multiple case predictive leverage statistics via the ERW, are deduced.
- Based on one of the leverage results, we propose a RQ multiple case predictive weighted leverage statistic and determine its cut-off value using simulation studies.
- We correct the original result of Hawkins *et al.* (1984), which relates the OLS single case residuals sum of squares statistic to the elemental predicted residual (EPR) sum of squares.
- We use the EPR sum of squares as a RQ "prediction" measure and determine its cut-off value using simulation studies, both based on the sinusoidal model as well as applying a robust loss function to the RQ predicted residuals.
- We extend the single case covariance ratio to the RQ scenario as an influence measure.
- Lastly, we give some areas of further research which include, *inter alia*, variable selection. We further propose using the lasso shrinkage procedures as variable selection procedures in the RQ scenario and give some tentative results based on them.

We now discuss these points in more detail.

Although it is well known that OLS statistics can be expressed as weighted averages of ES statistics (see section 2.2), a holistic picture of the relationship between the OLS statistics and ES statistics has not been fully exposed in published research. As our point of departure we investigated the ERW since it is the vehicle through which ES statistics relate to OLS

statistics. The ERW is based on the predictor matrix information. We show both the collinearity (of the predictor matrix) view and the leverage (variability) view of this statistic. In subsection 4.2.2 we give Theorem 4.1 and its proof based on matrix algebra. Also, for interest sake, we give another proof in the appendix based on the principle of mathematical induction. The consequence of this theorem is that the ERW can expressed as a product of a constant and two factors, *viz.*, the collinearity component and the variability component. It is shown that the collinearity component involves various usual OLS collinearity diagnostics such as the determinant of the correlation matrix, the variance inflation factor, *etc*. On one hand, the collinearity component can be viewed as the ratio of the degree of collinearity at the RQ/ES level to that at the full design matrix. On the other hand, the variability component can be viewed as the ratio of the variability at the RQ/ES level to that at the full design matrix.

We illustrate the dynamics between the variability view and the leverage view using artificial data sets in Chapter 5 (see section 5.4.1). These data sets consist of collinearity influential points (see section 4.6), *i.e.*, type A leverage points which induce collinearity and type B leverage points which hide it. Both scenarios result in a large ERW due to the fact that RQs have a high affinity for leverage points hence they tend to include them. However, in the presence of type A leverage points (collinearity influential points) the ERW is often relatively smaller due to a smaller collinearity factor.

We give Theorem 4.2 in section 4.4 and use it to show various multiple leverage views of the ERW. Also, we show that the ERW can be viewed as an extension of the complement of OLS single case leverage if $n = p + 1$. Actually, we show that the ERW appears in many generalized leverage and influential regression diagnostics, *e.g.*, the multiple case version of the Cook's distance (see Cook, 1977).

In section 4.5 we give Theorem 4.3 (and its proof) which relates single case leverage to RQ (ES) multiple case leverage via the ERW. This theorem consists of three results. The first result in item (i) was given by Hawkins *et al.* (1984) while the other two results are derived in this thesis. The last result in item (iii) was mainly made use of to derive the RQ/ES multiple case predictive leverage statistic, which is an analogue of a single case leverage statistic. Actually, we show that the OLS single case leverage statistic is a weighted average of the ES (RQs included) multiple case predictive leverage statistics. RQ multiple case predictive leverage is often small due to the fact that RQs tend to include leverage points in them (rather than predicting them). As a consequence the size of the RQ multiple case leverage statistic is predominantly determined by the size of the ERW. Using the fact that the ERWs sum to 1, we

illustrate using the artificial data sets the contribution of RQ statistics to OLS statistics. Although the number of ESs corresponding to RQs is substantially smaller (their proportion is extremely small) than the total number of ESs, the proportion of the ERWs corresponding to RQs, can be very large compared to those corresponding to their complement. Therefore RQs can contribute much more to OLS statistics than their complement.

We had originally suggested the extension of the single case leverage cut-off value to the RQ scenario. However, due to the fact that the total number of ESs is usually extremely large compared to the number of ESs corresponding to RQs, and the fact that RQs have a high affinity for leverage points (which results in large ERWs), the RQ multiple leverage cut-off value's direct analogue of the single case of leverage cut-off value is practically too small as exhibited by the artificial data sets in section 5.4. Therefore, in section 5.8 we determined reasonable cut-off values using a simulation study. The leverage picture exhibited was that the number of RQs being flagged increases as the number of leverage points increases (approaches p). This is so because RQs have a high affinity for leverage points and therefore they tend to include them in the corresponding ESs. As a consequence, if the number of leverage points is close to p, all the leverage points are likely to be included, implying that the ESs corresponding to RQs contain almost the same design matrix information.

In section 4.7 we give Theorem 4.4 which consists of two results. Both of these results were originally given by Hawkins *et al.* (1984). These relate OLS residuals and residual sum of squares to the elemental predicted residuals (EPR's) and EPR's sum of squares, via the ERW, respectively. However, the second result which expresses the OLS residual sum of squares as a weighted sum of EPR's sum of squares, is incorrect in the Hawkins *et al.* (1984) paper. We correct this result and it is the one that we mainly make use of in this thesis to study RQ multiple case outliers.

The single case cut-off is based on the ratio of the EPR's sum of squares to that of the OLS residual sum of squares. It is not practically reasonable to extend the analogy of the single case predicted residuals sum of squares' cut-off value to the EPR's sum of squares as one would be forced to compute the whole set of ESs. Also, it is clear that extreme RQs exhibit poor "prediction" compared to the "middle" ones. Therefore, in section 5.9 we use the robust loss function (see Ronchetti, Field and Blanchard, 1997) that bounds the influence function of the RQs in the response space as well as simulation studies using the sinusoidal model to

determine reasonable cut-off values for the RQ predicted residual sum of squares.

Although RQs are fairly robust to outliers we need to have an outlier diagnostic component in order to get a holistic influence picture, since regression influence diagnostics normally comprise both the leverage component and the outlier component. In section 4.8 we show that influence measures which are volumes of confidence ellipsoids such as the covariance ratio (see, Belsley *et al.*, 1980) generalise into a product of a factor that is a function of the ERW and another one that is a function of the ratio of the EPR's sum of squares to that of the OLS residual sum of squares. Using the leverage (ERW) cut-off values and the EPR's sum of squares cut-off values, we deduce the cut-off values of the RQ multiple case covariance ratio as an influence measure. The influence picture exhibited reveals that RQs are more adversely affected by leverage points under the normal distribution. Actually, the number of RQs flagged becomes less and less as the error distribution becomes heavier. This is due to the fact that RQs have influence functions that are bounded in the response variable but unbounded in the predictor variable, hence they have a high exclusion (repulsion) of outliers and a high affinity for leverage points. Some points can be both leverage points and outliers. In this case the resulting influence picture will be a trade-off between these two antagonistic forces. Actually, it has been observed that RQs may not be affected by leverage points to the same degree as the OLS estimators (see *e.g.*, Koenker and Hallock, 2001). But here we show, using a simulation study, that this may be attributable to the trade-off between the RQs affinity for leverage points and their exclusion of outliers (see section 5.10). So the researcher needs to take note of the underlying error distribution in the presence of leverage points in the design matrix.

Most regression model selection techniques, both OLS and robust procedures, involve some estimate of the variance, *e.g.*, the Mallows C_p statistic and its robust version (see Mallows, 1973 and Ronchetti *et al.*, 1997 respectively). However, we could not use procedures that involve some estimate of the variance when using an ES procedure since they exhibit the exact fit property which results in the estimate of the variance being zero (no degrees of freedom to measure error). More recently, model selection based on cross-validation has been found to be more appealing in many regression scenarios. This procedure could be adopted in the face of collinearity by employing the lasso RQ shrinkage technique. The lasso shrinkage technique in general was first proposed by Tibshirani (1996). However, in using the lasso penalty there is the added advantage that it ties in nicely with the linear programming

structure of RQs. Also, having obtained the ES corresponding to RQs, we employ the OLS and a lasso penalty using cross-validation on p observations present in the ESs corresponding to RQs. This procedure and the lasso procedure do not always select the same model. Actually, the lasso RQ procedure selects the same model over a number of RQ levels while the OLS plus a lasso penalty procedure is more likely to select a different model at a different RQ level.

In the literature, the ordinary ridge regression procedure is shown to be ineffective in the presence of collinearity influential points (see *e.g.* Mason and Gunst, 1985). However, applications of the lasso procedures in the RQ scenario show that they have the potential to be effective in the presence of collinearity influential points. We discuss these aspects in Chapter 7 and point out RQ variable selection using ES procedures as a potential area of further research, amongst others.

1.3 Thesis layout

Chapter 2 gives an overview of the development of ES regressions tracing as far back as 1755, before the advent of least squares. Also, we include the beginning of a renewed interest in recent years (see *e.g.*, Mayo and Gray, 1997) up to present day and their many applications.

In Chapter 3 we delve into the RQ literature. Also, an overview of the available computational software and some areas of their applications are given. We also give a number of other related regression estimators.

In Chapter 4 we give an overview of the various regression diagnostics in the literature and also develop new RQ multiple diagnostics based on ES procedures. These include collinearity, leverage, outlier and influential diagnostics. The lasso RQ procedure based on cross-validation is also proposed as a possible variable selection method. This is also further discussed in Chapter 7.

In Chapter 5 we investigate the properties and the cut-off values of the determined RQs statistics using artificial data sets as well as simulation studies.

In Chapter 6 we investigate the effectiveness of the cut-off values of the different RQs statistics using some standard data sets from the literature. Also, we show that the lasso RQ procedure fails in the presence of collinearity influential points using these data sets.

In Chapter 7 we give conclusions and some proposed areas for further research.

1.4 Notation

This section introduces the notation used throughout this thesis to serve as a quick reference for the reader.Vectors and matrices are denoted using bold faced letters.
Tables 1.1 to 1.3 contain the general, ESs and RQs notations respectively.

Table 1.1: General notation

SYMBOL/EXPRESSION	DESCRIPTION
N	Number of simulation replicates.
n	Sample size.
p	Number of predictors including the constant term.
$\mathbf{1}_n$	A vector of ones (constant term predictor).
X_j	The j^{th} predictor.
$\mathbf{X}$	An $n\times(p-1)$ design matrix without the constant predictor.
$\mathbf{X}^s$	An $n\times(p-1)$ design matrix without the constant predictor, standardized to correlation form.
$\tilde{\mathbf{X}}$	An $n\times p$ design matrix obtained by augmenting $\mathbf{1}_n$ with $\mathbf{X}$ (*i.e.* including the constant term predictor).
$\tilde{\mathbf{X}}_{(i)}$	$\tilde{\mathbf{X}}$ with the i^{th} row deleted.
$\tilde{\mathbf{X}}^s$	An $n\times p$ design matrix obtained by augmenting $\mathbf{1}_n$ with $\mathbf{X}^s$.
$\mathbf{x}_i'$, $\mathbf{x}_i^{s\prime}$, $\tilde{\mathbf{x}}_i^{s\prime}$	The i^{th} rows of $\mathbf{X}$, $\mathbf{X}^s$ and $\tilde{\mathbf{X}}^s$ respectively.
$\mathbf{X}_j$, $\mathbf{X}_j^s$	The j^{th} column vectors of $\mathbf{X}$ and $\mathbf{X}^s$

	respectively.
$\mathbf{H}$	The projection (Hat) matrix, $\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$ based on $\tilde{\mathbf{X}}$.
h_i	The diagonal elements of $\mathbf{H}$.
$\mathbf{H}_{(i)}$	The predictive Hat (projection) matrix, $\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}_{(i)}\tilde{\mathbf{X}}_{(i)}\right)^{-1}\tilde{\mathbf{X}}'$.
$h_{(i)}$	The diagonal elements of $\mathbf{H}_{(i)}$.
Y_i	The i^{th} response observation.
$\mathbf{Y}$	The response vector.
β_0	The intercept term based on $\tilde{\mathbf{X}}$.
$\boldsymbol{\beta}$	The slope coefficient.
β_0^s	The transformed intercept term based on $\tilde{\mathbf{X}}^s$.
$\boldsymbol{\beta}^s$	The transformed slope coefficient based on $\tilde{\mathbf{X}}^s$.
$\tilde{\boldsymbol{\beta}}$	$\left(\beta_0 \quad \boldsymbol{\beta}'\right)'$.
$e_i, 1 \le i \le n$	OLS residuals.
SSE	Residual sum of squares.
$\mathbf{C}$	The correlation matrix of $p-1$ non constant predictors.
$\lvert\mathbf{C}\rvert$	Determinant of the correlation matrix.
$\mathbf{X}_{(l)}$	Design matrix $\mathbf{X}$ with the l^{th} predictor deleted.

$R^2_{X_l\|\mathbf{x}_{(l)}}$	The coefficient of determination of X_l on the remaining variables, *viz.*, $\mathbf{X}_{(l)}$.
VIF_j	The variance inflation factor of the j^{th} predictor.
λ_j	The j^{th} eigenvalue such that $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_j \leq ... \leq \lambda_{p-1}$.
$\mathbf{u}_j$	The eigenvector corresponding to λ_j.
$s_j'^2$	The squared deviation from the mean of the j^{th} predictor.
$TOL_{X_l\|\mathbf{x}_{(l)}}$	The **tolerance** of the l^{th} predictor X_l on the remaining variables $\mathbf{X}_{(l)}$.
$(1-\alpha)N(0,1)+\alpha N(0,\sigma^2) \equiv CN\left(\alpha,\sigma^2\right)$	Contaminated normal distribution.
F	The distribution function (df).

Table 1.2: Elemental subset regression notation

SYMBOL/EXPRESSION	DESCRIPTION
K	Total number of elemental subsets=$\binom{n}{p}$.
$\tilde{\mathbf{X}}_J$, $\tilde{\mathbf{X}}_I$	A $p \times p$ nonsingular submatrix of the design matrix $\tilde{\mathbf{X}}$ and the $(n-p)\times p$ complement (submatrix) of $\tilde{\mathbf{X}}_J$ respectively.

$\mathbf{Y}_J$, $\mathbf{Y}_I$	A $p \times 1$ sub vector of the response vector and the $(n-p)\times p$ complement (sub vector) of $\mathbf{Y}_J$ respectively.
$\sum_{J \ni i}$, $\sum_{J \not\ni i}$	Summing over all ESs containing observation i and summing over all ESs not containing observation i respectively.
$\sum_{i \in J}$, $\sum_{i \notin J}$	Summing over all observations contained in ES J and summing over all observations not contained in ES J respectively.
$\hat{\tilde{\boldsymbol{\beta}}}_J$	The LS estimator obtained using $(\mathbf{Y}_J, \tilde{\mathbf{X}}_J)$.
e_{jJ} (EPR)	The j^{th} residual based on the fit using elemental set J, the elemental predicted residual.
PRESS	The leave one out usual predicted residual sum of squares, *PRESS*.
$PRESS_J$	The elemental "predicted" residual sum of squares which is the analogue of the leave one out usual predicted residual sum of squares, *PRESS*.
$\mathbf{H}_J$, $\mathbf{H}_{IJ}$ and $\mathbf{H}_I$	The matrices $\tilde{\mathbf{X}}(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}'$, $\tilde{\mathbf{X}}_I(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}_I'$ and $\tilde{\mathbf{X}}_I(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}_I'$ respectively.
$h_{jJ} \equiv R_{jJ}$ (Hawkins' notation)	The diagonal elements of $\tilde{\mathbf{X}}(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}'$ with $h_{iJ} = R_{iJ} = 1, \ \textit{for } i \in J$ and $h_{jJ} = diag\{\mathbf{H}_{IJ}\}, \ \textit{for } i \notin J$, the ES predictive leverage.
T_J	The weighted ES predictive leverage.

ω_J	The elemental regression weight, ERW.
$\mathbf{C}_J$	The correlation matrix of the $p-1$ predictors for **elemental set** J.
$s'^2_{J,j}$	The squared deviation from the mean of the j^{th} predictor in the elemental set J.
γ_J	The variability factor in ERW.
ρ_J	The collinearity factor in ERW.
$VIF_{l\,J}$	The variance inflation factor of the l^{th} predictor in the J^{th} elemental set.
$SSE_{X_j \mid X_1 \ldots X_{j-1}}$	The residual sum of squares from the regression of X_j on $X_1,\ldots,X_{j-1}$.

Table 1.3: Regression quantile notation

SYMBOL/EXPRESSION	DESCRIPTION
τ	Regression Quantile level: $0 < \tau < 1$.
q_τ	The τ^{th} sample quantile.
ρ_τ	A robust loss (check) function.
$Q_{Y\mid\mathbf{x}}$	The conditional quantile function of Y given the covariate $\mathbf{x}$.
RQ	Regression Quantile.
$\boldsymbol{\beta}(\tau) = \left[\beta_0 + F^{-1}(\tau), \beta_1, \beta_2, \ldots, \beta_p\right]'$	The τ^{th} RQ parameter.
$\hat{\boldsymbol{\beta}}_{TM}(\alpha)$, $0 < \alpha < 1$	The $100\alpha\%$ regression trimmed mean estimator.
RRQ	Restricted RQ.
$\hat{\boldsymbol{\beta}}^{(w)}(\tau)$	Bounded influence RQs (BI-RQs).

$\hat{\boldsymbol{\beta}}_{TM}^{(w)}(\alpha)$, $0 < \alpha < 1$	The 100α % bounded influence regression trimmed mean.
λ	The lasso shrinkage parameter.
$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau (Y_i - \beta_0 - \mathbf{x}_i' \boldsymbol{\beta})$	LP solution giving regression quantiles.
$\hat{\boldsymbol{\beta}}^{(\lambda)}(\tau) = \arg\min_{\beta_0, \boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \rho_\tau (Y_i - \beta_0 - \mathbf{x}_i' \boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} \lvert \beta_j \rvert \right\}$	The lasso regression quantile solution.

CHAPTER 2 ELEMENTAL SUBSET REGRESSION

2.1 Introduction

In this chapter we give a brief overview of elemental regression results in the literature. We will use and further develop some results based on elemental subsets in later chapters.

As a starting point, we consider the usual linear regression model,

$$\mathbf{Y} = \mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

with

$$\mathbf{Y}: n\times 1,\ \mathbf{X}: n\times(p-1), \mathbf{1}_n: n\times 1,\ \beta_0: \text{a constant}, \boldsymbol{\beta}: (p-1)\times 1, \boldsymbol{\varepsilon}: n\times 1,$$

where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$ and $\mathbf{1}_n$ is the vector of ones. Let

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{1}_n & \mathbf{X} \end{pmatrix}, \text{ and } \tilde{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 & \boldsymbol{\beta}' \end{pmatrix}'.$$

Partition $\tilde{\mathbf{X}}$ and $\mathbf{Y}$ as

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{X}}_J \\ \tilde{\mathbf{X}}_I \end{pmatrix} \text{ and } \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_J \\ \mathbf{Y}_I \end{pmatrix}$$

with $\mathbf{Y}_J : p\times 1$ and $\tilde{\mathbf{X}}_J : p\times p$. Without loss of generality $\begin{pmatrix} \tilde{\mathbf{X}}_J & \mathbf{Y}_J \end{pmatrix}$ can be viewed as the **elemental set.** The J^{th} **elemental regression** is obtained as

$$\hat{\tilde{\boldsymbol{\beta}}}_J = \left(\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J\right)^{-1} \tilde{\mathbf{X}}_J' \mathbf{Y}_J = \tilde{\mathbf{X}}_J^{-1} \mathbf{Y}_J, \qquad (2.1.1)$$

where $\tilde{\mathbf{X}}_J$ is square and assumed to be nonsingular. Thus the J^{th} **elemental regression** consists of a subset of observations of minimum size to estimate the necessary parameters of the above model. Let the total number of elemental subsets be denoted by

$$K = \binom{n}{p},$$

which increases rapidly when both n and p (as p approaches $\frac{n}{2}$) increase and hence, the computational load.

Early methods of regression estimation were based on combining the results of these so called elemental regressions, *e.g.*, Boscovich in 1755, 50 years before the advent of least squares, used this approach to estimate $\tilde{\boldsymbol{\beta}}$. He and Maire (see Mayo and Gray, 1997) were attempting to find the length of the median arc near Rome. During these early years such an approach

was severely limited because of the computational load. This problem and the advent of least squares resulted in this approach losing acceptance (see Sheynin, 1973; Stigler, 1986 for more details). Approximately two centuries later, Theil (1950) and Sen (1968) described simple linear regression estimators of both slope and intercept based on the elemental regressions. However with the computer power available today there has been renewed interest in the use of elemental regressions.

More recently elemental regressions have been proposed as a computational device to approximate estimators in areas of high breakdown regression and multivariate location/scale estimation, *e.g.*, the least median of squares (LMS) estimator (Hampel, 1985), the least trimmed squares (LTS) estimator (Rousseeuw, 1984), the best elemental estimator (BEE) (Hawkins, 1993), the least trimmed absolute deviations (LTA) estimator (Hössjer, 1994), the least quantile differences (LQD) estimator (Croux, Rousseeuw, and Hössjer, 1994) and the regression depth (RD) estimator (Rousseeuw and Hubert, 1997). In these estimators the criterion functions are not convex but multimodal and therefore not amenable to standard iterative methods. Hawkins (1993) and Hawkins and Olive (1999, 2002) investigated the accuracy of elemental set approximations for regression and showed that they provide excellent approximations for LMS, LTS, and ordinary least squares criteria.

Primarily elemental set methods were developed to provide an estimator for $\tilde{\boldsymbol{\beta}}$ but the idea has been extended to handle other regression problems as we will discuss subsequently, *e.g.*, outlier problems.

In the following section we will firstly discuss the relationship between elemental set methods for estimating $\tilde{\boldsymbol{\beta}}$ and some related estimators for $\tilde{\boldsymbol{\beta}}$. In section 2.3 the extension of the idea of elemental sets to handling outlier problems in multiple regression and in section 2.4 the statistics based on elemental regressions are considered. The last section gives some concluding remarks.

2.2 Relationship between elemental sets and some regression estimators

Many regression estimators can be expressed in terms of elemental regressions (see Hawkins, Bradu and Kass, 1984; Hawkins, 1993; Mayo and Gray, 1997). Here we will discuss estimators based on two broader classes of estimators, *viz.*, the least squares (least squares based) and leverage-residual weighted elemental estimators (LRWE), which encompasses regression quantile (regression quantile based) estimators. In the following subsection we

discuss least squares estimators.

2.2.1 Least squares estimators

The **elemental regression weight is** defined by

$$\omega_J = \frac{\left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right|}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|} = \frac{\left|\tilde{\mathbf{X}}_J\right|^2}{\sum_J\left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right|} = \frac{\left|\tilde{\mathbf{X}}_J\right|^2}{\sum_J\left|\tilde{\mathbf{X}}_J\right|^2}, \tag{2.2.1}$$

where $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$, $0 \le \omega_J \le 1$ and the summation is over all the elemental sets. These weights play a pivotal role in the construction of least squares estimators. The third form of the elemental regression weight is obtained by invoking the Cauchy-Binet Theorem given in the Appendix A.

Jacobi, in 1841 showed that the least squares estimators can be expressed as weighted averages of the elemental regressions (see Sheynin, 1973; Hoerl and Kennard, 1980). These weighted averages of the elemental regressions include the ordinary least squares (OLS) estimator as well as weighted least squares estimators.

In terms of the weights ω_J, the OLS estimator of $\tilde{\boldsymbol{\beta}}$ is given by

$$\hat{\tilde{\boldsymbol{\beta}}}_{OLS} = \sum_J \omega_J \hat{\tilde{\boldsymbol{\beta}}}_J. \tag{2.2.2}$$

Since $0 \le \omega_J \le 1$ and $\sum_J \omega_J = 1$, it follows that $\hat{\tilde{\boldsymbol{\beta}}}_{OLS}$ is a weighted average of the elemental regressions.

Based on the weights $\mathbf{V} = diag\left(v_1, v_2, ..., v_n\right)$, the weighted least squares estimator is given by

$$\hat{\tilde{\boldsymbol{\beta}}}_{WLS} = \left(\tilde{\mathbf{X}}'\mathbf{V}^{-1}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{V}^{-1}\mathbf{Y}. \tag{2.2.3}$$

It can be shown that

$$\hat{\tilde{\boldsymbol{\beta}}}_{WLS} = \frac{\sum_J\left|\tilde{\mathbf{X}}_J'\mathbf{V}_J\tilde{\mathbf{X}}_J\right|}{\left|\tilde{\mathbf{X}}'\mathbf{V}\,\tilde{\mathbf{X}}\right|}.\hat{\tilde{\boldsymbol{\beta}}}_J. \tag{2.2.4}$$

(See Mayo and Gray, 1997 and Chapter 4, where we construct this estimator using the elemental weight (2.2.1).)

In the next subsection other variants of weighted least squares based on leverage and residuals are presented.

2.2.2 Leverage-residual weighted elemental estimators (LRWE)

LRWE were proposed by Mayo and Gray (1997) and have the general form

$$\hat{\tilde{\beta}}(\lambda,\rho)=\frac{\sum_J \lambda_J.\rho_J.\hat{\tilde{\beta}}_J}{\sum_J \lambda_J.\rho_J} \qquad (2.2.5)$$

where

- λ_J is a weight function based on leverage, and

- ρ_J is a weight function based on the residual (degree of fit information).

Examples of LRWE estimators are :

1. For $\lambda_J = \left|\tilde{X}_J'\tilde{X}_J\right|$ and $\rho_J = 1,\ \forall J$, then $\hat{\tilde{\boldsymbol{\beta}}}(\lambda,\rho)=\hat{\tilde{\boldsymbol{\beta}}}_{OLS}$.

2. If all the weight is given to one elemental set that satisfies an appropriate fitting criterion, the elemental regression can be:

   (i) BEE (Best Elemental estimator) (Hawkins, 1993).

   (ii) Least Absolute Deviation (LAD) estimator or L_1 (see, *e.g.*, Barrodale and Roberts, 1973, 1974).

   (iii) Regression quantiles (RQs) (Koenker and Bassett, 1978).

Remark: Koenker and Bassett (1978) generalized the concept of a quantile from the univariate case to the regression case by defining a τ^{th} regression quantile (RQ) $\hat{\boldsymbol{\beta}}(\tau),\ 0<\tau<1,$ (see Chapter 3, section 3.1-3.3). This estimator is related to the L_1, the trimean (Koenker and Bassett, 1978) and the regression trimmed mean (Ruppert and Carroll, 1980) estimators as follows:

- The L_1 estimator is equivalent to $\hat{\boldsymbol{\beta}}(0.5)$, the middle RQ. In Chapter 3 we will see that RQs (and thus $\hat{\boldsymbol{\beta}}(0.5)$) can be obtained as solutions to linear programming (LP) problems.

- The trimean estimator is a linear combination of regression quantiles $(\tau = 0.25,\ 0.5,\ 0.75)$ and therefore they are also functions of elemental regressions.

- For the relationship between $\hat{\boldsymbol{\beta}}(\tau)$ and the regression trimmed mean, see the remark in section 3.2.

3. TEE (Trimmed Elemental Estimators) trim out those elemental regressions that poorly fit the full data and/or have extremely high or low leverage. The weight functions for TEE are:

- $\rho_J = \begin{cases} 1, & \textit{condition } 1 \\ 0, & \textit{otherwise,} \end{cases}$

  where *condition* 1 is a criterion based on a function of residuals, $g(e_J)$. For example, an elemental regression might be trimmed out if $g(e_J)$ is greater than or less than a specified value R.

- $\lambda_J = \lambda(J),$

  where $\lambda(J)$ is solely based on X information ("leverage"), which can be interpreted to be dispersion of the rows (observations) in $\mathbf{X}_J$.

In the literature the proposals of *condition* 1 are numerous, with $g(e_J)$ taking the forms, $\sum_{i=1}^{n} e_{iJ}^2$ or $\sum_{i=1}^{n} \left| e_{iJ} \right|$ amongst others, while $\lambda(J)$ is usually equal to $\left| \tilde{X}_J' \tilde{X}_J \right|$ (see Mayo and Gray, 1997).

In the next section we discuss the extension of elemental set methods to handling outlier problems and diagnostics based on elemental sets.

2.3 Handling outlier problems in multiple regression based on elemental sets and statistics based on elemental sets

Although elemental set methods were initially intended to provide an estimator for $\tilde{\boldsymbol{\beta}}$, the idea was later extended to handle outlier problems in multiple regression. This was first employed independently by Rousseeuw (1984) and by Hawkins, Bradu and Kass (1984). The latter proposed a robust method giving two summary statistics: an unweighted median, which is of bounded influence, and a weighted median, which is more efficient but less robust. This method, as a byproduct yields useful information on the influence (or leverage) of cases and

mutual masking (which makes outliers appear inlying) of high leverage points.

Elemental sets also arise naturally in the diagnostics of OLS, as we will discuss in section 2.4. That section also covers an extensive exploration of statistics based on elemental regressions.

2.4 Statistics based on elemental regressions

Statistics derived from elemental regressions are either based on the training set (elemental set) or on its complement (the validation set). Those derived from the training set are referred to as **internal statistics** while those based on the validation set are referred to as **external statistics**. We use J to index the elemental set and I its complement.

The number of useful statistics from an elemental regression are (not surprisingly) few, because of the exact fit property, *i.e.*

i) $\hat{Y}_{jJ} = Y_j$ for $j \in J$,

ii) $e_{jJ} = 0$ for $j \in J$,

where $e_{jJ} = \hat{Y}_{jJ} - \tilde{\mathbf{x}}_j' \hat{\tilde{\boldsymbol{\beta}}}_J$ is the j^{th} residual based on the fit using elemental set J,

iii) $SSE_J = 0$,

where SSE_J is the sum of squares of the elemental residuals,

iv) $h_{jJ} = 1$ for $j \in J$,

where h_{jJ} is the diagonal element of $\mathbf{H}_J = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J)^{-1} \tilde{\mathbf{X}}'$,

v) $R_J^2 = 1$,

where R_J^2 is the coefficient of determination from the elemental fit.

Apart from the elemental regression estimate $\hat{\tilde{\boldsymbol{\beta}}}_J$, the useful statistics are the **external statistics**. In the next subsection it will be shown that these external statistics are related to the information in $\tilde{\mathbf{X}}_J$ via the **elemental regression weight** ω_J as defined in (2.2.1). This formulation of ω_J leads to the crude conclusion that $\omega_J = 0$ when $\tilde{\mathbf{X}}_J$ is singular, while it is "large" when the design has large dispersion in the $\tilde{\mathbf{X}}_J$ space (the determinant can be viewed as a measure of volume and thus of dispersion). We devote the following subsection to external statistics since they give the bulk of useful statistics.

2.4.1 External statistics (based on the validation set)

The important statistics for the detection of influential subsets are based on the validation set. These statistics are usually based on the residuals (the response's degree of outlyingness), defined in (i) and leverage (the predictor's degree of outlyingness), defined in (ii). The original results were proved in Hawkins *et al.* (1984).

(i) **Residuals**

An **elemental predicted residual (EPR)** is defined as

$$e_{iJ} = Y_i - \tilde{\mathbf{x}}_i' \hat{\tilde{\boldsymbol{\beta}}}_J \quad i \in I \ \ (i.e.\ i \notin J),$$

and the **elemental predicted residual sum of squares** as

$$PRESS_J = \sum_{i \in I} e_{iJ}^2.$$

Here there is no harm in summing over all the $i's, 1 \le i \le n$ since $e_{iJ} = 0$ for $i \in J$.

The OLS residuals can be expressed as a sum of weighted predicted residuals, *viz.*,

$$e_i = \frac{\sum_J \left|\tilde{\mathbf{X}}_J\right|^2 e_{iJ}}{\sum_J \left|\tilde{\mathbf{X}}_J\right|^2}, 1 \le i \le n.$$

From (2.2.1) this can clearly be written as

$$e_i = \sum_{J \not\ni i} \omega_J e_{iJ}, 1 \le i \le n.$$

The OLS error sum of squares can be written as

$$SSE = \frac{\sum_J \left|\tilde{\mathbf{X}}_J\right|^2 \sum_{i \in I} e_{iJ}^2}{(p+1)\sum_J \left|\tilde{\mathbf{X}}_J\right|^2}$$

(see Theorem 4.4 in section 4.7).

Furthermore, using (2.2.1) this can be written as

$$SSE = \frac{\sum_J \omega_J . PRESS_J}{p+1},$$

However, we have corrected the original result by dividing by a factor of $p+1$.

Remark: In the case when $n = p+1$, the **elemental predicted residuals** become the usual OLS **predicted residuals** given by

$$e_{(i)} = Y_i - \tilde{\mathbf{x}}_i' \hat{\tilde{\boldsymbol{\beta}}}_{(i)} = \frac{e_i}{1-h_i},\ 1 \le i \le n$$

and the usual **predicted residual sum of squares** as

$$PRESS = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} \left(\frac{e_i}{1-h_i} \right)^2,$$

where the subscript notation "(i)" indicates the deletion of the i^{th} observation and h_i is as defined in (ii) below.

(ii) **Leverage and Residual Freedom**

The projection (hat) matrix $\mathbf{H} = \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}'$ and its variants play a very important role in leverage diagnostics, as we will now briefly discuss. A diagonal element of $\mathbf{H}$ is denoted by

$$h_i = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{x}}_i,$$

which can be thought of as the amount of leverage of the response value Y_i on the corresponding value $\hat{Y}_i$. Use the subscript notation "(i)" to indicate the deletion of the i^{th} observation. Then, another variant of h_i is

$$h_{i(i)} = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}_{(i)}' \tilde{\mathbf{X}}_{(i)} \right)^{-1} \tilde{\mathbf{x}}_i,$$

where $\tilde{\mathbf{X}}_{(i)}$ denotes $\tilde{\mathbf{X}}$ with the ith row left out.

This can be thought of as the amount of leverage of the response value Y_i on the corresponding predicted value $\hat{Y}_{i(i)}$.

In the full model, the i^{th} predicted residual is given by

$$e_{(i)} = Y_i - \tilde{\mathbf{x}}_i' \hat{\tilde{\boldsymbol{\beta}}}_{(i)}.$$

It has variance $\sigma_\varepsilon^2 (1 + h_{i(i)})$.

(See, Chatterjee and Hadi, 1986 for detail on applications of $\mathbf{H}$ and statistics calculated with the i^{th} observation omitted.)

In terms of the OLS, leverage values become

$$h_{i(i)} = \frac{h_i}{(1-h_i)} \qquad \text{or} \qquad h_i = \frac{h_{i(i)}}{1+h_{i(i)}} \tag{2.4.1}$$

(see *e.g.* Chatterjee and Hadi, 1986; Hawkins, Bradu and Kass, 1984, page 199).
Define

$$R_{iJ} = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J \right)^{-1} \tilde{\mathbf{x}}_i , \qquad i \notin J \qquad \left(\equiv h_{jJ} \text{ defined in section 2.4} \right).$$

Hawkins *et al.* (1984) refer to R_{iJ} as the ***residual freedom***, to "convey the impression of its property of measuring the extent to which the elemental set J fails to predict Y_i." This follows from the variance which is given by

$$Var\left(e_{iJ}\right) = \sigma_\varepsilon^2 \left(1 + R_{iJ}\right) \text{ for } i \in I.$$

2.5 Concluding remarks

In this chapter we gave an overview of elemental subset regression as well as its relationship to OLS regression. Also, in subsection 2.2.2 we briefly elaborated on leverage-residual weighted elemental estimators which comprise RQ based estimators amongst others. As a consequence there is therefore an inherent relationship amongst ESs, RQs (see Chapter 3) and OLS procedures. In subsection 4.5.1 we further elaborate on the relationship between ESs and RQs. While the relationship between ESs and OLS has been fairly widely explored and used to solve various OLS problems (see section 2.3) the one between ESs and RQs has been observed almost "casually" in the literature. Actually, we will see that a RQ corresponds to a specific ES of size p, in Chapter 4. Therefore by using the existing relationships between ESs estimators (statistics) and OLS estimators (statistics) as well as "new" ones, problems arising in the RQ scenario can be investigated fruitfully. ES based diagnostics (and hence RQ based) are viewed as multiple case diagnostics. Multiple case diagnostics are important since there may be situations where observations are jointly influential, but not individually. Not only is joint influence more difficult to detect, it can also be more serious. One practical problem that usually arises is to determine the size of the influential set. This however is not a problem as far as ES are concerned, multiple cases of the the $n - p$ observations not in the ES corresponding to a RQ are used as predictive (validation) statistics.

CHAPTER 3 REGRESSION QUANTILES

3.1 Introduction

Regression Quantiles (RQs), first proposed by Koenker and Bassett (1978), are natural extensions of order statistics, to the linear model. To define RQs, we begin with the location model (unstructured case) as our point of departure. Let $Y_1, Y_2, ..., Y_n$, be *iid* with distribution function (df) F, assumed to be continuous and strictly increasing. Denote the τ^{th} population quantile by

$$q_\tau \equiv F^{-1}(\tau), \quad 0 < \tau < 1,$$

where $F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$. Denote the order statistics of the sample by $Y_{(1)} \leq Y_{(2)} \leq ,..., \leq Y_{(n)}$, and the empirical distribution function (edf) by

$$F_n(y) = n^{-1} \sum_{i=1}^{n} I(Y_i \leq y).$$

Since F_n is an estimator for F, a natural estimator for q_τ is the τ^{th} sample quantile

$$\hat{q}_\tau \equiv F_n^{-1}(\tau).$$

Note that we have

$$\hat{q}_\tau \; Y_{([n\tau])},$$

where $[x]$ denotes the largest integer less than or equal to x. In order to circumvent this inherent relationship of sample quantiles to the ordered observations, Koenker and Bassett (1978) used a (then) perhaps less well known result of writing a population quantile as a solution to a minimization problem.

Define the function ρ_τ as

$$\rho_\tau(u) = u[\tau - I(u < 0)] \equiv u[\tau . I(u \geq 0) + (\tau - 1) . I(u < 0)]. \qquad (3.1.1)$$

Then it follows easily that

$$q_\tau = \arg\min_{\xi \in \Re} E[\rho_\tau(Y - \xi)]$$

where Y has df F. This then naturally leads to defining the sample quantiles, $\hat{q}_\tau$ as the solution to the corresponding minimization problem based on the sample , *viz.*,

$$\hat{q}_\tau = \arg\min_{\xi \in \Re} \sum_{i=1}^{n} \big(\rho_\tau(Y - \xi)\big). \qquad (3.1.2)$$

This minimization problem may be reformulated as

$$\begin{aligned} &\min\left[\tau \mathbf{1}_n' \mathbf{u}^+ + (1-\tau)\mathbf{1}_n' \mathbf{u}^-\right] \\ &\text{subject to } \mathbf{Y} = \mathbf{1}_n' \xi + \mathbf{1}_n' \mathbf{u}^+ - \mathbf{1}_n' \mathbf{u}^-, \\ &\qquad\qquad \mathbf{u}^+, \mathbf{u}^- \geq \mathbf{0} \end{aligned} \tag{3.1.3}$$

where $\mathbf{1}_n$ is the vector of ones and $\left\{u_i^+, u_i^- : i = 1,...,n\right\}$ represent the positive and the negative parts of residuals respectively. In this formulation it is clearly a linear programming (LP) problem to which the available LP tools could be applied (see *e.g.* Koenker, 2005).

Viewing quantiles as solutions to a minimization problem, Koenker and Bassett (1978) then extended this in a natural fashion to the regression case as we will show in the next section. Some discussion is also given there of the wide applicability of these so called regression quantiles. In section 3.3 estimation and computational aspects are considered and in section 3.4 restricted regression quantiles (RRQs) are introduced. Section 3.5 discusses bounded influence regression quantiles (BIRQs) and the last section gives some concluding remarks.

3.2 Regression case

Consider the usual linear regression model,

$$Y_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \text{ with df } F .$$

In the unstructured case (location model), it is possible to order the data whereas in the structured case the data cannot be ordered. However using the minimization approach of the previous section we can easily generalize to the regression situation as follows:

In an analogy to (3.1.2) define the τ^{th} regression quantile based on the sample $(Y_i, \mathbf{x}_i),\ i = 1,...,n,$ as

$$\hat{\boldsymbol{\beta}}(\tau) \;=\; \arg\min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau\left(Y_i - (\beta_0 + \mathbf{x}_i'\boldsymbol{\beta})\right), \tag{3.2.1}$$

where $\mathbf{x}_i'$ is the i^{th} row of the design matrix $\mathbf{X}$ **without** the constant covariate, β_0 is the intercept term, $\boldsymbol{\beta}$ is the slope coefficient and $\rho_\tau(u)$ as defined in (3.1.1).

What does $\hat{\boldsymbol{\beta}}(\tau)$ estimate? We consider this as follows :

Let $Q_{Y|\mathbf{x}}$ denote the conditional quantile function of Y given the covariate $\mathbf{x}$. Since we have the linear shift model,

$$Q_{Y|\mathbf{x}}(u) = F^{-1}(u) + \beta_0 + \mathbf{x}'\boldsymbol{\beta}.$$

This can be written as

$$Q_{Y|\mathbf{x}}(u) = \begin{pmatrix} 1 & \mathbf{x}' \end{pmatrix} \boldsymbol{\beta}(u),$$

with

$$\boldsymbol{\beta}(u) = \begin{pmatrix} F^{-1}(u) + \beta_0 \\ \boldsymbol{\beta} \end{pmatrix}.$$

Clearly $\hat{\boldsymbol{\beta}}(\tau)$ estimates $\boldsymbol{\beta}(\tau)$. The latter will be called the τ^{th} population regression quantile. Note that $\hat{\boldsymbol{\beta}}(\tau)$ is an M-estimator (see *e.g.* Huber, 1981) with check function ρ_τ. Also, for $\tau = 0.5$ we obtain the usual L_1 regression estimator.

Remark: Based on their definition of RQs, Koenker and Bassett (1978) also defined a $100\alpha\,\%$ regression trimmed mean for $0 < \alpha < 1$ as follows:

- For $0 < \alpha < 1$, determine the regression quantile hyperplanes

  $\begin{pmatrix} 1 & \mathbf{x}' \end{pmatrix} \hat{\boldsymbol{\beta}}(\alpha)$ and $\begin{pmatrix} 1 & \mathbf{x}' \end{pmatrix} \hat{\boldsymbol{\beta}}(1-\alpha)$.

- Discard those observations lying 'below' $\begin{pmatrix} 1 & \mathbf{x}' \end{pmatrix} \hat{\boldsymbol{\beta}}(\alpha)$ or 'above' $\begin{pmatrix} 1 & \mathbf{x}' \end{pmatrix} \hat{\boldsymbol{\beta}}(1-\alpha)$.

- Find the least squares estimates of the remaining observations. Call this $\hat{\boldsymbol{\beta}}_{TM}(\alpha)$, the $100\alpha\,\%$ regression trimmed mean estimator.

$\hat{\boldsymbol{\beta}}_{TM}(\alpha)$ is a robust estimator of $(\beta_0, \boldsymbol{\beta})$ with properties similar to the trimmed mean in the location case (see also Ruppert and Carroll, 1980).

Since the pioneering work of Koenker and Bassett (1978), RQs have been developed in many directions and applied in a variety of situations. An early paper was that of Ruppert and Carroll (1980) where they also derived the limiting distribution of $\hat{\boldsymbol{\beta}}(\tau)$ as well as giving a Bahadur type result for it. A recent paper by Yu, Lu and Stander (2003) gives an overview of recent and current research areas and applications of RQs. They conclude that quantile regression is emerging as a comprehensive approach to the statistical analysis of linear and non-linear models, partly because classical theory is essentially a theory for models of conditional expectations. The ability of RQs to handle conditionally skew distributions and their robustness in cases of error distributions heavier than the Gaussian, give them an edge against the least squares estimator. Some published applications are to medical reference

charts (Cole and Green, 1992; Royston and Altman, 1994); survival analysis (Koenker and Geling, 2001; Yang, 1999); financial and economics research (Taylor, 1999; Lauridsen, 2000; Bassett and Chen, 2001); economics (Buchinsky, 1995; Hendricks and Koenker, 1992); applications to environmental modelling (Pandey and Nguyen, 1999; Hendricks and Koenker, 1992) and applications to ecological modeling (Cade and Noon, 2003). Finally, see also the monograph by Koenker (2005) giving an extensive and authoritative overview of the field.

In the next section we will consider estimation methods and algorithms used for computation.

3.3 Estimation and computational methods

Note again that the minimization problem (3.2.1) can be reformulated as an LP problem. This leads to very efficient computational algorithms which we will now discuss. Consider the vector of residuals

$$\mathbf{r}(\mathbf{b}) = \mathbf{Y} - (\mathbf{1}_n \quad \mathbf{X})\mathbf{b} \equiv \mathbf{r}^+(\mathbf{b}) - \mathbf{r}^-(\mathbf{b}),$$

then we can write

$$\sum_{i=1}^{n} \rho_\tau(r_i(\mathbf{b})) = \tau \sum_{i=1}^{n} r_i^+(\mathbf{b}) + (1-\tau)\sum_{i=1}^{n} r_i^-(\mathbf{b}).$$

Hence in vector-matrix notation the minimization problem becomes

$$\begin{aligned} &\min\left[\tau \mathbf{1}_n' \mathbf{r}^+(\mathbf{b}) + (1-\tau)\mathbf{1}_n' \mathbf{r}^-(\mathbf{b})\right] \\ &\text{subject to } \mathbf{Y} = (\mathbf{1}_n \quad \mathbf{X})\mathbf{b} + \mathbf{r}^+(\mathbf{b}) - \mathbf{r}^-(\mathbf{b}), \\ &\qquad\qquad \mathbf{r}^+(\mathbf{b}), \mathbf{r}^-(\mathbf{b}) \geq \mathbf{0}. \end{aligned}$$

Although ordinary LP algorithms can be applied, some special ones have been developed, *e.g.*, Koenker and D'Orey (1987) and Portnoy and Koenker (1997).

Two major linear programming techniques exist for solving the above linear programming problem, *viz.*, exterior and interior methods. The exterior method is based on the Barrodale and Roberts (1974) simplex algorithm for L_1 estimation. The most widely used version of this is a slight modification due to Koenker and D'Orey (1994). The interior method is based on a modified version of the Frisch-Newton algorithm for quantile regression, *e.g.* see Koenker and Park (1996), Portnoy and Koenker (1997). This method further enhanced the computational speed for large sample size problems. Currently the literature on interior point methods for solving linear programming problems is growing explosively.

A suite of functions for quantile regression in both R (quantreg package) and S-PLUS are available from

http://cran.r-project.org/bin/windows/base/

or

http://www.lib.stat.cmu.edu/

or

http://www.econ.uiuc.edu/~roger/research/ rq/rq.html.

The latter also contains recent information on quantile regression fitting. A suite of functions for general parametric quantile regression in the S-PLUS package is also available on these sites *e.g.*, Koenker and Park (1996). However, despite these advances there are some challenges that still exist in fitting RQs. For instance, while the regression quantile hyperplanes are theoretically parallel, in practice they are often not parallel or cross each other. In the next section we discuss restricted regression quantiles (RRQs) proposed by Koenker (1984) and He (1997) as a solution to this problem.

3.4 Restricted regression quantiles

Regression quantile (RQ) plane crossing is a common problem especially in the lower and upper tails. Unfortunately RQs in the tails are the ones often used to construct prediction intervals. The asymptotic covariance of the RQ estimator is proportional to $\frac{\tau(1-\tau)}{\left[f(F^{-1}(\tau))\right]^2}$ (Koenker and Bassett, 1978; Ruppert and Carroll, 1980) which is often large when τ is close to zero or one. This seems to be the main reason for RQ planes crossing for linear models. Another reason for crossing of RQ planes or at least their failure to be parallel, is the presence of heteroscedasticity. This fact can also be used as a diagnostic tool to detect the presence of heteroscedasticity. However, we shall see in Chapter 5, section 5.2, that the presence of leverage points can be another reason for crossing of RQ planes.

In order to deal with the problem of the crossing of regression quantile planes, Koenker (1984) considered restricted regression quantiles (RRQs) by computing parallel regression quantiles simultaneously. He (1997) (see also Zhao, 2000) proposed a multi-step strategy for this problem and also applied it to a broader class of models including heteroscedastic linear models. We outline He's (1997) procedure for linear and linear heteroscedastic models in the paragraph below.

Koenker (1984) considered RRQs by simultaneously computing multiple parallel regression quantiles for the usual linear model which ensures that the full set of RQs is n.

He (1997) on the other hand, considered a multi-step procedure for both the usual linear model and the linear heteroscedastic model. The asymptotic properties of these estimators were studied by Zhao (2000). We now briefly consider his approach applied to the linear model.

$$Y_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, ..., n,$$

where the ε_i's are *i.i.d.* with df F and the $(1 \quad \mathbf{x}_i')$'s are either non random or random but independent of ε_i. The τ^{th} population regression quantile is given by $\boldsymbol{\beta}(\tau) = \left(\beta_0 + F^{-1}(\tau), \boldsymbol{\beta}\right)'$ of which only the first component depends on τ and the slope is independent of τ. Hence the slope can be estimated separately. Using this fact, He (1997) proposed a two step algorithm to estimate the RRQ planes, *viz.*,

Step1: Estimate the slope parameter by the median estimator, $\hat{\boldsymbol{\beta}}(0.5) = \hat{\boldsymbol{\beta}}$.

Step2: Estimate the intercept at different quantile levels by the sample quantile $\hat{\beta}_0(\tau)$. Denoting the residuals obtained from the first step by

$$r_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}},$$

it follows that

$$\hat{\beta}_0(\tau) = r_{([n\tau])},$$

the $[n\tau]^{th}$ order statistic.

Estimation of the common slope estimator in step 1 ensures that all the quantile planes are parallel. The estimator $\hat{\boldsymbol{\beta}}(\tau) = (\hat{\beta}_0(\tau), \hat{\boldsymbol{\beta}})$ is called the restricted regression quantile estimator (RRQ). The properties of the RRQ estimator were studied by Zhao (2000) who showed that it is consistent, asymptotically normal and more efficient than the ordinary RQ estimator.

3.5 Bounded influence regression quantiles

From the asymptotic distribution of RQs (see Koenker and Bassett, 1978; Ruppert and Carroll, 1980) it is clear that the influence function of $\hat{\boldsymbol{\beta}}(\tau)$ is bounded in the Y space, but unbounded in the X space. The latter shortcoming can be addressed by defining bounded

influence RQs (BI-RQs) as follows:

Let $\{w_i\}$ be appropriate weights (see below) depending on the design. Then the BI-RQ $\hat{\boldsymbol{\beta}}^{(w)}(\tau)$ is the solution to the minimization problem

$$\arg\min_{\beta_0,\boldsymbol{\beta}} \sum_{i=1}^{n} w_i \rho_\tau \left(Y_i - (\beta_0 + \mathbf{x}_i'\boldsymbol{\beta})\right).$$

The weights are typically chosen so as to bound the influence function also in the X space. A well known set of weights are *e.g.*, the Mallows weights (see Mallows, 1973) given by

$$w_i = \prod_{j=2}^{p} w_{ij},$$

where

$$w_{ij} = \begin{cases} 1 & L \le r_{ij} \le U \\ \left(x_{(L)j} - x_{ij}\right)/D_{ij} & r_{ij} < L \\ \left(x_{ij} - x_{(U)j}\right)/D_{ij} & r_{ij} > U, \end{cases}$$

$x_{(L)j}$ and $x_{(U)j}$ are the L^{th} and U^{th} order statistics of the observations $x_{1j},...,x_{nj}$ on the $(j-1)^{th}$ independent variable, for $j = 2,3,...,p$. Here r_{ij} denote the ranks of $x_{1j},...,x_{nj}$, $L = [n\tau]+1$, $U = n+1-L$ and $D_{ij} = 2x_{ij} - x_{(U)j} - x_{(L)j}$ $(i = 1,2,...,n)$.

BI-RQs can also be used to define a bounded influence regression trimmed mean (see the remark in section 3.2). The 100α % bounded influence regression trimmed mean $\hat{\boldsymbol{\beta}}_{TM}^{(w)}(\alpha)$ is defined as the solution to:

$$\arg\min_{\beta_0,\boldsymbol{\beta}\in R^p} \sum_{i=1}^{n} d_i^{(w)}(\alpha)\left(Y_i - (\beta_0 + \mathbf{x}_i'\boldsymbol{\beta})\right)^2,$$

with, for $i = 1,2,...,n$,

$$d_i^{(w)}(\alpha) = \begin{cases} w_i & (1 \quad \mathbf{x}_i')\hat{\boldsymbol{\beta}}^{(w)}(\alpha) \le Y_i \le (1 \quad \mathbf{x}_i')\hat{\boldsymbol{\beta}}^{(w)}(1-\alpha) \\ 0 & \textit{otherwise.} \end{cases}$$

Clearly this is a weighted least squares estimator with weights $d_i^{(w)}(\alpha)$. Let

$$\mathbf{D}^{(w)}(\alpha) = diag\left\{d_i^{(w)}(\alpha)\right\},$$

then

$$\hat{\boldsymbol{\beta}}_{TM}^{(w)}(\alpha) = \left(\mathbf{X}'\mathbf{D}_{\alpha}^{(w)}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_{\alpha}^{(w)}\mathbf{Y},$$

i.e., the weighted least squares estimator of the observations remaining after discarding observations lying outside the α^{th} and $(1-\alpha)^{th}$ bounded influence regression quantiles. Clearly, for the above Mallows choice of weights, observations with high leverage are down weighted. On the other hand, if $w_i = 1 \quad (1 \le i \le n)$, then $\hat{\boldsymbol{\beta}}_{TM}^{(w)}(\alpha) = \hat{\boldsymbol{\beta}}_{TM}(\alpha)$. For a further reparameterization of this problem and asymptotics of the bounded influence regression trimmed means, see *e.g.* De Jongh, De Wet and Welsh (1988).

3.6 Concluding remarks

In this chapter we consider RQs. In section 3.1 we considered the location model as a precursor to considering the regression model (structured case). In the location model, it is shown that the ordering through which sample quantiles are obtained can be replaced by solving a minimization problem. In the regression case, the RQs are then defined as solutions to a similar minimization problem. Therefore RQs can be viewed as extensions of order statistics to the linear model. Also, we mentioned the regression trimmed mean estimator as well as current research areas where RQs are being applied. In section 3.3 we considered some computational methods as well as computational software. Of particular importance to note is the availability of RQ algorithms as free add-on packages to the free R software at a number of websites. This, and the ability of RQs to handle various distributional scenarios led to RQs' emergence as a better, complementary or alternative procedure for statistical analysis in many situations.

In section 3.4 we considered restricted RQs which are often proposed as a remedy for the crossing of RQ hyperplanes. We elaborated on the literature that discussed the causes of RQ hyperplane crossing. In Chapter 5 we shall see that leverage points also result in the crossing

of RQ hyperplanes. This is further exacerbated by the fact that RQs have a high affinity for leverage points (see Chapter 5).

CHAPTER 4 COLLINEARITY, LEVERAGE, OUTLIERS, INFLUENTIAL POINTS AND ASSOCIATED DIAGNOSTICS

4.1 Introduction

In this chapter we give an overview of the problems of collinearity, leverage points (outliers in the design space), outliers, influential points and the challenges they present to multiple linear regression techniques. Furthermore, we give an overview of the OLS diagnostics of each of these problems in the literature as well as propose new diagnostics based on the elemental sets. Both the elemental set diagnostics in the literature and the ones we derive in this thesis, will be expressed in terms of the elemental set weights (2.2.1), where possible.

In the next section we consider collinearity. Leverage points are considered in section 4.3, an alternative view of the ERW based on multiple leverage points in section 4.4, multiple case leverage diagnostics for ESs in section 4.5, collinearity influential observations in section 4.6, "prediction" in RQs in section 4.7, influential observations in section 4.8 and shrinkage techniques in section 4.9. The last section gives some concluding remarks.

4.2 Collinearity

The approximate linear dependencies amongst the columns of predictor variables in multiple linear regression is referred to as collinearity (Silvey 1969; Gunst 1983). Hocking and Pendleton (1983) referred to the exact (or near exact) linear dependences amongst two or more input variables as multicollinearity. In this research, collinearity and multicollinearity are taken to mean the same phenomenon. This phenomenon can have harmful effects on various regression procedures and estimators. However, in the literature there is no consensus as to what constitutes harmful collinearity. As a consequence, for both the overview of the diagnostics in the literature (see subsections 4.2.1 and 4.2.2) and the ones we propose (see subsection 4.2.3), one needs to be cautious on the cut-off values.

While the source of exact linear dependencies amongst predictor variables may be known and avoided, the cause of near exact linear dependences are not clear (Hocking and Pendleton, 1983) and therefore continue to present challenges. The causes of exact linear dependence are mistakes or lack of understanding (*e.g.*, including a variable $X_3 = X_1 + X_2$ in the model when X_1 and X_2 are already in the model). The major sources of near exact linear dependencies

include model constraints, restrictions on the values of the regressors due to the nature of the population, sampling deviancies, leverage outliers (Mason and Gunst, 1985), and whether or not the 'true' theoretical model is known and fully specified (Mela and Kopalle, 2002).

Leverage points do not necessarily induce collinearity. However the occurrence of two or more extremely large in magnitude predictor values for an observation $\mathbf{x}_i'$, may result in collinearity among the corresponding columns of $\mathbf{X}$. On the other hand, in many cases, the researcher does not know the true model or does not have all the predictors needed to correctly specify the model. Misspecification of the model results in biased estimates (see *e.g.* Johnston, 1984; Johnson and Wichern, 1988). When the model is properly specified and estimation is based on sample data, the variances of parameter estimates increase (Lehmann *et al.*, 1988; Wittink, 1988). Mela and Kopalle (2002) point out that this can also, surprisingly, result in the decrease of parameter variance estimates and alluded to the fact that whatever the cause of these near exact linear dependences, their manifestations seem unclear. Amongst others, collinearity increases the variances estimates of parameters, yields high R^2 in the face of low parameter significance, and results in parameter estimates with incorrect signs and large in magnitude (Belsley *et al.*, 1980; Kmenta, 1986; Greene, 1990). Some of these manifestations of collinearity can be exhibited in RQ estimation to a greater degree as the degree of collinearity can be higher at the RQ level as we will see in Chapter 5.

Furthermore, Mela and Kopalle (2002) point out that there exist few studies that explicitly and analytically link the correlation matrix to the problems that these correlations cause and the ability of collinearity diagnostics to detect them. Although different collinearity structures may have different effects on regression, they may yield identical diagnostics. For instance, correlations of equal magnitude but opposite signs may have very different effects on variable omission bias and variance inflation but often yield the same collinearity diagnostic. Secondly, the same correlation structure may affect variable omission bias severely while having little effect on variance inflation (or vice versa). Hence it is recommended not to rely on a single diagnostic to assess the resulting harmful collinearity. Finally, the relationship between the regressors and the dependent variable moderates the effects of collinearity.

In the next section we discuss the ordinary diagnostics for collinearity.

4.2.1 Ordinary collinearity diagnostics

The ordinary diagnostics for collinearity do not always lead to the same conclusion. Furthermore, the same diagnostic often has different cut off values, *e.g.* Green *et al.* (1988),

Fox and Monette (1992) and Lehmann *et al.* (1988) respectively suggest 0.9, 0.35 and 0.7 as a threshold of bivariate correlations for the harmful effect of collinearity.

The sample correlation matrix of the input variables $\mathbf{C}=\left(c_{ij}\right)$ and its derivatives such as its inverse $\mathbf{C}^{-1}=\left(c^{ij}\right)$, eigenvalues, eigenvectors and the determinant $\left|\mathbf{C}\right|$, provide the bulk of ordinary diagnostics for collinearity. Some of these common indicators of collinearity are:

(i) $\left|c_{ij}\right| \approx 1$,

indicates that X_i and X_j are nearly proportional.

(ii) The squared multiple correlation coefficient for the regression of X_j on the remaining X variables,

$$R_j^2 = 1 - \frac{1}{c^{jj}} > 0.9,$$

indicates the degree to which X_j is explained by the linear combination of all the other remaining variables (Hocking and Pendleton, 1983).

Remark: In some situations in the future sections we will use the notation $R^2_{X_l|\mathbf{X}_{(l)}}$ for R_l^2 to explicitly indicate the fact that it is the coefficient of determination of X_l on the remaining variables, *viz.*, $\mathbf{X}_{(l)}$.

(iii) The variance inflation factors,

$$VIF_j = c^{jj} > 10,$$

where c^{jj} is proportional to the variance of the least squares estimate of the j^{th} coefficient, *i.e.*, $\sigma^2_{\beta_j} = \frac{\sigma_e^2}{\mathbf{x}_j'\mathbf{x}_j} VIF_j$ (Hair *et al.*, 1995). Thus a large VIF_j increases the variance of the estimate of the j^{th} coefficient. Also, the VIF can be expressed as

$$VIF_j = \frac{1}{1 - R_j^2},$$

(see Belsley, 1991 where R_j^2 is as defined in (ii)).

(iv) The occurrence of a near zero smallest eigenvalue, *i.e.*,

$\lambda_{\min} \approx 0$,

where

$\lambda_{\min} = \mathbf{u}_1' \mathbf{C} \mathbf{u}_1 = (\mathbf{X}^s \mathbf{u}_1)'(\mathbf{X}^s \mathbf{u}_1)$,

$\mathbf{X}^s$ denotes the predictor matrix standardized to correlation form and $\mathbf{u}_1$ is the eigenvector corresponding to λ_1 (see Hocking and Pendleton, 1983).

(v) The occurrence of elements in the eigenvector corresponding to the smallest near zero eigenvalue ($\lambda_{\min}$) that are large compared to other elements in that vector, identifies predictors that are involved in the multicollinearities (Gunst and Mason, 1980).

(vi) The condition number $\kappa(\mathbf{X}^s) = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of $\mathbf{X}^s$ respectively, provides a summary information on the potential difficulties to be encountered in calculations based on $\mathbf{X}^s$; the larger the condition number, the more ill conditioned $\mathbf{X}^s$ is.

(vii) Define the j^{th} condition index as

$\tilde{\eta}_j = \sqrt{\frac{\tilde{\lambda}_{\max}}{\tilde{\lambda}_j}}$,

where $\tilde{\lambda}_j = \mathbf{u}_j' \tilde{\mathbf{C}} \mathbf{u}_j = (\tilde{\mathbf{X}}^{st} \mathbf{u}_j)'(\tilde{\mathbf{X}}^{st} \mathbf{u}_j)$, and $\tilde{\mathbf{X}}^{st}$ includes the constant column and scaling such that each column has unit sum of squares but the columns are not centred about the mean (as in $\mathbf{X}^s$). Weak collinearities are associated with condition indices of about 5 to 10, whereas moderate to strong collinearities are associated with condition indices of 30 to 100 (see Belsley *et al.*, 1980; Johnston, 1984).

(viii) A near zero value of the determinant of the correlation matrix, $|\mathbf{C}|$ (Johnston, 1984) is indicative of collinearity. We will consider this approach extensively in our research.

Remark:

Hocking and Pendleton (1983) point out that a zero eigenvalue in $\mathbf{C}$ implies a zero eigenvalue in $\tilde{\mathbf{C}}$ and conversely and that the linear relations identified will be identical. They also suggested a singular value decomposition of $\mathbf{X}^s$ or $\mathbf{X}^{st}$ to avoid the disadvantages of forming $\mathbf{C}$ and $\tilde{\mathbf{C}}$ and unstable eigen-analysis if $\mathbf{X}$ is nearly degenerate.

In the next subsection the variance inflation factor (VIF) and its inverse (the tolerance) and

the determinant of the correlation matrix, $|\mathbf{C}|$ are further extended to the ES scenario.

4.2.2 Some useful expressions of VIF (tolerance), $|\mathbf{C}|$ and ERW

In this subsection some expressions for the VIF, tolerance and the determinant of the correlation matrix, $|\mathbf{C}|$ are given. These are further extended to ES, leading to alternative expressions for the ERW for use in later sections. The following theorem and its corollary are central to the derivation of these expressions.

Theorem 4.1: Let $\tilde{\mathbf{X}} = \left[\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_{p\text{-}1}\right]$ be the design matrix of a data set with $p-1$ predictors with first column entry, a column of ones (for the constant term) included and suppose $\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}$ exists. Then

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = n.s'^{2}_{p-1}.s'^{2}_{p-2}...s'^{2}_{1}.|\mathbf{C}| \ ,$$

where

$$s'^{2}_{j} = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2, \ j = 1,..., p-1,$$

is the squared deviation from the mean of the j^{th} predictor and $\mathbf{C}$ is the correlation matrix of the $p-1$ predictors.

Proof: Partitioning $\tilde{\mathbf{X}} = \left[\mathbf{1}_n \quad \mathbf{X}\right]$, then

$$\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \begin{bmatrix} n & \mathbf{1}_n'\mathbf{X} \\ \mathbf{X}'\mathbf{1}_n & \mathbf{X}'\mathbf{X} \end{bmatrix}.$$

Using results for the determinants of partitioned matrices (see *e.g.* Harville, 1997) it follows that

$$\begin{aligned} \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| &= n.\left|\mathbf{X}'\mathbf{X} - \frac{1}{n}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\right| \\ &= n.\left|\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}\right| \end{aligned}$$

Since the correlation matrix of the predictors can be written as

$$\mathbf{C} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}\mathbf{D}^{-\frac{1}{2}},$$

it follows that

$$\begin{aligned}\left|\mathbf{C}\right| &= \left|\mathbf{D}^{-\frac{1}{2}}\right|^2 . \left|\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}\right|, \\ &= \left|\mathbf{D}\right|^{-1}\frac{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|}{n},\end{aligned}$$

where

$$\mathbf{D} = diag\left\{s_j'^2\right\}.$$

Hence

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = n.\prod_j^{p-1} s_j'^2\left|\mathbf{C}\right|. \qquad \square$$

Remark: For interest sake, another proof for this theorem using the principle of mathematical induction is given in Appendix A.

Now, from Mela and Kopalle (2002), the coefficient of determination from the regression of X_l on $\mathbf{X}_{(l)}$ is

$$R^2_{X_l|\mathbf{X}_{(l)}} = 1 - \frac{SSE_{X_l|\mathbf{X}_{(l)}}}{SST_{X_l}} = 1 - \frac{\left|\mathbf{C}\right|}{\left|\mathbf{C}_{(l)}\right|} \qquad (4.2.1),$$

where $\mathbf{X}_{(l)}$ denotes the predictor matrix with the l^{th} predictor deleted, $SSE_{X_l|\mathbf{X}_{(l)}}$, SST_{X_l} and $\mathbf{C}_{(l)}$ are the residual sum of squares from the regression of X_l on $\mathbf{X}_{(l)}$, the total corrected sum of squares of X_l and the correlation matrix of $\mathbf{X}_{(l)}$ respectively.

The tolerance of the l^{th} predictor X_l on the remaining variables $\mathbf{X}_{(l)}$ is given by

$$TOL_{X_l|\mathbf{X}_{(l)}} \equiv 1 - R^2_{X_l|\mathbf{X}_{(l)}}$$

(see *e.g.* Berk, 1977; Raveh, 1985).

Many regression programs include a tolerance test that does not allow a variable to enter the regression if its correlation with the previously entered variables exceeds a prescribed level.

From the definitions of VIF and tolerance, we clearly have

$$VIF_l \equiv \frac{1}{1-R^2_{X_l|\mathbf{X}_{(l)}}} = \frac{1}{TOL_{X_l|\mathbf{X}_{(l)}}}$$

(see also SAS Institute, 1996; Fox, 1991). Thus, the more the variation in X_l is explained by the remaining predictors the less the tolerance becomes and the larger VIF_l becomes.

Also, clearly from (4.2.1), the VIF of the l^{th} predictor can be expressed as

$$VIF_l = \frac{\left|\mathbf{C}_{(l)}\right|}{\left|\mathbf{C}\right|} = \frac{\left|\tilde{\mathbf{X}}'_{(l)}\tilde{\mathbf{X}}_{(l)}\right|}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|} s_l'^2,$$

with the last equality following from Theorem 4.1.

A useful property of the determinant of the correlation matrix is that it can be written as a product of $p-1$ tolerances. To see this, write $\mathbf{C}_{p-1}$ for $\mathbf{C}$, then the determinant of the correlation matrix can be expressed as

$$\begin{aligned}\left|\mathbf{C}_{p-1}\right| &= \frac{\left|\mathbf{C}_1\right|}{1}.\frac{\left|\mathbf{C}_2\right|}{\left|\mathbf{C}_1\right|}...\frac{\left|\mathbf{C}_{p-2}\right|}{\left|\mathbf{C}_{p-3}\right|}.\frac{\left|\mathbf{C}_{p-1}\right|}{\left|\mathbf{C}_{p-2}\right|} \\ &= 1.(1-R^2_{X_2|X_i})...(1-R^2_{X_{p-2}|X_1...X_{p-3}}).(1-R^2_{X_{p-1}|X_1...X_{p-2}}) \\ &= 1.TOL_{X_2|X_1}...TOL_{X_{p-2}|X_1...X_{p-3}}.TOL_{X_{p-1}|X_1...X_{p-2}}\end{aligned}$$

(see Berk, 1977).

Note that the tolerance exposes the amount of "overlap" between the j^{th} predictor and the remaining $p-2$.

Since a variance inflation factor,

$$VIF_j = c^{jj} > 10,$$

is indicative of collinearity (Hair *et al.*, 1995), we can deduce that

$$Tolerance < 0.1$$

is equivalently indicative of collinearity.

Since

$$\left|\mathbf{C}_{p-1}\right| = 1.TOL_{X_2|X_1}...TOL_{X_{p-2}|X_1...X_{p-3}}.TOL_{X_{p-1}|X_1...X_{p-2}},$$

one tolerance less than 0.1 and the remaining ones close to 1, an indication of harmful collinearity can also be taken as,

$$\left|\mathbf{C}_{p-1}\right| < 0.1.$$

Furthermore,

$$\left|\mathbf{C}_{p-1}\right| < (0.1)^{p-2},$$

implying "all" the tolerances less than one, could be indicative of severe collinearity. Hence, an indication of moderately harmful collinearity can be taken as a value between these two "extremes", *e.g.*

$$\left|\mathbf{C}_{p-1}\right| < \frac{0.1+(0.1)^{p-2}}{2}.$$

The following corollary of Theorem 4.1 leads to some further useful results.

Corollary 4.1: Let $\tilde{\mathbf{X}}$ be as in Theorem 4.1 and suppose $\tilde{\mathbf{X}}_J^{-1}$ exists. Then

$$\left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right| = \left|\tilde{\mathbf{X}}_J\right|^2 = p.s'^2_{J,p-1}.s'^2_{J,p-2}...s'^2_{J,1}.\left|\mathbf{C}_J\right|,$$

where

$$s'^2_{J,j} = \sum_{i=1}^{p}(x_{J,ij} - \overline{x}_{J,j})^2,\ j = 1,..., p-1,$$

with $x_{J,ij}$ the $(i,j)^{th}$ element of $\mathbf{X}_J$, $s'^2_{J,j}$ the squared deviation from the mean of the j^{th} predictor in the elemental set J and $\mathbf{C}_J$ is the correlation matrix of the $p-1$ predictors for elemental set J.

Proof: The proof follows directly from Theorem 4.1.

Using corollary 4.1, clearly

$$VIF_{lJ} = \frac{\left|\mathbf{C}_{(l)J}\right|}{\left|\mathbf{C}_J\right|} = \frac{\left|\tilde{\mathbf{X}}'_{(l)J}\tilde{\mathbf{X}}_{(l)J}\right|}{\left|\tilde{\mathbf{X}}_J\right|^2}.s'^2_{J,l},$$

gives the variance inflation factor of the l^{th} predictor in elemental set J.

Theorem 4.1 and Corollary 4.1 give rise to some useful expressions. We mention the following three.

1. An expression for the ERW is

$$\omega_J = \frac{p}{n}\frac{\left|\mathbf{C}_J\right|}{\left|\mathbf{C}\right|}\prod_{j=1}^{p-1}\left(\frac{s'^2_{J,j}}{s'^2_j}\right) \equiv \frac{p}{n}\rho_J\gamma_J. \qquad (4.2.2)$$

Here $0 \le \gamma_J \le 1$, measures the product of the proportion of the ratio of the variabilities of the $p-1$ predictors in the ES to that of the design matrix, and thus gives us the variability view of the ERW. Also, $0 \le \rho_J < \infty$, measures the proportion of $|\mathbf{C}_J|$ to $|\mathbf{C}|$ and thus gives the collinearity view of the ERW. If $\rho_J = \frac{|\mathbf{C}_J|}{|\mathbf{C}|} < 1$, the degree of collinearity is higher at the RQ level compared to that at the design matrix level and vice versa.

Now, since $\sum_{J=1}^{K} \omega_J = 1$, on average each elemental set is expected to contribute $\frac{1}{K}$. A large value of the elemental weight ω_J (greater than $2\left(\frac{1}{K}\right)$, say) can approximately be ascribed to

(i) large value(-s) of $s'^2_{J,j}$ (large dispersion relative to s'^2_j) implying most variation is in set J, or

(ii) a value of $|\mathbf{C}|$ close to 0 (collinearity) or

(iii) a value of $|\mathbf{C}_J|$ close to 1 (little collinearity in elemental set J) or

(iv) a favourable combination of the three.

In Chapter 5 we shall use the above decomposition of ω_J, into the variability and the collinearity views to give us insight into the phenomenon of small ERW.

Remark: If $\mathbf{X}$ is standardized to correlation form (excluding the constant term), the ERW is given by

$$\omega_J = \frac{|\mathbf{C}_J|}{|\mathbf{C}|} = \rho_J .$$

2. Using (4.2.2) and our previously obtained expression,

$$|\mathbf{C}_{p-1}| = 1.(1 - R^2_{X_2|X_i})...(1 - R^2_{X_{p-2}|X_1 \ldots X_{p-3}}).(1 - R^2_{X_{p-1}|X_1 \ldots X_{p-2}}),$$

we find the following alternative expression for the ERW *viz.*

$$\omega_J = \frac{p.s'^2_{1J}.(1-R^2_{X_{2J}|X_{1J}}).s'^2_{2J}...s'^2_{p-2J}(1-R^2_{X_{p-2J}|X_{1J}...X_{p-3J}}).s'^2_{p-1J}(1-R^2_{X_{p-1J}|X_{1J}...X_{p-2J}})}{n.s'^2_1.(1-R^2_{X_2|X_1}).s'^2_2...s'^2_{p-2}(1-R^2_{X_{p-2}|X_1...X_{p-3}}).s'^2_{p-1}(1-R^2_{X_{p-1}|X_1...X_{p-2}})}$$

$$= \frac{p}{n}\prod_{j=1}^{p-1}\left(\frac{s'^2_{jJ}}{s'^2_j}\right)\prod_{j=2}^{p-1}\left(\frac{TOL_{X_{jJ}|X_{1J}...X_{j-1J}}}{TOL_{X_j|X_1...X_{j-1}}}\right). \qquad (4.2.3)$$

Here $R^2_{X_j|X_1...X_{j-1}}$ is the coefficient of determination from the regression of X_j on $X_1,...,X_{j-1}$. Thus the ERW can be expressed as a constant multiplied by the ratio of the products of the tolerances and variabilities based on the ES to those on the full design matrix.

3. Clearly (4.2.3) can also be expressed as

$$\omega_J = \frac{p}{n}\prod_{j=1}^{p-1}\left(\frac{SSE_{X_{jJ}|X_{1J}...X_{j-1J}}}{SSE_{X_j|X_1...X_{j-1}})}\right), \qquad (4.2.4)$$

where $SSE_{X_j|X_1...X_{j-1}}$ is the residual sum of squares from the regression of X_j on $X_1,...,X_{j-1}$. Note that $SSE_{X_1|X_0} \equiv s'^2_1$ and $SSE_{X_{1J}|X_{0J}} \equiv s'^2_{1J}$. Thus, the ERW can also be expressed as a constant multiplied by the ratio of the product of the residual sum of squares from the regression of X_j on $X_1,...,X_{j-1}$ based on the ES to that based on the full design matrix.

The next subsection considers the detection of collinearity using the determinant of the correlation matrix via the ES weight (2.2.1).

4.2.3 Detection of collinearity in elemental sets (RQs)

Elemental set methods have recently become very popular in the arenas of outlier detection and robust regression due to rapidly increasing computer power (see Chapter 2). However, these methods have apparently not been applied to the problem of collinearity to the same extent. In this section we tackle the problem of collinearity using the determinants of the correlation matrix $|\mathbf{C}|$ and the determinant of the correlation matrix based on the ES, $|\mathbf{C}_J|$. In the literature resampling algorithms are often used to obtain a representative sample of all the ESs, but in this study we are interested in each ES (at an individual level) that corresponds to a RQ. We approach the problem via the ES weight as given in the form of (4.2.2).

Using again $\sum_{J=1}^{K}\omega_J = 1$, gives

$$|\mathbf{C}| = \sum_J \frac{p}{n} \cdot \prod_{j=1}^{p} \left(\frac{s'^2_{jJ}}{s'^2_j} \right) \cdot |\mathbf{C}_J|$$
$$= \sum_{J=1}^{K} \omega^*_J |\mathbf{C}_J|,$$

where

$$\omega^*_J = \frac{p}{n} \cdot \prod_{j=1}^{p} \left(\frac{s'^2_{jJ}}{s'^2_j} \right).$$

Note that $0 \le \omega^*_J \le 1$, but $\sum_J \omega^*_J \ne 1$.

From Theorem 4.1 and Corollary 4.1 we can write

$$\left| \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right| = n. \left| \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n \right) \mathbf{X} \right|$$

and

$$\left| \tilde{\mathbf{X}}'_J \tilde{\mathbf{X}}_J \right| = p. \left| \mathbf{X}'_J \left(\mathbf{I} - \frac{1}{p} \mathbf{1}_p \mathbf{1}'_p \right) \mathbf{X}_J \right|,$$

respectively. Note that, $\mathbf{X}_J$ is $p \times (p-1)$ since the constant term is deleted. Now, the ERW is given by

$$\omega_J = \frac{p. \left| \mathbf{X}'_J \left(\mathbf{I} - \frac{1}{p} \mathbf{1}_p \mathbf{1}'_p \right) \mathbf{X}_J \right|}{n. \left| \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n \right) \mathbf{X} \right|}.$$

This expression is useful at times to calculate ω_J. However, we often need a bound on ω_J; this is given in the following lemma.

Lemma: For any elemental set J, $\omega_J \le \frac{p}{n}$.

Proof:

Let, as before, $\tilde{\mathbf{X}} = (\mathbf{1} \quad \mathbf{X})$, and denote by $\mathbf{H}$ the projection matrix of $\tilde{\mathbf{X}}$, *i.e.*,

$$\mathbf{H} = \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}'$$
$$= \mathbf{H}_0 + \mathbf{H}_1,$$

where

$$\mathbf{H}_0 = \frac{1}{n}\mathbf{J}_n ,$$

and

$$\mathbf{H} = (\mathbf{I}_n - \mathbf{H}_0)\mathbf{X}(\mathbf{X}'(\mathbf{I}_n - \mathbf{H}_0)\mathbf{X}_I)^{-1}\mathbf{X}'(\mathbf{I}_n - \mathbf{H}_0)$$

(see *e.g.* Chatterjee and Hadi, 1988, property 2.4, p16).

Consider the following equality

$$\left(\mathbf{I}_n - \mathbf{H}_0\right) = \left(\mathbf{I}_n - \mathbf{H}\right) + (\mathbf{H} - \mathbf{H}_0) .$$

Clearly all three matrices (each one in brackets) are nonnegative definite (since they are projection matrices).

Now, taking the corresponding $(n-p)\times(n-p)$ submatrices indexed by I from the above three matrices, we can write

$$\left(\mathbf{I}_n - \mathbf{H}_0\right)_I = \left(\mathbf{I}_n - \mathbf{H}\right)_I + (\mathbf{H} - \mathbf{H}_0)_I .$$

Again all three matrices are nonnegative definite (see *e.g.* van Vuuren, 1998, section 2.5.5). Using corollary 18.1.8 in Harville (1997), p 418, this gives

$$\left|\left(\mathbf{I}_n - \mathbf{H}_0\right)_I\right| \geq \left|\left(\mathbf{I}_n - \mathbf{H}\right)_I\right| .$$

Clearly

$$\left(\mathbf{I}_n - \mathbf{H}_0\right)_I = \mathbf{I}_{n-p} - \tfrac{1}{n}\mathbf{J}_{n-p} .$$

Also, since

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{X}}_J \\ \tilde{\mathbf{X}}_I \end{pmatrix},$$

we have

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} \tilde{\mathbf{X}}_J \\ \tilde{\mathbf{X}}_I \end{pmatrix} (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-!} \begin{pmatrix} \tilde{\mathbf{X}}'_J & \tilde{\mathbf{X}}'_I \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\mathbf{X}}_J (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-!}\tilde{\mathbf{X}}'_J & \tilde{\mathbf{X}}_J (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-!}\tilde{\mathbf{X}}'_I \\ \tilde{\mathbf{X}}_I (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-!}\tilde{\mathbf{X}}'_J & \tilde{\mathbf{X}}_I (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-!}\tilde{\mathbf{X}}'_I \end{pmatrix} \end{aligned}$$

From which

$$\left(\mathbf{H} - \mathbf{H}_0\right)_I = \mathbf{H}_I - \mathbf{H}_{0I}$$

and

$$\left(\mathbf{I}_n - \mathbf{H}\right)_I = \mathbf{I}_{n-p} - \mathbf{H}_I,$$

with

$$\mathbf{H}_{0I} = \frac{1}{n}\mathbf{J}_{n-p}.$$

The above inequality can then be written as

$$\left|\mathbf{I}_{n-p} - \frac{1}{n}\mathbf{J}_{n-p}\right| \geq \left|\mathbf{I}_{n-p} - \mathbf{H}_I\right| = \omega_J$$

(see Theorem 4.2).

Thus

$$\omega_J \leq \left|\mathbf{I}_{n-p} - \frac{1}{n}\mathbf{J}_{n-p}\right|.$$

The proof is completed by calculation of the determinant on the right hand side.

Now,

$$\begin{aligned}\left|\mathbf{I}_{n-p} - \frac{1}{n}\mathbf{J}_{n-p}\right| &= \begin{vmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{vmatrix} \\ &= 1 - \frac{(n-p)}{n} \\ &= \frac{p}{n}\end{aligned}$$

(see *e.g.* Searle, 1982, p 322).

This completes the proof of the lemma. □

Now, since $\omega_J = \omega_J^* \frac{|\mathbf{C}_J|}{|\mathbf{C}|}$, we have ω_J close to $\frac{p}{n}$ when

- ω_J^* close to $\frac{p}{n}$ and $\frac{|\mathbf{C}_J|}{|\mathbf{C}|}$ close to 1.

or

- $\frac{|\mathbf{C}_J|}{|\mathbf{C}|} > 1$ and ω_J^* proportionately smaller than $\frac{p}{n}$.

Alternatively ω_J close to 0 when

- $\frac{|\mathbf{C}_J|}{|\mathbf{C}|}$ close to 0 and/or ω_J^* close to 0.

It is clear that a holistically, appealing measure of collinearity for a RQ (or ES) is the determinant of the correlation matrix since it is bounded between 0 (singularity) and 1 (orthogonality) and furthermore, it can be expressed in terms of other collinearity measures such as the VIFs (tolerances) (see subsection 4.2.2). Like the VIFs, it is involved in the construction of confidence regions (ellipsoids). It can be shown that the ratio of the bound of the confidence ellipsoid based on an orthogonal reference design to that based on the actual design is $|\boldsymbol{C}|^{1/2}$ (see Willan and Watts, 1978). This volume ratio tells us how much smaller the confidence ellipsoid could have been if an orthogonal reference design had been used instead of the actual design.

Although the determinant of the correlation matrix $|\boldsymbol{C}|$ has the attractive property that it is bounded between 0 (singularity) and 1 (orthogonality), it has a drawback due to the absence of the usual cut-off values. This is exacerbated by the fact that there is no consensus as to what constitutes harmful collinearity (see subsection 4.2). One approach is to determine cut-off values based on those of the VIFs (see subsection 4.2.2). Another approach is to determine cut-off values based on the desired efficiency factor $|\boldsymbol{C}|^{1/2}$. So $|\boldsymbol{C}|$ and ρ_J can be viewed as the square of the efficiency factor and the square of the ratio of the efficiency factor at the RQ (ES) level to that at the full design matrix level, respectively.

The condition number $\kappa(\boldsymbol{X}^S)$ (see subsection 4.2.1) is another holistic measure of collinearity that can be extended to the RQ (ES) scenario. A number of multiple case influence measures based on the condition number have been proposed in the literature (see *e.g.* Hadi and Wells, 1990; Sengupta and Bhimasankaram, 1997). Such multiple case collinearity measures based on $\kappa(\boldsymbol{X}^S)$ have a number of shortcomings which, *inter alia*, include the cumbersome

computing load of all the $K = \binom{n}{p}$ cases. However, in the RQ scenario we are only interested in $\pm n$ $(<< K)$ ESs corresponding to specific RQs which results in a much reduced computational load. When the researcher is interested in collinearity measures that detect the effect of collinearity on the variance of individual parameter estimators, a natural influence measure is the VIF (see section 4.2.1). In a similar fashion to the extension of $\kappa(\boldsymbol{X}^S)$ to the multiple case scenario, the VIF can also be extended to the multiple case scenario (see Sengupta and Bhimasankaram, 1997). They proposed the statistic

$$f_{jI} = \left(1 + \frac{\boldsymbol{d}_{Ij}' \left|\boldsymbol{I} - \boldsymbol{H}_I\right|^{-1} \boldsymbol{d}_{Ij}}{((\boldsymbol{X}'\boldsymbol{X})^{-1})_{jj}}\right)\left(1 - \frac{\sum_{i\in I} x_{ij}^2}{\sum_{i=1}^n x_{lj}^2}\right), \quad 1 \le j \le p-1,$$

where $\boldsymbol{d}_{Ij}$ is the j^{th} column of the matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $((\boldsymbol{X}'\boldsymbol{X})^{-1})_{jj}$ denotes the diagonal entry of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

We can express this statistic as

$$f_{jI} = \left(1 + \frac{\boldsymbol{d}_{Ij}' \omega_J^{-1} \boldsymbol{d}_{Ij}}{((\boldsymbol{X}'\boldsymbol{X})^{-1})_{jj}}\right)\left(\frac{\sum_{i\in J} x_{ij}^2}{\sum_{i=1}^n x_{lj}^2}\right), \quad 1 \le j \le p-1$$

where $\omega_J = \left|\boldsymbol{I} - \boldsymbol{H}_I\right|$ (see Theorem 4.2) if the x_{ij}'s in the numerator of the last factor have the same standardisation as the ones in its denominator since the sets I and J are complementary.

They also proposed the statistic

$$\psi_I = \left|\boldsymbol{I} - \boldsymbol{H}_I\right|^{-1/2} \prod_{j=1}^{p-1}\left(1 - \frac{\sum_{i\in I} x_{ij}^2}{\sum_{i=1}^n x_{lj}^2}\right)^{1/2},$$

if the researcher's interest is the whole slope parameter vector. Again we can write this statistic as

$$\psi_I = \omega_J^{-1/2} \prod_{j=1}^{p-1}\left(\frac{\sum_{i\in J} x_{ij}^2}{\sum_{i=1}^n x_{lj}^2}\right)^{1/2}.$$

We mention the above measures for completeness to show how the ERW, ω_J, features frequently in multiple case collinearity diagnostics. However, in Chapters 5 and 6 we make

use of the collinearity diagnostics $|\boldsymbol{C}_J|$, $|\boldsymbol{C}|$ and ρ_J.

4.3 Leverage diagnostics

We first give a brief overview of single case leverage diagnostics. The leverage of the i^{th} observation is given by diagonal element, $h_i = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{x}}_i$ of the projection matrix $\mathbf{H} = \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}'$. This matrix is sometimes referred to as the Hat matrix because it maps $\mathbf{Y}$ into $\hat{\mathbf{Y}}$, *i.e.*,

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y} .$$

The relationship between the leverage and residuals is

$$h_i + \frac{e_i^2}{\sum e_i^2} \leq 1$$

(see Velleman and Welsch, 1981).

The matrix $\mathbf{H}$ has a number of important properties (see the remarks below). From the first remark it is clear that whenever $h_i = 0$ ($\hat{Y}_i$ is fixed at zero by the design) or $h_i = 1$ ($\hat{Y}_i = Y_i$), we have $h_{ij} = 0$ for all $j \neq i$. Also from the last remark below it is clear that the average leverage value is $\frac{p}{n}$.

Determining the points with high leverage involves checking whether the diagonal elements, $h_i = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{x}}_i$ of the projection matrix $\mathbf{H} = \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}'$ exceed a certain threshold value. According to Hoaglin and Welsch (1978), experience suggests a reasonable rule of thumb is $\frac{2p}{n}$ as a threshold value for h_i. This is the most popular one used in the literature.

They also suggest

$$\frac{\left[(n-p)(h_i - 1/n) \right]}{\left[(p-1)(1-h_i) \right]}, \qquad (4.3.1)$$

with an F distribution on $p-1$ and $n-p$ degrees of freedom as a more refined screening statistic when the model includes the constant predictor and the rows of $\mathbf{X}$ are sampled from a $p-1$ variate Gaussian distribution.

Belsley *et al.* (1980) further showed that for $p > 15$ and $n - p > 30$, $\frac{2p}{n}$ corresponds to an appropriate 95% point for this distribution while for $p > 6$ and $n - p > 12$, $\frac{3p}{n}$ is more appropriate as a threshold value. On the other hand, according to Huber (1981), $\frac{1}{h_i}$ may be viewed as the "equivalent number of observations" entering into the determination of $\hat{Y}_i$. Hence, a point with unit leverage determines its predicted value alone and $h_i > 0.5$ (2 equivalent observations) is clearly large, and $h_i > 0.2$ (5 equivalent observations) calls for special attention.

Remarks:

The projection matrix $\mathbf{H}$ has the following important mathematical properties.

- It is symmetric and idempotent ($\mathbf{H}^2 = \mathbf{H}$).
- Some properties of $\mathbf{H}$ which have a bearing on the degree of leverage and prediction are:

  $h_i = \sum_{j=1}^{n} h_{ij}^2 = h_i^2 + \sum_{j \neq i} h_{ij}^2$,

  $0 \leq h_i \leq 1$,

  and hence,

  $h_i = 0$ implies $h_{ij} = 0$ and

  $h_i = 1$ implies $h_{ij} = 0$.

- Its eigenvalues are either zero or one and the number of non-zero eigenvalues is equal to the rank of the design matrix $\tilde{\mathbf{X}}$.
- This leads to $rank(\mathbf{H}) = rank(\tilde{\mathbf{X}}) = p$ and hence $trace(\mathbf{H}) \equiv \sum_{i=1}^{n} h_i = p$.

In this thesis we are more interested in the linear model that includes the constant predictor. As a consequence it is important to outline the leverage bounds for this model, starting with the location model.

For this case we have the model,

$$\mathbf{Y} = \mathbf{1}_n \beta_0 + \boldsymbol{\varepsilon},$$

the diagonals of $\mathbf{H}$ are given by

$$h_i = \frac{1}{n}.$$

Adding a nonconstant term, and then two nonconstant predictors to the location case the diagonals of $\mathbf{H}$ are given by

$$h_i = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_{j=1}^{n}(x_j - \overline{x})^2}$$

and

$$h_i = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_{j=1}^{n}(x_j - \overline{x})^2} + \frac{e^2_{iX_2|\mathbf{X}_{(2)}}}{SSE_{X_2|\mathbf{X}_{(2)}}}$$

respectively. Here $e^2_{iX_2|\mathbf{X}_{(2)}}$ and $SSE_{X_2|\mathbf{X}_{(2)}}$ denote the i^{th} residual and the residual sum of squares from the regression of X_2 on $\mathbf{X}_{(2)}$.

It can be shown that for the model with the constant term the diagonals of $\mathbf{H}$ are given by $\frac{1}{n}$ plus a small positive number (less than 1) each time a nonconstant predictor is added to the model. Hence, for these models ($\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{X\beta} + \boldsymbol{\varepsilon}$) we have

$$\frac{1}{n} \le h_i \le 1 \qquad (4.3.2)$$

(see also van Vuuren, 1998; Hocking and Pendleton, 1983).

It can also be shown that for this model the diagonals of $\mathbf{H}$ are given by

$$\frac{1}{n} + \text{Mahalanobis distance of } \begin{pmatrix} x_{1i} \\ \vdots \\ x_{pi} \end{pmatrix} \text{ from } \begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix} \Bigg/ (n-1),$$

where $\overline{x}_k$ denotes the mean of the k^{th} predictor (see Ellis and Morgenthaler, 1992).

According to Huber (1981), if $c = \frac{1}{h_i}$ is the number of times the i^{th} row of $\mathbf{X}$ is replicated, *i.e.*, the maximum "equivalent number of cases" entering into the determination of $\hat{Y}_i$, then tighter bounds for this class of models are

$$\frac{1}{n} \le h_i \le \frac{1}{c},\ c = 2,3,...,n-1, \tag{4.3.3}$$

and $c = 1$ gives the same bounds as in (4.3.2).

Some of the relationships involving h_i are:

- $\text{var}(\hat{Y}_i) = h_i\sigma^2$, and $\text{var}(e_i) = (1-h_i)\sigma^2$,
- h_i is large if the i^{th} row of $\mathbf{X}$ is far removed from the other rows in the X space and
- $\frac{1}{h_i}$ is an approximation to the number of cases determining $\hat{Y}_i$.

Furthermore, h_i is utilized in residual outlier and influence diagnostics, *e.g.*,

$$t_i = e_i(\hat{\sigma}) = \frac{e_i}{\hat{\sigma}\sqrt{1-h_i}}, \tag{4.3.4}$$

and

$$t_i^* = e_i(\hat{\sigma}_{(i)}) = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-h_i}}, \tag{4.3.5}$$

where $\hat{\sigma}_{(i)} = \frac{(n-p)\hat{\sigma}}{n-p-1}$.

In the literature t_i^* is preferred to t_i (see *e.g.* Belsley, Kuh and Welsch, 1980) and is sometimes referred to as the "jackknife" residuals (see Atkinson, 1981) or the "studentized" residuals (see Velleman and Welsch, 1981).

Also, in Cook's distance proposed by Cook (1977), h_i is utilized, *viz.*,

$$D_i = \frac{n-p}{p} \cdot \frac{h_i q_i}{(1-h_i)^2}.$$

Here $q_i = e_i^2 \Big/ \left[(n-p)s^2\right]$ is the diagonal element of the residual matrix form $\mathbf{Q} = \mathbf{e}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'$ (Barrett and Ling,1992).

Also it relates the OLS residuals $e_i, 1 \le i \le n$ to the jackknife predicted residuals $e_{(i)}, 1 \le i \le n$, *i.e.*,

$$e_{(i)} = \frac{e_i}{(1-h_i)}, 1 \le i \le n .$$

Note that the predictive (jackknife) leverage, given by the i^{th} diagonal element $h_{i(i)} = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}_{(i)}' \tilde{\mathbf{X}}_{(i)} \right)^{-1} \tilde{\mathbf{x}}_i$, of the matrix $\mathbf{H}_{(i)} = \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}_{(i)}' \tilde{\mathbf{X}}_{(i)} \right)^{-1} \tilde{\mathbf{X}}'$, is monotonically related to ordinary leverage h_i via (2.4.1) and threshold values for $h_{i(i)}$ can easily be deduced from those of h_i.

In the next section we present an alternative view of the elemental weight based on leverage.

4.4 An alternative view of the elemental regression weight (ERW) based on multiple leverage points

For the ES, leverage values are equal to one (which is not useful). As a consequence, we only need to focus on those observations that are outside the ES J, *i.e.*, $i \in I$ (multiple "predictive leverage"). We loosely refer to the leverage of points outside the ES as multiple "predictive leverage". The ERW can be expressed in terms of multiple leverage via two different matrices, *viz.*,

$$\mathbf{H}_{IJ} = \tilde{\mathbf{X}}_I \left(\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J \right)^{-1} \tilde{\mathbf{X}}_I' \text{ and } \mathbf{H}_I = \tilde{\mathbf{X}}_I \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}_I'.$$

The following theorem gives these alternative expressions for the ERW.

Theorem 4.2: For $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ nonsingular, we have

(i) $\omega_J = \left| \mathbf{I}_{n-p} + \mathbf{H}_{IJ} \right|^{-1}$

and

(ii) $\omega_J = \left| \mathbf{I} - \mathbf{H}_I \right|$.

Proof of (i): Note from Chapter 2 that $\omega_J = \frac{\left| \tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J \right|}{\left| \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right|}$. We use this to prove the theorem.

From the well known result (see *e.g.* Harville, 1997)

$$\left| \mathbf{R} + \mathbf{STU} \right| = \left| \mathbf{R} \right| \left| \mathbf{T} \right| \left| \mathbf{T}^{-1} + \mathbf{UR}^{-1}\mathbf{S} \right|,$$

where $\mathbf{R}$ and $\mathbf{T}$ are nonsingular, we have

$$\begin{aligned}\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| &= \left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J + \tilde{\mathbf{X}}_I'\mathbf{I}_{n-p}\tilde{\mathbf{X}}_I\right| \\ &= \left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right|\left|\mathbf{I}_{n-p}\right|\left|\mathbf{I}_{n-p} + \tilde{\mathbf{X}}_I(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}_I'\right|\end{aligned}$$

and (i) follows.

Proof of (ii): Let $\mathbf{Z} = \begin{pmatrix} \mathbf{I} & \tilde{\mathbf{X}}_I \\ \tilde{\mathbf{X}}_I' & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{pmatrix}$. Now, for any $k \times k$ nonsingular matrix $\mathbf{A}_{11}$, $k \times m$ matrix $\mathbf{A}_{12}$, $m \times k$ matrix $\mathbf{A}_{21}$ and $m \times m$ nonsingular matrix $\mathbf{A}_{22}$, it is well known that the determinant of the partitioned matrix (see also appendix A, Theorem A.1(i)) is given by

$$\begin{aligned}\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} &= \left|\mathbf{A}_{22}\right|\left|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right| \\ &= \left|\mathbf{A}_{11}\right|\left|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right|.\end{aligned}$$

Applying the first form to $\mathbf{Z}$, gives

$$\begin{aligned}\left|\mathbf{Z}\right| &= \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|\left|\mathbf{I} - \tilde{\mathbf{X}}_I\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}_I'\right| \\ &= \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|\left|\mathbf{I} - \mathbf{H}_I\right|,\end{aligned}$$

and using the second form, gives

$$\begin{aligned}\left|\mathbf{Z}\right| &= \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}} - \tilde{\mathbf{X}}_I'\tilde{\mathbf{X}}_I\right| \\ &= \left|\tilde{\mathbf{X}}_{(I)}'\tilde{\mathbf{X}}_{(I)}\right| \\ &= \left|\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right|.\end{aligned}$$

From these two expressions, (ii) follows. This completes the proof of the theorem. □

The variance-covariance matrix of the elemental predicted residuals (see subsection 2.4.1) is given by

$$\begin{aligned}Var(\mathbf{e}_{IJ}) &= \sigma^2\left(\mathbf{I}_{n-p} + \tilde{\mathbf{X}}_I\left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right)^{-1}\tilde{\mathbf{X}}_I'\right) \\ &= \sigma^2\left(\mathbf{I}_{n-p} + \mathbf{H}_{IJ}\right).\end{aligned}$$

Extending the definition of the generalized variance of the jackknife predicted residuals, as given in Remark 1 below, to the elemental predicted residuals, we find that

$$\begin{aligned}gVar(\mathbf{e}_{IJ}) &\equiv \sigma^2\left|\mathbf{I}_{n-p} + \mathbf{H}_{IJ}\right| \\ &= \sigma^2\omega_J^{-1}.\end{aligned}$$

Clearly, the larger the generalized variance, the smaller the elemental weight ω_J and vice versa. Again on average each elemental set is expected to contribute $\frac{1}{K}$. Hence a large value of the elemental weight ω_J (greater than $2\left(\frac{1}{K}\right)$, say) implies a small generalized variance and vice versa.

Remark 1: If $n = p+1,$ then the jackknife (predictive) leverage and the elemental predictive leverage coincide since $\tilde{\mathbf{X}}_{(i)} = \tilde{\mathbf{X}}_J$ and hence

$$\tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}_{(i)}' \tilde{\mathbf{X}}_{(i)}\right)^{-1} \tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i' \left(\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J\right)^{-1} \tilde{\mathbf{x}}_i, \quad i \notin J\,.$$

Also,

$$\omega_J = \frac{\left|\tilde{\mathbf{X}}_J' \tilde{\mathbf{X}}_J\right|}{\left|\mathbf{X}'\mathbf{X}\right|} = \frac{\left|\tilde{\mathbf{X}}_{(i)}' \tilde{\mathbf{X}}_{(i)}\right|}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|} = 1 - h_i, \qquad i \notin J$$

which implies that a single leverage point can result in a small weight for the corresponding elemental set, J.

Theorem 4.2 (ii) is an analogue of the multiple case to the simple case, *i.e.*,

$$\begin{aligned}\omega_J &= \left|\mathbf{I} - \tilde{\mathbf{X}}_I \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}_I'\right| \\ &= \left|\mathbf{I} - \mathbf{H}_I\right|.\end{aligned}$$

Remark 2: Note that from Hawkins *et al*. (1984), it follows that the generalized variance of the predicted (jackknife) residuals is given by

$$gVar(e_{(i)}) = \left|\mathbf{I} + \mathbf{H}_{(i)}\right| \sigma^2\,.$$

Remark 3: Besides its determinantal relationship to ω_J, the $(n-p)\times(n-p)$ matrix $\left(\mathbf{I} - \mathbf{H}_I\right) = \left(\mathbf{I}_{n-p} + \mathbf{H}_{IJ}\right)^{-1}$ has, *inter alia*, also the following properties of use in our further work:

- Its eigenvalues are bounded by zero and one, hence $0 \le \omega_J = \left|\mathbf{I} - \mathbf{H}_I\right| = \left|\mathbf{I}_{n-p} + \mathbf{H}_{IJ}\right|^{-1} \le 1,$

  since it is the product of the eigenvalues.

- The tighter upper bound for ω_J is $\frac{p}{n}$, *i.e.*, $0 \le \omega_J < \frac{p}{n}$ (see subsection 4.2.3).

- It is positive definite whenever the maximum eigenvalue of $\mathbf{H}_I$ is less than one.

- $gVar(\mathbf{e}_{IJ}) = \sigma^2 \left|\mathbf{I} - \mathbf{H}_I\right|^{-1}$,

  can be deduced from Theorem 4.2.

Remark 4: The **elemental regression matrix** $\left(\mathbf{I} - \mathbf{H}_I\right) = \left(\mathbf{I}_{n-p} + \mathbf{H}_{IJ}\right)^{-1}$ plays important roles in determining the influence of multiple cases. The two most important and well known are:

- Cook's distance D_i (see section 4.3) generalizes to

  $$D_I = \frac{\mathbf{e}_I' \left(\mathbf{I} - \mathbf{H}_I\right)^{-1} \mathbf{H}_I \left(\mathbf{I} - \mathbf{H}_I\right)^{-1} \mathbf{e}_I}{p\hat{\sigma}^2}.$$

  Also, this diagnostic can be expressed in terms of leverage, residual and interaction. *E.g.*, Barrett and Ling (1992) expressed Cook's distance in the form

  $$D_I \approx \frac{n-p}{p} \frac{tr(\mathbf{H}_I \mathbf{Q}_I)}{\left(1 - \left\|\mathbf{H}_I\right\|\right)^2} = \frac{n-p}{p} \frac{\left\|\mathbf{H}_I\right\| \cdot \left\|\mathbf{Q}_I\right\| \cos\alpha_I}{\left(1 - \left\|\mathbf{H}_I\right\|\right)^2},$$

  where $\mathbf{Q}_I = \mathbf{e}_I \left(\mathbf{e}'\mathbf{e}\right)^{-1} \mathbf{e}_I'$ is a submatrix of the residual matrix (see section 4.3) and α_I is the "interaction" between elements of $\mathbf{H}_I$ and $\mathbf{Q}_I$, where $\cos\alpha_I$ is the cosine of the angle between $vec(\mathbf{H}_I)$ $vec(\mathbf{Q}_I)$. If $\cos\alpha_I \approx 0$, the joint influence of the subset is nearly zero. This phenomenon is referred to as joint cancelling influence.

- The estimated generalized variance of $\hat{\boldsymbol{\beta}}$ is given by $\left|\hat{\sigma}^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|$ and that of $\hat{\boldsymbol{\beta}}_{(i)}$ is given by $\left|\hat{\sigma}^2_{(i)}(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|$. Based on the measure (the covariance ratio)

$$CVR_i \equiv \frac{\left|\hat{\sigma}^2_{(i)}(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|}{\left|\hat{\sigma}^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|} = \frac{\hat{\sigma}^{2p}_{(i)}\left|(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|}{\hat{\sigma}^{2p}\left|(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|},$$

(see Belsley *et al.*, 1980) and applying Theorem 4.2(ii), an extension of this diagnostic, if σ^2 is known, becomes

$$\begin{aligned} CVR_I &= \frac{g\ \mathrm{var}(\boldsymbol{\beta}_J)}{g\ \mathrm{var}(\boldsymbol{\beta})} \\ &= \frac{\sigma^{2p}\left|\tilde{\mathbf{X}}'_{(I)}\tilde{\mathbf{X}}_{(I)}\right|^{-1}}{\sigma^{2p}\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|^{-1}} \\ &= \frac{\sigma^{2p}\left|\tilde{\mathbf{X}}'_J\tilde{\mathbf{X}}_J\right|^{-1}}{\sigma^{2p}\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|^{-1}} \\ &= \left|\mathbf{I}-\mathbf{H}\right|^{-1} = \omega_J^{-1}. \end{aligned}$$

We will, however, mainly focus on one derivative of $(\mathbf{I}-\mathbf{H}_I)$, *viz.* the ERW,

$$\omega_J = \left|\mathbf{I}-\mathbf{H}_I\right| = \left|\mathbf{I}_{n-p}+\mathbf{H}_{IJ}\right|^{-1}.$$

Other properties of $(\mathbf{I}-\mathbf{H}_I)$ can be fruitfully investigated in future research.

In the next section we consider multiple leverage diagnostics where the roles of ERW and $(\mathbf{I}-\mathbf{H}_I)$ are evident. These diagnostics are motivated by the relationship between $h_i/h_{i(i)}$ and ES predictive leverage via the ERW.

4.5 Multiple case leverage diagnostics for ESs

In this section we extend the single case leverage diagnostics discussed in section 2.4 to regression quantile/elemental regression prediction. These include the residual freedom $R_{iJ} = \tilde{\mathbf{x}}'_i\left(\tilde{\mathbf{X}}'_J\tilde{\mathbf{X}}_J\right)^{-1}\tilde{\mathbf{x}}_i$, $i \notin J$ and the OLS leverage values h_i and $h_{i(i)}$ given in equation (2.4.1).

Now, define

$$\tilde{R}_i = \frac{\sum_{J \not\ni i}\omega_J R_{iJ}}{\sum_{J \not\ni i}\omega_J},$$

and let h_i and $h_{i(i)}$ be as in (2.4.1). We then have

Theorem 4.3: For $\tilde{\mathbf{X}}_J$ nonsingular,

(i) $h_{i(i)} = \dfrac{\sum_{J \ni i} \omega_J R_{iJ}}{(n-p)\sum_{J \ni i} \omega_J} = \dfrac{\tilde{R}_i}{(n-p)}$.

(ii) $h_i = \dfrac{\tilde{R}_i}{(n-p)+\tilde{R}_i}$.

(iii) $h_i = \dfrac{\sum_{J \ni i} \omega_J R_{iJ}}{(n-p)}$.

Proof of (i).

For any $\mathbf{d} \in \Re^p$, $\left|\mathbf{I}_p + \mathbf{d}\mathbf{d}'\right| = 1 + \mathbf{d}'\mathbf{d}$. Similarly for any $(p \times p)$ symmetric and nonsingular matrix $\mathbf{A}$, $\mathbf{d} \in \Re^p$, $\left|\mathbf{A} + \mathbf{d}\mathbf{d}'\right| = \left|\mathbf{A}\right|\left(1 + \mathbf{d}'\mathbf{A}^{-1}\mathbf{d}\right)$.

Thus, using the Cauchy-Binet Theorem (see Theorem A.2 in appendix A),

$$
\begin{aligned}
1 + h_{i(i)} &= 1 + \tilde{\mathbf{x}}_i'\left(\tilde{\mathbf{X}}_{(i)}'\tilde{\mathbf{X}}_{(i)}\right)^{-1}\mathbf{x}_i \\
&= \frac{\left|\left(\tilde{\mathbf{X}}_{(i)}'\tilde{\mathbf{X}}_{(i)}\right) + \tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right|}{\left|\left(\tilde{\mathbf{X}}_{(i)}'\tilde{\mathbf{X}}_{(i)}\right)\right|} \quad = \frac{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|}{\left|\left(\tilde{\mathbf{X}}_{(i)}'\tilde{\mathbf{X}}_{(i)}\right)\right|} \\
&= \frac{\sum_J \left|\tilde{\mathbf{X}}_J\right|^2}{\sum_{J \ni i}\left|\tilde{\mathbf{X}}_J\right|^2} \\
h_{i(i)} &= \frac{\sum_{J \ni i}\left|\tilde{\mathbf{X}}_J\right|^2}{\sum_{J \ni i}\left|\tilde{\mathbf{X}}_J\right|^2}. \qquad (4.5.1)
\end{aligned}
$$

Similarly, again using the Cauchy-Binet Theorem,

$$
\begin{aligned}
1 + R_{iJ} &= 1 + \tilde{\mathbf{x}}_i'\left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right)^{-1}\tilde{\mathbf{x}}_i \\
&= \frac{\left|\left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right) + \tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right|}{\left|\left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right)\right|} \\
&= \frac{\sum_{H \subset \{J,i\}}\left|\tilde{\mathbf{X}}_H\right|^2}{\left|\tilde{\mathbf{X}}_J\right|^2}.
\end{aligned}
$$

Thus

$$R_{iJ} = \frac{\sum_{H \subset \{J,i\}, H \ni i} \left|\tilde{\mathbf{X}}_H\right|^2}{\left|\tilde{\mathbf{X}}_J\right|^2}.$$

Multiplying both sides by $\left|\tilde{\mathbf{X}}_J\right|^2$ and summing over $J \not\ni i$,

$$\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2 R_{iJ} = \sum_{J \not\ni i} \sum_{H \subset \{J,i\}, H \ni i} \left|\tilde{\mathbf{X}}_H\right|^2.$$

Now, for any given set H, set J, which does not include i, must contain the other $p-1$ elements of H. Thus J will have only one free element, which will run over the $n-p$ values outside H. Consequently any term $\left|\tilde{\mathbf{X}}_H\right|^2$ will appear $n-p$ times, *i.e.*,

$$\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2 R_{iJ} = (n-p) \sum_{J \ni i} \left|\tilde{\mathbf{X}}_J\right|^2, \qquad (4.5.2)$$

(for this result, see also Hawkins, Bradu and Kass, 1984).

Dividing both sides by $(n-p)\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2$, the above equation becomes

$$\frac{\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2 R_{iJ}}{(n-p)\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2} = \frac{\sum_{J \ni i} \left|\tilde{\mathbf{X}}_J\right|^2}{\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2}.$$

Thus, using (4.5.1)

$$h_{i(i)} = \frac{\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2 R_{iJ}}{(n-p)\sum_{J \not\ni i} \left|\tilde{\mathbf{X}}_J\right|^2}.$$

Dividing the numerator and denominator by $\sum_J \left|\tilde{\mathbf{X}}_J\right|^2$, the relation becomes

$$h_{i(i)} = \frac{\sum_{J \not\ni i} \omega_J R_{iJ}}{(n-p)\sum_{J \not\ni i} \omega_J}. \qquad (4.5.3)$$

The first result of (i) is proved.

Dividing the numerator and denominator by $\sum_{J \not\ni i} \omega_J$, the relation becomes

$$h_{i(i)} = \frac{\tilde{R}_i}{n-p}, \quad \text{where } \tilde{R}_i = \frac{\sum_{J \not\ni i} \omega_J R_{iJ}}{\sum_{J \not\ni i} \omega_J}, \qquad (4.5.4)$$

proving the second result of (i).

Proof of (ii)

From (2.4.1), we have

$$h_i = \frac{h_{i(i)}}{1+h_{i(i)}}$$

Substituting $h_{i(i)}$ from (4.5.4), (ii) follows immediately.

Proof of (iii)

Using (2.4.1) again with (4.5.3) gives

$$h_i = \frac{\sum_{J \not\ni i} \omega_J R_{iJ}}{(n-p)\sum_{J \not\ni i} \omega_J + \sum_{J \not\ni i} \omega_J R_{iJ}} \qquad (4.5.5)$$

Dividing (4.3.2) by $\sum_J |\mathbf{X}_J|^2$, the relation becomes

$$\sum_{J \not\ni i} \omega_J R_{iJ} = (n-p)\sum_{J \ni i} \omega_J \ .$$

Substituting $(n-p)\sum_{J \ni i} \omega_J$ for $\sum_{J \not\ni i} \omega_J R_{iJ}$ the denominator of (4.3.5) becomes

$$(n-p)\left(\sum_{J \not\ni i} \omega_J + \sum_{J \ni i} \omega_J\right) \text{ but } \sum_{J \not\ni i} \omega_J + \sum_{J \ni i} \omega_J = 1$$

$$\therefore\ h_i = \frac{\sum_{J \not\ni i} \omega_J R_{iJ}}{(n-p)},$$

(iii) is proved. □

Remark:

In (i) we rewrote the crude expression for $h_{i(i)}$ given by Hawkins *et al.* (1984) in terms of the ERW and $\tilde{R}_i$. In (ii) and (iii) we extended the same ideas to h_i. However, for the purpose of determining multiple leverage points, it is more convenient to use (iii).

In the following subsection we will now use these results and the fact that RQs are specific ESs to derive multiple case leverage diagnostics for RQs.

4.5.1 Connection between ESs and RQs and implementation of multiple case diagnostics to RQs

Koenker and Bassett (1978) explicitly showed that each RQ corresponds to a specific ES (see also Theorem A.3) as follows:

- Let $\mathrm{B}(\tau) = \left\{\hat{\boldsymbol{\beta}}(\tau_1), \hat{\boldsymbol{\beta}}(\tau_2), ..., \hat{\boldsymbol{\beta}}(\tau_{K^*})\right\}$, for $0 < \tau_1 < \tau_2 < ... < \tau_{K^*-1} < \tau_{K^*} < 1,$ be the complete set of solutions to the LP problem (3.2.1) giving K^* RQs, where $K^* < K$.
- The solutions to the LP problem (3.2.1) do not change over the intervals $[\tau_{k-1}, \tau_k),$ for $k = 2, ..., K^*$. Thus, the lowest RQ will be obtained at τ_1 and the highest at τ_{K^*}
- If $\tilde{\mathbf{X}}$ has rank $p,$ it can be shown that the set of RQs, $\mathrm{B}(\tau)$ has at least one element of the form $\hat{\boldsymbol{\beta}}(\tau_k) = \tilde{\mathbf{X}}_{J_k}^{-1}\mathbf{Y}_{J_k}$, an expression similar to (2.1.1).
- Thus, every basic optimal solution of (3.2.1) in the simplex tableau which corresponds to a RQ, is a specific elemental set.

Alternatively, RQs can be determined using ES methods.

Defining the J^{th} elemental regression as

$$\hat{\tilde{\boldsymbol{\beta}}}_J = \left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right)^{-1}\tilde{\mathbf{X}}_J'\tilde{\mathbf{Y}}_J = \tilde{\mathbf{X}}_J^{-1}\tilde{\mathbf{Y}}_J,$$

where $\mathbf{X}_J$ is square and assumed to be non-singular, and the ES residuals as

$$e_i = \begin{cases} 0, & i \in J \\ Y_i - \tilde{\mathbf{x}}_i'\hat{\tilde{\boldsymbol{\beta}}}_J, & i \in I \end{cases}, \; i = 1, 2, ..., n.$$

RQs can obtained by minimizing the weighted sum of the ES residuals.

Actually, based on the ES approach, the θ^{th} regression quantile (RQ), $0 < \theta < 1,$ is defined as any solution to the minimization problem:

$$\min_J \left[\sum_i \theta e_i^+ + \sum_i (1-\theta) e_i^-\right], \tag{4.5.6}$$

where $e_i^+, \; e_i^- \geq 0$. Thus, again we see that a RQ is a specific ES.

Now, instead of doing the expensive diagnostics for all the K ESs, we use the property that the ES predictive weighted leverage sum to a constant (see Theorem 4.3 (iii)) in order to

diagnose the K^* RQs (ESs). Note that K^* is much smaller than K.

In section (2.4) we defined

$$\mathbf{H}_J = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}'.$$

Clearly the diagonal elements h_{jJ} (see section 2.4) of $\mathbf{H}_J$ are given by

$$h_{jJ} = \tilde{\mathbf{x}}_j'\left(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J\right)^{-1}\tilde{\mathbf{x}}_j, \quad j = 1,...,n.$$

For $j \notin J$, these were previously defined as R_{iJ}, the residual freedom (see section 2.4 (ii)). Also, for $j \in J$ it is easy to see that

$$h_{jJ} = 1$$

(see also section 2.4).

Multiplying these diagonal elements by ω_J and summing over j and then over J, gives

$$\begin{aligned}\sum_J \sum_j \omega_J h_{jJ} &= \sum_J \sum_{j\notin J} \omega_J h_{jJ} + \sum_J \sum_{j\in J} \omega_J h_{jJ} \\ &= \sum_j \sum_{J\not\ni j} \omega_J h_{jJ} + \sum_J \omega_J \sum_{j\in J} 1 \\ &= \sum_j (n-p)h_j + \sum_J \omega_J p \\ &= p(n-p) + p \\ &= p(n-p+1),\end{aligned}$$

since $\sum_j h_j = p$ and using Theorem 4.3 (iii).

Thus, excluding $h_{jJ} = 1,\ j \in J$, we have

$$\sum_J \left(\frac{\sum_{j\notin J} \omega_J R_{jJ}}{n-p} \right) = p,$$

since $R_{jJ} = h_{jJ},\ j \notin J$.

We propose the **ES predictive weighted leverage statistic** to be

$$T_J = \frac{\sum_{i\notin J} \omega_J R_{iJ}}{n-p}, \qquad (4.5.7)$$

given in table 4.1 as indicative of how large ES (RQ) leverage is.

Remarks: 1. Following table 4.1 below, we consider single case leverage diagnostics and use that in order to motivate the use of T_J as a **multiple case leverage** statistic.

2. Note that $(n-p)T_J$ is the weighted sum of the diagonal elements of $H \equiv \tilde{\mathbf{X}}_I(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\tilde{\mathbf{X}}_I'$, $J = 1, 2, ..., K$.

In table 4.1 below we give a schematic representation of all the various elements that make up this last result as the grand total, p.

Table 4.1: Relationship between ES predictive weighted leverage, T_J and OLS Leverage, h_i

Obs	ELEMENTAL SETS						Leverage
	1	2	.	.	.	K	
$i=1$	$\frac{\omega_1 R_{11}}{n-p}$	0	.	.	.	$\frac{\omega_K R_{1K}}{n-p}$	h_1
$i=2$	$\frac{\omega_1 R_{21}}{n-p}$	$\frac{\omega_2 R_{22}}{n-p}$	.	.	.	$\frac{\omega_K R_{2K}}{n-p}$	h_2
$i=3$	0	0	.	.	.	$\frac{\omega_K R_{3K}}{n-p}$	h_3
$\vdots$	$\vdots$	$\vdots$	.	.	.	$\vdots$	$\vdots$
$i=n$	$\frac{\omega_1 R_{n1}}{n-p}$	0	.	.	.	$\frac{\omega_K R_{nK}}{n-p}$	h_n
Sum	T_1	T_2	.	.	.	T_K	p

Remark: Note that, in order to make the statistics T_J a weighted average rather than just a linear combination of the R_{iJ}'s we used the rule

$$R_{iJ} = \begin{cases} h_{iJ} & for\ i \notin J \\ 0 & for\ i \in J \end{cases}.$$

The component, $T_J = \frac{\sum_{i \notin J} \omega_J R_{iJ}}{n-p}$ in the column sums is a sum of the ES predictive weighted

leverage of a particular ES J. For the p observations in ES J, we have $h_{jJ} = 1$ (these are not interesting and we replaced them with 0's in the above table).

Remark: Table 4.1 gives both row sums (weighted leverage values corresponding to a specific observation) and column sums (predictive weighted leverage values corresponding to a specific ES). However, in this thesis we are interested in column sums (ES) corresponding to the entire LP solution set giving RQs, since they are specific ESs (see, Theorem A.3 in appendix A).

As a point of departure, we first consider the single case leverage diagnosis and then extend the methodology analogously to multiple (ES) case leverage diagnosis.

The usual procedure for determining single case leverage values is as follows:

- we have n observations,
- h_i, the diagonal element of the hat matrix, $\mathbf{H} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$ is the indicator of leverage,
- the leverage values sum to a constant, *i.e.*, $\sum_{i=1}^{n} h_i = p$,
- therefore the average leverage value is given by $\frac{p}{n}$, and
- if the leverage value h_i is greater than a threshold value (say, $\frac{2p}{n}$ or $\frac{3p}{n}$) (2 times or 3 times the average leverage value), then h_i is classified as large (see, *e.g.*, Hoaglin and Welsch, 1978 and Belsley *et al.*, 1980 respectively).

Remark: In this thesis we refer to ES leverage values as multiple case leverage values. Furthermore, we use the fact that the leverage values of all the ESs sum to a constant in order to detect multiple leverage cases for those ESs corresponding to the set of RQs.

We extend the single case leverage diagnostics to the multiple case in the following procedure, with the column sums (ESs) as the analogues of the observations in the single case.

- We take the K column sums corresponding to the number of ESs as the analogue of the number of observations n, in the single case,
- the sum of the column sums in table 4.1 sum to a constant, *i.e.*,

$$\sum_J \left(\frac{\sum_{i \notin J} \omega_J R_{iJ}}{n-p} \right) = p ,$$

- therefore the average ES predicted leverage value is given by $\frac{p}{K}$, and
- if the ES predictive weighted value $T_J = \frac{\sum_{i \notin J} \omega_J R_{iJ}}{n-p}$ is greater than a threshold value (say, $\frac{2p}{K}$ or $\frac{3p}{K}$) then ES J could be viewed as a high leverage set. (Note that the same then holds for the corresponding RQ, see also Chapter 5.)

Remark: The Huber (1981) single leverage points, having $h_i > 0.2$ and $h_i > 0.5$ as threshold values (see section 4.3), can be extended to the multiple case leverage points to obtain $\frac{0.2n}{K}$ or $\frac{0.5n}{K}$ as threshold values. Furthermore, threshold values $\frac{4p}{K}$ *or* $\frac{5p}{K}$ can be used if there is need for the researcher to be more aggressive.

Using $0 \le \omega_J \le \frac{p}{n}$ the upper bounds of the values of predictive weighted leverages $\omega_J \frac{\sum_{i \notin J} R_{iJ}}{n-p}$ can be obtained easily using the relation (Theorem 4.3 (i)) of the OLS jackknife (with observation i deleted), $h_{i(i)}$ leverage to ES predictive weighted leverage, *viz.*,

$$\frac{\sum_{J \not\ni i} \omega_J R_{iJ}}{n-p} = h_{i(i)} \sum_{J \not\ni i} \omega_J .$$

Recall from (2.4.1) that $h_{i(i)}$ and h_i are monotonically related, *i.e.*,

$$h_{i(i)} = \frac{h_i}{1-h_i} .$$

Using the threshold values of h_i, say $\frac{p}{n}$ and $\frac{2p}{n}$, the threshold values of $h_{i(i)}$ can be obtained as

$$\frac{p}{n-p} \text{ and } \frac{2p}{n-2p}$$

respectively.

The threshold for the ERW, ω_J can be taken to be $\frac{p}{2n}$, say since $0 \le \omega_J \le \frac{p}{n}$.

From these we deduce the threshold values of $T_J = \omega_J \frac{\sum_{i \notin J} R_{iJ}}{n-p}$.

Using $\frac{p}{n}$, for instance as a threshold value of h_i, the upper bound of the row sums (over ESs) for each i is given by

$$\frac{\sum_{J \ni i} \omega_J R_{iJ}}{n-p} = \frac{p^2}{n(n-p)} \binom{n-1}{p}.$$

Summing over i gives

$$\sum_i \left(\frac{\sum_{J \ni i} \omega_J R_{iJ}}{n-p} \right) = \frac{p^2}{(n-p)} \binom{n-1}{p}.$$

The upper bound of the average of column sums, $T_J = \frac{\omega_J \sum_{i \notin J} R_{iJ}}{n-p}$ is given by

$$\frac{p^2}{n-p} \binom{n-1}{p} \Big/ \binom{n}{p} = \frac{p^2}{n}.$$

However, this statistic tends to suffer from the usual extremely large number of ESs, needed in order to obtain cut-off values.

We will apply this to our examples in Chapter 5. In the next section we discuss outliers and influential observations.

Remark: Since the **ES predictive weighted leverage** statistics sum to p, *i.e.* $\sum_J T_J = p$, T_J can be standardized by dividing it by p to obtain the statistic T_J^s such that $0 \le T_J^s \le 1$ and $\sum_J^K T_J^s = 1$. However, in the simulation studies the unstandardized T_J will be used.

4.6 Collinearity-influential observations in RQs (ES)

Outliers in the regressor space are often referred to as leverage observations (Mason and Gunst, 1985). Chatterjee and Hadi (1988) remark that high-leverage observations tend to influence the eigenstructure of the regressor matrix **X** and hence its collinearity structure. They refer to observations that create or obscure collinearity as **collinearity-influential observations**. Therefore a small condition index (a large determinant of the correlation matrix or variance inflation factor (VIF)) does not necessarily imply absence of collinearity. On the other hand, one or two observations may be responsible for a small or large determinant of the correlation matrix. We refer to observations that create collinearity as **type A observations** and to observations that obscure collinearity as **type B observations**. Type A observations inflate variances, covariances and correlations. The converse is true for type B observations. Type B observations sustain spurious dimensions which may lead to non-significant regressors being included in the regression model. This is a potential danger in automated regression procedures. Note that while collinearity-influential observations usually have high leverage, not all high leverage observations are collinearity-influential observations (Chatterjee and Hadi, 1988). These observations, a combination of type A and type B observations, are labelled type AB observations. They have little or no effect on the correlation structure of the regressors.

Mason and Gunst (1985) used the Gunst and Mason data set (1980, Appendix C) (see Chapter 5) to illustrate collinearity influential observations as shown in figure 4.1 below.

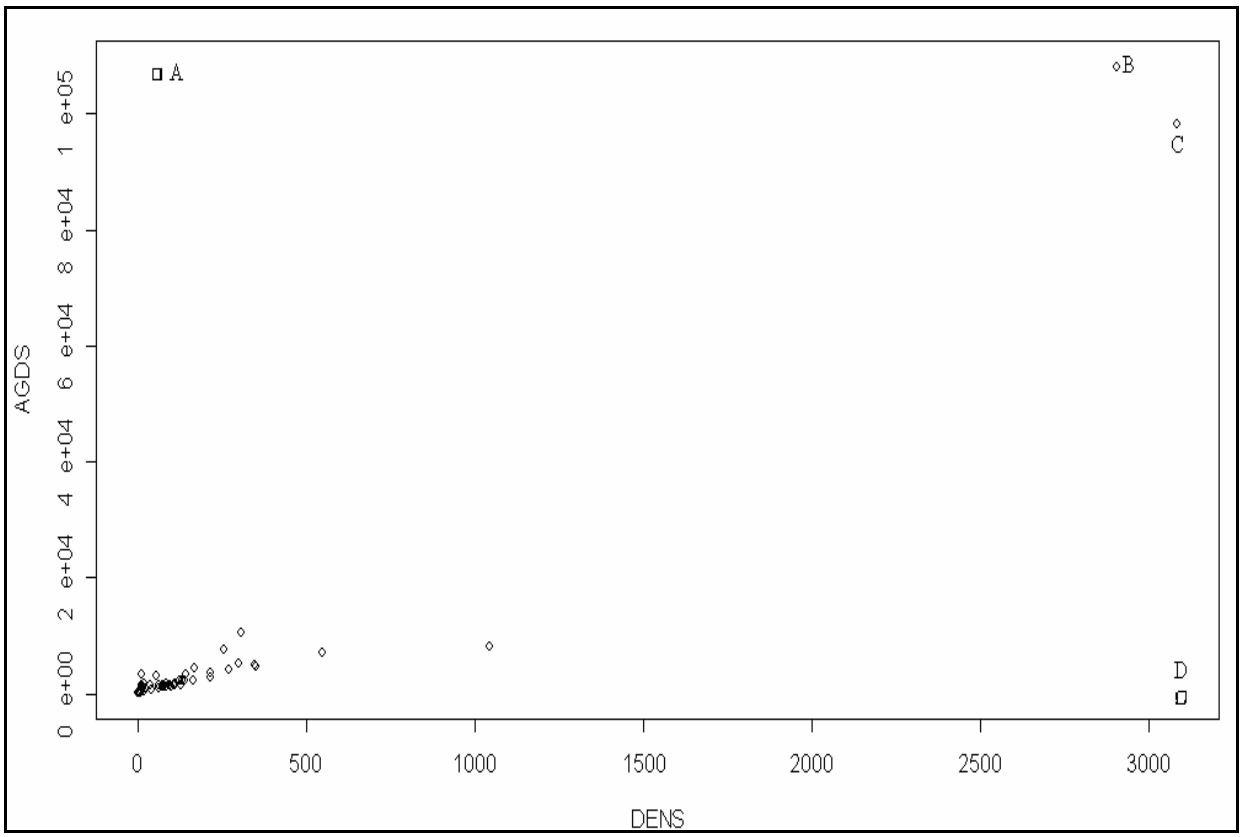


Figure 4.1: Scatterplot of two predictor variable (DENS and AGDS) in the Gunst and Mason Data set.

Remark: A and D are fictitious points that were added to stress that leverage points need not necessarily induce collinearities.

The occurrence of observations A and D (individually or together in a data set) without B and C would not necessarily constitute collinearity influential observations. However, both observations will be diagnosed as leverage points. On the other hand, the occurrence of observations B and C close to one another without A and D would induce a collinearity in $\mathbf{X}$. However, if they occur individually in a data set they might only be detectable as leverage points.

The occurrence of type B points is best illustrated by the Hocking and Pendleton (1983) data set. Observation 24 is a high leverage point with $h_i = 0.927 > \frac{2p}{n} = 0.308$ and hides (distorts) the collinearity in this data set. The bivariate scatter plot of $(X_1 \; vs \; X_2)$ does not suggest any extreme behaviour as shown in figure 4.2 below.

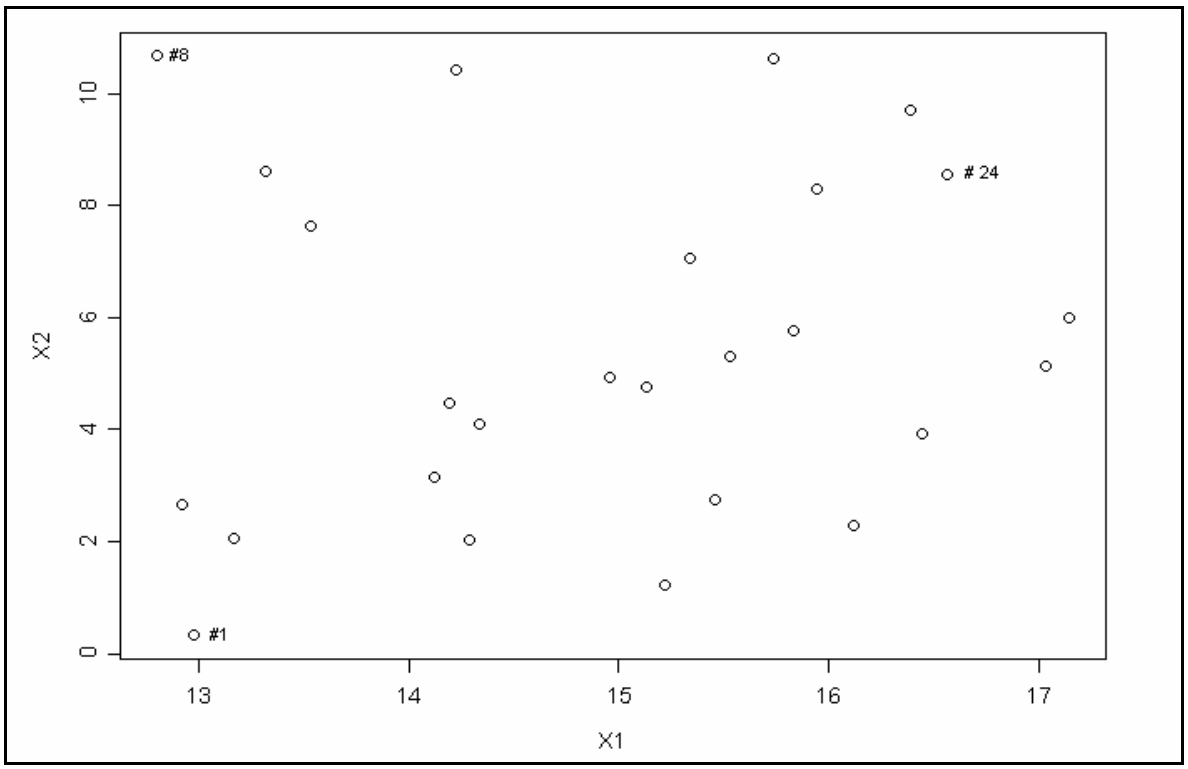


Figure 4.2: The Hocking data set scatter plot of X_1 vs X_2.

On the other hand, the bivariate scatter plots of ($(X_2 \ vs \ X_3)$ and $(X_1 \ vs \ X_2)$), given in figures 4.3 and 4.4 below, do not suggest any extreme behaviour, except that observation 24 is somewhat isolated.

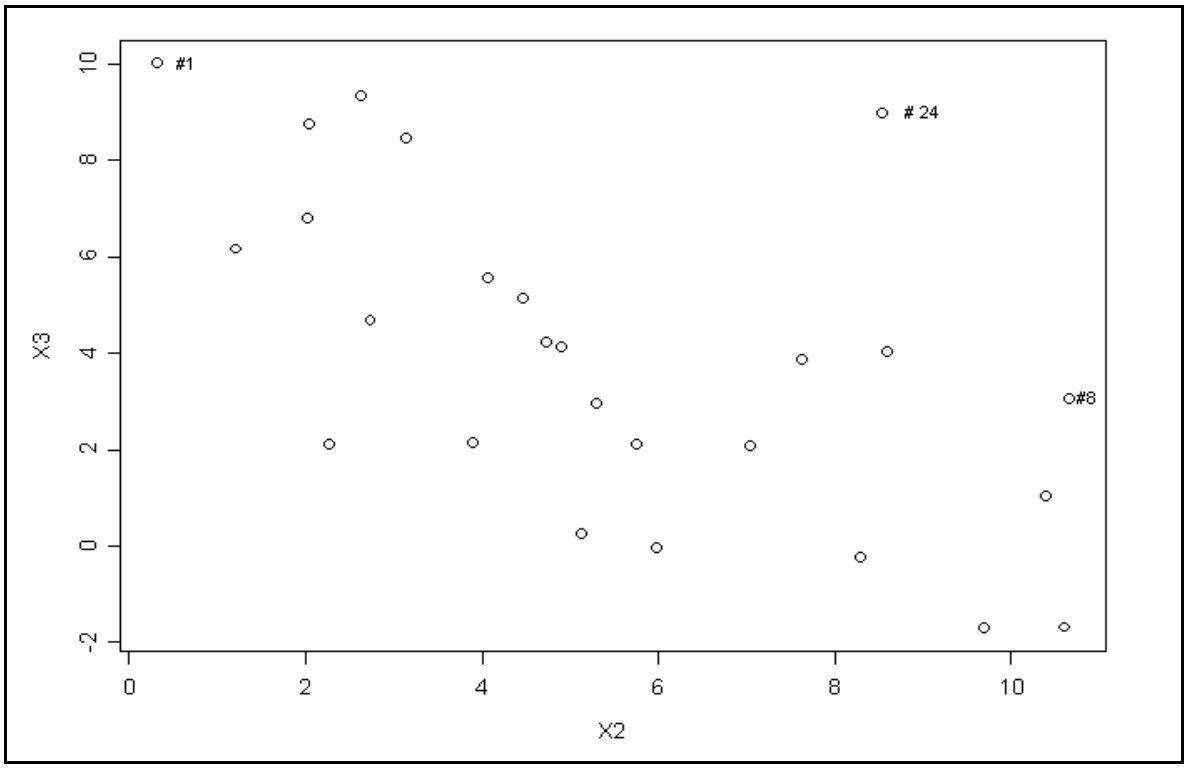


Figure 4.3: The Hocking data set scatter plot of X_2 vs X_3.

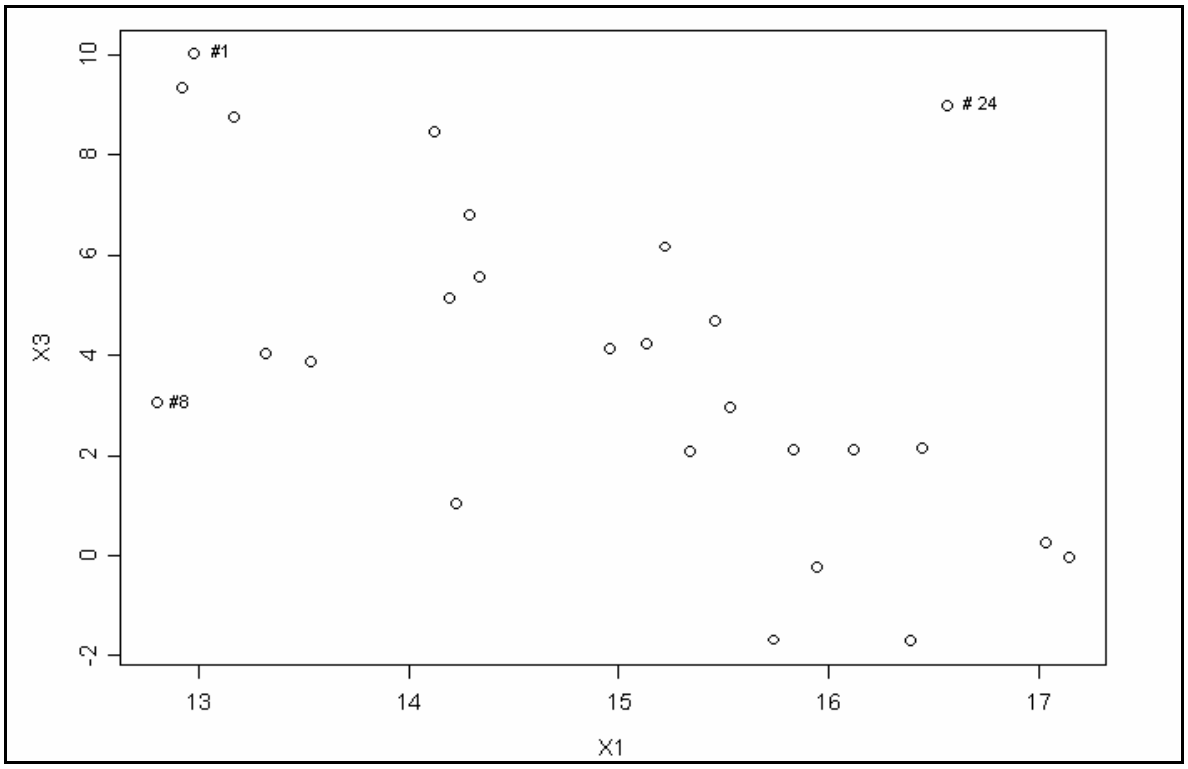


Figure 4.4: The Hocking data set plot of X_1 vs X_3.

In Chapter 5 we assess multiple leverage points and collinearity influential observations and their effect on RQs via elemental set methods. In the next section we focus on prediction and assess the effect of collinearity in RQ prediction.

4.7 Prediction in RQs (ESs)

Predictions are the values of the response variable, $\hat{Y}_{(i)}$ obtained when the prediction equation

$$Y_j = \beta_0 + \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_j, \text{ with } \varepsilon_j \text{ having df } F,$$

is applied to the data, with observation i deleted from the data when fitting the model.

In this single (usual) case scenario, identification of poor prediction is done by comparing the predicted residuals sum of squares, $PRESS$ (see remark in subsection 2.4.1) to the ordinary residual sum of squares SSE.

Poor prediction occurs if

$$PRESS \geq 2SSE, \tag{4.7.1}$$

(see *e.g.* Freund and Wilson, 1998).

Our original idea was to extend this procedure in an analogous fashion to the RQ/ES situation, *viz.*,

$$PRESS_J \geq \tilde{a}_J SSE, \tag{4.7.2}$$

for some constant $\tilde{a}_J$, where $PRESS_J$ is the elemental predicted residual sum of squares. However, for RQs prediction is normally poor at extreme τ levels and excellent at the middle ones for symmetrical error distributions. Hence a_J is a function of τ rather than a constant. An attractive candidate which mimics this function is a robust loss function which does not penalize the models that fit the majority of the data despite having big prediction errors at a few outlying points in the validation sample, is

$$\sum_{k \notin J} \rho(e_{kJ}).$$

This criterion can be viewed as a robust criterion for a location problem between the observed values and the fitted values. We make use of one such typical loss function suggested by Ronchetti *et al.* (1997), *viz.*

$$\rho(t) = \min\{t^2, b^2 \hat{\sigma}_e^2\}, \tag{4.7.3}$$

where $b = 1.345$ and $\hat{\sigma}_2 = 1.483 med_{k \notin J} \left| e_k - med_{i \notin J}(e_{iJ}) \right|$.

In Chapter 5, we determine reasonable cut-off values of a_J via a simulation study using a sinusoidal function. We compare the cut-off values from the robust loss function (4.7.3) and those from the sinusoidal loss function.

Also, we exploit the relationship (*i.e.*, SSE is a weighted sum of $PRESS_J$'s) between ordinary residual sum of squares, SSE and elemental predicted residual sum of squares, $PRESS_J$ via the ES weight to assess the effect of collinearity on ES/RQ prediction. Item (i) of Theorem 4.4 below relates a single OLS residual via all the ESs. However, since we are only interested in the specific ESs corresponding to RQs, item (i) is therefore not useful for the purposes of this research. On the other hand, item (ii) of the theorem plays a very important role in evaluating the effect of collinearity (small ERW) since it relates ordinary residuals and elemental predicted residuals (EPR) via the ES weight.

Theorem 4.4: Let $\mathbf{X}_J$ be non-singular. Then,

(i) the OLS residual can be expressed as a sum of weighted EPR's, *viz.*,

$$e_i = \sum_{J \ni i} \omega_J e_{iJ}, \; 1 \le i \le n.$$

(ii) the OLS error sum of squares can be written as

$$SSE = \frac{\sum_J \omega_J . PRESS_J}{p+1}.$$

Proof: The proof of (i) is given by Hawkins *et al.* (1984). The latter gave an incorrect result corresponding to (ii), *viz.*

$$SSE = \sum_J \omega_J . PRESS_J \; .$$

The proofs of both (i) and (ii) are given in appendix A. □

Schematically, this theorem can be represented as in table 4.2.

Table 4.2: Relationship between the EPR Sum of Squares, $PRESS_J$ and SSE.

Obs	ELEMENTAL SETS						Sum
	1	2	.	.	.	K	
$i=1$	$\frac{\omega_1 e_{11}^2}{p+1}$	$\frac{\omega_2 e_{12}^2}{p+1}$	.	.	.	$\frac{\omega_K e_{1K}^2}{p+1}$	$\sum_J \frac{\omega_J e_{1J}^2}{p+1}$
$i=2$	$\frac{\omega_1 e_{21}^2}{p+1}$	$\frac{\omega_2 e_{22}^2}{p+1}$	.	.	.	$\frac{\omega_K e_{2K}^2}{p+1}$	$\sum_J \frac{\omega_J e_{2J}^2}{p+1}$
$i=3$	⋮	⋮	.	.	.	⋮	⋮
⋮	⋮	⋮	.	.	.	⋮	⋮
$i=n$	$\frac{\omega_1 e_{n1}^2}{p+1}$	.	.	.	.	$\frac{\omega_K e_{nK}^2}{p+1}$	$\sum_J \frac{\omega_J e_{nJ}^2}{p+1}$
Sum	U_1	U_2	.	.	.	U_K	SSE

In the prediction scenario the analogue of the **predictive weighted leverage** statistic, T_J is the

weighted predicted residual sum of squares

$$U_J = \frac{\omega_J PRESS_J}{p+1},$$

or its standardized version

$$U_J^s = \frac{\omega_J PRESS_J}{(p+1)SSE},$$

such that $\sum_J^K U_J^s = 1$. In this form, U_J^s is bounded between zero and one and therefore can be viewed as the proportion of poor prediction contributed to by ES J. However, this statistic tends to suffer from the usual extremely large number of ESs, needed in order to obtain cut-off values. Furthermore, this statistic comprises both X information ("leverage") in the ERW, ω_J and Y information in the ratio, $PRESS_J / SSE$ which may cause one to view this statistic as an influential measure rather than a prediction measure. In order to circumvent this we propose as a prediction measure, the statistic

$$a_J = \frac{PRESS_J}{SSE}$$

and determined its cut-off values using simulation studies in Chapter 5.

Remark: In future research we plan to return to U_J in order to further investigate its potential use to detect designs sensitive to outliers.

In the following section we give a brief overview of infuential diagnostics and propose "new" ones for the RQ/ES situation. We will see that influential diagnostics are functions of leverage and residuals ($PRESS_J$ in the case of RQs/ESs).

4.8 Influential observations

In the classical linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, it is assumed that $\varepsilon \sim N(\mathbf{0}, \sigma \mathbf{I}_n)$ so that $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$. However, in practice data often deviate from the Gaussian assumption due to the prevalence of outliers in the response variable. An outlier is a point that is markedly different from the other observations. Often extreme observations in the independent (X-space) variable also occur. These are called leverage points (see section 4.3).

Early methods of locating and testing for the significance of a single outlier involved the least squares residuals,

$$e_i = Y_i - \tilde{\mathbf{x}}'\hat{\tilde{\boldsymbol{\beta}}}$$

or its scaled version,

$$e_i(\sigma) = \frac{e_i}{\sigma\sqrt{1-h_i}}.$$

Single case influence diagnostics involve mainly scaled residuals from which two special statistics are computed, *viz.*, t_i (4.3.4) and t_i^* (4.3.5).

In the literature t_i^* is preferred to t_i (see *e.g.* Belsley, Kuh and Welsch, 1980) and are sometimes referred to as the "jackknife" residuals (see Atkinson, 1981) or the "studentized" residuals (see Velleman and Welsch, 1981).

In regression analysis it is of interest to identify observations having an influential effect on

(i) estimation (thus on $\hat{\tilde{\boldsymbol{\beta}}}$),

(ii) prediction of the fit (estimated variance, that is $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ and $\hat{\mathbf{Y}}$).

An observation which is a leverage (see section 4.3) and/or an outlier (see section 4.7) and has undesirable effects (in terms of (i) and/or (ii)) on various regression statistics, is referred to as an influential point (see, *e.g.*, Chatterjee and Hadi, 1986).

Remark: Not all leverage and/or outlier points are necessarily influential.

In the literature there are several influence measures (Belsley *et al.*, 1980 or Cook and Weisberg, 1982). Cases that are influential according to one influence measure may not be influential with respect to another measure.

While single outlier problems have been effectively dealt with in the literature, multiple outliers problems continue to be challenge (see, *e.g.* Barrett and Gray, 1997b) due the following reasons:

(i) masking, which makes outliers appear inlying, and
(ii) swamping, which makes inliers appear outlying.

These problems can occur individually or simultaneously and single case diagnostics have

been found to be ineffective in the presence of masking and swamping (see, *e.g.*, Rousseeuw and van Zomeren, 1990; Barrett and Gray, 1997a). As a consequence, the identification of multiple cases has been an important research area since the 1980's (see, Barrett and Gray, 1997b) and several procedures have been proposed (see, *e.g.*, Cook and Weisberg, 1982; Gray and Ling, 1984; Barrett and Gray, 1992, 1995; Hadi and Simonoff, 1993). Most of these procedures involve the extension of existing single case diagnostics to subsets of cases (multiple cases), *e.g.*, Cook's distance D_i (see section 4.3) was extended to D_I (see section 4.4).

In this study we are interested in identifying multiple case influential points in the RQ situation. Hence, we are interested in those multiple case diagnostics that can (easily) be useful in the RQ case. This is despite the fact that a RQ is an ES and thus suffers from the shortcomings that affect ESs (see section 2.4). One such class of diagnostics that "conform" to the RQ situation are the volumes of confidence ellipsoids which are functions of determinants. These measures tie in nicely in the RQ situation via the ERW, ω_J which is the ratio of the determinant of the ES design matrix to that of the entire design matrix. Here we extend two such diagnostics, *viz.*, the covariance ratio, CVR_i suggested by Belsley *et al.* (1980) and the Andrews-Pregibon statistic, AP_i, suggested by Andrews and Pregibon (1978).

The covariance ratio, CVR_i is the ratio of the generalized variance of $\hat{\boldsymbol{\beta}}$ given by $\left|\sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|$ to that of $\hat{\boldsymbol{\beta}}_{(i)}$ given by $\left|\hat{\sigma}^2_{(i)}(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|$, *viz.*,

$$CVR_i \equiv \frac{\left|\hat{\sigma}^2_{(i)}(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|}{\left|\hat{\sigma}^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|} = \frac{\hat{\sigma}^{2p}_{(i)}\left|(\tilde{\mathbf{X}}'_{(i)}\tilde{\mathbf{X}}_{(i)})^{-1}\right|}{\hat{\sigma}^{2p}\left|(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|} = (\hat{\sigma}^2_{(i)}/\hat{\sigma}^2)^p/(1-h_i) = \left(\frac{n-p-t_i^2}{n-p-1}\right)^p \Big/ (1-h_i).$$

Here $\hat{\sigma}$, $\hat{\sigma}_{(i)}$, and t_i are as in (4.3.4) (see, also Belsley, Kuh, and Welsch, 1980). A rough indication of an influential point is $\left|CVR_i - 1\right| > \frac{3p}{n}$.

Recall that for the case $n = p+1$ (leave one out), $\omega_J^{-1} = 1/(1-h_i)$ (see Remark 1, section 4.4). Hence, when $n = p+1$

$$CVR_i = (\hat{\sigma}^2_{(i)}/\hat{\sigma}^2)^p/\omega_J = \left(\frac{n-p-t_i^2}{n-p-1}\right)^p \Big/ \omega_J .$$

Note that, viewing a RQ as an ES in the case where σ is unknown, we cannot estimate the

variance based on the ES procedure, because of the exact fit property ($SSE_J = 0$) (see section 2.4). We make use of the fact that the OLS residual sum of squares, SSE can be expressed as a weighted average of elemental predicted residual sum of squares (see Theorem 4.4, section 4.7) to propose a "pseudo" estimate of σ based on ES, given by

$$\hat{\sigma}_J^2 = \frac{\omega_J PRESS_J}{(p+1)(n-p)}.$$

(Thus, $\hat{\sigma}^2 = \sum_J \hat{\sigma}_J^2$).

Dropping the scaling factor, $1/(p+1)$ for convenience, this will lead us to an analogue of the covariance ratio for the RQ/ES situation given by

$$CVR_J \equiv \frac{\left|\hat{\sigma}_J^2(\tilde{\mathbf{X}}_J'\tilde{\mathbf{X}}_J)^{-1}\right|}{\left|\hat{\sigma}^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\right|} = \left(\frac{\omega_J PRESS_J}{SSE}\right)^p \omega_J^{-1}, \qquad (4.8.1)$$

with cut-off value $\left(\frac{\tilde{c}\tilde{a}}{K}\right)^p \left(\frac{K}{\tilde{c}}\right)$.

The Andrews-Pregibon statistic is given by

$$\begin{aligned} AP_i &\equiv \frac{\left|\mathbf{X}_{(i)}^{*\prime}\mathbf{X}_{(i)}^{*}\right|}{\left|\mathbf{X}^{*\prime}\mathbf{X}^{*}\right|} \\ &= 1 - h_i - e_i^2/SSE, \\ &= 1 - h_i^* \end{aligned}$$

where $\mathbf{X}^* = (\mathbf{X} : \mathbf{Y})$ and $h_i^* = h_i - e_i^2/SSE$ measures the relative change in $\left|\mathbf{X}^{*\prime}\mathbf{X}^{*}\right|$ when the i^{th} observation is deleted. Small values of AP_i call for special attention. Note that, AP_i does not distinguish between a leverage point and an outlier.

The extension of this statistic to the RQ situation again suffers a drawback from the exact fit property (see section 2.4) as follows:

The denominator in AP_i, can be expressed as $\left|\mathbf{X}^{*\prime}\mathbf{X}^{*}\right| = SSE.\left|\mathbf{X}'\mathbf{X}\right|$. For the multiple case analogue of AP_i, the numerator can be expressed as $\left|\mathbf{X}_J^{*\prime}\mathbf{X}_J^{*}\right| = SSE_J\left|\mathbf{X}_J'\mathbf{X}_J\right|$ (see also, Chatterjee and Hadi, 1986). However, since $\left|\mathbf{X}_J^{*\prime}\mathbf{X}_J^{*}\right| = SSE_J\left|\mathbf{X}_J'\mathbf{X}_J\right| = 0$, we rather propose as

an analogue to AP_i, the statistic

$$AP_J \equiv \frac{PRESS_J \left|\mathbf{X}_J'\mathbf{X}_J\right|}{SSE\left|\mathbf{X}'\mathbf{X}\right|} = \frac{PRESS_J}{SSE}\rho_J. \qquad (4.8.2)$$

Here the statistic ρ_J is as defined in subsection 4.2.2, equation 4.2.2. Recall that this statistic gives the collinearity view of the ERW.

Remark: Since the Andrews-Pregibon statistic does not distinguish between high leverage points and outliers (see *e.g.* Chatterjee and Hadi, 1986), we only considered the covariance ratio in further simulations.

In view of collinearity influential observations, which create collinearity in the design space, shrinkage procedures can be proposed to remedy the effect of collinearity. These comprise, *inter alia*, ridge regression, first proposed by Hoerl (1962). In the next section we give an overview of these shrinkage procedures.

4.9 Shrinkage techniques

To counter the problems of high degrees of collinearity in the ordinary least squares scenario, many shrinkage procedures alternatives have been proposed. These include ridge regression, amongst others.

Traditionally, the ridge estimator is defined as,

$$\hat{\boldsymbol{\beta}}^{ridge} = (\mathbf{X}^{s'}\mathbf{X}^{s} + \lambda\mathbf{I})^{-1}\mathbf{X}^{s'}\mathbf{Y}, \qquad (4.9.1)$$

where $\mathbf{X}^s$ is the design matrix (without the constant covariate) standardized to correlation form (see Hoerl *et al.* 1975) and $\lambda \geq 0$ (normally $0 < \lambda < 1$), is a complexity parameter that controls the amount of shrinkage.

Thus, adding a small positive constant λ stabilizes $\mathbf{X}'\mathbf{X}$ making it nonsingular (see Hoerl and Kennard, 1970). This was the main motivation for ridge regression when it was first introduced, but since then, its scope has increased extensively in the literature.

Delving into singular value decomposition (SVD) of the standardized predictor matrix $\mathbf{X}^s$, clearly shows that the ridge beta coefficients are directly related to the eigenstructure of the

correlation matrix (see Hastie, Tibshirani and Friedman, 2001).

The ridge coefficients minimize a penalised residual sum of squares:

$$\hat{\boldsymbol{\beta}}^{ridge} = \underset{\boldsymbol{\beta}}{\arg\min} \left\{ \sum_{j=1}^{N} \left(Y_i - \beta_0 - \sum_{j=1}^{p-1} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p-1} \beta_j^2 \right\}, \tag{4.9.2}$$

where x_{ij} are standardized. The coefficients are shrunk towards zero.

Equivalently, (4.9.2) can be expressed as

$$\hat{\boldsymbol{\beta}}^{ridge} = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{j=1}^{N} \left(Y_i - \beta_0 - \sum_{j=1}^{p-1} x_{ij}\beta_j \right)^2 \tag{4.9.3}$$

$$\text{subject to } \sum_{j=1}^{p-1} \beta_j^2 \leq s,$$

where there is a one-one correspondence between the parameters λ and s. Hence (4.9.2) and (4.9.3) are equivalent.

The ridge solutions are not equivariant under scaling of the inputs, and so the inputs are normally standardized (to correlation form) before solving. The intercept β_0 is left out of the penalty term since penalization of the intercept would make the procedure depend on the origin chosen for Y, *i.e.* adding a constant d to each of the responses Y_i would not simply result in a shift of predictions by the same amount d. Ridge estimation shrinks towards zero some of the least squares coefficients, which will then result in biased parameter estimates. However, such biased estimates are sometimes useful for prediction purposes.

Since the OLS method is sensitive to outliers in the Y space, we consider combining RQs and ridge regression in situations where there is a simultaneous occurrence of Y outliers and collinearity in the data in the next subsection.

4.9.1 Combining RQs and ridge regression

We combine the minimization problem that leads to the RQs with ridge shrinkage in a natural fashion, *viz.* as

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \rho_\tau (Y_i - \beta_0 - \mathbf{x}_i'\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p-1} \left| \beta_j \right| \right\}, \ 0 < \tau < 1 \text{ and } \lambda > 0. \tag{4.9.4}$$

This type of shrinkage is referred to as the lasso (least absolute shrinkage and selection

operator) penalty $\left(\sum_{j=1}^{p-1} \left| \beta_j \right| \right)$, and it ties in better with RQs than the quadratic (ordinary ridge) penalty $\sum_{j=1}^{p} \beta_j^2 = \boldsymbol{\beta}'\boldsymbol{\beta}$ since the former penalty involves absolute values like RQs themselves. Here we refer to this process as the RQ-Lasso estimation. This minimization problem is reduced to an LP formulation by writing it in a data augmentation form. Adapting the data augmentation of Marquardt (1970) and Askin and Montgomery (1980), we need only augment the original data with p points, *viz.* $\left(\mathbf{X_{aug}}, \mathbf{Y_{aug}} \right) = \left(\lambda \mathbf{I_p}, \mathbf{0} \right)$ to solve (4.9.4) using standard linear programming techniques.

Determining the shrinkage (biasing) parameter λ has been the subject of much debate in the literature due to the data dependency of the procedures. As a result many proposals have been put forward over the last 35 years. Most of these involve some variance estimate, $\hat{\sigma}^2$ (see subsection 4.9.2) which is equal to zero in the ES scenario. Hence we make use of cross-validation.

Clearly, the ES approach already leaves us in a cross-validation situation in the sense that the p observations in the ES can be viewed as the training set while the $n - p$ observations can be viewed as the validation set. As a consequence, for (4.9.4) we take advantage of this fact since the prediction error $PRESS_J^{\lambda}$ is based on observations outside the ES corresponding to a RQ. Therefore we take the optimal value of λ to be the one that results in the lowest prediction error $PRESS_J^{\lambda}$.

In the next subsection we give a brief overview of some of these procedures used in the estimation of λ in the literature.

4.9.2 Shrinkage parameters

Hoerl and Kennard (1970) originally suggested that λ be determined from the inspection of the "ridge trace" (a plot of the $\beta_j(\lambda)'s$ versus λ) using certain guidelines involving sign reversals, stability and increase in residual sum of squares . However, this procedure, like all others, is inherently subjective. Various proposals for estimating the optimal values of λ or λ_j (in the case of generalized ridge regression) exist in the literature. For instance, McDonald and Galarneau (1975) suggest Monte Carlo studies while Hoerl and Kennard (1970) have shown that the values of the λ_i's which minimize the mean square error of $\hat{\beta}^{\lambda}$, are given by

$$\lambda_i(opt) = \frac{\sigma^2}{\beta_j^2} \approx \frac{\hat{\sigma}^2}{\hat{\beta}_j^2}, \quad j = 1, 2, ..., p,$$

where $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p}$ and $\hat{\beta}_j$ is the least squares estimate of β_j.

The majority of the proposed optimal shrinkage parameters involve the residual sum of squares, SSE or (or the variance estimate $\hat{\sigma}^2$). However, in this thesis we study RQs using the ES approach which exhibit the exact fit property resulting in $SSE_J = 0$. A more recent procedure to estimate the optimal shrinkage parameter without involving $\hat{\sigma}^2$ or SSE_J is cross validation based on the Least Angle Regressions Shrinkage (lars) procedure (see Efron *et al.*, 2004). The value of t or λ that yields the lowest prediction error is selected as the optimal one. After the determination of the lasso shrinkage parameter λ or t, (4.9.2) or (4.9.3) are evaluated using these optimal shrinkage values (see *e.g.*, Tibshirani, 1996). Since the shrinkage parameter will have been determined using the lars procedure, the estimation exhibited in (4.9.2) and (4.9.3) is referred to as lars-lasso estimation. The computational software for these procedures can be downloaded as add-on R packages at

http://cran.r-project.org/bin/windows/base/

or

http://www-stat.stanford.edu/~tibs/lasso .

In Chapter 7 we will compare results based on (4.9.3) to those based on (4.9.4).

In the literature, shrinkage procedures have been applied fairly successfully in variable selection. However, Mason and Gunst (1985), showed using an empirical study, that the ordinary ridge procedure and wide range of robust estimators fail in this situation (the presence of collinearity influential observations). In Chapter 7 we apply the RQ lasso shrinkage procedure to some well known data sets.

Remark: Note that shrinkage procedures require one to standardize the non constant term predictor matrix $\mathbf{X}$ to a one in correlation form $\mathbf{X}^s$ (see *e.g.* Marquardt and Snee, 1975; Marquardt, 1980).

4.10 Conclusions

In this chapter, existing OLS regression diagnostics were explored and further elaborated on, new relationships between OLS diagnostics and ES ones were deduced and the relationship between ESs and RQs was further elaborated on. It is upon this three-tier relationship that the thesis is based on to address the problems of collinearity, leverage and outliers in the RQ scenario.

Section 4.2 defined collinearity and considered some sources of it. Various undesirable effects of collinearity were given. From the literature survey, it was concluded that some of the manifestations of collinearity are still unclear and the diagnostic tools used often give different interpretations. Therefore it is recommended that the researcher must not rely on a single diagnostic. An overview of the single case collinearity diagnostics and their respective "cut-off values" for the detection of harmful effect of collinearity were given in subsection 4.2.1. In subsection 4.2.2 we elaborated on the relationship of the VIF to other collinearity diagnostics such as the coefficient of determination (the squared correlation coefficient), the tolerance and the determinant of the correlation matrix. Also, the ERW, a multiple case statistic was expressed in terms of these single case ones. Furthermore, the ERW was expressed as a product of a constant and two ratios. One is the ratio of the determinant of the correlation matrix of the $p-1$ non constant predictors at the ES (RQ) level to that at the full design matrix level while the other is the ratio of the product of the variabilities of the $p-1$ non constant predictors at the ES (RQ) level to that at the full design matrix level. The first one is the ratio of the degree of collinearity at the ES (RQ) level to that at the full design matrix level while the second one is the degree of variability ("leverage") at the ES (RQ) level to that at the full design matrix level.

In section 4.2.3 we showed that existing multiple case collinearity diagnostics are related to the single case ones via the ERW. Thus, cut-off values of the single case ones can be extended to the multiple case ones. Section 4.2 can also be viewed as considering the collinearity view of the ERW apart from considering other collinearity diagnostics in general.

Section 4.3 considered single case leverage diagnostics and their cut-off values. The main diagnostics used for detecting leverage in the single case scenario are the diagonal entries of the hat (projection) matrices $\mathbf{H}$ and $\mathbf{H}_{(i)}$ (see table 1.2). While in section 4.2 we considered the collinearity view of the ERW in section 4.4 we considered the leverage view of the ERW.

Furthermore, we showed that many generalized regression statistics comprise the ERW. Section 4.5 considered multiple case leverage diagnostics for ESs (RQs) showing the existing relations to single case ones as well as deducing new ones. The relationship between ESs and RQs was further elaborated on showing that a RQ corresponds to a specific ES. Based on this background we proposed a multiple case leverage statistic for ESs (RQs) and proposed some cut-off values. However, in Chapter 5 (section 5.4), it is shown that the cut-off values derived using the single case analogue are not applicable the RQ scenario. Therefore reasonable cut-off are determined using simulation studies (see section 5.8).

Clearly, leverage points which are referred to as collinearity influential points can induce or hide collinearity. Therefore in section 4.6 we considered collinearity influential points, illustrating these phenomena using well known standard data sets.

Section 4.7 considered prediction. We corrected and used Hawkins *et al.*'s (1984) original result which related ES regression prediction to OLS regression prediction to derive the ES predicted weighted residual sum of squares. Because this statistic comprises a scaling factor, a leverage factor and a prediction factor, we only made use of a prediction factor which is the ratio of RQ (ES regression) prediction to OLS regression prediction. This statistic will be very usefully employed in the simulation studies in Chapter 5.

In the RQ scenario prediction can only meaningfully be done at $\tau = 0.5$ since prediction using non central τ's is bound to be poorer as the corresponding RQ will be further away from the centre of the data (by definition). However, we considered all the τ levels using cut-off values that mimic the prediction pattern of RQs in order to have a holistic picture. Such cut-off values ensure that extreme RQ's are not classified as influential on the basis that they are extreme ones.

Like in the single case scenario, subsets of cases which are leverage points and/or outliers can be influential. One practical problem that arises in the multiple case scenario is the determination of the size of the influential case. However, as mentioned before, in the ES (RQ) scenario this is not a problem since influential sets of size $n-p$ not in the ES, corresponding to a RQ, are used to construct predictive (validation) statistics.

Section 4.8 considered influential observations. An overview of single case influential observations was first considered as a precursor to the ES (RQ) one. It was shown that influential diagnostics generally comprise a leverage component and a residual component. In the RQ (ES) scenario influential diagnostics which are functions of determinants such as the covariance ratio tie in nicely with the expression for the ERW. Actually, the ERW forms the

leverage (X information) component of the RQ (ES) diagnostic. The covariance ratio was mainly considered as a RQ (ES) influential diagnostic. In Chapter 6 the cut-off values of the multiple case covariance ratio are determined from those of the ERW and the RQ predicted residual sum of squares.

Lastly, section 4.9 briefly considered variable selection in the RQ scenario. An overview of variable selection procedures in the literature was given. However, shrinkage procedures were mainly considered there since they tend to be more appealing to the RQ scenario. In particular, we considered the lasso shrinkage procedure based on two variants. The lars-lasso which combines the quadratic objective function and lasso (a constant times the absolute values of non constant term predictors) penalty and RQ Lasso which consists of absolute values in the objective function and the penalty (lasso) function. The later is more convenient to use since it ties in nicely with the linear programming structure of RQs. In Chapter 7 we give some tentative results based on these shrinkage procedures.

CHAPTER 5 RQ MULTIPLE CASE DIAGNOSTICS: A SIMULATION STUDY

5.1 Introduction

In this Chapter we present a Monte Carlo simulation to investigate the effects of leverage, outliers and influential cases on RQs. We consider both the single case and the RQ (multiple) case scenarios. As a point of departure we first analyse the computer generated artificial data sets before we undertake the simulation study based on them. These data are of size n=10 and the number of predictors p=3 at different leverage and outlier scenarios. In order to illustrate that the cut-off value of the RQ (multiple case) leverage diagnostic directly derived from the single case one is often too small, we also considered one scenario of n=20 and p=5. For these artificial data, we consider the ES view of RQs in section 5.3, multiple (RQs) case leverage in section 5.4, the EPR sum of squares in section 5.5, the covariance ratio in section 5.6 and discussions, conclusions and possible further work in section 5.7. Simulation studies to determine cut-off values for the RQ leverage, EPR sum of squares and the covariance ratio are considered in section 5.8, section 5.9 and section 5.10, respectively. In the main simulation a sample size of n=20, p=3, 4,…, 8 were considered. In order to check for any changes we also considered sample sizes n=30, 50 at these sizes of p. Discussions, conclusions and further work are considered in section 5.11.

5.2 Leverage, residual and influence diagnosis in RQs: artificial data

RQs have influence functions that are bounded in the Y-space but unbounded in the X-space. As a consequence, they are fairly robust to residual outliers but can be susceptible to leverage points (X space outliers). In order to understand the extent of the effect of leverage, residual and influential points in RQs, we study them under different leverage and error distribution scenarios using artificial data to get an ad hoc picture as a preliminary step to a simulation study.

We consider the following scenarios;

- Error distributions
  - Gaussian,
  - $CN(\alpha,\sigma^2)$ with (α,σ^2) choices $(0.1,9)$, $(0.2,9)$, $(0.1,100)$,
- Design matrices choices

- $D1 - x_{ij}$ iid $N(0,1)$ for $i = 1,...n$ and $j = 2,3,..., p$.
- $D2$ – As in $D1$, but one point is moved 10 units in the X space.
- $D3$ – As in $D1$, but two points are moved 10 units in the X space.
- $D4$ – As in $D1$, but one point is moved 100 units in the X space.
- $D5$ – As in $D1$, but two points are moved 100 units in the X space.

- Covariates: $p = 3$.
- Regression coefficients: $\beta_0 = 0, \boldsymbol{\beta}$ as the eigenvector corresponding to the largest eigenvalue of $\mathbf{X}'\mathbf{X}$,
- Choices of τ: the entire RQ solution set.
- Sample size: n =10.

The design matrix $D1$ is orthogonalized so that $\mathbf{X}'\mathbf{X} = n\mathbf{I}$. For design matrices $D2$ to $D5$, the first (second) point to be moved out M units in the design matrix $D1$ was chosen as the point having the largest (second largest) Euclidian distance from the centre of the design space. The chosen vector was then extended M units in the X space. This choice is motivated by the fact that from simple linear regression, the most interesting behaviour of estimators was observed when points on the edges of the design were moved out. For example, estimators become inefficient when the design matrix contains leverage value(s) (see, de Jongh, de Wet and Welsh, 1988). Also, it is of interest to investigate the effect of different leverage and error distribution scenarios on prediction. We give some single case leverage and influence diagnostics for each scenario.

The following table gives the leverage values, h_i and the studentized residuals, t_i^* (see section 4.3). The statistic t_i^* follows a student t distribution on $n - p - 1$ degrees of freedom in the case of the Gaussian distribution. In the literature, it is preferred to t_i (see section 4.3).

Table 5.1: Single case leverage and influential diagnostics.

D1 MATRIX-OLS DIAGNOSTICS					
Obs	h_i	$t_i^*, N(0,1)$	$t_i^*, CN(0.1,9)$	$t_i^*, CN(0.2,9)$	$t_i^*, CN(0.1,100)$
1	0.2854	-1.3847	-0.9604	-0.9637	-0.7261
2	0.2394	-0.1943	-0.1010	-0.1938	-0.0482
3	0.1789	4.2292	10.0646	9.7049	18.4474
4	0.1927	-0.2694	-0.1776	-0.1624	-0.1227
5	0.4809	0.5030	0.0827	0.0827	-0.1373
6	0.3982	0.9318	0.3003	0.3345	-0.0157
7	0.3847	-0.3522	-0.0743	-0.0354	0.0734
8	0.4397	0.7267	0.3733	0.4266	-0.7702
9	0.4724	0.7846	0.1404	0.2202	-0.2060
10	0.5052	0.8426	-0.0925	0.0138	-0.8296
SSE		0.0831	0.4005	0.4023	3.6505
D2 MATRIX-OLS DIAGNOSTICS					
Obs	h_i	$t_i^*, N(0,1)$	$t_i^*, CN(0.1,9)$	$t_i^*, CN(0.2,9)$	$t_i^*, CN(0.1,100)$
1	0.2887	-1.2657	-0.9119	-0.9129	-0.7206
2	0.2213	-0.2471	-0.1398	-0.2332	-0.0804
3	0.1806	6.2705	17.3652	15.9668	23.7130
4	0.1524	0.0760	-0.0258	-0.0052	-0.0782
5	0.9856	-0.3575	-0.2675	-0.2812	-0.2139
6	0.1553	0.3656	0.0576	0.0771	-0.1026
7	0.3945	-0.2954	-0.0655	-0.0265	0.0551
8	0.3539	-0.9342	-0.8386	-0.8415	-0.7659
9	0.1270	-0.3840	-0.2686	-0.2499	-0.2020
10	0.1408	0.0266	-0.4506	-0.4444	-0.7039
SSE		0.0814	0.4048	0.4072	3.6669
D3 MATRIX-OLS DIAGNOSTICS					
Obs	h_i	$t_i^*, N(0,1)$	$t_i^*, CN(0.1,9)$	$t_i^*, CN(0.2,9)$	$t_i^*, CN(0.1,100)$
1	0.1495	-1.4911	-0.8435	-0.8251	-0.5036
2	0.1137	0.0738	-0.1247	-0.2243	-0.2254
3	0.1344	3.2402	8.2765	8.5635	17.0183
4	0.1043	0.2820	-0.0218	-0.0098	-0.1843
5	0.9852	-0.4620	-0.2736	-0.2789	-0.1578
6	0.1355	0.5183	0.0580	0.0717	-0.1888
7	0.9688	-0.8121	-0.0609	-0.0087	0.3296
8	0.1752	-1.1935	-0.7536	-0.7347	-0.4972
9	0.1118	-0.2135	-0.2654	-0.2520	-0.2779
10	0.1215	-0.0888	-0.4549	-0.4407	-0.6353
SSE		0.0891	0.4048	0.4071	3.7314

Table 5.1 (continued).

D4 MATRIX-OLS DIAGNOSTICS					
Obs	h_i	$t_i^*, N(0,1)$	$t_i^*, CN(0.1,9)$	$t_i^*, CN(0.2,9)$	$t_i^*, CN(0.1,100)$
1	0.2898	-1.2412	-0.9082	-0.9089	-0.7230
2	0.2166	-0.2725	-0.1517	-0.2452	-0.0845
3	0.1829	6.2485	18.3514	16.6185	25.3593
4	0.1612	0.1246	-0.0099	0.0113	-0.0803
5	0.9999	-0.4851	-0.3133	-0.3285	-0.2161
6	0.1341	0.2691	0.0227	0.0406	-0.1068
7	0.3941	-0.3043	-0.0725	-0.0339	0.0503
8	0.3446	-1.0066	-0.8630	-0.8669	-0.7645
9	0.1299	-0.3707	-0.2665	-0.2478	-0.2049
10	0.1469	0.0919	-0.4274	-0.4203	-0.7068
SSE		0.0829	0.4066	0.4091	3.6674
D5 MATRIX-OLS DIAGNOSTICS					
Obs	h_i	$t_i^*, N(0,1)$	$t_i^*, CN(0.1,9)$	$t_i^*, CN(0.2,9)$	$t_i^*, CN(0.1,100)$
1	0.1271	-1.5126	-0.8332	-0.8139	-0.4697
2	0.1231	0.0746	-0.1361	-0.2350	-0.2436
3	0.1259	2.9355	7.6435	7.8976	16.6088
4	0.1221	0.3485	-0.0060	0.0077	-0.1994
5	0.9998	-0.6228	-0.3193	-0.3249	-0.1375
6	0.1255	0.4180	0.0232	0.0371	-0.1929
7	0.9997	-0.8969	-0.0698	-0.0196	0.3637
8	0.1291	-1.3151	-0.7637	-0.7449	-0.4509
9	0.1234	-0.1932	-0.2652	-0.2503	-0.2869
10	0.1243	-0.0530	-0.4342	-0.4184	-0.6240
SSE		0.0925	0.4066	0.4091	3.7467

For single case leverage values, the threshold value is $2pn^{-1}$, *i.e.*, a leverage point has $h_i > 2pn^{-1}$ (=0.6). In general, we use $t_{0.05,n-p-1}$ as a cut-off value to identify single case influential observations, *i.e.*, observation i is influential if $\left|t_i^*\right| > t_{0.05,n-p-1}$. In our case $n=10$, $p=3$, giving $t_{0.05,6}=1.86$. The leverage points are: none in D1, the points at the edges of the design matrix that were moved out, *viz.*, 5 in D2 and D4 and 5 and 7 in D3 and D5 as expected.

Figure 5.1 below gives the scatter plot representation of matrices D2 to D5.

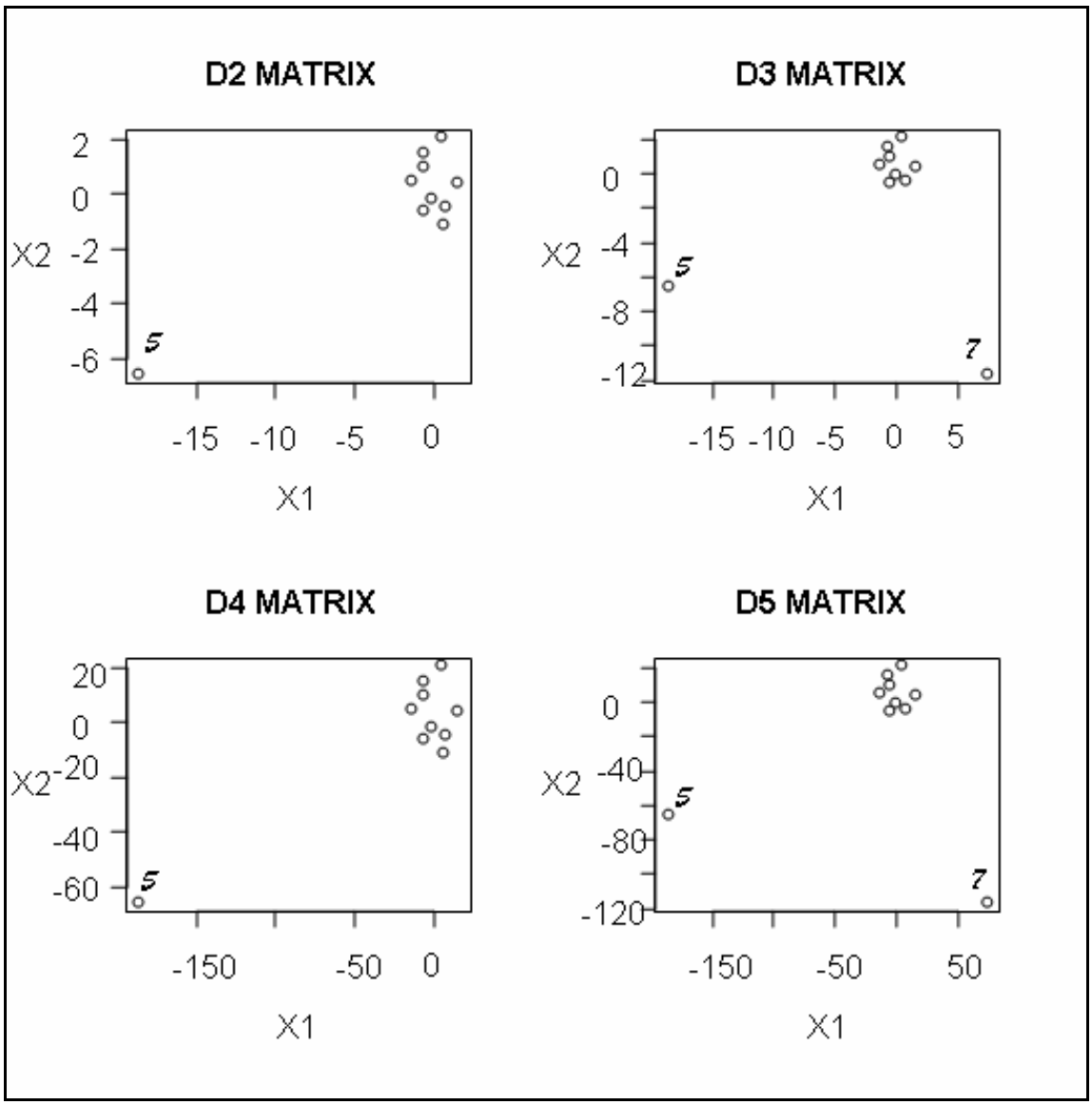


Figure 5.1: Scatter plots for design matrices D2 to D5

Note: The scale in the scatter plot of D4 and D5 is 10 times that in D2 and D3, *i.e.*, point 5 with coordinates (-18.581,-6.587) in D2 has coordinates (-185.812, -65.869) in D4 and points 5 and 7 with coordinates (-18.581, -6.587) and (7.458, -11.641) in D3 have (-185.812, -65.869) and (74.582,-116.407) in D5 respectively.

There is only one influential point, *viz.*, observation 3, across all design matrices and distributions. The influence of this point increases as point 5's leverage increases and as the error distribution becomes heavier. However, the addition of another leverage point (observation 7), reduces the influence of this point. For example, comparing the t_i^* values for D2 and D3 in table 5.1.

The OLS residual sum of squares, SSE is minimal at the normal error distribution as expected. However, there seems to be very little difference from D1 to D4 with a small increase at D5. The same picture is exhibited at heavier tails except that the SSE increases

approximately 4.5 times at $CN(0.1,9)$ and $CN(0.2,9)$ and approximately 45 times at $CN(0.1,100)$ (a dramatic increase!).

It is important to see what this single case diagnosis picture translates to at the RQ level (multiple case diagnosis) corresponding to RQs. In the following section we compute RQs and explore their nature for the artificial data starting with the leverage aspect.

5.3 An ES view of RQs obtained from artificial data

Viewing RQs as ESs, it is of essence to identify p observations that make up these RQs. Furthermore, it is of interest to understand why RQs have an affinity for these observations while they tend to exclude others. Or are we going to get a random pattern in the way in which observations get included in the RQs? To answer this question we obtain the solution of the LP problem (3.2.1) giving the RQs. Note that from this entire solution we only give those τ's at which the ES corresponding to the RQ changes - the ES remains the same between the consecutive τ values. The following table gives the RQs obtained at different design matrices and under different error distributions.

Remark: Note that we have approximately $n = 10$ ESs corresponding to RQs, but this number varies and is often rather smaller than n.

Table 5.2: Unique ESs corresponding to RQs obtained for different design matrices and error distributions scenarios.

NORMAL

D1		D2		D3		D4		D5	
τ	ES	τ	ES	τ	ES	τ	ES	τ	ES
0.2008	4 7 9	0.1988	5 7 8	0.1348	5 7 8	0.1994	5 7 8	0.1260	5 7 8
0.2679	4 7 10	0.2480	5 8 9	0.2788	5 7 9	0.2474	2 5 8	0.2530	5 7 9
0.2803	7 8 10	0.5288	2 5 7	0.3960	5 7 10	0.4338	5 8 9	0.3772	5 7 10
0.3772	4 8 10	0.5417	5 7 10	0.5176	2 5 7	0.4844	2 5 9	0.5018	2 5 7
0.5153	2 8 10	0.6107	4 5 10	0.6330	4 5 7	0.5255	2 5 7	0.6258	4 5 7
0.6915	2 5 7	0.7468	4 5 7	0.8708	5 6 7	0.5375	5 7 10	0.8746	5 6 7
0.8453	5 6 7	0.8342	5 6 7			0.6073	4 5 10		
						0.7515	4 5 7		
						0.8329	5 6 7		

CN1

τ	ES	τ	ES	τ	ES	τ	ES	τ	ES
0.2231	7 8 10	0.1988	5 7 8	0.1348	5 7 8	0.1994	5 7 8	0.1260	5 7 8
0.2396	8 9 10	0.2480	5 8 9	0.2788	5 7 10	0.2474	2 5 8	0.2530	5 7 10
0.4837	2 8 10	0.3913	2 5 8	0.4004	5 7 9	0.4338	5 8 9	0.3777	5 7 9
0.6462	2 5 8	0.6107	5 8 10	0.5176	2 5 7	0.5255	2 5 9	0.5018	2 5 7
0.6915	2 5 7	0.6402	4 5 8	0.6330	4 5 7	0.6073	4 5 10	0.6258	4 5 7
0.8453	5 6 7	0.7468	4 5 7	0.8708	5 6 7	0.7515	4 5 7	0.8746	5 6 7
		0.8342	5 6 7			0.8329	5 6 7		

CN2

τ	ES	τ	ES	τ	ES	τ	ES	τ	ES
0.2231	2 7 10	0.1978	1 2 5	0.1348	5 7 8	0.1964	1 2 5	0.1260	5 7 8
0.3289	7 8 10	0.2438	5 7 8	0.2788	5 7 10	0.2430	5 7 8	0.2530	5 7 10
0.3868	7 8 9	0.4409	5 8 9	0.4004	5 7 9	0.4338	5 8 9	0.3777	5 7 9
0.6407	5 7 8	0.6107	5 8 10	0.5176	2 5 7	0.5375	5 7 9	0.5018	2 5 7
0.8453	5 6 7	0.6402	4 5 8	0.6330	4 5 7	0.6073	4 5 10	0.6258	4 5 7
		0.7468	4 5 7	0.8708	5 6 7	0.7515	4 5 7	0.8746	5 6 7
		0.8342	5 6 7			0.8329	5 6 7		

CN3

τ	ES	τ	ES	τ	ES	τ	ES	τ	ES
0.1797	6 7 8	0.1458	1 5 7	0.1216	1 5 7	0.1507	1 5 7	0.1246	1 5 7
0.2506	1 7 8	0.3569	2 5 9	0.2564	5 7 8	0.3613	2 5 8	0.2507	5 7 8
0.3910	1 2 8	0.3972	5 8 9	0.4004	5 7 9	0.4996	5 8 9	0.3777	5 7 9
0.4837	2 8 9	0.4473	2 5 8	0.5176	2 5 7	0.6426	4 5 8	0.5018	2 5 7
0.6462	2 5 8	0.6402	4 5 8	0.6330	4 5 7	0.7515	4 5 7	0.6258	4 5 7
0.6915	2 5 7	0.7468	4 5 7	0.8708	5 6 7	0.8329	5 6 7	0.8746	5 6 7
0.8453	5 6 7	0.8342	5 6 7						

Although there are no leverage points under design matrix D1 (in the sense that there is no $h_i > 2pn^{-1} = 0.6$), observations 10, 5, 9 and 8 have relatively high leverages followed by 6 and 7 which have more or less the same degree of leverage while observation 3 has the least leverage. Also, observation 5 is furthest away from the centre of the design space while observation 7 is second furthest. It is interesting to note that although there is no consistent pattern as we move from the normal distribution to heavier ones at design matrix D1, observation 3 (with the least leverage) is never included in the ESs corresponding to RQs.

At the remaining design matrices, at least one of the leverage points 5 and 7 shown in violet colour have a 100% inclusion in the computed RQs. Also, the more the severity of leverage and the more the number of leverage points the less the effect of the error distributions on the choice of observations included in RQs. Since leverage is the dominant factor here, it is important to be able to "quantify" it at the RQ level. This leads us to a multiple case leverage view of RQs. In the next section we discuss this aspect.

5.4 Regression quantile (multiple case) leverage diagnosis

Most multiple case regression diagnostics in the literature involve the extension of existing single case diagnostics. Similarly, the extension of the analogy of the procedure used in identifying large leverage ($h_i > 2pn^{-1}$) values in the single case scenario to the RQ situation is quite appealing here. Based on the ES view of RQs, an attractive and natural extension of the h_i statistic to the RQ (ES) scenario is the ES predictive weighted leverage T_J (see (4.5.7) in subsection 4.5.1) since the sum of OLS leverage values, the h_i's, is equal to the sum of the T_J's, *i.e.*,

$$\sum_{i=1}^{n} h_i = \sum_{i=1}^{K} T_J = p\,.$$

Hence, our original idea was to extend the single case leverage diagnostic procedure of Hoaglin and Welsch (1978) to the multiple (RQ) case. In their method, they used twice the average of the total leverage as a cut-off point, thus using $2pn^{-1}$. Since $\sum_J T_J = p$, and the sum extends over K $(=\binom{n}{p})$ elements, this would give $2pK^{-1}$ as a cut-off point for the RQ case , *i.e.*, flag $T_J > 2pK^{-1}$ in the RQ case. However, K is typically very large and RQs have a high affinity for leverage points resulting in very large ω_J's (see 4.2.2) and hence T_J's, leading to useless cut-off values as revealed in the following table in the analysis of the artificial data.

Table 5.3: RQ predictive weighted leverage, T_J at different design matrices and distribution scenarios.

NORMAL

D1		D2		D3		D4		D5	
τ	T_J	τ	T_J	τ	T_J	τ	T_J	τ	T_J
0.2008	0.0085	0.1988	0.0717	0.1348	0.1164	0.1994	0.0727	0.1260	0.1244
0.2679	0.0227	0.2480	0.0556	0.2788	0.1225	0.2474	0.0594	0.2530	0.1252
0.2803	0.0366	0.5288	0.0852	0.3960	0.1216	0.4338	0.0556	0.3772	0.1251
0.3772	0.0280	0.5417	0.0599	0.5176	0.1237	0.4844	0.0484	0.5018	0.1252
0.5153	0.0293	0.6107	0.0356	0.6330	0.1248	0.5255	0.0859	0.6258	0.1254
0.6915	0.0420	0.7468	0.0743	0.8708	0.1206	0.5375	0.0608	0.8746	0.1249
0.8453	0.0517	0.8342	0.0716			0.6073	0.0385		
						0.7515	0.0764		
						0.8329	0.0706		

CN1

τ	T_J	τ	T_J	τ	T_J	τ	T_J	τ	T_J
0.2231	0.0366	0.1988	0.0717	0.1348	0.1164	0.1994	0.0727	0.1260	0.1244
0.2396	0.0225	0.2480	0.0556	0.2788	0.1216	0.2474	0.0594	0.2530	0.1251
0.4837	0.0293	0.3913	0.0594	0.4004	0.1225	0.4338	0.0556	0.3777	0.1252
0.6462	0.0410	0.6107	0.0674	0.5176	0.1237	0.5255	0.0484	0.5018	0.1252
0.6915	0.0420	0.6402	0.0560	0.6330	0.1248	0.6073	0.0385	0.6258	0.1254
0.8453	0.0517	0.7468	0.0743	0.8708	0.1206	0.7515	0.0764	0.8746	0.1249
		0.8342	0.0716			0.8329	0.0706		

CN2

τ	T_J	τ	T_J	τ	T_J	τ	T_J	τ	T_J
0.2231	0.0218	0.1978	0.0541	0.1348	0.1164	0.1964	0.0546	0.1260	0.1244
0.3289	0.0366	0.2438	0.0717	0.2788	0.1216	0.2430	0.0727	0.2530	0.1251
0.3868	0.0326	0.4409	0.0556	0.4004	0.1225	0.4338	0.0556	0.3777	0.1252
0.6407	0.0483	0.6107	0.0674	0.5176	0.1237	0.5375	0.0699	0.5018	0.1252
0.8453	0.0517	0.6402	0.0560	0.6330	0.1248	0.6073	0.0385	0.6258	0.1254
		0.7468	0.0743	0.8708	0.1206	0.7515	0.0764	0.8746	0.1249
		0.8342	0.0716			0.8329	0.0706		

CN3

τ	T_J	τ	T_J	τ	T_J	τ	T_J	τ	T_J
0.1797	0.0540	0.1458	0.0681	0.1216	0.1183	0.1507	0.0688	0.1246	0.1247
0.2506	0.0381	0.3569	0.0480	0.2564	0.1164	0.3613	0.0594	0.2507	0.1244
0.3910	0.0282	0.3972	0.0556	0.4004	0.1225	0.4996	0.0556	0.3777	0.1252
0.4837	0.0236	0.4473	0.0594	0.5176	0.1237	0.6426	0.0564	0.5018	0.1252
0.6462	0.0410	0.6402	0.0560	0.6330	0.1248	0.7515	0.0764	0.6258	0.1254
0.6915	0.0420	0.7468	0.0743	0.8708	0.1206	0.8329	0.0706	0.8746	0.1249
0.8453	0.0517	0.8342	0.0716						

For T_J, we flag ($T_J \geq 2pK^{-1}$=0.050) shown in violet colour; the RQ at the highest level, $\tau = 0.8453$ under all distributions and the RQ at the lowest level under CN3 at D1. The T_J's are more or less of the same size at D1, the "clean" case. All T_J's are flagged at D2 except at $\tau = 0.2506$ under CN3 while we fail to flag at two RQ levels under the normal, CN1 and CN2 and at one RQ level at D4. At D3 and D5 we flag all T_J's.

Remark: Using $T_J > 4pK^{-1}$=0.1000 determined in the simulation study in section 5.8, we flag no RQs at D1, D2 and D4 but flag all RQs at D3 and D5.

A small sample size like n=10 gives a false impression that $2pK^{-1}$ is "large enough" as a cut-off value. Let us digress a bit from n=10 and p=3 in order to see what the picture translates to for a sample of size n=20, p=5, giving $2pK^{-1} = 0.00064$ under the normal distribution as shown in table 5.4 below.

Table 5.4: RQ predictive weighted leverage at different design matrices under the Normal distribution using a sample size of n=20.

D1		D2		D3		D4		D5	
τ	T_J	τ	T_J	τ	T_J	τ	T_J	τ	T_J
0.0952	0.0007	0.1038	0.0045	0.0646	0.0179	0.1031	0.0037	0.0634	0.0187
0.1429	0.0013	0.2029	0.0036	0.2133	0.0039	0.2015	0.0029	0.1681	0.0135
0.1905	0.0019	0.2231	0.0029	0.2359	0.0049	0.2460	0.0010	0.2128	0.0041
0.2381	0.0018	0.2468	0.0009	0.3320	0.0041	0.3245	0.0017	0.2363	0.0052
0.2857	0.0004	0.3455	0.0017	0.3582	0.0046	0.3456	0.0016	0.3332	0.0045
0.3333	0.0007	0.4269	0.0015	0.4268	0.0050	0.4297	0.0022	0.3963	0.0043
0.3810	0.0008	0.4676	0.0023	0.5126	0.0045	0.5478	0.0038	0.4243	0.0100
0.4286	0.0007	0.5483	0.0039	0.5366	0.0048	0.5588	0.0037	0.4632	0.0094
0.4762	0.0011	0.5596	0.0036	0.5864	0.0109	0.6482	0.0047	0.4989	0.0015
0.5238	0.0014	0.6482	0.0046	0.6377	0.0110	0.6654	0.0047	0.5057	0.0045
0.5714	0.0013	0.7012	0.0040	0.7346	0.0163	0.7018	0.0040	0.5309	0.0051
0.6190	0.0012	0.8096	0.0027	0.8589	0.0087	0.8581	0.0027	0.5845	0.0094
0.6667	0.0017	0.8585	0.0027	0.9025	0.0049	0.8963	0.0014	0.6336	0.0115
0.7143	0.0024	0.8965	0.0013					0.7323	0.0171
0.7619	0.0024							0.8586	0.0053
0.8095	0.0014								
0.8571	0.0006								

It is clear that $2pK^{-1}$ can be extremely small for most data sets and we further pursue this avenue in a slightly modified fashion, *viz.*, search for the value of $\tilde{c}$ via a simulation study such that $\tilde{c}pK^{-1}$ is reasonable ("large enough") (see section 5.8).

Since T_J is a sum of elemental predictive weighted leverage, analysis of the ERW, ω_J (see 4.2.2) might provide us with an insight into the nature of this RQ leverage statistic. In the next subsection we focus on ω_J.

Remark: Using $T_J > 15pK^{-1}$ =0.0048 determined in the simulation study in section 5.8, we flag no RQs at D1, D2 and D4 but flag 4 RQs at D3 and 3 RQs at D5.

5.4.1 The elemental Regression weight

The statistic T_J is a sum of the elemental predictive weighted leverage with weight ω_J. Hence, the size of T_J partly depends on the size of ω_J. Table 5.5 gives the ω_J's.

Table 5.5: ERW, ω_J at different design matrices and distribution scenarios.

NORMAL									
D1		D2		D3		D4		D5	
τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J
0.2008	0.0008	0.1988	0.1181	0.1348	0.1573	0.1994	0.1149	0.1260	0.1289
0.2679	0.0011	0.2480	0.0415	0.2788	0.1042	0.2474	0.0727	0.2530	0.1233
0.2803	0.0395	0.5288	0.0051	0.3960	0.1122	0.4338	0.0428	0.3772	0.1242
0.3772	0.0103	0.5417	0.0533	0.5176	0.1010	0.4844	0.0038	0.5018	0.1229
0.5153	0.0260	0.6107	0.0172	0.6330	0.0926	0.5255	0.0048	0.6258	0.1219
0.6915	0.0036	0.7468	0.0088	0.8708	0.1228	0.5375	0.0578	0.8746	0.1254
0.8453	0.0213	0.8342	0.0191			0.6073	0.0193		
						0.7515	0.0102		
						0.8329	0.0169		
CN1									
τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J
0.2231	0.0395	0.1988	0.1181	0.1348	0.1573	0.1994	0.1149	0.1260	0.1289
0.2396	0.0102	0.2480	0.0415	0.2788	0.1122	0.2474	0.0727	0.2530	0.1242
0.4837	0.0260	0.3913	0.0752	0.4004	0.1042	0.4338	0.0428	0.3777	0.1233
0.6462	0.0481	0.6107	0.0063	0.5176	0.1010	0.5255	0.0038	0.5018	0.1229
0.6915	0.0036	0.6402	0.0519	0.6330	0.0926	0.6073	0.0193	0.6258	0.1219
0.8453	0.0213	0.7468	0.0088	0.8708	0.1228	0.7515	0.0102	0.8746	0.1254
		0.8342	0.0191			0.8329	0.0169		
CN2									
τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J
0.2231	0.0024	0.1978	0.0622	0.1348	0.1573	0.1964	0.0621	0.1260	0.1289
0.3289	0.0395	0.2438	0.1181	0.2788	0.1122	0.2430	0.1149	0.2530	0.1242
0.3868	0.0051	0.4409	0.0415	0.4004	0.1042	0.4338	0.0428	0.3777	0.1233
0.6407	0.0719	0.6107	0.0063	0.5176	0.1010	0.5375	0.0171	0.5018	0.1229
0.8453	0.0213	0.6402	0.0519	0.6330	0.0926	0.6073	0.0193	0.6258	0.1219
		0.7468	0.0088	0.8708	0.1228	0.7515	0.0102	0.8746	0.1254
		0.8342	0.0191			0.8329	0.0169		
CN3									
τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J	τ	ω_J
0.1797	0.0102	0.1458	0.0991	0.1216	0.1377	0.1507	0.1012	0.1246	0.1270
0.2506	0.0141	0.3569	0.0036	0.2564	0.1573	0.3613	0.0727	0.2507	0.1289
0.3910	0.0090	0.3972	0.0415	0.4004	0.1042	0.4996	0.0428	0.3777	0.1233
0.4837	0.0040	0.4473	0.0752	0.5176	0.1010	0.6426	0.0555	0.5018	0.1229
0.6462	0.0481	0.6402	0.0519	0.6330	0.0926	0.7515	0.0102	0.6258	0.1219
0.6915	0.0036	0.7468	0.0088	0.8708	0.1228	0.8329	0.0169	0.8746	0.1254
0.8453	0.0213	0.8342	0.0191						

For the ERWs, we know that $\sum_J \omega_J = 1$, so we take ω_J as very small if it is less than the average value, *i.e.*, $\omega_J < K^{-1}$ and $\omega_J > 2K^{-1}$ as tentatively large. For $n = 10$, $p = 3$, we have $K^{-1} = 0.0083$ and $2K^{-1} = 0.0167$.

At D1 small ω_J are flagged (shown in orange colour) at the lowest 2 and the second last RQ levels under the normal distribution, the second last under $CN(0.1,9)$, the first and the 2 middle one under $CN(0.2,9)$, the middle one and the 2nd last under $CN(0.1,100)$. Also, we have few "small" ω_J at the remaining design matrices where more of the weights are generally large ($\omega_J > 2K^{-1}$). In fact the number of small weights decrease as the number of leverage points and degree of leverage increase.

Large ω_J's mimic T_J's especially at high leverage. Thus we have "large" ω_J's at high leverage (shown in violet colour).

In order to investigate what influences the size of ω_J, we decompose it as before in a way that gives two views, *viz.*, the **variability** and the **collinearity** views, as follows:

$$\omega_J = \frac{p}{n}\prod_{j=1}^{p-1}\left(\frac{s'^2_{J,j}}{s'^2_j}\right)\frac{|\mathbf{C}_J|}{|\mathbf{C}|}$$
$$= \frac{p}{n}.\gamma_J.\rho_J$$

(see equation (4.2.2)).

Clearly, the size of T_J depends on ω_J while on the other hand, the size of the ERW ω_J depends on the two factors γ_J and ρ_J. The factor $0 \le \gamma_J \le 1$, measuring the product of the proportion of the ratio of the variabilities of the $p-1$ predictors in the ES to that of the design matrix, gives us the **variability view** of the ERW. Also, $0 \le \rho_J < \infty$ measures the proportion of $|\mathbf{C}_J|$ to $|\mathbf{C}|$. If $\rho_J = \frac{|\mathbf{C}_J|}{|\mathbf{C}|} < 1$, the degree of collinearity is higher at the RQ level compared to that at the design matrix level and vice versa. Hence, the factor ρ_J raises the **collinearity view** aspect of the ERW. Small values of γ_J and $\rho_J < 1$ would result in a small ERW.

Remark: Recall that if the $\mathbf{X}$ matrix consists of the $p-1$ nonconstant term predictors and is standardized to correlation form, then the ERW exposes the collinearity view only, *i.e.*,

$$\omega_J = \frac{|\mathbf{C}_J|}{|\mathbf{C}|}.$$

In order to have a clear picture of the interaction between these two factors we present both of them in table 5.6 below.

Table 5.6: ERW variability Factor, γ_J and ERW collinearity Factor, ρ_J at different design matrices and distribution scenarios.

NORMAL

D1			D2			D3			D4			D5		
τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J
0.201	0.005	0.623	0.199	0.566	0.695	0.135	0.528	0.993	0.199	0.553	0.692	0.126	0.445	0.967
0.268	0.031	0.118	0.248	0.576	0.240	0.279	0.362	0.961	0.247	0.559	0.433	0.253	0.427	0.962
0.280	0.133	0.987	0.529	0.339	0.050	0.396	0.394	0.948	0.434	0.559	0.255	0.377	0.431	0.961
0.377	0.062	0.560	0.542	0.373	0.476	0.518	0.348	0.968	0.484	0.539	0.024	0.502	0.426	0.963
0.515	0.087	1.000	0.611	0.364	0.158	0.633	0.328	0.941	0.525	0.533	0.030	0.626	0.423	0.960
0.691	0.013	0.943	0.747	0.308	0.096	0.871	0.413	0.990	0.537	0.535	0.360	0.875	0.433	0.966
0.845	0.087	0.822	0.834	0.423	0.151				0.607	0.535	0.120			
									0.752	0.528	0.065			
									0.833	0.542	0.104			

CN1

τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J
0.223	0.133	0.987	0.199	0.566	0.695	0.135	0.528	0.993	0.199	0.553	0.692	0.126	0.445	0.967
0.240	0.046	0.744	0.248	0.576	0.240	0.279	0.394	0.948	0.247	0.559	0.433	0.253	0.431	0.961
0.484	0.087	1.000	0.391	0.588	0.426	0.400	0.362	0.961	0.434	0.559	0.255	0.378	0.427	0.962
0.646	0.047	0.783	0.611	0.568	0.037	0.518	0.348	0.968	0.525	0.539	0.024	0.502	0.426	0.963
0.691	0.013	0.943	0.640	0.544	0.318	0.633	0.328	0.941	0.607	0.535	0.120	0.626	0.423	0.960
0.845	0.087	0.822	0.747	0.308	0.096	0.871	0.413	0.990	0.752	0.528	0.065	0.875	0.433	0.966
			0.834	0.423	0.151				0.833	0.542	0.104			

CN2

τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J
0.223	0.040	0.197	0.198	0.500	0.415	0.135	0.528	0.993	0.196	0.551	0.376	0.126	0.445	0.967
0.329	0.133	0.987	0.244	0.566	0.695	0.279	0.394	0.948	0.243	0.553	0.692	0.253	0.431	0.961
0.387	0.017	0.990	0.441	0.576	0.240	0.400	0.362	0.961	0.434	0.559	0.255	0.378	0.427	0.962
0.641	0.058	0.917	0.611	0.568	0.037	0.518	0.348	0.968	0.537	0.533	0.107	0.502	0.426	0.963
0.845	0.087	0.822	0.640	0.544	0.318	0.633	0.328	0.941	0.607	0.535	0.120	0.626	0.423	0.960
			0.747	0.308	0.096	0.871	0.413	0.990	0.752	0.528	0.065	0.875	0.433	0.966
			0.834	0.423	0.151				0.833	0.542	0.104			

CN3

τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J	τ	γ_J	ρ_J
0.180	0.036	0.949	0.146	0.475	0.695	0.122	0.471	0.975	0.151	0.545	0.619	0.125	0.439	0.964
0.251	0.063	0.752	0.357	0.383	0.031	0.256	0.528	0.993	0.361	0.559	0.433	0.251	0.445	0.967
0.391	0.039	0.769	0.397	0.576	0.240	0.400	0.362	0.961	0.500	0.559	0.255	0.378	0.427	0.962
0.484	0.013	0.999	0.447	0.588	0.426	0.518	0.348	0.968	0.643	0.554	0.334	0.502	0.426	0.963
0.646	0.047	0.783	0.640	0.544	0.318	0.633	0.328	0.941	0.752	0.528	0.065	0.626	0.423	0.960
0.691	0.013	0.943	0.747	0.308	0.096	0.871	0.413	0.990	0.833	0.542	0.104	0.875	0.433	0.966
0.845	0.087	0.822	0.834	0.423	0.151									

At D1, γ_J is generally responsible for the size of ω_J as ρ_J is generally close to 1. In particular, ω_J is very small at the first τ level due to a small γ_J (=0.005) . Also, at this design matrix the values of γ_J are relatively small (shown in orange colour) with a minimum value of 0.005 and a maximum value of 0.087 and ρ_J is relatively large (shown in violet colour) with all ρ_J's greater than 0.5 except for the 2nd RQ under the normal distribution and first one under CN2. Actually, at D1 the level of collinearity at the ES corresponding to a RQ is more or less the same as that exhibited as evidenced by $\rho \approx 1$.

At D2 and D4 the inducement of leverage on point 5 induces collinearity at the RQ level as ρ_J becomes smaller. However, there is an "antagonistic" relationship between γ_J and ρ_J at these two design matrices as a smaller ρ_J is counteracted by a dramatic increase in γ_J . Furthermore, at D4 γ_J is generally larger and has less variability than at D2.

Remark: It seems this "antagonistic" relationship between γ_J and ρ_J is the main reason why leverage does not affect RQ prediction adversely despite the presence of collinearity. Actually, the size of ω_J does not decrease.

A further inducement of leverage on point 7 (D3 and D5) hides the collinearity induced at point 5 and ρ_J becomes close to 1 while γ_J remains fairly large. Observations 5 and 7 are referred to as collinearity influential points (see section 4.6). It seems highly unlikely that the presence of collinearity induced by collinearity influential results in the size of ω_J being small. Hence, ω_J remains large at designs that contain leverage points irrespective of the presence of collinearity.

A clear picture of collinearity inducement and hiding is given by the statistic $|\mathbf{C}_J|$, given in the following table.

Table 5.7: RQs Collinearity View based on, ρ_J and $|C_J|$ at different design matrices and distribution scenarios.

NORMAL

D1			D2			D3			D4			D5		
τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$
0.201	0.623	0.622	0.199	0.695	0.142	0.135	0.993	0.988	0.199	0.692	0.002	0.126	0.967	0.962
0.268	0.118	0.118	0.248	0.240	0.049	0.279	0.961	0.955	0.247	0.433	0.001	0.253	0.962	0.958
0.280	0.987	0.986	0.529	0.050	0.010	0.396	0.948	0.943	0.434	0.255	0.001	0.377	0.961	0.957
0.377	0.560	0.560	0.542	0.476	0.097	0.518	0.968	0.963	0.484	0.024	0.000	0.502	0.963	0.959
0.515	1.000	1.000	0.611	0.158	0.032	0.633	0.941	0.935	0.525	0.030	0.000	0.626	0.960	0.956
0.691	0.943	0.942	0.747	0.096	0.020	0.871	0.990	0.985	0.537	0.360	0.001	0.875	0.966	0.961
0.845	0.822	0.821	0.834	0.151	0.031				0.607	0.120	0.000			
									0.752	0.065	0.000			
									0.833	0.104	0.000			

CN1

τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$
0.223	0.987	0.986	0.199	0.695	0.142	0.135	0.993	0.988	0.199	0.692	0.002	0.126	0.967	0.962
0.240	0.744	0.743	0.248	0.240	0.049	0.279	0.948	0.943	0.247	0.433	0.001	0.253	0.961	0.957
0.484	1.000	1.000	0.391	0.426	0.087	0.400	0.961	0.955	0.434	0.255	0.001	0.378	0.962	0.958
0.646	0.783	0.783	0.611	0.037	0.008	0.518	0.968	0.963	0.525	0.024	0.000	0.502	0.963	0.959
0.691	0.943	0.942	0.640	0.318	0.065	0.633	0.941	0.935	0.607	0.120	0.000	0.626	0.960	0.956
0.845	0.822	0.821	0.747	0.096	0.020	0.871	0.990	0.985	0.752	0.065	0.000	0.875	0.966	0.961
			0.834	0.151	0.031				0.833	0.104	0.000			

CN2

τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$
0.223	0.197	0.197	0.198	0.415	0.085	0.135	0.993	0.988	0.196	0.376	0.001	0.126	0.967	0.962
0.329	0.987	0.986	0.244	0.695	0.142	0.279	0.948	0.943	0.243	0.692	0.002	0.253	0.961	0.957
0.387	0.990	0.989	0.441	0.240	0.049	0.400	0.961	0.955	0.434	0.255	0.001	0.378	0.962	0.958
0.641	0.917	0.917	0.611	0.037	0.008	0.518	0.968	0.963	0.537	0.107	0.000	0.502	0.963	0.959
0.845	0.822	0.821	0.640	0.318	0.065	0.633	0.941	0.935	0.607	0.120	0.000	0.626	0.960	0.956
			0.747	0.096	0.020	0.871	0.990	0.985	0.752	0.065	0.000	0.875	0.966	0.961
			0.834	0.151	0.031				0.833	0.104	0.000			

CN3

τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$	τ	ρ_J	$\|\mathbf{C}_J\|$
0.180	0.949	0.949	0.146	0.695	0.142	0.122	0.975	0.970	0.151	0.619	0.002	0.125	0.964	0.960
0.251	0.752	0.751	0.357	0.031	0.006	0.256	0.993	0.988	0.361	0.433	0.001	0.251	0.967	0.962
0.391	0.769	0.768	0.397	0.240	0.049	0.400	0.961	0.955	0.500	0.255	0.001	0.378	0.962	0.958
0.484	0.999	0.998	0.447	0.426	0.087	0.518	0.968	0.963	0.643	0.334	0.001	0.502	0.963	0.959
0.646	0.783	0.783	0.640	0.318	0.065	0.633	0.941	0.935	0.752	0.065	0.000	0.626	0.960	0.956
0.691	0.943	0.942	0.747	0.096	0.020	0.871	0.990	0.985	0.833	0.104	0.000	0.875	0.966	0.961
0.845	0.822	0.821	0.834	0.151	0.031									

There is a dramatic increase in the level of collinearity as evidenced by a sharp decrease of $|\mathbf{C}_J|$ from D1 to D2 and D4. The level of collinearity at the RQ level is always far less than the level of collinearity at the design matrix level. Hence, the moving of observation 5 along the X-axis has induced collinearity at D2 and D4. However, at D3 and D5 the moving of

observation 7 along the X-axis has hidden the collinearity induced by observation 5 resulting in $|\mathbf{C}_J| \approx 1$ and $\rho_J \approx 1$. However, $|\mathbf{C}_J|$ is slightly less than $|\mathbf{C}|$.

Clearly the degree of collinearity increases as we move observation 5 along the X-axis as shown by the decrease in $|\mathbf{C}_J|$ from D2 to D4. Also, moving observation 7 along the X-axis hides the collinearity.

Another statistic which is involved in the RQ predictive leverage statistic T_J is the residual freedom R_{iJ}. We present this statistic in the following section.

5.4.2 RQ/ES predictive leverage

The RQ leverage statistic T_J is a weighted average of ES predictive leverage R_{iJ} also referred to as the **residual freedom** by Hawkins *et al.* (1984) (see subsection 2.4.1). Table 5.8 gives the sum of the residual freedoms, $\sum_{i \notin J} R_{iJ}$.

Remark: Ironically the RQ predictive weighted leverage, T_J is generally largest at D3 and D5 where the sum of the residual freedoms, $\sum_{i \notin J} R_{iJ}$ is small. This clearly suggests that the ERW, ω_J is the dominant factor in determining the size of T_J.

Clearly, on average, there is no marked distinction between the R_{iJ}'s at D1 (no leverage points) and D2 and D4 (one leverage point). However, there is a remarkable decrease of the R_{iJ}'s at D3 and D5 (two leverage points). It seems the R_{iJ}'s decrease as the number of leverage points increases.

Table 5.8: Unweighted RQ Predictive Leverage at different design matrices and distribution scenarios.

NORMAL

D1		D2		D3		D4		D5	
τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$
0.201	70.349	0.199	4.249	0.135	5.179	0.199	4.430	0.126	6.754
0.268	145.989	0.248	9.392	0.279	8.232	0.247	5.722	0.253	7.106
0.280	6.489	0.529	117.737	0.396	7.592	0.434	9.099	0.377	7.047
0.377	18.980	0.542	7.874	0.518	8.577	0.484	88.880	0.502	7.132
0.515	7.891	0.611	14.462	0.633	9.430	0.525	124.319	0.626	7.199
0.691	82.555	0.747	58.819	0.871	6.873	0.537	7.357	0.875	6.972
0.845	16.946	0.834	26.205			0.607	13.991		
						0.752	52.314		
						0.833	29.278		

CN1

D1		D2		D3		D4		D5	
τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$
0.223	6.489	0.199	4.249	0.135	5.179	0.199	4.430	0.126	6.754
0.240	15.367	0.248	9.392	0.279	7.592	0.247	5.722	0.253	7.047
0.484	7.891	0.391	5.525	0.400	8.232	0.434	9.099	0.378	7.106
0.646	15.628	0.611	74.327	0.518	8.577	0.525	88.880	0.502	7.132
0.691	82.555	0.640	7.553	0.633	9.430	0.607	13.991	0.626	7.199
0.845	16.946	0.747	58.819	0.871	6.873	0.752	52.314	0.875	6.972
		0.834	26.205			0.833	29.278		

CN2

D1		D2		D3		D4		D5	
τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$
0.223	64.962	0.198	6.092	0.135	5.179	0.196	6.154	0.126	6.754
0.329	6.489	0.244	4.249	0.279	7.592	0.243	4.430	0.253	7.047
0.387	44.948	0.441	9.392	0.400	8.232	0.434	9.099	0.378	7.106
0.641	4.696	0.611	74.327	0.518	8.577	0.537	28.552	0.502	7.132
0.845	16.946	0.640	7.553	0.633	9.430	0.607	13.991	0.626	7.199
		0.747	58.819	0.871	6.873	0.752	52.314	0.875	6.972
		0.834	26.205			0.833	29.278		

CN3

D1		D2		D3		D4		D5	
τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$	τ	$\sum_{i \notin J} R_{iJ}$
0.180	37.206	0.146	4.808	0.122	6.014	0.151	4.756	0.125	6.873
0.251	18.885	0.357	93.054	0.256	5.179	0.361	5.722	0.251	6.754
0.391	21.806	0.397	9.392	0.400	8.232	0.500	9.099	0.378	7.106
0.484	41.262	0.447	5.525	0.518	8.577	0.643	7.109	0.502	7.132
0.646	15.628	0.640	7.553	0.633	9.430	0.752	52.314	0.626	7.199
0.691	82.555	0.747	58.819	0.871	6.873	0.833	29.278	0.875	6.972
0.845	16.946	0.834	26.205						

In the next section we focus on the prediction aspect of RQs.

5.5 The elemental predicted residual sum of squares

The elemental predicted residuals, e_{iJ}'s defined in subsection 2.4.1 are useful in the single case scenario. For the RQ (multiple) case, the natural statistic to use is the elemental predicted sum of squares, $PRESS_J$ (see subsection 2.4.1). Table 5.9 below gives these statistics for the artificial data.

Table 5.9: EPR Sum of Squares, $PRESS_J$ at different design matrices and distribution scenarios.

NORMAL

D1		D2		D3		D4		D5	
τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$
0.2008	0.1461	0.1988	0.1107	0.1348	0.1776	0.1994	0.1182	0.1260	0.2324
0.2679	0.1112	0.2480	0.1079	0.2788	0.0945	0.2474	0.1131	0.2530	0.0966
0.2803	0.0993	0.5288	0.0861	0.3960	0.0903	0.4338	0.1122	0.3772	0.0929
0.3772	0.0965	0.5417	0.0833	0.5176	0.0904	0.4844	0.1090	0.5018	0.0932
0.5153	0.0951	0.6107	0.0819	0.6330	0.1012	0.5255	0.0904	0.6258	0.1058
0.6915	0.0951	0.7468	0.1075	0.8708	0.1155	0.5375	0.0856	0.8746	0.1108
0.8453	0.2224	0.8342	0.1379			0.6073	0.0841		
						0.7515	0.1185		
						0.8329	0.1270		

CN1

τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$
0.2231	0.5177	0.1988	0.5013	0.1348	0.5816	0.1994	0.5145	0.1260	0.6498
0.2396	0.5164	0.2480	0.4976	0.2788	0.5050	0.2474	0.5074	0.2530	0.4939
0.4837	0.5130	0.3913	0.4947	0.4004	0.4427	0.4338	0.5062	0.3777	0.4400
0.6462	0.4399	0.6107	0.4885	0.5176	0.4135	0.5255	0.4996	0.5018	0.4155
0.6915	0.4064	0.6402	0.4835	0.6330	0.4051	0.6073	0.4907	0.6258	0.4066
0.8453	0.4541	0.7468	0.4055	0.8708	0.4063	0.7515	0.4087	0.8746	0.4068
		0.8342	0.4134			0.8329	0.4108		

CN2

τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$
0.2231	0.5503	0.1978	0.5522	0.1348	0.5765	0.1964	0.5554	0.1260	0.6429
0.3289	0.5177	0.2438	0.5014	0.2788	0.5017	0.2430	0.5146	0.2530	0.4907
0.3868	0.5112	0.4409	0.4980	0.4004	0.4417	0.4338	0.5071	0.3777	0.4391
0.6407	0.4438	0.6107	0.4902	0.5176	0.4356	0.5375	0.4834	0.5018	0.4356
0.8453	0.4571	0.6402	0.4868	0.6330	0.4072	0.6073	0.4945	0.6258	0.4091
		0.7468	0.4075	0.8708	0.4094	0.7515	0.4110	0.8746	0.4097
		0.8342	0.4159			0.8329	0.4133		

CN3

τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$	τ	$PRESS_J$
0.1797	4.9194	0.1458	4.4498	0.1216	4.6090	0.1507	4.4511	0.1246	4.6597
0.2506	4.5160	0.3569	4.3767	0.2564	4.4606	0.3613	4.3524	0.2507	4.5756
0.3910	4.5039	0.3972	4.3263	0.4004	4.1265	0.4996	4.3499	0.3777	4.1070
0.4837	4.3722	0.4473	4.3204	0.5176	4.0096	0.6426	4.3049	0.5018	4.0086
0.6462	4.1253	0.6402	4.2865	0.6330	3.9337	0.7515	3.8890	0.6258	3.9244
0.6915	3.9607	0.7468	3.9132	0.8708	3.8892	0.8329	3.8690	0.8746	3.9078
0.8453	3.7298	0.8342	3.8424						

We tentatively take $PRESS_J > 2.SSE$ as indicative of a "large" $PRESS_J$ from the analogy of OLS predicted residual sum of squares (leave one out) $PRESS$ (see section 2.4.1).

At D1 the RQ predicted residual sum of squares, $PRESS_J$ s are smallest under the normal distribution, increase by about the same factor under $CN(0.1,9)$ and then increase dramatically under $CN(0.1,100)$. At the remaining design matrices, the predicted residual sum of squares, $PRESS_J$ s follow a similar pattern to that exhibited at D1. Actually, the effect of the choice of design matrices is not visible on prediction.

Multiple leverage and/or outliers may be influential. To get an insight into the influence picture, we focus on the covariance ratio as an influential measure in the next section.

5.6 The covariance ratio

The multiple case covariance ratio, CVR_J is a statistic obtained by extension of the single case CVR_i statistic (see section 4.8). Recall from (4.8.1) that

$$CVR_J = \left(\frac{\omega_J PRESS_J}{SSE} \right)^p \omega_J^{-1}.$$

We need to find reasonable cut-off values for CVR_J. Tentatively taking $PRESS_J > 2.SSE$ from the analogy of OLS predicted residual (leave one out) situation as indicative of a "large" $PRESS_J$ and $\omega_J > 2K^{-1}$ as indicative of a "large" ω_J, *i.e.*, $\tilde{c} = \tilde{a} = 2$, our cut-off value for CVR_J, $\left(\frac{\tilde{c}\tilde{a}}{K} \right)^p \frac{K}{2}$ becomes $\left(\frac{4}{K} \right)^p \left(\frac{K}{2} \right)$ (see section 4.8). Hence, for n=20 and p=5 we tentatively flag $CVR_J > \left(\frac{4}{K} \right)^p \left(\frac{K}{2} \right)$=8.861e-15 as an influential set. Again we see from the simulation that this cut-off value is too small (see table 5.10 below).

Table 5.10: RQ Covariance Ratio, CVR_J at different design matrices under the normal distribution using a sample size of n =20.

D1		D2		D3		D4		D5	
τ	CVR_J	τ	CVR_J	τ	CVR_J	τ	CVR_J	τ	CVR_J
0.1446	3.39E-13	0.1038	1.53E-08	0.0646	2.23E-04	0.1031	1.64E-08	0.0634	2.74E-04
0.1567	3.95E-14	0.2029	1.46E-08	0.2133	1.53E-15	0.2015	9.80E-10	0.1681	1.52E-06
0.2442	3.07E-10	0.2231	8.88E-10	0.2359	3.92E-10	0.2460	1.57E-13	0.2128	2.05E-15
0.2512	7.35E-12	0.2468	4.03E-13	0.3320	4.67E-16	0.3245	3.16E-12	0.2363	4.26E-10
0.2699	7.05E-15	0.3247	1.71E-13	0.3582	1.01E-12	0.3456	4.81E-13	0.3332	9.01E-13
0.3626	4.10E-14	0.3455	3.36E-12	0.4268	1.56E-10	0.4297	8.74E-12	0.3963	1.06E-15
0.3904	1.32E-13	0.4269	4.18E-13	0.5126	1.03E-11	0.5478	1.20E-10	0.4243	1.56E-08
0.4192	1.69E-14	0.4676	9.85E-12	0.5366	1.61E-10	0.5588	1.92E-10	0.4632	2.91E-10
0.4366	2.59E-13	0.5483	1.32E-10	0.5864	9.07E-08	0.6482	4.02E-10	0.4989	8.22E-20
0.5181	1.49E-15	0.5596	1.67E-10	0.6377	1.34E-07	0.7018	5.67E-10	0.5057	8.77E-12
0.5442	4.53E-14	0.6482	3.16E-10	0.7346	2.11E-07	0.8094	5.67E-10	0.5309	2.78E-10
0.5990	4.53E-14	0.7012	5.87E-10	0.8589	3.00E-07	0.8581	3.39E-10	0.5845	5.28E-09
0.6339	3.28E-13	0.8096	3.39E-10	0.9025	3.76E-08	0.8963	1.80E-12	0.6336	1.87E-07
0.6909	2.30E-13	0.8965	1.67E-12					0.7323	3.16E-07
0.6957	1.32E-10							0.8586	3.78E-09
0.8111	5.38E-11								
0.8623	1.91E-11								
0.8987	4.27E-14								

Clearly, from this example, we flag almost all RQs based on $CVR_J > \left(\frac{4}{K}\right)^p \left(\frac{K}{2}\right)$=8.861e-15, even at the ideal design. Thus we need to obtain more realistic cut-off values using a simulation study. This will be done in section 5.10, based also on the simulations in sections 5.8 and 5.9.

5.7 Discussions, conclusions and further work

Clearly, from the artificial data, the single case leverage analogy of the Hoaglin and Welsch (1978) threshold value idea (see, Chapter 4) cannot be extended directly to the multiple (RQ) case leverage. This may be due to the fact that RQs have a high affinity for leverage points resulting in generally larger values of the leverage statistic, T_J. Hence, it may not be sensible to compare the RQs leverage cases with other ESs that do not correspond to RQs since RQs would have higher values of the statistic T_J and the ERW ω_J. Also, the total number of unique RQs (ESs) from the solution of LP problem 3.2.1 is approximately n. Actually, if we impose the equality of slopes to ensure that the RQ-planes are parallel, the total number of

unique RQs is n (see *e.g.*, Koenker, 1984). On the other hand, the total number of ESs is $K = \binom{n}{p}$ implying the number of unique RQ's<<K. This results in the threshold value of T_J, $2pK^{-1}$ being too small for the intended purpose of flagging high leverage RQs. In view of all the ESs (including RQs) it is important to understand what differentiates RQs from the other ESs. One way that could be informative is to study the ERW, ω_J.

In subsection 2.2.1, the OLS estimator $\hat{\hat{\boldsymbol{\beta}}}_{OLS}$, given in (2.2.2), is a weighted average of elemental regression estimators $\hat{\hat{\boldsymbol{\beta}}}_J$, $J = 1,...,K$, *viz.*,

$$\begin{aligned}\hat{\hat{\boldsymbol{\beta}}}_{OLS} &= \sum\nolimits_J \omega_J \hat{\hat{\boldsymbol{\beta}}}_J \\ &= \sum\nolimits_{J \in B(\tau)} \omega_J \hat{\hat{\boldsymbol{\beta}}}_J + \sum\nolimits_{J \notin B(\tau)} \omega_J \hat{\hat{\boldsymbol{\beta}}}_J .\end{aligned}$$

Here $B(\tau)$ is the convex hull of solutions to the LP problem (3.2.1) corresponding to RQs (see subsection 4.5.1 or Theorem A.3, appendix A). Recall that in the construction of the design matrices, D2 and D4 are similar and D3 and D5 are similar, so it is reasonable to compare their results in these respective pairs. It is of interest to see how much contribution the ESs corresponding to the entire RQs LP solution make towards building $\hat{\hat{\boldsymbol{\beta}}}_{OLS}$ for the different design matrices, since $\sum_J \omega_J = 1$. These contributions are approximately 10.26% for D1, 26.30% for D2, 69% for D3, 34.30% for D4 and 75% for D5. Clearly, leverage points resulted in the increase of the ω_J's and hence the contribution of $\hat{\hat{\boldsymbol{\beta}}}_J$ (the ESs corresponding to RQs) to $\hat{\hat{\boldsymbol{\beta}}}_{OLS}$ is much more compared to other ESs which do not correspond to RQs. As a consequence, the adverse effects of leverage in RQs are inherent in the OLS estimators.

Actually, all the OLS statistics can be expressed as weighted averages of elemental regressions and suffer the adverse effects of leverage via the ERW, ω_J. As an example, SSE is a scaled weighted average of $PRESS_J$.

As explained in section 4.7, leverage influential points can induce collinearity or hide (mask) it. Intuitively, the induction of collinearity should result in a small ω_J but this is not the case since the effect of a small collinearity factor ρ_J is counteracted with a relatively large value of the scale factor γ_J due to the presence of leverage points in the RQs. So, ω_J remains relatively large. It looks as though pertaining to leverage, RQs are in "the league of their own"; they are not quite comparable to their complement.

From the results based on $|C_J|$ and ρ_J it is clear that points 5 and 7 are **collinearity influential points** with the degree of collinearity being severe at the RQ level in a considerable number of cases as evidenced by $\rho_J < 1$. RQs are susceptible to leverage points (including collinearity influential ones) and the residual freedoms, R_{iJ} s, which measure the extent to which ES J fails to predict points outside it, are generally more or less the same. This may be due to the RQs' high affinity for leverage points, *i.e.*, RQs would include leverage points instead of "predicting" them (having them in the "predictive" set). Hence, for the ES corresponding to RQs we are not likely to encounter large R_{iJ} 's.

As an influence measure in the RQ case, CVR_J is reasonable as it comprises both the leverage and "prediction" components, ω_J and $PRESS_J$ respectively. Although by their nature RQs exhibit best prediction at $\tau = 0.5$ and poor "prediction" at extreme τ 's, the CVR_J values in table 5.10 do not portray this picture, thus emphasizing that the leverage component dominates the prediction component ($PRESS_J$) in this influential diagnostic. This is very encouraging since RQs can be susceptible to leverage but are fairly robust to residual outliers. However, similar to the leverage case, the cut-off value $CVR_J > \left(\frac{4}{K}\right)^p \left(\frac{K}{2}\right)$ is very small again. Furthermore, another drawback of this cut-off value is that the cut-off value for $PRESS_J$ is a constant, $\tilde{a} = 2$, and does not take into account the poor "prediction" exhibited by RQs at extreme τ levels. Hence, we need to determine reasonable cut-off values for this statistic (see section 5.10).

In the following sections we determine the cut-off values of these various statistics using a simulation study. We start with the leverage statistic T_J .

5.8 Determining the threshold (cut-off) values for the RQ leverage, T_J using simulation studies

In the previous section we chose a small sample of size n =10 in order to conveniently get an ad hoc picture of RQ diagnostics. Our simulation design is as in section 5.2 except that now $p = 3,...,8$, the sample size n =20, 30 and 50 and the number of simulations is 200. We now choose larger sample sizes in the simulations to study these for more realistic sample sizes.

As exhibited by the artificial data above, the threshold value for the RQ case, T_J based on the

Hoaglin and Welsch (1978) methodology is too small. Therefore we use a simulation study to determine threshold values for the RQ case of the form $\frac{cp}{K}$, $c = 3,4,5,....$ We start with the orthogonal design $D1$ under the three error distributions, *viz.*, the Gaussian, $CN(0.1,9)$ and $CN(0.1,100)$. At this design we want to find the minimum value of c, say $\tilde{c}$ for which we would flag no T_J's corresponding to RQs. This is so because D1 is our "ideal" design and we are not expected to flag any RQ at this design matrix. We then propose $\frac{\tilde{c}p}{K}$ as a threshold value.

Remark: Since T_J is entirely a function of X and is therefore not distribution dependent, we only give results for the Gaussian case.

We start our search for reasonable cut-off values by investigating the multiple (RQ) predictive leverage statistic T_J. For each design matrix we show the display depicted both in table form and graphically. The graphical representation gives us some idea of the distribution of the T_J's since we plot the τ levels (and p sizes) against the quantiles of the simulated T_J's.

We take values of $\tilde{c}$ to be the set $\{4,7,15,21,26,34\}$ corresponding to the sizes of p in the set $\{3,4,5,6,7,8\}$ respectively. Hence for these sizes of p, the threshold values are $\left\{\frac{4p}{K},\frac{7p}{K},\frac{15p}{K},\frac{21p}{K},\frac{26p}{K},\frac{34p}{K}\right\}$ respectively.

Results for the D1 matrix

Table 5.11 below gives $\tilde{c}$ and $\tilde{c}-1$ for $p = 3,...,8$ and n =20.

Table 5.11: Number of T_J 's flagged for sample size n =20 and D1 matrix under the Gaussian distribution based on $\tilde{c}$ and $\tilde{c}-1$ and 200 simulation runs

τ	p =3		p =4		p =5		p =6		p =7		p =8	
	$\tilde{c}-1$	$\tilde{c}$ =4	$\tilde{c}-1$	$\tilde{c}$ =7	$\tilde{c}-1$	$\tilde{c}$ =15	$\tilde{c}-1$	$\tilde{c}$ =21	$\tilde{c}-1$	$\tilde{c}$ =26	$\tilde{c}-1$	$\tilde{c}$ =34
0.0952	8	0	0	0	1	0	1	0	0	0	1	0
0.1429	14	0	4	0	2	0	1	0	0	0	1	0
0.1905	14	0	5	0	0	0	2	0	0	0	0	0
0.2381	11	0	5	0	0	0	2	0	0	0	0	0
0.2857	11	0	6	0	1	0	0	0	0	0	0	0
0.3333	9	0	5	0	1	0	0	0	1	0	0	0
0.3810	9	0	7	0	1	0	0	0	1	0	0	0
0.4286	9	0	7	0	0	0	1	0	0	0	0	0
0.4762	8	0	6	0	0	0	0	0	1	0	0	0
0.5238	15	0	3	0	1	0	1	0	0	0	1	0
0.5714	15	0	7	0	2	0	1	0	0	0	1	0
0.6190	15	0	8	0	2	0	1	0	0	0	0	0
0.6667	18	0	6	0	0	0	1	0	0	0	0	0
0.7143	16	0	7	0	1	0	1	0	0	0	0	0
0.7619	17	0	2	0	2	0	1	0	0	0	0	0
0.8095	9	0	1	0	1	0	0	0	0	0	0	0
0.8571	7	0	3	0	0	0	0	0	1	0	0	0
0.9048			2	0	0	0	0	0	1	0	0	0
Mean	12.1	0.0	4.7	0.0	0.8	0.0	0.7	0.0	0.2	0.0	0.2	0.0

Based on the orthogonal ("clean") D1 we do not flag any T_J 's at the given $\tilde{c}$ for a given size of p.

Graphically the plot of the τ levels (and p sizes) against the quantiles of the 200 simulated T_J values, $\underset{1\le i\le 200}{quant}(T_{Ji})$ is as in figure 5.2 below:

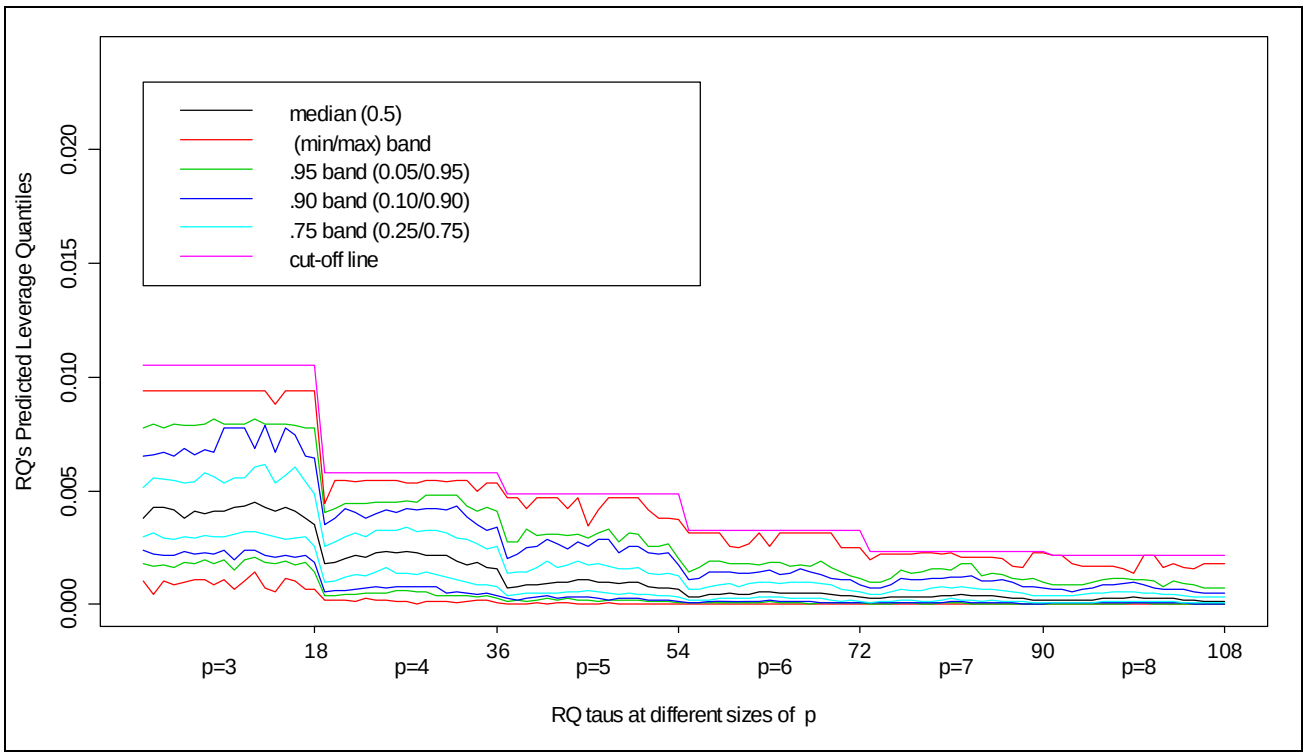


Figure 5.2: RQ Leverage based on D1 under the Normal distribution at different sizes of p

At D1 under the Gaussian distribution, the entire solution to the LP problem (3.2.1), $B(\tau)$ consists of 18 unique ESs corresponding to RQs at the given sizes of p, *i.e.*, $\{3,4,...,8\}$, thus the horizontal axis consists of 18 τ levels per given size of p. The vertical axis consists of the (min/max), 95%, 90%, 75% bands, the median and the cut-off values for the 200 simulated T_J values as shown in the legend.

Remark: For the graphical representations that will follow and the various statistics we are investigating, the horizontal axis will stay the "same" as in figure 5.2, *i.e.*, we will fix the τ's at equally spaced 18 levels at the given p sizes. This helps us to carry out convenient comparisons amongst the different scenarios as each scenario may result in a different number of unique RQs.

Note that the distribution of the T_J's generally decreases in an approximate stepwise fashion. That is, as K increases (or as p approaches $\frac{n}{2}$), the statistic T_J generally becomes smaller and smaller. However, as shown in the artificial data set, $2pK^{-1}$ remains too small a cut-off

value. Hence we determined the new cut-off values shown in violet colour. Note that the cut-off values flag no RQs at this "clean" case as expected. It is interesting to see if these cut-off values will result in the flagging of RQs in the presence of different leverage scenarios. We first give results for one leverage point scenario and D2 and D4.

Results for D2 and D4 matrices

Since D2 and D4 matrices are similar, we present their results together here.

Table 5.12 below gives results for the D2 matrix.

Table 5.12: Number of T_J 's flagged for sample size n =20 and D2 matrix under the Gaussian distribution based on $\tilde{c}$ and c (=$\tilde{c}$ for $p+1$) and 200 simulation runs

	p =3		p =4		p =5		p =6		p =7		p =8	
τ	$\tilde{c}$ =4	c =7	$\tilde{c}$ =7	c =15	$\tilde{c}$ =15	c =21	$\tilde{c}$ =21	c =26	$\tilde{c}$ =26	c =34	c =26	$\tilde{c}$ =34
0.0952	151	28	73	0	25	6	10	2	0	0	5	1
0.1429	157	33	67	0	30	8	15	4	0	0	3	1
0.1905	163	15	67	0	29	10	23	10	0	0	4	1
0.2381	158	18	71	0	31	12	19	9	0	0	5	2
0.2857	158	25	62	0	41	6	15	7	0	0	15	6
0.3333	158	21	76	0	34	4	17	6	0	0	18	8
0.3810	165	24	76	0	31	3	18	9	0	0	20	3
0.4286	162	30	81	0	34	6	20	10	0	0	21	5
0.4762	156	33	84	0	32	4	24	11	0	0	15	5
0.5238	163	31	79	0	41	12	18	8	0	0	15	6
0.5714	157	27	74	0	26	10	20	10	0	0	16	6
0.6190	161	28	84	0	30	13	18	8	0	0	13	2
0.6667	154	36	88	0	39	10	17	5	0	0	18	6
0.7143	146	30	74	0	24	4	20	6	0	0	12	6
0.7619	147	28	74	0	20	3	11	5	0	0	7	2
0.8095	145	26	64	0	24	3	17	6	0	0	6	3
0.8571	145	25	67	0	20	4	8	2	0	0	4	2
0.9048	146	21	60	0	12	1	3	1	0	0	6	2
Mean	155.1	26.6	73.4	0.0	29.1	6.61	16.3	6.61	0.0	0.0	11.3	3.7

On average we flag 78 % for p =3, 37 % for p =4, 15 % for p =5, 8 % for p =6, 0.0 % for p =7, 6 % for p =8 of the RQs using the threshold values based on $\tilde{c}$ for this design matrix. Generally, the proportion of RQs being flagged decreases as p increases except for p =8.

The graphical display for D2 matrix is given in figure 5.3.

Table 5.13 gives results for D4 matrix.

Table 5.13: Number of T_J 's flagged for sample size n =20 and D4 matrix under the Gaussian distribution based on $\tilde{c}$ and c (=$\tilde{c}$ for $p+1$) and 200 simulation runs

τ	p =3		p =4		p =5		p =6		p =7		p =8	
	$\tilde{c}$ =4	c =7	$\tilde{c}$ =7	c =15	$\tilde{c}$ =15	c =21	$\tilde{c}$ =21	c =26	$\tilde{c}$ =26	c =34	c =26	$\tilde{c}$ =34
0.0952	156	36	70	0	26	6	12	2	5	0	5	1
0.1429	161	35	64	0	30	9	16	4	6	0	3	1
0.1905	162	26	74	0	36	12	23	10	7	2	4	2
0.2381	163	25	70	0	37	14	19	9	9	2	6	2
0.2857	162	35	61	0	45	10	16	7	9	2	14	7
0.3333	161	32	74	0	37	5	17	6	9	2	15	7
0.3810	165	31	75	0	38	5	20	9	5	2	19	3
0.4286	160	30	82	0	39	7	20	11	5	2	20	4
0.4762	157	39	88	0	33	6	23	11	7	1	15	5
0.5238	159	40	89	0	41	16	19	8	13	3	15	7
0.5714	165	35	75	0	29	11	24	11	14	6	18	7
0.6190	160	26	84	0	33	12	22	10	10	3	12	2
0.6667	161	37	87	0	37	9	17	5	12	4	19	7
0.7143	151	35	81	0	32	5	17	5	7	2	12	7
0.7619	149	38	70	0	27	3	13	7	3	1	8	2
0.8095	150	29	66	0	26	3	14	7	2	0	5	2
0.8571	149	34	69	0	24	5	9	4	4	0	3	1
0.9048	153	26	60	0	14	3	8	1	2	0	4	1
Mean	158.0	32.7	74.4	0.0	32.4	7.8	17.2	7.1	7.2	1.8	10.9	3.8

On average we flag 79 % for p =3, 37 % for p =4, 16 % for p =5, 9 % for p =6, 4 % for p =7, 5 % for p =8 of the RQs using the threshold values based on $\tilde{c}$ for this design matrix. Generally, the proportion of RQs being flagged decreases as p increases. This pattern is more or less as that exhibited for D2. This is understandable since the observation that was moved 10 units in the X direction in D2 is moved 100 units in the same direction in D4.

In order to see the effect of the degree of leverage on the distribution of the T_J we give the graphs of the τ levels (and p sizes) against the quantiles of the 200 simulated T_J 's for D2 and D4 together since they are similar (see the simulations design in section 5.2) in figure 5.3 below.

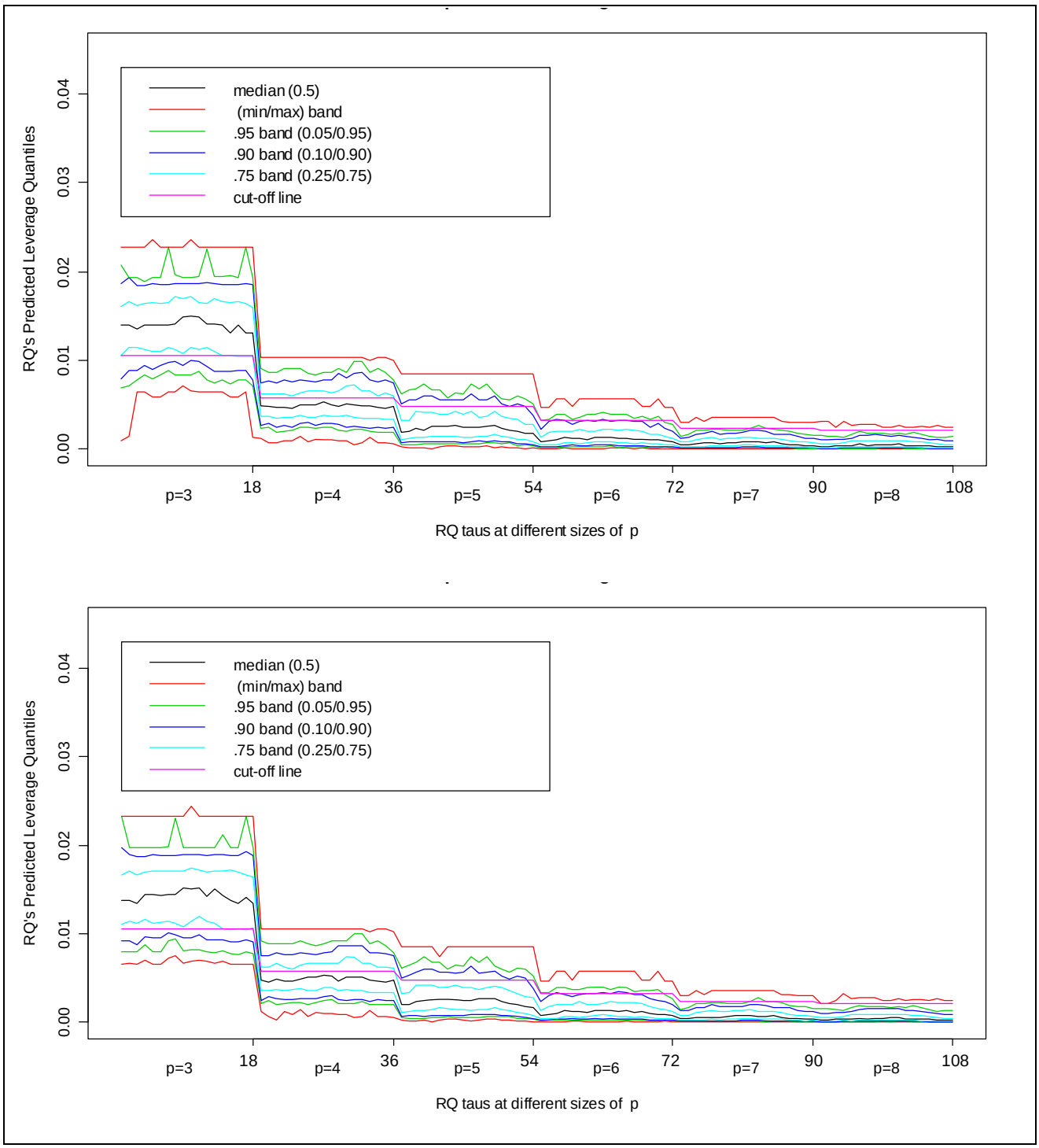


Figure 5.3: RQ Leverage based on design matrices D2 and D4 under the Normal distribution at different sizes of p . Upper panel D2; Lower panel D4

For both these design matrices (D2 and D4), T_J values approximately above the 10% quantile curve at $p = 3$, approximately above the 50% at $p = 4$ and approximately above the 75% of the T_J's at $p = 5$ are flagged. For larger values of p, the quantile curve above which one flags becomes difficult to distinguish. The general trend is clearly that one flags less as p increases.

Results for D3 and D5 matrices

Again since the D3 and D5 matrices are similar, we present their results together here.

Table 5.14 below gives results for the D3 matrix.

Table 5.14: Number of T_J 's flagged for sample size n =20 and D3 matrix under the Gaussian distribution based on $\tilde{c}$ and c (=$\tilde{c}$ for $p+1$) and 200 simulation runs

τ	p =3		p =4		p =5		p =6		p =7		p =8	
	$\tilde{c}$ =4	c =7	$\tilde{c}$ =7	c =15	$\tilde{c}$ =15	c =21	$\tilde{c}$ =21	c =26	$\tilde{c}$ =26	c =34	c =26	$\tilde{c}$ =34
0.0952	181	181	198	136	116	71	58	36	16	6	30	16
0.1429	191	191	200	138	120	75	66	52	26	10	37	17
0.1905	196	196	200	134	137	89	74	57	46	19	38	18
0.2381	200	200	200	132	130	84	75	53	46	21	38	19
0.2857	200	200	200	133	135	86	80	55	49	20	48	24
0.3333	200	200	200	146	134	94	76	54	48	24	55	34
0.3810	200	200	200	138	144	91	72	57	55	28	55	37
0.4286	200	200	200	137	138	89	89	72	51	22	61	32
0.4762	200	200	200	149	130	86	83	58	63	31	57	36
0.5238	200	200	200	151	126	76	92	62	63	35	50	32
0.5714	200	200	200	143	129	70	88	63	59	28	49	31
0.6190	200	200	200	144	131	80	95	75	56	21	54	32
0.6667	200	200	200	142	128	83	88	74	55	24	46	26
0.7143	200	200	200	126	130	73	82	62	45	16	48	24
0.7619	199	199	200	127	124	74	65	49	43	14	44	17
0.8095	198	198	200	133	111	70	54	41	40	11	33	19
0.8571	196	196	199	124	112	61	49	36	29	9	31	20
0.9048	180	180	199	133	101	52	42	27	20	5	23	15
Mean	196.7	196.7	199.8	137.0	126.4	78.0	73.8	54.6	45.0	19.1	44.3	24.9

On average we flag 98 % for p =3, 100 % for p =4, 63 % for p =5, 37 % for p =6, 23% for p =7, 22 % for p =8 of the RQs using the threshold values based on $\tilde{c}$ for this design matrix. Generally, the proportion of RQs being flagged decreases as p increases.

Table 5.15 below gives results for the D5 matrix.

Table 5.15: Number of T_J 's flagged for sample size n =20 and D5 matrix under the Gaussian distribution based on $\tilde{c}$ and c (= $\tilde{c}$ for $p+1$) and 200 simulation runs

τ	p =3		p =4		p =5		p =6		p =7		p =8	
	$\tilde{c}$ =4	c =7	$\tilde{c}$ =7	c =15	$\tilde{c}$ =15	c =21	$\tilde{c}$ =21	c =26	$\tilde{c}$ =26	c =34	c =26	$\tilde{c}$ =34
0.0952	200	200	200	156	119	76	67	34	19	7	31	21
0.1429	200	200	200	150	118	85	69	51	24	11	38	22
0.1905	200	200	200	148	140	97	74	61	45	19	37	19
0.2381	200	200	200	149	128	90	80	53	46	29	40	17
0.2857	200	200	200	140	133	92	79	53	49	22	49	22
0.3333	200	200	200	160	133	94	76	56	44	26	58	33
0.3810	200	200	200	151	141	89	73	63	52	25	58	42
0.4286	200	200	200	158	136	82	88	69	55	24	59	35
0.4762	200	200	200	161	130	82	89	61	62	32	57	36
0.5238	200	200	200	159	127	79	89	60	64	35	57	36
0.5714	200	200	200	153	125	72	88	64	59	28	53	30
0.6190	200	200	200	161	136	84	95	74	55	20	58	31
0.6667	200	200	200	148	138	90	95	73	53	26	49	28
0.7143	200	200	200	151	129	84	79	61	47	18	43	21
0.7619	200	200	200	137	124	75	64	47	42	15	40	16
0.8095	200	200	200	138	114	73	60	39	42	12	30	16
0.8571	200	200	200	144	114	67	51	37	31	14	27	14
0.9048	200	200	200	145	104	59	44	27	18	10	24	13
Mean	200.0	200.0	200.0	150.5	127.2	81.7	75.6	54.6	44.8	20.7	44.9	25.1

On average we flag 100 % for p =3, 100 % for p =4, 64 % for p =5, 38 % for p =6, 22% for p =7, 23 % for p =8 of the RQs using the threshold values based on $\tilde{c}$ for this design matrix. Generally, the proportion of RQs being flagged decreases as p increases. This pattern is similar to that exhibited at D3. This is understandable since the two observations that were moved 10 units in the X direction are moved 100 units in the same direction.

In order to see the effect of the degree of leverage on the distribution of the T_J we give the plots of the τ levels (and p sizes) against the quantiles of the T_J 's for D3 and D5 together since they are similar (see section 5.2) as shown in figure 5.4 below.

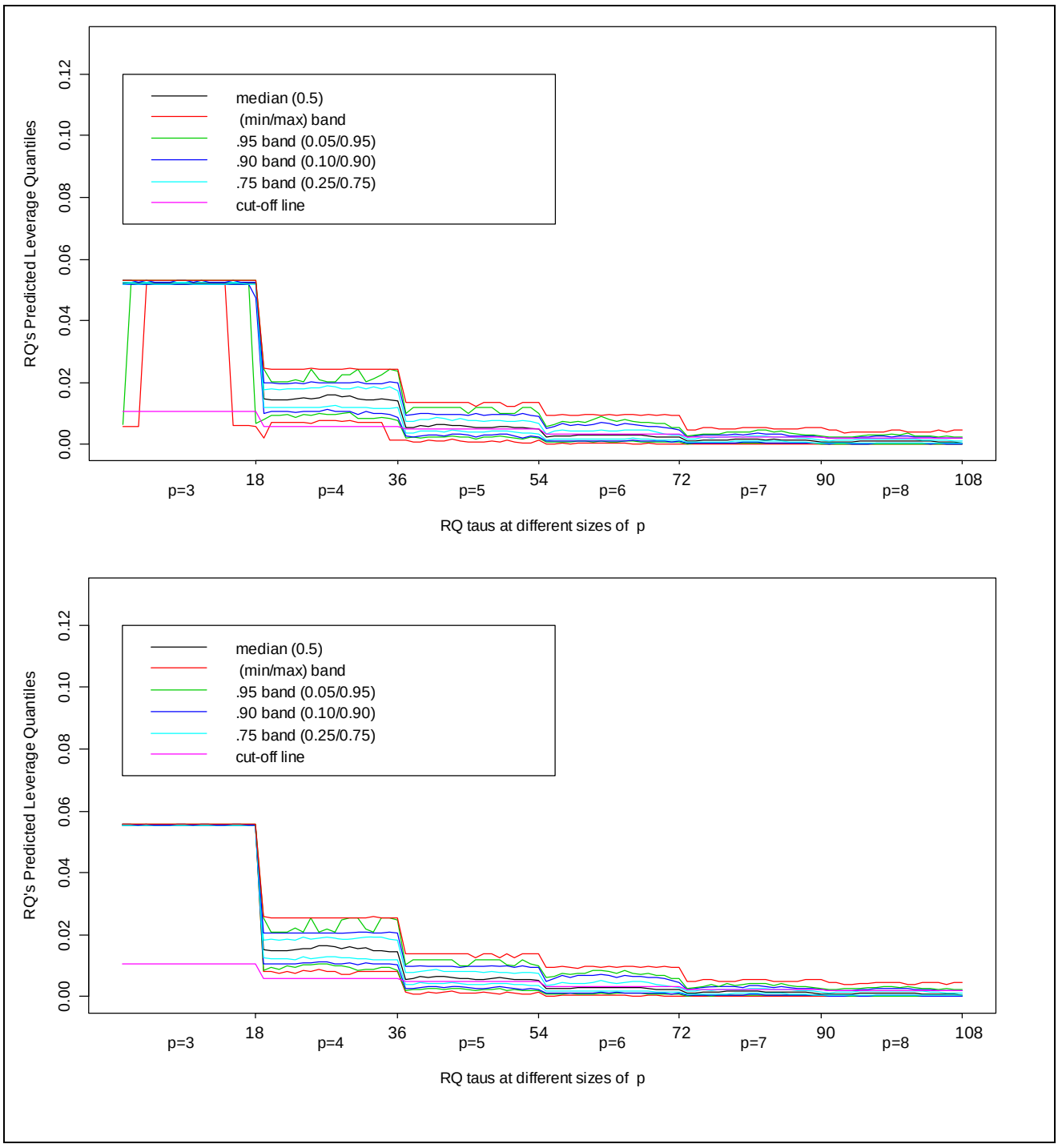


Figure 5.4: RQ Leverage based on D3 and D5 under the Normal distribution under the Normal distribution at different sizes of p . Upper panel D3; Lower panel D5

Again the results conveyed at D3 are similar to those conveyed at D5. In the same way the results conveyed at D2 are similar to those conveyed at D4 due to the similarity of D3 and D5 and D2 and D4 respectively. Approximately 98% T_J 's are flagged at D3 while100% are flagged at D5 for $p = 3$, approximately 100% at $p = 4$ for both design matrices. To get the complete picture of the proportions flagged, see the two tables above. Actually, the proportion of the T_J values flagged generally decreases in an approximate stepwise form.

It is of interest to see the summary picture of these simulation studies for the cut-off values T_J for different sizes of n. In the next section we focus on this aspect.

5.8.1 Summary picture for the cut-off values for T_J

It is important to investigate whether there are serious deviations at the cut-off values proposed above at different sample sizes. The following table gives a summary of the simulation results. It consists of the proportions of T_J's flagged at different sample sizes, number of predictors and $\tilde{c}$'s as shown in table 5.16 below.

Table 5.16: Proportions of T_J's flagged for different sample sizes and design matrices under the Gaussian distribution based on $\tilde{c}$ and 200 simulation runs

		p =3	p =4	p =5	p =6	p =7	p =8
		$\tilde{c}$ =4	$\tilde{c}$ =7	$\tilde{c}$ =15	$\tilde{c}$ =21	$\tilde{c}$ =26	$\tilde{c}$ =34
D1	n =20	0.0000	0.000	0.000	0.0000	0.0000	0.0000
	n =30	0.0432	0.043	0.027	0.0341	0.0300	0.0288
	n =50	0.0002	0.048	0.0258	0.0396	0.0708	0.0652
D2	n =20	0.7756	0.367	0.145	0.0814	0.0350	0.0186
	n =30	0.9671	0.721	0.324	0.2771	0.2489	0.1934
	n =50	0.945	0.894	0.594	0.4943	0.0708	0.3786
D3	n =20	0.9836	0.999	0.632	0.3689	0.2250	0.1247
	n =30	0.9611	0.987	0.953	0.8486	0.7259	0.6579
	n =50	0.9922	0.954	0.952	0.9419	0.9108	0.8688
D4	n =20	0.7900	0.372	0.162	0.0858	0.0358	0.0189
	n =30	0.9907	0.733	0.336	0.2870	0.2452	0.1946
	n =50	1.0000	0.919	0.613	0.5011	0.5107	0.3833
D5	n =20	1.0000	1.000	0.636	0.3778	0.2242	0.1256
	n =30	1.0000	1.000	0.971	0.8723	0.7368	0.6668
	n =50	0.9986	0.919	0.998	0.9824	0.9484	0.8883

At D1 0% of the T_J's are flagged for n =20 and all combinations of p's and $\tilde{c}$. However, at n =30 and n =50 the ranges of the T_J's flagged are from 2.7% to 4.3% and from 0% to 7.1% respectively. There is no particular way in which the flagging occurs at this design matrix. Although D1 is the "clean" case, it is not surprising that a small proportions of T_J's are flagged (due to chance) since observations which are not individually outlying can be

outlying jointly (see, *e.g.* Barrett and Gray, 1997b).

At the remaining design matrices the ranges of the T_J 's flagged are 77.6% to 100% for p =3 and $\tilde{c}$ =4, 36.7% to 100% p =4 and $\tilde{c}$ =7, 14.5% to 99.8% for p =5 and $\tilde{c}$ =15, 8.1% to 98.2% for p =6 and $\tilde{c}$ =21, 3.5% to 94.8% for p =7 and $\tilde{c}$ =26 and 1.9% to 88% for p =8 and $\tilde{c}$ =34.

Remark: Note that using the same idea of the average of the ERWs, the cut-off value for the ERW, ω_J is $\tilde{c}K^{-1}$ since $\sum_J \omega_J = 1$.

5.8.2 Conclusions on the statistic T_J

Based on the statistic T_J and the suggested cut-off values more RQs are flagged when the number of predictors is small. This is so because RQs have a high affinity for leverage points, implying that almost all RQs contain the leverage points present in the data. For instance, if p =3 and there are 2 or 3 leverage points, RQs will "force" these points into them. This results in large ω_J s and T_J s and lack of variability in both statistics. Also, in some leverage instances such as D3 and D5, the residual freedoms, R_{iJ} s become smaller and with little variability, further rendering the lack of variability in the statistic T_J . Hence, we are likely to flag more T_J s when the number of leverage points is closer to the number of predictors, p as there is less variability in the T_J s. Our simulation study showed that the proposed algorithm is reasonable for flagging high leverage RQs. It is interesting to see what the situation is if we increase the sample size. Taking sample sizes n =30 and n =50 resulted in the flagging of very minimal proportions ($\leq 7\%$) of T_J s at D1 under the Normal distribution. Such small proportions are acceptable since in the literature there is evidence that points that are not individually outlying may be jointly outlying and vice versa. Hence this should be the case for the small proportions of the T_J s that are flagged at this "clean case" scenario.

While these simulation results show that a large ω_J statistic reasonably identifies RQs multiple case leverage, a small value may also be indicative of some undesirable phenomenon. A small value of T_J can be a result of a small ERW, ω_J (most likely) and small sum of the residual freedoms, $\sum_{i \notin J} R_{iJ}$. In turn a small ERW may be a result of a small

product $\gamma_J.\rho_J$, where γ_J and ρ_J raise the variability and collinearity views respectively as defined in (4.2.2). However, in the presence of leverage points this phenomenon is almost nonexistent. Also, in the presence of leverage collinearity influential points γ_J and ρ_J are often "antagonistic" in size and may not result in a small ERW and hence T_J. This interaction behaviour may not be exhibited if the collinearity arises from another source other than collinearity influential points. Clearly, the cut-off values for a small value of ω_J which may cause poor prediction in RQs (see section 4.7) and T_J need further investigation.

In order to get a more complete picture of RQs multiple diagnostics, we focus in the next subsection on the estimation ("prediction") aspect of RQs. Although we are estimating the population τ^{th} conditional quantile of Y given x, $Q_{Y|x}(\tau)$ by $\hat{Q}_{Y|x}(\tau)$ rather than predicting $\hat{Y}$ (except at $\tau = 0.5$) we shall refer to this procedure as prediction throughout. This terminology is used due to the fact that we use the RQ/ES $PRESS_J$, the analogue to the leave one observation out statistic, $PRESS$.

5.9 Determining the threshold (cut-off) values for the RQ prediction statistic, $PRESS_J$, using simulation studies

In the literature, single (usual) case poor prediction based on the OLS predicted (leave one out) residuals sum of squares $PRESS$, is identified when

$$\frac{PRESS}{SSE} \geq 2 \text{ (see, section 4.7).}$$

A proper analogy of this procedure to the RQ (ES) situation would require us to evaluate all the $K = \binom{n}{p}$ ESs and hence $PRESS_J$'s in order to determine $\tilde{a}_J$ such that

$$\frac{PRESS_J}{SSE} > \tilde{a}_J$$

identifies poor prediction.

However, the limitations of this analogy are that $\tilde{a}_J$ is not a constant (see, section 4.7) and considering all the ESs is an expensive process. Therefore for the multiple (RQ) case, we determine the cut-off values via a simulation study.

We define the ratio,

$$a_{Ji} = \frac{PRESS_{Ji}}{SSE_i},\ i = 1,...200 \text{ and } J = 1,...,18 \qquad (5.9.1)$$

Where, J corresponds to a particular τ, and obtain the plots of the τ's (and p's) against the quantiles, $\underset{1\le i\le 200}{quant}(a_{Ji})$, of the 200 simulated values of this ratio for different design matrix-error distribution combinations. Also, it is clear that the plots (in figure 5.5) of this ratio follow a sinusoidal wave fashion, approximately. Hence, we propose obtaining cut-off values using the model,

$$Y_J = \beta_0 + r\cos(\theta + \phi) + \varepsilon_i, \quad J = 1,....,18, \qquad (5.9.2)$$

where Y_J are the values, $\max\limits_{1\le i\le 200}(a_{Ji})$ determined from the "clean" scenario, *i.e.*, D1 under the Gaussian distribution, β_0 is the intercept, r is the amplitude, $\theta \in [\pi/2, 3\pi/2]$, ϕ is the phase (horizontal shift) and J corresponds to a particular τ. The rationale for using the maximum values of the 200 simulated values of the a_J's from the "clean" case scenario is that any a_J's from other scenarios greater than these values are perceived as indicating poor "prediction". In order to be as conservative (obtain the large values from model (5.2)) as possible, we apply nonlinear quantile regression estimation to this model at $\tau = 0.995$ to get the cut-off values for $PRESS_J$'s, *viz.*,

$$\tilde{a}_J = \hat{Y}_J,\ J = 1,...,18 \qquad (5.9.3)$$

We give the plots of the τ's (and p's) against the quantiles of the 200 simulated values of the a_J's, $\underset{1\le i\le 200}{quant}(a_{Ji})$, and $\tilde{a}_J$ based on the "clean" design matrix, D1 under the Normal distribution as reasonable cut-off values (shown in violet colour) for each scenario. The rationale here is that under this scenario prediction should be "good".

In order examine the effect of the distribution we give the 3 plots at D1 corresponding to the Normal distribution, CN1 and CN3 respectively as shown in figure 5.5.

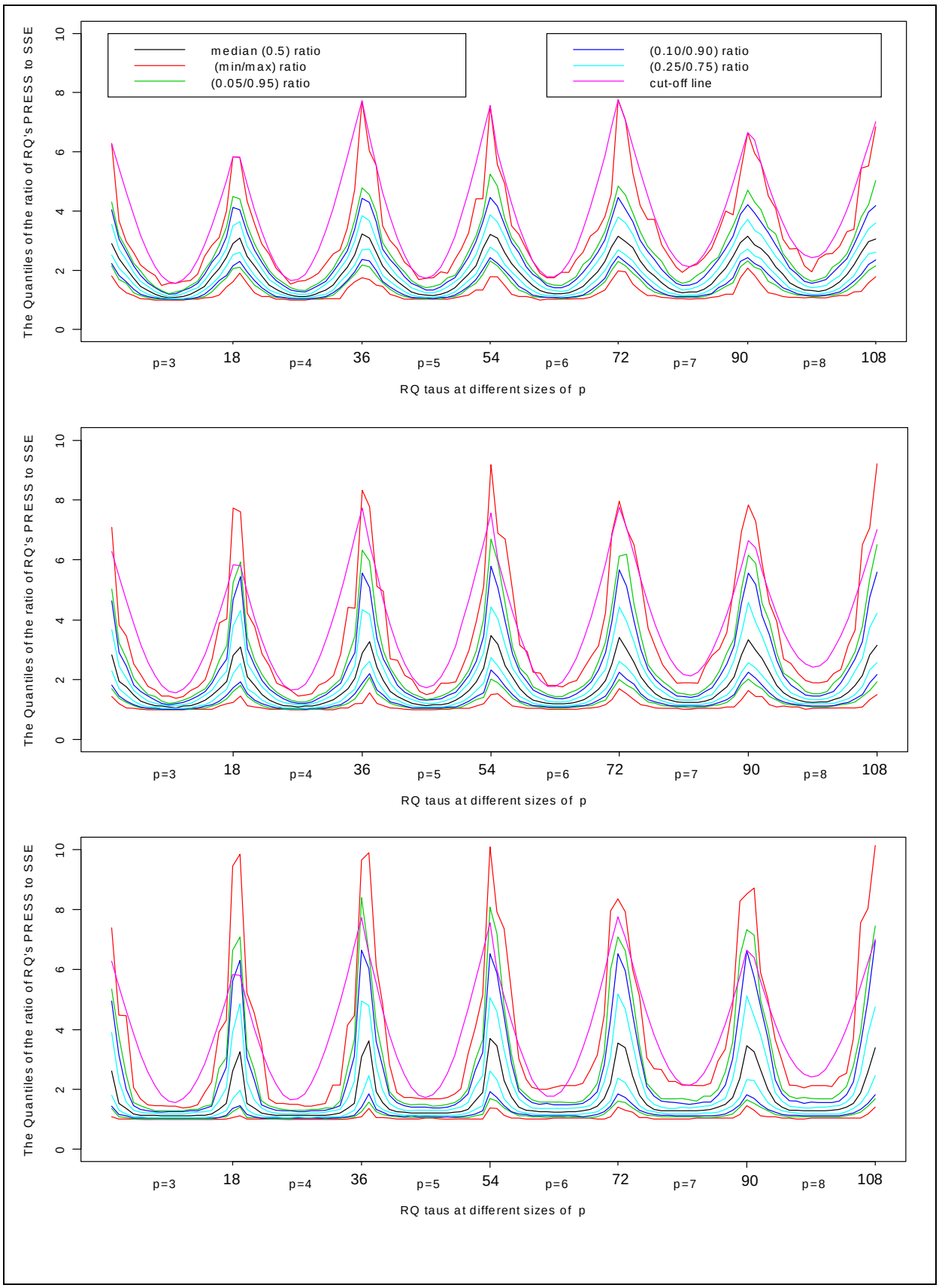


Figure 5.5: RQ "Prediction" at D1 under different error distributions. The first panel is under the Normal distribution, the second panel is under CN1 and the third panel is under CN3 distribution

Taking $\tilde{a}_J$'s obtained from 5.3 is reasonable since it ensures that $PRESS_J$ will not be flagged

at our “clean” scenario (D1 under the Normal distribution) as desired. However, under CN1 and CN3 at least the 90% quantiles of the $PRESS_J$ s are flagged. Actually, the heavier the underlying distribution the more the flagging of the $PRESS_J$ s. This is expected since CN1 and CN3 are heavy tailed error distributions which imply outliers.

On the other hand, it is important to see the effect of different leverage scenarios alone (in the absence of outliers) on prediction. In order to understand this, we plot the τ levels (and $p's$) against a_{Ji} at different design scenarios under the Normal distribution as shown in figure 5.6.

The colour legend is as in figure 5.5.

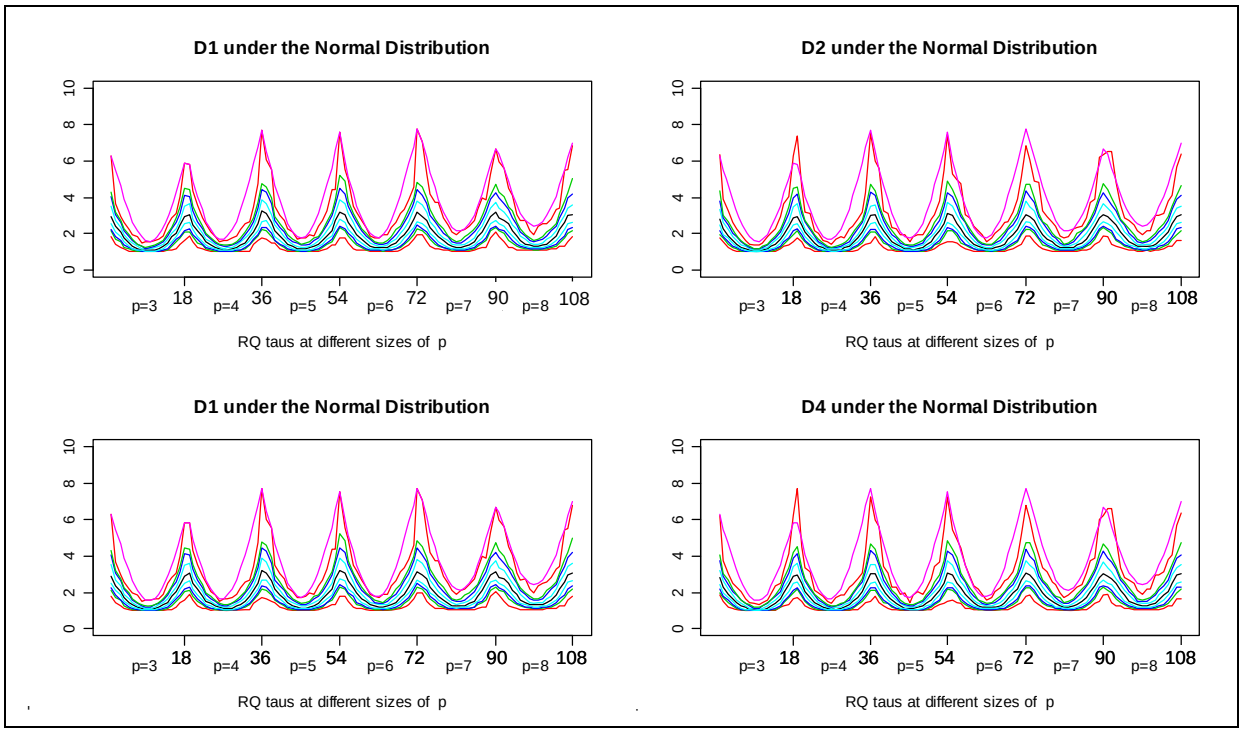


Figure 5.6: Effect of design matrices D2 and D4 on RQ "Prediction" showing D1 as the control. The Y axis represents the quantiles of the ratio of the RQ PRESS to SSE

We give the plots for D2 and D4 together in figure 5.6 since they only differ in the degree of leverage. There is not much difference in prediction at a clean design matrix and at the matrices contaminated with type A outliers (D2 and D4).

In the next plot given in figure 5.7 below we observe the effect on prediction of the kind of leverage induced by D3 and D5. The colour legend is as in figure 5.5. Again we plot them together in the following plots since they are similar.

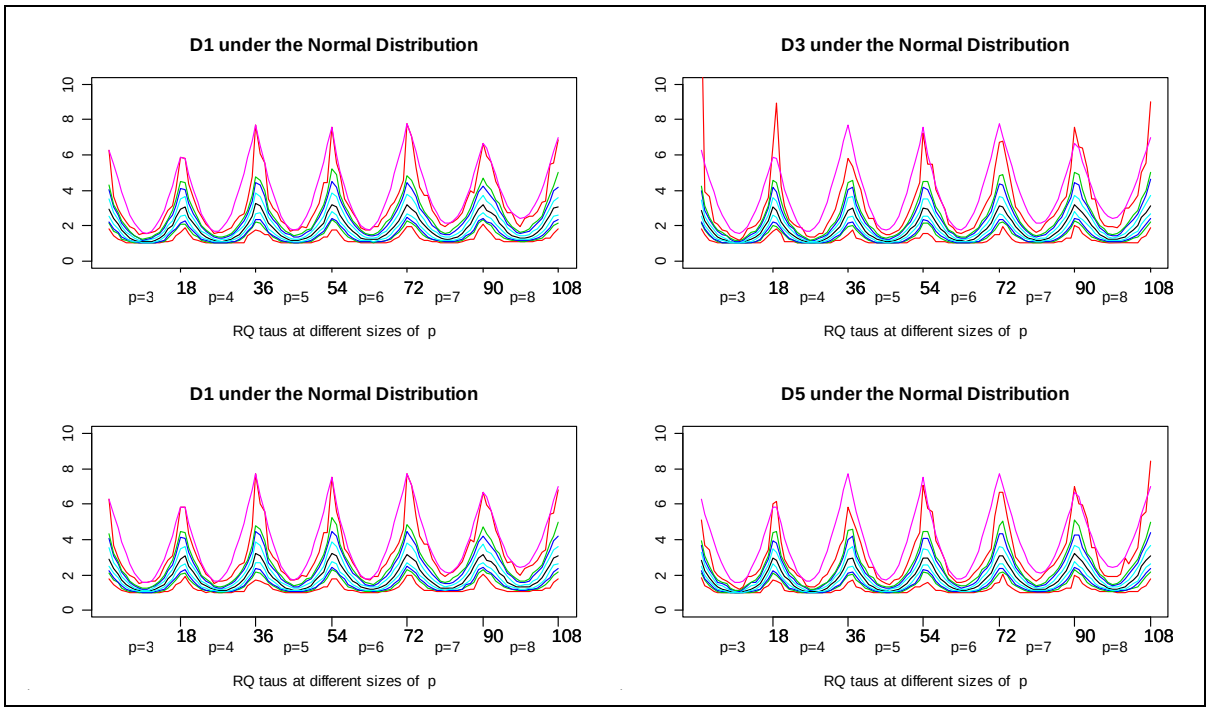


Figure 5.7: Effect of design matrices D3 and D5 on RQ "Prediction" showing D1 as the control. The Y axis represents the quantiles of the ratio of the RQ PRESS to SSE

Again, we give the plots for D3 and D5 together in figure 5.7 since they only differ in the degree of leverage. From the comparison of figures 5.6 and 5.7, type B X outliers (see section 4.6) at D3 and D5 result in poorer prediction compared to type A X outliers at D2 and D4.

So, leverage may result in RQ poor prediction but the degree of poor prediction depends on the nature of the leverage points present in the data, *i.e.*, whether the leverage points involved are type B or type A X outliers. But what if both leverage and outliers (heavy tailed error distributions) are present in the data, do the pictures exhibit poor RQ prediction? The answer is in the affirmative for all scenarios (see the plots given in appendix B). Actually, the prediction picture for type A leverage points (at D2 and D4) under CN1 and CN3 is more or less to the same degree as that exhibited without leverage points (at D1) while for type B leverage points (at D3 and D5) under CN1 and CN3 poor prediction increases drastically.

In order to evaluate the usefulness of the cut-off values derived from the robust loss function (4.7.3) we plot them together with the sinusoidal cut-offs in figure 5.8 below. Since we are

concerned with “prediction” here we consider different error distributions at the D1 matrix.

Clearly, under the normal distribution the robust loss function cut-off values are more like the sinusoidal cut-off values. However, there is an increase in the size of the cut-off values at extreme RQs as the underlying error distribution becomes heavier. This is expected since the robust loss function (4.7.3) is designed not to penalize the model due to a few extremely outlying observations (see section 4.7).

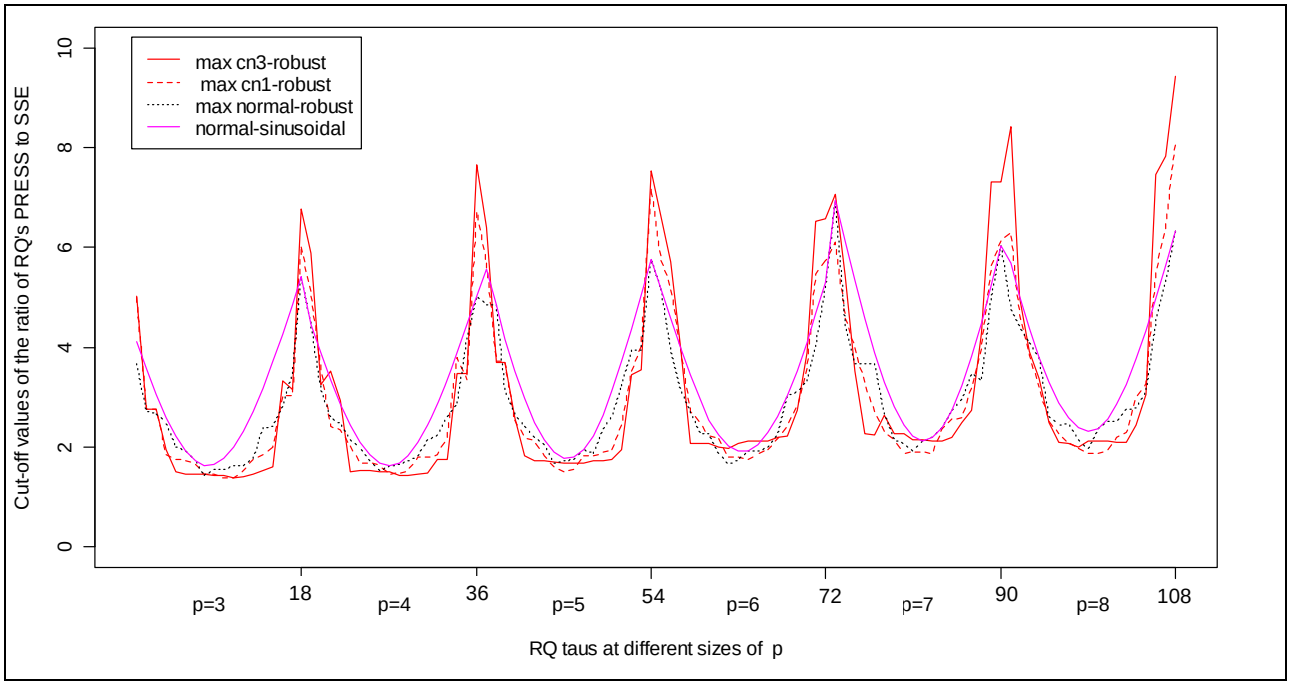


Figure 5.8: Comparison of the sinusoidal model cut-off values and the robust loss function cut-off values at D1

In the following subsection we give the conclusions on “prediction”.

5.9.1 Conclusions on the statistic $PRESS_J$

Although RQs are fairly robust to response outliers we determined the prediction picture in order to get a complete view point. The RQ (ES) predicted residual sum of squares, $PRESS_J$ s exhibited a consistent pattern in the sense that larger values of this statistic are always associated with extreme τ levels and prediction becomes poorer as the underlying error distribution becomes heavier despite the design matrix used. This is expected since all the error distributions used are symmetrical heavy tailed, implying outliers. However, there are minimal differences in RQ prediction between type A leverage points (using D2 and D4) and

absence of leverage points (D1 matrix), *i.e.*, the D1 matrix resulted in slightly smaller $PRESS_J$ s than those obtained using D2 and D4 (see figure 5.6). However, type B leverage points (using D3 and D5) resulted in slightly larger $PRESS_J$ s under the given underlying error distribution (see figure 5.7)

However, the degree of poor prediction as a result of the presence of leverage points alone is generally less compared to that which results from the presence of outliers. This could be attributed to the fact that RQs have a high affinity for leverage points but a high exclusion rate of outliers and thus predicting them (outliers) poorly. Furthermore, the degree of poor prediction as a result of design matrices depends on the type of leverage points present in them.

Type A leverage points (D2 and D4) tend to result in slightly poor prediction (figure 5.6; only a few times above the cut-off curve) while type B leverage points (D3 and D5) results in much poorer prediction (figure 5.7; more often above the cut-off curve than in the D2 and D4 case). As a consequence, the combined effect of both leverage points and outliers is that prediction is poorest at design matrices D3 and D5 (type B outliers) under the heaviest error distribution (CN3).

In the previous section and this section we have focused on RQ leverage and "prediction" respectively. However, it is important to see whether RQs multiple leverage and/or outliers are influential. In the next subsection we focus on the RQ covariance ratio, CVR_J as an influence measure.

5.10 Determining the threshold (cut-off) values for the RQ influence statistic, CVR_J, using simulation studies

We have treated the leverage and outlier cases separately earlier on in sections 5.8 and 5.9 respectively. However, multiple leverage and/or multiple outliers may/may not be influential jointly. Therefore it is important to look at both their individual and combined influential cases in the RQ situation. We use the covariance ratio, CVR_J for this purpose because of the attractiveness of its determinantal relationship with the ERW, ω_J (and T_J).

Using the two cut-off values, $\tilde{c}$ (determined in the case of T_J (ω_J)) and $\tilde{a}_J$ (determined in the case of $PRESS_J$), we deduce that reasonable cut-off values to identify RQ influential

points for the CVR_J's to be $\left(\frac{\tilde{c}\tilde{a}_J}{K}\right)^p \frac{K}{\tilde{c}}$ (see (4.8.1) in section 4.8). In figure 5.9 we give the plots of the τ's (and p's) against the quantiles of the 200 simulated values of the CVR_J's, $\underset{1\le i\le 200}{quant}(CVR_{Ji})$, that are greater than the cut-off values, $\left(\frac{\tilde{c}\tilde{a}_J}{K}\right)^p \frac{K}{\tilde{c}}$, under the different leverage and outlier scenarios. Note that only the $\underset{1\le i\le 200}{quant}(CVR_{Ji})$'s that are flagged appear in the plots.

At D1 under the normal distribution, the entire solution to LP problem (3.2.1), $B(\tau)$ yielded 18 τ levels at the given sizes of p's, *i.e.*, $\{3,4,...,8\}$ that resulted in RQs that correspond to unique ESs. Recall that the RQ/ES only changes at the next τ level. However, for the other remaining scenarios we fix the 18 τ levels at equally spaced intervals at the given p's to facilitate easier comparisons. The horizontal axis consists of the maximum, 95%, 90%, 75%, 50%, 25%, 10%, 5% and the minimum quantiles of the 200 simulated CVR_J values. The rationale here is to find the quantile levels to which the cut-off values can flag. For instance, if only the maximum values are flagged we expect to see the red circles in the plots; if at least the 95% quantiles are flagged then we expect to see both the red circles and the green circles, etc, as given by the legend in figure 5.9 below.

Therefore the picture depicted by the plots can be viewed as that of the quantiles of the CVR_J's that are being flagged using these cut-off values. The plot at D1 under the Normal distribution is the "benchmark" (ideal case) since D1 is orthogonal and the underlying distribution is the Gaussian which imply absence of leverage points and outliers respectively. Under this scenario we ideally expect no flagging of RQs except for the few odd ones due to chance. At the other design matrices (D2 to D5) under the Normal distribution, the flagging of RQs as influential cases is viewed as due to leverage points only (in the absence of outliers). On the other hand, the plots based on D1 under CN1 and CN3 are viewed as showing the flagging of RQs due to the underlying heavy tailed error distribution (outliers) in the absence of leverage points. We expect to flag more RQs due to leverage points than due to outliers since RQs are susceptible to leverage points but fairly robust (see section 5.2). However, the remaining scenarios which include both leverage points and outliers can be a paradox since on one hand RQs have a high affinity for leverage points while, on the other hand, they have a high repulsion (exclusion) of outliers. The resulting influential picture could be viewed as the net result of the trade off between these two antagonistic forces. We give first the set of plots (at D1) in figure 5.9 to show the effect of the underlying distribution.

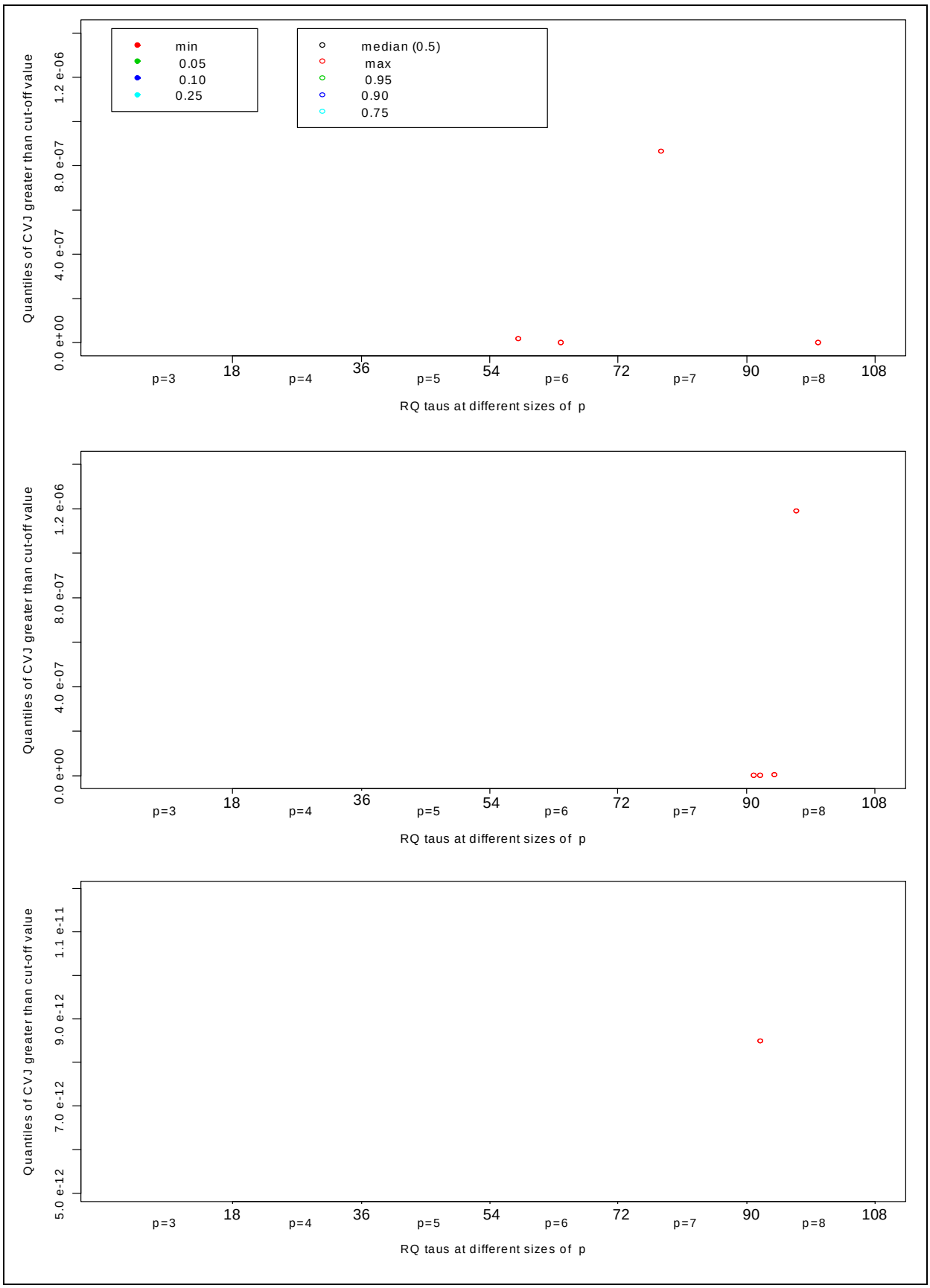


Figure 5.9: RQs flagged using the influence measure CVR_J at D1 under different error distributions. The first panel is under the Normal distribution, the second panel is under the CN1 and the third panel is under the CN3 distribution

At this design matrix (D1) we flag 4 RQs corresponding to $\max_{1\le i\le 200}(CVR_{Ji})$ under the normal and CN1 distributions. Under CN3 distribution we only flag one RQ corresponding to $\max_{1\le i\le 200}(CVR_{Ji})$, due to the RQs' robustness to outliers. To gain insight into the effectiveness of $\left(\frac{\tilde{c}\tilde{a}_J}{K}\right)^p \frac{K}{\tilde{c}}$ based on leverage as cut-off values, we analyse the following plots at different design matrices under the Normal distribution. The design matrices D2 and D4 are similar (they only differ in the severity of leverage) and likewise D3 and D5 are similar. As a consequence, throughout we give the plots of D2 and D4 together and also those of D3 and D5 for purposes of comparisons. We firstly give those of D2 and D4 under the Normal distribution in figure 5.10 below. Throughout the colour of the legend is as in figure 5.9.

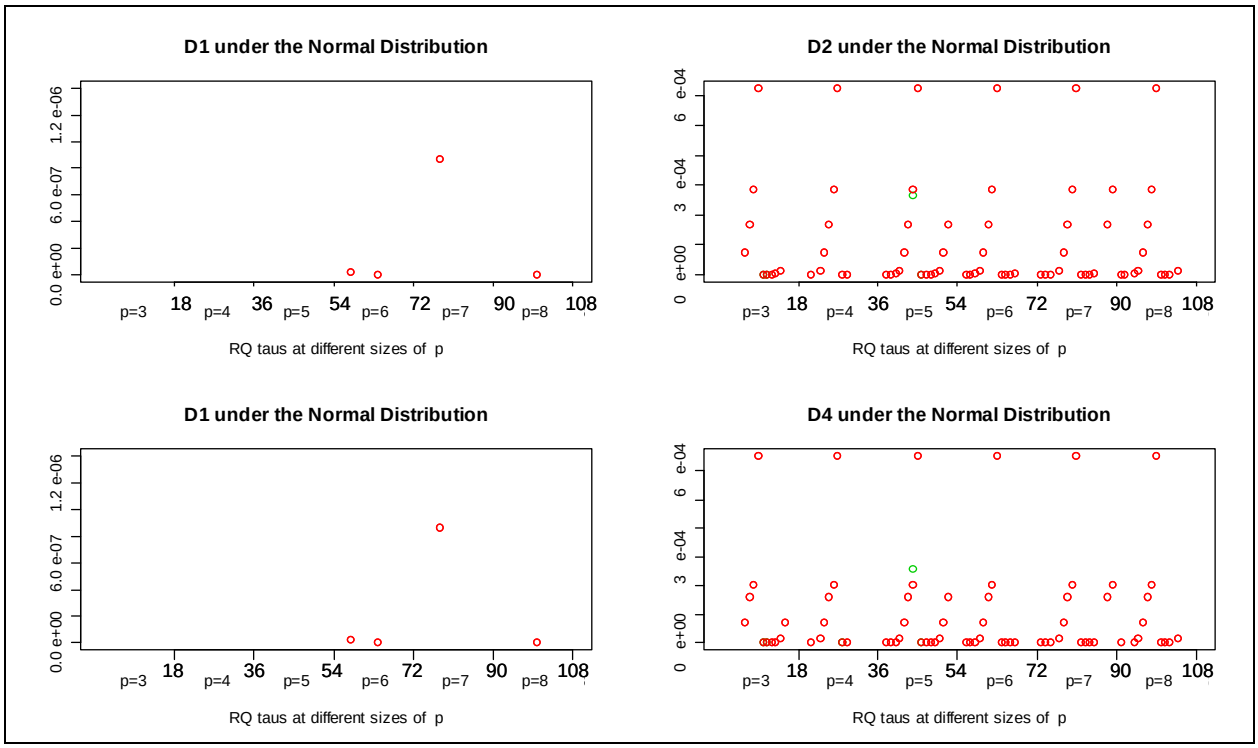


Figure 5.10: RQs flagged using the influence measure CVR_J at D2 and D4 under the Normal distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

Similar to the leverage aspect, the similarity between D2 and D4 is evident here. It is encouraging to see the somewhat uniform flagging of the RQs corresponding to $\max_{1\le i\le 200}(CVR_J)$ across all RQ levels and $p's$. This shows that our cut-off values for CVR_J are effective at all RQ levels regardless of the fact that extreme RQ levels have large $PRESS_J$'s.

Another leverage scenario is exhibited by D3 and D5. The following plots in figure 5.11 show us what the picture is in this case.

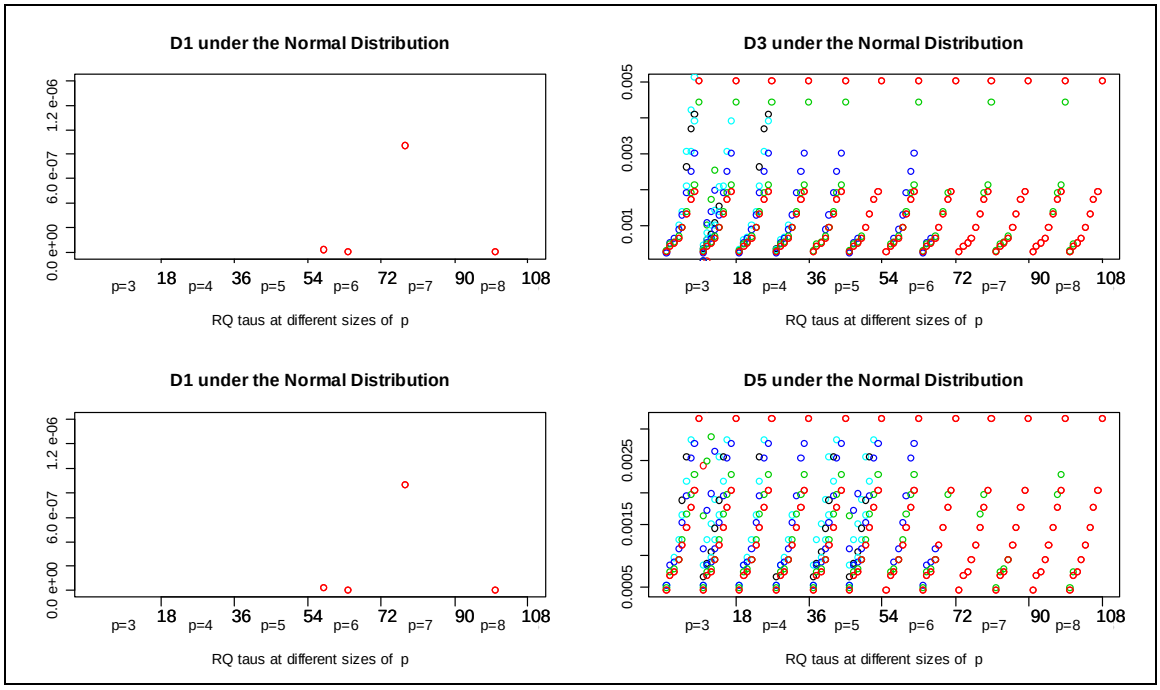


Figure 5.11: RQs flagged using the influence measure CVR_J at D3 and D5 under the Normal distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

Again, due to the similarity between D3 and D5, they expose a similar picture. In both cases at least the 75% quantile of the CVR_J's are flagged at $p \leq 6$, at least the 95% quantiles at $p \geq 7$. More quantiles (including lower ones) of the CVR_J's are flagged here as compared to D2 and D4.

Remark: Recall that, previously there were also higher proportions of flagging RQ cases using the statistic T_J (in both the artificial data and leverage simulation) for lower values of p. This is a desirable outcome as this outcome clearly exhibits the leverage picture (since RQs are susceptible to leverage points but robust to outliers).

The remaining scenarios include both leverage and error distribution influences. It is important to see what the picture is for these scenarios. We start our analyses of these scenarios with D2 and D4 under the CN1 distribution given in the following plots in figure

5.12. It is only the maximum quantiles of the CVR_J's that are generally flagged at the D2 and D4 . There is a clear reduction in the number of RQs that are flagged as we move from the Normal distribution to CN1 distribution at D2 and D4.

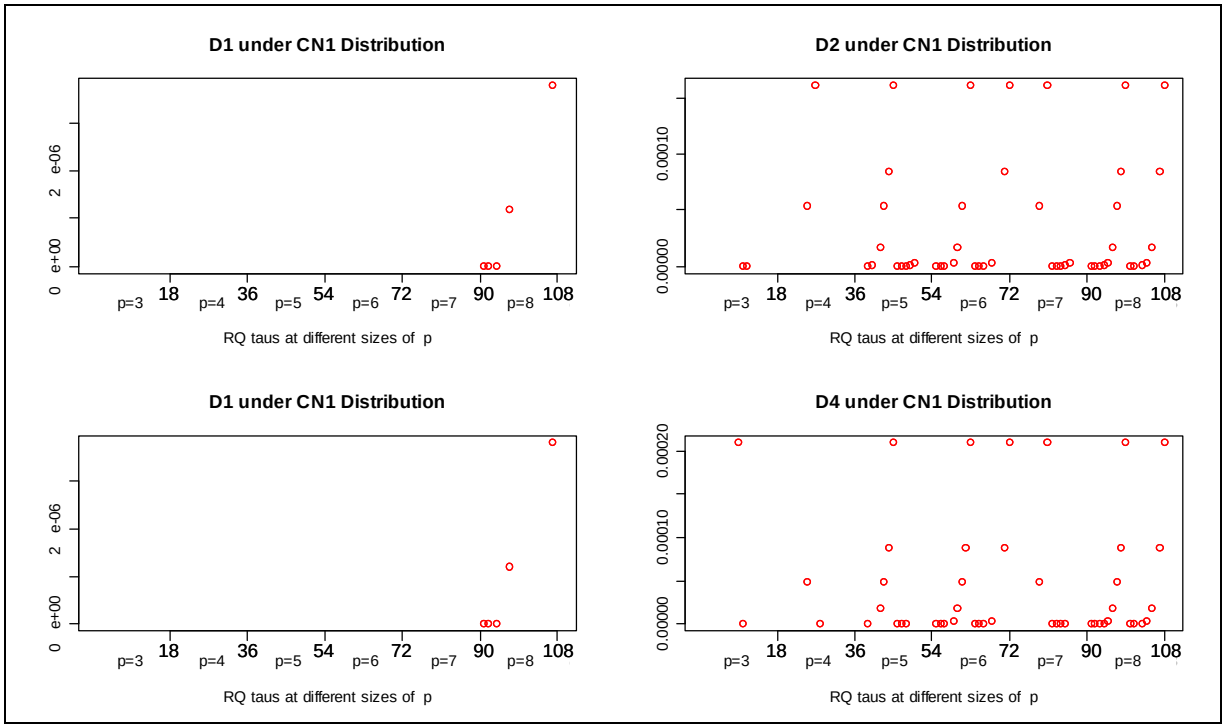


Figure 5.12: RQs flagged using the influence measure CVR_J at D2 and D4 under the CN1 distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

As noted before, this is due to the antagonistic nature between leverage and outliers in the RQ situation, *i.e.*, on one hand RQs have a great affinity for leverage points while on the other they repel (exclude) outliers. The picture here is the net result of the trade-off between these two opposing phenomena.

Next we give the diagrams for D3 and D5 under the CN1 distribution in figure 5.13 below.

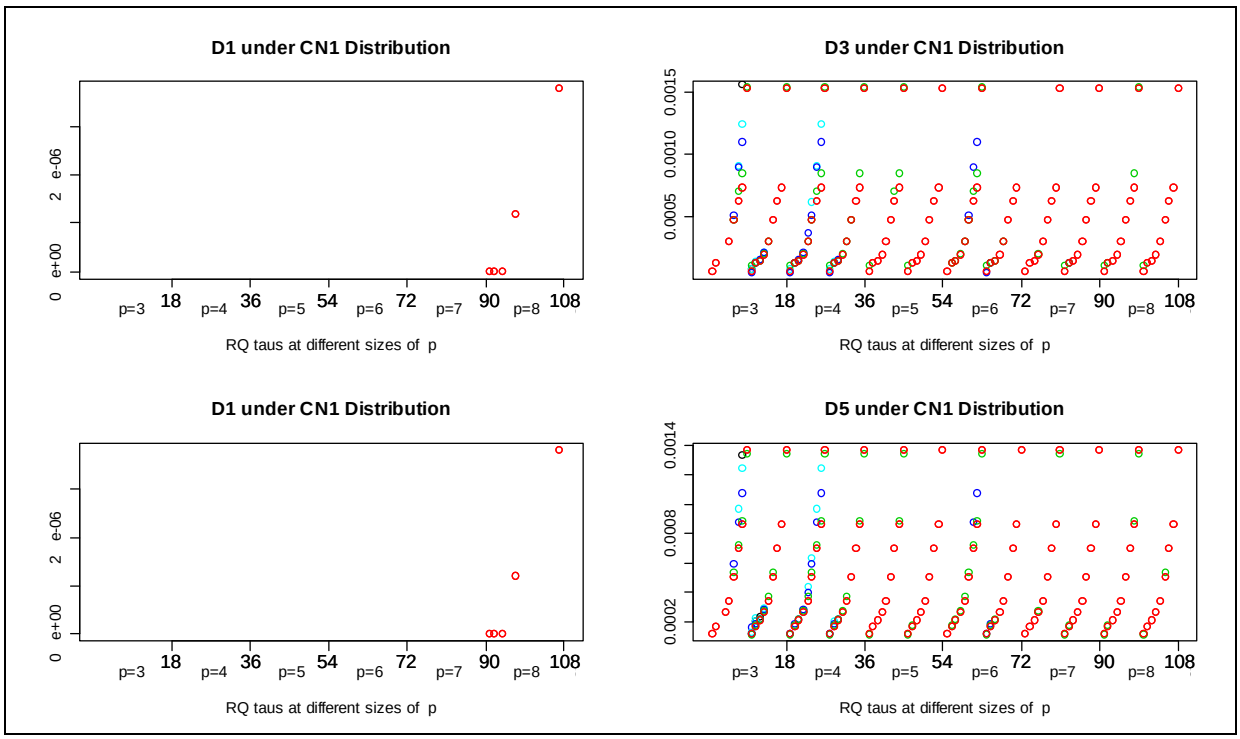


Figure 5.13: RQs flagged using the influence measure CVR_J at D3 and D5 under the CN1 distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

The diagrams given here look like those for D3 and D5 under the normal distribution at first glance (figure 5.11), but differs to some extent. Actually, the number of RQs being flagged here is less than the number being flagged under the Normal distribution at the same design matrices, again confirming the antagonistic nature between leverage and outliers for RQs.

Lastly, we consider the different leverage scenarios under the heaviest error distribution CN3. We give the diagrams exposed at D2 and D4 in figure 5.14 below.

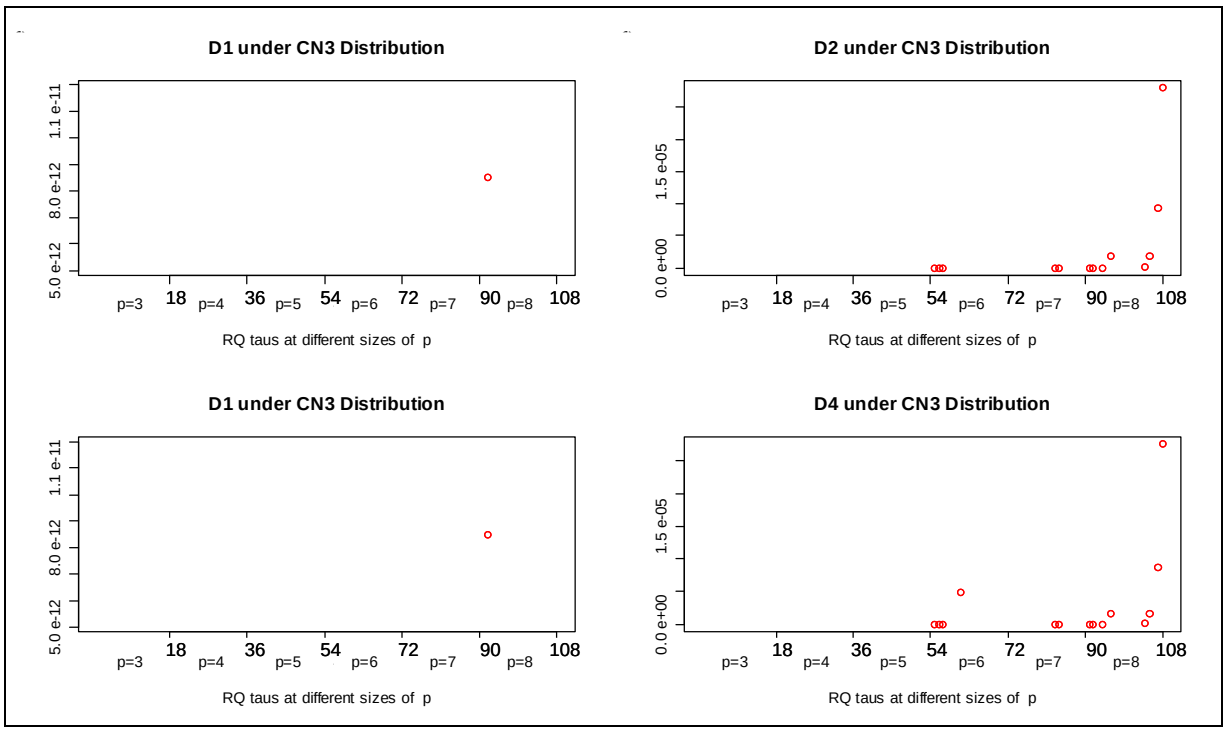


Figure 5.14: RQs flagged using the influence measure CVR_J at D2 and D4 under the CN3 distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

It is clear that the quantiles of CVR_J flagged here are fewer than those flagged under the normal distribution and under CN1 at D2 and D4. However, in all these cases only the maximum quantiles of the CVR_J's are generally flagged at $p \geq 5$.

We give the diagrams exposed at D3 and D5 in the following plots in figure 5.15.

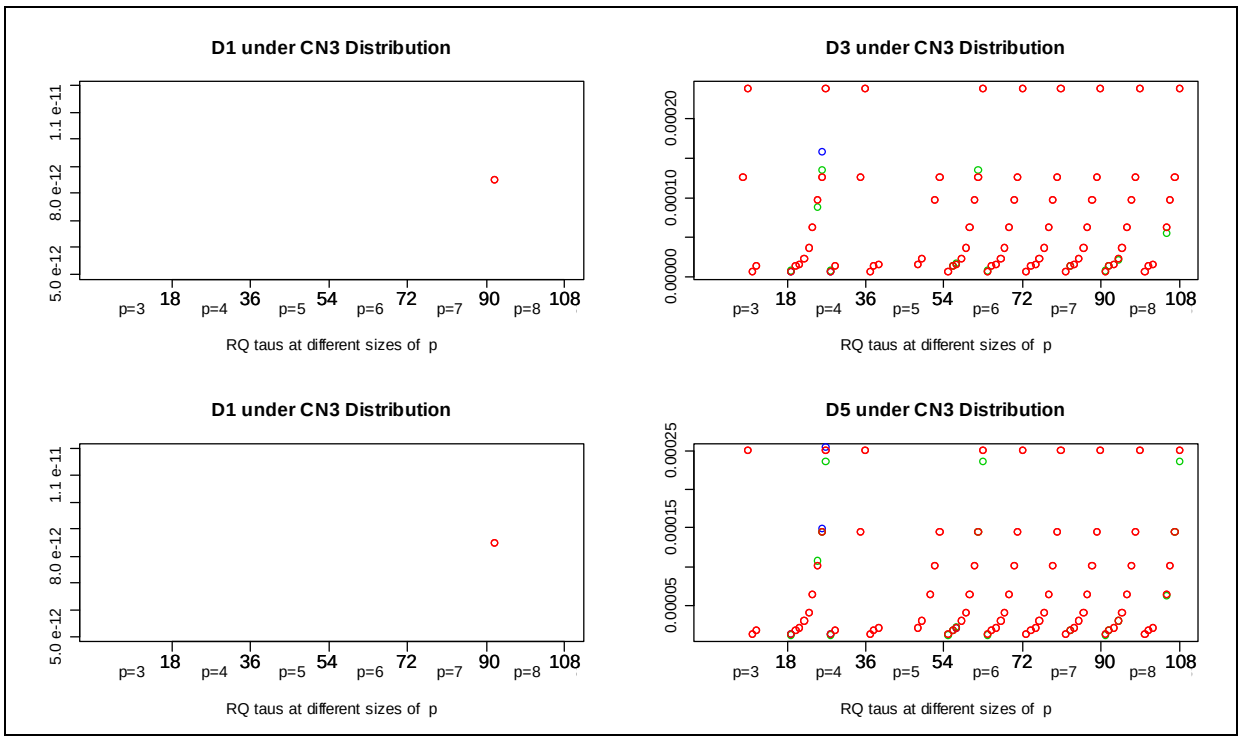


Figure 5.15: RQs flagged using the influence measure CVR_J at D3 and D5 under the CN3 distribution showing D1 as the control. The Y axis represents the quantiles of the CVR_J values

In figure 5.15 at least 95% of the quantiles of the CVR_J's are generally flagged at D3 and D5 (see figure 5.9 for the colour legend). In most instances it is at the maximum quantile where flagging takes place.

Again, it is clear that the number of quantile values of the CVR_J's flagged here are fewer than those flagged under the Normal distribution and under CN1 at D3 and D5.

In summary, we find that on the one hand outliers on their own are not that influential in the RQ situation while on the other hand leverage points on their own are highly influential. This is expected since RQs have influence functions that are bounded in the Y space but unbounded in the X space and thus are susceptible to X outliers.

5.10.1 Conclusions on the statistic CVR_J

Leverage points and/or outliers can have undesirable effects on prediction or on the fit. In order to identify RQ influential cases, we suggested the RQ covariance ratio, CVR_J because of the attractiveness of its determinantal relationship with the ERW, ω_J (and T_J). Furthermore, this statistic is made up of the leverage component ω_J and the prediction

component $PRESS_J$.

In the absence of leverage points (at D1), under the CN1 and CN3, the flagging of RQs is attributed to outliers, while in the absence of outliers (under the Normal distribution), at D2 to D5, the flagging of RQs is attributed to leverage points. As expected, more RQs were flagged due to leverage points than due to outliers since RQs are susceptible to leverage points but fairly robust to outliers. While flagging of influential RQs due to outliers (only) depends on the heaviness of the underlying error distribution, flagging of RQs influential cases due to leverage (only) depends mainly on the nature (kind) of leverage rather than the degree of leverage. Actually, more RQs influential sets are flagged due to type B leverage points (D3 and D5) than type A leverage points (D2 and D4). This may be attributed to the presence of collinearity that results in a slightly smaller T_J statistic value. Furthermore, at D3 and D5 under the normal distribution, the flagging of RQs due to the CVR_J cut-off values is similar to that due the T_J cut-off values. Thus a considerable proportion of maximum values of the CVR_J's are flagged at D2 and D4 while at D3 and D5 the quantile values of the CVR_J's are massively flagged as from the 75% quantile especially for lower p values under the Normal distribution. However, the flagging of influential points is drastically reduced at designs that contain leverage under heavy tailed error distributions. Again, this is attributable to the two antagonistic forces between leverage and outliers in the RQ scenario, *viz.*, the RQs affinity for leverage points and their exclusion of outliers.

This picture of flagging more RQs due to leverage points than due to outliers is welcome since RQs are more susceptible to leverage points due the unbounded nature their influence functions in the X space, unlike that in the Y space. At D2 and D4 the number of CVR_J simulation quantiles flagged decreased as one moves from the normal distribution to CN1 and CN3. This picture is somewhat the same for D3 and D5 although a larger amount of the quantile values of the CVR_J's are still flagged at CN3 than in the previous case. The reasons for this unexpected phenomenon may be due the interplay of the influence on the observations to be included in the RQ between leverage points and outliers. In the artificial data set we have seen that at D1 a change in the distribution resulted in a change in the observations included in the RQ set due to the exclusion of outlier observations by RQs. Also, a change in the design matrix under the Normal distribution also resulted in the change in the observations included in the RQ set due to the inclusion of high leverage. However, the presence of both a design matrix with leverage points and outliers also resulted in a change in the observations included in the RQ set. However, the effect of increasing leverage seemed to be stronger than

the effect of outlyingness (heavier distributions) with regards to determining the observations included in the ESs corresponding to RQs. Naturally, a change in the observations included in the RQ set will result in a change in the statistic CVR_J and hence the subsequent flagging pattern.

In the next section we give overall discussions, conclusions and prospects for further work for all the statistics involved here.

5.11 Overall discussions and conclusions of Chapter 5

The RQs multiple case diagnostics presented here can be classified into 3 main classes, *viz.*, leverage, outlier and influential aspects. As a consequence we presented 3 main diagnostic measures, *viz.*, RQs predictive leverage T_J, the RQ (ES) predicted residual sum of squares, $PRESS_J$ and the RQ (ES) covariance ratio, CVR_J as measures of these 3 aspects respectively.

The leverage aspect is more central to this research since it is well known in the literature that although RQs are fairly robust to outliers, they are susceptible to leverage points (see section 5.2). Influential leverage points have 2 forms, *viz.*, type A observations (D2 and D4) and type B observations (D3 and D5) (see also the discussion in section 4.6). Type A observations induce collinearity while type B observations obscure it. Consequently, type A observations result in the usual harmful effects of collinearity to various regression statistics while type B ones pose a danger to model selection procedures (see section 4.2). In the case of type A observations (D2 and D4), we expected that the ERW will be small due to the inducement of collinearity. However, the ERW was only marginally smaller at D2 and D4. The reason for this outcome could be due to the fact that despite the presence of collinearity at these design matrices as evidenced by the collinearity factor, $\rho_J \to 0$, the variability factor, γ_J becomes large and thus off-setting ρ_J (see section 5.4). It will be interesting to investigate the effect of collinearity from other sources on the ERW and RQ prediction. Surprisingly, the effect of type B observations on RQ prediction was slightly worse compared to type A observations. However, prediction generally gets poorer as the error distribution becomes heavier despite the design matrix used, although it is slightly worse under both leverage and outliers compared to outliers only.

The influential case is the most interesting of the three aspects. The two leverage scenarios

show that type A observations (D2 and D4) are less influential than type B observations (D3 and D5) despite the underlying error distribution. However, outliers on their own (in the absence of leverage points) are barely influential as expected due to the robustness of RQs. It is interesting to see that the presence of both leverage points and outliers resulted in a less influential situation compared to leverage points on their own (in the absence of outliers). Actually, the degree of influence decreases at the given leverage scenarios as the underlying error distribution becomes heavier. Therefore the resulting influence picture is the net result between these two antagonistic forces. Actually, the presence of outliers reduces the influence of leverage points.

Although there exists extensive research in the area of robust model selection, model selection in the RQ arena in particular has not been studied exhaustively, especially using the ES approach. In view of the collinearity induced by type A leverage points (D2 and D4) one class of appealing procedures that could be useful are the shrinkage techniques (see section 4.9) since they are less greedy than automated model selection techniques (see *e.g.*, Tibshirani, 1996). The lasso shrinkage technique (see subsection 4.9.1) is particularly appealing as it ties in nicely with the linear programming structure of RQs. Also, variants of automated model selection techniques based on R^2 could be developed since SSE is a weighted sum of the $PRESS_J$'s. Some of these aspects are discussed further in Chapter 7.

In the next chapter we apply the diagnostics discussed thus far to some real and well known data sets from the literature.

CHAPTER 6 APPLICATIONS

6.1 Introduction

In this chapter we apply the RQs (multiple) case diagnostics derived in Chapter 4 to some well known data sets in the literature. These are the Gunst and Mason data set (see Gunst and Mason, 1980), the Hocking data set (see Hocking and Pendleton, 1983) and the Hald data set (Montgomery and Peck, 1982), contaminated with type A leverage points, type B leverage points and collinearity, respectively. These data sets are given in appendix C. We use the cut-off values determined in the simulation study in Chapter 5.

Since leverage points also induce collinearities individually or jointly, multiple leverage diagnostics should be used together with single case diagnostics. As a consequence, we give some single case regression diagnostics in order to understand what the picture translates to when we move from a single case scenario to the RQ (multiple) case scenario. These applications will provide a better judgment on the performance of the diagnostics and their respective cut-off values.

6.2 Gunst and Mason data set

This data set provides a good example of how X outliers (leverage points) could induce collinearities in data sets. Actually the collinearity here is mainly induced by the two observations, *viz.*, 17 (Hong Kong) and 39 (Singapore) which are in "close proximity" to each other. The solution for collinearity influential points is not that simple since leverage points induced collinearities require different remedies from the "ordinary" ones proposed in the literature. For instance, biased estimators like ridge regression or robust estimators like M-estimators fail in this situation. Thus it is very important to always first identify the source of collinearities before one proposes a remedy.

The data set consists of the gross national product (GNP) of 49 countries along with six additional socio-economic predictors: an infant death rate (INFD), a physician/population ratio (PHYS), a population density (DENS), density as a function of agricultural land area (AGDS), a literacy measure (LIT), and an index of higher education (HIED). Residual analysis suggests that the correctly specified model is the regression of the natural logarithm of GNP on the above mentioned six socio-economic predictors.

We first give the single case diagnostics results based on the regression of the natural logarithm of GNP on the six socio-economic predictors model before we move on to the RQ

(multiple) case scenario. Table 6.1 below gives the number of occurrences (frequencies) of each observation in all the 43 specific ESs corresponding to RQs as well as some single case leverage, outlier and influential diagnostics (see sections 4.3, 4.7 and 4.8, respectively).

Table 6.1: Some single case regression diagnostics for the Gunst and Mason data set

	frequency	h_i	t_i	t_i^*	D_i	CVR_i	$\lvert CVR_i - 1 \rvert$
1. Australia	3	0.0641	0.4769	0.4725	0.0022	0.9535	0.0465
2. Austria	3	0.0372	-0.3632	-0.3594	0.0007	0.9832	0.0168
3. Barbados	9	0.2376	-1.9536	-2.0243	0.1699	0.7101	0.2899
4. Belgium	0	0.0433	1.2022	1.2088	0.0093	0.9463	0.0537
5. Brit. Guiana	7	0.1108	-0.9169	-0.9151	0.0150	0.8926	0.1074
6. Bulgaria	2	0.0408	-1.1057	-1.1087	0.0074	0.9540	0.0460
7. Canada	0	0.0416	1.9420	2.0111	0.0234	0.8936	0.1064
8. Chile	2	0.2107	1.0545	1.0560	0.0424	0.7871	0.2129
9. Costa Rica	12	0.0527	-0.2200	-0.2175	0.0004	0.9693	0.0307
10. Cyprus	15	0.1647	0.7057	0.7014	0.0140	0.8456	0.1544
11. Czechoslovakia	2	0.0394	-0.1717	-0.1697	0.0002	0.9833	0.0167
12. Denmark	0	0.0376	0.4855	0.4811	0.0013	0.9803	0.0197
13. El Salvador	7	0.1523	0.3042	0.3009	0.0024	0.8664	0.1336
14. Finland	3	0.0602	-0.3091	-0.3057	0.0009	0.9606	0.0394
15. France	4	0.0384	0.1812	0.1791	0.0002	0.9843	0.0157
16. Guatemala	5	0.2226	0.4584	0.4540	0.0086	0.7924	0.2076
17. Hong Kong	19	0.5113	-0.1071	-0.1058	0.0017	0.5005	0.4995
18. Hungary	1	0.0571	-0.4588	-0.4545	0.0018	0.9611	0.0389
19. Iceland	3	0.0452	-0.9251	-0.9235	0.0058	0.9582	0.0418
20. India	18	0.5581	1.3246	1.3369	0.3165	0.4338	0.5662
21. Ireland	3	0.0437	-0.7321	-0.7280	0.0035	0.9671	0.0329
22. Italy	1	0.0392	-0.0804	-0.0794	0.0000	0.9841	0.0159
23. Jamaica	8	0.1548	-0.4369	-0.4326	0.0050	0.8619	0.1381
24. Japan	1	0.0487	-2.5956	-2.7987	0.0492	0.8182	0.1818
25. Luxembourg	1	0.0840	2.2380	2.3562	0.0656	0.8265	0.1735
26. Malaya	6	0.2808	0.8035	0.8001	0.0360	0.7254	0.2746
27. Malta	14	0.6880	1.4826	1.5048	0.6925	0.3029	0.6971
28. Mauritius	3	0.1205	0.0210	0.0207	0.0000	0.9009	0.0991
29. Mexico	7	0.1506	0.2180	0.2155	0.0012	0.8691	0.1309
30. Netherlands	6	0.1058	-0.6892	-0.6848	0.0080	0.9057	0.0943
31. New Zealand	1	0.0595	0.5603	0.5557	0.0028	0.9562	0.0438
32. Nicaragua	13	0.1775	-0.5855	-0.5809	0.0106	0.8357	0.1643
33. Norway	1	0.0850	0.9843	0.9840	0.0129	0.9157	0.0843
34. Panama	8	0.0520	-0.4755	-0.4711	0.0018	0.9659	0.0341
35. Poland	3	0.0800	-0.0986	-0.0974	0.0001	0.9422	0.0578
36. Portugal	2	0.1247	-0.3880	-0.3840	0.0031	0.8934	0.1066
37. Puerto Rico	9	0.1803	-0.8948	-0.8926	0.0252	0.8237	0.1763
38. Romania	10	0.0835	-0.3638	-0.3600	0.0017	0.9359	0.0641
39. Singapore	36	0.6321	0.5658	0.5612	0.0786	0.3741	0.6259
40. Spain	5	0.0418	-1.4721	-1.4936	0.0135	0.9309	0.0691
41. Sweden	1	0.0589	1.2093	1.2162	0.0131	0.9305	0.0695
42. Switzerland	1	0.0429	1.4531	1.4732	0.0135	0.9312	0.0688
43. Taiwan	4	0.1783	-2.2768	-2.4027	0.1607	0.7379	0.2621
44. Taiwan	1	0.0514	0.2315	0.2288	0.0004	0.9705	0.0295
45. United Kingdom	1	0.0453	1.0045	1.0047	0.0068	0.9545	0.0455
46. United States	23	0.4897	0.8072	0.8037	0.0893	0.5146	0.4854
47. USSR	4	0.0400	-0.5698	-0.5652	0.0019	0.9758	0.0242
48. West Germany	1	0.0391	0.4540	0.4497	0.0012	0.9795	0.0205
49. Yugoslavia	12	0.0963	-0.7297	-0.7256	0.0081	0.9140	0.0860

Values of these statistics greater than the cut-off value ($|CVR_i - 1| > 0.4286$) are given in violet colour. Malta (27), Singapore (39), India (20), Hong Kong (17) and the United States (46) are classified as leverage points since they have leverage values $h_i > \frac{2p}{n}$ (=0.286).

Clearly, the situation depicted here is that RQs have a high affinity for leverage points as Malta (27), Singapore (39), India (20), Hong Kong (17) and the United States (46) have high frequencies. Note that although Malta (27) has the highest leverage value and is therefore expected to have the highest frequency, it has the lowest one.

Also, these observations (17, 20, 27, 39 and 46) are influential points as determined by the covariance ratio, CVR_i (with cut-off value, $|CVR_i - 1| > \frac{3p}{n}$ =0.4286). The values of the other remaining outlier and influential measures (h_i, t_i and t_i^*) that are greater than their cut-off values, are shown in violet colour.

It is important to see what this translates to at the RQ (multiple case) level. We begin our investigation with the RQ leverage aspect. In the following table 6.2 we give the statistics that are involved in the RQ leverage, *viz.*, the RQ (ES) predictive weighted leverage, T_J (see 4.5.1) , the ERW, ω_J (see 4.2.2), the residual freedom, R_{iJ} (see subsection 2.4.1) as well as the observations that comprise each ES corresponding to a particular RQ.

Table 6.2: RQs case leverage picture and the observations contained in the ESs corresponding to the RQs for the Gunst and Mason data set

τ	T_J	ω_J	ES						
0.050	6.12E-08	2.12E-09	3	5	20	24	32	40	43
0.054	2.49E-07	1.29E-09	3	5	20	37	40	43	49
0.057	1.06E-06	1.34E-07	3	5	20	39	40	43	49
0.060	7.85E-07	2.31E-07	3	5	37	39	40	43	49
0.106	1.66E-06	2.78E-07	3	5	20	37	39	40	49
0.107	5.13E-06	2.69E-06	3	5	20	32	37	39	49
0.140	8.71E-07	3.42E-07	3	5	6	32	37	39	49
0.147	4.39E-07	1.35E-07	3	6	32	34	37	39	49
0.175	6.57E-07	2.20E-07	3	19	32	34	37	39	49
0.178	3.48E-07	2.77E-08	19	23	32	34	37	39	49
0.196	2.18E-06	1.30E-06	17	19	23	32	37	39	49
0.205	3.75E-07	8.17E-08	17	21	23	32	34	39	49
0.221	4.60E-07	1.84E-07	17	21	30	32	34	39	49
0.244	2.19E-06	9.39E-07	17	20	21	30	32	34	39
0.260	1.76E-06	6.38E-07	17	20	30	32	34	39	47
0.270	3.91E-07	4.75E-08	17	20	30	32	34	38	47
0.285	3.13E-07	3.77E-08	9	17	20	30	32	38	47
0.323	4.62E-08	7.59E-09	9	17	23	30	36	38	47
0.342	1.29E-07	2.16E-08	2	9	17	23	36	38	39
0.366	1.58E-07	2.67E-08	2	9	17	23	29	38	39
0.373	1.60E-06	4.78E-07	2	9	17	23	29	39	46
0.382	1.55E-06	4.52E-07	9	14	17	23	29	39	46
0.449	5.33E-07	1.73E-07	9	14	17	18	28	29	46
0.469	6.25E-07	1.49E-07	9	14	17	28	29	38	46
0.496	7.21E-07	1.60E-07	9	11	17	28	38	39	46
0.519	1.12E-06	2.40E-07	9	11	13	17	38	39	46
0.532	2.25E-06	4.35E-07	13	17	22	29	38	39	46
0.567	2.79E-06	9.81E-07	13	15	17	29	38	39	46
0.580	2.18E-06	5.08E-07	9	10	13	15	17	39	46
0.580	6.72E-06	2.13E-06	9	10	13	15	27	39	46
0.588	1.03E-05	3.78E-06	10	13	15	27	35	39	46
0.629	9.53E-06	1.71E-06	10	13	27	35	39	44	46
0.635	1.54E-05	5.65E-06	1	10	16	27	35	39	46
0.674	7.43E-06	4.39E-07	1	10	16	27	39	46	48
0.698	8.89E-05	4.01E-05	1	10	16	20	27	39	46
0.726	8.93E-05	4.11E-05	10	16	20	27	31	39	46
0.748	1.32E-04	9.38E-05	8	10	16	20	27	39	46
0.772	1.57E-04	1.22E-04	8	10	20	26	27	39	46
0.814	1.38E-04	1.29E-04	10	20	26	27	33	39	46
0.845	1.21E-04	1.13E-04	10	20	26	27	39	45	46
0.875	1.26E-04	1.11E-04	10	20	26	27	39	41	46
0.937	1.21E-04	1.09E-04	10	20	26	27	39	42	46
0.957	1.42E-04	1.63E-04	10	20	25	26	27	39	46

Based on the statistic $T_J \geq 26pK^{-1}$=2.12E-06 (see section 5.8.1 for this cut-off value) we flag 3 τ levels below the middle one and fail to flag only one τ level above it while based on $\omega_J \geq 26K^{-1}$=3.03e-07 (see the last remark in section 5.8.1 for this cut-off value) the picture is more or less similar to that exhibited for T_J. It is important to see what the situation is for the reduced data set (without observation 17 and 39, which we will see below are collinearity influential points). We give the leverage picture of this reduced data set in table 6.3 below.

Table 6.3: RQs case leverage statistics and the observations contained in the ESs corresponding to the RQs for the Gunst and Mason data set with observations 17 and 39 deleted

τ	T_J	ω_J	ES						
0.042	5.04E-07	4.73E-08	3	5	23	31	36	38	41
0.099	2.15E-07	1.10E-07	3	5	6	18	23	31	36
0.147	6.75E-07	1.63E-07	3	5	18	23	31	36	47
0.158	4.60E-07	1.15E-06	3	5	23	31	35	36	47
0.160	2.32E-08	9.64E-07	3	5	20	23	31	35	45
0.161	6.88E-07	8.02E-06	3	5	23	31	35	44	45
0.184	6.11E-07	7.38E-07	3	23	31	35	44	45	47
0.190	3.31E-07	1.21E-05	3	20	23	35	44	45	47
0.215	6.87E-07	7.20E-06	3	20	23	33	35	44	47
0.226	2.53E-06	2.57E-08	20	23	26	33	35	44	47
0.228	5.30E-07	1.51E-07	20	26	29	33	35	44	47
0.239	4.17E-07	1.19E-07	26	29	33	35	44	45	47
0.244	2.98E-08	1.80E-06	20	29	33	35	44	45	47
0.249	6.17E-08	2.61E-06	20	29	33	35	41	45	47
0.258	2.20E-06	4.78E-06	13	26	29	35	41	45	47
0.276	4.21E-06	3.91E-07	13	22	26	29	35	41	47
0.334	3.92E-06	1.01E-05	13	17	22	26	29	41	47
0.335	1.22E-07	4.38E-05	13	14	17	22	35	41	47
0.339	2.54E-06	3.43E-07	13	14	22	26	35	41	47
0.345	1.32E-05	4.42E-09	13	14	22	26	35	41	44
0.370	5.24E-06	6.57E-10	2	13	14	22	26	41	44
0.415	6.72E-06	2.95E-09	9	13	14	22	26	41	44
0.435	1.69E-05	7.80E-06	14	16	22	26	37	41	44
0.442	1.22E-05	2.58E-06	11	22	26	28	37	41	44
0.503	1.95E-05	3.36E-06	11	16	26	28	37	41	44
0.526	1.83E-05	1.11E-05	15	16	26	28	37	41	44
0.564	1.02E-05	3.57E-08	15	26	27	28	34	41	44
0.588	1.81E-05	2.84E-07	16	21	26	28	34	41	44
0.614	2.41E-05	7.47E-09	1	16	26	28	34	41	44
0.647	1.28E-06	7.22E-07	1	16	26	28	42	44	46
0.678	3.84E-06	5.00E-06	1	16	19	26	42	44	46
0.709	3.09E-06	1.52E-05	12	16	19	26	42	44	46
0.719	1.10E-05	9.00E-06	10	16	19	26	30	44	46
0.721	1.92E-05	1.04E-05	10	16	19	26	30	32	44
0.753	6.57E-06	5.54E-10	8	16	19	26	30	32	44
0.779	2.61E-05	1.16E-06	8	10	16	19	26	32	44
0.809	3.16E-05	4.48E-07	8	10	19	25	26	32	44
0.813	2.57E-05	4.30E-06	10	19	25	26	32	43	44
0.848	2.24E-05	8.45E-06	4	10	19	25	26	43	44
0.880	2.30E-05	1.83E-07	4	10	19	25	26	39	44
0.913	2.19E-05	5.33E-07	4	10	19	25	26	40	44
0.951	8.73E-05	6.89E-06	4	19	24	25	26	41	44

There is no marked improvement in the leverage picture after deleting observations 17 and 39. Actually, there is a slight increase in the number of RQs being flagged here. This is due to the change in the correlation structure of the design matrix since observations 17 and 39 are type A observations which induce collinearity which in turn could result in small ρ_J's and hence ω_J and T_J. In the next subsection we focus on this collinearity aspect. The following table 6.4 below gives the RQ leverage view excluding all the leverage points, *viz.* Malta (27), Singapore (39), India (20), Hong Kong (17) and the United States (46).

Table 6.4: RQs case leverage statistics for the Gunst and Mason data set with observations 17, 20, 27, 39 and 46 deleted

τ	T_J	ω_J	τ	T_J	ω_J
0.0443	3.36E-04	2.85E-06	0.5487	3.26E-05	2.67E-08
0.0694	9.38E-05	4.86E-07	0.5720	2.60E-05	1.47E-07
0.148	9.39E-05	8.49E-07	0.5993	6.56E-05	2.29E-07
0.1482	4.46E-05	3.90E-07	0.6349	1.38E-05	7.36E-08
0.1744	6.17E-05	3.99E-07	0.6477	1.20E-05	1.02E-08
0.1919	1.30E-04	1.90E-06	0.6769	9.24E-05	1.53E-06
0.1996	3.33E-04	5.06E-06	0.6832	9.71E-05	4.01E-07
0.2508	2.87E-04	4.03E-06	0.6906	1.32E-04	1.88E-06
0.2551	3.02E-04	1.79E-06	0.7466	1.06E-04	1.01E-06
0.3319	1.13E-04	1.40E-06	0.7615	3.31E-05	1.87E-07
0.334	1.28E-04	2.87E-08	0.7636	6.80E-05	3.68E-07
0.3506	1.47E-05	8.34E-09	0.8032	1.82E-05	1.04E-07
0.3764	2.26E-04	2.69E-06	0.8099	6.67E-05	2.47E-07
0.4613	2.57E-04	3.71E-06	0.8250	9.37E-04	5.11E-06
0.4823	1.19E-04	1.13E-06	0.8300	1.22E-03	2.05E-05
0.4856	1.44E-04	3.14E-07	0.8427	1.67E-03	3.30E-05
0.4927	1.36E-04	1.13E-06	0.8645	1.48E-04	1.70E-07
0.4969	9.89E-05	7.13E-07	0.8712	1.87E-05	8.23E-08
0.4976	5.60E-05	1.92E-07	0.8818	1.47E-05	4.99E-08
0.5278	1.79E-05	5.81E-09	0.9082	4.00E-05	2.91E-07
0.5451	2.48E-04	1.50E-06	0.9088	2.93E-05	1.28E-08
0.5473	6.03E-04	6.81E-06	0.9131	2.81E-05	4.46E-08

Based on the statistic $T_J \geq 26pK^{-1}$=4.749e-06 we fail to flag any τ level and flag fewer τ levels based on $\omega_J \geq 26K^{-1}$=6.785e-07 compared to the data sets with leverage points (as shown in tables 6.2 and 6.3). It is desirable that no RQs and fewer RQs are flagged based on the statistic T_J and ω_J, respectively, in the absence of the leverage points. This is in line with the simulation results.

In the next subsection we focus on this collinearity aspect.

6.2.1 The collinearity and variability view

The Gunst and Mason data set has leverage points which induce collinearity in the design matrix. Such leverage points are called collinearity influential points also referred to as type A observations (see section 4.6) as shown in figure 4.1 (see section 4.6).

Although the determinant of the correlation matrix is more appealing in our situation, we also give various collinearity measures, since one measure may be able to detect the collinearity where another one fails.

The following table 6.5 gives the correlation matrix of the six socio-economic predictors.

Table 6.5: Correlation matrix

	Full Data set					
	INFD	PHYS	DENS	AGDS	LIT	HIED
INFD	1.0000	0.5686	-0.0904	-0.0903	-0.6277	-0.3142
PHYS	0.5686	1.0000	0.1217	0.1311	-0.7824	-0.3720
DENS	-0.0904	0.1217	1.0000	0.9724	-0.2563	-0.1007
AGDS	-0.0903	0.1311	0.9724	1.0000	-0.2490	-0.0712
LIT	-0.6277	-0.7824	-0.2563	-0.2490	1.0000	0.4174
HIED	-0.3142	-0.3720	-0.1007	-0.0712	0.4174	1.0000

The pairwise correlations show that the highest correlation is between DENS and ARGS, *i.e.*, $c_{34} = 0.9724$.

In the following table we give some collinearity diagnostics based on the eigenvalues and eigenvectors as well as the variance inflation factors (VIF).

Table 6.6: Some Collinearity Measures for the full data set

Eigenvector	Eigenvalues						VIF
(Predictors)	2.6992	1.9257	0.7443	0.4121	0.1920	0.0267	
INFD	0.4099	0.3555	-0.2597	0.7698	0.2134	0.0066	1.8950
PHYS	0.5082	0.1869	-0.2116	-0.5851	0.5644	0.0340	2.7064
DENS	0.2725	-0.6360	-0.0069	0.1262	0.0508	0.7090	19.0840
AGDS	0.2689	-0.6375	-0.0460	0.1114	0.1095	-0.7034	18.8324
LIT	-0.5519	-0.1084	0.1731	0.1839	0.7868	0.0275	3.4941
HIED	-0.3500	-0.1264	-0.9250	-0.0544	-0.0476	0.0252	1.2460
Condition	100.995	72.051	27.848	15.420	7.1857	1.0000	

The existence of a very small eigenvalue ($0.0267 \approx 0$) relative to the largest one (2.6992) and high condition numbers are indicative of collinearity. Furthermore, large elements in the eigenvector corresponding to the near-zero eigenvalue identify predictors that are involved in the multicollinearities. Hence, in this case DENS and AGDS are collinear. This is further confirmed by the large variance inflation factors (VIF).

To gain insight into the change in the correlation structure due to the deletion of observations 17 and 39, we present table 6.7 based on the reduced data set below.

Table 6.7: Some collinearity measures for the reduced data set

Eigenvector	Eigenvalues						VIF
(Predictors)	2.6300	1.8089	0.7288	0.4318	0.2204	0.1800	
INFD	-0.4892	0.1817	-0.2189	0.7736	0.2677	-0.0980	1.9395
PHYS	-0.5332	0.0957	-0.2160	-0.5664	0.2239	-0.5375	2.7525
DENS	-0.0901	-0.6896	-0.0788	0.2082	-0.5233	-0.4391	2.6882
AGDS	-0.0923	-0.6896	-0.1193	-0.0919	0.6007	0.3638	2.6774
LIT	0.5654	-0.0262	0.1854	0.1686	0.4908	-0.6132	3.4165
HIED	0.3743	0.0773	-0.9223	-0.0230	-0.0503	0.0163	1.2379
Condition	14.6070	10.0470	4.0480	2.3980	1.2240	1.0000	

There is a dramatic change in the collinearity measures after observations 17 and 39 are removed. This is confirmed by the absence of a near zero eigenvalue, a dramatic drop in condition numbers, absence of inordinately large elements in the eigenvector corresponding to the smallest eigenvalue and the absence of large VIF's.

We also present the variability and collinearity picture exposed by the statistics γ_J, ρ_J and $|\mathbf{C}_J|$ (see section 4.2.2 and equation 4.2.2). The collinearity picture for the full data set is shown in table 6.8 below.

Table 6.8: RQs case variability and collinearity picture and the observations contained in the ESs corresponding to the RQs for the Gunst and Mason data set

τ	γ_J	ρ_J	$\lvert\mathbf{C}_J\rvert$	ω_J	ES						
0.050	5.48E-08	2.71E-01	2.216E-03	2.12E-09	3	5	20	24	32	40	43
0.054	6.85E-08	1.32E-01	1.077E-03	1.29E-09	3	5	20	37	40	43	49
0.057	5.63E-05	1.67E-02	1.364E-04	1.34E-07	3	5	20	39	40	43	49
0.060	5.05E-06	3.20E-01	2.617E-03	2.31E-07	3	5	37	39	40	43	49
0.106	2.22E-04	8.76E-03	7.165E-05	2.78E-07	3	5	20	37	39	40	49
0.107	1.67E-04	1.13E-01	9.227E-04	2.69E-06	3	5	20	32	37	39	49
0.140	7.45E-06	3.21E-01	2.629E-03	3.42E-07	3	5	6	32	37	39	49
0.147	4.55E-06	2.07E-01	1.694E-03	1.35E-07	3	6	32	34	37	39	49
0.175	7.50E-06	2.06E-01	1.683E-03	2.20E-07	3	19	32	34	37	39	49
0.178	1.17E-05	1.65E-02	1.350E-04	2.77E-08	19	23	32	34	37	39	49
0.196	3.75E-05	2.42E-01	1.980E-03	1.30E-06	17	19	23	32	37	39	49
0.205	6.62E-06	8.65E-02	7.075E-04	8.17E-08	17	21	23	32	34	39	49
0.221	2.14E-05	6.01E-02	4.918E-04	1.84E-07	17	21	30	32	34	39	49
0.244	4.79E-04	1.37E-02	1.122E-04	9.39E-07	17	20	21	30	32	34	39
0.260	5.36E-04	8.33E-03	6.813E-05	6.38E-07	17	20	30	32	34	39	47
0.270	2.44E-04	1.37E-03	1.117E-05	4.75E-08	17	20	30	32	34	38	47
0.285	2.38E-04	1.11E-03	9.084E-06	3.77E-08	9	17	20	30	32	38	47
0.323	7.02E-06	7.57E-03	6.191E-05	7.59E-09	9	17	23	30	36	38	47
0.342	4.74E-06	3.18E-02	2.603E-04	2.16E-08	2	9	17	23	36	38	39
0.366	5.18E-06	3.61E-02	2.956E-04	2.67E-08	2	9	17	23	29	38	39
0.373	1.06E-04	3.16E-02	2.585E-04	4.78E-07	2	9	17	23	29	39	46
0.382	1.05E-04	3.00E-02	2.458E-04	4.52E-07	9	14	17	23	29	39	46
0.449	6.06E-05	1.99E-02	1.630E-04	1.73E-07	9	14	17	18	28	29	46
0.469	6.02E-05	1.74E-02	1.420E-04	1.49E-07	9	14	17	28	29	38	46
0.496	1.57E-04	7.17E-03	5.870E-05	1.60E-07	9	11	17	28	38	39	46
0.519	2.99E-04	5.62E-03	4.602E-05	2.40E-07	9	11	13	17	38	39	46
0.532	2.43E-04	1.25E-02	1.024E-04	4.35E-07	13	17	22	29	38	39	46
0.567	3.23E-04	2.13E-02	1.739E-04	9.81E-07	13	15	17	29	38	39	46
0.580	2.25E-04	1.58E-02	1.292E-04	5.08E-07	9	10	13	15	17	39	46
0.580	8.67E-05	1.72E-01	1.407E-03	2.13E-06	9	10	13	15	27	39	46
0.588	9.35E-05	2.83E-01	2.318E-03	3.78E-06	10	13	15	27	35	39	46
0.629	6.65E-05	1.80E-01	1.473E-03	1.71E-06	10	13	27	35	39	44	46
0.635	2.79E-04	1.42E-01	1.160E-03	5.65E-06	1	10	16	27	35	39	46
0.674	2.21E-04	1.39E-02	1.134E-04	4.39E-07	1	10	16	27	39	46	48
0.698	3.33E-03	8.43E-02	6.898E-04	4.01E-05	1	10	16	20	27	39	46
0.726	3.32E-03	8.66E-02	7.088E-04	4.11E-05	10	16	20	27	31	39	46
0.748	2.37E-03	2.77E-01	2.267E-03	9.38E-05	8	10	16	20	27	39	46
0.772	2.03E-03	4.19E-01	3.430E-03	1.22E-04	8	10	20	26	27	39	46
0.814	2.69E-03	3.36E-01	2.752E-03	1.29E-04	10	20	26	27	33	39	46
0.845	2.52E-03	3.13E-01	2.562E-03	1.13E-04	10	20	26	27	39	45	46
0.875	2.63E-03	2.96E-01	2.424E-03	1.11E-04	10	20	26	27	39	41	46
0.937	2.64E-03	2.89E-01	2.361E-03	1.09E-04	10	20	26	27	39	42	46
0.957	2.66E-03	4.29E-01	3.511E-03	1.63E-04	10	20	25	26	27	39	46

Table 6.9 below gives the variability and collinearity picture based on the reduced data sets.

Table 6.9: RQs case variability and collinearity picture and the observations contained in the ESs corresponding to RQs for the Gunst and Mason data set with observations 17 and 39 deleted

τ	γ_J	ρ_J	$\lvert\mathbf{C}_J\rvert$	ω_J	ES						
0.0425	2.25E-06	1.41E-01	8.39E-03	4.73E-08	3	5	23	31	36	38	41
0.0992	2.36E-06	7.30E-02	4.34E-03	1.10E-07	3	5	6	18	23	31	36
0.1467	4.05E-06	5.69E-01	3.38E-02	1.63E-07	3	5	18	23	31	36	47
0.1581	1.77E-06	1.35E-01	8.02E-03	1.15E-06	3	5	23	31	35	36	47
0.1596	1.64E-06	2.27E-03	1.35E-04	9.64E-07	3	5	20	23	31	35	45
0.1609	1.24E-05	5.96E-02	3.54E-03	8.02E-06	3	5	23	31	35	44	45
0.1842	9.38E-06	1.08E-01	6.41E-03	7.38E-07	3	23	31	35	44	45	47
0.1899	3.05E-06	9.73E-03	5.78E-04	1.21E-05	3	20	23	35	44	45	47
0.2154	4.93E-06	3.87E-01	2.30E-02	7.20E-06	3	20	23	33	35	44	47
0.2257	1.82E-05	4.29E-01	2.55E-02	2.57E-08	20	23	26	33	35	44	47
0.2283	1.71E-05	6.41E-02	3.81E-03	1.51E-07	20	26	29	33	35	44	47
0.2393	1.76E-05	4.55E-02	2.70E-03	1.19E-07	26	29	33	35	44	45	47
0.2444	4.34E-07	1.02E-02	6.04E-04	1.80E-06	20	29	33	35	44	45	47
0.2491	4.85E-07	1.03E-01	6.14E-03	2.61E-06	20	29	33	35	41	45	47
0.2582	3.04E-05	9.89E-02	5.88E-03	4.78E-06	13	26	29	35	41	45	47
0.2764	2.35E-05	3.28E-01	1.95E-02	3.91E-07	13	22	26	29	35	41	47
0.3341	2.96E-05	4.09E-01	2.43E-02	1.01E-05	13	17	22	26	29	41	47
0.3352	6.11E-07	3.24E-02	1.92E-03	4.38E-05	13	14	17	22	35	41	47
0.3390	9.93E-06	4.88E-01	2.90E-02	3.43E-07	13	14	22	26	35	41	47
0.3447	1.31E-04	2.21E-01	1.31E-02	4.42E-09	13	14	22	26	35	41	44
0.3697	1.06E-04	6.10E-02	3.62E-03	6.57E-10	2	13	14	22	26	41	44
0.4154	1.28E-04	1.37E-01	8.16E-03	2.95E-09	9	13	14	22	26	41	44
0.4354	3.25E-04	1.61E-01	9.59E-03	7.80E-06	14	16	22	26	37	41	44
0.4424	4.87E-05	6.90E-01	4.10E-02	2.58E-06	11	22	26	28	37	41	44
0.5027	3.41E-04	1.66E-01	9.88E-03	3.36E-06	11	16	26	28	37	41	44
0.5261	3.48E-04	1.55E-01	9.20E-03	1.11E-05	15	16	26	28	37	41	44
0.5636	7.08E-05	4.53E-01	2.69E-02	3.57E-08	15	26	27	28	34	41	44
0.5883	2.61E-04	6.64E-02	3.95E-03	2.84E-07	16	21	26	28	34	41	44
0.6136	4.48E-04	2.27E-01	1.35E-02	7.47E-09	1	16	26	28	34	41	44
0.6466	1.91E-04	6.41E-03	3.81E-04	7.22E-07	1	16	26	28	42	44	46
0.6779	2.37E-03	2.09E-03	1.24E-04	5.00E-06	1	16	19	26	42	44	46
0.7095	1.85E-03	1.42E-03	8.41E-05	1.52E-05	12	16	19	26	42	44	46
0.7188	2.48E-03	9.07E-03	5.39E-04	9.00E-06	10	16	19	26	30	44	46
0.7212	2.69E-03	2.24E-02	1.33E-03	1.04E-05	10	16	19	26	30	32	44
0.7533	2.71E-03	1.32E-03	7.85E-05	5.54E-10	8	16	19	26	30	32	44
0.7785	1.97E-03	4.11E-02	2.44E-03	1.16E-06	8	10	16	19	26	32	44
0.8093	1.64E-03	4.12E-02	2.45E-03	4.48E-07	8	10	19	25	26	32	44
0.8134	1.67E-03	4.44E-02	2.64E-03	4.30E-06	10	19	25	26	32	43	44
0.8484	1.65E-03	4.23E-02	2.51E-03	8.45E-06	4	10	19	25	26	43	44
0.8803	1.86E-03	2.48E-02	1.47E-03	1.83E-07	4	10	19	25	26	39	44
0.9128	1.75E-03	2.77E-02	1.65E-03	5.33E-07	4	10	19	25	26	40	44
0.9514	3.01E-03	9.77E-02	5.81E-03	6.89E-06	4	19	24	25	26	41	44

It is clear that there is a general increment in the sizes of the statistics, γ_J, ρ_J, $\lvert\mathbf{C}_J\rvert$ and hence ω_J, which shows that the collinearity which is a result of type A observations have been reduced. This is in line with the picture depicted by the other collinearity measures above.

In the next subsection we explore the "prediction" and the influential pictures of the ESs corresponding to RQs.

6.2.2 RQ case outlier and influential view

In order to have a complete picture of the RQs (multiple) case prediction and influence we use the prediction statistic, $PRESS_J$ (see subsection 2.4.1) and covariance ratio statistic, CVR_J (see 4.8.1) respectively. We compute these statistics for both the full data set and the reduced data sets. Figure 6.1 below gives the plots of these statistics based on the full a data set.

Poor prediction is only exhibited at the lowest few RQs. Using the cut-off value of CVR_J we flag 18 RQs out of 43 as influential. It might give us some insight into the prediction and influential pictures if we delete observations 17 and 39 responsible for the collinearity.

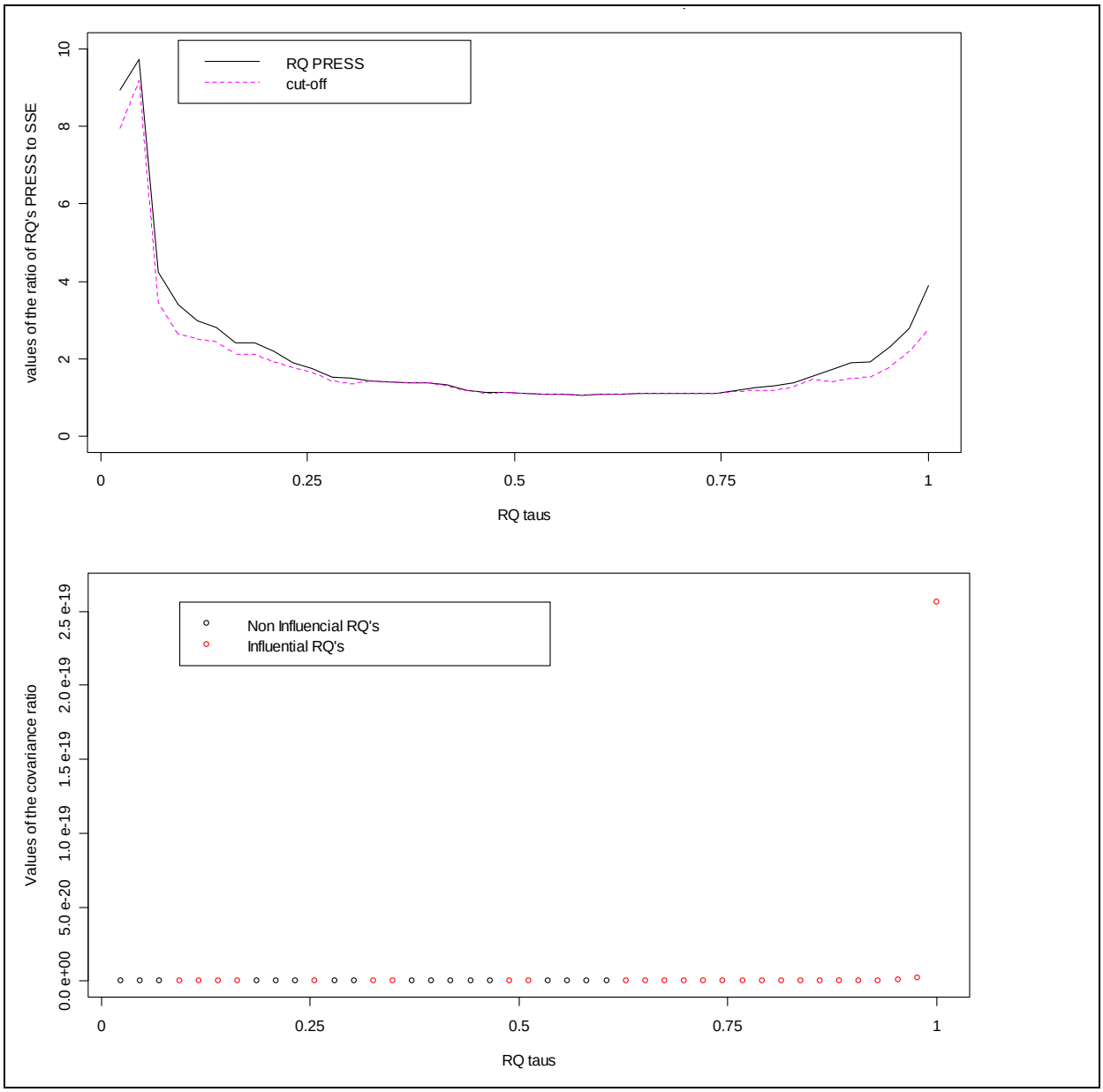


Figure 6.1: The prediction and influential diagrams exhibited by the Gunst and Mason full data set

Figure 6.2 below gives these pictures based on the reduced data set.

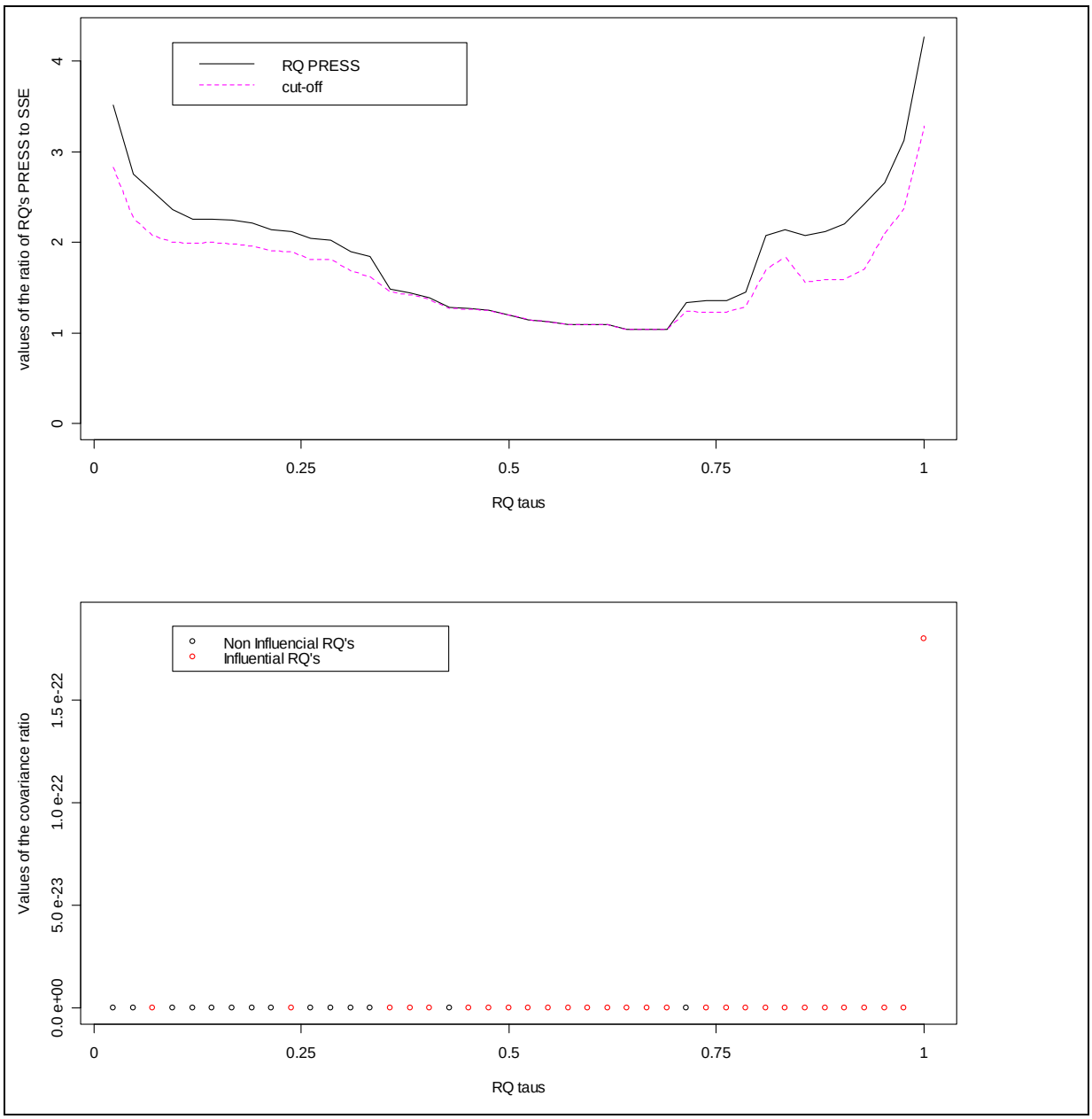


Figure 6.2: The prediction and influential diagrams exhibited by the Gunst and Mason data set with observations 17 and 39 deleted

Prediction has moderately improved in the sense that the $PRESS_J$'s from the reduced data set are generally relatively smaller than those from the full data set. However, more $PRESS_J$ are flagged for the reduced data set. More RQs based on CVR_J are flagged in the full data set than in the reduced on, *i.e.*, 17 and 14 respectively.

6.2.3 Conclusions

The diagnostics picture exhibited by the Gunst and Mason data set is very much in line with the simulation results as well as with what one would expect, especially the leverage and

collinearity picture. In the full data set the values of the T_J's and ω_J's are smaller due to the collinearity induced by observations 17 and 39. Hence, more RQs are flagged for the reduced data set with the two observations deleted than for the full data set. However, deleting all the leverage points, 17, 20, 27, 39 and 46 resulted in a drastic decrease in the values of the T_J's, thereby failing to flag any RQ. So, leverage points, whether they are type A leverage points (collinearity influential points) or not, resulted in RQs (multiple) leverage cases. On the other hand, the prediction picture also improved in the absence of type A leverage points (and leverage points in general) in the sense that the $PRESS_J$ values are smaller for the reduced data sets than the full data sets. However, more $PRESS_J$ s are flagged in the reduced data sets than the full data set. The RQs case influence picture is similar to the RQs leverage picture. This is desirable since RQs are more susceptible to leverage points than outliers due to the fact that they are unbounded in the predictor space, but bounded in the response variable.

The Gunst and Mason data set is a good example for exploring the effect of type A observations (leverage points) which induce collinearity. On the other hand, type B observations are leverage points that obscure collinearity. We explore this aspect in the next section using the Hocking and Pendleton data set.

6.3 The Hocking data set

The Hocking data set consists of 26 observations with three non constant term predictors, X_1, X_2 and X_3 with the last predictor being an approximate linear combination of the first two, resulting in the collinearity present in this data set. However, the presence of a type B leverage point (observation 24) hides the collinearity. Thus, the Hocking data set is a good example of how X outliers (type B leverage points) can obscure collinearities in a data set. Hence, this data set may be viewed as the "converse" of the Gunst and Mason data set. For all the single case regression diagnostics of this data set and the criticisms of the various proposals for the remedy of type B observations and collinearities, see Hocking and Pendleton (1983).

We start our analysis with a brief overview of some single case regression diagnostics of this data set. These are given in table 6.10 below.

Table 6.10: Some single case regression diagnostics for the Hocking data set

observation	frequency	h_i	t_i	t_i^*	D_i
1	5	0.2154	0.2838	0.2778	0.0055
2	1	0.0932	0.7252	0.7172	0.0135
3	2	0.0476	0.2112	0.2065	0.0006
4	1	0.0420	0.0235	0.0229	0.0000
5	1	0.0525	-0.1467	-0.1434	0.0003
6	5	0.1547	0.6057	0.5967	0.0168
7	1	0.0815	0.2234	0.2185	0.0011
8	14	0.3010	0.0778	0.0760	0.0007
9	5	0.1550	0.2434	0.2381	0.0027
10	6	0.1470	0.3485	0.3414	0.0052
11	3	0.1736	-2.0037	-2.1651	0.2108
12	1	0.0531	-0.2315	-0.2265	0.0008
13	2	0.1629	0.1136	0.1110	0.0006
14	8	0.1744	-0.1701	-0.1663	0.0015
15	3	0.1218	0.4744	0.4659	0.0078
16	7	0.1772	-0.2968	-0.2906	0.0047
17	0	0.0400	2.2057	2.4419	0.0506
18	1	0.1135	-3.5604	-5.3435	0.4057
19	2	0.1602	-0.7247	-0.7167	0.0250
20	0	0.1138	0.6618	0.6531	0.0141
21	2	0.1190	0.2568	0.2513	0.0022
22	2	0.0554	0.4461	0.4378	0.0029
23	2	0.0587	-0.0343	-0.0335	0.0000
24	23	0.9268	1.1095	1.1157	3.8941
25	2	0.1590	0.1532	0.1498	0.0011
26	1	0.1010	-0.3562	-0.3490	0.0036

Using $h_i > \frac{2p}{n} = 0.308$, clearly observation 24 is an extreme leverage point and it has the highest frequency of being included in ESs corresponding to RQs. In table 6.11 we present the RQ (multiple) case leverage diagnostics. Notice also that those points which are relative outliers or non leverage points have extremely low frequencies, *e.g.* observations 17 and 20.

Table 6.11: RQs case leverage picture and the observations contained in the ESs corresponding to the RQs for the Hocking data set

τ	T_J	ω_J	ES			
0.085	6.178E-04	2.877E-04	8	11	16	18
0.093	4.053E-04	6.775E-06	8	11	16	19
0.123	3.795E-03	2.945E-03	8	11	19	24
0.186	2.419E-03	6.237E-04	8	12	13	24
0.205	3.199E-03	4.963E-04	8	13	14	24
0.253	2.482E-03	1.612E-03	1	14	24	26
0.259	2.052E-03	8.210E-04	1	5	14	24
0.305	2.962E-03	9.577E-04	1	14	16	24
0.366	1.682E-03	3.310E-04	1	4	16	24
0.402	2.196E-03	1.784E-03	1	14	23	24
0.441	1.020E-03	2.189E-04	14	16	23	24
0.469	2.317E-03	6.505E-04	10	14	16	24
0.537	1.839E-03	1.333E-03	7	10	14	24
0.545	1.251E-03	9.378E-05	3	9	10	24
0.551	2.385E-03	1.699E-03	3	8	10	24
0.622	3.650E-03	4.869E-03	8	9	10	24
0.631	2.608E-03	1.576E-04	8	9	24	25
0.684	2.008E-03	1.758E-03	9	15	24	25
0.723	3.075E-03	2.929E-03	8	9	15	24
0.730	3.288E-03	1.186E-03	8	10	15	24
0.739	2.523E-03	3.410E-04	6	8	21	24
0.766	1.149E-03	3.973E-04	6	21	22	24
0.828	2.410E-03	1.518E-03	6	8	22	24
0.955	3.013E-03	3.455E-03	2	6	8	24
0.957	3.948E-03	7.228E-04	6	8	16	24

Using $T_J > 7\,pK^{-1}$=1.87E-03 and $\omega_J > 7K^{-1}$=4.68E-04 we flag almost all RQs and also observation 24 is included in almost all RQs. It is interesting to see what the picture translates to for the reduced data set (without observation 24). We present these statistics for the reduced data set in the table 6.12.

Remark: It seems the RQs' affinity for leverage points to some extent offsets their "repulsion" of outliers in the event that a leverage point is also an outlier if the leverage of that point is high enough as evidenced by a 92% inclusion of observation 24 despite the fact that it is both a leverage point and a moderate outlier.

Table 6.12: RQs case leverage picture and the observations contained in the ESs corresponding to the RQs for the Hocking data set with observation 24 deleted

τ	T_J	ω_J	ES			
0.122637	2.058E-03	9.25E-05	8	11	16	19
0.132835	5.100E-04	3.72E-05	1	8	12	19
0.174595	4.225E-04	7.12E-05	1	8	12	16
0.195535	5.475E-04	1.17E-05	1	8	16	25
0.198855	4.130E-04	1.94E-06	1	8	14	25
0.205002	5.208E-04	8.43E-05	1	8	13	14
0.212238	3.811E-04	1.46E-04	1	13	14	25
0.26945	3.914E-04	7.44E-05	1	13	14	16
0.310236	3.310E-04	1.10E-04	1	5	13	14
0.337018	1.857E-04	4.24E-05	5	10	13	14
0.360633	2.445E-04	7.73E-05	10	13	14	23
0.429872	2.367E-04	7.12E-05	7	10	13	14
0.499821	2.136E-04	5.09E-05	4	10	13	14
0.507684	3.406E-04	4.61E-05	4	9	13	14
0.517091	4.230E-04	3.71E-05	9	13	14	15
0.544614	2.475E-04	3.28E-05	4	9	13	21
0.64217	1.823E-04	6.86E-05	4	9	15	21
0.654719	8.971E-05	2.64E-05	4	9	15	22
0.657185	2.688E-04	1.80E-05	8	13	15	22
0.715805	3.220E-04	8.33E-05	3	8	13	15
0.747377	2.723E-04	1.03E-04	3	8	13	22
0.780144	3.870E-04	1.08E-04	8	13	20	22
0.833112	9.873E-05	1.12E-05	8	20	22	24
0.852453	2.488E-04	4.22E-06	6	8	20	22
0.912969	4.217E-04	2.35E-05	2	6	8	20
0.930884	5.659E-04	2.80E-04	2	11	20	21
0.933721	9.731E-04	4.45E-04	11	16	20	21
0.949244	1.264E-03	4.55E-05	8	11	16	21
0.951055	1.236E-03	7.07E-04	8	11	16	17

The RQ (multiple) case leverage picture changes dramatically at the deletion of observation 24. Using $T_J > 7pK^{-1}$=2.21E-03 and $\omega_J > 7K^{-1}$=5.53E-04 we can only flag one RQ based on the statistic ω_J. The reason for the smaller values of both T_J and ω_J may be due to smaller values of ρ_J caused by the exposure of collinearity due the deletion of observation 24. Also, deletion of observation 24 may result in smaller values of γ_J which result in smaller values of both T_J and ω_J. We focus on these aspects in the next section.

6.3.1 The collinearity and variability view

Hocking and Pendleton (1983) gave the 3 bivariate scatter plots; X_1 *vs* X_2, X_1 *vs* X_3 and X_2 *vs* X_3 (reproduced in section 4.6 as figures 4.2, 4.4 and 4.3) as a preliminary step in investigating the collinearity structure of this data set. However, inspection of these plots did not suggest any extreme behaviour except that case 24 is somewhat isolated on the X_1 *vs* X_2

and X_2 *vs* X_3 plots.

It is also important to see if the collinearity structure changes upon the deletion of observation 24. Firstly we present the collinearity view based on the full data set in table 6.13 below.

Table 6.13: RQs case variability and collinearity picture and the observations contained in the ESs corresponding to RQs for the Hocking data set

τ	γ_J	ρ_J	$\lvert\mathbf{C}_J\rvert$	ω_J	ES			
0.0853	0.0079	0.2363	0.0819	2.877E-04	8	11	16	18
0.0930	0.0026	0.0170	0.0059	6.775E-06	8	11	16	19
0.1232	0.0082	2.3418	0.8118	2.945E-03	8	11	19	24
0.1861	0.0035	1.1531	0.3997	6.237E-04	8	12	13	24
0.2046	0.0068	0.4733	0.1641	4.963E-04	8	13	14	24
0.2528	0.0137	0.7658	0.2655	1.612E-03	1	14	24	26
0.2593	0.0088	0.6042	0.2095	8.210E-04	1	5	14	24
0.3053	0.0183	0.3401	0.1179	9.577E-04	1	14	16	24
0.3659	0.0126	0.1710	0.0593	3.310E-04	1	4	16	24
0.4018	0.0097	1.1910	0.4129	1.784E-03	1	14	23	24
0.4412	0.0009	1.5581	0.5402	2.189E-04	14	16	23	24
0.4686	0.0099	0.4250	0.1473	6.505E-04	10	14	16	24
0.5370	0.0049	1.7676	0.6128	1.333E-03	7	10	14	24
0.5448	0.0036	0.1698	0.0589	9.378E-05	3	9	10	24
0.5512	0.0050	2.2284	0.7725	1.699E-03	3	8	10	24
0.6215	0.0120	2.6306	0.9119	4.869E-03	8	9	10	24
0.6315	0.0030	0.3408	0.1182	1.576E-04	8	9	24	25
0.6839	0.0042	2.7288	0.9460	1.758E-03	9	15	24	25
0.7227	0.0080	2.3722	0.8224	2.929E-03	8	9	15	24
0.7304	0.0048	1.5930	0.5523	1.186E-03	8	10	15	24
0.7385	0.0023	0.9691	0.3360	3.410E-04	6	8	21	24
0.7660	0.0013	2.0233	0.7014	3.973E-04	6	21	22	24
0.8276	0.0041	2.4238	0.8403	1.518E-03	6	8	22	24
0.9549	0.0082	2.7356	0.9483	3.455E-03	2	6	8	24
0.9570	0.0037	1.2789	0.4434	7.228E-04	6	8	16	24

There are a number of cases where $\lvert\mathbf{C}_J\rvert < \lvert\mathbf{C}\rvert$ as shown by $\rho_J > 1$ indicating a higher degree of collinearity at full design matrix than at the RQ (ES) level.

In the following table we give the collinearity and variability quantities of the reduced data set.

It is important to see the RQ (multiple) case influence case. We present this for the reduced data set in table 6.14 below.

Table 6.14: RQs case variability and collinearity picture and the observations contained in the ESs corresponding to RQs for the Hocking data set with observation 24 deleted

τ	γ_J	ρ_J	$\lvert\mathbf{C}_J\rvert$	ω_J	ES			
0.1226	0.0032	0.1821	5.89E-03	9.25E-05	8	11	16	19
0.1328	0.0175	0.0133	4.30E-04	3.72E-05	1	8	12	19
0.1746	0.0130	0.0342	1.11E-03	7.12E-05	1	8	12	16
0.1955	0.0212	0.0035	1.12E-04	1.17E-05	1	8	16	25
0.1989	0.0116	0.0010	3.39E-05	1.94E-06	1	8	14	25
0.2050	0.0028	0.1854	6.00E-03	8.43E-05	1	8	13	14
0.2122	0.0146	0.0628	2.03E-03	1.46E-04	1	13	14	25
0.2694	0.0191	0.0243	7.88E-04	7.44E-05	1	13	14	16
0.3102	0.0067	0.1031	3.34E-03	1.10E-04	1	5	13	14
0.3370	0.0040	0.0659	2.13E-03	4.24E-05	5	10	13	14
0.3606	0.0054	0.0892	2.89E-03	7.73E-05	10	13	14	23
0.4299	0.0036	0.1251	4.05E-03	7.12E-05	7	10	13	14
0.4998	0.0026	0.1224	3.96E-03	5.09E-05	4	10	13	14
0.5077	0.0062	0.0466	1.51E-03	4.61E-05	4	9	13	14
0.5171	0.0102	0.0227	7.36E-04	3.71E-05	9	13	14	15
0.5446	0.0022	0.0928	3.01E-03	3.28E-05	4	9	13	21
0.6422	0.0010	0.4453	1.44E-02	6.86E-05	4	9	15	21
0.6547	0.0003	0.4854	1.57E-02	2.64E-05	4	9	15	22
0.6572	0.0017	0.0643	2.08E-03	1.80E-05	8	13	15	22
0.7158	0.0038	0.1375	4.45E-03	8.33E-05	3	8	13	15
0.7474	0.0019	0.3330	1.08E-02	1.03E-04	3	8	13	22
0.7801	0.0042	0.1599	5.18E-03	1.08E-04	8	13	20	22
0.8331	0.0005	0.1393	4.51E-03	1.12E-05	8	20	22	24
0.8525	0.0019	0.0136	4.39E-04	4.22E-06	6	8	20	22
0.9130	0.0050	0.0295	9.55E-04	2.35E-05	2	6	8	20
0.9309	0.0007	2.4441	7.91E-02	2.80E-04	2	11	20	21
0.9337	0.0015	1.8608	6.02E-02	4.45E-04	11	16	20	21
0.9492	0.0028	0.1019	3.30E-03	4.55E-05	8	11	16	21
0.9511	0.0028	1.5861	5.13E-02	7.07E-04	8	11	16	17

There are 3 cases where $\lvert\mathbf{C}_J\rvert < \lvert\mathbf{C}\rvert$ as shown by $\rho_J > 1$ indicating a higher degree of collinearity at full design matrix than at the RQ (ES) level.

Clearly, there is a remarkable decrease in the statistics, ρ_J, $\lvert\mathbf{C}_J\rvert$ and γ_J based on the reduced data set. Therefore the deletion of observation 24 has exposed the two data weaknesses, *viz.* the collinearity aspect and the lack of variability at the RQ level.

In the next section we tackle the effect of type on prediction and influence.

6.3.2 RQ case outlier and influential view

It is also important to see if type B observations result in RQ (multiple) case poor prediction or/and RQ (multiple) case influential points. Figure 6.3 below gives the RQ prediction and influence picture based on the full data set.

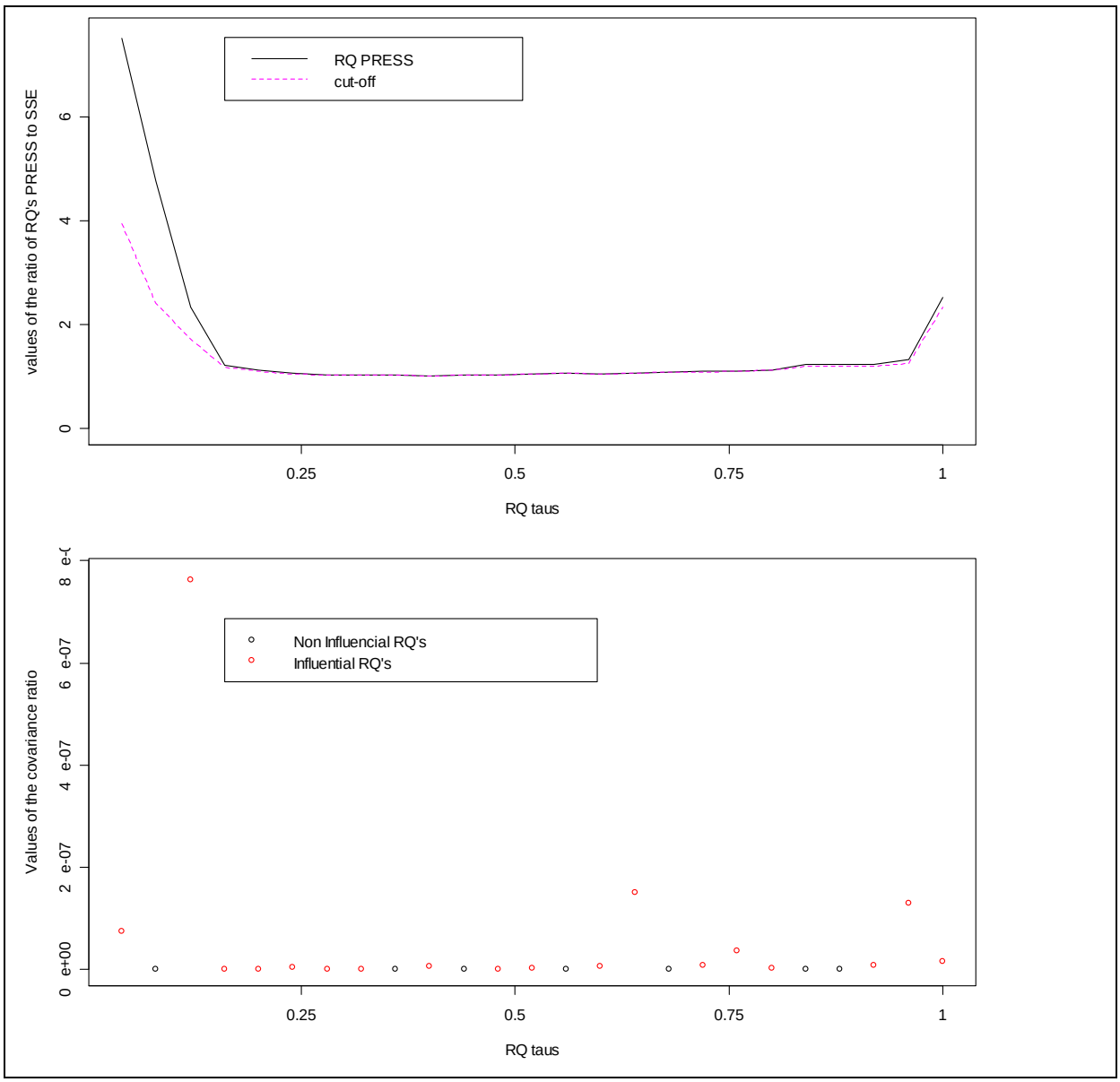


Figure 6.3: The prediction and influential diagrams exhibited by the Hocking data set

Poor prediction is mostly exhibited at the lowest few RQs. We flag 12 RQs out of 25 based on the cut-off value of CVR_J. In order to see if there is any change in the RQ prediction and influential picture due to the deletion of observation 24 we present the prediction and influential picture based on the reduced data set in figure 6.4 below.

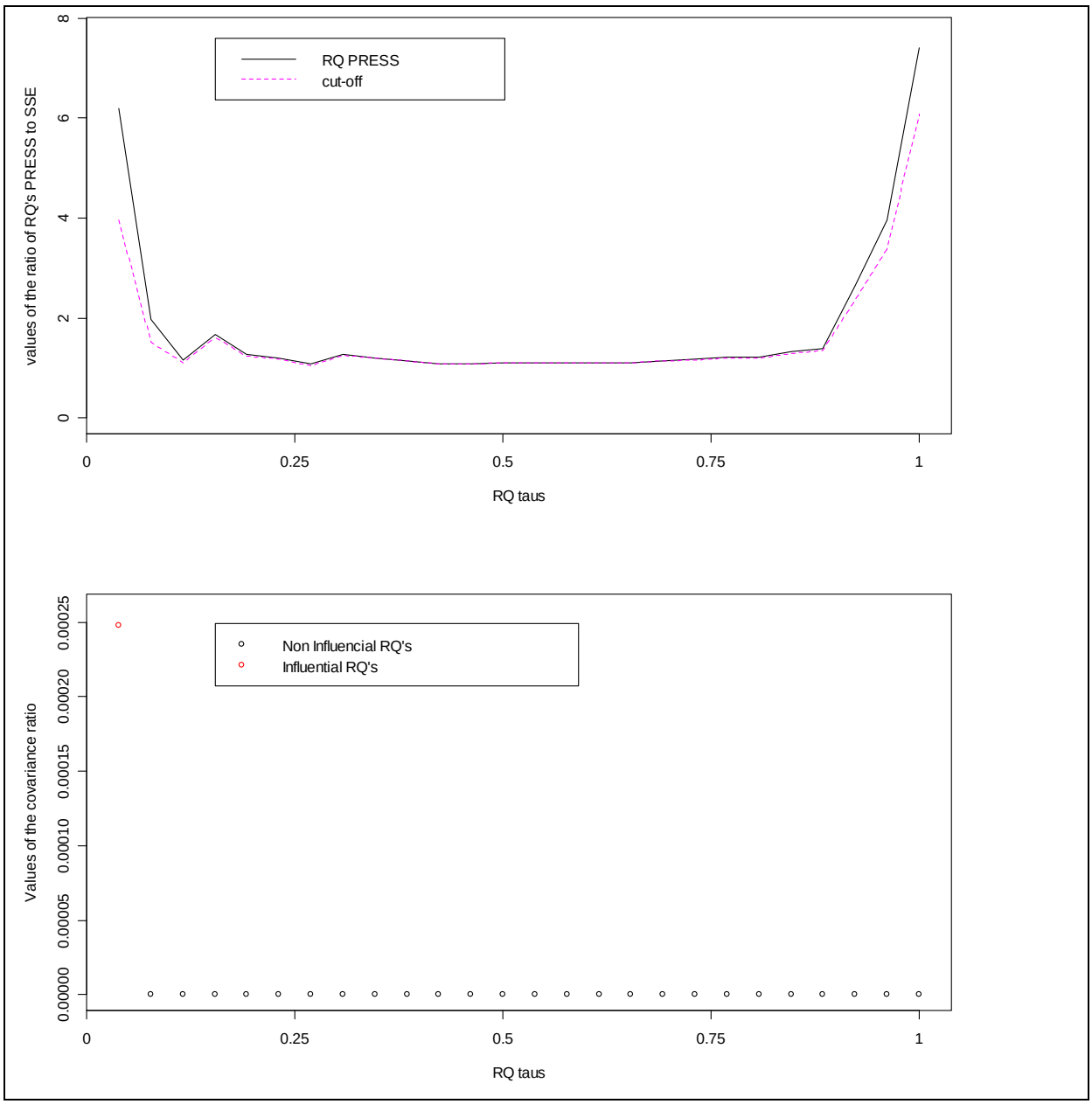


Figure 6.4: The prediction and influential picture exhibited by the Hocking reduced data set

Poor prediction is exhibited at both the lower and the upper extreme RQs although prediction has moderately improved upon the deletion of the leverage point 24, similar to the Gunst and Mason reduced data set (figure 6.2). However, the deletion of observation 24 results in dramatic change in the influence picture. Using the cut-off values of CVR_J only one RQ (the first one) is flagged.

6.3.3 Conclusions

As expected the RQs' leverage and influence picture is in line with the simulation results. The presence of a type B leverage point, observation 24, resulted in a massive flagging of RQs

based on multiple leverage and influential statistics T_J and CVR_J. However, there is not much difference in the prediction picture save that the $PRESS_J$ values follow a skewed distribution in the full data set while they follow an approximate symmetric one for the reduced data set. Also, the presence of collinearity was evident in the reduced data set. Thus observation 24 was responsible for hiding the collinearity.

The two data sets, *viz.* the Gunst and Mason data set and the Hocking data set exhibit the collinearity aspect due to the presence or the absence of leverage points (collinearity influential points). It might be enlightening to see the effect of the presence of collinearity in the absence of leverage points. One typical example of such a data set exhibiting this type of collinearity in the literature is the Hald data set. We explore this data set in the next section.

6.4 The Hald data set

The Hald data set (cement hardening data set) is a good example of a data set that has no leverage points, but exhibiting collinearities. The data set has $n = 13$ observations and 4 predictors that are almost linearly related (see Montgomery and Peck, 1982). Actually, X_1 and X_3 are highly correlated ($c_{13} = -0.824$) as well as X_2 and X_4 ($c_{24} = -0.975$). In the literature this data set has been used for variable selection. However, here we want to investigate the effectiveness of our RQ multiple diagnostics in the absence of leverage points (to which RQs can be susceptible), since it does not have leverage points as exhibited by the single case diagnostics in table 6.15 below.

Table 6.15 Some single case regression diagnostics for the Hald data set

observation	frequency	h_i	t_i	t_i^*	D_i
1	9	0.5503	0.0029	0.0027	0.0000
2	2	0.3332	0.7566	0.7345	0.0572
3	7	0.5769	-1.0503	-1.0581	0.3009
4	0	0.2952	-0.8411	-0.8240	0.0593
5	2	0.3576	0.1279	0.1198	0.0018
6	0	0.1242	1.7148	2.0170	0.0834
7	4	0.3671	-0.7445	-0.7218	0.0643
8	4	0.4085	-1.6878	-1.9675	0.3935
9	2	0.2943	0.6708	0.6459	0.0375
10	9	0.7004	0.2103	0.1973	0.0207
11	6	0.4255	1.0739	1.0859	0.1708
12	7	0.2630	0.4634	0.4394	0.0153
13	0	0.3037	-1.1241	-1.1459	0.1102

Clearly, using $h_i > \frac{2p}{n} = 0.769$ there are neither leverage points in this data set nor inordinately large outlier diagnostics. However, observations with relatively high leverage values and relatively low outlier diagnostic values have higher frequencies of inclusion in the RQs as shown in table 6.15 above and table 6.16 below which presents the RQ leverage picture.

Table 6.16 RQs case leverage picture and the observations contained in the ESs corresponding to the RQs for the Hald data set

τ	T_J	ω_J	ES				
0.2302	0.0049	0.0002	1	3	4	5	8
0.2526	0.0066	0.0004	3	4	5	7	8
0.3365	0.0148	0.0135	3	5	7	8	10
0.4258	0.0162	0.0038	1	3	5	7	10
0.4285	0.0122	0.0006	1	3	5	10	12
0.5144	0.0167	0.0158	1	3	10	11	12
0.5470	0.0092	0.0006	1	5	10	11	12
0.7519	0.0100	0.0067	1	9	10	11	12
0.8245	0.0096	0.0045	1	2	10	11	12
0.8601	0.0030	0.0003	1	2	5	11	12

Using $T_J > 15pK^{-1} = 0.0583$ and $\omega_J > 15K^{-1} = 0.0117$ we flag no RQ.

The small ω_J may be a result of a small γ_J or a small ρ_J (see 4.2.2). We present these statistics in the following table.

Table 6.17: RQs case variability and collinearity picture and the observations contained in the ESs corresponding to the RQs for the Hald data set

τ	γ_J	ρ_J	$\lvert\mathbf{C}_J\rvert$	ω_J	ES				
0.2302	0.0043	0.0928	0.0001	0.0002	1	3	4	5	8
0.2526	0.0112	0.0919	0.0001	0.0004	3	4	5	7	8
0.3365	0.0205	1.7118	0.0018	0.0135	3	5	7	8	10
0.4258	0.0168	0.5899	0.0006	0.0038	1	3	5	7	10
0.4285	0.0012	1.3867	0.0015	0.0006	1	3	5	10	12
0.5144	0.0304	1.3538	0.0014	0.0158	1	3	10	11	12
0.5470	0.0277	0.0595	0.0001	0.0006	1	5	10	11	12
0.7519	0.0415	0.4218	0.0005	0.0067	1	9	10	11	12
0.8245	0.0511	0.2274	0.0002	0.0045	1	2	10	11	12
0.8601	0.0122	0.0642	0.0001	0.0003	1	2	5	11	12

The small ω_J is generally a result of small γ_J values of rather than ρ_J. The collinearity picture is exposed by the small values of $\lvert\mathbf{C}_J\rvert$. However, there are few cases where $\rho_J > 1$ indicating a higher degree of collinearity at the full design matrix level than at the RQ (ES) level.

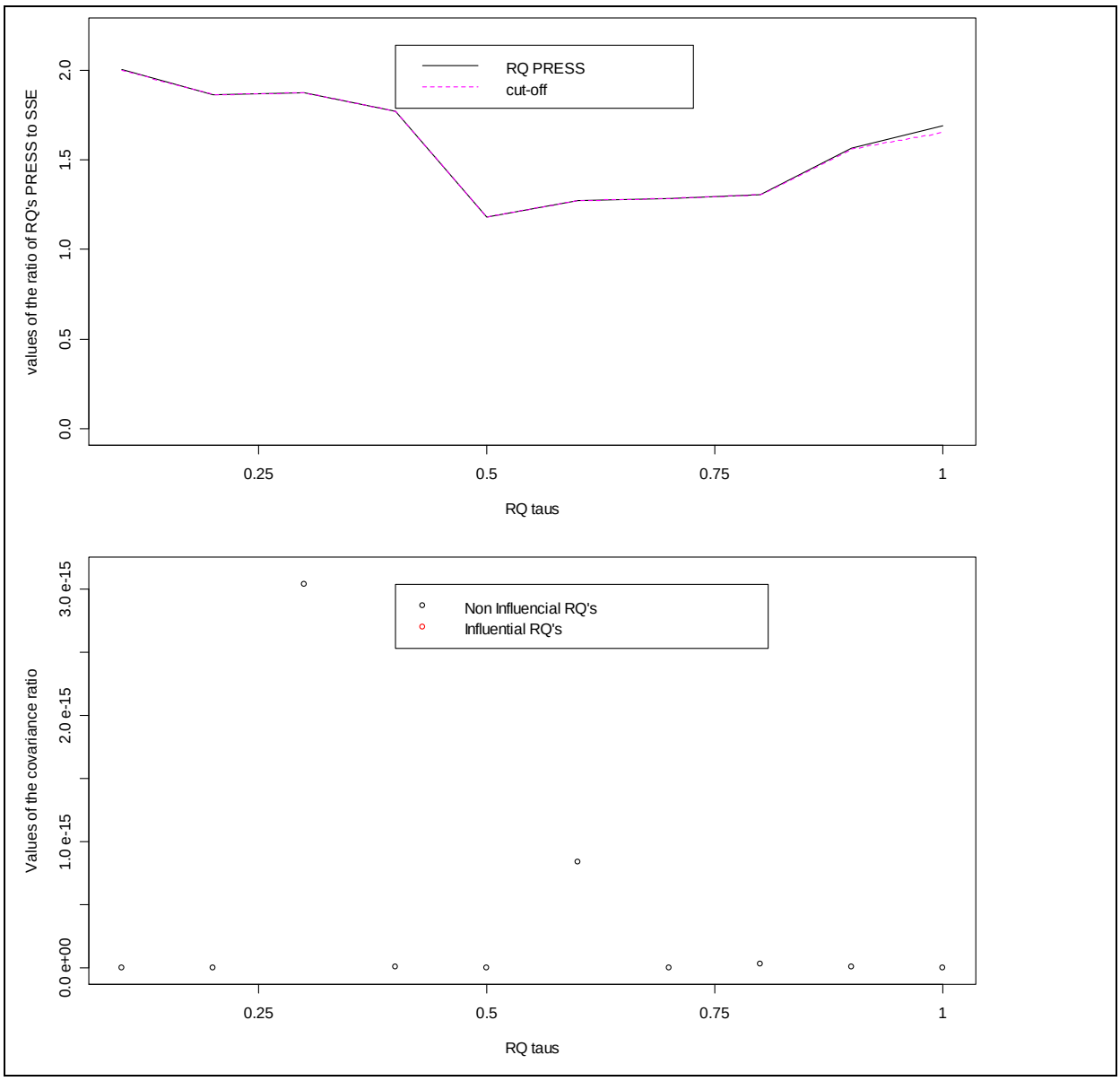


Figure 6.5: The prediction and influential picture exhibited by the Hald data set

Clearly, prediction is not that poor. Using the cut-off values of CVR_J, no RQs are flagged. This is expected since there are neither leverage points nor outliers in this data set. Furthermore, this might be due to the presence of collinearity which results in small ρ_J's which in turn result in small CVR_J's.

Remark: The phenomenon whereby the ERW is small needs further scrutiny especially in the RQ scenario.

6.5 Discussions and conclusions

The applications to the data sets reveal that the derived RQs (multiple) case diagnostics and their cut-off values are generally in line with the simulation results as one would expect. However, the prediction picture exhibited using the Hald data set is not bad at all contrary to what would be expected. The picture exhibited by the leverage and outlier diagnostics is clearly that of a trade-off between leverage points and outliers.

On the one hand, the RQs' affinity for leverage points was consistently evident, while on the other hand, their repulsion of outliers was also consistent. In the event that a leverage point is also an outlier the net result is a trade-off between these two antagonistic forces. So the researcher must be careful as to the nature of the underlying error distribution when he/she wants to determine RQ (multiple) influential cases. However, it seems the RQs' affinity for leverage points will always off-set their repulsion of outliers to some degree.

As expected RQs (multiple) leverage statistics, T_J and ω_J, become larger as the degree of leverage and the number of leverage points increase as evidenced when leverage points were deleted. However, this picture happens to a lesser extent for outliers since RQs have influence functions that are bounded in the Y space but unbounded in the X space. As a consequence most RQs were flagged as influential cases because of the leverage points rather than the outlier points. In the Hald data set, where there were neither leverage points nor outliers, as was expected, no RQ was flagged. Furthermore, leverage points we encountered in the data sets affected the correlation structure of the design matrix in addition to being influential.

CHAPTER 7 CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

7.1 Introduction

In this chapter we make some concluding remarks on the contribution of this thesis to the understanding of the three-tier relationship amongst ES statistics (estimators), RQ statistics (estimators) and OLS statistics (estimators). Also, we give some areas of further research. One such area of study is that of variable selection in the RQ scenario, especially using ES procedures. We mainly explore this dimension of study in the current chapter, giving some tentative results.

In the next subsection we give the overall conclusions. In section 7.3 we give some areas of further research. In section 7.4 we apply the RQ Lasso and Lars-Lasso procedures based on cross-validation to some real life data sets.

7.2 Conclusions

The major contribution of this thesis can be broadly viewed as extending the single case OLS regression diagnostics to the RQ scenario using ES methods. This is made possible by the fact that many OLS statistics (estimators) can be expressed as weighted averages of ES statistics (estimators) via the ERW, ω_J. As a consequence, we gave an extensive overview of both the existing OLS regression diagnostics and ES statistics in Chapter 4. Building on this existing foundation we further deduced some "new" relations between the OLS statistics and ES statistics via the ERW. Some of the relations between the OLS statistics and ES (RQ) statistics are not quite useful to the RQ situation while others are. We used the latter to further understand and address the problems of collinearity, leverage points and outliers in the RQ scenario.

In subsection 4.2.2 we give Theorem 4.1 and corollary 4.1 which show the relationship between the determinants of the scatter matrices $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{X}}_J'\tilde{\boldsymbol{X}}_J$ and the determinants of the correlation matrices $\boldsymbol{C}$ and $\boldsymbol{C}_J$ respectively. The consequence of this theorem and corollary is another very useful statistic, the ERW. We explored various views of this statistic, including the collinearity view based on the statistic $\rho_J \equiv \frac{|\boldsymbol{C}_J|}{|\boldsymbol{C}|}$.

In both subsections 4.2.2 and 4.2.3 we elaborate on the relationships between the determinant

of the design correlation matrix, the EWR and some well known measures of collinearity such as the VIFs. Actually, we showed that some existing multiple case collinearity measures based on the VIFs comprise the ERW. Therefore there is an inherent relationship amongst the determinant of the design correlation matrix, the VIFs and the ERW. Based on this three-tier relationship, the fact that the determinant of the correlation matrix is bounded between singularity (0) and nonsingularity (1) and the fact that cut-off values based on the other collinearity measures (like VIFs and tolerances) can be extended to the determinant of the correlation matrix, we conclude that it is the most appealing measure to use in the RQ situation. Hence we mainly made use of the statistics $|\boldsymbol{C}_J|$, $|\boldsymbol{C}|$ and ρ_J in Chapters 5 and 6 as collinearity diagnostic tools.

Since the three areas of study *viz.* OLS, ESs and RQs are related via the ERW, a detailed understanding of this statistic is imperative. The ERW consists totally of design space (X space) information. Decomposing this statistic into a product of a constant and two factors (the collinearity one, ρ_J and the variability one, γ_J), we found that more information with regards to the two major design space aberrations *viz.* collinearity and variability ("leverage") is exposed. The collinearity factor ρ_J exposes the fact that collinearity can be worse at the RQ (ES) level. Therefore more care must be taken when dealing with collinearity in the RQ scenario. It is expected that a smaller value of the collinearity factor will result in a markedly smaller ERW. However, in the presence of leverage points which are also collinearity influential points, this is not the case.

Collinearities arise from various sources. Leverage points which are also type A outliers, also induce collinearities. RQs have a high affinity for leverage points since their influence functions are unbounded in the design space but bounded in the response variable. As a consequence, an ES corresponding to a RQ tends to contain the leverage points (including collinearity influential points) present in the design matrix. Thus the leverage points included result in a relatively large variability factor that offsets a smaller collinearity factor. This was clearly exposed by both the artificial data sets and the real life ones. Also, this tendency of "fatally" attracting leverage points by RQs, results in RQ statistics (estimators) contributing immensely towards the building of OLS statistics (estimators) compared to other ESs that do not correspond to RQs. Using small artificial data sets, we have found that the ERWs corresponding to RQs can contribute as much as 75% in the building of OLS statistics yet they are far fewer than the other ESs that do not correspond to RQs. This "fatal" affinity for

leverage points by RQs may explain why RQ based procedures can be more adversely affected by collinearity influential points (type A outliers) than OLS procedures. Generally, the global picture is that the problems that affect OLS procedures are more likely to affect RQ's since there is an inherent relationship between the two.

The multiple (RQ) case diagnostics studied here can be categorized into 3 main classes *viz.*, leverage, outliers and influential cases. However, we also considered the usual OLS regression diagnostics in order to understand how the picture changes from the single case scenario to the RQ scenario. The inclusion rate of observations into ES corresponding to RQs interplay between leverage points and outliers is antagonistic in nature. There is a high inclusion rate of leverage points into ESs corresponding to RQs, in the presence of leverage points alone (without outliers being present). Also, the higher the leverage value of an individual observation, the higher is its rate of inclusion into the ESs corresponding to RQs. However, in the presence of outliers, the inclusion rate of leverage points into ESs corresponding to RQs is drastically reduced. Actually, the inclusion rate of leverage points into ESs corresponding to RQs decrease as the number of outliers and the degree of outlyingness increase. The overall inclusion rate of leverage points into ESs corresponding to RQs is the trade-off between the RQs affinity for leverage points and their exclusion (repulsion) of outliers. This phenomenon again is attributable to the fact that RQs have influence functions that are unbounded in the X space but bounded in the Y space. The overall conclusion is that the inclusion rate of leverage points into ESs corresponding to RQs depends on the abundance ratio of leverage points to outliers as well the degrees of outlyingness of both leverage points and outliers. Also, we have seen that the influence picture depends on the inclusion and exclusion of leverage points and outliers respectively. Since leverage and/or outliers can be influential we first consider the RQ (multiple) case leverage and outlier picture before we elaborate on the resulting influential picture. We consider the leverage picture in the next paragraph.

Since RQs have influence functions that are bounded in the Y space but unbounded in the X space they are fairly robust to outliers but susceptible to leverage points. It is therefore imperative to address the leverage problem in the RQ scenario. In section 4.4 we give Theorem 4.2 and use it to elaborate the leverage view of the ERW. We showed that the most generalized regression diagnostics comprise ERW and some other factor. In section 4.5 we deduced item (iii) of Theorem 4.3 which relates the usual OLS leverage to RQ (ES) case leverage. Using this result, our original idea was to directly extend the Hoaglin and Welsch

(1978) procedure to the RQ scenario. However, this is not reasonable due to large ERWs which in turn lead to large values of the leverage statistic T_J and the fact that the number of ESs corresponding to RQs is far fewer than the total number of ESs (see section 5.7). The flagging of RQs picture shows that more RQs are flagged as the number of leverage points approaches $p-1$, the number of non constant term predictors, due to the lack of variability in the statistic T_J. A small value of T_J can be a result of a small ERW (most likely) and small sum of the residual freedoms, $\sum_{i\notin J} R_{iJ}$. In turn a small ERW may be a result of a small product $\gamma_J.\rho_J$, where γ_J and ρ_J raise the variability and collinearity views respectively as defined in (4.2.5). However, in the presence of leverage points (including collinearity influential points) this phenomenon is almost nonexistent. Actually, in the presence of collinearity influential points γ_J and ρ_J are often "antagonistic" in size and may not result in a small ERW and hence T_J to the extent that one would expect. As a consequence the number of RQs flagged in the presence of type A leverage points (collinearity influential points) is not much less than those flagged in the presence of type B leverage points (collinearity hiding points). However, the difference is more noticeable as the number of predictors p increases. Actually, more RQs are flagged in the presence of type B leverage points than in the presence of type A leverage points as p increases. This interaction behaviour may not be exhibited if the collinearity arises from another source other than collinearity influential points.

Although RQs are fairly robust to outliers due to the fact that their influence functions are bounded in the Y space we considered the outlier aspect in order to have a holistic picture. Viewing a RQ as an ES, we considered "prediction" based on the statistic $PRESS_J$. Based on Hawkins *et al.* (1984)'s original incorrect result relating SSE to $PRESS_J$ we deduced the correct one (see Theorem 4.4, item (ii)). Our original idea was to extend the usual leave one observation out predicted residual sum of squares $PRESS$ to the RQ predicted residual sum of squares $PRESS_J$. However, it is not reasonable to extend the usual leave one observation out predicted residual sum of squares $PRESS$ to the RQ predicted residual sum of squares $PRESS_J$ (see section 4.7). Hence, we determined the cut-off values of both these statistics (T_J and $PRESS_J$) using simulation studies. The RQ (ES) predicted residual sums of squares, $PRESS_J$s, exhibited a consistent pattern in the sense that larger values of this statistic are always associated with extreme τ levels and prediction becomes poorer as the underlying

error distribution becomes heavier despite the design matrix used. This is expected since all the error distributions used are symmetrical heavy tailed distributions which imply outliers. However, there are minimal differences in RQ prediction between type A leverage points and absence of leverage points *i.e.*, the absence of leverage points resulted in slightly smaller $PRESS_J$ s than those obtained in the presence of type A outliers, while type B leverage points resulted in the slightly larger $PRESS_J$ s under the given underlying error distribution. Furthermore, the degree of poor prediction as a result of design matrices depends on the type of leverage points present in them. Type A leverage points tend to slightly cause poor prediction while type B leverage points (D3 and D5) results in slightly poorer prediction. As a consequence, the combined effect of both leverage points and outliers is that prediction is poorest at design matrices in the presence of type B outliers under the heaviest error distribution.

Leverage points and/or outliers can have undesirable effects on prediction or on the fit. In view of the fact that the inclusion rate of observations into ESs corresponding to RQs depends on the antagonistic interplay between leverage points and outliers it is important to understand the influential dynamics picture between leverage points and outliers. In order to identify RQ influential cases we suggested the RQ covariance ratio, CVR_J because of the attractiveness of its determinantal relationship with the ERW (and T_J). To determine the cut-off values of CVR_J , we used the cut-off values of the statistics ω_J and $PRESS_J$ since CVR_J is made up of these statistics.

As expected more RQs were flagged due to leverage points than due to outliers since RQs are susceptible leverage points but fairly robust to outliers. Thus the flagging of influential RQs due to outliers (only) slightly increases as the heaviness of the underlying error distribution increases while flagging of RQs influential cases due to leverage (only) depends mainly on the nature (kind) of leverage rather than the degree of leverage. Actually, more RQs influential sets are flagged due to type B leverage points than type A leverage points. This may be attributed to the presence of collinearity that results in a relatively smallish T_J statistic value. Furthermore, in the presence of type B leverage points under the Normal distribution the flagging of RQs due to the CVR_J cut-off values is similar to that due the T_J cut-off values. This is expected since RQs have influence functions that are bounded in the Y space but unbounded in the X space. As a consequence leverage points are bound to have a more influential effect than outliers. However, the flagging of influential points is drastically reduced at designs that contain leverage under heavy tailed error distributions (which imply

outliers). Again, this is attributable to the two antagonistic forces between leverage and outliers in the RQ scenario, *viz.*, the RQs affinity for leverage points and their exclusion of outliers. In a data set containing both leverage points and outliers, RQs may be proposed as a remedy for outliers but the presence of outliers may reduce the influence of leverage points since naturally, a change in the observations included in the RQ set will result in the change in the statistic CVR_J and hence the subsequent flagging. The overall influence picture is summed up in the following remark.

Remark: RQs have influence functions that are bounded in the Y space but unbounded in the X space. As a consequence they have a great affinity for leverage points but a great repulsion (exclusion) of outliers. So the resulting influence picture is the net resultant between these two antagonistic forces.

The performance of these statistics and their cut-off values is quite satisfactory both in the simulation studies and in applications to real life data sets. The contributions of this study to the understanding of the three-tier relationship amongst OLS statistics (estimators), ES statistics (estimators) and RQ statistics (estimators) give insight into further research such as variable selection in the RQ arena.

Remark: The flagging of certain RQs based on RQ multiple case statistics cut-off values at different sizes of p , can also be viewed as a variable selection procedure.

In this thesis we only considered using RQs in the linear model situation. It would be interesting to find out how far ES methods are applicable to other variants of RQ scenarios. In the next section we consider areas of further research and give some tentative results on variable selection.

7.3 Further research

In this thesis the model upon which we based our analyses is the linear one. As a consequence, these analyses are applicable to transformations that would result in a linear model. The linear parametric quantile regression model has no explicit solution. However,

this model can be formulated as a linear programming problem (LP) and solved using recent efficient LP algorithms (see Chapter 3). The resulting basic optimal solution corresponds to an ES.

In this thesis the ERW, ω_J was found to play an important role in diagnostics. Since the ERW is simply the determinant of the elemental regression matrix, $\left(\mathbf{I}-\mathbf{H}_I\right)=\left(\mathbf{I}_{n-p}+\mathbf{H}_{IJ}\right)^{-1}$ (see remark 4, section, 4.4), one would expect other derivatives derived from this matrix also to have important diagnostic information. This is an important area for future research.

In subsection 4.5.1 T_J was defined as an ES predictive weighted leverage statistic. The statistic U_J, and its standardized version U_J/SSE, were also introduced in section 4.7. However, in the subsequent studies we focussed only on T_J. Further research on the uses of U_J, still needs to be done.

While, on the one hand, the linear regression model is widely used in applied statistics, on the other hand, the list of research areas where non linear models and other forms of regression models are applicable is ever growing (see *e.g.* Yu *et al.*, 2003). Therefore, further research on the use of ES and RQ based techniques in such areas is of vital importance. These research areas include the following amongst others: Bayesian Quantile Regression (see Chen *et al.*, 2001), GeneChip Microarray studies (see Wang and He, 2007), kernel based estimation (see Li, Liu and Zhu, 2007), Transformations in the Box-Cox transformations and Power transformations (see Mu and He, 2007). In some situations the application of non linear quantile regression is more favourable than linear quantile regression. Therefore the applicability of ES methods in non linear quantile regression estimation needs further research.

Another area that needs further research is that of variable selection in the RQ scenario, especially using ES procedures. Mosteller and Tukey (1977) point out that the OLS procedure which is based on the conditional mean of Y given $\boldsymbol{x}$, $E(Y\mid\boldsymbol{x})$ gives an incomplete picture of the relationships between variables, unlike the conditional quantile function of Y given $\boldsymbol{x}$, $Q_{Y|\boldsymbol{x}}(u)$ (see section 3.2). Actually, there may be a weak or no predictive relationship between the mean of the response variable, Y and the predictive factors, X. However, there may be stronger, useful predictive relationships at some quantile levels, of the response variable distribution. Hence variables selected using the OLS procedure can differ to those selected using the RQ procedure. Thus using all the three procedures (*i.e.* OLS, ES and RQ

procedures) in a complementary fashion is certainly more beneficial to the researcher, as it gives a more complete picture of the conditional distribution of Y given $\boldsymbol{x}$.

One approach to variable selection in the RQ arena is that of step wise procedures based on the R^2 (SSE) since the SSE is a weighted average of the $PRESS_J$'s. However, more appealing and popular modern procedures based on cross-validation exist. Two such procedures are the RQ Lasso and the Lars-Lasso procedures (see section 4.9).

In section 7.4 we apply RQ Lasso and the Lars-Lasso procedures to two data sets that exhibit collinearities, *viz.* the Gunst and Mason data set as well as the Hald data set (see sections 6.2 and 6.4).

Remark: It is not reasonable to apply shrinkage techniques to the Hocking data set which has type B leverage points that hide the collinearity as these procedures are aimed at the effects of collinearity. Type B leverage points are still a problem in as far as variable selection procedures are concerned.

7.4 Applications of the RQ lasso and lars-lasso to real life data sets

The collinearity exhibited in the Gunst and Mason data set is induced by leverage points (also referred to as type A or collinearity influential points) while the one exhibited in the Hald data set arises from a different source. Variable selection in the presence of collinearities from sources other than leverage points, has been fairly successfully dealt with in the literature. It is important to find out what the degree of success in using the RQ Lasso (see equation 4.9.4) and the Lars-Lasso (see equation 4.9.3) procedures will be for both the Gunst and Mason data set and the Hald data set. We consider the Gunst and Mason data set firstly.

Table 7.1 below shows the L_1 parameter estimates and the OLS parameter estimates for the full data set (n=49) and the reduced data set (n=47) with observations 17 and 39, corresponding to Hong Kong and Singapore respectively, deleted.

Table 7.1: Comparison of RQ ($\tau = 0.5$) Estimation and OLS Estimation

	RQ (τ =0.5)				OLS			
	$\hat{\boldsymbol{\beta}}$ (n =49)	t-value	$\hat{\boldsymbol{\beta}}$ (n =47)	t-value	$\hat{\boldsymbol{\beta}}$ (n =49)	t-value	$\hat{\boldsymbol{\beta}}$ (n =47)	t-value
Intercept	6.192	89.184	6.1529	62.9579	6.199	105.595	6.2161	108.9530
INFD	-2.447	-2.324	-2.5787	-3.0963	-1.870	-3.306	-2.0762	-3.8120
PHYS	-0.027	-0.038	0.5651	0.7323	0.169	0.250	0.3348	0.5160
DENS	-1.472	-0.738	1.2172	1.1781	-1.095	-0.610	0.6219	0.9700
AGDS	1.042	0.560	-2.0806	-1.2406	0.863	0.484	-1.4475	-2.2620
LIT	1.120	1.214	1.6963	1.8243	2.296	2.989	2.2033	3.0480
HIED	1.899	5.355	1.7476	9.0747	1.454	3.170	1.3958	3.2070

The presence of collinearity may result in the parameter estimates exhibiting wrong signs (see section 4.2). The two observations 17 and 39 are mainly responsible for collinearity inducement. Therefore, the signs of the parameter estimates from the reduced data set with observations 17 and 39 deleted are deemed to be the correct signs of the parameter estimates. The L_1 (RQ at $\tau = 0.5$) parameter estimates have wrong signs in 3 cases, *viz.*, PHYS, DENS and AGDS while OLS parameter estimates have wrong signs in 2 cases, *viz.*, DENS and AGDS. Clearly L_1 parameter estimates exhibit signs contrary to OLS, M-estimation and ridge regression given by Gunst and Mason (1980) which showed great similarity among the three sets of parameter estimates and their corresponding t-values. This further highlights the fact that the effect of collinearity can be worse at the RQ level as shown by the statistic ρ_J in Chapter 6. Clearly L_1 (and RQ) estimation can be more adversely affected by collinearities than OLS procedures.

We judge the effectiveness of the shrinkage procedures based on the correct signs of the parameter estimates and the pairwise correlations between the given two predictors. From table 6.5 (see Chapter 6), the highest pairwise correlation coefficients in descending order of their magnitudes are; 0.9724 for DENS and ARGS, -0.7824 for PHYS and LIT, -0.6277 for INFD and LIT and 0.5686 for INFD and PHYS. We would naturally expect at least one of the variables involved in each of these pairwise high correlation coefficients to be excluded (or heavily shrunk) from the most parsimonious model after applying variable selection procedures. This is especially true of DENS and ARGS with the highest pairwise correlation coefficients. Actually, we expect at least 3 of the variables involved in the high correlation

coefficients to be excluded (heavily shrunk). The parameter estimates obtained from applying the RQ Lasso estimation and the Lars-Lasso OLS estimation are given in table 7.2 below with the shaded columns as in table 7.1. The optimal value of the optimal shrinkage parameter λ is determined from the lasso profiles corresponding to the minimum value of the RQ predicted residual sum of squares, $\min_{0\leq\lambda\leq1}(PRESS_J^{\lambda})$. The lasso profiles for the determination of the optimal value of λ are given in figure 7.1 below.

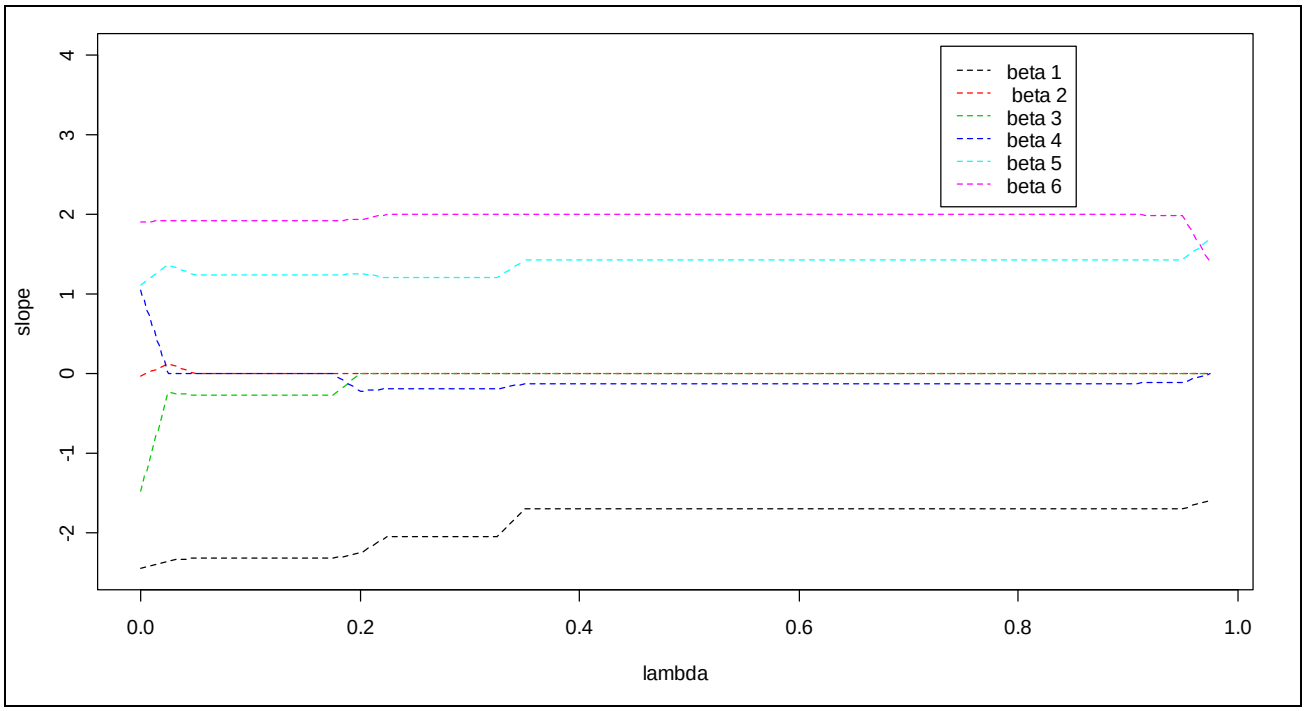


Figure 7.1: Lasso profile for RQ Lasso variable selection at $\tau = 0.5$ with optimal value, $\lambda = 0.975$

The coefficients β_2, β_3 and β_4 are shrunk to zero when the optimal shrinkage parameter is $\lambda = 0.975$. This is also depicted in table 7.2 below.

Table 7.2: Comparison of the RQ Lasso and Lars-Lasso Estimations

	RQ($\tau = 0.5$)			OLS			
	(n =49)	Lasso $\lambda = 0.975$	(n =47)	(n =49)	Lars-lasso (n =49)	Lars-Lasso ($n = p$ =7)	(n =47)
Intercept	6.192	6.18E+00	6.1529	6.199	6.199	6.130	6.2161
INFD	-2.447	-1.59E+00	-2.5787	-1.870	-0.175	0.000	-2.0762
PHYS	-0.027	-9.49E-11	0.5651	0.169	0.000	0.000	0.3348
DENS	-1.472	-5.05E-10	1.2172	-1.095	0.000	0.000	0.6219
AGDS	1.042	-6.32E-09	-2.0806	0.863	0.000	0.000	-1.4475
LIT	1.120	1.69E+00	1.6963	2.296	1.140	0.000	2.2033
HIED	1.899	1.40E+00	1.7476	1.454	0.000	0.688	1.3958

Note: The column denoted $n = p$ (Lars-Lasso) represents the variable selection based on the p observations in the ES corresponding the RQ at τ =0.5.

Three variables are excluded from the resulting model using the RQ Lasso procedure while 4 variables and 5 variables are excluded using the Lars-Lasso procedure both at the full data set and at the ES corresponding to $\tau = 0.5$ respectively. Also, all the coefficients with wrong signs are shrunk to zero using both procedures. Clearly the Lars-Lasso is a more greedy (shrinks more coefficients to zero) procedure. In the literature Lars-Lasso which minimizes the residual sum of squares (a quadratic objective function), subject to the sum of the absolute values of the non constant predictor coefficients being bounded by a constant, is found to be more greedy than the ordinary ridge regression where both the objective function and the constraints are quadratic (see Tibshirani, 1996). However, when both the objective function to be minimized and the constraints are absolute values as in the case of the RQ Lasso procedure (see equation (4.9.4)), the results are clearly different in the sense that the Lars-Lasso procedure is greedier than the RQ Lasso procedure. Mason and Gunst (1985) showed that ordinary ridge regression and robust estimators like M-estimators fail to deal with collinearity influential points. However, the results we obtained here show that the RQ Lasso and Lars-Lasso estimation could both be solutions to collinearity influential points which are type A leverage points.

The collinearities exhibited by the Gunst and Mason data set are induced by leverage points (collinearity influential points). It is important to find out how the RQ Lasso and the Lars-Lasso procedures compare for the data set exhibiting collinearities in the absence of leverage points. A typical one is the Hald data set with X_1 and X_3 highly correlated ($c_{13} = -0.824$) as

well as X_2 and X_4 ($c_{24} = -0.975$). As a consequence after applying the variable selection procedures we expect that at least one of the variables involved in the pairs of variables exhibiting high correlation coefficients to be heavily shrunk or excluded from the chosen variables (model).

Figure 7.2 below gives the RQ lasso profiles for $\tau = 0.5$. The optimal value of λ is 0.275.

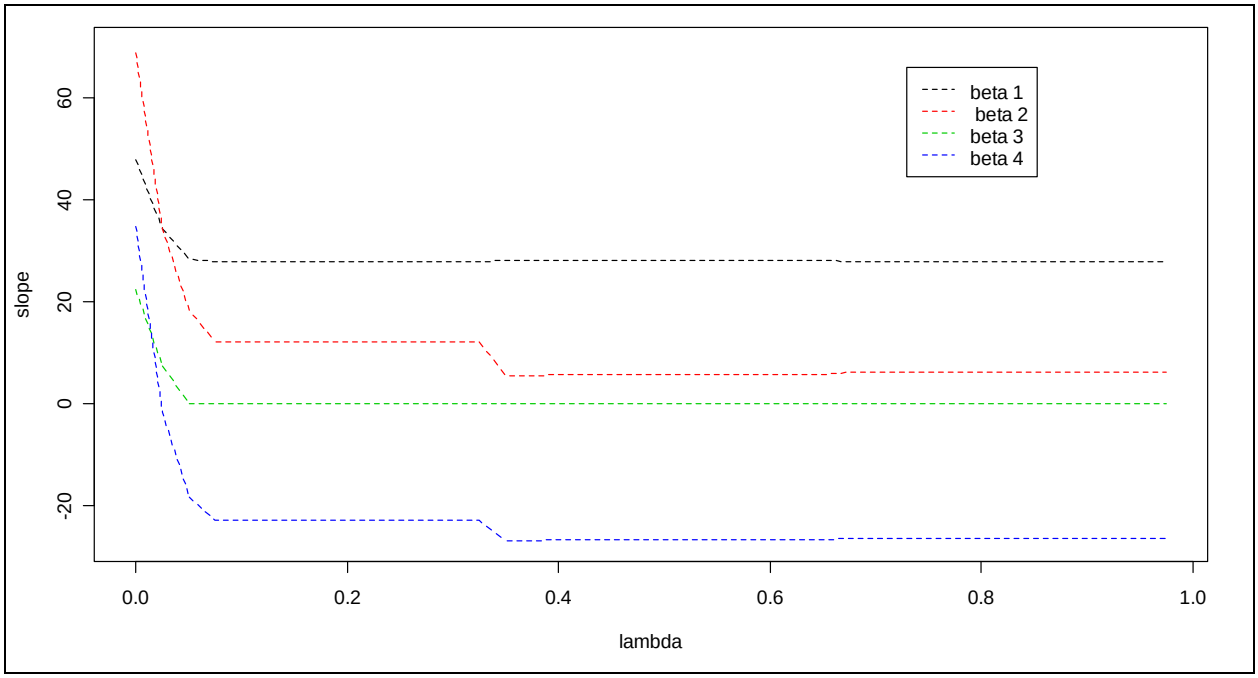


Figure 7.2: Lasso profile for RQ Lasso variable selection procedure at $\tau = 0.5$ with optimal value, $\lambda = 0.275$

The resulting parameter estimates from both the RQ Lasso and the Lars-Lasso are shown in table 7.3 below. The information in the ES can be based on fewer observations, p. As a consequence, we recommend that optimal shrinkage parameter s (see equation (4.9.3)) in the Lars-Lasso procedure should be based on the full data set rather than ES data set. The RQ Lasso procedure results in only β_3 being shrunk to zero. However, the other coefficients, β_2 and β_4, corresponding to predictors involved in a large correlation coefficient, are hugely shrunk in magnitude with β_4 changing sign from a positive one to a negative one.

The Lars-Lasso results in β_3 being shrunk to zero at the full data set and both β_2 and β_3 being shrunk to zero at the ES corresponding to the RQ at $\tau = 0.5$.

Remark: Note that the effect of collinearity on RQ estimation can be different from that on OLS estimation as evidenced by the existence of opposite signs in β_4.

Table 7.3: Comparison of the RQ Lasso estimates to the Lars-Lasso estimates

	RQ (τ =0.5)		OLS		
	RQ	RQ Lasso	OLS(n =13)	Lars-Lasso (n =13)	Lars-Lasso (n = p =5) s
β_0	95.731	9.606E+01	95.423	95.423	95.731
β_1	47.976	2.785E+01	31.607	11.161	10.496
β_2	68.985	1.217E+01	27.500	7.371	0.000
β_3	22.352	-3.125E-07	2.261	0.000	0.000
β_4	34.826	-2.306E+01	-8.353	-10.146	-14.244

Note: The column denoted $n = p$ (Lars-Lasso) represents the variable selection based on the p observations in the ES corresponding the RQ at τ =0.5.

It is important to find out if the same variables selected at $\tau = 0.5$ will also be selected at different τ levels. For this purpose it is convenient to consider the Hald data set since it is smaller. The variable selection picture from the RQ Lasso for all the τ levels is displayed in table 7.4 below. The parameter estimate β_3 is never shrunk to zero up to the 3rd τ level. However, it is consistently shrunk to zero thereafter at the other remaining τ levels. In some cases the RQ Lasso parameter estimates have opposite signs to the usual RQ parameter estimates. Actually, β_3 at $\tau = 0.253$ and β_4 at all τ levels have opposite signs.

Table 7.4: RQ Lasso parameter estimates for all τ levels for the Hald data set

τ	λ	β_0	β_1	β_2	β_3	β_4
0.230	0.000	93.763	29.801	36.186	-0.739	0.177
	0.400	93.780	17.624	1.193E+01	-9.346E+00	-2.398E+01
0.253	0.000	93.918	34.136	44.537	3.315	8.495
	0.400	93.780	17.624	1.193E+01	-9.346E+00	-2.398E+01
0.337	0.000	94.272	42.869	59.505	11.159	24.248
	0.825	94.631	24.850	1.918E+01	7.381E-10	-1.485E+01
0.426	0.000	94.860	45.352	65.322	16.003	30.295
	0.125	95.345	28.433	6.513E+00	-4.332E-11	-2.614E+01
0.428	0.000	95.434	47.437	69.385	20.472	34.750
	0.125	95.345	28.433	6.513E+00	1.856E-11	-2.614E+01
0.514	0.000	95.974	42.040	53.977	15.839	18.902
	0.275	96.123	27.888	1.292E+01	-2.701E-08	-2.211E+01
0.547	0.000	96.206	36.383	39.666	9.631	3.720
	0.700	96.123	27.888	1.292E+01	-2.665E-08	-2.211E+01
0.752	0.000	96.409	31.423	27.132	4.189	-9.582
	0.875	96.354	28.022	1.587E+01	-1.946E-09	-1.838E+01
0.824	0.000	96.714	24.599	8.359	-3.611	-29.301
	0.950	96.939	26.760	8.199E+00	-1.677E-08	-1.968E+01
0.860	0.000	98.142	18.418	-34.951	-17.214	-73.569
	0.975	107.900	12.041	3.101E-09	-2.192E-08	-6.940E-07

It is important to see the Lars-Lasso estimation picture at all the τ levels also. The variable selection picture for the Lars-Lasso procedure is displayed in table 7.5 below. The overall variable selection picture is that the Lars-Lasso procedure is greedier than the RQ Lasso procedure. The parameter estimates β_3 and β_4 at 2 extreme τ levels and β_4 is shrunk to zero at the 2^{nd} τ level. At all the other remaining τ levels β_2 and β_3 are shrunk to zero.

Table 7.5: Lars-Lasso parameter estimates for all τ levels for the Hald data set, s

τ	s	β_0	β_1	β_2	β_3	β_4
0.230	0.000	93.795	30.516	37.733	-0.056	1.718
	8.278	91.481	13.075	22.166	0.000	0.000
0.253	0.000	93.987	36.179	48.377	5.217	12.319
	4.251	92.749	1.660	19.897	-4.950	0.000
0.337	0.000	94.277	42.865	59.475	11.170	24.239
	8.278	98.098	9.678	0.000	0.000	-14.559
0.426	0.000	95.285	47.167	69.586	19.528	34.712
	8.278	100.204	7.469	0.000	0.000	-19.512
0.428	0.000	95.731	47.976	68.985	22.352	34.826
	8.278	97.848	10.496	0.000	0.000	-14.244
0.514	0.000	96.204	36.421	39.769	9.673	3.827
	8.278	97.283	10.280	0.000	0.000	-15.086
0.547	0.000	96.259	34.416	34.361	7.457	-1.775
	8.278	94.331	13.672	0.000	0.000	-11.412
0.752	0.000	96.270	34.812	35.700	7.907	-0.490
	8.278	96.604	10.710	0.000	0.000	-12.831
0.824	0.000	96.680	24.804	10.395	-3.075	-27.340
	8.278	94.350	9.538	0.000	0.000	-13.090
0.860	0.000	96.850	23.779	0.190	-5.762	-37.170
	8.278	91.929	0.000	26.871	0.000	0.000

In the next subsection we give some conclusions.

7.4.1 Conclusions on the lasso procedures

Although one of the shortcomings of the lasso procedure is that it is a data driven technique, it is applicable in various research areas (see *e.g.* Tibshirani, 1997). In this section we gave some data driven results on variable selection based on lasso shrinkage procedures. We plan to pursue these issues in our future research. However, much more research needs to be done on this topic in order to build a clearer picture. In particular, there are still more challenges in dealing with type B leverage points. It does seem that there is a potential for more success in the use of shrinkage procedures for variable selection in the RQ scenario even in the presence collinearity influential points (type A leverage points). This is expected since shrinkage procedures are designed to deal with collinearities which are concealed in the presence of type B leverage points.

We claim some degree of success since at least one of the variables involved in the pairs with high correlation coefficients was heavily shrunk or excluded from the chosen parsimonious model. Also, the coefficients estimated with wrong signs were often heavily shrunk or shrunk

to zero and/or estimated with the correct sign. However, there is need for further investigation since the two lasso procedures do not always coincide. It seems the RQ Lasso is more stable (robust) in the sense that the set of variables chosen (not shrunk to zero) does not change drastically from one τ level to another.

7.5 Overall conclusions of the thesis

This thesis contributed to further understanding the three-tier relationship amongst OLS, ES and RQ procedures. There is an inherent relationship amongst the three. The one between OLS and ES procedures has been fairly widely explored (see Hawkins *et al.*, 1984) while the one between ES and RQ procedures has been noted almost "casually" in the literature (see Koenker and Bassett, 1978; Mayo and Gray, 1997). Using this three-tier relationship we investigated the problems of collinearity, leverage and outliers in the RQ scenario. However, because of the complexity and interdependence between the different aspects, the "solutions" are often situation dependent, and it is not a simple matter to clear these issues in general. The thesis did provide some general guidelines and conclusions where possible. Clearly, much research still needs to be done on many of these issues.

Chapter 2 was devoted to ES regression tracing its origins as far back 1755 before the advent of OLS regression. We elaborated on the history of regression in general, how and why ES regression lost acceptance to OLS regression, its subsequent re-emergence and its involvement in the building of various OLS estimators and RQ based ones. Based on this foundation further ES-OLS relations were deduced. The existing relations and the ones deduced were then used to address the problems of collinearity, leverage and outliers arising in the RQ situation.

RQs were dealt with in Chapter 3. The chapter considered the location model and elaborated on how RQs can be viewed as extensions of order statistics, to the linear model. Also, computational procedures and software were considered in this chapter as well as the competitive advantage RQs have. The problem of the unboundedness of RQs' influence functions in the X space was mentioned and its proposed remedy (in the literature) of using bounded influence RQs. Also, the problem of the crossing of RQ hyperplanes was considered. The causes of the crossing of RQ hyperplanes was elaborated on. In Chapter 5 we clearly showed that leverage points are responsible for the crossing of RQ hyperplanes.

Existing regression diagnostics as well as new ones were considered in Chapter 4. Also, we briefly discussed variable selection procedures. We first considered the single case collinearity, leverage and outlier diagnostics as a preliminary step to considering the multiple case ones. Existing relationships between single case diagnostics and multiple case ones as well as new ones were considered. Although we mainly focused on developing RQ (multiple case) diagnostics we also showed that some existing multiple case diagnostics are related to their single case counterparts via the ERW. Furthermore, we showed that the ERW itself can either be viewed (used) as a collinearity diagnostic or a leverage one. It is important to note that some leverage points induce or hide collinearity. Hence, it is recommended that the problems of collinearity and leverage be addressed together. We proposed RQ predictive weighted leverage as an analogue to the single case one.

Although prediction can only meaningfully be done at the central RQ (see section 1.1) we considered prediction for all the RQ levels in order to have a holistic view. Also, influential diagnostics comprise a leverage component and a residual (prediction) component. Multiple case influential diagnostics that suited the RQ scenario are those that involve determinants, since the ERW which is the leverage part, is a function of determinants. We mainly considered the covariance ratio as an influence diagnostic. Lastly, in this chapter we briefly considered the lasso shrinkage technique as a more appealing procedure in the RQ scenario.

Our original idea was to extend the single case diagnostics to the multiple case one. However, the cut-off value of the single case diagnostics could not be directly extended to the RQ case. Therefore in Chapter 5 we determined the cut-off values using simulation studies. In order validate the performance of RQ case diagnostics and their cut-off values we applied them to some standard data sets in the literature in Chapter 6. Their performance was found to be satisfactory.

In Chapter 7 we concluded by giving some conclusions on the regression diagnostics as well as potential areas of further research.

REFERENCES

AITKEN A.G. (1964): *Determinants and Matrices*. Oliver and Boyd, Edinburgh.

ANDREWS D.F. AND PREGIBON D. (1978): Finding the outliers that matter. *Journal of the Royal Statistical Society,* Series B 40, 85-93.

ASKIN R.G. AND MONTGOMERY D.C. (1980): Augmented robust estimators. *Technometrics*, 22, 333-341.

ATKINSON A.C. (1981): Likelihood ratios, posterior odds and information criteria. *Journal of Econometrics,* 16, 15-20.

BARRETT B.E. AND GRAY B.J. (1992): Efficient Computation of Subset Influence in Regression. *Journal of Computational and Graphical Statistics*, 1, 271-286.

BARRETT B.E. AND GRAY J.B. (1995): Computation of determinantal subset influence in regression. *Statistics and Computing,* 6, 131-138.

BARRETT B.E. AND GRAY J.B. (1997a): On the use of robust diagnostics in least squares regression analysis. *Proceedings of the Statistical Computing Section, The American Statistical Association*, 130-135.

BARRETT B.E. AND GRAY J.B. (1997b): Leverage, residual and interaction diagnostics for subsets of cases in least squares regression. *Computational Statistics and Data Analysis,* 26, 39-52.

BARRETT B.E. AND LING R.F. (1992): General Classes of Influence Measures for Multivariate Regression. *Journal of the American Statistical Association,* 87, 184-191.

BARRODALE I. AND ROBERTS F. D. K. (1973): An improved algorithm for discrete L_1 linear approximation. *SIAM Journal on Numerical Analysis,* 10, 839-848.

BARRODALE I. AND ROBERTS F.D.K. (1974): Algorithm 478: Solution of an overdetermined system of equations in the L_1 norm. *Communications of the Association for Computing Machinery, 17, 319-320.*

BASSETT G. W. AND CHEN H. (2001): Portfolio style: return-based attribution using quantile regression. *Empirical Economics*, 26, 293-305.

BELSLEY D.A. (1991): *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. John Wiley & Sons, New York.

BELSLEY D.A., KUH E. AND WELSCH R.E. (1980): *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons, New York.

BERK K.N. (1977): Tolerance and Condition in Regression Computations. *Journal of the American Statistical Association,* 72 , 863-866.

BUCHINSKY M. (1995): Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics*, 65, 109-154.

CADE B.S. AND NOON B. R. (2003): A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment,* 1, 412-420.

CHATTERJEE S. AND HADI A.S. (1986): Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science,* 1, 379-416.

CHATTERJEE S. AND HADI A.S. (1988): *Sensitivity Analysis in Linear Regression*. John Wiley & Sons, New York.

CHEN M.H., SHAO Q.M. AND IBRAHIM J.G. (2001): *Monte Carlo Methods in Bayesian Computation*. Springer, New York.

COLE T.J. AND GREEN P.J. (1992): Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, 11,1305-19.

COOK R.D. (1977): Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

COOK R.D. AND WEISBERG S. (1982): *Residuals and influence in regression*. Chapman and Hall, New York.

CROUX C., ROUSSEEUW P.J. AND HÖSSJER O. L. A. (1994): Generalized S-Estimators. *Journal of the American Statistical Association,* 89, 1271-1281.

DE JONGH P. J., DE WET T. AND WELSH A. H. (1988): Mallows-Type Bounded-Influence Regression Trimmed Means. *Journal of the American Statistical Association*, 83, 805-810.

EFRON B., JOHNSTONE I., HASTIE T. AND TIBSHIRANI R. (2004): Least angle regression. *Annals of Statistics*, 32, 407-499.

ELLIS S.P. AND MORGENTHALER S. (1992): Leverage and breakdown in L1 regression. *Journal of the American Statistical Association,* 87, 143-148.

FOX J. (1991): *Regression Diagnostics*. Sage Publications, Thousand Oaks, California.

FOX J. AND MONETTE G. (1992): Generalized collinearity diagnosis. *Journal of the American Statistical Association,* 87, 178-83.

FREUND R.J. AND WILSON W.J. (1998): *Regression Analysis: Statistical Modeling of a Response Variable*. Academic Press, San Diego.

GRAY B.J. AND LING R.F. (1984): K-Clustering as a Detection Tool for Influential Subsets in Regression. *Technometrics*. 26, 305-318.

GRAYBILL F.A. (1969): *Introduction to matrices with applications to Statistics*. Wadsworth, Belmont, California.

GREEN E., TUIL D. S. AND ALBAUM G. (1988): *Research for Marketing Decisions*, Fifth Edition. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

GREENE W.H. (1990): *Econometric Analysis*. Macmillan Publishing Company, New York.

GUNST R.F. (1983): Regression Analysis With Multicollinear Predictor Variables: Definition, Detection, and Effects. *Communications in Statistics, Part A-Theory and Methods*, 12, 2217-2260.

GUNST R.F. AND MASON R.L. (1980): *Regression Analysis and Its Application: A Data Oriented Approach*. Marcel Dekker, New York.

HADI A. S. AND SIMONOFF J. S. (1993): Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, 88, 1264-1272.

HADI A. S. AND WELLS M.T. (1990): Assessing the effects of multiple rows on the condition number of a matrix. *Journal of the American Statistical Association*, 85, 786-792.

HAIR J. F., JR., ANDERSON R. E., TATHAM R. L. AND BLACK W. C.(1995): *Multivariate Data Analysis*, Third Edition. Macmillan Publishing Company, New York.

HAMPEL F. (1985): The breakdown points of the mean combined with some rejection rules. *Technometrics*, 27, 95–107.

HARVILLE D.A.(1997): *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

HASTIE T., TIBSHIRANI R. AND FRIEDMAN J. (2001): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.

HAWKINS, D.M. (1993): The Accuracy of Elemental Set Approximations for Regression. *Journal of the American Statistical Association*, 88, 580-589.

HAWKINS D.M., BRADU D. AND KASS G.V. (1984): Location of several outliers in multiple regression data using elemental sets. *Technometrics*, 26, 197-208.

HAWKINS, D.M. AND OLIVE D.J. (1999): Improved feasible solution algorithms for high breakdown estimation. *Computational Statistics and Data Analysis*, 30, 1-11.

HAWKINS D.M. AND OLIVE D.J. (2002): Inconsistency of Resampling Algorithms for High-Breakdown Regression Estimators and a New Algorithm. *Journal of the American Statistical Association*, 97, 136-159.

HENDRICKS, W. AND KOENKER, R. (1992): Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 93, 58–68.

HE X. (1997): Quantile curves without crossing. *The American Statistician* , 51, 186-192.

HOAGLIN D.C. AND WELSCH R.E. (1978): The Hat Matrix and Anova. *The American Statistician*, 32, 17-22.

HOCKING R.R. AND PENDLETON O.J. (1983): The Regression dilemma. *Communications in Statistics Part A-Theory and Methods*, 12, 497-527.

HOERL A. E. (1962): Applications of ridge analysis to regression problems. *Chemical Engineering Progress, 58*, 54-59.

HOERL A.E. AND KENNARD R.W. (1970): Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* , 12, 55-67.

HOERL A.E. AND KENNARD R. W. (1980): Ridge regression, advances, algorithms, and applications, *American Journal of Mathematical and Management Sciences*, 1, 5-83.

HOERL A.E., KENNARD R.W. AND BALDWIN K.F. (1975): Ridge regression: some simulations. *Communications in Statistics*, 4, 105-123.

HÖSSJER O. (1994): Rank-Based Estimates in the Linear Model with High Breakdown Point. *Journal of the American Statistical Association*, 89, 149-158.

HUBER P.J. (1981): *Robust Statistics*, Second Edition. John Wiley & Sons, New York.

JOHNSON R. A. AND WICHERN D. W. (1988): *Applied Multivariate Statistical Analysis*, Second Edition Prentice Hall, Englewood Cliffs, New Jersey.

JOHNSTON J. (1984): *Econometric Methods*, Third Edition. McGraw-Hill Publishing Company, New York.

KMENTA J. (1986): *Elements of Econometrics*, Second Edition. Macmillan Publishing Company, New York.

KOENKER R. (1984): A note on L-estimates for linear models. *Statistics and Probability Letters*, 2, 323-325.

KOENKER R. (2005): *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, New York.

KOENKER R. AND BASSETT G. JNR (1978): Regression Quantiles. *Econometrica*, 46, 33-50.

KOENKER R.W. AND D'OREY V. (1987): Algorithm AS 229: Computing Regression Quantiles, *Applied Statistics*, 36, 383-393.

KOENKER R. AND D'OREY V. (1994): Remark AS R92: A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores. *Applied Statistics*, 43, 410-414.

KOENKER R. AND GELING O. (2001): Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, 96, 458-468.

KOENKER R. AND HALLOCK K.F. (2001): Quantile Regression, An Introduction. *Journal of Economic Perspectives*, 15, 143-156.

KOENKER R. AND PARK B.J. (1996): An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71, 265-283.

LAURIDSEN S. (2000): Estimation of value at risk by extreme value methods. *Extremes*, 3, 107–144.

LEHMANN D. R., GUPTA, S. AND STECKEL, J. (1988): *Marketing Research*. Addison-Wesley, Reading, Massachusetts.

LI Y., LIU Y. AND ZHU J. (2007): Quantile Regression in Reproducing Kernel Hilbert Spaces. *Journal of the American Statistical Association*, 102, 255-268.

MALLOWS C.L. (1973): Influence functions. Unpublished paper presented at a Conference on Robust Regression held at Cambridge, MA. Sponsored by the National Bureau for Economic Research.

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

MARQUARDT D.W. (1980): You should standardize the predictor variable in your regression models (discussion of a paper by G. Smith and F. Campbell). *Journal of the*

American Statistical Association, 75, 87-91.

MARQUARDT D.W. AND SNEE R.D. (1975): Ridge regression in practice. *The American Statistician*, 29, 3-20.

MASON R.L. AND GUNST R.F. (1985): Outlier-Induced Collinearities. *Technometrics*, 27, 401-408.

MAYO M.S. AND GRAY J.B. (1997): Elemental subsets: The Building Blocks of Regression. *The American Statistician,* 51, 122-129.

McDONALD G.C. AND GALARNEAU D. I. (1975): A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association,* 70, 407-416.

MELA C. F. AND KOPALLE P. (2002): The Impact of Collinearity on Regression Analysis: The Asymmetric Effect of Positive and Negative Correlations. *Applied Economics*, 34, 667-677.

MONTGOMERY D.C. AND PECK E.A. (1982): *Introduction to linear regression analysis*. John Wiley & Sons, New York.

MOSTELLER F. AND TUKEY J.W. (1977): *Data Analysis and Regression*. Addison - Wesley, New York.

MU Y. AND HE X. (2007): Power Transformation Toward a Linear Regression Quantile. *Journal of the American Statistical Association,* 102, 269-279.

PANDEY, G. R. AND NGUYEN, V. T. (1999): A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225, 92-101.

PORTNOY S. AND KOENKER R. (1997): The Gaussian hare and the Laplacian tortoise: computability of squared-error estimators (with discussion). *Statistical Science*, 12, 279-300.

RAVEH A. (1985): On the use of the Inverse of the Correlation Matrix in Multivariate Data Analysis. *The American Statistician*, 39, 39-42.

RONCHETTI E., FIELD C. AND BLANCHARD W. (1997): Robust Linear Model Selection by Cross-Validation. *Journal of the American Statistical Association,* 92, 1017-1023.

ROUSSEEUW P.J. (1984): Least median of squares regression. *Journal of the American Statistical Association*, 79, 871–880.

ROUSSEEUW P.J. AND HUBERT M. (1997): Recent developments in PROGRESS. *L1-Statistical Procedures and Related Topics* (edited by Y. Dodge). Institute of Mathematical Statistics Lecture Notes-Monograph Series, 31, 201-214.

ROUSSEEUW P.J. AND VAN ZOMEREN B.C. (1990): Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85, 633-639.

ROYSTON P. AND ALTMAN D.G. (1994): Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling. *Applied Statistics*, 43, 429-467.

RUPPERT D. AND CARROLL R.J. (1980): Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association*, 75, 828-838.

SAS INSTITUTE (1996): *SAS Manual*. SAS Institute, Cary, North Carolina.

SEARLE S.R. (1982): *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York.

SENGUPTA D. AND BHIMASANKARAM (1997): On the Roles of Observations in Collinearity in the Linear model. *Journal of the American Statistical Association*, 92, 1024-1032.

SEN P.K. (1968): Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, 63, 1379-1389.

SHEYNIN O. B. (1973): R. J. Boscovich's work on probability. *Archive for History of Exact Sciences*, 9, 306-324.

SILVEY S.D. (1969): Multicollinearity and imprecise estimation. *Journal of the Royal Statistical Society*, Series B, 31, 539-552.

STIGLER S. (1986): *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press, Cambridge, Massachusetts.

TAYLOR J. (1999): A quantile regression approach to estimating the distribution of multiperiod returns. *Journal of Derivatives*, 24, 64-78.

THEIL H. (1950): A rank-invariant method of linear and polynomial regression analysis, 1,2, and 3. *Nederlandse Akademie van Wetenschappen*. Proc., 53, 386-392, 521-525, 1397-1412.

TIBSHIRANI R. (1996): Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, Series B, 58, 267-288.

TIBSHIRANI R. (1997): The lasso method for variable selection in the Cox model. *Statistics in Medicine*, 16, 385-395.

VAN VUUREN J.O. (1998): *Collinearity Influential Observations and Outliers in the Multiple Linear Regression Model*. PhD Thesis, University of Stellenbosch.

VELLEMAN P.F. AND WELSCH R.E. (1981): Efficient Computing of Regression Diagnostics. *The American Statistician*, 35, 234-242.

WANG H. AND HE X. (2007): Detecting differential Expressions in GeneChip Microarray studies: A Quantile Approach. *Journal of the American Statistical Association*, 102, 104-112.

WILLAN A. R. AND WATTS D.G. (1978): Meaningful Multicollinearity Measures. *Technometrics*, 20, 407-412.

WITTINK D.K. (1988): *The Application of Regression Analysis*. Simon & Schuster, Needham Heights, Massachusetts.

YANG S. (1999): Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94, 137-145.

YU K., LU Z. AND STANDER J. (2003): Quantile regression: applications and current research areas. *The Statistician*, 52, 331-350.

ZHAO Q. (2000): Restricted Regression Quantiles. *Journal of Multivariate Analysis*, 72, 78-99.

APPENDIX A: SOME THEOREMS AND PROOFS

The following theorem (see Graybill, 1969) is central to the proof of **Theorem 4.1**.

Theorem A.1: Let $\mathbf{B}$ be an $n \times n$ matrix that is partitioned as follows:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \text{ where } \mathbf{B}_{ij} \text{ has size } n_i \times n_j,\ i, j = 1, 2 \text{ and } n_1 + n_2 = n.$$

Suppose $|\mathbf{B}| \neq 0, |\mathbf{B}_{11}| \neq 0$ *and* $|\mathbf{B}_{22}| \neq 0,$ then

(i) $|\mathbf{B}| = |\mathbf{B}_{11}||\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}|$ or $|\mathbf{B}| = |\mathbf{B}_{22}||\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}|$ and

(ii) $$\mathbf{B}^{-1} = \begin{pmatrix} \left[\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\right]^{-1} & -\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\left[\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\right]^{-1} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\left[\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\right]^{-1} & \left[\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\right]^{-1} \end{pmatrix}.$$

Proof of Theorem 4.1:

We invoke the principle of Mathematical Induction for the proof of this theorem, starting with $p = k + 1 = 2$ predictors.

Now taking $k = 1,$ we have

$$\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{pmatrix}.$$

Letting $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \begin{pmatrix} n & a & b \\ a & c & d \\ b & d & e \end{pmatrix}$, then

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = 2abd + cne - b^2c - d^2n - a^2e$$

and

$$2abd = -n\left(d - \frac{ab}{n}\right)^2 + \frac{a^2b^2}{n} + nd^2.$$

Hence,

$$\begin{aligned}
\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| &= -n\left(d-\frac{ab}{n}\right)^2+\frac{a^2b^2}{n}+nd^2+cne-b^2c-d^2n-a^2e \\
&= -n\left(d-\frac{ab}{n}\right)^2-a^2\left(e-\frac{b^2}{n}\right)+c\left(ne-b^2\right) \\
&= -ns_{12}'^2-a^2s_2'^2+cns_2'^2 \\
&= \left(cn-a^2\right)s_2'^2-ns_{12}'^2 \\
&= ns_1'^2s_2'^2-ns_{12}'^2 \\
&= ns_1'^2s_2'^2\left(1-\frac{ns_{12}'^2}{ns_1'^2s_2'^2}\right) \\
&= ns_1'^2s_2'^2(1-c_{12}) \\
&= ns_1'^2s_2'^2\left|\mathbf{C}\right|,
\end{aligned}$$

where

$$s_j'^2=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2,\ j=1,...,k \text{ and } s_{jl}'=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{il}-\bar{x}_l),\ l,j=1,...,k \text{ and } l\neq j.$$

Therefore the formula

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = n.s_{p-1}'^2.s_{p-2}'^2...s_1'^2.\left|\mathbf{C}\right| \ ,$$

is true for $k=1$.

Now, assuming the formula is true for $k=k^*$ where $k^*\in\square$, then

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = n.s_{k^*}'^2.s_{k^*-1}'^2...s_1'^2.\left|\mathbf{C}\right|.$$

It is required to prove that the formula is true for $k=k^*+1$

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = \begin{vmatrix}
n & \sum x_1 & \sum x_2 & \cdots & \sum x_{k^*+1} & \sum x_{k^*+2} \\
\sum x_1 & \sum x_1^2 & \sum x_1x_2 & \cdots & \sum x_1x_{k^*+1} & \sum x_1x_{k^*+2} \\
\sum x_2 & \sum x_1x_2 & \sum x_2^2 & \cdots & \sum x_2x_{k^*+1} & \sum x_2x_{k^*+2} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
\sum x_{k^*+2} & \sum x_1x_{k^*+2} & \sum x_2x_{k^*+2} & \cdots & \sum x_{k^*+1}x_{k^*+2} & \sum x_{k^*+2}^2
\end{vmatrix}.$$

Applying Theorem A.1(i), we have

$$\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right| = \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|_{(k^*+2)}\left|\sum x_{k^*+2}^2-\mathbf{a}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_{(k^*+2)}^{-1}\mathbf{a}\right|$$

where

$$\mathbf{a}' = \left[\sum x_{k^*+2} \quad \sum x_1 x_{k^*+2} \quad \sum x_2 x_{k^*+2} \quad \ldots \quad \sum x_{k^*+1} x_{k^*+2} \right]$$

and $(k^* + 2)$ denotes the design matrix that excludes the $(k^* + 2)^{th}$ predictor.

The R.H.S. can be written as

$$n.s'^2_{k^*+2}.s'^2_{k^*+1}...s'^2_1.|\mathbf{C}| = n.s'^2_{k^*+1}\left(n.s'^2_{k^*+1}.s'^2_{k^*}...s'^2_1\right).|\mathbf{C}|$$
$$= \left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|_{(k^*+2)} . \frac{|\mathbf{C}|}{|\mathbf{C}|_{(k^*+2)}}.$$

Now, we need to prove that

$$\frac{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|_{(k^*+2)}} = s'^2_{k^*+1} \frac{|\mathbf{C}|}{|\mathbf{C}|_{(k^*+2)}} .$$

The determinant of the correlation matrix can be expressed as

$$|\mathbf{C}| = \begin{vmatrix} 1 & c_{1,2} & c_{1,3} & \ldots & c_{1,k^*+1} & c_{1,k^*+2} \\ c_{2,1} & 1 & c_{2,3} & \ldots & c_{2,k^*+1} & c_{2,k^*+2} \\ c_{3,1} & c_{3,2} & 1 & \ldots & c_{3,k^*+1} & c_{3,k^*+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{k^*+1,1} & c_{k^*+1,2} & & \ldots & 1 & c_{k^*+1,k^*+2} \\ c_{k^*+2,1} & c_{k^*+2,2} & & \ldots & c_{k^*+2,k^*+1} & 1 \end{vmatrix}.$$

Applying Theorem A.1(i) again, the correlation matrix can be expressed as

$$|\mathbf{C}| = |\mathbf{C}|_{(k^*+2)} \left| 1 - \mathbf{c}'_{k^*+2} \mathbf{C}^{-1}_{(k^*+2)} \mathbf{c}_{k^*+2} \right|$$

where

$$\mathbf{c}'_{k^*+2} = \left[c_{k^*+2,1} \quad c_{k^*+2,2} \quad \cdots \quad c_{k^*+2,k^*+1} \right].$$

Now we need to prove that the L.H.S =R.H.S., *i.e.*,

$$\sum x^2_{k^*+2} - \mathbf{a}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}_{(k^*+2)} \mathbf{a} = s'^2_{k^*+2}.(1 - \mathbf{c}'_{k^*+2} \mathbf{C}^{-1}_{(k^*+2)} \mathbf{c}_{k^*+2}) \qquad (A.4.1).$$

Let $\mathbf{S}'_{(k^*+2)}$ be the matrix whose entries are s'_{jl}, $l, j = 1,...,k+1$, *i.e.*,

$$\mathbf{S}_{(k^*+2)} = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_{k^*+1} \\ \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_2 x_{k^*+1} \\ \vdots & \vdots & & \vdots \\ \sum x_1 x_{k^*+2} & \sum x_2 x_{k^*+2} & \dots & \sum x_{k^*+2}^2 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} \sum x_1 \\ \sum x_2 \\ \vdots \\ \sum x_{k^*+2} \end{bmatrix} \begin{bmatrix} \sum x_1 & \sum x_2 & \dots & \sum x_{k^*+2} \end{bmatrix}$$

and

$$\mathbf{B}_{22} = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_{k^*+1} \\ \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_2 x_{k^*+1} \\ \vdots & \vdots & & \vdots \\ \sum x_1 x_{k^*+2} & \sum x_2 x_{k^*+2} & \dots & \sum x_{k^*+2}^2 \end{bmatrix},$$

$$\mathbf{B}_{12} = \begin{bmatrix} \sum x_1 & \sum x_2 & \dots & \sum x_{k^*+2} \end{bmatrix}$$

and hence, we can express $\mathbf{S}_{(k^*+2)}$ as

$$\mathbf{S}_{(k^*+2)} = \mathbf{B}_{22} - \frac{1}{n}\mathbf{B}_{12}\mathbf{B}_{21}, \text{ since } \mathbf{B}_{12}' = \mathbf{B}_{21}.$$

Invoking Theorem A.1 (ii), we have

$$(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}_{(k^*+2)} = \begin{pmatrix} \left[n - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\right]^{-1} & -\frac{1}{n}\mathbf{B}_{12}\left[\mathbf{B}_{22} - \mathbf{B}_{21}\frac{1}{n}\mathbf{B}_{12}\right]^{-1} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\left[n - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\right]^{-1} & \left[\mathbf{B}_{22} - \mathbf{B}_{21}\frac{1}{n}\mathbf{B}_{12}\right]^{-1} \end{pmatrix}.$$

$$= \begin{pmatrix} \left[n - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\right]^{-1} & -\frac{1}{n}\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)} \\ -\frac{1}{n}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} & \mathbf{S}^{-1}_{(k^*+2)} \end{pmatrix}$$

From the well known result for symmetrical matrices (see Harville, 1997)

$$(\mathbf{R} + \mathbf{WTU})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}(\mathbf{T} + \mathbf{UR}^{-1}\mathbf{W})^{-1}\mathbf{UR}^{-1}$$

where $\mathbf{R}$ is an $n \times n$ matrix, $\mathbf{W}$ is an $n \times m$ matrix, $\mathbf{T}$ is an $m \times m$ matrix and $\mathbf{U}$ is an $m \times n$ matrix, we have

$$(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}_{(k^*+2)} = \begin{pmatrix} \frac{1}{n} - \frac{1}{n}\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} & -\frac{1}{n}\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)} \\ -\frac{1}{n}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} & \mathbf{S}^{-1}_{(k^*+2)} \end{pmatrix}.$$

We can write

$$\mathbf{a}' = \begin{bmatrix} \sum x_{k^*+2} & \sum x_1 x_{k^*+2} & \sum x_2 x_{k^*+2} & \ldots & \sum x_{k^*+1} x_{k^*+2} \end{bmatrix}$$
$$= \begin{bmatrix} \sum x_{k^*+2} & \mathbf{b}' \end{bmatrix}$$

and $\sum x_{k^*+2}^2 - \mathbf{a}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}_{(k^*+2)}\mathbf{a}$ can be expressed as

$$\sum x_{k^*+2}^2 - \begin{bmatrix} \left(\frac{1}{n}\sum x_{k^*+2} - \frac{1}{n^2}\sum x_{k^*+2}\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} - \frac{1}{n}\mathbf{b}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} \right) \\ -\frac{1}{n}\sum x_{k^*+2}\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)} + \mathbf{b}'\mathbf{S}^{-1}_{(k^*+2)} \end{bmatrix}' \begin{bmatrix} \sum x_{k^*+2} \\ \mathbf{b} \end{bmatrix}.$$

This expression can be expressed as

$$\sum x_{k^*+2}^2 - \frac{1}{n}\left(\sum x_{k^*+2}\right)^2 + \frac{1}{n^2}\left(\sum x_{k^*+2}\right)^2 \mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} - \mathbf{b}'\mathbf{S}^{-1}_{(k^*+2)}\mathbf{b}$$

which can further can expressed as

$$s'^2_{k^*+2} + \frac{1}{n^2}\left(\sum x_{k^*+2}\right)^2 \mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} - \mathbf{b}'\mathbf{S}^{-1}_{(k^*+2)}\mathbf{b},$$

which is the L.H.S. of (*A*.4.1).

Now, we need to show that the R.H.S. is equal to this expression.

The correlation matrix can be expressed as

$$\mathbf{C}_{(k^*+2)} = \mathbf{D}_s^{-\frac{1}{2}}\mathbf{S}_{k^*+2)}\mathbf{D}_s^{-\frac{1}{2}}$$

implying that

$$\mathbf{C}^{-1}_{(k^*+2)} = \mathbf{D}_s^{\frac{1}{2}}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{D}_s^{\frac{1}{2}}$$

Where $\mathbf{D}_s = diag\left\{s'^2_j\right\}$, $s'^2_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$, $j = 1,...,k$ and $\mathbf{S}_{(k^*+2)}$ is the variance-covariance matrix based on the design matrix that excludes the $(k^*+2)^{th}$ predictor.

The $(k^*+2, j)^{th}$ entry of $\mathbf{C}$ is given by

$$c_{k^*+2,j} = \frac{s'_{k^*+2,j}}{s'_{k^*+2}s'_j}, \quad j = 1,2,...,k+1$$

and therefore

$$\mathbf{c}'_{k^*+2}\mathbf{D}_s^{\frac{1}{2}} = \frac{1}{s'_{k^*+2}}\begin{bmatrix} c_{k^*+2,1} & c_{k^*+2,2} & \cdots & c_{k^*+2,k^*+1} \end{bmatrix}$$

$$= \frac{1}{s'_{k^*+2}}\begin{bmatrix} \sum x_{k^*+2}x_1 - \frac{\sum x_{k^*+2}.\sum x_1}{n} & \cdots & \sum x_{k^*+2}x_{k^*+1} - \frac{\sum x_{k^*+2}.\sum x_{k^*+1}}{n} \end{bmatrix}$$

$$= \frac{1}{s'_{k^*+2}}\left[\mathbf{b}' - \frac{1}{n}\sum x_{k^*+2}\mathbf{B}_{12}\right].$$

Hence,

$$\mathbf{c}'_{k^*+2}\mathbf{C}^{-1}_{(k^*+2)}\mathbf{c}_{k^*+2} = \mathbf{c}'_{k^*+2}\mathbf{D}_s^{\frac{1}{2}}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{D}_s^{\frac{1}{2}}\mathbf{c}_{k^*+2}$$

$$= \frac{1}{s'_{k^*+2}}\left[\mathbf{b}' - \frac{1}{n}\sum x_{k^*+2}\mathbf{B}_{12}\right]\mathbf{S}^{-1}_{(k^*+2)}\left[\mathbf{b} - \frac{1}{n}\sum x_{k^*+2}\mathbf{B}_{21}\right]\frac{1}{s'_{k^*+2}}$$

$$= \frac{1}{s'^2_{k^*+2}}\left\{\mathbf{b}'\mathbf{S}^{-1}_{(k^*+2)}\mathbf{b} - \frac{1}{n^2}\left(\sum x_{k^*+2}\right)^2\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21}\right\}.$$

The R.H.S. of $(A.4.1)$ is given by

$$s'^2_{k^*+2}.(1 - \mathbf{c}'_{k^*+2}\mathbf{C}^{-1}_{(k^*+2)}\mathbf{c}_{k^*+2})$$

can be expressed as

$$s'^2_{k^*+2} + \frac{1}{n^2}\left(\sum x_{k^*+2}\right)^2\mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21} - \mathbf{B}_{12}\mathbf{S}^{-1}_{(k^*+2)}\mathbf{B}_{21}$$

which is equal to the L.H.S. of $(A.4.1)$.

Therefore the formula is true for all $k^* \in \square$. □

Proof of Theorem 4.4:

Proof of (i)

This is given by Hawkins *et al.* (1984) as follows:

$$e_{iJ} = \frac{\begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \mathbf{Y}_J & \tilde{\mathbf{X}}_J \end{vmatrix}}{\left|\tilde{\mathbf{X}}_J\right|}$$

and

$$e_i = Y_i - \tilde{\mathbf{x}}'_i\hat{\tilde{\boldsymbol{\beta}}} = Y_i - \tilde{\mathbf{x}}'_i(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y} = \frac{\begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix}}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|}$$

by Theorem A.1 (i) (Appendix A). The numerator in the last equality can be expressed as

$$\begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix} = \left| \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ & & \tilde{\mathbf{X}}' & & \end{pmatrix} \begin{pmatrix} \mathbf{Y} & \vdots & \tilde{\mathbf{X}} \end{pmatrix} \right|.$$

The one in the first row of the first factor matrix is at position i. The single $(p+1)(p+1)$ nonzero minors of this first factor are those that include column i and the p columns of any set $J \not\ni i$; the value of any such minor will be $\omega|\mathbf{X}'| = \omega|\mathbf{X}|$, where $\omega = \pm 1$. If we now take the same rows i and J in the second matrix factor, the corresponding minor will be

$$\omega \begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \mathbf{Y}_J & \tilde{\mathbf{X}}_J \end{vmatrix} = \omega |\tilde{\mathbf{X}}_J| (Y_i - \tilde{\mathbf{x}}'\tilde{\mathbf{X}}_J^{-1}\mathbf{Y}_J) = \omega |\tilde{\mathbf{X}}_J| e_{iJ},$$

with the same ω as the minor of the first factor.

Invoking the Cauchy- Binet Theorem (Theorem A.2, Appendix A), which gives the minors of a product of two matrices, one gets

$$\begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \mathbf{Y}_J & \tilde{\mathbf{X}}_J \end{vmatrix} = \sum_{J \not\ni i} \omega^2 |\tilde{\mathbf{X}}_J|^2 e_{iJ} = \sum_{J \not\ni i} |\tilde{\mathbf{X}}_J|^2 e_{iJ}.$$

The same theorem gives for $\mathbf{X}'\mathbf{X}$

$$|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}| = \sum_J |\tilde{\mathbf{X}}_J| |\tilde{\mathbf{X}}_J| = \sum_J |\tilde{\mathbf{X}}_J|^2,$$

from which substituting into (4.4.2) gives

$$e_i = \frac{\sum_{J \not\ni i} |\tilde{\mathbf{X}}_J|^2 e_{iJ}}{\sum_J |\tilde{\mathbf{X}}_J|^2} = \sum_{J \not\ni i} \omega_J e_{iJ}, \; 1 \le i \le n.$$

Here there is no harm in summing over all the $i's, 1 \le i \le n$ since $e_{iJ} = 0$ *for* $i \in J$.

Proof of (ii)

Here we have corrected the original result Hawkins *et al.* (1984).

The residual sum of squares is given by

$$SSE = \sum_{i=1}^{n} (Y_i - \tilde{\mathbf{x}}_i'\hat{\boldsymbol{\beta}})^2 = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y},$$

which can expressed as

$$SSE = \frac{\begin{vmatrix} \mathbf{Y}\mathbf{Y}' & \mathbf{Y}'\tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix}}{\left|\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right|}.$$

Again by Theorem A.2, the numerator can be expressed as

$$\begin{vmatrix} \mathbf{Y}\mathbf{Y}' & \mathbf{Y}'\tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix} = \begin{pmatrix} \mathbf{Y}' \\ \tilde{\mathbf{X}}' \end{pmatrix} \begin{pmatrix} \mathbf{Y} & \tilde{\mathbf{X}} \end{pmatrix}.$$

Now, taking the $(p+1)\times(p+1)$ minors along the first row of the first submatrix for the columns corresponding to each observation i and the rows corresponding to each observation i on the L.H.S, ensures that each resulting $p \times p$ ES is repeated $(p+1)$ times. This is so because the total number of minors are $\begin{pmatrix} n \\ p+1 \end{pmatrix}$, while the number of ESs that do not contain observation i is $\begin{pmatrix} n \\ p \end{pmatrix}\begin{pmatrix} n-p \\ 1 \end{pmatrix}$ since an ES can be chosen in $\begin{pmatrix} n \\ p \end{pmatrix}$ ways and i^{th} observation can be chosen from the remaining $n-p$ observations in $\begin{pmatrix} n-p \\ 1 \end{pmatrix}$ ways. Dividing $\begin{pmatrix} n \\ p \end{pmatrix}\begin{pmatrix} n-p \\ 1 \end{pmatrix}$ by $\begin{pmatrix} n \\ p+1 \end{pmatrix}$ shows that the L.H.S. must be divided by $(p+1)$. Invoking the Cauchy-Binet Theorem (Theorem A.2) requires that the $p \times p$ ESs be chosen without repetitions, *i.e.*,

$$\begin{vmatrix} \mathbf{Y}\mathbf{Y}' & \mathbf{Y}'\tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix} = \frac{1}{p+1}\sum_{J \ni i} \begin{vmatrix} Y_i & \mathbf{Y}_J' \\ \mathbf{x}' & \tilde{\mathbf{X}}_J \end{vmatrix} \begin{vmatrix} Y_i & \tilde{\mathbf{x}}' \\ \mathbf{Y}_J & \tilde{\mathbf{X}}_J \end{vmatrix} \qquad (A.4.2)$$

Each of the two minors appearing in the sum (A.4.2) equals

$$\left|\tilde{\mathbf{X}}_J\right|(Y_i - \tilde{\mathbf{x}}'\tilde{\mathbf{X}}_J^{-1}\mathbf{Y}_J) = \left|\tilde{\mathbf{X}}_J\right| e_{iJ},$$

and (A.4.2) becomes

$$\begin{vmatrix} \mathbf{Y}\mathbf{Y}' & \mathbf{Y}'\tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}'\mathbf{Y} & \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \end{vmatrix} = \frac{1}{p+1}\sum_J \sum_{J \ni i} \left|\tilde{\mathbf{X}}_J\right|^2 e_{iJ}^2 .$$

Thus dividing both sides by $\left|\mathbf{X}'\mathbf{X}\right|$ gives

$$SSE = \frac{\sum_J \omega_J PRESS_J}{p+1}, \text{ proving (ii).} \qquad \square$$

The following theorem, referred to as the **Cauchy-Binet formula** (Aitken, 1964, p. 86), generalizes the multiplicativity of the determinant (the fact that the determinant of a product of two square matrices is equal to the product of the two determinants) to non-square matrices.

Theorem A.2: Suppose $\mathbf{A}$ is an $p \times n$ matrix and $\mathbf{B}$ is an $n \times p$ matrix. If J is a subset of $\{1, ..., n\}$ with p elements, we write $\mathbf{A}_J$ for the $p \times p$ matrix whose columns are those columns of $\mathbf{A}$ that have indices from J. Similarly, we write $\mathbf{B}_J$ for the $p \times p$ matrix whose rows are those rows of $\mathbf{B}$ that have indices from J. The Cauchy-Binet formula then states

$$\left|\mathbf{AB}\right| = \sum_J \left|\mathbf{A}_J\right|\left|\mathbf{B}_J\right|$$

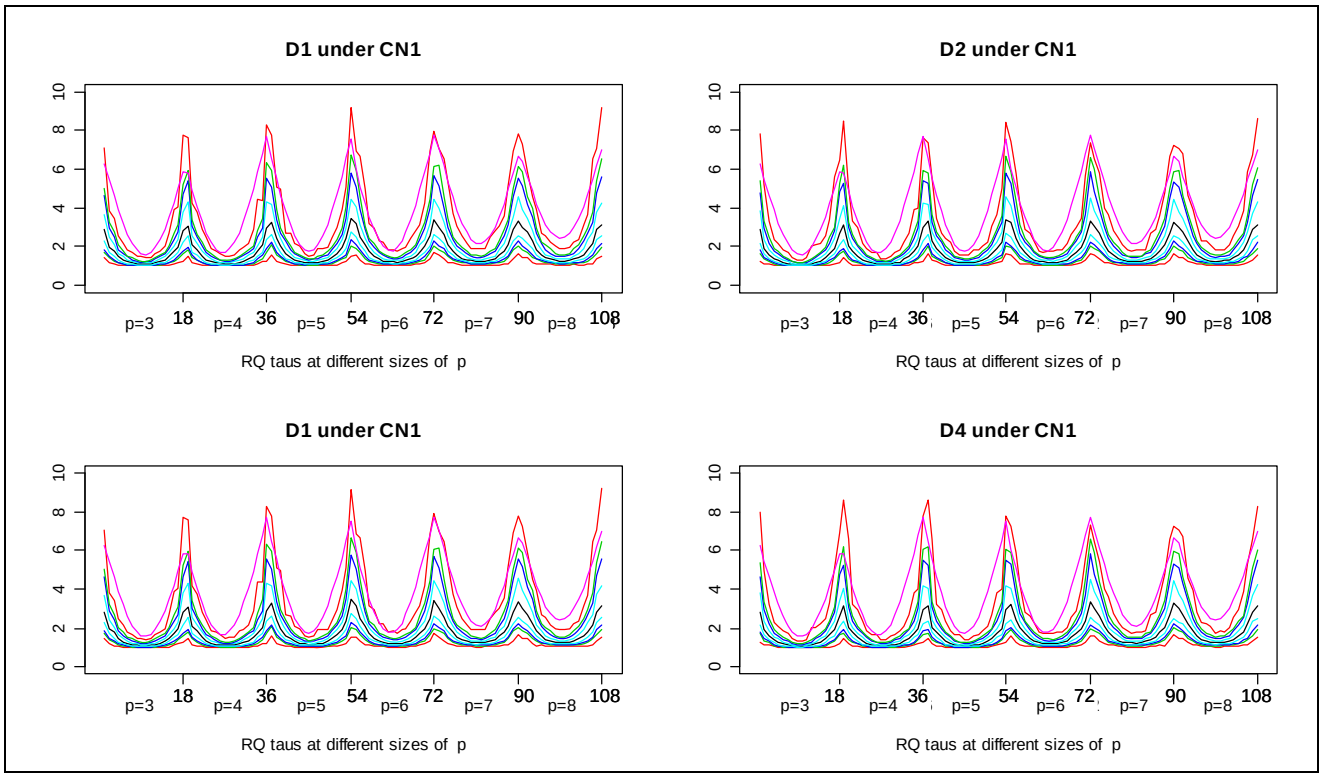
where the sum extends over all possible subsets J of $\{1, ..., n\}$ containing p elements (there are $\binom{n}{p}$ of them).

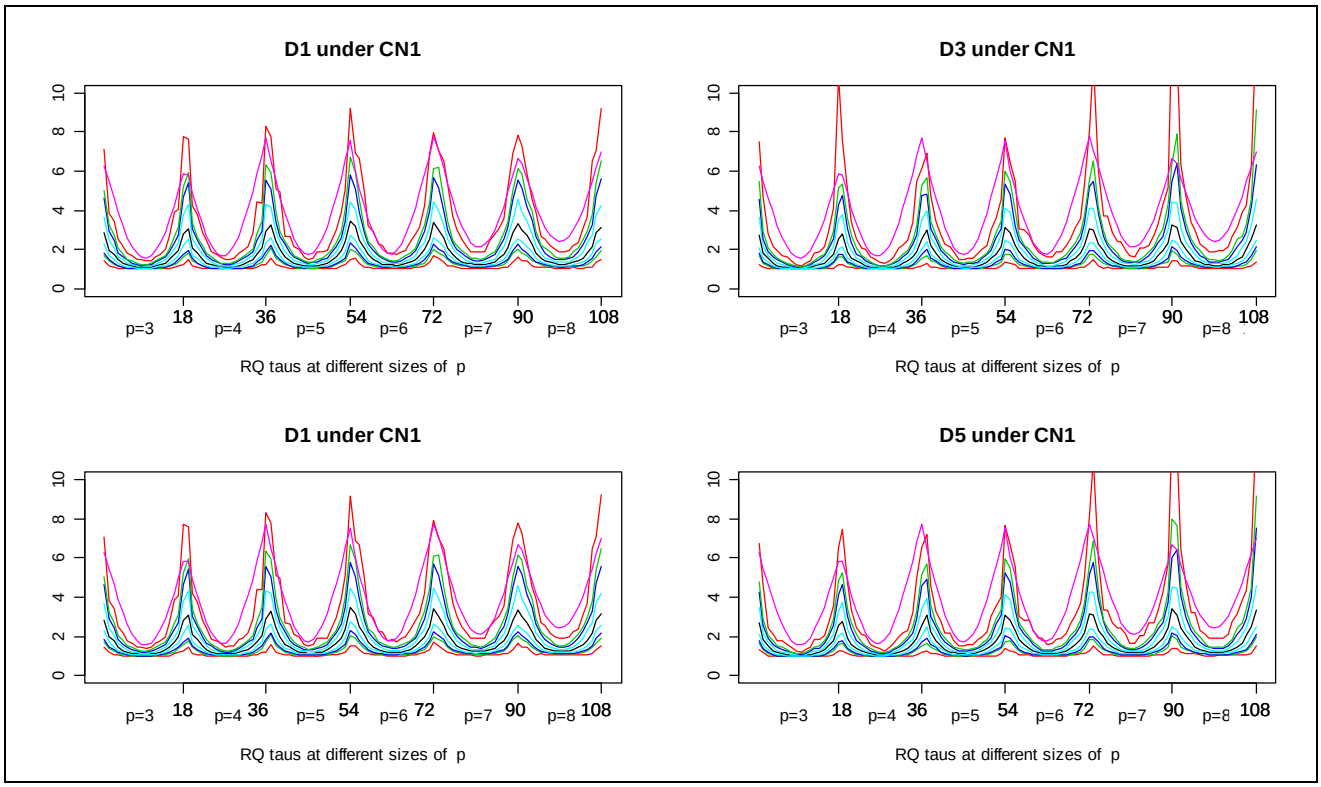
If $p = n$, *i.e.* if $\mathbf{A}$ and $\mathbf{B}$ are square matrices of the same format, then there is only a single admissible set J, and the Cauchy-Binet formula reduces to the ordinary multiplicativity of the determinant. If $p = 1$ then there are n admissible sets J and the formula reduces to that for the dot product. If $p > n$, then there is no admissible set J and the determinant $\left|\mathbf{AB}\right|$ is zero.

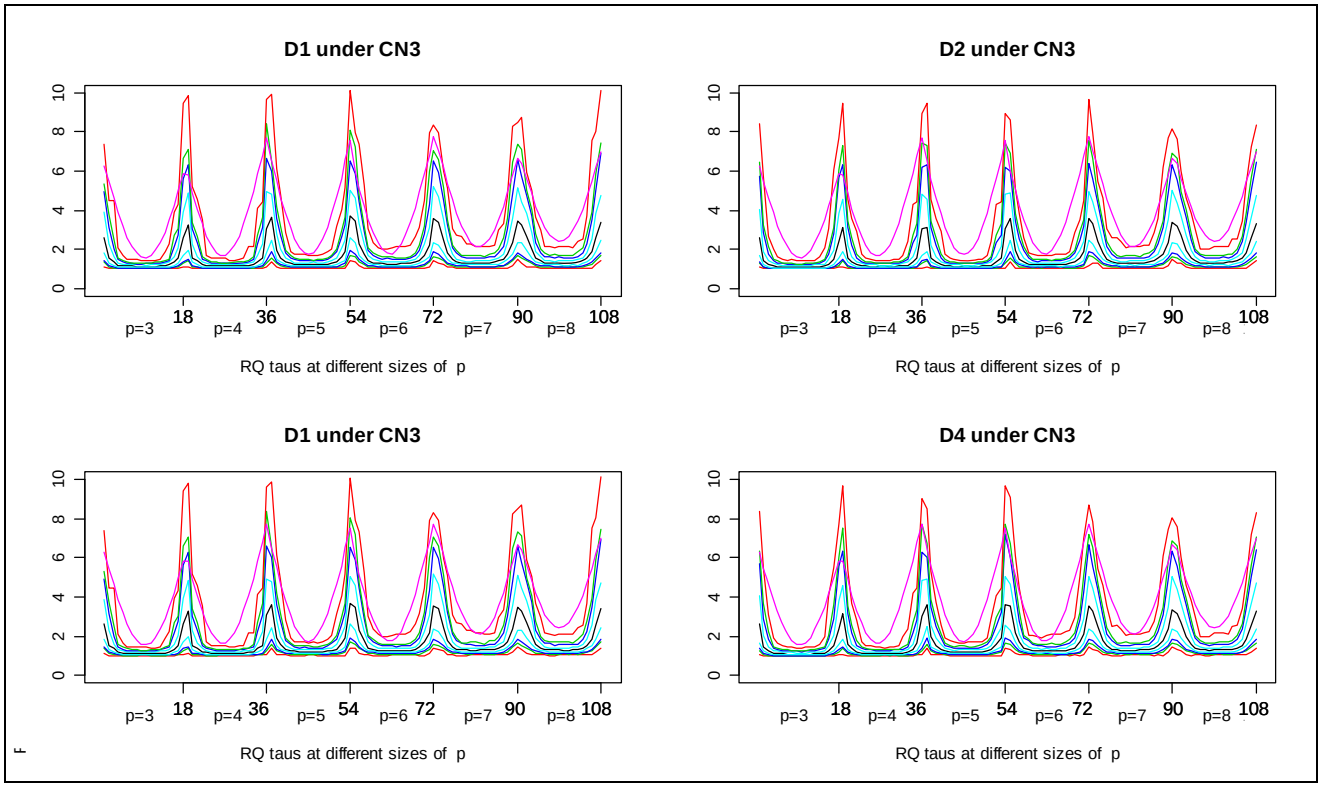
The following theorem from Koenker and Bassett (1978) points out to the fact that a RQ corresponds to a specific ES.

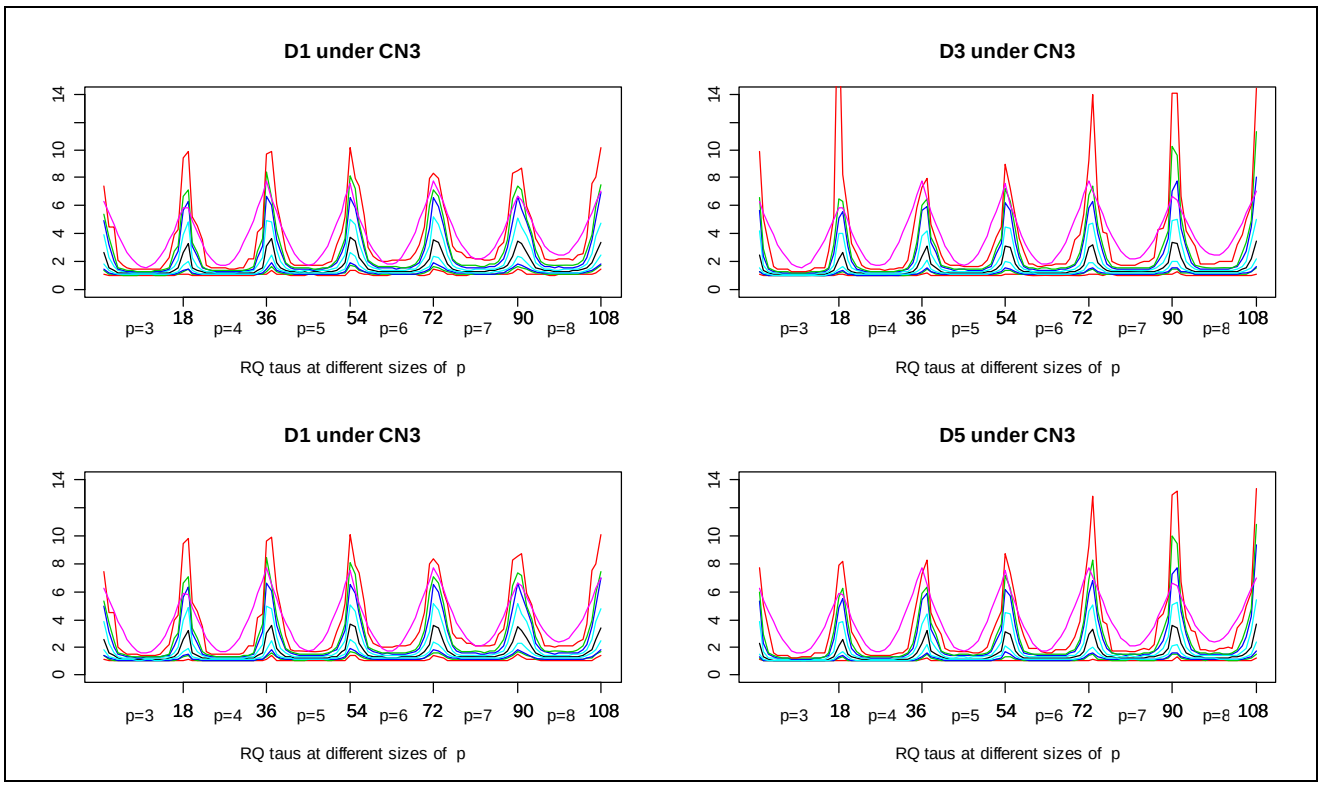
Theorem A.3: If $\tilde{\mathbf{X}}$ has rank p, the set of RQs, $\mathbf{B}(\tau) = \left\{\hat{\boldsymbol{\beta}}(\tau_1), \hat{\boldsymbol{\beta}}(\tau_2), ..., \hat{\boldsymbol{\beta}}(\tau_{K^*})\right\}$, for $0 < \boldsymbol{\tau}_1 < \boldsymbol{\tau}_2 < ... < \boldsymbol{\tau}_{K^*-1} < \boldsymbol{\tau}_{K^*} < 1$, has at least one element of the form $\hat{\boldsymbol{\beta}}(\tau_k) = \tilde{\mathbf{X}}_{J_k}^{-1}\mathbf{Y}_{J_k}$ for some $J_k \in H = \left\{J_k \in S \mid rank(\tilde{\mathbf{X}}_{J_k}) = p\right\}$. Moreover $\mathbf{B}(\tau)$ is the convex hull of all solutions having this form.

APPENDIX B: RQs PRESS GRAPHS






D1 under CN3
D2 under CN3
D1 under CN3
D4 under CN3
0 2 4 6 8 10
p=3 18 p=4 36 p=5 54 p=6 72 p=7 90 p=8 108
RQ taus at different sizes of p
F


D1 under CN3
D3 under CN3
D1 under CN3
D5 under CN3
0 2 4 6 8 10 14
p=3 18 p=4 36 p=5 54 p=6 72 p=7 90 p=8 108
RQ taus at different sizes of p

APPENDIX C: DATA SETS

Table C.1: The Gunst and Mason data set

Obs	Country	GNP	INFD	PHYS	DENS	AGDS	LIT	HIED
1	Australia	1316	19.5	806	1	21	98.5	856
2	Austria	670	37.5	695	84	1720	98.5	546
3	Barbados	200	60.4	3000	548	7121	91.1	24
4	Belgium	1196	35.4	819	301	5257	96.7	536
5	Brit Guiana	235	67.1	3900	3	192	74	27
6	Bulgaria	365	45.1	740	72	1380	85	456
7	Canada	1947	27.3	900	2	257	97.5	645
8	Chile	379	127.9	1700	11	1164	80.1	257
9	Costa Rica	357	78.9	2600	24	948	79.4	326
10	Cyprus	467	29.9	1400	62	1042	60.5	78
11	Czechoslovakia	680	31	620	108	1821	97.5	398
12	Denmark	1057	23.7	830	107	1434	98.5	570
13	El Salvador	219	76.3	5400	127	1497	39.4	89
14	Finland	794	21	1600	13	1512	98.5	529
15	France	943	27.4	1014	83	1288	96.4	667
16	Guatamala	189	91.9	6400	36	1365	29.4	135
17	Hong Kong	272	41.5	3300	3082	98143	57.5	176
18	Hungary	490	47.6	650	108	1370	97.5	258
19	Iceland	572	22.4	840	2	79	98.5	445
20	India	73	225	5200	138	2279	19.3	220
21	Ireland	550	30.5	1000	40	598	98.5	362
22	Italy	516	48.7	746	164	2323	87.5	362
23	Jamaica	316	58.7	4300	143	3410	77	42
24	Japan	306	37.7	930	254	7563	98	750
25	Luxemborg	1388	31.5	910	123	2286	96.5	36
26	Malaya	356	68.9	6400	54	2980	38.4	475
27	Malta	377	38.3	980	1041	8050	57.6	142
28	Mauritius	225	69.5	4500	352	4711	51.8	14
29	Mexico	262	77.7	1700	18	296	50	258
30	Netherlands	836	16.5	900	346	4855	98.5	923
31	New Zealand	1310	22.8	700	9	170	98.5	839
32	Nicaragua	160	71.7	2800	10	824	38.4	110
33	Norway	1130	20.2	946	11	3420	98.5	258
34	Panama	329	54.8	3200	15	838	65.7	371
35	Poland	475	74.7	1100	96	1411	95	351
36	Portugal	224	77.5	1394	100	1087	55.9	272
37	Puerto Rico	563	52.4	2200	271	4030	81	1192
38	Rumania	360	75.7	788	78	1248	89	226
39	Singapore	400	32.3	2400	2904	108214	50	437
40	Spain	293	43.5	1000	61	1347	87	258
41	Sweden	1380	16.6	1089	17	1705	98.5	401
42	Switzerland	1428	21.1	765	133	2320	98.5	398
43	Taiwan	161	30.5	1500	305	10446	54	329
44	Trinidad	423	45.4	2300	168	4383	73.8	61
45	United Kingdom	1189	24.1	935	217	2677	98.5	460
46	United States	2577	26.4	780	20	399	98	1983
47	USSR	600	35	578	10	339	95	539
48	West Germany	927	33.8	798	217	3631	98.5	528
49	Yugoslavia	265	100	1637	73	1215	77	524

Table C.2: The Hocking data set

Y	X_1	X_2	X_3
57.702	12.980	0.317	9.998
59.296	14.295	2.028	6.776
56.166	15.531	5.305	2.947
55.767	15.133	4.738	4.201
51.722	15.342	7.038	2.053
60.446	17.149	5.982	-0.055
60.715	15.462	2.737	4.657
37.447	12.801	10.663	3.048
60.947	17.039	5.132	0.257
55.270	13.172	2.039	8.738
59.289	16.125	2.271	2.101
54.027	14.340	4.077	5.545
53.199	12.923	2.643	9.331
41.896	14.231	10.401	1.041
63.264	15.222	1.220	6.149
45.798	15.740	10.612	-1.691
58.699	14.958	4.915	4.111
50.086	14.125	3.153	8.453
48.890	16.391	9.698	-1.714
62.213	16.452	3.912	2.145
45.625	13.535	7.625	3.851
53.923	14.199	4.474	5.112
55.799	15.837	5.753	2.087
56.741	16.565	8.549	8.974
43.145	13.322	8.598	4.011
50.706	15.949	8.290	-0.248

Table C.3: The Hald data set

Y	X_1	X_2	X_3	X_4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12