Chapter 14

Other Techniques

14.1 Resistant Regression

Ellipsoidal trimming can be used to create resistant multiple linear regression (MLR) estimators. To perform ellipsoidal trimming, an estimator $(T, \boldsymbol{C})$ is computed and used to create the squared Mahalanobis distances D_i^2 for each vector of observed predictors $\boldsymbol{x}_i$. If the ordered distance $D_{(j)}$ is unique, then j of the $\boldsymbol{x}_i$'s are in the ellipsoid

$$\{\boldsymbol{x} : (\boldsymbol{x} - T)^T \boldsymbol{C}^{-1}(\boldsymbol{x} - T) \le D_{(j)}^2\}. \tag{14.1}$$

The ith case $(Y_i, \boldsymbol{x}_i^T)^T$ is trimmed if $D_i > D_{(j)}$. Then an estimator of $\boldsymbol{\beta}$ is computed from the remaining cases. For example, if $j \approx 0.9n$, then about 10% of the cases are trimmed, and OLS or L_1 could be used on the cases that remain.

Recall that a response plot is a plot of the fitted values $\hat{Y}_i$ versus the response Y_i and is very useful for detecting outliers. If the MLR model holds and the MLR estimator is good, then the plotted points will scatter about the identity line that has unit slope and zero intercept. The identity line is added to the plot as a visual aid, and the vertical deviations from the identity line are equal to the residuals since $Y_i - \hat{Y}_i = r_i$.

The resistant trimmed views estimator combines ellipsoidal trimming and the response plot. First compute $(T, \boldsymbol{C})$, perhaps using the RFCH estimator or the *R/Splus* function `cov.mcd`. Trim the $M\%$ of the cases with the largest Mahalanobis distances, and then compute the MLR estimator $\hat{\boldsymbol{\beta}}_M$ from the

remaining cases. Use $M = 0, 10, 20, 30, 40, 50, 60, 70, 80,$ and 90 to generate ten response plots of the fitted values $\hat{\boldsymbol{\beta}}_M^T \boldsymbol{x}_i$ versus y_i using all n cases. (Fewer plots are used for small data sets if $\hat{\boldsymbol{\beta}}_M$ can not be computed for large M.) These plots are called "trimmed views."

Definition 14.1. The trimmed views (TV) estimator $\hat{\boldsymbol{\beta}}_{T,n}$ corresponds to the trimmed view where the bulk of the plotted points follow the identity line with smallest variance function, ignoring any outliers.

Example 14.1. For the Buxton (1920) data, *height* was the response variable while an intercept, *head length, nasal height, bigonal breadth,* and *cephalic index* were used as predictors in the multiple linear regression model. Observation 9 was deleted since it had missing values. Five individuals, cases 61–65, were reported to be about 0.75 inches tall with head lengths well over five feet! OLS was used on the cases remaining after trimming, and Figure 14.1 shows four trimmed views corresponding to 90%, 70%, 40% and 0% trimming. The OLS TV estimator used 70% trimming since this trimmed view was best. Since the vertical distance from a plotted point to the identity line is equal to the case's residual, the outliers had massive residuals for 90%, 70% and 40% trimming. Notice that the OLS trimmed view with 0% trimming "passed through the outliers" since the cluster of outliers is scattered about the identity line.

The TV estimator $\hat{\boldsymbol{\beta}}_{T,n}$ has good statistical properties if an estimator with good statistical properties is applied to the cases $(\boldsymbol{X}_{M,n}, \boldsymbol{Y}_{M,n})$ that remain after trimming. Candidates include OLS, L_1, Huber's M–estimator, Mallows' GM–estimator or the Wilcoxon rank estimator. See Rousseeuw and Leroy (1987, p. 12-13, 150). The basic idea is that if an estimator with $O_P(n^{-1/2})$ convergence rate is applied to a set of $n_M \propto n$ cases, then the resulting estimator $\hat{\boldsymbol{\beta}}_{M,n}$ also has $O_P(n^{-1/2})$ rate provided that the response Y was not used to select the n_M cases in the set. If $\|\hat{\boldsymbol{\beta}}_{M,n} - \boldsymbol{\beta}\| = O_P(n^{-1/2})$ for $M = 0, ..., 90$ then $\|\hat{\boldsymbol{\beta}}_{T,n} - \boldsymbol{\beta}\| = O_P(n^{-1/2})$ by Pratt (1959).

Let $\boldsymbol{X}_n = \boldsymbol{X}_{0,n}$ denote the full design matrix. Often when proving asymptotic normality of an MLR estimator $\hat{\boldsymbol{\beta}}_{0,n}$, it is assumed that

$$\frac{\boldsymbol{X}_n^T \boldsymbol{X}_n}{n} \to \boldsymbol{W}^{-1}.$$

If $\hat{\boldsymbol{\beta}}_{0,n}$ has $O_P(n^{-1/2})$ rate and if for big enough n all of the diagonal elements
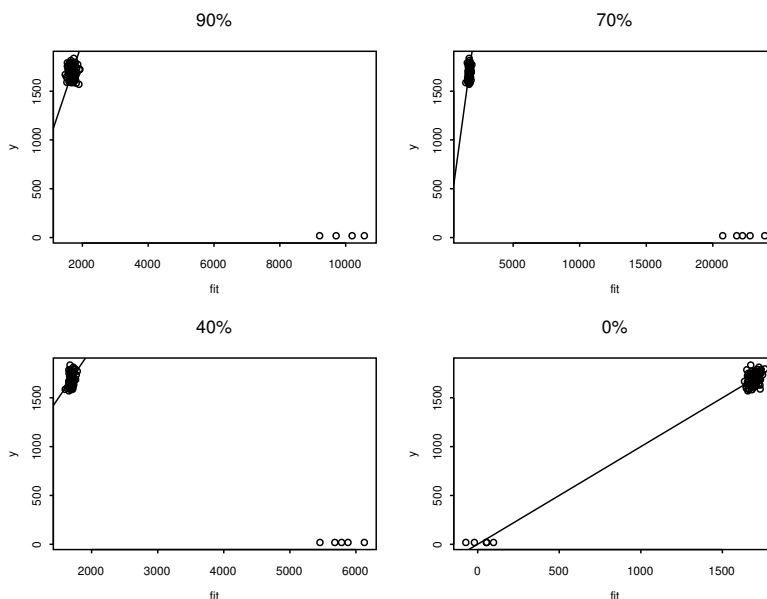
Figure 14.1: 4 Trimmed Views for the Buxton Data

of

$$\left(\frac{\boldsymbol{X}_{M,n}^{T}\boldsymbol{X}_{M,n}}{n}\right)^{-1}$$

are all contained in an interval $[0, B)$ for some $B > 0$, then $\|\hat{\boldsymbol{\beta}}_{M,n} - \boldsymbol{\beta}\| = O_P(n^{-1/2})$.

The distribution of the estimator $\hat{\boldsymbol{\beta}}_{M,n}$ is especially simple when OLS is used and the errors are iid $N(0, \sigma^2)$. Then

$$\hat{\boldsymbol{\beta}}_{M,n} = (\boldsymbol{X}_{M,n}^{T}\boldsymbol{X}_{M,n})^{-1}\boldsymbol{X}_{M,n}^{T}\boldsymbol{Y}_{M,n} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}_{M,n}^{T}\boldsymbol{X}_{M,n})^{-1})$$

and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{M,n} - \boldsymbol{\beta}) \sim N_p(\boldsymbol{0}, \sigma^2(\boldsymbol{X}_{M,n}^{T}\boldsymbol{X}_{M,n}/n)^{-1})$. Notice that this result does not imply that the distribution of $\hat{\boldsymbol{\beta}}_{T,n}$ is normal.

Table 14.1 compares the TV, MBA (for MLR), lmsreg, ltsreg, L_1 and OLS estimators on 7 data sets available from the text's website. The column headers give the file name while the remaining rows of the table give the sample size n, the number of predictors p, the amount of trimming M used by the TV estimator, the correlation of the residuals from the TV estimator with

Table 14.1: Summaries for Seven Data Sets, the Correlations of the Residuals from TV(M) and the Alternative Method are Given in the 1st 5 Rows

Method	Buxton	Gladstone	glado	hbk	major	nasty	wood
MBA	0.997	1.0	0.455	0.960	1.0	-0.004	0.9997
LMSREG	-0.114	0.671	0.938	0.977	0.981	0.9999	0.9995
LTSREG	-0.048	0.973	0.468	0.272	0.941	0.028	0.214
L1	-0.016	0.983	0.459	0.316	0.979	0.007	0.178
OLS	0.011	1.0	0.459	0.780	1.0	0.009	0.227
outliers	61-65	none	119	1-10	3,44	2,6,...,30	4,6,8,19
n	87	274	274	75	112	32	20
p	5	7	7	4	6	5	6
M	70	0	30	90	0	90	20

the corresponding alternative estimator, and the cases that were outliers. If the correlation was greater than 0.9, then the method was effective in detecting the outliers, and the method failed, otherwise. Sometimes the trimming percentage M for the TV estimator was picked after fitting the bulk of the data in order to find the good leverage points and outliers.

Notice that the TV, MBA and OLS estimators were the same for the Gladstone data and for the *major* data (Tremearne 1911) which had two small Y–outliers. For the Gladstone data, there is a cluster of infants that are good leverage points, and we attempt to predict *brain weight* with the head measurements *height, length, breadth, size* and *cephalic index*. Originally, the variable *length* was incorrectly entered as 109 instead of 199 for case 119, and the *glado* data contains this outlier. In 1997, `lmsreg` was not able to detect the outlier while `ltsreg` did. Due to changes in the *Splus* 2000 code, `lmsreg` now detects the outlier but `ltsreg` does not.

The TV estimator can be modified to create a resistant weighted MLR estimator. To see this, recall that the weighted least squares (WLS) estimator using weights W_i can be found using the ordinary least squares (OLS) regression (without intercept) of $\sqrt{W_i}Y_i$ on $\sqrt{W_i}\boldsymbol{x}_i$. This idea can be used for categorical data analysis since the minimum chi-square estimator is often computed using WLS. Let $\boldsymbol{x}_i = (1, x_{i,2}, ..., x_{i,p})^T$, let $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ and let

$\tilde{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$.

Definition 14.2. For a multiple linear regression model with weights W_i, a **weighted response plot** is a plot of $\sqrt{W_i}\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}$ versus $\sqrt{W_i}Y_i$. The **weighted residual plot** is a plot of $\sqrt{W_i}\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}$ versus the WMLR residuals $r_{Wi} = \sqrt{W_i}Y_i - \sqrt{W_i}\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}$.

Application 14.1. For resistant weighted MLR, use the WTV estimator which is selected from ten weighted response plots.

14.2   1D Regression

Regression is the study of the conditional distribution $Y|\boldsymbol{x}$ of the response Y given the $k \times 1$ vector of nontrivial predictors $\boldsymbol{x}$. The scalar Y is a random variable and $\boldsymbol{x}$ is a random vector. A special case of regression was the multiple linear regression model $Y = \alpha + x_1\beta_1 + \cdots + x_k\beta_k + e = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$ where $k = p-1$ and the nontrivial predictors are collected in the $k \times 1$ vector $\boldsymbol{x}$.

Definition 14.3: Cook and Weisberg (1999a, p. 414). In a *1D regression model*, the response Y is conditionally independent of $\boldsymbol{x}$ given a single linear combination $\boldsymbol{\beta}^T\boldsymbol{x}$ of the predictors, written

$$Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x} \ \ \text{or} \ \ Y \perp\!\!\!\perp \boldsymbol{x}|(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}). \tag{14.2}$$

An important 1D regression model, introduced by Li and Duan (1989), has the form

$$Y = g(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}, e) \tag{14.3}$$

where g is a bivariate (inverse link) function and e is a zero mean error that is independent of $\boldsymbol{x}$. The constant term α may be absorbed by g if desired.

Special cases of the 1D regression model (14.2) include many important *generalized linear models* (GLMs) and the additive error *single index model*

$$Y = m(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}) + e. \tag{14.4}$$

Typically m is the conditional mean or median function. For example if all of the expectations exist, then

$$E[Y|\boldsymbol{x}] = E[m(\alpha + \boldsymbol{\beta}^T\boldsymbol{x})|\boldsymbol{x}] + E[e|\boldsymbol{x}] = m(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}).$$

The *multiple linear regression model* is an important special case where m is the identity function: $m(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$. Another important special case of 1D regression is the *response transformation model* where

$$g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}, e) = t^{-1}(\alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e) \qquad (14.5)$$

and t^{-1} is a one to one (typically monotone) function. Hence

$$t(Y) = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e.$$

Definition 14.4. *Regression* is the study of the conditional distribution of $Y|\boldsymbol{x}$. Focus is often on the *mean function* $E(Y|\boldsymbol{x})$ and/or the *variance function* $\text{VAR}(Y|\boldsymbol{x})$. There is a distribution for each value of $\boldsymbol{x} = \boldsymbol{x}_o$ such that $Y|\boldsymbol{x} = \boldsymbol{x}_o$ is defined. For a 1D regression,

$$E(Y|\boldsymbol{x} = \boldsymbol{x}_o) = E(Y|\boldsymbol{\beta}^T \boldsymbol{x} = \boldsymbol{\beta}^T \boldsymbol{x}_o) \equiv M(\boldsymbol{\beta}^T \boldsymbol{x}_o)$$

and

$$\text{VAR}(Y|\boldsymbol{x} = \boldsymbol{x}_o) = \text{VAR}(Y|\boldsymbol{\beta}^T \boldsymbol{x} = \boldsymbol{\beta}^T \boldsymbol{x}_o) \equiv V(\boldsymbol{\beta}^T \boldsymbol{x}_o)$$

where M is the *kernel mean function* and V is the *kernel variance function.*

Notice that the mean and variance functions depend on the *same* linear combination if the 1D regression model is valid. This dependence is typical of GLMs where M and V are known kernel mean and variance functions that depend on the family of GLMs. See Cook and Weisberg (1999a, section 23.1). A *heteroscedastic regression model*

$$Y = M(\boldsymbol{\beta}_1^T \boldsymbol{x}) + \sqrt{V(\boldsymbol{\beta}_2^T \boldsymbol{x})} \; e \qquad (14.6)$$

is a 1D regression model if $\boldsymbol{\beta}_2 = c\boldsymbol{\beta}_1$ for some scalar c.

Dimension reduction can greatly simplify our understanding of the conditional distribution $Y|\boldsymbol{x}$. If a 1D regression model is appropriate, then the k–dimensional vector $\boldsymbol{x}$ can be replaced by the 1–dimensional scalar $\boldsymbol{\beta}^T \boldsymbol{x}$ with *"no loss of information about the conditional distribution."* Cook and Weisberg (1999a, p. 411) define a *sufficient summary plot* (SSP) to be a plot that contains all the sample regression information about the conditional distribution $Y|\boldsymbol{x}$ of the response given the predictors.

Definition 14.5: If the 1D regression model holds, then $Y \perp\!\!\!\perp \boldsymbol{x} | (a + c\boldsymbol{\beta}^T \boldsymbol{x})$ for any constants a and $c \neq 0$. The quantity $a + c\boldsymbol{\beta}^T \boldsymbol{x}$ is called a *sufficient predictor* (SP), and a *sufficient summary plot* is a plot of any SP versus Y. An *estimated sufficient predictor* (ESP) is $\tilde{\alpha} + \tilde{\boldsymbol{\beta}}^T \boldsymbol{x}$ where $\tilde{\boldsymbol{\beta}}$ is an estimator of $c\boldsymbol{\beta}$ for some nonzero constant c. A *response plot* or *estimated sufficient summary plot* (ESSP) is a plot of any ESP versus Y.

If there is only one predictor x, then the plot of x versus Y is both a sufficient summary plot and a response plot, but generally only a response plot can be made. Since a can be any constant, $a = 0$ is often used. The following section shows how to use the OLS regression of Y on $\boldsymbol{x}$ to obtain an ESP. If we plot the fitted values and the ESP versus Y, the plots are called fit–response and ESP-response plots. For multiple linear regression, these two plots are the same.

14.3 Visualizing 1D Regression

Consider the OLS estimator $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$. Li and Duan (1989, p. 1031) show that under regularity conditions, $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$ consistent estimator of $c\boldsymbol{\beta}$ for some constant c. If $\hat{\boldsymbol{\beta}} \approx c\boldsymbol{\beta}$ when model (14.2) holds, then the response plot of

$$\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x} \quad \text{versus} \quad Y$$

can be used to visualize the conditional distribution $Y | (\alpha + \boldsymbol{\beta}^T \boldsymbol{x})$ provided that $c \neq 0$. **Often if no strong nonlinearities are present among the predictors,** the bias vector is small enough so that $\hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ is a useful ESP.

Suppose $Y = m(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) + e$ and the errors e are small. Suppose $\hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ is a good estimator of $c\boldsymbol{\beta}^T \boldsymbol{x}$. Then m can be visualized with both a plot of $ESP = a + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ versus Y if $c \neq 0$. If $c > 0$ then the plot of ESP versus Y is similar to the plot of $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$ versus Y: except the labels of the horizontal axis change. (The two plots are usually not exactly identical since plotting controls to "fill space" depend on several factors and will change slightly.) If $c < 0$, then the plot appears to be flipped about the vertical axis. OLS often provides a useful estimator of $c\boldsymbol{\beta}$ where $c \neq 0$, but OLS can result in $c = 0$ if g is symmetric about the population median of $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$.

Definition 14.6. If the 1D regression model (14.2) holds, and OLS is

used, then the ESP may be called the *OLS ESP* and the response plot may be called the *OLS response plot*. Other estimators, such as SIR, may have similar labels.
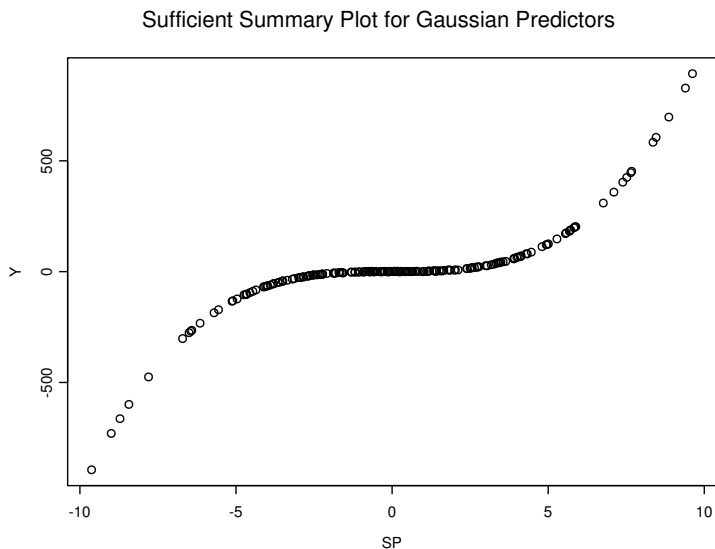

Sufficient Summary Plot for Gaussian Predictors

Figure 14.2: SSP for $m(u) = u^3$

Example 14.2. Suppose that $\boldsymbol{x}_i \sim N_3(\boldsymbol{0}, \boldsymbol{I}_3)$ and that

$$Y = m(\boldsymbol{\beta}^T \boldsymbol{x}) + e = (x_1 + 2x_2 + 3x_3)^3 + e.$$

Then a 1D regression model holds with $\boldsymbol{\beta} = (1, 2, 3)^T$. Figure 14.2 shows the sufficient summary plot of $\boldsymbol{\beta}^T \boldsymbol{x}$ versus Y, and Figure 14.3 shows the sufficient summary plot of $-\boldsymbol{\beta}^T \boldsymbol{x}$ versus Y. Notice that the functional form m appears to be cubic in both plots and that both plots can be smoothed by eye or with a scatterplot smoother such as *lowess*. The two figures were generated with the following *R/Splus* commands.

```
X <- matrix(rnorm(300),nrow=100,ncol=3)
SP <- X%*%1:3
Y <- (SP)^3 + rnorm(100)
    plot(SP,Y)
    plot(-SP,Y)
```
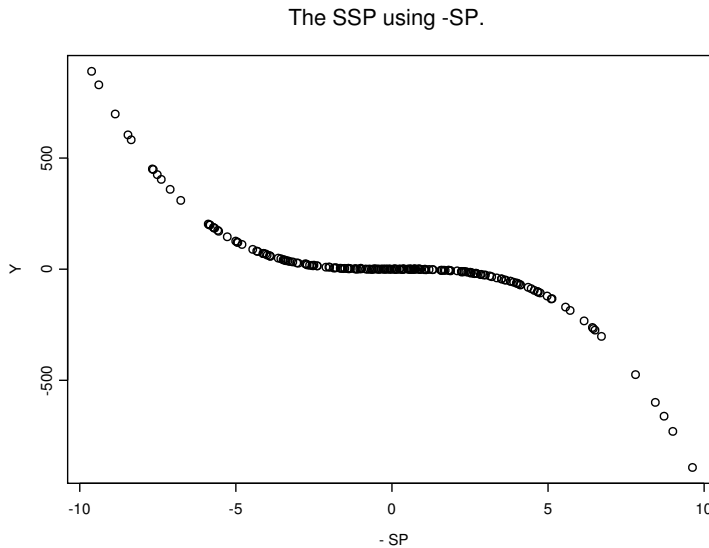
The SSP using -SP.

Figure 14.3: Another SSP for $m(u) = u^3$

We particularly want to use the OLS estimator $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ to produce an estimated sufficient summary plot. This estimator is obtained from the usual multiple linear regression of Y_i on $\boldsymbol{x}_i$, but *we are not assuming that the multiple linear regression model holds*; however, we are hoping that the 1D regression model $Y \perp\!\!\!\perp \boldsymbol{x} | \boldsymbol{\beta}^T \boldsymbol{x}$ is a useful approximation to the data and that $\hat{\boldsymbol{\beta}} \approx c\boldsymbol{\beta}$ for some nonzero constant c. Nice results exist if the single index model is appropriate. Recall that

$$\text{Cov}(\boldsymbol{x}, \boldsymbol{Y}) = E[(\boldsymbol{x} - E(\boldsymbol{x}))((\boldsymbol{Y} - E(\boldsymbol{Y}))^T].$$

Definition 14.7. Suppose that $(Y_i, \boldsymbol{x}_i^T)^T$ are iid observations and that the positive definite $k \times k$ matrix $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_X$ and the $k \times 1$ vector $\text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{X,Y}$. Let the OLS estimator $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ be computed from the multiple linear regression of Y on $\boldsymbol{x}$ plus a constant. Then $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ estimates the population quantity $(\alpha_{OLS}, \boldsymbol{\beta}_{OLS})$ where

$$\boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Sigma}_{X,Y}. \tag{14.7}$$

317

The following notation will be useful for studying the OLS estimator. Let the sufficient predictor $z = \boldsymbol{\beta}^T \boldsymbol{x}$ and let $\boldsymbol{w} = \boldsymbol{x} - E(\boldsymbol{x})$. Let $\boldsymbol{r} = \boldsymbol{w} - (\boldsymbol{\Sigma}_X \boldsymbol{\beta}) \boldsymbol{\beta}^T \boldsymbol{w}$.

Theorem 14.1. In addition to the conditions of Definition 14.7, also assume that $Y_i = m(\boldsymbol{\beta}^T \boldsymbol{x}_i) + e_i$ where the zero mean constant variance iid errors e_i are independent of the predictors $\boldsymbol{x}_i$. Then

$$\boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Sigma}_{X,Y} = c_{m,X} \boldsymbol{\beta} + \boldsymbol{u}_{m,X} \tag{14.8}$$

where the scalar

$$c_{m,X} = E[\boldsymbol{\beta}^T (\boldsymbol{x} - E(\boldsymbol{x}))\ m(\boldsymbol{\beta}^T \boldsymbol{x})] \tag{14.9}$$

and the bias vector

$$\boldsymbol{u}_{m,X} = \boldsymbol{\Sigma}_X^{-1} E[m(\boldsymbol{\beta}^T \boldsymbol{x})\boldsymbol{r}]. \tag{14.10}$$

Moreover, $\boldsymbol{u}_{m,X} = \boldsymbol{0}$ if $\boldsymbol{x}$ is from an EC distribution with nonsingular $\boldsymbol{\Sigma}_X$, and $c_{m,X} \neq 0$ unless $\text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{0}$. If the multiple linear regression model holds, then $c_{m,X} = 1$, and $\boldsymbol{u}_{m,X} = \boldsymbol{0}$.

The proof of the above result is outlined in Problem 14.1 using an argument due to Aldrin, Bølviken, and Schweder (1993). See related results in Cook, Hawkins, and Weisberg (1992). If the 1D regression model is appropriate, then typically $\text{Cov}(\boldsymbol{x}, Y) \neq \boldsymbol{0}$ unless $\boldsymbol{\beta}^T \boldsymbol{x}$ follows a symmetric distribution and m is symmetric about the median of $\boldsymbol{\beta}^T \boldsymbol{x}$.

Definition 14.8. Let $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ denote the OLS estimate obtained from the OLS multiple linear regression of Y on $\boldsymbol{x}$. The *OLS view* is a response plot of $a + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ versus Y. Typically $a = 0$ or $a = \hat{\alpha}$.

Remark 14.1. All of this awkward notation and theory leads to a remarkable result, perhaps first noted by Brillinger (1977, 1983) and called the *1D Estimation Result* by Cook and Weisberg (1999a, p. 432). The result is that if the 1D regression model is appropriate, then *the OLS view will frequently be a useful estimated sufficient summary plot* (ESSP). Hence the OLS predictor $\hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ is a useful *estimated sufficient predictor* (ESP).

Although the OLS view is frequently a good ESSP if no strong nonlinearities are present in the predictors and if $c_{m,X} \neq 0$ (eg the sufficient summary plot of $\boldsymbol{\beta}^T \boldsymbol{x}$ versus Y is not approximately symmetric), even better estimated sufficient summary plots can be obtained by using ellipsoidal trimming. This topic is discussed in the following section and follows Olive (2002) closely.

To perform ellipsoidal trimming, an estimator $(T, \boldsymbol{C})$ is computed where T is a $k \times 1$ multivariate location estimator and $\boldsymbol{C}$ is a $k \times k$ symmetric positive definite dispersion estimator. Then the ith squared Mahalanobis distance is the random variable

$$D_i^2 = (\boldsymbol{x}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - T) \qquad (14.11)$$

for each vector of observed predictors $\boldsymbol{x}_i$. If the ordered distances $D_{(j)}$ are unique, then j of the $\boldsymbol{x}_i$ are in the hyperellipsoid

$$\{\boldsymbol{x} : (\boldsymbol{x} - T)^T C^{-1} (\boldsymbol{x} - T) \leq D_{(j)}^2\}. \qquad (14.12)$$

The ith case $(Y_i, \boldsymbol{x}_i^T)^T$ is trimmed if $D_i > D_{(j)}$. Thus if $j \approx 0.9n$, then about 10% of the cases are trimmed.

We suggest that the estimator $(T, \boldsymbol{C})$ should be the classical sample mean and covariance matrix $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ or a robust multivariate location and dispersion estimator such as RFCH. See Section 4.4. When $j \approx n/2$, the RFCH estimator attempts to make the volume of the hyperellipsoid given by Equation (14.12) small.

Ellipsoidal trimming seems to work for at least three reasons. The trimming divides the data into two groups: the *trimmed cases* and the *remaining cases* $(\boldsymbol{x}_M, Y_M)$ where $M\%$ is the amount of trimming, eg $M = 10$ for 10% trimming. If the distribution of the predictors $\boldsymbol{x}$ is EC then the distribution of $\boldsymbol{x}_M$ still retains enough symmetry so that the bias vector is approximately zero. If the distribution of $\boldsymbol{x}$ is not EC, then the distribution of $\boldsymbol{x}_M$ will often have enough symmetry so that the bias vector is small. In particular, trimming often removes strong nonlinearities from the predictors and the weighted predictor distribution is more nearly elliptically symmetric than the predictor distribution of the entire data set (recall Winsor's principle: "all data are roughly Gaussian in the middle"). Secondly, under heavy trimming, the mean function of the remaining cases may be more linear than the mean function of the entire data set. Thirdly, if $|c|$ is very large, then the bias vector may be small relative to $c\boldsymbol{\beta}$. Trimming sometimes inflates $|c|$. From Theorem 14.1, any of these three reasons should produce a better estimated sufficient predictor.

For example, examine Figure 5.4. The data are not EC, but the data within the resistant covering ellipsoid are approximately EC.

Example 14.3. Cook and Weisberg (1999a, p. 351, 433, 447) gave a data set on 82 mussels sampled off the coast of New Zealand. The variables
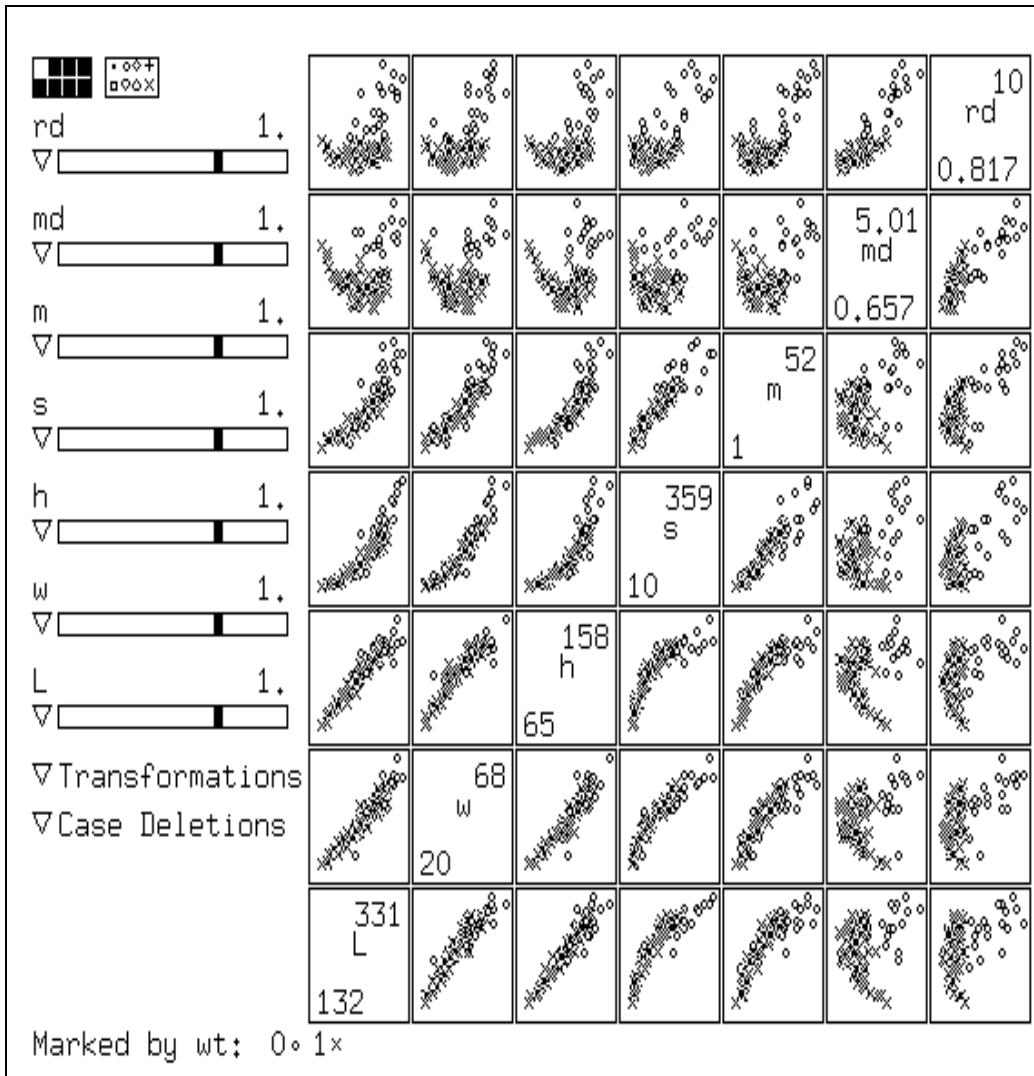
319

Figure 14.4: Scatterplot for Mussel Data, o Corresponds to Trimmed Cases

are the *muscle mass M* in grams, the *length L* and *height H* of the shell in mm, the *shell width W* and the *shell mass S*. The robust and classical Mahalanobis distances were calculated, and Figure 14.4 shows a scatterplot matrix of the mussel data, the RD_i's, and the MD_i's. Notice that many of the subplots are nonlinear. The cases marked by open circles were given weight zero by the FMCD algorithm, and the linearity of the retained cases has increased. Note that only one trimming proportion is shown and that a heavier trimming proportion would increase the linearity of the cases that were not trimmed.

The two ideas of using ellipsoidal trimming to reduce the bias and choosing a view with a smooth mean function and smallest variance function can be combined into a graphical method for finding the estimated sufficient summary plot and the estimated sufficient predictor. Trim the $M\%$ of the cases with the largest Mahalanobis distances, and then compute the OLS estimator $(\hat{\alpha}_M, \hat{\boldsymbol{\beta}}_M)$ from the cases that remain. Use $M = 0$, 10, 20, 30, 40, 50, 60, 70, 80, and 90 to generate ten plots of $\hat{\boldsymbol{\beta}}_M^T \boldsymbol{x}$ versus Y using all n cases. In analogy with the Cook and Weisberg procedure for visualizing 1D structure with two predictors, the plots will be called "trimmed views." Notice that $M = 0$ corresponds to the OLS view.

Definition 14.9. The *best trimmed view* is the trimmed view with a smooth mean function and the smallest variance function and is the estimated sufficient summary plot. If $M^* = E$ is the percentage of cases trimmed that corresponds to the best trimmed view, then $\hat{\boldsymbol{\beta}}_E^T \boldsymbol{x}$ is the estimated sufficient predictor.

The following examples illustrate the *R/Splus* function `trviews` that is used to produce the ESSP. If R is used instead of *Splus*, the command

$$\texttt{library(MASS)}$$

needs to be entered to access the function `cov.mcd` called by `trviews`. The function `trviews` is used in Problem 14.2. Also notice the `trviews` estimator is basically the same as the `tvreg` estimator described in Section 14.1. The `tvreg` estimator can be used to simultaneously detect whether the data is following a multiple linear regression model or some other single index model. Plot $\hat{\alpha}_E + \hat{\boldsymbol{\beta}}_E^T \boldsymbol{x}$ versus Y and add the identity line. If the plotted points follow the identity line then the MLR model is reasonable, but if the plotted points

follow a nonlinear mean function, then a nonlinear single index model may be reasonable.

Example 14.2 continued. The command

$$\texttt{trviews(X, Y)}$$

produced the following output.

```
Intercept        X1        X2        X3
0.6701255 3.133926 4.031048 7.593501
Intercept        X1        X2        X3
 1.101398 8.873677 12.99655 18.29054
Intercept        X1        X2        X3
0.9702788 10.71646 15.40126 23.35055
Intercept        X1        X2        X3
0.5937255 13.44889 23.47785 32.74164
Intercept        X1        X2        X3
 1.086138 12.60514 25.06613 37.25504
Intercept        X1        X2        X3
 4.621724 19.54774 34.87627 48.79709
Intercept        X1        X2        X3
 3.165427 22.85721 36.09381 53.15153
Intercept        X1        X2        X3
 5.829141 31.63738 56.56191 82.94031
Intercept        X1        X2        X3
 4.241797 36.24316 70.94507 105.3816
Intercept        X1        X2        X3
 6.485165 41.67623 87.39663 120.8251
```

The function generates 10 trimmed views. The first plot trims 90% of the cases while the last plot does not trim any of the cases and is the OLS view. To advance a plot, press the right button on the mouse (in R, highlight stop rather than continue). After all of the trimmed views have been generated, the output is presented. For example, the 5th line of numbers in the output corresponds to $\hat{\alpha}_{50} = 1.086138$ and $\hat{\boldsymbol{\beta}}_{50}^{T}$ where 50% trimming was used. The second line of numbers corresponds to 80% trimming while the last line corresponds to 0% trimming and gives the OLS estimate $(\hat{\alpha}_0, \hat{\boldsymbol{\beta}}_0^{T}) = (\hat{a}, \hat{\boldsymbol{b}})$. The trimmed views with 50% and 90% trimming were very good.
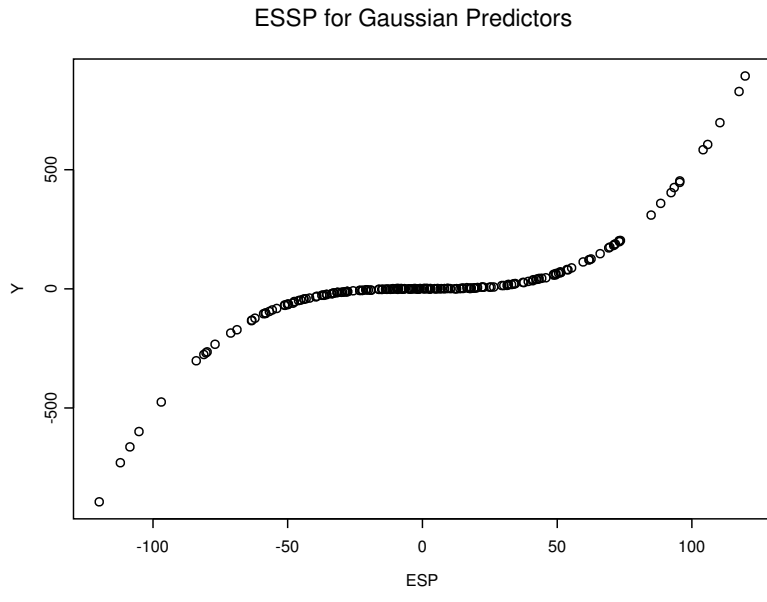
Figure 14.5: Best View for Estimating $m(u) = u^3$
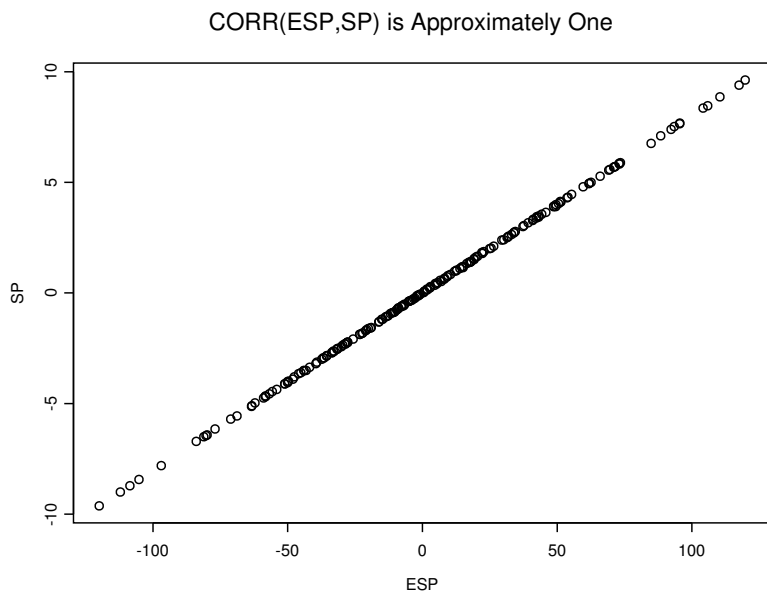

Figure 14.6: The angle between the SP and the ESP is nearly zero.

We decided that the view with 50% trimming was the best. Hence $\hat{\boldsymbol{\beta}}_E = (12.60514, 25.06613, 37.25504)^T \approx 12.5\boldsymbol{\beta}$. The best view is shown in Figure 14.5 and is nearly identical to the sufficient summary plot shown in Figure 14.2. Notice that the OLS estimate $= (41.68, 87.40, 120.83)^T \approx 42\boldsymbol{\beta}$.

The plot of the estimated sufficient predictor versus the sufficient predictor is also informative. Of course this plot can usually only be generated for simulated data since $\boldsymbol{\beta}$ is generally unknown. If the plotted points are highly correlated (with $|\text{corr(ESP,SP)}| > 0.95$) and follow a line through the origin, then the estimated sufficient summary plot is nearly as good as the sufficient summary plot. The simulated data used $\boldsymbol{\beta} = (1, 2, 3)^T$, and the commands

```
SP <- X %*% 1:3
ESP <- X %*% c(12.60514, 25.06613, 37.25504)
plot(ESP,SP)
```

generated the plot shown in Figure 14.6.

Example 14.5. An artificial data set with 200 trivariate vectors $\boldsymbol{x}_i$ was generated. The marginal distributions of $x_{i,j}$ are iid lognormal for $j = 1, 2,$ and 3. Since the response $Y_i = \sin(\boldsymbol{\beta}^T \boldsymbol{x}_i)/\boldsymbol{\beta}^T \boldsymbol{x}_i$ where $\boldsymbol{\beta} = (1, 2, 3)^T$, the random vector $\boldsymbol{x}_i$ is not elliptically contoured and the function m is strongly nonlinear. Figure 14.7d shows the OLS view and Figure 14.8d shows the best trimmed view. Notice that it is difficult to visualize the mean function with the OLS view, and notice that the correlation between Y and the ESP is very low. By focusing on a part of the data where the correlation is high, it may be possible to improve the estimated sufficient summary plot. For example, in Figure 14.8d, temporarily omit cases that have ESP less than 0.3 and greater than 0.75. From the untrimmed cases, obtained the ten trimmed estimates $\hat{\boldsymbol{\beta}}_{90}, ..., \hat{\boldsymbol{\beta}}_0$. Then using *all of the data*, obtain the ten views. The best view could be used as the ESSP.

Application 14.2. Suppose that a 1D regression analysis is desired on a data set, use the trimmed views as an exploratory data analysis technique to visualize the conditional distribution $Y|\boldsymbol{\beta}^T\boldsymbol{x}$. The best trimmed view is an estimated sufficient summary plot. If the single index model (14.4) holds, the function m can be estimated from this plot using parametric models or scatterplot smoothers such as `lowess`. Notice that Y can be predicted visually using *up and over lines*.
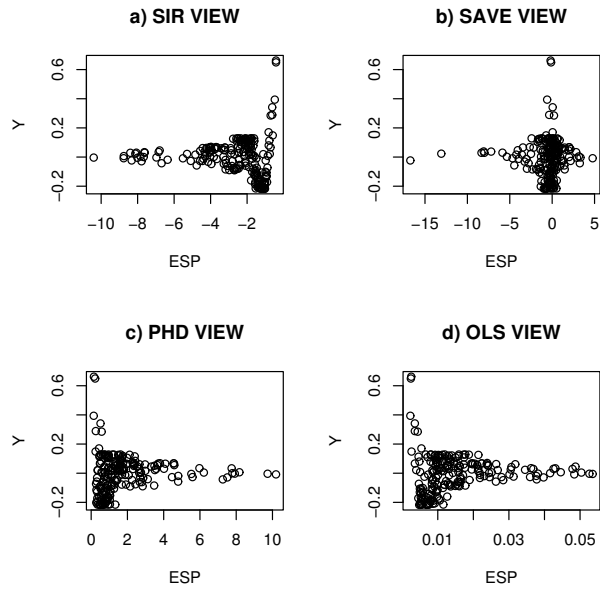
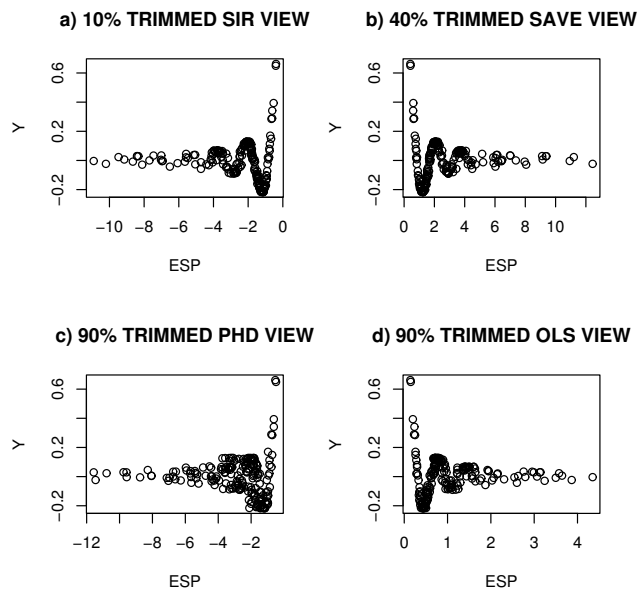Figure 14.7: Estimated Sufficient Summary Plots Without Trimming

Figure 14.8: 1D Regression with Trimmed Views

Table 14.2: Estimated Sufficient Predictors Coefficients Estimating $c(1, 2, 3)^T$

method	b_1	b_2	b_3
OLS View	0.0032	0.0011	0.0047
90% Trimmed OLS View	0.086	0.182	0.338
SIR View	-0.394	-0.361	-0.845
10% Trimmed SIR VIEW	-0.284	-0.473	-0.834
SAVE View	-1.09	0.870	-0.480
40% Trimmed SAVE VIEW	0.256	0.591	0.765
PHD View	-0.072	-0.029	-0.0097
90% Trimmed PHD VIEW	-0.558	-0.499	-0.664
LMSREG VIEW	-0.003	-0.005	-0.059
70% Trimmed LMSREG VIEW	0.143	0.287	0.428

Application 14.4. The best trimmed view can also be used as a diagnostic for linearity and monotonicity.

For example in Figure 14.5, if ESP = 0, then $\hat{Y} = 0$ and if ESP = 100, then $\hat{Y} = 500$. Figure 14.5 suggests that the mean function is monotone but not linear, and Figure 14.8 suggests that the mean function is neither linear nor monotone.

Application 14.4. Assume that a known 1D regression model is assumed for the data. Then the best trimmed view is a model checking plot and can be used as a diagnostic for whether the assumed model is appropriate.

The trimmed views are sometimes useful even when the assumption of linearly related predictors fails. Cook and Li (2002) summarize when competing methods such as the OLS view, sliced inverse regression (SIR), principal Hessian directions (PHD), and sliced average variance estimation (SAVE) can fail. All four methods frequently perform well if there are no strong nonlinearities present in the predictors.

Example 14.5 (continued). Figure 14.7 shows that the response plots for SIR, PHD, SAVE, and OLS are not very good while Figure 14.8 shows that trimming improved the SIR, SAVE and OLS methods.
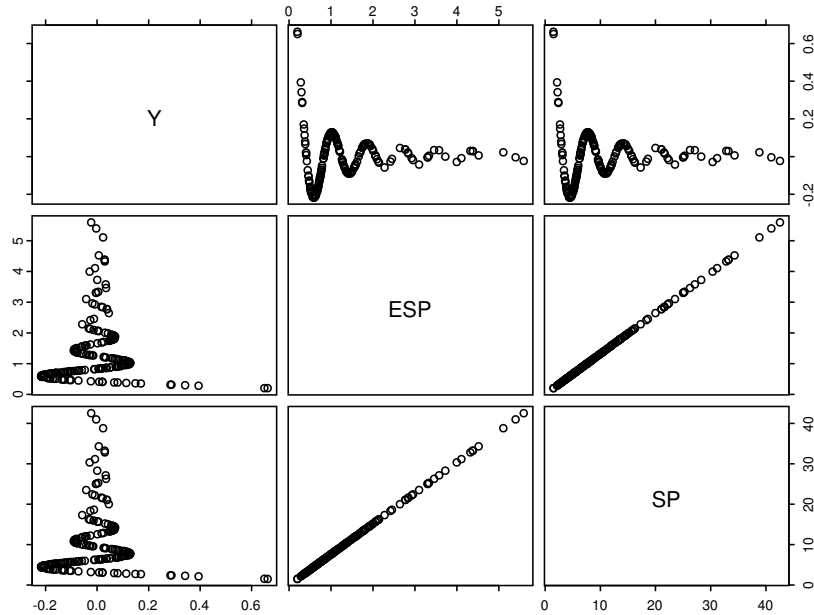
Figure 14.9: 1D Regression with `lmsreg`

One goal for future research is to develop better methods for visualizing 1D regression. Trimmed views seem to become less effective as the number of predictors $k = p - 1$ increases. Consider the sufficient predictor SP $=$ $x_1 + \cdots + x_k$. With the sin(SP)/SP data, several trimming proportions gave good views with $k = 3$, but only one of the ten trimming proportions gave a good view with $k = 10$. In addition to problems with dimension, it is not clear which covariance estimator and which regression estimator should be used. We suggest using the RFCH estimator with OLS, and preliminary investigations suggest that the classical covariance estimator gives better estimates than `cov.mcd`. But among the many *Splus* regression estimators, `lmsreg` often worked well. There is OLS theory, but there is no theory for the robust regression estimators.

Example 14.5 continued. Replacing the OLS trimmed views by alternative MLR estimators often produced good response plots, and for single index models, the `lmsreg` estimator often worked the best. Figure 14.9 shows a scatterplot matrix of Y, ESP and SP where the sufficient predictor SP $=$ $\boldsymbol{\beta}^T \boldsymbol{x}$. The ESP used ellipsoidal trimming with `cov.mcd` and with `lmsreg`
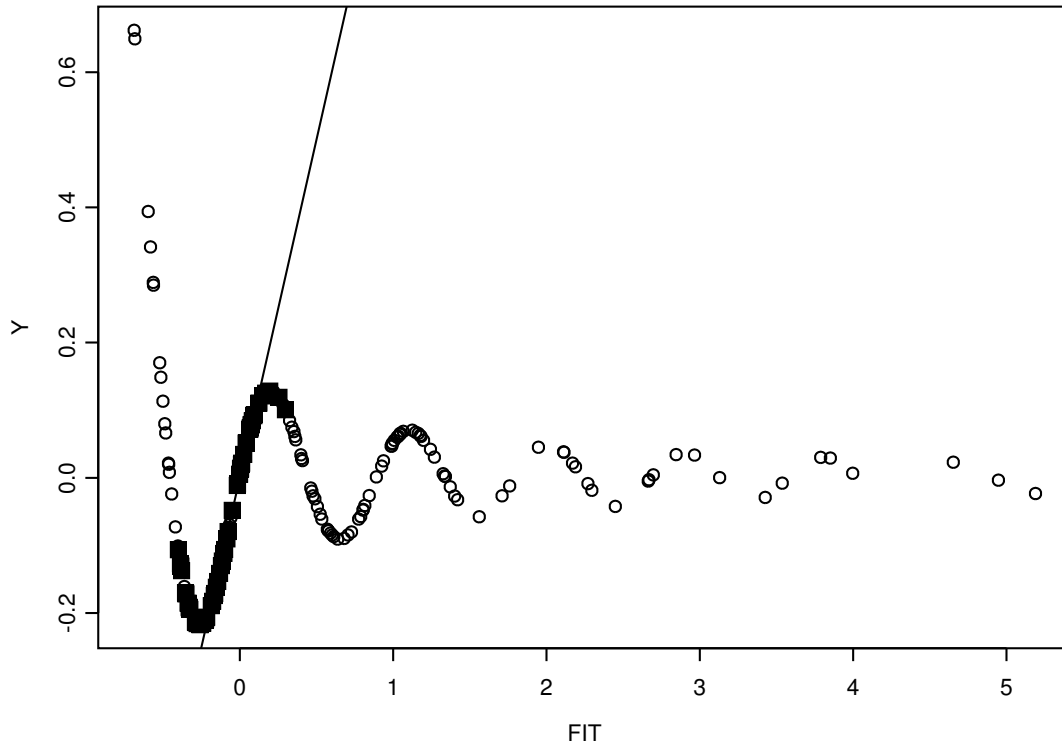
327

LMSREG TRIMMED VIEW

Figure 14.10: The Weighted `lmsreg` Fitted Values Versus Y

instead of OLS. The top row of Figure 14.9 shows that the estimated sufficient summary plot and the sufficient summary plot are nearly identical. Also the correlation of the ESP and the SP is nearly one. Table 14.2 shows the estimated sufficient predictor coefficients b when the sufficient predictor coefficients are $c(1, 2, 3)^T$. Only the SIR, SAVE, OLS and `lmsreg` trimmed views produce estimated sufficient predictors that are highly correlated with the sufficient predictor.

Figure 14.10 helps illustrate why ellipsoidal trimming works. This view used 70% trimming and the open circles denote cases that were trimmed. The highlighted squares correspond to the cases (x_{70}, Y_{70}) that were not trimmed. Note that the highlighted cases are far more linear than the data set as a

whole. Also `lmsreg` will give half of the highlighted cases zero weight, further linearizing the function. In Figure 14.10, the `lmsreg` constant $\hat{\alpha}_{70}$ is included, and the plot is simply the response plot of the weighted `lmsreg` fitted values versus Y. The vertical deviations from the line through the origin are the "residuals" $Y_i - \hat{\alpha}_{70} - \hat{\boldsymbol{\beta}}_{70}^T \boldsymbol{x}$ and at least half of the highlighted cases have small residuals.

Example 14.6. This insulation data was contributed by Ms. Spector. A box with insulation was heated for 20 minutes then allowed to cool down. The response variable $Y = temperature$ in middle of box was taken at *time* 0, 5, ..., 40. The *type* of insulation was a factor with type $1 =$ no insulation, $2 =$ corn pith, $3 =$ fiberglass, $4 =$ styrofoam and $5 =$ bubbles. There were 45 temperature measurements, one for each time type combination. The measurements were averages of ten trials and starting temperatures were close but not exactly equal.

The model using time, $(time)^2$, type, and the interactions type:time and type:$(time)^2$ had $E(Y|\boldsymbol{x}) \approx (\boldsymbol{x}^T\boldsymbol{\beta})^2$. A second model used time, $(time)^2$ and type, and rather awkward R code for producing the response plot in Figure 14.11 is shown below. The solid curve corresponds to $(\boldsymbol{x}^T\hat{\boldsymbol{\beta}}, (\boldsymbol{x}^T\hat{\boldsymbol{\beta}})^2) = (FIT, (FIT)^3)$ where $\hat{\boldsymbol{\beta}}$ is the OLS estimator from regressing Y on $\boldsymbol{x}^T = (1, time, (time)^2, type)$. The thin curve corresponds to lowess. Since the two lines correspond, $E(Y|\boldsymbol{x}) \approx (\boldsymbol{x}^T\boldsymbol{\beta})^3$ or $Y = m(\boldsymbol{x}^T\boldsymbol{\beta}) + e$ where $m(w) = w^3$. See Problem 14.7 for producing the response plot in *Arc*.

```
#assume the insulation data is loaded
ftype <- as.factor(insulation[,2])
zi <- as.data.frame(insulation)
iout <- lm(ytemp~time+I(time^2)+ftype,data=zi)
FIT <- iout$fit
Y <- insulation[,1]
plot(FIT,Y)
lines(lowess(FIT,Y))  #get (FIT,(FIT)^3) curve
zx <- FIT
z <- lsfit(cbind(zx,zx^2,zx^3),Y)
zfit <- Y-z$resid
lines(FIT,zfit)
```
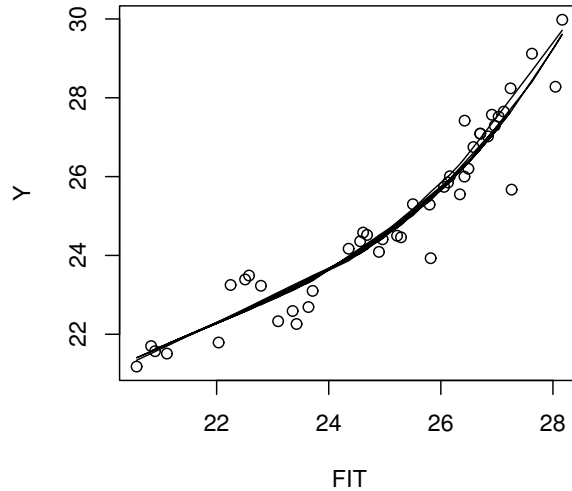
Figure 14.11: Response Plot for Insulation Data

14.4   Complements

The TV estimator was proposed by Olive (2002, 2005) and is similar to an estimator proposed by Rousseeuw and van Zomeren (1992). Although both the TV and MBA estimators have the good $O_P(n^{-1/2})$ convergence rate, their efficiency under normality may be very low. Chang and Olive (2007, 2010) suggest a method of adaptive trimming such that the resulting estimator is asymptotically equivalent to the OLS estimator.

Introduction to 1D regression and regression graphics are Cook and Weisberg (1999a, ch. 18, 19, and 20) and Cook and Weisberg (1999b), while Olive (2010) considers 1D regression. Also see Olive (2013, ch. 12).

14.5   Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USE-FUL.

14.1*. (Aldrin, Bφlviken, and Schweder 1993). Suppose

$$Y = m(\boldsymbol{\beta}^T \boldsymbol{x}) + e \qquad\qquad (14.13)$$

where m is a possibly unknown function and the zero mean errors e are independent of the predictors. Let $z = \boldsymbol{\beta}^T \boldsymbol{x}$ and let $\boldsymbol{w} = \boldsymbol{x} - E(\boldsymbol{x})$. Let $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \text{Cov}(\boldsymbol{x}, Y)$, and let $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \text{Cov}(\boldsymbol{x}) = \text{Cov}(\boldsymbol{w})$. Let $\boldsymbol{r} = \boldsymbol{w} - (\boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}) \boldsymbol{\beta}^T \boldsymbol{w}$.

a) Recall that $\text{Cov}(\boldsymbol{x}, \boldsymbol{Y}) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{Y} - E(\boldsymbol{Y}))^T]$ and show that $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = E(\boldsymbol{w}Y)$.

b) Show that $E(\boldsymbol{w}Y) = \boldsymbol{\Sigma}_{\boldsymbol{x},Y} = E[(\boldsymbol{r} + (\boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}) \boldsymbol{\beta}^T \boldsymbol{w}) \, m(z)] =$

$$E[m(z)\boldsymbol{r}] + E[\boldsymbol{\beta}^T \boldsymbol{w} \, m(z)]\boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}.$$

c) Using $\boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$, show that $\boldsymbol{\beta}_{OLS} = c(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{u}(\boldsymbol{x})$ where the constant

$$c(\boldsymbol{x}) = E[\boldsymbol{\beta}^T(\boldsymbol{x} - E(\boldsymbol{x}))m(\boldsymbol{\beta}^T \boldsymbol{x})]$$

and the bias vector $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} E[m(\boldsymbol{\beta}^T \boldsymbol{x})\boldsymbol{r}]$.

d) Show that $E(\boldsymbol{w}z) = \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}$. (Hint: Use $E(\boldsymbol{w}z) = E[(\boldsymbol{x} - E(\boldsymbol{x}))\boldsymbol{x}^T \boldsymbol{\beta}] = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x}^T - E(\boldsymbol{x}^T) + E(\boldsymbol{x}^T))\boldsymbol{\beta}]$.)

e) Assume $m(z) = z$. Using d), show that $c(\boldsymbol{x}) = 1$ if $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta} = 1$.

f) Assume that $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta} = 1$. Show that $E(z\boldsymbol{r}) = E(\boldsymbol{r}z) = \boldsymbol{0}$. (Hint: Find $E(\boldsymbol{r}z)$ and use d).)

g) Suppose that $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta} = 1$ and that the distribution of $\boldsymbol{x}$ is multivariate normal. Then the joint distribution of z and $\boldsymbol{r}$ is multivariate normal. Using the fact that $E(z\boldsymbol{r}) = \boldsymbol{0}$, show $\text{Cov}(\boldsymbol{r}, z) = 0$ so that z and $\boldsymbol{r}$ are independent. Then show that $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{0}$.

(Note: the assumption $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta} = 1$ can be made without loss of generality since if $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta} = d^2 > 0$ (assuming $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is positive definite), then $y = m(d(\boldsymbol{\beta}/d)^T \boldsymbol{x}) + e \equiv m_d(\boldsymbol{\eta}^T \boldsymbol{x}) + e$ where $m_d(u) = m(du)$, $\boldsymbol{\eta} = \boldsymbol{\beta}/d$ and $\boldsymbol{\eta}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\eta} = 1$.)

R/Splus Problems

Warning: Use the command *source("G:/mpack.txt")* **to download the programs. See Preface or Section 15.2.** Typing the name of the

`mpack` function, eg *trviews*, will display the code for the function. Use the `args` command, eg *args(trviews)*, to display the needed arguments for the function.

14.2. Use the following *R/Splus* commands to make 100 $N_3(\mathbf{0}, I_3)$ cases and 100 trivariate non-EC cases.

```
n3x <- matrix(rnorm(300),nrow=100,ncol=3)
ln3x <- exp(n3x)
```

In *R*, type the command *library(MASS)*.

a) Using the commands *pairs(n3x)* and *pairs(ln3x)* and include both scatterplot matrices in *Word*. (Click on the plot and hit *Ctrl* and *c* at the same time. Then go to *file* in the *Word* menu and select *paste*.) Are strong nonlinearities present among the MVN predictors? How about the non-EC predictors? (Hint: a box or ball shaped plot is linear.)

b) Make a single index model and the sufficient summary plot with the following commands

```
 ncy <- (n3x%*%1:3)^3 + 0.1*rnorm(100)
plot(n3x%*%(1:3),ncy)
```

and include the plot in *Word*.

c) The command *trviews(n3x, ncy)* will produce ten plots. To advance the plots, click on the *rightmost mouse button* (and in *R* select *stop*) to advance to the next plot. The last plot is the OLS view. Include this plot in *Word*.

d) After all 10 plots have been looked at the output will show 10 estimated predictors. The last estimate is the OLS (least squares) view and might look like

```
Intercept         X1         X2         X3
 4.417988 22.468779 61.242178 75.284664
```

If the OLS view is a good estimated sufficient summary plot, then the plot created from the command (leave out the intercept)

```
plot(n3x%*%c(22.469,61.242,75.285),n3x%*%1:3)
```

should cluster tightly about some line. Your linear combination will be different than the one used above. Using your OLS view, include the plot using the command above (but with your linear combination) in *Word*. Was this plot linear? Did some of the other trimmed views seem to be better than the OLS view, that is, did one of the trimmed views seem to have a smooth mean function with a smaller variance function than the OLS view?

e) Now type the *R/Splus* command

```
lncy <- (ln3x%*%1:3)^3 + 0.1*rnorm(100).
```

Use the command *trviews(ln3x,lncy)* to find the best view with a smooth mean function and the smallest variance function. This view should not be the OLS view. Include your best view in *Word*.

f) Get the linear combination from your view, say $(94.848, 216.719, 328.444)^T$, and obtain a plot with the command

```
 plot(ln3x%*%c(94.848,216.719,328.444),ln3x%*%1:3).
```

Include the plot in *Word*. If the plot is linear with high correlation, then your response plot in e) should be good.

14.3. (At the beginning of your *R/Splus* session, use *source("G:/rpack.txt")* command (and *library(MASS)* in *R*.))

a) Perform the commands

```
> nx <- matrix(rnorm(300),nrow=100,ncol=3)
> lnx <- exp(nx)
> SP <- lnx%*%1:3
> lnsincy <- sin(SP)/SP + 0.01*rnorm(100)
```

For parts b), c) and d) below, to make the best trimmed view with `trviews`, `ctrviews` or `lmsviews`, you may need to use the function twice. The first view trims 90% of the data, the next view trims 80%, etc. The last view trims 0% and is the OLS view (or `lmsreg` view). Remember to advance the view with the rightmost mouse button (and in *R*, highlight "stop"). Then click on the plot and next simultaneously hit *Ctrl* and *c*. This makes a copy of the plot. Then in *Word*, use the menu commands "Copy>paste."

b) Find the best trimmed view with `OLS` and `covfch` with the following commands and include the view in *Word*.

```
> trviews(lnx,lnsincy)
```

(With `trviews`, suppose that 40% trimming gave the best view. Then instead of using the procedure above b), you can use the command

```
> essp(lnx,lnsincy,M=40)
```

to make the best trimmed view. Then click on the plot and next simultaneously hit *Ctrl* and *c*. This makes a copy of the plot. Then in *Word*, use the menu commands "Copy>paste". Click the rightmost mouse button (and in *R*, highlight "stop") to return the command prompt.)

c) Find the best trimmed view with `OLS` and $(\overline{x}, S)$ using the following commands and include the view in *Word.* See the paragraph above b).

```
> ctrviews(lnx,lnsincy)
```

d) Find the best trimmed view with `lmsreg` and `cov.mcd` using the following commands and include the view in *Word.* See the paragraph above b).

```
> lmsviews(lnx,lnsincy)
```

e) Which method or methods gave the best response plot? Explain briefly.

14.4. Warning: this problem may take too much time. This problem is like Problem 14.3 but uses many more single index models.
a) Make some prototype functions with the following commands.

```
> nx <- matrix(rnorm(300),nrow=100,ncol=3)
> SP <- nx%*%1:3
> ncuby <- SP^3 + rnorm(100)
> nexpy <- exp(SP) + rnorm(100)
> nlinsy <- SP + 4*sin(SP) + 0.1*rnorm(100)
> nsincy <- sin(SP)/SP + 0.01*rnorm(100)
> nsiny <- sin(SP) + 0.1*rnorm(100)
> nsqrty <- sqrt(abs(SP)) + 0.1*rnorm(100)
> nsqy <- SP^2 + rnorm(100)
```

b) Make sufficient summary plots similar to Figures 14.2 and 14.3 with the following commands and include both plots in *Word.*

```
> plot(SP,ncuby)
> plot(-SP,ncuby)
```

c) Find the best trimmed view with the following commands (first type `library(MASS)` if you are using *R*). Include the view in *Word.*

```
> trviews(nx,ncuby)
```

You may need to use the function twice. The first view trims 90% of the data, the next view trims 80%, etc. The last view trims 0% and is the OLS view. Remember to advance the view with the rightmost mouse button (and in *R*, highlight "stop"). Suppose that 40% trimming gave the best view. Then use the command

```
> essp(nx,ncuby, M=40)
```

to make the best trimmed view. Then click on the plot and next simultaneously hit *Ctrl* and *c*. This makes a copy of the plot. Then in *Word*, use the menu commands "Copy>paste".

d) To make a plot like Figure 14.6, use the following commands. Let *tem* = $\hat{\boldsymbol{\beta}}$ obtained from the *trviews* output. In Example 14.2 (continued), *tem* can be obtained with the following command.

```
> tem <- c(12.60514, 25.06613, 37.25504)
```

Include the plot in *Word.*

```
> ESP <- nx%*%tem
> plot(ESP,SP)
```

e) Repeat b), c) and d) with the following data sets.
i) nx and nexpy
ii) nx and nlinsy
iii) nx and nsincy
iv) nx and nsiny
v) nx and nsqrty
vi) nx and nsqy
Enter the following commands to do parts vii) to x).

```
> lnx <- exp(nx)
> SP <- lnx%*%1:3
> lncuby <- (SP/3)^3 + rnorm(100)
> lnlinsy <- SP + 10*sin(SP) + 0.1*rnorm(100)
> lnsincy <- sin(SP)/SP + 0.01*rnorm(100)
> lnsiny <- sin(SP/3) + 0.1*rnorm(100)
> ESP <- lnx%*%tem
```

vii) lnx and lncuby

viii) lnx and lnlinsy

ix) lnx and lnsincy

x) lnx and lnsiny

14.5. Warning: this problem may take too much time. Repeat Problem 14.4 but replace `trviews` with a) `lmsviews`, b) `symviews` (that creates views that sometimes work even when symmetry is present), c) `ctrviews` and d) `sirviews`.

Except for part a), the *essp* command will not work. Instead, for the best trimmed view, click on the plot and next simultaneously hit *Ctrl* and *c*. This makes a copy of the plot. Then in *Word*, use the menu commands "Copy>paste".

14.6. a) In addition to the *source("G:/mpack.txt")* command, also use the *source("G:/mrobdata.txt")* command (and in *R*, type the *library(MASS)* command).

b) Type the command *tvreg(buxx,buxy,ii=1)*. Click the rightmost mouse button (and in *R*, highlight *Stop*). The response plot should appear. Repeat 10 times and remember which plot percentage *M* (say M = 0) had the best response plot. Then type the command *tvreg2(buxx,buxy, M = 0)* (except use your value of M, not 0). Again, click the rightmost mouse button (and in *R*, highlight *Stop*). The response plot should appear. Hold down the *Ctrl* and *c* keys to make a copy of the plot. Then paste the plot in *Word*.

c) The estimated coefficients $\hat{\boldsymbol{\beta}}_{TV}$ from the best plot should have appeared on the screen. Copy and paste these coefficients into *Word*.

Problem using ARC

14.7. a) Activate the *insulation.lsp* dataset of Example 14.6 with the menu commands "File > Load > Removable Disk (G:) > insulation.lsp." Scroll up the screen to read the data description.

b) From the insulation menu select *Transform*, click on *time*, change the number in the *p box* to 2 and click on OK to add time^2 to the variable list. From the insulation menu select *Make factors*, click on *type* and click on OK to make the factor {F}type. From the insulation menu select *Make interactions*, click on {F}type and time, then click on OK. Again from the insulation menu select *Make interactions*, click on {F}type and time^2, then click on OK.

c) From the Graph&Fit menu select *Fit linear LS*, place *y* in the *response*

box and time, time^2 and {F}type in the *Terms/Predictors box.* Click on OK and copy and paste the output into *Word.*

d) To make a response plot use the menu commands "Graph&Fit >Plot of". Select y for the V-box and L1:Fit-Values for the H-box. Click on *OK.* When the graph appears, move the OLS slider bar to 3 and the lowess slider bar to 0.5. Since the lowess curve and the OLS cubic fit to $\boldsymbol{x}^T\hat{\boldsymbol{\beta}}$ nearly coincide, the approximation $E(Y|\boldsymbol{x}) \approx (\boldsymbol{x}^T\boldsymbol{\beta})^3$ seems to be good. Copy the plot into *Word.*

e) From the Graph&Fit menu select *Fit linear LS*, place y in the *response box* and time, time^2, {F}type and From the Graph&Fit menu select *Fit linear LS*, place y in the *response box* and time, time^2, {F}type, {F}type∗time and {F}type∗time^2 in the *Terms/Predictors box.* Click on OK and copy and paste the output into *Word.*

f) To make a response plot for a second 1D regression model use the menu commands "Graph&Fit >Plot of". Select y for the V-box and L2:Fit-Values for the H-box. Click on *OK.* When the graph appears, move the OLS slider bar to 2 and the lowess slider bar to 0.5. Since the lowess curve and the OLS quadratic fit to $\boldsymbol{x}^T\hat{\boldsymbol{\beta}}$ nearly coincide, the approximation $E(Y|\boldsymbol{x}) \approx (\boldsymbol{x}^T\boldsymbol{\beta})^2$ seems to be good. Copy the plot into *Word.*