SEGMENTED REGRESSION:
A ROBUST APPROACH

BRIAN HEALEY

Segmented Regression: A Robust Approach

A practicum submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Applied Statistics.

Department of Mathematics and Statistics,
Memorial University of Newfoundland

2004

St.John's **Newfoundland** **Canada**

Abstract

Robust estimators are developed for the *segmented regression model*, a model consisting of two linear segments separated by a *change-point*. Julious (2001) introduced a method to estimate parameters in the case of unknown change-point. The focus of this practicum is on robustifying the Julious algorithm via iteratively re-weighting, extending the work of Julious (2001). Simulation studies are conducted to assess the performance of the iterative re-weighting. The methods are applied to a physiological data set studied by Julious, and to two stock-recruit data sets from fisheries science.

Acknowledgments

I would like to thank my co-supervisors, Drs. Charles Lee and Noel Cadigan, for advice, discussion and constructive criticism during the course of this work. Your assistance is greatly appreciated. I also wish to thank the statistics graduate students and statistics faculty at Memorial for many informative and helpful discussions over the past two years.

I am indebted to Drs. Gary Sneddon and Nicholas Barrowman for providing comments on an earlier draft leading to improvements in the clarity of this report.

I would also like to thank many colleagues at the Department of Fisheries and Oceans for their patience and assistance as I completed a graduate degree. Special thanks to my colleagues in the Gadoids section, and also to library staff.

To all, your contributions are greatly appreciated; any errors are mine alone.

Finally, and most importantly, I wish to deeply express my thanks to my wife Charleen for her complete love, support, and her understanding. I also thank her for the sacrifices made as I completed this research. I dedicate this work to her, and to our precious daughter Sarah.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Change-Point Problem

In the world of parametric modelling, smooth[1] functions are predominantly selected when postulating functional forms for statistical models. Further, most non-parametric frameworks focus upon smooth functions. However, in practice, it is possible that an abrupt transition may exist at some critical point and that the true dynamics of the underlying process are not smooth at this critical point. In statistics, a large field of research on such *change-point models* exists. Weighted regression in change-point models is the focus of this practicum.

Generally speaking, change-point regression is a regression problem in which the expected value of the dependent variable or response is assumed to have a different functional form in several neighbourhoods of the explanatory vari-

[1] At least once differentiable.

able space. This model can be expressed as:

$$E(Y) = \begin{cases} f_1(x_1, \ldots, x_n) \\ \qquad \vdots \\ f_N(x_1, \ldots, x_n), \end{cases} \tag{1.1}$$

where $f_1, \ldots, f_N$ are distinct functions of the explanatory variables $x_1, \ldots, x_n$.

There are several considerations within the change-point problem: there may be multiple change-points, the location of the change-point(s) may be unknown, and the model may or may not be constrained to meet at all change-points. A schematic of a hypothetical change-point model (with $N-1$ change-points) is presented in Figure 1.1.

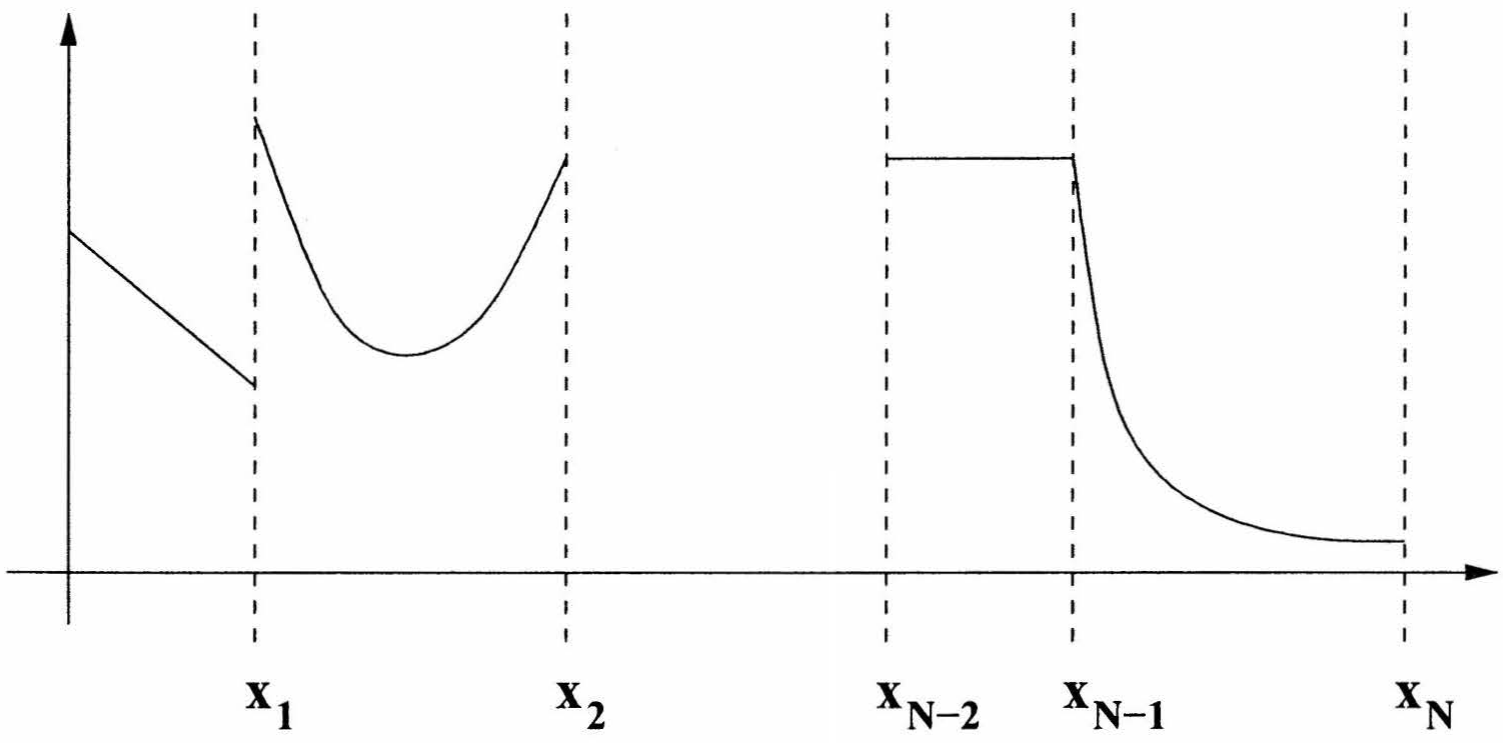

Figure 1.1: General Change-Point Model.

The following two examples are situations in which change-point models may

be applicable:

- A psychologist may expect that given a drug, patient attentiveness may increase with the drug dosage. However, one would expect a saturation point, as there are finite limits to both patient attentiveness and also to safe drug dosage levels. At this saturation point, there may be a dis-continuity in the model.

- Often, buying bulk quantities of merchandise results in savings. Consider an scenario in which purchasing large quantities of some item gives a discount per item, up to some maximum discount, after which the price is constant.

In each of these simple examples, there is reason to postulate that an abrupt transition will occur in the underlying process. For the first case, once a critical dosage of drug is administered, the patient would cease to show improvement, and would likely recess abruptly as dosage increased further. In the latter example, merchandise cost must eventually reach a minimum and become constant.

In the literature, the change-point model is also referred to as two- or multi-phase regression, segmented regression, two-stage least squares (Shaban, 1980), and broken-line regression (Feder, 1975a).

1.2 Practicum Scope

This practicum will focus on simple linear regression within the change-point problem. The models consist of two linear segments with an intercept and a single regressor and a single, unknown change-point. Hereafter, this model shall be simply referred to as the *segmented* regression model.

$$E(Y|x) = \begin{cases} \alpha_1 + \beta_1 x & \text{for } x \leq \delta \\ \alpha_2 + \beta_2 x & \text{for } x \geq \delta \end{cases} \tag{1.2}$$

The equality constraint on each segment of model (1.2) ensures that the expected mean $E(Y|X)$ is continuous at the change-point (δ). It follows that at $x = \delta$, $\alpha_1 + \beta_1\delta = \alpha_2 + \beta_2\delta$. Two illustrative depictions of the segmented regression model are presented in Figure 1.2.

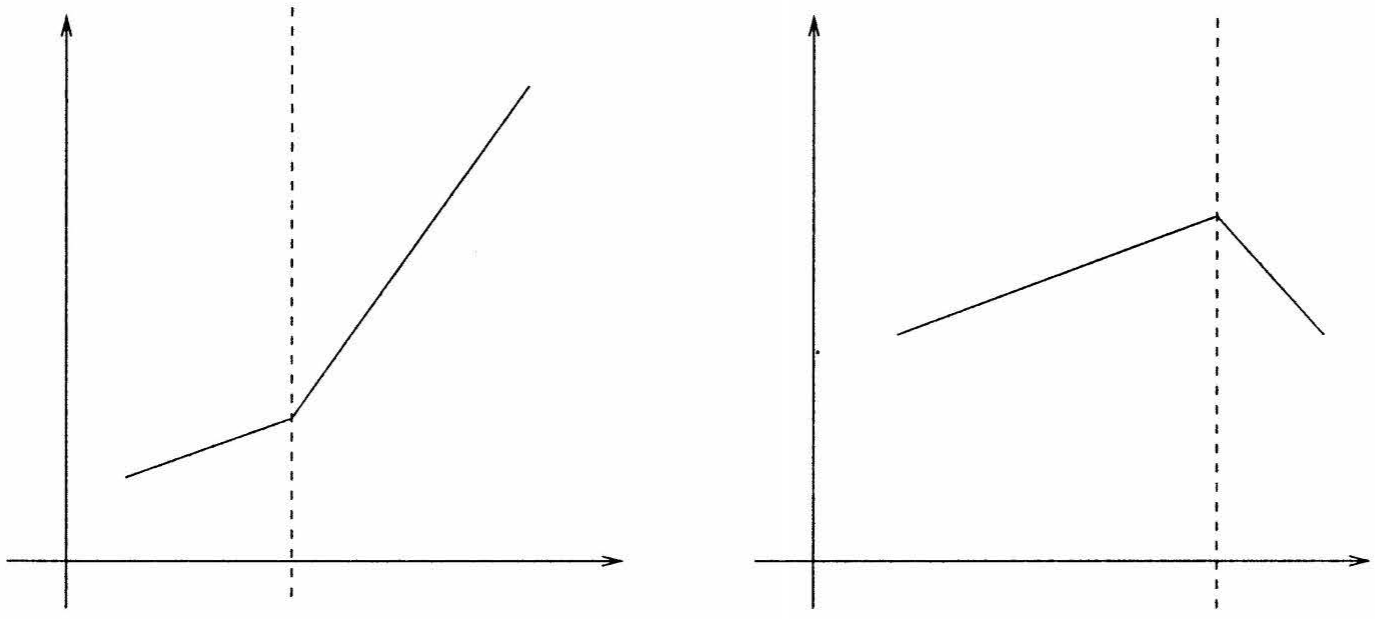

Figure 1.2: Example Segmented Regression Models.

In the event of known change-points, it is straightforward to analyse a seg-

mented linear model (1.2) and obtain estimates. One could, for example, construct a model matrix and use ordinary least squares (OLS) to estimate parameters and obtain inferences. However, if the change-point is also an unknown parameter to estimate, the model is no longer linear in the parameters and estimation becomes more involved. Recent research conducted by Julious (2001) provides an efficient algorithm for estimating the change-point in the segmented regression model, and the approach of Julious is followed in this research.

Consideration is also given to *robustification* of parameter estimates. In applying his algorithm, Julious' research implements OLS on a physiological dataset which has low variability. The sensitivity of OLS to deviations from model assumptions is well-documented (see Barnett and Lewis, 1995). We focus on methods which limit the influence of large outliers in the response variable (i.e. large residuals).

In situations with extreme outliers, one may consider eliminating these prior to analysis, i.e. case deletion. Robust estimation has greatest utility when there are one or more outliers present which are difficult (if not impossible) to detect visually.

Outliers in the covariates are not considered in this research. Ryan (1997) defines an x-outlier in regression simply as "a point that is outlying only in regard to the x co-ordinate". A class of bounded influence estimators (GM-estimators) can be applied to overcome the sensitivities of other methods to

outliers in the covariates (Ryan, 1997).

If the algorithm of Julious is applied to data which includes one or more outliers, the estimates determined by the algorithm may be poor. Outlying points may cause the least-squares estimates to differ from the trends in the remainder of the dataset (see Rousseeuw and Leroy (1987), Figure 2(b) for an illustration). The application considered in this research (see §1.3) examines data with variability considerably larger than that in Julious' dataset, so influential points and their effects on estimation are an issue of concern. The robustification is accomplished using weighted least squares, within an iteratively re-weighted framework.

A simulation study compares the performance of Julious' algorithm to that used in conjuction with iterative re-weighting.

1.3 Motivation

Change-point models are currently being used in the field of fisheries science as an approach to model the numbers of recruits[2] from the size of the spawning stock. This research has been motivated by the fact that traditional recruitment models often fit observed stock-recruit data poorly. Needle (2002) provides a review of recruitment research, including discussion on difficulties encountered in estimating parameters of traditional stock-recruit models.

[2]Typically, the number of "new" fish entering the fishable population in a given year.

An important concept in current fisheries management is the *precautionary approach* (PA), which, as the name suggests, involves managing fisheries in a risk-averse manner. Under past management approaches, there have been substantial stock collapses, perhaps none as dramatic as that of the Northern Cod Stock. This stock, historically ranging over the Grand Banks of Newfoundland and Labrador, is estimated to have declined from 3000 Kt in 1962 to 210 Kt in 1992 (age 3+ stock biomass, Bishop et *al.* 1993). In an attempt at avoiding such stock collapes in the future, Canada is currently developing the PA as an approach to managing fisheries (Shelton and Rivard, 2003). Within the PA framework, there are thresholds for various critical quantities, each of which specify a particular management response. One such set of "critical quantities" are so-called biological limit reference points. Of interest in this work is B_{lim}, defined as some threshold biomass, below which stock size should not be reduced. A review of precautionary approach reference points in Canadian fisheries management is given in Shelton and Rice (2002). The interest in the change-point method in stock-recruit modelling is that if a change-point model can suitably describe the underlying stock-recruit dynamics, then the estimated change-point (δ) provides a clear candidate for B_{lim}. Stock sizes below the change-point correspond to those which may produce impaired recruitment. If the change-point model becomes an established method to predict recruitment, then the spawning stock size at which the change-point occurs becomes quite important in terms of both fisheries management and the precautionary approach. Thus, it is important that accurate estimates of this quantity are obtained.

This fisheries research and potential applications are the motivation for this practicum.

1.4 Practicum Outline

Chapter 2 of this practicum report provides a review of the literature on segmented regression, robust methods for the linear model, and recent applications of the change-point method in fisheries science. Chapter 3 details the model considered in this project, and parameter estimation methods. Chapter 3 also contains work on robust methods, with relevant details for obtaining robust estimates in the change-point model. Chapter 4 reports on simulation studies performed to assess the performance of the estimators. Chapter 5 provides an application of the method, and Chapter 6 contains conclusions and some directions for future research.

1.5 Computational Details

Practicum computing was conducted using R software (Ihaka and Gentleman, 1996), a statistical computing package freely available on the internet[3]. The routines used in this analysis are available from the author upon request.

[3] http://www.r-project.org

Chapter 2

Literature Review

To begin this research project, an in-depth review of literature was conducted. A synopsis of relevant literature on the change-point model and on robust statistics are given in this chapter.

2.1 Change-Point Models and Estimation

The literature on change-point modelling is extensive. Originating with the work of Page (1954, 1955) on mean-shift models and Quandt (1958, 1960) in linear models, considerable effort and research has been devoted to this topic. Applications range from physiology (Bennett 1988, Julious 2001), to fisheries science (Butterworth and Bergh, 1993, Barrowman and Myers 2000, ICES 2002a), and a wide range of methodologies have been applied to estimate model parameters and derive inferences. A review paper by Krishnaiah and Miao (1988) and a bibliography compiled by Shaban (1980) provide excellent sources of information on change-point models. Research from a

conference studying change-point problems (Carlstein et *al.*, 1994) details recent progress in this field.

Numerous methodological approaches have been implemented in examining change-point models. Maximum-likelihood estimation (Robison, 1964, Hudson 1966), Bayesian estimation (Chernoff and Zacks, 1964, Bacon and Watts, 1971, Ogden and Lynch, 1999), isotonic regression (Wu et *al.* 2001), piecewise regression (McGee and Carelton, 1970), quasi-likelihood (Braun et *al.*, 2000) and non-parametric regression (Loader, 1996) are among the methods which have been applied to change-point models. Grid searching approaches have also been used to examine the change-point problem (Lerman, 1980). Change-point methods appear in many statistics research fields, including generalized linear models (Stansinopoulos and Rigby, 1992), hazard function models (Müller and Wang, 1994), time-series (Ray and Tsay, 2002), non-parametrics (Carlstein et *al.*, 1994, Csörgó and Horváth, 1988) and longitudinal studies (Piepho and Ogutu, 2003). The following paragraph considers some results relevant to change-point research in linear models.

Sprent (1961) provides extensive detail on the segmented regression problem, using least-squares methods to estimate coefficients. However, Sprent's methods include the restrictive assumption that it is known between which 2 covariate values the true change-point occurs. Robison (1964) considered the problem of estimating the intersection of two polynomial regressions. Assuming that there are N_1 observations associated with the first segment and N_2 with the second (and $N = N_1 + N_2$ is known), Robison examined

two possibilites: N_1 (i) known, and (ii) unknown. For N_1 known, maximum likelihood estimates are developed for model parameters. In the case of N_1 unknown, Robison maximizes the likelihood conditional on "plausible pairs" of (N_1, N_2). The choice of (N_1, N_2) which maximizes the conditional likelihood are taken as the solution, and the theory of the N_1 known case is applied to estimate the change-point. Robison's methods consider data which come from given ordered times. Hudson (1966) provides thorough work on finding the "overall" least-squares solution in fitting two or more constrained models. Hudson's work considers three "model join" possibilites at the change-point: (i) the models intersect at one of the covariate values, (ii) not intersecting at a data point, and model segments meet with unequal slopes, and (iii) not intersecting at a data point, and model segments meet with equal slopes. Hudson develops estimators for each type, and if the model join is of unknown type, the overall solution is the best fitting of the three possible join types. Grid-search estimation is also discussed by Hudson for determining the change-point estimate. Change-point inference is developed by Hinkley (1969), in which an asymptotic distribution for the maximum likelihood estimate of the change-point is derived. Bacon and Watts (1971) extended segmented regression and change-point modeling by introducing a "transition function" to provide a smooth transition between two intersecting straight lines. A transition parameter controls the length of the transition region, which includes abrupt transitions as a limiting case. Asymptotic distribution work on likelihood ratio statistics for testing hypotheses on parameters can be found in the work of Feder (1975a, 1975b). Feder (1975b) also discusses the relationship between change-point models and spline ap-

proximation, but identifies a key difference in interpretation. In spline theory, the knots are "chosen merely for analytical convenience", but in segmented regression, change-points "have intrinsic physical meaning". An F-test for comparing a two-line segmented regression against a single regression is given by Worsley (1983). However, as noted by Julious (2001), there are difficulties with this test if the change-point is being estimated.

2.1.1 The Julious Algorithm

For this research project, the most important article on changepoint modelling in the literature is recent work by Julious (2001). Julious has proposed a simple, but effective algorithm for estimating changepoints in the constrained segmented regression model (1.2). This algorithm is based upon three insightful notes attributed to Hudson (1966) dealing with the residual sum of squares in the model (1.2). Mathematical details of estimation in the algorithm are deferred to **Chapter 3**. Without loss of generality, we can consider the residual sum of squares of two change-point models: one constrained to meet at a given (fixed) changepoint, and the other unconstrained. Following the notation of Julious, suppose we estimate the parameters of an unconstrained segemented regression model (1.2) from the ordered data $(x_1, y_1), \ldots, (x_t, y_t)$ and $(x_{t+1}, y_{t+1}), \ldots, (x_T, y_T)$. Specifically, suppose there are T points, t of which are assumed to belong to the first model segment, and the remainder are associated with the second model segment.

The three considerations of Hudson which form the basis for Julious' algo-

rithm (from Julious 2001) are:

1. If two fitted lines meet between the adjacent extreme points of each model (x_t, x_{t+1}), then this model has residual sum of squares (RSS) that is no larger than that for any other constrained model for these two sets of points constrained to meet between (x_t, x_{t+1}).

2. If the two lines do not meet between x_t and x_{t+1}, then the constrained model with the smallest RSS will have a changepoint at either x_t or x_{t+1}.

3. Constraining a model to meet at a required point will not decrease the RSS.

Julious takes advantage of these considerations to produce the algorithm in Figure 2.1.1, hereafter referred to as the *Julious algorithm*.

Julious' algorithm is used in this project to estimate model parameters using least-squares. However, this research extends the Julious algorithm by incorporating weights and using a *weighted Julious algorithm* as a method to offer outlier robustness.

2.1.2 Relevant Applications of Change-Point Methods

As noted in §1.3, recent research on change-point models in fisheries science motivated this practicum. The change-point model has been explored by several researchers to examine its utility in modeling stock-recruit dynamics. Recent work within the International Council for the Exploration of the

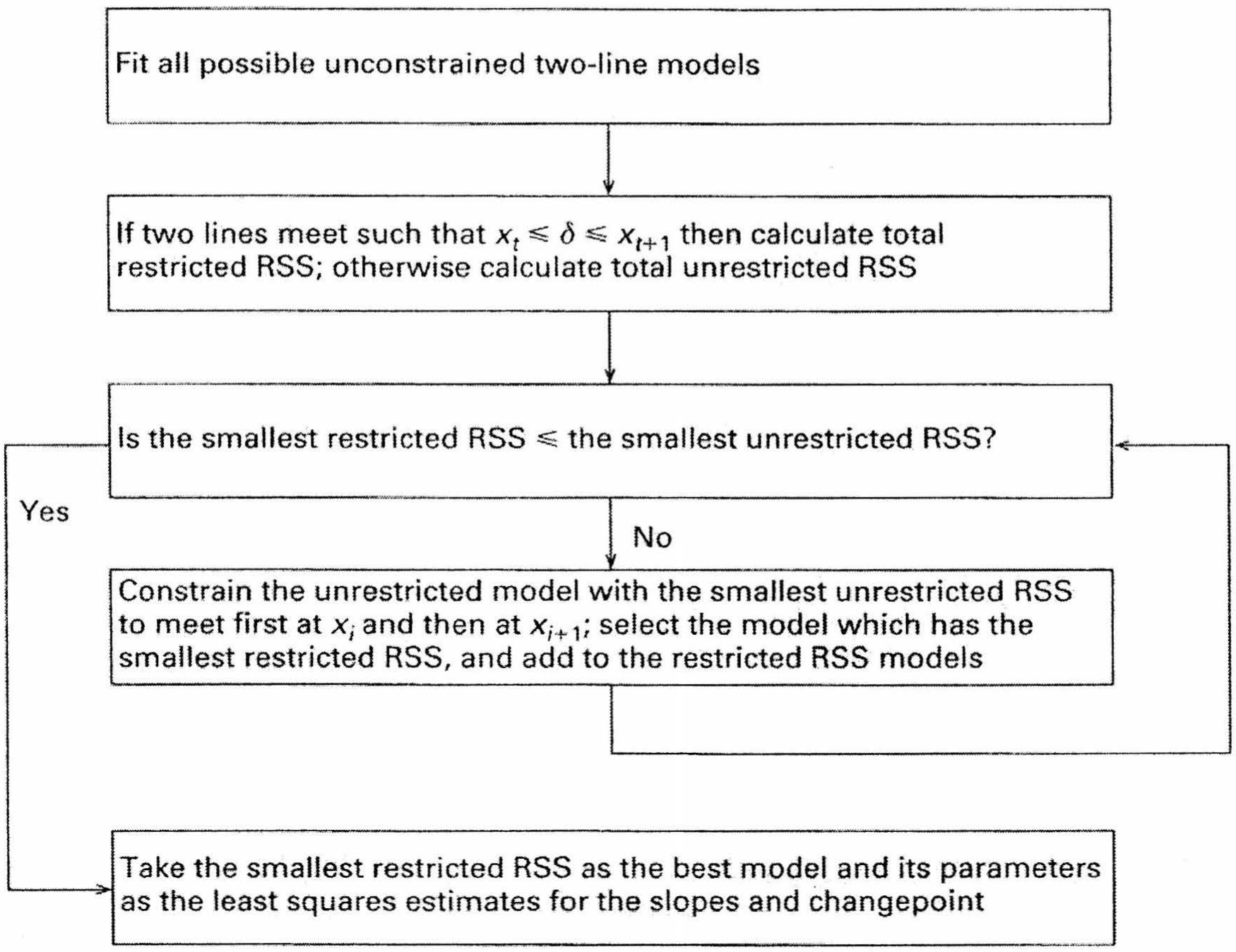

Figure 2.1: The Julious algorithm. Reproduced from Julious (2001).

Seas (ICES, 2002a) focused on application of Julious' algorithm to several fish stocks. Barrowman and Myers (2000) compare a sub-model of (1.2), the hockey-stick model (see 3.6), to the Beverton-Holt stock-recruitment model (Beverton and Holt, 1957), a cornerstone of fisheries population modelling. Barrowman and Myers also consider "generalized hockey-stick" models, an extension of the hockey-stick model. The generalized hockey-stick models involve additional parameter(s), which provide the ability to have a smooth transition (cf. Bacon and Watts, 1971) between the linear segments of the hockey-stick model. Meta-analysis methods are used to compare the esti-

mated parameters for the hockey-stick models and the Beverton-Holt model. Independent research on segmented models and smooth transitions arise in the recent doctoral thesis of Chiu (2002).

2.1.3 Robust Change-point Methods

No literature seems to exist on change-point methods using robust techniques. However, the estimators discussed by Hudson (1966), Lerman (1980), and Krishnaiah and Miao (1988) include consideration of weights assigned to each observation, which could be applied in a robust context. Application of these methods has been confined to situations with equal weights.

2.2 Robustification - Theory and Methods

2.2.1 Terminology and Preliminaries

In **Chapter 1**, it was noted that one of the primary goals of this research was to obtain robust parameter estimates in a change-point model. In statistics, the term 'robust' is a vague term - it has different interpretations in different applications. Hampel et *al.* (1986) capture this vagueness by noting: "Robust statistics, in a loose, nontechnical sense, is concerned with the fact that many assumptions commonly made in statistics (such as normality, linearity, independence) are at most approximations to reality". Developing robust statistics is necessary since slight departures from these assumptions may lead to severe problems in estimation.

An important concept in robustness and particularly robustification within linear models is that of *outliers*. Like "robustness", the term outlier does not have any formal or universal definition. Outliers can be described as data points which deviate from the trend or pattern evident in the entire dataset. The *Encyclopedia of Statistical Sciences* (Kotz and Johnson, 1988) gives the following description of outliers: "The intuitive definition of an outlier ... is some observation whose discordancy from the majority of the sample is excessive in relation to the assumed distributional model for the sample, thereby leading to the suspicison that it is not generated by this model".

The *breakdown point* of an estimator provides a measure of its sensitivity to outlying points. Informally, this is the proportion of the dataset which can be contaminated (replacing sample values with arbitrary values) without affecting the estimator. More formally, following the notation of Rousseeuw and Leroy (1987), suppose that given a random sample $X = (x_1, \ldots, x_n)$ of size n, the estimator $T(X) = \hat{\boldsymbol{\theta}}$ can be computed. We are concerned with changes in $\hat{\boldsymbol{\theta}}$ for all possible contaminated samples X'. Define bias$(m; T, X)$ as the maximum bias which can be introduced by contamination of m $(m \leq n)$ of the n points in X:

$$\text{bias}(m; T, X) = \sup_{X'} \|T(X') - T(X)\|, \tag{2.1}$$

over all possible contaminated samples X'. If this quantity is infinite, the estimator $\hat{\boldsymbol{\theta}}$ has "broken down" for the value of m considered. Since (2.1) includes the possibility of an infinite bias, Rousseeuw and Leroy (1987) propose

the finite-sample breakdown point (ε_n^* for a sample of size n) to be:

$$\varepsilon_n^*(T, X) = \min\{\frac{m}{n}; \text{bias}(m; T, X) \text{ is infinite}\}. \tag{2.2}$$

As an example, the breakdown point of the sample mean is $1/n$, which tends to 0 for large n; so we say the breakdown point of the mean is 0. That is, a single member of the original sample X could be contaminated so that the bias in (2.1) becomes arbitrarily large. However, the same cannot be said for the median of a sample; up to 50% of the sample could be changed without affecting the median. These examples represent the extreme cases for the breakdown point - the best we can hope for is a breakdown point of 0.5, whereas a breakdown point of 0 implies a potentially precarious situation - any change of the sample will change the value of the estimator.

Robust methods are typically developed to reduce the influence outlying points have on estimators. Most often, the driving reason for consideration and development of robust methods in regression is to provide the best fit to most, but not all, of the data. Robust methods are used to accurately describe the underlying trend in the data, even in the presence of outliers.

2.2.2 Classical Developments in Robustness

Sensitivity to outliers in least-squares regression is a well-recognized problem in statistics. Developments in least-squares date back to Gauss and Legendre (see Plackett, 1972 for a historical account of least-squares). Edgeworth (1887, as cited in Rouseeuuw and Leroy, 1987) noted that in least-squares

regression, minimizing the L_2 norm $||\boldsymbol{y} - X\boldsymbol{\beta}||$ could lead to an unrealistic fit if one or more of the points differed substantially from the trends of other points. Since this time, several suggestions have been made to produce regression estimates which are less susceptible to outliers. Many robust regression procedures now exist, and very brief synopses of some of these are given below. A more thorough treatment and description of such procedures can be found in Hampel et *al.* (1986) and Huber (1977).

One of the earliest suggestions to remedy the sensitivity of least-squares regression was L_1 regression, which is generally attributed to Edgeworth (1887, as cited in Rouseeuuw and Leroy, 1987). In least-squares, the estimated parameters are determined by minimizing the sum of squared regression residuals, $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$, where $\hat{\boldsymbol{y}}$ is the predicted value of $\boldsymbol{y}$:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2}. \tag{2.3}$$

Since the residuals are squared, it is evident that outlying points will contribute significantly to this objective function. In L_1 regression, the absolute value of the residuals are considered when computing estimates - thus outlying points have reduced impact on the objective:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} |\varepsilon_i|. \tag{2.4}$$

In many situations with an outlier or an influential point, L_1 regression can provide improved parameter estimates. Although it provides an improvement over least-squares (in the sense of robustness), L_1 regression is not

without problems. In the case of simple regression with a single outlier in the x-direction, L_1 regression estimators offer no relief - it fails, as will least-squares in this situation. Observe that in least-squares, the sum of squared residuals (2.3) is minimized. Dividing by n, notice that (2.3) is equivalent to minimizing the mean squared residual. As noted in §2.1, the mean has a breakdown point of 0, and thus is extremely sensitive to outliers. By using the median in the objective function, we obtain an estimator which is more robust to outliers. This is the least median of squares (LMS) estimator of Rousseeuw (1984).

In LMS, Rousseeuw suggested estimating regression parameters by not considering the sum, but the median, of the squared residuals because the median is a robust estimator of location. Thus the objective is:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}}[\text{Median } (\hat{\varepsilon_i}^2)]. \tag{2.5}$$

In trimmed regression, instead of examining all residuals to estimate parameters, only a subset of the residuals considered in (2.3) are used. The ordered residuals $\varepsilon_{(i)}$ are used to compute:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \sum_{i=1}^{l<n} \hat{\varepsilon}^2_{(i)}. \tag{2.6}$$

for some l, similar to a trimmed mean (Casella and Berger, 1990). That is, the $(n - l)$ largest of the ordered squared residuals are trimmed from the total residual sum of squares, and thus the fit is less susceptible to the

most extreme outliers. More generally, one may consider L_p-norm regression, and minimize the sum of the residuals raised to the p^{th} power (Birkes and Dodge, 1993). The LMS and trimmed methods perform well under most circumstances, but the potential to obtain an "exact-fit" exists (Rousseeuw and Leroy, 1987). An exact-fit can be obtained if a subset of the data exactly satisfy $\boldsymbol{y} = X\boldsymbol{\beta}$. If this subset is sufficiently large, the remaining data will be excluded from (2.5) or (2.6). Further, LMS estimates have low efficiency, and are not robust against "small changes in centrally located points" (Birkes and Dodge, 1993).

2.2.3 Other Classes of Robust Estimators

In addition to the ad-hoc estimators of the previous section, several classes of robust estimators have been developed and applied to regression problems. Extensive theory can be found in Huber (1977), Hampel et *al.* (1986), Roussseeuw and Leroy (1987), Jurečková and Sen (1995), and references therein.

M-estimators (Huber, 1977) are estimators of the form:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho(\varepsilon_i) \tag{2.7}$$

for some function ρ. Huber (1977) proposed

$$\rho(\varepsilon_i) = \begin{cases} \varepsilon_i^2/2 & \text{for } |\varepsilon_i| \leq c \\ c|\varepsilon_i| - c^2/2 & \text{for } |\varepsilon_i| > c \end{cases} \tag{2.8}$$

to yield robustness. (This choice for ρ leads to the Huber weights described in §3.3.2.)

Alternatively, we may write:

$$\sum_{i=1}^{n} \boldsymbol{x_i} \psi(\varepsilon_i) = 0, \tag{2.9}$$

where ψ is defined as:

$$\psi(\varepsilon_i) = \frac{\partial}{\partial \boldsymbol{\beta}} \rho(\varepsilon_i). \tag{2.10}$$

For example, if $\rho(\varepsilon_i) = {\varepsilon_i}^2/2$, the M-estimate becomes the least-squares estimate, and if $\rho(\varepsilon_i) = |\varepsilon_i|$, the M-estimate is equivalent to L_1 regression. Iterative re-weighting (see §2.2.4) is often used to compute M-estimates, via weights $w_i = \psi(\varepsilon_i)/\varepsilon_i$. Venables and Ripley (1999) describe additional technical estimation details.

R-estimates (Huber, 1977) are a class of methods which use the ranks of the residuals to obtain robust estimates. Work by Jaeckel (see §3.6 of Rousseeuw and Leroy, 1987) leads to the objective function:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} a_n(R_i)\varepsilon_i \tag{2.11}$$

where R_i is the rank of ε_i, and $a_n(.)$ is a score function. Typical score functions are listed in Rouseeuuew and Leroy (1987).

L-estimates (Huber, 1977, Jurečková and Sen, 1995) are constructed from linear combinations of the ordered residuals to estimate parameters. One

obvious example of an L-estimate is to use a binary step function $a_n(.)$ to obtain the LTS estimator (2.6) of Rousseeuw.

S-estimators minimize dispersion in the residuals (Yohai, 1987). S-estimates are computed from:

$$\min_{\beta} S(\beta), \tag{2.12}$$

where $S(\beta)$, the dispersion, is an M-estimate of the scale of the residuals.

MM-estimates, developed by Yohai (1987), are computed in three stages. MM-estimates have improved efficiency in high breakdown estimators. First, a robust regression estimate β^* is computed. Next, an M-estimate of scale of the residuals $\varepsilon_i = \varepsilon_i(\beta^*)$ is computed. An M-estimate is then obtained using the results of the two initial steps (see Rousseeuw and Leroy, 1987 for additional details).

Several classes of estimators discussed above require a robust estimate of scale. Huber (1981) provides detailed discussion of methods for robustifying scale estimates. In this research, we use the robust estimator MAD (Hampel et *al.*, 1986) to robustify the scale estimate (see §3.3.2).

2.2.4 Iterative Re-Weighting

Weighted regression was introduced by Cotes in unpublished work during the eighteenth century (Plackett, 1972). Weighted (or generalized) regression (Draper and Smith, 1981) can be used to offer outlier protection. Iteratively re-weighting regressions is a recent approach, attributed to Beaton and Tukey (1974). It is conceptually simple in that regressions are continually re-weighted until some convergence criteria is attained. The weighting method and convergence criteria of iteratively re-weighted least-squares (IRLS) are subjective choices in application. Computational details for IRLS are given in Holland and Welsch (1977). Additional properties of IRLS, including convergence properties, are discussed in Birch (1980a, 1980b).

The IRLS algorithm (following Neter et *al.* 1996) is:

1. Choose a weight function for weighting the cases.

2. Obtain starting weights for all cases.

3. Use the weights in weighted least squares and obtain the residuals from the fitted regression function.

4. Use the residuals in step 3 to obtain revised weights.

5. Continue steps 3 and 4 the until convergence is obtained.

Some of the robust estimation methods discussed in the previous section can be solved iteratively (e.g. Huber M-estimators); hence they may be formulated as IRLS problems (see Carroll and Ruppert, 1988).

Chapter 3

The Model and Estimation

This chapter contains the methodologies used to obtain parameter estimates for the segmented regression model, and other related models.

3.1 Segmented Regression Model - Notation and Preliminaries

Recall from **Chapter 1** the segmented regression model:

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & \text{for } x_i \leq \delta \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & \text{for } x_i \geq \delta \end{cases} \tag{3.1}$$

Suppose there are T data points (x_i, y_i) such that $x_1 < x_2 < \ldots < x_T$. Further, suppose that the true change-point is known. Define t such that the change-point belongs to the set $[x_t, x_{t+1})$. Thus t observations satisfy $x_i \leq \delta$, and there are $(T - t)$ points such that $x_i > \delta$. Following Julious (2001), we

may write (3.1) in the conventional form of a linear model, $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. At the change-point, we must have $\alpha_1 + \beta_1\delta = \alpha_2 + \beta_2\delta$, or re-arranging,

$$\delta = \frac{(\alpha_2 - \alpha_1)}{(\beta_1 - \beta_2)}. \tag{3.2}$$

Letting:

$\boldsymbol{y} = (y_1, y_2, \ldots, y_T)^T$, $\boldsymbol{\beta} = (\alpha_2, \beta_1, \beta_2)^T$, and

$$X = \begin{pmatrix} 1 & x_1 - \delta & \delta \\ \vdots & \vdots & \vdots \\ 1 & x_t - \delta & \delta \\ 1 & 0 & x_{t+1} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_T \end{pmatrix},$$

we expand $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, to obtain:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_t \\ y_{t+1} \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \alpha_2 + (x_1 - \delta)\beta_1 + \beta_2\delta \\ \vdots \\ \alpha_2 + (x_t - \delta)\beta_1 + \beta_2\delta \\ \alpha_2 + \beta_2 x_{t+1} \\ \vdots \\ \alpha_2 + \beta_2 x_T \end{pmatrix}.$$

If we replace δ as in (3.2), it follows that for $i = 1, \ldots, t$,

$$\begin{aligned} y_i &= \alpha_2 + \beta_1 x_i + (\beta_2 - \beta_1)\delta + \varepsilon_i \\ &= \alpha_2 + \beta_1 x_i + (\beta_2 - \beta_1)\left(\frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2}\right) + \varepsilon_i \\ &= \alpha_2 + \beta_1 x_i + \alpha_1 - \alpha_2 + \varepsilon_i \\ &= \alpha_1 + \beta_1 x_i + \varepsilon_i. \end{aligned}$$

If we obtain estimates of $\hat{\boldsymbol{\beta}} = \left(\hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2\right)^T$, then we can estimate α_1 using: $\hat{\alpha}_1 = \hat{\alpha}_2 + (\hat{\beta}_2 - \hat{\beta}_1)\delta$. Estimates of $\boldsymbol{\beta}$ can be obtained using any of the estimation methods described in **Chapter 2**.

As an alternative to the model construction outlined above, Julious (2001) showed how the least-squares estimate of parameters constrained to meet at δ could be obtained from the unconstrained estimates, i.e. those obtained from simple regression for each model segment. We follow this approach in the next section. Consideration is given to weighted estimation assuming (i) normal and (ii) lognormal error distributions. Iterative re-weighting is applied to offer robustness against outliers.

3.2 Estimation of Parameters - Equal Weighting

3.2.1 Least-Squares Estimates for known Change-point (δ)

If the model is not constrained to meet at the change-point and the location of the change-point is known exactly, least-squares (LS) estimates are trivial to obtain. Let X_1, X_2 be the design matrices for each model segment (such that X_1 corresponds to $x_i \leq \delta$, X_2 to $x_i > \delta$). Each model segment is a simple linear model with an intercept and one covariate, x:

$$X_i = (\boldsymbol{j}, \boldsymbol{x_i}), \qquad \text{for } i = 1, 2,$$

where $\boldsymbol{j}$ is a vector of 1's, and $\boldsymbol{x_i}$ is the covariate vector for each segment.

The LS estimates, which are also the maximum likelihood estimates if $y \overset{iid}{\sim} N(0, \sigma^2)$, follow directly from the normal equations (Julious 2001):

$$\hat{\boldsymbol{\beta}}_i = (X_i^T X_i)^{-1} X_i^T \boldsymbol{y_i}, \qquad \text{for } i = 1, 2. \tag{3.3}$$

To constrain the model segments to intersect at the change-point, δ, estimates can be obtained via the method of Lagrange multipliers (Rencher 2000, Julious 2001). Generally, for the linear model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, suppose it is required to constrain the estimate $\hat{\boldsymbol{\beta}}$ to satisfy the p constraints specified by $H\boldsymbol{\beta} = 0$, where $\boldsymbol{\beta} \in \mathbb{R}^p, H \in \mathbb{R}^{k \times p}$. The constrained estimate, $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$, is:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} &= \hat{\boldsymbol{\beta}}_{LS} - (X^T X)^{-1} H^T [H(X^T X)^{-1} H^T]^{-1} H \hat{\boldsymbol{\beta}}_{LS} \qquad (3.4)\\ &= (I_p - A)\hat{\boldsymbol{\beta}}_{LS}\end{aligned}$$

where $\hat{\boldsymbol{\beta}}_{LS} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{1}} \\ \hat{\boldsymbol{\beta}}_{\mathbf{2}} \end{pmatrix}$ are the LS estimates from (3.3), I_p is the $p \times p$ identity matrix, and $A = (X^T X)^{-1} H^T [H(X^T X)^{-1} H^T]^{-1} H$. In this case, with a single constraint, $H = \boldsymbol{h}^t$ is a vector, so that:

$$A = \frac{(X^T X)^{-1} \boldsymbol{h}\boldsymbol{h}^T}{\boldsymbol{h}^T (X^T X)^{-1} \boldsymbol{h}}.$$

We may write the constraint as: $\alpha_1 - \alpha_2 + (\beta_1 - \beta_2)\delta = 0$. Since $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)^T$, it follows $\boldsymbol{h} = (1, \delta, -1, -\delta)^T$.

The least squares estimates $\hat{\boldsymbol{\beta}}$ minimize the residual sum of squares (RSS):

$$RSS = \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^T (\boldsymbol{y} - X\hat{\boldsymbol{\beta}}). \qquad (3.5)$$

Two sub-models of the segmented regression model (3.1) are of interest in this research. If $\alpha_1 = \beta_2 = 0$, then we have the "hockey-stick" model (3.6) of Barrowman and Myers (2000),

$$y_i = \begin{cases} \beta_1 x_i + \varepsilon_i & \text{for } x \leq \delta \\ \alpha_2 + \varepsilon_i & \text{for } x \geq \delta \end{cases}, \qquad (3.6)$$

(also see Butterworth and Bergh, 1993). Retaining both slope parameters,

($\alpha_1 = 0$), then the "doorhinge" model (3.7) is

$$y_i = \begin{cases} \beta_1 x_i + \varepsilon_i & \text{for } x \leq \delta \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & \text{for } x \geq \delta \end{cases} \tag{3.7}$$

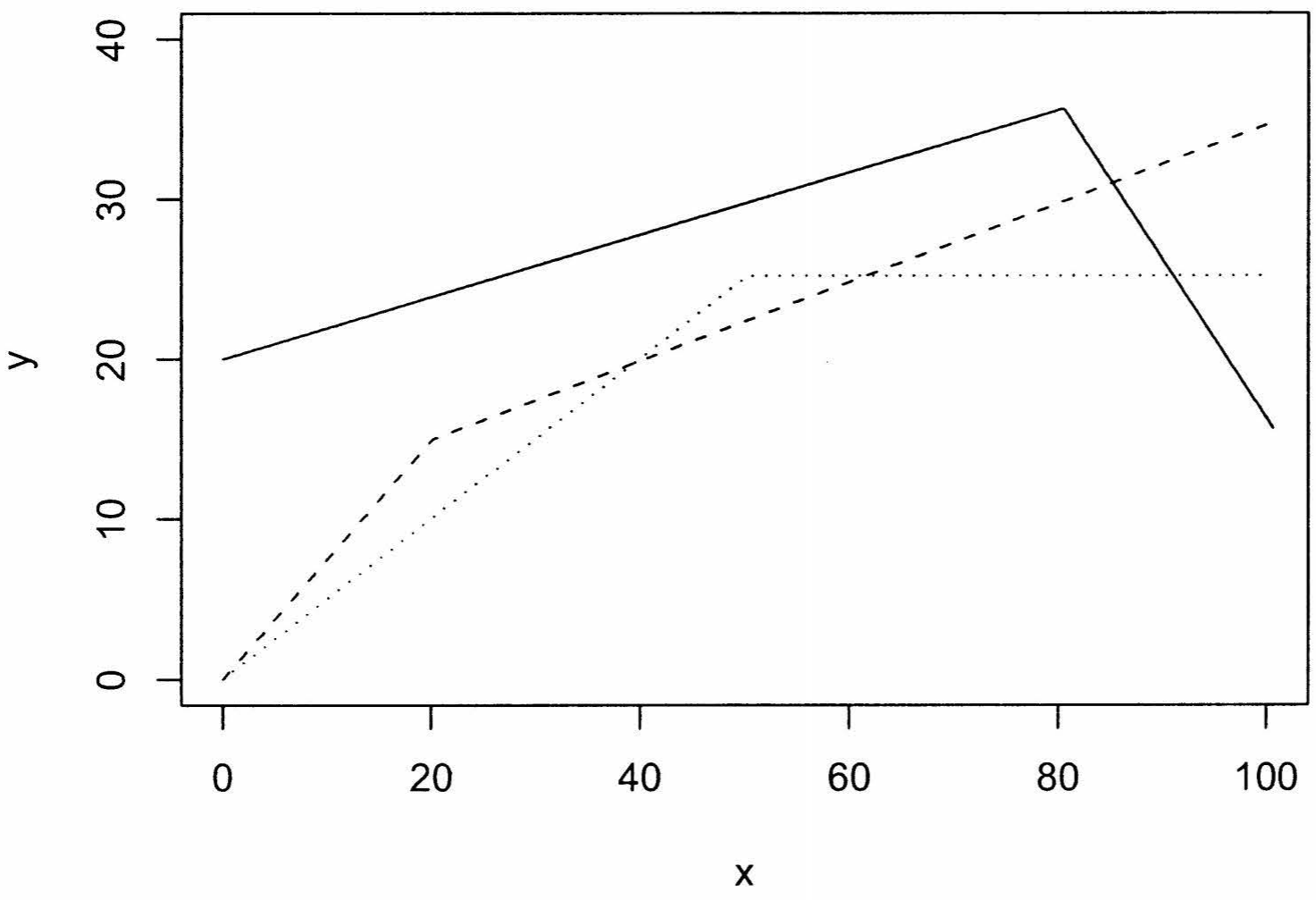

Figure 3.1: Segmented (3.1, solid line), doorhinge (3.7, dashed line) and hockey-stick (3.6, dotted line) regression models.

an intermediate case between the segmented regression, and the hockey-stick model. Note that the constraint in (3.6) implies $\beta_1 \delta = \alpha_2$. For (3.7), the constraint is $\beta_1 \delta = \alpha_2 + \beta_2 \delta$.

3.2.2 Estimation if Change-point (δ) unknown

If the change-point is unknown, the segmented regression model (3.1) is non-linear. Parameter estimation could be conducted using numerical analysis techniques, but these may have problems with local minima or false convergence (Julious 2001). Further, the lack of smoothness in the model limits the utility of such techniques. The simple framework described in the previous section is dependent on knowing the value of the change-point. For unknown change-point, one could conduct a grid-search over δ (Lerman, 1980). The value of δ which results in the best fit could be used as the estimated change-point, $\hat{\delta}$. The estimated regression parameters $\hat{\boldsymbol{\beta}}$ are those corresponding to the model fit using $\hat{\delta}$ (as a fixed parameter). Another approach is the Julious algorithm (see §2.1). We show two examples to illustrate the algorithm in the next section.

3.2.3 Illustrative Examples using Julious Algorithm

The Julious algorithm successively divides the ordered covariates x_i into two groups: $x_1, \ldots, x_t$ and $x_{t+1}, \ldots, x_T$. For each such grouping (i.e. for all possible t), the fit of the regression model is assessed to determine the plausibility that $\hat{\delta}$ belongs to the interval $[x_t, x_{t+1}]$.

Example (i) Julious' Data

Julious (2001) used the algorithm to estimate the parameters of the segmented regression model assuming normally distributed errors (3.1) to model the relationship between the volume of carbon dioxide exhaled and oxygen inhalation during exercise. Section 5.3 summarizes this and additional analysis of this data set. Table 3.1 shows the algorithm computations for this data set. The rows of Tables 3.1 and 3.2 represent the results for each of the possible data groupings.

The columns of the table are: x_t, x_{t+1}, the endpoints of the two data subsets ($x_1, \ldots, x_t$ and $x_{t+1}, \ldots, x_T$) considered; RSS is the residual sum of squares for an unconstrained model for any fixed δ within (x_t, x_{t+1}), and $\hat{\delta}$ is the intersection point of the model segments from the unconstrained fit. "Con.", i.e. constrained is true if $x_t \leq \hat{\delta} \leq x_{t+1}$. The Julious algorithm requires fitting constrained models for both $\delta = x_t$ and $\delta = x_{t+1}$. The column "x_t Best" is true if $\text{RSS}_{(\delta=x_t)} < \text{RSS}_{(\delta=x_{t+1})}$, and RSS_2 is the residual sum of squares of the best fitting of these two constrained models. Finally, r_i gives the rank of RSS, the unconstrained residual sum of squares. Since we are considering the segmented regression model, we consider $t = 2$ to $(T-2)$, with repeated covariate values omitted.

Having computed the required model fits from the data (Table 3.1), we can step through the Julious algorithm (Figure 2.1.1) to determine the estimate of the change-point and the regression parameters. Step 2 of the Julious algorithm requires testing if $x_t \leq \hat{\delta} \leq x_{t+1}$. If this test is true (column

x_t	x_{t+1}	RSS	$\hat{\delta}$	Con.	x_t Best	RSS_2	r_i
21.5	24.8	1.160	26.952	F	F	1.194	20
24.8	26.2	1.104	26.930	F	F	1.116	18
26.2	27.4	1.090	28.402	F	F	1.102	16
27.4	27.9	1.104	28.971	F	F	1.127	17
27.9	29.2	0.816	31.458	F	F	0.892	15
29.2	31.1	0.727	33.316	F	F	0.717	12
31.1	32.6	0.750	33.465	F	F	0.742	14
32.6	34.6	0.740	34.608	F	F	0.720	13
34.6	34.9	0.468	34.604	T	F	0.466	10
34.9	35.2	0.452	36.293	F	F	0.457	9
35.2	36.3	0.416	37.219	F	F	0.420	6
36.3	37.6	0.403	38.414	F	F	0.409	4
37.6	**40.1**	**0.389**	**39.463**	**T**	**F**	**0.391**	**1**
40.1	42.7	0.391	40.469	T	F	0.391	2
42.7	43.4	0.399	40.951	F	F	0.411	3
43.4	44.2	0.409	41.422	F	F	0.423	5
44.2	47.9	0.418	41.644	F	F	0.438	7
47.9	48.1	0.418	41.059	F	F	0.563	8
48.1	48.4	0.476	41.489	F	F	0.580	11
48.4	49.9	1.157	49.894	T	F	1.069	19
49.9	51.7	1.192	49.902	T	F	1.192	21
51.7	51.8	1.241	50.488	F	F	1.321	22
51.8	54.9	1.443	52.031	T	F	1.431	23
54.9	55.5	1.499	53.072	F	F	1.648	24
55.5	57.0	1.697	52.511	F	F	1.882	25
57.0	57.9	1.849	52.357	F	F	2.148	26
57.9	58.2	1.975	51.319	F	F	2.321	27
58.2	58.3	2.122	53.273	F	F	2.395	28
58.3	59.5	2.342	51.722	F	F	2.543	29
59.5	59.7	2.434	52.270	F	F	2.579	30

Table 3.1: Intermediate calculations for Julious algorithm applied to data from Julious (2001).

"Con." in Table 3.1), the RSS is considered to be a restricted RSS, otherwise RSS is considered an unrestricted RSS. This test is satisfied in six (x_t, x_{t+1})

intervals. At Step 3 of the Julious algorithm, we test if the minimum restricted RSS $\leq$ the minimum unrestricted RSS. Since Con.=T for the row corresponding to $r_i = 1$, we have that the minimum RSS is in fact restricted, and this test is satisfied, thus we proceed to the terminal step of the Julious algorithm. The "best model" is the one corresponding to the smallest restricted RSS (RSS=0.389), and the parameters from this model, $\hat{\delta} = 39.463$, $\hat{\boldsymbol{\beta}} = (0.0765, 0.0423, -1.6594, 0.0863)^T$, are the least-squares parameter estimates. The row in Table 3.1 in boldface indicates the best model.

Example (ii) Simulated Data

We now consider a second example, fitting the hockey-stick model to a simulated data set. Twenty-five data points were computed assuming the true model parameters are $\boldsymbol{\beta} = (\beta_1, \alpha_2)^T = (2, 100)^T$, with $\delta = 50$. Normally distributed errors $\varepsilon_i \sim N(0, 9)$ were used to generate random observations. Table 3.2 lists the results from fitting the various constrained and unconstrained models. The data and the fitted hockey-stick model using Julious' algorithm are presented in Figure 3.2. Having completed Step 1 of the Julious algorithm, observe Con.=T in only one (x_t, x_{t+1}) interval. At this point, the minimum restricted residual sum of squares (RSS=238.1; see row with Con.=T) is larger than the minimum unrestricted RSS (RSS=167.5, see the row with $r_i = 1$). Thus the iterative part of the algorithm is required (Step 4). In the row with $r_i = 1$, we see from x_t Best and RSS_2 that the best fitting restricted model between the constrained models with $\delta = x_t$ and $\delta = x_{t+1}$ is the model assuming $\delta = x_{t+1}$, with associated RSS=372.8. Thus the minimum restricted RSS remains as RSS=238.1. We eliminate the row

x_t	x_{t+1}	RSS	$\hat{\delta}$	Con.	x_t Best	RSS_2	r_i
3.85	7.69	19343.3	26.12	F	F	20098.1	24
7.69	11.54	15185.8	37.40	F	F	16889.1	23
11.54	15.38	11944.3	38.22	F	F	14042.3	22
15.38	19.23	8830.1	42.57	F	F	11363.6	21
19.23	23.08	6401.9	44.20	F	F	8944.7	18
23.08	26.92	4098.0	47.28	F	F	6698.2	15
26.92	30.77	2900.7	46.60	F	F	4845.3	13
30.77	34.62	1603.3	48.52	F	F	3229.0	10
34.62	38.46	984.6	48.69	F	F	1978.7	8
38.46	42.31	395.8	50.06	F	F	1001.9	5
42.31	46.15	167.5	50.51	F	F	372.8	1
46.15	**50.00**	**238.1**	**49.31**	**T**	**F**	**243.9**	**3**
50.00	53.85	236.6	49.28	F	T	243.9	2
53.85	57.69	261.1	49.72	F	T	497.3	4
57.69	61.54	407.4	50.70	F	T	1062.2	6
61.54	65.38	602.2	51.63	F	T	1852.8	7
65.38	69.23	1156.1	53.40	F	T	2837.9	9
69.23	73.08	2001.7	55.76	F	T	3903.1	11
73.08	76.92	2833.4	57.88	F	T	4968.5	12
76.92	80.77	3793.5	60.14	F	T	6024.7	14
80.77	84.62	4785.3	62.26	F	T	7028.5	16
84.62	88.46	6107.2	65.50	F	T	7950.3	17
88.46	92.31	6978.5	66.36	F	T	8700.0	19
92.31	96.15	8354.2	69.46	F	T	9310.5	20

Table 3.2: Intermediate calculations for Julious algorithm applied to simulated data.

corresponding to $r_i = 1$ from consideration. Now the minimum unrestricted RSS=236.6 (row with $r_i = 2$). Since the minimum restricted RSS=238.1, we again require Step 4 of the algorithm. The best fitting constrained model between $\delta = x_t$ and $\delta = x_{t+1}$ is that of $\delta = x_t$, having RSS=243.9 (see x_t Best and RSS_2 in the $r_i = 2$ row). The minimum restricted RSS is unchanged from 238.1. Returning to Step 3 of the Julious algorithm, the test is still

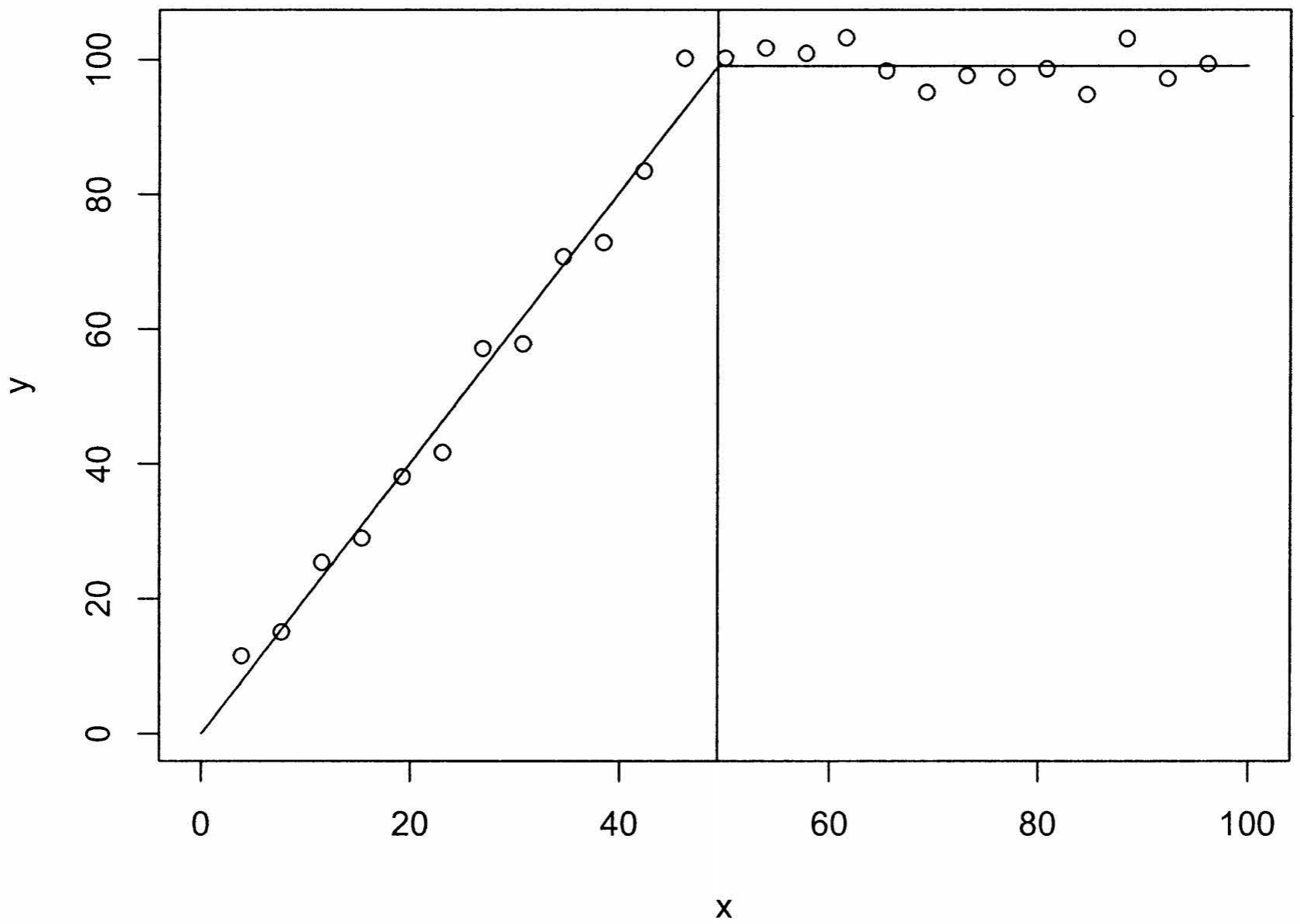

Figure 3.2: Data and fitted model of Example (ii).

false, since the minimum unrestricted RSS (236.6) is less than the minimum restricted RSS (238.1). Thus we continue and the $r_i = 2$ row is eliminated from consideration. This leaves minimum unrestricted RSS of 261.1 (see $r_i = 4$; we skip $r_i = 3$ because Con.=T, and this is a restricted model). At this point, we have minimum restricted RSS=238.1 and minimum unrestricted RSS=261.1. Hence our test at Step 3 of the Julious algorithm is satisfied. Thus, we proceed to Step 5 of the Julious algorithm, with the "best model" being that corresponding to RSS=238.1, with $\hat{\delta} = 49.31$, with $\boldsymbol{\beta} = (2.009, 99.049)^T$. The row in Table 3.2 in boldface indicates the best

model.

3.3 Estimation of Parameters - Weighting and Robustification

3.3.1 Constrained, Weighted Least Squares Estimates

We now discuss estimators $\hat{\boldsymbol{\beta}}$ which attempt to prevent outlying points from exerting undue influence during estimation. In order to prevent unrealistic parameter estimates arising as a result of outliers, we weight each observation and perform weighted (or generalized) least-squares (WLS, Draper and Smith, 1981). If $w_i = 1 \; \forall i$, then these estimators become the LS estimators of the previous section. In WLS, the model is:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{with } \varepsilon_i \sim N(0, \frac{\sigma^2}{w_i}). \tag{3.8}$$

We consider only the case of unequal variances with uncorrelated errors - so that $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. From (3.8), observe that the weight for each residual is inversely proportional to its variance. The matrix $W = \text{diag}(\frac{1}{w_1}, \ldots, \frac{1}{w_T})$ is the weight matrix, where w_i is the weight corresponding to (x_i, y_i). An equivalent formulation to (3.8) is:

$$\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \tag{3.9}$$

with $\boldsymbol{y}^* = W^{1/2}\boldsymbol{y}$, $X^* = W^{1/2}X$, and $\boldsymbol{\varepsilon}^* \sim N(0, \sigma^2 I)$, as:

$$\begin{aligned} var(\boldsymbol{\varepsilon}^*) &= var(\boldsymbol{y}^*) \\ &= var(W^{1/2}\boldsymbol{y}) \\ &= W^{1/2} var(\boldsymbol{y}) W^{1/2} \\ &= W^{1/2}\sigma^2 W^{-1} W^{1/2} \\ &= \sigma^2. \end{aligned}$$

(Note $W^{1/2^T} = W^{1/2}$ since W is diagonal.) Thus, WLS is a LS problem using the transformed data $\boldsymbol{y}^*$ and X^*. Therefore, the WLS estimate of the parameter vector $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{WLS}$, is the LS estimate of (3.9):

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{WLS} &= (X^{*T}X^*)^{-1}X^{*T}\boldsymbol{y}^* \\ &= (X^T W^{1/2^T} W^{1/2} X)^{-1} X^T W^{1/2^T} W^{1/2} \boldsymbol{y} \\ &= (X^T W X)^{-1} X^T W \boldsymbol{y} \end{aligned}$$

The weighted RSS for the WLS model follows from (3.5) :

$$RSS = \hat{\boldsymbol{\varepsilon}}^T W^{-1} \hat{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_{WLS})^T W^{-1} (\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_{WLS}). \qquad (3.10)$$

Constrained, weighted least-squares estimates are also required to ensure the model segments meet at the change-point. Following directly from (3.4), replacing X by $W^{1/2}X$, the WLS estimates constrained to satisfy the constraint $H\boldsymbol{\beta} = 0$ are:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_c &= \hat{\boldsymbol{\beta}}_{WLS} - (X^T W X)^{-1} H^T [H(X^T W X)^{-1} H^T]^{-1} H \hat{\boldsymbol{\beta}}_{WLS} \quad (3.11)\\ &= (I_p - A)\hat{\boldsymbol{\beta}}_{WLS}\end{aligned}$$

where I_p is the $p \times p$ identity matrix, $A = (X^T W X)^{-1} H^T [H(X^T W X)^{-1} H^T]^{-1} H$, and $\hat{\boldsymbol{\beta}}_{WLS}$ is the unconstrained WLS estimate.

3.3.2 Weighting the Observations

Use of appropriate weights can introduce robustness to parameter estimation in the linear model. In this research, we focus on weighting the regression residuals, an approach commonly followed in robust regression (Neter et al., 1996, Huber, 1977) and iteratively re-weighted least-squares (Beaton and Tukey, 1974). Estimates are computed by minimizing the weighted RSS (3.10). Numerous weighting functions have been proposed in practice; some examples are the Cauchy, bi-square, and Huber weighting functions (Heiberger and Becker, 1992, Figure 3.3).

For this research, Huber's weighting function is preferred. Huber's weighting function is computed as:

$$w(z) = \begin{cases} 1 & \text{if } |z| < c \\ \frac{c}{|z|} & \text{if } |z| \geq c, \end{cases} \quad (3.12)$$

where c is a fixed constant. In this practicum, weights are computed using scale-standardized residuals, $z_i = \varepsilon_i / \text{MAD}(\boldsymbol{\varepsilon})$, where MAD is the median

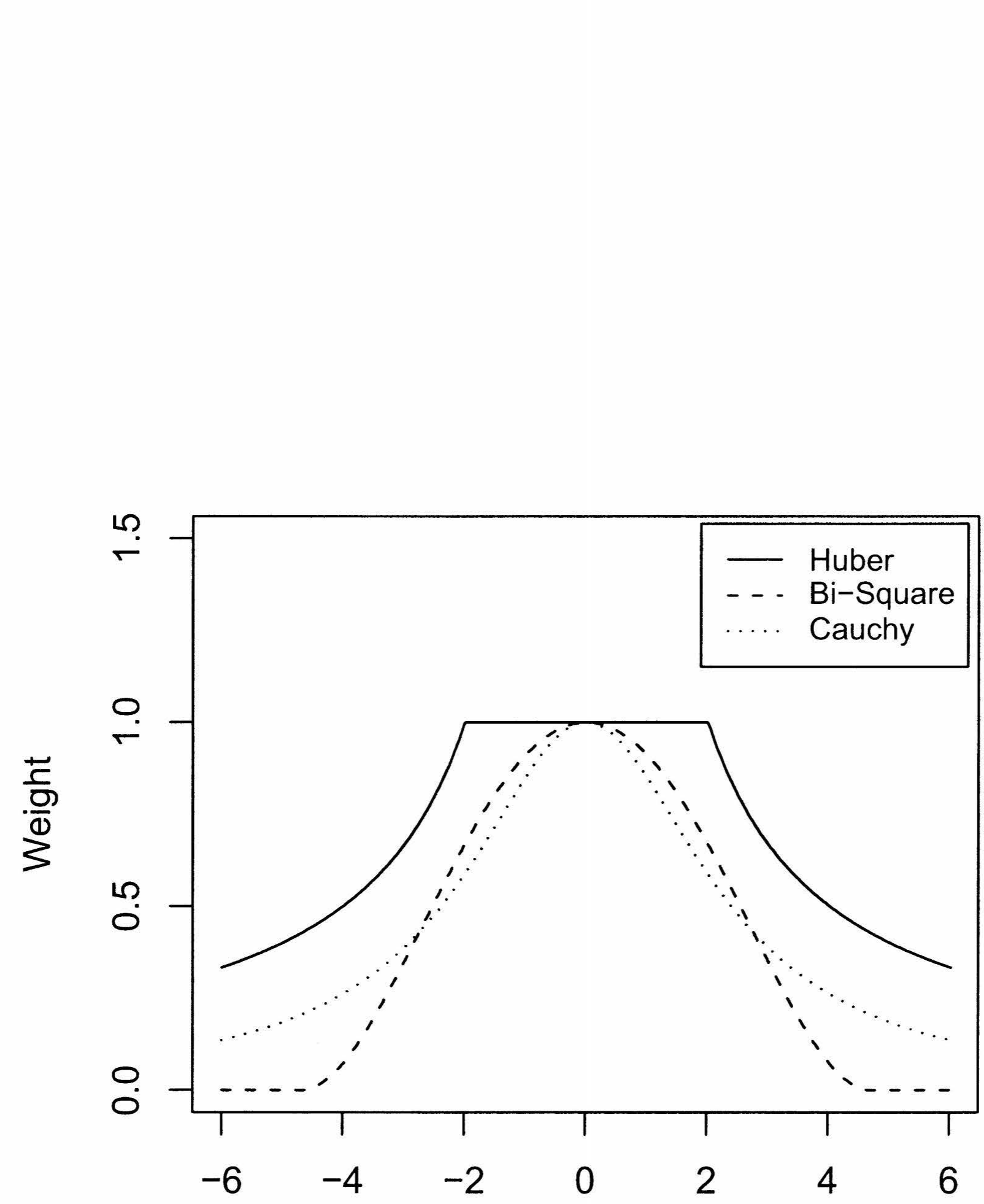

Figure 3.3: Comparison of three weighting functions.

absolute deviation function, a robust estimate of scale (Hampel *et al.*, 1986). Notice that Huber weights are constant provided $|z| < c$. Only if $|z| \geq c$ are the weights reduced. Huber's weighting function was selected over alternative functions which downweight *every* residual. With Huber's weighting function, only residuals large in magnitude are down-weighted. A weighting function proposed by Hampel (Heiberger and Becker, 1992) has similar properties to Huber's, but is dependent on three constants.

Although decreasing values of c lead to increased protection against outliers and thus a more robust estimate, a robustness/efficiency trade-off exists. As robustness is increased, the cost is loss of efficiency in the estimators. Observe that as $c \to \infty$, observations are equally weighted, and LS estimators are obtained. As $c \to 0$, all weight is assigned to a single point, and the weighted estimate reduces to LMS.

3.3.3 Weighting the Julious Algorithm and IRLS

To offer protection against outliers during estimation, iteratively re-weighted least squares (IRLS, Beaton and Tukey, 1974) is employed. We combine IRLS with the Julious algorithm, to produce an iteratively re-weighted Julious algorithm. During each step of the iterative process, the Julious algorithm is applied using WLS. Each call to the Julious algorithm uses fixed weights. Weights are updated independent of the Julious algorithm, once the estimates $\hat{\boldsymbol{\beta}}$ (and hence residuals $\hat{\boldsymbol{\varepsilon}}$) from the Julious algorithm have been determined. This process continues, continually updating the weights.

It is crucial that the weights remain fixed within the Julious algorithm. One could use IRLS to fit "all possible unconstrained two-line models" in the first step of the Julious algorithm, iterating to convergence for each of these fits. However, the RSS for each model fit (with IRLS) would no longer be comparable, and the Julious algorithm is not applicable. Fixing the weights within each call to the Julious algorithm allows comparison of model fits and estimates of the change-point for each iteration step. At each step, the

weights were scaled so that the sum of the weights equals the sample size. The convergence criteria is a subjective choice in IRLS. Some options are to iterate until convergence (to some fixed tolerance) in the RSS, or to iterate until convergence in parameters is attained. Figure 3.4 details the iterative algorithm used.

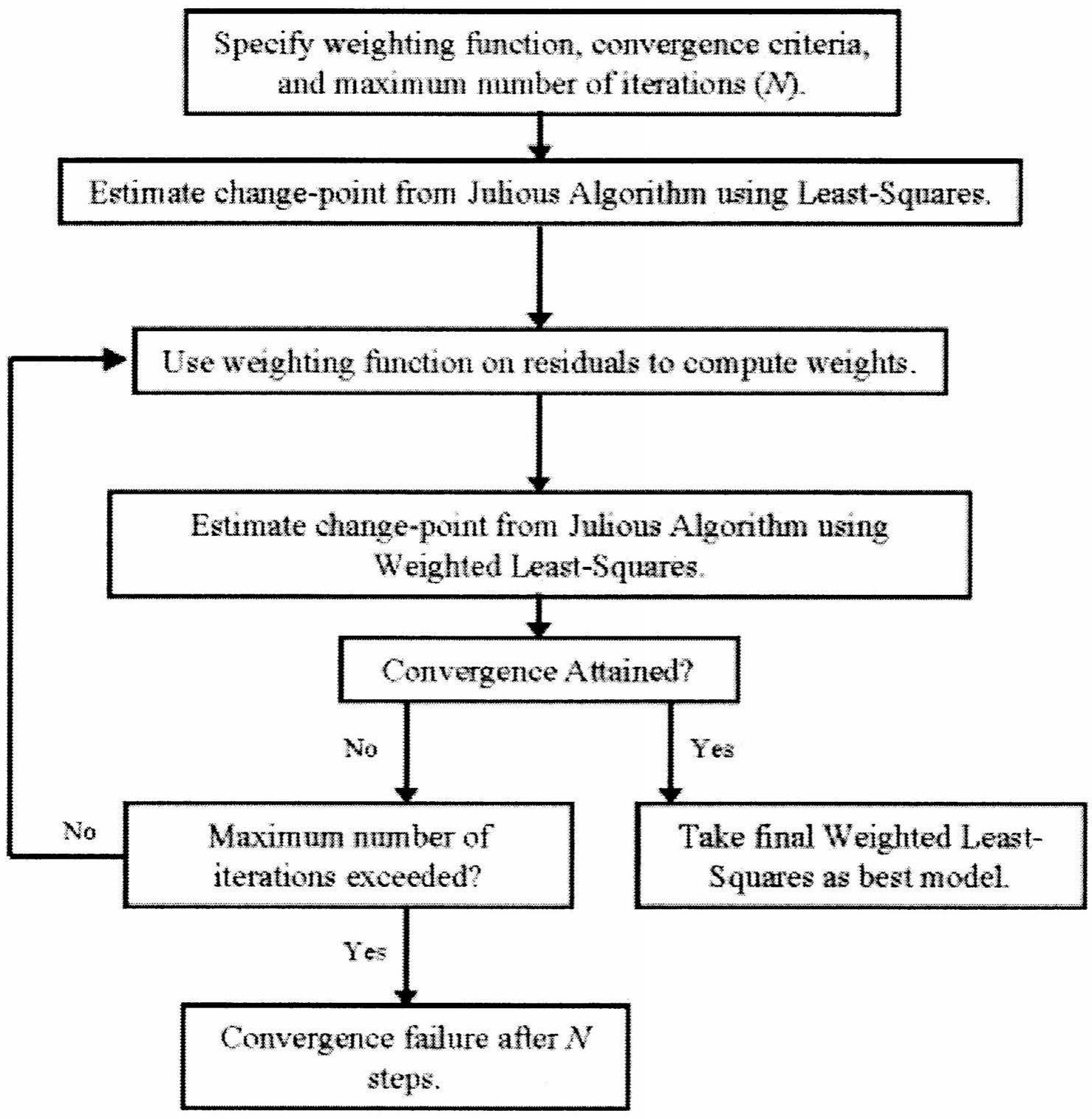

Figure 3.4: Algorithm to iteratively re-weight Julious' Algorithm.

3.4 Estimation of parameters assuming log-normal error

3.4.1 Hockey-Stick Model

We now consider least-squares estimation for the hockey-stick model assuming errors are lognormally distributed. To begin, assume the change-point is known. The model is:

$$E(y_i) = \begin{cases} \beta_1 x_i e^{\sigma^2/2-1} & \text{for } x_i \leq \delta \\ \alpha_2 e^{\sigma^2/2-1} & \text{for } x_i > \delta, \end{cases} \tag{3.13}$$

or, on the log-scale, we may write:

$$\log(y_i) = \begin{cases} \log(\beta_1) + \log(x_i) + \varepsilon_i & \text{for } x_i \leq \delta \\ \log(\alpha_2) + \varepsilon_i & \text{for } x_i > \delta, \end{cases} \tag{3.14}$$

where $\varepsilon_i \sim N(0, \sigma^2)$. Note that $\log(x)$ has slope fixed at 1. As a result, each segment of the model is constant:

$$\begin{aligned} \Rightarrow \log(y_i) - \log(x_i) &= \log(\beta_1) + \varepsilon_i \quad \text{for } x_i \leq \delta \\ \log(y_i) &= \log(\alpha_2) + \varepsilon_i \quad \text{for } x_i > \delta. \end{aligned}$$

Following §3.1, suppose there are T observations, t of which are modeled by the first linear segment, i.e. define t such that $x_t \leq \delta$ and $x_{t+1} > \delta$. To formulate a linear model, we write:

$$\boldsymbol{y}_* = \begin{pmatrix} \log(\frac{y_1}{x_1}) \\ \vdots \\ \log(\frac{y_t}{x_t}) \\ \log(y_{t+1}) \\ \vdots \\ \log(y_T) \end{pmatrix}, X = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}, \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \log(\beta_1) \\ \log(\alpha_2) \end{pmatrix}.$$

Since $X^T X = \begin{pmatrix} t & 0 \\ 0 & T-t \end{pmatrix}$, the least-squares estimate of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}_* = \begin{pmatrix} \frac{1}{t} \sum_{j=1}^{t} \log(\frac{y_j}{x_j}) \\ \frac{1}{T-t} \sum_{j=t+1}^{T} \log(y_j) \end{pmatrix}, \tag{3.15}$$

which is simply the mean of the elements of $\boldsymbol{y}_*$ associated with each model segment.

To constrain the model segments to meet at the change-point, we require:

$$\begin{aligned} \beta_1 \delta &= \alpha_2 \\ \Leftrightarrow \log(\beta_1) + \log(\delta) &= \log(\alpha_2) \\ \Leftrightarrow \log(\beta_1) - \log(\alpha_2) &= -\log(\delta). \end{aligned}$$

If the model is constrained to meet at the change-point, and if δ is known, we can reparameterize (3.14) replacing α_2 by $\beta_1 \delta$. Write:

$$\log(y_i) = \begin{cases} \log(\beta_1) + \log(x_i) + \varepsilon & \text{for } x_i \leq \delta \\ \log(\beta_1) + \log(\delta) + \varepsilon & \text{for } x_i \geq \delta, \end{cases} \tag{3.16}$$

$$\begin{aligned} \Leftrightarrow \quad \log(y_i/x_i) &= \log(\beta_1) + \varepsilon \quad \text{for } x_i \leq \delta \\ \log(y_i/\delta) &= \log(\beta_1) + \varepsilon \quad \text{for } x_i \geq \delta. \end{aligned}$$

Then clearly, the MLE is:

$$\widehat{\log(\beta_1)} = \frac{1}{T}\left(\sum_{i \leq t} \log(\frac{y_i}{x_i}) + \sum_{i > t} \log(\frac{y_i}{\delta})\right). \tag{3.17}$$

Below, we verify this result using the theory of linear models as in §3.2 and 3.3.

In terms of the parameter vector $\boldsymbol{\beta} = (\log(\beta_1), \log(\alpha_2))^T$, we may write the constraint as $\boldsymbol{h}^T\boldsymbol{\beta} = k$ with $\boldsymbol{h} = (1, -1)^T$, and $k = -\log(\delta)$.

The least-squares estimate of $\boldsymbol{\beta}$ subject to $H\boldsymbol{\beta} = k$ (Rencher, 2000) is given by:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \hat{\boldsymbol{\beta}}_{LS} - (X^TX)^{-1}H^T[H(X^TX)^{-1}H^T]^{-1}(H\hat{\boldsymbol{\beta}}_{LS} - k) \tag{3.18}$$

where $\hat{\boldsymbol{\beta}}_{LS}$ is the unconstrained LS estimate. For a single constraint, $H = \boldsymbol{h}^T$, and thus:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \hat{\boldsymbol{\beta}}_{LS} - \frac{(X^TX)^{-1}\boldsymbol{h}(\boldsymbol{h}^T\hat{\boldsymbol{\beta}}_{LS} - k)}{\boldsymbol{h}^T(X^TX)^{-1}\boldsymbol{h}} \tag{3.19}$$

Since

$$(X^TX)^{-1}\boldsymbol{h} = \begin{pmatrix} \frac{1}{t} & 0 \\ 0 & \frac{1}{T-t} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{t} \\ \frac{-1}{T-t} \end{pmatrix},$$

$$\boldsymbol{h}^T(X^TX)^{-1}\boldsymbol{h} = \begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}\frac{1}{t} \\ \\ \frac{-1}{T-t}\end{pmatrix}, \text{ and}$$

$$\begin{aligned}\boldsymbol{h}^T\hat{\boldsymbol{\beta}}_{LS} - k &= \begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}\widehat{\log(\beta_1)} \\ \widehat{\log(\alpha_2)}\end{pmatrix} - (-\log(\delta)) \\ &= \widehat{\log(\beta_1)} - \widehat{\log(\alpha_2)} + \log(\delta) \\ &= \frac{1}{t}\sum_{i=1}^{t}\log(\frac{y_i}{x_i}) - \frac{1}{T-t}\sum_{i=t+1}^{T}\log(y_i) + \log(\delta),\end{aligned}$$

it follows after some algebra that :

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \begin{pmatrix}\frac{1}{T}\sum_{i=1}^{t}\log(\frac{y_i}{x_i}) + \frac{1}{T}\sum_{i=t+1}^{T}\log(y_i) - \frac{T-t}{T}\log(\delta) \\ \frac{1}{T}\sum_{i=1}^{t}\log(\frac{y_i}{x_i}) + \frac{1}{T}\sum_{i=t+1}^{T}\log(y_i) + \frac{t}{T}\log(\delta)\end{pmatrix};$$

note the first element matches the result for $\widehat{\log(\beta_1)}$ in (3.17). As $\log(x_i)$ is an offset in the first model segment, it is obvious that the constraint is satisfied. Estimates of β_1 and α_2 are obtained by exponentiating the elements of $\boldsymbol{\beta}$. Note that the first two terms in each element of $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ equate to $\bar{\boldsymbol{y}}_*$.

If the observations are to be weighted, then the WLS estimates follow readily as in the normal-errors case. Referring to the LS estimates given by (3.15) and (3.18), we have for the unconstrained case:

$$\hat{\boldsymbol{\beta}}_{WLS} = (X^TWX)^{-1}X^TW\boldsymbol{y}_* = \begin{pmatrix}c_1\sum_{j=1}^{t} w_j\log(\frac{y_j}{x_j}) \\ c_2\sum_{j=t+1}^{T} w_j\log(y_j)\end{pmatrix} \qquad (3.20)$$

where $c_1 = \frac{1}{\sum_{j=1}^{t} w_j}$, and $c_2 = \frac{1}{\sum_{j=t+1}^{T} w_j}$. For the constrained, WLS estimates, we have:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \hat{\boldsymbol{\beta}}_{WLS} - (X^T W X)^{-1} H^T [H(X^T W X)^{-1} H^T]^{-1} (H\hat{\beta}_{WLS} - k). \quad (3.21)$$

Again, with the single constraint, $H = \boldsymbol{h}^T$, and (3.21) simplifies to:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \hat{\boldsymbol{\beta}}_{WLS} - \frac{(X^T W X)^{-1} \boldsymbol{h} (\boldsymbol{h}^T \hat{\beta}_{WLS} - k)}{\boldsymbol{h}(X^T W X)^{-1} \boldsymbol{h}}. \quad (3.22)$$

Noting that

$$X^T W X = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} \mathrm{diag}(w_1, \ldots, w_T) \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{t} w_j & 0 \\ 0 & \sum_{j=t+1}^{T} w_j \end{pmatrix},$$

we have $(X^T W X)^{-1} = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$.

Further, $(X^T W X)^{-1} \boldsymbol{h} = \begin{pmatrix} c_1 \\ -c_2 \end{pmatrix}$, $\boldsymbol{h}^T (X^T W X)^{-1} \boldsymbol{h} = c_1 + c_2$. Finally, using expression (3.20),

$$\boldsymbol{h}^T \hat{\boldsymbol{\beta}}_{WLS} - k = c_1 \sum_{j=1}^{t} w_j \log(\frac{y_j}{x_j}) - c_2 \sum_{j=t+1}^{T} w_j \log(y_j) + \log(\delta).$$

Substituting these expressions into (3.21), the constrained, weighted estimator is:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \begin{pmatrix} \frac{1}{\sum_1^T w_i}\sum_{i=1}^{t} w_i \log(\frac{y_i}{x_i}) + \frac{1}{\sum_1^T w_i}\sum_{i=t+1}^{T} w_i \log(y_i) - \left(\frac{\sum_{t+1}^T w_i}{\sum_1^T w_i}\right)\log(\delta) \\ \frac{1}{\sum_1^T w_i}\sum_{i=1}^{t} w_i \log(\frac{y_i}{x_i}) + \frac{1}{\sum_1^T w_i}\sum_{i=t+1}^{T} w_i \log(y_i) + \left(\frac{\sum_1^t w_i}{\sum_1^T w_i}\right)\log(\delta) \end{pmatrix}.$$

If the change-point is unknown and to be estimated, the Julious algorithm (or iteratively re-weighted Julious algorithm) is applied. See §3.2.2 and §3.3.3.

3.4.2 Estimation of parameters in other models

In the previous section, estimators are developed for the hockey-stick model assuming lognormal errors. However, if we wish to estimate the parameters of the doorhinge (3.7) or segmented regression (3.1) models assuming lognormally-distributed errors, we cannot ordinary use least-squares, since on the log-scale (see 3.14), these models are not linear in the parameters. As this practicum is focused on using LS and the Julious algorithm, alternate estimation methods would be required. Some options which may offer outlier protection are weighted maximum likelihood estimators (Field and Smith, 1994), or even robustified maximum likelihood estimators (Carroll and Ruppert, 1988).

Chapter 4

Simulation Studies

4.1 Introduction

Simulation studies were conducted to evaluate the estimation procedure, and to compare results obtained using least-squares (LS) and iteratively re-weighted least squares (IRLS). In order to assess the impact outliers have on each procedure, simulated data with and without outliers were used as input data for parameter estimation.

Details on the simulations and a summary of the results are presented in this chapter. The focus of the simulation work is on the estimation of the change-point, δ. The models examined were the normal and lognormal forms of the hockey-stick model (3.6) and (3.13).

4.2 Simulation Design

For each model, thirty-six simulations were conducted. There are four simulation parameters: sample size (N), the change-point (δ), the amount of noise used to generate the sample (σ^2), and the proportion of outliers in the data set (p). Sample size and change-point location each have three levels, whereas σ^2 and the proportion of outliers each have two levels.

Sample-size was varied in simulations to determine how the number of observations used in fitting the model affects the algorithm with respect to estimating the change-point. Contrasting results across three values of N also allows conjecture about the asymptotic behaviour of the estimators. Altering the change-point permits comparison of how estimation behaves when the change-point is in the center of the data (which we may expect to be the least difficult case) to that of off-center change-points. Different values of δ also permit study of how the constraint of passing through the origin using the hockey-stick model affects the LS and IRLS estimators. Two values of σ^2 were used to consider the effect of changing the amount of noise in the data. In the application for which the IRLS using Julious' algorithm was developed, stock-recruit modeling in fisheries science, there are often several outlying points. By including outliers in the simulation data sets ($p > 0$ cases), we can compare the LS and IRLS results, and also compare estimates to the analogous situation with no outliers ($p = 0$).

Each simulation has 3000 iterations, and the design points are equally spaced within $[0, 100]$. Each simulation iteration stores $\hat{\boldsymbol{\beta}}$, $\hat{\delta}$, and the residual sum of

squares (3.5) from both LS and IRLS. The IRLS estimates are computed using Huber's weighting function (3.12), with iterative re-weighting continuing until successive estimates of δ agree to three decimal places. The constant in Huber's weighting function was held constant in all simulations at $c = 2$. Increasing the value of the constant will lead to IRLS which is more "tolerant" of outlying points, and thus the difference between the LS and IRLS parameter estimates would decrease. Conversely, decreasing the value of c may increase the discrepancy between the LS and IRLS estimates. For this reason, and also to moderate the total number of simulations, Huber's c was held fixed. Using $c = 2$, the asymptotic relative efficiency of the sample mean to the robust estimate of the mean as estimated from normal data is 1.01 (see Table 5.2, Barnett and Lewis, 1995).

For each simulation, the mean squared error (MSE) and average bias of the estimated parameters were computed from (considering δ as an example):

$$\text{Bias}(\hat{\delta}) = \sum_{l=1}^{L} \frac{(\hat{\delta}_l - \delta)}{L} \tag{4.1}$$

$$\text{MSE}(\hat{\delta}) = \sum_{l=1}^{L} \frac{(\hat{\delta}_l - \delta)^2}{L} \tag{4.2}$$

To compute the variance of the simulation estimates of parameters, we use the fact that:

$$\begin{aligned} \mathrm{MSE}(\hat{\delta}) &= \mathrm{V}ar(\hat{\delta}) + \mathrm{Bias}(\hat{\delta})^2. \\ \Rightarrow \mathrm{V}ar(\hat{\delta}) &= \mathrm{MSE}(\hat{\delta}) - \mathrm{Bias}(\hat{\delta})^2. \end{aligned} \tag{4.3}$$

Outliers are included in half of the simulation runs. Outliers are generated according to the mixture model:

$$H = (1-p)F + pG. \tag{4.4}$$

This mixture distribution (Barnett and Lewis, 1995) is a simple combination of two distributions, F and G, with $P(\varepsilon \in G) = p$ representing the probability that ε is generated by G, the distribution used to generate outliers. In this research, G will be a distribution that is identical to F except with an inflated scale parameter.

4.3 Non-converging IRLS

IRLS estimation using the Julious algorithm did not converge for all simulated data sets. The number of non-convergent cases was infrequent (approximately 0.3% of all simulation runs), and as such these cases were excluded from further analyses and simulation summaries.

In the unconverging cases, problems arose with "cycling" of the regression weights. Formally, suppose at step i of IRLS, the estimated parameters are $\hat{\boldsymbol{\beta}}_i$, $\hat{\delta}_i$, with associated weights $\boldsymbol{w}_i = \{w_{i1}, \ldots, w_{iT}\}$, and residuals

$\boldsymbol{\varepsilon_i} = \{\varepsilon_{i1}, \ldots, \varepsilon_{iT}\}$. The weights $\boldsymbol{w}_i$ are used to obtain estimates in the $(i+1)^{\text{st}}$ step, from which $\hat{\boldsymbol{\varepsilon}}_{i+1}$ and subsequently $\boldsymbol{w}_{i+1}$ are computed. For the non-converging cases, the weights from step $(i+1)$ reproduce the estimated parameters from step i. As a result, $\hat{\boldsymbol{\varepsilon}}_{i+2} = \hat{\boldsymbol{\varepsilon}}_i$, and $\boldsymbol{w}_{i+2} = \boldsymbol{w}_i$, etc. The estimated parameters and weights will cycle indefinitely.

In practice, if cycling of weights were to occur, one possible solution would be to alter the IRLS termination criteria. One could also select another weighting function to use, or, alter the IRLS estimates at each step, computing a partial-step estimate, $\tilde{\boldsymbol{\beta}}_{i+1}$:

$$\boldsymbol{\beta}_{i+1} = \gamma\boldsymbol{\beta}_i + (1-\gamma)\boldsymbol{\beta}_{i+1}$$

where $\gamma \ni (0 \leq \gamma \leq 1)$ controls step-size, $\tilde{\boldsymbol{\beta}}_i$ is the estimate of step i, and $\hat{\boldsymbol{\beta}}_i$ is the WLS estimate from the $(i+1)^{\text{st}}$ step.

4.4 Simulation Models and Results

4.4.1 Hockey Stick Model, Normal Errors

A simulation exercise was conducted for the hockey-stick model with normal errors. In this simulation, the parameter values used were: $N = \{25, 50, 100\}$, $p = \{0, 0.15\}$, $\sigma^2 = \{1, 9\}$, and $\delta = \{25, 50, 75\}$. Quantities held fixed over simulation trials were: (Huber's) $c = 2$, $\tau = 5$, $F = N(0, \sigma^2)$, and $G = N(0, \tau^2\sigma^2)$. Using $\tau = 5$, 31% of the data generated from G will lie within two standard errors of the mean, when $\sigma^2 = 1$.

The true model from which random data were generated is:

$$y_i = \begin{cases} \beta_1 x_i + \varepsilon_i & x_i \leq \delta \\ \\ \alpha_2 + \varepsilon_i & x_i \geq \delta \end{cases},$$

with $\beta_1 = 1$, and $\alpha_2 = \delta$.

The random error ε_i comes from the mixture distribution H (4.4). Observe that τ is the contamination parameter. In the simulations with $p > 0$, τ controls the degree to which the variance of F is inflated.

Tables of bias and MSE for $\hat{\delta}$, $\hat{\beta}_1$, and $\hat{\alpha_2}$ are given in **Appendix A** for each simulation conducted. A run identification number (sequential from 1-36) is associated with each simulation to aide in discussion and identifying simulations in the summary graphs. Since assessing the estimation of the change-point (δ) is of primary interest, graphical summaries of Bias($\hat{\delta}$) and MSE($\hat{\delta}$) for the simulations are presented, along with discussion and conclusion.

Figure 4.1 displays two simulation datasets and the estimated regression lines using both the LS and IRLS methods. The data are from the run 4 simulation scheme, and the filled points demarcate those coming from distribution G in (4.4), which has inflated variance.

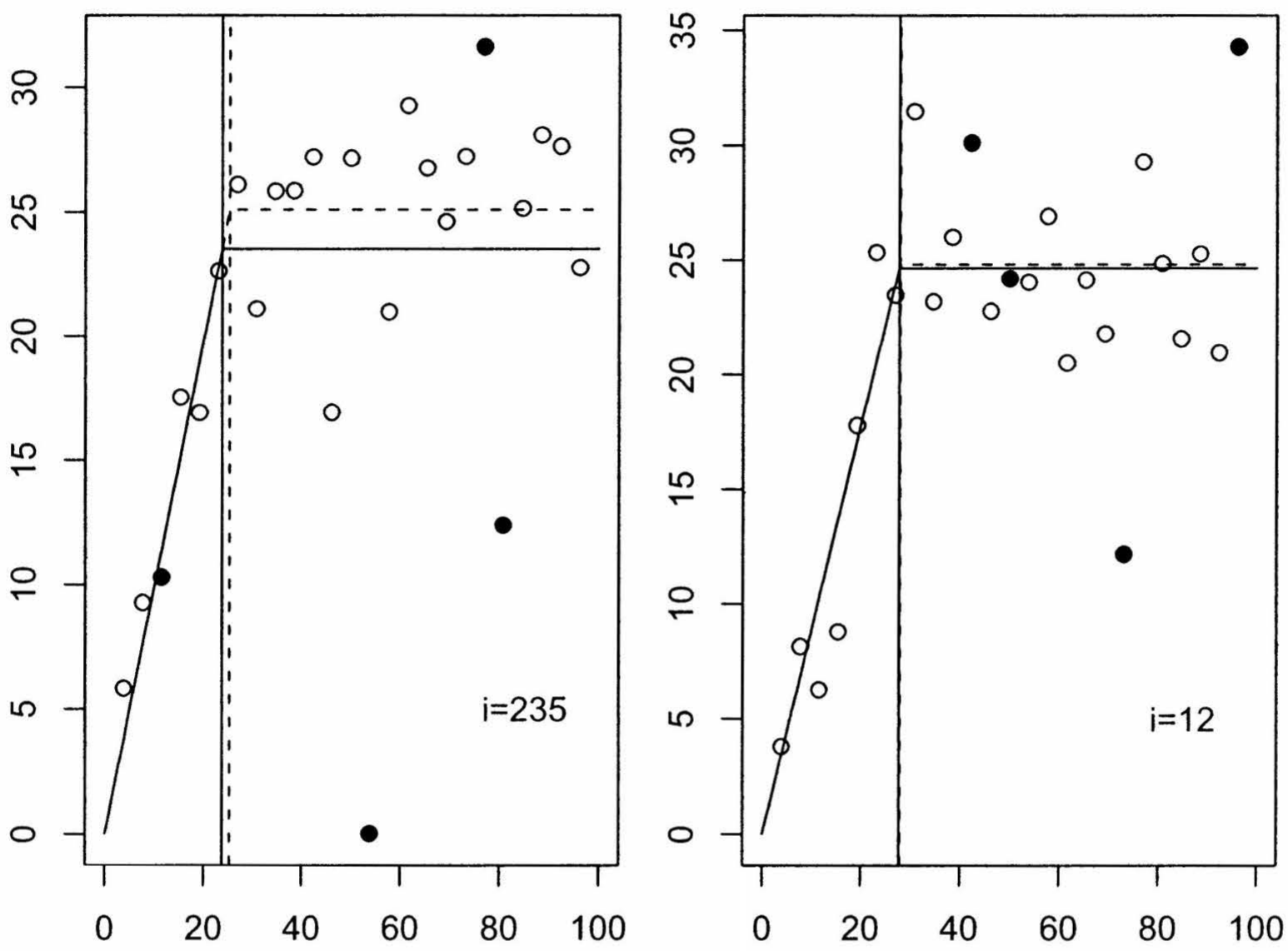

Figure 4.1: Two simulation datasets. Data come from run4, and numbers reference the iteration number. Filled points are realizations from the inflated variance distribution in 4.4.

Bias of $\hat{\delta}$

An examination of Bias($\hat{\delta}$) (Figure 4.2, Table A1) indicates all simulation results have values of Bias($\hat{\delta}$) which are small in magnitude. This suggests that the estimate of δ from both LS and IRLS within the Julious algorithm are unbiased. In the plot of Bias($\hat{\delta}$), it is obvious that three of the bias values are much larger in magnitude than in other simulations. These are $\text{Bias}_{IRLS}(\hat{\delta})$ in run # 4 (with simulation parameters $N = 25$, $\delta = 25$, $p = 0.15$, and

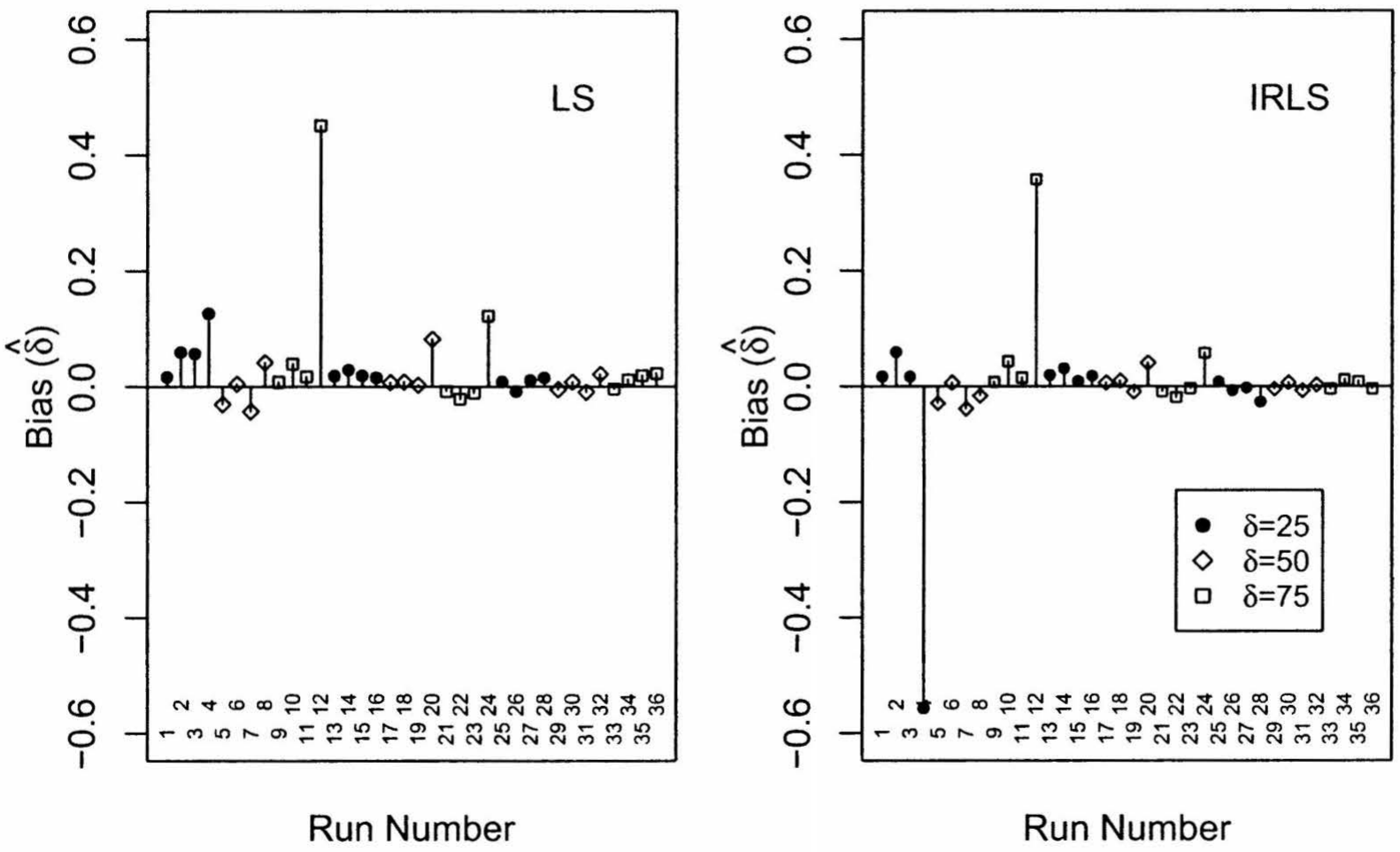

Figure 4.2: Comparison of Bias$(\hat{\delta})$ for simulation runs.

$\sigma^2 = 9$), and both $\text{Bias}_{LS}(\hat{\delta})$ and $\text{Bias}_{IRLS}(\hat{\delta})$ in run # 12 (simulation parameters $N = 25$, $\delta = 75$, $p = 0.15$ and $\sigma^2 = 9$). In each of these cases, two new simulation data sets were generated, and identical simulations were conducted. The results of the additional work corroborated the initial results. These two simulation scenarios are among the most difficult situations for estimation in the simulations: sample size is small, the true change-point is off-center, outliers are present, and σ^2 is large.

The results of run # 4 were examined in depth to gain insight into the large negative bias when using IRLS in contrast to the small positive bias observed when using the LS estimators. In Figure 4.3, we consider $\text{Bias}_{IRLS}(\hat{\delta})$ and $\text{Bias}_{LS}(\hat{\delta})$. Note that if $\text{Bias}_{LS}(\hat{\delta}) > 0$, iterative re-weighting reduces

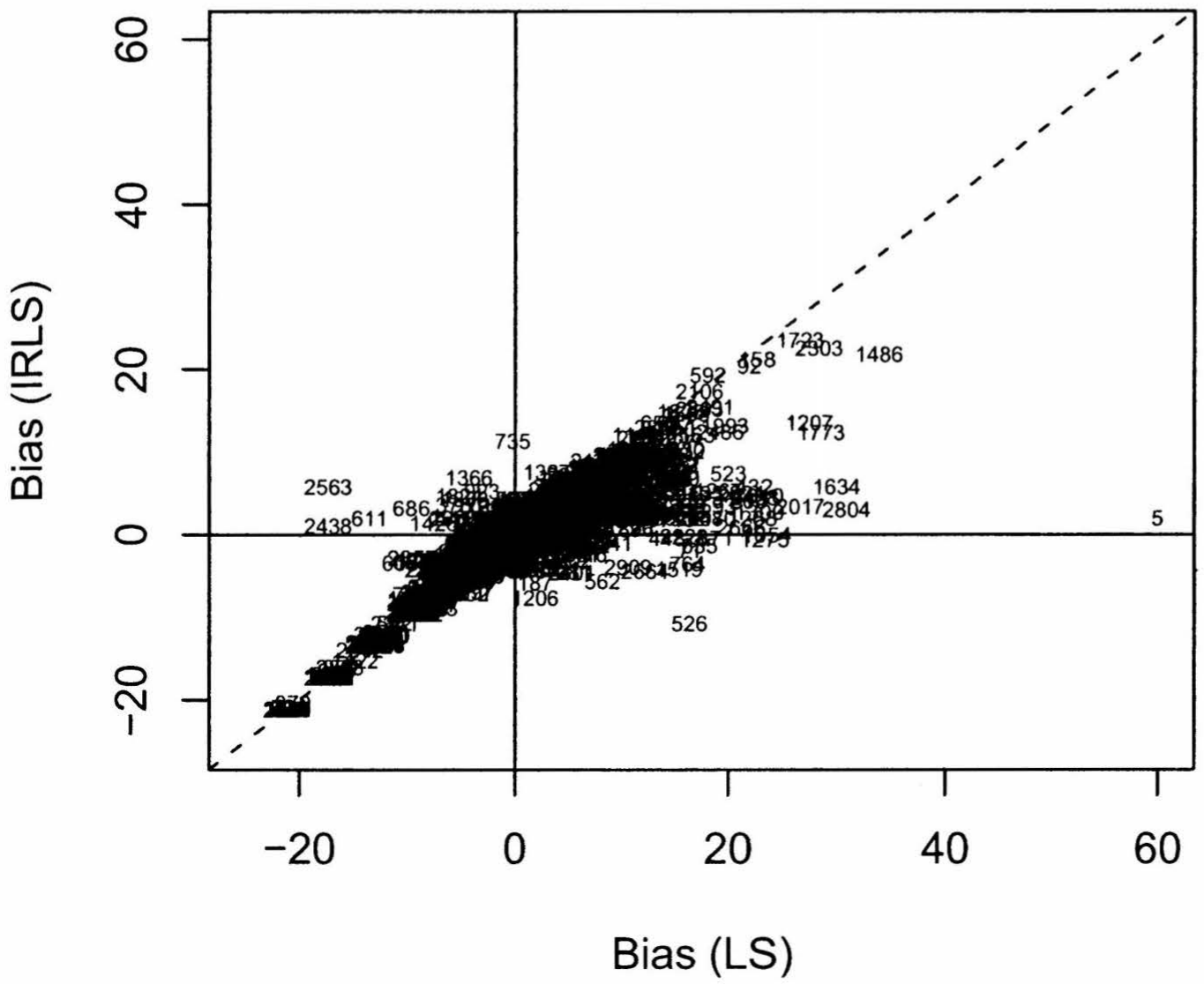

Figure 4.3: Bias($\hat{\delta}$) for simulation run # 4. (Numbers indicate simulation iteration.)

the bias in $\hat{\delta}$ in most cases (e.g. simulation iteration 5). However, in the $\text{Bias}_{LS}(\hat{\delta}) < 0$ cases, we find that reduction in bias by re-weighting is minimal or zero. This yields a mean simulation bias (IRLS) of -0.558 (see Figure 4.2). Thus $\text{Bias}_{LS}(\delta) < 0$ cases in which IRLS reduced the bias were compared to selected cases in which there was no improvement. For the cases exhibiting improvement, in general, the simulation datasets had extreme positive outliers to the left of the true change-point, or had large negative outliers towards the upper range of the data. Robust estimation via re-weighting

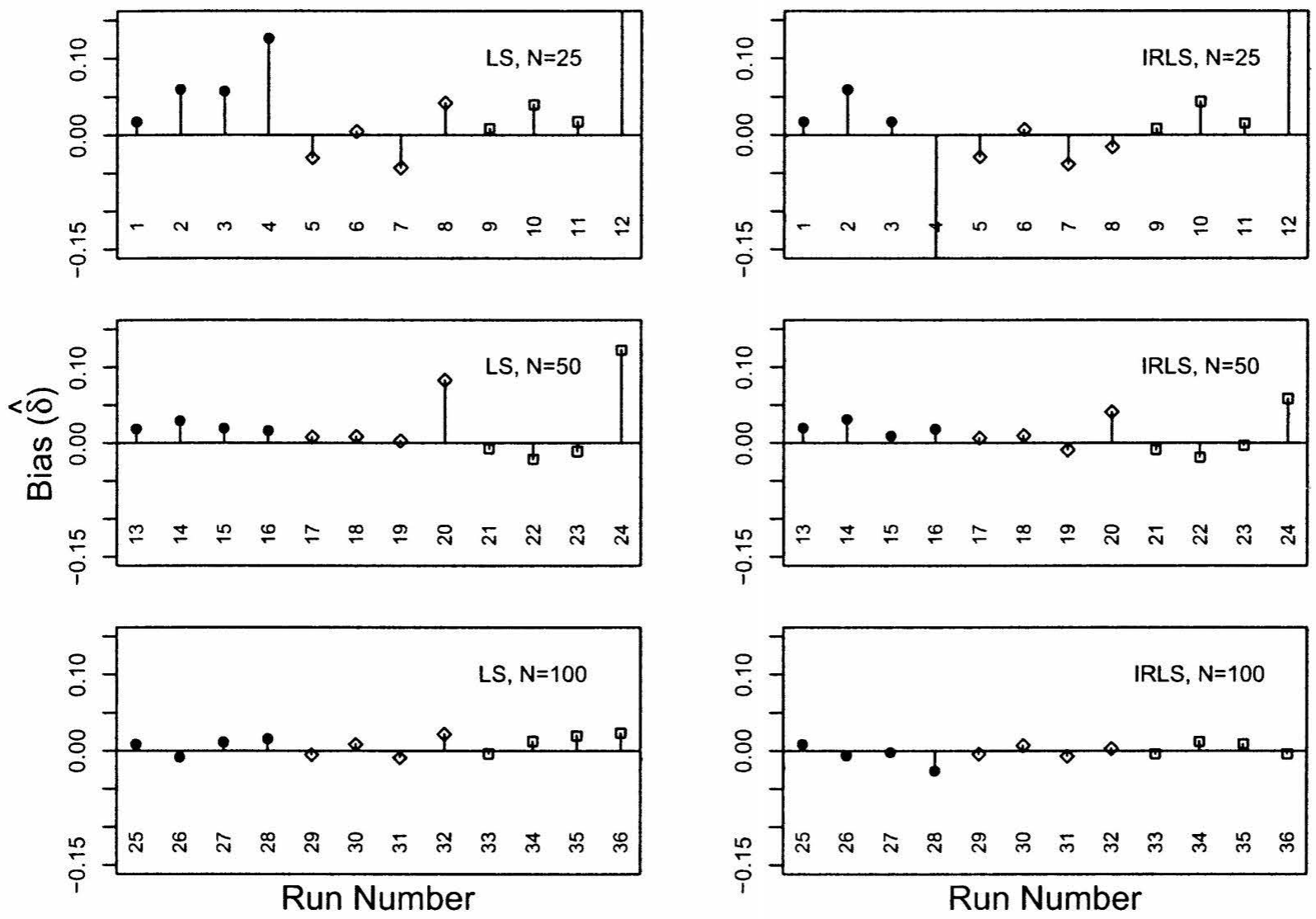

Figure 4.4: Comparison of Bias($\hat{\delta}$) for each value of N.

reduces the influence of these points, yielding reduced Bias($\hat{\delta}$). For the cases with little or no reduction in Bias($\hat{\delta}$), outliers were positioned such that re-weighting did not substantially alter the estimate of δ from the LS estimate.

If we compare Bias($\hat{\delta}$) at each level of N (Figure 4.4), as N increases, the size of Bias($\hat{\delta}$) decreases. Further, note that the magnitude of $\text{Bias}_{IRLS}(\hat{\delta})$ is usually either similar to or smaller than $\text{Bias}_{LS}(\hat{\delta})$, and are particularly smaller if $N = 25$. (Observe that the three large bias values discussed earlier exceed the scale of this and the subsequent bias plots.)

Examining Bias($\hat{\delta}$) for each value of δ (Figure 4.5), the true location of the change-point, the bias for the off-center change-points $\delta = \{25, 75\}$, is larger

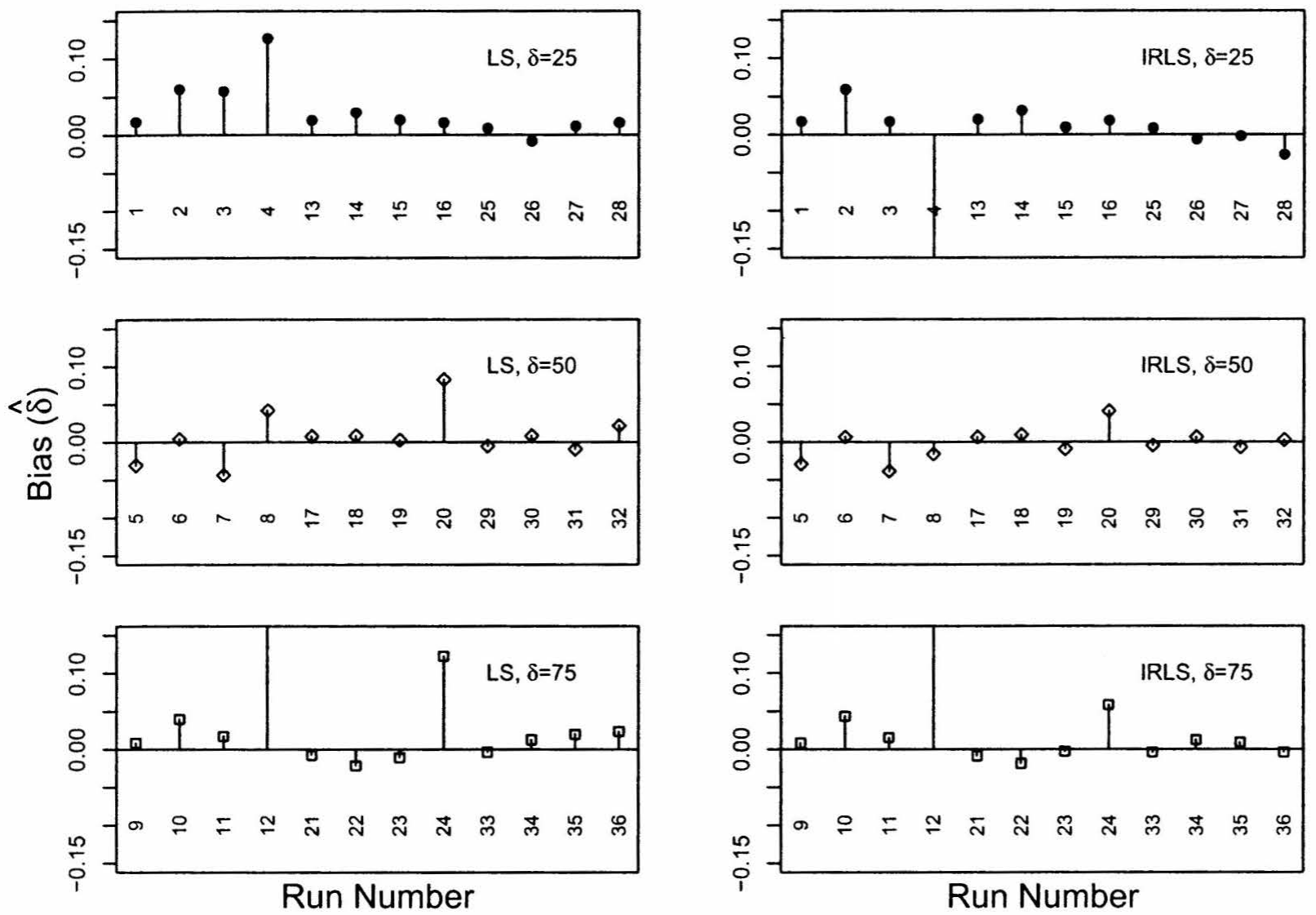

Figure 4.5: Comparison of Bias($\hat{\delta}$) across each value of δ.

than that for $\delta = 50$. Separating Bias($\hat{\delta}$) by p (Figure 4.6), it is obvious that Bias($\hat{\delta}$) is usually larger when outliers are present. In these cases, the use of IRLS generally reduces Bias($\hat{\delta}$) (with run #4 being a notable exception). Finally, inspection of Bias($\hat{\delta}$) at each level of σ^2 (Figure 4.7), we see that Bias($\hat{\delta}$) for $\sigma^2 = 1$ is typically smaller than that for $\sigma^2 = 9$. All values of Bias($\hat{\delta}$) using IRLS under $\sigma^2 = 1$ are quite small.

In summary, the magnitude of Bias($\hat{\delta}$) is quite small in all cases, and for most simulations, we find Bias($\hat{\delta}$) to be negligible. Given the simulation design, the "bias" referred to here is a direct reflection of the random error used in generating the simulation data sets. That is, given that a symmetric error distribution is used to generate noise in the data, and that outliers are

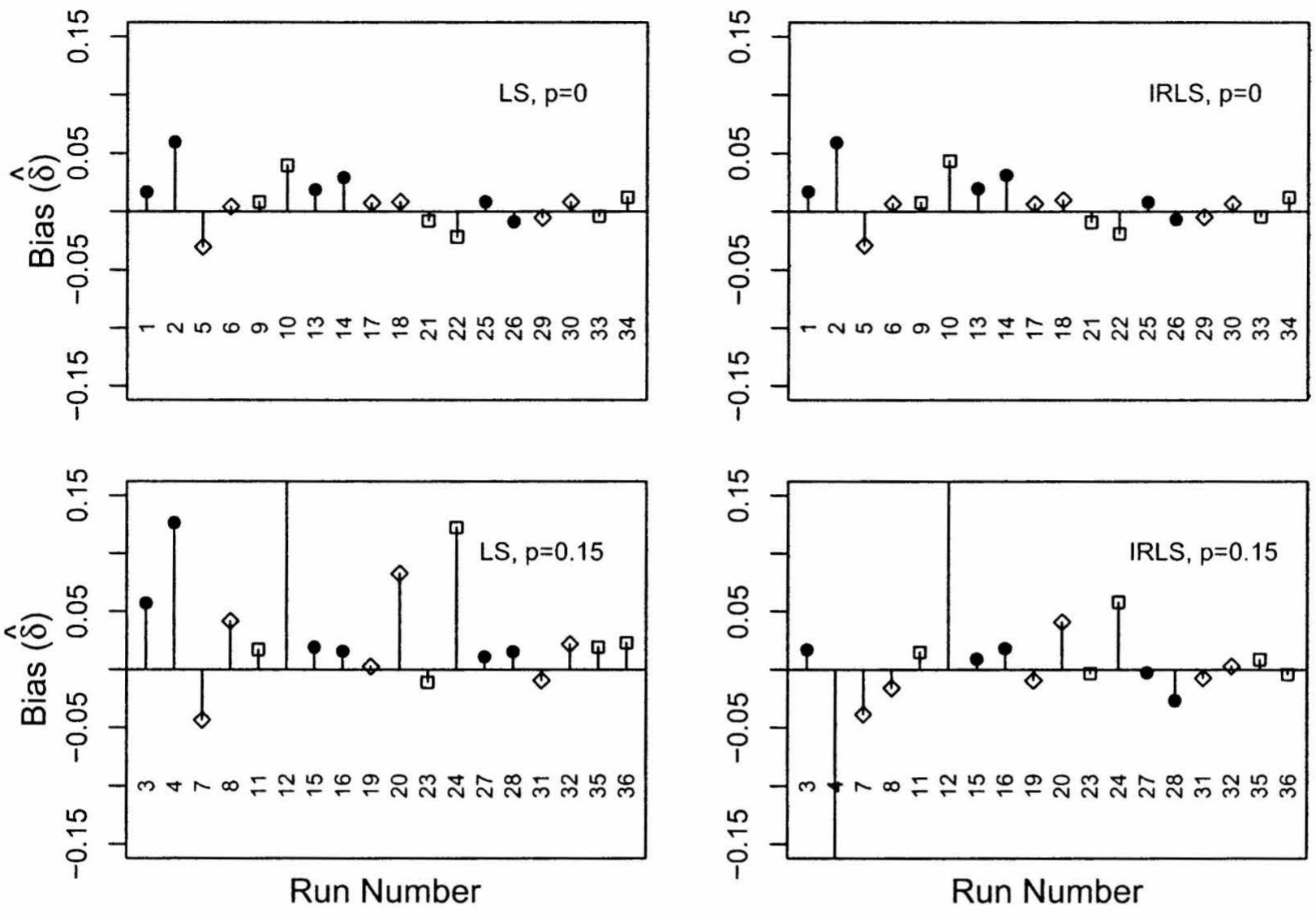

Figure 4.6: Comparison of Bias($\hat{\delta}$) across each value of p.

inserted at randomly selected design points, it appears that Bias($\hat{\delta}$) is simply tracking the noise in the data.

MSE of $\hat{\delta}$

Studying MSE($\hat{\delta}$) from the simulations (Table A1), clear patterns are evident. Note that MSE($\hat{\delta}$) is approximating $Var(\hat{\delta})$ since the values of Bias($\hat{\delta}$) are small, making (Bias($\hat{\delta}$))^2 negligible in (4.3). In the figures which compare MSE, we examine $\text{MSE}_{LS}(\hat{\delta})$ and the ratio $\text{MSE}_{LS}(\hat{\delta})/\text{MSE}_{IRLS}(\hat{\delta})$ (e.g. Figure 4.8). Comparing the ratio of MSE($\hat{\delta}$) from each method provides a simple metric to gauge the performance of IRLS relative to LS. If $\text{MSE}_{LS}(\hat{\delta}) > \text{MSE}_{IRLS}(\hat{\delta})$, then the ratio is greater than 1, and measured by MSE, the IRLS would be considered an improvement over LS.

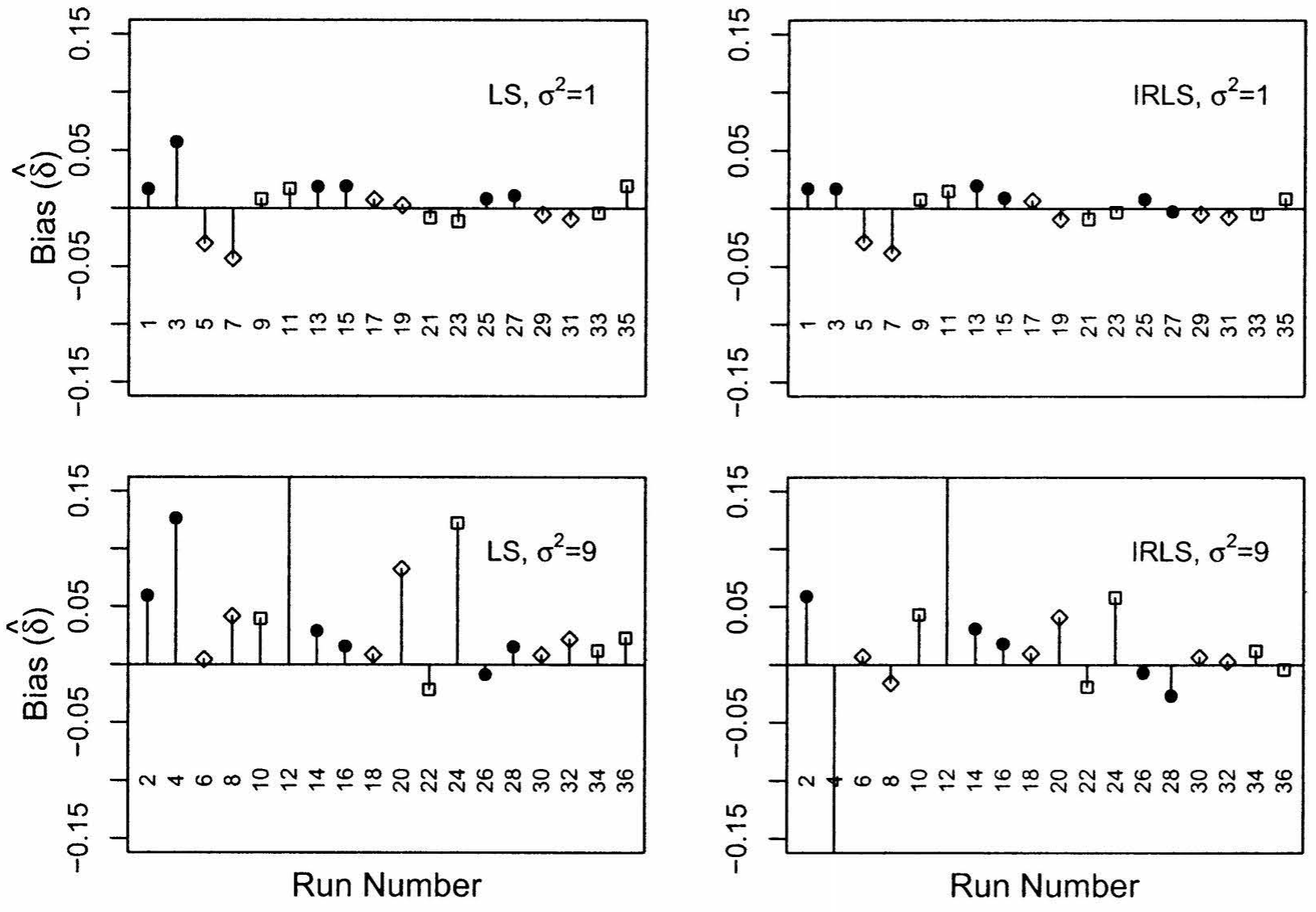

Figure 4.7: Comparison of Bias($\hat{\delta}$) across each value of σ^2.

Considering MSE($\hat{\delta}$) with respect to sample size N (Figure 4.8), observe that all other simulation parameters being equal, we have $\text{MSE}(\hat{\delta})_{N=25} > \text{MSE}(\hat{\delta})_{N=50} > \text{MSE}(\hat{\delta})_{N=100}$. In fact, the rate of decrease in MSE($\hat{\delta}$) with N is inversely proportional to the factor by which sample size is increased. That is, (again, all other simulation parameters equivalent), $1/2(\text{MSE}(\hat{\delta})_{N=25}) \approx \text{MSE}(\hat{\delta})_{N=50}$, $1/2(\text{MSE}(\hat{\delta})_{N=50}) \approx \text{MSE}(\hat{\delta})_{N=100}$, and thus $1/4(\text{MSE}(\hat{\delta})_{N=25}) \approx \text{MSE}(\hat{\delta})_{N=100}$. Displaying MSE($\hat{\delta}$) over each value of δ (Figure 4.9), we see that for similarly structured runs, $\text{MSE}(\hat{\delta})_{\delta=25} > \text{MSE}(\hat{\delta})_{\delta=75} > \text{MSE}(\hat{\delta})_{\delta=50}$. That is, MSE($\hat{\delta}$) is smallest in the situation for which δ is centered in the data set, and the off-center values of δ lead to larger MSE($\hat{\delta}$). In particular, MSE($\hat{\delta}$) is smaller for $\delta = 75$ than $\delta = 25$. Recalling (3.6), note that

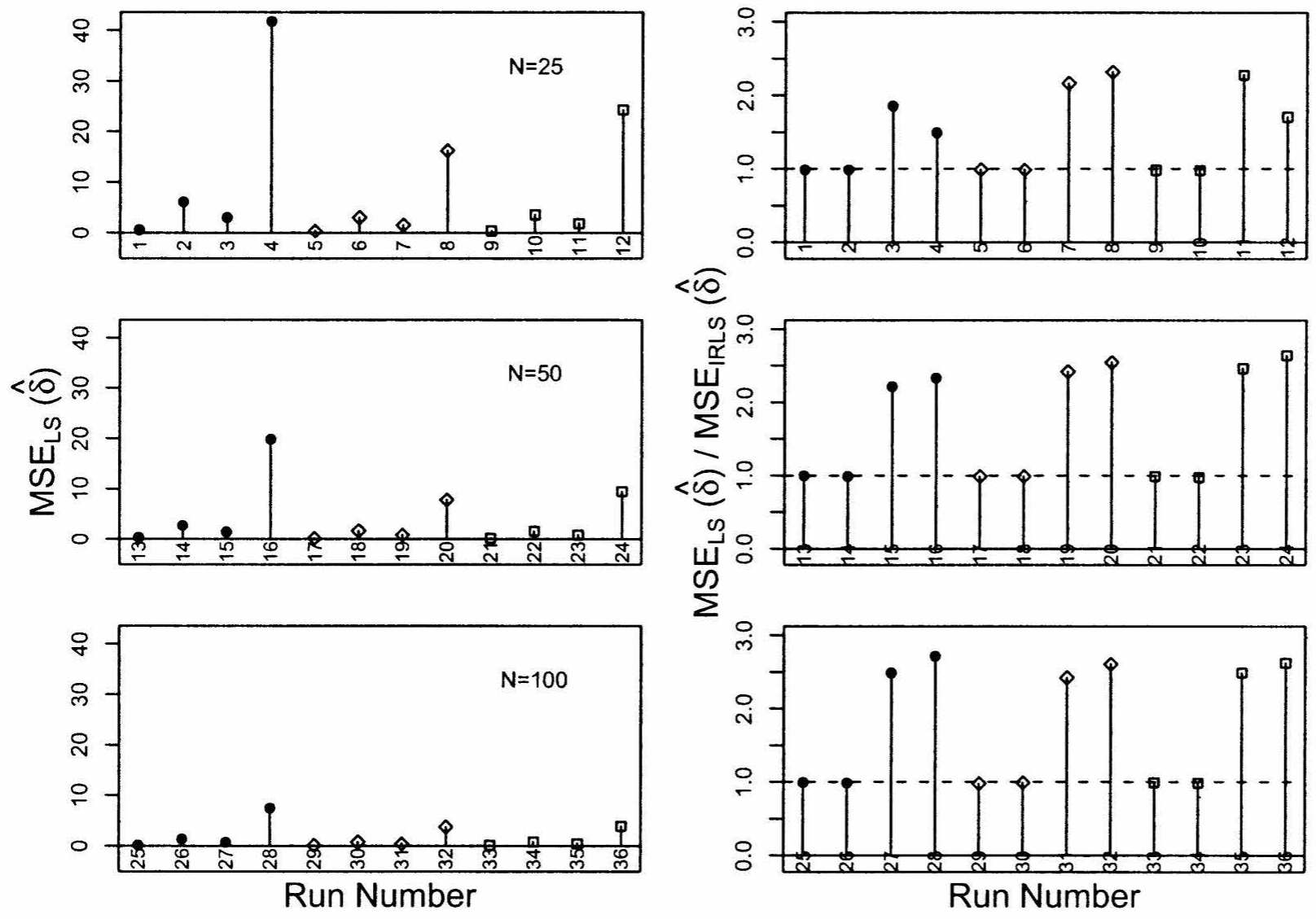

Figure 4.8: Comparison of $\text{MSE}(\hat{\delta})$ across each value of N.

the model is constrained not only to meet at the change-point, but it is also constrained to pass through the origin. This fact likely explains why $\text{MSE}(\hat{\delta})_{\delta=25} > \text{MSE}(\hat{\delta})_{\delta=75}$.

An examination of $\text{MSE}(\hat{\delta})$ considering the presence of outliers (Figure 4.10) reveals a critical result from this simulation exercise. In the upper panel, which includes only runs without outliers, observe that $\text{MSE}_{LS}(\hat{\delta}) \approx \text{MSE}_{IRLS}(\hat{\delta})$. In fact, with $p = 0$, $\text{MSE}_{LS}(\hat{\delta}) \leq \text{MSE}_{IRLS}(\hat{\delta})$ but the actual difference in the two quantities is negligible. Conversely, in the cases which include outliers, examining the MSE ratio plot (Figure 4.10, lower panel), indicates the advantage of IRLS over LS is evident. Finally, if we consider $\text{MSE}(\hat{\delta})$ across the two levels of σ^2 (Figure 4.11), observe as σ^2 is increased, $\text{MSE}(\hat{\delta})$ also

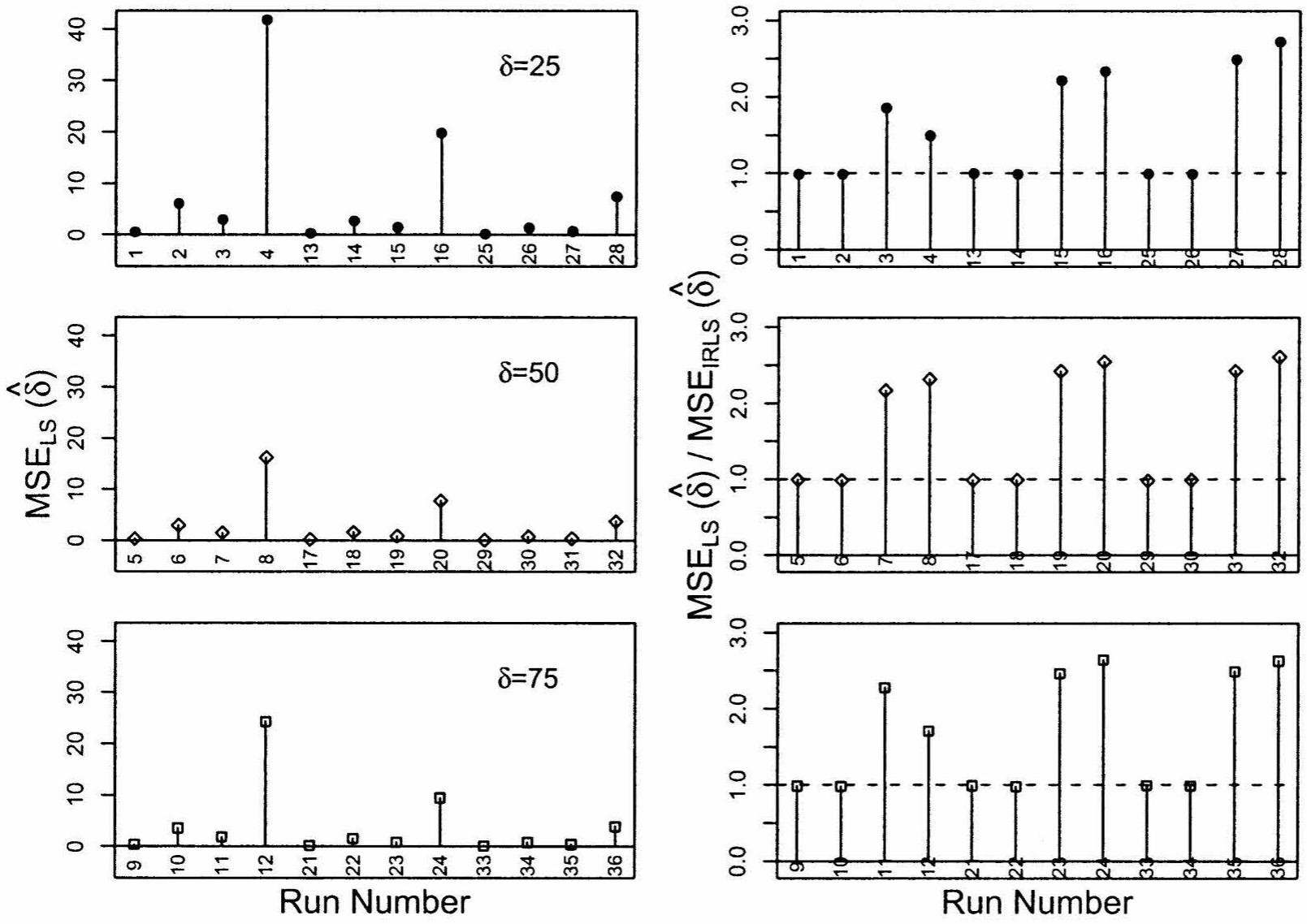

Figure 4.9: Comparison of MSE($\hat{\delta}$) across each value of δ.

increases; not an unexpected result.

For this simulation, summaries of MSE($\hat{\delta}$) are more informative than those of Bias($\hat{\delta}$) in determining the effect of each simulation parameter on the estimation and in enabling comparison between the two estimation methods. In particular, examination of MSE($\hat{\delta}$) points to a clear benefit of using IRLS with the Julious algorithm. If outliers are not present, the LS and IRLS estimation using Julious' algorithm are practically equivalent. However, should the data be contaminated with outliers, IRLS, as we might suspect *a priori*, performs substantially better than LS. Thus, in practice, to guard against problems arising from the presence of outliers when using the Julious algorithm to estimate the change-point, simulation results indicate that with

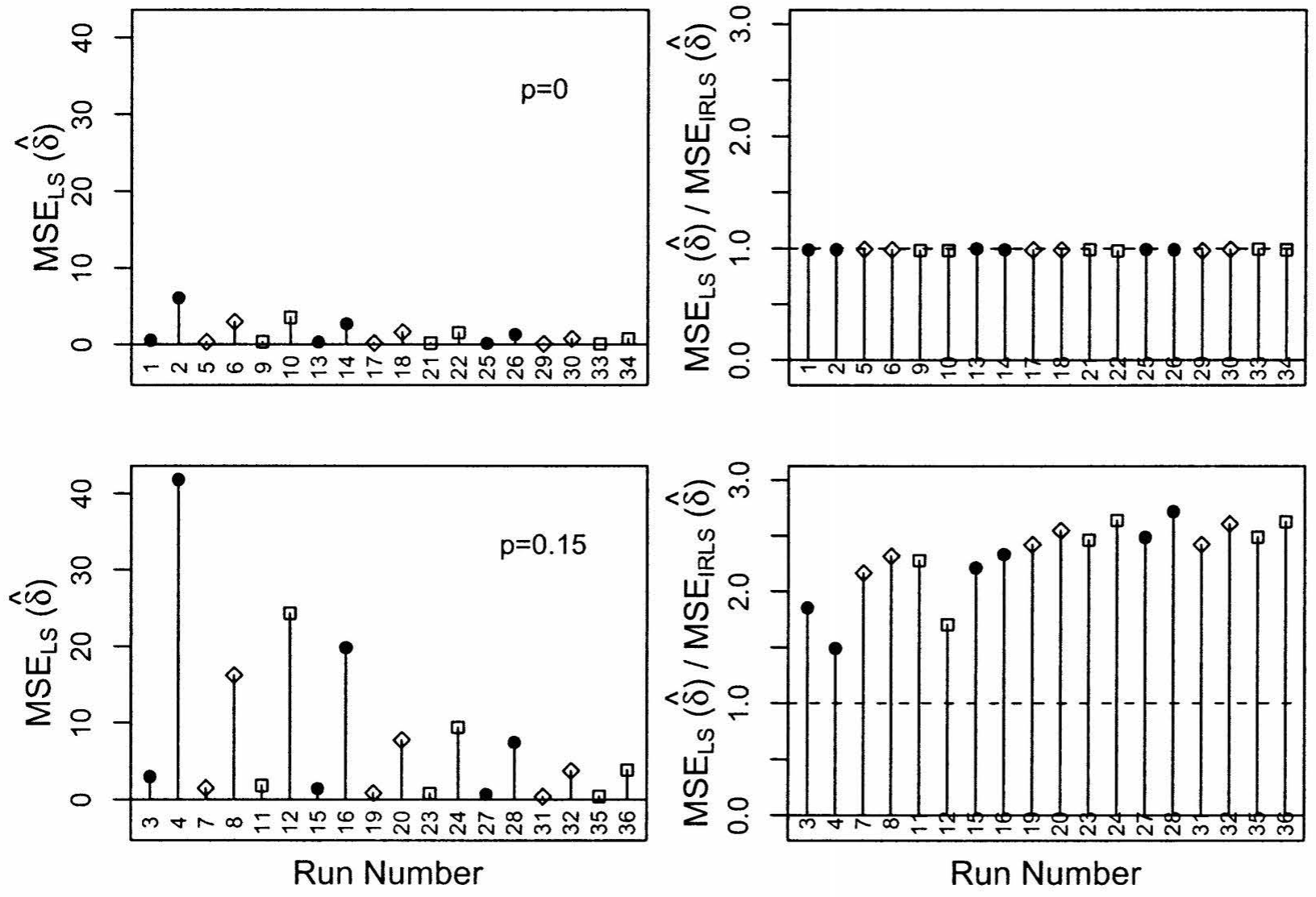

Figure 4.10: Comparison of MSE($\hat{\delta}$) across each value of p.

respect to bias and MSE of the change-point δ, iterative re-weighting is preferable under the assumption of normally distributed errors.

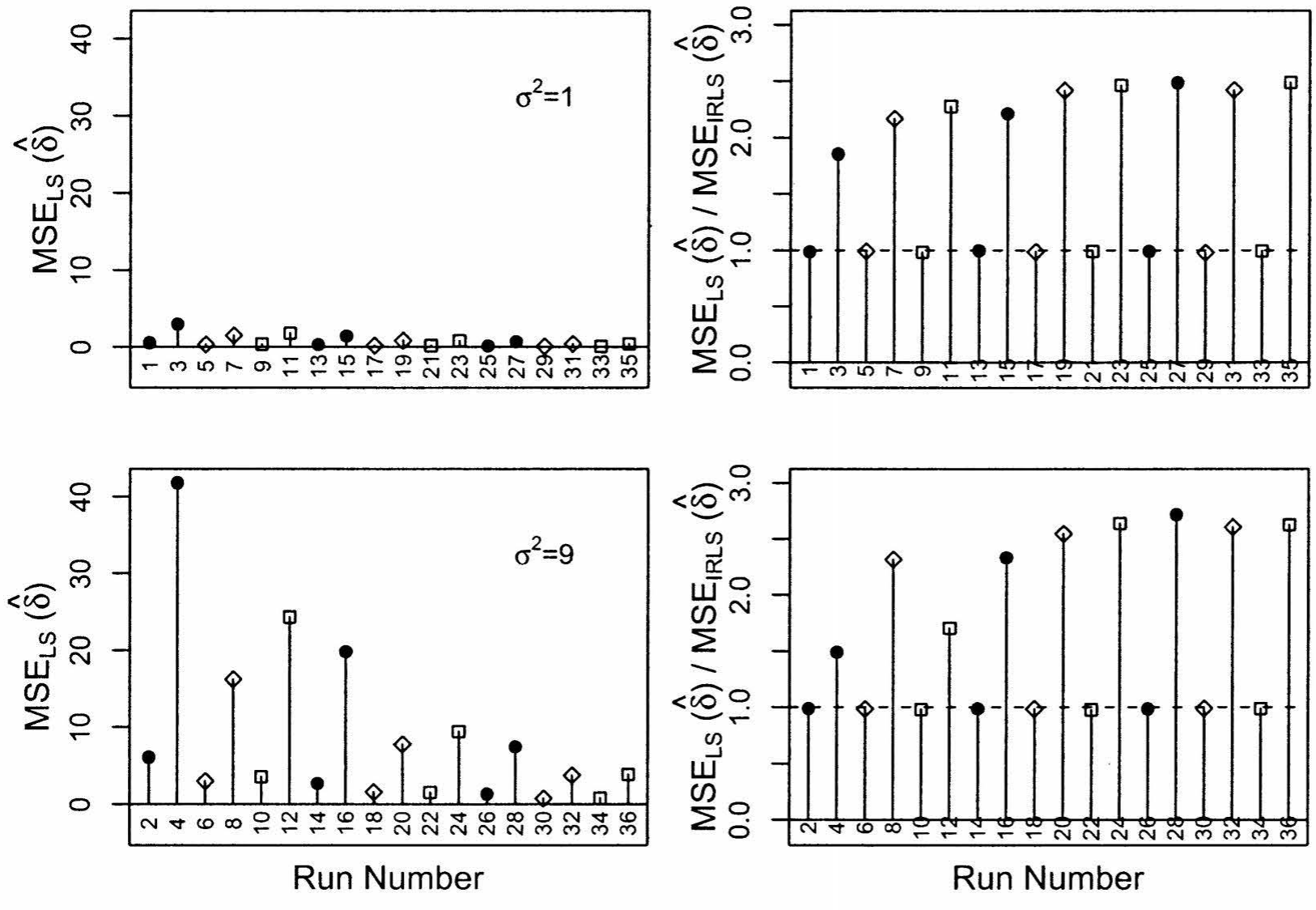

Figure 4.11: Comparison of MSE($\hat{\delta}$) across each value of σ^2.

4.4.2 Hockey Stick Model, Lognormal Errors

A second simulation exercise was conducted, studying the hockey-stick model under the assumption of lognormal errors. In this simulation, the parameter values used were: $N = \{25, 50, 100\}$, $p = \{0, 0.15\}$, $\sigma^2 = \{0.025, 0.1\}$, and $\delta = \{25, 50, 75\}$. Quantities held fixed over simulation trials were: (Huber's) $c = 2$, $\tau = 3$, $F = N(0, \sigma^2)$, and $G = N(0, \tau^2\sigma^2)$. With $\tau = 3$, 50% of the realizations from G lie within two standard errors of the mean when $\sigma^2 = 1$.

The 'true' model from which random data were generated was:

$$\log(y_i) = \begin{cases} \log(\beta_1) + \log(x_i) + \varepsilon & x_i \leq \delta \\ \log(\alpha_2) + \varepsilon & x_i \geq \delta \end{cases},$$

with $\beta_1 = 1$, and $\alpha_2 = \delta$. In the normal-error simulation, by setting $\delta = \{25, 50, 75\}$, we could assess the effect of change-points in the center of the data against off-center points. This is still true in this simulation exercise yet estimation takes place on the log-scale (3.14). On the log-scale, the values $\{25, 50, 75\}$ no longer represent center and off-center cases for the covariate $\log(x)$, i.e. $\log(50)$ is not the midpoint between $\log(25)$ and $\log(75)$. In fact, the midpoint of the data is now a function of sample size, because the simulation design selects N equally-spaced points *within* $[0, 100]$. In particular, note x_1 and x_N are functions of sample size. Thus, on a logarithmic scale, the mid-point of $\log(x_1)$ and $\log(x_N)$ re-transformed to the linear scale is computed as $\exp([\log(x_1) + \log(x_N)]/2) = \sqrt{x_1 x_N}$. Setting $N = \{25, 50, 100\}$, this mid-point evaluates to $\{19.2, 13.9, 9.9\}$, respectively. We discuss this further when examining simulation results across the true change-point values.

The random error ε again comes from the mixture distribution H (4.4), with τ as the contamination parameter.

Tables of bias and MSE for $\hat{\delta}$, $\hat{\beta}_1$, and $\hat{\alpha_2}$ are given in **Appendix A** for each simulation conducted. Although estimation takes place on the log-scale, the bias and MSE are evaluated on the linear scale, since interest is on estimates of $\hat{\delta}$ as opposed to $\log(\hat{\delta})$. As in the normal model simulations, a run identification number is associated with each simulation, and focus is restricted to estimation of the change-point, δ.

Bias of $\hat{\delta}$

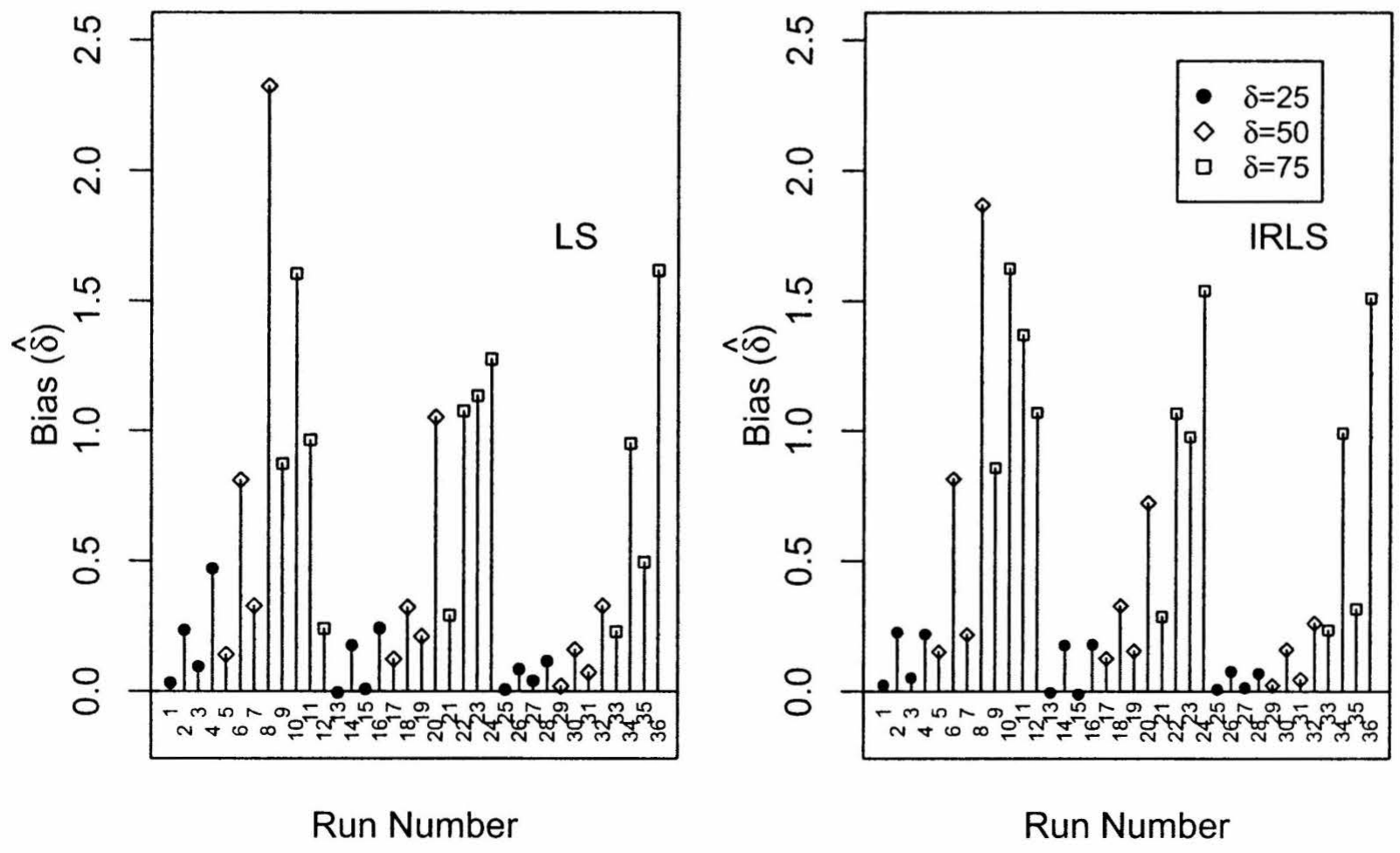

Figure 4.12: Comparison of Bias($\hat{\delta}$) for simulation runs.

The least-squares and iteratively re-weighted estimators of δ exhibited substantial simulation bias (Figure 4.12). In contrast to the normal-error simulations, in which most cases had negligible bias, there are now just a few cases which have low bias. Most cases have positive bias - the bias of $\hat{\delta}$ depends on the location of the true value of δ. The largest biases correspond to the cases in which the true change-point is greater than the center of the covariates (i.e. to the right of center).

Plotting Bias($\hat{\delta}$) for each value of N (Figure 4.13), we see that Bias($\hat{\delta}$) is generally decreasing as N increases. However, the reduction in bias seems to

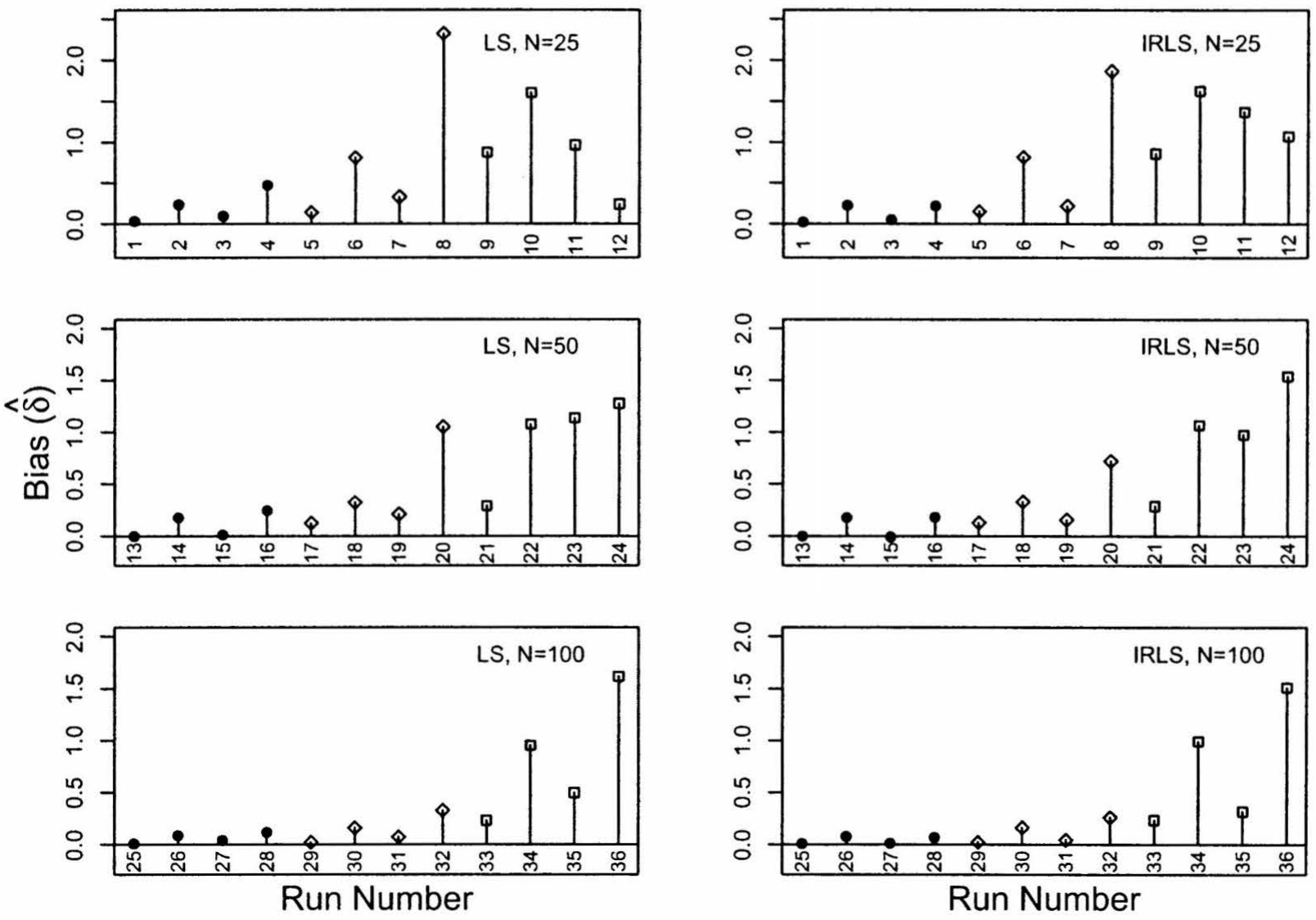

Figure 4.13: Comparison of Bias($\hat{\delta}$) across each value of N.

occur at a slower rate than in the normal model (Figure 4.4).

Examining Bias($\hat{\delta}$) at each level of δ (Figure 4.14), we see that the relation $\text{Bias}(\hat{\delta})_{\delta=25} < \text{Bias}(\hat{\delta})_{\delta=50} < \text{Bias}(\hat{\delta})_{\delta=75}$ holds for the majority of cases (all other simulation parameters being equal).

The effect of outliers on Bias($\hat{\delta}$) (Figure 4.15) is obvious. Although the no-outlier cases (top row of Figure 4.15) show little difference between $\text{Bias}_{LS}(\hat{\delta})$ and $\text{Bias}_{IRLS}(\hat{\delta})$, we see that if outliers are present, then generally, $\text{Bias}_{IRLS}(\hat{\delta}) < \text{Bias}_{LS}(\hat{\delta})$. For some cases (e.g. run #4), iterative re-weighting leads to considerable reduction in bias.

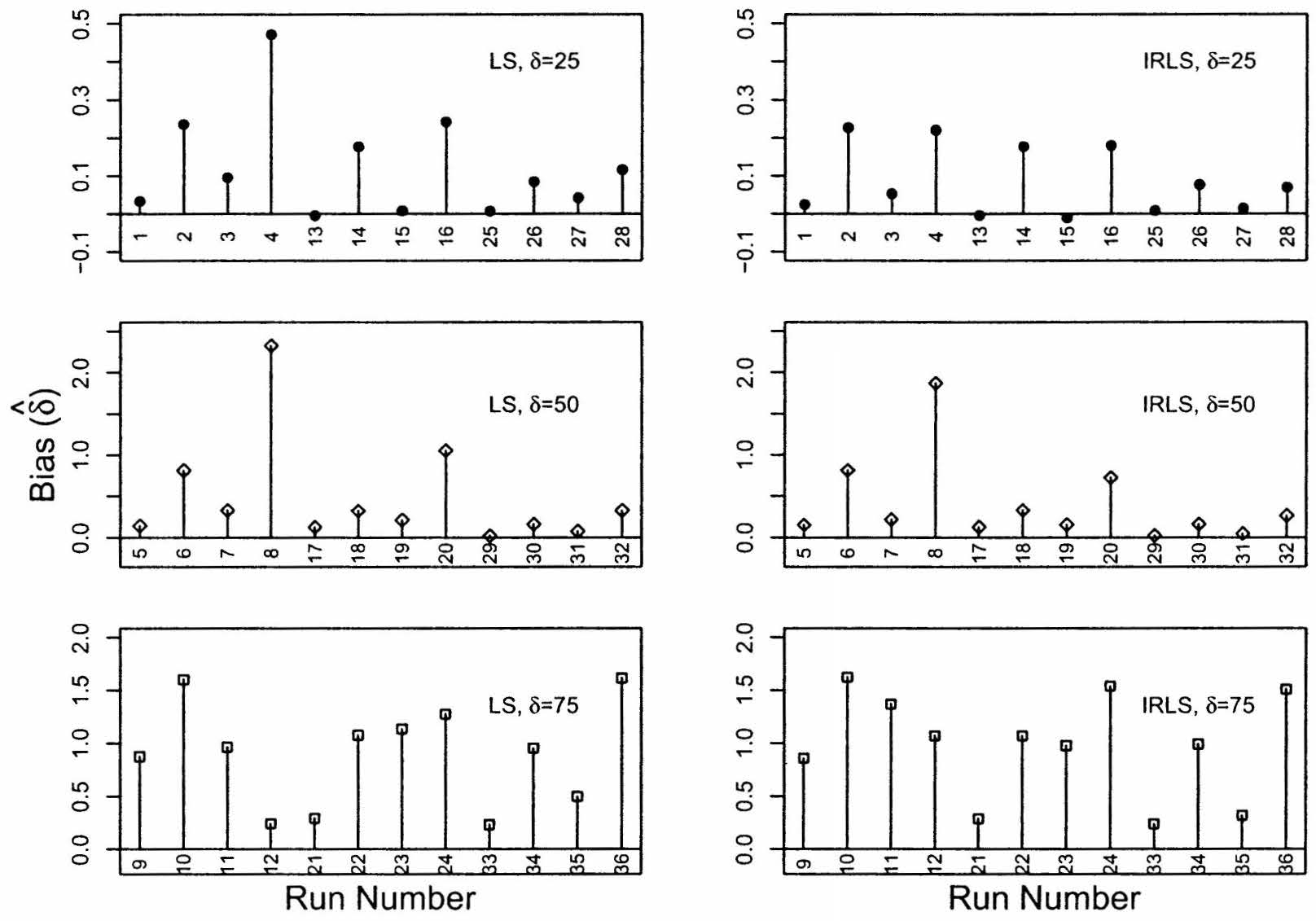

Figure 4.14: Comparison of Bias($\hat{\delta}$) across each value of δ.

Studying the results of Bias($\hat{\delta}$) at each level of σ, we observe that Bias$(\hat{\delta})_{\sigma=0.025}$ < Bias$(\hat{\delta})_{\sigma=0.1}$ (Figure 4.16) for most situations.

The results for Bias($\hat{\delta}$) are generally consistent with those in the normal model. One important finding in the simulations using the lognormal errors is that the values of Bias($\hat{\delta}$) are not negligible; thus the estimation method appears to be biased.

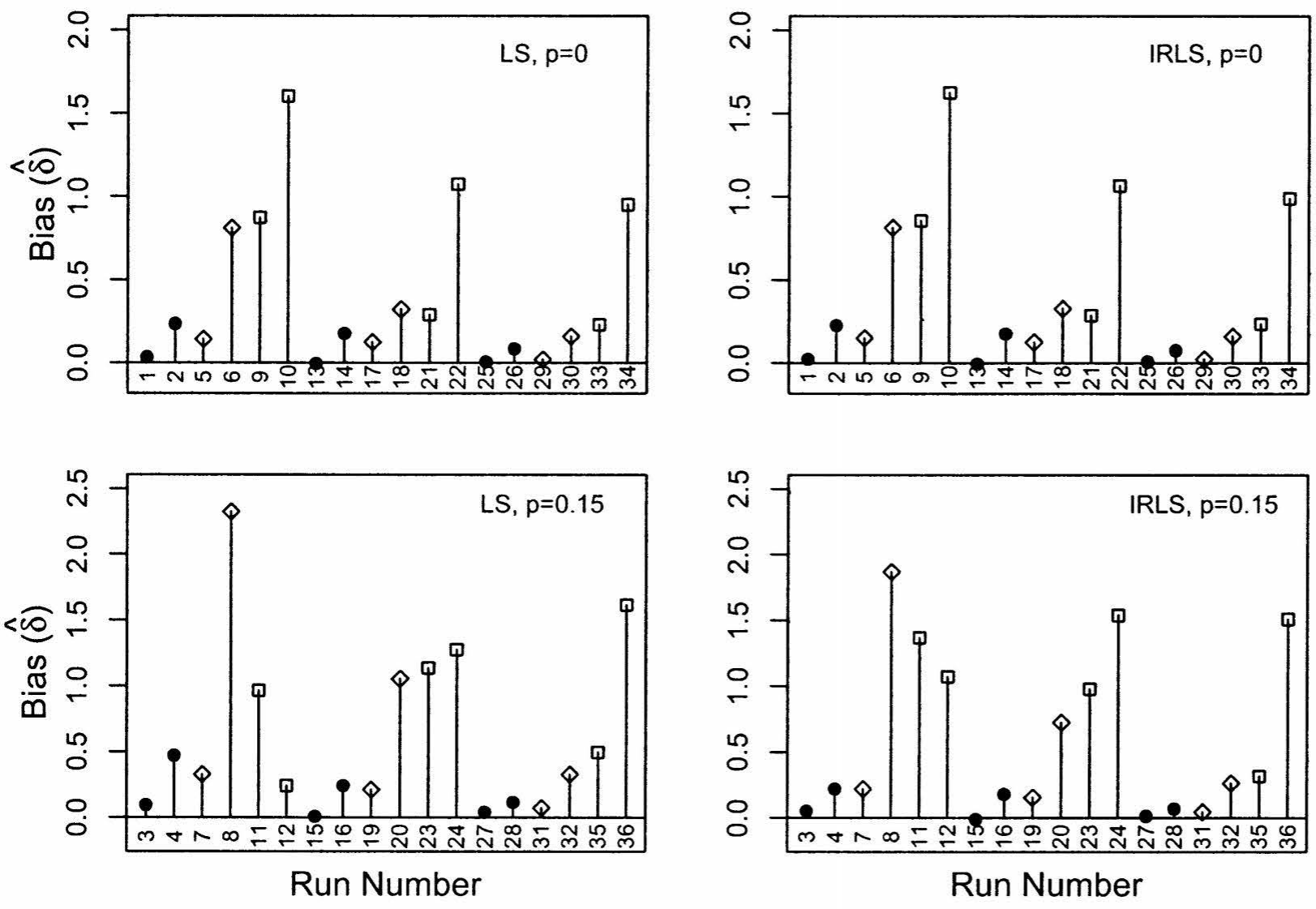

Figure 4.15: Comparison of Bias($\hat{\delta}$) across each value of p.

MSE of $\hat{\delta}$

Next, we assess MSE($\hat{\delta}$) in the lognormal simulations. Although "large" values of Bias($\hat{\delta}$) were noted in this simulation study, it is still true that $\text{Var}(\hat{\delta}) \approx \text{MSE}(\hat{\delta})$ in (4.3) since $\text{MSE}(\hat{\delta}) \gg (\text{Bias}(\hat{\delta}))^2$.

A plot of MSE($\hat{\delta}$) for each value of N (Figure 4.17) reveals that MSE($\hat{\delta}$) decreases as N increases. In most situations, the rate of decrease in MSE($\hat{\delta}$) is again inversely proportional to the change in sample size, as in the normal-error simulations.

Examination of MSE($\hat{\delta}$) at each value of δ (Figure 4.18) indicates that MSE($\hat{\delta}$)

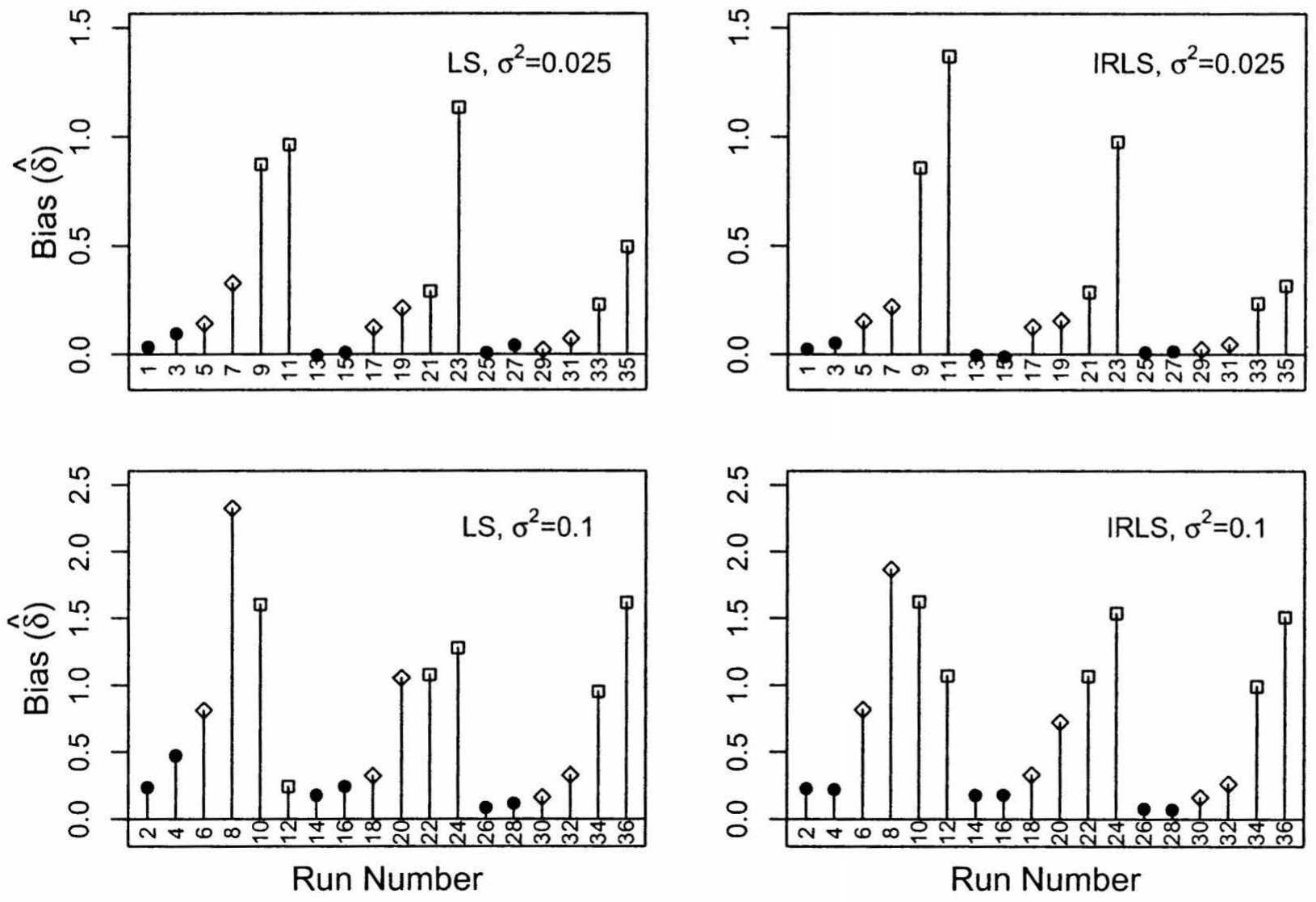

Figure 4.16: Comparison of Bias($\hat{\delta}$) across each value of σ^2.

increases as the true-value of the change-point increases. This is not surprising considering the earlier discussion on how the change-points $\delta = \{25, 50, 75\}$ translate to the log-scale. The true change-point values are greater than the mid-point of $\log(x)$, which we may expect to increase MSE.

Evaluating MSE($\hat{\delta}$) at each value of p (Figure 4.19), we note that $\text{MSE}_{IRLS}(\hat{\delta}) < \text{MSE}_{LS}(\hat{\delta})$ for $p > 0$ (bottom row of figure). Considerable reduction in MSE($\hat{\delta}$) can occur by robustifying the estimation (compare ratio plots in right-hand column of Figure 4.19). However, in the cases without outliers, we find $\text{MSE}_{IRLS}(\hat{\delta}) > \text{MSE}_{LS}(\hat{\delta})$, with little difference in the magnitude of MSE($\hat{\delta}$) from either method. This is consistent with the result of §4.4.1.

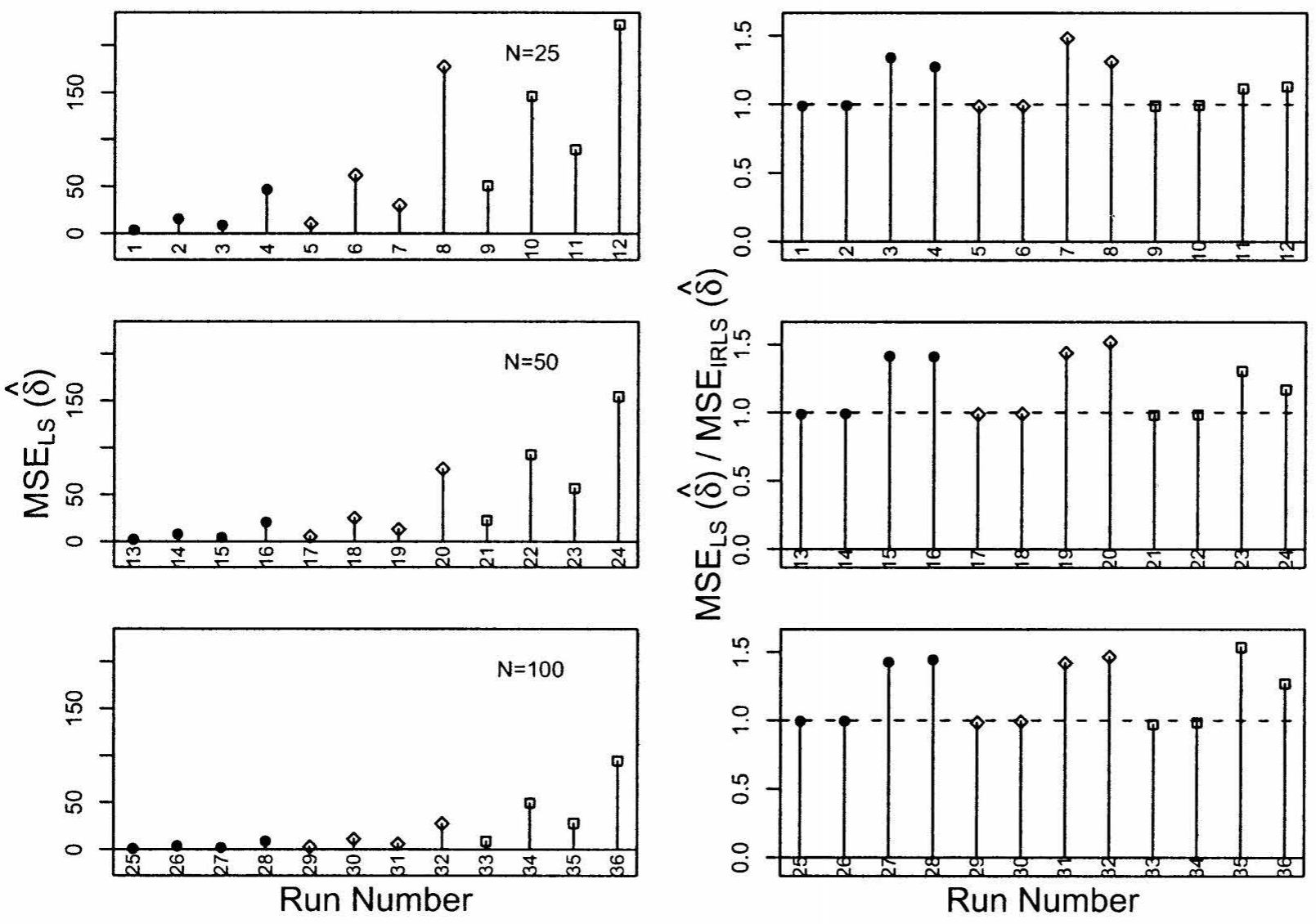

Figure 4.17: Comparison of MSE($\hat{\delta}$) across each value of N.

Values of MSE($\hat{\delta}$) examined at each level of σ (Figure 4.20) reveal similar patterns. At each of the low and high levels of σ, we observe a scale change in values of MSE($\hat{\delta}$).

Based upon simulation results under the lognormal model, Bias($\hat{\delta}$) and MSE($\hat{\delta}$) verify that the use of weighted Julious algorithm is preferred over the LS alternative. Findings indicate that the LS and IRLS methods perform equally well if outliers are not present, but if outliers exist in the data, then IRLS performs better than LS. It has been demonstrated that the estimates of change-point using IRLS have improved bias and MSE. Thus, IRLS results should be evaluated against the results obtained from LS in practice.

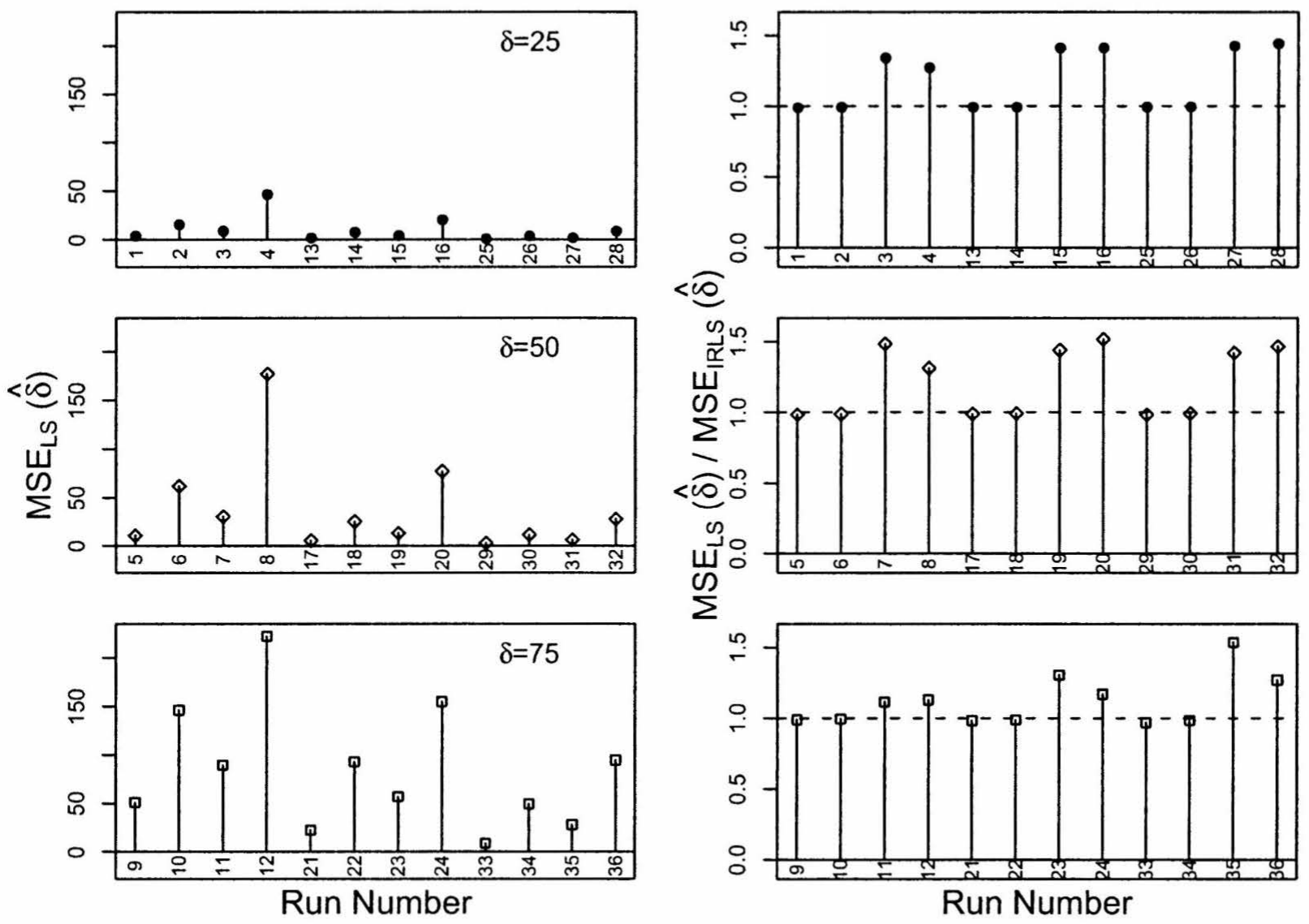

Figure 4.18: Comparison of MSE($\hat{\delta}$) across each value of δ.

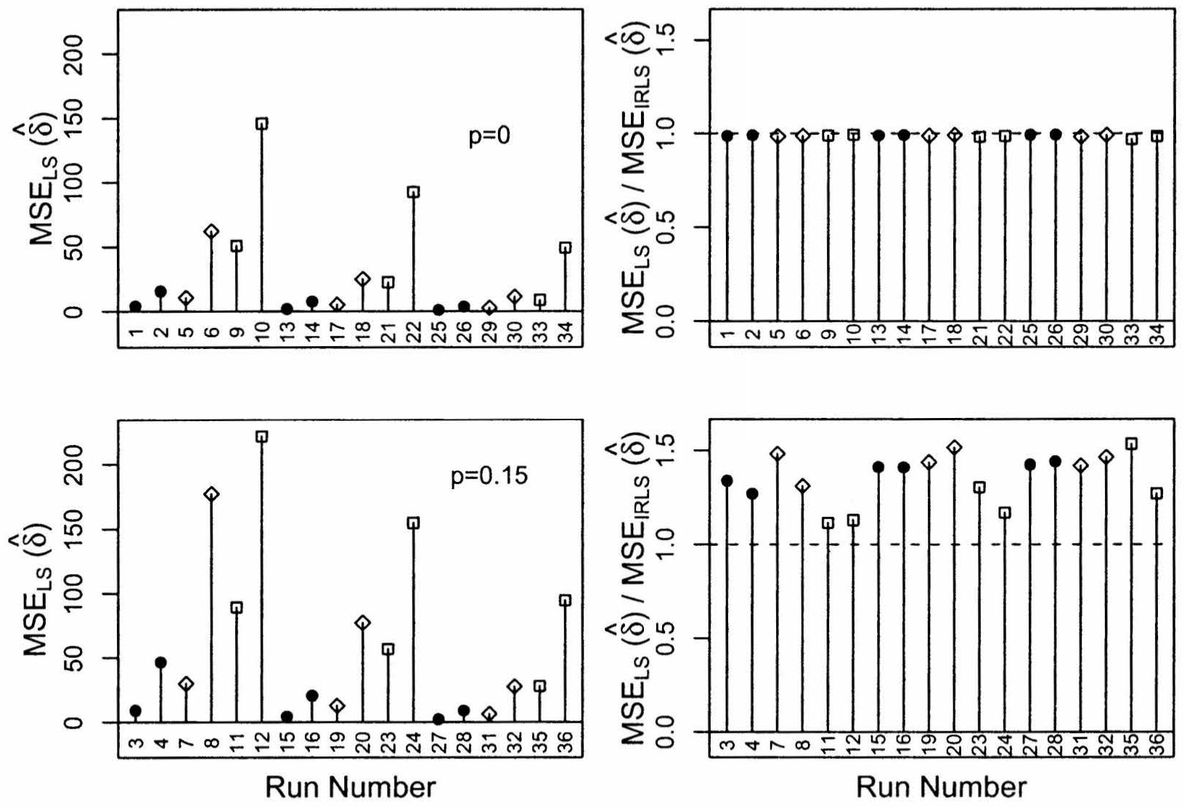

Figure 4.19: Comparison of MSE($\hat{\delta}$) across each value of p.

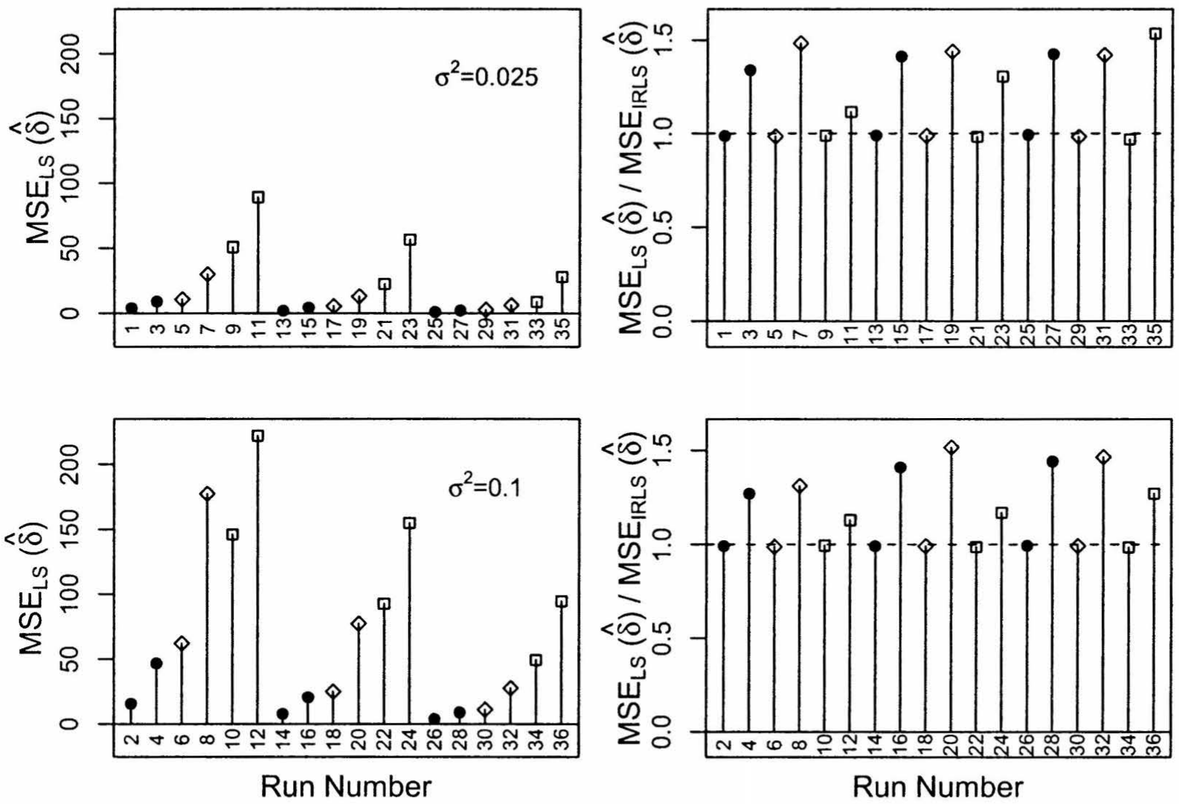

Figure 4.20: Comparison of MSE($\hat{\delta}$) across each value of σ^2.

Chapter 5

Application

5.1 Preliminaries

In this chapter, we apply the estimators developed in **Chapter 3** to three data sets. Of primary concern is the location of the change-point under the Julious algorithm (JA) and the iteratively re-weighted Julious algorithm (IRJA). Model estimates are produced for the following data: (i) the physiological data set from Julious (2001), used to demonstrate the JA in Julious (2001), and two fisheries science examples: stock and recruitment data for the (ii) 3LNO American Plaice (NAFO, 2003) and (iii) North Sea Plaice fish stocks (ICES 2002b). Each of these data sets are tabled in **Appendix B**.

In the following section, we consider inference about the change-point, δ.

5.2 Inferences via Bootstrap

Given the indirect manner in which $\hat{\delta}_{LS}$ and $\hat{\delta}_{IRLS}$ are obtained (see Figures 2.1 and 3.4), inference for the change-point is not straightforward. A simple empircal approach to provide a measure of uncertainty for the estimated change-point is to use the bootstrap (Efron and Tibshirani, 1993). For each of $\hat{\delta}_{LS}$ and $\hat{\delta}_{IRLS}$, we may apply the following bootstrap algorithm:

1. Given the n observations (x_i, y_i), compute the estimate of change-point from the JA (IRJA if estimating $\hat{\delta}_{IRLS}$).

2. From the model fit in step 1, compute the residual vector:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}.$$

3. Randomly sample n values from the elements of $\hat{\boldsymbol{\varepsilon}}$, with replacement. From the n re-sampled values, form the vector $\boldsymbol{\varepsilon}_j$.

4. Construct a bootstrap sample $\boldsymbol{y}_j$ as:

$$\boldsymbol{y}_j = \hat{\boldsymbol{y}} + \boldsymbol{\varepsilon}_j.$$

5. Use the data $\boldsymbol{x}$, $\boldsymbol{y}_j$ to compute the j^{th} bootstrap estimate of $\hat{\delta}_{LS}$ or $\hat{\delta}_{IRLS}$ as appropriate (denote as $\hat{\delta}_j$).

6. Repeat steps 3-5 B times.

7. Compute:

$$\hat{\text{se}}_B(\hat{\delta}) = \sqrt{\frac{\sum_{j=1}^{B}(\bar{\delta} - \hat{\delta}_j)^2}{(B-1)}},$$

where $\bar{\delta} = \frac{\sum_{j=1}^{B} \hat{\delta}_j}{B}$.

In the applications in this chapter, we consider $B = 1000$ bootstrap replicates.

It may be of interest to evaluate whether $\hat{\delta}_{JA}$ and $\hat{\delta}_{IRJA}$ are statistically different. Such comparisons are not straightforward in that the estimators are derived using OLS and WLS (e.g. a nested hypothesis test cannot be constructed). One formal method to evaluate the significance of the difference between the two estimators is to conduct a bootstrap exercise (Efron and Tibshirani, 1993). A bootstrap distribution of the difference estimator $\hat{\delta}_{JA} - \hat{\delta}_{IRJA}$ can be used to determine empirically if the least-squares and robust estimate of the change-point are statistically different. However, application of the bootstrap would require the assumption that errors are identically and independently distributed. This is not the case in the simulation work of **Chapter 4**, and significance tests evaluating the difference between $\hat{\delta}_{JA} - \hat{\delta}_{IRJA}$ are not considered in this practicum.

5.3 Julious' Data Set

Recall (§3.2.3) the physiological data set from Julious (2001), from illustration of the JA in Chapter 3. We now study this data set using iterative re-weighting and compare the estimates and model fit. We apply the full segmented regression model (3.1). Computing estimates using the JA, we found:

$$\hat{\boldsymbol{\beta}} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)^T = (0.0765, 0.0423, -1.6595, 0.0863)^T,$$

with $\hat{\delta}_{LS} = 39.463$, and $\text{RSS} = 0.3895$.

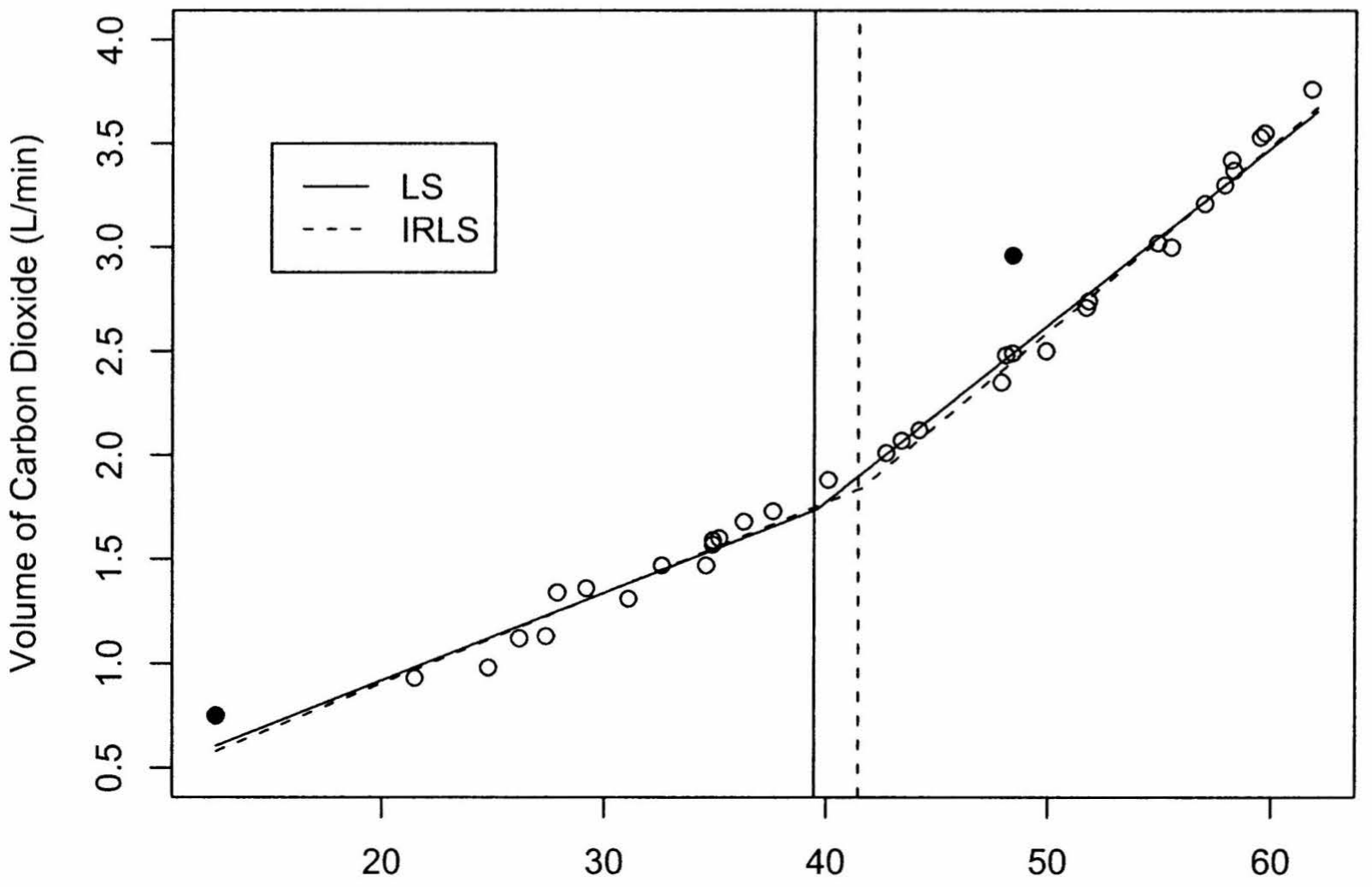

Figure 5.1: LS and IRLS estimates (normal model) for Julious' data set. Filled points are down-weighted in IRLS.

If we apply the IRJA using Huber's $c = 2$, and iterating until $\hat{\delta}$ is constant to three decimal places, we obtain the following estimates after 8 iterations:

$$\hat{\boldsymbol{\beta}} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)^T = (0.0296, 0.0440, -1.8725, 0.0899)^T,$$

$$\hat{\delta}_{IRLS} = 41.442, \text{ and } \text{RSS}_W = 0.2467.$$

A plot of the data and the estimated regression line under LS and IRLS is presented as Figure 5.1. Vertical lines are used to demarcate the estimated

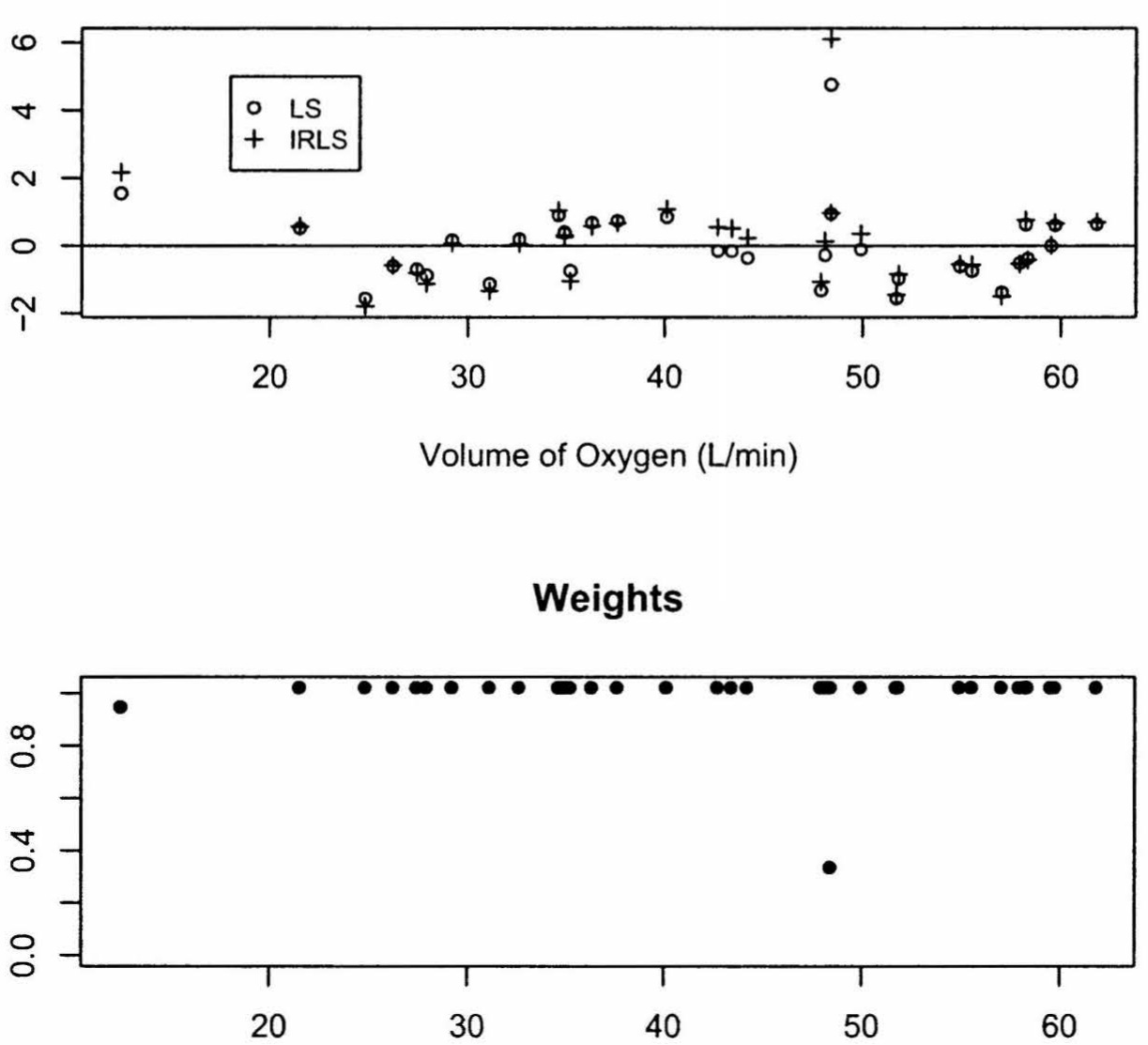

Figure 5.2: Standardized Residuals for LS and IRLS estimates (upper panel); Weights from IRLS (lower panel).

change-point for each method. The solid points in the figure are those which have been down-weighted in the final IRLS iteration. Observe that $\hat{\delta}_{IRLS}$ is approximately 5% larger than $\hat{\delta}_{LS}$. This is potentially important in interpretation, since here the "change-point represents the point at which a subject switches metabolic pathways, from aerobic to anaerobic" (from Julious 2001).

Figure 5.2 displays the standardized residuals from the LS and IRLS estimates (top panel) and the weightings in the final IRLS iteration (lower

panel). The residuals are standardized by MAD($\boldsymbol{\varepsilon}$) as described in §3.3.2. As the fitted regression lines are quite similar, there are no large-scale changes in the residuals. However, for the oxygen volume values inside the interval $[40, 55]$, we see that the IRLS residuals are generally smaller due to the down-weighting of the y-outlier at $x = 48.4$. The first point (at $x = 12.5$) is also downweighted, but only marginally so.

Bootstrapping the IRLS residuals, we obtain the following estimate of uncertainty for the change-point:

$$\hat{\text{se}}_{1000}(\hat{\delta}) = 1.698.$$

5.4 Stock-Recruit Data

We now consider two fisheries data sets which are estimated using the hockey-stick model assuming lognormal errors (3.13). This model is assessed in the simulation studies of **Chapter 4**.

5.4.1 American Plaice in 3LNO

American Plaice (a flounder species) in divisions 3LNO of the Northwest Atlantic Fisheries Organization (NAFO) regulatory area are distributed across the Grand Banks of Newfoundland. The data used in this analysis are age 0 recruits. (See NAFO, 2003 for additional details).

The results obtained using the LS JA are:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\alpha}_2)^T = (19.0739, 589.1886)^T,$$

$$\hat{\delta}_{LS} = 30.8898, \text{ and RSS} = 2.7438.$$

Estimates obtained using the IRJA (after 6 iterations, using Huber's c=2 and iterating $\hat{\delta}$ until convergence to three decimal places) are:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\alpha}_2)^T = (17.9318, 587.0866)^T,$$

$$\hat{\delta}_{IRLS} = 32.7399, \text{ and RSS}_W = 2.4798.$$

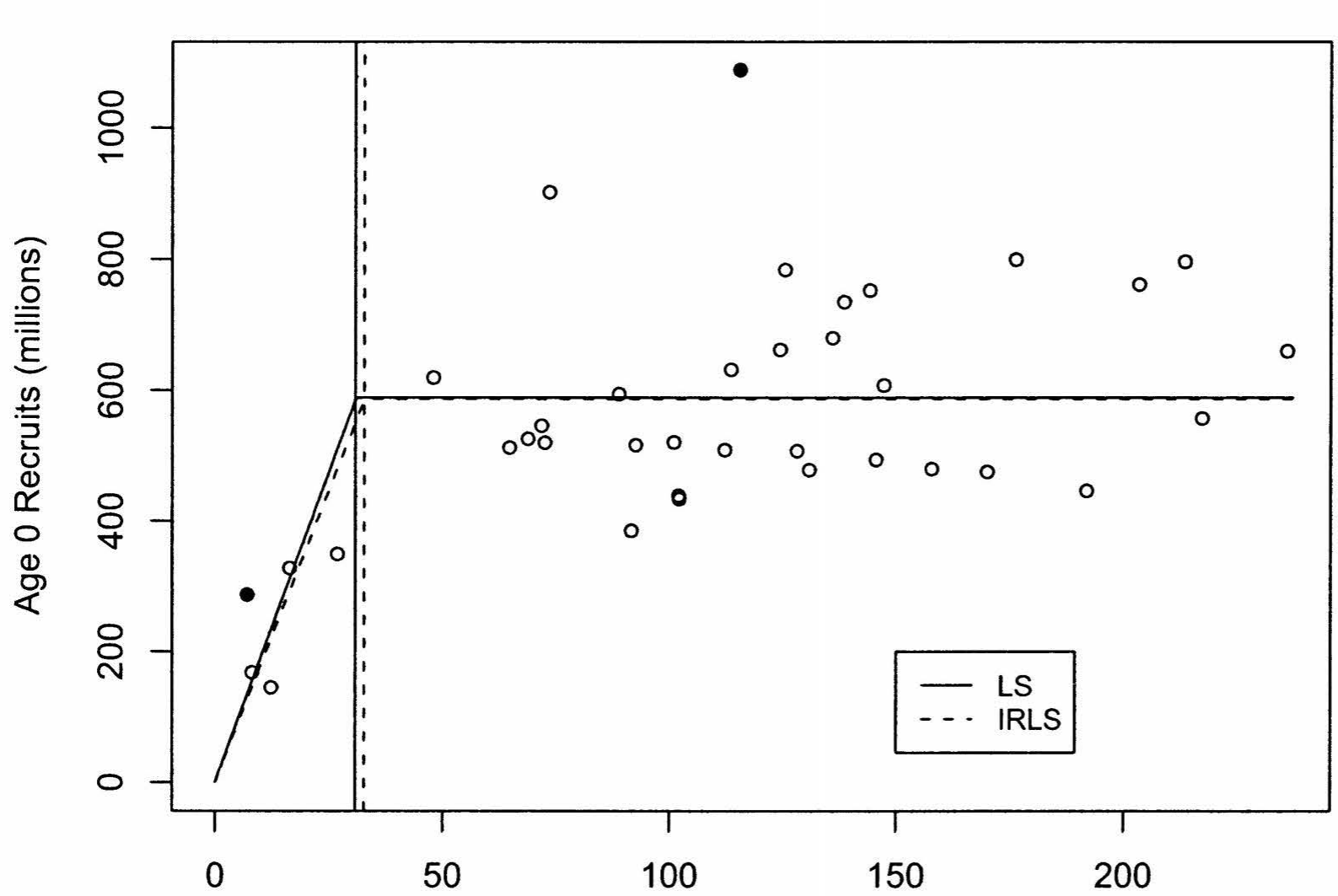

Figure 5.3: LS and IRLS estimates for 3LNO American Plaice stock-recruit data (lognormal model). Filled points are down-weighted in IRLS.

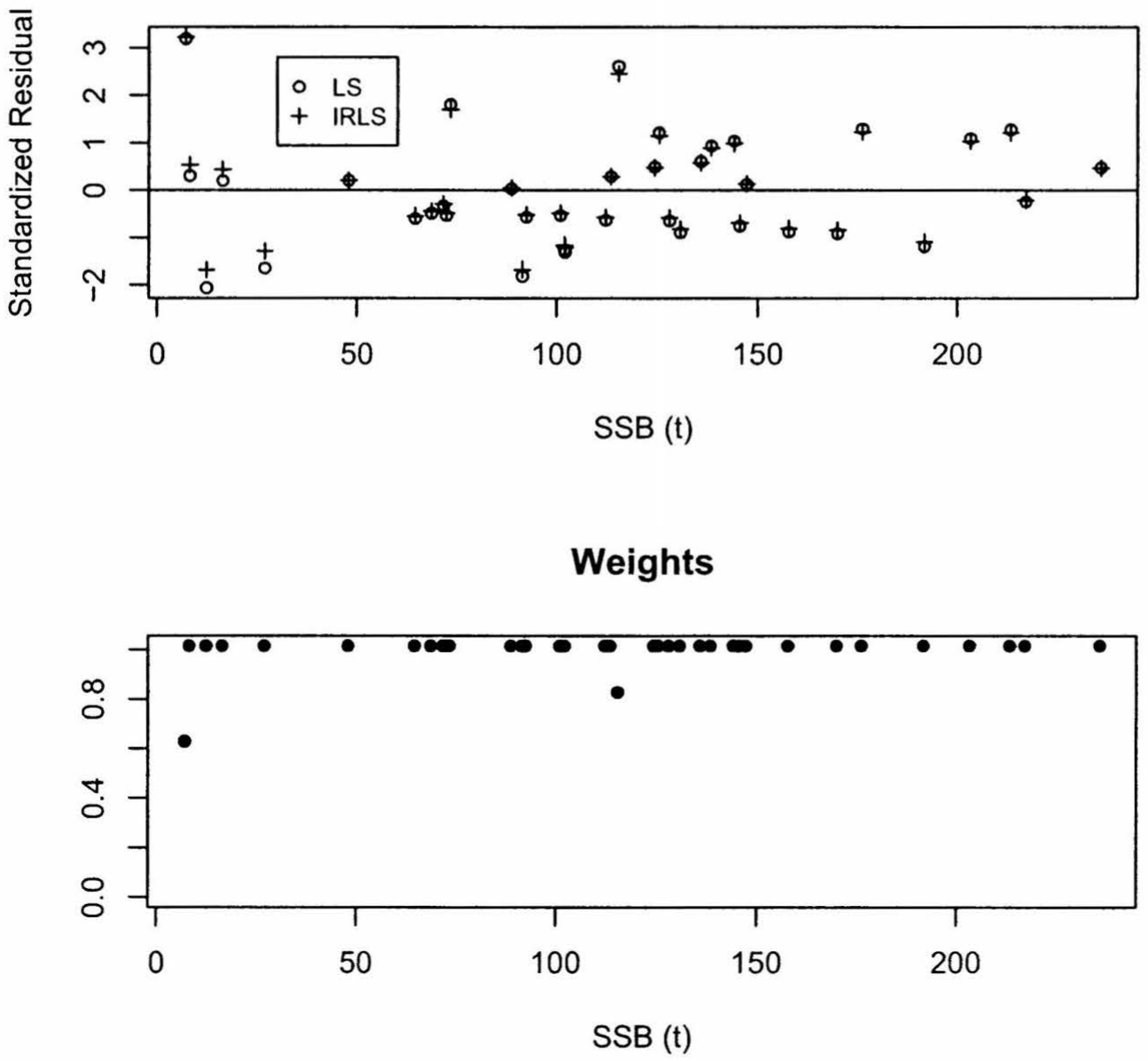

Figure 5.4: Standardized Residuals for LS and IRLS estimates (upper panel); Weights from IRLS (lower panel).

Figure 5.3 illustrates the fitted lines (median recruitment) from LS and IRLS, and we see there is little difference in the two estimates. From a practical perspective (see §1.3 for precautionary approach discussion), the difference in the estimates of change-point is inconseqential. We see that two points (filled circles in Figure 5.3) are down-weighted, however, these points are not close to the estimated change-points or are not extreme enough to substantially change the estimates of $\hat{\boldsymbol{\beta}}$ or $\hat{\delta}$, keeping in mind estimation takes place on the log-scale. In the figure of standardized residuals, we see the

residuals (log-scale) are near-identical, and the weight reduction of the two down-weighted points is slight.

A bootstrap exercise yields $\hat{\text{se}}_{1000}(\hat{\delta}) = 4.3295$ as a variability measure of $\hat{\delta}_{IRLS}$.

5.4.2 North Sea Plaice

The North Sea Plaice stock is a flounder stock, habituated off Western Europe and managed by the International Council for the Exploration of the Seas (ICES). This data set (see **Appendix B**, Figure 5.5) comes from the 2001 assessment of North Sea Plaice (ICES, 2002b). Inspection of the stock-recruit scatter (Figure 5.5) indicates that some of the recruitment values are considerably larger than the bulk of the data, points which may adversely affect the LS estimates. Two iteratively re-weighted runs were conducted using two values of Huber's c to assess the sensitivity of the results of the value of c. The LS JA results are:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\alpha}_2)^T = (1.7833, 421.2836)^T,$$

$$\hat{\delta}_{LS} = 236.2327, \text{ and RSS} = 6.8376.$$

Using the IRJA, with $c = 2$ (after 7 iterations) we have the estimates:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\alpha}_2)^T = (1.5440, 420.6233)^T,$$

$$\hat{\delta}_{IRLS} = 272.4235, \text{ and RSS}_W = 5.9647.$$

And for the IRJA with $c = 1.5$ (requires 8 iterations):

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\alpha}_2)^T = (1.4148, 418.0098)^T,$$

$$\hat{\delta}_{IRLS} = 295.4544, \text{ and } \mathrm{RSS}_W = 5.1579.$$

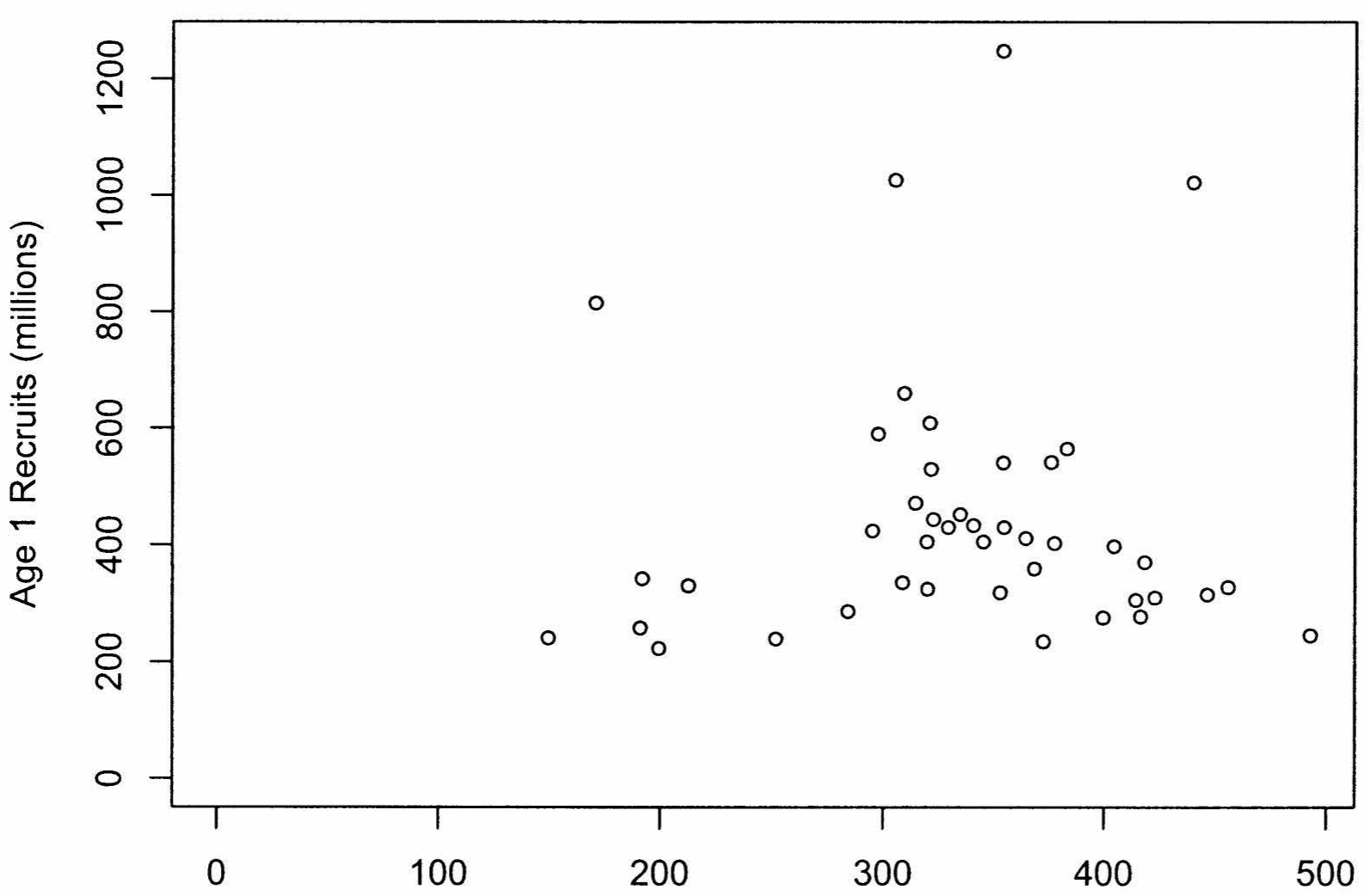

Figure 5.5: North Sea Plaice stock-recruit data.

Figure 5.6 displays the the JA estimate of change-point, together with two IRJA estimates. The model fits indicate the estimated median recruitment. In this example, there are substantial differences across the three sets of estimates, particularly in the estimated change-point. In this case, it is apparent that the estimates are highly sensitive to the value of Huber's c. The filled points in Figure 5.6 indicate those which are downweighted in the iteratively re-weighted estimates using $c = 1.5$. For $c = 2$, just four of these points are down-weighted - the (373, 234) point is added to the down-weighted cases

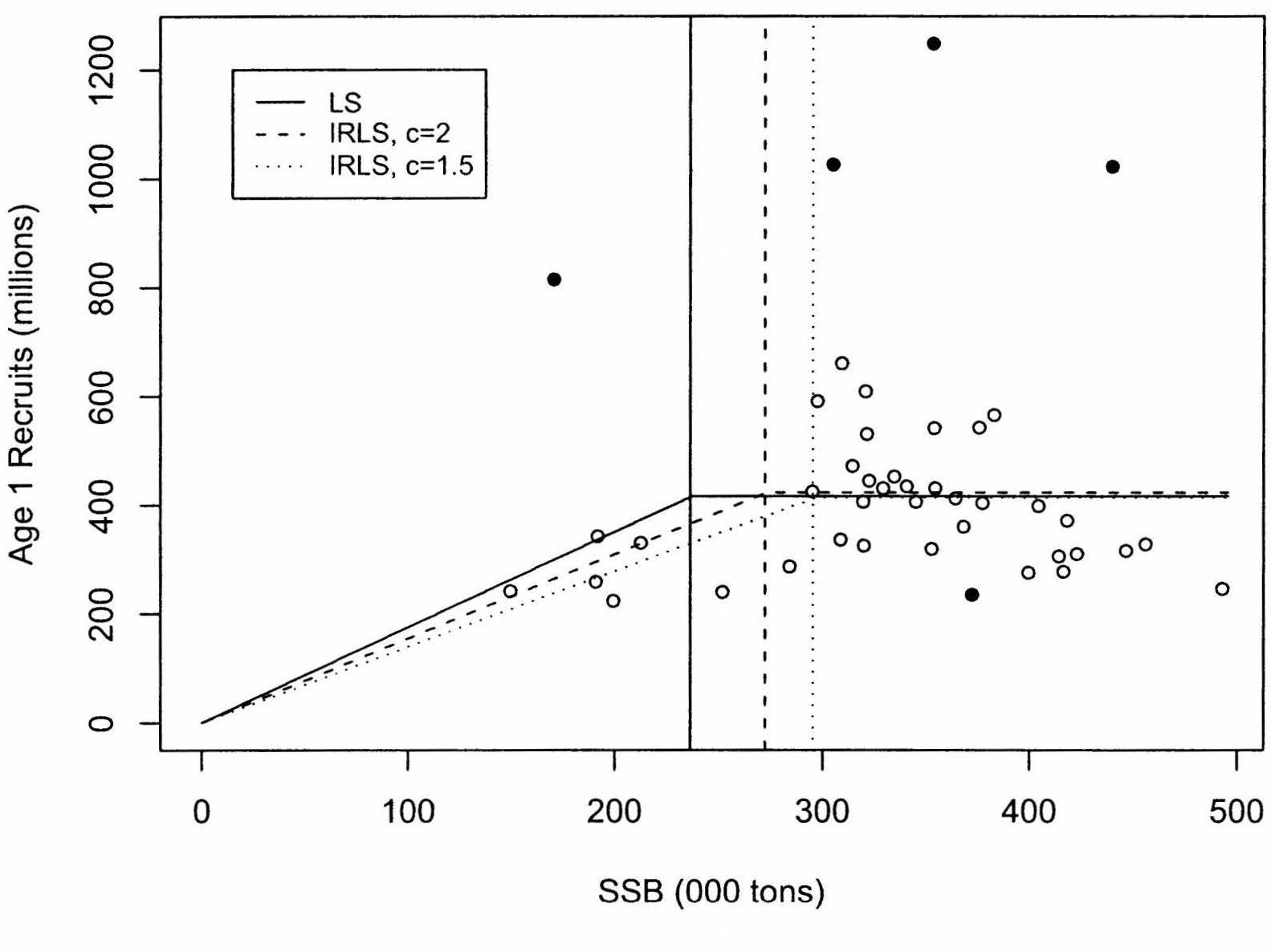

Figure 5.6: LS and IRLS estimates for North Sea Plaice stock-recruit data (lognormal model). Filled points are down-weighted in IRLS using $c = 1.5$. The filled point $(372, 234)$ is not down-weighted when $c = 2$.

when c is reduced to 1.5.

The implication in application of the IRJA to this stock is that depending on the estimate selected as the most appropriate fit, we have different estimates for the point below which impaired recruiment is observed. Comparing LS and $\text{IRLS}_{c=1.5}$, there is a 25% difference is the estimate of $\hat{\delta}$, which would correspond to a 25% difference in the reference point B_{lim} (see discussion in §1.3). Figure 5.7 gives the standardized residuals (log-scale) for each fit,

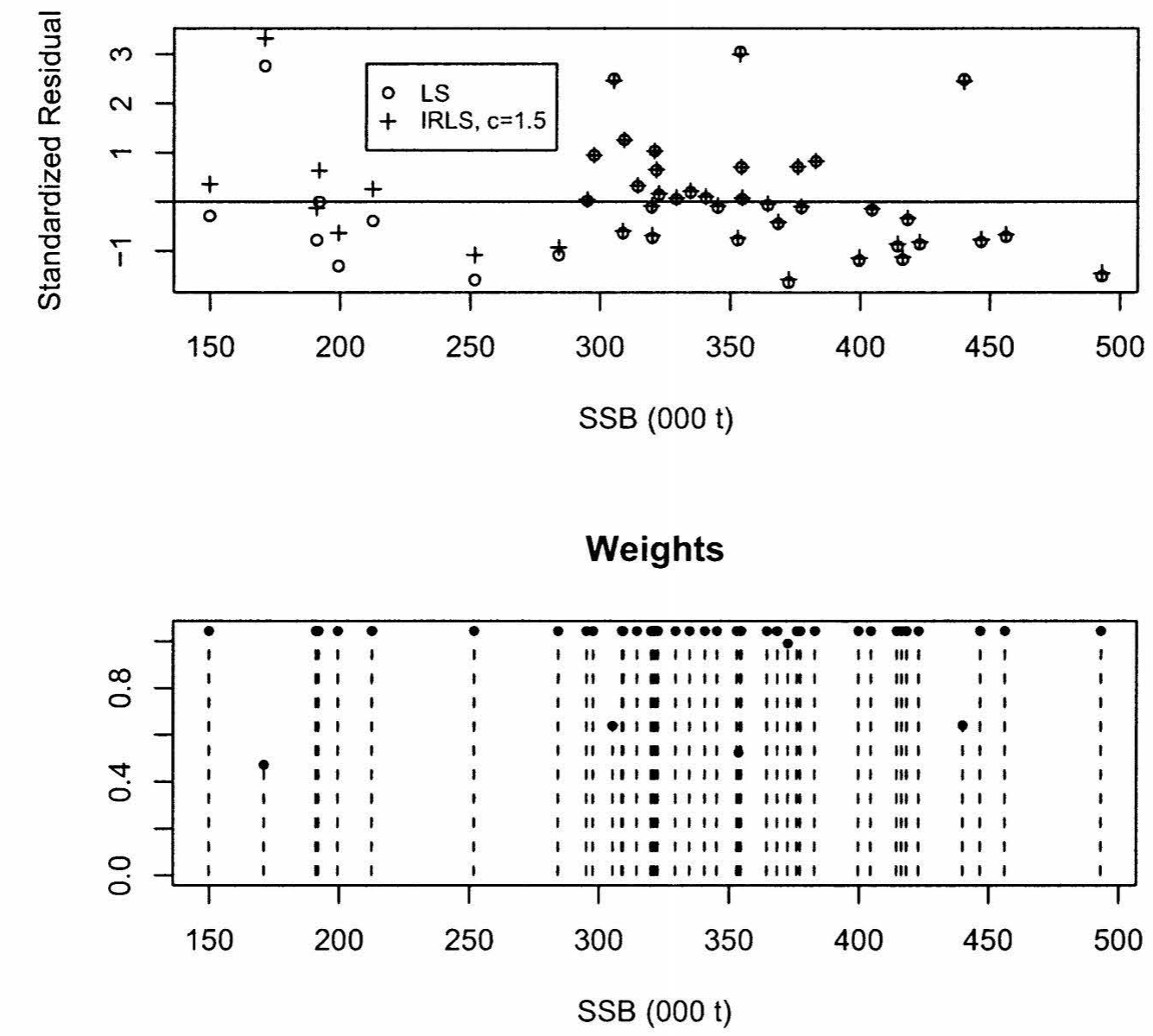

Figure 5.7: Standardized Residuals for LS and IRLS estimates (upper panel); Weights from IRLS (lower panel).

and the weights for the $c = 1.5$ case. In the estimated parameters for each of these runs, there is little change in $\hat{\alpha_2}$. However, robustifying the estimation, the estimated change-point is increased, and accordingly, the slope of the first most segment decreases. (Recall that the constraint to meet is $\beta_1\delta = \alpha_2$.) Thus, the LS and IRLS residuals are considerably different for those SSB values which are less than the estimated change-point. In the lower panel, observe that the four large recruitment values discussed earlier are considerably down-weighted.

Bootstrapping the residuals of the $c = 1.5$ IRLS fit, we estimate $\hat{\text{se}}_{1000}(\hat{\delta}) = 64.8217$.

For several other stock-recruit data sets, the LS JA and IRJA were applied, and it was found that the estimates of $\hat{\delta}_{LS}$ and $\hat{\delta}_{IRLS}$ were quite similar, as with the 3LNO American Plaice. It is also possible that the estimates of $\hat{\delta}_{LS}$ and $\hat{\delta}_{IRLS}$ are equal. However, for the North Sea Plaice, the data seem consistent with a hockey-stick formulation, and given that several recruitment values appear to be outlying, the IRJA offers robustification over the LS JA.

Chapter 6

Conclusion

6.1 Concluding Remarks

In this practicum, robust estimators are developed for the segmented regression model. The focus is on robustifying the Julious algorithm via iteratively re-weighting, extending the work of Julious (2001). Simulation studies were conducted to assess the performance of the iterative re-weighting for the hockey stick model. The methods are applied to estimate parameters of a segmented regression model from a physiological data set and the parameters of the hockey-stick model for two stock-recruit data sets from fisheries science. Key results are given below, and directions for future research with discussion focussed on fisheries science are found in the following section.

In developing robust estimators for the segmented regression problem, we apply an iterative re-weighting scheme, following the iteratively re-weighted least squares work of Holland and Welsch (1977). In practice, these methods

involve subjective choices: the type of weighting function, the parameter(s) of the weighting function, and the test for convergence are all choices which affect the results of the robust estimation, and should be selected by the researcher carefully. The development of the robust estimators for segmented regression models under normal and lognormal errors is the primary result of this practicum.

Simulation studies conclusively demonstrate that in the presence of outliers, the robust estimators are preferable in terms of bias and MSE. If no outliers are present, then the least squares and iteratively re-weighted least squares perform equally well. In practice, when the presence of outliers may be impossible to quantify, the robust estimation is advocated in addition to LS analysis. Simulation studies conducted for the hockey-stick model under log-normal errors indicate that the estimates of the change-point are biased, for both least squares and iteratively re-weighted least squares.

Application results demonstrate the potential of the method. For the physiological data set considered by Julious (2001), we find that the robustified estimate is approximately 5% larger than the least squares estimate, which could have important implications in practice. The robust estimators were applied to two fisheries science data sets, and we find that for one of the cases, there is little difference in the least squares and the robust estimates. However, in the second case considered, we find that the estimated regression lines from the least squares and the iteratively re-weighted methods are not at all similar, and further, that the robust estimate is sensitive to the

parameter used in the weighting function.

6.2 Future Research and Limitations

Despite several decades of research and methodological advances, stock-recruitment relationships and related estimation are still a topic of current fisheries research. On a broad scale, no one parametric model is universally applicable to all fish stocks, including the hockey-stick model studied in this practicum. Research continues on existing parametric and non-parametric methods, such as accounting for additional covariates which may affect recruitment. For several fish stocks, there is no evident stock and recruitment relationship. In such cases, blind application of existing methods is inappropriate. Sprent (1961) stated: "A biologist will often postulate a two-phase linear regression rather than some alternative such as a parabolic one on largely intuitive grounds, and his decision on this point must be to some extent a matter of experience and common sense, as is generally the case in selecting appropriate hypotheses and models for statistical examination. In many cases a two-phase regression can only be a reasonable approximation, adequate for many purposes, but by no means a complete description of what is taking place." This quote befits the application of the hockey-stick model to stock and recruitment data.

For the hockey-stick model, future work should include an appropriate method for determining the value of the weighting function parameter(s). In this

work, the choice of c in Huber's weight function was selected based upon efficiency of an estimator of the mean from normal samples. An examination of the efficiency of the estimators (JA versus IRJA) for several values of c would be a useful contribution.

In certain applications, it may be useful to adapt the method to provide robust estimates allowing for outliers in the independent variable (see §1.2). Such analysis was not considered in this practicum.

Additional research could determine the conditions under which cyclical non-convergence of the IRJA occurs. This problem (as described in §4.3) occurred infrequently and was not considered a focal point of this research.

In addition, future research for stock and recruitment application should explore alternate estimation methods, moving away from least squares, which could avoid weighting on the log-scale. Alternate estimation methods may allow further exploration of the doorhinge model (3.7) as a potential model for stock and recruitment processes.

Bibliography

[1] Bacon, D.W., and Watts, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, **58**, p. 525-534.

[2] Barnett, V., and Lewis, T. (1995). *Outliers in Statistical Data*, Third Edition. Wiley, New York.

[3] Barrowman, N.J., and Myers, R.A. (2000). Still more spawner-recruitment curves: the hockey stick and its generalizations. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, p. 665-676.

[4] Beaton, A.E., and Tukey, J.W. (1974). The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data. *Technometrics*, **16**, p. 147-185.

[5] Bennett, G.W. (1988). Determination of Anaerobic Threshold. *Canadian Journal of Statistics*, **16**, p. 307-316.

[6] Beverton, R.J.H., and Holt, S.J. (1957). *On the Dynamics of Exploited Fish Populations*. Chapman & Hall, London.

[7] Birch, J.B. (1980a). Effects of the Starting Value and Stopping Rule on Robust Estimates Obtained by Iterated Weighted Least Squares.

Communications in Statistics: Simulation and Computation, **B9**(2), p. 141-154.

[8] Birch, J.B. (1980b). Some Convergence Properties of Iterated Weighted Least Squares in the Location Model. *Communications in Statistics: Simulation and Computation*, **B9**(4), p. 359-369.

[9] Birkes, D., and Dodge, Y. (1993). *Alternative Methods of Regression.* John Wiley & Sons, New York.

[10] Bishop, C.A., Murphy, E.F., Davis, M.B., Baird, J.W., and Rose, G.A. (1993). An Assessment of the Cod Stock in NAFO Divisions 2J+3KL. *Northwest Atlantic Fisheries Organization SCR Document* 93/86. Ser. No. N2271.

[11] Braun, J.V., Braun, R.K., and Müller, H.-G. (2000). Multiple change-point fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika*, **87**, p. 301-314.

[12] Butterworth, D.S., and Bergh, M.O. (1993). The development of a management procedure for the South African anchovy resource, p. 83-99. In: Smith, S.J., Hunt, J.J., and Rivard, D. (eds.) Risk evaluation and biological reference points for fisheries management. *Canadian Special Publication of Fisheries and Aquatic Sciences*, **120**.

[13] Carlstein, E., Müller, H.-G., and Siegmund, D. (1994). *Change-point Problems. IMS Lecture Notes-Monograph Series*, **23**. IMS, Hayward.

[14] Carroll, R.J., and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, London.

[15] Casella, G., and Berger, R.L. (1990). *Statistical Inference.* Duxbury Press, Belmont, CA.

[16] Chernoff, H., and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, p. 999-1018.

[17] Chiu, G.A. (2002). MS. *Bent-Cable Regression for Assessing Abruptness of Change.* Ph.D. Thesis, Simon Fraser University. `http://www.stat.sfu.ca/alumni/theses/chiu.pdf`

[18] Csörgó, M., and Horváth, L. (1988). Nonparametric Methods for Changepoint Problems. In Krishnaiah, P.R., and Rao, C.R. (Eds.) *Handbook of Statistics*, **7**, p. 403-425.

[19] Draper, N.R., and Smith, H. (1981). *Applied Regression Analysis.* Second Edition. Wiley, New York.

[20] Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

[21] Feder, P.I. (1975a). On Asymptotic Distribution Theory in Segmented Regression Problems - Identified Case. *The Annals of Statistics*, **3**, p. 43-83.

[22] Feder, P.I. (1975b). The Log Likelihood Ratio in Segmented Regression. *The Annals of Statistics*, **3**, p. 84-97.

[23] Field, C., and Smith, B. (1994). Robust Estimation: A weighted maximum likelihood approach. *International Statistical Review*, **62**, p. 405-424.

[24] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* John Wiley & Sons, New York.

[25] Heiberger, R.M., and Becker, R.A. (1992). Design of an S function for Robust Regression using Iteratively Reweighted Least Squares. *Journal of Computational and Graphical Statistics*, **1**, p. 181-196.

[26] Hinkley, D.V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, **56**, p. 495-504.

[27] Holland, P.W., and Welsch, R.E. (1977). Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics: Theory and Methods*, **A6**, p. 813-827.

[28] Hudson, D.J. (1966). Fitting Segmented Curves whose Join Points have to be Estimated. *Journal of the American Statistical Association*, **61**, p. 1097-1129.

[29] Huber, P.J. (1977). *Robust Statistical Procedures.* SIAM, Philadelphia.

[30] Huber, P.J. (1981). *Robust Statistics.* John Wiley & Sons, New York.

[31] ICES. (2002a). Report of the Study Group on the Further Development of the Precautionary Approach to Fishery Management. *ICES CM 2002/ACFM:10.*

[32] ICES. (2002b). Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak. *ICES CM 2002/ACFM:01.*

[33] Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, **5**, p. 299-314.

[34] Julious, S.A. (2001). Estimation and Inference in a changepoint regression problem. *The Statistician*, **50**, p.51-62.

[35] Jurečková, J., and Sen, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* John Wiley & Sons, New York.

[36] Krishnaiah, P.R., and Miao, B.Q. (1988). Review About Estimation of Change Points. In: Krishnaiah P.R., and Rao, C.R. (eds.) *Handbook of Statistics, v.7, Quality Control and Reliability.* Elsevier Science Publishers B.V., p. 375-402.

[37] Kotz, S., and Johnson, N.L., (Eds.) (1988). *Encyclopedia of Statistical Sciences*, **8**. John Wiley & Sons, New York.

[38] Lerman, P.M. (1980). Fitting segmented regression models by grid search. *Applied Statistics*, **29**, p. 77-84.

[39] Loader, C.R. (1996). Changepiont estimation using nonparameteric regression. *The Annals of Statistics*, **24**, p. 1667-1678.

[40] McGee, V.E., and Carleton, W.T. (1970). Piecewise Regression. *Journal of the American Statistical Association*, **65**, p. 1109-1124.

[41] Müller, H.G., and Wang, J.-L. (1994). Change-point models for hazard functions. In: Carstein, E., Müller, H.-G., and Siegmund, D. (eds.) *Change-point Problems. IMS Lecture Notes-Monograph Series* **23**. IMS, Hayward.

[42] NAFO. (2003). Report of NAFO Scientific Council Workshop of the Precautionary Approach to Fisheries Management. *Northwest Atlantic Fisheries Organization SCS Document* 03/05.

[43] Needle, C.M. (2002). Recruitment models: diagnosis and prognosis. *Reviews in Fish Biology and Fisheries*, **11**, p. 95-111.

[44] Neter, J., Dutner M.H., Nachtsheim, C.J., and Wasserman, W. (1996). *Applied Linear Statistical Models*, Fourth Edition. IRWIN, Chicago.

[45] Ogden, R.T., and Lynch, J.D. (1999). Bayesian analysis of change-point models. In: Müller, P., and Vidakovic, B. (eds.) *Bayesian Inference in Wavelet Based Models*. John Wiley & Sons, New York, p. 67-82.

[46] Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika*, **41**, p. 100-115.

[47] Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, p. 523-526.

[48] Piepho, H.P., and Ogutu, J.O. (2003). Inference for the break point in segmented regression with application to longitudinal data. *Biometrical Journal*, **45**, p. 591-601.

[49] Plackett, R.L. (1972). The discovery of the method of least squares. *Biometrika*, **59**, p. 239-251.

[50] Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, p. 873-880.

[51] Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, **55**, p. 324-330.

[52] Ray, B.K., and Tsay, R.S. (2002). Bayesian Methods for Change-point Detection in Long-range Dependent Processes. *Journal of Time Series Analysis*, **23**, p. 687-705.

[53] Rencher, A.C. (2000). *Linear Models in Statistics.* John Wiley & Sons, New York.

[54] Robison, D.E. (1964). Estimates for the points of intersection of two polynomial regressions. *Journal of the American Statistical Association*, **59**, p. 214-224.

[55] Rousseeuw, P.J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, **79**, p. 871-880.

[56] Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons, New York.

[57] Ryan, T.P. (1997). *Modern Regression Methods.* John Wiley & Sons, New York.

[58] Shaban, S.A. (1980). Change Point Problem and Two-Phase Regression: an Annotated Bibliography. *International Statistical Review*, **48**, p. 83-93.

[59] Shelton, P.A., and Rice, J.C. (2002). Limits to overfishing: reference points in the context of the Canadian persepctive on the precautionary approach. *Canadian Science Advisory Secretariat Research Document* 2002/084.

[60] Shelton, P.A., and Rivard, D. (2003). Developing a Precautionary Approach to Fisheries Management in Canada - the Decade following the Cod Collapses. *Northwest Atlantic Fisheries Organization SCR Document* 03/1. Ser.No. N4808.

[61] Sprent, P. (1961) Some Hypotheses Concerning Two Phase Regression Lines. *Biometrics*, **17**, p. 634-645.

[62] Stansinopoulos, M., and Rigby, R. (1992). Detecting break points in generalised linear models. *Computational Statistics & Data Analysis*, **13**, p. 461-471.

[63] Venables, W.N., and Ripley, B.D. (1999). *Modern Applied Statistics with S-PLUS*, Third Edition. Springer-Verlag, New York.

[64] Worsley, K.J. (1983). Testing for an two phase multiple regression. *Technometrics*, **25**, p. 35-42.

[65] Wu. W.B., Woodroofe, M., and Mentz, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika*, **88**, p. 793-804.

[66] Yohai, V.J. (1987). High Breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, p. 642-656.

Appendix A: Simulation Summary Tables

Refer to **Chapter 4** for details.

Table A1: Hockey-Stick Model, Normal Errors. Summaries of $\hat{\delta}$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
1	25	25	0	1	0.017	0.017	0.533	0.541
2	25	25	0	9	0.059	0.059	6.075	6.148
3	25	25	0.15	1	0.057	0.017	2.959	1.596
4	25	25	0.15	9	0.126	-0.558	41.776	28.005
5	25	50	0	1	-0.030	-0.029	0.320	0.323
6	25	50	0	9	0.004	0.007	2.959	3.001
7	25	50	0.15	1	-0.043	-0.039	1.492	0.689
8	25	50	0.15	9	0.042	-0.016	16.205	6.999
9	25	75	0	1	0.008	0.008	0.343	0.349
10	25	75	0	9	0.040	0.043	3.529	3.599
11	25	75	0.15	1	0.017	0.015	1.802	0.792
12	25	75	0.15	9	0.451	0.358	24.301	14.244

Table A1, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
13	50	25	0	1	0.019	0.020	0.290	0.291
14	50	25	0	9	0.029	0.031	2.684	2.717
15	50	25	0.15	1	0.019	0.009	1.412	0.638
16	50	25	0.15	9	0.016	0.018	19.800	8.486
17	50	50	0	1	0.007	0.006	0.165	0.167
18	50	50	0	9	0.008	0.010	1.599	1.621
19	50	50	0.15	1	0.003	-0.009	0.832	0.344
20	50	50	0.15	9	0.082	0.041	7.760	3.050
21	50	75	0	1	-0.008	-0.009	0.163	0.164
22	50	75	0	9	-0.022	-0.019	1.538	1.572
23	50	75	0.15	1	-0.011	-0.003	0.798	0.324
24	50	75	0.15	9	0.122	0.058	9.430	3.573

Table A1, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
25	100	25	0	1	0.008	0.008	0.135	0.136
26	100	25	0	9	-0.009	-0.007	1.309	1.326
27	100	25	0.15	1	0.011	-0.003	0.673	0.271
28	100	25	0.15	9	0.015	-0.027	7.432	2.737
29	100	50	0	1	-0.005	-0.005	0.082	0.084
30	100	50	0	9	0.008	0.007	0.750	0.758
31	100	50	0.15	1	-0.009	-0.007	0.390	0.161
32	100	50	0.15	9	0.022	0.003	3.715	1.425
33	100	75	0	1	-0.004	-0.004	0.081	0.081
34	100	75	0	9	0.012	0.012	0.761	0.771
35	100	75	0.15	1	0.019	0.009	0.395	0.159
36	100	75	0.15	9	0.023	-0.004	3.855	1.468

Table A2: Hockey-Stick Model, Normal Errors. Summaries of $\hat{\beta}_1$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
1	25	25	0	1	0.0000	0.0000	0.0008	0.0008
2	25	25	0	3	0.0073	0.0074	0.0090	0.0090
3	25	25	0.15	1	0.0020	0.0016	0.0043	0.0024
4	25	25	0.15	3	0.1180	0.1321	0.4498	0.4558
5	25	50	0	1	0.0008	0.0008	0.0001	0.0001
6	25	50	0	3	0.0016	0.0016	0.0009	0.0009
7	25	50	0.15	1	0.0013	0.0011	0.0005	0.0002
8	25	50	0.15	3	0.0054	0.0033	0.0050	0.0021
9	25	75	0	1	0.0000	0.0000	0.0000	0.0000
10	25	75	0	3	0.0004	0.0004	0.0003	0.0003
11	25	75	0.15	1	0.0000	-0.0001	0.0001	0.0001
12	25	75	0.15	3	0.0006	-0.0001	0.0014	0.0006

Table A2, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
13	50	25	0	1	-0.0002	-0.0002	0.0004	0.0004
14	50	25	0	3	0.0026	0.0026	0.0039	0.0040
15	50	25	0.15	1	0.0017	0.0007	0.0021	0.0010
16	50	25	0.15	3	0.0566	0.0216	0.2664	0.0959
17	50	50	0	1	-0.0001	-0.0001	0.0000	0.0000
18	50	50	0	3	0.0009	0.0009	0.0005	0.0005
19	50	50	0.15	1	0.0001	0.0000	0.0002	0.0001
20	50	50	0.15	3	0.0026	0.0009	0.0023	0.0009
21	50	75	0	1	0.0000	0.0001	0.0000	0.0000
22	50	75	0	3	0.0005	0.0005	0.0001	0.0001
23	50	75	0.15	1	0.0003	0.0001	0.0001	0.0000
24	50	75	0.15	3	0.0008	0.0004	0.0007	0.0003

Table A2, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
25	100	25	0	1	-0.0001	-0.0001	0.0002	0.0002
26	100	25	0	3	0.0024	0.0023	0.0019	0.0019
27	100	25	0.15	1	0.0006	0.0005	0.0010	0.0004
28	100	25	0.15	3	0.0114	0.0050	0.0123	0.0041
29	100	50	0	1	0.0002	0.0002	0.0000	0.0000
30	100	50	0	3	0.0002	0.0003	0.0002	0.0002
31	100	50	0.15	1	0.0003	0.0002	0.0001	0.0000
32	100	50	0.15	3	0.0011	0.0003	0.0011	0.0004
33	100	75	0	1	0.0001	0.0001	0.0000	0.0000
34	100	75	0	3	0.0001	0.0001	0.0001	0.0001
35	100	75	0.15	1	0.0000	0.0000	0.0000	0.0000
36	100	75	0.15	3	0.0004	0.0003	0.0003	0.0001

Table A3: Hockey-Stick Model, Normal Errors. Summaries of $\hat{\alpha}_2$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
1	25	25	0	1	-0.0031	-0.0035	0.0527	0.0540
2	25	25	0	3	0.0216	0.0218	0.4884	0.4964
3	25	25	0.15	1	0.0010	-0.0036	0.2597	0.1008
4	25	25	0.15	3	0.0581	-0.0072	2.4352	1.0149
5	25	50	0	1	0.0040	0.0054	0.0803	0.0821
6	25	50	0	3	0.0428	0.0411	0.7511	0.7644
7	25	50	0.15	1	0.0014	0.0050	0.3942	0.1630
8	25	50	0.15	3	0.0656	0.0440	3.6326	1.5498
9	25	75	0	1	0.0043	0.0052	0.1697	0.1721
10	25	75	0	3	0.0497	0.0515	1.6162	1.6427
11	25	75	0.15	1	0.0089	0.0055	0.8804	0.3977
12	25	75	0.15	3	0.3744	0.2860	12.8314	7.9082

Table A3, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
13	50	25	0	1	0.0039	0.0043	0.0269	0.0270
14	50	25	0	3	-0.0024	-0.0027	0.2279	0.2315
15	50	25	0.15	1	0.0096	0.0033	0.1324	0.0516
16	50	25	0.15	3	0.0130	-0.0031	1.1911	0.4804
17	50	50	0	1	0.0010	0.0005	0.0414	0.0420
18	50	50	0	3	0.0288	0.0294	0.3847	0.3872
19	50	50	0.15	1	-0.0041	-0.0124	0.2028	0.0793
20	50	50	0.15	3	0.0939	0.0399	1.7196	0.7129
21	50	75	0	1	-0.0054	-0.0062	0.0818	0.0822
22	50	75	0	3	0.0079	0.0092	0.7625	0.7790
23	50	75	0.15	1	0.0051	0.0015	0.4006	0.1611
24	50	75	0.15	3	0.1282	0.0691	4.6121	1.9029

Table A3, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
25	100	25	0	1	0.0007	0.0006	0.0133	0.0134
26	100	25	0	3	0.0042	0.0040	0.1252	0.1264
27	100	25	0.15	1	0.0020	-0.0007	0.0612	0.0245
28	100	25	0.15	3	0.0186	-0.0017	0.5757	0.2243
29	100	50	0	1	0.0014	0.0017	0.0206	0.0211
30	100	50	0	3	0.0070	0.0080	0.1847	0.1871
31	100	50	0.15	1	0.0000	-0.0005	0.0965	0.0385
32	100	50	0.15	3	0.0219	-0.0023	0.8303	0.3206
33	100	75	0	1	-0.0005	-0.0004	0.0394	0.0396
34	100	75	0	3	0.0117	0.0110	0.3827	0.3859
35	100	75	0.15	1	0.0173	0.0081	0.1869	0.0760
36	100	75	0.15	3	0.0236	0.0106	1.7726	0.7041

Table A4: Hockey-Stick Model, Lognormal Errors. Summaries of $\hat{\delta}$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
1	25	25	0	0.025	0.033	0.024	3.925	3.974
2	25	25	0	0.1	0.235	0.226	15.577	15.716
3	25	25	0.15	0.025	0.095	0.052	8.942	6.675
4	25	25	0.15	0.1	0.471	0.219	46.606	36.656
5	25	50	0	0.025	0.141	0.151	10.550	10.720
6	25	50	0	0.1	0.811	0.815	62.050	62.804
7	25	50	0.15	0.025	0.327	0.218	29.860	20.143
8	25	50	0.15	0.1	2.323	1.867	177.087	134.993
9	25	75	0	0.025	0.874	0.858	50.848	51.379
10	25	75	0	0.1	1.603	1.624	145.861	146.595
11	25	75	0.15	0.025	0.964	1.369	89.247	80.032
12	25	75	0.15	0.1	0.240	1.069	221.764	196.202

Table A4, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
13	50	25	0	0.025	-0.006	-0.005	1.839	1.856
14	50	25	0	0.1	0.176	0.176	7.595	7.659
15	50	25	0.15	0.025	0.008	-0.011	4.211	2.981
16	50	25	0.15	0.1	0.242	0.179	20.448	14.493
17	50	50	0	0.025	0.124	0.126	5.307	5.370
18	50	50	0	0.1	0.322	0.328	24.876	25.095
19	50	50	0.15	0.025	0.211	0.154	12.856	8.934
20	50	50	0.15	0.1	1.052	0.722	77.206	50.903
21	50	75	0	0.025	0.291	0.286	22.472	22.863
22	50	75	0	0.1	1.075	1.067	92.641	93.843
23	50	75	0.15	0.025	1.135	0.976	56.641	43.378
24	50	75	0.15	0.1	1.275	1.538	154.684	132.225

Table A4, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\delta})$	$Bias_{IRLS}(\hat{\delta})$	$MSE_{LS}(\hat{\delta})$	$MSE_{IRLS}(\hat{\delta})$
25	100	25	0	0.025	0.006	0.008	0.916	0.922
26	100	25	0	0.1	0.084	0.076	3.691	3.715
27	100	25	0.15	0.025	0.041	0.013	2.065	1.449
28	100	25	0.15	0.1	0.115	0.069	8.770	6.080
29	100	50	0	0.025	0.020	0.023	2.680	2.724
30	100	50	0	0.1	0.160	0.161	11.086	11.177
31	100	50	0.15	0.025	0.072	0.046	6.070	4.277
32	100	50	0.15	0.1	0.327	0.261	27.452	18.744
33	100	75	0	0.025	0.229	0.235	8.715	8.989
34	100	75	0	0.1	0.952	0.991	49.311	50.086
35	100	75	0.15	0.025	0.496	0.316	27.954	18.201
36	100	75	0.15	0.1	1.616	1.509	94.447	74.361

Table A5: Hockey-Stick Model, Lognormal Errors. Summaries of $\hat{\beta}_1$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
1	25	25	0	0.025	0.004	0.004	0.005	0.005
2	25	25	0	0.1	0.015	0.015	0.021	0.022
3	25	25	0.15	0.025	0.010	0.008	0.012	0.009
4	25	25	0.15	0.1	0.081	0.078	0.246	0.234
5	25	50	0	0.025	0.001	0.001	0.002	0.002
6	25	50	0	0.1	0.009	0.010	0.010	0.010
7	25	50	0.15	0.025	0.004	0.003	0.005	0.003
8	25	50	0.15	0.1	0.020	0.013	0.030	0.019
9	25	75	0	0.025	0.002	0.002	0.001	0.001
10	25	75	0	0.1	0.010	0.010	0.006	0.006
11	25	75	0.15	0.025	0.008	0.005	0.003	0.002
12	25	75	0.15	0.1	0.032	0.023	0.046	0.041

Table A5, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
13	50	25	0	0.025	0.002	0.002	0.002	0.002
14	50	25	0	0.1	0.005	0.005	0.009	0.010
15	50	25	0.15	0.025	0.006	0.005	0.005	0.004
16	50	25	0.15	0.1	0.024	0.019	0.055	0.054
17	50	50	0	0.025	0.000	0.000	0.001	0.001
18	50	50	0	0.1	0.004	0.004	0.004	0.004
19	50	50	0.15	0.025	0.002	0.002	0.002	0.002
20	50	50	0.15	0.1	0.009	0.005	0.011	0.008
21	50	75	0	0.025	0.002	0.002	0.001	0.001
22	50	75	0	0.1	0.007	0.007	0.003	0.003
23	50	75	0.15	0.025	0.002	0.001	0.002	0.001
24	50	75	0.15	0.1	0.012	0.009	0.007	0.005

Table A5, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\beta}_1)$	$Bias_{IRLS}(\hat{\beta}_1)$	$MSE_{LS}(\hat{\beta}_1)$	$MSE_{IRLS}(\hat{\beta}_1)$
25	100	25	0	0.025	0.001	0.001	0.001	0.001
26	100	25	0	0.1	0.003	0.004	0.004	0.004
27	100	25	0.15	0.025	0.002	0.002	0.002	0.002
28	100	25	0.15	0.1	0.009	0.007	0.011	0.008
29	100	50	0	0.025	0.001	0.001	0.001	0.001
30	100	50	0	0.1	0.001	0.001	0.002	0.002
31	100	50	0.15	0.025	0.001	0.001	0.001	0.001
32	100	50	0.15	0.1	0.004	0.002	0.005	0.003
33	100	75	0	0.025	0.000	0.000	0.000	0.000
34	100	75	0	0.1	0.002	0.002	0.001	0.001
35	100	75	0.15	0.025	0.001	0.001	0.001	0.001
36	100	75	0.15	0.1	0.003	0.001	0.003	0.002

Table A6: Hockey-Stick Model, Lognormal Errors. Summaries of $\hat{\alpha}_2$.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
1	25	25	0	0.025	0.011	0.006	0.828	0.845
2	25	25	0	0.1	0.114	0.110	3.533	3.552
3	25	25	0.15	0.025	0.054	0.031	1.965	1.388
4	25	25	0.15	0.1	0.390	0.282	9.409	6.442
5	25	50	0	0.025	0.099	0.108	5.250	5.308
6	25	50	0	0.1	0.744	0.749	28.895	29.310
7	25	50	0.15	0.025	0.263	0.192	14.405	9.563
8	25	50	0.15	0.1	1.875	1.519	87.814	66.202
9	25	75	0	0.025	0.864	0.856	38.466	38.789
10	25	75	0	0.1	1.928	1.954	115.466	116.439
11	25	75	0.15	0.025	1.324	1.555	69.951	63.107
12	25	75	0.15	0.1	1.409	1.853	179.871	159.518

Table A6, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
13	50	25	0	0.025	-0.004	-0.005	0.411	0.415
14	50	25	0	0.1	0.068	0.067	1.635	1.656
15	50	25	0.15	0.025	0.028	0.016	0.907	0.635
16	50	25	0.15	0.1	0.133	0.105	4.208	2.882
17	50	50	0	0.025	0.062	0.063	2.579	2.600
18	50	50	0	0.1	0.310	0.309	11.994	12.168
19	50	50	0.15	0.025	0.186	0.150	6.185	4.353
20	50	50	0.15	0.1	0.876	0.573	38.128	25.071
21	50	75	0	0.025	0.339	0.342	16.395	16.758
22	50	75	0	0.1	1.352	1.348	70.705	71.762
23	50	75	0.15	0.025	1.140	0.962	42.741	32.876
24	50	75	0.15	0.1	1.684	1.806	120.547	104.192

Table A6, continued.

Run #	N	δ	p	σ^2	$Bias_{LS}(\hat{\alpha}_2)$	$Bias_{IRLS}(\hat{\alpha}_2)$	$MSE_{LS}(\hat{\alpha}_2)$	$MSE_{IRLS}(\hat{\alpha}_2)$
25	100	25	0	0.025	-0.002	-0.001	0.211	0.213
26	100	25	0	0.1	0.057	0.053	0.870	0.879
27	100	25	0.15	0.025	0.026	0.014	0.458	0.316
28	100	25	0.15	0.1	0.079	0.054	1.930	1.318
29	100	50	0	0.025	0.024	0.026	1.299	1.319
30	100	50	0	0.1	0.116	0.120	5.413	5.472
31	100	50	0.15	0.025	0.055	0.043	2.824	1.972
32	100	50	0.15	0.1	0.275	0.207	13.213	9.206
33	100	75	0	0.025	0.200	0.204	6.633	6.807
34	100	75	0	0.1	0.967	1.005	36.882	37.614
35	100	75	0.15	0.025	0.483	0.321	20.518	13.288
36	100	75	0.15	0.1	1.561	1.421	71.873	56.141

Appendix B: Application Data Sets

Refer to **Chapter 5** for details.

Table B1: Physiological data ($n = 35$) from Julious (2001).

Oxygen=Volume of Oxygen (L/min), and Carbon Dioxide=Volume of Carbon Dioxide (L/min)

Oxygen	Carbon Dioxide	Oxygen	Carbon Dioxide
12.5	0.75	44.2	2.12
26.2	1.12	47.9	2.35
24.8	0.98	49.9	2.50
27.4	1.13	48.1	2.48
31.1	1.31	48.4	2.49
34.6	1.47	51.7	2.71
21.5	0.93	51.8	2.74
27.9	1.34	55.5	3.00
29.2	1.36	54.9	3.02
35.2	1.60	57.0	3.21
32.6	1.47	57.9	3.30
34.9	1.57	58.3	3.37
34.9	1.59	58.2	3.42
37.6	1.73	59.5	3.53
36.3	1.68	59.7	3.55
40.1	1.88	61.8	3.76
42.7	2.01	48.4	2.96
43.4	2.07		

Table B2: Stock-recruit data ($n = 37$) for 3LNO American Plaice. Spawner biomass (SSB) units are thousands of tons, and the recruitment values are in millions of fish. The corresponding year is also tabled. Data rounded to three decimal places.

Year	SSB (000 t)	Rec (millions)	Year	SSB (000 t)	Rec (millions)
1960	125.530	783.463	1979	100.982	520.249
1961	135.984	679.684	1980	112.252	508.438
1962	147.358	607.180	1981	102.201	433.575
1963	157.850	479.781	1982	91.542	384.962
1964	170.034	475.168	1983	102.010	438.498
1965	191.805	446.019	1984	128.157	506.875
1966	217.279	556.770	1985	138.567	734.456
1967	236.082	659.795	1986	145.634	493.425
1968	213.499	795.917	1987	130.813	478.201
1969	203.359	761.297	1988	124.431	661.824
1970	176.259	799.137	1989	115.509	1088.944
1971	144.235	752.060	1990	73.466	901.637
1972	113.622	631.194	1991	47.983	619.074
1973	88.815	594.013	1992	26.949	349.556
1974	71.766	545.643	1993	16.390	327.817
1975	68.760	526.005	1994	7.108	287.121
1976	64.712	512.253	1995	8.192	167.977
1977	72.509	519.830	1996	12.374	145.381
1978	92.532	515.880			

Table B3: Stock-recruit data ($n = 43$) for North Sea Plaice. Spawner biomass (SSB) units are tons, and the recruitment values are in thousands of fish. The corresponding year is also tabled. Note that the estimation in **Chapter 5** scales SSB and recruits by a factor of 1000.

Year	SSB (t)	Rec (thousands)	Year	SSB (t)	Rec (thousands)
1957	354624	429984	1979	309364	659903
1958	340636	433436	1980	295050	424238
1959	345187	405323	1981	305205	1025826
1960	368311	359381	1982	297576	590479
1961	352877	318800	1983	320905	608707
1962	446570	315181	1984	321505	529685
1963	439975	1021880	1985	353680	1247920
1964	422933	309565	1986	354173	540636
1965	414353	305370	1987	382881	564989
1966	416386	277225	1988	364401	411573
1967	493006	245503	1989	404435	397584
1968	456101	327474	1990	377262	402941
1969	418277	370438	1991	319786	405305
1970	399572	275478	1992	284117	285978
1971	372352	234584	1993	251706	239231
1972	375802	541889	1994	212554	330003
1973	334724	451919	1995	190933	258095
1974	308823	335732	1996	170895	814830
1975	320041	324585	1997	149718	241100
1976	314520	471354	1998	199344	223262
1977	329233	430024	1999	191916	342057
1978	322622	443809			