Model-Robust Quantal Regression

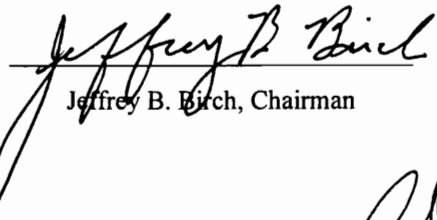
by

Quinton J. Nottingham

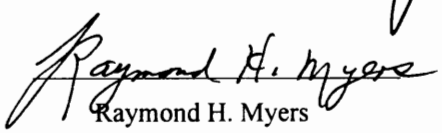
Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
IN
STATISTICS

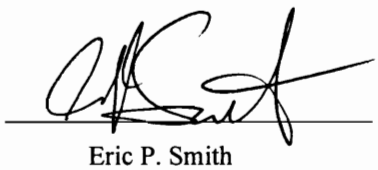
APPROVED:



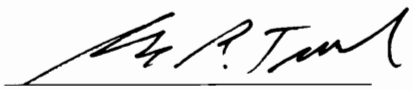
Jeffrey B. Birch, Chairman



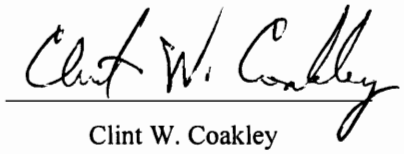
Raymond H. Myers



Eric P. Smith



George R. Terrell



Clint W. Coakley

August 22, 1995
Blacksburg, Virginia

Key Words: Nonparametric, Logistic, Dose-response

Model-Robust Quantal Regression

by

Quinton J. Nottingham

Jeffrey B. Birch, Chairman

Statistics

Abstract

In the analysis of quantal dose-response data, the most commonly used parametric procedure is logistic regression, commonly referred to as "logit analysis." The adequacy of the fit by the logistic regression curve is tested using the chi-square lack-of-fit test. If the lack-of-fit test is not significant, then the logistic model is assumed to be adequate and estimation of effective doses and confidence intervals on the effective doses can be made. When the tolerance distribution of the dose-response data is not known and cannot be assumed by the user, one can use nonparametric methods, such as kernel regression or local linear regression, to estimate the dose-response curve, effective doses, and confidence intervals. This research proposes another alternative to analyzing quantal dose-response data called model-robust quantal regression (MRQR). MRQR linearly combines the parametric and nonparametric predictions with the use of a mixing parameter. MRQR uses logistic regression as the parametric portion of the model and either kernel or local linear regression as the nonparametric portion of the model. Research has shown that the MRQR procedure can improve the fit of the dose-response curve by producing narrower confidence intervals for predictions, while providing improved precision of estimates of the effective doses with respect to logit analysis.

Acknowledgements

My sincere thanks goes to my advisor, Jeffrey Birch, whose expertise in regression analysis and teaching ability have given me a solid foundation. I would also like to thank Jeffrey for his countless hours of editing and suggestions to make this a more readable dissertation. I would like to express my sincere thanks to all of the faculty, staff, and graduate students of the Department of Statistics for their support over the past five years. I would also like to thank each member of my committee, Raymond H. Myers, Eric P. Smith, George R. Terrell, and Clint Coakley for taking an interest in my research.

I would like to thank the State Council of Higher Education for Virginia for partially supporting my graduate work by awarding me the SCHEV Fellowship during my first two years of graduate school.

I would like to thank my wife Jacqueline for her love, support and understanding throughout my graduate studies. I would also like to thank my mother Jennie who has always encouraged me to reach for the stars, as well as constantly reminding me of how proud she is that her youngest son will be a doctor.

Table of Contents

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Tables (continued)

List of Figures

List of Figures (continued)

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

§1.1 Introduction and Literature Review

Biological assays are methods for the determination of the potency of a substance by biological means. There are two types of biological assays—direct and indirect. The direct biological assay is one in which the doses of the standard and test preparations sufficient to produce a specified response is *directly* measured. The direct assay is seldom used in biological applications due to its dependence upon the ability to measure *just* the dose needed for a particular response and *not* one that is *at least* large enough, as in an indirect quantal assay. Indirect quantal assays are methods for the estimation of the potency of a dose by means of the observed response. The response variable of interest measured for each dose can be either quantal or quantitative. For the quantal response, the experiment can only determine whether a subject responds or fails to respond to a particular dose of the drug; that is, an all or nothing response such as death. Thus, if the variable of interest is quantal, it is recorded as a "1" for a response or a "0" for nonresponse. In quantal assays, the primary concern is with the distribution of the tolerance of the subjects to a particular drug or chemical being studied, where the "tolerance of the subject" is the dose just sufficient to elicit a response from the subject. The quantitative response measures some property of the subject such as weight, time of survival, or blood calcium. Quantal assays are the subject of this research.

In quantal assays, the response is a Bernoulli random variable with parameter P_i , the probability that a subject will respond to dose x_i. Then, in a given sequence of doses $x_1, x_2, \ldots, x_d$, we observe the random sequence of 0's and 1's. The experimental results may be summarized as:

Dose	Sample Size	Number Responding	Sample Proportion
x_1	n_1	r_1	p_1
x_2	n_2	r_2	p_2
.	.	.	.
.	.	.	.
.	.	.	.
x_d	n_d	r_d	p_d

where x_i represents the i^{th} dose in increasing order, n_i the number of subjects receiving the i^{th} dose, r_i is the number of subjects responding to the i^{th} dose, and $p_i = \frac{r_i}{n_i}$, the sample proportion, estimates the true probability, P_i, that a randomly selected subject will respond to the i^{th} dose, $i = 1,2,...,d$.

The dose-response relationship is a model of the probability of response, P_i, at dose x_i as

$$P_i = F(x_i)$$

where $F(x_i)$ is the unknown cumulative distribution function (cdf) for the tolerance of the subjects in the population to the drug or chemical. As in many applications in regression, the exact form of F is unknown and the user supplies models that are approximations to the true underlying F. In the analysis of quantal dose-response data, the most commonly used parametric procedures are probit and logit analysis. Probit analysis assumes that the tolerance follows a normal distribution and thus F is a normal cdf. In logit analysis, it is assumed that tolerance follows a logistic distribution and that F is a logistic cdf. For most problems, there is relatively little difference between the normal and logistic specifications of the model since both distributions have similar shape characteristics, differing slightly in the tails of the distributions.

When the user is satisfied with the chosen model, the primary concern then shifts to the estimation of effective doses. The effective dose, denoted by $ED_{100\alpha}$, is the dose at level α where $100\alpha\%$ of the subjects, on the average, show a response. That is, the $ED_{100\alpha}$, is simply the 100α percentile of the distribution F. The most commonly used value of α is 0.5, which indicates the effective dose required to yield 50% response, denoted by ED_{50}. In some cases, experiments are designed to estimate more extreme dose levels. Extreme dose levels, such as the ED_{10} and the ED_{90}, may be regarded as more suitable for characterizing the tolerance distribution, F. For example, when fitting data using probit and logit analyses, the ED_{50} values may be quite similar, but the ED_{10} and ED_{90} values may be different due to the difference in the tails of the logistic and standard normal distributions.

For any estimate of the $ED_{100\alpha}$, appropriate confidence intervals should be given as an interval estimate of the effective dose. In this research, the parametric confidence interval on effective dose will be computed using Fieller's theorem (Finney, 1978). It should be noted that Fieller's intervals do not possess the feature of symmetry. Other methods that can be used for confidence intervals on effective dose are the delta method (Goedhart, 1985) and the likelihood-ratio interval (Cox, 1990).

The material in this research will emphasize use of logit analysis, but the setup will be such that any general form of F may also be used such as probit analysis, using the normal distribution which has slightly lighter tails than the logistic form, the Gompertz distribution (sometimes called the extreme value

distribution), and the Cauchy distribution function. Other models used in the analysis of quantal dose-response data can be found in Ashton (1972), Cox and Snell (1989) and Morgan (1992).

While such values of the $ED_{100\alpha}$ are easily estimated from the parametric procedures, it may be argued that a nonparametric procedure could be more relevant. Until recently, much of the nonparametric analysis of quantal dose-response data has centered on median effective dose or ED_{50} estimation. Classical nonparametric procedures only yield estimates of the ED_{50}. Such procedures include the Spearman-Kärber estimator (Spearman (1908) and Kärber (1931)), the Reed-Muench estimator (Reed and Muench (1938)), and the moving average estimator due to Thompson (1947). The methods mentioned above center on the median effective dose estimation, but not curve estimation. Therefore, if one is interested in estimating the $ED_{100\alpha}$ for $\alpha \neq 0.5$, other nonparametric methods are necessary.

Staniswalis and Cooper (1988) and Muller and Schmitt (1988) used kernel regression methods to estimate dose-response relationships. Staniswalis and Cooper (1988) used kernel estimates to develop approximate confidence intervals for the optimal combination dose of a drug with therapeutic effects at the low doses and toxic effects at high doses, as well as to determine the lethal dose levels of a toxic chemical. The procedure proposed by Staniswalis and Cooper (1988) was implemented on real and simulated data. Muller and Schmitt (1988) introduced kernel regression for ungrouped data in which they derived the bias, variance, asymptotic normality, uniform consistency, and rates of convergence of the kernel estimates. Both Staniswalis and Cooper (1988) and Muller and Schmitt (1988) gave methods for finding $ED_{100\alpha}$ for $0 < \alpha < 1$.

Einsporn (1987) and Einsporn and Birch (1992) established a bridge between nonparametric kernel regression and the standard parametric approach called model-robust regression (MRR). In one application of Einsporn's (1987) method, the least squares "hat" matrix is mixed or linked with the weight matrix corresponding to nonparametric kernel regression. This specific application of MRR is termed "HATLINK." By using a mixing parameter, λ, the HATLINK prediction is a linear combination of parametric and nonparametric predictions. If the user's parametric model is "correct," then the procedure should place emphasis on the parametric predictions. On the other hand, if the user's model appears to be different from the "true" model, the procedure should place greater emphasis on the nonparametric predictions. That is, more weight will be given to the kernel prediction where predictions at a given location x_0 will be influenced more by observations in a small neighborhood around x_0 rather than observations far from x_0. The goals of MRR are to use the magnitude of the mixing parameter to determine if F has been misspecified, and to provide a model-robust method for making the usual regression inferences.

Other work combining nonparametric and parametric methods for data analysis, called semiparametric modeling, has been proposed by Olkin and Spiegelman (1987), Speckman (1988), and Severini and Staniswalis (1994). Olkin and Spiegelman (1987) suggests a linear combination of parametric and nonparametric methods in density estimation. Speckman (1988) introduces "partial linear regression" models in which the nonparametric and parametric portions of the model are additive. Severini and Staniswalis (1994) propose a semiparametric model using an additive function of parametric and nonparametric components, with estimates based on quasi-likelihood methods. Sutherland (1992) extended Einsporn's (1987) HATLINK procedure to response surface methodology and design of experiments.

The procedure considered in this research is an extension of the results of Einsporn (1987) and Einsporn and Birch (1993) to quantal dose-response data where the classical parametric analysis, such as logit analysis, is combined with a nonparametric procedure. Two nonparametric methods will be studied in this proposal: kernel regression, which was used in the work by Einsporn (1987) and Einsporn and Birch (1993), and local linear regression, a procedure first introduced by Cleveland (1979) and minimax properties developed by Fan (1993). Local linear regression has not been used to study quantal dose-response applications. The proposed method will be called model-robust quantal regression (MRQR) because it "robustifies" the quantal parametric model by adjusting the parametric predictions using a mixing parameter in the same manner suggested by Einsporn (1987). Also, another model-robust procedure will be studied in this research, termed "MRQR2" in which one of the nonparametric procedures in this research will be used to fit the residuals of the parametric fit and then add this residual fit to the parametric model using a mixing parameter between zero and one.

The chi-square goodness-of-fit test is the statistical yardstick used to quantify the measure of fit in probit and logit analysis. But the chi-square test often lacks power, especially when the model has been misspecified. The model-robust procedures introduced in this research, with the use of the mixing parameter, improves the fit by decreasing the chi-square statistic, and give improved estimates of the $ED_{100\alpha}$ with narrower confidence intervals. The magnitude of the mixing parameter may also confirm that the parametric procedure is appropriate (for small values of λ) or that the nonparametric procedure may provide a better estimate of the dose-response curve, indicated by large values of λ.

This dissertation is constructed as follows: **Chapter 2** reviews quantal bioassay and some more common cumulative distribution functions used in the analysis of quantal bioassay. **Chapter 3** gives a brief review of ordinary least squares and some of the nonparametric methods used in this research. **Chapter 4** applies the nonparametric regression methods introduced in Chapter 3 to estimating quantal dose-response curves along with effective dose. In **Chapter 5**, details of the MRR method are presented in a general context for linear models and detailed in **Chapter 6** for its use in the analysis of quantal bioassay,

the MRQR procedure. **Chapter 7** contains the properties of the model-robust procedures, but begins with the asymptotic properties of the parametric and nonparametric methods. In **Chapter 8**, some results of the Monte Carlo simulations are given for the procedures presented in this research. **Chapter 9** applies the model-robust quantal regression procedure to real and simulated data, and **Chapter 10** summarizes the conclusions and lists some future considerations.

CHAPTER 2
INTRODUCTION TO QUANTAL BIOASSAY

§2.1 Quantal Bioassay

In applications in the biological and health sciences where the response is binary (pass/fail, defective/nondefective, success/failure), one is interested in modeling the relationship between the response variable and a set of regressor variables. In the typical dose-response setting, there is only one regressor variable, the dose of the drug. Situations where doses are combinations of several drugs will not be considered in this research, but will be suggested for future research. It will be assumed that the data are grouped, in that for each dose, several subjects or experimental units are observed. The ungrouped case, where a single subject is tested at each dose, can be considered as a special case of the grouped data case.

Under the usual assumptions of a binomial trial, if the probability of response to dose x_i is P_i, then the probability of r_i responses in n_i trials is given by

$$\binom{n_i}{r_i} P_i^{r_i} (1-P_i)^{n_i - r_i}$$

The method of maximum likelihood will be used to estimate the unknown parameters. In the quantal dose-response setting, we have an experiment with d doses of a substance, and the i^{th} dose, x_i, is given to n_i individuals, of whom r_i respond. If P_i is the probability that any individual responds to the i^{th} dose, then the likelihood of $r_1, r_2, \ldots, r_d$ is the product of the binomial terms:

$$\prod_{i=1}^{d} \binom{n_i}{r_i} P_i^{r_i} (1-P_i)^{n_i - r_i}$$

The log likelihood is then

$$\sum_{i=1}^{d} r_i ln(P_i) + \sum_{i=1}^{d} (n_i - r_i) ln(1-P_i) + \sum_{i=1}^{d} ln\binom{n_i}{r_i} \tag{2.1.1}$$

The probabilities of response, P_i, can be modeled as

$$P_i = F(\beta_0 + \beta_1 x_i) \tag{2.1.2}$$

where F is the tolerance cumulative distribution function. If F is a member of the location-scale family of distributions, the unknown parameters β_0 and β_1 can be expressed as functions of the location and scale parameters. Several commonly used forms of F are the standard normal cdf, given by

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-u^2} du$$

and the logistic cdf given by

$$F(u) = \frac{1}{1+e^{-u}}$$

Other models frequently used in the analysis of quantal response data such as the Gompertz, Cauchy, and Sine curves can be found in Morgan (1992), Govindarajula (1988), and Cox and Snell (1989).

Given the model in (2.1.2), it is of primary interest to estimate β_0 and β_1, the unknown parameters. One possibility is to "linearize" the sample proportions by using the transformation

$$y_i = F^{-1}(p_i)$$

and the model

$$y_i = \beta_0 + \beta_1 x + \varepsilon_i$$

Then ordinary least squares may be used to estimate β_0 and β_1. However,

$$Var(y_i) = Var\left(F^{-1}(p_i)\right) \approx Var(p_i)\left(\frac{\partial F^{-1}(p_i)}{\partial p_i}\right)^2$$

and

$$Var(p_i) = \frac{P_i(1-P_i)}{n_i}$$

with P_i defined in (2.1.2). Thus, the y_i's are heteroscedastic and the usual method of ordinary least squares is no longer appropriate for estimating β_0 and β_1. The correct procedure for estimating β_0 and β_1 is to use the method of maximum likelihood.

The likelihood function in (2.1.1) can be viewed as a function of the two parameters β_0 and β_1, $L(\beta_0, \beta_1)$. Taking the partial derivative of (2.1.1) with respect to β_0 and with respect to β_1, and setting equal to zero yields the following equations:

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{d} \frac{n_i z_i^2}{P_i(1-P_i)} \frac{(p_i - P_i)}{z_i} = 0 \qquad (2.1.3)$$

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{d} \frac{n_i z_i^2 x_i}{P_i(1-P_i)} \frac{(p_i - P_i)}{z_i} = 0 \qquad (2.1.4)$$

where $z_i = f(\beta_0 + \beta_1 x_i)$ is the probability density function of F and p_i is the observed proportion responding at dose x_i. Notice that equations (2.1.3) and (2.1.4), called the maximum likelihood equations, are nonlinear and need to be solved to obtain maximum likelihood estimates (MLEs) of β_0 and β_1.

To solve equations (2.1.3) and (2.1.4), it is necessary to use numerical techniques such as the Gauss-Newton algorithm along with iterated re-weighted least squares (IRLS). This is done by rewriting equations (2.1.3) and (2.1.4) as

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{d} w_i \left(y_i^* - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right) = 0 \tag{2.1.5}$$

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{d} w_i x_i \left(y_i^* - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right) = 0 \tag{2.1.6}$$

respectively, where w_i is given by

$$w_i = \frac{n_i \hat{z}_i^2}{\hat{P}_i \hat{Q}_i}$$

where

$$\hat{P}_i = F\left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right),$$

$$\hat{Q}_i = 1 - \hat{P}_i,$$

$$\hat{z}_i = f\left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right),$$

and

$$y_i^* = \frac{p_i - \hat{P}_i}{\hat{z}_i} + \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Since equations (2.1.5) and (2.1.6) now appear linear in the unknown parameters, they can be solved to obtain estimates of β_0 and β_1 using the weighted least squares procedure. Unfortunately, w_i and y_i^* are also functions of the unknown parameters, and consequently, initial estimates of the parameters must be obtained and an iterative scheme must be used to solve for the MLEs. The IRLS algorithm is as follows:

1. Obtain initial estimates of the parameters. Usually the initial estimates are obtained by regressing y_i vs. x_i using ordinary least squares. Since $y_i = F^{-1}(p_i)$, it follows that, initially

$$y_i = log\left(\frac{p_i}{1 - p_i} \right)$$

or

$$y_i = \Phi^{-1}(p_i)$$

where Φ is the standard normal cdf, depending on whether F is assumed to be logistic or the normal cdf, respectively.

2. Compute the following statistics using the current estimates of β_0 and β_1:

$$\hat{y}_{i,0} = \hat{\beta}_{00} + \hat{\beta}_{10} x_i,$$

$$\hat{P}_{i,0} = F\left(\hat{y}_{i,0}\right),$$

$$\hat{Q}_{i,0} = 1 - \hat{P}_{i,0},$$

$$\hat{z}_{i,0} = f\left(\hat{\beta}_{00} + \hat{\beta}_{10} x_i\right) = f\left(\hat{y}_{i,0}\right),$$

$$w_i = \frac{n_i \hat{z}_{i,0}^2}{\hat{P}_{i,0}\hat{Q}_{i,0}}, \text{ and}$$

$$y_{i,0}^* = \hat{y}_{i,0} + \frac{p_i - \hat{P}_{i,0}}{\hat{z}_{i,0}}.$$

3. Using the estimates from Step 2, compute

$$\hat{\beta}_{11} = \frac{S_{xwy^*}}{S_{xwx}} \text{ and } \hat{\beta}_{01} = \bar{y}_w^* - \hat{\beta}_1 \bar{x}_w$$

where

$$\bar{x}_w = \frac{\sum_{i=1}^{d} w_i x_i}{\sum_{i=1}^{d} w_i}, \quad \bar{y}_w^* = \frac{\sum_{i=1}^{d} w_i y_i^*}{\sum_{i=1}^{d} w_i}, \quad S_{xwy^*} = \sum_{i=1}^{d} \left(x_i - \bar{x}_w\right) y_i^* w_i, \text{ and}$$

$$S_{xwx} = \sum_{i=1}^{d} \left(x_i - \bar{x}_w\right)^2 w_i$$

4. If $\hat{\beta}_{11} \approx \hat{\beta}_{10}$ and $\hat{\beta}_{01} \approx \hat{\beta}_{00}$ then stop. $\hat{\beta}_{11}$ and $\hat{\beta}_{01}$ are the MLEs, $\hat{\beta}_1$ and $\hat{\beta}_0$, respectively. Otherwise, replace $\hat{\beta}_{10}$ by $\hat{\beta}_{11}$ and $\hat{\beta}_{00}$ by $\hat{\beta}_{01}$ and return to Step 2.

At convergence, the maximum likelihood estimates have been obtained. In particular, we have

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{W}\mathbf{y}^* \tag{2.1.7}$$

where $\mathbf{y}^*$ is a $dx1$ vector of y_i^* computed in Step 2 and $\mathbf{W}$ is a dxd diagonal matrix of weights, with

$$w_i = \frac{n_i \hat{z}_i^2}{\hat{P}_i\left(1 - \hat{P}_i\right)}$$

being the i^{th} diagonal element of $\mathbf{W}$. Maximum likelihood theory states that the asymptotic variance of $\hat{\beta}$ is given by $I^{-1}\left(\hat{\beta}\right)$ where

$$I\left(\hat{\beta}\right) = -E\left(\frac{\partial^2 \ln L\left(\hat{\beta}\right)}{\partial\hat{\beta}\,\partial\hat{\beta}'}\right)$$

is the Fisher information matrix. $I\left(\hat{\beta}\right)$ is estimated by

$$\hat{I}\left(\hat{\beta}\right) = -\frac{\partial^2 \ln L\left(\hat{\beta}\right)}{\partial\hat{\beta}\,\partial\hat{\beta}'}.$$

It can be shown that the estimated variance of $\hat{\beta}$, $I^{-1}\left(\hat{\beta}\right)$, can be expressed as

$$(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{2.1.8}$$

Quite similar to classical linear regression, some goals of quantal bioassay are model building, hypothesis testing of model parameters and confidence intervals on parameters. The two elements of primary concern in the analysis of dose-response data is the test of lack-of-fit and the estimation of effective doses. The lack-of-fit test examines whether the proper F has been chosen. The most common test used to examine fit is the chi-square goodness-of-fit test (called the test of homogeneity by Finney (1978)). The goodness-of-fit test measures the departure of the observed data from the assumed model under the assumption that the proper F was chosen. The chi-square statistic is given by

$$\chi^2 = \sum_{i=1}^{d} \frac{n_i\left(p_i - \hat{P}_i\right)^2}{\hat{P}_i\left(1 - \hat{P}_i\right)} \tag{2.1.9}$$

which has an approximate chi-square distribution with $(d-2)$ degrees of freedom. If the lack-of-fit test is insignificant, it is assumed that the correct form of F has been specified and the analysis may proceed. If the lack-of-fit test is significant, another F should be used. When a proper F cannot be chosen, another option when one has significant lack-of-fit is to make a transformation on the dose. Eggar (1979) and Guerrero and Johnson (1982) have applied the family of Box-Cox transformations to the dose scale.

Once it has been determined that F is a good representation of the true dose-response relationship, confidence intervals may be obtained for P_i at each dose x_i by straightforward application of the delta method. That is,

$$\hat{P}_i = F\left(\mathbf{x}_i'\hat{\beta}\right)$$

where $\mathbf{x}_i' = (1\ x_i)$. Thus,

$$Var\left(\hat{P}_i\right) = \mathbf{x}_i'\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{x}_i f^2\left(\mathbf{x}_i'\hat{\beta}\right), \tag{2.1.10}$$

at each dose, x_i. However, our main concern is point and confidence interval estimation of the effective doses. The effective dose, denoted by $ED_{100\alpha}$, is the dose where, on average, $100\alpha\%$ of the subjects

respond to treatment. Unless an experiment has been designed to estimate extreme dose levels, the percentage of responses usually span the 50% value or the median, usually denoted by ED_{50}. Sometimes, more extreme dose values such as ED_{10} or ED_{90} are better suited in characterizing the data since many different cdfs could give quite similar estimates of the ED_{50}, but could give different estimates of the extreme dose levels.

In general, for any quantal assay model where $\hat{P}_i = F\left(\hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_i\right)$, it follows that $\hat{\mathbf{x}}_\alpha$ is the $ED_{100\alpha}$ satisfying $\alpha = F\left(\hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_\alpha\right)$. Then $\hat{\mathbf{x}}_\alpha$ may be readily obtained as

$$\hat{x}_\alpha = \frac{F^{-1}(\alpha) - \hat{\beta}_0}{\hat{\beta}_1} \tag{2.1.11}$$

For the logistic model, if $\alpha = F\left(\hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_\alpha\right)$, it follows that

$$F^{-1}(\alpha) = log\left(\frac{\alpha}{1-\alpha}\right) = logit(\alpha) = \hat{\beta}_0 + \hat{\beta}_1 x_\alpha$$

where $\hat{\mathbf{x}}_\alpha$ is the $ED_{100\alpha}$ value. So,

$$\hat{x}_\alpha = \frac{1}{\hat{\beta}_1}\left\{log\left(\frac{\alpha}{1-\alpha}\right) - \hat{\beta}_0\right\} \tag{2.1.12}$$

In particular,

$$ED_{50} = \frac{-\hat{\beta}_0}{\hat{\beta}_1} \tag{2.1.13}$$

The three most common methods for estimating confidence intervals on the effective dose are the delta method, likelihood-ratio interval, and Fieller's Theorem. For more information on the delta method and the likelihood-ratio interval see Morgan (1992). Fieller's theorem will be used to compute confidence intervals in this research. Actually, Fieller's theorem results in a "fiducial interval" for the parameter of interest. In a manner similar to Finney (1978) we will consider confidence intervals and fiducial intervals as equivalent. Note that $\hat{\mathbf{x}}_\alpha$ in (2.1.11) is a ratio of estimates of parameters.

To motivate Fieller's theorem, consider any ratio of the form $\mu = \frac{\alpha}{\beta}$, where μ, α, and β are parameters. Let a, b, and m be estimates of α, β, and μ, respectively. Assume that $(a, b)'$ follows a bivariate normal distribution with mean $(\alpha, \beta)'$ and variance-covariance matrix Σ given by

$$\Sigma = \sigma^2 \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

where V_{11}, V_{22}, V_{12}, *and* V_{21} are appropriate constants. Since both a and b are normal, the distribution of $(a-\mu b)$ is

$$(a-\mu b)\sim N\left(0,\ \sigma^2\left(V_{11}-2\mu V_{12}+\mu^2 V_{22}\right)\right)$$

Thus, a confidence interval for μ with confidence level of (1-γ)x100% is given by the set of values of μ satisfying

$$(a-\mu b)\le t^2\hat{V}^2 \tag{2.1.14}$$

where

$$t=t_{1-\frac{\gamma}{2}}(f)$$

is the $\left(1-\frac{\gamma}{2}\right)$ percent point of the Student's t distribution, with f being the degrees of freedom associated with the appropriate estimate of σ^2, s^2, and

$$\hat{V}=s\left(V_{11}-2\mu V_{12}+\mu^2 V_{22}\right)$$

Using the quadratic equation of (2.1.14) and solving for μ, one obtains the (1-γ)x100% confidence limits of μ as

$$\left(\frac{1}{1-g}\right)\left(m-g\frac{V_{12}}{V_{22}}\right)\pm\ \frac{ts}{b}\left(V_{11}-2mV_{12}+m^2V_{22}-g\left(V_{11}-\frac{V_{12}^2}{V_{22}}\right)\right)^{\frac{1}{2}} \tag{2.1.15}$$

where g, called the "index of significance of the difference of b from zero" is defined by

$$g=\frac{t^2s^2V_{22}}{b^2}$$

Usually s^2 is the mean squared error (MSE) from the analysis of variance (ANOVA) table. In the quantal dose-response setting, if we have significant lack-of-fit, s^2 is set equal to the "heterogeneity factor," $\chi^2/(d-2)$, where χ^2 is the goodness-of-fit test statistic. If the lack-of-fit is insignificant, then s^2 is set equal to 1.

In our quantal dose-response setting, replacing m by

$$\frac{F^{-1}(\alpha)-\hat{\beta}_0}{\hat{\beta}_1}$$

in (2.1.15), and using the estimated variance for $\hat{\beta}_0$ and $\hat{\beta}_1$, obtained from $s^2\ (\mathbf{X'WX})^{-1}$, the Fieller interval for the $ED_{100\alpha}$ value can be computed. For a more complete description of Fieller's theorem, see Finney (1978).

Another statistic used in the analysis of dose-response data will be the prediction error sums of squares, the *PRESS* statistic. In ordinary least squares, the *PRESS* statistic is given by

$$PRESS = \sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}\right)^2 = \sum_{i=1}^{n} e_{i,-i}^2$$

where

$$e_{i,-i} = \frac{e_i}{1-h_{ii}}$$

is the residual at the i^{th} data point with the i^{th} observation removed. For a development of the *PRESS* statistic in OLS, see Myers (1990).

A "minus-i" statistic has yet to be developed for nonlinear models in the literature. Because the *PRESS* statistic is used a great deal to compare procedures and parameter selection in the nonparametric methods presented later, a "minus-i" statistic is developed in this research (see **Appendix A**). It has been shown that through the use of the Sherman-Morrison-Woodbury theorem, an approximation to the "minus-i" predicted values in logistic regression is

$$\hat{P}_{i,-i} = \hat{P}_{i,-i}^{Logistic} \approx \frac{1}{1+e^{-\mathbf{x}_i'\mathbf{b}}e^{h_{ii}^L e_i/\left(1-h_{ii}^L\right)}} \tag{2.1.16}$$

where h_{ii}^L is the i^{th} diagonal element of $\mathbf{H}^L = \mathbf{X}\left(\mathbf{X}'\mathbf{W}^*\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^*$, the "hat-like" matrix used in weighted least squares for the one-step estimation of the unknown coefficients of the logit model, and $e_i = logit\left(p_i\right) - \mathbf{x}_i'\mathbf{b}$. Because the logit transformation is used to linearize the logistic model, $\mathbf{W}^*$ is the diagonal weight matrix with the i^{th} diagonal element as $w_i^* = n_i p_i\left(1-p_i\right)$. Given the "minus-i" predicted values, one can compute the *PRESS* statistics for logistic regression. Since the homogeneous variance assumption is not valid with quantal dose-response data, the i^{th} *PRESS* residual will be weighted by the reciprocal of the variance of the i^{th} observed response such that another form for the *PRESS* statistic for logistic regression is

$$PRESS = \sum_{i=1}^{d}\frac{n_i\left(p_i - \hat{P}_{i,-i}\right)^2}{p_i\left(1-p_i\right)} \tag{2.1.17}$$

This weighting scheme on the *PRESS*-type statistics will be used throughout this research for uniformity and consistency.

§2.2 Examples

In this section examples will be presented illustrating the application of logit analysis on quantal dose-response data. In the analysis, the log transformation on the dose scale will be used, and the method of maximum likelihood is used to obtain parameter estimates. Confidence intervals for proportion responding will be obtained using the delta method and Fieller's theorem will be used to compute the confidence intervals on the effective doses.

§2.2.A Example 1

The following data set , taken from Huber (1994), in **Table 2.2.1** is an example of a single agent quantal bioassay of a typical toxicity experiment. The results show the effect of different doses of nicotine on the common fruit fly. Notice, that this is an indirect assay because the exact concentration of nicotine required to kill an insect cannot be observed by the experimenter and the response (the mortality rate) is being observed. That is, if an insect dies at a given concentration it is known that the tolerance level of that insect was exceeded by that concentration.

Table 2.2.1 Toxicity Data on Common Fruit Fly

Concentration (g/100 cc)	Number of Insects	Number Killed	Percent Killed
0.10	47	8	17.0
0.15	53	14	26.4
0.20	55	24	43.6
0.30	52	32	61.5
0.50	46	38	82.6
0.70	54	50	92.6
0.95	52	50	96.2

As the dose of the nicotine increases, the proportion of flies responding (killed) also increases. The logistic model is given by

$$P_i = F\left(\beta_0 + \beta_1 log(dose)\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 log(dose))}}$$

where P_i is the proportion of flies responding to the nicotine concentration, β_0 and β_1 being the unknown coefficients. Using IRLS to obtain the maximum likelihood estimates, the estimated logistic curve for the fly data is

$$\hat{P}_i = \frac{1}{1+e^{-(3.12+4.899\,log(dose))}}$$

Of primary interest is the chi-square lack-of-fit test given by (2.1.9). For this data set, χ^2 is 0.7351, with 5 degrees of freedom, which is significant at the 0.9810 level. Thus, the logistic function adequately fits the fly data set. Assuming that the best possible model has been fit to the data and there is no significant lack-of-fit, estimates of the ED_{50} and a confidence interval on the ED_{50} are in order. Using Fieller's theorem, $\hat{x}_{50}$ = 0.2304 g/100cc with a 95% confidence interval of (0.2019 g/100cc, 0.2610 g/100cc), yielding a width of 0.0590 g/100 cc. **Figure 2.2.1** contains a scatter of the raw data with the logistic fit accompanied by a 95% confidence band on the dose-response curve. Note that the logistic model does an excellent job in fitting these data, which is indicated in the small size of the chi-square statistic as well as the width of the confidence interval on the ED_{50}. Extreme dose levels of interest in this research will be the ED_{20} and ED_{80} values. For this particular data set, the ED_{20} = 0.1201 g/100cc with a 95% confidence interval of (0.0951 g/100cc, 0.1423 g/100cc). The estimate that yields a 80% response is 0.4420 g/100cc with a 95% confidence interval on the ED_{80} of (0.3802 g/100cc, 0.5398 g/100cc).

The confidence interval for the ED_{50}, or any other α% effective dose, can be graphically represented on the dose-response curve by using "inverse" estimation. That is, a horizontal line is extended from 0.5, or α%, to intersect with the dose-response curve and the upper and lower confidence interval curves. Three vertical lines are extended downward toward the abscissa from these intersections until they intersect with the "dose axis." The lower confidence limit for the ED_{50} occurs where the vertical line from the upper confidence level intersects with the dose axis; the ED_{50} is where the vertical line from the dose-response curve intersects with the dose axis, and finally, the upper confidence limit for the ED_{50} occurs where the vertical line from the lower confidence level intersects with the dose axis. An illustration of inverse regression is given in **Figure 2.2.1** in which LCL and UCL denote the lower and upper confidence limits on the ED_{50}, respectively, and ED50 represents the median effective dose. The inverse regression method for estimating $ED_{100\alpha}$ values and confidence limits on effective doses will also be used in the nonparametric procedures introduced later in this proposal.

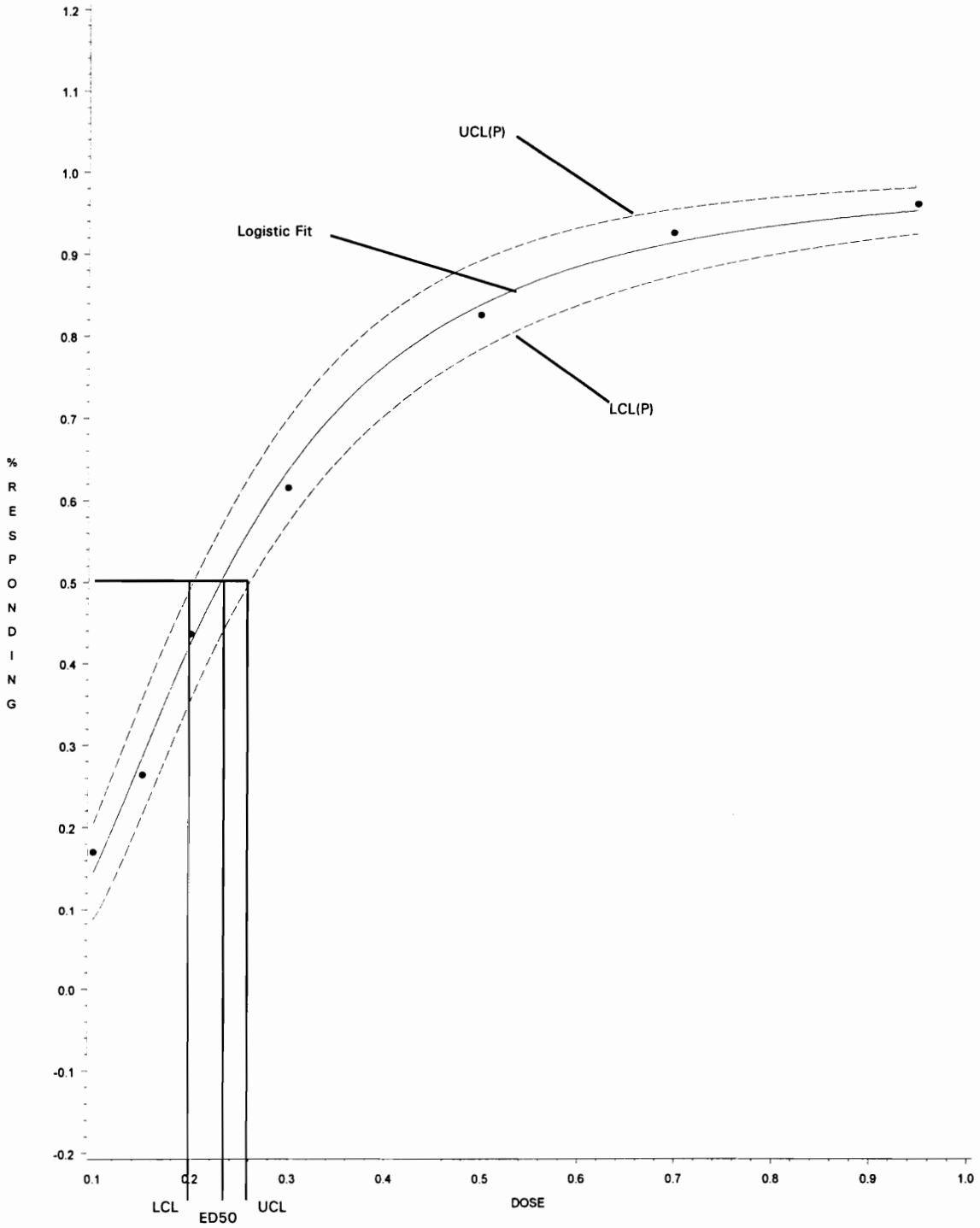


Figure 2.2.1 Logistic Fit to Fly Dose-response Data

• • • Raw data, ___ Logistic fit, - - - 95% Confidence band

§2.2.B Example 2

The data in this example has been created from a mixture of logistic cumulative distribution functions of the form

$$\mathbf{P}(x) = (1-\gamma)L(0.5,\ 0.1) + \gamma\left[\delta L(0.25,\ 0.05) + (1-\delta)L(0.75,\ 0.05)\right] \qquad (2.2.B.1)$$

where $L(\mu, \tau)$ is the logistic cdf written as

$$L(\mu, \tau) = \frac{1}{1 + \mathbf{e}^{-\left(\frac{x-\mu}{\tau}\right)}}$$

with μ and τ being the location and scale parameters of the logistic distribution, respectively. The value of δ will be 0.5 throughout this research. The value of γ, which will denote the degree of model misspecification, will range from zero to one. If γ=0, then the model will be a logistic cdf with location and scale of 0.5 and 0.1, respectively, and the user's model will be correct. As the value of γ increases to one, the degree of model-misspecification also increases. **Figure 2.2.2** is a plot of equation (2.2.B.1) for γ=0, the logistic cdf. In this example, data will be randomly generated for the case in which there are five equally spaced doses of the drug, ranging from 0.1 to 0.9, with twenty subjects receiving each dose, and γ=0.5. **Figure 2.2.2** is a plot of the curve with γ=0.5 overlaid on the γ=0 curve. The data generated for this example is given in **Table 2.2.2**.

Table 2.2.2 Logit Mixture Data

Dose	Number of Subjects	Number Responding
0.1	20	1
0.3	20	5
0.5	20	11
0.7	20	15
0.9	20	19

Using the logit mixture data, the logistic regression model will be evaluated with respect to its performance in estimating the effective doses. Fitting the logistic model to the data yields

$$\hat{P}_i = \frac{1}{1 + e^{-(-3.20 + 6.54 dose)}}$$

The value of the chi-square statistic is 0.6682, which is significant at the 0.8806 level, indicating that there is no lack-of-fit present for this data when using the logistic model. Hence, the lack-of-fit test fails to detect that the logistic model is incorrect here, a type II error. The estimate of the ED_{50} is 0.4896 with a 95% confidence interval of (0.4035, 0.5742). For this data, the true value of the ED_{50} is 0.5, which is

covered by the 95% confidence interval. The estimate of the ED_{20} is 0.2771 with a 95% confidence interval of (0.1197, 0.3685) yielding a width of 0.2588. The 80% effective dose is 0.7014, with a 95% confidence interval of (0.6108, 0.8564) having a width of 0.2456. The true values of the ED_{20} and ED_{80} are covered by their 95% confidence intervals. The true ED_{20} value is 0.2730 and the true ED_{80} value is 0.7270. A scatter of the data of **Table 2.2.2** and the estimated logistic curve accompanied with 95% confidence bands can be seen in **Figure 2.2.3**.

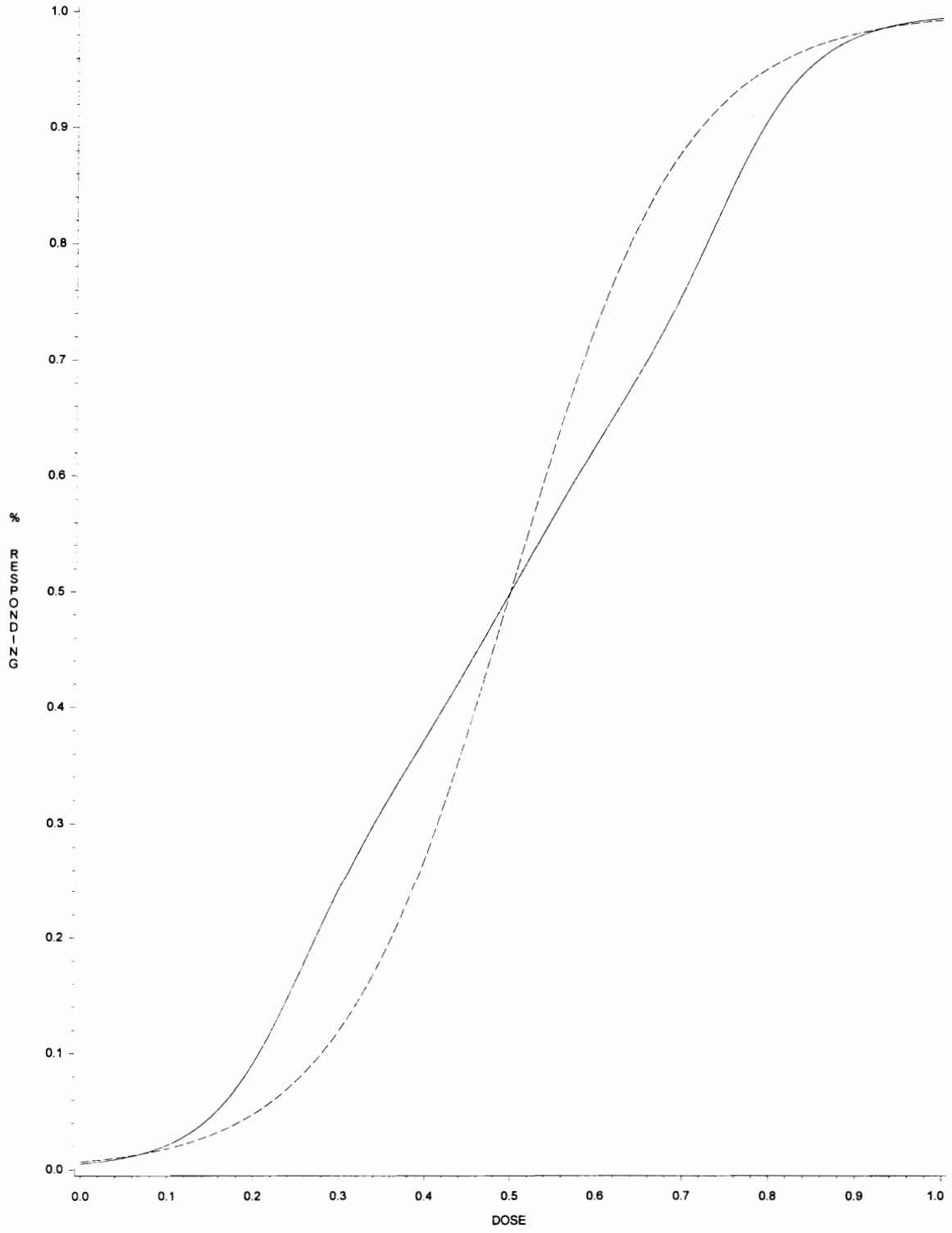


Figure 2.2.2 Plot of Logit Mixture Model for $\gamma = 0$ (dashed) and $\gamma = 0.5$ (solid).

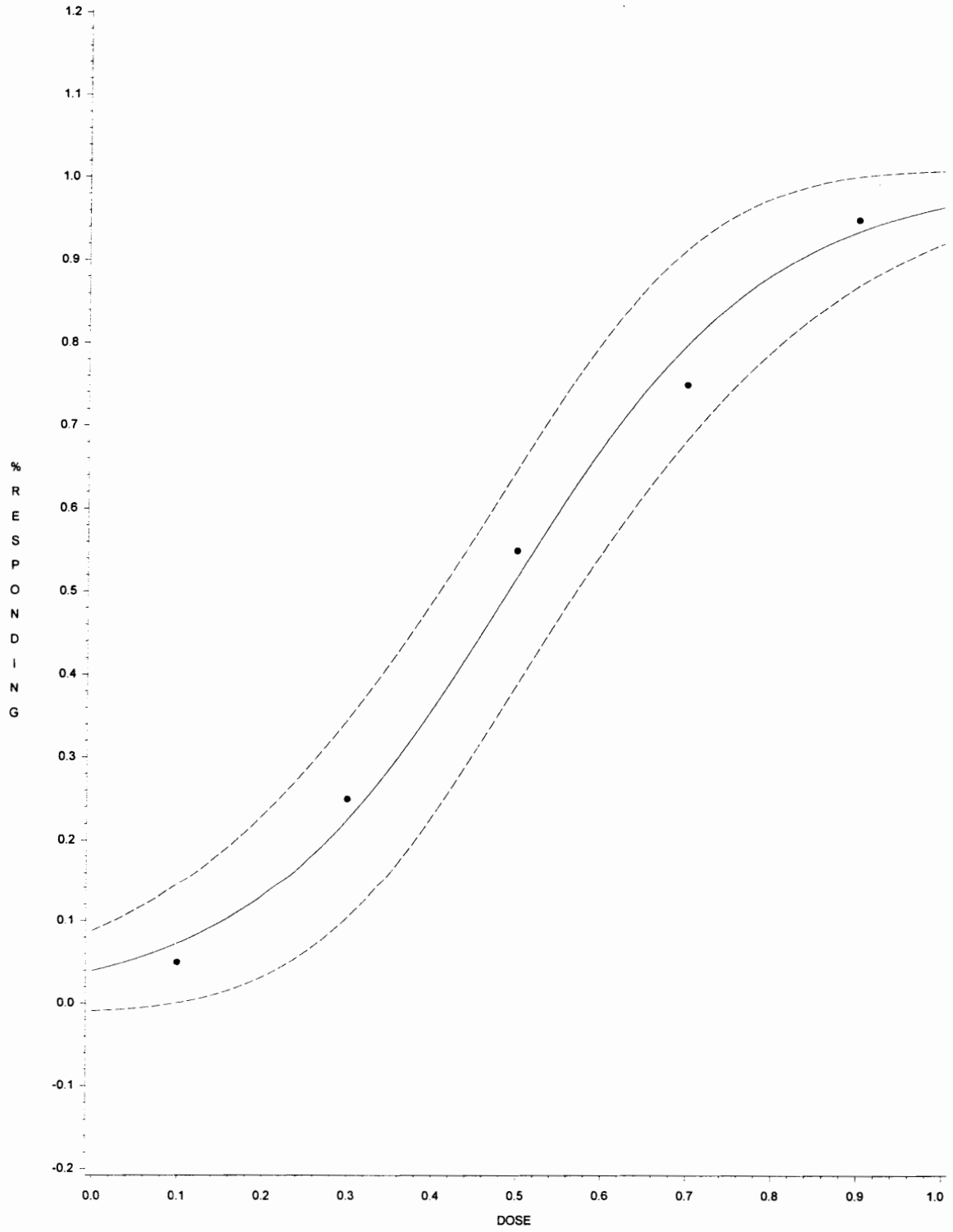


Figure 2.2.3 Logistic fit to the Logit Mixture Data from Equation 2.2.B.1

. . . Raw data, ___ Logistic fit, - - - 95% Confidence band

CHAPTER 3

ORDINARY LEAST SQUARES AND SOME NONPARAMETRIC METHODS

Regression curves describe the relationship between the k regressor variables, $x_1, x_2, \ldots, x_k$, and the response variable y. The regression relationship is modeled as $y_i = f(x_1, x_2, \ldots, x_k) + \varepsilon_i$, i =1,2,...,n, where f is the unknown regression function and ε_i are the random observation errors. The goal in regression is to approximate f, the unknown function. If the parametric approach is used, the user assumes f has some prespecified functional form. An alternative to the parametric approach is nonparametric regression which attempts to estimate f without reference to any specific form or distribution of the errors. In the following sections, a brief introduction to ordinary least squares is given, along with some of its properties, followed by a discussion of some nonparametric regression methods.

§3.1 Ordinary Least Squares

Consider the model

$$y_i = f(x_{1i}, x_{2i}, \ldots, x_{ki}) + \varepsilon_i, \; i = 1,2,\ldots,n; \; n \geq k+1$$

where ε_i are independent and identically distributed with mean zero and constant variance σ^2. The x_j, $j = 1,2,\ldots,k$, represent the regressor variables which are assumed to be non-stochastic and measured with negligible error. In linear regression it is assumed that the correct form of f can be specified by the user as a model linear in the unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$ as

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i \tag{3.1.1}$$

The vector form of equation (3.1.1) can be written as

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{3.1.2}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

is the nx1 vector of observed responses,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{k1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{k2} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & x_{1n} & x_{2n} & . & . & . & x_{kn} \end{bmatrix} = \begin{bmatrix} 1 \; \mathbf{x}_1 \; \mathbf{x}_2 \; \ldots \; \mathbf{x}_k \end{bmatrix}$$

is the model matrix with 1 being the nx1 vector of ones, the i^{th} row of the $\mathbf{X}$ matrix is written as

$$\mathbf{x}_i' = \begin{bmatrix} 1 \; x_{1i} \; \ldots \; x_{ki} \end{bmatrix}$$

and

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

is the px1 vector of unknown coefficients, where $p = k+1$. The estimate of β, obtained by the method of least squares, is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.1.4}$$

Estimation of the mean response at data locations are obtained as

$$\hat{\mathbf{y}} = \mathbf{X}\beta = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \tag{3.1.5}$$

where

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} = [h_{ij}] \tag{3.1.6}$$

is commonly called the "hat" matrix or the projection matrix because it projects the observed responses, $\mathbf{y}$, to the predicted values, $\hat{\mathbf{y}}$. The hat matrix plays a very important role in general linear models theory as stressed by Hoaglin and Welch (1978), who presented many of its mathematical properties. Some of these properties are as follows:

1. $\mathbf{H}$ is symmetric idempotent
2. $-1 \le h_{ij} \le 1$
3. $\sum_{i=1}^{n} h_{ii} = p$, where $p = k + 1$ is the number of parameters to be estimated
4. $\sum_{i=1}^{n} h_{ij} = 1$, for each i.

By some straightforward manipulation, a number of results in general linear models theory can be expressed in terms of the hat matrix or its elements. Several such results are given below:

1. $Var(\hat{\mathbf{y}}) = Var(\mathbf{H}\mathbf{y}) = \sigma^2\mathbf{H}$
2. $Var(\hat{\mathbf{y}}_i) = \sigma^2[\mathbf{H}]_{ii} = \sigma^2 h_{ii}$
3. $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta} = [\mathbf{I} - \mathbf{H}]\mathbf{y}$
4. $Var(\mathbf{e}) = \sigma^2[\mathbf{I} - \mathbf{H}]$
5. $Var(e_i) = \sigma^2(1 - h_{ii})$
6. $\hat{\sigma}^2 = \dfrac{\sum e_i^2}{n-p} = \dfrac{\sum e_i^2}{tr[(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H})']}$

where $\mathbf{I}$ represents the nxn identity matrix and $\mathbf{e}$ is the nx1 vector of residuals in the above expressions. It is important to note that equation (3.1.5) reduces to

$$\hat{y}_i = \sum_{j=1}^{n} h_{ij} y_j \tag{3.1.7}$$

for predicting y at the particular data location $\mathbf{x}_i'$. From equation (3.1.7), it can be seen that the prediction at data location $\mathbf{x}_i'$ is a weighted average of the observations y_j, with the weights given by the i^{th} row of the hat matrix. An observation $(y_i, \mathbf{x}_i')$ with a large h_{ii} would have large influence on the prediction $\hat{\mathbf{y}}_i$.

In such a case, the i^{th} set of regressors, $\mathbf{x}_i'$, is called a "high leverage point" and h_{ii} is called the "leverage value." The leverage value measures the "standardized" distance of $\mathbf{x}_i'$ from the center of the x data.

For example, in simple linear regression (SLR) with $k = 1$, the hat matrix can be expressed as

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}$$

where

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

In SLR, an observation x_i is said to be a high leverage point if its distance from the center of the data is remote when compared to the rest of the x's; in other words, h_{ii} is large. If the data point x_j is far removed from x_i then h_{ij} will be large (in magnitude) and the response y_j will receive a large "weight" in the estimation of $\hat{y}_i$. Conversely, observations taken close to x_i receive less weight in predicting $\hat{y}_i$ since the corresponding h_{ij}'s can be approximately zero. This weighting scheme used for making predictions by the method of least squares is heavily dependent upon the user's choice of the parametric model.

If the user's model is correct over the entire range of the observed x's, then the method of least squares is desirable and the weighting scheme which assigns larger weights to far distance points is appropriate. But when the user's model differs from the true model, then the least squares weighting scheme could produce poor predictions at many x locations. In this situation, when predicting at an arbitrary point, x_0, it is no longer sensible to give more weight to observations y_j taken at x locations far from x_0. Rather, giving more weight to observations closer to x_0 seems more logical.

This idea of giving emphasis to points closer to x_0 is accomplished by nonparametric regression methods, such as kernel and local linear regression (LLR). The nonparametric regression model assumes that f is entirely unknown. If f is well-behaved, then observations y_j at locations nearest x_0 should be given more weight than observations taken at locations far from x_0. This weighting scheme weights the y_j's according to some decreasing function of their distance from x_0. This is the weighting scheme used in kernel regression and LLR which will be discussed in the following sections.

§3.2 Kernel Regression

Consider the usual regression setting with response variable y_i and one regressor x. In the nonparametric regression model setting, the response is modeled as

$$y_i = f(x_i) + \varepsilon_i,\ i = 1,2,\ldots,n \tag{3.2.1}$$

for some arbitrary smooth function f. Kernel regression or kernel smoothing is a method of approximating f using a weighting sequence on the response variable where the weights are functions of the distances between the values of the regressor variable x. One form of the kernel weighting sequence was proposed by Nadaraya (1964) and Watson (1964) and assigns weights of $w_j(x)$ to y_j when estimating the response at x by

$$w_j(x) = \frac{K\left(\frac{x - x_j}{b}\right)}{\sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right)} \tag{3.2.2}$$

where $K(\ )$ represents the kernel function and b is the parameter of the kernel function called the bandwidth. The form of K is assumed to be continuous, bounded, and a symmetric real function which integrates to one. Then, the kernel estimate of response at $x = x_i$, $\hat{y}^k(x_i)$, is given by

$$\hat{y}^k(x_i) = \sum_{j=1}^{n} w_{ij} y_j \tag{3.2.3}$$

Just as in ordinary least squares, where the hat matrix projects the observed responses to the predicted ones, the kernel weight matrix serves the same role. Equation (3.2.3) can also be expressed as

$$\hat{y}^k(x_i) = \sum_{j=1}^{n} h_{ij}^k y_j \tag{3.2.4}$$

where h_{ij}^k are the kernel weights on the observations y_j for estimating the mean response at location x_i using the kernel function $K(\ )$, with h_{ij}^k given by

$$h_{ij}^k = \frac{K\left(\frac{x_i - x_j}{b}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x_j}{b}\right)} \tag{3.2.5}$$

Another useful expression of (3.2.4) is

$$\hat{\mathbf{y}}^k = \mathbf{H}^k \mathbf{y} \tag{3.2.6}$$

where $\hat{\mathbf{y}}^k$ is the nx1 vector of kernel predictions, $\mathbf{H}^k$ is the nxn matrix of weights, h_{ij}^k. The matrix $\mathbf{H}^k$, called the kernel hat matrix, kernel weight matrix, or smoother matrix, plays a role in kernel regression

similar to that of $\mathbf{H}$ in linear regression. Note that unlike its OLS counterpart, the kernel hat matrix is neither symmetric nor idempotent. As an example of using $\mathbf{H}^k$, one can express the variance of $\hat{\mathbf{y}}$ as

$$Var\left(\hat{\mathbf{y}}^k\right) = \sigma^2 \mathbf{H}^k \mathbf{H}^{k'} \tag{3.2.7}$$

an exact expression if the bandwidth b is non-stochastic. If one is interested in predicting at a non-data point x_0, then the kernel prediction may be obtained by

$$\hat{y}^k(x_0) = \hat{f}(x_0) = \sum_{i=1}^{n} h_{0j}^k y_j \tag{3.2.8}$$

where

$$h_{0j}^k = \frac{K\left(\frac{x_0 - x_j}{b}\right)}{\sum_{j=1}^{n} K\left(\frac{x_0 - x_j}{b}\right)} \tag{3.2.9}$$

The main disadvantage to using kernel regression is that it is not possible to obtain a closed functional form of $\hat{f}$. The lack of a closed functional form of $\hat{f}$ also makes applying kernel regression, along with many other nonparametric methods, computationally intensive.

As can be seen in equation (3.2.2), the kernel estimate of f depends on the choice of the kernel function $K(\ )$ and the bandwidth b. There are a wide variety of kernel functions available in the literature (see Härdle (1990)) and Chu and Marron (1991). It is recommended by Chu and Marron (1991) that the kernel function be nonnegative, symmetric about zero, and be continuous with at least two derivatives. Some of the more commonly used kernels are the quartic kernel, Gaussian kernel, and the Epanechnikov kernel. Copas (1983) found that in practice, the choice of kernel has little effect on the resulting estimate of f, so for convenience the Gaussian kernel will be used throughout this research, where

$$K(u) = e^{-u^2}.$$

Though the choice of kernel may not be crucial in estimating f, the method for selecting the bandwidth is extremely critical. The magnitude of the bandwidth determines the smoothness, or lack thereof, of the regression function. The kernel fit $\hat{y}^k(x_i)$ may be regarded as a weighted average of the observed y_i's. If b is very small, then the majority of the weight is given to the y_j's close to x_i, resulting in $\hat{y}^k(x_i) = \hat{f}(x_i) \approx y_i$. Similarly, if b is very large, then the weights are spread evenly throughout all the n points resulting in $\hat{y}^k(x_i) \approx \bar{y}$. Therefore, small values of b tend to give $\hat{f}$ that simply reproduces the data, often resulting in a curve with considerable oscillation. Large values of b tend to give an over-smooth regression curve $\hat{f}$ such that the fit is merely the average of the responses. The precision of the

kernel estimates of mean response then are also dependent on b with the $Var\left(\hat{y}^k(x_i)\right)$ increasing as b tends toward zero.

One of the more common methods is to select the bandwidth using the data. Such methods are called data-driven selectors or automatic bandwidth selectors. A plethora of methods exist in the literature for data-driven bandwidth selectors, see, for example, Chapter 5 of Härdle (1990) for a description and summary of several of these methods, some of which are discussed below.

The method of cross-validation, also called the "leave-one-out" method, is analogous to the *PRESS* procedure introduced by Allen (1974) and is one selection method used in this proposal. The bandwidth is found by minimizing

$$PRESS(b) = \sum_{i=1}^{n}\left(y_i - \hat{y}^k_{i,-i}(b)\right)^2 \tag{3.2.10}$$

with respect to b, where $\hat{y}^k_{i,-i}(b)$ denotes the predicted value of y for the current value of b, where the prediction is made without the i^{th} observation. The "minus-i" prediction, $\hat{y}^k_{i,-i}(b)$, does not require performing n different regressions, each with the i^{th} observation deleted. The "minus-i" predictions can be computed by

$$\hat{y}^k_{i,-i}(b) = \sum_{i \neq j} h^k_{ij,-i} y_j \tag{3.2.11}$$

where

$$h^k_{ij,-i} = \frac{h^h_{ij}}{1 - h^k_{ii}} \tag{3.2.12}$$

Another procedure, quite similar to $PRESS(b)$, is one in which the bandwidth is found using a penalized $PRESS(b)$ statistic, denoted $PRESS^*(b)$, that finds the value of b minimizing

$$PRESS^*(b) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}^k_{i,-i}(b)\right)^2}{n - tr\left[\mathbf{H}^k\right]} \tag{3.2.13}$$

where n is the sample size and $tr\left[\mathbf{H}^k\right]$ is the trace of the nxn kernel weight matrix $\mathbf{H}^k$. $PRESS(b)$ is similar to the method of generalized cross-validation (GCV), first introduced by Wahba and Wold (1975).

Other methods employed for data-driven bandwidth selection are based on finding the value of the bandwidth that optimize quadratic error measures for the regression curve. One form of quadratic error is the mean squared error (MSE) given by

$$MSE(b) = E\left[\hat{f}(x) - f(x)\right]^2$$

which is the sum of the variance and a squared bias term. That is,

$$MSE(b) = Var\left(\hat{f}(x)\right) + bias^2\left(\hat{f}(x)\right)$$

Other forms of quadratic error measures are the mean integrated squared error (MISE) given by

$$MISE(b) = E\int\left[\hat{f}(x) - f(x)\right]^2 w(x)dx$$

and the average squared error (ASE), expressed as

$$ASE(b) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{f}(x_i) - f(x_i)\right]^2 w(x_i)$$

where $w(x_i)$ is a nonnegative weight function. Because the distances above are the sums of the variance and the squared bias, the data driven bandwidth selectors seek to find the optimal smoothing parameter b that balances the variance and the bias. Härdle (1990) devotes an entire chapter to choosing the optimal bandwidth.

One of the more paramount problems in kernel regression is the selection of the smoothing parameter, as well as how one automatically chosen bandwidth compares with another. A reason why so many bandwidth selection methods exist is due to the purpose of the smooth. As stated in Härdle (1990), "If the purpose of smoothing is to increase the "signal-to-noise ratio" for presentation, or to suggest a simple (parametric) model, then a slightly "oversmoothed" curve with a subjectively chosen smoothing parameter might be desirable. On the other hand, when the interest is purely in estimating the regression curve itself with an emphasis on local structures then a slightly "undersmoothed" curve may be appropriate." Therefore, since the goals differ in a variety of platforms, there isn't an outright "winner" of one bandwidth selection method over another. As stated earlier, the data-driven bandwidth selectors are chosen to balance variance and bias, sometimes resulting in an oversmoothed curve and other times resulting in an undersmoothed curve.

§3.3 Local Linear Regression

In preliminary studies, use of kernel regression usually gives reasonable estimates of f in the interior dose region, but its biggest drawback has been bias at the extreme doses caused by the asymmetry of observations at the boundaries. Rice (1984) suggested using a modified kernel in the boundaries to remedy the bias problem, which would be different from the kernel used in the interior region. Another alternative to using different kernels in the boundaries is the local linear regression (LLR) estimate, first proposed by Cleveland (1979). Quite similar to kernel regression, the model for local linear regression is $y_i = f(x_i) + \varepsilon_i$, where f is assumed to be an unknown smooth function with ε_i independent with mean zero. In LLR, for each point x, $f(x)$ is estimated using simple linear regression through the weighted least squares method, with weights assigned to the observations using some nonparametric method, such as the kernel weight matrix. Kernel regression may be thought of as a special case of LLR, where in kernel regression the fit at each point x is just the weighted average of the n responses. Thus, kernel regression estimates mean response using a location model at each data point x and obtains the fits using weighted least squares. In LLR, mean response is estimated using a simple linear regression model and the fits are obtained using weighted least squares. The weights are the same as those used in kernel regression.

Under the assumption of homogeneous variance, the local linear regression estimate at each observation x_i is found by minimizing

$$\sum_{j=1}^{n}\left(y_j - \beta_0 - \beta_1\left(x_i - x_j\right)\right)^2 K\left(\frac{x_i - x_j}{b}\right) \tag{3.3.1}$$

where $K(\ )$ is the kernel function and b is the bandwidth. In matrix notation, (3.3.1) can be written as

$$L(x_i) = \left(\mathbf{y} - \mathbf{X}\beta_i\right)' \mathbf{H}_i^k \left(\mathbf{y} - \mathbf{X}\beta_i\right) \tag{3.3.2}$$

where $\mathbf{H}_i^k$ represents the diagonal matrix formed from the i^{th} row of the kernel weight matrix $\mathbf{H}^k$ and β_i is the unknown coefficient vector for the i^{th} data point; that is, $\mathbf{H}^k$ can be written as

$$\mathbf{H}^k = \begin{bmatrix} \mathbf{h}_1^{k'} \\ \mathbf{h}_2^{k'} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{h}_n^{k'} \end{bmatrix}$$

and

$$\mathbf{H}_i^k = \begin{bmatrix} h_{i1}^k & & & \\ & h_{i2}^k & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & h_{in}^k \end{bmatrix} \tag{3.3.3}$$

Minimizing (3.3.2) yields

$$\hat{\beta}_i = \left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_i^k\mathbf{y} \tag{3.3.4}$$

and thus

$$\hat{y}^{LLR}(x_i) = \mathbf{x}_i'\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_i^k\mathbf{y} \tag{3.3.5}$$

is the LLR estimate at dose x_i. So, in LLR, one fits a simple linear regression model at each $\mathbf{x}_i$, yielding n sets of

$$\hat{\beta}_i = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \end{bmatrix}, i=1,2,...,n$$

Now, the variance properties of the LLR estimate are obtained just as in ordinary least squares. So, under the assumption of homogeneous variance and a non-stochastic bandwidth, this research suggests

$$Var\left(\hat{\mathbf{y}}^{LLR}(\mathbf{x}_i)\right) = \sigma^2\mathbf{x}_i'\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_i^k\mathbf{H}_i^{k'}\mathbf{X}\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{x}_i \tag{3.3.6}$$

Fan (1993) and Fan and Gijbels (1992) studied LLR and established some asymptotic minimax efficiency properties, as well as the fact that LLR for curve estimation does not have the boundary effects, and hence does not require modifications at the boundary as suggested by Rice (1984) for kernel regression. Just as one can fit LLR with $\mathbf{x}_i' = (1,\ x_i)$, it is just as easy to fit local quadratic or local cubic regressions by modifying $\mathbf{x}_i'$ to $\mathbf{x}_i' = (1,\ x_i,\ x_i^2)$ or $\mathbf{x}_i' = (1,\ x_i,\ x_i^3)$, respectively. The higher the order of the polynomials being fit, the further the reduction in bias is obtained. However, more variability in the estimate occurs when higher-order polynomials are fit. Therefore, Hastie and Loader (1993), Fan (1993), and Cleveland and Devlin (1998) have found that the local linear and local quadratic fits are most useful.

CHAPTER 4
NONPARAMETRIC METHODS APPLIED TO QUANTAL BIOASSAY

In the previous chapter an introduction to two nonparametric procedures, kernel regression and local linear regression, was given. In this section, these two nonparametric procedures will be applied to quantal dose-response data. In addition, this section contains a brief presentation of two classical nonparametric procedures used in quantal bioassays: The Spearman-Kärber and Thompson's moving average methods. The classical methods were used mainly in estimating ED_{50}'s, thus limiting one who wishes to estimate other effective doses, say ED_{10} or ED_{90}. Over the past decade, much work has been done in nonparametric regression in the general context, but little attention has been given to estimating the dose-response relationship via nonparametric methods, with the exception of the Spearman-Kärber estimate. Copas (1983) introduced kernel regression applied to bioassay, in which a nonparametric binary regression function was proposed as a method of plotting p, the probability of an event (response), against x, the quantitative variable. Kappenman (1987) applied kernel regression to dose-response curves as a means of estimating the ED_{50}. Staniswalis and Cooper (1988) introduced their method for grouped data with one drug as well as a combination of drugs in search of antagonism or synergism between two or more drugs. Kelley and Rice (1990) used a spline-based procedure to estimate dose-response curves and to assess possible synergism and antagonism of the two drugs. Work has also been introduced using local linear regression in conjunction with generalized linear models by Severini and Staniswalis (1994) and Tibshirani and Hastie (1987).

§4.1 Classical Nonparametric Methods Applied to Quantal Bioassays

Classical nonparametric methods emphasize estimation of the ED_{50}, the effective dose required to yield, on average, a response in fifty percent of the subjects. The nonparametric approach to estimating the ED_{50} has its origin in papers by Spearman (1908), Dragstedt and Lang (1928), Behrens (1929), Kärber (1931), Reed and Muench (1938), and Thompson (1947).

The Reed-Muench estimator of the ED_{50} was introduced by Reed and Muench (1938) as a computationally simple method of estimating the median-effective dose. Similar to the Reed-Muench estimator is the Dragstedt-Behrens estimator of the ED_{50} introduced by Dragstedt and Lang (1928) and Behrens (1929). The Reed-Muench and the Dragstedt-Behrens methods do not have a sound theoretical basis, but due to their computational simplicity they were widely accepted in many circles. Neither method is used very often today.

Two of the classical nonparametric procedures that continue to be widely used and accepted in the estimation of the ED_{50} are the Moving Average method and the Spearman-Kärber estimate. The Moving Average method is due to Thompson (1947) and like the aforementioned nonparametric methods, does not require a distributional assumption. Suppose that x_i is the dose administered to n_i subjects and r_i subjects respond, for $i = 1,2,\ldots,d$ The proportion of subjects responding is given by

$$p_i = \frac{r_i}{n_i}$$

For the given doses in the experiment, calculate the moving average of response

$$p_i^* = \left(p_i + p_{i+1} + \ldots + p_{i+j-1}\right)/j$$

for some integer j, and associated with each moving average is

$$x_i^* = \left(x_i + x_{i+1} + \ldots + x_{i+j-1}\right)/j$$

Select the two p_i^* such that $p_i^* < 0.50 < p_{i+1}^*$ and linearly interpolate between x_i^* and x_{i+1}^* to obtain the estimate of the ED_{50}. As far as the value of j is concerned, Thompson (1947) suggests $j = 3$ as a reasonable span of the moving average. Variance estimates of Thompson's Moving Average method can be found in Finney (1950, 1953).

The Spearman-Kärber (S-K) estimate is the most commonly used of the classical methods due to its computational simplicity and sound theoretical properties. Since this estimate will also be used in this research, a more detailed presentation of it will be made. Let Δ_i represent the spacing between consecutive doses given by

$$\Delta_i = \left(x_{i+1} - x_i\right)$$

The S-K estimate of theED_{50} is defined as

$$\hat{\mu} = p_1\left(x_1 - \frac{\Delta_1}{2}\right) + \sum_{i=1}^{d-1}(p_{i+1} - p_i)\left(x_i + \frac{\Delta_i}{2}\right) + (1 - p_d)\left(x_d + \frac{\Delta_{d-1}}{2}\right) \tag{4.1.1}$$

where d is the number of doses and x_i is the i^{th} dose, for $i = 1,2,...,d$. If $p_1 = 0$ and $p_d = 1$, then equation (4.1.1) reduces to

$$\hat{\mu} = \sum_{i=1}^{d-1}(p_{i+1} - p_i)\left(x_i + \frac{\Delta_i}{2}\right) = \frac{1}{2}\sum_{i=1}^{d-1}(p_{i+1} - p_i)(x_i + x_{i+1}) \tag{4.1.2}$$

An alternative expression of equation (4.1.2) is

$$\hat{\mu} = \frac{1}{2}\left\{\sum_{i=2}^{d-1} p_i(x_{i-1} - x_{i+1})\right\} + (x_{i-1} + x_i) \tag{4.1.3}$$

Expression (4.1.3) is convenient for evaluating the properties of the S-K estimate. Assuming that $p_1 = 0$ and $p_d = 1$, the variance of the S-K estimate is given by

$$Var(\hat{\mu}) = \frac{1}{4}\sum_{i=2}^{d-1} P_i(1 - P_i)(x_{i-1} - x_{i+1})^2 / n_i \tag{4.1.4}$$

Since the S-K estimate estimates the mean of the tolerance distribution it is an appropriate estimate of the ED_{50} when the tolerance distribution is symmetric. The general formula for the S-K estimate of the ED_{50} is given by

$$\hat{\mu} = \sum_{i=1}^{d-1}\frac{(x_{i+1} + x_i)}{2}(p_{i+1} - p_i) \tag{4.1.5}$$

As can be seen in the variance properties of the S-K estimate, it is often desired that $p_1 = 0$ and $p_d = 1$. If this is not the case in the experiment, it is often assumed that $p_0 = 0$ and $p_{d+1} = 1$. Although this assumption does not harm the analysis and occurs in practice quite often, it can be avoided by using the trimmed S-K estimate introduced by Hamilton et al. (1977), Miller and Halpern (1980), and Hoekstra (1989). This trimming is done in the same manner as that of the trimmed mean in the normal univariate situation.

§4.2 Kernel Regression Applied to Quantal Bioassays

Using the form of the kernel weights by Nadaraya (1964) and Watson (1964), the kernel estimate of the probability of response at dose x is given by

$$\hat{P}^k(x) = \frac{\sum_{j=1}^{d} K\left(\frac{x - x_j}{b}\right) p_j}{\sum_{j=1}^{d} K\left(\frac{x - x_j}{b}\right)} \tag{4.2.1}$$

where b is the bandwidth and $K(\ )$ is the kernel function, which is a function of the bandwidth. Using the Gaussian kernel, the kernel estimate of the probability of response at dose x_i can be written as

$$\hat{P}^k(x_i) = \frac{\sum_{j=1}^{d} e^{-\left(\frac{x_i - x_j}{b}\right)^2} p_j}{\sum_{j=1}^{d} e^{-\left(\frac{x_i - x_j}{b}\right)^2}} = \sum_{j=1}^{d} h_{ij}^k p_j \tag{4.2.2}$$

where

$$h_{ij}^k = \frac{e^{-\left(\frac{x_i - x_j}{b}\right)^2}}{\sum_{j=1}^{d} e^{-\left(\frac{x_i - x_j}{b}\right)^2}} \tag{4.2.3}$$

In matrix notation, the kernel estimate of the probability of response at the d doses can be expressed as

$$\hat{\mathbf{P}}^k = \mathbf{H}^k \mathbf{p} \tag{4.2.4}$$

with the elements of $\mathbf{H}^k$ given in (4.2.3) and $\mathbf{p}$ is the dx1 vector of observed probabilities of response.

$PRESS^*(b)$ will be used initially to select the bandwidth, b. However, other methods for bandwidth selection will be proposed and studied during this research. Einsporn (1987) and Sutherland (1992) have demonstrated that $PRESS^*(b)$ tends to yield more conservative values of the bandwidth, thus protecting the nonparametric regression procedure from either oversmoothing (the bandwidth is too large) or undersmoothing (the bandwidth is too small) the regression curve.

Applying the aforementioned method of bandwidth selection to dose-response data, this research proposes a weighted $PRESS^*(b)$ statistic given by

$$PRESS^*(b) = \frac{\sum_{i=1}^{d} w_i \left(p_i - \hat{P}_{i,-i}^k(b)\right)^2}{d - tr\left[\mathbf{H}^k\right]} \tag{4.2.5}$$

where $w_i = \frac{n_i}{p_i(1-p_i)}$, serving as the weight, is the reciprocal of the variance at each dose, d is the number of doses used in the experiment, and $\hat{P}^k_{i,-i}(b)$ is the "minus-i" predicted proportion of subjects responding to the i^{th} dose x_i for the current value of b with the i^{th} observation removed. The trace of the dxd weight matrix is denoted by $tr[\mathbf{H}^k]$. The quantity $tr[\mathbf{H}^k]$ is commonly referred to as the model degrees of freedom, and thus $d - tr[\mathbf{H}^k]$ represents the kernel regression error degrees of freedom. For kernel regression, computing $\hat{P}^k_{i,-i}(b)$ does not require performing d different regressions, each with the i^{th} data point deleted. Instead, "minus-i" predictions can be computed by

$$\hat{P}^k_{i,-i}(b) = \sum_{i \neq j}^{d} h^k_{ij,-i} p_j \tag{4.2.6}$$

where

$$h^k_{ij,-i} = \frac{h^k_{ij}}{1 - h^k_{ii}} \tag{4.2.7}$$

Once the optimal bandwidth has been selected, the dose-response curve can be fitted to the data and the $ED_{100\alpha}$ values as well as confidence intervals for the $ED_{100\alpha}$ values can be computed. Because it is not possible to obtain a closed functional form for the estimate of f, the computation of the $ED_{100\alpha}$ value will be done iteratively. That is, predictions will be computed for the dose-response relationship along the dose metameter until the dose $\hat{x}_a$ is found corresponding to $\hat{P}^k(x_a) = a$.

With the value of $\hat{x}_a$ computed, the sampling distribution of $\hat{x}_a$ must be determined before confidence intervals on the $ED_{100\alpha}$ value can be determined. Instead of deriving a distributional form for the dose metameter, the inverse estimation method will be applied to the confidence limits on estimated probabilities to put bounds on the $ED_{100\alpha}$ value. This is done by first computing confidence intervals for $\hat{\mathbf{P}}^k(\mathbf{x})$. In this research, the variance estimate of $\hat{\mathbf{P}}^k(\mathbf{x})$ will be obtained as a function of the dxd kernel weight matrix. One form for the approximate variance of $\hat{P}^k(x_i)$ for dose x_i is given by

$$Var\left(\hat{P}^k(x_i)\right) = \sum h^{k^2}_{ij} \hat{P}^k(x_i)\left(1 - \hat{P}^k(x_i)\right) / n_i \tag{4.2.8}$$

Using (4.2.8), approximate upper and lower confidence 95% intervals on P can be constructed of the form

$$\hat{P}^k(x_i) \pm 1.96\, se\left(\hat{P}^k(x_i)\right) \tag{4.2.9}$$

where $se\left(\hat{P}^k(x_i)\right)$ is the standard error of the nonparametric estimate at dose x_i, which is the square root of the variance of $\hat{P}^k(x_i)$, $Var\left(\hat{P}^k(x_i)\right)$, given in equation (4.2.8). Using the method of "inverse" estimation

presented in **Section 2.2.A Example 1**, and illustrated in **Figure 2.2.1**, the kernel estimates of effective dose and confidence intervals on effective dose can be computed. One thing of note is that the confidence intervals on $\hat{\mathbf{x}}_\alpha$ *will not* be symmetric, although the confidence bounds on $\hat{\mathbf{P}}^{\mathbf{k}}$ are symmetric.

This research proposes the following chi-square goodness-of-fit statistic to measure the adequacy of the kernel fit to the observed data:

$$\chi_k^2 = \sum_{i=1}^{d} \frac{n_i\left(p_i - \hat{P}_i^k\right)^2}{\hat{P}_i^k\left(1-\hat{P}_i^k\right)} \tag{4.2.10}$$

Because the distributional form of this statistic is not known, χ_k^2 will be used only as a method of comparison with the logistic regression analysis and the other nonparametric procedures. Studying the distribution of the nonparametric chi-square statistics will be suggested for future research.

§4.2.A Example 3

Using the dose-response data of **Example 2.2.A** (the fly data), kernel regression was applied with the *PRESS*$^*(b)$ procedure as the method of bandwidth selection and the Gaussian kernel. The data-driven bandwidth for the fly data is 0.1866 with a χ^2 statistic of 1.7692. As opposed to the $p = 2$ model degrees of freedom associated with the parametric logistic regression model, the model degrees of freedom associated with the kernel fit is given by $tr[\mathbf{H}^k] = 3.8265$, implying that it takes kernel regression approximately 3.8 "parameters" to fit the fly data set, almost 2 more model degrees of freedom than the logistic regression analysis. The estimate of the ED_{50} is 0.2461 g/100cc with an approximate 95% confidence interval on the ED_{50} estimate of (0.2059 g/100cc, 0.2956 g/100cc), yielding a width of 0.0896 g/100cc. Note that the kernel estimate of the ED_{50} is slightly larger than that of the logistic estimate of the ED_{50} = 0.2304 g/100cc with a width of 0.0590 g/100cc. Concerning the extreme dose levels, the kernel procedure estimates the ED_{80} to be 0.4558 g/100cc with a 95% confidence interval of (0.3898 g/100cc, 0.5467 g/100cc) having a width of 0.1569 g/100cc. The kernel procedure estimates the ED_{20} as 0.0946 g/100cc with a 95% confidence interval of (0.0324 g/100cc, 0.1321 g/100cc), yielding a width of 0.0997 g/100cc. Also of interest are the PRESS and PRESS* statistics. For the logistic procedure, the PRESS and PRESS* statistics are 14.4272 and 2.8854, respectively, whereas for the kernel procedure, the PRESS(b) and PRESS*(b) statistics are 12.4278 and 3.9161, respectively. So, in terms of "minus-i" predictions, kernel regression yields the better PRESS statistic, and logistic regression produces the smallest PRESS* statistic. **Figure 4.2.1** contains a graph of the kernel fit with a 95% confidence band for the fly data. **Figure 2.2.1** and **Figure 4.2.1** are very similar, although it seems that the kernel fit has more variance, which can be measured by the mean squared error, which is 0.1470 for the logistic fit and nearly four times larger in the kernel fit at 0.5575. Fits at the left and right boundaries seem to overestimate and underestimate, respectively, the observed probabilities of response. All things considered, the kernel regression fits the data reasonably well, though not as well as the logistic analysis.

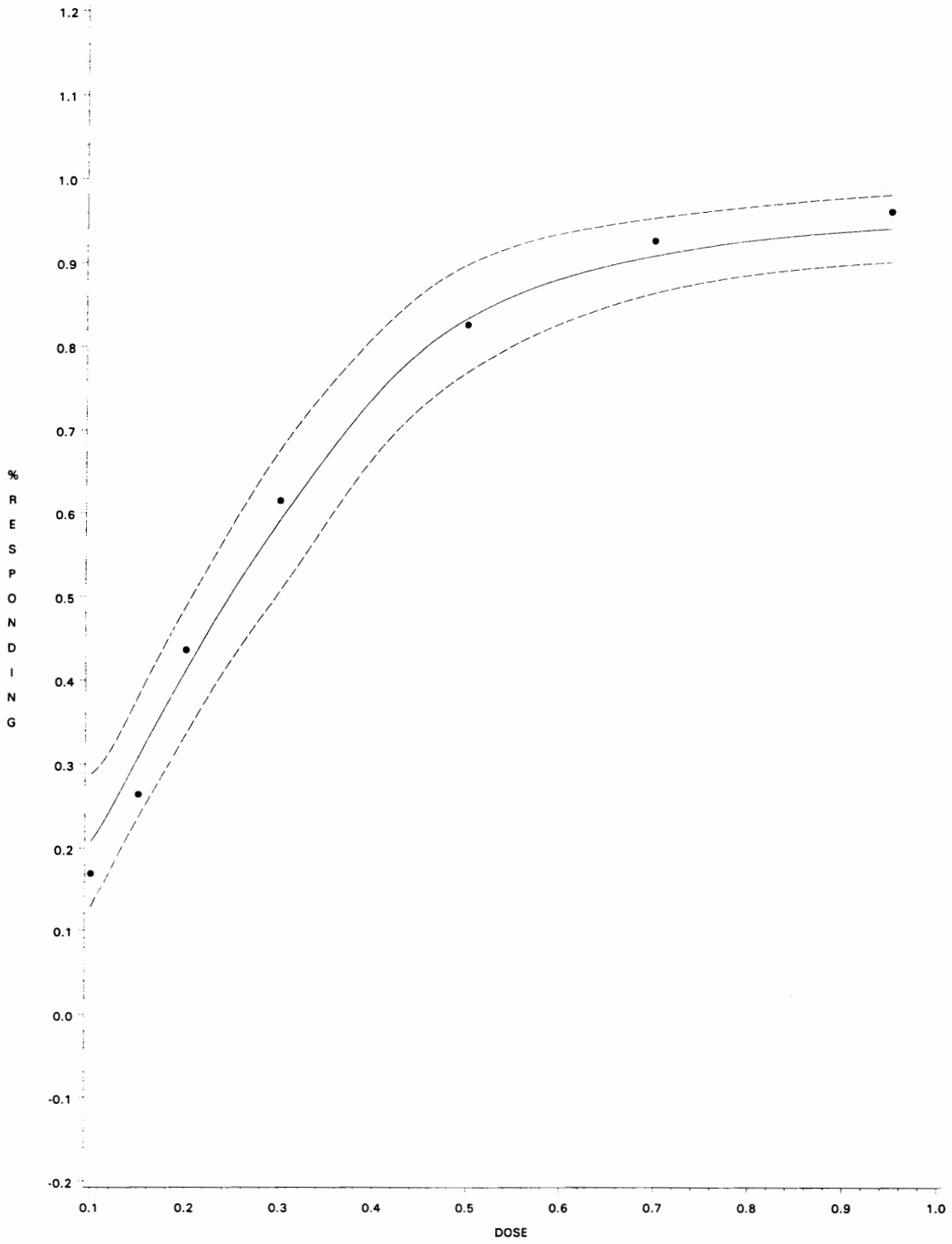


Figure 4.2.1 Kernel Fit to the Fly Dose-response Data

. . . Raw data, ___ Kernel fit, - - - 95% Confidence band

§4.2.B Example 4

The data given in **Table 2.2.2** was created from the logit mixture model given in equation 2.2.B.1. The kernel fit to the data yielded a bandwidth of 0.2314 and a chi-square statistic of 2.3633, which is larger than the chi-square statistic from the logit analysis of 0.6682. Also, the kernel mean squared error of 1.0742 is approximately five times as large as that of the logit analysis of 0.2227. The kernel estimate of the median effective dose is 0.4797 with a 95% confidence interval of (0.3832, 0.5834) yielding a width of 0.2003. Recall that the true ED_{50} is 0.5, thus the logit procedure estimates the median effective dose more accurately than the kernel procedure. The kernel estimate of the ED_{20} and ED_{80} are 0.2084 and 0.7765, respectively. The true value of the ED_{20} is 0.2730, which the kernel procedure estimates with only 76% accuracy, though the true ED_{20} value is covered by the 95% confidence interval of (0.0696, 0.3239). The true ED_{80} over-estimated by nearly 7%, with the true value being 0.7270. All of the true effective doses (ED_{20}, ED_{50}, and ED_{80}) are covered by their respective 95% confidence interval estimates. See **Figure 4.2.2** for a plot of the kernel fit to the logit mixture data used in this example accompanied by a 95% confidence band.

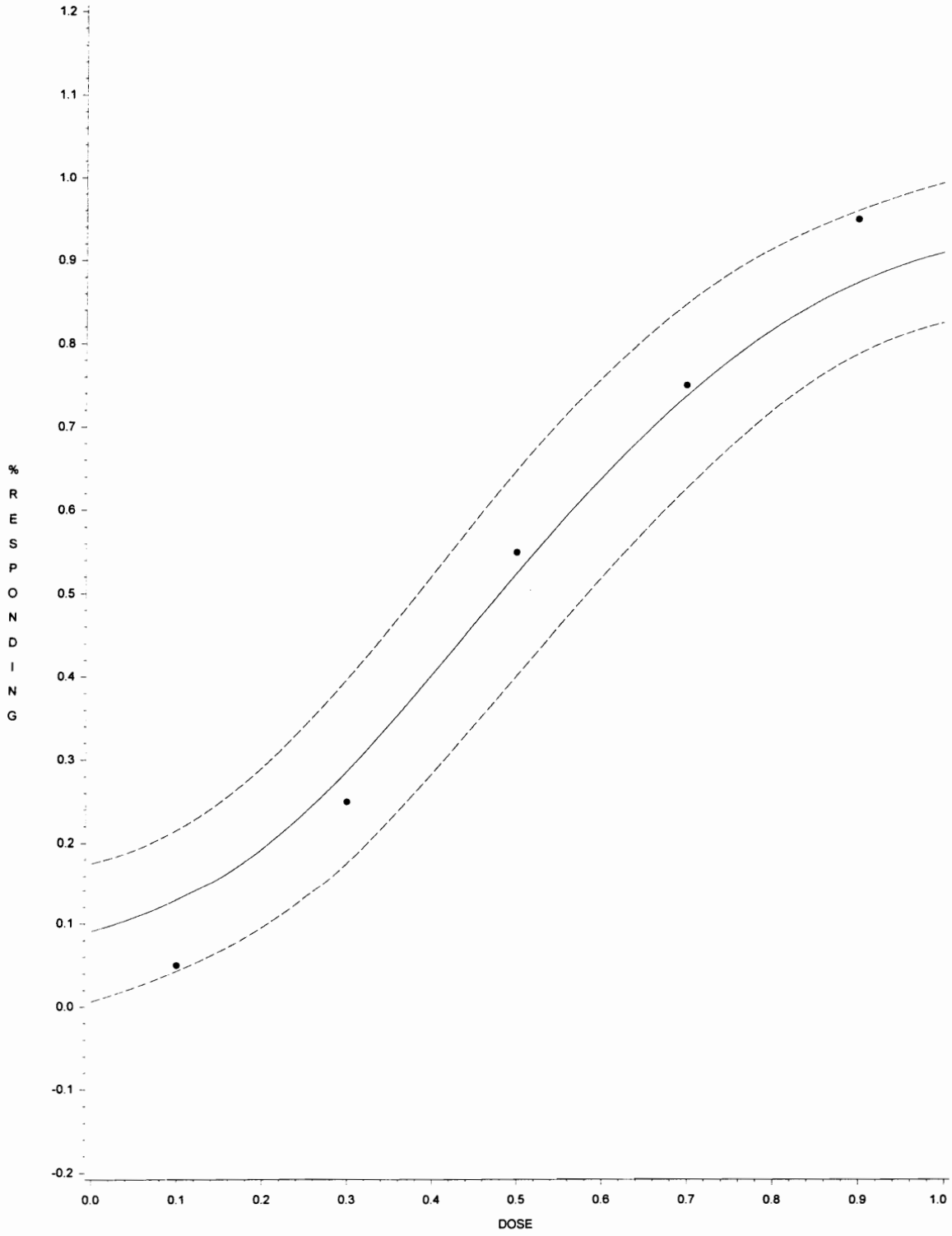


Figure 4.2.2 Kernel fit to Logit Mixture Data

. . . Raw data, ___ Kernel fit, - - - 95% Confidence band

§4.3 Local Linear Regression Applied to Dose-Response Data

In a manner quite similar to kernel regression applied to quantal bioassays, local linear regression may be applied as well. Recall that LLR estimation is analogous to weighted least squares (WLS) estimation in which the rows of the kernel weight matrix are used to construct the weight matrix in WLS. In the quantal dose-response setting, it is known that the variances at each observation are heterogeneous. Therefore, the LLR estimate at dose x_i is given by

$$\hat{P}^{LLR}(\mathbf{x}_i) = \mathbf{x}_i'\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_i^k\mathbf{p} \tag{4.3.1}$$

where $\mathbf{H}_i^k$ is given in (3.3.3) and p is the $dx1$ vector of observed proportions. One can express the vector of predictions at the d doses as

$$\hat{\mathbf{P}}^{LLR} = \begin{bmatrix} \mathbf{x}_1'\left(\mathbf{X}'\mathbf{H}_1^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_1^k \\ \mathbf{x}_2'\left(\mathbf{X}'\mathbf{H}_2^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_2^k \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_d'\left(\mathbf{X}'\mathbf{H}_d^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_d^k \end{bmatrix}\mathbf{p} \tag{4.3.2}$$

which when written as a linear predictor can be expressed as

$$\hat{\mathbf{P}}^{LLR} = \mathbf{H}^{LLR}\mathbf{p} \tag{4.3.3}$$

Using the form of equation (4.2.3), the bandwidth will be chosen using the $PRESS^*(b)$ procedure described in **Chapter 3**, with the "minus-i" predictions obtained through methods analogous to those of the kernel regression procedure.

One form for the approximate variance of the LLR estimate, assuming that the bandwidth is fixed, is analogous in equation (3.3.6). So,

$$Var\left(\hat{P}^{LLR}(\mathbf{x}_i)\right) = \mathbf{x}_i'\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}_i^k\mathbf{V}^{-1}\mathbf{H}_i^k\mathbf{X}\left(\mathbf{X}'\mathbf{H}_i^k\mathbf{X}\right)^{-1}\mathbf{x}_i \tag{4.3.4}$$

Using the estimate of the variance in (4.3.4), one can compute confidence intervals for the LLR estimate. Also of interest is the model degrees of freedom associated with the LLR fit, which in effect can be obtained as the trace of the LLR weight matrix, denoted by $tr\left[\mathbf{H}^{LLR}\right]$. The error degrees of freedom will be expressed as $d - tr\left[\mathbf{H}^{LLR}\right]$. Given the variance estimate, confidence intervals may be obtained using the inverse estimation method described in the previous section for kernel regression.

§4.3.A Example 5

The local linear regression procedure was performed on the fly dose-response data given in **Table 2.2.1**. The LLR procedure has a data-driven bandwidth, selected by $PRESS^{*}(b)$, of 0.3199 and a chi-square statistic of 2.003, which is slightly larger than the chi-square statistic of both the logistic and kernel regression procedures. Though the LLR procedure fits a series of linear regressions to the data, note that the curve, given in **Figure 4.3.1** along with a 95% confidence band, is not linear at all. In fact, the curve fits very well to the data, especially in the boundaries, a problem area for kernel regression. The model degrees of freedom for this LLR fit is 3.7125, slightly less than that of the kernel regression fit. The estimate of the ED_{50} is 0.23715 g/100cc with an approximate 95% confidence interval of (0.2056 g/100cc, 0.2728 g/100cc) and a width of 0.06718. The width of the LLR fit is smaller than the width of the kernel confidence interval on the ED_{50}, but slightly larger than the ED_{50} from the logistic regression procedure. The ED_{20} estimate is 0.1186 g/100cc with a 95% confidence interval of (0.0324 g/100cc, 0.1381 g/100cc). The ED_{80} estimate is 0.4925 g/100cc with a 95% confidence interval of (0.4388 g/100cc, 0.5675 g/100cc) yielding a width of 0.1287 g/100cc. The *PRESS* and $PRESS^{*}$ statistics from the LLR fit are 15.3984 and 4.1477, respectively, which are larger than those from the kernel and the logistic procedures. The mean squared error of 0.5395 in the LLR procedure is slightly less than that of the kernel procedure, but much larger than the logistic mean squared error. The excellent fit in the boundaries supports the contention that a modified kernel, as suggested by Rice (1984) is not necessary.

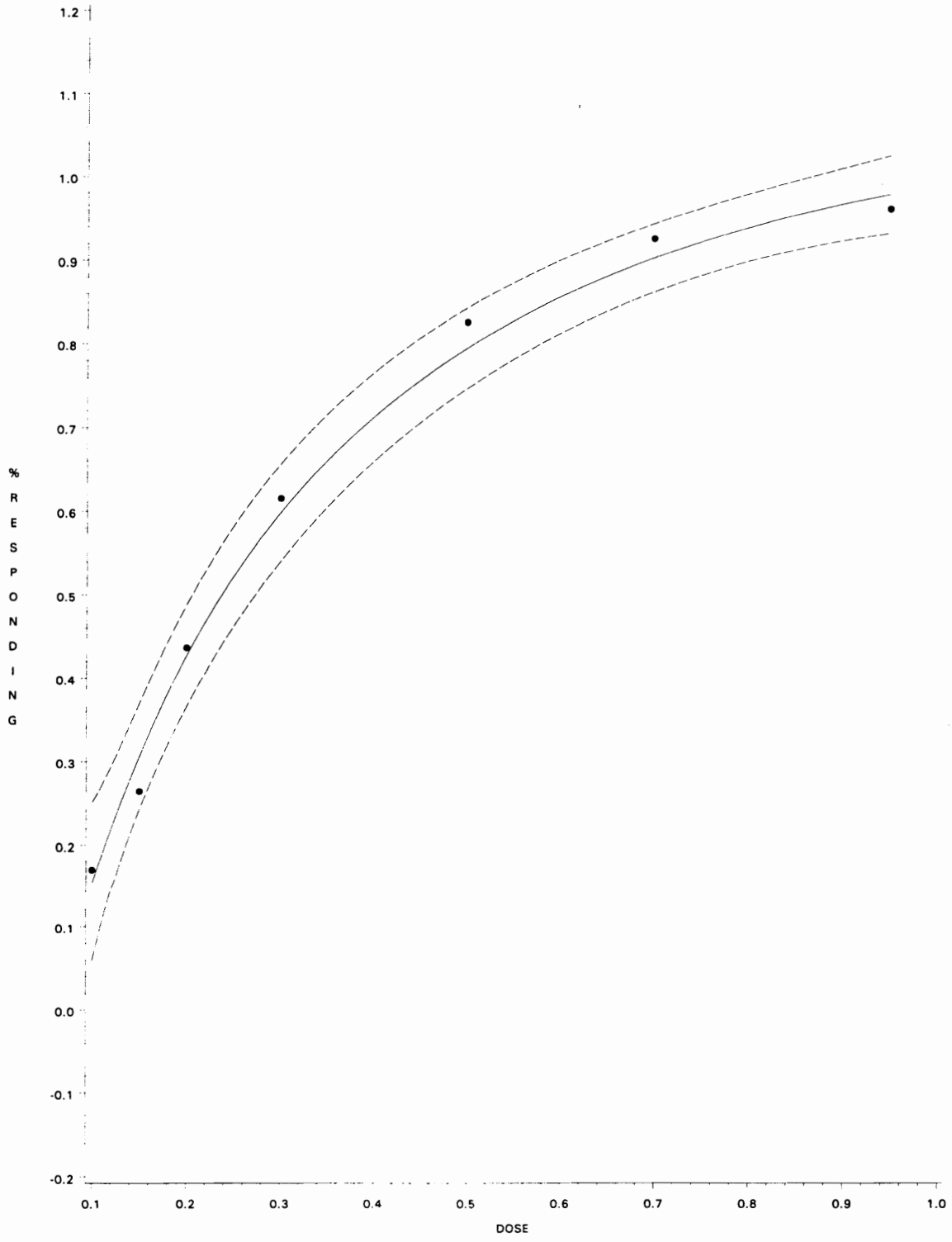


Figure 4.3.1 Local Linear Regression Fit to Fly Data

. . . Raw data, ___ LLR fit, - - - 95% Confidence band

§4.3.A Example 6

The local linear regression procedure was fit to the logit mixture data set given in **Table 2.2.2**. The LLR was fit to the data using a data-driven bandwidth of 1.0000, producing a chi-square statistic of 0.4422 and a mean squared error of 0.1531. The chi-square statistic for the LLR procedure is much smaller than those of the logit and kernel analyses. The bandwidth of one implies that the LLR procedure fits a single line through the entire data with very little curvature, thus in **Figure 4.3.2** the LLR fit is extremely linear. The LLR procedure manages to fit the data quite well, with an ED_{50} estimate of 0.4906, with a 95% confidence interval of (0.4229, 0.5581), yielding a width of 0.1353. Note that the width of the 95% confidence interval on the ED_{50} for the LLR procedure is substantially smaller than the widths of the 95% confidence intervals of the logit and kernel methods. The extreme dose estimates of the ED_{20} and ED_{80} are 0.2315 and 0.7521, respectively, with the true ED_{20} being under-estimated by 8.5%, and the true ED_{80} being under-estimated by 3.3% . The true effective doses for the logit mixture model for γ=0.5 are: ED_{20}=0.2730, ED_{50}=0.5, and ED_{80}=0.7270. The $PRESS^*$ statistic for the LLR procedure is 0.5291, which is nearly three times as small as the $PRESS^*$ from the logit procedure of 1.7716. Also note that with the linear structure of the LLR fit, the procedure used only 2.1119 model degrees of freedom. Although the LLR fits the data extremely well, the dose-response structure, which asymptotes to zero in the left tail and one in the right tail, is still desired.

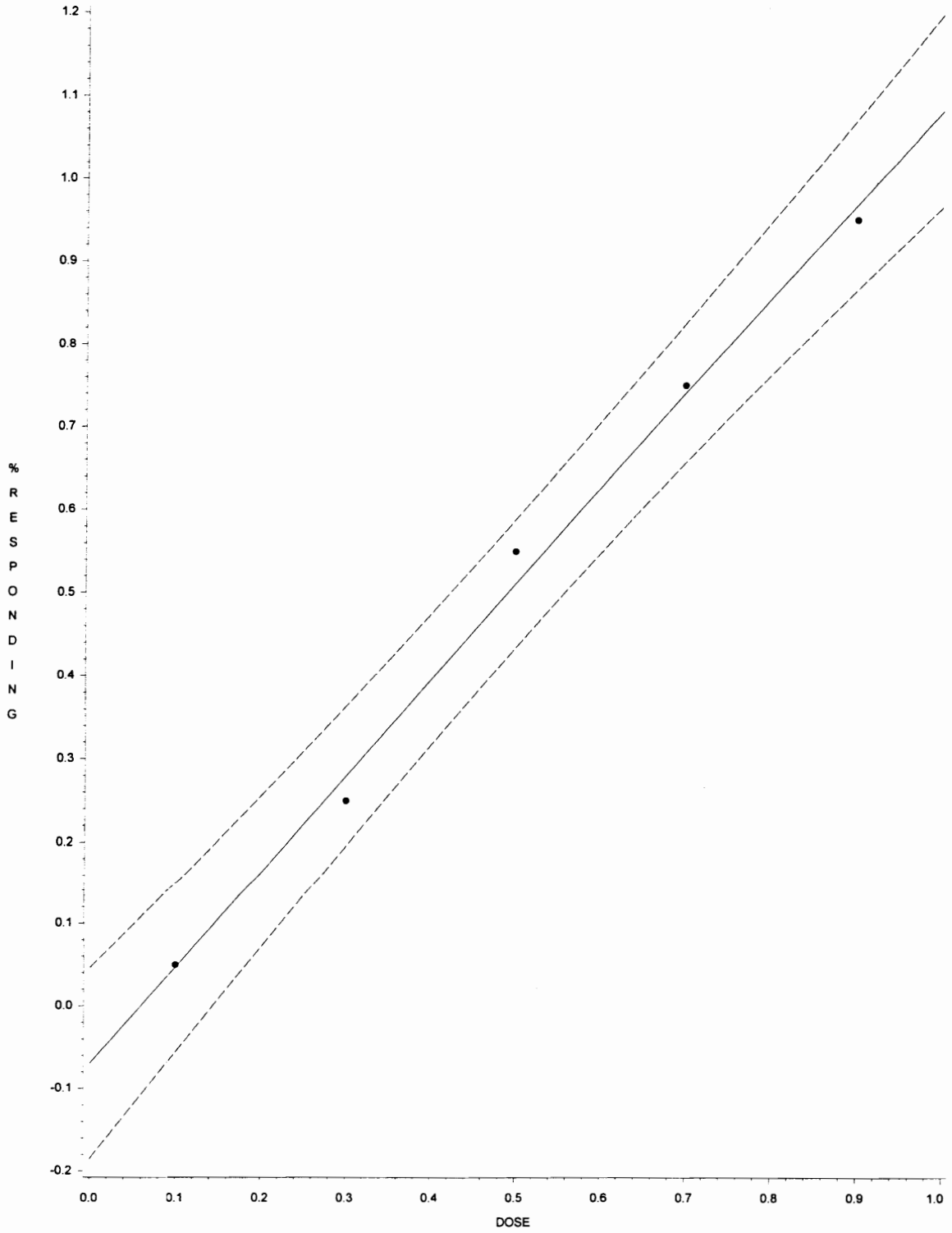


Figure 4.3.2 Local Linear Regression fit to the Logit Mixture Data

. . . Raw data, ___ LLR fit, - - - 95% Confidence band

CHAPTER 5
MODEL-ROBUST REGRESSION FOR LINEAR MODELS

§5.1 Model-Robust Regression for Linear Models

Model-robust regression (MRR) provides a link between parametric and nonparametric curve estimation. The MRR procedure attempts to improve predictions in the regression setting by combining parametric and nonparametric predictions using a mixing parameter λ. That is, the MRR prediction is obtained by

$$\hat{\mathbf{y}}^{MRR}(\lambda) = \lambda\hat{\mathbf{y}}^{NP} + (1-\lambda)\hat{\mathbf{y}}^{P} \tag{5.1.1}$$

where $\hat{\mathbf{y}}^{NP}$ is the nx1 vector of nonparametric predictions and $\hat{\mathbf{y}}^{P}$ is the nx1 vector of parametric predictions. The nonparametric prediction can be obtained from any one of several smoothing techniques such as kernel smoothing, local linear regression, k-nearest neighbor estimates, or spline smoothing. For more information on nonparametric smoothers not covered in this proposal, see Härdle (1990). For the parametric prediction, ordinary least squares can be used in the classical regression setting, although other parametric methods, such as robust regression based on M-estimators, could also be used.

As for the parametric portion of (5.1.1), the OLS predictions are obtained by

$$\hat{\mathbf{y}}^{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}^{P}\mathbf{y} \tag{5.1.2}$$

where $\mathbf{H}^{P}$ is the OLS hat matrix and $\mathbf{y}$ is the nx1 vector of observed responses at the n data points, $\mathbf{x}_i'$, i=1, 2,...,n. If the nonparametric predictions are obtained using kernel regression and the Gaussian kernel, as proposed by Einsporn (1987) and Einsporn and Birch (1993), the estimate is given by

$$\hat{\mathbf{y}}^{NP} = \mathbf{H}^{k}\mathbf{y} \tag{5.1.3}$$

where $\mathbf{H}^{k}$ is the kernel weight matrix with elements h_{ij}^{k} defined in (4.2.4). Methods for finding the bandwidth, b, were given in **Section 3.2**.

There are several methods in which one can use to choose the mixing parameter, λ for the linear model. Einsporn (1987) compares many forms of selecting λ including a variety of "C_p-type" statistics

and "*PRESS*-type" statistics. The work of Einsporn (1987) and that of Sutherland (1992) suggests that minimizing $PRESS^*(\lambda)$ with respect to λ works very well across a wide variety of situations where

$$PRESS^*(\lambda) = \frac{\sum \left(y_i - \hat{y}_{i,-i}(\lambda)\right)^2}{n - tr\left[\mathbf{H}(\lambda)\right]} \tag{5.1.4}$$

where $\hat{y}_{i,-i}(\lambda)$ is the linear combination of "minus-i" predictions obtained from parametric and nonparametric regression, given by

$$\hat{y}_{i,-i}(\lambda) = \lambda\, \hat{y}^{NP}_{i,-i} + (1-\lambda)\, \hat{y}^{P}_{i,-i} \tag{5.1.5}$$

In equation (5.1.4), $\mathbf{H}(\lambda)$ is a linear combination of the nonparametric and parametric weight matrices given by

$$\mathbf{H}(\lambda) = \lambda \mathbf{H}^{NP} + (1-\lambda)\mathbf{H}^{P} \tag{5.1.6}$$

In particular, Einsporn (1987) defined his MRR estimate as

$$\hat{\mathbf{y}}^{MRR}(b,\lambda) = \lambda \mathbf{H}^{k}\mathbf{y} + (1-\lambda)\mathbf{H}\mathbf{y} \tag{5.1.7}$$

$$= \mathbf{H}(\lambda)\mathbf{y} \tag{5.1.8}$$

where

$$\mathbf{H}(\lambda) = \lambda \mathbf{H}^{k} + (1-\lambda)\mathbf{H} \tag{5.1.9}$$

with $\mathbf{H}^k$ being the kernel weight matrix and $\mathbf{H}$ being the OLS "hat" matrix. The b, needed to obtain $\mathbf{H}^k$, and λ used on the right-hand side of (5.1.7) emphasize the dependence of the MRR predictions on these two parameters. The phrase "HATLINK" was coined by Einsporn (1987) to reflect the fact that the MRR predictions used a "hat" matrix formed from a linear combination of the OLS and kernel "hat" matrices. The value of λ will be between 0 and 1.

Once the MRR equation has been established by proper choice of the smoothing parameter, b, and the mixing parameter, λ, then an estimate of variance, σ^2, can be obtained as

$$s^2(\lambda) = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i(\lambda)\right)^2}{n - tr\left(\mathbf{H}(\lambda)\right)} \tag{5.1.10}$$

where $tr\left(\mathbf{H}(\lambda)\right)$ represents, in effect, the model degrees of freedom or "number of parameters" used to fit the regression curve to the data, and the quantity $\left\{n - tr\left[\mathbf{H}(\lambda)\right]\right\}$ represents the "error degrees of freedom" for the MRR fit. For a thorough presentation of model-robust regression applied to the linear model, the reader is referred to Einsporn (1987) and Einsporn and Birch (1993). If the reader is interested in applying MRR for design augmentation with model misspecification, see Sutherland (1992) and Sutherland and Birch (1993).

CHAPTER 6
MODEL-ROBUST REGRESSION APPLIED TO QUANTAL BIOASSAY

§6.1 Model-Robust Quantal Regression

Model-robust regression can be extended to quantal bioassays in an attempt to combine the parametric logistic regression analysis with the nonparametric regression methods in fitting dose-response data. As mentioned earlier, the chi-square lack-of-fit test is proposed as a measure of the performance of a parametric regression analysis, such as logistic regression, on dose-response data. The model-robust quantal regression (MRQR) may provide an improved analysis of dose-response data, either by improving the fit thereby decreasing the chi-square statistic, improving the estimate of the $ED_{100\alpha}$, or by narrower confidence bands resulting in narrower confidence intervals on the $ED_{100\alpha}$ values. The model-robust model for fitting quantal dose-response data is

$$\hat{P}_i^{MRQR} = \lambda \hat{P}_i^{NP} + (1-\lambda) \hat{P}_i^{P} \tag{6.1.1}$$

where $\hat{P}_i^{MRQR}$ is the model-robust estimate of probability of response to dose x_i , λ is the mixing parameter, $\hat{P}_i^{NP}$ represents the estimated probability that the subject will respond to dose x_i obtained through nonparametric methods, and $\hat{P}_i^{P}$ is the estimated probability of response using parametric methods. In this research, kernel and local linear regression will be the nonparametric methods used and logistic regression will be used for the parametric estimates. All applications and results are based on the grouped data case, but the ungrouped data case can be considered with slight modifications to the procedure.

The value of the mixing parameter λ determines the degree to which the predictions are adjusted, with $\lambda = 0$ corresponding to logistic regression and $\lambda = 1$ resulting in a pure nonparametric regression fit. To make predictions using MRQR, an appropriate value of λ must be chosen. The method of selecting λ in this proposal is by cross-validation, the $PRESS^{*}(b)$ procedure of Section 3.2. It should be noted that in the classical linear regression framework, the assumption of homogeneous variance is used when estimating coefficients and making predictions. That is, when fitting data to the model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

it is assumed that ε is independent and identically distributed (iid) with a mean of zero and common variance $\sigma^2\mathbf{I}$. But when working with quantal dose-response data, this assumption is no longer valid. In fact, at each response r_i, the variance of r_i is $n_i p_i(1-p_i)$, and thus the variance of the proportion responding to dose x_i is

$$Var(p_i) = \frac{P_i(1-P_i)}{n_i} \tag{6.1.2}$$

Therefore, due to the heterogeneity of variance, λ is chosen to minimize an adjusted or weighted $PRESS^*(\lambda)$ given by

$$PRESS^*(\lambda) = \frac{\sum_{i=1}^{d} w_i\left(p_i - \hat{P}_{i,-i}(\lambda)\right)^2}{d - \left\{\lambda tr\left[\mathbf{H}^{NP}\right] + (1-\lambda)2\right\}} \tag{6.1.3}$$

where w_i is the reciprocal of the variance at the observation defined as

$$w_i = \frac{n_i}{P_i(1-P_i)}, \tag{6.1.4}$$

$\mathbf{H}^{NP}$ is the nonparametric weight matrix, and the 2 in the denominator is the number of parameters being estimated by the parametric (logistic) regression procedure. The "minus-i" prediction, $\hat{P}_{i,-i}(\lambda)$, is the linear combination of nonparametric and parametric "minus-i" predictions given by

$$\hat{P}_{i,-i}(\lambda) = \lambda\hat{P}^{NP}_{i,-i}(b) + (1-\lambda)\hat{P}^{P}_{i,-i} \tag{6.1.5}$$

For kernel and local linear regression, the "minus-i" predictions are given by

$$\hat{P}^{NP}_{i,-i} = \sum_{j \neq i} h^{NP}_{ij,-i} y_j$$

where

$$h^{NP}_{ij,-i} = \frac{h^{NP}_{ij}}{1 - h^{NP}_{ii}}$$

Recall that the "minus-i" predictions for logistic regression were given in **Chapter 2**. So by simply using the weight matrix for either the kernel or local linear regression procedures, which were introduced in **Chapter 3**, one can calculate the "minus-i" predictions. The adjusted $PRESS^*(\lambda)$ is used to protect against over-fitting some data sets. Einsporn (1987) has shown that using the conventional *PRESS* procedure, as well as error sums of squares criteria, tend to select λ near zero, favoring the parametric regression method.

In many dose-response applications, interest is in predicting the probability of a subject's response to a specific dose of a drug. Prediction at any dose x_0 is obtained by

$$\hat{P}^{MRQR}(x_0) = \lambda\hat{P}^{NP}(x_0) + (1-\lambda)\hat{P}^{P}(x_0) \tag{6.1.6}$$

where $\hat{P}^{P}(x_0)$ represents the parametric prediction at dose x_0, which will be based on logistic regression for this research, and $\hat{P}^{NP}(x_0)$ is the nonparametric prediction, kernel or local linear regression, at dose x_0. The logistic prediction at dose x_0 is given by

$$\hat{P}^{P}(x_0) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 x_0)}}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the MLEs. The nonparametric prediction obtained at dose x_0 using kernel regression is

$$\hat{P}^{k}(x_0) = \sum_j \frac{K\left(\frac{x_0 - x_j}{b}\right)}{\sum_j K\left(\frac{x_0 - x_j}{b}\right)} p_j = \mathbf{h}^{k'}(x_0)\mathbf{p}$$

and the local linear regression estimate at dose x_0 can be obtained from

$$\hat{P}^{LLR}(x_0) = \mathbf{x}_0'\left(\mathbf{X}'\mathbf{H}^{k}(x_0)\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{H}^{k}(x_0)\mathbf{p}$$

where $\mathbf{x}_0' = (1,\ \mathbf{x}_0)$ and $\mathbf{H}^{k}(x_0)$ is the diagonal matrix formed from the vector $\mathbf{h}^{k'}(x_0)$.

Once the MRQR estimate has been computed, confidence bands on the MRQR curve can be calculated. From these, the $ED_{100\alpha}$ value can be computed along with confidence intervals on the $ED_{100\alpha}$ value. The effective doses will be obtained iteratively by the method illustrated in **Section 3.2.** That is, iterate over the range of the doses until x_α is found such that

$$\hat{P}^{MPQR}(x_\alpha) = \lambda\hat{P}^{NP}(x_\alpha) + (1-\lambda)\hat{P}^{P}(x_\alpha) = \alpha \qquad (6.1.7)$$

Then x_α is the appropriate $ED_{100\alpha}$.

If $\hat{\mathbf{P}}^{MRQR}$ represents the dx1vector of estimated probabilities of response at doses $x_1, x_2, \ldots, x_d$, then one proposed form of the variance of the MRQR estimate is given by

$$Var\left(\hat{\mathbf{P}}^{MRQR}\right) = Var\left[\lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{P}\right] \qquad (6.1.8)$$

$$= \left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]\mathbf{V}\left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]' \qquad (6.1.9)$$

where $\mathbf{V} = diag\left[Var(\mathbf{p})\right]$ and $\mathbf{B} = \left\langle f(\mathbf{X}\beta)\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1}$, where $\mathbf{W}$ is a diagonal weight matrix with the i^{th} diagonal element given by

$$w_i = \frac{n_i f^2\left(\mathbf{x}_i'\hat{\beta}\right)}{\hat{P}_i\left(1-\hat{P}_i\right)} = n_i\hat{P}_i\left(1-\hat{P}_i\right) \qquad (6.1.10)$$

with f being the probability density function of the logistic model and $\hat{P}_i = F\left(\mathbf{x}_i'\hat{\beta}\right)$. See **Appendix B** for the derivation of this variance expression for the MRQR estimate.

Given the variance of the MRQR estimate, a 95% confidence interval for the true probability of response at any dose x_0 can be obtained by

$$\hat{P}^{MRQR}(x_0) \pm 1.96\sqrt{Var\left(\hat{P}^{MRQR}(x_0)\right)} \tag{6.1.11}$$

where one proposed expression for $Var\left(\hat{\mathbf{P}}^{MRQR}(x_0)\right)$ is

$$Var\left(\hat{P}^{MRQR}(x_0)\right) = \left[\lambda\mathbf{h}^{NP}(x_0) + (1-\lambda)\mathbf{B}(x_0)\right]' \mathbf{V}\left[\lambda\mathbf{h}^{NP}(x_0) + (1-\lambda)\mathbf{B}(x_0)\right] \tag{6.1.12}$$

where $\mathbf{h}(x_0)$ is the dx1 vector of nonparametric weights given by $\mathbf{h}(x_0) = (h_{01},\ h_{02},\ldots,\ h_{0d})'$ where

$$h_{0j} = \frac{K\left(\frac{x_0 - x_j}{b}\right)}{\sum_{j=1}^{d} K\left(\frac{x_0 - x_j}{b}\right)},\ j = 1,2,\ldots,d$$

and $\mathbf{B}(x_0) = f\left(\mathbf{x}_0'\beta\right)\mathbf{x}_0'\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1}$. The derivation of formula (6.1.12) also appears in **Appendix B**. Note, that the critical value of 1.96 is used to be consistent with the logistic regression analysis, not necessarily as a distributional assumption.

In evaluating the fit of the MRQR procedure, a chi-square statistic can be computed by

$$\chi^2_{MRQR} = \sum_{i=1}^{d} \frac{n_i\left(p_i - \hat{P}_i^{MRQR}\right)^2}{\hat{P}_i^{MRQR}\left(1 - \hat{P}_i^{MRQR}\right)} \tag{6.1.13}$$

Just like the chi-square statistics from the kernel and LLR procedures, the distributional form of the MRQR chi-square statistic needs to be studied. Until further research has been done on the distribution of the nonparametric chi-square statistics, they will be used for comparison only and not as a formal test. That is, since the method of maximum likelihood finds the unknown coefficients that minimize the chi-square statistic, the procedure yielding the minimum chi-square statistic, will be taken into consideration as providing the better fit to the data.

CHAPTER 7
PROPERTIES OF MRQR UNDER MODEL MISSPECIFICATION

§7.1 Bias, Variance, and Mean Squared Error Properties

For the procedures presented in this research, bias, variance, and mean squared error expressions have been derived (See Appendix B for more details). In this section, a brief summary of these derivations will be given. Following the presentation of the mean squared error properties, tables and graphs containing mean squared error efficiencies with respect to the logistic regression procedure will be presented. For the nonparametric and model-robust procedures given in this research, the bandwidth and mixing parameters are obtained which minimize the mean squared error formulas.

For the parametric procedure, specifically logit analysis, it will be assumed that the true model can be written as

$$\mathbf{P} = \mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{X}\beta) + \mathbf{H}(\mathbf{x}) \tag{7.1.1}$$

where $\mathbf{P}$ is the dx1 vector of probabilities of response, $\mathbf{G}$ is the true cumulative distribution function, a function of the d doses, $\mathbf{F}$ is the user's fitted model, specifically the logistic cdf, and $\mathbf{H}(\mathbf{x})$ denotes the difference between the user's postulated model and the true model. That is, $\mathbf{H}(\mathbf{x}) = \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{X}\beta)$.

The bias, variance and mean squared error formulas that follow are all asymptotic in nature and, consequently, they are only approximate formulas for finite samples. Using the model given in (7.1.1), the bias for the parametric procedure can be expressed as

$$Bias\left(\hat{\mathbf{P}}^{\mathbf{P}}\right) \approx \left[\left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x}) \tag{7.1.2}$$

where $\hat{\mathbf{P}}^{\mathbf{P}}$ is the dx1 parametric estimate of proportion responding across the d doses, with $\left\langle f(\mathbf{X}\beta)\right\rangle$ being a diagonal matrix with elements of $\left(f(\mathbf{x}_1'\beta), f(\mathbf{x}_2'\beta),..., f(\mathbf{x}_d'\beta)\right)$. The variance expression for the parametric prediction is given by

$$Var\left(\hat{\mathbf{P}}^{\mathbf{P}}\right) \approx \left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}\mathbf{V}_{\hat{\beta}}\mathbf{X}'\left\langle f(\mathbf{X}\beta)\right\rangle \tag{7.1.3}$$

where $\mathbf{V}_{\hat{\beta}}$ is an expression for the variance of the coefficient vector β written as

$$Var\left(\hat{\beta}\right) \approx (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle\frac{\mathbf{G}}{\mathbf{F}}\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{7.1.4}$$

with

$$\left\langle \frac{\mathbf{G}}{\mathbf{F}} \right\rangle = \left\langle \frac{G(\mathbf{x}_1)(1-G(\mathbf{x}_1))}{F(\mathbf{x}_1'\beta)(1-F(\mathbf{x}_1'\beta))}, \cdots, \frac{G(\mathbf{x}_d)(1-G(\mathbf{x}_d))}{F(\mathbf{x}_d'\beta)(1-F(\mathbf{x}_d'\beta))} \right\rangle \tag{7.1.5}$$

being a diagonal matrix. Given the bias and the variance of the parametric procedure, the approximate mean squared error is computed as the sum of the variance and the squared bias. A noteworthy observation is that if the user's model is correct, that is if **F**=**G**, then $\mathbf{H}(\mathbf{x}) = 0$ and $\left\langle \frac{\mathbf{G}}{\mathbf{F}} \right\rangle = \mathbf{I}$, which yields zero bias for the parametric procedure, the variance expression of (7.1.3) is equivalent to expression (2.1.10), and the mean squared error expression is comprised of only the variance term. The mean squared error for any point x_0 can be expressed as the following:

$$MSE\left(\hat{P}(x_0)\right) \approx f^2(\mathbf{x}_0'\beta)\mathbf{x}_0'\mathbf{V}_{\hat{\beta}}\mathbf{x}_0 + \left[f(\mathbf{x}_0'\beta)\mathbf{x}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{H}(\mathbf{x}) - \mathbf{H}(x_0)\right]^2.$$

For the two nonparametric procedures, kernel regression and local linear regression, the mean squared error properties will be developed for the model

$$\mathbf{P} = \mathbf{G}(\mathbf{x})$$

where **G** is any arbitrary cdf. The bias for the nonparametric prediction, based on a fixed bandwidth, is given by

$$Bias\left(\hat{\mathbf{P}}^{NP}\right) = \left(\mathbf{H}^{NP} - \mathbf{I}\right)\mathbf{P} \tag{7.1.6}$$

where $\mathbf{H}^{NP}$ corresponds to the nonparametric weight matrix for either kernel or local linear regression. Similarly, the variance can be expressed as

$$Var\left(\hat{\mathbf{P}}^{NP}\right) = \mathbf{H}^{NP}\mathbf{V}_{\mathbf{G}}\mathbf{H}^{NP'} \tag{7.1.7}$$

where

$$\mathbf{V}_{\mathbf{G}} = Var(\mathbf{p}) = \left\langle \frac{\mathbf{G}(\mathbf{x}_1)(1-\mathbf{G}(\mathbf{x}_1))}{\mathbf{n}_1}, \cdots, \frac{\mathbf{G}(\mathbf{x}_d)(1-\mathbf{G}(\mathbf{x}_d))}{\mathbf{n}_d} \right\rangle$$

Having the bias and the variance of the nonparametric procedure, the mean squared error can be easily obtained. The mean squared error at any point x_0 is given by

$$MSE\left(\hat{P}^{NP}(x_0)\right) = \mathbf{h}'(x_0)\mathbf{V}_{\mathbf{G}}\mathbf{h}(x_0) + \left[\mathbf{h}'(x_0)\mathbf{G}(\mathbf{x}) - \mathbf{G}(x_0)\right]^2$$

where **h** is the dx1 vector of weights associated with x_0, and $\mathbf{G}(x_0)$ is the true proportion responding at x_0.

For the model-robust quantal regression procedure, which is a linear combination of the parametric and nonparametric predictions, written as

$$\hat{\mathbf{P}}^{MRQR} = \lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{P}, \tag{7.1.8}$$

the properties will not be strictly additive. The variance expression is approximated by

$$Var\left(\hat{\mathbf{P}}^{MRQR}\right) = Var\left[\lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{P}\right]$$
$$\approx \left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]\mathbf{V}_{\mathbf{G}}\left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]'$$

where $\mathbf{B} = \langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}$. The bias for the MRQR procedure can be expressed as

$$Bias\left(\hat{\mathbf{P}}^{MRQR}\right) \approx \lambda\left[\mathbf{H}^{NP} - \mathbf{I}\right]\mathbf{P} + (1-\lambda)\left[\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})$$

which turns out to be a linear combination of the biases of the parametric and nonparametric procedures. As for the mean squared error, note that under the correctly or near correctly specified parametric model $\mathbf{H}(\mathbf{x}) = 0$ and $\lambda \approx 0$ such that the mean squared error is comprised of primarily the parametric variance. On the other hand, if the user's postulated model has been grossly misspecified, then $\lambda \approx 1$ and $\mathbf{H}(\mathbf{x}) \neq 0$, and thus the mean squared error will be made up of mainly the nonparametric fit to the data. For any given data point, x_0, the variance and bias for the model-robust quantal regression procedure can be expressed as

$$Var\left(\hat{P}^{MRQR}(x_0)\right) \approx \left[\lambda\mathbf{h}'(x_0) + (1-\lambda)\mathbf{B}(\mathbf{x}_0)\right]\mathbf{V}_{\mathbf{G}}\left[\lambda\mathbf{h}'(x_0) + (1-\lambda)\mathbf{B}(\mathbf{x}_0)\right]'$$

and

$$Bias\left(\hat{P}^{MRQR}(x_0)\right) \approx \lambda\left[\mathbf{h}'(x_0)\mathbf{G}(\mathbf{x}) - \mathbf{G}(x_0)\right] + (1-\lambda)\left[f(\mathbf{x}_0'\beta)\mathbf{x}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{H}(\mathbf{x}) - \mathbf{H}(x_0)\right]$$

where

$$\mathbf{B}(\mathbf{x}_0) = f(\mathbf{x}_0'\beta)\mathbf{x}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}.$$

The alternative model-robust procedure, deemed MRQR2 in this research, in which a nonparametric method is used to obtain a fit to the residuals from the parametric model and added to the parametric fit using a mixing parameter, also denoted by λ. The MRQR2 procedure is analogous to the work done in Speckman (1988). In Speckman (1988) the nonparametric method was used to obtain the fit of partial residuals of another variable, and not solely the residuals as is the case for MRQR2. Mays (1995) has found some good results in linear models when using the MRQR2 procedure. The MRQR2 procedure is also analogous to the resmoothing technique of Tukey (1977), where the final "smoothed" fit is the sum of the initial smooth fit to the data plus the smooth of the residuals.

The MRQR2 estimate is given by

$$\hat{\mathbf{P}}^{MRQR2} = F\left(\mathbf{X}\hat{\beta}\right) + \lambda\mathbf{H}^{NP}\mathbf{e} \qquad (7.1.9)$$

where $\mathbf{e} = \mathbf{p} - F\left(\mathbf{X}\hat{\beta}\right)$ is the vector of residuals from the parametric fit, $\mathbf{H}^{NP}$ is the nonparametric weight matrix used to fit the residuals, and λ is the mixing parameter denoting the amount of residual information being added to the parametric fit. The bias for this MRQR2 procedure is written as

$$Bias\left(\hat{\mathbf{P}}^{MRQR2}\right) \approx \left(\mathbf{I} - \lambda\mathbf{H}^{NP}\right)\left[\left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x}) \qquad (7.1.10)$$

and the variance is approximated by

$$Var\left(\hat{\mathbf{P}}^{MRQR2}\right) \approx \left[\mathbf{B} + \lambda\mathbf{H}^{NP}(\mathbf{I} - \mathbf{B})\right]\mathbf{V}_{\mathbf{G}}\left[\mathbf{B} + \lambda\mathbf{H}^{NP}(\mathbf{I} - \mathbf{B})\right]' \qquad (7.1.11)$$

The variance and bias for the MRQR2 procedure at any point x_0 can be expressed in the following forms:

$$Var\left(\hat{P}^{MRQR2}(x_0)\right) \approx \left[\mathbf{B}(\mathbf{x}_0) + \lambda\mathbf{h}'(x_0)(\mathbf{I} - \mathbf{B})\right]\mathbf{V}_{\mathbf{G}}\left[\mathbf{B}(\mathbf{x}_0) + \lambda\mathbf{h}'(x_0)(\mathbf{I} - \mathbf{B})\right]'$$

and

$$Bias\left(\hat{P}^{MRQR2}(x_0)\right) \approx -\mathbf{H}(x_0) + \left[\mathbf{B}(\mathbf{x}_0) + \lambda\mathbf{h}'(\mathbf{x}_0)(\mathbf{I} - \mathbf{B})\right]\mathbf{H}(\mathbf{x}).$$

One rationale behind adding a portion of the residuals to the parametric fit is that if the parametric model has been misspecified to some degree, then there should be "structure" in the residuals, and therefore the residual fit will compensate for the misspecification. That is, if the residuals are a random scatter of points about zero, then the nonparametric fit will be a flat (horizontal) line, revealing no structure. But if the residuals are not random and show some structure, the intent is for the nonparametric fit to find the structure and add some fraction (determined by λ) of it back to the parametric fit.

§7.2 Comparison of Theoretical Average MSE

Given the properties derived in the previous section, the optimal values of the bandwidth and mixing parameters for the nonparametric and model-robust procedures were computed to minimize the average mean squared error at the design points. For the parametric procedure, the coefficient vector β was computed by the method of maximum likelihood to fit the data points. That is, assuming that **G** is the true cumulative distribution function, then the true response, P, can be obtained at each of the dose levels. Having the "true" responses at *d* evenly spaced dose levels, the coefficient vector, β, is now obtained via the method of maximum likelihood.

The domain, x, of the function **G** will be scaled to be between zero and one. The dose levels (or design points) will be evenly spaced in a manner such that almost the entire curve will be covered. The general formula for the spacing between doses is given as

$$spacing = \frac{max(dose) - min(dose)}{d-1}$$

where max(dose) is the maximum dose level and min(dose) is the minimum dose level used. In this research, the minimum dose level used will be 0.1 and the maximum dose level will be 0.9. For example, if d=3 (three doses or design points), then the spacing between doses is 0.4 and the dose levels will be 0.1, 0.5, and 0.9. Similarly, if d=5, then the spacing between design points will be 0.2 and the design points will be 0.1, 0.3, 0.5, 0.7, and 0.9.

In order to evaluate the theoretical mean squared error for a given procedure, several "model" parameters must first be set. These parameters include the form of the true cdf, **G(x)**, the number of doses, *d*, and the number of replicates at each dose, n_i. The model used to obtain mean squared error estimates was

$$\mathbf{G}(x) = (1-\gamma)L(x;0.5,\ 0.1) + \gamma\left[\delta L(x;0.25,\ 0.05) + (1-\delta)L(x;0.75,\ 0.05)\right] \qquad (7.2.1)$$

which, depending on the value of γ, is a mixture of logistic cumulative distribution functions. The notation $L(x;\mu,\tau)$ is the logistic cdf written as

$$L(x;\mu,\tau) = \frac{1}{1+e^{-\left(\frac{x-\mu}{\tau}\right)}} \qquad (7.2.2)$$

with μ and τ being the location and scale parameters of the logistic distribution, respectively. The value of δ will be 0.5 throughout this research. With the value of δ = 0.5, **G** is a mixture of a logistic cdf and a symmetric bimodal cdf. Other values of δ, not considered in this research, will yield asymmetric distributions. This form of **G(x)** was chosen because it provides a broad variety of reasonable cdfs ranging from the logistic cdf to cdfs very different from the logistic. The value of γ, which will denote the degree

of model misspecification, will range from zero to one. If γ=0, then the model will be a logistic cdf with location and scale of 0.5 and 0.1, respectively, and the user's model will be correct since it is assumed that the user always selects the logistic model. As the value of γ increases to one, the degree of model-misspecification increases. **Figure 7.2.1** contains plots of $\mathbf{G}(\mathbf{x})$ for values of γ=0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, and 1.0. Note that for γ=0, $\mathbf{G}(\mathbf{x})$ is the logistic cdf, then γ=0.1 denotes slight misspecification, and as γ increases to one, the degree of misspecification increases. The misspecification is noticeable in the plot, especially in the left and right portions, where the curves tend to separate as γ increases.

Finney (1978) suggests that the number of doses (or design points) be equally spaced and that the same number of subjects (or replicates) be assigned to each dose. Thus, along with the range of values indicated for γ, there will also be three doses used (d=3, 5, and 7) and three sets of replicates at each dose level (n_i=10, 20, and 50). The dose levels and replicates at each dose level were chosen to represent typical quantal bioassay situaions found in practice.

The measure by which the procedures in this research will be judged is the average mean squared error statistic computed either over the design points (when finding the optimal bandwidth or optimal mixing parameter) or over the entire curve (for comparing the procedures). Note that the mse formulas given in the previous section are either *dxd* matrices or scalars representing the mse at an individual dose, x_0. To quantify the mean squared errors at the *d* design points, the trace of the mean squared error matrices is taken, which is the sum of the mses across the *d* dose levels. Then the optimal bandwidth or optimal mixing parameter were chosen to minimize this sum (or equivalently, their average). Once the optimal bandwidth and optimal mixing parameter (if appropriate) were chosen, the average mse (amse) was computed for each procedure by averaging the individual mses at each dose over 100 evenly spaced doses ranging from the minimum dose level of the design to the maximum dose level. The amse for each procedure represents that procedure's ability to fit the curve $\mathbf{G}(\mathbf{x})$ by taking it into account both average bias and average variance of the fitted responses over the entire curve. Naturally, the smaller the amse the better a procedure is able to fit $\mathbf{G}(\mathbf{x})$.

To compare procedures across all "model" parameters, the average mse efficiency (amse-efficiency) for any nonparametric or model-robust procedure ("other") with respect to the logit method is computed, where

$$amse-efficiency = \frac{amse(Logit)}{amse(other)} \tag{7.2.3}$$

Therefore, the yardstick by which to assess the mse-efficiencies is one. So, if the amse-efficiency for a given model situation is greater than one, then the logit procedure has a larger average mean squared error and the "other" procedure is preferred; similarly, if the amse-efficiency is less than

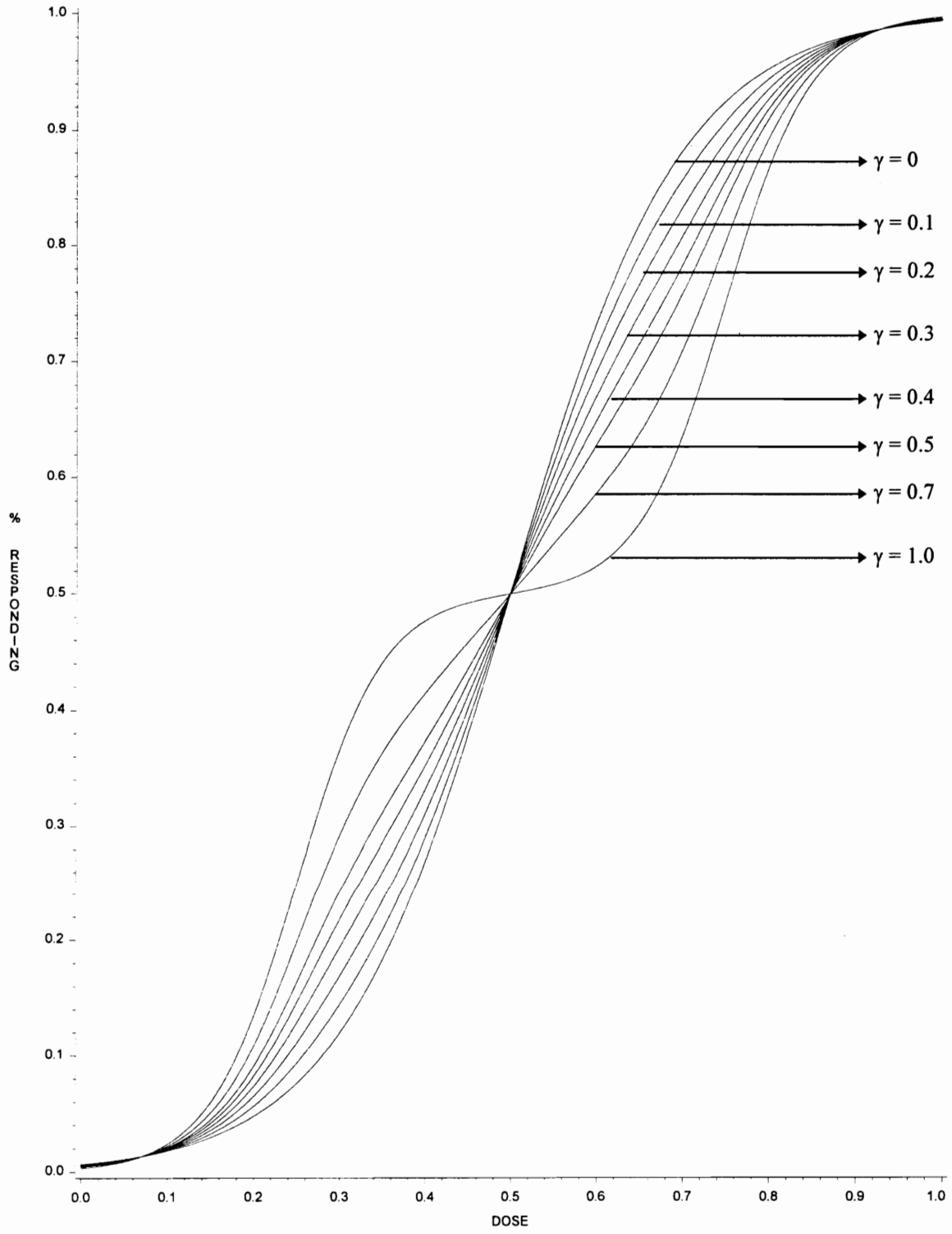


Figure 7.2.1 Plot of the model function **G(x)** as the degree of model-misspecification, γ, increases from 0 to 1.

one, then the logit procedure has the smaller average mse and is preferred over the "other" procedure for a given model situation.

Tables 7.2.1-7.2.3 contain mean squared error efficiencies for each of the (d, n) combinations for the values of γ given above. Listed below are some of the more salient observations regarding the theoretical mean squared error efficiencies.

1. For three doses (d=3), the kernel efficiencies are less than one for γ=0, 0.1, and 0.2 at 10 and 20 subjects. For d=3, n=50, the kernel efficiencies are less than one for γ=0, 0.1, 0.2, 0.3, and 0.4. So, for a small number of doses and at mild model-misspecification, the kernel procedure does not have a smaller amse than the logit procedure.

2. The MRQR Kernel procedure has all efficiencies equal to or greater than one at the (d=3, n=10) combination and all but one efficiency equal to or greater than one at the (d=3, n=20) and (d=3, n=50) scenarios. This indicates that the MRQR Kernel procedure is detecting even the slightest model-misspecification with (d=3, n=10,20).

3. For d=3, when the number of subjects increases to 50, the MRQR Kernel fails to detect any model-misspecification since all efficiencies are one. This implies that the optimal λ is zero and the fits from the MRQR Kernel technique are obtained solely from the logit model. The MRQR Kernel method then fails to take advantage of the increased efficiency from the kernel method. Therefore, for a large number of subjects at each dose, the logit procedure and the MRQR Kernel procedure are equal. Also noteworthy, at d=3, is that the Kernel procedure has the smaller amse when compared to MLE when γ is 0.4 or higher (mild to more extreme misspecification), regardless of the number of subjects at each dose.

4. When applying the local linear regression (LLR) procedure to a small number of doses, d=3, the efficiencies with respect to the logit procedure are greater than one for n=10 with the exception of γ=0 (the correct model). At the (d=3, n=20) combination, the LLR procedure outperforms the logit procedure for all values of γ greater than 0.2. Similarly, for the (d=3, n=50) scenario, the LLR procedure has a smaller mse than logit for values of γ greater than 0.3.

5. As the sample size (n) increases, the logit procedure performs well under slight model misspecification.

6. For d=3, the model-robust quantal regression procedure using LLR (MRQR LLR) has a mixing parameter of one (λ=1) for all (d, n, γ) combinations.

7. For the MRQR2 procedures (using Kernel and LLR as the nonparametric method to fit the residuals), with d=3, the nonparametric fit on the residuals is unable to detect (or fit) any structure in the residuals, therefore yielding a mixing parameter of zero and an efficiency of one. So, even with gross model-misspecification, the MRQR2 procedures will not detect any structure in the residuals when there are only three doses available.

8. In comparing the efficiencies for the (d=3, n, γ) combinations, the MRQR LLR procedure has much larger values than the MRQR Kernel procedure. So, for three doses, the MRQR LLR (or LLR) procedure outperforms all of the other new procedures presented in this research. **Figures 7.2.2-7.2.4** are mse efficiency plots for (d=3, n=10), (d=3, n=20) and (d=3, n=50) combinations, respectively, against γ. It is evident in the plots that as the degree of model-misspecification increases (i.e., γ increases), the MRQR LLR efficiencies are greater than the other efficiencies. In some of the efficiency plots, if the value of λ is zero then the parametric and model-robust plots will overlap. Similarly, if the value of λ is one, then the nonparametric and model-robust procedures will overlap. This overlapping will not occur with MRQR2 when λ=1.

9. For d=5 and n=10, the MRQR Kernel and MRQR LLR procedures have efficiencies greater than one for all values of γ. Note that for Kernel regression, at the (d=5, n=10) combination, the efficiencies are less than one with the exception of at γ=0.7 and 1.0. Thus, the MRQR Kernel efficiencies all greater than one imply that the model-robust procedure (mixing) has improved upon the logit as well as the kernel procedures. Also, the efficiencies for the MRQR LLR procedure are always greater than one or equal to those for LLR. Consequently, it is apparent that the MRQR method is working as anticipated, improving the fits are over those obtained by either the parametric or nonparametric methods.

10. For d=5, n=10, the MRQR2 procedures, λ=0 for all values of γ except for γ=0.7 and 1.0. Note that the efficiencies of the MRQR LLR procedure is smaller than the other nonparametric and model-robust procedures.

11. As the number of doses increases and the number of subjects at each dose increases, the model-robust procedures (MRQR Kernel and MRQR LLR) tend to detect model-misspecification for

small values of γ, i.e., at small values of γ, the value of λ is nonzero. For example, with the (d=5, n=20) combination, the MRQR Kernel and MRQR LLR procedures have efficiencies greater than one, with the exception of when the model has been correctly specified for MRQR Kernel. The MRQR Kernel procedure contains efficiencies greater than the logit and kernel procedures alone, thus indicating that mixing the two procedures is improving on the amse as the degree of model-misspecification increases.

12. For the MRQR LLR procedure, the results are the same as those of the MRQR Kernel procedure, except efficiencies are higher. Therefore, of the two nonparametric procedures used in the model-robust procedure, the local linear regression procedure is more efficient in terms of amse.

13. The MRQR2 procedures do not improve the amse except for large degrees of misspecification (γ=0.5, 0.7, 1.0). The MRQR2 LLR procedure has a higher mse efficiency than the MRQR LLR procedure at γ=1.0.

14. Increasing the number of subjects to 50 with five dose, the MRQR Kernel and MRQR LLR procedures tend to have efficiencies ranging from one to 2.38 and 2.81, respectively. Under the correctly specified model or slight model-misspecification (γ=0.1), the model-robust efficiencies are one. But as γ increases from 0.1, the efficiencies rise above one, again indicating that as the number of doses increases and the number of subjects increase at each dose, the model-robust procedures detects the misspecification for small values of γ.

15. Similar to the previous scenario of (d=5, n=20), the MRQR LLR procedure has higher efficiencies than the other procedures with the exception of MRQR2 LLR at γ=1.0 (the highest degree of misspecification). **Figures 7.2.5-7.2.7** provide graphical representation of the mse efficiencies for d=5 and n=10, 20, and 50, respectively. Figure 7.2.5 shows that as the degree of misspecification increases, the MRQR LLR has the highest efficiency for all values of γ. But MRQR LLR is outperformed by MRQR2 LLR for the (d=5, n=20) and (d=5, n=50) combinations at γ=1.0.

16. When the number of doses increases to seven, both model-robust procedures (MRQR and MRQR2) tend to perform extremely well with respect to amse efficiencies. The MRQR procedures have amse efficiencies ranging from one to 2.05 for n=10 subjects at each dose. For no or slight model-misspecification (γ=0, 0.1, and 0.2) the efficiencies are one, but as γ increases from 0.3 and above, the efficiencies increase above one, again indicating that mixing is helping to

improve on the individual parametric and nonparametric procedures. The same observations can be made for the (d=7, n=20) and (d=7, 50) combinations. Also, the MRQR2 LLR procedure only outperforms the MRQR LLR procedure at γ=1.0 as was previously observed. **Figures 7.2.8-7.2.10** are plots of the mse efficiencies for the (d=7, n) scenarios.

In conclusion, for all the (d, n, γ) combinations mentioned in this research, the mse efficiencies for the MRQR LLR are greater than the mse efficiencies for LLR with one exception. This is a clear indication that the model-robust procedure works and is mixing the parametric and nonparametric procedures in a manner to obtain a balance between the variance and squared bias.

Table 7.2.1 Average MSE Efficiencies with respect to the logit procedure for d=3 doses.

D	N	γ	KERNEL	MRQR KERNEL	LLR	MRQR LLR	MRQR2 KERNEL	MRQR2 LLR
3	10	0.0	0.6051	1.0261	0.8677	0.8677	1.0000	1.0000
		0.1	0.7317	1.1651	1.1215	1.1215	1.0000	1.0000
		0.2	0.9203	1.3229	1.5119	1.5119	1.0000	1.0000
		0.3	1.1780	1.4787	2.1385	2.1385	1.0000	1.0000
		0.4	1.5196	1.5816	2.9245	2.9245	1.0000	1.0000
		0.5	1.8925	1.6449	3.9111	3.9111	1.0000	1.0000
		0.7	2.6180	1.6409	5.0652	5.0652	1.0000	1.0000
		1.0	2.8480	1.5214	4.0455	4.0455	1.0000	1.0000
	20	0.0	0.3734	0.9672	0.5000	0.5000	1.0000	1.0000
		0.1	0.4841	1.1509	0.6932	0.6932	1.0000	1.0000
		0.2	0.6800	1.3333	1.0462	1.0462	1.0000	1.0000
		0.3	1.0126	1.4815	1.7021	1.7021	1.0000	1.0000
		0.4	1.5238	1.5484	2.9091	2.9091	1.0000	1.0000
		0.5	2.2264	1.5733	4.5385	4.5385	1.0000	1.0000
		0.7	3.6458	1.4831	6.4815	6.4815	1.0000	1.0000
		1.0	3.6220	1.3439	4.3044	4.3044	1.0000	1.0000
	50	0.0	0.1678	1.0000	0.2264	0.2264	1.0000	1.0000
		0.1	0.2364	1.0000	0.3377	0.3377	1.0000	1.0000
		0.2	0.3976	1.0000	0.6111	0.6111	1.0000	1.0000
		0.3	0.7377	1.0000	1.2857	1.2857	1.0000	1.0000
		0.4	1.3864	1.0000	2.7727	2.7727	1.0000	1.0000
		0.5	2.5938	1.0000	5.5333	5.5333	1.0000	1.0000
		0.7	5.3846	1.0072	9.3333	9.3333	1.0000	1.0000
		1.0	4.6607	1.0000	4.5790	4.5790	1.0000	1.0000

Table 7.2.2 Average MSE Efficiencies with respect to the logit procedure for d=5 doses.

D	N	γ	KERNEL	MRQR KERNEL	LLR	MRQR LLR	MRQR2 KERNEL	MRQR2 LLR
5	10	0.0	0.8333	1.0526	0.8955	1.1111	1.0000	1.0000
		0.1	0.8219	1.0526	0.9677	1.1539	1.0000	1.0000
		0.2	0.8378	1.0690	1.1273	1.2653	1.0000	1.0000
		0.3	0.8553	1.0656	1.3542	1.4444	1.0000	1.0000
		0.4	0.8734	1.0781	1.6429	1.6429	1.0000	1.0000
		0.5	0.9146	1.0870	1.8750	1.8750	1.0000	1.0000
		0.7	1.011	1.1463	2.0000	2.0000	1.0217	1.2368
		1.0	1.2069	1.2727	1.7284	1.7284	1.1765	1.5909
	20	0.0	0.7143	0.7895	0.7317	1.0000	1.0000	1.0000
		0.1	0.7143	1.0000	0.7500	1.0714	1.0000	1.0000
		0.2	0.7442	1.0000	0.8889	1.1852	1.0000	1.0000
		0.3	0.7955	1.0938	1.1290	1.4000	1.0000	1.0000
		0.4	0.8667	1.1143	1.5000	1.6957	1.0000	1.0833
		0.5	0.9375	1.1539	1.9565	2.0455	1.0227	1.2162
		0.7	1.1634	1.2800	2.2857	2.2857	1.1636	1.4884
		1.0	1.5857	1.6324	2.0556	2.0556	1.5417	2.0943
	50	0.0	0.5455	1.0000	0.5714	1.0000	1.0000	1.0000
		0.1	0.5714	1.0000	0.5714	1.0000	1.0000	1.0000
		0.2	0.6667	1.0769	0.7000	1.1667	1.0000	1.0000
		0.3	0.8095	1.2143	0.9444	1.4167	1.0625	1.1333
		0.4	0.9546	1.3125	1.4000	1.7500	1.1053	1.3125
		0.5	1.1250	1.3500	2.2500	2.4546	1.2273	1.5882
		0.7	1.6429	1.7037	2.7059	2.7059	1.6429	2.1905
		1.0	2.3846	2.3846	2.8182	2.8182	2.4474	3.1000

Table 7.2.3 Average MSE Efficiencies with respect to the logit procedure for d=7 doses.

D	N	γ	KERNEL	MRQR KERNEL	LLR	MRQR LLR	MRQR2 KERNEL	MRQR2 LLR
7	10	0.0	0.7736	1.0000	0.8367	1.0790	1.0000	1.0000
		0.1	0.7593	1.0000	0.8913	1.1081	1.0000	1.0000
		0.2	0.7818	1.0000	1.0488	1.2286	1.0000	1.0000
		0.3	0.8070	1.0455	1.2778	1.3939	1.0000	1.0000
		0.4	0.8644	1.0851	1.5938	1.5938	1.0000	1.0000
		0.5	0.9194	1.0962	1.9000	1.8387	1.0000	1.0000
		0.7	1.0556	1.1875	2.0541	2.0541	1.0857	1.2258
		1.0	1.3596	1.3908	1.7794	1.7794	1.3297	1.6351
	20	0.0	0.6452	1.0000	0.6897	1.0000	1.0000	1.0000
		0.1	0.6774	1.0000	0.7500	1.0500	1.0000	1.0000
		0.2	0.7097	1.0000	0.8462	1.1579	1.0000	1.0000
		0.3	0.7813	1.0870	1.1364	1.3158	1.0000	1.0000
		0.4	0.9091	1.1538	1.5789	1.6667	1.0345	1.1111
		0.5	1.0000	1.2000	2.2500	2.1176	1.0909	1.2000
		0.7	1.2857	1.3846	2.3478	2.3478	1.2857	1.5429
		1.0	1.8868	1.9231	2.2727	2.2727	1.8519	2.3256
	50	0.0	0.5333	1.0000	0.5333	1.0000	1.0000	1.0000
		0.1	0.6000	1.0000	0.6429	1.0000	1.0000	1.0000
		0.2	0.6667	1.0000	0.7692	1.1111	1.0000	1.0000
		0.3	0.8667	1.1818	1.0833	1.6250	1.0833	1.1818
		0.4	1.0625	1.3077	1.7000	1.8889	1.2143	1.3077
		0.5	1.3529	1.5333	2.8750	2.5556	1.3529	1.6429
		0.7	2.1000	2.1000	3.0000	3.0000	2.0000	2.4706
		1.0	3.2222	3.2222	3.6250	3.6250	3.2222	3.9546

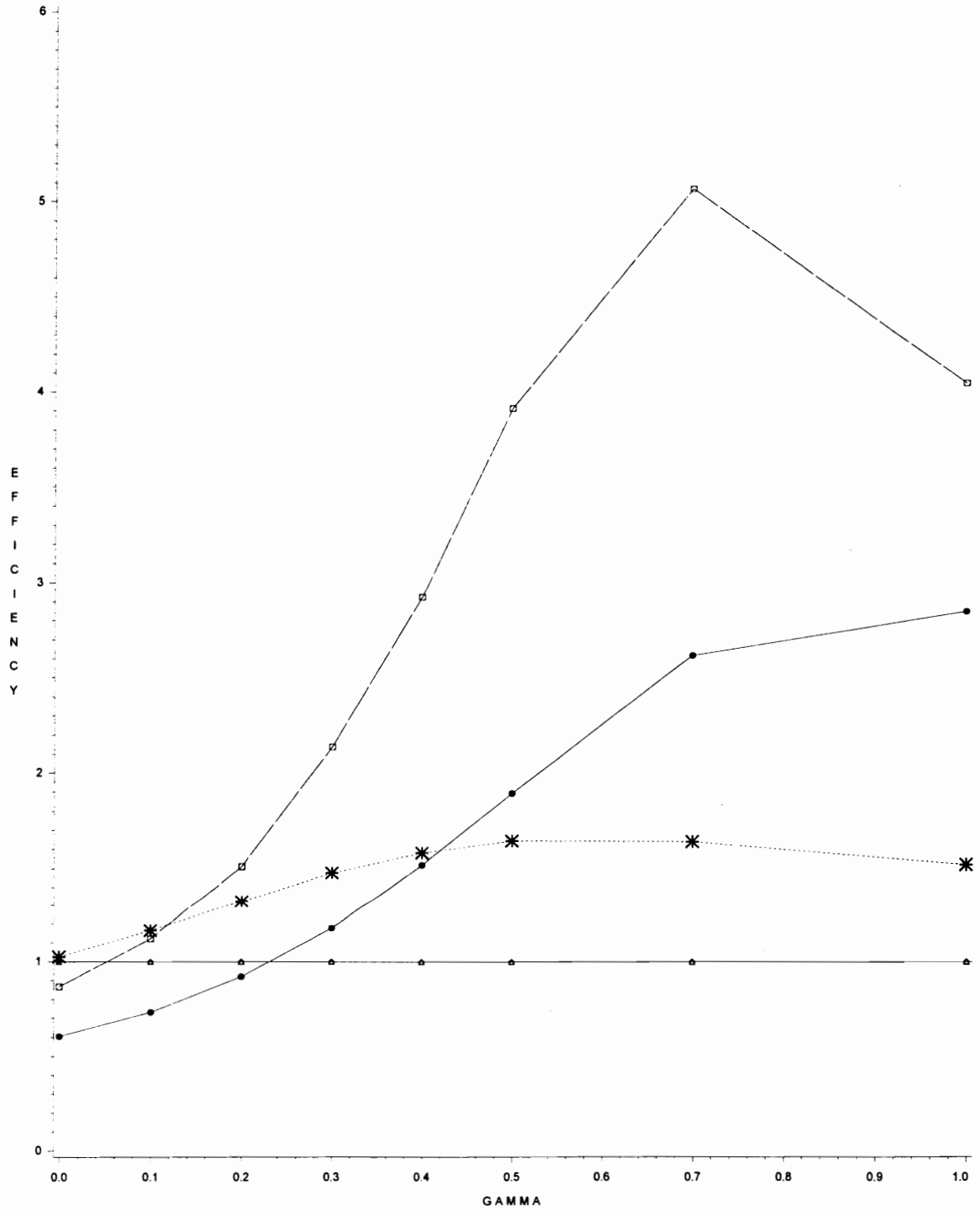


Figure 7.2.2 Average MSE Efficiency Plot for (d=3, n=10) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

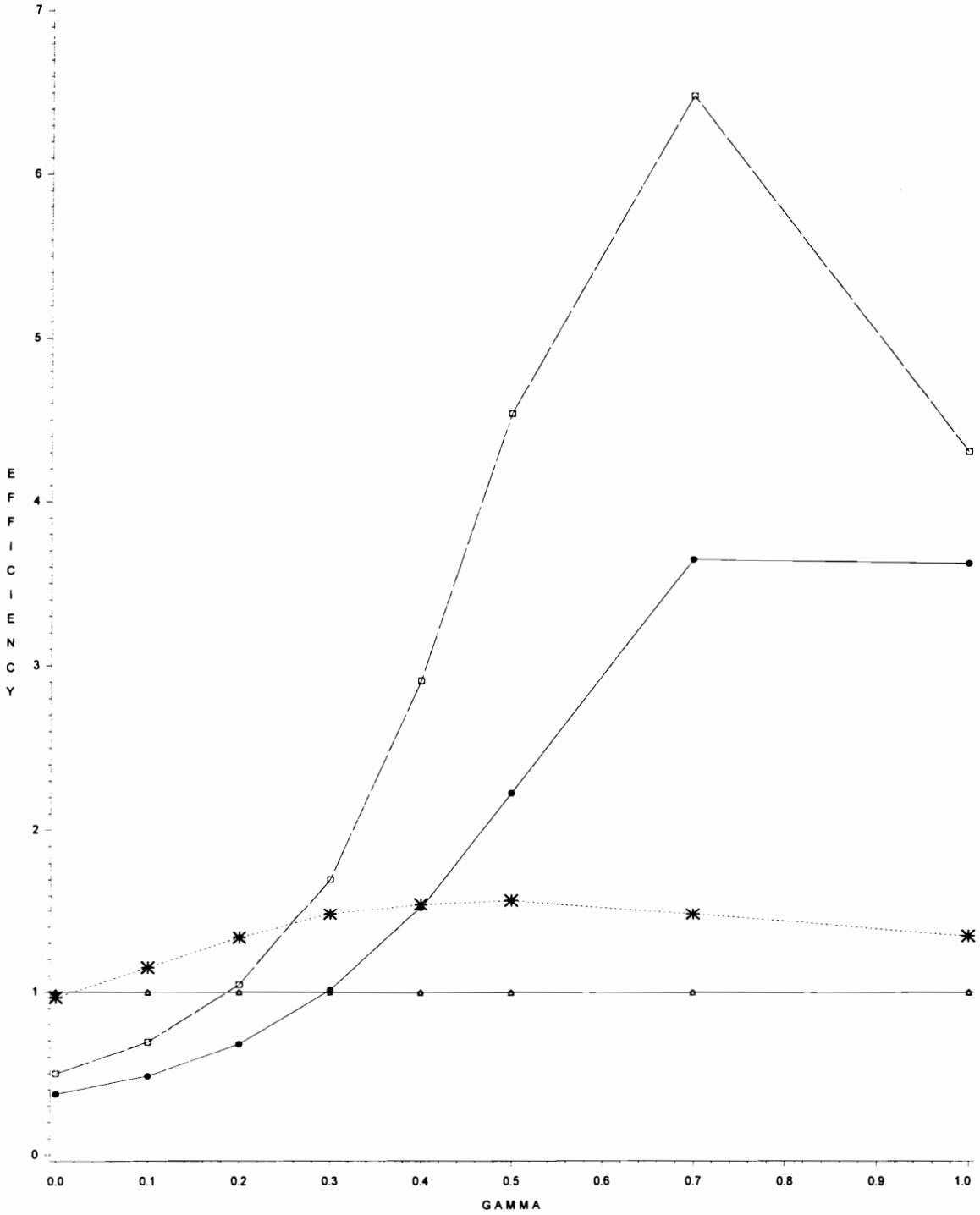


Figure 7.2.3 Average MSE Efficiency Plot for (d=3, n=20) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

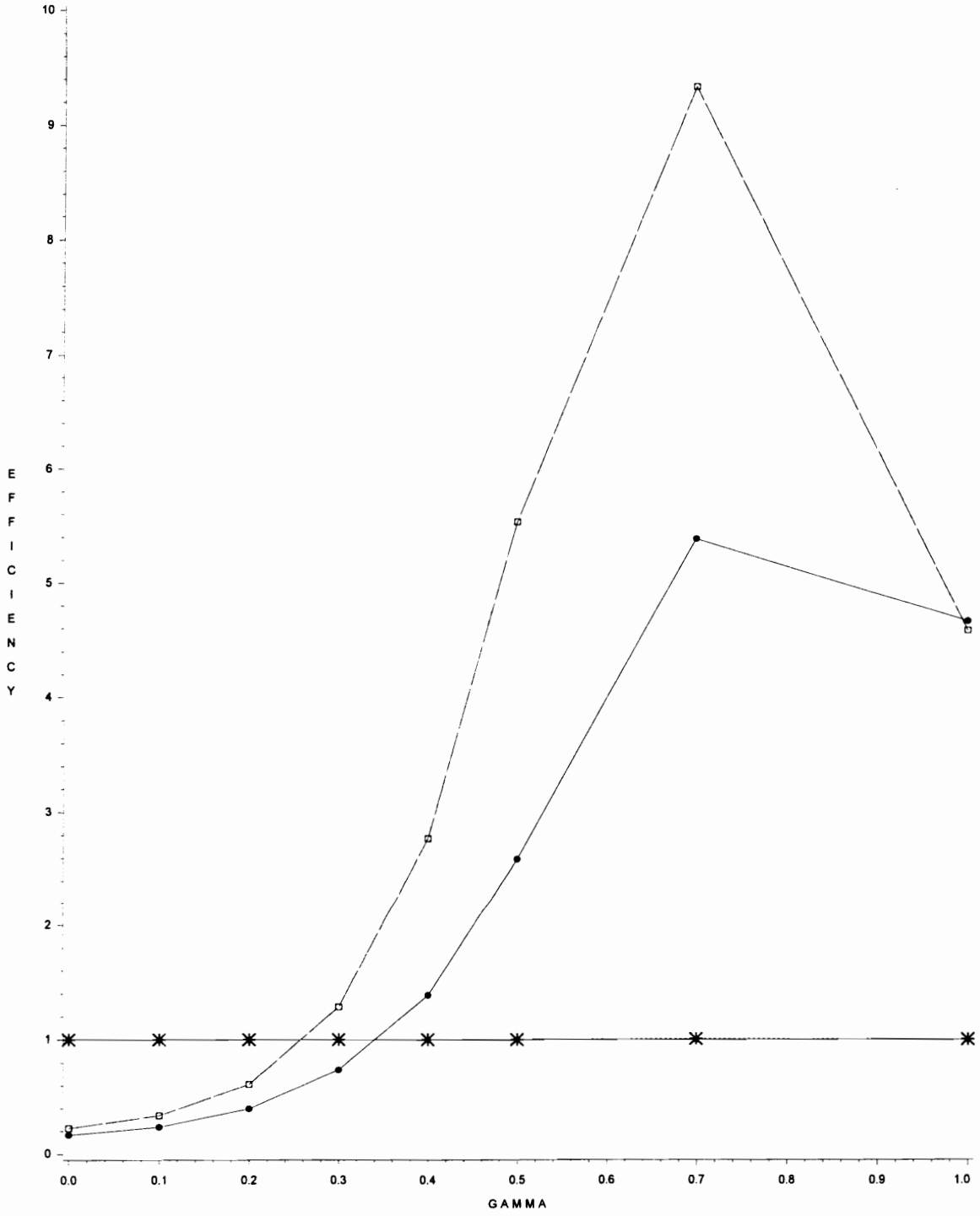


Figure 7.2.4 Average MSE Efficiency Plot for (d=3, n=50) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

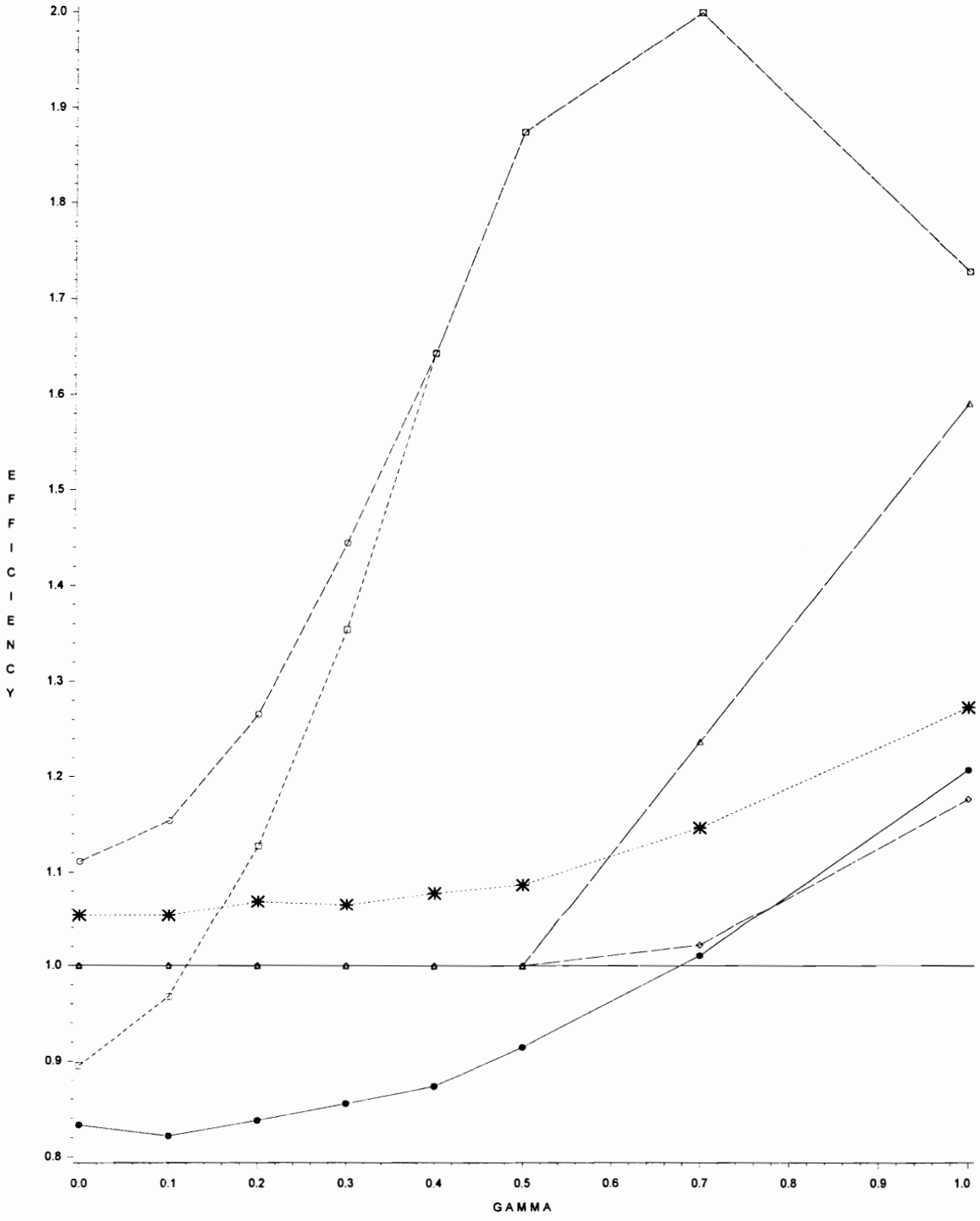


Figure 7.2.5 Average MSE Efficiency Plot for (d=5, n=10) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

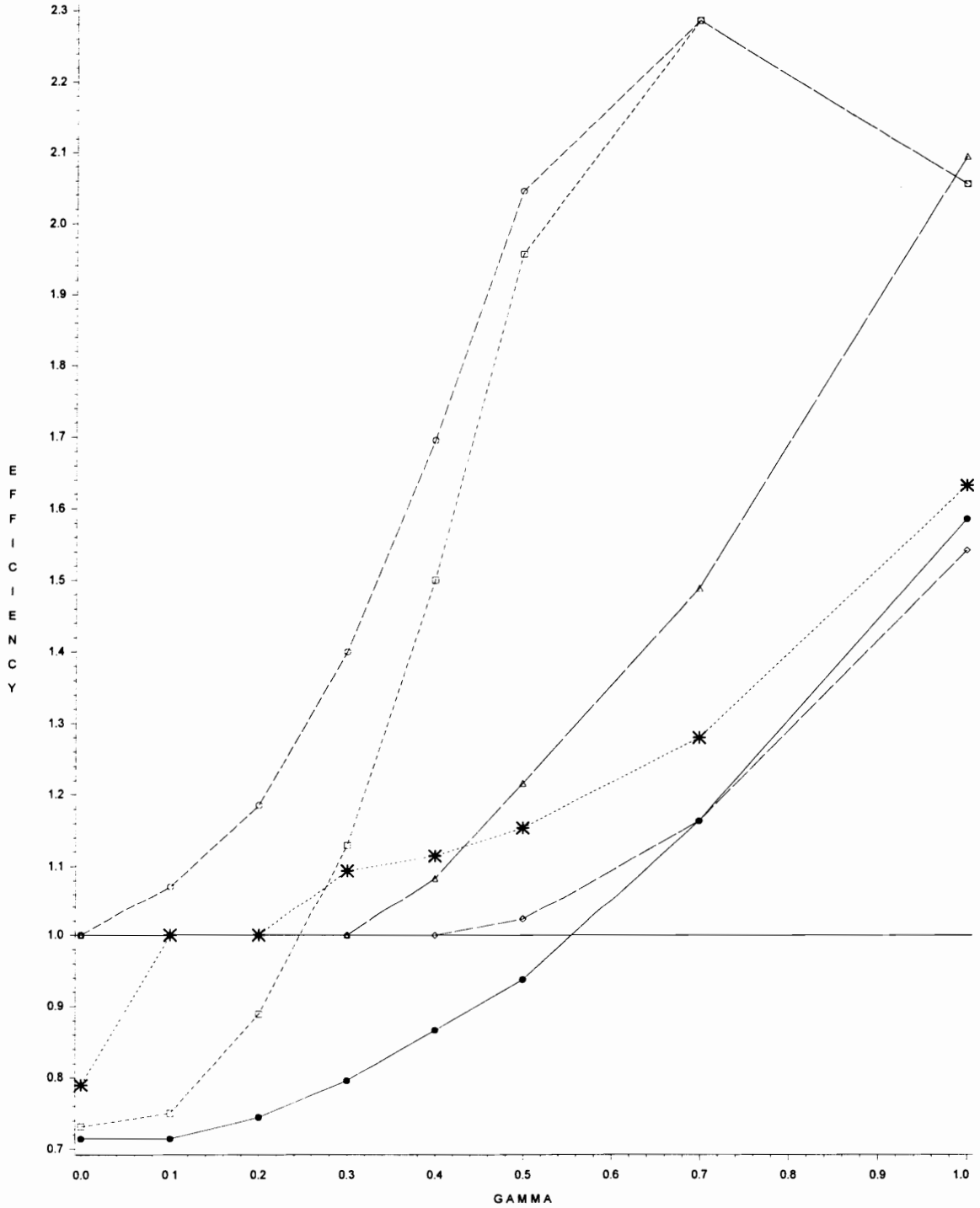


Figure 7.2.6 Average MSE Efficiency Plot for (d=5, n=20) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

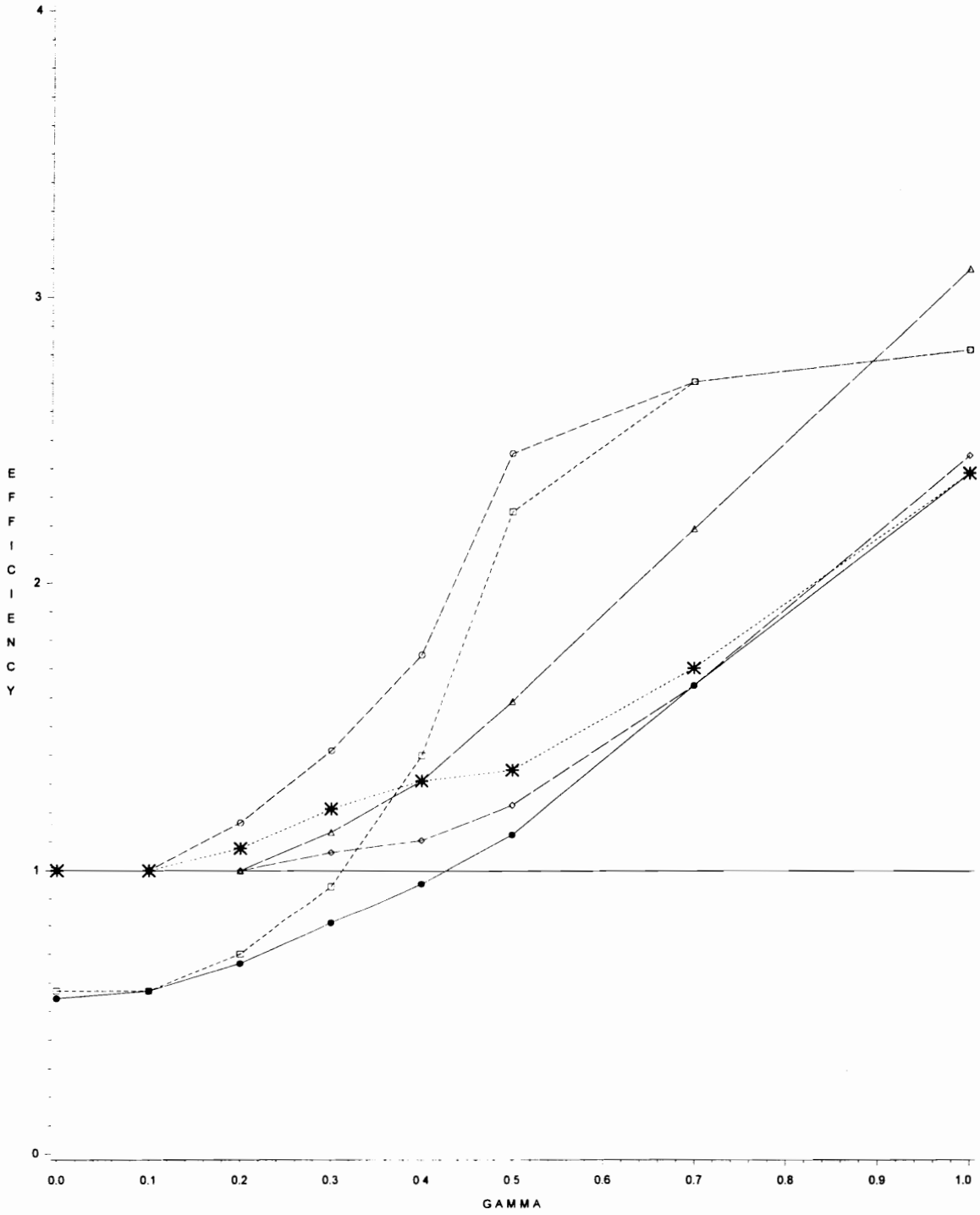


Figure 7.2.7 Average MSE Efficiency Plot for (d=5, n=50) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

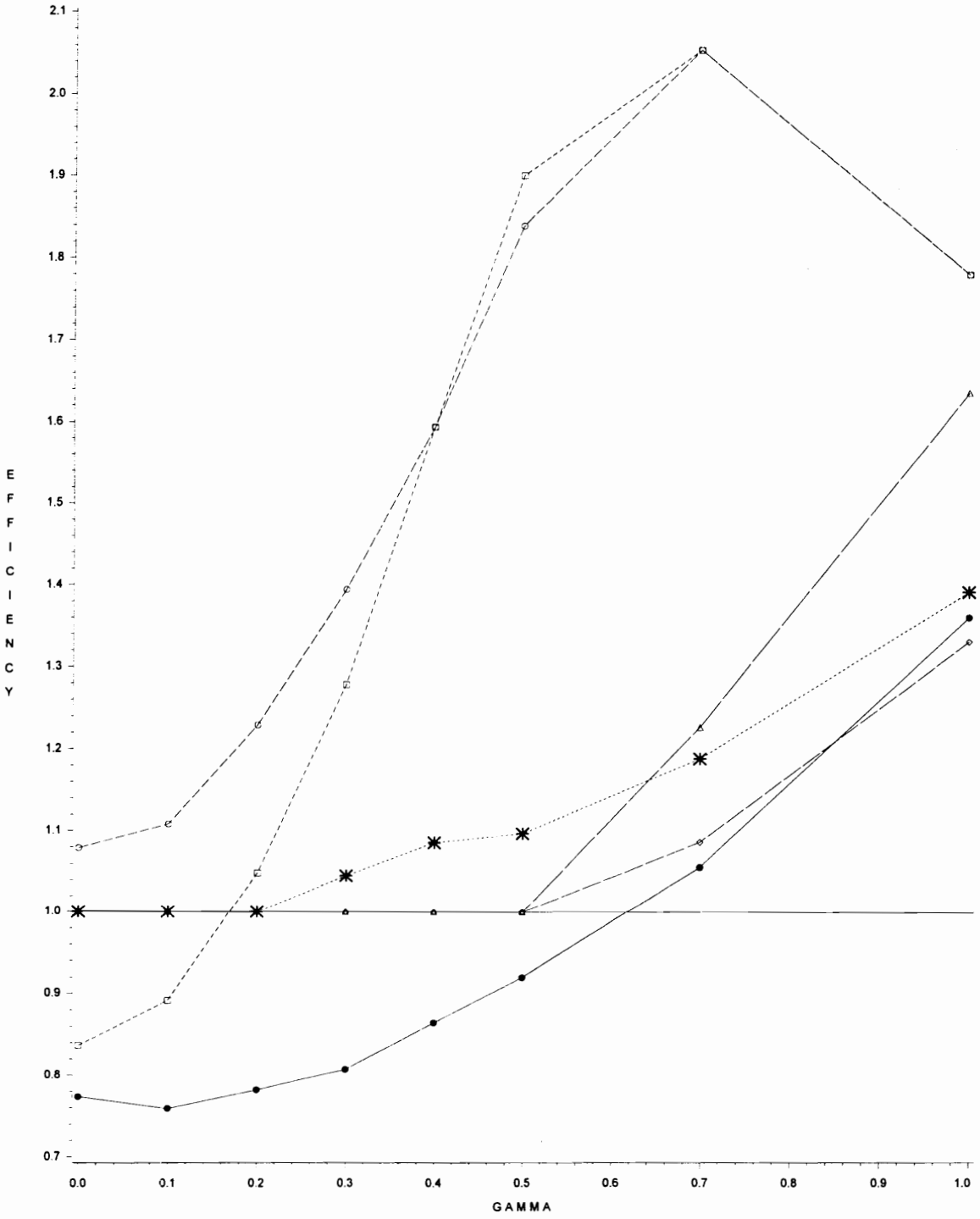


Figure 7.2.8 Average MSE Efficiency Plot for (d=7, n=10) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

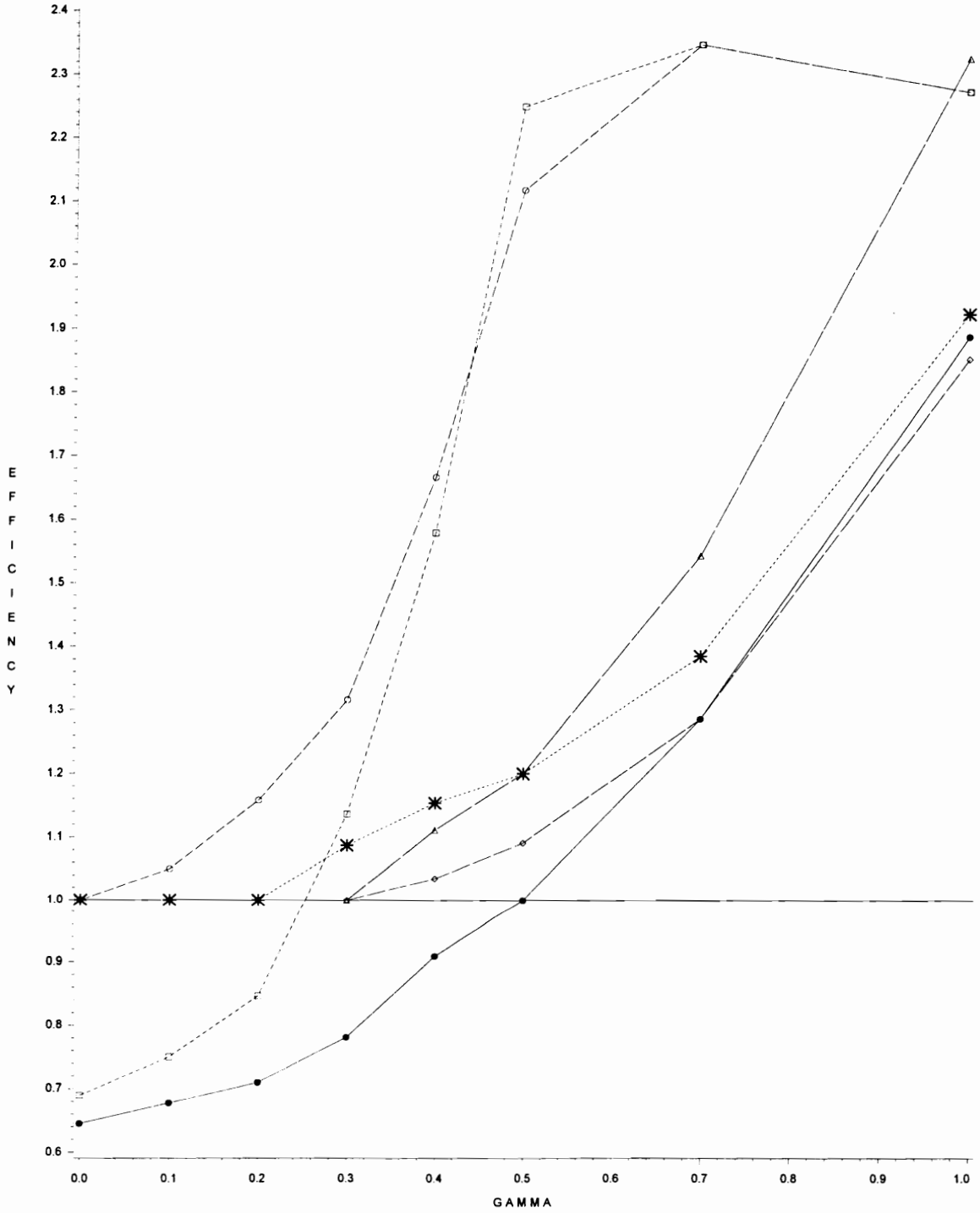


Figure 7.2.9 Average MSE Efficiency Plot for (d=7, n=20) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

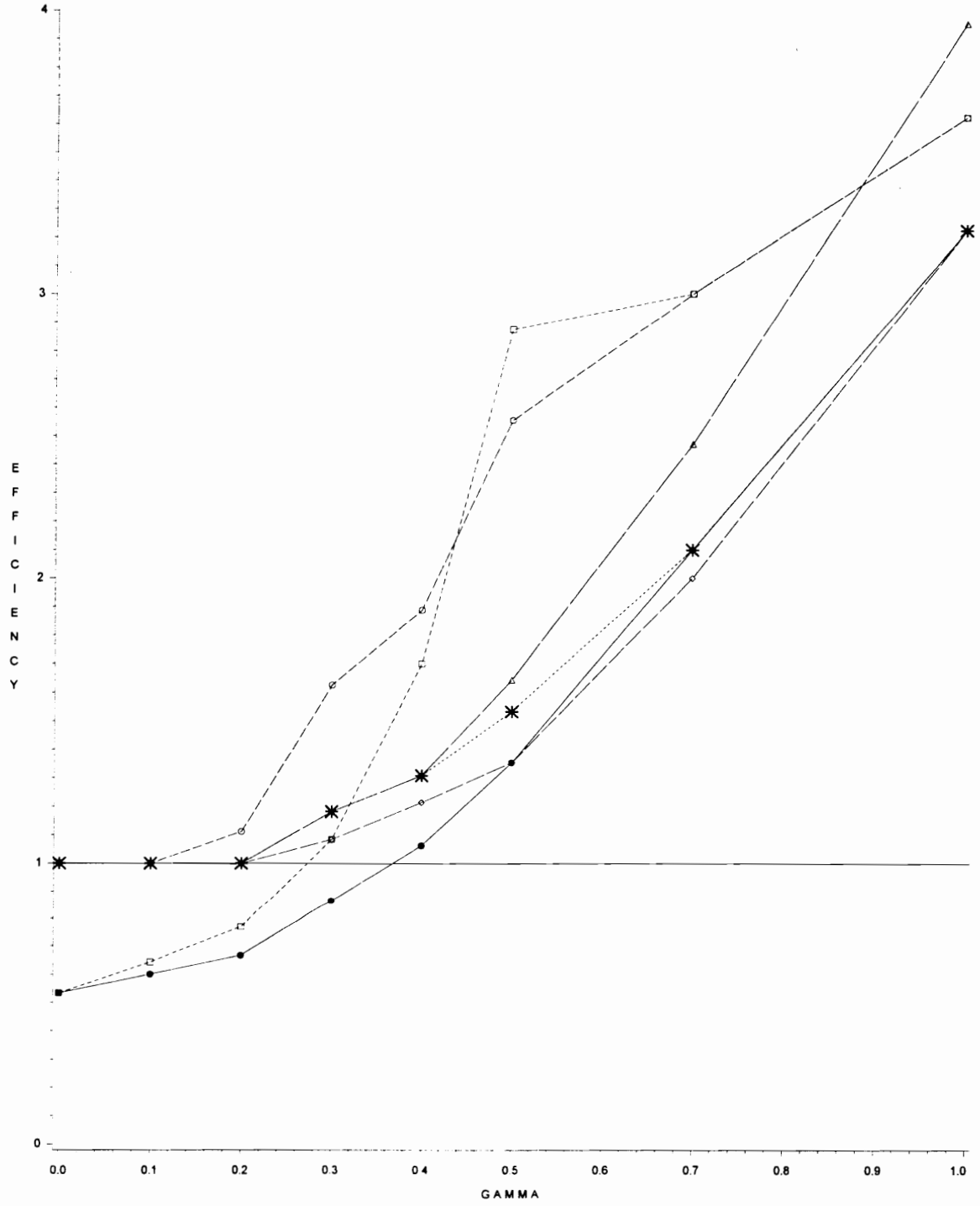


Figure 7.2.10 Average MSE Efficiency Plot for (d=7, n=50) Combination.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

◊ ◊ ◊ MRQR2 Kernel, Δ Δ Δ MRQR2 LLR

§7.3 Optimal Bandwidth and Mixing Parameters

Another parameter of interest when using the model robust regression procedures is the mixing parameter λ. As indicated earlier in this research, if the user's model is correctly specified, then it is desired that λ will be approximately zero, thus giving much of the weight to the parametric fit. On the other hand, if the user's model has been misspecified, then as the degree of misspecification increases, so should λ. Also, recall that the most important parameter when using nonparametric regression, in particular, kernel and local linear regression, is the bandwidth or smoothing parameter. The size of the bandwidth determines the smoothness of the regression function. **Tables 7.2.4-7.2.6** represent the optimal values of the bandwidth and mixing parameter for all the (d, n, γ) combinations.

Some of the key points are noted below with respect to the mixing parameter for the model-robust procedures.

1. At the smallest dose (d=3), the only procedure that mixes the parametric and the nonparametric methods is the MRQR Kernel at sample sizes n=10 and n=20. The MRQR LLR procedure chooses λ=1 for all (d=3, n, γ) combinations and the MRQR2 procedures yields λ=0 for all (d=3, n, γ) combinations. It is interesting to note that for MRQR Kernel and both (d=3, n=10) and (d=3, n=20) combinations, as γ increases (more model-misspecification) the mixing parameter λ decreases (less mixing). This is perhaps due to the lack of information gained when using kernel regression with three doses, especially since it will be heavily biased in the boundaries. Also, at the (d=3, n=50) combination, for the model-robust procedures other than MRQR LLR the mixing parameter is zero for all values of γ.

2. When increasing the number of doses to five, the model-robust procedures tend to mix more of the nonparametric methods as the degree of model-misspecification increases. In particular, for the (d=5, n=10) scenario, the MRQR procedures mix when the user's model has been correctly specified, resulting in smaller amses than MLE, as seen in **Table 7.2.2**.

3. As the degree of model-misspecification increases, the value of the mixing parameter associated with the MRQR LLR procedure increases at a faster rate than the MRQR Kernel and the MRQR2 procedures (see **Figure 7.2.10**). The MRQR2 procedures begin to mix (i.e., find structure in the residuals) when the value of γ is 0.5.

4. Increasing the number of subjects from n=10 to n=20 tends to decrease the rate at which the mixing parameter increases. That is, for the correctly specified model or slightly misspecified, the value of the mixing parameter is near zero. For the MRQR Kernel procedure, the model-misspecification is not detected until γ=0.3, whereas with the MRQR LLR procedure, some mixing is involved at γ=0.1.

5. The MRQR2 procedures begin to detect some "significant" structure in the residuals at γ=0.4. **Figure 7.2.11** is a plot of the mixing parameter versus the degree of misspecification for the (d=5, n=20) combination.

6. For d=7, if the number of subjects increases from 20 to 50, the MRQR Kernel and the MRQR LLR procedure begins to mix at γ=0.2. Note that for (d=5, n=20), the MRQR LLR procedure began mixing at γ=0.1. Thus, the increase in the number of subjects at each dose enabled the parametric procedure (logit analysis) to fit quite well at slight model-misspecification. The MRQR2 procedures found some structure in the residuals at γ=0.3 (see **Figure 7.2.12**). The same observations can be made for the (d=7, n, γ) combinations (see Figures **7.2.13-7.2.15**).

The obvious trend noticeable in examining the mixing parameters is that as the degree of model-misspecification increases, the value of the mixing parameter increases. Also of note is that as the sample size (number of subjects at each dose) increases from 10, 20, to 50, the MRQR LLR procedure takes longer to detect misspecification (i.e., yield a nonzero mixing parameter). For example, with the d=5 case, when n=10 the MRQR LLR mixes at γ=0, then for n=20 MRQR LLR mixes at γ=0.1, and finally at n=50, the procedure mixes at γ=0.2. The MRQR2 procedures, on the other hand, tend to mix sooner as the number of subjects at each dose increases.

A review of the bandwidth, or smoothing parameter, yields some very interesting observations. Some observations with respect to the bandwidth for the kernel and local linear regression procedures are given below.

1. For the kernel regression procedure, the bandwidth stays relatively constant across all values of γ for fixed (d, n) combinations. This is also the case for the LLR procedure, but only for the three dose-level situation. This is perhaps due to the fact that the bandwidth is a function of the distance between the dose levels moreso that it is a function of the responses, P. That is, as γ increases, the distance between the dose levels remains the same for fixed (d, n) combinations although the responses do change slightly.

2. For each (d, n, γ) combination, the bandwidth for kernel regression is always smaller than the bandwidth for the LLR procedure. This seems to imply that the size of the bandwidth plays is not as important in LLR as it is in kernel regression. Note that in kernel regression, a large bandwidth, say approximately one, will give a fitted line through the mean of the responses. Recall that the LLR procedure fits a series of regression lines via weighted least squares. Therefore, for the LLR procedure, a bandwidth near one implies that a single regression line will be fit through the entire data, rather than a series of lines.

3. Another observation concerning the bandwidth is that for both kernel and LLR, the bandwidth gets smaller as the number of dose levels increases, and as the number of replications, n_i, at each dose increases. This seems to confirm the notion that the more data points involved in the analysis, the closer the points will be and thus a smaller bandwidth is needed to fit the data. In terms of the bandwidth decreasing as n_i increases, the decrease is very subtle in most instances, but it is a reflection of having more information available to fit the data.

Table 7.3.1 Optimal values of the bandwidth and mixing parameter for the model-robust

			MRQR KERNEL		MRQR LLR		MRQR2 KERNEL		MRQR2 LLR	
D	N	γ	b	λ	b	λ	b	λ	b	λ
3	10	0.0	0.2540	0.4553	0.6310	1.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.2541	0.4477	0.6319	1.0000	1.0000	0.0000	1.0000	0.0000
		0.2	0.2544	0.4402	0.6328	1.0000	1.0000	0.0000	1.0000	0.0000
		0.3	0.2546	0.4330	0.6337	1.0000	1.0000	0.0000	1.0000	0.0000
		0.4	0.2547	0.4263	0.6346	1.0000	1.0000	0.0000	1.0000	0.0000
		0.5	0.2549	0.4197	0.6355	1.0000	1.0000	0.0000	1.0000	0.0000
		0.7	0.2553	0.4070	0.6373	1.0000	1.0000	0.0000	1.0000	0.0000
		1.0	0.2559	0.3898	0.6400	1.0000	1.0000	0.0000	1.0000	0.0000
	20	0.0	0.2295	0.3062	0.6304	1.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.2297	0.2994	0.6319	1.0000	1.0000	0.0000	1.0000	0.0000
		0.2	0.2299	0.2935	0.6329	1.0000	1.0000	0.0000	1.0000	0.0000
		0.3	0.2300	0.2877	0.6336	1.0000	1.0000	0.0000	1.0000	0.0000
		0.4	0.2302	0.2822	0.6346	1.0000	1.0000	0.0000	1.0000	0.0000
		0.5	0.2303	0.2768	0.6355	1.0000	1.0000	0.0000	1.0000	0.0000
		0.7	0.2306	0.2669	0.6373	1.0000	1.0000	0.0000	1.0000	0.0000
		1.0	0.2311	0.2534	0.6400	1.0000	1.0000	0.0000	1.0000	0.0000
	50	0.0	0.2309	0.0000	0.6313	1.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.2039	0.0000	0.6319	1.0000	1.0000	0.0000	1.0000	0.0000
		0.2	0.2041	0.0000	0.6328	1.0000	1.0000	0.0000	1.0000	0.0000
		0.3	0.2042	0.0000	0.6337	1.0000	1.0000	0.0000	1.0000	0.0000
		0.4	0.2043	0.0000	0.6346	1.0000	1.0000	0.0000	1.0000	0.0000
		0.5	0.2044	0.0000	0.6356	1.0000	1.0000	0.0000	1.0000	0.0000
		0.7	0.2046	0.0000	0.6373	1.0000	1.0000	0.0000	1.0000	0.0000
		1.0	0.2050	0.0000	0.6401	1.0000	1.0000	0.0000	1.0000	0.0000

Table 7.3.2 Optimal values of the bandwidth and mixing parameter for the model-robust

			MRQR KERNEL		MRQR LLR		MRQR2 KERNEL		MRQR2 LLR	
D	N	γ	b	λ	b	λ	b	λ	b	λ
5	10	0.0	0.1754	0.2594	0.2087	0.3496	1.0000	0.0000	1.0000	0.0000
		0.1	0.1817	0.2622	0.2468	0.4345	1.0000	0.0000	1.0000	0.0000
		0.2	0.1862	0.2765	0.2942	0.5356	1.0000	0.0000	1.0000	0.0000
		0.3	0.1888	0.3011	0.3589	0.6480	1.0000	0.0000	1.0000	0.0000
		0.4	0.1896	0.3366	0.4513	0.7811	1.000	0.0000	1.0000	0.0000
		0.5	0.1885	0.3851	0.5220	0.9496	0.2607	0.0960	1.0000	0.0000
		0.7	0.1815	0.5235	0.4714	1.0000	0.2058	0.5497	0.2372	0.7020
		1.0	0.1647	0.7619	0.2848	1.0000	0.1724	0.8318	0.1906	0.8563
	20	0.0	0.1547	0.0000	0.1665	0.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.1619	0.0000	0.1920	0.3063	1.0000	0.0000	1.0000	0.0000
		0.2	0.1668	0.0000	0.2325	0.4358	1.0000	0.0000	1.0000	0.0000
		0.3	0.1695	0.3108	0.2899	0.5698	1.0000	0.0000	1.0000	0.0000
		0.4	0.1699	0.3886	0.3792	0.7118	0.2362	0.3504	0.2668	0.5331
		0.5	0.1682	0.4796	0.5038	0.8904	0.2035	0.5857	0.2338	0.6974
		0.7	0.1601	0.6799	0.4255	1.0000	0.1739	0.8288	0.1925	0.8525
		1.0	0.1445	0.8912	0.2327	1.0000	0.1485	0.9510	0.1558	0.9361
	50	0.0	0.1302	0.0000	0.1339	0.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.1371	0.0000	0.1482	0.0000	1.0000	0.0000	1.0000	0.0000
		0.2	0.1426	0.2659	0.1707	0.3794	1.000	0.0000	1.0000	0.0000
		0.3	0.1457	0.3908	0.2133	0.5351	0.2127	0.5854	0.2396	0.6608
		0.4	0.1462	0.5208	0.2912	0.6794	0.1831	0.7911	0.2043	0.8192
		0.5	0.1444	0.6480	0.4510	0.8486	0.1653	0.8936	0.1809	0.8935
		0.7	0.1367	0.8472	0.3477	1.0000	0.1433	0.9693	0.1529	0.9563
		1.0	0.1238	0.9685	0.1732	1.0000	0.1251	0.9941	0.1308	0.9827

Table 7.3.3 Optimal values of the bandwidth and mixing parameter for the model-robust

			MRQR KERNEL		MRQR LLR		MRQR2 KERNEL		MRQR2 LLR	
D	N	γ	b	λ	b	λ	b	λ	b	λ
7	10	0.0	0.1511	0.0000	0.1781	0.2877	1.0000	0.0000	1.0000	0.0000
		0.1	0.1566	0.0000	0.2089	0.3619	1.0000	0.0000	1.0000	0.0000
		0.2	0.1606	0.0000	0.2515	0.4595	1.0000	0.0000	1.0000	0.0000
		0.3	0.1629	0.2541	0.3051	0.5642	1.0000	0.0000	1.0000	0.0000
		0.4	0.1635	0.2924	0.3922	0.6819	1.0000	0.0000	1.0000	0.0000
		0.5	0.1622	0.3487	0.5301	0.8313	0.2591	0.1703	1.0000	0.0000
		0.7	0.1549	0.5026	0.5904	1.0000	0.1876	0.5927	0.2015	0.6421
		1.0	0.1389	0.7626	0.2532	1.0000	0.1497	0.8581	0.1610	0.8493
	20	0.0	0.1335	0.0000	0.1454	0.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.1389	0.0000	0.1642	0.2591	1.0000	0.0000	1.0000	0.0000
		0.2	0.1426	0.0000	0.1933	0.3799	1.0000	0.0000	1.0000	0.0000
		0.3	0.1446	0.2817	0.2401	0.5089	1.0000	0.0000	1.0000	0.0000
		0.4	0.1445	0.3636	0.3099	0.6368	0.2300	0.3923	0.2605	0.4776
		0.5	0.1426	0.4608	0.4605	0.7862	0.1892	0.6097	0.2021	0.6332
		0.7	0.1349	0.6755	0.5612	1.0000	0.1527	0.8456	0.1635	0.8313
		1.0	0.1209	0.9154	0.7799	1.0000	0.1280	1.0000	0.1358	0.9698
	50	0.0	0.1135	0.0000	0.1179	0.0000	1.0000	0.0000	1.0000	0.0000
		0.1	0.1191	0.0000	0.1303	0.0000	1.0000	0.0000	1.0000	0.0000
		0.2	0.1230	0.0000	0.1471	0.3506	1.0000	0.0000	1.0000	0.0000
		0.3	0.1247	0.3719	0.1739	0.5034	0.2016	0.5888	0.2156	0.5880
		0.4	0.1243	0.5059	0.2242	0.6392	0.1648	0.7882	0.1759	0.7667
		0.5	0.1221	0.6414	0.3348	0.7705	0.1457	0.9096	0.1557	0.8793
		0.7	0.1143	0.8633	0.4528	1.0000	0.1243	1.0000	0.1314	0.9908
		1.0	0.1014	1.0000	0.1256	1.0000	0.1057	1.0000	0.1101	1.0000

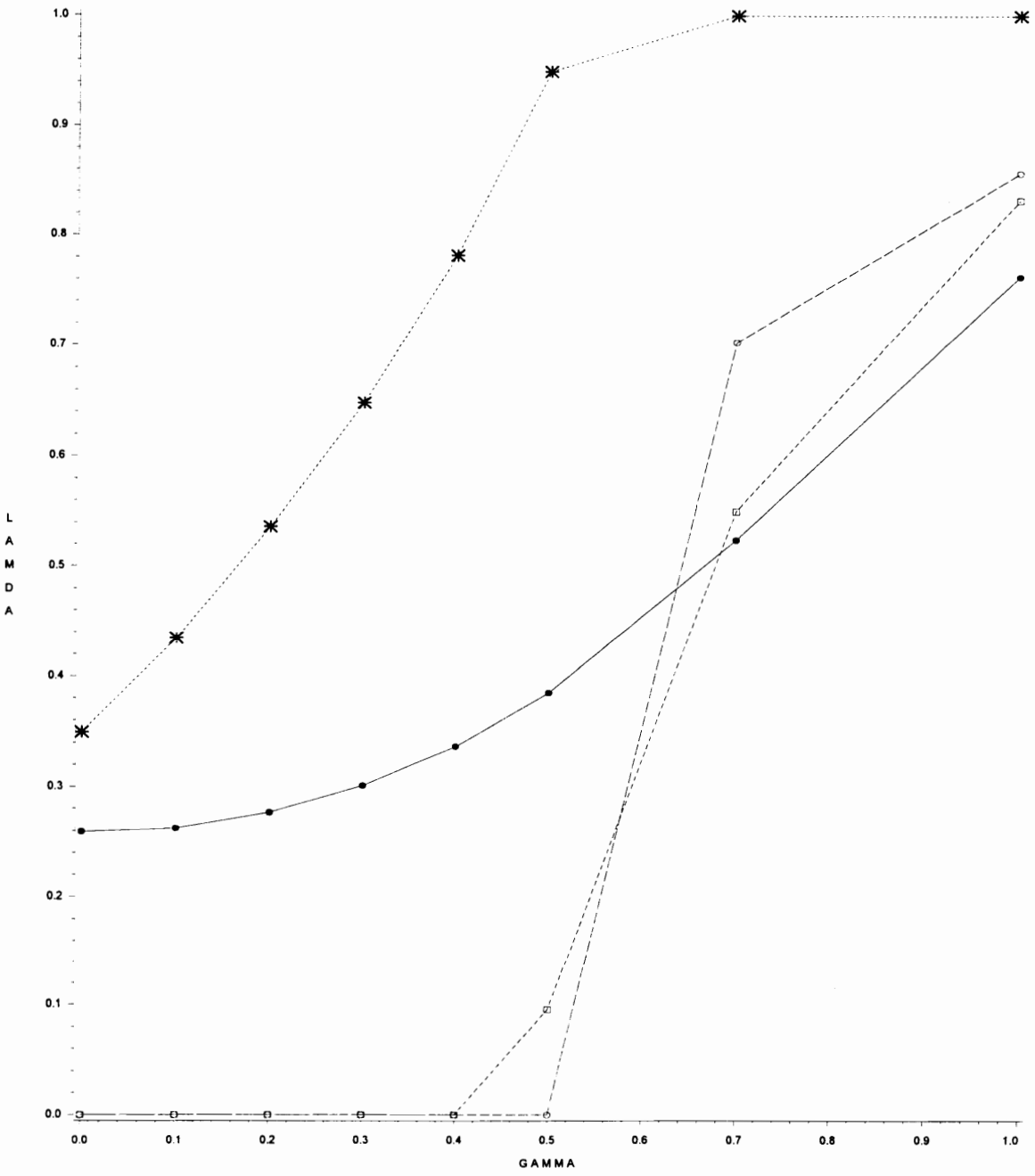


Figure 7.3.1 Mixing Parameter vs. γ Plots for (d=5, n=10).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

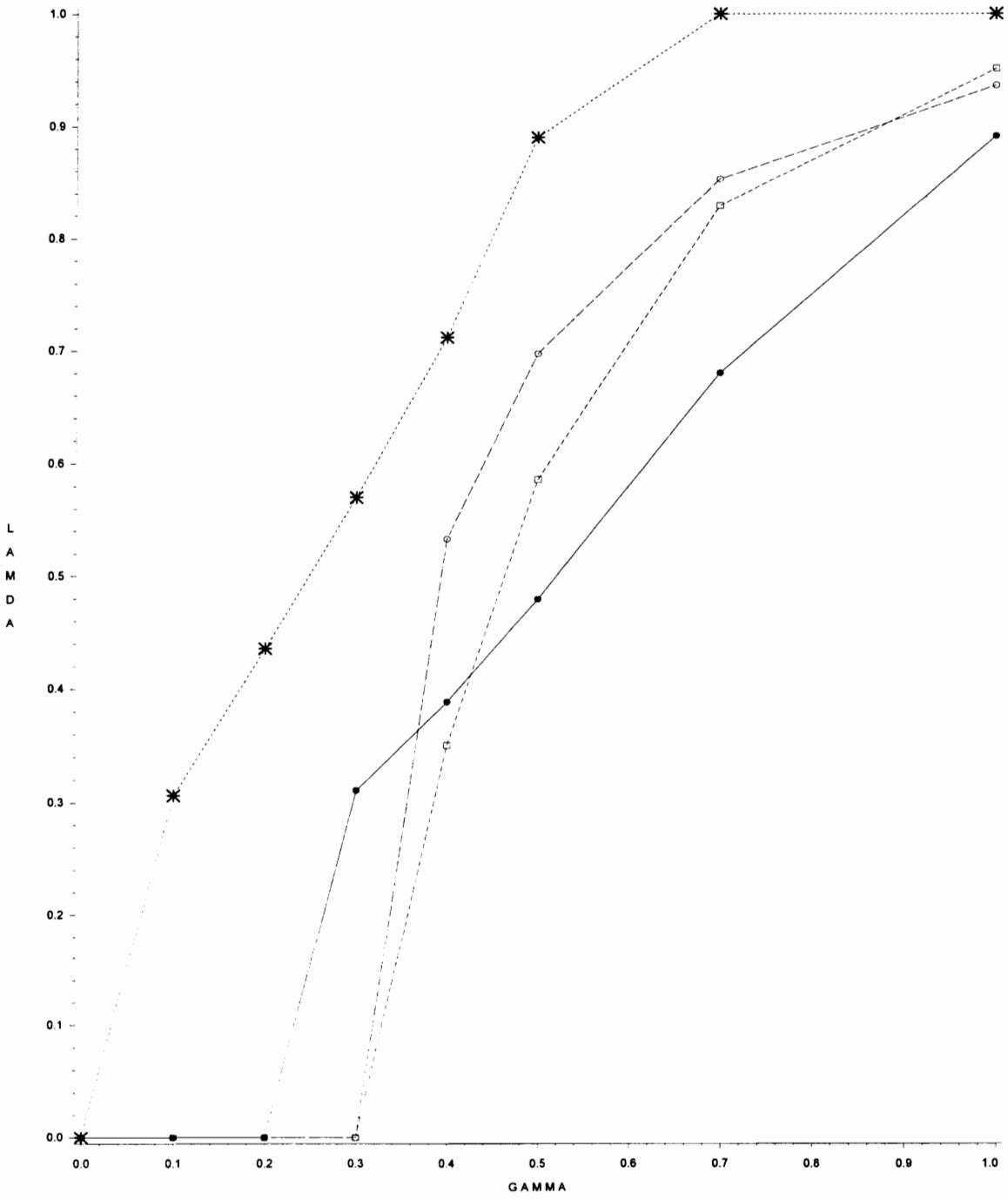


Figure 7.3.2 Mixing Parameter vs. γ Plots for (d=5, n=20).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

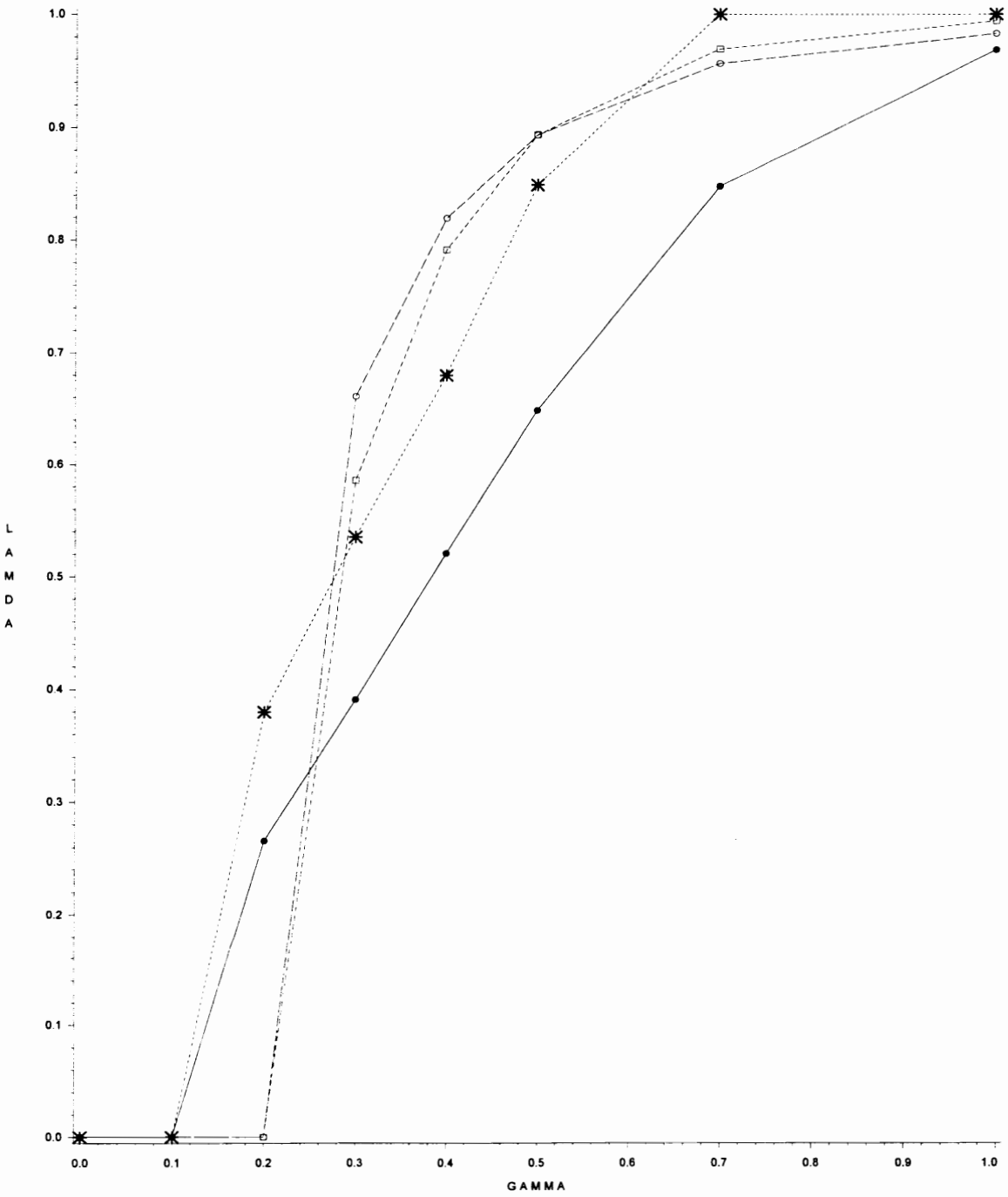


Figure 7.3.3 Mixing Parameter vs. γ Plots for (d=5, n=50).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

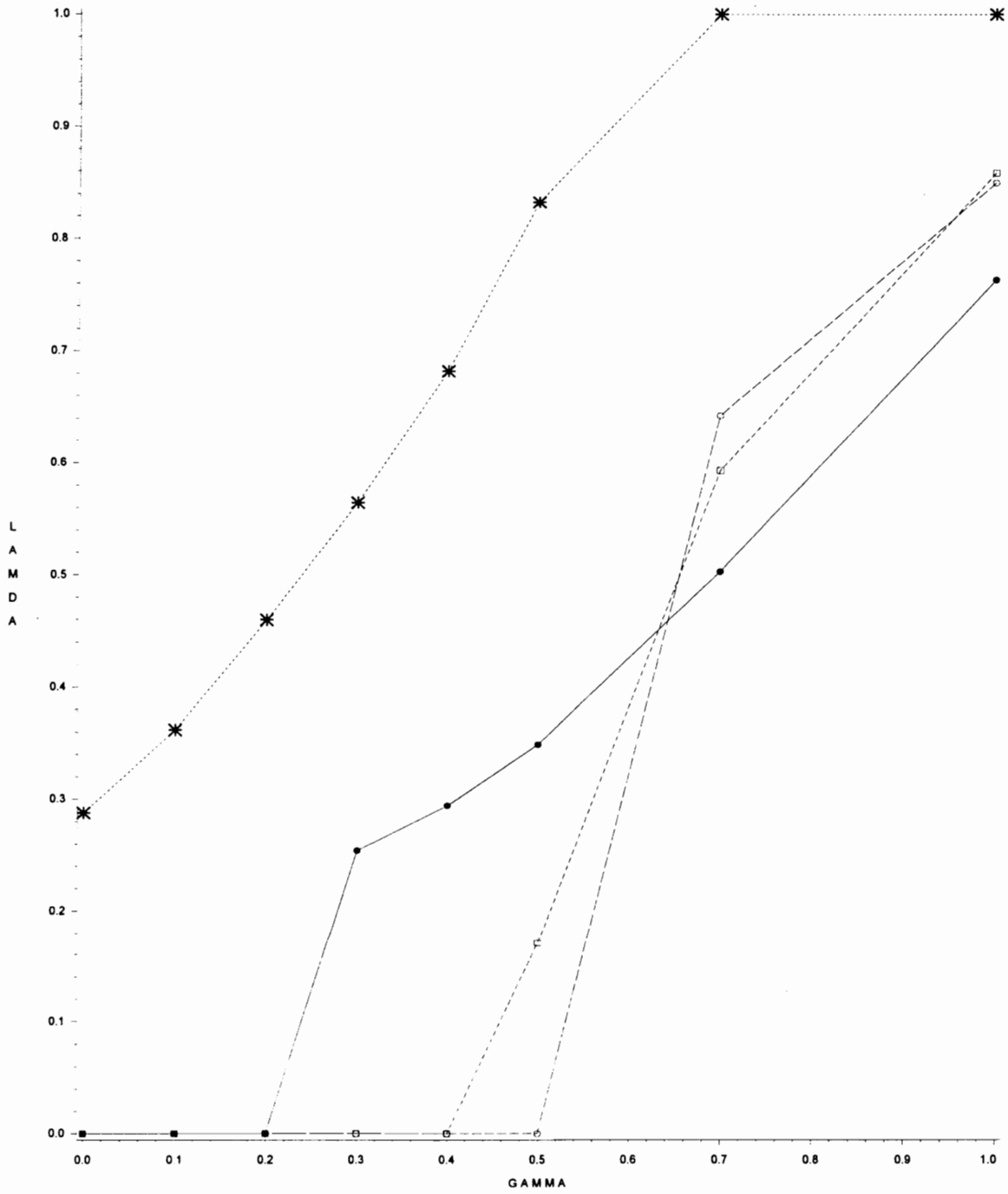


Figure 7.3.4 Mixing Parameter vs. γ Plots for (d=7, n=10).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

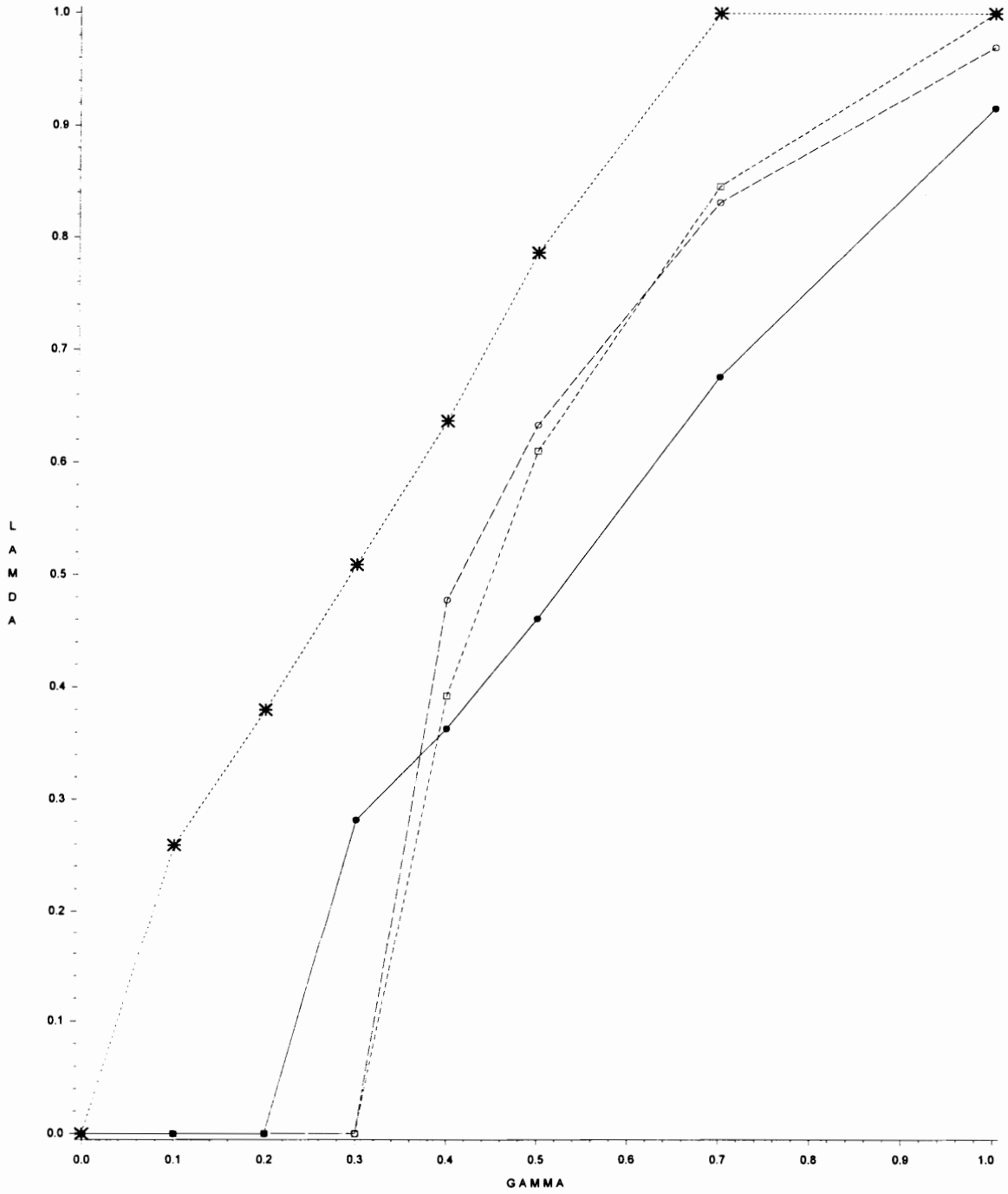


Figure 7.3.5 Mixing Parameter vs. γ Plots for (d=7, n=20).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

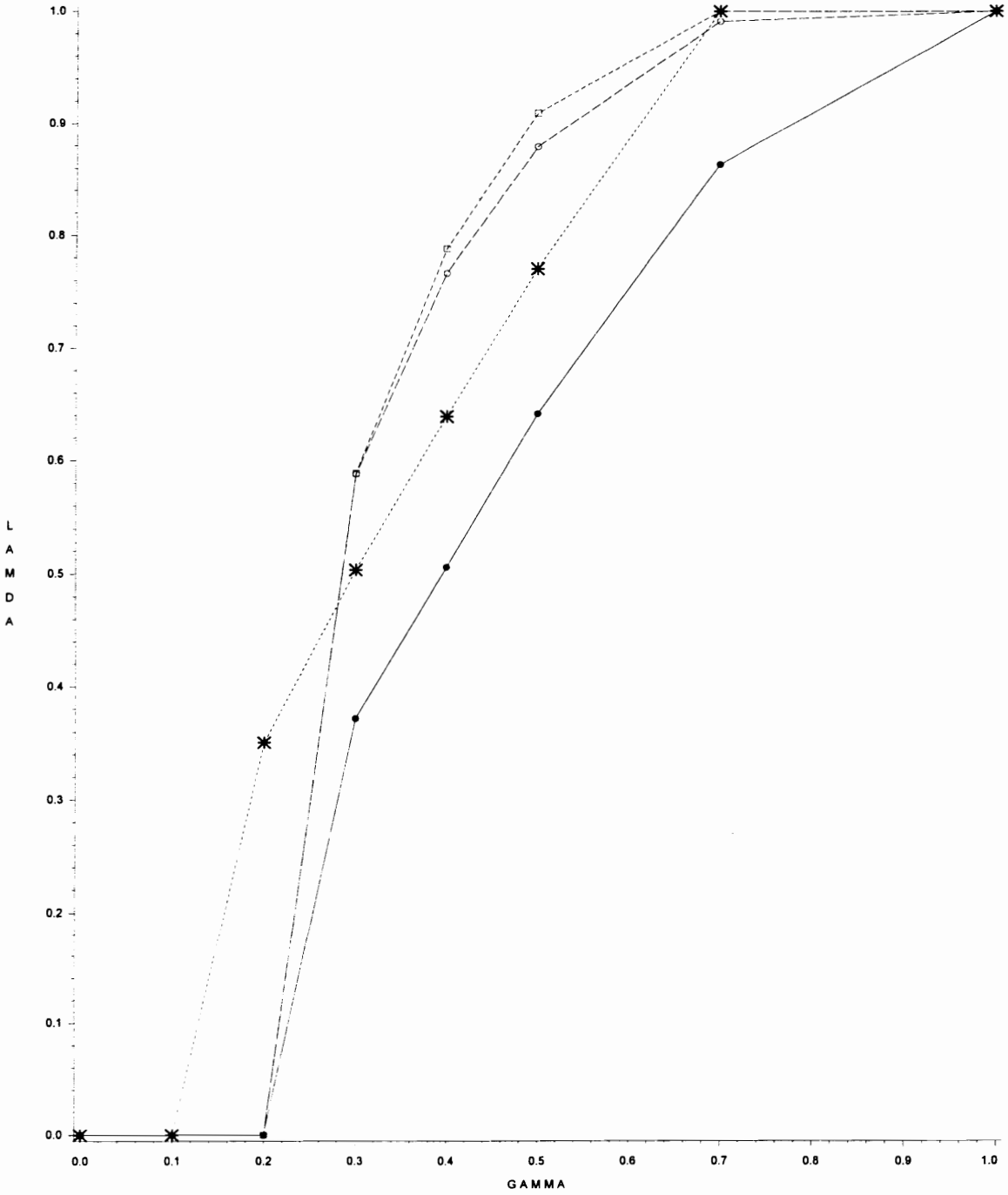


Figure 7.3.6 Mixing Parameter vs. γ Plots for (d=7, n=50).

. . . MRQR Kernel, * * * MRQR LLR, □ □ □

MRQR2 Kernel, o o o MRQR2 LLR

CHAPTER 8
RESULTS OF SIMULATIONS

§ 8.1 Setup of Simulations

Evaluations of the relative performance of the parametric and nonparametric procedures presented in this research will be done through a Monte Carlo simulation study. Throughout this research the model-robust methods were presented as alternatives to analyzing quantal dose-response data for the one-regressor case. The simulations were designed to emulate the models given in Chapter 7, but due the large number of possible (d, n, γ) combinations (3x3x8=72), only a subset of the combinations was considered. Dose-response data were simulated under the model given in (7.2.1) which represented a mixture of logistic distributions. In summarizing the average mean squared error properties using the optimal values of the bandwidth and mixing parameter, the following (d, n) combinations were used for γ=0. 0.1, 0.2, 0.3. 0.4, 0.5, 0.7, and 1.0: (d=3, n=10, 20), (d=5, n=10, 20, 50), and (d=7, n=20). For effective dose estimation with the optimal and data-driven bandwidths and mixing parameters, the combinations used will be (d=3, 5, 7, n=20) at values of γ=0, 0.1, 0.3, 0.5, and 1.0. Note that the values of γ range from 0 for no model-misspecification to 1.0, indicating a large degree of model-misspecification. The number of Monte Carlo repetitions was 500 for each combination considered.

Two sets of Monte Carlo simulations were obtained. The first set of simulations will be used to examine the theoretical properties (bias, variance, and mean squared error) of the model-robust procedures. The theoretical properties will be evaluated by comparing the simulated mean squared errors based on the optimal values of the bandwidth and mixing parameters for the various procedures presented with those from the theoretical results obtained by the formulas developed in **Chapter 7** using the same optimal bandwidth and mixing parameters. Recall that the optimal values of the bandwidth and mixing parameters were obtained to minimize the theoretical average mean squared error. Also, using the optimal values of the bandwidth and mixing parameters, the model-robust procedures will be evaluated with respect to their ability to estimate effective doses, width of 95% confidence intervals on effective doses, and coverage probabilities. Similarly, the model-robust procedures will be evaluated with respect to their abilities to estimate the proportion responding, P, 95% confidence intervals on P, and coverage probabilities as well.

After studying the model-robust procedures using the optimal values of the bandwidth and mixing parameters, the second set of Monte Carlo simulations were performed to evaluate the effectiveness of the PRESS* procedure for selecting the bandwidth and mixing parameter.

§ 8.2 MSE Results Using Optimal b and λ

The simulated average mean squared error is computed as follows:

1. The dose-response data is simulated for a given (d, n γ) combination.
2. For the logit procedure, the coefficients are obtained using the method of maximum likelihood. Using the estimated coefficients, the response, P, can be estimated at any dose level, x_0, as

$$\hat{P}(x_0) = \frac{1}{1 + e^{-(x_0\hat{\beta})}}$$

3. For the nonparametric and model-robust procedures, the optimal bandwidth and mixing parameter are used to obtain the proper "hat matrices."
4. The fit, $\hat{P}_i$, for 100 points over the dose range were obtained for the parametric, nonparametric, and model-robust procedures.
5. Compute the squared error (se) given by

$$se = \frac{\sum_{i=1}^{100}\left(P_i - \hat{P}_i\right)^2}{100}$$

where P_i represents the true proportion responding at a given dose level x_i.

To obtain a representation of how the procedures fit to the true curve, note that the amse is computed over the entire curve and not just at the *d* design points. The amse is computed for each Monte Carlo run (500 times in this research), and thus the Monte Carlo average mean squared error is given by

$$amse = \frac{\sum_{i=1}^{500} se_i}{500}.$$

The average mean squared error results, given in **Tables 8.2.1-8.2.3** are obtained using the optimal value of the bandwidth and mixing parameter and generating random quantal dose-response data. **Table 8.2.1** represents the mse results for the three doses with ten and twenty replications at each dose level. The **bold** values in the tables represent the theoretical mean squared errors used to compute amse-efficiencies in **Table 7.2.1**. Note that for the (d=3, n=10) scenario, the average mses for the simulated results are often substantially larger than the average mses from the theoretical results. This is mainly due to the fact that not very much information is obtained with only three doses. Notice that increasing the number of subjects at each dose level to n=20, the amse results from the simulations match fairly well with the amse from the theoretical results, although there are some instances in which the two amses are quite different, for example at γ=0.7. Some of these differences may be contributed to Monte Carlo error. It is

clear from this table that the asymptotic mean squared error formulas presented in **Chapter 7** do not give very accurate results when d=3.

When the number of doses are increased from d=3 to d=5, many of amses from the simulations match the amses from the theoretical results to four decimal places. In fact, all of the simulated amses are extremely close to the theoretical amses, apart from simulation error. This is especially noticeable for the (d=5, n=50) scenario in **Table 8.2.2**. The amses also match in the (d=7, n=20) combination (see **Table 8.2.3**), thus reinforcing the validity of the theoretical mse via simulation. Thus, it is also clear that for d=5 and n as small as 10 that the asymptotic formulas for the mse from **Chapter 7** provide for accurate estimation of the true amse.

If one were to examine the amse efficiency plots for the scenarios presented above, they would be nearly identical to those given in **Chapter 7**. **Figures 8.2.1-8.2.6** are plots of the amse efficiencies for the (d=3, n=10), (d=3, n=20), (d=5, n=10, 20, 50), and (d=7, n=20) combinations, respectively. From the figures, it is evident that the MRQR LLR and the LLR procedures have the largest efficiencies, thus the smallest amses with respect to logit analysis.

Table 8.2.1 Mean squared error estimates from 500 Monte Carlo simulations for three dose levels. N indicates the number of subjects receiving each dose of the drug. The **bold** numbers indicate the theoretical mean squared errors used to obtain amse-efficiencies in Table 7.2.1.

D	N	γ	Logistic	Kernel	MRQR Kernel	LLR	MRQR LLR
3	10	0.0	0.0125	0.0236	0.0162	0.0181	0.0181
			0.0118	**0.0195**	**0.0115**	**0.0136**	**0.0136**
		0.1	0.0111	0.0204	0.0139	0.0148	0.0148
			0.0120	**0.0164**	**0.0103**	**0.0107**	**0.0107**
		0.2	0.0105	0.0176	0.0123	0.0128	0.0128
			0.0127	**0.0138**	**0.0096**	**0.0084**	**0.0084**
		0.3	0.0111	0.0162	0.0120	0.0106	0.0106
			0.0080	**0.0079**	**0.0054**	**0.0047**	**0.0047**
		0.4	0.0104	0.0136	0.0104	0.0084	0.0084
			0.0155	**0.0102**	**0.0098**	**0.0053**	**0.0053**
		0.5	0.0101	0.0113	0.0093	0.0067	0.0067
			0.0176	**0.0093**	**0.0107**	**0.0045**	**0.0045**
		0.7	0.0120	0.0096	0.0097	0.0057	0.0057
			0.0233	**0.0089**	**0.0142**	**0.0046**	**0.0046**
		1.0	0.0198	0.0121	0.0155	0.0085	0.0085
			0.0356	**0.0125**	**0.0234**	**0.0088**	**0.0088**
	20	0.0	0.0051	0.0172	0.0071	0.0133	0.0133
			0.0059	**0.0158**	**0.0061**	**0.0118**	**0.0118**
		0.1	0.0046	0.0139	0.0057	0.0102	0.0102
			0.0061	**0.0126**	**0.0053**	**0.0088**	**0.0088**
		0.2	0.0045	0.0108	0.0047	0.0077	0.0077
			0.0068	**0.0100**	**0.0051**	**0.0065**	**0.0065**
		0.3	0.0048	0.0088	0.0044	0.0056	0.0056
			0.0080	**0.0079**	**0.0054**	**0.0047**	**0.0047**
		0.4	0.0056	0.0069	0.0044	0.0041	0.0041
			0.0096	**0.0063**	**0.0062**	**0.0033**	**0.0033**
		0.5	0.0070	0.0053	0.0050	0.0030	0.0030
			0.0118	**0.0053**	**0.0075**	**0.0026**	**0.0026**
		0.7	0.0117	0.0050	0.0086	0.0029	0.0029
			0.0297	**0.0082**	**0.0221**	**0.0069**	**0.0069**
		1.0	0.0225	0.0075	0.0173	0.0064	0.0064
			0.0297	**0.0082**	**0.0221**	**0.0069**	**0.0069**

Table 8.2.2 Mean squared error estimates from 500 Monte Carlo simulations for five dose levels. N indicates the number of subjects receiving each dose of the drug. The **bold** numbers indicate the theoretical mean squared errors used to obtain amse-efficiencies in Table 7.2.2.

D	N	γ	Logistic	Kernel	MRQR Kernel	LLR	MRQR LLR
5	10	0.0	0.0053	0.0069	0.0050	0.0065	0.0048
			0.0060	**0.0072**	**0.0057**	**0.0067**	**0.0054**
		0.1	0.0059	0.0071	0.0055	0.0061	0.0050
			0.0060	**0.0073**	**0.0057**	**0.0062**	**0.0052**
		0.2	0.0064	0.0073	0.0059	0.0055	0.0048
			0.0062	**0.0074**	**0.0058**	**0.0055**	**0.0049**
		0.3	0.0065	0.0071	0.0059	0.0045	0.0042
			0.0065	**0.0076**	**0.0061**	**0.0048**	**0.0045**
		0.4	0.0078	0.0082	0.0070	0.0044	0.0044
			0.0069	**0.0079**	**0.0064**	**0.0042**	**0.0042**
		0.5	0.0081	0.0085	0.0073	0.0041	0.0042
			0.0075	**0.0082**	**0.0069**	**0.0040**	**0.0040**
		0.7	0.0099	0.0093	0.0084	0.0047	0.0047
			0.0094	**0.0093**	**0.0082**	**0.0047**	**0.0047**
		1.0	0.0147	0.0134	0.0131	0.0101	0.0101
			0.0140	**0.0116**	**0.0110**	**0.0080**	**0.0080**
	20	0.0	0.0033	0.0042	0.0033	0.0041	0.0033
			0.0030	**0.0042**	**0.0038**	**0.0041**	**0.0030**
		0.1	0.0030	0.0040	0.0030	0.0038	0.0027
			0.0030	**0.0042**	**0.0030**	**0.0040**	**0.0028**
		0.2	0.0032	0.0043	0.0032	0.0036	0.0030
			0.0032	**0.0043**	**0.0032**	**0.0036**	**0.0027**
		0.3	0.0036	0.0042	0.0032	0.0030	0.0025
			0.0035	**0.0044**	**0.0032**	**0.0031**	**0.0025**
		0.4	0.0042	0.0044	0.0036	0.0025	0.0023
			0.0039	**0.0045**	**0.0035**	**0.0026**	**0.0023**
		0.5	0.0046	0.0048	0.0040	0.0023	0.0023
			0.0045	**0.0048**	**0.0039**	**0.0023**	**0.0022**
		0.7	0.0067	0.0056	0.0051	0.0029	0.0029
			0.0064	**0.0055**	**0.0050**	**0.0028**	**0.0028**
		1.0	0.0111	0.0067	0.0066	0.0052	0.0052
			0.0111	**0.0070**	**0.0068**	**0.0054**	**0.0054**

Table 8.2.2 (continued)

D	N	γ	Logistic	Kernel	MRQR Kernel	LLR	MRQR LLR
5	50	0.0	0.0012	0.0023	0.0012	0.0024	0.0012
			0.0012	**0.0022**	**0.0012**	**0.0021**	**0.0012**
		0.1	0.0013	0.0020	0.0013	0.0020	0.0013
			0.0012	**0.0021**	**0.0012**	**0.0021**	**0.0012**
		0.2	0.0014	0.0022	0.0013	0.0021	0.0013
			0.0014	**0.0021**	**0.0013**	**0.0020**	**0.0012**
		0.3	0.0017	0.0021	0.0014	0.0018	0.0012
			0.0017	**0.0021**	**0.0014**	**0.0018**	**0.0012**
		0.4	0.0021	0.0022	0.0016	0.0015	0.0011
			0.0021	**0.0022**	**0.0016**	**0.0015**	**0.0012**
		0.5	0.0028	0.0024	0.0020	0.0012	0.0011
			0.0027	**0.0024**	**0.0020**	**0.0012**	**0.0011**
		0.7	0.0046	0.0027	0.0026	0.0017	0.0017
			0.0046	**0.0028**	**0.0027**	**0.0017**	**0.0017**
		1.0	0.0094	0.0040	0.0039	0.0034	0.0034
			0.0093	**0.0039**	**0.0039**	**0.0033**	**0.0033**

Table 8.2.3 Mean squared error estimates from 500 Monte Carlo simulations for seven dose levels. N indicates the number of subjects receiving each dose of the drug. The **bold** numbers indicate the theoretical mean squared errors used to obtain amse-efficiencies in Table 7.2.3.

D	N	γ	Logistic	Kernel	MRQR Kernel	LLR	MRQR LLR
7	20	0.0	0.0019	0.0030	0.0019	0.0029	0.0019
			0.0020	**0.0031**	**0.0020**	**0.0029**	**0.0020**
		0.1	0.0020	0.0029	0.0020	0.0026	0.0018
			0.0021	**0.0031**	**0.0021**	**0.0028**	**0.0020**
		0.2	0.0023	0.0031	0.0023	0.0026	0.0020
			0.0022	**0.0031**	**0.0022**	**0.0026**	**0.0019**
		0.3	0.0027	0.0032	0.0024	0.0023	0.0019
			0.0025	**0.0032**	**0.0023**	**0.0022**	**0.0019**
		0.4	0.0030	0.0033	0.0026	0.0018	0.0017
			0.0030	**0.0033**	**0.0026**	**0.0019**	**0.0018**
		0.5	0.0038	0.0036	0.0031	0.0017	0.0018
			0.0036	**0.0036**	**0.0030**	**0.0016**	**0.0017**
		0.7	0.0054	0.0041	0.0038	0.0022	0.0022
			0.0054	0.0042	**0.0039**	**0.0023**	**0.0023**
		1.0	0.0101	0.0051	0.0051	0.0043	0.0043
			0.0100	**0.0053**	**0.0052**	**0.0044**	**0.0044**

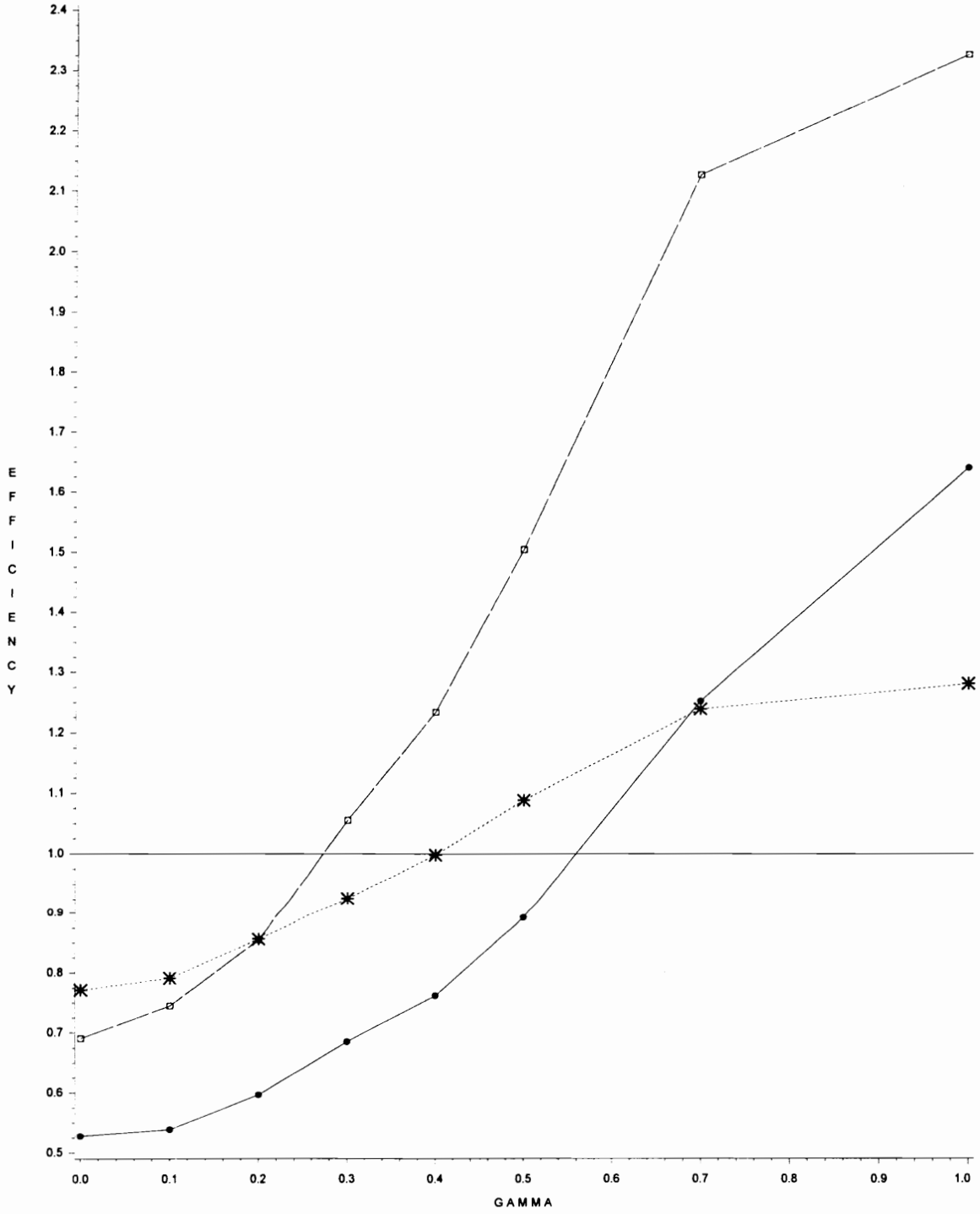


Figure 8.2.1 Average MSE Efficiency Plot for (d=3, n=10) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

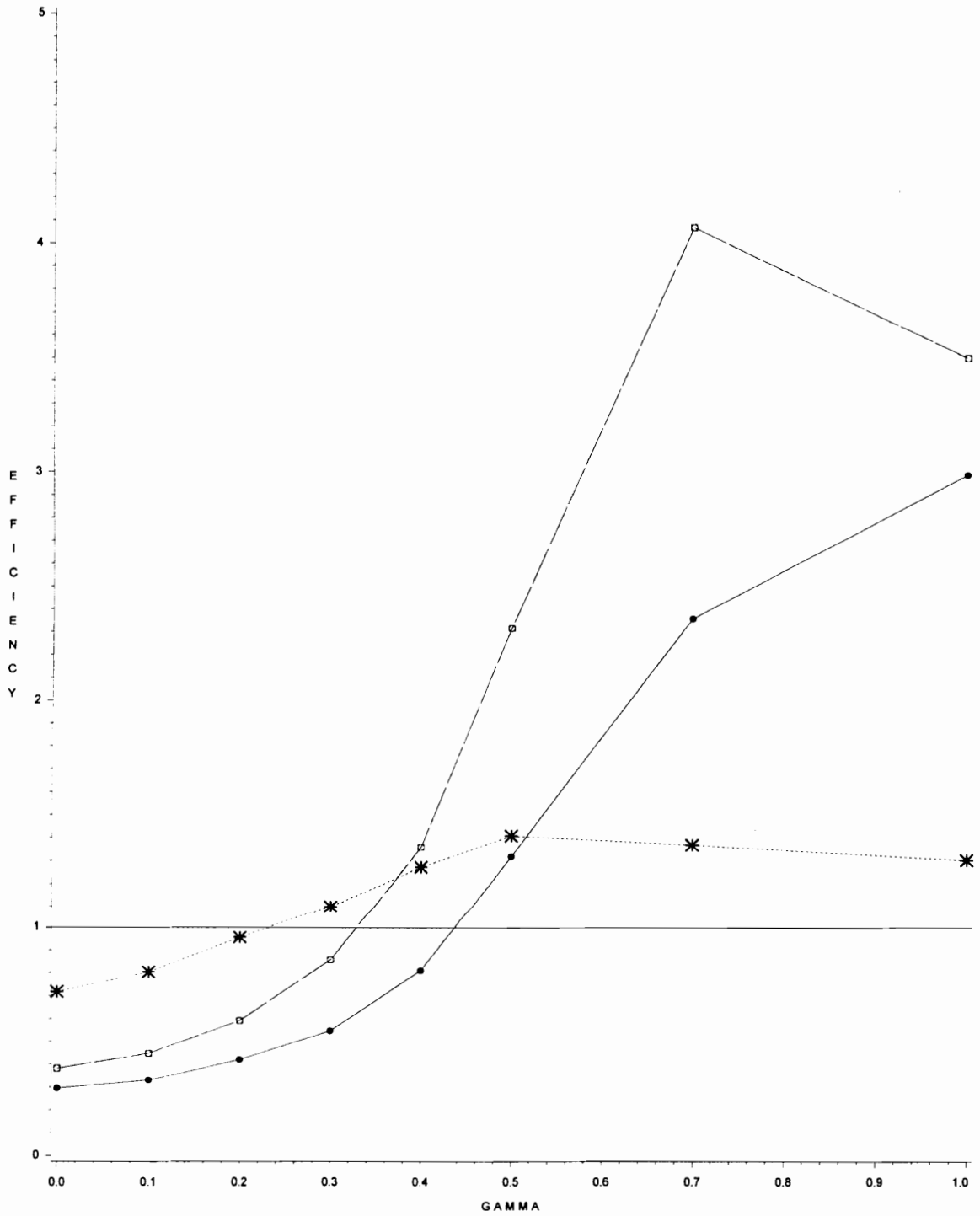


Figure 8.2.2 Average MSE Efficiency Plot for (d=3, n=20) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

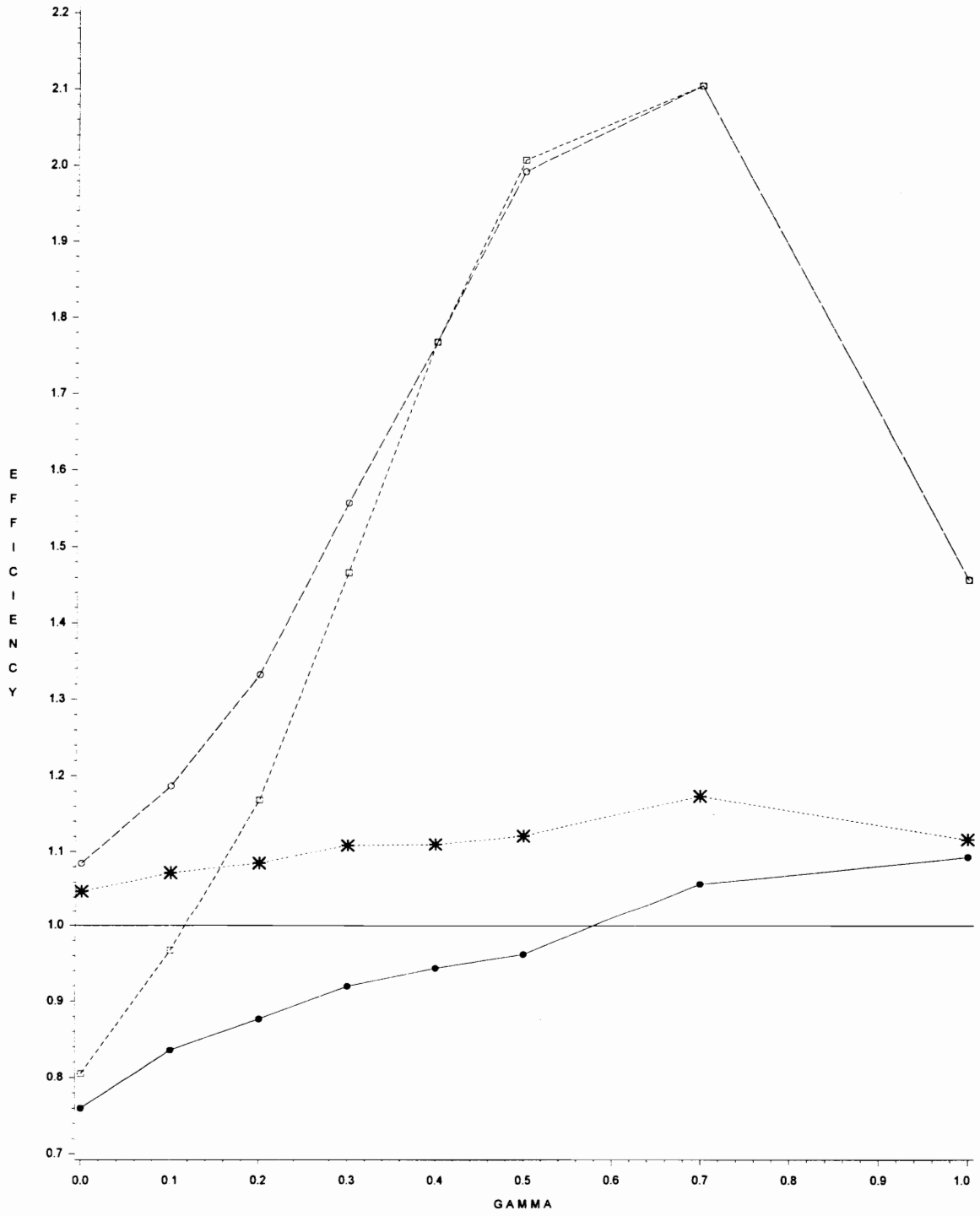


Figure 8.2.3 Average MSE Efficiency Plot for (d=5, n=10) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

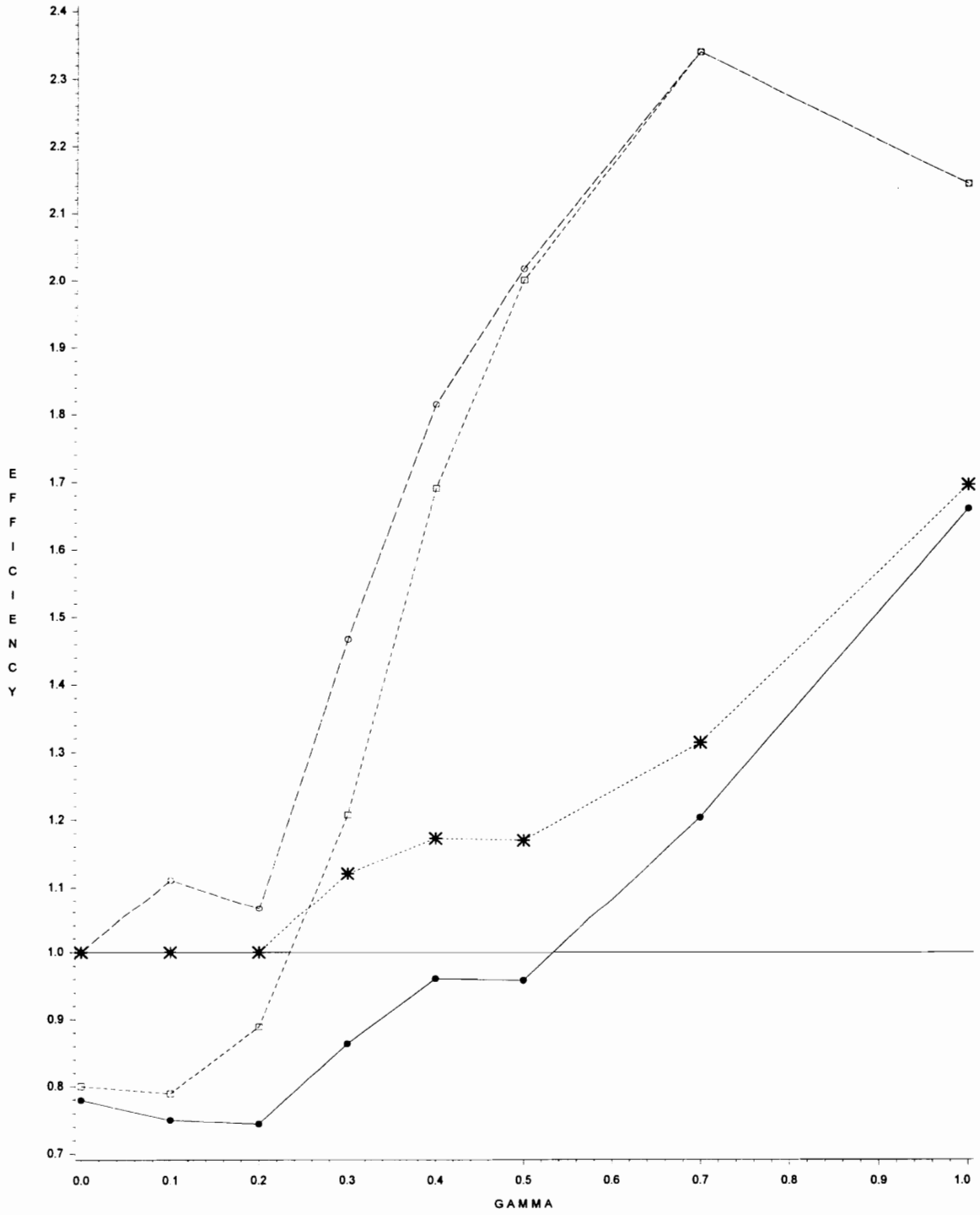


Figure 8.2.4 Average MSE Efficiency Plot for (d=5, n=20) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

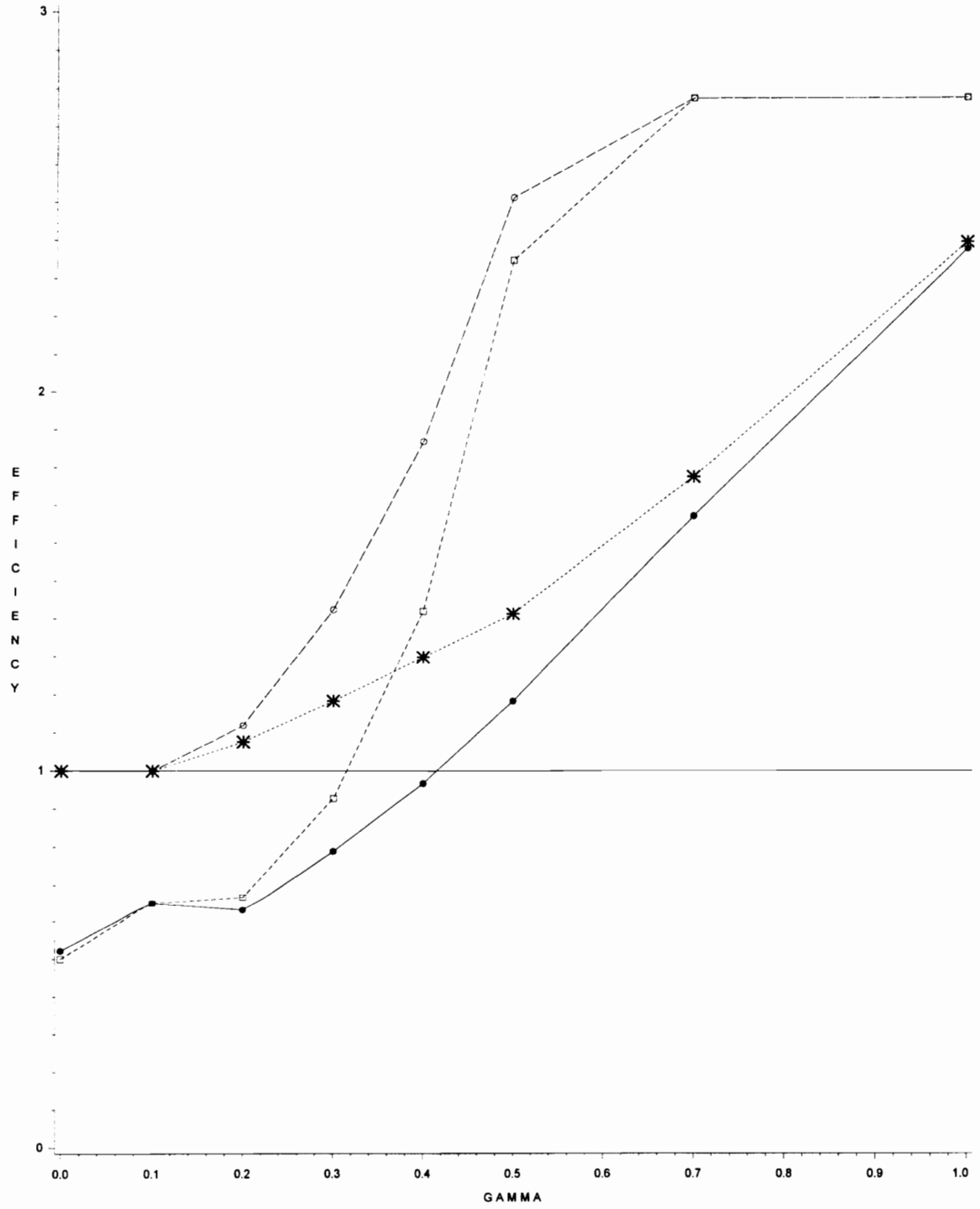


Figure 8.2.5 Average MSE Efficiency Plot for (d=5, n=50) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

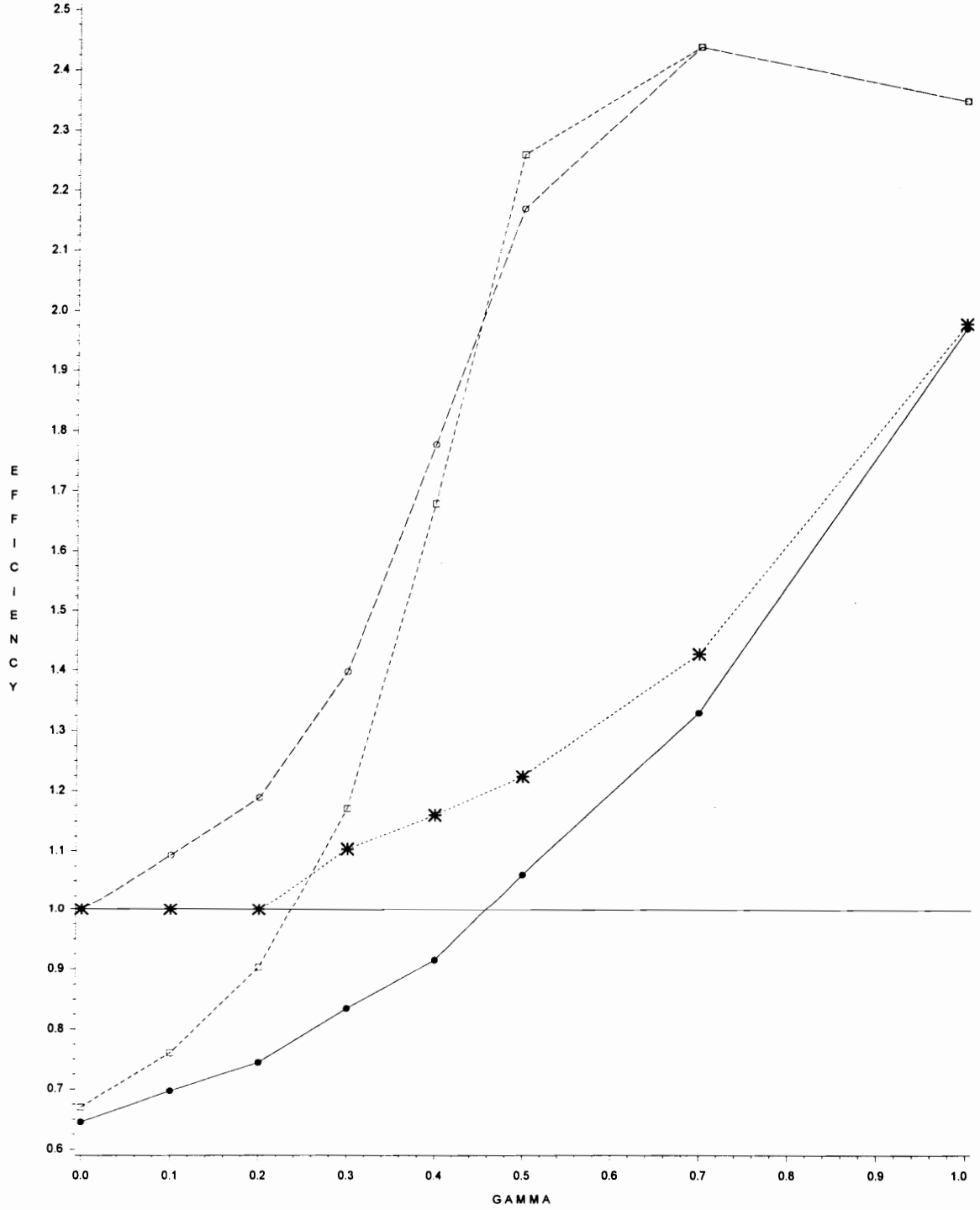


Figure 8.2.6 Average MSE Efficiency Plot for (d=7, n=20) combination with optimal bandwidth and mixing parameter.

. . . Kernel, * * * MRQR Kernel, □ □ □ LLR, o o o MRQR LLR

§ 8.3 Summary of Effective Dose Estimation with Optimal b and λ

In analyzing the results for estimating effective doses, the optimal values of the bandwidth and mixing parameter will be used to estimate effective doses, 95% confidence intervals on the effective doses, average widths of the 95% confidence intervals, and their coverage probabilities for the parametric, nonparametric, and model-robust procedures. The effective doses of interest in this research are the ED_{20}, ED_{50}, and ED_{80}. The (d=3, 5, 7, n=20) combinations will be summarized at γ=0, 0.1, 0.3, 0.5, and 1.0. These 15 (3x1x5) design combinations were analyzed to get a representation of the complete number of 72 possible Monte Carlo combinations. For each (d, n, γ) combination considered, 500 Monte Carlo repetitions were obtained and for each combination, the mean effective dose, average 95% confidence intervals on the effective doses, and average width of the 95% confidence intervals were obtained. Also, the coverage probability of the 95% confidence intervals were estimated as the proportion of repetitions among the 500 where the 95% confidence interval contained the true effective dose. Along with the nonparametric regression methods, the Spearman-Karber and Thompson Moving Averages were used to estimate the median effective doses.

For the (d=3, n=20) scenario, **Tables 8.3.1-8.3.5** represent the effective dose summaries for γ=0, 0.1, 0.3, 0.5, and 1.0, respectively. The following are some of the more noteworthy observations for the (d=3, n=20, γ) combinations:

1. Under the correct model (γ=0, **Table 8.3.1**), it is interesting to note that the logit procedure with estimates obtained via maximum likelihood estimation, the coverages of the 95% confidence Fieller confidence interval on the ED_{20}, ED_{50}, and ED_{80} are 78.4, 86.0, and 78.0, respectively. Although they should have 95% coverage, the small coverage probabilities are due to the lack of information obtained when using only three doses in the study.

2. The estimates of the median effective doses, ED_{50}, for each of the procedures are relatively the same with the nonparametric and model-robust procedures having better coverage probabilities (all above 93%) with slightly larger widths in the confidence intervals than the logit procedure.

3. The coverage probability of the confidence intervals for the ED_{20} and ED_{80} estimates for kernel and local liner regression are not very good. But, the MRQR Kernel procedure mixes well enough so that the coverage probabilities of the ED_{20} and ED_{80} increase to 91.2 and 89.8, respectively. On the other hand, with the local linear regression and the MRQR LLR procedures (recall that the optimal mixing parameter is one for three dose across all values of γ), the fit is linear in nature and

thus over-estimating the ED_{80} (the average ED_{80} is 0.7570 compared to the true ED_{80} of 0.6385) and under-estimating the ED_{20} (the average ED_{20} is 0.2443 with the true estimate of 0.3615). The coverage probabilities for the ED_{20} and ED_{80} by the MRQR LLR procedure are 75.7 and 17.8, respectively. Under the correct model, the logit procedure has the best estimates of effective doses, but the MRQR Kernel procedure yields the highest coverage probabilities.

4. The Spearman-Karber and Thompson's Moving Average estimates of the ED_{50} are 0.4667 and 0.5013, respectively, with coverage probabilities of 86.0 and 95.2. The widths of the two classical nonparametric procedures are slightly larger than the logit procedure, but smaller than the nonparametric and model-robust regression methods.

5. As the degree of model-misspecification increases to 0.1 and 0.3 for (d=3, n=20), the results are quite the same as above. The MRQR Kernel procedure has the highest coverage probabilities on the effective doses (see **Tables 8.3.2** and **Table 8.3.3**), but the logit procedure has the best estimates of the effective doses at γ=0.1, whereas the MRQR Kernel procedure has the highest coverage probabilities and, thus, closer to the nominal 95% value, and more accurate estimates of the effective doses at γ=0.3. Also, the LLR and the MRQR LLR procedures continue to fit poorly in the extremes (ED_{20} and ED_{80}). The S-K and the TMA estimates have coverage probabilities on the ED_{50} of 85.2 and 94.6, respectively. The ED_{50} estimates are approximately the same for all of the methods presented, with the exception of the S-K method which seems to underestimate the ED_{50}.

6. When the degree of model-misspecification is γ=0.5, the kernel procedure has the highest coverage probabilities as well as the most accurate estimates of the effective doses. The increase in coverage probabilities between the logit and kernel procedures is perhaps due to the increase in the width of the confidence intervals for the ED_{50} estimates from 0.1608 for the logit procedure to 0.2794 for the kernel procedure (see **Table 8.3.4**). The LLR procedure also does well in covering the true ED_{20} and ED_{80} values, although it slightly underestimates the ED_{20} value and slightly overestimates the ED_{80} value. The S-K procedure does not estimate the ED_{50} as well as the TMA procedure.

7. At the highest degree of misspecification (γ=1.0, **Table 8.3.5**), the LLR and the MRQR LLR procedure (recall that the mixing parameter is one) has the highest coverage probabilities as well

as the most accurate estimates of the effective doses. The MRQR LLR procedure also has the smallest widths on the confidence intervals for the effective doses.

It seems that for three doses, the MRQR Kernel procedure tends to perform quite well under slight model-misspecification, but as the degree of model-misspecification increases to 1.0, the MRQR LLR procedure begins to outperform the other nonparametric and the MRQR Kernel procedures. For all values of γ, the logit, nonparametric, and model-robust procedures perform well in estimating the median effective dose, but the model-robust procedures perform better when estimating the extreme values (ED_{20} and ED_{50}) at values of γ between 0.3 and 1.0.

Increasing the number of doses to d=5, the model-robust procedures begin to outperform the logit and nonparametric procedures with respect to coverage probabilities and effective dose estimation. The main points of emphasis for (d=5, n=20) combination are listed below:

1. **Table 8.3.6** is a summary of the effective dose estimation for γ=0. Under the correct model, the logit procedure, MRQR Kernel, and MRQR LLR procedures have more accurate estimates of effective doses, shorter confidence interval widths, and higher coverage probabilities than the nonparametric procedures. Note that when γ=0, the optimal value of λ=0, which implies that the model-robust procedures are choosing the proper value of λ.

2. Increasing γ to 0.1 (see **Table 8.3.7**), the MRQR Kernel procedure did not detect the slight misspecification but the misspecification was detected by the MRQR LLR procedure. The MRQR LLR procedure has coverage probabilities of 88.6%, 94.0%, and 93.4% for the ED_{20}, ED_{50}, and ED_{80} estimates, respectively. In addition, the MRQR LLR procedure has higher coverage probability, shorter average widths and more accurate estimates of effective doses than either MLE or LLR, again indicating that the concept of mixing does give improved results. The average widths of the confidence intervals for the logit, nonparametric, and model-robust procedures are relatively the same, but the coverages of the MRQR LLR procedure is slightly higher. Also, the MRQR procedure estimates the ED_{20} and ED_{50} quite well, but over estimates the ED_{80} value. Though the S-K and the TMA estimates have narrower confidence intervals, this seems to contribute to the low coverage probability with respect to the other procedures.

3. **Tables 8.3.8-8.3.10** are for the (d=5, n=20, γ=0.3, 0.5, 1.0) combinations. For these scenarios, the MRQR LLR procedure has higher coverage probabilities, narrower confidence intervals, and estimates the effective doses more accurately than the other procedures. The logit procedure

overestimates the ED_{20} and underestimates the ED_{80} values. Like the case with three doses, the MRQR LLR procedure tends to outperform the logit procedure, nonparametric methods, and the MRQR Kernel procedures as the degree of model-misspecification increases.

For the (d=7, n=20, γ) combinations, **Tables 8.3.11-8.3.15** contains the effective dose summaries. In the seven dose case, the MRQR LLR procedure outperforms the other procedures. It is also interesting that the S-K and TMA coverage probabilities are low with very narrow confidence intervals. Also, with seven doses, the logit procedure has high coverage probability on the effective doses, but the confidence interval widths for MRQR LLR for ED_{20} and ED_{80} are substantially smaller. Again, MLE does a great job of estimating the ED_{50}, even at gross model-misspecification.

With respect to effective dose estimation, the median effective dose is well estimated by all of the methods presented. Regardless of the degree of model-misspecification, the ED_{50} is estimated well by the logit procedure. This is perhaps due to the fact that for all values of γ, the true ED_{50} value is 0.5, which is the center of the data. Because the median effective dose is estimated so well by all of the procedures, the extreme dose value estimation, such as the ED_{20} and ED_{80} yields a better representation of how the methods are fitting the data. With five or seven dose-levels, the MRQR LLR procedure has the highest coverage probabilities on the extreme dose levels, narrower 95% confidence intervals, and more accurate estimates.

TABLE 8.3.1 Effective dose estimation summary for the (d=3, n=20, γ=0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3615; ED_{50}=0.5; and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3392	0.2296	78.4
	0.5	0	0.5009	0.1588	86.0
	0.8	0	0.6626	0.2304	78.0
Kernel	0.2	44	0.2644	0.1769	44.7
	0.5	0	0.5040	0.2803	93.8
	0.8	43	0.7363	0.1715	39.6
MRQR Kernel	0.2	0	0.3134	0.2189	91.2
	0.5	0	0.5015	0.2082	93.6
	0.8	0	0.6883	0.2164	89.8
LLR	0.2	0	0.2443	0.1540	75.7
	0.5	0	0.5011	0.1647	93.6
	0.8	0	0.7570	0.1521	17.8
MRQR LLR	0.2	0	0.2443	0.1540	75.7
	0.5	0	0.5011	0.1647	93.6
	0.8	0	0.7570	0.1521	17.8
Spearman-Karber	0.5	0	0.4667	0.1712	86.0
TMA	0.5	0	0.5013	0.2138	95.2

TABLE 8.3.2 Effective dose estimation summary for the (d=3, n=20, γ=0.1) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3445; ED_{50}=0.5; and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3392	0.2310	85.8
	0.5	0	0.5030	0.1604	85.8
	0.8	0	0.6668	0.2330	85.6
Kernel	0.2	39	0.2647	0.1796	56.0
	0.5	0	0.5068	0.2822	93.6
	0.8	53	0.7384	0.1700	43.8
MRQR Kernel	0.2	0	0.3140	0.2198	93.4
	0.5	0	0.5035	0.2094	94.0
	0.8	0	0.6913	0.2193	94.0
LLR	0.2	0	0.2451	0.1541	33.8
	0.5	0	0.5026	0.1658	93.6
	0.8	0	0.7595	0.1552	26.6
MRQR LLR	0.2	0	0.2451	0.1541	33.8
	0.5	0	0.5026	0.1658	93.6
	0.8	0	0.7595	0.1522	26.6
Spearman-Karber	0.5	0	0.4653	0.1714	85.2
TMA	0.5	0	0.5028	0.2154	94.6

TABLE 8.3.3 Effective dose estimation summary for the (d=3, n=20, γ=0.3) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3055; ED_{50}=0.5; and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3393	0.2300	94.2
	0.5	0	0.5015	0.1588	88.0
	0.8	0	0.6636	0.2309	94.8
Kernel	0.2	38	0.2640	0.1725	77.5
	0.5	0	0.5030	0.2830	94.6
	0.8	38	0.7375	0.1723	82.3
MRQR Kernel	0.2	0	0.3146	0.2183	95.0
	0.5	0	0.5017	0.2078	94.8
	0.8	0	0.6879	0.2182	94.6
LLR	0.2	0	0.2444	0.1533	66.2
	0.5	0	0.5012	0.1651	94.6
	0.8	0	0.7577	0.1533	64.4
MRQR LLR	0.2	0	0.2444	0.1533	66.2
	0.5	0	0.5012	0.1651	94.6
	0.8	0	0.7577	0.1533	64.4
Spearman-Karber	0.5	0	0.4657	0.1717	88.0
TMA	0.5	0	0.5013	0.2140	95.8

TABLE 8.3.4 Effective dose estimation summary for the (d=3, n=20, γ=0.5) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2730; ED_{50}=0.5; and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3353	0.2324	93.4
	0.5	0	0.4992	0.1608	84.0
	0.8	0	0.6630	0.2318	91.6
Kernel	0.2	52	0.2624	0.1739	93.8
	0.5	0	0.4981	0.2794	91.8
	0.8	55	0.7371	0.1780	94.3
MRQR Kernel	0.2	0	0.3125	0.2204	83.2
	0.5	0	0.4991	0.2074	92.2
	0.8	0	0.6857	0.2217	80.8
LLR	0.2	0	0.2421	0.1546	86.8
	0.5	0	0.4992	0.1657	91.8
	0.8	0	0.7567	0.1561	87.4
MRQR LLR	0.2	0	0.2421	0.1546	86.8
	0.5	0	0.4992	0.1656	91.8
	0.8	0	0.7567	0.1561	87.4
Spearman-Karber	0.5	0	0.4622	0.1711	85.4
TMA	0.5	0	0.4990	0.2145	94.0

TABLE 8.3.5 Effective dose estimation summary for the (d=3, n=20, γ=1.0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2300; ED_{50}=0.5; and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3387	0.2337	87.0
	0.5	0	0.5054	0.1621	87.0
	0.8	0	0.6721	0.2360	85.0
Kernel	0.2	55	0.2646	0.1859	72.4
	0.5	0	0.5121	0.2823	93.6
	0.8	65	0.7447	0.1643	73.1
MRQR Kernel	0.2	0	0.3178	0.2262	54.8
	0.5	0	0.5065	0.2076	93.6
	0.8	0	0.6929	0.2209	57.0
LLR	0.2	0	0.2453	0.1590	87.4
	0.5	0	0.5052	0.1669	93.4
	0.8	0	0.7626	0.1562	91.2
MRQR LLR	0.2	0	0.2453	0.1590	87.4
	0.5	0	0.5052	0.1669	93.4
	0.8	0	0.7626	0.1562	91.2
Spearman-Karber	0.5	0	0.4674	0.1715	88.6
TMA	0.5	0	0.5062	0.2191	97.2

TABLE 8.3.6 Effective dose estimation summary for the (d=5, n=20, γ=0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3615; ED_{50}=0.5; and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3662	0.1609	90.8
	0.5	0	0.5020	0.1183	92.8
	0.8	0	0.6378	0.1612	91.6
Kernel	0.2	27	0.3318	0.1632	91.8
	0.5	0	0.5041	0.1451	93.4
	0.8	33	0.6715	0.1534	90.4
MRQR Kernel	0.2	0	0.3662	0.1609	90.8
	0.5	0	0.5020	0.1183	92.8
	0.8	0	0.6378	0.1612	91.6
LLR	0.2	36	0.3276	0.1577	88.4
	0.5	0	0.5038	0.1411	93.4
	0.8	43	0.6757	0.1488	84.0
MRQR LLR	0.2	0	0.3662	0.1609	90.8
	0.5	0	0.5020	0.1183	92.8
	0.8	0	0.6378	0.1612	91.6
Spearman Karber	0.5	0	0.4841	0.1228	87.2
TMA	0.5	0	0.5024	0.1061	88.6

TABLE 8.3.7 Effective dose estimation summary for the (d=5, n=20, γ=0.1) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3455; ED_{50}=0.5; and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3534	0.1675	86.4
	0.5	0	0.4993	0.1222	93.0
	0.8	0	0.6452	0.1672	92.6
Kernel	0.2	30	0.3154	0.1669	90.6
	0.5	0	0.4993	0.1526	92.8
	0.8	29	0.6852	0.1676	92.6
MRQR Kernel	0.2	0	0.3534	0.1675	86.4
	0.5	0	0.4993	0.1222	93.0
	0.8	0	0.6452	0.1672	92.6
LLR	0.2	34	0.3076	0.1519	86.3
	0.5	0	0.4993	0.1459	93.0
	0.8	42	0.6928	0.1534	88.0
MRQR LLR	0.2	0	0.3403	0.1558	88.6
	0.5	0	0.4993	0.1292	94.0
	0.8	0	0.6586	0.1559	93.4
Spearman-Karber	0.5	0	0.4813	0.1253	83.2
TMA	0.5	0	0.4993	0.1049	85.2

TABLE 8.3.8 Effective dose estimation summary for the (d=5, n=20, γ=0.3) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3055; ED_{50}=0.5; and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3365	0.1851	90.0
	0.5	0	0.5011	0.1327	94.6
	0.8	0	0.6656	0.1850	93.2
Kernel	0.2	29	0.2886	0.1819	91.3
	0.5	0	0.5008	0.1769	92.0
	0.8	21	0.7157	0.1834	94.8
MRQR Kernel	0.2	0	0.3223	0.1720	87.8
	0.5	0	0.5009	0.1436	94.8
	0.8	0	0.6804	0.1723	91.0
LLR	0.2	19	0.2753	0.1361	88.6
	0.5	0	0.5011	0.1470	93.4
	0.8	13	0.7269	0.1356	90.1
MRQR LLR	0.2	4	0.3019	0.1538	92.5
	0.5	0	0.5011	0.1399	94.8
	0.8	4	0.7004	0.1535	95.2
Spearman-Karber	0.5	0	0.4839	0.1357	88.0
TMA	0.5	0	0.5016	0.1066	83.2

TABLE 8.3.9 Effective dose estimation summary for the (d=5, n=20, γ=0.5) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2730; ED_{50}=0.5; and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3229	0.2072	88.8
	0.5	0	0.5051	0.1450	92.8
	0.8	0	0.6873	0.2083	90.8
Kernel	0.2	21	0.2688	0.1883	92.1
	0.5	0	0.5028	0.2113	93.0
	0.8	33	0.7442	0.1927	94.9
MRQR Kernel	0.2	0	0.2963	0.1846	86.0
	0.5	0	0.5042	0.1675	93.0
	0.8	0	0.7145	0.1892	91.2
LLR	0.2	0	0.2601	0.1301	93.2
	0.5	0	0.5042	0.1344	94.2
	0.8	0	0.7479	0.1340	91.4
MRQR LLR	0.2	0	0.2663	0.1359	93.8
	0.5	0	0.5044	0.1358	94.2
	0.8	0	0.7420	0.1300	92.2
Spearman-Karber	0.5	0	0.4838	0.1485	87.8
TMA	0.5	0	0.5049	0.1094	75.8

TABLE 8.3.10 Effective dose estimation summary for the (d=5, n=20, γ=1.0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2300; ED_{50}=0.5; and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.2683	0.3164	87.6
	0.5	0	0.5011	0.1950	95.4
	0.8	0	0.7340	0.3169	87.4
Kernel	0.2	42	0.2129	0.1747	92.4
	0.5	0	0.5008	0.3564	93.8
	0.8	37	0.7884	0.1691	90.7
MRQR Kernel	0.2	42	0.2166	0.1729	91.3
	0.5	0	0.5006	0.3336	93.8
	0.8	37	0.7848	0.1699	90.5
LLR	0.2	15	0.2198	0.1490	92.0
	0.5	0	0.5013	0.2584	94.4
	0.8	15	0.7818	0.1463	93.0
MRQR LLR	0.2	15	0.2198	0.1490	92.0
	0.5	0	0.5013	0.2584	94.4
	0.8	105	0.7818	0.1376	92.7
Spearman-Karber	0.5	0	0.4785	0.2012	91.8
TMA	0.5	0	0.5015	0.1276	74.4

TABLE 8.3.11 Effective dose estimation summary for the (d=7, n=20, γ=0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3615; ED_{50}=0.5; and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3643	0.1309	92.0
	0.5	0	0.4990	0.0977	93.2
	0.8	0	0.6336	0.1309	91.8
Kernel	0.2	28	0.3390	0.1377	92.4
	0.5	0	0.4981	0.1195	92.6
	0.8	22	0.6596	0.1442	91.8
MRQR Kernel	0.2	0	0.3643	0.1309	92.0
	0.5	0	0.4990	0.0977	93..2
	0.8	0	0.6336	0.1309	91.8
LLR	0.2	124	0.3355	0.1337	89.6
	0.5	0	0.4982	0.1167	93.0
	0.8	100	0.6631	0.1399	90.0
MRQR LLR	0.2	0	0.3643	0.1309	92.0
	0.5	0	0.4990	0.0977	93.2
	0.8	0	0.6336	0.1309	91.8
Spearman-Karber	0.5	0	0.4983	0.0973	78.8
TMA	0.5	0	0.4817	0.0674	76.0

TABLE 8.3.12 Effective dose estimation summary for the (d=7, n=20, γ=0.1) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3445; ED_{50}=0.5; and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3517	0.1391	92.4
	0.5	0	0.4993	0.1025	90.0
	0.8	0	0.6468	0.1389	92.6
Kernel	0.2	27	0.3215	0.1512	90.5
	0.5	0	0.4995	0.1319	92.6
	0.8	28	0.6754	0.1506	90.9
MRQR Kernel	0.2	0	0.3517	0.1391	92.4
	0.5	0	0.4993	0.1025	90.0
	0.8	0	0.6468	0.1389	92.6
LLR	0.2	86	0.3156	0.1385	87.4
	0.5	0	0.4993	0.1250	92.2
	0.8	85	0.6819	0.1401	88.0
MRQR LLR	0.2	0	0.3429	0.1316	90.4
	0.5	0	0.4993	0.1064	90.8
	0.8	0	0.6555	0.1317	91.6
Spearman-Karber	0.5	0	0.4791	0.1011	77.0
TMA	0.5	0	0.4987	0.0065	68.0

TABLE 8.3.13 Effective dose estimation summary for the (d=7, n=20, γ=0.3) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.3055; ED_{50}=0.5; and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3311	0.1531	93.2
	0.5	0	0.4980	0.1104	93.4
	0.8	0	0.6649	0.1529	89.0
Kernel	0.2	33	0.2937	0.1626	93.8
	0.5	0	0.4972	0.1544	93.8
	0.8	27	0.7012	0.1649	94.7
MRQR Kernel	0.2	0	0.3207	0.1455	90.0
	0.5	0	0.4979	0.1182	93.8
	0.8	0	0.6750	0.1461	86.0
LLR	0.2	31	0.2822	0.1207	90.0
	0.5	0	0.4975	0.1268	93.6
	0.8	20	0.7144	0.1241	92.1
MRQR LLR	0.2	1	0.3061	0.1345	95.0
	0.5	0	0.4978	0.1167	93.6
	0.8	1	0.6900	0.1356	90.8
Spearman-Karber	0.5	0	0.4770	0.1075	79.6
TMA	0.5	0	0.4972	0.0640	63.6

TABLE 8.3.14 Effective dose estimation summary for the (d=7, n=20, γ=0.5) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2730; ED_{50}=0.5; and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3137	0.1711	87.4
	0.5	0	0.5014	0.1198	93.4
	0.8	0	0.6891	0.1712	92.0
Kernel	0.2	33	0.2706	0.1704	93.6
	0.5	0	0.5022	0.1913	93.8
	0.8	25	0.7325	0.1685	93.9
MRQR Kernel	0.2	0	0.2929	0.1601	86.8
	0.5	0	0.5016	0.1432	95.0
	0.8	0	0.7101	0.1593	90.0
LLR	0.2	0	0.2618	0.1167	93.0
	0.5	0	0.5014	0.1141	94.8
	0.8	0	0.7401	0.1160	94.2
MRQR LLR	0.2	0	0.2719	0.1255	93.0
	0.5	0	0.5014	0.1157	94.8
	0.8	16	0.7303	0.1180	91.9
Spearman Karber	0.5	0	0.4807	0.1145	81.4
TMA	0.5	0	0.5019	0.0608	54.0

TABLE 8.3.15 Effective dose estimation summary for the (d=7, n=20, γ=1.0) combination using optimal values of bandwidth and mixing parameter. The true effective dose values are: ED_{20}=0.2300; ED_{50}=0.5; and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.2598	0.2404	94.4
	0.5	0	0.4999	0.1531	93.8
	0.8	0	0.7401	0.2405	92.0
Kernel	0.2	38	0.2191	0.1512	92.0
	0.5	0	0.4976	0.3241	93.2
	0.8	33	0.7807	0.1506	92.9
MRQR Kernel	0.2	38	0.2226	0.1513	92.2
	0.5	0	0.4977	0.3080	93.0
	0.8	33	0.7778	0.1509	92.9
LLR	0.2	30	0.2230	0.1288	93.2
	0.5	0	0.4983	0.2630	92.0
	0.8	27	0.7770	0.1281	92.2
MRQR LLR	0.2	30	0.2230	0.1288	93.2
	0.5	0	0.4983	0.2630	92.0
	0.8	27	0.7770	0.1236	93.2
Spearman Karber	0.5	0	0.4781	0.1375	84.0
TMA	0.5	0	0.4992	0.0582	28.8

§ 8.4 Estimation of Proportion Responding Using Optimal b and λ

Another aspect of interest in the analysis of quantal dose-response data is how well the parametric, nonparametric, and model-robust procedures estimate the proportion responding, P. It has been shown in the previous section that the model-robust procedures (especially MRQR LLR) perform extremely well in estimating effective doses, especially as the number of doses increases. **Tables 8.4.1-8.4.10** are summaries of the dose levels accompanied by the average percent responding, the mean width of the 95% confidence intervals on P, and the coverage probabilities at each dose level. The tables below summarize the (d=3, 5, n=20) combinations with γ=0, 0.1, 0.3, 0.5, and 1.0.

For the (d=3, n=20) combination, at each value of γ, the model-robust procedures yield the highest coverage probabilities on the true response at the given dose levels. At γ=0, the logit procedure has the largest width on the 95% confidence interval at the middle dose (0.5), but a narrower width at the extreme points (dose level 0.1 and 0.9). With respect to predicting the proportion responding at the three dose levels, all of the procedures presented predict well in the middle of the data (at dose level 0.5), but the MRQR LLR procedure more accurately predicts the true responses in the extremes. As the degree of model-misspecification (γ) increases, the results are quite similar.

Although there is not a strong distinction between the parametric, nonparametric, and model-robust procedures for three dose levels, the results are very interesting when increasing to five doses. First of all, under the correctly specified model (γ=0) and the slightly misspecified model (γ=0.1), the logistic regression procedure only yields 95% coverage of true proportion responding at only one of the doses (dose level 0.5). Interestingly, the nonparametric procedures of kernel and local linear regression have narrower confidence intervals and higher coverage probabilities on P, the true proportion responding, than the logit and model-robust procedures.

Increasing the degree of model-misspecification to γ=0.3, the LLR and MRQR LLR procedures tend to yield higher coverage probabilities on the true proportion responding as well as narrower confidence intervals than the logit, kernel, and MRQR Kernel procedures. The MRQR LLR predictions best estimate the true proportion responding over the five dose levels. Although both model-robust procedures overestimate in the boundaries, the MRQR Kernel procedure has the larger bias in the extremes, thus overestimating in the left boundary and under estimating in the right boundary. Again, these results are indicative of the boundary problem inherent with kernel regression. The LLR procedure does not suffer from this problem as expected.

At γ=0.5, the LLR and MRQR LLR procedures continue to outperform the logit, kernel, and MRQR Kernel procedures. The logit procedure predicts well at the 0.5 dose level, but does not predict well elsewhere and its coverage probabilities are generally very poor. The kernel procedure predicts well

for the middle three doses, but still does not fit well in the boundaries. With neither the kernel or logit procedure fitting the data very well, thus the MRQR Kernel procedure will not fit the data well. The LLR and MRQR LLR procedures have high coverage probabilities on the true proportion responding as well as excellent predictions at the dose levels.

At the largest degree of model-misspecification, the widths of the 95% confidence interval for the logit procedure are more than twice as wide as the confidence intervals of the nonparametric and model-robust procedures, thus yielding fairly good coverage. With a significant reduction in confidence interval widths, the LLR and MRQR LLR procedures have good coverage. At dose levels 0.3 and 0.7, the Kernel and MRQR Kernel procedures has the higher coverage probabilities than the LLR and MRQR LLR procedures. This may be due to the ability of the Kernel procedure to detect the curvature at these points, whereas the LLR procedure tends to be linear in nature. Also, with respect to predicting the true response, the nonparametric and model-robust procedures have approximately the same predictions in the boundaries, but the LLR and MRQR LLR procedures have narrower confidence intervals.

Table 8.4.1 Summary of mean response at the (d=3, n=20, γ=0) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0346	0.0180	0.1202	99.0
	0.5	0.4978	0.5000	0.3855	92.8
	0.9	0.9649	0.9820	0.1217	99.0
Kernel	0.1	0.0564	0.0180	0.1359	99.0
	0.5	0.4968	0.5000	0.3907	93.8
	0.9	0.9441	0.9820	0.1339	99.8
MRQR Kernel	0.1	0.0413	0.0180	0.1212	99.8
	0.5	0.4975	0.5000	0.3783	93.6
	0.9	0.9586	0.9820	0.1210	99.8
LLR	0.1	0.0377	0.0180	0.1538	100.0
	0.5	0.4988	0.5000	0.1943	93.6
	0.9	0.9621	0.9820	0.1522	99.8
MRQR LLR	0.1	0.0377	0.0180	0.1538	100.0
	0.5	0.4988	0.5000	0.1943	93.6
	0.9	0.9621	0.9820	0.1522	99.8

Table 8.4.2 Summary of mean response at the (d=3, n=20, γ=0.1) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0356	0.0186	0.1212	99.4
	0.5	0.4931	0.5000	0.3841	92.4
	0.9	0.9631	0.9814	0.1260	99.2
Kernel	0.1	0.0560	0.0186	0.1354	99.4
	0.5	0.4950	0.5000	0.3910	93.6
	0.9	0.9408	0.9814	0.1417	99.6
MRQR Kernel	0.1	0.0417	0.0186	0.1210	99.8
	0.5	0.5000	0.5000	0.3771	94.0
	0.9	0.9564	0.9814	0.1270	99.6
LLR	0.1	0.0373	0.0186	0.1535	99.6
	0.5	0.4969	0.5000	0.1947	93.6
	0.9	0.9589	0.9814	0.1586	100.0
MRQR LLR	0.1	0.0373	0.0186	0.1535	99.6
	0.5	0.4969	0.5000	0.1947	93.6
	0.9	0.9589	0.9814	0.1586	100.0

Table 8.4.3 Summary of mean response at the (d=3, n=20, γ=0.3) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0347	0.0197	0.1200	99.0
	0.5	0.4964	0.5000	0.3860	93.6
	0.9	0.9649	0.9803	0.1223	99.6
Kernel	0.1	0.0555	0.0197	0.1323	99.4
	0.5	0.4978	0.5000	0.3913	94.6
	0.9	0.9426	0.9803	0.1384	99.2
MRQR Kernel	0.1	0.0407	0.0197	0.1193	99.2
	0.5	0.4968	0.5000	0.3788	94.8
	0.9	0.9585	0.9803	0.1233	99.8
LLR	0.1	0.0366	0.0197	0.1512	99.8
	0.5	0.4985	0.5000	0.1944	94.6
	0.9	0.9615	0.9803	0.1559	99.6
MRQR LLR	0.1	0.0366	0.0197	0.1513	99.8
	0.5	0.4985	0.5000	0.1944	94.6
	0.9	0.9615	0.9803	0.1559	99.6

Table 8.4.4 Summary of mean response at the (d=3, n=20, γ=0.5) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0365	0.0209	0.1245	98.8
	0.5	0.5027	0.5000	0.3830	90.6
	0.9	0.9633	0.9791	0.1234	98.4
Kernel	0.1	0.0582	0.0209	0.1381	99.8
	0.5	0.5026	0.5000	0.3897	91.8
	0.9	0.9417	0.9791	0.1380	99.2
MRQR Kernel	0.1	0.0425	0.0209	0.1244	98.8
	0.5	0.5027	0.5000	0.3759	92.2
	0.9	0.9573	0.9791	0.1237	98.2
LLR	0.1	0.0399	0.0209	0.1561	100.0
	0.5	0.5011	0.5000	0.1940	91.8
	0.9	0.9601	0.9791	0.1561	99.6
MRQR LLR	0.1	0.0399	0.0209	0.1561	100.0
	0.5	0.5011	0.5000	0.1940	91.8
	0.9	0.9601	0.9791	0.1561	99.6

Table 8.4.5 Summary of mean response at the (d=3, n=20, γ=1.0) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0367	0.0237	0.1229	99.2
	0.5	0.4893	0.5000	0.3828	92.8
	0.9	0.9600	0.9763	0.1309	99.6
Kernel	0.1	0.0598	0.0237	0.1430	100.0
	0.5	0.4870	0.5000	0.3900	93.6
	0.9	0.9391	0.9763	0.1399	99.2
MRQR Kernel	0.1	0.0426	0.0237	0.1251	99.2
	0.5	0.4887	0.5000	0.3759	93.6
	0.9	0.9547	0.9763	0.1285	99.4
LLR	0.1	0.0389	0.0237	0.1604	100.0
	0.5	0.4940	0.5000	0.1945	93.4
	0.9	0.9572	0.9763	0.1584	99.4
MRQR LLR	0.1	0.0389	0.0237	0.1604	100.0
	0.5	0.4940	0.5000	0.1945	93.4
	0.9	0.9572	0.9763	0.1584	99.4

Table 8.4.6 Summary of mean response at the (d=5, n=20, γ=0) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0193	0.0180	0.0652	86.0
	0.3	0.1156	0.1192	0.2269	89.2
	0.5	0.4950	0.5000	0.3716	95.0
	0.7	0.8805	0.8808	0.2322	91.2
	0.9	0.9799	0.9820	0.0679	86.6
Kernel	0.1	0.0351	0.0180	0.0941	99.6
	0.3	0.1581	0.1192	0.2038	89.0
	0.5	0.4918	0.5000	0.3144	93.4
	0.7	0.8398	0.8808	0.2027	90.8
	0.9	0.9655	0.9820	0.0921	99.2
MRQR Kernel	0.1	0.0193	0.0180	0.0652	86.0
	0.3	0.1156	0.1192	0.2269	89.2
	0.5	0.4950	0.5000	0.3716	95.0
	0.7	0.8805	0.8808	0.2322	91.2
	0.9	0.9799	0.9820	0.0679	86.6
LLR	0.1	0.0197	0.0180	0.0931	100.0
	0.3	0.1642	0.1192	0.1953	89.4
	0.5	0.4924	0.5000	0.2956	93.4
	0.7	0.8331	0.8808	0.1943	91.6
	0.9	0.9808	0.9820	0.0906	100.0
MRQR LLR	0.1	0.0193	0.0180	0.0652	86.0
	0.3	0.1156	0.1192	0.2269	89.2
	0.5	0.4950	0.5000	0.3716	95.0
	0.7	0.8805	0.8808	0.2322	91.2
	0.9	0.9799	0.9820	0.0679	86.6

Table 8.4.7 Summary of mean response at the (d=5, n=20, γ=0.1) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0262	0.0186	0.0783	88.2
	0.3	0.1330	0.1438	0.2297	87.0
	0.5	0.4992	0.5000	0.3485	95.0
	0.7	0.8693	0.8562	0.2318	90.2
	0.9	0.9752	0.9814	0.0769	89.8
Kernel	0.1	0.0439	0.0186	0.1033	98.8
	0.3	0.1801	0.1438	0.2122	89.2
	0.5	0.5000	0.5000	0.3029	92.8
	0.7	0.8202	0.8562	0.2125	91.2
	0.9	0.9586	0.9814	0.0974	97.6
MRQR Kernel	0.1	0.0262	0.0186	0.0783	88.2
	0.3	0.1330	0.1438	0.2297	87.0
	0.5	0.4992	0.5000	0.3485	95.0
	0.7	0.8693	0.8562	0.2318	90.2
	0.9	0.9752	0.9814	0.0769	89.8
LLR	0.1	0.0217	0.0186	0.1007	100.0
	0.3	0.1921	0.1438	0.1927	86.0
	0.5	0.5000	0.5000	0.2641	93.0
	0.7	0.8084	0.8562	0.1929	88.6
	0.9	0.9813	0.9814	0.0915	100.0
MRQR LLR	0.1	0.0248	0.0186	0.0689	86.4
	0.3	0.1511	0.1438	0.1927	87.4
	0.5	0.4994	0.5000	0.2844	94.0
	0.7	0.8506	0.8562	0.1931	92.4
	0.9	0.9770	0.9814	0.0650	87.2

Table 8.4.8 Summary of mean response at the (d=5, n=20, γ=0.3) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0364	0.0197	0.1503	96.4
	0.3	0.1565	0.1931	0.2523	84.0
	0.5	0.4963	0.5000	0.3374	95.6
	0.7	0.8414	0.8069	0.2560	88.2
	0.9	0.9638	0.9803	0.1060	97.4
Kernel	0.1	0.0548	0.0197	0.1075	89.8
	0.3	0.2145	0.1931	0.2307	91.2
	0.5	0.4981	0.5000	0.2942	92.0
	0.7	0.7810	0.8069	0.2347	94.8
	0.9	0.9457	0.9803	0.1049	90.4
MRQR Kernel	0.1	0.0421	0.0197	0.0858	94.2
	0.3	0.1746	0.1931	0.2085	87.6
	0.5	0.4969	0.5000	0.2796	94.8
	0.7	0.8226	0.8069	0.2117	91.0
	0.9	0.9582	0.9803	0.0852	96.2
LLR	0.1	0.0172	0.0197	0.1120	99.8
	0.3	0.2324	0.1931	0.1738	88.4
	0.5	0.4980	0.5000	0.2046	93.4
	0.7	0.7646	0.8069	0.1755	89.2
	0.9	0.9841	0.9803	0.1083	100.0
MRQR LLR	0.1	0.0254	0.0197	0.0919	97.8
	0.3	0.1998	0.1931	0.1850	92.2
	0.5	0.4973	0.5000	0.2381	94.8
	0.7	0.7977	0.8069	0.1871	95.0
	0.9	0.9753	0.9803	0.0899	99.2

Table 8.4.9 Summary of mean response at the (d=5, n=20, γ=0.5) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0467	0.0209	0.1402	98.6
	0.3	0.1751	0.2424	0.2865	75.6
	0.5	0.4903	0.5000	0.3522	95.2
	0.7	0.8138	0.7576	0.2938	81.4
	0.9	0.9498	0.9792	0.1461	98.4
Kernel	0.1	0.0597	0.0209	0.1064	83.0
	0.3	0.2401	0.2424	0.2485	90.8
	0.5	0.4972	0.5000	0.2981	93.0
	0.7	0.7454	0.7576	0.2570	94.6
	0.9	0.9329	0.9792	0.1147	76.0
MRQR Kernel	0.1	0.0529	0.0209	0.0955	89.4
	0.3	0.2062	0.2424	0.2219	83.2
	0.5	0.4937	0.5000	0.2731	93.0
	0.7	0.7810	0.7576	0.2303	88.4
	0.9	0.9417	0.9792	0.1025	85.4
LLR	0.1	0.0200	0.0209	0.1488	99.6
	0.3	0.2493	0.2424	0.1571	94.4
	0.5	0.4949	0.5000	0.1643	94.2
	0.7	0.7409	0.7576	0.1615	93.2
	0.9	0.9744	0.9792	0.1569	99.2
MRQR LLR	0.1	0.0229	0.0209	0.1421	99.6
	0.3	0.2411	0.2424	0.1619	92.4
	0.5	0.4944	0.5000	0.1762	94.2
	0.7	0.7489	0.7576	0.1668	93.2
	0.9	0.9717	0.9792	0.1499	99.2

Table 8.4.10 Summary of mean response at the (d=5, n=20, γ=1.0) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability (%)
MLE	0.1	0.0861	0.0237	0.3609	97.0
	0.3	0.2318	0.3656	0.5289	74.6
	0.5	0.4983	0.5000	0.5568	98.6
	0.7	0.7657	0.6344	0.5325	73.4
	0.9	0.9129	0.9763	0.3649	96.8
Kernel	0.1	0.0661	0.0237	0.1102	77.6
	0.3	0.3368	0.3656	0.3192	89.8
	0.5	0.4981	0.5000	0.3364	93.8
	0.7	0.6601	0.6344	0.3203	90.0
	0.9	0.9336	0.9763	0.1106	78.2
MRQR Kernel	0.1	0.0682	0.0237	0.1098	76.4
	0.3	0.3254	0.3656	0.3050	87.8
	0.5	0.4981	0.5000	0.3198	93.8
	0.7	0.6716	0.6344	0.3061	88.2
	0.9	0.9313	0.9763	0.1102	75.8
LLR	0.1	0.0309	0.0237	0.1083	100.0
	0.3	0.3077	0.3656	0.2272	80.0
	0.5	0.4982	0.5000	0.2473	94.4
	0.7	0.6897	0.6344	0.2278	80.4
	0.9	0.9690	0.9763	0.1089	100.0
MRQR LLR	0.1	0.0309	0.0237	0.1083	100.0
	0.3	0.3077	0.3656	0.2272	80.0
	0.5	0.4982	0.5000	0.2473	94.4
	0.7	0.6897	0.6344	0.2278	80.4
	0.9	0.9690	0.9763	0.1089	100.0

§8.5 Results of Simulations Using Data-Driven Bandwidth and Mixing Parameter

In this section, the parametric, nonparametric, and model-robust procedures will be assessed using the values of the bandwidth and mixing parameter chosen by the PRESS* procedure. Recall that the PRESS* procedure selects the bandwidth and mixing parameter via a penalized cross-validation technique (see Chapter 3 for more details). Using these data-driven parameters, the mean squared errors, effective dose estimation, proportion responding estimation, and other statistics such as the χ^2 and PRESS*, will be compared across the procedures presented in this research. As mentioned in the introduction to nonparametric regression techniques, the bandwidth is the most important element when applying nonparametric methods. Though other elements will be compared, the main focus on this section will be the efficiency of the PRESS* procedure to select the optimal bandwidth and the optimal mixing parameter.

§ 8.5.1 Bandwidth and Mixing Parameter Selection

Table 8.5.1.1 lists the average values of the bandwidth and mixing parameters chosen using PRESS* and listed in **bold** are the optimal values, chosen by minimizing the theoretical average mean squared error. The table represents the (d=3, 5, 7, n=20) combinations for γ=0, 0.1, 0.3, 0.5, and 1.0.

For three doses, at all values of γ, the data-driven bandwidth is at least twice the size of the optimal bandwidth for the kernel regression case. The data-driven bandwidth selected for local linear regression overestimates the optimal values over all values of γ, but at a smaller relative error rate compared to the kernel regression procedure. The mixing parameter, on the other hand, is underestimated by the data-driven technique when using kernel and local linear regression techniques. For the two model-robust procedures, the optimal mixing parameter is most of the time twice as large and the data-driven one.

When increasing the number of doses to five, the data-driven technique for the kernel regression procedure only slightly overestimates the optimal bandwidth for slight model-misspecification. But at the highest level of misspecification, the data-driven bandwidth for the kernel procedure is twice that of the optimal bandwidth. The data-driven bandwidth for the local linear regression procedure is approximately one for all values of γ, often over-estimating the optimal values of the bandwidth by two to four times. As for the data-driven mixing parameter for five doses, the same trend is apparent that was evident when studying the optimal values of the mixing parameter: as the degree of model-misspecification increases, so does the value of the data-driven mixing parameter. For the MRQR Kernel procedure, the data-driven mixing parameter overestimates the optimal value at low degrees of misspecification, but at γ=0.5 the optimal value is underestimated, and at γ=1.0, the data-driven value of the mixing parameter

underestimates the optimal value by approximately 20%. The data-driven mixing parameter for the MRQR LLR procedure does an excellent job in estimating the optimal values of the bandwidth for values of γ greater than zero. At γ=0.1, the optimal value is overestimated by 14%, the data-driven value of the mixing parameter is estimated to within 4%, 15%, and 2% error, relative to the optimal mixing parameter.

For kernel regression, as the number of doses increases, it seems that the efficiency for the data-driven technique to estimate the optimal bandwidth also increases. With seven doses and twenty subjects receiving each dose of the drug, the data-driven bandwidth overestimates the optimal bandwidth for kernel regression, but by a range of 8% to 15% for γ=0 through 0.5. At γ=1.0, the optimal bandwidth is overestimated by the kernel data-driven technique by 45%. As for the mixing parameter for the MRQR Kernel procedure, at γ=0, 0.1, the optimal value of the mixing parameter is overestimated. For γ=0.3, 0.5, and 1.0, the data-driven technique estimates the optimal mixing parameter with 86%, 76% and 90% efficiency, respectively.

The local linear regression procedure seems to yield the same results for seven doses as they did for five doses. That is, the data-driven technique for the bandwidth selection overestimates the optimal values across all degrees of misspecification. But the data-driven technique for the mixing parameter does extremely well for γ=0.3, 0.5, and 1.0.

Table 8.5.1.1 Average values of the bandwidth (b) and mixing parameter (λ) for N=20, and d=3, 5, and 7 doses. The **bold** values are the optimal ones chosen to minimize the theoretical mean squared errors.

			MRQR KERNEL		MRQR LLR	
N	D	γ	b	λ	b	λ
20	3	0.0	0.6260	0.0780	0.9769	0.4836
			0.2295	**0.3062**	**0.6304**	**1.0000**
		0.1	0.6369	0.0798	0.9637	0.5470
			0.2297	**0.2994**	**0.6319**	**1.0000**
		0.3	0.6074	0.0951	0.9724	0.4873
			0.2300	**0.2877**	**0.6336**	**1.0000**
		0.5	0.6414	0.1100	0.9637	0.5409
			0.2303	**0.2768**	**0.6355**	**1.0000**
		1.0	0.6145	0.1012	0.9721	0.5270
			0.2311	**0.2534**	**0.6400**	**1.0000**
	5	0.0	0.2089	0.1511	0.9687	0.3228
			0.1547	**0.0000**	**0.1665**	**0.0000**
		0.1	0.2127	0.1156	0.9558	0.3477
			0.1619	**0.0000**	**0.1920**	**0.3063**
		0.3	0.2300	0.1480	0.8837	0.5454
			0.1695	**0.3108**	**0.2899**	**0.5698**
		0.5	0.2524	0.2395	0.8339	0.7554
			0.1682	**0.4796**	**0.5038**	**0.8904**
		1.0	0.3394	0.7066	0.9180	0.9767
			0.1445	**0.8912**	**0.2327**	**1.0000**

Table 8.5.1.1 (continued)

			MRQR KERNEL		MRQR LLR	
N	D	γ	b	λ	b	λ
20	7	0.0	0.1518	0.2233	0.7866	0.3416
			0.1335	**0.0000**	**0.1454**	**0.0000**
		0.1	0.1541	0.2112	0.8309	0.4262
			0.1389	**0.0000**	**0.1642**	**0.2591**
		0.3	0.1563	0.2417	0.8375	0.5778
			0.1446	**0.2817**	**0.2401**	**0.5089**
		0.5	0.1639	0.3485	0.7938	0.7890
			0.1426	**0.4608**	**0.4605**	**0.7862**
		1.0	0.1754	0.8251	0.7799	0.9840
			0.1209	**0.9154**	**0.1723**	**1.0000**

§ 8.5.2 Data-Driven MSE Summary

Having obtained the data-driven and optimal bandwidths and mixing parameters for the procedures presented in this research, it is necessary to examine the effect of using the data-driven bandwidth on the average mean squared error (amse) over the true curve. **Table 8.5.2.1** lists the average mses for the (d=3, 5, 7, n=20) combinations for values of γ=0, 0.1, 0.3, 0.5, and 1.0. The bold values indicate the optimal average mses obtained using the optimal bandwidths and mixing parameters by minimizing the theoretical mse formulas given in **Chapter 7** and derived in Appendix B. There are several situations where the data-driven mses are actually smaller than the optimal values (γ=0, 0.1 for MLE and MRQR LLR). This apparent abnormality can be explained by the two primary sources of variability of these numbers: Monte Carlo variability for the simulated, data-driven values and small sample error involved with using asymptotic formulas to obtain the optimal values.

For the (d=3, n=20) combination the average mean squared errors using the data-driven bandwidths and mixing parameters for the nonparametric and model-robust procedures tend to overestimate the optimal amses. The mses obtained from the data-driven parameters for the Kernel, MRQR Kernel procedures overestimates the mses using the optimal parameters. As for the LLR procedure, the data-driven mses overestimate the optimal ones, but not by very much. In fact, across all values of γ, the data-driven mses for the LLR procedure nearly matches the optimal mses. The data-driven mses for the MRQR LLR procedure tend to overestimate the optimal values, as expected. It seems that although the data-driven bandwidths for the LLR procedure overestimates the optimal bandwidths, the mses are not effected, whereas use of the mixing parameter for the MRQR LLR procedure is not as robust.

As the number of doses increases to five, the amses obtained using the data-driven nonparametric procedures continue to overestimate the optimal ones, moreso for the Kernel procedure than the LLR procedure. That is, the LLR procedure continues to be robust to bandwidth selection, whereas the bandwidth is having a larger effect on the Kernel regression procedure to estimate the optimal mse. Also, for the MRQR Kernel procedure, the value of the data-driven mixing parameter overestimates the optimal mixing parameters by as much as 1.5 to 5 times. Since the data-driven values of the bandwidth and mixing parameters were well estimated by the LLR and MRQR Procedures, so were the mean squared errors. Just as the data-driven bandwidths and mixing parameters better estimated the optimal ones as the number of doses increased from five to seven, the same holds true for the mean squared errors.

Two additional features are worth noting. One, for d greater than 5 (and n=20) the simulated and theoretical mse results are nearly identical over all values of γ, indicating that the asymptotic formulas work extremely well under these conditions -- a pleasant surprise. Two, the PRESS* procedure used to select the bandwidth and mixing parameter performs well, especially at large values of γ. For small values

of γ, the MRQR LLR method using data-driven values, is no longer the optimal method. It is slightly inferior to MLE at γ=0.1 and slightly inferior to LLR at γ=0.3. Perhaps another selection procedure can be developed to eliminate this problem.

Table 8.5.2.1 Average mean squared error obtained using the data-driven bandwidths and mixing parameters. The **bold** values are the optimal mses.

N	D	γ	MLE	KERNEL	MRQR KERNEL	LLR	MRQR LLR
20	3	0.0	0.0051	0.0582	0.0119	0.0133	0.0090
			0.0059	**0.0158**	**0.0061**	**0.0118**	**0.0118**
		0.1	0.0043	0.0533	0.0108	0.0102	0.0075
			0.0061	**0.0126**	**0.0053**	**0.0088**	**0.0088**
		0.3	0.0053	0.0404	0.0117	0.0059	0.0055
			0.0080	**0.0079**	**0.0054**	**0.0047**	**0.0047**
		0.5	0.0069	0.0351	0.0130	0.0030	0.0047
			0.0118	**0.0053**	**0.0075**	**0.0026**	**0.0026**
		1.0	0.0229	0.0228	0.0253	0.0068	0.0145
			0.0297	**0.0082**	**0.0221**	**0.0069**	**0.0069**
	5	0.0	0.0030	0.0059	0.0041	0.0045	0.0037
			0.0030	**0.0042**	**0.0038**	**0.0041**	**0.0030**
		0.1	0.0030	0.0053	0.0051	0.0037	0.0034
			0.0030	**0.0042**	**0.0030**	**0.0040**	**0.0028**
		0.3	0.0037	0.0063	0.0052	0.0035	0.0040
			0.0035	**0.0044**	**0.0032**	**0.0031**	**0.0025**
		0.5	0.0047	0.0072	0.0053	0.0029	0.0037
			0.0045	**0.0048**	**0.0039**	**0.0023**	**0.0022**
		1.0	0.0114	0.0148	0.0138	0.0059	0.0061
			0.0111	**0.0070**	**0.0068**	**0.0054**	**0.0054**
	7	0.0	0.0021	0.0032	0.0027	0.0029	0.0025
			0.0020	**0.0031**	**0.0020**	**0.0028**	**0.0020**
		0.1	0.0020	0.0034	0.0028	0.0029	0.0026
			0.0021	**0.0031**	**0.0021**	**0.0028**	**0.0020**
		0.3	0.0028	0.0037	0.0030	0.0030	0.0032
			0.0025	**0.0032**	**0.0023**	**0.0022**	**0.0019**
		0.5	0.0037	0.0039	0.0038	0.0028	0.0031
			0.0036	**0.0036**	**0.0030**	**0.0016**	**0.0017**
		1.0	0.0101	0.0066	0.0072	0.0043	0.0044
			0.0100	**0.0053**	**0.0052**	**0.0044**	**0.0044**

§ 8.5.3 Effective Dose Estimation using Data-Driven Bandwidth and Mixing Parameter

In this section, the estimation of the effective doses will be compared across the procedures presented in this research in which the bandwidths and mixing parameters are chosen to minimize the PRESS* statistic. Along with effective dose estimation, the widths of the 95% confidence interval on the effective doses and the coverage probabilities of the true $ED_{100\alpha}$ will be given.

Tables 8.5.3.1-8.5.3.5 summarizes the effective dose estimation with three doses for γ=0, 0.1, 0.3, 0.5, and 1.0, respectively. Recall that with three doses and twenty subjects receiving each dose, the PRESS* did not estimate accurately the bandwidth or mixing parameter for either the nonparametric or model-robust procedures. Thus, it is of no surprise that the coverage probabilities are not as high as when the optimal bandwidths and mixing parameters were used earlier. Although, the median effective dose is well estimated by the logit, nonparametric, and model-robust procedures, the extreme doses, ED_{20} and ED_{80}, are not covered very well. Also at three doses, due to the bias problem in the kernel regression procedure, the coverage probability on the extreme doses is very low. In fact, the highest coverage probability across all values of γ for the ED_{20} is 29.9% and the highest coverage probability for the ED_{80} is 40.7%. The local linear regression procedure fairs only slightly better than the kernel procedure in estimating extreme doses, most notably at the larger values of γ. Because the coverage probabilities are poor for the nonparametric procedures, the mixing of the two procedures does not show any significant improvement in the coverages at the extreme doses. Note that with the nonparametric procedures, some of the average ED_{20} estimates are less than zero and some of the average ED_{80} estimates are greater than one. This is because there are no constraints used when applying the nonparametric procedures.

As for estimating effective doses using only three dose levels, the logit procedure performs very well at γ=0 and 0.1. Increasing γ to 0.3 and 0.5, the MRQR LLR procedure estimates the effective doses quite well. At γ=1.0, the LLR procedure estimates the extreme doses well, while the other procedures tend to overestimate the ED_{20} and underestimate the ED_{80}. Again, this inability of the procedures to estimate extreme doses is a result of the inability of the PRESS* procedure to choose the proper bandwidth. Recall that when the optimal bandwidth and mixing parameters were used (see **Table 8.3.5**), the coverages for the MRQR LLR procedure were very high and it overestimated the ED_{20} by 6% and estimated the ED_{80} with 99% accuracy.

Tables 8.5.3.6-8.5.3.10 are summaries of the effective dose estimation for the (d=5, n=20) combination for all values of γ. Increasing to five doses, the same observations are apparent. But with five doses, the Kernel procedure has relatively reasonable coverage on the effective doses for small degrees of misspecification (γ=0-0.5), and the LLR procedure has poor coverage probability on the extreme doses at γ=0 and 0.1, but increases its coverage to at least 75% at γ=0.3. If the model is correctly specified or

slightly misspecified (γ=0, 0.1, and 0.3) the logit procedure yields the highest coverage probability with very good estimates of the effective doses. But when the degree of misspecification increases to γ=0.5 and 1.0, the LLR and MRQR LLR has the highest coverage probabilities on the effective doses. In almost all instances, the median effective dose is covered extremely well by all of the procedures, but the extremes dose estimation tend to indicate how the curves are being fit to the data (whether it is over- or under-fitting in the boundaries). Note that for γ=0.3, 0.5 and 1.0, the MRQR LLR procedure estimates the effective doses to nearly 100% efficiency, slightly over-estimating each of them. In addition, for MRQR LLR, the confidence intervals obtained using the inverse estimation technique are quite conservative (intervals too narrow and the coverage probability less than the nominal 95% level) at the extremes, especially for smaller values of γ. The MLE method does a good job of maintaining high coverage probabilities (greater than 90%), but, as γ increases, the width of the intervals tend to be too wide (compared to MRQR LLR) and the ED_{20} is substantially overestimated and the ED_{80} is substantially underestimated.

At seven doses and twenty subjects receiving each dose, all of the procedures perform fairly well with respect to coverages and effective dose estimation except the LLR procedure at small values of γ. Again, this is a result of poor bandwidth selection. Although it seemed that the LLR procedure was robust to the size of the bandwidth when comparing the mean squared errors, the bandwidth and mixing parameter is having a significant impact on effective dose estimation and coverage probability. See **Tables 8.5.3.11-8.5.3.15** for summaries of the effective dose estimation with d=7 and n=20.

Table 8.5.3.1 Effective dose estimation summary for the (d=3, n=20, γ=0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3615, ED_{50}=0.5, and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3430	0.2267	82.0
	0.5	0	0.5015	0.1565	86.2
	0.8	0	0.6601	0.2275	78.6
Kernel	0.2	81	-0.0572	0.4163	10.5
	0.5	0	0.5042	0.3751	93.6
	0.8	84	1.0589	0.4147	10.3
MRQR Kernel	0.2	35	0.3383	0.2272	80.5
	0.5	0	0.5031	0.1943	86.2
	0.8	38	0.6604	0.2272	77.7
LLR	0.2	0	0.2451	0.1515	8.4
	0.5	0	0.5014	0.1511	94.0
	0.8	0	0.7572	0.1485	7.8
MRQR LLR	0.2	0	0.2969	0.1920	57.8
	0.5	0	0.5015	0.1673	92.2
	0.8	0	0.7059	0.1908	57.4
Spearman Karber	0.5	0	0.4710	0.1713	89.2
TMA	0.5	0	0.5024	0.2125	95.2

Table 8.5.3.2 Effective dose estimation summary for the (d=3, n=20, γ=0.1) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3445, ED_{50}=0.5, and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3364	0.2317	88.2
	0.5	0	0.4997	0.1596	90.4
	0.8	0	0.6631	0.2313	87.2
Kernel	0.2	213	-0.0596	0.4312	7.7
	0.5	1	0.5009	0.3904	95.0
	0.8	215	1.0597	0.4217	8.8
MRQR Kernel	0.2	37	0.3335	0.2332	87.3
	0.5	1	0.5005	0.1977	90.4
	0.8	38	0.6643	0.2320	86.6
LLR	0.2	0	0.2417	0.1549	18.0
	0.5	0	0.7578	0.1550	95.2
	0.8	0	0.4998	0.1540	18.0
MRQR LLR	0.2	0	0.2857	0.1909	55.6
	0.5	0	0.4996	0.1702	94.4
	0.8	0	0.7133	0.1903	55.2
Spearman Karber	0.5	0	0.4647	0.1720	88.8
TMA	0.5	0	0.4999	0.2118	96.8

Table 8.5.3.3 Effective dose estimation summary for the (d=3, n=20, γ=0.3) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3055, ED_{50}=0.5, and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3402	0.2301	94.4
	0.5	0	0.5025	0.1594	95.0
	0.8	0	0.6648	0.2310	95.0
Kernel	0.2	180	-0.0423	0.4142	23.1
	0.5	0	0.5037	0.3660	94.0
	0.8	187	1.0554	0.4145	20.4
MRQR Kernel	0.2	45	0.3363	0.2311	94.3
	0.5	0	0.5031	0.2042	85.0
	0.8	45	0.6674	0.2318	94.7
LLR	0.2	0	0.2440	0.1536	64.0
	0.5	0	0.5020	0.1542	94.2
	0.8	0	0.6400	0.1546	61.2
MRQR LLR	0.2	0	0.2957	0.1941	75.4
	0.5	0	0.5025	0.1711	92.6
	0.8	0	0.7093	0.1560	77.6
Spearman Karber	0.5	0	0.4662	0.1715	88.0
TMA	0.5	0	0.5025	0.2157	96.6

Table 8.5.3.4 Effective dose estimation summary for the (d=3, n=20, γ=0.5) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2730, ED_{50}=0.5, and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3334	0.2326	96.0
	0.5	0	0.4972	0.1597	89.0
	0.8	0	0.6609	0.2312	95.0
Kernel	0.2	221	-0.0776	0.4324	14.7
	0.5	0	0.4970	0.3947	95.2
	0.8	218	1.0647	0.4131	21.6
MRQR Kernel	0.2	54	0.3292	0.2326	95.3
	0.5	0	0.4976	0.2134	89.0
	0.8	52	0.6635	0.2322	93.8
LLR	0.2	0	0.2395	0.1566	89.2
	0.5	0	0.4978	0.1553	96.0
	0.8	0	0.7561	0.1531	92.4
MRQR LLR	0.2	0	0.2842	0.1933	91.0
	0.5	0	0.4973	0.1692	94.4
	0.8	0	0.7105	0.1898	87.8
Spearman Karber	0.5	0	0.4637	0.1720	86.6
TMA	0.5	0	0.4977	0.2109	95.8

Table 8.5.3.5 Effective dose estimation summary for the (d=3, n=20, γ=1.0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2300, ED_{50}=0.5, and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3319	0.2344	88.6
	0.5	0	0.4973	0.1614	86.8
	0.8	0	0.6626	0.2327	75.4
Kernel	0.2	192	-0.0630	0.4243	29.9
	0.5	4	0.4934	0.3707	93.3
	0.8	215	1.0542	0.4092	40.7
MRQR Kernel	0.2	49	0.3299	0.2357	87.6
	0.5	4	0.4960	0.2072	86.7
	0.8	50	0.6641	0.2331	76.7
LLR	0.2	0	0.2387	0.1570	94.4
	0.5	0	0.4978	0.1559	94.0
	0.8	0	0.7571	0.1565	93.0
MRQR LLR	0.2	0	0.2849	0.1954	81.2
	0.5	0	0.4975	0.1717	93.0
	0.8	0	0.7102	0.1933	77.4
Spearman Karber	0.5	0	0.4600	0.1715	86.0
TMA	0.5	0	0.4971	0.2142	95.4

Table 8.5.3.6 Effective dose estimation summary for the (d=5, n=20, γ=0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3615, ED_{50}=0.5, and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3622	0.1587	90.0
	0.5	0	0.4983	0.1170	95.0
	0.8	0	0.6344	0.1587	94.0
Kernel	0.2	25	0.3045	0.1609	72.2
	0.5	0	0.4992	0.1363	95.0
	0.8	16	0.6936	0.1576	73.1
MRQR Kernel	0.2	6	0.3532	0.1536	86.6
	0.5	0	0.4978	0.1161	94.0
	0.8	4	0.6428	0.1554	91.7
LLR	0.2	0	0.2773	0.1015	3.6
	0.5	0	0.4986	0.0965	95.8
	0.8	0	0.7200	0.1001	6.2
MRQR LLR	0.2	0	0.3383	0.1367	73.4
	0.5	0	0.4984	0.1148	95.4
	0.8	0	0.6585	0.1351	72.4
Spearman Karber	0.5	0	0.4835	0.1214	86.4
TMA	0.5	0	0.4990	0.1042	88.2

Table 8.5.3.7 Effective dose estimation summary for the (d=5, n=20, γ=0.1) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3455, ED_{50}=0.5, and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.2434	0.1669	93.2
	0.5	0	0.5004	0.1222	95.0
	0.8	0	0.6455	0.1670	93.0
Kernel	0.2	20	0.2935	0.1747	81.0
	0.5	0	0.5015	0.1439	93.4
	0.8	26	0.7068	0.1671	81.4
MRQR Kernel	0.2	3	0.3489	0.1636	92.2
	0.5	0	0.5003	0.1212	94.6
	0.8	5	0.6522	0.1630	92.1
LLR	0.2	0	0.2750	0.1075	20.0
	0.5	0	0.5004	0.1018	95.8
	0.8	0	0.7256	0.1069	20.8
MRQR LLR	0.2	0	0.3307	0.1416	77.6
	0.5	0	0.5003	0.1197	96.0
	0.8	0	0.6700	0.1406	76.2
Spearman Karber	0.5	0	0.4814	0.1260	86.6
TMA	0.5	0	0.5003	0.1055	88.0

Table 8.5.3.8 Effective dose estimation summary for the (d=5, n=20, γ=0.3) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3055, ED_{50}=0.5, and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3351	0.1871	92.8
	0.5	0	0.5002	0.1340	91.4
	0.8	0	0.6653	0.1874	91.4
Kernel	0.2	37	0.2595	0.1962	85.1
	0.5	0	0.5013	0.1622	90.0
	0.8	27	0.7410	0.1954	83.5
MRQR Kernel	0.2	11	0.3213	0.1846	89.2
	0.5	0	0.5005	0.1369	90.2
	0.8	7	0.6797	0.1859	86.2
LLR	0.2	0	0.2652	0.1222	76.0
	0.5	0	0.5006	0.1147	92.0
	0.8	0	0.7352	0.1203	74.6
MRQR LLR	0.2	0	0.3009	0.1461	83.0
	0.5	0	0.5004	0.1272	93.0
	0.8	0	0.6989	0.1443	76.2
Spearman Karber	0.5	0	0.4809	0.1368	83.8
TMA	0.5	0	0.5006	0.1070	79.2

Table 8.5.3.9 Effective dose estimation summary for the (d=5, n=20, γ=0.5) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2730, ED_{50}=0.5, and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3177	0.2070	89.4
	0.5	0	0.5019	0.1443	92.4
	0.8	0	0.6861	0.2075	91.0
Kernel	0.2	39	0.2181	0.2254	84.6
	0.5	0	0.5036	0.1864	93.4
	0.8	28	0.7870	0.2244	81.4
MRQR Kernel	0.2	16	0.2822	0.2077	81.0
	0.5	0	0.5031	0.1555	92.0
	0.8	11	0.7227	0.2073	81.2
LLR	0.2	2	0.2557	0.1350	93.4
	0.5	0	0.5019	0.1278	93.4
	0.8	1	0.7470	0.1339	91.2
MRQR LLR	0.2	2	0.2739	0.1481	82.7
	0.5	0	0.5020	0.1342	93.8
	0.8	1	0.7292	0.1464	84.6
Spearman Karber	0.5	0	0.4819	0.1460	86.2
TMA	0.5	0	0.5027	0.1078	78.4

Table 8.5.3.10 Effective dose estimation summary for the (d=5, n=20, γ=1.0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2300, ED_{50}=0.5, and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.2662	0.3173	85.2
	0.5	0	0.5010	0.1945	95.0
	0.8	0	0.7358	0.3180	83.6
Kernel	0.2	86	0.0561	0.3088	59.7
	0.5	2	0.5025	0.2662	93.4
	0.8	76	0.9482	0.3138	59.7
MRQR Kernel	0.2	73	0.0935	0.2933	60.2
	0.5	2	0.5025	0.2481	93.0
	0.8	66	0.9107	0.2952	60.8
LLR	0.2	4	0.2218	0.1684	94.6
	0.5	0	0.5010	0.1522	94.0
	0.8	1	0.7796	0.1677	93.6
MRQR LLR	0.2	4	0.2233	0.1697	93.8
	0.5	0	0.5010	0.1528	94.0
	0.8	1	0.7781	0.1689	92.0
Spearman Karber	0.5	0	0.4781	0.1994	92.8
TMA	0.5	0	0.5016	0.1254	73.0

Table 8.5.3.11 Effective dose estimation summary for the (d=7, n=20, γ=0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3615, ED_{50}=0.5, and ED_{80}=0.6385.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3653	0.1306	94.4
	0.5	0	0.5006	0.0976	93.0
	0.8	0	0.6359	0.1307	92.2
Kernel	0.2	21	0.3329	0.1374	89.6
	0.5	0	0.5001	0.1167	94.8
	0.8	20	0.6682	0.1391	87.1
MRQR Kernel	0.2	12	0.3598	0.1286	93.2
	0.5	0	0.5005	0.0995	92.6
	0.8	10	0.6414	0.1296	91.0
LLR	0.2	20	0.3011	0.0930	20.2
	0.5	0	0.5004	0.0844	93.8
	0.8	14	0.7000	0.0942	21.2
MRQR LLR	0.2	8	0.3461	0.1152	76.0
	0.5	0	0.5006	0.0942	93.8
	0.8	5	0.6552	0.1161	75.4
Spearman Karber	0.5	0	0.4822	0.0970	80.0
TMA	0.5	0	0.4998	0.0669	74.4

Table 8.5.3.12 Effective dose estimation summary for the (d=7, n=20, γ=0.1) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3455, ED_{50}=0.5, and ED_{80}=0.6355.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3528	0.1393	94.4
	0.5	0	0.5003	0.1028	94.6
	0.8	0	0.6478	0.1393	96.0
Kernel	0.2	18	0.3157	0.1483	89.6
	0.5	0	0.5005	0.1300	93.0
	0.8	22	0.6841	0.1481	90.0
MRQR Kernel	0.2	7	0.3467	0.1369	92.9
	0.5	0	0.5005	0.1065	93.0
	0.8	10	0.6537	0.1365	93.7
LLR	0.2	10	0.2894	0.0977	26.7
	0.5	0	0.5000	0.0889	94.8
	0.8	15	0.7110	0.0980	26.0
MRQR LLR	0.2	5	0.3289	0.1183	73.9
	0.5	0	0.5002	0.0985	94.8
	0.8	5	0.6721	0.1190	74.7
Spearman Karber	0.5	0	0.4823	0.1017	80.8
TMA	0.5	0	0.5000	0.0654	70.4

Table 8.5.3.13 Effective dose estimation summary for the (d=7, n=20, γ=0.3) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.3055, ED_{50}=0.5, and ED_{80}=0.6955.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3330	0.1524	93.4
	0.5	0	0.5001	0.1100	92.2
	0.8	0	0.6673	0.1527	86.8
Kernel	0.2	32	0.2902	0.1600	92.3
	0.5	0	0.5008	0.1578	94.0
	0.8	27	0.7100	0.1602	90.5
MRQR Kernel	0.2	15	0.3236	0.1506	90.9
	0.5	0	0.5003	0.1218	91.6
	0.8	13	0.6772	0.1494	85.4
LLR	0.2	2	0.2743	0.1070	79.5
	0.5	0	0.5002	0.0987	93.0
	0.8	4	0.7256	0.1070	75.6
MRQR LLR	0.2	2	0.3013	0.1227	82.1
	0.5	0	0.5001	0.1060	92.6
	0.8	1	0.6982	0.1229	73.9
Spearman Karber	0.5	0	0.4796	0.1070	80.2
TMA	0.5	0	0.5001	0.0630	60.4

Table 8.5.3.14 Effective dose estimation summary for the (d=7, n=20, γ=0.5) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2730, ED_{50}=0.5, and ED_{80}=0.7270.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.3124	0.1718	90.4
	0.5	0	0.4993	0.1204	93.0
	0.8	0	0.6862	0.1715	88.2
Kernel	0.2	28	0.2602	0.1706	90.9
	0.5	0	0.4988	0.1826	92.0
	0.8	22	0.7388	0.1743	92.7
MRQR Kernel	0.2	18	0.2935	0.1637	86.1
	0.5	0	0.4990	0.1452	92.4
	0.8	14	0.7057	0.1651	87.2
LLR	0.2	0	0.2612	0.1164	94.4
	0.5	0	0.4988	0.1100	93.4
	0.8	2	0.7385	0.1180	95.2
MRQR LLR	0.2	2	0.2737	0.1247	87.5
	0.5	0	0.4987	0.1138	93.4
	0.8	0	0.7260	0.1264	86.6
Spearman Karber	0.5	0	0.4798	0.1155	82.0
TMA	0.5	0	0.4987	0.0609	50.2

Table 8.5.3.15 Effective dose estimation summary for the (d=7, n=20, γ=1.0) combination using PRESS* to select the bandwidth and mixing parameter. The true effective doses are: ED_{20}=0.2300, ED_{50}=0.5, and ED_{80}=0.7705.

Method	α for $ED_{100\alpha}$	# Incalculable	Mean $ED_{100\alpha}$	Mean Width	Observed Coverage Probability (%)
MLE	0.2	0	0.2564	0.2450	94.2
	0.5	0	0.4990	0.1548	94.8
	0.8	0	0.7415	0.2444	95.8
Kernel	0.2	39	0.1897	0.1802	87.0
	0.5	0	0.4989	0.2787	93.8
	0.8	36	0.8068	0.1767	85.8
MRQR Kernel	0.2	34	0.1996	0.1816	86.1
	0.5	0	0.4983	0.2610	93.6
	0.8	30	0.7973	0.1784	84.7
LLR	0.2	5	0.2213	0.1460	93.9
	0.5	0	0.5000	0.1521	93.6
	0.8	6	0.7760	0.1450	95.5
MRQR LLR	0.2	5	0.2221	0.1467	93.9
	0.5	0	0.5000	0.1523	93.6
	0.8	6	0.7752	0.1456	93.9
Spearman Karber	0.5	0	0.4758	0.1384	83.2
TMA	0.5	0	0.4992	0.0575	29.4

§ 8.5.4 Estimation of Proportion Responding Using Data-Driven b and λ

Just as the summaries before, when examining the ability of the logit, nonparametric, and model-robust procedures to predict the proportion responding using the data-driven bandwidth and mixing parameter, it is really an inquiry in how well the PRESS* procedure selects the bandwidth and mixing parameter and what effect it has on predicting the proportion responding. In this section, **Tables 8.5.4.1-8.5.4.5** represent a summary of the mean response for the (d=5, n=20) scenario, for γ=0, 0.1, 0.3, 0.5, and 1.0, respectively. Following are a few interesting points regarding the estimation of the proportion responding.

1. At low degrees of model-misspecification, $\gamma \leq 0.3$, the MRQR Kernel procedure has an average λ near zero, thus giving more weight to the logistic regression fit to the data. Note that the coverage probabilities are high in each case (γ=0, 0.1, and 0.3).

2. The MRQR LLR procedure performs well at γ=0, in which the mixing parameter is zero, but at γ=0.1, the MRQR LLR procedure has very poor coverage probabilities at dose levels 0.3 and 0.7. This is perhaps due to the linear nature in the curve, as mentioned earlier.

3. At γ=0.5, the MRQR LLR procedure has the highest coverage probabilities in the extremes of the data. At the 0.5 dose level, the logit procedure has a slightly higher coverage probability, but all of the procedures continue to fit well in the middle of the data. The MRQR LLR procedure also predicts the true response extremely well for this particular combination.

4. At the highest degree of model-misspecification, none of the procedures tend to have reasonable coverage probabilities. This may be a reflection on the PRESS* procedure to select the bandwidth and mixing parameter. Another possible explanation for the poor coverage probabilities is in the variance formulas used by the nonparametric and model-robust methods. The formulas, given in **Chapter 7**, are all based on a fixed bandwidth and mixing parameters. Neither is fixed in the case when the bandwidth and mixing parameters are chosen by data-driven methods, as they are in the simulation. Note in **Table 8.4.10** that the MRQR LLR procedure has good coverage probabilities when using the optimal (and fixed) values of the bandwidth and mixing parameters.

Table 8.5.4.1 Summary of mean response at the (d=5, n=20, γ=0) combination for 500 Monte Carlo simulations and the data-driven values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability
MLE	0.1	0.0201	0.0180	0.0646	86.2
	0.3	0.1199	0.1192	0.2213	91.0
	0.5	0.5037	0.5000	0.3543	95.8
	0.7	0.8841	0.8808	0.2172	90.2
	0.9	0.9806	0.9820	0.0624	84.2
Kernel	0.1	0.0590	0.0180	0.1086	82.0
	0.3	0.1939	0.1192	0.1796	66.2
	0.5	0.5011	0.5000	0.2490	95.0
	0.7	0.8094	0.8808	0.1781	71.4
	0.9	0.9450	0.9820	0.1026	84.4
MRQR Kernel	0.1	0.0269	0.0180	0.0646	86.2
	0.3	0.1303	0.1192	0.2213	91.0
	0.5	0.5043	0.5000	0.3542	95.8
	0.7	0.8735	0.8808	0.2172	90.2
	0.9	0.9733	0.9820	0.0624	84.2
LLR	0.1	-0.0393	0.0180	0.1460	100.0
	0.3	0.2305	0.1192	0.1337	4.4
	0.5	0.5020	0.5000	0.1292	95.8
	0.7	0.7730	0.8808	0.1315	6.8
	0.9	1.0418	0.9820	0.1418	100.0
MRQR LLR	0.1	0.0135	0.0180	0.0646	86.2
	0.3	0.1514	0.1192	0.2213	91.0
	0.5	0.5031	0.5000	0.3541	95.8
	0.7	0.8525	0.8808	0.2171	90.2
	0.9	0.9874	0.9820	0.0625	84.2

Table 8.5.4.2 Summary of mean response at the (d=5, n=20, γ=0.1) combination for 500 Monte Carlo simulations and the data-driven values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability
MLE	0.1	0.0244	0.0186	0.0759	91.6
	0.3	0.1303	0.1438	0.2331	90.8
	0.5	0.4992	0.5000	0.3531	96.2
	0.7	0.8689	0.8562	0.2339	89.6
	0.9	0.9753	0.9814	0.0764	91.2
Kernel	0.1	0.0659	0.0186	0.1131	75.8
	0.3	0.2046	0.1438	0.1879	79.0
	0.5	0.4975	0.5000	0.2481	93.4
	0.7	0.7955	0.8562	0.1849	76.6
	0.9	0.9350	0.9814	0.1121	76.2
MRQR Kernel	0.1	0.0301	0.0186	0.0759	91.6
	0.3	0.1381	0.1438	0.2331	90.8
	0.5	0.4994	0.5000	0.3531	96.2
	0.7	0.8613	0.8562	0.2339	89.6
	0.9	0.9693	0.9814	0.0764	91.2
LLR	0.1	-0.0310	0.0186	0.1508	100.0
	0.3	0.2331	0.1438	0.1384	23.6
	0.5	0.4995	0.5000	0.1340	95.8
	0.7	0.7662	0.8562	0.1374	23.6
	0.9	1.0309	0.9814	0.1497	100.0
MRQR LLR	0.1	0.0160	0.0186	0.1498	100.0
	0.3	0.1619	0.1438	0.1391	24.8
	0.5	0.4996	0.5000	0.1357	95.8
	0.7	0.8375	0.8562	0.1379	24.0
	0.9	0.9833	0.9814	0.1488	100.0

Table 8.5.4.3 Summary of mean response at the (d=5, n=20, γ=0.3) combination for 500 Monte Carlo simulations and the data-driven values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability
MLE	0.1	0.0367	0.0197	0.1084	96.2
	0.3	0.1583	0.1931	0.2624	86.0
	0.5	0.4999	0.5000	0.3498	94.0
	0.7	0.8416	0.8069	0.2623	84.6
	0.9	0.9360	0.9803	0.1056	97.4
Kernel	0.1	0.0914	0.0197	0.1291	50.8
	0.3	0.2363	0.1931	0.1973	82.8
	0.5	0.4989	0.5000	0.2395	90.0
	0.7	0.7633	0.8069	0.1962	83.0
	0.9	0.9090	0.9803	0.1272	54.2
MRQR Kernel	0.1	0.0472	0.0197	0.1084	96.2
	0.3	0.1699	0.1931	0.2624	86.0
	0.5	0.5010	0.5000	0.3498	94.0
	0.7	0.8296	0.8069	0.2623	84.6
	0.9	0.9509	0.9803	0.1086	97.4
LLR	0.1	-0.0086	0.0197	0.1608	99.4
	0.3	0.2445	0.1931	0.1488	76.8
	0.5	0.4995	0.5000	0.1450	92.0
	0.7	0.7548	0.8069	0.1483	74.4
	0.9	1.0093	0.9803	0.1577	99.8
MRQR LLR	0.1	0.0217	0.0197	0.1598	99.4
	0.3	0.1998	0.1931	0.1495	76.8
	0.5	0.5002	0.5000	0.1468	92.0
	0.7	0.8001	0.8069	0.1489	74.4
	0.9	0.9781	0.9803	0.1568	99.8

Table 8.5.4.4 Summary of mean response at the (d=5, n=20, γ=0.5) combination for 500 Monte Carlo simulations and the data-driven values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability
MLE	0.1	0.0493	0.0209	0.1395	97.4
	0.3	0.1799	0.2424	0.2796	77.0
	0.5	0.4963	0.5000	0.3412	95.6
	0.7	0.8160	0.7576	0.2830	80.6
	0.9	0.9495	0.9792	0.1425	97.8
Kernel	0.1	0.1196	0.0209	0.1418	27.2
	0.3	0.2654	0.2424	0.2025	88.4
	0.5	0.4955	0.5000	0.2325	93.4
	0.7	0.7301	0.7576	0.2028	87.2
	0.9	0.8795	0.9792	0.1398	28.4
MRQR Kernel	0.1	0.0737	0.0209	0.1418	27.2
	0.3	0.2026	0.2424	0.2025	88.4
	0.5	0.4962	0.5000	0.2325	93.4
	0.7	0.7929	0.7576	0.2028	87.2
	0.9	0.9249	0.9792	0.1398	28.4
LLR	0.1	0.0234	0.0209	0.1684	99.2
	0.3	0.2539	0.2424	0.1581	94.2
	0.5	0.4977	0.5000	0.1548	93.4
	0.7	0.7423	0.7576	0.1578	93.4
	0.9	0.9764	0.9792	0.1657	99.2
MRQR LLR	0.1	0.0284	0.0209	0.1684	99.2
	0.3	0.2314	0.2424	0.1581	94.2
	0.5	0.4970	0.5000	0.1548	93.4
	0.7	0.7646	0.7576	0.1578	93.4
	0.9	0.9712	0.9792	0.1657	99.2

Table 8.5.4.5 Summary of mean response at the (d=5, n=20, γ=1.0) combination for 500 Monte Carlo simulations and the data-driven values of the bandwidth and mixing parameter.

Method	Dose	Mean Response	True Response	Mean Width	Observed Coverage Probability
MLE	0.1	0.0877	0.0237	0.3670	97.0
	0.3	0.2336	0.3656	0.5269	71.0
	0.5	0.4986	0.5000	0.5489	98.4
	0.7	0.7644	0.6344	0.5263	74.0
	0.9	0.9114	0.9763	0.3666	97.2
Kernel	0.1	0.2124	0.0237	0.1647	3.0
	0.3	0.3396	0.3656	0.1984	83.4
	0.5	0.4987	0.5000	0.2102	93.4
	0.7	0.6584	0.6344	0.1990	81.4
	0.9	0.7862	0.9763	0.1642	3.4
MRQR Kernel	0.1	0.1878	0.0237	0.1647	3.0
	0.3	0.3135	0.3656	0.1984	83.4
	0.5	0.4982	0.5000	0.2102	93.4
	0.7	0.6845	0.6344	0.1990	81.4
	0.9	0.8113	0.9763	0.1642	3.4
LLR	0.1	0.0665	0.0237	0.1898	88.6
	0.3	0.2834	0.3656	0.1689	52.2
	0.5	0.4989	0.5000	0.1613	94.0
	0.7	0.7147	0.6344	0.1689	55.0
	0.9	0.9330	0.9763	0.1887	87.8
MRQR LLR	0.1	0.0685	0.0237	0.1898	88.2
	0.3	0.2816	0.3656	0.1689	52.2
	0.5	0.4990	0.5000	0.1613	94.0
	0.7	0.7165	0.6344	0.1689	55.0
	0.9	0.9311	0.9763	0.1887	87.8

§ 8.5.5 The Chi-square Statistic

Table 8.5.5.1 is a summary of the average chi-square statistic for the (d=3, 5, 7, n=20) combinations for γ=0, 0.1, 0.3, 0.5, and 1.0. Recall that the chi-square statistic is a weighted error sums of squares given by

$$\chi^2 = \Sigma \frac{n_i\left(p_i - \hat{P}_i\right)^2}{\hat{P}_i\left(1-\hat{P}_i\right)}$$

which is approximately distributed with a chi-square distribution with (*d-2*) degrees of freedom.
This section is primarily an examination of the chi-square lack-of-fit test under a variety of conditions.

First of all, for the (d=3, n=20) scenario, it is interesting that for the MLE method for all of the values of γ considered, there is not one occurrence of a significant lack-of-fit out of 500 Monte Carlo repetitions for each value of γ. At α=0.10, one would expect to see (0.10x500)=50 out of 500 tests to detect lack-of-fit for the MLE. This is a reflection of the lack of power of the chi-square test as well as the lack of information available when using only three dose levels. Apparently, the chi-square distribution is not the appropriate distribution for this statistic at d=3. Also note that the logit procedure has the smallest average chi-square statistic at three dose levels.

Increasing the number of doses to d=5, the lack-of-fit occurrences for the MLE method increases as the degree of model-misspecification increases. Note that for γ=1.0, the highest degree of misspecification, the chi-square lack-of-fit test was significant 303 out of a possible 500 Monte Carlo repetitions (61%). Note that for (d=5, n=20), at low degrees of model-misspecification (γ=0, 0.1), the MRQR Kernel procedure has the smallest average chi-square statistic, and the MRQR Kernel and the logit procedure has approximately the same average chi-square statistic at γ=0.3. As the degree of model-misspecification increases to γ=0.5 and 1.0, the MRQR LLR procedure has the smallest average chi-square statistic.

For the (d=7, n=20) combination, the MRQR Kernel procedure has the smallest chi-square statistic across all values of γ. Also, the number of occurrences in which the chi-square test was significant differed slightly from the (d=5, n=20) scenario. Especially at γ=1.0, there were only 271 Monte Carlo situations which had significant lack-of-fit as opposed to the 303 with five doses.

Also, note that for d=5 and d=7 under the correctly specified model (γ=0), the chi-square test detects slightly fewer significant lack-of-fits for the MLE method than it should at the 10% level of significance. The LLR and MRQR LLR procedures would detect far too many significant lack-of-fits than would be desirable at both d=5 and d=7 design points. For example, with d=5, the average chi-square statistic for the LLR procedure is 14.4775, compared with the critical value of 6.25.

Table 8.5.5.1 Summary of the average chi-square statistics for the (d=3, 5, 7, n=20) combinations. γ denotes the degree of misspecification, #LOF indicates the number of times the chi-square test had significant lack-of-fit out of 500 Monte Carlo runs.

N	D	γ	# LOF	MLE	KERNEL	MRQR KERNEL	LLR	MRQR LLR
20	3	0.0	0	0.8665	13.2899	2.5185	1.8995	1.3497
		0.1	0	0.8844	13.4645	2.4804	1.4647	1.1950
		0.3	0	0.8378	12.6011	2.7256	1.6656	1.2270
		0.5	0	0.9403	13.6470	3.1643	1.8792	1.4729
		1.0	0	0.8534	12.7130	2.9737	1.7912	1.3755
	5	0.0	37	2.7493	3.5769	2.6244	14.4775	5.9434
		0.1	43	2.7380	3.5873	2.5778	13.7198	6.1300
		0.3	61	3.3807	4.3143	3.4005	16.4820	4.5264
		0.5	72	3.8563	5.2827	4.2836	5.3359	3.3953
		1.0	303	7.9375	10.4735	9.9232	5.7631	4.9484
	7	0.0	41	4.8306	3.4600	4.0716	17.8514	9.5509
		0.1	45	4.9625	3.4957	4.2072	16.0589	7.8044
		0.3	35	4.9734	3.5294	4.2342	14.4419	8.0644
		0.5	83	6.1246	4.2785	5.0681	9.5867	6.9814
		1.0	271	10.5878	6.3149	6.6768	7.7295	7.5382

§ 8.5.6 The PRESS* Statistic

Recall that the PRESS* statistic was used in this research as a penalized cross-validation procedure in selecting the bandwidth for the nonparametric procedures and the mixing parameter for the model-robust procedures. **Table 8.5.6.1** lists the average PRESS* statistics for the 500 Monte Carlo repetitions at the (d=3, 5, 7, n=20) combinations for γ=0, 0.1, 0.3, 0.5, and 1.0. Note that for all of the (d, n, γ) combinations listed in Table 8.5.6.1, the MRQR LLR procedure has the smallest PRESS* statistic. As the number of dose levels increase, the PRESS* statistic for MRQR LLR decreases at the corresponding values of γ. The PRESS* statistic for the logit procedure tends to increase as the degree of model misspecification increases, but decreases as the number of dose levels increases at the corresponding values of γ.

For the (d=3, n=20) combination, the kernel procedure has extremely large PRESS* statistics. Recall that the kernel procedure has boundary problems which could be a direct reflection on the PRESS* statistic. As the number of doses increases, the PRESS* statistic for the kernel and MRQR Kernel procedures tend to get smaller.

Note that the model-robust procedures continue to reduce the PRESS* statistic with respect to its corresponding nonparametric procedure and the logit procedure. That is, for each value of (d, n, γ) combination, the MRQR Kernel and MRQR LLR procedures have smaller PRESS* statistics than the logit and kernel, and the logit and LLR procedures, respectively, showing once again that mixing is beneficial to fitting a curve to the data.

Table 8.5.6.1 Summary of the average PRESS* statistics for the (d=3, 5, 7, n=20) combinations. γ denotes the degree of misspecification.

N	D	γ	MLE	KERNEL	MRQR KERNEL	LLR	MRQR LLR
20	3	0.0	340.64	1594.75	281.96	319.54	150.00
		0.1	364.82	1544.69	302.71	283.90	155.18
		0.3	350.57	1511.48	284.52	283.61	129.63
		0.5	435.72	1580.89	352.76	275.24	167.66
		1.0	377.52	1468.82	306.75	282.81	140.04
	5	0.0	178.70	206.41	66.69	153.75	45.75
		0.1	158.50	211.73	58.65	116.70	33.67
		0.3	211.83	287.46	110.78	63.69	22.19
		0.5	402.56	365.67	185.99	46.30	29.05
		1.0	1622.43	583.22	533.53	129.84	129.02
	7	0.0	92.03	60.48	30.71	71.72	27.56
		0.1	83.19	70.62	33.54	55.37	24.18
		0.3	114.07	78.58	41.73	34.15	17.22
		0.5	164.26	110.43	73.98	20.23	14.65
		1.0	722.86	184.60	179.01	30.96	30.65

CHAPTER 9
EXAMPLES OF MRQR APPLIED TO QUANTAL BIOASSAY

§9.1 Examples of MRQR Applied to Quantal Bioassay

In this section, the MRQR procedure will be applied to the two data sets used previously for the logit and nonparametric procedures presented in this research. Because the nonparametric portion of the model can be either kernel or local linear regression, the logistic model will be combined with both procedures. That is, logit procedure with kernel regression will be denoted MRQR Kernel, and logit procedure combined with LLR will be denoted MRQR LLR. In each procedure the bandwidth and the mixing parameter will be selected to minimize the $PRESS^{*}(b)$ statistic. Comparisons will be made between the logistic, kernel, LLR, MRQR Kernel, and MRQR LLR with respect to the model degrees of freedom, the chi-square statistic, mean squared error and width of the confidence intervals on the ED_{20}, ED_{50}, and ED_{80}. For the logit mixture data, one can base the adequacy of the fit on how close the estimate of the effective doses are with respect to their true values.

§9.2.A Example 7

The fly data set in **Table 2.2.1** was fit using the MRQR procedure. **Table 9.2.1** contains summary information from the different fits to the data. Plots of the MRQR Kernel and MRQR LLR fits can be found in **Figure 9.2.1** and **Figure 9.2.2**, respectively.

Table 9.2. 1 Summary information of fits to fly dose-response data. The 95% confidence interval estimates on effective doses are given in parentheses.

Procedure	χ^2	Dfmodel	MSE	ED_{20}	ED_{50}	ED_{80}
Logistic	0.7351	2.0000	0.1470	0.1201 (0.0951,0.1423)	0.2304 (0.2019,0.2610)	0.4420 (0.3802,0.5398)
Kernel	1.7692	3.8265	0.5575	0.0973 (0.0324,0.1321)	0.2457 (0.2059,0.2956)	0.4403 (0.3898,0.5467)
LLR	2.0030	3.7125	0.5395	0.1186 (0.0324,0.1360)	0.2371 (0.2056,0.2728)	0.4925 (0.4403,0.5786)
MRQR Kernel	0.7250	2.8526	0.2021	0.1125 (0.0789,0.1391)	0.2372 (0.2051,0.2720)	0.4285 (0.3871,0.5292)
MRQR LLR	0.6315	2.5281	0.1412	0.1205 (0.0324,0.1408)	0.2328 (0.2037,0.2640)	0.4417 (0.4025,0.5460)

Note that the chi-square statistic is smallest for the MRQR LLR procedure, with the chi-square statistic from the logistic procedure only slightly larger. The value of the mixing parameter λ for the MRQR Kernel procedure is 0.4668, and for MRQR LLR, $\lambda = 0.4102$. The magnitude of the mixing parameter for both MRQR procedures implies approximately an equal amount of weight is being given to the parametric and nonparametric fits to this data. As shown in the theoretical results of **Chapter 7** with the logit mixture model, a mixing parameter approximately equal to 0.4 corresponds to a degree of model-misspecification on the 0.3 to 0.4 range. Though the chi-square test does not indicate lack-of-fit in this example, it can be very misleading since the chi-square lack-of-fit test does not reject enough. Also, the model degrees of freedom for the MRQR LLR procedure is less than the model degrees of freedom for the MRQR Kernel fit, but only slightly. The mean squared error is smallest for the MRQR LLR procedure, only slightly less than the mean squared error for the logistic fit. Note that the estimates of the ED_{50} values are approximately the same for the five procedures, with the logit procedure yielding the smallest width of the 95% confidence interval on the ED_{50}, only a few thousandths less than the width of the MRQR LLR procedure. Not only do the procedures have approximately the same estimate on the median effective dose, their ED_{20} and ED_{80} estimates are approximately the same as well, with the exception of the LLR procedure which has a substantially larger estimate of the ED_{20} with respect to the other procedures.

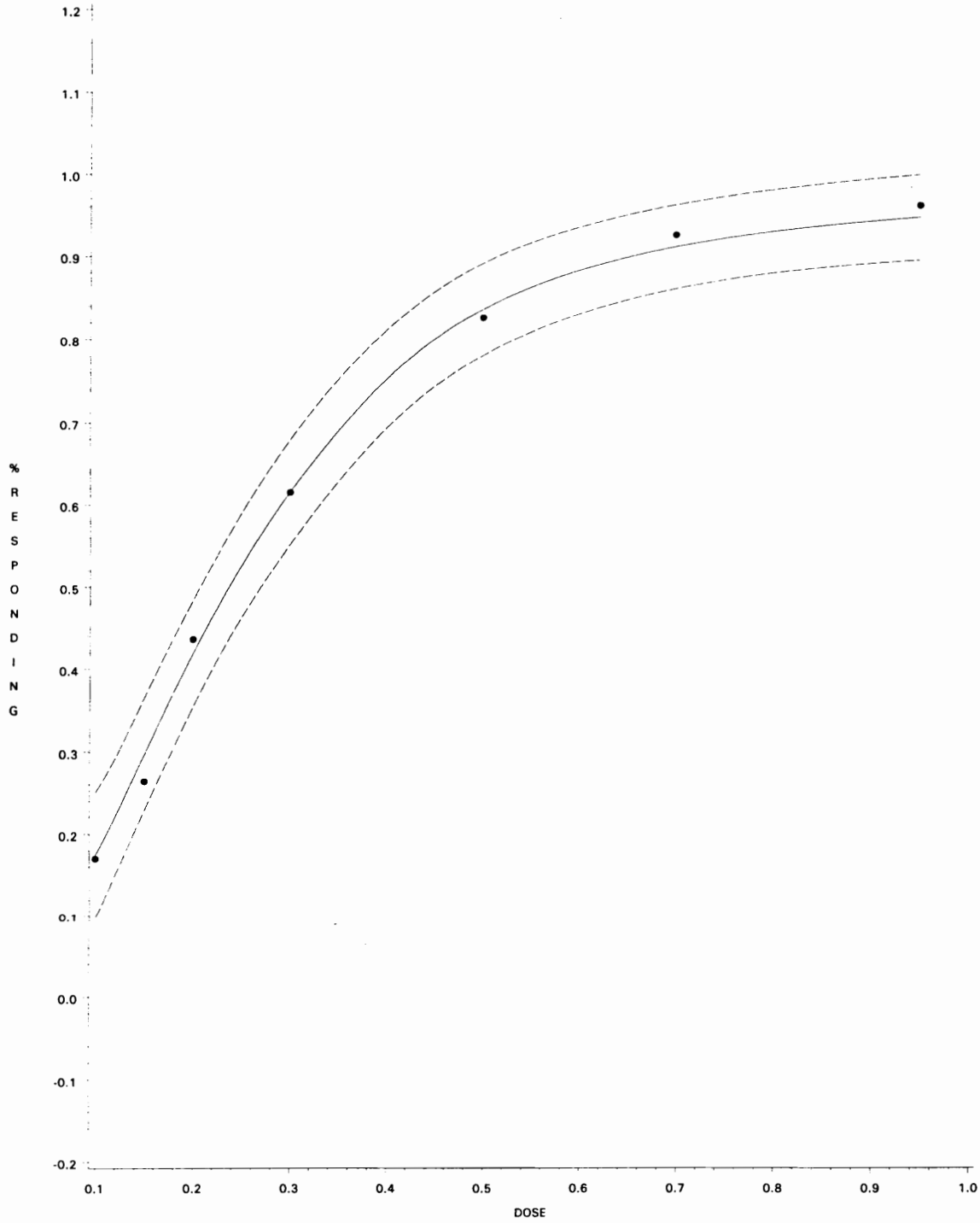


Figure 9.2.1 Model-robust Quantal Regression Using Kernel Regression as the Nonparametric Portion to Fit the Fly Data.

. . . Raw data, ___ MRQR Kernel fit, - - - 95% Confidence band

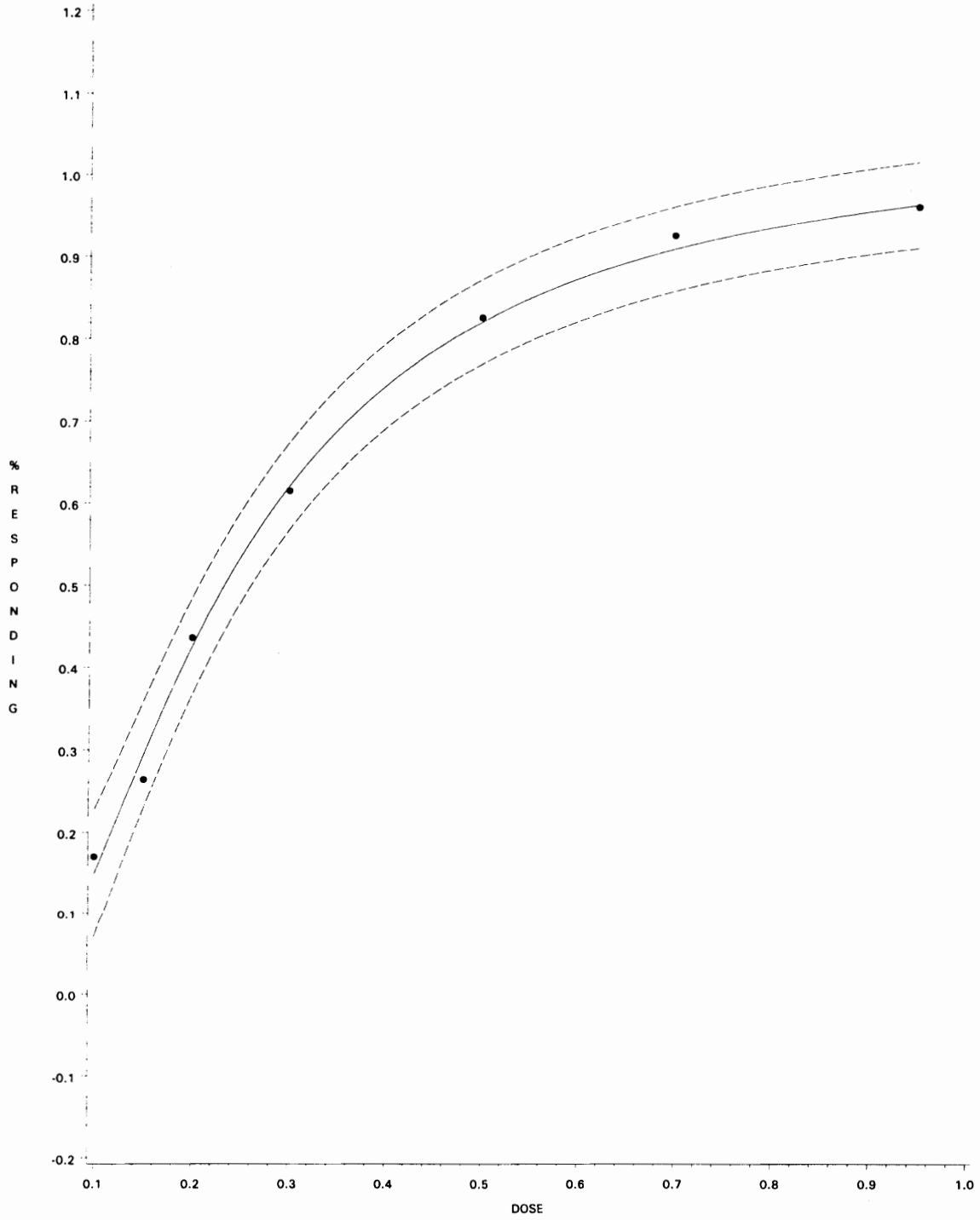


Figure 9.2.2 Model-robust Quantal Regression Using LLR Regression as the Nonparametric Portion to Fit the Fly Data.

. . . Raw data, ___ MRQR LLR fit, - - - 95% Confidence band

§9.2.B Example 8

The previous example was used to illustrate the use of the model-robust procedures using real data. In this section, the model-robust procedures will be applied to the logit mixture data set from **Table 2.2.2.** The MRQR LLR procedure has a chi-square statistic of 0.1885, which is much smaller than the chi-square statistic of the MRQR Kernel and the logit procedure. Note that the statistics in Table 2.2.2 for the logistic and MRQR Kernel procedures are approximately the same. This is because the mixing parameter for the MRQR Kernel procedure is approximately zero (λ=0.0086), thus giving almost all of the weight to the logistic fit.

Table 9.2. 2 Summary information of fits to logit mixture dose-response data. The 95% confidence interval estimates on effective doses are in parentheses.

Procedure	χ^2	Dfmodel	MSE	ED_{20}	ED_{50}	ED_{80}
Logistic	0.6682	2.0000	0.2227	0.2771 (0.1197,0.3685)	0.4893 (0.4035,0.5742)	0.7014 (0.6108,0.8564)
Kernel	2.3633	2.7998	1.0742	0.2084 (0.0696,0.3239)	0.4797 (0.3832,0.5834)	0.7765 (0.6447,0.9252)
LLR	0.4422	2.1119	0.1531	0.2315 (0.1475,0.3037)	0.4906 (0.4229,0.5581)	0.7521 (0.6796,0.8363)
MRQR Kernel	0.6682	2.0075	0.2227	0.2771 (0.1363,0.3519)	0.4893 (0.4149,0.5628)	0.7014 (0.6272,0.8400)
MRQR LLR	0.1885	2.0806	0.0646	0.2427 (0.1544,0.3242)	0.4901 (0.4182,0.5623)	0.7395 (0.6570,0.8287)

The mixing parameter for the MRQR LLR procedure is 0.7202, which implies more weight is given to the local linear regression fit than the logistic fit. The median effective dose for the procedures in **Table 9.2.2** are approximately the same, but the extreme doses are substantially different. Recall that the true ED_{20} is 0.2730 and the true ED_{80} is 0.7270. The logistic and MRQR Kernel procedure estimates the ED_{20} well, overestimating it by approximately 2%. But the ED_{80} value is more accurately estimated by the MRQR LLR procedure, overestimating the true ED_{80} by only 2%. All of the true effective doses are covered by their respective 95% confidence intervals for each of the procedures, but the MRQR LLR procedure has the narrowest intervals on the ED_{20} and ED_{80}. The width of the 95% confidence interval for the ED_{20} is 0.1697, compared to a width of 0.2156 by the MRQR Kernel and logit procedures. The MRQR LLR procedure has a 95% confidence interval width on the ED_{80} estimate of 0.1718, compared to a width of 0.2128 by the MRQR Kernel and logit procedures.

Another statistic of interest is the model degrees of freedom. Note that with only a slight increase in the model degrees of freedom, the MRQR LLR procedure reduces the chi-square statistic with respect to

the LLR and logit procedures. Also, the sample mean square error for the MRQR LLR procedure is much smaller than that of the other methods.

Figure 9.2.3 and **Figure 9.2.4** are plots of the MRQR Kernel and MRQR LLR fits to the logit mixture dose response data, respectively. Because the value of the mixing parameter is approximately zero for the MRQR Kernel procedure, **Figure 9.2.3** is almost identical to logit plot in **Figure 2.2.3**. The MRQR LLR procedure fits this data extremely well, even detecting the slight curvature at dose levels 0.3 and 0.7.

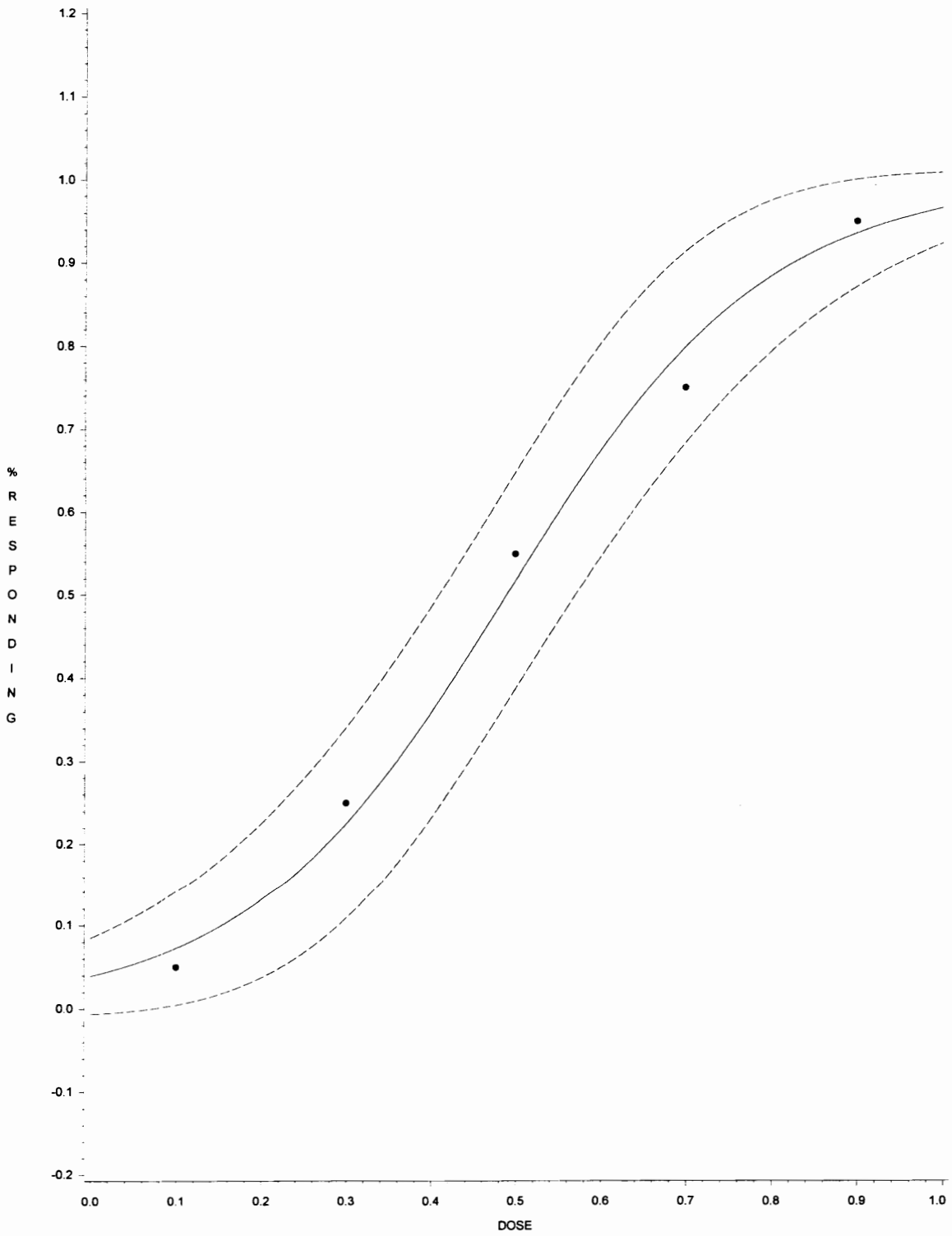


Figure 9.2.3 Model-robust Quantal Regression Using Kernel Regression as the Nonparametric Portion to Fit the Logit Mixture data.

. . . Raw data, ___ MRQR LLR fit, - - - 95% Confidence band

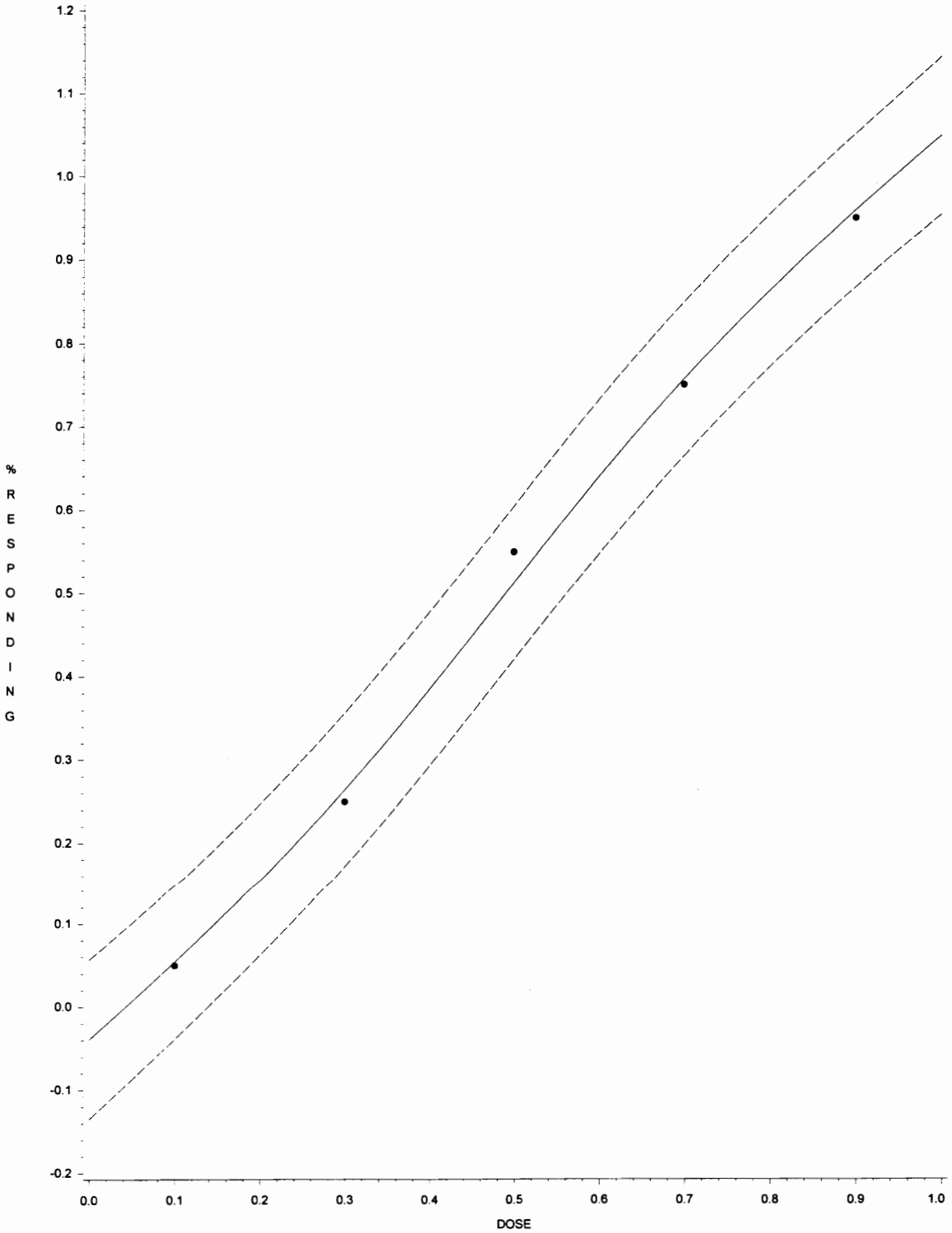


Figure 9.2.4 Model-robust Quantal Regression Using LLR Regression as the Nonparametric Portion to Fit the Logit Mixture data.

. . . Raw data, ___ MRQR LLR fit, - - - 95% Confidence band

CHAPTER 10
SUMMARY AND FUTURE RESEARCH

§10.1 Summary and Future Research

The nonparametric regression procedures presented in the previous chapters show promise as methods that can be used alongside the parametric analysis of quantal dose-response data, or as a tool for curve fitting and effective dose estimation when the user does not know the parametric form of the true underlying model. The theoretical results of **Chapter 7** indicate that the model-robust procedures, especially MRQR LLR, improves the average mean squared error with respect to the parametric (logit) and nonparametric methods. This improvement can be seen when using the optimal bandwidth and mixing parameters as well as the data-driven ones chosen using the PRESS* statistic. Although both model-robust quantal regression procedures are attractive, the method using local linear regression as the nonparametric portion seems to be better due to its lack of boundary problems.

Based on the theoretical results of **Chapter 7** and the simulation results presented in **Chapter 8**, the model-robust procedures improves the prediction when the user is not certain as to the form of the true underlying model. Summarizing **Chapter 7** and **Chapter 8**, the following is a list of the more pertinent features of the model-robust procedures:

1. With a small number of doses, such as d=3, the MRQR LLR procedure yields as small or smaller average mean squared error than the other procedures, with the exception of the MLE when γ=0, for n=10, 20, and 50 replicates at each dose level when using the optimal bandwidth and mixing parameter.

2. Using the optimal bandwidth and mixing parameters, the MRQR LLR procedure had higher mean squared error efficiencies than the other procedures for the (d=5, 7; n=10, 20, 50) combinations across almost all values of γ. This implies that with the optimal values of the bandwidth and mixing parameter, the MRQR LLR procedure performs better than the logit, kernel, LLR, and MRQR Kernel procedures.

3. Using the optimal values of the bandwidth, the model-robust procedures mixes appropriately. That is, as the degree of model-misspecification increases, the value of the mixing parameter increases as well, giving more weight to the nonparametric procedure.

4. Based on 500 Monte Carlo repetitions, the asymptotic theoretical properties of the bias, variance, and mean squared error formulas for the nonparametric and model-robust procedures are valid, especially when $d \geq 5$.

5. Although the estimated median effective dose was approximately the same for all the procedures presented, as the degree of model-misspecification increased, the model-robust procedures estimated the extreme doses more accurately, and with higher coverage probabilities, when using the optimal bandwidth and mixing parameter.

6. Using the optimal bandwidth and mixing parameter for the nonparametric and model-robust procedures, respectively, the true response, P, is also more accurately estimated by the model-robust procedures.

7. The PRESS* procedure, which was used to estimate the bandwidth and mixing parameter, tends to overestimate these parameters, especially the bandwidth.

8. When applying the model-robust procedure using the PRESS* statistic as the selection criteria for the bandwidth and mixing parameter, the estimates of effective doses, coverage probabilities, and estimates of the proportion responding tend to be adversely effected by shortcomings of the PRESS* procedure. But the MRQR LLR procedure still performed well when compared to the other procedures even when the bandwidth and mixing parameter are chosen by PRESS*.

It is evident from the above summary that the model-robust procedure applied to quantal dose-response data is capable of improving the fit obtained by either the parametric and nonparametric methods with the proper values of the bandwidth and mixing parameter. Although there are several areas for future research with respect to the model-robust procedures, the primary area is that of bandwidth and mixing parameter selection. Other statistics were studied for bandwidth and mixing parameter selection, such as the PRESS and generalized cross validation (GCV), but these methods had a tendency to pick the bandwidth too small or too large.

Also worthy of future study is the MRQR2 procedure in conjunction with quantal dose response data. The method is performing quite well in the linear models case (see Mays (1995)), but did not perform well in this research. This may have been due to the type of curves studied as well as the method for bandwidth and mixing parameter selection.

In this research, one of the nonparametric methods used was local linear regression. Another possibility would be to use local quadratic regression in conjunction with a parametric method. The local quadratic regression may fit better in the tails of the data, since many dose-response curves/data asymptotes in the extremes. That is, small doses yield a low responses (near zero), and large doses yield high responses (near one).

Just as local linear regression was used to fit the dose-response data, it seems feasible to fit a local logistic regression as well. This would entail fitting a logistic regression at each dose, just as in the local linear regression, a linear regression model is used at each dose. This can also be generalized such that any generalized linear model (Cauchy, Gompertz, Weibull, etc.) can be used to fit the dose-response curve locally. Also, instead of using the logistic regression model as the parametric function in the model-robust model, on can also use any generalized linear model such as the Cauchy, Gompertz, and Weibull, to name a few. Using any generalized linear model would make the model-robust procedure more applicable across a variety of fields. This is similar to the method proposed by Severini and Staniswalis (1994), in which they are using quasi-likelihood method to fit semiparametric models.

Although this research has shown promise, a number of avenues are incomplete and still need to be studied. First of all, the distribution on the mixing parameter may serve as a method of testing the adequacy of the models. In other words, if one can obtain a formal test on λ, then it can be used to determine if the parametric model supplied by the user is adequate or if the nonparametric method is appropriate. Finally, continued research is necessary for the proper bandwidth and mixing parameter selection. The performance of the model-robust procedure is quite good when the bandwidth and mixing parameter are chosen by the PRESS* statistic, but use of the optimal bandwidth and mixing parameters substantially improves performance.

References

Allen, D. A. (1974) The relationship between variable selection and data augmentation and a method of prediction. Technometrics, 16, 125-127.

Ashton, W. D. (1972) The Logit Transformation With Special References To Its Uses In Bioassay. Griffin, London.

Chu, C.-K. and Marron, J. S. (1991) Choosing a kernel regression estimator. *Statistical Science*, **6**, 4, 404-436.

Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 368, 829-836.

Cleveland, W. S. and Devlin, S. J. (1988) Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 403, 596-610.

Copas, J. B. (1983) Plotting p against x. *Applied Statistics*, **32**, 1, 25-31.

Cox, C. (1990) Fieller's Theorem, the likelihood and the delta method. *Biometrics*, **46**, 709-718.

Cox, D. R. and Snell, E. J. (1989) *Analysis of Binary Data* (2nd edition), Chapman and Hall, London.

Dragstedt, C. A. and Lang, V. F. (1928) Respiratory stimulants in acute cocaine poisoning in rabbits. *Journal of Pharmaceutical Experimental Therapy*, **32**, 215-222.

Eggar, M. J. (1979) Power transformations to achieve symmetry in quantal bioassay. Technical Report No. 47, Division of Biostatistics, Stanford University, California.

Einsporn, R. (1987) A link between least squares regression and nonparametric curve estimation. Unpublished Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Einsporn, R. L. and Birch, J. B. (1993) Model -robust regression: Using nonparametric regression to improve parametric regression analyses. *Technical Report Number* 93-5, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Fan, J. (1993) Local linear regression smoothers and their minimax efficiency. *The Annals of Statistics*, **21**, 196-216.

Fan, J. and Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, **20**, 2008-2036.

Finney, D. J. (1950) The estimation of the mean of a normal tolerance distribution. *Sankhya*, **10**, 341-360.

Finney, D. J. (1953) The estimation of the ED50 for a logistic response curve. *Sankhya*, **12**, 121-136.

Finney, D. J. (1978) *Statistical Method in Biological Assay* (3rd edition), Griffin, London.

Goedhart, P. (1985) Statistical analysis of quantal-assay data. Unpublished M. Sc. dissertation, University of Kent, Canterbury, Kent, England.

Govindarajula, Z. (1988) *Statistical Techniques in Bioassay*. Karger, Basel, Switzerland.

Guerrero, V. M. and Johnson, R. A. (1982) Use of Box-Cox transformation with binary response models. *Biometrika*, **69**, 2, 309-314.

Hamilton, M. A., Russo, P. C., and Thurston, R. V. (1971) Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. *Environmental Science and Technology*, **11**, 714-719. Correction, 1978, **12**, 417.

Härdle, W. (1990) *Applied Nonparametric Regression*. Cambridge University Press, New York.

Hastie, T. and Loader, C. (1993) Local regression: Automatic kernel carpentry. *Statistical Science*, **8**, 2, 120-143.

Hastie, T. and Tibshirani, R. (1987) Generalized additive models. *Statistical Science*,1, 3, 297-318.

Hastie, T. and Tibshirani, R. (1987) Generalized additive models: Some applications. *Journal of the American Statistical Association*, **82**, 398, 371-386.

Hoaglin, D. C. and Welch, R. (1978) The hat matrix in regression and ANOVA. *The American Statistician*, **32**, 17-22.

Hoekstra, J. A. (1989) Estimation of the ED_{50}. *Biometrics*, **45**, 1, 337-338.

Huber, J. J. (1994) Environmental Risk Assessment - Lecture Notes. Department of Mathematics and Statistics, University of Guelph.

Irwin, J. O. (1937) Statistical methods applied to biological assays. *Journal of the Royal Statistical Society, Supplement*, **4**, 1-60.

Kappenman, R. F. (1987) Nonparametric estimation of dose-response curves with applications to ED_{50} estimation. *Journal of Statistical Computation and Simulation*, **28**, 1-13.

Kelley, C. and Rice, J. (1990) Monotone smoothing with applications to dose-response curves and the assessment of synergism. *Biometrics*, **46**, 1071-1085.

Mays, James (1995) Model-robust regression combining parametric, nonparametric, and semiparametric methods. Unpublished Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Miller, R. G. and Halpern, J. W. (1980) Robust estimators for quantal bioassay. *Biometrika*, **67**, 103-110.

Morgan, B. J. T. (1992) *Analysis of Quantal Response Data*, Chapman and Hall, London.

Müller, H.-G. and Schmitt, T. (1988) Kernel and Probit estimates in quantal bioassay. *Journal of the American Statistical Association*, **83**, 403, 750-759.

Myers, R. H. (1990) Classical and Modern Regression with Applications (2nd edition), PWS-Kent, Boston, Massachusetts.

Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability and its Applications*, **9**, 141-142.

Olkin, I. and Spiegelman, C. H. (1987) A semiparametric approach to density estimation. *Journal of the American Statistical Association*, **82**, 399, 858-865.

Rao, C. R. (1973) *Linear Statistical Inferences and its Applications* (2nd edition). New York: John Wiley.

Reed, L. J. and Muench, H. (1938) A simple method of estimating fifty percent endpoints. *American Journal of Hygiene*, **27**, 493-497.

Rice, J. (1984) Boundary modification for kernel regression. *Communications in Statistics: Theory and Methods*, **13**, 7, 893-900.

Schmoyer, R. L. (1984) Sigmoidally constrained maximum likelihood estimation in quantal bioassay. *Journal of the American Statistical Association*, **79**, 448-453.

Severini, T. A. and Staniswalis, J. G. (1994) Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, **89**, 426, 501-511.

Spearman, C. (1908) The method of 'right and wrong cases' ('constant stimuli') without Gauss's formulae. *British Journal of Psychology*, **2**, 227-242.

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistics Society, Series B*, **50**, 3,413-436.

Staniswalis, J. G. and Cooper, V. (1988) Kernel estimates of dose response. *Biometrics*, **44**, 4, 1103-1120.

Sutherland, S. (1992) Sequential design augmentation with model misspecification. Unpublished Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Sutherland, S. and Birch, J. B. (1993) Sequential design augmentation under model misspecification using the BIIV criteria. *Technical Report Number* 93-10, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Thompson, W. R. (1947) Use of moving averages and interpolation to estimate median-effective doses. I: Fundamental formulas, estimation of error and relation to other methods. *Bacteriological Reviews*, **11**, 115-145.

Tukey, John W. (1977) Exploratory Data Analysis, Addison-Wesley Publishing Company, Redding, Massachusettes.

Wahba, G. and Wold, S. (1975) A completely automatic French curve: fitting spline functions by cross-validation. *Communications in Statistics, Series A*, **4**, 1-17.

Watson, G. S. (1964) Smooth regression analysis. *Sankhya, Series A*, **26**, 359-372.

APPENDICES

Appendix A

Recall that the *PRESS* statistic is the predicted error sums of squares in which predictions are made with the i^{th} data point removed, for i=1,2,...,n. This "minus-i" prediction is straightforward in the linear regression framework. However, when fitting a nonlinear function such as logistic regression, the predictions cannot be obtained via a linear predictor. In this section, an approximation will be given for the "minus-i" statistic for the logistic regression procedure.

The ordinary least squares predictions are given by

$$\hat{\mathbf{y}} = \mathbf{X}\beta = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \tag{A.1}$$

where $\mathbf{H}$ is the OLS "hat" matrix. It has also been shown in this research that the kernel and local linear regression estimates may be obtained in the same manner as the OLS predictions of the form

$$\hat{\mathbf{y}}^{NP} = \mathbf{H}^{NP}\mathbf{y} \tag{A.2}$$

where $\hat{\mathbf{y}}^{NP}$ is the nx1 vector of nonparametric prediction and $\mathbf{H}^{NP}$ is the nxn nonparametric weight or "hat" matrix. So, the "minus-i" predictions are given by

$$\hat{y}_{i,-i} = \sum_{i \neq j} h_{ij,-i} y_j \tag{A.3}$$

where

$$h_{ij,-i} = \frac{h_{ij}}{1 - h_{ii}} \tag{A.4}$$

The "minus-i" predictions for the nonparametric procedures are analogous, simply substituting the elements of the nonparametric weight matrix in equations (A.3) and (A.4). In this research, the *PRESS* statistic has been used as a method of comparison between the various models in fitting quantal dose-response data, and as a method of selecting the bandwidth and the mixing parameter in the nonparametric regression procedures and the model-robust regression procedures, respectively.

For logistic regression, and nonlinear models in particular, no such deletion formulas or "minus-i" prediction formulas exist, so equation (2.1.16) was derived in this research. The logistic model is given by

$$P_i = \frac{1}{1 + e^{-\mathbf{x}_i'\beta}} \tag{A.5}$$

If the logistic model is linearized using the logit transformation such that

$$logit(\mathbf{p}) = \mathbf{X}\beta \tag{A.6}$$

where $logit(\mathbf{p})$ is the dx1 vector of logits, then by applying the Sherman-Morrison-Woodbury theorem (see Rao (1973)), a "minus-i" prediction may be obtained.

The SMW theorem states

$$(\mathbf{X}'\mathbf{X})_{-i}^{-1} = (\mathbf{X}'\mathbf{X} - \mathbf{x}_i\mathbf{x}_i')^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}}{1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i} \tag{A.7}$$

The left-hand side of the equation above is the $\mathbf{X}'\mathbf{X}$ matrix with the i^{th} data point removed. In the context of logistic regression, if the logistic regression model is linearized using the logit transformation, then the *PRESS* residual is given by

$$e_{i,-i} = logit(p_i) - \mathbf{x}_i'\mathbf{b}_{-i} \tag{A.8}$$

where $\mathbf{b}_{-i}$ is the vector of estimated regression coefficients with the i^{th} data point set aside. Now, it is easy to show that the SMW theorem applies to the $(X'W^*X)^{-1}$ matrix results in

$$(X'W^*X)_{-i}^{-1} = (X'W^*X - x_i w_i^* x_i') = \frac{(X'W^*X)x_i w_i^* x_i'(X'W^*X)^{-1}}{1 - x_i'(X'W^*X)^{-1}x_i w_i^*}$$

Thus, it follows that

$$\boldsymbol{b}_{-i} = \boldsymbol{b} - \frac{(X'W^*X)^{-1}x_i w_i^* e_i}{1 - h_{ii}} \tag{A.9}$$

where $w_i^* = n_i p_i(1-p_i)$, $e_i = logit(p_i) - \mathbf{x}_i'\mathbf{b}$, and h_{ii} is the i^{th} diagonal element of $H^L = X(X'W^*X)^{-1}X'W^*$, i.e. $h_{ii}^L = x_i'(X'W^*X)^{-1}x_i w_i^*$, with $\mathbf{W}^*$ being the diagonal matrix with elements w_i^*. It should be noted that $\mathbf{b}$ is the one-step weighted least squares estimate of β. That is, $\mathbf{b}$ *is not* the maximum-likelihood estimate of β given in **Chapter 2**, in which the weight matrix $\mathbf{W}$ changes in each iteration until the maximum likelihood estimate of β is obtained.

So,

$$e_{i,-i} = logit(p_i) - logit(\hat{P}_{i,-i}) \tag{A.10}$$

with

$$\begin{aligned} \hat{logit}\left(P_{i,-i}\right) &= \boldsymbol{x}_i'\left[\boldsymbol{b} - \frac{\left(\boldsymbol{X}\boldsymbol{W}^*\boldsymbol{X}\right)^{-1}\boldsymbol{x}_i w_i^* e_i}{1-h_{ii}^L}\right] \\ &= \boldsymbol{x}_i'\boldsymbol{b} - \frac{\boldsymbol{x}_i'\left(\boldsymbol{X}\boldsymbol{W}^*\boldsymbol{X}\right)^{-1}\boldsymbol{x}_i w_i^* e_i}{1-h_{ii}^L} \\ &= \boldsymbol{x}_i'\boldsymbol{b} - \frac{h_{ii}^L e_i}{1-h_{ii}^L} \end{aligned} \tag{A.11}$$

So, transforming back to the original scale, the "minus-i" prediction is approximated by

$$\hat{P}_{i,-i} \approx \frac{1}{1+e^{-\mathbf{x}_i'\mathbf{b}}e^{h_{ii}^L e_i/\left(1-h_{ii}^L\right)}} \tag{A.12}$$

where $\boldsymbol{b} = \left(\boldsymbol{X}\boldsymbol{W}^*\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{W}^*\, logit(\boldsymbol{p})$ and $e_i = logit(p_i) - \mathbf{x}_i'\mathbf{b}$.

Due to the small size of many dose-response data sets, it is possible to compute the true "minus-i" values and compare them to the approximate ones. Using 21 real data sets, the approximate "minus-i" values were very close to the true ones, with an average difference of -0.0026, and a mean absolute difference of 0.0244. Due to the precision of the approximate "minus-i" values, it seems reasonable to use them in the computation of the *PRESS* statistic for logistic regression.

Appendix B

Model-robust quantal regression (MRQR), a linear combination of parametric and nonparametric estimates, presents several difficulties when developing variance expressions for estimates. One problem is the nonlinear nature of the parametric procedure. Variance expressions here are asymptotic in nature. A second problem concerns the variance for the nonparametric estimates. There are numerous methods available in the literature for estimating the variance for a nonparametric regression procedure. Most of them are based on asymptotic results and their accuracy for small samples has yet to be ascertained. One method for the derivation of the bias and variance expressions of the MRQR procedure will be presented here. Before the bias and variance estimates for the MRQR procedures are obtained, bias and variance expressions for the logistic and nonparametric regression methods will be derived.

Though asymptotic expressions of the variance of prediction are available for the logistic method, a linearization of the logistic estimate will be performed to ease the development of the prediction variance. Let $\hat{\mathbf{P}}^{Logistic} = F\left(\mathbf{X}\hat{\beta}\right)$ denote the dx1 vector of predictions at the d doses. That is, the function $F\left(\mathbf{X}\hat{\beta}\right)$ is the dx1 vector $\left(F\left(\mathbf{x}_1'\hat{\beta}\right), F\left(\mathbf{x}_2'\hat{\beta}\right),..., F\left(\mathbf{x}_d'\hat{\beta}\right)\right)'$. Expanding $F\left(\mathbf{X}\hat{\beta}\right)$ by a first-order multivariate Taylor series about the true parameter vector, β, results in

$$F\left(\mathbf{X}\hat{\beta}\right) \approx F\left(\mathbf{X}\beta\right) + \left.\frac{\partial F\left(\mathbf{X}\beta\right)}{\partial \hat{\beta}}\right|_{\hat{\beta}=\beta} \left(\hat{\beta}-\beta\right), \tag{B.1}$$

$$= F\left(\mathbf{X}\beta\right) + \left\langle f\left(\mathbf{X}\beta\right)\right\rangle \mathbf{X}\left(\hat{\beta}-\beta\right) \tag{B.2}$$

where $f(u) = \frac{df(u)}{du}$ is the logistic probability density function, and $\left\langle f\left(\mathbf{X}\beta\right)\right\rangle$ is the dxd diagonal matrix whose diagonal elements are $\left(f\left(\mathbf{x}_1'\beta\right), f\left(\mathbf{x}_2'\beta\right),..., f\left(\mathbf{x}_d'\beta\right)\right)$. Algebraic manipulation of the right-hand side of the above equation yields

$$
\begin{aligned}
\hat{\mathbf{P}}^{Logistic} &\approx F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\hat{\beta}-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\beta \\
&= F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{p}^{*}-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\beta \\
&= F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left[\langle f(\mathbf{X}\beta)\rangle^{-1}(\mathbf{p}-F(\mathbf{X}\beta))+\mathbf{X}\beta\right]-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\beta \\
&= F(\mathbf{X}\beta)-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}F(\mathbf{X}\beta) \\
&\quad+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{p} \\
&= \left[\mathbf{I}-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\right]F(\mathbf{X}\beta) \\
&\quad+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{p}
\end{aligned}
$$

where here, **W** is the *dxd* diagonal matrix of "true weights" written as

$$w_i = \frac{n_i f^2(\mathbf{x}_i'\beta)}{P_i Q_i}$$

with $P_i = F(\mathbf{x}_i'\beta)$ and $Q_i = 1 - P_i$. Thus, one obtains

$$\hat{\mathbf{P}}^{Logistic} \approx \mathbf{A} + \mathbf{B}\mathbf{p}, \tag{B.3}$$

where $\mathbf{A} = [\mathbf{I} - \mathbf{B}]F(\mathbf{X}\beta)$ and $\mathbf{B} = \langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}$.

Before detailing any further derivations, it will be assumed, without loss of generality, that the true model is given by

$$\mathbf{P} = \mathbf{G}(x) = \mathbf{F}(\mathbf{X}\beta) + \mathbf{H}(\mathbf{x}) \tag{B.4}$$

where $\mathbf{H}(x)$ is a *d*x1 vector representing the difference between the user's parametric model, **F**, and the true model, **G**. Therefore, if **F**=**G**, the $\mathbf{H}(\mathbf{x}) = 0$ implying that the users model has been correctly specified.

The expectation of $\hat{\beta}$ will now be derived due to its necessity later for the expectation of the logistic predictions. From the maximum likelihood equation and equation (B.2), one can write the maximum likelihood estimate, $\hat{\beta}$, as $\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{p}^{*}$, where **W** is the "true weight" matrix defined above. So, the expected value of $\hat{\beta}$ under the true model **G** is given by

$$
\begin{aligned}
\mathbf{E}\left(\hat{\beta}\right) &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}E\left(p^{*}\right) \\
&= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}E\left[\langle f(\mathbf{X}\beta)\rangle^{-1}(\mathbf{p}-F(\mathbf{X}\beta))+\mathbf{X}\beta\right] \\
&= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}E(\mathbf{p})-(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}F(\mathbf{X}\beta) \\
&\quad +(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\beta \\
&= \beta+(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\left[F(\mathbf{X}\beta)+\mathbf{H}(\mathbf{x})\right]-(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}F(\mathbf{X}\beta) \\
&= \beta+(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{H}(\mathbf{x})
\end{aligned}
$$

Thus, under the correct model, i.e. **F**=**G** and $\mathbf{H}(\mathbf{x})=0$, $E\left(\hat{\beta}\right)=\beta$ which is an unbiased estimate.

The expectation of $\hat{\mathbf{P}}^{Logistic}$ can now be obtained as follows:

$$
\begin{aligned}
E\left(\hat{\mathbf{P}}^{Logistic}\right) &= E\left[F\left(\mathbf{X}\hat{\beta}\right)\right] \\
&\approx E\left[F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\left(\hat{\beta}-\beta\right)\right] \\
&= E\left[F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}E\left(\hat{\beta}\right)-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\beta\right] \\
&= F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\left[\beta+(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{H}(\mathbf{x})\right]-\langle f(\mathbf{X}\beta)\rangle\mathbf{X}\beta \\
&= F(\mathbf{X}\beta)+\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}\mathbf{H}(\mathbf{x})
\end{aligned}
$$

Manipulating the above expression, it can be shown that

$$
E\left(\hat{\mathbf{P}}^{Logistic}\right)=\mathbf{P}+\left[\langle f(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\beta)\rangle^{-1}-\mathbf{I}\right]\mathbf{H}(\mathbf{x}) \tag{B.5}
$$

Again, it is noted that if $\mathbf{H}(\mathbf{x})=0$, then $E\left(\hat{\mathbf{P}}^{Logistic}\right)=\mathbf{P}=F(\mathbf{X}\beta)$, the true response curve. Using equation (B.6), the bias of $\hat{\mathbf{P}}^{Logistic}$ can be obtained as

$$
\begin{aligned}
Bias\left(\hat{\mathbf{P}}^{Logistic}\right) &= E\left(\hat{\mathbf{P}}^{Logistic}\right)-\mathbf{P} \\
&= \mathbf{P}+\left[\langle \mathbf{f}(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle \mathbf{f}(\mathbf{X}\beta)\rangle^{-1}-\mathbf{I}\right]\mathbf{H}(\mathbf{x})-\mathbf{P} \\
&= \left[\langle \mathbf{f}(\mathbf{X}\beta)\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle \mathbf{f}(\mathbf{X}\beta)\rangle^{-1}-\mathbf{I}\right]\mathbf{H}(\mathbf{x}) \qquad \text{(B.6)}
\end{aligned}
$$

It is now of interest to obtain the variance of $\hat{\mathbf{P}}^{Logistic}$, but first the variance of $\hat{\beta}$ must be obtained. Using the MLE expression of $\hat{\beta}$ from the previous page, the variance is given by

$$
\begin{aligned}
Var\left(\hat{\beta}\right) &= Var\left[\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{p}^{\bullet}\right] \\
&= \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}Var\left(\mathbf{p}^{\bullet}\right)\mathbf{W}'\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}Var\left[\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\left(\mathbf{p}-F\left(\mathbf{X}\beta\right)\right)+\mathbf{X}\beta\right]\mathbf{W}'\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}Var\left(\mathbf{p}\right)\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\mathbf{W}'\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\mathbf{V}_{\mathbf{G}}\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\mathbf{W}'\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}
\end{aligned}
$$

where

$$
\mathbf{V}_{\mathbf{G}} = Var\left(\mathbf{p}\right) = \left\langle \frac{\mathbf{G}\left(\mathbf{x}_1\right)\left(1-\mathbf{G}\left(\mathbf{x}_1\right)\right)}{\mathbf{n}_1},\cdots,\frac{\mathbf{G}\left(\mathbf{x}_d\right)\left(1-\mathbf{G}\left(\mathbf{x}_d\right)\right)}{\mathbf{n}_d}\right\rangle \tag{B.7}
$$

for the model given in equation (B.4). Note that

$$
\begin{aligned}
\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\mathbf{V}_{\mathbf{G}}\left\langle f\left(\mathbf{X}\beta\right)\right\rangle^{-1}\mathbf{W}' &= \left\langle w_1^2 f_1^2\left(\mathbf{x}_1'\beta\right)G\left(\mathbf{x}_1\right)\left(1-G\left(\mathbf{x}_1\right)\right),\cdots,w_d^2 f_d^2\left(\mathbf{x}_d'\beta\right)G\left(\mathbf{x}_d\right)\left(1-G\left(\mathbf{x}_d\right)\right)\right\rangle \\
&= \left\langle \frac{n_1 f_1^2\left(\mathbf{x}_1'\beta\right)f_1^{-2}\left(\mathbf{x}_1'\beta\right)G\left(\mathbf{x}_1\right)\left(1-G\left(\mathbf{x}_1\right)\right)}{n_1 F\left(\mathbf{x}_1'\beta\right)\left(1-F\left(\mathbf{x}_1'\beta\right)\right)},\cdots,\frac{n_d f_d^2\left(\mathbf{x}_d'\beta\right)f_d^{-2}\left(\mathbf{x}_d'\beta\right)G\left(\mathbf{x}_d\right)\left(1-G\left(\mathbf{x}_d\right)\right)}{n_d F\left(\mathbf{x}_d'\beta\right)\left(1-F\left(\mathbf{x}_d'\beta\right)\right)}\right\rangle \\
&= \left\langle \frac{G\left(\mathbf{x}_1\right)\left(1-G\left(\mathbf{x}_1\right)\right)}{F\left(\mathbf{x}_1'\beta\right)\left(1-F\left(\mathbf{x}_1'\beta\right)\right)},\cdots,\frac{G\left(\mathbf{x}_d\right)\left(1-G\left(\mathbf{x}_d\right)\right)}{F\left(\mathbf{x}_d'\beta\right)\left(1-F\left(\mathbf{x}_d'\beta\right)\right)}\right\rangle \\
&= \left\langle \frac{G}{F}\right\rangle
\end{aligned}
$$

Thus,

$$
Var\left(\hat{\beta}\right) = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\left\langle \frac{G}{F}\right\rangle\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} \tag{B.8}
$$

Under the correct model, $\left\langle \frac{\mathbf{G}}{\mathbf{F}}\right\rangle = \mathbf{I}$ which yields $\mathbf{Var}\left(\hat{\beta}\right) = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}$, the asymptotic variance of the MLE. From this point on, the variance of $\hat{\beta}$ will be denoted by $\mathbf{V}_{\hat{\beta}}$.

Now, the variance of the logistic estimate for the proportion responding is given by

$$
\begin{aligned}
Var\left(\hat{\mathbf{P}}^{Logistic}\right) &\approx Var\left[\mathbf{F}\left(\mathbf{X}\beta\right)+\left\langle f\left(\mathbf{X}\beta\right)\right\rangle\mathbf{X}\hat{\beta}-\left\langle f\left(\mathbf{X}\beta\right)\right\rangle\mathbf{X}\beta\right] \\
&= \left\langle f\left(\mathbf{X}\beta\right)\right\rangle\mathbf{X}Var\left(\hat{\beta}\right)\mathbf{X}'\left\langle f\left(\mathbf{X}\beta\right)\right\rangle \\
&= \left\langle f\left(\mathbf{X}\beta\right)\right\rangle\mathbf{X}\mathbf{V}_{\hat{\beta}}\mathbf{X}'\left\langle f\left(\mathbf{X}\beta\right)\right\rangle
\end{aligned} \tag{B.9}
$$

Similarly, if **G**=**F**, then $\left\langle \frac{\mathbf{G}}{\mathbf{F}}\right\rangle = \mathbf{I}$ and (B.9 yields

$$
Var\left(\hat{\mathbf{P}}^{Logistic}\right) \approx \left\langle f\left(\mathbf{X}\beta\right)\right\rangle\mathbf{X}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\left\langle f\left(\mathbf{X}\beta\right)\right\rangle \tag{B.10}
$$

which is the asymptotic variance of estimated response when using maximum likelihood estimation. The mean squared error can be expressed as the variance plus the squared bias. In the logistic regression case, if the user's model is correct, then the mse is composed of the variance expression only (equation B.10), but if the user's model is misspecified, then the mse is comprised of the variance expression in (B.9) and the squared bias of equation (B.7).

In a similar manner, variance, bias, and mean square error expressions can be obtained for the nonparametric procedures (kernel and local linear regression) for fixed bandwidth. The expected value of the nonparametric prediction for a fixed bandwidth is given by

$$\begin{aligned} E\left(\hat{\mathbf{P}}^{NP}\right) &= E\left(\mathbf{H}^{NP}\mathbf{p}\right) \\ &= \mathbf{H}^{NP}E(\mathbf{p}) \\ &= \mathbf{H}^{NP}\mathbf{P} \end{aligned}$$

where $\mathbf{H}^{NP}$, which is assumed fixed for all derivations, corresponds to the nonparametric weight matrix for either kernel or local linear regression. The bias of $\hat{\mathbf{P}}^{NP}$ can be obtained via

$$\begin{aligned} Bias\left(\hat{\mathbf{P}}^{NP}\right) &= E\left(\hat{\mathbf{P}}^{NP}\right) - \mathbf{P} \\ &= \mathbf{H}^{NP}\mathbf{P} - \mathbf{P} \\ &= \left(\mathbf{H}^{NP} - \mathbf{I}\right)\mathbf{P} \end{aligned}$$

The variance for the nonparametric regression estimate, assuming that the bandwidth is constant, is given by

$$\begin{aligned} Var\left(\hat{\mathbf{P}}^{NP}\right) &= Var\left[\mathbf{H}^{NP}\mathbf{p}\right] \\ &= \mathbf{H}^{NP}Var(\mathbf{p})\mathbf{H}^{NP'} \\ &= \mathbf{H}^{NP}\mathbf{V}_{\mathbf{G}}\mathbf{H}^{NP'} \end{aligned}$$

where $\mathbf{V}_{\mathbf{G}}$ is given in equation (B.8). Thus, the mean squared error of the nonparametric estimate is given by

$$\mathbf{MSE}\left(\hat{\mathbf{P}}^{NP}\right) = \mathbf{H}^{NP}\mathbf{V}_{\mathbf{G}}\mathbf{H}^{NP'} + \left[\left(\mathbf{H}^{NP} - \mathbf{I}\right)\mathbf{P}\right]\left[\left(\mathbf{H}^{NP} - \mathbf{I}\right)\mathbf{P}\right]' \tag{B.11}$$

Given the form of the MRQR model in (6.1.1), the variance of the MRQR estimated probabilities of response of the d doses can be written as the variance of the linear combination:

$$Var\left(\hat{\mathbf{P}}^{MRQR}\right) = Var\left[\lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{P}\right] \tag{B.12}$$

where $\hat{\mathbf{P}}^{NP}$ is the nonparametric estimate of the probability of response, and $\hat{\mathbf{P}}^{P}$ denotes the parametric estimate of the probability of response, and λ is the mixing parameter. $\hat{\mathbf{P}}^{NP}$ is given in equation (4.2.4) for kernel regression, and in equation (4.3.4) for the local linear regression estimate, with both estimates being

written as linear predictors. As noted throughout this proposal, the logistic regression procedure will be the parametric analysis used on the quantal dose-response data. It should also be noted that the logistic estimate cannot be written as a linear predictor, as is the case for ordinary least squares, kernel regression, and local linear regression. With the estimated variance for the parametric procedure given in equation (B.9 and the estimated variance for the nonparametric method given above, the variance of the MRQR estimate of probabilities of response for a fixed mixing parameter, λ, and bandwidth, b, can be expressed as

$$\begin{aligned} Var\left(\hat{\mathbf{P}}^{MRQR}\right) &= Var\left[\lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{Logistic}\right] \\ &\approx Var\left\{\lambda\mathbf{H}^{NP}\mathbf{p} + (1-\lambda)[\mathbf{A} + \mathbf{B}\mathbf{p}]\right\} \\ &= Var\left\{\lambda\mathbf{H}^{NP}\mathbf{p} + (1-\lambda)\mathbf{A} + (1-\lambda)\mathbf{B}\mathbf{p}\right\} \\ &= Var\left\{\left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]\mathbf{p} + (1-\lambda)\mathbf{A}\right\} \\ &= \left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]\mathbf{V}_G\left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{B}\right]' \end{aligned}$$

where **A** and **B** are defined in (B.3). Using the MRQR model, the expected value of $\hat{\mathbf{P}}^{MRQR}$ can be obtained as

$$\begin{aligned} E\left(\hat{\mathbf{P}}^{MRQR}\right) &= E\left[\lambda\hat{\mathbf{P}}^{NP} + (1-\lambda)\hat{\mathbf{P}}^{P}\right] \\ &= \lambda E\left(\hat{\mathbf{P}}^{NP}\right) + (1-\lambda)E\left(\hat{\mathbf{P}}^{P}\right) \end{aligned}$$

Since $E\left(\hat{\mathbf{P}}^{P}\right) = E\left(\hat{\mathbf{P}}^{Logistic}\right)$ is given in (B.6) and $E\left(\hat{\mathbf{P}}^{NP}\right) = \mathbf{H}^{NP}\mathbf{P}$, it follows that

$$\begin{aligned} E\left(\hat{\mathbf{P}}^{MRQR}\right) &= \lambda\mathbf{H}^{NP}\mathbf{P} + (1-\lambda)\left\{\mathbf{P} + \left[\left\langle f(\mathbf{X}\beta)\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})\right\} \\ &= \left[\lambda\mathbf{H}^{NP} + (1-\lambda)\mathbf{I}\right]\mathbf{P} + (1-\lambda)\left[\left\langle f(\mathbf{X}\beta)\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x}) \end{aligned}$$

where **I** is the *dxd* identity matrix. Given the expectation of $\hat{\mathbf{P}}^{MRQR}$, the bias of the MRQR estimate is

$$\begin{aligned} Bias\left(\hat{\mathbf{P}}^{MRQR}\right) &= \lambda\mathbf{H}^{NP}\mathbf{P} + (1-\lambda)\left\{\mathbf{P} + \left[\left\langle f(\mathbf{X}\beta)\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})\right\} - \mathbf{P} \\ &= \lambda\left[\mathbf{H}^{NP} - \mathbf{I}\right]\mathbf{P} + (1-\lambda)\left[\left\langle f(\mathbf{X}\beta)\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x}) \end{aligned}$$

With the variance and bias expressions for the MRQR procedure, the mean squared error can be expressed as the variance and the squared bias. Examining the bias and variance expressions for the MRQR procedure, under the correct parametric model, it is expected that $\mathbf{H}(\mathbf{x}) \approx 0$ and $\lambda \approx 0$, thus the contribution to the mse is supplied mainly by the variance of the parametric model with small bias. Alternatively, if the users model is incorrect, then $\mathbf{H}(\mathbf{x}) \neq 0$, which contributes bias to the mse, and λ may be close to 1, depending upon the degree of misspecification. Thus, the mse will be composed of a larger contribution from bias.

For the alternative MRQR procedure, deemed MRQR2 in **Chapter 6**, the properties can be derived in a similar manner. The true model is assumed to be given in (B.4) and the user fits

$$\hat{\mathbf{P}}^{MRQR2} = F\left(\mathbf{X}\hat{\beta}\right) + \lambda \mathbf{H}^{NP}\mathbf{e} \tag{B.13}$$

where $\mathbf{e} = \mathbf{p} - F\left(\mathbf{X}\hat{\beta}\right)$ is the vector of residuals from the parametric fit, $\mathbf{H}^{NP}$ is the nonparametric weight matrix used to fit the residuals, and λ is the mixing parameter denoting the amount of residuals being added to the parametric fit. For the MRQR2 procedure, the estimate of proportion responding can be written as

$$\begin{aligned}\hat{\mathbf{P}}^{MRQR2} &= F\left(\mathbf{X}\hat{\beta}\right) + \lambda \mathbf{H}^{NP}\mathbf{e} \\ &= F\left(\mathbf{X}\hat{\beta}\right) + \lambda \mathbf{H}^{NP}\left(\mathbf{p} - F\left(\mathbf{X}\hat{\beta}\right)\right) \\ &= \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)F\left(\mathbf{X}\hat{\beta}\right) + \lambda \mathbf{H}^{NP}\mathbf{p}\end{aligned}$$

With the expression above, the expectation of $\hat{\mathbf{P}}^{MRQR2}$ can be obtained as follows:

$$\begin{aligned}E\left(\hat{\mathbf{P}}^{MRQR2}\right) &= \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)E\left(F\left(\mathbf{X}\hat{\beta}\right)\right) + \lambda \mathbf{H}^{NP}E(\mathbf{p}) \\ &\approx \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\left\{\mathbf{P} + \left[\left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})\right\} + \lambda \mathbf{H}^{NP}\mathbf{P} \\ &= \mathbf{P} + \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\left[\left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})\end{aligned}$$

Given the expectation of $\hat{\mathbf{P}}^{MRQR2}$, the bias is

$$Bias\left(\hat{\mathbf{P}}^{MRQR2}\right) = \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\left[\left\langle f(\mathbf{X}\beta)\right\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle f(\mathbf{X}\beta)\right\rangle^{-1} - \mathbf{I}\right]\mathbf{H}(\mathbf{x})$$

Note that under the correctly specified model, $\hat{\mathbf{P}}^{MRQR2}$ is an unbiased estimate of $\mathbf{P}$.

Using the derivations previously obtained for the parametric procedure, the variance of the MRQR2 estimate can be obtained by expanding $\hat{\mathbf{P}}^{MRQR2}$ in the following manner:

$$\begin{aligned}\hat{\mathbf{P}}^{MRQR2} &\approx \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\left[\mathbf{A} + \mathbf{B}\mathbf{p}\right] + \lambda \mathbf{H}^{NP}\mathbf{p} \\ &= \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\mathbf{A} + \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\mathbf{B}\mathbf{p} + \lambda \mathbf{H}^{NP}\mathbf{p} \\ &= \left(\mathbf{I} - \lambda \mathbf{H}^{NP}\right)\mathbf{A} + \left[\mathbf{B} + \lambda \mathbf{H}^{NP}\left(\mathbf{I} - \mathbf{B}\right)\right]\mathbf{p}\end{aligned}$$

Since the first term of the above expression is constant, the variance of $\hat{\mathbf{P}}^{MRQR2}$ is

$$Var\left(\hat{\mathbf{P}}^{MRQR2}\right) \approx \left[\mathbf{B} + \lambda \mathbf{H}^{NP}\left(\mathbf{I} - \mathbf{B}\right)\right]\mathbf{V}_{\mathbf{G}}\left[\mathbf{B} + \lambda \mathbf{H}^{NP}\left(\mathbf{I} - \mathbf{B}\right)\right]' \tag{B.14}$$

Having mse formulas for the nonparametric and model-robust procedures introduced in this research allows one to obtain optimal values of the mixing parameter, λ, and the smoothing parameter, b, by minimizing the mse. These mse properties will also be a useful tool in comparing the various methods.

Vita

Quinton Jacob Nottingham, son of Edward L. Nottingham and Jennie B. Nottingham, was born on January 17, 1966 in Machipongo, Virginia. He graduated from Northampton High School, Eastville, in 1984, and enrolled in Virginia Polytechnic Institute and State University (VPI&SU), Blscksburg, Virginia. From there he graduated with a Bachelor of Science degree in Statistics in 1989. After graduation, he worked for Hoffmann-La Roche Pharmaceuticals in Nutley, New Jersey, as a Data Manager. In 1990, he enrolled in graduate school at VPI&SU in the Department of Statistics with after being awarded the State Council of Higher Education of Virginia (SCHEV) Fellowship. In 1991, he received his Master of Science degree in Statistics. He has accepted the position of Assistant Professor in the Management Science Department in the R. B. Pamplin School of Business at VPI&SU, starting August 15, 1995. He is married to Jacqueline Little (James) Nottingham.