

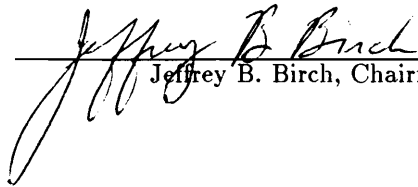
**Sequential Design Augmentation With Model Misspecification**

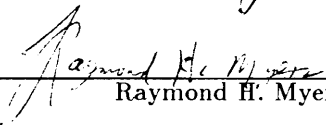
by

Sindee S. Sutherland

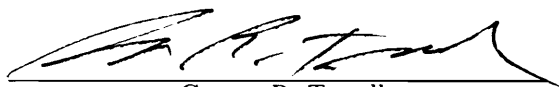
Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in  
Statistics

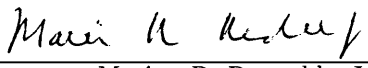
APPROVED:

  
\_\_\_\_\_  
Jeffrey B. Birch, Chairman

  
\_\_\_\_\_  
Raymond H. Myers

  
\_\_\_\_\_  
Eric P. Smith

  
\_\_\_\_\_  
George R. Terrell

  
\_\_\_\_\_  
Marion R. Reynolds, Jr.

Blacksburg, Virginia

December 14, 1992

C.2

LJ  
5655  
V856  
1992  
5884  
C.2

## **Sequential Design Augmentation With Model Misspecification**

by

Sindee S. Sutherland

Jeffrey B. Birch, Chairman

Statistics

(ABSTRACT)

In Response Surface Methodology (RSM) one attempts to model some variable of interest, usually as a known function of design variables. Subsequent analysis often indicates a need to move to a new region of interest. Many times the design is augmented by adding points sequentially to this new region of interest. Current methods of sequential design augmentation are used under the assumption of either correctly specified models or misspecification in the user's model that can be quantified, such as using a first order model when a second order model is correct. However, under model misspecification the sequential placement of points in the new region of interest using usual augmentation techniques may not be optimal, especially if the misspecification in the model is not due to polynomial terms.

A new methodology, based on a modified kernel regression procedure called HATLINK, is presented that incorporates model misspecification into the sequential augmentation of points in the new region. HATLINK is a combination of parametric and nonparametric regressions and is designed to perform best when the user has specified a reasonable approximate model. Parametric regression supplies a basic fit, while nonparametric regression allows adjustments to

compensate for some misspecification in the parametric model. The mixing parameter is determined adaptively through cross-validation. The augmentation is performed by a new technique called BIIV, the bias-influenced integrated prediction variance. BIIV attempts to select points that both minimizes the integrated prediction variance and the location where the current fit is the worst. Thus, BIIV incorporates an estimate of the bias due to misspecification of the parametric model into the augmentation procedure. It is shown that the designs generated by sequential design augmentation using HATLINK and BIIV are superior to designs from other methods.

## Acknowledgements

First, I would like to thank my husband, Keith Selander, who encouraged me to undertake the task of pursuing a doctoral degree. His love and support, especially during the difficult times, helped me throughout the five years it has taken. I would also like to thank him for all the discussions and suggestions made on my behalf. His help and emotional support was quite definitely a major factor in my completion of this degree.

I would also like to thank both my parents and my parents-in-law for their assistance and support, especially at the end. They helped in many ways too numerous to mention.

I would like to express my appreciation to all my professors in the Department of Statistics at Virginia Tech. I would like to thank all the members of my committee, but also express special appreciation to two members whose contributions have been greatest. I would like to especially thank my advisor, Jeffrey Birch, for his excellent editing and guidance throughout my research. Ray Myers, who is one of the best teachers in the statistics department, is appreciated for his inspiring courses and assistance in the rough places. This dissertation is very dependent on the material taught in his classes.

# TABLE OF CONTENTS

<b>I. INTRODUCTION .....</b>	<b>1</b>
<b>II. RESPONSE SURFACE METHODOLOGY</b>	
II.1 Description .....	6
II.2 Example of Response Surface Methodology .....	9
II.3 Some Current Methods of Design Augmentation .....	12
<b>III. KERNEL REGRESSION AND HATLINK</b>	
III.1 Description of Kernel Regression.....	15
III.2 Bandwidth Selection .....	18
III.3 Variations of Kernel Regression .....	19
III.4 Description of the Hatlink Procedure .....	20
III.5 Mixing Parameter Selection .....	23
III.5.A Press Methods for Selecting $\lambda$ .....	23
III.5.B $C_P$ Methods for Selecting $\lambda$ .....	24
III.6 Example of Fitted Hatlink Surface in RSM.....	28

**IV. SEQUENTIAL BATCH AUGMENTATION WITH HATLINK**

IV.1 Description of the Augmentation Procedure..... 31

IV.2 Sequential Batch Augmentation ..... 34

IV.3 Example of Sequential Design Augmentation with Batch Size of One ..... 37

IV.4 HATLINK Augmentation with BIIV vs IPV ..... 43

IV.5 L. S. Augmentation with BIIV vs IPV..... 46

IV.6 Example of Sequential Design Augmentation with Batch Size of Two..... 50

IV.7 Bias in the Model ..... 54

IV.8 Numerical Considerations ..... 57

IV.9 Analytical Considerations ..... 58

IV.10 Conclusions ..... 61

**V. RESULTS OF INITIAL SIMULATIONS**

V.1 Description of Study ..... 63

V.2 Augmentation Using the BIIV Method ..... 70

V.3 Effect of Batch Size Upon the HATLINK Method..... 77

V.4 Introduction to Repetitions..... 81

    V.4.A Case 3:  $CT = 0.7$  &  $\sigma = 1$  ..... 83

    V.4.B Case 4:  $CT = 1$  &  $\sigma = 0.025$  ..... 89

    V.4.C Case 5:  $CT = 1$  &  $\sigma = 0.1$  ..... 94

    V.4.D Case 6:  $CT = 1$  &  $\sigma = 0.2$  ..... 100

    V.4.E Case 7:  $CT = 1$  &  $\sigma = 0.5$  ..... 106

    V.4.F Case 8:  $CT = 1$  &  $\sigma = 1$  ..... 111

V.5 Summary of Results from Repetitions ..... 116

## **VI. THE HATLINK PROCEDURE IN MULTIPLE DIMENSIONS**

VI.1 Introduction .....	118
VI.2 First Example in Two Dimensions.....	119
VI.2.1 Numerical Comparisions of First Example.....	120
VI.2.2 Comparision of Plots of Fitted Surfaces for the First Example.....	125
VI.2.3 Comparision of Contour Plots for the First Example .....	137
VI.3 Second Example in Two Dimensions .....	149
VI.4 Introduction to Three Dimensions .....	180
VI.5 Example in Three Dimensions .....	181
VI.6 Conclusions .....	189

## **VII. SUMMARY AND AREAS FOR FURTHER RESEARCH**

VII.1 Summary of the Performance of Augmentation with HATLINK .....	191
VII.2 Areas for Further Research .....	193

## **VIII. REFERENCES .....**

196

## **APPENDIX 1. ANALYTICAL CONSIDERATIONS – FURTHER DETAILS .....**

199

## **APPENDIX 2. NUMERICAL CONSIDERATIONS**

A2.1 Numerical Considerations .....	205
A2.2 Examples of the Prediction Variance Plot with Different Bandwidths and Lambdas.....	206
A2.3 Approximating Integrals .....	210



**APPENDIX 3. ADDITIONAL CONSIDERATIONS**

A3.1 Augmentation Using Integrated Prediction Variance .....211

    A3.1.A IPV &  $C_P$  Methods for Selecting  $\lambda$  .....212

    A3.1.B IPV & Other  $C_P$  Criteria for Selecting  $\lambda$  .....214

    A3.1.C IPV & PRESS\* Criteria for Selecting  $\lambda$  .....216

    A3.1.D Generalized Cross Validation for Selecting  $\lambda$  .....217

    A3.1.E Summary of Augmentation Using IPV .....219

A3.2 Variations to the BIIV Method of Augmentation .....221

A3.3 Conclusions.....229

## LIST OF TABLES

II.2.1. Design Matrix and Observation Vector for Initial Region of Interest .....	10
IV.3.1. Six Points Augmented Using HATLINK and BIIV .....	41
IV.3.2. Six Points Augmented Using L. S. and BIIV .....	41
IV.4.1. Six Points Augmented Using HATLINK and BIIV .....	44
IV.4.2. Six Points Augmented Using HATLINK and Integrated Pred. Var. ....	44
IV.5.1. Six Points Augmented Using L. S. and BIIV .....	48
IV.5.2. Six Points Augmented Using L. S. and Integrated Pred. Var. ....	48
IV.6.1. Six Points Augmented Using HATLINK and BIIV – Batch Size of Two .....	52
IV.6.2. Six Points Augmented Using L. S. and BIIV – Batch Size of Two .....	52
V.3.1. ESE Values with Batch Size of One .....	79
V.3.2. ESE Values with Batch Size of Two.....	79
V.3.3. ESE Values with Batch Size of Three.....	80
V.3.4. ESE Values with Batch Size of Four .....	80
V.4.1. Values of $\sigma$ and CT Studied for Ten Repetitions .....	82

V.4.2. Average Distance and Fitted Maximum with $\sigma=1$ & $CT=0.7$ .....	88
V.4.3. Average ESE values with $\sigma = 1$ and $CT = 0.7$ .....	88
V.4.4. Average Distance and Fitted Maximum with $\sigma=0.025$ & $CT=1$ .....	93
V.4.5. Average ESE values with $\sigma = 0.025$ and $CT = 1$ .....	93
V.4.6. Average Distance and Fitted Maximum with $\sigma=0.1$ and $CT=1$ .....	99
V.4.7. Average ESE values with $\sigma = 0.1$ and $CT = 1$ .....	99
V.4.8. Average Distance and Fitted Maximum with $\sigma=0.2$ & $CT=1$ .....	105
V.4.9. Average ESE values with $\sigma = 0.2$ and $CT = 1$ .....	105
V.4.10. Average Distance and Fitted Maximum with $\sigma=0.5$ and $CT=1$ .....	110
V.4.11. Average ESE values with $\sigma = 0.5$ and $CT = 1$ .....	110
V.4.12. Average Distance and Fitted Maximum with $\sigma=1$ and $CT=1$ .....	115
V.4.13. Average ESE values with $\sigma = 1$ and $CT = 1$ .....	115
VI.2.1. Twelve Points Augmented Using HATLINK and BIIV – Example 1 .....	122
VI.2.2. Twelve Points Augmented Using L. S. and BIIV – Example 1 .....	123
VI.2.3. Twelve Points Augmented Using Kernel and BIIV – Example 1 .....	124
VI.3.1. Twelve Points Augmented Using HATLINK and BIIV – Example 2 .....	154
VI.3.2. Twelve Points Augmented Using L. S. and BIIV – Example 2 .....	155
VI.3.3. Twelve Points Augmented Using Kernel and BIIV – Example 2 .....	156
VI.5.1. Design Matrix & Responses in Initial Region of Interest – Three Variables .....	184
VI.5.2. Ten Points Augmented Using HATLINK and BIIV .....	186
VI.5.3. Ten Points Augmented Using L. S. and BIIV .....	187
VI.5.4. Ten Points Augmented Using Kernel and BIIV .....	188
A3.2.1. Variations to the Bias Part of BIIV .....	226

## LIST OF FIGURES

II.2.1. Fitted L. S. Surface in Old Region of Interest .....	11
III.6.1. Fitted HATLINK Surface in Old Region of Interest .....	29
III.6.2. True Surface in Old Region of Interest .....	30
IV.3.1. True Surface in New Region of Interest .....	40
IV.3.2. Plot of Six Points Augmented Using HATLINK & BIIV .....	42
IV.3.3. Plot of Six Points Augmented Using L. S. & BIIV .....	42
IV.4.1. Plot of Six Points Augmented Using HATLINK & BIIV .....	45
IV.4.2. Plot of Six Points Augmented Using HATLINK & Integrated Pred. Var. ....	45
IV.5.1. Plot of Six Points Augmented Using L. S. and BIIV .....	49
IV.5.2. Plot of Six Points Augmented Using L. S. and Integrated Pred. Var. ....	49
IV.6.1. Plot of Six Points Augmented Using HATLINK & BIIV – Batch Size of Two .....	53
IV.6.2. Plot of Six Points Augmented Using L. S. & BIIV – Batch Size of Two .....	53
V.1.1. True Surface in Old Region with $CT = .1$ .....	66
V.1.2. True Surface in New Region with $CT = .1$ .....	67

V.1.3. True Surface in Old Region with $CT = 1$ .....	68
V.1.4. True Surface in New Region with $CT = 1$ .....	69
V.2.1. Plot of ESE Values with $CT = 1$ & $\sigma = .1$ .....	72
V.2.2. Plot of ESE Values with $CT = 1$ & $\sigma = .2$ .....	73
V.2.3. Plot of ESE Values with $CT = 1$ & $\sigma = .3$ .....	74
V.2.4. Plot of ESE Values with $CT = 1$ & $\sigma = .5$ .....	75
V.2.5. Plot of ESE Values with $CT = 1$ & $\sigma = 1$ .....	76
V.4.1. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=0.7$ .....	85
V.4.2. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=0.7$ .....	86
V.4.3. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=0.7$ .....	87
V.4.4. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=0.025$ .....	90
V.4.5. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=0.025$ .....	91
V.4.6. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=0.025$ .....	92
V.4.7. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=0.1$ .....	96
V.4.8. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=0.1$ .....	97
V.4.9. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=0.1$ .....	98
V.4.10. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=0.2$ .....	102
V.4.11. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=0.2$ .....	103
V.4.12. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=0.2$ .....	104
V.4.13. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=0.5$ .....	107
V.4.14. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=0.5$ .....	108
V.4.15. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=0.5$ .....	109
V.4.16. Plot of Twelve Points with HATLINK, $\sigma=1$ & $CT=1$ .....	112
V.4.17. Plot of Twelve Points with L. S., $\sigma=1$ & $CT=1$ .....	113
V.4.18. Plot of Twelve Points with Kernel, $\sigma=1$ & $CT=1$ .....	114

VI.2.1. True Surface in New Region – Example 1 .....	126
VI.2.2. Fitted HATLINK Surface with Initial Points – Example 1 .....	127
VI.2.3. Fitted L. S. Surface with Initial Points – Example 1 .....	128
VI.2.4. Fitted Kernel Surface with Initial Points – Example 1 .....	129
VI.2.5. Fitted HATLINK Surface with Six Points – Example 1 .....	130
VI.2.6. Fitted L. S. Surface with Six Points – Example 1 .....	131
VI.2.7. Fitted Kernel Surface with Six Points – Example 1 .....	132
VI.2.8. Fitted HATLINK Surface with Twelve Points – Example 1 .....	133
VI.2.9. Fitted L. S. Surface with Twelve Points – Example 1 .....	134
VI.2.10. Fitted Kernel Surface with Twelve Points – Example 1 .....	135
VI.2.11. Contour Plots of True Surface in New Region – Example 1 .....	139
VI.2.12. Contour Plots of HATLINK Surface with Initial Points – Example 1 .....	140
VI.2.13. Contour Plots of L. S. Surface with Initial Points – Example 1 .....	141
VI.2.14. Contour Plots of Kernel Surface with Initial Points – Example 1 .....	142
VI.2.15. Contour Plots of HATLINK Surface with Six Points – Example 1 .....	143
VI.2.16. Contour Plots of L. S. Surface with Six Points – Example 1 .....	144
VI.2.17. Contour Plots of Kernel Surface with Six Points – Example 1 .....	145
VI.2.18. Contour Plots of HATLINK Surface with Twelve Points – Example 1 .....	146
VI.2.19. Contour Plots of L. S. Surface with Twelve Points – Example 1 .....	147
VI.2.20. Contour Plots of Kernel Surface with Twelve Points – Example 1 .....	148
VI.3.1. True Surface in Old Region – Example 2 .....	150
VI.3.2. True Surface in New Region – Example 2 .....	151
VI.3.3. Plot of Twelve Points Augmented using HATLINK & BIIV – Example 2 .....	157
VI.3.4. Plot of Twelve Points Augmented using L. S. & BIIV – Example 2 .....	158
VI.3.5. Plot of Twelve Points Augmented using KERNEL & BIIV – Example 2 .....	159

VI.3.6. True Surface in New Region – Example 2 .....	160
VI.3.7. Fitted HATLINK Surface with Initial Points – Example 2 .....	161
VI.3.8. Fitted L. S. Surface with Initial Points – Example 2 .....	162
VI.3.9. Fitted Kernel Surface with Initial Points – Example 2 .....	163
VI.3.10. Fitted HATLINK Surface with Six Points – Example 2 .....	164
VI.3.11. Fitted L. S. Surface with Six Points – Example 2 .....	165
VI.3.12. Fitted Kernel Surface with Six Points – Example 2 .....	166
VI.3.13. Fitted HATLINK Surface with Twelve Points – Example 2 .....	167
VI.3.14. Fitted L. S. Surface with Twelve Points – Example 2 .....	168
VI.3.15. Fitted Kernel Surface with Twelve Points – Example 2 .....	169
VI.3.16. Contour Plots of True Surface in New Region – Example 2 .....	170
VI.3.17. Contour Plots of HATLINK Surface with Initial Points – Example 2 .....	171
VI.3.18. Contour Plots of L. S. Surface with Initial Points – Example 2 .....	172
VI.3.19. Contour Plots of Kernel Surface with Initial Points – Example 2 .....	173
VI.3.20. Contour Plots of HATLINK Surface with Six Points – Example 2 .....	174
VI.3.21. Contour Plots of L. S. Surface with Six Points – Example 2 .....	175
VI.3.22. Contour Plots of Kernel Surface with Six Points – Example 2 .....	176
VI.3.23. Contour Plots of HATLINK Surface with Twelve Points – Example 2 .....	177
VI.3.24. Contour Plots of L. S. Surface with Twelve Points – Example 2 .....	178
VI.3.25. Contour Plots of Kernel Surface with Twelve Points – Example 2 .....	179
A2.2.1. Plot of Prediction Variance – $h$ & $\lambda$ selected by data.....	207
A2.2.2. Plot of Prediction Variance – $h=12.5$ & $\lambda=0.90$ .....	208
A2.2.3. Plot of Prediction Variance – $h=0.20$ & $\lambda=0.90$ .....	209

# CHAPTER I

## INTRODUCTION

In Response Surface Methodology (RSM) one desires to model some variable of interest  $\eta$  or  $Y$ , called the response variable, as a function of  $k$  different variables  $x_1, x_2, x_3, \dots, x_k$ , called the design variables. The true relationship between the response variable and the design variables is usually unknown and may be rather complex. The purpose of RSM is to approximate this relationship with a known function of the design variables. Thus the response variable can be written as

$$\eta = Y = f(x_1, x_2, x_3, \dots, x_k) + \epsilon,$$

where  $f$  is some function of  $(x_1, x_2, x_3, \dots, x_k)$  and  $\epsilon$  is random error. Usually either a first order polynomial of the form

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon \quad (I.1)$$

or a second order polynomial of the form

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon \quad (I.2)$$



is used as the estimate of the function  $f$ ,  $\hat{f}$ .

Once the form of  $\hat{f}$  is chosen, the data is then used to estimate the unknown parameters, customarily by the method of least squares. RSM techniques are then used to both locate the values of the design variables that will either maximize or minimize the response variable and to explore the surface. Often the data are originally taken in a particular region and subsequent analysis reflects a need to move to a new region. Many times this new region somewhat overlaps the original region. In this case, additional points need to be augmented in the new region in order to fit a new model.

Assuming the two regions overlap somewhat, there are currently three main ways to augment with these additional points. One common way is to ignore all the previous data points, choose an experimental design, rescale and center the new region and take new data points based on the chosen design. This method completely ignores any data from the old region, especially those in the overlap region. Another way is to choose an experimental design, rescale and center the new region and take the data points based on the design augmented with the points from the initial region that lie in the overlap region. A different method proposed by Rozum (1991) suggests taking all data from the overlap region and augmenting with points sequentially in the new region by either D-optimality (or one of the other alphabetic optimality criteria) or integrated spherical variance. The method proposed by Rozum takes advantage of some of the existing data. These methods tend to assume there is minimal model misspecification and that the first or second order model is a very good approximation to the function  $f$ .

While these methods work under correct model specification, it is desirable to have a method of augmentation that uses all the existing data and is more robust to model misspecification than the current methods. By applying Einsporn's (1987) concept of HATLINK,

one can achieve both of these goals. HATLINK fits a regression function that is a convex linear combination of the user's proposed parametric model and some nonparametric model. By estimating a mixing parameter, a regression function estimate,  $\hat{f}$ , can be obtained that is a mixture of both the parametric and nonparametric models. Throughout this dissertation, the parameters of the parametric model will be estimated using the method of least squares, while kernel regression will be used to fit the nonparametric model.

In nonparametric kernel regression, it is assumed that the true response function is unknown and no parametric form for  $f$  need be specified by the user. If one wishes to predict the mean response of  $Y$  at a point  $\underline{x}_0 = (x_{01}, \dots, x_{0k})'$ , kernel regression takes a weighted average of the observed  $Y$ 's where the weights are related to the distance between the observed points  $\underline{x}_i$  and the prediction point  $\underline{x}_0$ . In general, this method gives more weight to the points closer to  $\underline{x}_0$  and less weight to those further from  $\underline{x}_0$ . Since one can predict the mean response of  $Y$  at any given  $\underline{x}$  location an estimated (nonlinear) regression curve can be produced. Unfortunately, no closed form expression for this curve exists, but the graph of this curve may suggest the true form of the function  $f$ .

Thus, by adjusting the mixing parameter, HATLINK can estimate a response surface by using either the parametric model, the nonparametric model, or a combination of the two. This parameter is estimated by either PRESS-type or  $C_P$ -type criteria. If the parametric function is a good approximation to the true unknown function, the mixing parameter will tend to produce an estimate of the response surface based more on this function. If, on the other hand, the user has misspecified the model too much, the mixing parameter will yield an estimate based more on the nonparametric model. Thus, HATLINK should be more robust to model misspecification and give a better idea of the true response surface in question.

Once the bandwidth and mixing parameter have been estimated, the method proposed

here suggests retaining all the original data points and augmenting by batches of points individually in the new region. This is called sequential batch design augmentation. Because of the nature of kernel regression, the proposed method requires at least one starting point that lies completely in the new region. The individual data points will be augmented to minimize some type of criteria. Various criteria, such as integrated prediction variance (IPV) or integrated mean squared error (IMSE), could be used in augmenting with these points sequentially. One procedure proposed in this work suggests augmenting with points to minimize integrated prediction variance over the new region. That is, place the new point  $\underline{x}_0$  in the new region to minimize

$$\int_{\text{new region}} \text{Var}[\hat{Y}(\underline{x})] d\underline{x} . \quad (\text{I.3})$$

A second method, which seems to perform better in simulations, is augmenting by points to minimize the bias-influenced integrated prediction variance, BIIV. In this case, place the new point  $\underline{x}_0$  in the new region to minimize

$$- [\text{bias}(\underline{x}_0)]^2 + \int_{\text{new region}} \text{Var}[\hat{Y}(\underline{x})] d\underline{x} . \quad (\text{I.4})$$

This attempts to select the point  $\underline{x}_0$  that both minimizes the integrated prediction variance and finds the location where the fit is currently the worst. Under certain circumstances this procedure appears to yield a better fitted surface than ones obtained using other techniques. Both methods will be considered further in subsequent chapters.

A third method is augmentation to minimize IMSE, that is select the new point  $\underline{x}_0$  in

the new region of interest to minimize

$$\int_{\text{new region}} \left( [\text{bias}(\underline{x})]^2 d\underline{x} + \text{Var}[\hat{Y}(\underline{x})] \right) d\underline{x}. \quad (1.5)$$

The use of IMSE as the augmentation criterion yielded results very similar to those for BIIV, while requiring extra computational effort. Because of this, IMSE as an augmentation criterion will not be considered further.

After moving to a new region, most augmentation procedures use at most only the data points in the new region. This proposed method is different since it uses all existing data points, both in the old and new regions. This is achieved through the HATLINK procedure.

Chapter II presents background information on parametric response surfaces and least squares regression. Chapter III presents background information on kernel regression and the HATLINK method. Chapter III also contains a brief literature review on kernel regression and some other nonparametric regression techniques. Chapter IV describes the general HATLINK augmentation procedure and includes some examples of its use. Also some associated technical matters are briefly discussed.

To better understand the behavior of the HATLINK procedure, in particular the selection of the mixing parameter, the results of some exploratory simulations are presented in Chapter V. The results of simulations for a variety of situations are presented in Chapter VI. Chapter VII contains a summary of the findings and suggests some areas for further study. Some mathematical developments are included in the Appendix, along with the computer program developed for this research.

# CHAPTER II

## RESPONSE SURFACE METHODOLOGY

### II.1 DESCRIPTION

In RSM an experimental design is chosen to achieve optimality in some sense. Usually these designs are orthogonal and, depending upon the design, enables either a first or second order model to be fit. The general procedure used in RSM is to start with a region of interest and choose an appropriate experimental design. The natural units of the design variables,  $\zeta_i$ , are transformed into design units,  $x_i$  by  $x_i = \frac{\zeta_i - \bar{\zeta}}{d_i}$  where  $\bar{\zeta}$  is the arithmetic mean of the  $\zeta_i$  and  $d_i$  is the scaled spacing between the high and low levels, that is  $d_i = (\zeta_{i,high} - \zeta_{i,low})/2$ . This standardizes the experimental region. After choosing the design, N experimental runs are performed.

The usual assumption is that the response surface function is well approximated by a low order polynomial function. Then the response surface is fit with either the first or second order model as stated in (I.1) and (I.2). Both models can be written as

$$\underline{Y} = X\underline{\beta} + \epsilon \tag{II.1.1}$$

where  $\underline{Y}$  is the vector of response values, X is an (N x p) matrix and p is the number of

parameters to be estimated,  $\underline{\beta}$  is the  $(p \times 1)$  vector of model parameters and  $\underline{\epsilon}$  is the  $(N \times 1)$  vector of error terms. Here each column of  $X$  represents a term in the model and each row of  $X$  represents an experimental run. It is further assumed that  $\underline{\epsilon} \sim N(0, \sigma^2)$ .

The least squares estimate of  $\underline{\beta}$  is given by

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{Y} \quad (\text{II.1.2})$$

and thus the vector of predicted response values is given by

$$\hat{\underline{Y}} = X \hat{\underline{\beta}}. \quad (\text{II.1.3})$$

Prediction at a particular data point  $\underline{x}_i$  is given by  $\hat{Y}(\underline{x}_i) = \underline{x}_i' \hat{\underline{\beta}}$  and is commonly denoted by  $\hat{Y}_i$ . The variance of this prediction is  $\text{Var}(\hat{Y}(\underline{x}_i)) = \sigma^2 \underline{x}_i' (X'X)^{-1} \underline{x}_i$ . The properties of  $\hat{\underline{\beta}}$  and  $\hat{\underline{Y}}$  are well-known and are given by

$$\begin{aligned} E(\hat{\underline{\beta}}) &= \underline{\beta} \\ \text{Var}(\hat{\underline{\beta}}) &= \sigma^2 (X'X)^{-1} \\ E(\hat{\underline{Y}}) &= X \underline{\beta} \\ \text{Var}(\hat{\underline{Y}}) &= \sigma^2 X (X'X)^{-1} X' \end{aligned}$$

Note that under the assumption of  $f$  being well approximated by a low order polynomial function there is no bias in either the estimated coefficients or the predicted values.

Using (II.1.2), one can write (II.1.3) as  $\hat{\underline{Y}} = X(X'X)^{-1}X'\hat{\underline{Y}} = H\underline{Y}$ , where  $H$  is the so-called hat matrix, so named due to its role in making predictions. As detailed in Myers (1990), the  $H$  matrix has several very important properties. Some of these are

- (i) H is symmetric and idempotent
- (ii)  $-1 \leq h_{ij} \leq 1$
- (iii)  $\sum_{i=1}^n h_{ii} = p$ , where p is the number of parameters to be estimated
- (iv)  $\sum_{j=1}^n h_{ij} = 1 \quad \forall i$ .

Many regression results can be written in terms of the hat matrix or its elements. Two very important results that are used extensively in this work are

- (i)  $\text{Var}(\hat{\underline{Y}}) = \text{Var}(H \underline{Y}) = \sigma^2 H$
- (ii)  $\text{Var}(\hat{Y}_i) = \sigma^2 (H)_{ii} = \sigma^2 h_{ii}$ .

These results are derived from the properties of the hat matrix.

Many times one wishes to predict at regressor locations not in the design. Then for such a point  $\underline{x}_0$ , where  $\underline{x}_0 = (1, x_{01}, x_{02}, x_{03}, \dots, x_{0k})'$ , the predicted response based upon the least squares estimate is given by

$$\hat{Y}(\underline{x}_0) = \underline{x}_0' \hat{\underline{\beta}}$$

and the variance for this prediction is given by

$$\text{Var}(\hat{Y}(\underline{x}_0)) = \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0.$$

Thus one may calculate the predicted response and the corresponding variance for any location on the response surface.

If a first order model is to be fit, it is customary to add a few center runs to the original design. This permits one to test if a second order model would fit the desired region better than a first order model. If the first order model is not rejected, then the method of steepest ascent may be used to move sequentially on a path away from the center of the region in a direction where one expects either the greatest increase or decrease in response. Frequently steepest ascent travels on a path to the edge of the region, consequently indicating a new region of interest. Subsequent design augmentation takes place in the new region and enough data points to estimate the parameters for the model in the new region are required. As mentioned in the introduction, at this point several ways of proceeding exist.

## II.2 EXAMPLE OF RESPONSE SURFACE METHODOLOGY

In order to illustrate the RSM procedure, a simple example in two design variables is presented. Suppose in an experiment it is desired to study the influence of temperature ( $X_1$ ) and epoxy additive ( $X_2$ ) upon the strength of a particular plastic. The user wishes to maximize the strength. Three levels of the variables were used in the experimental design. After centering and scaling, the design matrix,  $D$  and the corresponding vector,  $\underline{Y}$ , of observations are given in Table II.2.1. If a first order model is fit, then the  $X$  matrix is  $[\underline{1} \mid D]$ , where  $\underline{1}$  is a (20 x 1) vector of 1's.

It is assumed that the user has prescribed a first order model,  $Y = \beta_0 + \sum_{i=1}^2 \beta_i x_i + \epsilon$ . A graph of the fitted first order response surface model is included as Figure II.2.1. The fitted model was  $\hat{Y} = 30.829 + 6.091x_1 + 5.611x_2$  and produced  $R^2 = 0.845$ . Note that to maximize



**Table II.2.1. Design Matrix and Response Vector for Initial Region of Interest.** In this example,  $X_1$  is temperature,  $X_2$  is epoxy additive and  $\underline{Y}$  is the strength of a plastic. It is desired to maximize the strength of the plastic.

	$X_1$	$X_2$		$\underline{Y}$	
$D =$	{	1	1	{	39.5831
		1	1		38.6040
		1	-1		33.3899
		1	-1		32.6781
		-1	1		33.9784
		-1	1		33.3788
		-1	-1		13.2757
		-1	-1		17.3391
		1	0		32.8431
		1	0		34.3771
		-1	0		23.0180
		-1	0		23.1481
		0	1		39.0582
		0	1		39.3992
		0	-1		26.8865
		0	-1		27.3446
		0	0		30.2539
		0	0		33.6762
		0	0		33.2740
		0	0		31.0812

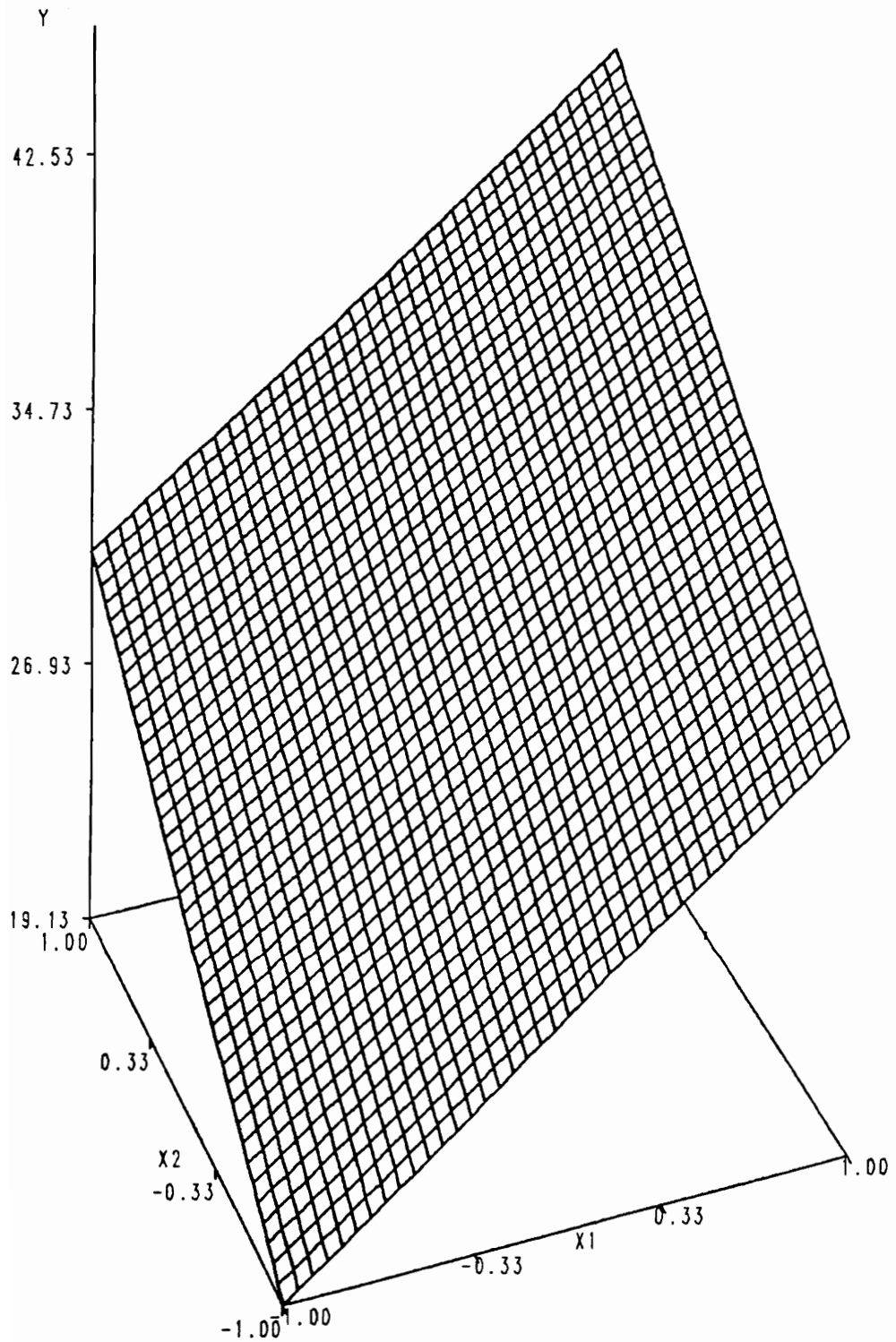


FIG. 11.2.1. A FITTED FIRST ORDER LEAST SQUARES SURFACE  
 WITH  $CT=0.7$  &  $RESPONSE = 30.829 + 6.091 \cdot X1 + 5.611 \cdot X2$

the response, analysis by steepest ascent would indicate moving to the (1, 1) corner of the initial region. This indicates a need, perhaps, to move to a new region to locate the maximum response.

### II.3 SOME CURRENT METHODS OF DESIGN AUGMENTATION

One very common method of exploring the new region is to ignore the old region and the original data points completely. The procedure in the new region is identical to that followed in the old region. That is, an experimental design is chosen, the region centered and rescaled, and the new data points are taken. Using the data, the model parameters are estimated. This can become very expensive and wasteful of perfectly good data.

A second method is to choose a design for the new region, center and rescale the new region and take the data points. Now any points in the overlap region (if there is one) are added to the design. This gives more information to use in estimating parameters but this method too can become wasteful since some data from the old region are ignored.

A third method is to retain the data points in any overlap region and augment with additional points. These points can be augmented by several different criteria. These criteria all belong to the family of alphabetic optimality criteria. All are conditional optimality criteria since the methods depend upon the initial design points. These criteria include conditional D-optimality, conditional A-optimality, conditional E-optimality and conditional  $I_\lambda$ - optimality.

For conditional D-optimality one augments by the point  $\underline{x}_j$  where the prediction variance is the largest. This maximizes  $|X'X|$ , where  $X$  is composed of both the design points in the overlap region and new design points. Conditional D-optimality is not very robust to model misspecification and tends to put points on the edge of the design. It is probably the most

commonly used method of augmentation. Several different methods of augmenting with conditional D-optimality have been studied in some detail. First order design augmentation, using D-optimality, has been studied by many people including Dykstra (1966), Gaylor and Merrill (1968), and Wynn (1970). Using D-optimality for second order design augmentation has been studied by Dykstra (1971) and Hebble and Mitchell (1972) among others. Simultaneous multiple point design augmentation in the first and second order case, using D-optimality, has been studied by Welch (1982) and Evans (1979).

Conditional G-optimality augments with a point to minimize the maximum predicted variance over all locations  $\underline{x}$  in the region of interest. Conditional A-optimality augments by a point to minimize the trace of  $(X'X)^{-1}$ . It can be shown that this is equivalent to minimizing the sum of the variances of the estimated coefficients. Another type of criteria is conditional E-optimality. This augments within the current region of interest to minimize the maximum eigenvalue of  $(X'X)^{-1}$ . Further discussions regarding D, G, A, and E optimality may be found in Kiefer (1959) and Kiefer and Wolfowitz (1959, 1960).

Conditional  $I_{\lambda}$ -optimality augments with a point to minimize  $\int_{\underline{x}} \underline{x}'(X'X)^{-1}\underline{x} \lambda(\underline{x})$  where  $\lambda(\bullet)$  is a probability measure on the space of predictor variables. This criterion also belongs to a more general class of L-optimality criteria discussed by Federov(1972). There are two particular  $\lambda(\underline{x})$ 's that will be of interest and are discussed below. These are integrated prediction variance and spherical prediction variance on a sphere with radius r.

Integrated prediction variance is denoted as  $I_{IV}$  and conditional  $I_{IV}$ -optimality augments with a point to minimize

$$K \int_R \underline{x}'(X'X)^{-1}\underline{x} d\underline{x} \tag{II.3.1}$$

where  $K^{-1} = \int_R d\underline{x}$  and  $\lambda(\underline{x}) = K d\underline{x}$ .

Here,  $R$  represents the current region of interest which is very often either the hypercube  $\{\underline{x} : a_i \leq x_i \leq b_i \forall i\}$  or the hypersphere  $\{\underline{x} : \sum_{i=1}^k \tilde{x}_i^2 \leq \rho^2 \text{ and } \tilde{x}_i = (x_i - t_i)\}$  with center  $(t_1, t_2, \dots, t_k)$ . This method initially requires enough points in the current region of interest to estimate the model parameters. Consequently, one may be required to initially augment with a group of points to estimate the model parameters.

Augmenting by integrated prediction variance is the method closest to augmentation method (I.3) used in this dissertation. The integrated prediction variance method studied in this dissertation was suggested by Rozum (1990), who only used design points in the new region. Through HATLINK, this dissertation studies augmentation criterion using all the design points in both the old and new regions.

Similar to integrated prediction variance is a method called spherical prediction variance on a sphere with radius  $r$ , which is denoted as  $I_{sv_r}$ . This consists of augmenting by the point to minimize integrated prediction variance on a sphere with radius  $r$ , where  $r^2 \leq \rho^2$ , the radius of the hypersphere. This is contrasted to integrated prediction variance, which is calculated over the entire region. The concept of a spherical prediction variance is attributable to Hussey, Myers and Houck (1987). It was further explored and studied by Giovannitti-Jensen and Myers (1989). While this method is not utilized in this dissertation, it could be studied in future research.

When augmentation in a new region is desired, all commonly used augmentation criteria use the least-squares estimator and assume that the parametric model is correctly specified. Furthermore, if any of the original data points are used, only those that also lie in the new region are utilized. They also tend to yield designs that are not robust to model-misspecification. The method proposed here uses all the data points in both regions. And, as will be demonstrated, the method will also be more robust to model-misspecification through the nonparametric portion of the HATLINK method.

# CHAPTER III

## KERNEL REGRESSION AND HATLINK

### III.1 DESCRIPTION OF KERNEL REGRESSION

As stated in the introduction, kernel regression is a nonparametric regression technique based upon a linear combination of weighted observed response values. In the case of one regressor variable where the  $x_i$  are data locations and the  $Y_i$  represent the corresponding observed values, Watson (1964) and Nadaraya (1964) proposed an estimator of the unknown function  $f(x_i)$  to be

$$\hat{Y}^*(x_i) = \sum_{j=1}^n h_{ij}^* Y_j \quad (\text{III.1.1})$$

where

$$h_{ij}^* = \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_i - x_j}{h}\right)} \quad \text{for some } h \geq 0. \quad (\text{III.1.2})$$

The  $h_{ij}^*$  are the weights on the observed values and depend upon some constant  $h$ , referred to as the bandwidth, and some function  $K(u)$ . The function  $K(u)$  depends upon the distances between

the data point of interest  $x_i$  and the data points  $x_j$  and also depends upon the constant  $h$ .

By calculating (III.1.1) for all the data points, a vector of predicted response values can be achieved, that is

$$\hat{\underline{Y}}^* (X) = H^* \underline{Y} , \quad (III.1.3)$$

where  $H^* = (h_{ij}^*)$  is the matrix of weights associated with kernel regression. This creates a type of hat matrix similar to the hat matrix from least squares regression. Just as the rows of the regression hat matrix sum to one, the denominator in (III.1.2) makes the rows of  $H^*$  sum to one, that is, for each  $i$ ,  $\sum_{j=1}^n h_{ij}^* = 1$ . Unlike the regression hat matrix,  $H^*$  is rarely symmetric or idempotent. In many respects, this kernel hat matrix may be used like the hat matrix from regression.

Prediction at any non-data location  $x_0$  may be calculated by

$$\hat{Y}^* (x_0) = \sum_{j=1}^n h_{0j}^* Y_j \quad (III.1.4)$$

where

$$h_{0j}^* = \frac{K\left(\frac{x_0 - x_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_0 - x_j}{h}\right)} . \quad (III.1.5)$$

Thus, it is possible to obtain predictions at any location and so kernel regression can produce an estimated regression curve. Unfortunately, no closed form expression for  $\hat{f}(x)$  can be obtained.

The function  $K(u)$  which appears in (III.1.2) and (III.1.5) is called the kernel function and is often thought of as a probability density function. Butler (1975) recommends that  $K$  and

its first derivative be smooth and continuous and that  $K'(u)$  be zero when  $u = 0$ . Several authors have suggested that the form of  $K(u)$  is not crucial, thus most work has been done using only a few main kernel functions. Many kernel functions are non-negative and symmetric about zero. The kernel function used in this work is

$$K(u) = e^{-u^2} \quad (\text{III.1.6})$$

where  $u = \frac{(x_i - x_j)}{h}$  for any two regressor values. Because of the similarity of (III.1.6) to the normal distribution, this kernel function is often called a normal kernel function. The kernel hat matrix  $H^*$  using (III.1.6) then becomes

$$H_{ker} = (h_{ij}^{ker}) \quad (\text{III.1.7})$$

$$\text{where } h_{ij}^{ker} = \frac{e^{-(x_i - x_j)^2/h^2}}{\sum_{j=1}^n e^{-(x_i - x_j)^2/h^2}} .$$

For multiple regression the distance measure is the customary Euclidian measure and thus the kernel hat matrix becomes

$$(h_{ij}^{ker}) = \frac{e^{-\|\underline{x}_i - \underline{x}_j\|^2/h^2}}{\sum_{j=1}^n e^{-\|\underline{x}_i - \underline{x}_j\|^2/h^2}} \quad (\text{III.1.8})$$

$$\text{where } \|\underline{x}_i - \underline{x}_j\|^2 = (\underline{x}_i - \underline{x}_j)'(\underline{x}_i - \underline{x}_j).$$

The availability of the kernel hat matrix enables kernel regression to perform those procedures and techniques that utilize the least squares hat matrix.



### III.2 BANDWIDTH SELECTION

The bandwidth,  $h$ , is always positive and determines how much influence the observations have in making predictions. Small values of  $h$  tend to yield predictions based only on the closest observations. That is, as  $h \rightarrow 0$ ,  $\hat{Y}^*(x_0)$  converges to a kernel prediction based only on the observations at  $x_0$ . Conversely, large values of  $h$  tend to produce predictions based upon the average of all the observations. That is, as  $h \rightarrow \infty$ ,  $\hat{Y}^*(x_0)$  converges to  $\bar{Y}$ . Thus, when generating predicted regression curves based upon kernel regression, one can vary the smoothness of the curve from a “connect the dots” approach to a constant line by varying  $h$ .

For choosing the value of  $h$ , Wong (1983) and Rice (1984) show consistency results for kernel regression estimates with bandwidths obtained by a PRESS-type approach. By a consistent estimator of the bandwidth, they mean that, asymptotically,  $\hat{Y}^*(\underline{x}, h)$  converges to  $Y(\underline{x})$ . The PRESS procedure (Allen 1974) is a “leave-one-out” method and is commonly referred to in the literature as cross-validation. Stone (1974) and Geisser (1975) introduced the notion of PRESS in a very general framework. The PRESS procedure selects the bandwidth  $h$  so that the quantity

$$\text{PRESS}(h) = \sum_{i=1}^n \left( Y_i - \hat{Y}_{i,-i}^* \right)^2 \quad (\text{III.2.1})$$

is minimized. Here,  $\hat{Y}_{i,-i}^*$  represents the predicted value of  $Y$  at the location  $x_i$  based upon kernel regression with the  $i^{\text{th}}$  data point left out. Based upon empirical studies, Einsporn (1987) recommended using

$$\text{PRESS}^*(h) = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_{i,-i}^*)^2}{n - \text{tr}[H^*(h)]} \quad (\text{III.2.2})$$

instead of the original PRESS as defined in (III.2.1). Here,  $H^*(h)$  represents the kernel hat matrix for a specific value of  $h$ . Einsporn suggested that this version of PRESS yields better kernel prediction. The denominator helps to protect against overfitting to the data since  $\text{tr}[H^*]$  becomes large for very small bandwidths. Contrary to appearances, calculation of  $\hat{Y}_{i,-i}^*$  may easily be obtained by the formula

$$\hat{Y}_{i,-i}^* = \sum_{j \neq i} h_{ij,-i}^* Y_j, \quad (\text{III.2.3})$$

where  $h_{ij,-i}^* = \frac{h_{ij}^*}{(1 - h_{ii}^*)}$ .

The denominator is to ensure that  $\sum_{j=1}^n h_{ij,-i}^* = 1$ .

### III.3 VARIATIONS OF KERNEL REGRESSION

Several variations to the basic kernel regression method outlined previously have been suggested. One such variation is local linear regression and was introduced by Cleveland (1979). Müller (1987) demonstrates the close relationship between local linear regression and the kernel method. A second variation is nonparametric spline regression. In his paper, Silverman (1985) presents a review of spline regression and indicates a connection between spline regression and kernel regression. He gives the “effective” kernel function and bandwidth for spline regression.

The bandwidth depends upon the local density of the  $x$ 's and so this corresponds to a local bandwidth method. Both Müller and Stadtmüller (1985) and Marriot (1985) prefer kernel regression to spline regression and list several reasons for their preference.

There are other variations to the basic kernel regression method. To select the bandwidth, one could use a different version of (III.2.1). The PRESS-type methods for selecting the bandwidth outlined in section 2 are global procedures. A different approach for selecting the bandwidth is to use some type of local bandwidth selection method. Two distinct methods have been proposed in the literature. One method is used by both Cleveland (1979) and Georgiev and Greblicki (1986). The other method is recommended by Müller and Stadtmüller (1985, 1987).

A different approach to calculating (III.1.5) was proposed by Priestley and Chao (1972). Instead of dividing by  $\sum K\left(\frac{x_0 - x_j}{h}\right)$  in (III.1.5), they recommend multiplying by  $\left(\frac{x_{j+1} - x_j}{h}\right)$  for  $j=1, \dots, n-1$ . They assume that  $x_1 < x_2 < \dots < x_n$ . Benedetti (1975) found this estimate to have inferior properties to the usual kernel function. An improved version was suggested by Gasser and Müller (1979) and studied by Cheng and Lin (1981).

The kernel function used in this paper is, by no means, the only kernel function possible. In the area covered by this paper, future work could consider other kernel functions.

#### III.4 DESCRIPTION OF THE HATLINK PROCEDURE

The HATLINK procedure, as suggested by Einsporn (1987), is a linear combination of predictions from a parametric model and a nonparametric regression. The parameters for the parametric model are estimated by the method of least squares. It should be recognized at this point that other estimation methods could be used instead of least squares. For example,  $m$ -estimation could be used if outliers were of concern to the researcher. The role of alternative

parametric estimation procedures will be discussed in more detail in chapter VII as an area of further research. Analogously, kernel regression will be used to obtain predictions for the nonparametric regression. Chapter VII will also discuss the use of alternative nonparametric estimation methods. Let  $H_{ols} = X(X'X)^{-1}X' = (h_{ij}^{ols})$  be the usual hat matrix associated with least squares regression for the hypothesized model. Let  $H_{ker} = (h_{ij}^{ker})$  be the hat matrix associated with the basic method of kernel regression as discussed above, using the normal kernel function. Now  $\hat{\underline{Y}}_{ols} = H_{ols} \underline{Y}$  are the least squares predictions and  $\hat{\underline{Y}}_{ker} = H_{ker} \underline{Y}$  are the kernel predictions at the data locations. Furthermore predictions at an individual data location  $\underline{x}_i$  are given by  $\hat{Y}_i^{ols} = \sum_{j=1}^n h_{ij}^{ols} Y_j$  and  $\hat{Y}_i^{ker} = \sum_{j=1}^n h_{ij}^{ker} Y_j$  for least squares and kernel regression, respectively. If the hypothesized model is not an accurate approximation to the true function, then prediction may be improved by emphasizing the kernel prediction more than the least squares prediction. This is the idea behind the HATLINK method.

To apply the HATLINK method, define a new hat matrix  $H(\lambda)$  by

$$H(\lambda) = \lambda H_{ker} + (1 - \lambda) H_{ols} \quad (\text{III.4.1})$$

for some real number  $0 \leq \lambda \leq 1$ . The value  $\lambda$  is referred to as the mixing parameter. Prediction at the vector of data locations is given by

$$\begin{aligned} \hat{\underline{Y}}(\lambda) &= H(\lambda) \underline{Y} \\ &= \lambda \hat{\underline{Y}}_{ker} + (1 - \lambda) \hat{\underline{Y}}_{ols} \end{aligned} \quad (\text{III.4.2})$$

and at an individual data location  $\underline{x}_i$  by

$$\begin{aligned}
\hat{Y}(x_i, \lambda) &= \sum_{j=1}^n h_{ij}(\lambda) Y_j \\
&= \sum_{j=1}^n [\lambda h_{ij}^{ker} + (1-\lambda) h_{ij}^{ols}] Y_j .
\end{aligned} \tag{III.4.3}$$

Note that  $h_{ij}(\lambda)$  represents the  $ij^{th}$  entry of  $H(\lambda)$ . Also, to ease the notation,  $\hat{Y}(x_i, \lambda)$  will be written as  $\hat{Y}_i(\lambda)$ .

Prediction at a data location,  $\underline{x}_0$ , not in the design is given by

$$\begin{aligned}
\hat{Y}_0(\lambda) &= \sum_{j=1}^n h_{oj}(\lambda) Y_j \\
&= \sum_{j=1}^n [\lambda h_{oj}^{ker} + (1-\lambda) h_{oj}^{ols}] Y_j .
\end{aligned} \tag{III.4.4}$$

$$\text{where } h_{oj}^{hat} = \frac{e^{-\|\underline{x}_o - \underline{x}_j\|^2/h^2}}{\sum_{j=1}^n e^{-\|\underline{x}_o - \underline{x}_j\|^2/h^2}} \text{ and } \|\underline{x}_o - \underline{x}_j\|^2 = (\underline{x}_o - \underline{x}_j)'(\underline{x}_o - \underline{x}_j). \tag{III.4.5}$$

The prediction variance at  $\underline{x}_0$  is given by

$$\begin{aligned}
\text{Var}[\hat{Y}_0(\lambda)] &= \sigma^2 \sum_{j=1}^n [h_{oj}(\lambda)]^2 \\
&= \sigma^2 \sum_{j=1}^n [\lambda h_{oj}^{ker} + (1-\lambda) h_{oj}^{ols}]^2 .
\end{aligned} \tag{III.4.6}$$

### III.5 MIXING PARAMETER SELECTION

If  $\lambda = 0$ , then the predicted value is based entirely upon the parametric model and least squares regression and if  $\lambda = 1$ , then the prediction is based entirely upon the nonparametric model kernel regression. Thus, to some extent, the degree of model misspecification can be represented by the value of  $\lambda$ . Several methods of selecting  $\lambda$  have been suggested. One method is by cross-validation using the PRESS procedure where  $\lambda$  is chosen so that the quantity

$$\text{PRESS}(\lambda) = \sum_{i=1}^n [Y_i - \hat{Y}_{i,-i}(\lambda)]^2 \quad (\text{III.5.1})$$

is minimized. A second method is by a  $C_P$  - type of criteria.

#### III.5.A PRESS METHODS FOR SELECTING $\lambda$

Choosing  $\lambda$  by cross-validation requires the PRESS criterion to achieve a proper balance between overfitting and underfitting. In empirical studies, Einsporn (1987) notes that (III.5.1) tended to choose a  $\lambda$  that was too large. In such cases, the HATLINK fit may be very close to the kernel fit thus overfitting to random induced "trends" in the data. He recommends a more conservative method of calculating PRESS. In a similar fashion to the adjustment to PRESS for bandwidth selection, he suggests using

$$\text{PRESS}^*(\lambda) = \frac{\sum_{i=1}^n [Y_i - \hat{Y}_{i,-i}(\lambda)]^2}{n - \text{tr} [H(\lambda)]} \quad (\text{III.5.2})$$

A search routine must be implemented to search for the value of  $\lambda$  that minimizes (III.5.2). By convention, any value of  $\lambda > 1$  is set equal to 1 and any value of  $\lambda < 0$  is set equal to 0.

A method similar to that proposed in (III.5.2) is based upon a technique from linear regression called generalized cross validation (Myers, 1990). This method did not seem to perform better than (III.5.2) and was not considered for very long. The definition and summary of the results are presented in Appendix 3.

### III.5.B $C_P$ METHODS FOR SELECTING $\lambda$

Under the assumption of model misspecification, one would expect some bias in prediction in the model. Thus, one might like to have a method of choosing  $\lambda$  that incorporates both prediction variance and bias. A good choice of  $\lambda$  would have a reasonably small bias while also maintaining a fairly small prediction variance over all  $\underline{x}$  locations. This idea is incorporated in Mallows's  $C_P$  statistic (Mallows, 1973). The usual  $C_P$  statistic is defined as a estimate of

$$\frac{1}{\sigma^2} \sum_{i=1}^n \text{MSE } \hat{Y}(\underline{x}_i) = \frac{1}{\sigma^2} \sum_{i=1}^n \left[ \text{Var } \hat{Y}(\underline{x}_i) + \text{Bias } \hat{Y}(\underline{x}_i)^2 \right]. \quad (\text{III.5.3})$$

It can be shown (see for example Einsporn (1987) or Myers (1990)) that if least squares estimation is used then (III.5.3) reduces to

$$\frac{1}{\sigma^2} \sum_{i=1}^n \text{MSE } \hat{Y}(\underline{x}_i) = p + \frac{(n - p)}{\sigma^2} [s^2 - \sigma^2] \quad (\text{III.5.4})$$

where  $p$  is the number of parameters in the user's model and  $s^2$  is the usual MSE for this model. By estimating  $E(s^2)$  with  $s^2$  and using some appropriate estimate of  $\sigma^2$  ( $\hat{\sigma}^2$ ) the usual  $C_P$

statistic is defined by

$$C_P = p + \frac{(n - p)}{\hat{\sigma}^2} (s^2 - \hat{\sigma}^2). \quad (\text{III.5.5})$$

Obviously, increasing the number of parameters in the model increases the variance since the variance component of (III.5.3) can be written as

$$\begin{aligned} \frac{1}{\sigma^2} \sum_{i=1}^n \text{Var}(\hat{Y}(\underline{x}_i)) &= \frac{1}{\sigma^2} \sum_{i=1}^n \text{Var} \left( \sum_{j=1}^n h_{ij} Y_j \right) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \left( \sigma^2 \sum_{j=1}^n h_{ij}^2 \right) \\ &= \text{tr}[H'H] \\ &= p . \end{aligned} \quad (\text{III.5.6})$$

But increasing the number of parameters may lower the bias portion of (III.5.5). If the  $C_P$  for the higher order model is lower than that for the original model, then the increase in variance has been smaller than the reduction in bias. This indicates that the higher order model is better than the original model in terms of  $\text{MSE}(\hat{Y}(\underline{x}_i))$  over all  $\underline{x}$  locations.

By viewing nonparametric regression as fitting a model at least as complex as the assumed parametric model,  $C_P$  methods can be used to find the optimal  $\lambda$  in some sense. Recall that  $\lambda = 0$  means that the HATLINK model is just the assumed parametric model and  $\lambda = 1$  means that the HATLINK model is all nonparametric regression. Thus choosing a  $\lambda$  greater than zero is essentially adding to the complexity of the model. So increasing  $\lambda$  from 0 to 1 means that prediction variance will steadily increase while the bias may decrease at some point. Viewing  $C_P$  as a function of  $\lambda$ , the increase in variance versus decrease in bias can be measured



and the value of  $\lambda$  can be chosen that yields the minimum value of  $C_P(\lambda)$ . Since  $C_P$  is an estimate of the sum of mean squared errors at the data locations, this choice of  $\lambda$  will provide HATLINK with the lowest such value.

By estimating  $s^2$  and  $\hat{\sigma}^2$  in different ways, several versions of  $C_P$  can be used by HATLINK for estimating  $\lambda$ . One estimate of  $\hat{\sigma}^2$ , obtained by viewing kernel regression as the “full model” (i.e.  $\lambda=1$  is as complex as one is allowing) is

$$s_{ker}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i^{ker})^2}{\text{error df}} = \frac{\text{SSE(ker)}}{\text{error df}}. \quad (\text{III.5.7})$$

Three different ways of estimating error df have been suggested in the literature. Also  $s^2$  in (III.5.5) is replaced by  $s^2(\lambda)$ , an estimate of error variance based on the HATLINK fit.

Analogous to (III.5.5) and using  $(n - \text{tr}[H(\lambda)H'(\lambda)])$  to represent the error df for HATLINK,  $C_P(\lambda)$  may be calculated as

$$C_P(\lambda) = \text{tr}[H(\lambda)' H(\lambda)] + \frac{(s^2(\lambda) - s_{ker}^2)\text{tr}[(I - H(\lambda))'(I - H(\lambda))]}{s_{ker}^2}. \quad (\text{III.5.8})$$

Empirical studies indicate that the prediction performance of (III.5.8) is relatively poor for HATLINK. Four other versions of  $C_P$  that performed better in empirical studies have been studied by Einsporn (1987). All four versions of  $C_P$  were studied in this dissertation with various results. The details appear in Appendix 3. The first version is

$$C_{P1}(\lambda) = \text{tr}[H(\lambda)] + \frac{(s^2(\lambda) - s_{ker}^2)(n - \text{tr}[H(\lambda)])}{s_{ker}^2} \quad (\text{III.5.9})$$

$$\text{where } s_{ker}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i^{ker})^2}{n - \text{tr}[H_{ker}]} \quad \text{and } s^2(\lambda) = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i(\lambda))^2}{n - \text{tr}[H(\lambda)]} .$$

The second version is

$$C_{P2}(\lambda) = n - \text{tr}[(I - H(\lambda))'(I - H(\lambda))] + \frac{(s^2(\lambda) - s_{ker}^2)(\text{tr}[(I - H(\lambda))'(I - H(\lambda))])}{s_{ker}^2} \quad (\text{III.5.10})$$

where  $s_{ker}^2$  and  $s^2(\lambda)$  have denominators  $\text{tr}[(I - H_{ker})'(I - H_{ker})]$  and  $\text{tr}[(I - H(\lambda))'(I - H(\lambda))]$  respectively.

The third version is

$$C_{P3}(\lambda) = \text{tr}[H(\lambda)] + \frac{(s^2(\lambda) - s_{ols}^2)(n - \text{tr}[H(\lambda)])}{s_{ols}^2} \quad (\text{III.5.11})$$

where  $s_{ols}^2$  is the least squares estimate of variance for the user's model, and  $s^2(\lambda)$  is as in  $C_{P1}$ .

The fourth version of  $C_P$  studied is

$$C_{P4}(\lambda) = 2(\text{tr}[H(\lambda)] - p) + \frac{(\text{SSE}(\lambda) - \text{SSE}_{ols})(n - \text{tr}[H_{ker}])}{\text{SSE}_{ker}} \quad (\text{III.5.12})$$

All six methods of estimating  $\lambda$  performed differently, depending upon both the degree of model misspecification and the amount of variance. Over a broad range of model misspecification, the PRESS\* method appears to yield the most consistent estimates while  $C_{P3}$  yielded similar results under most conditions. This is further detailed in Appendix 3.

### III.6 EXAMPLE OF FITTED HATLINK SURFACE IN RSM

In order to illustrate the HATLINK method, consider the experiment proposed in Section II.2. For the same region of interest, a HATLINK response surface fit will be explored. It is still assumed that the user has prescribed a first order model,  $Y = \beta_0 + \sum_{i=1}^2 \beta_i x_i + \epsilon$ . The mixing parameter,  $\lambda = 0.9237$  was selected by  $\text{PRESS}^*(\lambda)$ . Since this value is close to one, some model misspecification has apparently occurred. In his work, Einsporn (1987) suggests an  $R^2$  statistic for HATLINK. Such a statistic is defined in the natural way as

$$R^2 = \frac{SSY - SSE(\lambda)}{SSY}, \quad (\text{III.6.1})$$

where  $SSY = \sum (Y_i - \bar{Y})^2$  and  $SSE(\lambda) = \sum (Y_i - \hat{Y}_i(\lambda))^2$ . For this example,  $R^2 = 0.978$ . Recall that  $R^2$  for the least squares model in chapter II was .845. Based upon  $R^2$ , the HATLINK fit seems to be an improvement over the least squares fitted model. A graph of the fitted HATLINK surface is included as Figure III.6.1, and a graph of the true surface is presented in Figure II.6.2. Note that the HATLINK surface seems to have some curvature, especially on the “edges” of the region. Compare this to the least squares fit as presented in Figure II.2.1. This further confirms the suspicion that the proposed first order model does not adequately represent the unknown response surface.

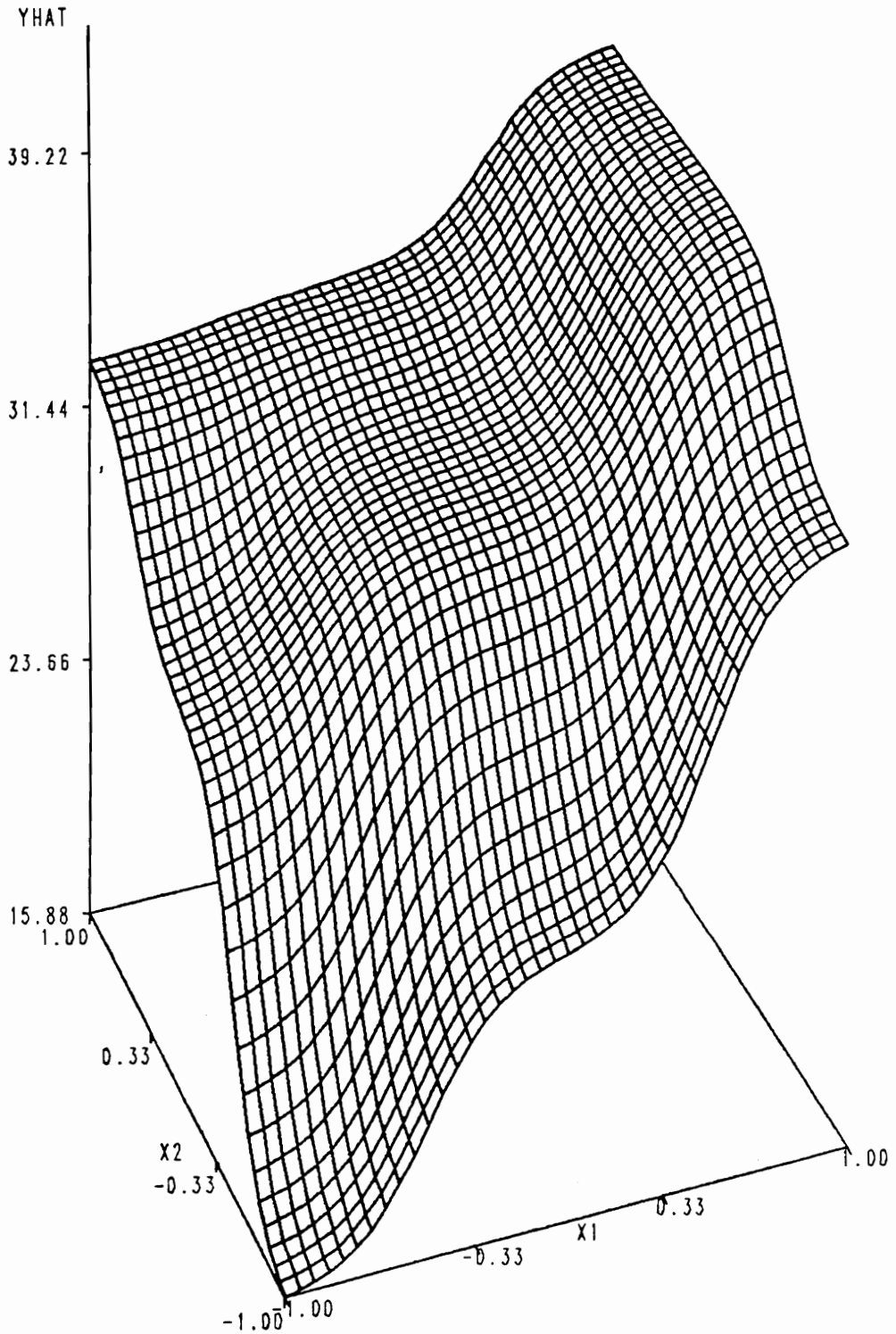


FIG. III.6.1. FITTED HATLINK SURFACE WITH FIRST ORDER LEAST SQUARES MODEL WITH  $CT=0.7$ ,  $H=.4992$  &  $LAM=.9375$

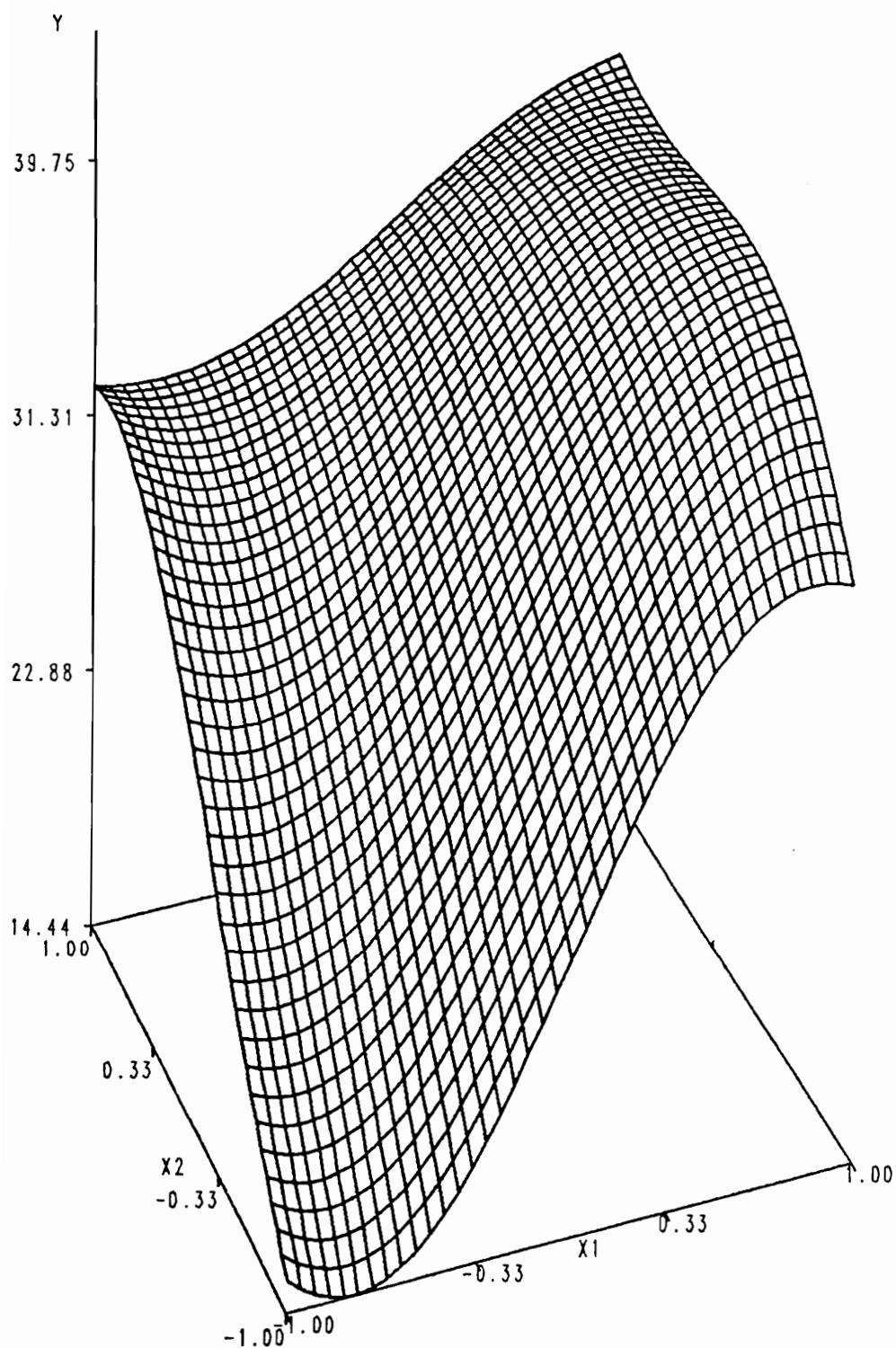


FIG. III.6.2. TRUE RESPONSE SURFACE IN OLD REGION  
WITH  $CT = 0.7$

# CHAPTER IV

## SEQUENTIAL DESIGN AUGMENTATION WITH HATLINK

### IV.1 DESCRIPTION OF THE AUGMENTATION PROCEDURE

As stated in the introduction, many times in response surface estimation the form of the true underlying function is unknown. Thus, a parametric model may be assumed that is not a good approximation to the true model. In this situation the user may unknowingly be introducing bias into the estimation process. The methods of design augmentation mentioned in Chapter II assume there is little or no misspecification in the model. But since HATLINK is more robust to model misspecification than simply using the parametric model (Einsporn, 1987), it could be used to augment points.

For the basic method of design augmentation studied in this dissertation, a region of interest needs to be determined. The candidate points for selection will come from this region. An augmentation criterion is evaluated for all the candidate points and a batch of points is selected for addition to the design matrix which best satisfies the desired criterion. After the batch of points is chosen, corresponding response values will be observed for all new selected points. This process will be referred to as sequential batch augmentation. The number of points in the batch, or batch size, may consist of one or more points. Only uniform batch size was studied here, variable batch size was not considered.

Recall that for HATLINK, once the bandwidth and mixing parameter are selected, predicted values and the corresponding prediction variances can be calculated at any point. Thus any of the augmentation methods mentioned in Chapter II that use prediction variance can be applied to HATLINK. In particular, points could be augmented by utilizing the integrated prediction variance criterion or conditional D-optimality criterion.

Sequential batch augmentation requires knowing where a new region of interest lies. Recall that for prediction at a point  $\underline{x}_i$ , kernel regression uses a measure of the distances between  $\underline{x}_i$  and the data locations. Usually the overlap between the new region of interest and the initial region of interest is fairly small. Consequently, prediction in the new region at locations far from the initial region would be based on data points relatively far from the location where prediction is desired. To help overcome this problem, one new data point is augmented somewhere in the new region of interest, which yields a new design matrix which will constitute the initial design matrix used by the proposed design augmentation procedure.

After augmenting by one extra point in the new region as described above, it was initially proposed to augment with further points individually by choosing the point,  $\underline{x}_0$ , to minimize the integrated prediction variance over the new region of interest, that is

$$\begin{aligned} \min_{\underline{x}_0 \in R_K} \int_{R_K} \text{Var}[\hat{Y}(\underline{x})] d\underline{x} &= \min_{\underline{x}_0 \in R_K} \sigma^2 \int_{R_K} \left\{ \sum_{j=1}^n [h_{0j}^{hat}]^2 \right\} d\underline{x} \\ &= \min_{\underline{x}_0 \in R_K} \sigma^2 \int_{R_K} \left\{ \sum_{j=1}^n [\lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols}]^2 \right\} d\underline{x} \end{aligned} \quad (IV.1.1)$$

where  $R_K$  is the new region of interest and  $h_{oj}^{hat}$  is defined as in (III.4.5). Note that all the original data points are contained in this hat matrix and any data points remote from the new region are less heavily weighted than those closer to the new region. Also, if the parametric model is a good approximation, then the choice of  $\lambda$  will tend to emphasize the least squares results through the parametric model over kernel regression through the nonparametric model. Conversely, if the model is not a good approximation, then  $\lambda$  will yield predicted values based more on the nonparametric model.

As stated in the introduction, a different criterion called BIIV, the bias-influenced integrated prediction variance criterion, seemed to perform better in simulation studies than (IV.1.1). It yielded results similar to those from IMSE while requiring less computational effort. This method also augments with points individually by choosing the point,  $\underline{x}_0$ , except now it is attempting to both minimize the integrated prediction variance criterion and evaluate where the fit is currently the worst, that is

$$\min_{\underline{x}_0 \in R_K} \left\{ - [\widehat{\text{bias}}(\underline{x}_0)]^2 + \int_{R_K} \text{Var}[\hat{Y}(\underline{x})] d\underline{x} \right\} . \quad (\text{IV.1.2})$$

The bias at the point  $\underline{x}_0$  must be estimated. It is proposed to estimate this quantity by

$$[\widehat{\text{bias}}(\underline{x}_0)]^2 = [\hat{Y}_{ker}(\underline{x}_0) - \hat{Y}_{hat}(\underline{x}_0)]^2 . \quad (\text{IV.1.3})$$

Equation (IV.1.3) was suggested because kernel regression is not a model based procedure, and as such should have less bias than least squares regression using an incorrectly specified model. The rationale for the negative sign in (IV.1.2) is that the bias measures where the fit is worst



and the quantity  $\left[ \widehat{\text{bias}}(\underline{x}_0) \right]^2$  estimates this measure. Recall that maximizing a number is the same as minimizing its negative. That procedure is performed in (IV.1.2). Note that if the negative in (IV.1.2) was not there, then the point with minimum bias would be found. This corresponds to finding the point where the fit is the best, obviously not a good procedure to follow.

Ideally, the same point would both minimize the integrated prediction variance and be the location where the fit is worst. In practice, this may not always happen, so the two quantities are combined in a straight linear fashion. One possible refinement might be some type of scaling parameter on at least one of the quantities.

The augmentation criterion in (IV.1.1) was suggested by the integrated prediction variance method of (II.3.1). The BIIV criterion in (IV.1.2) attempts to use IV.1.1, incorporating bias into the augmentation procedure. For conditional  $I_{IV}$ -optimality, the user's model is assumed to be a good approximation to the original model. With the HATLINK method, if this is not a valid assumption, the mixing parameter will be adjusted accordingly. Conditional  $I_{IV}$ -optimality as used by Rozum (1990) used only design points in the new region. Through nonparametric regression, the methods proposed in (IV.1.1) and in (IV.1.2) take advantage of all data points, regardless of location. Because the BIIV criterion seemed to perform better than integrated prediction variance in preliminary studies, unless specified otherwise, the BIIV criterion will always be used to augment points.

## IV.2 SEQUENTIAL BATCH AUGMENTATION

After choosing an additional design point to augment, the user would usually prefer to perform an experimental run at the new point and observe the corresponding response value. For

various reasons, sometimes performing only one experimental run is impractical. In such a situation, it may be desirable to wait until a batch of data points are obtained before performing experimental runs. Sequential batch design augmentation is based on such a situation.

For sequential batch design augmentation, it may be desired to add several batches of design points. After each batch of design points is augmented, the corresponding response values will be observed. Note that calculating the prediction variance (III.4.6) at a non-design point,  $\underline{x}_0$ , does not require a corresponding observed  $y$ -value. Recall that for any point,  $\underline{x}_0$ , calculating  $\hat{Y}(\underline{x}_0)$  using either kernel regression (III.1.4) or HATLINK (III.4.4) requires knowing the response values corresponding to all design locations. Because the PRESS\* and  $C_P$ -type methods in (II.2.2), (III.5.2), and in (III.5.9) through (III.5.12) all require calculating predicted response values at all design locations, finding the bandwidth and  $\lambda$  parameters require observing response values for all points. After the corresponding response values have been observed, the user may re-evaluate the bandwidth and  $\lambda$  parameters. This method allows HATLINK to take advantage of new data points to further explore the response surface and re-evaluate the appropriateness of the prescribed user's model.

The augmented points are individually selected based on minimizing either the integrated prediction variance as defined in (IV.1.1) or the BIIV criterion as defined in (IV.1.2). Duplications at any location are not permitted in order to encourage the design points to be spread around the region of interest as much as possible. As individual locations are selected for augmentation, they are saved as batch points. After selection of all the points for the batch is completed, the values are then added to the design matrix. The user next performs experimental runs for the batch of data locations augmented and observes the corresponding response values. After doing so, the bandwidth and  $\lambda$  parameters may be re-evaluated and updated.

If either the bandwidth or  $\lambda$  parameters change when updated, the predicted response

value at any point will change also. When the true model is known, an estimated squared error (ESE) type of criterion can be used for judging the fit of a particular model and for comparing two models. A grid of points is established over the new region of interest. If the true model is known, both the predicted value,  $\hat{Y}_i$ , and the true value,  $f(x_i)$ , for any point can be used to calculate

$$\text{ESE} = \sum_{\text{grid points}} [\hat{Y}_i - f(x_i)]^2. \quad (\text{IV.2.1})$$

This quantity may be calculated both for the original data and after a batch of points is augmented. If the ESE drops after augmenting with additional points, then the new design is considered to be better than the design before augmentation. Also, for two competing models, the model with the lower ESE is considered to be superior to the other model.

The methods of design augmentation mentioned in Chapter II always generate designs that do not depend upon the observed response values. That is, given an initial design matrix, the augmented design points would always be the same, regardless of what corresponding response values are observed. Sequential batch augmentation using HATLINK is a stochastic method. For this technique, the bandwidth and mixing parameters must be estimated and all of the proposed methods of estimating these parameters incorporate the observed response values. Consequently, given the same initial design matrix, the augmented design points can vary as the observed response values change. This is another main difference between many design augmentation methods and augmenting design points using HATLINK. When running simulations using the method proposed here, the augmented design points can change from repetition to repetition and this variability must be taken into consideration.

The choice of augmentation criterion, that is the selection of either BIIV or integrated

prediction variance, can affect the nature of the designs when the parametric method is used instead of HATLINK. When using only least squares as the parametric method, no bandwidth or mixing parameters must be estimated, thus any dependence upon the observed response values depends upon the augmentation criterion. If integrated prediction variance is considered as the augmentation criterion only the prediction variance must be calculated, and this is obviously independent of the observed response values. Thus, any new designs generated by this method of augmentation are non-stochastic in nature. That is, the choice of points does not depend upon the observed response values and will not vary when repeating runs for simulation studies. Conversely, the BIIV augmentation method yields stochastic designs, regardless of which method is used, HATLINK or least squares. In this sense, augmentation using BIIV is different from other more commonly used methods.

### **IV.3 EXAMPLE OF SEQUENTIAL DESIGN AUGMENTATION WITH BATCH SIZE OF ONE**

For the experiment in Section II.2, suppose after applying steepest ascent and further considerations, the new region of interest was determined to be the square with corners at (.5, 0), (2.5, 0), (.5, 2) and (2.5, 2). Further, because some curvature of the response surface is expected in this new region, a second order model will be proposed by the user. It is desired to augment with new points in this region using the BIIV criterion and the HATLINK method. Because of kernel regression, one experimental run at (2.5, 2) is performed and augmented to the original data set in Section II.2. This point was chosen because, based on distance, it is the farthest point from the old region of interest. The response value for the point (2.5, 2) is also obtained. Now points can be augmented using the BIIV criterion.

Suppose it is desired to augment by six points using sequential design augmentation with a batch size of one and the HATLINK method. In this case, response values will be immediately observed, allowing the bandwidth and  $\lambda$  values to be re-evaluated after augmenting with each point. The new data set used is the original data in Section II.2 plus the response at (2.5, 2). The response values are simulated observations from the model

$$Y = x_1^2 + x_1 - 3(x_2 - 1)^2 + 35 + \text{CT} \left\{ 8 \left[ \sin\left(\frac{\pi * x_1}{1.6}\right) \right] - 4 \left[ \sin\left(\frac{\pi * x_1 * x_2}{1.5}\right) \right] \right\} + \epsilon, \quad (\text{IV.3.1})$$

where  $\text{CT} = 0.7$  and  $\epsilon \sim N(0,1)$ . A plot of this surface over the new region of interest is presented in Figure IV.3.1. Of course, this model is unknown to the researcher. The ESE for both the HATLINK and parametric models can be calculated and compared. The results for the HATLINK method are given in Table IV.3.1.

These new points are illustrated in Figure IV.3.2. As can be seen, this method of augmentation tends to place some new points in the interior of the design region as well as placing some on the edges. As measured by the ESE, the additional points being augmented are steadily reducing the ESE until the fourth point augmented. After the fourth point, the ESE decreases and then increases somewhat. At the lowest ESE value, the ESE has dropped by more than a factor of ten of its original value after augmenting with the additional points. It is further noted that the value of  $\lambda$  is 1 after the first augmented point but then eventually decreases to 0.914 and lower, indicating that, by the fourth point, the HATLINK fit was based on both the parametric and nonparametric portions of HATLINK.

The results for using only a second order model and least squares estimation are very

different and are given in Table IV.3.2. The locations of these six points are illustrated in Figure IV.3.3. Notice that the augmented points tend to be placed on the edges of the region, especially initially. Based on the ESE values, the HATLINK fit would be considered to be better than the fit based solely on the parametric model.

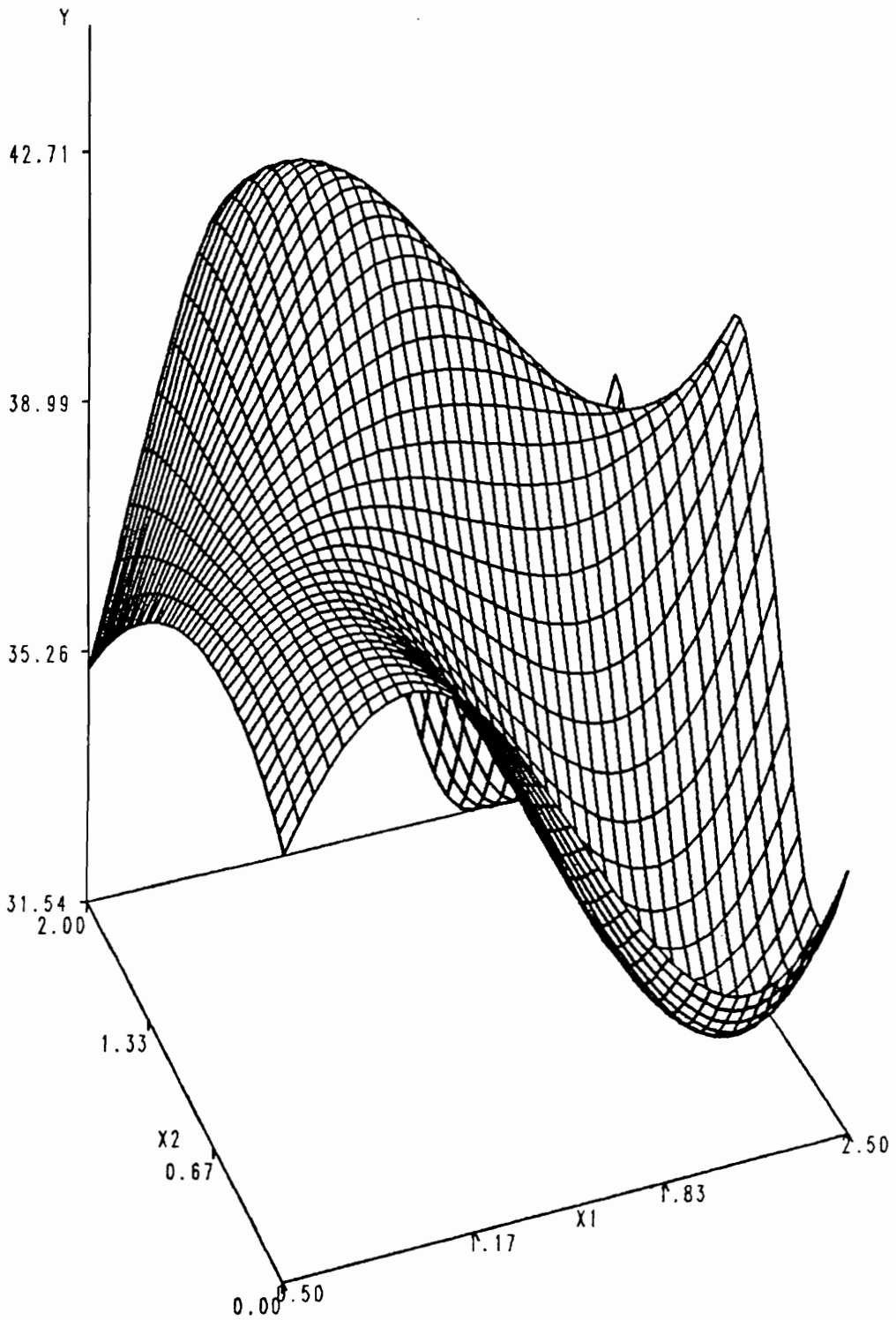


FIG. IV.3.1. TRUE RESPONSE SURFACE IN NEW REGION  
WITH  $CT = 0.7$

**Table IV.3.1. Results for Six Points Augmented Using IIATLINK and BIIV.** Using  $CT = 0.7$  and  $\sigma = 1$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAMBDA</u>	<u>ESE</u>
org. design	—	—	.507	0.00	1615.71
1	2.5000	0.0000	.525	1.00	118.39
2	2.5000	2.0000	.517	1.00	120.38
3	1.0000	1.0000	.519	1.00	110.29
4	2.5000	0.5000	.515	.937	139.43
5	2.5000	1.0000	.583	.625	127.55
6	0.5000	2.0000	.591	.875	129.97

**Table IV.3.2. Results for Six Points Augmented Using Least Squares and BIIV.** Using  $CT = 0.7$  and  $\sigma = 1$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>
org. design	—	—	1615.71
1	2.5000	0.0000	188.73
2	2.5000	0.5000	175.52
3	0.5000	2.0000	183.86
4	2.0000	0.5000	191.56
5	2.5000	2.0000	159.08
6	2.0000	1.0000	157.84



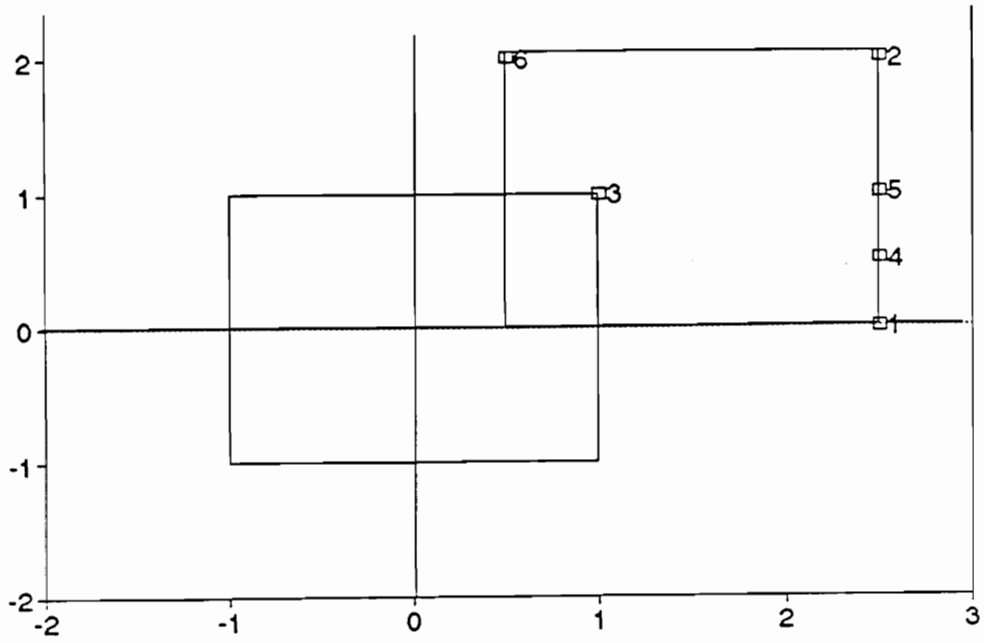


Fig. IV.3.2. Plot of the Six Augmented Points Using HATLINK & BIIV with  $CT=0.7$  &  $\text{Sigma}=1$ , Batch Size of One

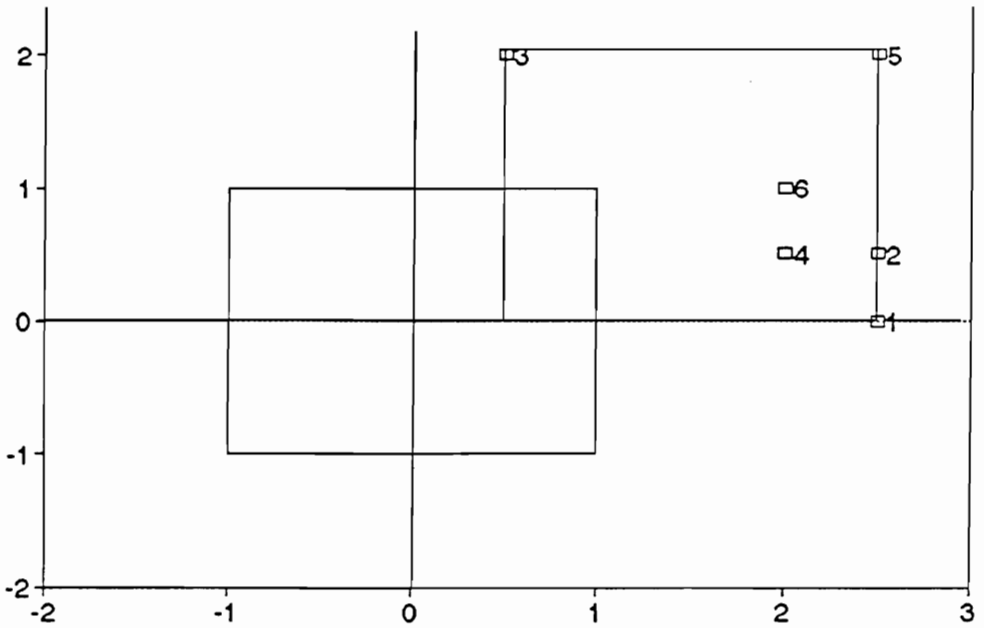


Fig. IV.3.3. Plot of the Six Augmented Points Using L.S. & BIIV with  $CT=0.7$  &  $\text{Sigma}=1$ , Batch Size of One

#### IV.4 HATLINK AUGMENTATION WITH BIIV VS. IPV

Recall that based upon preliminary investigations, the BIIV augmentation criterion method was considered superior to integrated prediction variance and examples were presented to illustrate the augmentation procedure. So far, no examples have been included that directly compare the behavior of augmentation using HATLINK and BIIV as compared to augmentation using HATLINK and integrated prediction variance. In this section, such a comparison will be made. Recall that the example in Section IV.3 augmented with six points using BIIV and a batch size of one. Using the same initial conditions, six points will be augmented using HATLINK, integrated prediction variance and a batch size of one. The fitted surfaces, summarized by their ESE values, will be compared for the two methods. The data set used is the original data in Section II.2 plus the response at (2.5, 2), and the true model is assumed to be the same as that in (IV.3.1) with  $CT = 1.0$  and  $\sigma = 0.2$ .

The results for HATLINK using BIIV are presented in Table IV.4.1. The results for HATLINK using integrated prediction variance are given in Table IV.4.2.

The new points from BIIV are illustrated in Figure IV.4.1, while the new points from integrated prediction variance are illustrated in Figure IV.4.2. Note that the two methods performed almost identically until the second point is augmented. With the augmentation of the second point, both the bandwidth and the lambda parameters started to vary somewhat between methods. Notice also that the ESE value for BIIV is significantly lower than the corresponding value for integrated prediction variance. This difference in ESE values persists with the augmentation of additional points. Notice the large difference in ESE values with each point augmented. These differences imply that the fit from the design generated by BIIV is superior to that for integrated prediction variance. This is obviously because BIIV was designed to

**Table IV.4.1. Results for Six Points Augmented Using HATLINK and BIIV.** Using  $CT = 1.0$  and  $\sigma = 0.2$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAMBDA</u>	<u>ESE</u>
org. design	—	—	0.410	1.000	484.30
1	2.5000	0.0000	0.363	0.750	785.67
2	2.5000	2.0000	0.143	0.992	281.50
3	1.0000	0.5000	0.139	1.000	279.94
4	2.0000	1.0000	0.137	1.000	250.26
5	1.0000	0.0000	0.132	1.000	245.96
6	1.5000	0.0000	0.429	1.000	221.85

**Table IV.4.2. Results for Six Points Augmented Using HATLINK and Integrated Prediction Variance.** Using  $CT = 1.0$  and  $\sigma = 0.2$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAMBDA</u>	<u>ESE</u>
org. design	—	—	0.410	1.000	484.30
1	2.5000	2.0000	0.363	0.750	785.67
2	1.0000	1.0000	0.364	0.750	780.37
3	1.0000	0.0000	0.372	0.750	789.92
4	1.0000	0.5000	0.384	0.750	768.10
5	0.5000	1.0000	0.291	1.000	457.34
6	0.5000	0.0000	0.348	0.750	748.44

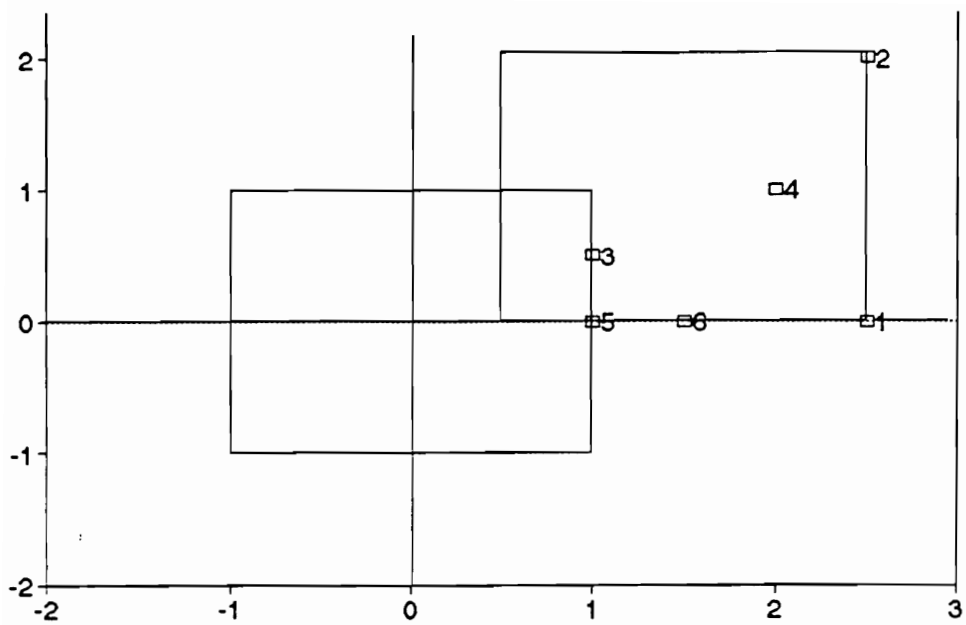


Fig. IV.4.1. Plot of the Six Augmented Points Using HATLINK & BIIV with  $CT=1$  &  $\text{Sigma}=0.2$

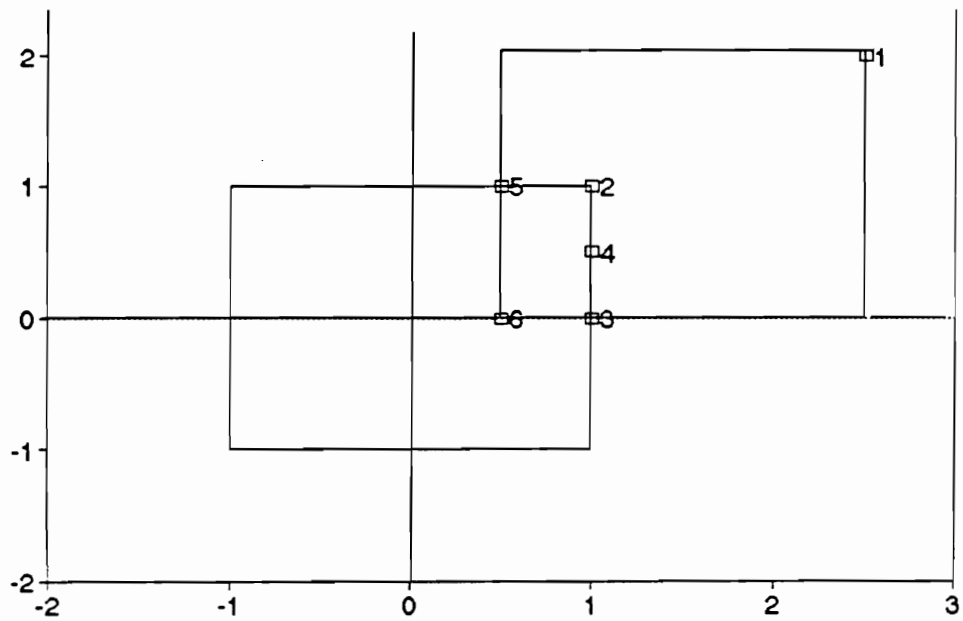


Fig. IV.4.2. Plot of the Six Augmented Points Using HATLINK & IPV with  $CT=1$  &  $\text{Sigma}=0.2$

incorporate bias into the augmentation procedure. With larger degrees of CT, integrated prediction variance can not account for the model misspecification present. Preliminary results indicated that this behavior is consistent, hence BIIV is the recommended method of augmentation.

#### IV.5 LEAST SQUARES AUGMENTATION WITH BIIV VS. IPV

This section deals with the behavior of least squares regression using BIIV as compared to integrated prediction variance. In this case, the extent of model misspecification suspected by the user could possibly affect which augmentation criterion is selected, that is, whether BIIV or integrated prediction variance is chosen. If it is suspected that there is minimal model misspecification, then little or no bias in the procedure would be expected. Thus, a BIIV type of criterion that attempts to incorporate bias might not be appropriate. In such a case, the user might chose to augment by integrated prediction variance. In practice, however, there is often more model misspecification than originally suspected. In such cases, it would be hoped that augmentation with BIIV would yield a better fit in terms of the ESE values then augmentation using integrated prediction variance. By using BIIV, an experimenter could protect against bias, with minimal extra computations, over other methods. This section illustrates the advantage of using the BIIV criterion.

Suppose the user has the same conditions as in Section IV.4. That is, the same old and new regions of interest, the same initial point in the new region and the same true model as in (IV.3.1) with  $CT = 1.0$  and  $\sigma = 0.2$ . Here, the user will augment with six points with batch size of one. In this case the results for a second order least squares model using BIIV are given in Table IV.5.1. In the second case, the results for a second order least squares model using

integrated prediction variance are presented in Table IV.5.2.

These points are illustrated in Figures IV.5.1 and IV.5.2. Notice that after the first augmented point, the ESE values are higher for the integrated prediction variance criterion, although sometimes not too much higher. In this case, some model misspecification has occurred, and the BIIV method is consistently performing better, as measured by the ESE fits, after the first point augmented. It does indeed appear that for the least squares method, augmentation by BIIV performs better than merely using integrated prediction variance. This further strengthens the claim that BIIV is superior in general to integrated prediction variance when the model has been misspecified.

**Table IV.5.1. Results for Six Points Augmented Using Least Squares and BIIV.** Using  $CT = 1.0$  and  $\sigma = 0.2$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>
org. design	—	—	2631.60
1	2.5000	0.5000	424.55
2	2.0000	0.5000	388.47
3	1.5000	0.5000	363.89
4	0.5000	2.0000	321.57
5	2.0000	0.0000	316.37
6	2.5000	2.0000	320.52

**Table IV.5.2 Results for Six Points Augmented Using Least Squares and Integrated Prediction Variance.** Using  $CT = 1.0$  and  $\sigma = 0.2$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>
org. design	—	—	2631.60
1	2.5000	0.0000	417.01
2	0.5000	2.0000	423.34
3	2.5000	1.5000	369.42
4	2.5000	0.5000	325.30
5	2.0000	2.0000	366.69
6	1.5000	2.0000	357.59

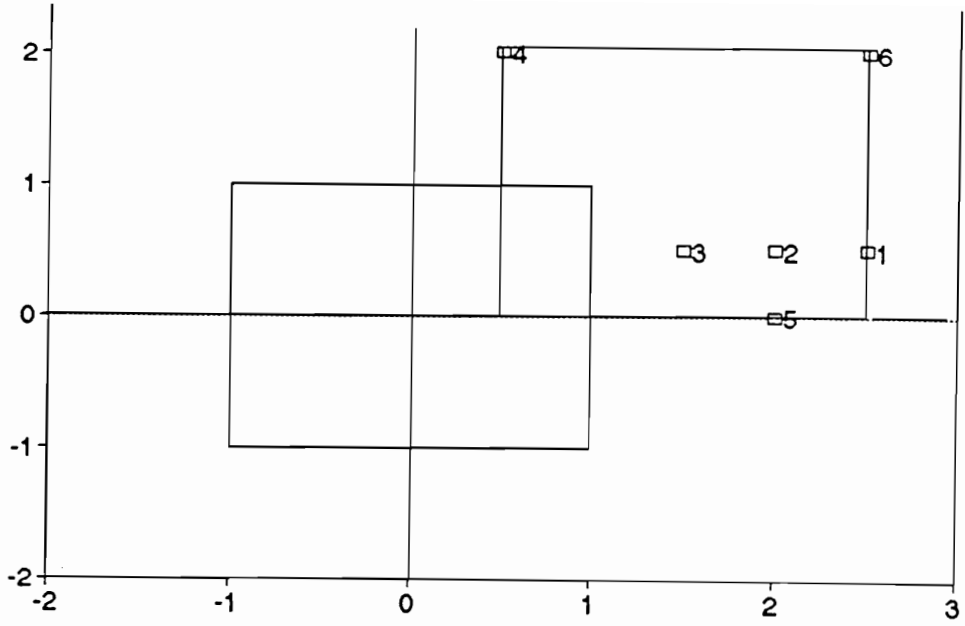


Fig. IV.5.1. Plot of the Six Augmented Points Using L.S. & BIIV with  $CT=1$  &  $\Sigma=0.2$

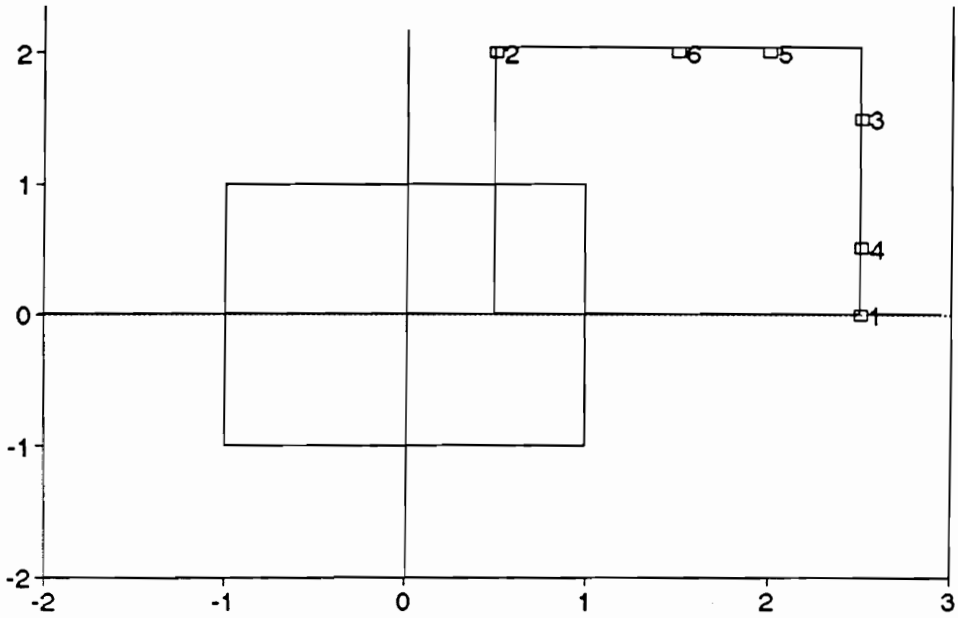


Fig. IV.5.2. Plot of the Six Augmented Points Using L.S. & IPV with  $CT=1$  &  $\Sigma=0.2$



## IV.6 EXAMPLE OF SEQUENTIAL DESIGN AUGMENTATION WITH BATCH SIZE OF TWO

In the previous example, response values were observed with each point augmented, thus allowing the user to re-evaluate the bandwidth and  $\lambda$  parameters each time. This section contains an example of the augmentation procedure with a batch size of two. In such a case, two points would be augmented, but corresponding response values will not be observed until the selection of the second point. This means that the bandwidth and  $\lambda$  parameters may only be re-evaluated with every other point augmented. Potentially, the HATLINK fit could be worse for the larger batch size, since the bandwidth and  $\lambda$  parameters will not be updated as often. Eventually, as points are augmented, it is considered likely that updating the parameters will not be as large a factor in evaluating the fit. Simulation studies have indicated that usually, regardless of the method of augmentation, the smaller the batch size, the quicker the fit improves.

Suppose the user has the same initial conditions as those in Section IV.3. That is, the user considers the same initial and new regions of interest, the same initial point in the new region, and the same true model as in (IV.3.1). Here, the user will augment by six points in batches of two points each. In this case the results for HATLINK are given in Table IV.6.1. These new points are illustrated in Figure IV.6.1.

As can be seen when compared to Figure IV.3.1, this method of augmentation tends to also place some new points in the interior of the new region as well as along the edges of the new region. Furthermore, the two examples tended to select the same points, although the order of selection was different. As measured by the ESE, the additional points being augmented are steadily reducing the ESE, although the ESE values for the batch size of two tended to be larger

than the corresponding ESE values for the batch size of one. As expected, after six points are augmented, the two methods have approximately the same bandwidth,  $\lambda$  and ESE values. It appears that the loss of information due to a larger batch size can be overcome to some extent by augmenting enough batches. This issue will be addressed in more detail in Chapter V.

Table IV.6.2 presents the results for using only a second order model and least squares estimation, which are very different from the results using a batch size of one. The augmented points are illustrated in Figure IV.6.2. As can be seen, augmentation with only the parametric model through least squares tended to place points mostly around the edges of the region. Although the ESE for both the parametric and the HATLINK methods dropped some with the first few points, using least squares for the parametric model did not perform as well as the HATLINK technique. In fact, the ESE for the least squares model even increased after the first two batches were augmented. Again, for both batch sizes studied in these two examples, after the six points were augmented, the ESE values were approximately the same.

For the examples in Section IV.3 and Section IV.6, the ESE values for the HATLINK method were always lower than the ESE values for least squares, indicating that the HATLINK technique is giving a better augmentation design than the parametric model. Augmenting points and updating the bandwidth and lambda parameters led HATLINK to a larger  $\lambda$  value after the first batch was augmented. This implies that a second order least squares model is somewhat misspecified for this example. The HATLINK method would be expected to out-perform least squares through the parametric model in such a case. Based upon the ESE values for both the parametric model and the HATLINK method, this has been observed.

**Table IV.6.1. Results for Six Points Augmented Using IIATLINK and BIIV and a Batch Size of Two.** This lists the points augmented, the bandwidth and lambda values selected and the ESE values after the two points are augmented. The bandwidths, lambdas and ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAMBDA</u>	<u>ESE</u>
org. design	—	—	.507	0.00	1615.71
1	2.5000	0.0000	—	—	—
2	2.5000	0.5000	.514	1.00	150.23
3	2.5000	2.0000	—	—	—
4	1.0000	1.0000	.515	.937	139.43
5	2.5000	1.0000	—	—	—
6	1.0000	0.5000	.592	.625	127.18

**Table IV.6.2. Results for Six Points Augmented Using Least Squares and BIIV and a Batch Size of Two.** This lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented.

<u>AUG. NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>
org. design	—	—	1615.71
1	2.5000	0.0000	—
2	2.5000	0.5000	175.52
3	0.5000	2.0000	—
4	2.0000	0.5000	191.56
5	2.5000	2.0000	—
6	1.0000	0.0000	154.87

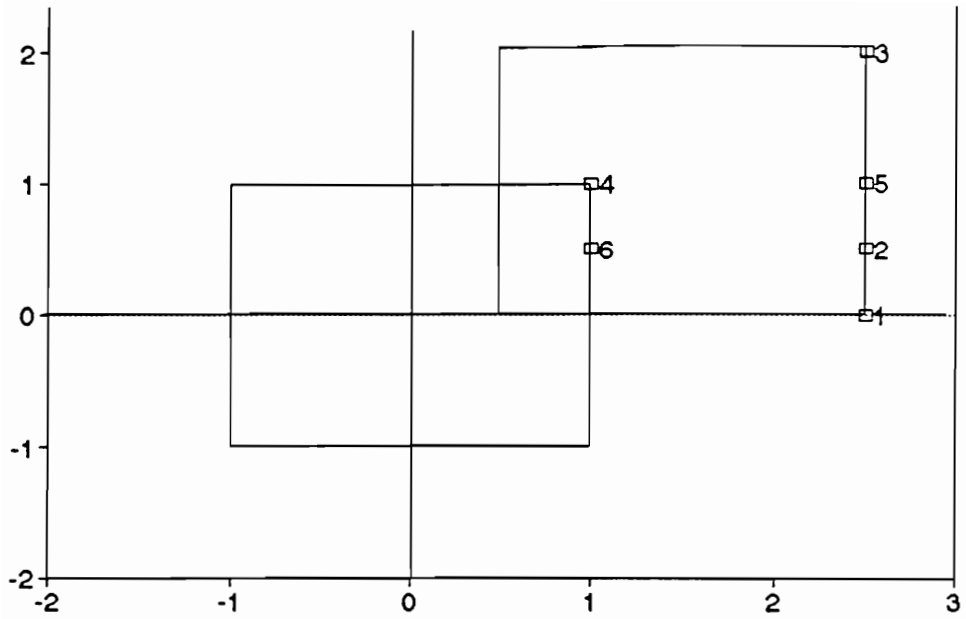


Fig. IV.6.1. Plot of the Six Augmented Points Using HATLINK & BIIV with  $CT=0.7$  &  $\Sigma=1$ , Batch Size of Two

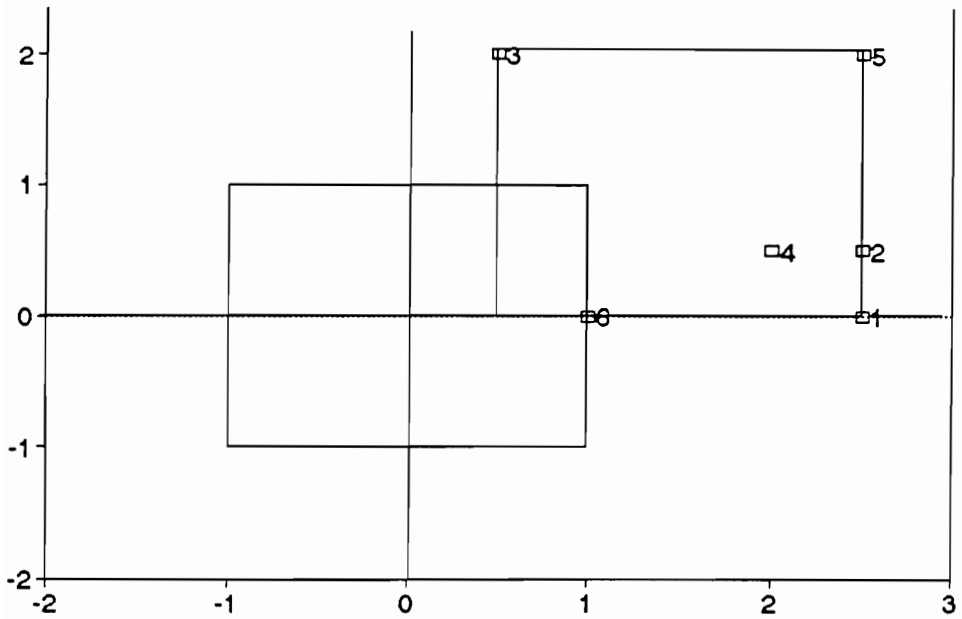


Fig. IV.6.2. Plot of the Six Augmented Points Using L.S. & BIIV with  $CT=0.7$  &  $\Sigma=1$ , Batch Size of Two

#### IV.7 BIAS IN THE MODEL

The HATLINK model has been proposed as one that is robust to model misspecification. If such exists, it is only natural to consider the bias due to the misspecification in the predicted value. Suppose the experimenter has fit the linear model

$$\underline{Y} = X_1 \underline{\beta}_1 + \epsilon$$

where  $X_1$  is the  $(n \times p_1)$  matrix of independent variables associated with the presumed model, and  $\underline{\beta}_1$  is the corresponding  $(p_1 \times 1)$  vector of unknown coefficients. Suppose  $p$  is the number of parameters in the correct model that should have been used, so  $p = p_1 + p_2$ . The true model that should have been used is written as

$$\underline{Y} = X_1 \underline{\beta}_1 + X_2 \underline{\beta}_2 + \epsilon^*$$

where  $\underline{\beta}_2$  is the  $(p_2 \times 1)$  vector of parameters not present in the presumed model and  $X_2$  is the  $(n \times p_2)$  matrix associated with these parameters. Let  $\underline{x}_0 = (\underline{x}_{1,0}, \underline{x}_{2,0})'$  be an arbitrary point where  $\underline{x}_{1,0}$  is the  $(p_1 \times 1)$  portion of  $\underline{x}_0$  associated with the parameters in the presumed model and  $\underline{x}_{2,0}$  is the  $(p_2 \times 1)$  portion of  $\underline{x}_0$  associated with the parameters not present. Recall that for predictions for HATLINK, the predicted value at  $\underline{x}_0$  is

$$\begin{aligned} \hat{Y}(\underline{x}_0)^{hat} &= \hat{Y}_0^{hat} = \sum_{j=1}^n h_{0j}^{hat} Y_j \\ &= \sum_{j=1}^n \left[ \lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols} \right] Y_j \end{aligned}$$

Under the presumed model, the expected value of the predicted value at the arbitrary point  $\underline{x}_0$  is given by

$$\begin{aligned}
 E[\hat{Y}_0^{hat}] &= E\left[\sum_{j=1}^n h_{0j}^{hat} Y_j\right] \\
 &= E\left[\sum_{j=1}^n \left(\lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols}\right) Y_j\right] \\
 &= \sum_{j=1}^n \left[\lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols}\right] E[Y_j] \\
 &= \sum_{j=1}^n \left[\lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols}\right] [\underline{x}_{1,j}' \underline{\beta}_1 + \underline{x}_{2,j}' \underline{\beta}_2].
 \end{aligned}$$

Now,  $E(Y_0) = \underline{x}_{1,0}' \underline{\beta}_1 + \underline{x}_{2,0}' \underline{\beta}_2$ . Then the bias can be written as

$$\begin{aligned}
 \text{bias}[\hat{Y}^{hat}(\underline{x}_0)] &= E(\hat{Y}_0^{hat}) - E(Y_0) \\
 &= \left\{ \sum_{j=1}^n \left[\lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols}\right] (\underline{x}_{1,j}' \underline{\beta}_1 + \underline{x}_{2,j}' \underline{\beta}_2) \right\} - \underline{x}_{1,0}' \underline{\beta}_1 + \underline{x}_{2,0}' \underline{\beta}_2. \quad (\text{IV.7.3})
 \end{aligned}$$

An alternate way of writing this in terms of an alias matrix is first by writing

$$\hat{Y}_0^{hat} = \sum_{j=1}^n \left(\lambda h_{0j}^{ker} Y_j\right) + (1 - \lambda) \underline{x}_{1,0}' (X_1' X_1)^{-1} X_1' \underline{Y}.$$

Then letting the alias matrix  $A = (X_1'X_1)^{-1}X_1'X_2$  the bias becomes

$$\text{bias } (\hat{Y}_0^{hat}) = \sum_{j=1}^n \left\{ \lambda h_{0j}^{ker} [\underline{x}_{1,j}' \underline{\beta}_1 + \underline{x}_{2,j}' \underline{\beta}_2] \right\} +$$

$$[\underline{x}_{1,0}' A - \underline{x}_{2,0}'] \underline{\beta}_2 - \lambda \underline{x}_{1,0}' [\underline{\beta}_1 + A \underline{\beta}_2]. \quad (\text{IV.7.1})$$

Note that if  $\lambda = 0$ , i.e. only the least squares results through the parametric model is being used, then (IV.7.1) reduces to the customary value from regression (Myers, 1990).

The variance for  $\hat{Y}(\underline{x}_0)$  is given by

$$\text{Var} [\hat{Y}(\underline{x}_0)] = \sigma^2 \sum_{j=1}^n \left[ \lambda h_{0j}^{ker} + (1 - \lambda) h_{0j}^{ols} \right]^2 \quad (\text{IV.7.2})$$

Now the mean squared error of the prediction at  $\underline{x}_0$ ,  $\text{MSE}(\hat{Y}_0)$  may be calculated from knowing that  $\text{MSE}(\hat{Y}_0) = \text{Var}(\hat{Y}_0) + \text{bias}(\hat{Y}_0)^2$  and using (IV.7.1) and (IV.7.2).

Integrating (IV.7.2) over the new region of interest would yield an integrated mean square error (IMSE). It would be most desirable to augment by the design point which minimizes the IMSE. In such a case, the bias given in (IV.7.1) cannot be calculated without knowing the degree and type of model misspecification. Thus the bias must be estimated. This was the motivation for the BIIV method of augmentation, which was suggested as a compromise between an IMSE approach and an integrated prediction variance approach. In future research, better methods of estimating the bias in (IV.7.1) could yield improvements in the HATLINK method of augmentation using BIIV.

## IV.8 NUMERICAL CONSIDERATIONS

As stated before, several problems arise in practice when applying either (IV.1.1) or (IV.1.2). First, since no closed form expression for the kernel function exists, there is no easy way to pick the point that minimizes either BIIV or the integrated prediction variance. The only way to achieve this in practice is to add individual points in the new region to the design matrix and calculate the integrated prediction variance, along with the estimated bias if BIIV is being performed. Then the location is augmented that yields either the minimum BIIV or minimum integrated prediction variance value. Obviously not all possible points in the new region can be checked, so a finite collection of candidate locations must be selected. Without any other information, a simple grid of points was selected for this stage. Thus each time an additional value was augmented, it came from the same collection of candidate points. The number of candidate locations is limited by the computer time available. For most applications considered here a five by five grid was utilized.

Another problem is evaluating the integral in (IV.1.1) and (IV.1.2). Again, without a closed form expression for the kernel function, the integral cannot be evaluated and it must be approximated by numerical methods. The most straightforward method for evaluating the integral contained in (IV.1.1) and (IV.1.2) is a Simpson's Rule approach mentioned in Stroud (1971) for higher order dimensions. This consists of evenly gridding the region of integration as required by Simpson's Rule, which is then applied in each dimension. For some dimension values, Stroud (1971) has formulas already worked out. This method may require the function to be evaluated at many points to attain a relatively high degree of accuracy. If the function is hard or difficult to evaluate, it may become very time consuming to apply a Simpson's Rule approach.



A more efficient method of numerically evaluating integrals incorporates Gaussian quadrature. This involves approximating the multiple integral by a weighted sum of points evaluated within the integrand. Selected formulas for Gaussian quadrature are presented in Stroud (1971) for standard regions, say a circle centered at zero with radius of one or a cube centered at zero with sides of length one. Any spherical or cuboidal region can be translated and/or transformed into a standard region. Thus the formulas given in Stroud (1971) may be applied to these regions with the appropriate changes. Some of these formulas may be used to approximate (IV.1.1) and (IV.1.2) with fewer points but the same accuracy as Simpson's Rule. These methods for approximating integrals require the integrand to be relatively smooth. A Gaussian quadrature equation from Stroud (1971) was selected to approximate the integrals. Using Gaussian quadrature resulted in approximations that were quicker to evaluate while retaining the same accuracy as a Simpson's Rule approach using a very fine grid. As can be seen in this section, numerical evaluation for the integrals in (IV.1.1) and (IV.1.2) was not a straightforward procedure. Other techniques for approximating multiple dimension integrals might be studied.

#### IV.9 ANALYTICAL CONSIDERATIONS

As stated above, many times the integrals that result from calculating the integrated prediction variance can never be evaluated exactly and must be approximated. An obvious question is how well do the approximation formulas work when performing sequential design augmentation using HATLINK. Recall that the kernel used throughout this work was the "normal" kernel function, that is

$$K(u) = \exp(-u^2) \quad -\infty < u < \infty \quad (\text{IV.9.1})$$

where  $u = (x - x_j)/h$  for the single variable case. Unfortunately, this leads to integrals in (IV.3.1) and (IV.3.2) that cannot be evaluated in closed form solution. For example, finding the point  $\underline{x}_0$  to minimize the integrated prediction variance must be done using approximation methods utilizing the numerical techniques discussed in the previous section. Instead of the kernel function in (IV.9.1), consider the function

$$K(u) = \begin{cases} 1 - \frac{1}{9} u^2 & |u| \leq 3 \\ 0 & \text{otherwise} \end{cases} \quad (\text{IV.9.2})$$

This kernel function is more tractable and leads, in simple cases, to minimizing integrated prediction variances that may be evaluated in a closed form solution.

As an example, assume the univariate case with  $[-1, 1]$  as the old region of interest for the design variable  $x$ . Some subsequent analysis reflects a need to move to the region  $[0, 2]$ , which is the new region of interest. Starting with a design as in Chapter IV, a new point will be augmented to minimize the integrated prediction variance criterion of IV.1.1. Using (IV.9.2) as the kernel function, the integral will be evaluated exactly and the point that minimizes the integrated prediction variance will be chosen.

For the design, use one point each at  $x = -1, 0,$  and  $1$  in the old region. To start the augmentation procedure, augment by the point  $x = 2$  in the new region. For notation, let  $x_1 = -1, x_2 = 0, x_3 = 1, x_4 = 2$ . It is desired to augment with the point  $x_0 \in [0, 2]$  to minimize the integrated prediction variance as stated in (IV.1.1). The details of the calculations for this particular situation are presented in Appendix 1.

Responses corresponding to each of the original four points were simulated, and the bandwidth and lambda values were computed using the PRESS\* criterion for the bandwidth and the  $C_P^2$  criterion for the lambda. The results were  $h = 0.86066$  and  $\lambda = 0.43158$ . These

values were used in evaluating (IV.1.1).

Formula Manipulation Compiler (FORMAC) is a language for symbolic manipulation of mathematical expressions. It was used to perform the sums and the algebra for the individual integrals in (IV.1.1). FORMAC was also used to evaluate the integrals at the limits of integration and simplify the answers somewhat. The algebra involved would be too tedious to perform correctly by hand! See Appendix 1 for more details. After considerable algebra, the results are given below.

The point in the new region that minimized (IV.1.1) was  $x_0 = 2$ . This produced

$$\int_0^2 \frac{\text{var} [\hat{y}(x)]}{\sigma^2} dx = 0.48482602 ,$$

which is the exact value for the integral. Using the same kernel function (IV.9.2) and HATLINK, the integrals were approximated and the minimizing point was selected by the computer program performing sequential design augmentation. The point selected was also  $x_0 = 2$ . The integrated prediction variance was calculated by Gaussian quadrature to be

$$\int_0^2 \frac{\text{var} [\hat{y}(x)]}{\sigma^2} dx = 0.4848431 ,$$

a value agreeing very closely with the exact value for the integrated prediction variance. Such results imply that the integral approximations performed on the computer are reasonably accurate, at least for the kernel function (IV.9.2). Lastly, the normal kernel function as given in (IV.9.1) was used to select the point to minimize integrated prediction variance, again implementing the computer program for sequential design augmentation with the HATLINK method. As for the other two cases, the point selected was  $x_0 = 2$ . In this case, the integrated

prediction variance was calculated numerically as

$$\int_0^2 \frac{\text{var}[\hat{y}(\underline{x})]}{\sigma^2} dx = 0.50980 .$$

This value also agrees fairly closely with the other two values, the slight difference being due to the different kernel function, supporting Butler's statement that the form of the kernel function is not too important. It also indicates that the methods presented to approximate integrals and find the point that minimizes integrated prediction variance appear to work in practice. It was through examples such as this that indicated that the methods of numerically approximating (IV.1.1) and (IV.1.2) employed in this dissertation are justified.

#### IV.10 CONCLUSIONS

This chapter presented the HATLINK method of design augmentation. Two different criteria for augmentation were discussed and examples were included. The results indicated that the HATLINK method appears to perform well, at least under some circumstances. Obviously further simulations must be studied in order to justify using the HATLINK method of design augmentation. In Chapter V, additional examples with several quantities varied, including the degree and type of model misspecification and the variance will be considered and presented. To demonstrate that augmentation with BIIV yielded a better design, Sections 4 and 5 compared BIIV and integrated prediction variance for HATLINK and for least squares.

Section IV.7 considered the bias in the model. When the user has misspecified the parametric model, some bias will occur when fitting a surface. This motivated the concept of the BIIV method of augmentation. Potential improvements in the BIIV criterion are suggested in

## Chapter VII.

Some numerical problems that arise when calculating the integrals in (IV.1.1) and (IV.1.2) were briefly considered. The nature of the HATLINK method means that analytic results are extremely difficult to obtain, except under very simplistic conditions. In Section IV.9, one such example was presented. This example augmented with one point using integrated prediction variance. The integral was both evaluated exactly and approximated by two methods, and it was shown that the final results were very similar. It was concluded that the approximations necessary to perform the augmentation procedure are reasonable.

Chapter VI has further examples of the HATLINK method of augmentation. It considers both two two-dimensional examples and one three-dimensional examples of the HATLINK method using the BIIV method of augmentation.

# CHAPTER V

## RESULTS OF INITIAL SIMULATIONS

### V.1 DESCRIPTION OF STUDY

Recall that Chapter II reviewed response surface methodology, particularly augmenting with design points in a new region of interest. Chapter III presented kernel regression as an example of a nonparametric regression technique. By use of a mixing parameter, Chapter III also developed a new method called HATLINK which was suggested as a compromise between parametric and nonparametric regressions. Chapter IV gave more detail on the HATLINK method and presented two examples showing the effect of the batch size upon the proposed augmentation method. Using least squares for the parametric model and kernel regression for the nonparametric model, it has been proposed to use HATLINK to perform design augmentation. This should result in a new design that yields a better fit and, consequentially, better prediction properties than currently used methods that are based upon least squares regression.

In order to study the behavior of the proposed method, some simulated examples of augmenting designs using various augmentation criteria are presented. Normally, such criteria are based upon least squares techniques as indicated by (II.3.1). Possible augmentation criteria suggested were either minimizing integrated prediction variance (IV.1.1) or BIIV (IV.1.2). In comparing the HATLINK, least squares, and kernel methods, it will be shown that for both

(IV.1.1) and (IV.1.2) the HATLINK method yielded a better fit as measured by the ESE criterion of (IV.2.1) over a broad range of conditions.

In order to define a particular response surface setting for simulation, the following items must be specified:

- (i) The initial design points.
- (ii) The new region of interest.
- (iii) The total number of design points to be augmented.
- (iv) The number of batches.
- (v) The distribution of the random error variable.
- (vi) The user's parametric model.
- (vii) The functional form of the true underlying model.
- (viii) The degree of departure between the true & user's models.
- (ix) The selection of the bandwidth parameter.
- (x) The selection of the lambda parameter.
- (xi) The choice of augmentation criterion.

Obviously, there are a infinite number of ways to vary these eleven items. In this work, restrictions are placed on some of the items. For example, the random error is always assumed to be distributed as  $N(0, \sigma^2)$  where  $\sigma^2$  must be specified by the experimenter. To allow the investigation to proceed in an orderly manner, all items are held constant except for the choice of  $\sigma^2$ , (iv), (viii), (x), and (xi). To study these cases, the behavior of the least squares, kernel and HATLINK methods will be observed when the degree of model misspecification is varied from where the user's model is nearly correct to a point where it becomes rather questionable.

It will be assumed that the user has specified a first order model in the initial region of interest and that a second order model is specified in the new region. After augmenting design points in the new region of interest, the ESE criterion will be used to measure how well the fitted surface is approximating the true response surface. As indicated in Chapter III, two methods of estimating  $\lambda$  appeared to work best over a broad range of degrees of model misspecification,  $C_P3$  (III.7.9) and PRESS\* (III.6.1). Both methods have been more thoroughly investigated in this study. Also, the behavior of both augmentation methods (IV.1.1) and (IV.1.2) will be considered.

In the following simulations, the initial design matrix is assumed to be the same as stated in Section II.2. The new region of interest is the same as in Section IV.3 and one point is initially observed at the location (2.5, 2). Suppose that the user is fitting a second order parametric model in the new region of interest, while the true underlying model has the same form as the example in Section IV.3, that is

$$y = x_1^2 + x_1 - 3(x_2 - 1)^2 + 35 + CT \left[ 8 \left\{ \sin\left(\frac{\pi * x_1}{1.6}\right) \right\} - 4 \sin\left(\frac{\pi * x_1 * x_2}{1.5}\right) \right] + \epsilon \quad (\text{V.1.1})$$

where  $\epsilon \sim N(0, \sigma^2)$ . For this study,  $\sigma$  was varied between 0.1 and 1, and for each value of  $\sigma$  the value of CT ranged between 0.1 and 1. By altering CT in this fashion, the degree of model misspecification can be varied from where the user's model is very nearly correct (CT = 0.1) to where the misspecification is fairly substantial (CT = 1). Figure V.1.1 and Figure V.1.2 illustrate the minimal model misspecification when CT = 0.1 for the old and new regions of interest, respectively. Figure V.1.3 and Figure V.1.4 illustrate the more pronounced degree of misspecification when CT = 1 for the same regions of interest. A second order model is not unreasonable when CT = 0.1, but is a poor model when CT = 1 although customary lack of fit tests fail to detect this.



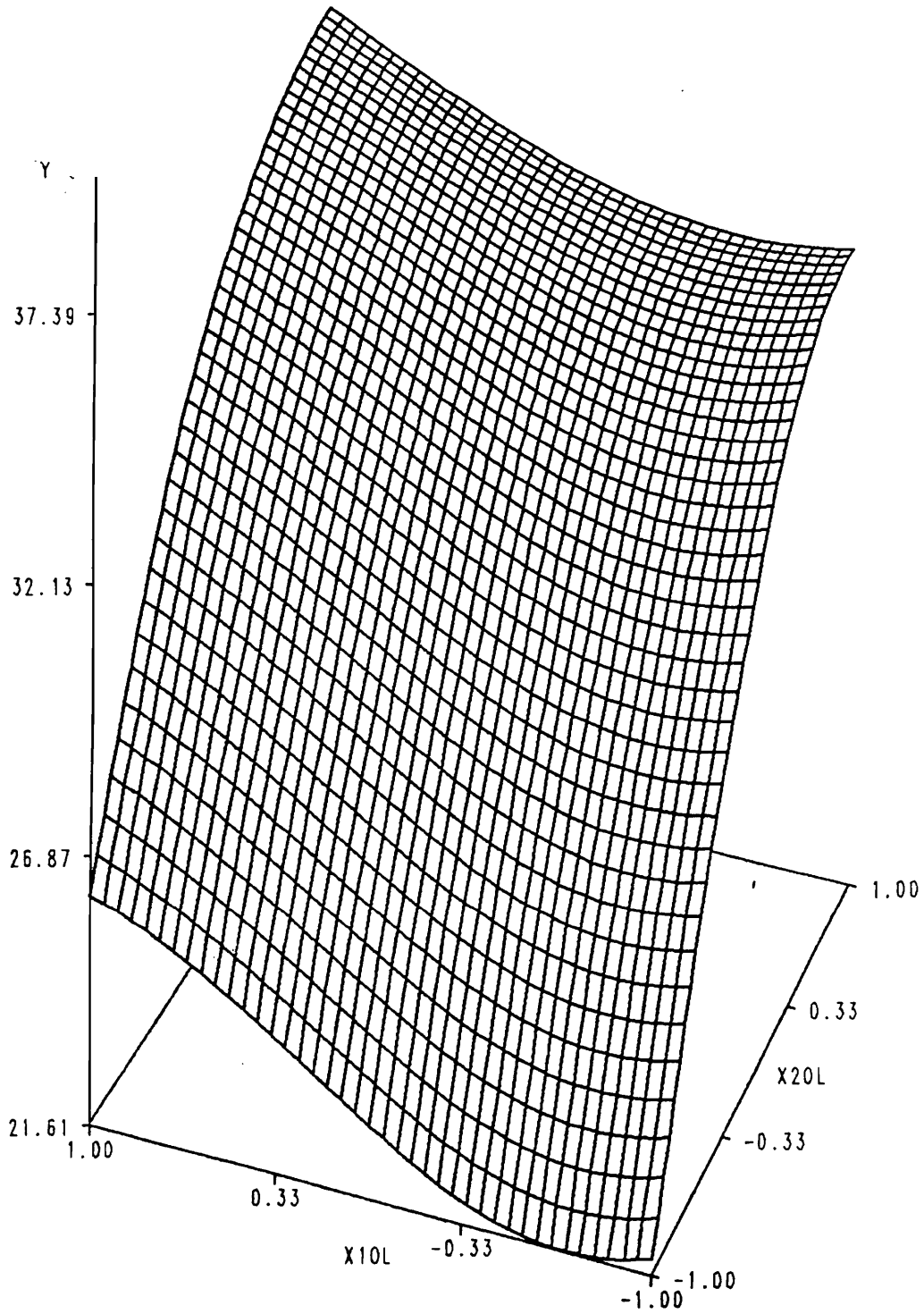


FIG. V.1.1. TRUE MODEL IN OLD REGION WITH  $CT=0.1$

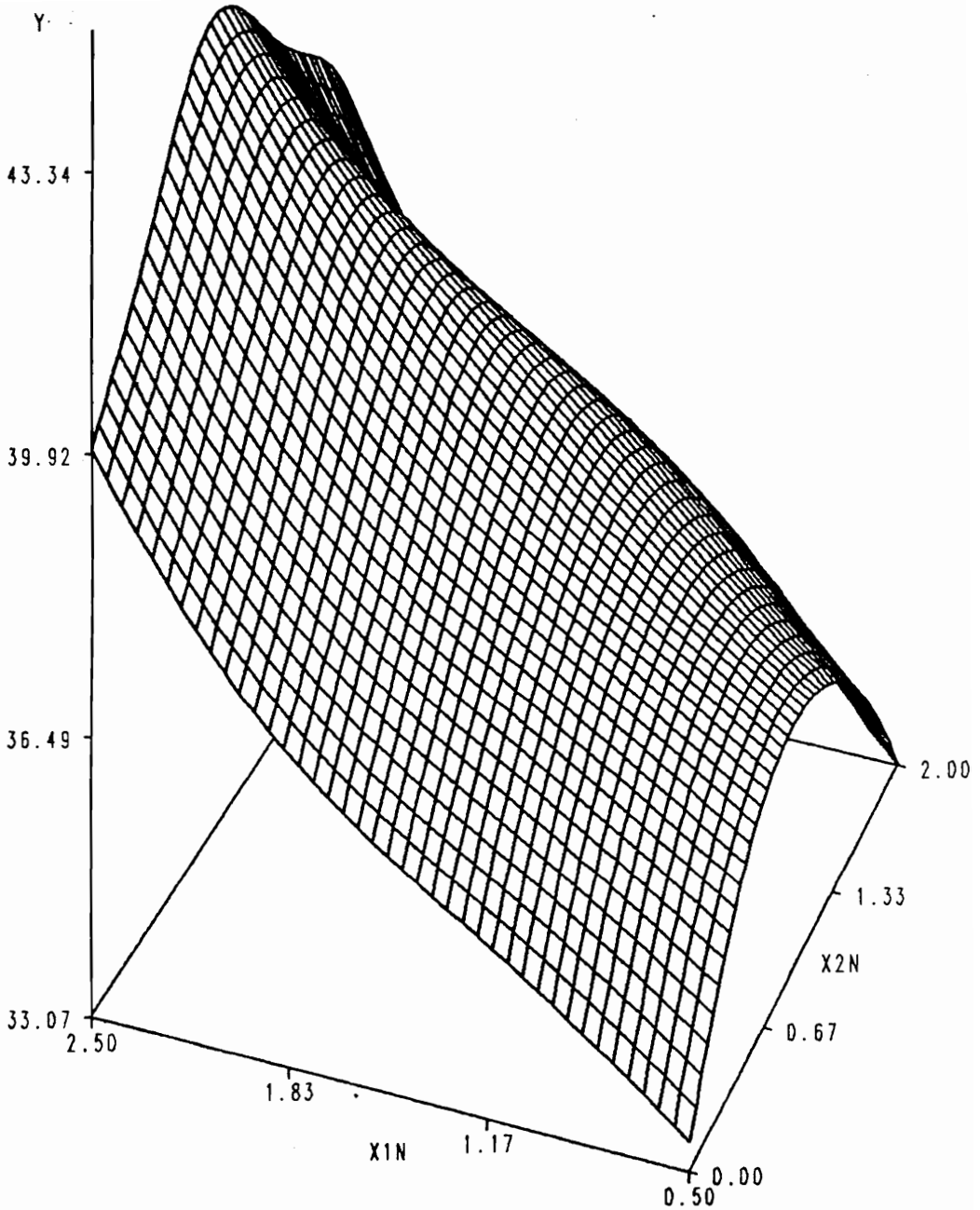


FIG. V.1.2. TRUE MODEL IN NEW REGION WITH CT=0.1

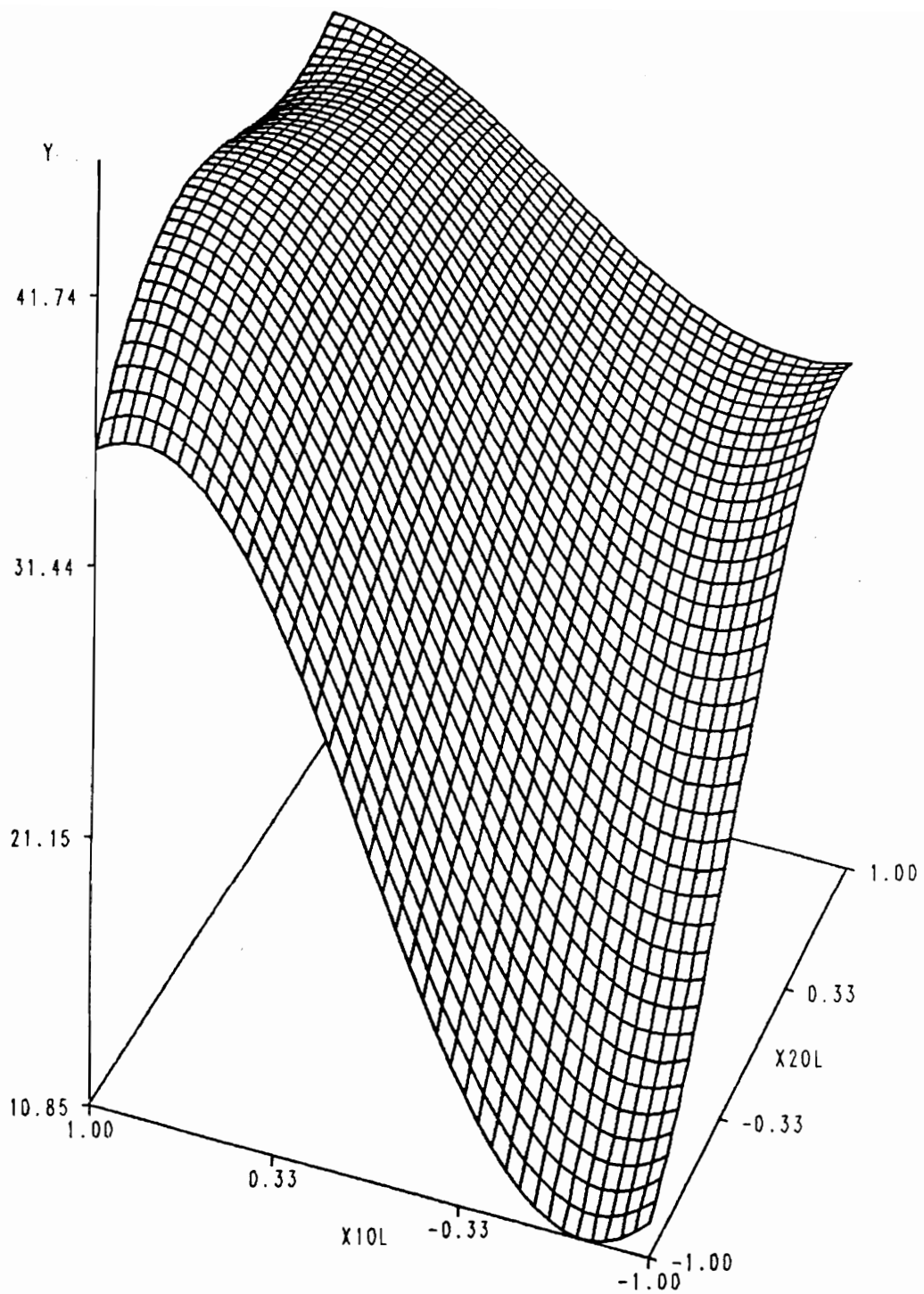


FIG. V.1.3. TRUE MODEL IN OLD REGION WITH CT=1  
68

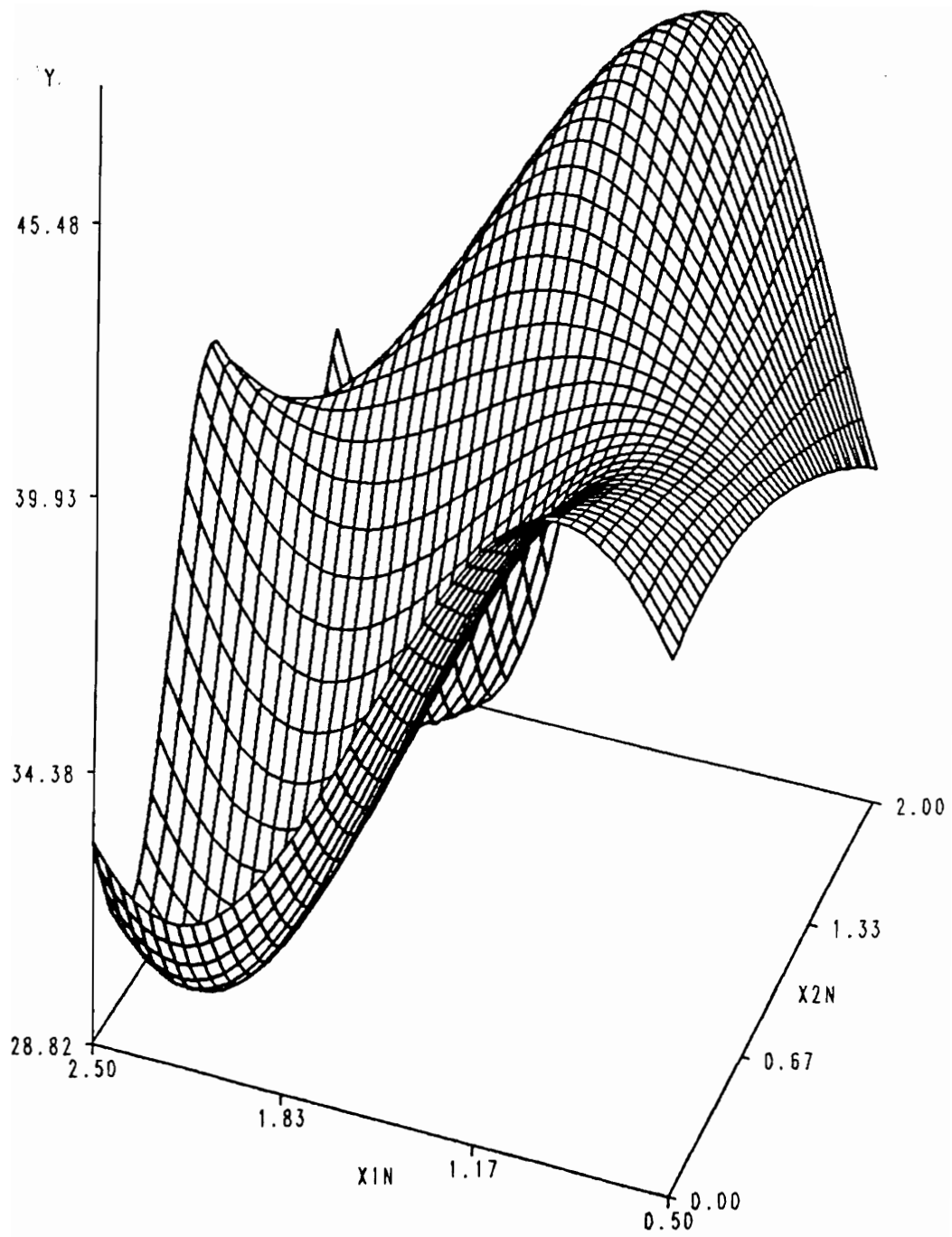


FIG. V.1.4. TRUE MODEL IN NEW REGION WITH CT=1

## V.2 AUGMENTATION USING THE BIIV METHOD

Preliminary work indicated that  $\lambda$  should be selected by the PRESS\* method and that the selected augmentation technique should incorporate bias into the augmentation procedure. The BIIV criterion is designed to attempt to achieve that objective. As stated in Section IV.3, the BIIV method is designed to both minimize integrated prediction variance, assuming the candidate point has been augmented, and to augment where the current fit is the worst. This produced very good results, in terms of the fitted surface as measured by the ESE values.

Initially, results were studied for both the degree of model misspecification represented by  $CT = 0.1$  and  $CT = 1$ , and for  $\sigma$  values of 0.1, 0.2, 0.3, 0.5 and 1. In all simulations performed at this stage, a basic procedure was consistently used. Always using the same seed to generate the random errors and for the specified values of both  $\sigma$  and  $CT$  mentioned above, twenty five points were augmented sequentially, observing responses after augmenting each point using both HATLINK and least squares regression. Recall that, when compared to a second order model, a  $CT$  value of 0.1 represented very little model misspecification, while a  $CT$  value of 1 represented a fairly substantial degree of model misspecification.

In the case when  $CT = 0.1$ , regardless of the  $\sigma$  value, the  $\lambda$  value was almost always zero. The only non-zero values occurred with the smallest  $\sigma$  values when augmenting with the first few points. Obviously, if the  $\lambda$  values are always zero when using HATLINK, it is really equivalent to using least squares regression. Consequently, in most of the cases studied when  $CT = 0.1$ , the fitted surfaces resulting from both HATLINK and least squares regression are the same. With the small amount of model misspecification present in this case, such a result is not unexpected.

A more interesting situation occurred when the  $CT$  value was increased to one.

Regardless of the  $\sigma$  value, the  $\lambda$  value was always between 0.8 and 1, indicative of a larger amount of model misspecification. Such values indicate that HATLINK is using results from both least squares and kernel regressions. Comparison of the ESE values reveal that, after the augmentation of at most one point, HATLINK is always fitting the surface better than the least squares method, as based upon the ESE values. As the graphs in Figure V.2.1 to Figure V.2.5 indicate, the ESE values for HATLINK are then at least one-third lower than the corresponding ESE values for the least squares method. Clearly, HATLINK is winning in these situations.

Based on the initial results, more simulations were performed using  $\sigma = 0.2$  and 1 and  $CT = 0.2, 0.5,$  and  $0.8$ . These were selected as intermediate degrees of misspecification between  $CT$  of 0.1 and 1. Here,  $\sigma = 0.2$  was selected because in previous simulations this appeared to be a troublesome number. When augmenting by integrated prediction variance as detailed in Appendix 3, this was one of the  $\sigma$  values that yielded unexpected results. With  $CT = 0.1$ , the mixing parameter was over-estimated for the first six or seven points augmented and likewise, when  $CT = 1$ , HATLINK required approximately sixteen points before it consistently outperformed least squares based upon the ESE values.

The results for the new values of  $\sigma$  and  $CT$  were very similar to those for the initial ten examples studied. The one main exception was the case when  $CT = 0.2$  and  $\sigma = 0.2$ . In this case, the  $\lambda$  parameter was estimated to be more than 0.6 for the first eight points augmented. With such a small  $CT$ , this is unexpected. Further augmentation yielded mixing parameters close to 0.3 to 0.4, more reasonable for this situation. By comparing the ESE results, it can be seen that estimates of  $\lambda$  over 0.6 are not very appropriate. For such numbers, the ESE values for HATLINK are significantly higher than those for the least squares method. The two methods become competitive only after  $\lambda$  drops below 0.6.

Otherwise, the simulations using the new values of  $\sigma$  and  $CT$  behave in a very

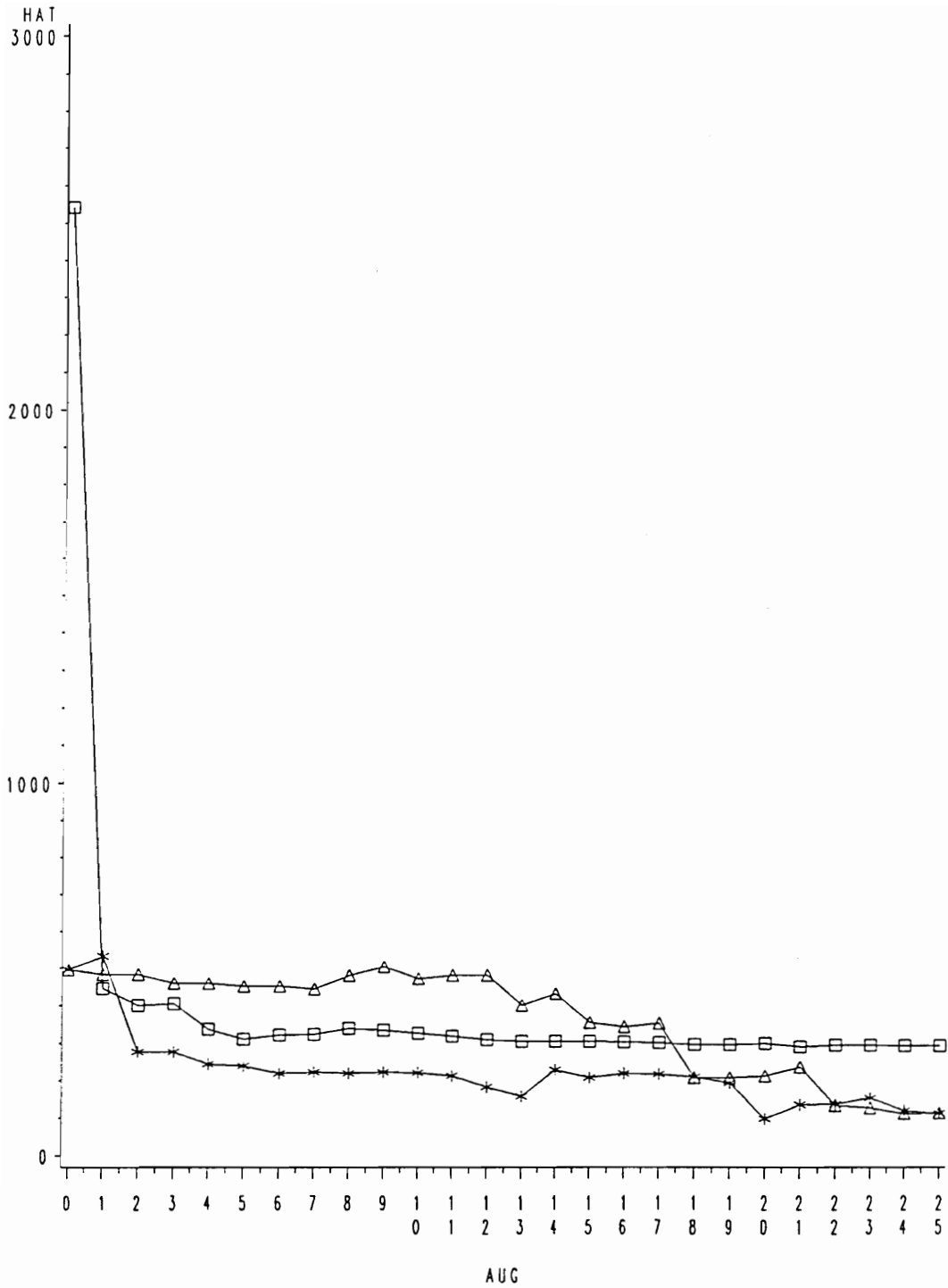


FIGURE V.2.1. ESE VALUES VS AUGMENTATION NUMBER FOR HATLINK (STAR), LEAST SQUARES REGRESSION (SQUARE) & KERNEL REGRESSION (TRIANGLE) WITH CT = 1 AND SIGMA = 0.1

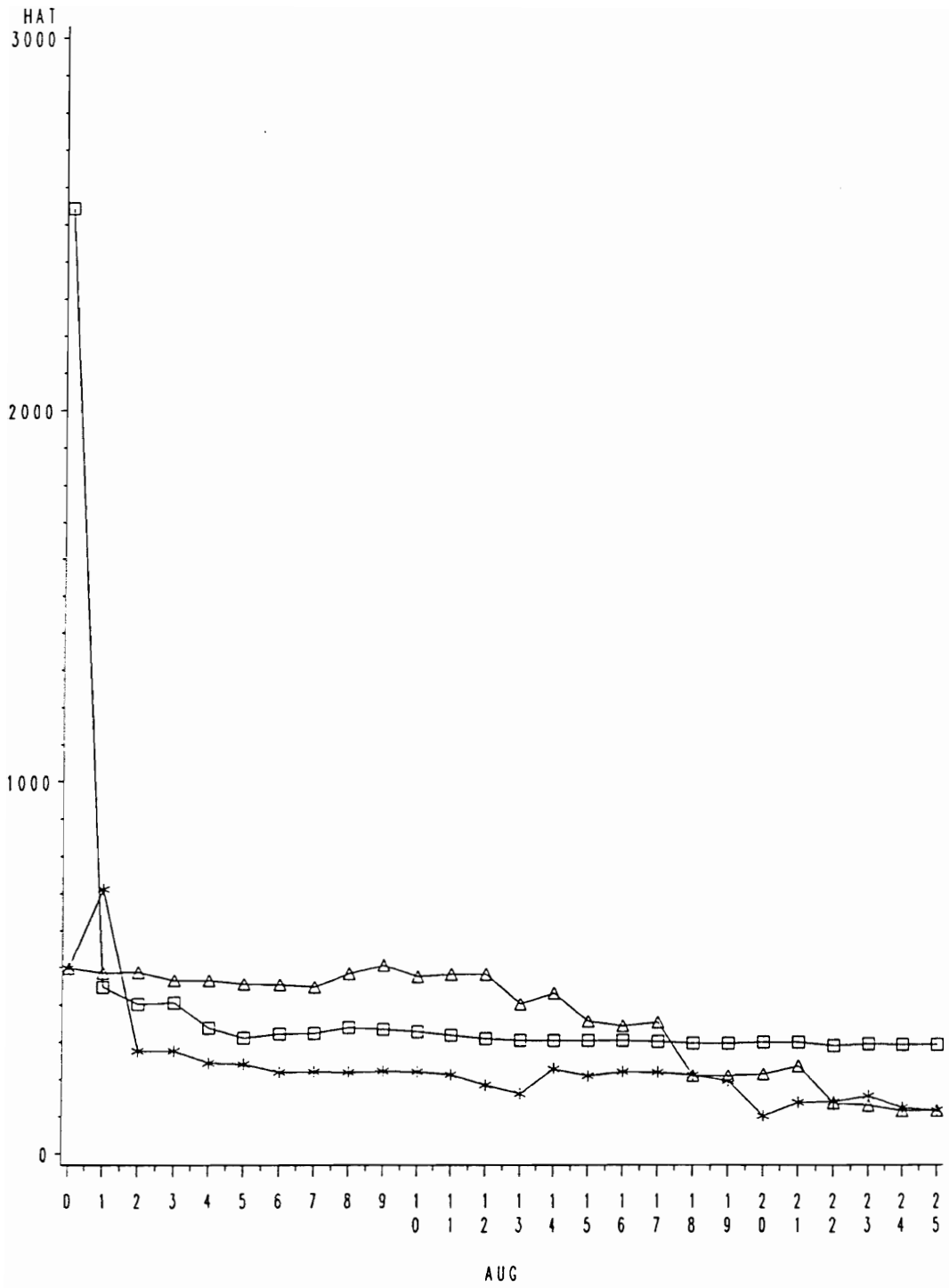


FIGURE V.2.2. ESE VALUES VS AUGMENTATION NUMBER FOR HATLINK (STAR), LEAST SQUARES REGRESSION (SQUARE) & KERNEL REGRESSION (TRIANGLE) WITH CT = 1 AND SIGMA = 0.2



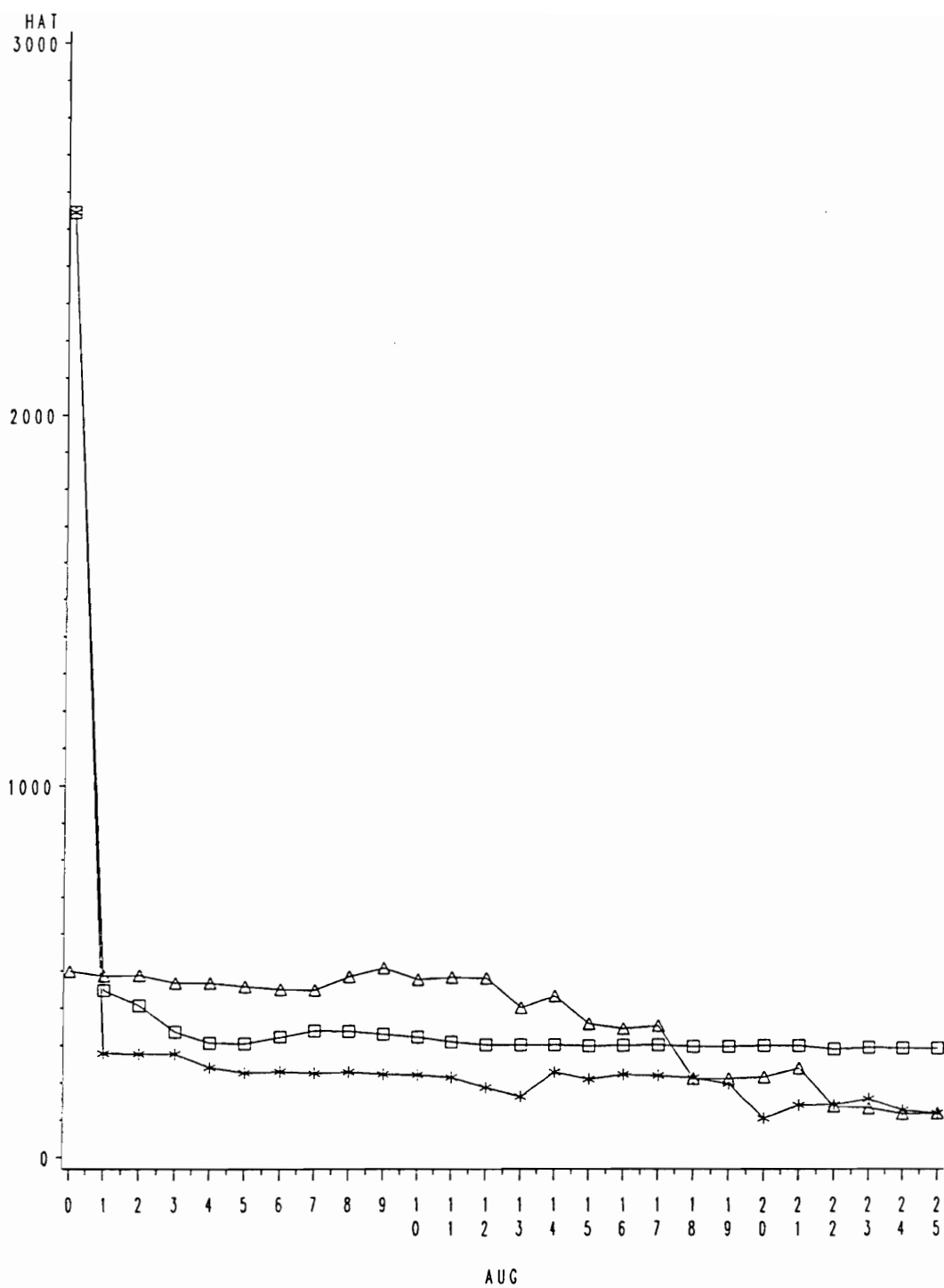


FIGURE V.2.3. ESE VALUES VS AUGMENTATION NUMBER FOR HATLINK (STAR), LEAST SQUARES REGRESSION (SQUARE) & KERNEL REGRESSION (TRIANGLE) WITH CT = 1 AND SIGMA = 0.3

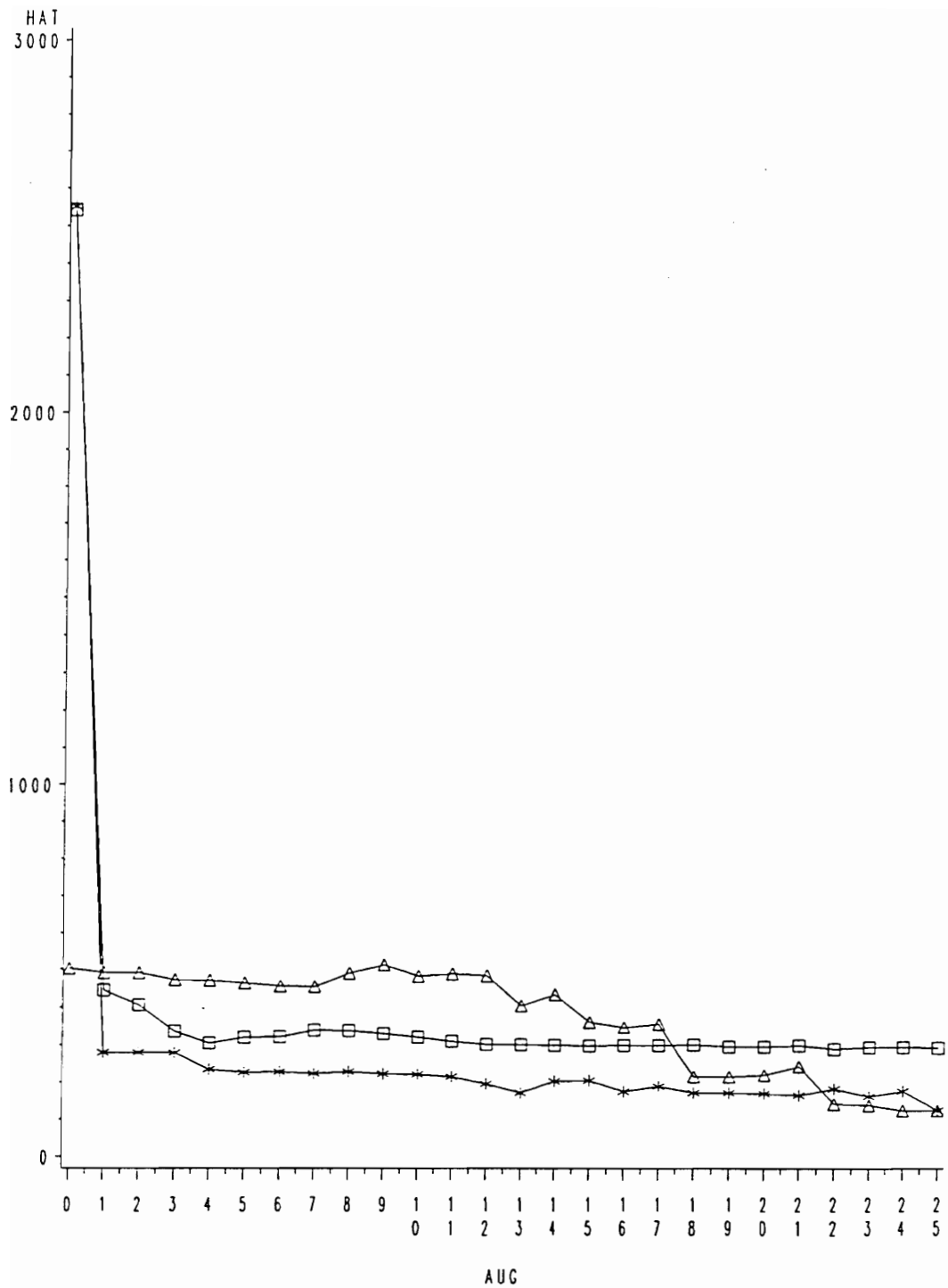


FIGURE V.2.4. ESE VALUES VS AUGMENTATION NUMBER FOR HATLINK (STAR), LEAST SQUARES REGRESSION (SQUARE) & KERNEL REGRESSION (TRIANGLE) WITH CT = 1 AND SIGMA = 0.5

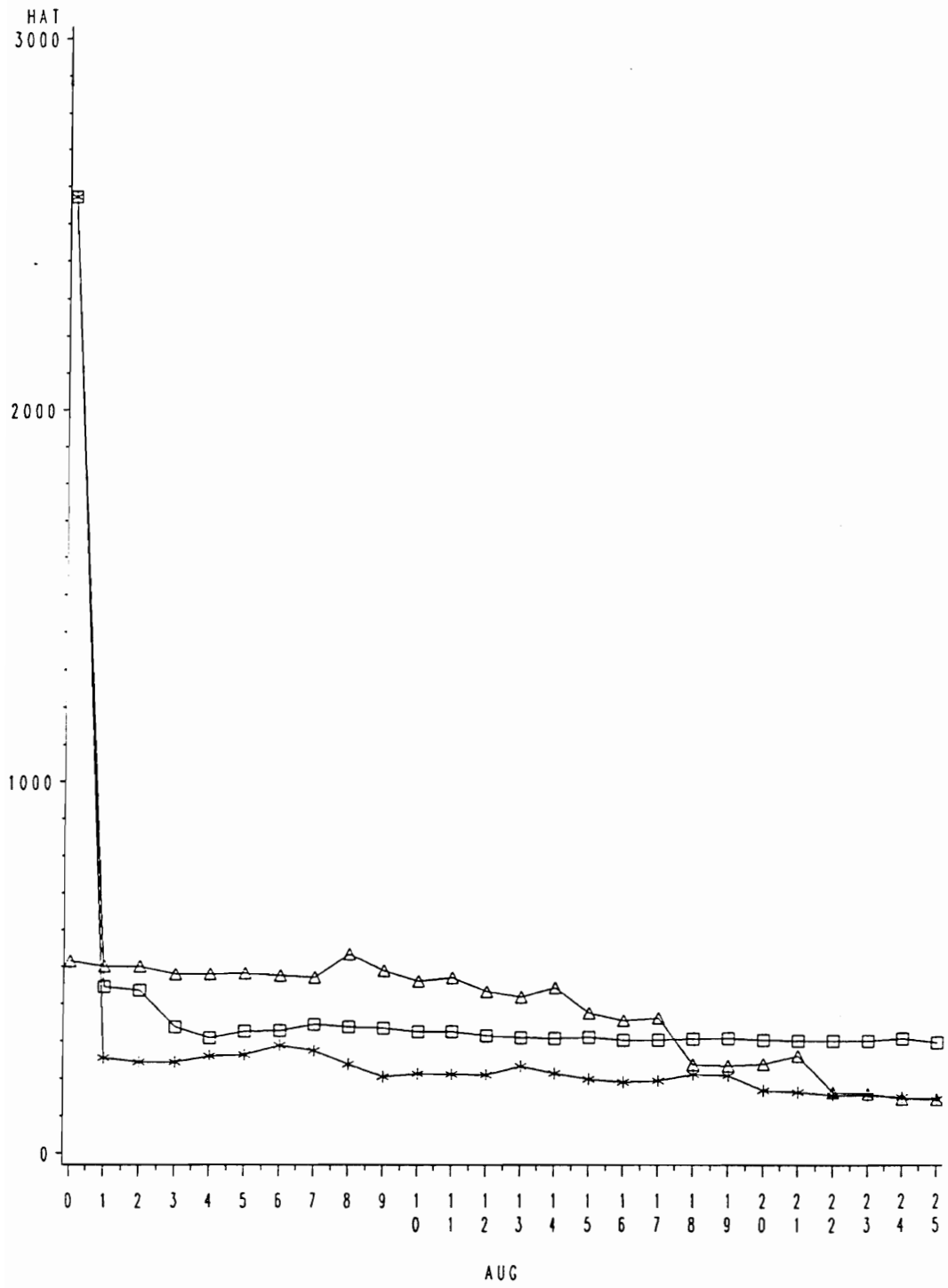


FIGURE V.2.5. ESE VALUES VS AUGMENTATION NUMBER FOR HATLINK (STAR), LEAST SQUARES REGRESSION (SQUARE) & KERNEL REGRESSION (TRIANGLE) WITH CT = 1 AND SIGMA = 1

reasonable fashion. When the model misspecification is small, the  $\lambda$  values tended to be very close to zero and thus HATLINK and least squares regression were basically the same. Conversely, when the model misspecification was larger, the  $\lambda$  values were very close to one and HATLINK was the winner over least squares regression, based upon the ESE values.

All of the above results looked very promising for the HATLINK method proposed in this dissertation, using the PRESS\* criterion for selecting  $\lambda$  and BIIV for augmenting points. From these initial studies, the BIIV method for augmenting points and the PRESS method of selecting  $\lambda$  were determined to be the best of the proposed methods. From the behavior of  $\lambda$  when  $\sigma$  is small, it appears there is some area for improvement in selecting  $\lambda$ . Some variations to the BIIV method of augmentation are presented in Appendix 3.

### V.3 EFFECT OF BATCH SIZE UPON THE HATLINK METHOD

Chapter IV introduced the concept of augmentation with batch size of  $n$ , where  $n$  is not necessarily one. To illustrate this procedure, an example with batch size of two was presented in Section IV.6. Now the effect of batch size when augmenting points will be explored further. Due to the nature of the PRESS\* methods, the bandwidth and mixing parameters can not be updated until experimental runs are performed. When augmenting by a set number of points, the larger the batch size the less often  $h$  and  $\lambda$  can be revised. Using a batch size of one should yield the most immediate gain in terms of ESE values since the two parameters are being updated more frequently. With the augmentation of several batches of points, the effect of batch size appears to diminish as the bandwidth and  $\lambda$  values converge separately to individual numbers.

Using equation (IV.3.1) with  $\sigma = 1$  and  $CT = 0.7$  for the true model, twelve points will

be augmented using batch sizes of one, two, three and four. The results are presented as follows: batch size of one in Table V.3.1, batch size of two in Table V.3.2, batch size of three in Table V.3.3 and batch size of four in Table V.3.4.

First, notice the behavior of the  $\lambda$  parameters in Table V.3.1 through Table V.3.4. Even though the numerical values varied somewhat depending upon the batch size, eventually  $\lambda$  converged to approximately 0.937. The bandwidth parameters were fairly consistent in these examples. Notice also that while the ESE values changed depending upon the batch size, with the augmentation of twelve points these numerical measures of fit indicated that the HATLINK surface was about the same regardless of how often the bandwidth and  $\lambda$  parameters were updated. This also applies to the curves generated by both least squares and kernel regressions. In most cases, with the augmentation of two or three of batches of points, the two parameters  $h$  and  $\lambda$  tended to converge fairly quickly. In such cases, the effect of batch size appears to diminish as measured by the ESE values.

**Table V.3.1. ESE values after augmentation of the batch of points.** Batch size of 1 and a total of 12 points augmented for all three methods with  $\sigma = 1$  and  $CT = 0.7$ . H and LAM refer to the bandwidth and lambda values for HATLINK.

ESE VALUES					
NUM	H	LAM	HATLINK	L.S	KERNEL
0	0.507	0.000	1615.7	1615.7	179.3
1	0.525	1.000	118.4	188.7	178.0
2	0.517	1.000	120.4	175.5	173.4
3	0.519	1.000	110.3	183.9	169.5
4	0.515	0.937	139.4	191.6	178.8
5	0.583	0.625	127.5	159.1	183.3
6	0.591	0.875	130.0	157.8	223.3
7	0.583	0.875	132.8	158.7	224.7
8	0.564	0.875	123.8	159.2	210.5
9	0.563	0.937	121.6	155.9	206.1
10	0.563	0.937	124.5	150.3	188.6
11	0.584	0.937	121.1	149.2	180.4
12	0.590	0.937	115.3	148.7	178.3

**Table V.3.2. ESE values after augmentation of the batch of points.** Batch size of 2 and a total of 12 points augmented for all three methods with  $\sigma = 1$  and  $CT = 0.7$ . H and LAM refer to the bandwidth and lambda values for HATLINK.

ESE VALUES					
NUM	H	LAM	HATLINK	L.S	KERNEL
0	0.507	0.000	1615.7	1615.7	179.3
1	—	—	—	—	—
2	0.514	1.000	150.2	175.5	173.4
3	—	—	—	—	—
4	0.515	0.937	139.4	191.6	178.8
5	—	—	—	—	—
6	0.592	0.625	127.2	154.9	223.3
7	—	—	—	—	—
8	0.591	0.875	130.5	156.4	223.4
9	—	—	—	—	—
10	0.570	0.937	118.4	154.7	198.1
11	—	—	—	—	—
12	0.591	0.937	116.9	148.7	178.2

**Table V.3.3. ESE values after augmentation of the batch of points.** Batch size of 3 and a total of 12 points augmented for all three methods with  $\sigma = 1$  and  $CT = 0.7$ . H and LAM refer to the bandwidth and lambda values for HATLINK.

ESE VALUES					
NUM	H	LAM	HATLINK	L.S	KERNEL
0	0.507	0.000	1615.7	1615.7	179.3
1	—	—	—	—	—
2	—	—	—	—	—
3	0.597	0.750	115.5	175.4	217.1
4	—	—	—	—	—
5	—	—	—	—	—
6	0.603	0.750	149.6	173.7	223.3
7	—	—	—	—	—
8	—	—	—	—	—
9	0.583	0.875	139.7	162.4	222.4
10	—	—	—	—	—
11	—	—	—	—	—
12	0.590	0.937	115.3	153.0	194.0

**Table V.3.4. ESE values after augmentation of the batch of points.** Batch size of 4 and a total of 12 points augmented for all three methods with  $\sigma = 1$  and  $CT = 0.7$ . H and LAM refer to the bandwidth and lambda values for HATLINK.

ESE VALUES					
NUM	H	LAM	HATLINK	L.S	KERNEL
0	0.507	0.000	1615.7	1615.7	179.3
1	—	—	—	—	—
2	—	—	—	—	—
3	—	—	—	—	—
4	0.596	0.625	112.8	169.9	210.9
5	—	—	—	—	—
6	—	—	—	—	—
7	—	—	—	—	—
8	0.592	0.625	131.0	165.9	223.4
9	—	—	—	—	—
10	—	—	—	—	—
11	—	—	—	—	—
12	0.624	0.875	109.9	151.8	178.2

## V.4 INTRODUCTION TO REPETITIONS

As mentioned in chapter 4, using HATLINK and/or the BIIV augmentation criterion resulted in new designs that were stochastic in nature. An obvious concern is the stability of the augmented design if the augmentation procedure is repeated using the same initial design. All of the results mentioned in chapter V looked very promising for the HATLINK method proposed in this dissertation, using the PRESS\* criterion for selecting  $\lambda$  and BIIV for augmenting points. In the next stage of investigation, 10 repetitions were performed for selected values of  $\sigma$  and CT, and each time, twelve points were augmented with a batch size of one. For the 10 repetitions, the mean and sample standard deviation were computed for certain values. These values are the coordinates of the location of the points augmented, the ESE values after augmenting with a point, the maximum fitted value, and the distance between the location of the fitted maximum and the true maximum. With each point augmented, all of the above quantities were calculated. The summary values were included to help judge the consistency of the HATLINK augmentation method as proposed above.

For all the repetitions performed here, the same initial design will be used that was introduced in Section IV.3 and consistently used in all the examples that have been presented so far. Also, the true model will be assumed known to be the same as (IV.3.1), except for the degree of model misspecification. The general equation is given by

$$Y = x_1^2 + x_1 - 3(x_2 - 1)^2 + 35 + \text{CT} \left\{ 8 \left[ \sin \left( \frac{\pi * x_1}{1.6} \right) \right] - 4 \left[ \sin \left( \frac{\pi * x_1 * x_2}{1.5} \right) \right] \right\} + \epsilon, \quad (\text{V.4.1})$$

where  $\epsilon \sim N(0, \sigma^2)$ . Further, the same initial seed will be used for all cases of  $\sigma$  and CT studied.



For 10 repetitions using the same  $\sigma$  and CT values, enough distinct random errors will be generated from the initial seed to yield sufficient observed response values for the candidate points.

For this phase of study, eight cases were considered and are given in Table V.4.1. These values of CT and  $\sigma$  were selected to attempt to cover a broad range of values for both variables. Cases 1, 2, 6, and 8 were studied first. For the cases where CT = 0.1, HATLINK consistently selected  $\lambda$  to either equal 0 or be very close to 0. After twelve points were augmented, the HATLINK and least squares methods were almost identical in terms of points augmented, ESE values, fitted maxima, and distances to the true maximum. Because of this behavior, no further examples with CT = 0.1 were studied.

Case 3 is the combination of CT and  $\sigma$  values studied the most extensively in examples in Chapter 3 through Chapter 5 and in Appendix 3. Cases 4, 5, and 7 were studied as further examples of the behavior of the HATLINK method with different values of  $\sigma$ . Cases 3 through 8 will be examined more thoroughly in the following sections. As stated above, the question of interest here is whether or not the performance of HATLINK is consistent across different repetitions.

**Table V.4.1. Values of  $\sigma$  and CT Studied with Ten Repetitions with the Corresponding Case Numbers.**

CASE	1	2	3	4	5	6	7	8
<u>CT</u>	0.1	0.1	0.7	1	1	1	1	1
<u><math>\sigma</math></u>	0.2	1	1	0.025	0.1	0.2	0.5	1

#### V.4.A CASE 3: $CT = 0.7$ & $\sigma = 1$

Since this combination of  $CT$  and  $\sigma$  values have already been studied in the previous chapters, some results for the first repetition have already been presented. For example, Section IV.3 presented graphs of the first six augmented points for both the HATLINK and least squares methods of augmentation. Section IV.3 indicated that the HATLINK method appeared to be outperforming both least squares and kernel regressions based upon the ESE values after the first point was augmented.

For each point augmented, average  $x_1$  and  $x_2$  values were computed for the ten repetitions. This yields what could be considered an “average” design. Plots of the twelve points for a typical design can be made for HATLINK, least squares and kernel regressions as presented in Figures V.4.1 to V.4.3 respectively. As expected, these appear to be very different designs. Comparison of the corresponding sample standard deviations for the average points reveal that these values are very similar, ranging between 0 and 0.9.

After twelve points are augmented, the fitted maximum and distance to the true maximum are calculated for each repetition and averaged. In this case, both the distance and the fitted value were about the same for all three methods. The average values for the three methods are summarized in Table V.4.2. The three methods tended to perform about the same in terms of estimating and locating the maximum. In Section IV.3 it was observed that after the first point was augmented, the ESE values for HATLINK dropped significantly compared to the least squares and kernel methods. It was concluded that HATLINK worked better in estimating the overall surface. The drop in the ESE values for HATLINK was consistently observed over the ten repetitions performed, and can be seen in the average ESE values. These values are presented in Table V.4.3.

As can be seen from Table V.4.3, after the second point has been augmented, HATLINK is always outperforming both least squares and kernel in terms of the ESE values. It appears that the behavior observed in the previous chapters is consistent with repetitions, implying that at least in this case, the method is relatively consistent across repetitions. Additional examples are presented to determine if this behavior is common over a broader range of CT and  $\sigma$  values.

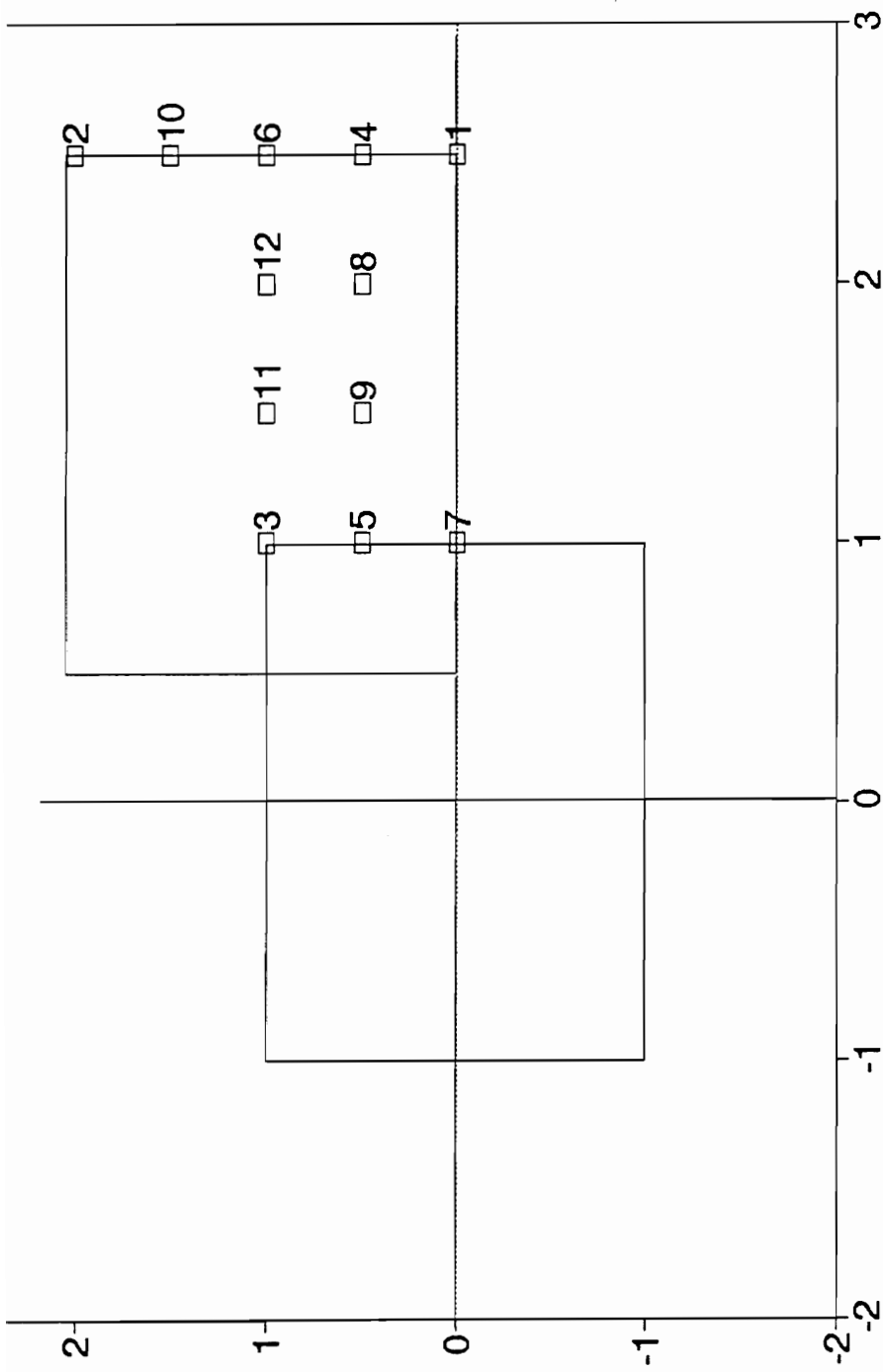


Fig. V.4.1. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with  $CT=0.7$  &  $\Sigma=1$

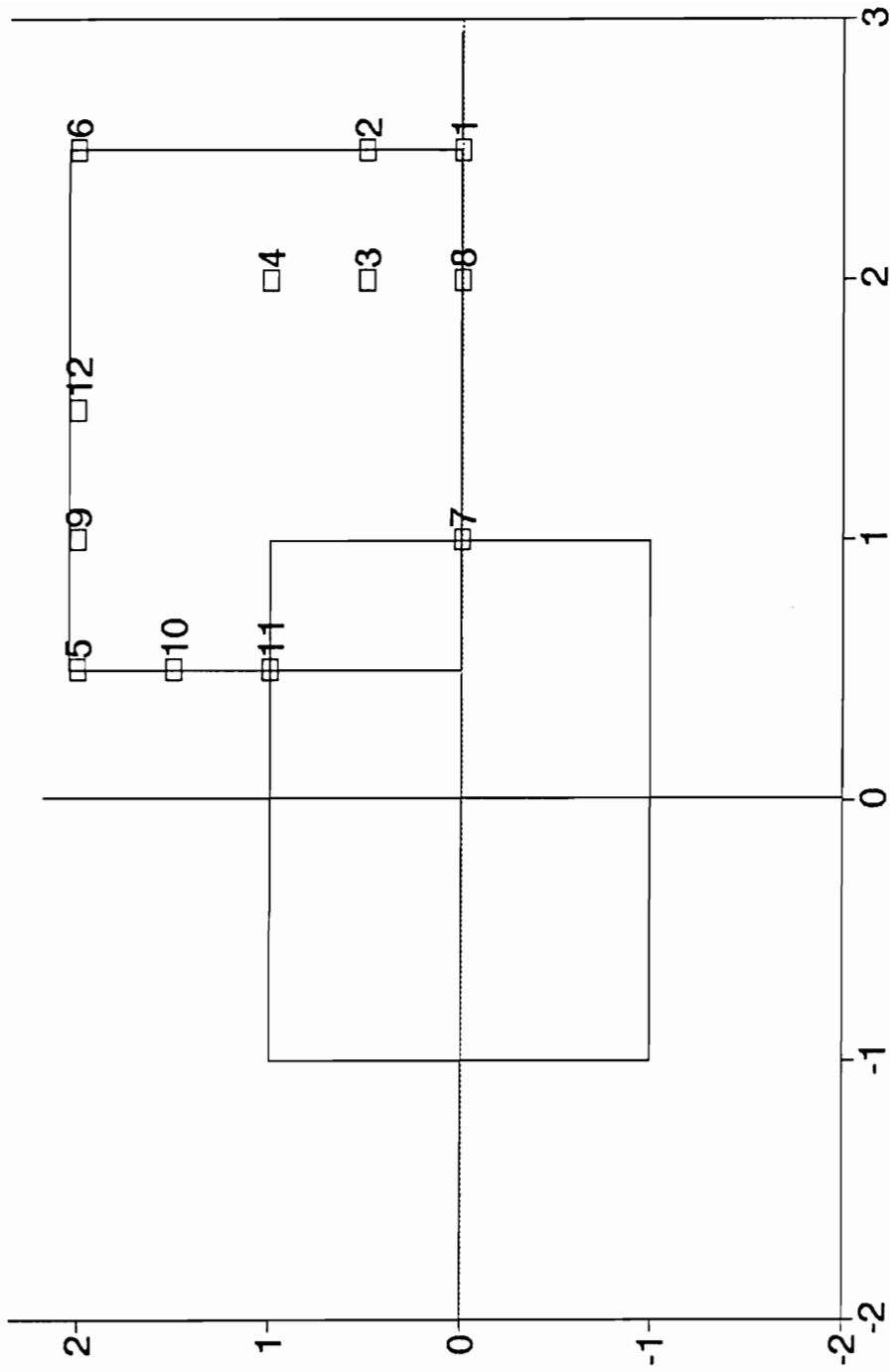


Fig. V.4.2. Twelve Points Augmented Using L.S. & BIIV, Typical Repetition with  $CT=0.7$  &  $\Sigma=1$

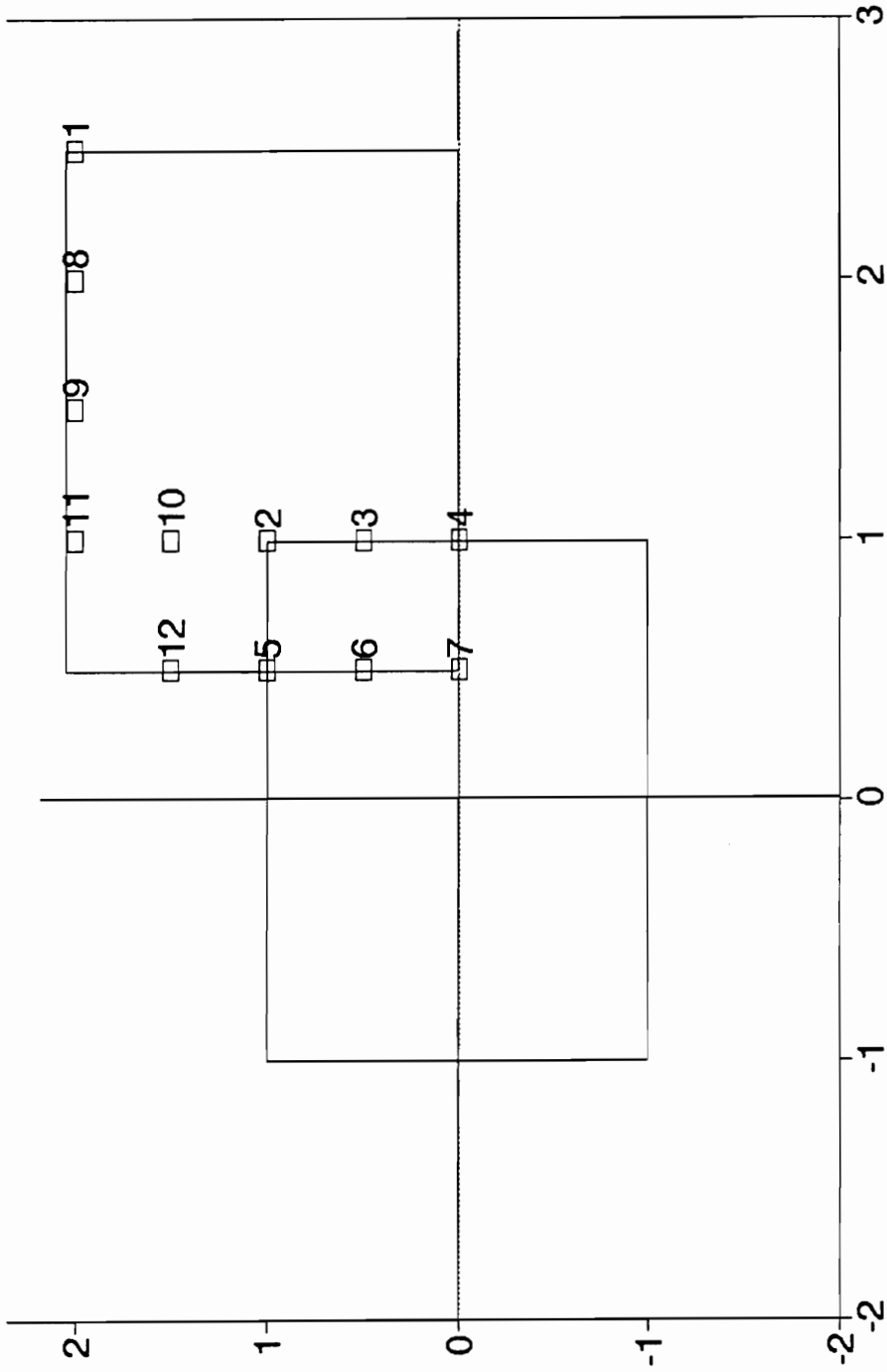


Fig. V.4.3. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with  $CT=0.7$  &  $\Sigma=1$

**Table V.4.2. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 42.71 located at (1.25, 1.6) with  $\sigma = 1$  and  $CT = 0.7$ .

METHOD	$\overline{\text{dis.}}$	$s_{dis}$	$\overline{\text{max.}}$	$s_{max}$
HATLINK	0.503	0.414	39.818	0.737
least squares	0.444	0.212	39.582	0.356
kernel	0.484	0.191	40.382	0.528

**Table V.4.3. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 1$  and  $CT = 0.7$ .

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	1130.7	546.0	1299.5	272.8	181.2	18.9
1	251.0	268.9	211.7	28.3	183.8	23.2
2	128.0	13.2	192.7	19.1	187.0	25.9
3	132.0	21.7	177.2	21.9	176.7	20.0
4	146.0	24.5	171.1	18.9	175.9	18.4
5	137.9	16.2	159.6	10.2	176.6	17.2
6	130.1	19.7	164.2	9.5	184.5	23.7
7	125.2	16.7	170.8	13.6	181.3	26.4
8	116.8	15.7	166.6	12.0	188.6	23.1
9	114.5	12.5	164.2	10.3	185.1	19.0
10	114.7	12.1	160.3	10.3	177.5	14.3
11	111.7	13.1	159.9	9.2	174.6	16.1
12	106.2	11.0	158.3	8.4	161.3	15.5

#### V.4.B CASE 4: $CT = 1$ & $\sigma = 0.025$

Here, it was desired to study the behavior of the augmentation method with a very small sigma value. When selecting data locations with integrated prediction variance as detailed in Appendix 3, small  $\sigma$  values meant that HATLINK required 15 or 16 points to be added to the design matrix before it yielded lower ESE values than the least squares method. It is hoped that augmentation with BIIV will improve this problem.

Plots of a typical design can be made for HATLINK, least squares and kernel regressions. The plot of the HATLINK design is given in Figure V.4.4, the least squares design in Figure V.4.5 and the kernel design in Figure V.4.6. If the sample standard deviations of the locations for the augmented points are all zero, then each repetition must have placed all the points in an identical order. This has occurred for both the least squares and kernel designs. The design generated by the HATLINK method was more variable, but comparisons of the actual points augmented reveal that in most cases the augmented points were the same. The augmented points varied occasionally.

After twelve points are augmented, the fitted maximum and distance to the true maximum are calculated for each repetition and averaged, as summarized in Table V.4.4. Again, both the distance and the fitted value were about the same for all three methods. For all ten repetitions, with the augmentation of the first point, the ESE values for HATLINK dropped significantly compared to the least squares and kernel methods. The average ESE values reflect this. Again, it appears that HATLINK worked better in estimating the overall surface. The average ESE values are presented in Table V.4.5.



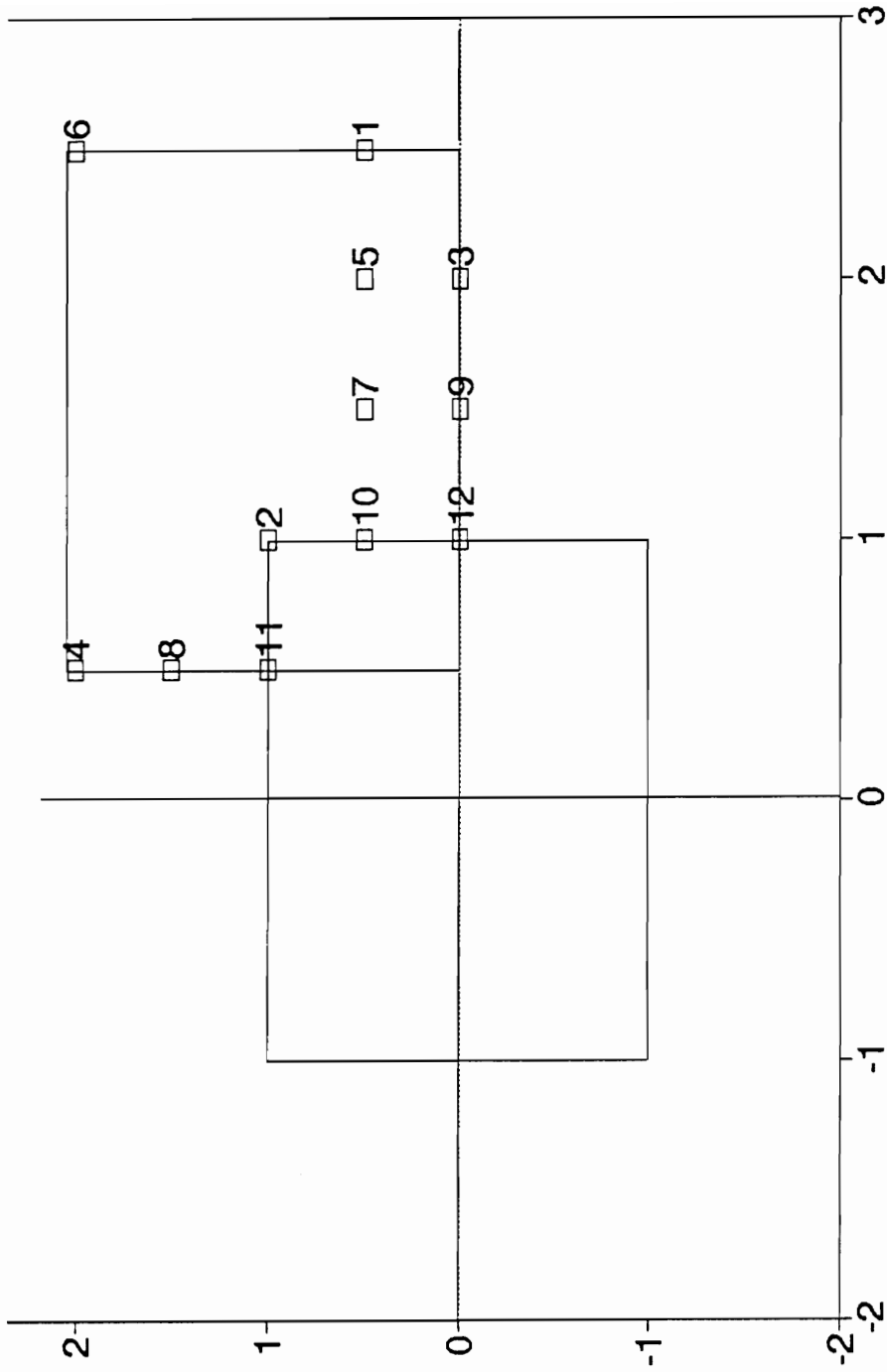


Fig. V.4.4. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.025$

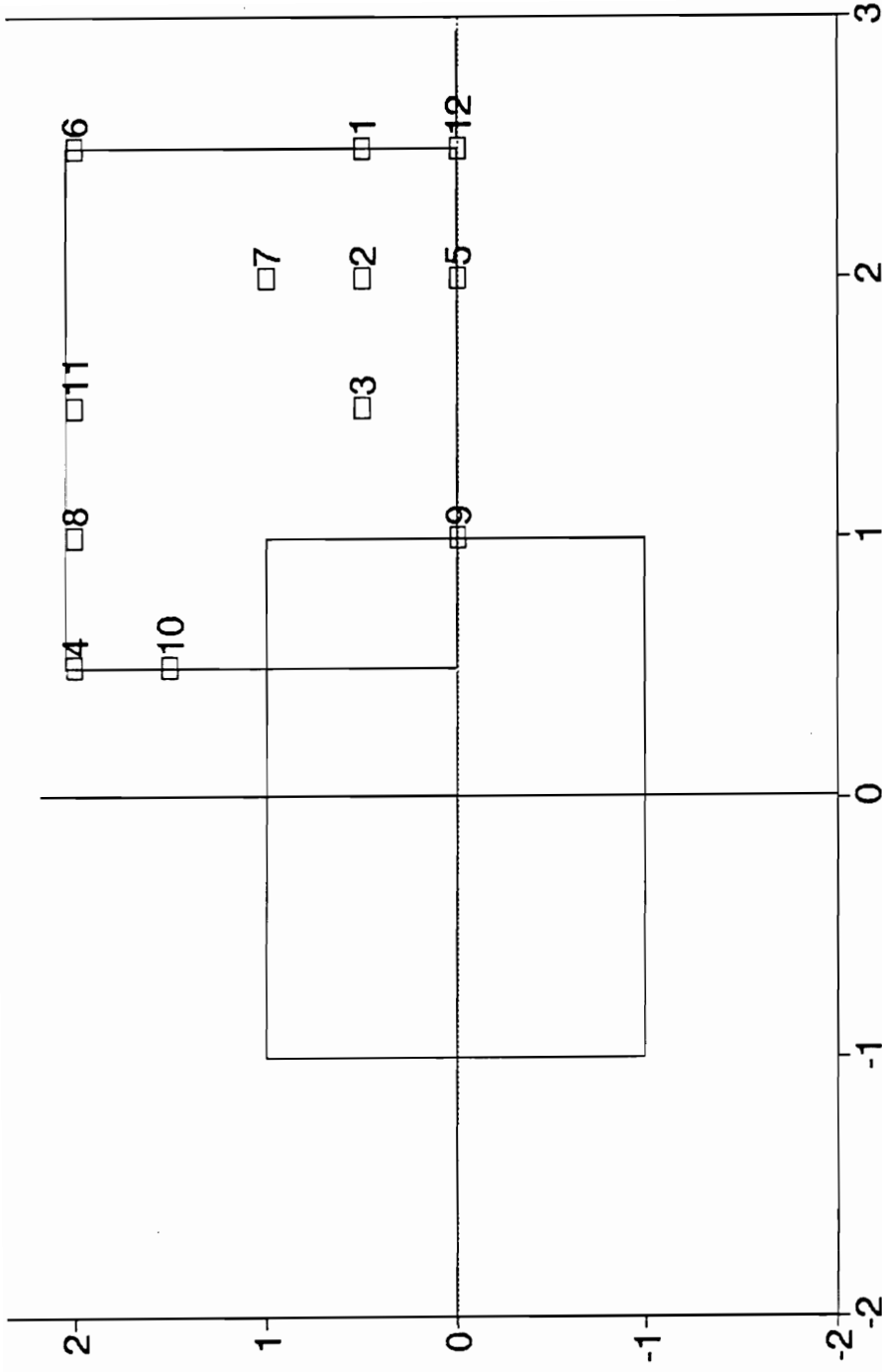


Fig. V.4.5. Twelve Points Augmented Using L.S. & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.025$

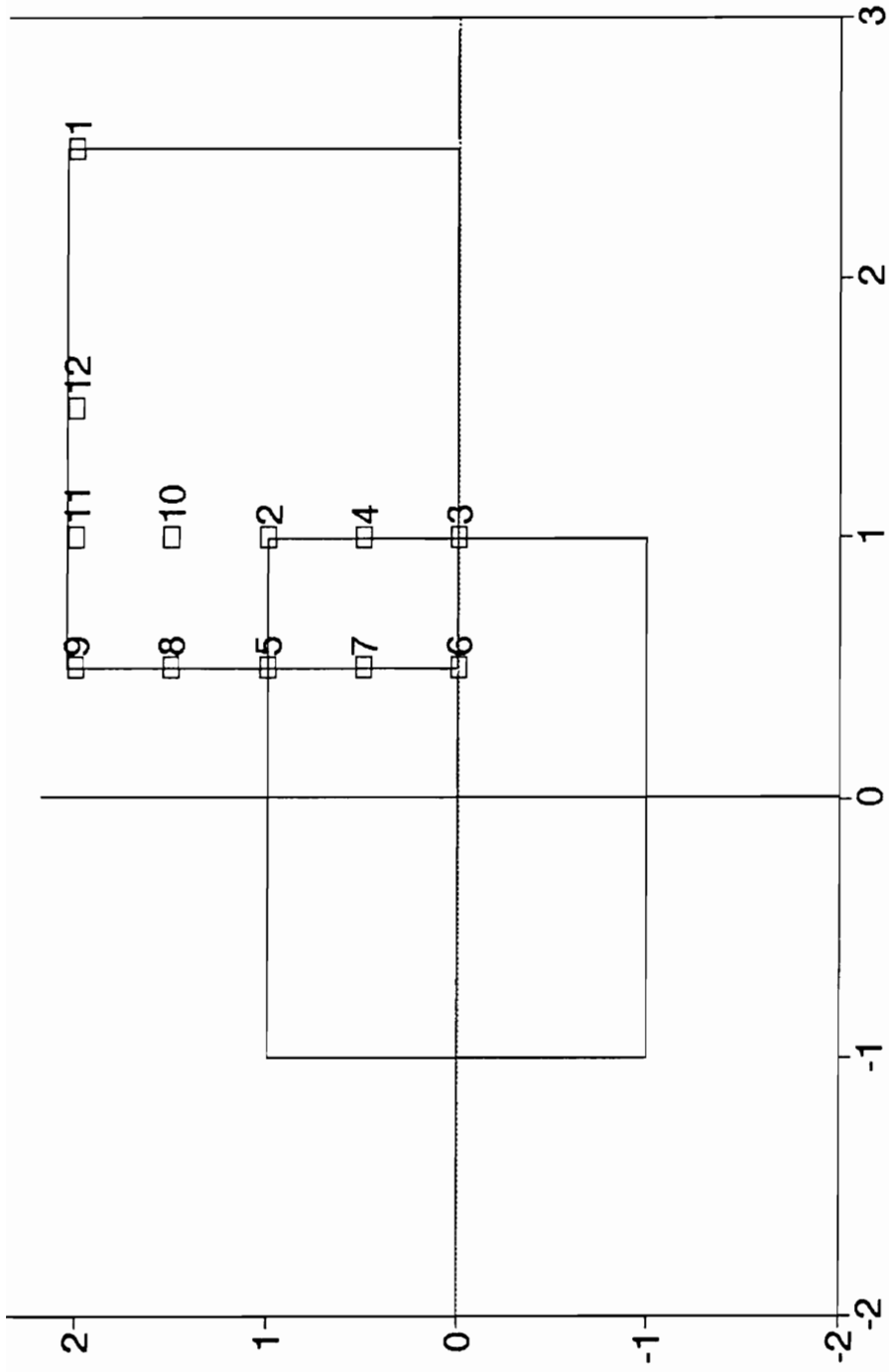


Fig. V.4.6. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.025$

**Table V.4.4. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 44.86 located at (1, 2) with  $\sigma = 0.025$  and  $CT = 1$ .

METHOD	$\overline{dis.}$	$s_{dis}$	$\overline{max.}$	$s_{max}$
HATLINK	1.87	0.41	41.19	0.13
least squares	0.40	0.21	41.15	0.01
kernel	0.00	0.00	41.23	0.01

**Table V.4.5. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 0.025$  and  $CT = 1$ .

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	627.5	1.11	2554.4	9.22	495.1	0.94
1	276.3	0.69	445.8	1.21	480.9	1.05
2	275.0	0.77	400.5	0.79	480.1	1.09
3	246.2	10.5	404.5	0.80	473.7	0.97
4	296.1	18.3	337.5	0.46	455.8	0.70
5	290.4	17.9	310.7	0.35	450.3	0.75
6	290.2	25.1	321.5	0.33	451.3	0.72
7	288.4	23.5	323.0	0.34	442.6	0.81
8	288.3	24.5	339.0	0.44	478.6	0.92
9	293.4	25.0	334.0	0.45	502.8	1.05
10	287.6	23.8	327.5	0.42	471.5	0.88
11	288.4	26.6	319.6	0.33	481.6	0.90
12	278.5	35.8	310.1	0.31	480.5	0.71

#### V.4.C CASE 5: $CT = 1$ & $\sigma = 0.1$

For this case, the  $\sigma$  value was selected to be slightly higher than the previous cases. This is to investigate further the behavior of the HATLINK augmentation procedure when the  $\sigma$  values are small.

Plots of the twelve points in a typical design can be made for HATLINK, least squares and kernel regressions as given in Figure V.4.7, Figure V.4.8, and Figure V.4.9. Comparison of the corresponding sample standard deviations for the average points reveal that the least squares and kernel designs were almost always the same, regardless of the number of repetitions, while the HATLINK design was more variable. For the HATLINK case, comparisons of the actual points augmented reveal that most times the augmented points were the same, the points varying only in a few cases.

After twelve points are augmented, Table V.4.6 summarizes the average of the fitted maximum and distance to the true maximum. Both the distance and the fitted value were about the same for all three methods and, in terms of estimating and locating the maximum, the three methods tended to perform about the same. The ESE values for HATLINK quickly dropped significantly in comparison to the least squares and kernel methods. Again, it appears that HATLINK was superior in estimating the overall surface. This behavior was consistently observed over the ten repetitions performed, and can be seen in the average ESE values as presented in Table V.4.7.

As can be seen from Table V.4.7, after the second point has been augmented, HATLINK is always outperforming both least squares and kernel in terms of ESE. Note the standard deviations for the three methods. The standard deviations for the HATLINK method are significantly higher than those for either of the other two. Comparisons of the original ESE

values for the individual repetitions reveal that those for the HATLINK method tended to vary between two distinct numbers over the ten repetitions, thus leading to the higher standard deviations. Both numbers from the HATLINK method were lower than the corresponding numbers for the least squares and kernel methods.

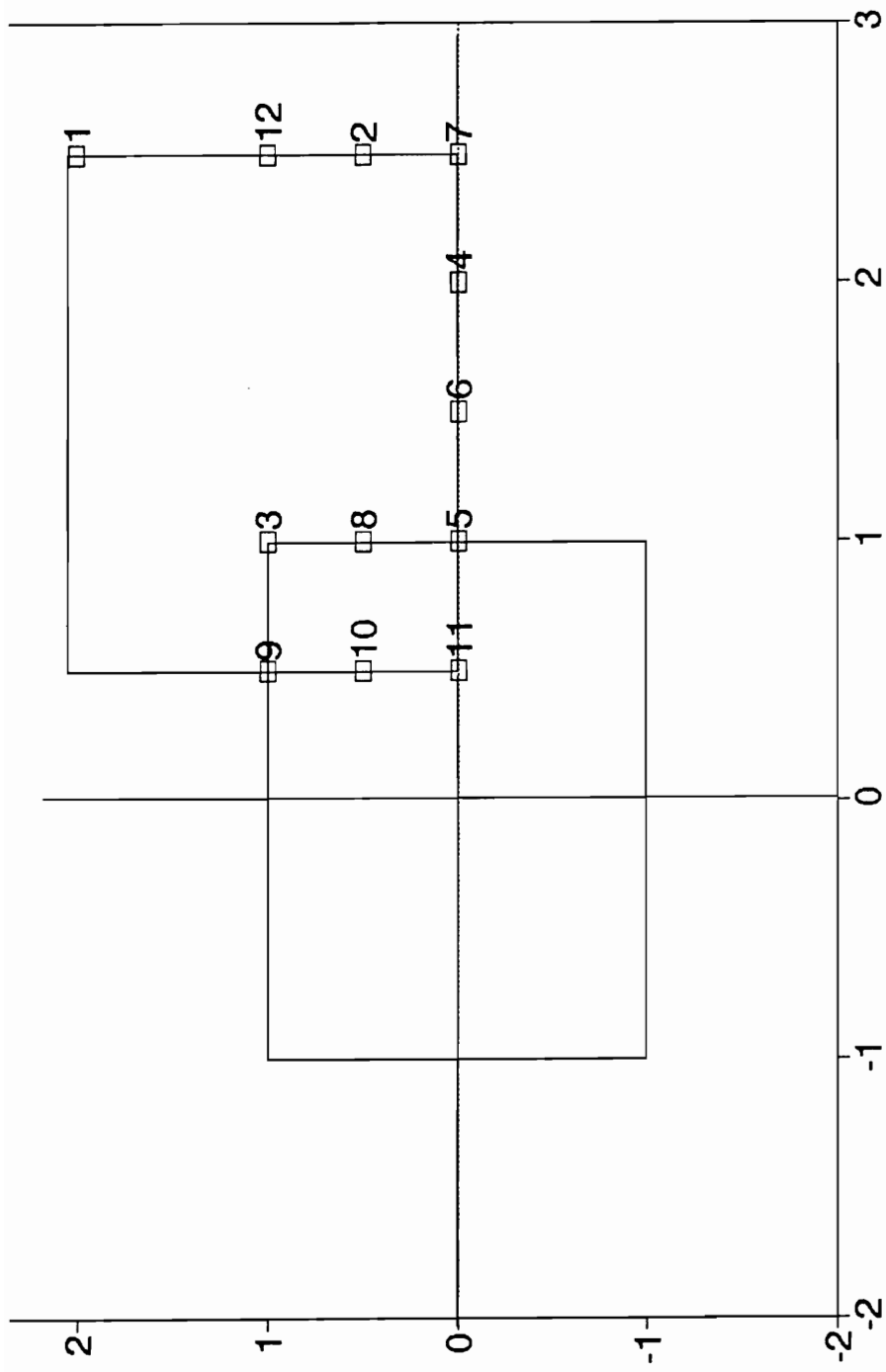


Fig. V.4.7. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.1$

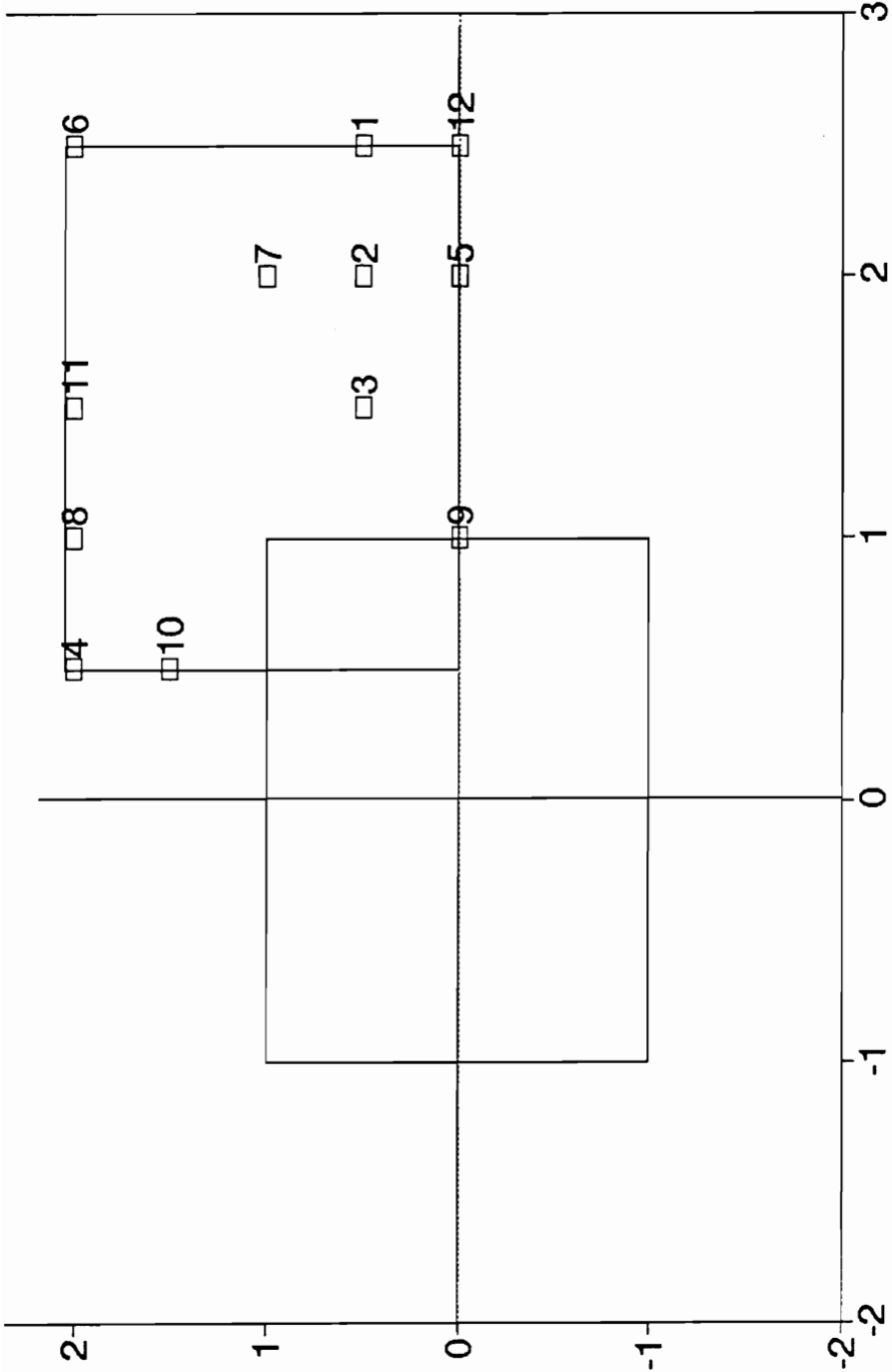


Fig. V.4.8. Twelve Points Augmented Using L.S. & BIV, Typical Repetition with CT=1 & Sigma=0.1



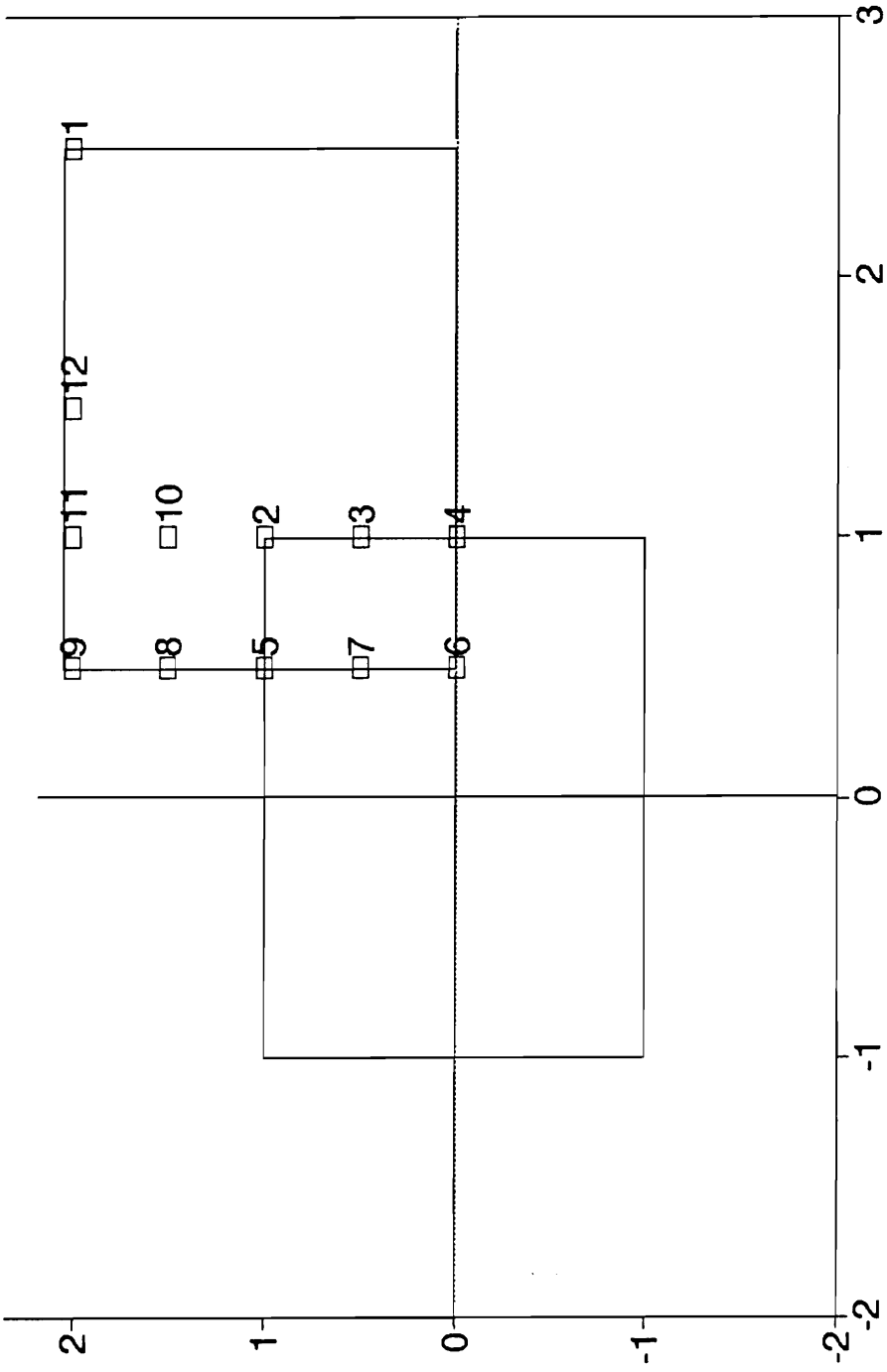


Fig. V.4.9. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.1$

**Table V.4.6. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 44.86 located at (1, 2) with  $\sigma = 0.1$  and  $CT = 1$ .

METHOD	$\overline{\text{dis.}}$	$s_{dis}$	$\overline{\text{max.}}$	$s_{max}$
HATLINK	0.92	0.57	41.45	1.11
least squares	0.45	0.16	41.17	0.05
kernel	0.30	0.79	41.24	0.04

**Table V.4.7. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 0.1$  and  $CT = 1$ .

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	572.5	69.8	2547.6	37.0	493.3	3.85
1	475.9	258.4	444.1	4.81	481.6	4.10
2	275.3	2.65	399.2	3.11	481.5	4.26
3	258.9	17.6	403.3	3.17	468.2	9.92
4	266.7	29.8	337.7	1.89	457.3	3.09
5	263.6	31.7	311.3	1.47	449.8	3.04
6	240.0	33.5	321.6	1.32	450.3	2.90
7	243.2	32.1	323.1	1.34	441.8	3.43
8	241.3	33.0	338.9	1.73	477.6	3.68
9	245.0	33.6	333.9	1.77	501.3	4.13
10	242.9	32.6	327.5	1.66	470.9	3.66
11	237.9	36.2	319.7	1.31	480.3	3.60
12	215.0	51.1	310.1	1.21	479.4	2.96

#### V.4.D CASE 6: $CT = 1$ & $\sigma = 0.2$

The value of  $\sigma$  selected for this case had problems when augmenting points using integrated prediction variance. It was anticipated that using BIIV as the augmentation criterion would improve the results.

Plots of the twelve points in a typical design can be made for HATLINK (Figure V.4.10), least squares (Figure V.4.11) and kernel regressions (Figure V.4.12). For this case, the HATLINK method consistently selected the same locations after the first three points were augmented. For all ten repetitions, the kernel method always selected the same values in the same order of augmentation. The least squares method was the most variable in terms of the points augmented and the order of augmentation.

After twelve points are augmented, the fitted maximum and distance to the true maximum are calculated for each repetition and averaged. Both the distance and the fitted maximum were about the same for HATLINK, least squares regression and kernel regression. The average values for the three methods are summarized in Table V.4.8. In terms of estimating and locating the maximum, the three techniques tended to perform about the same. All three methods performed about equally well in locating the maximum, although least squares regression had the lowest standard deviation. With the augmentation of the first two points, the ESE values for HATLINK dropped significantly compared to the least squares and kernel methods. Again, it appears that HATLINK was superior in estimating the overall surface. The average ESE values are presented in Table V.4.9.

As can be seen from Table V.4.9, after the second point has been augmented, HATLINK is always outperforming both least squares and kernel in terms of ESE. In this case, the standard deviations for the HATLINK method dropped further and faster than those for

either of the other two methods. With the third point augmented, the least squares standard deviation increased substantially as compared to the previous standard deviation.

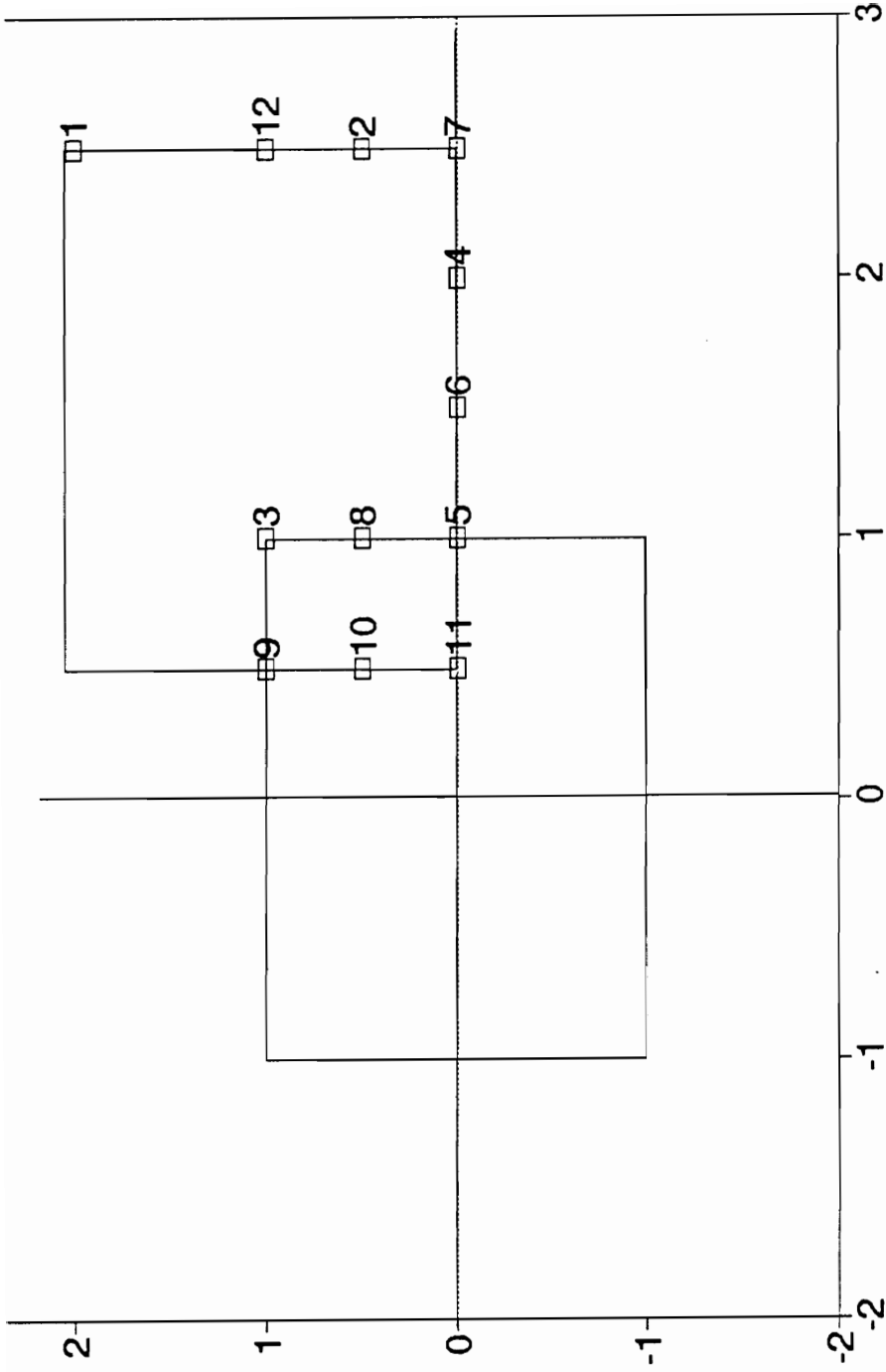


Fig. V.4.10. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.2$

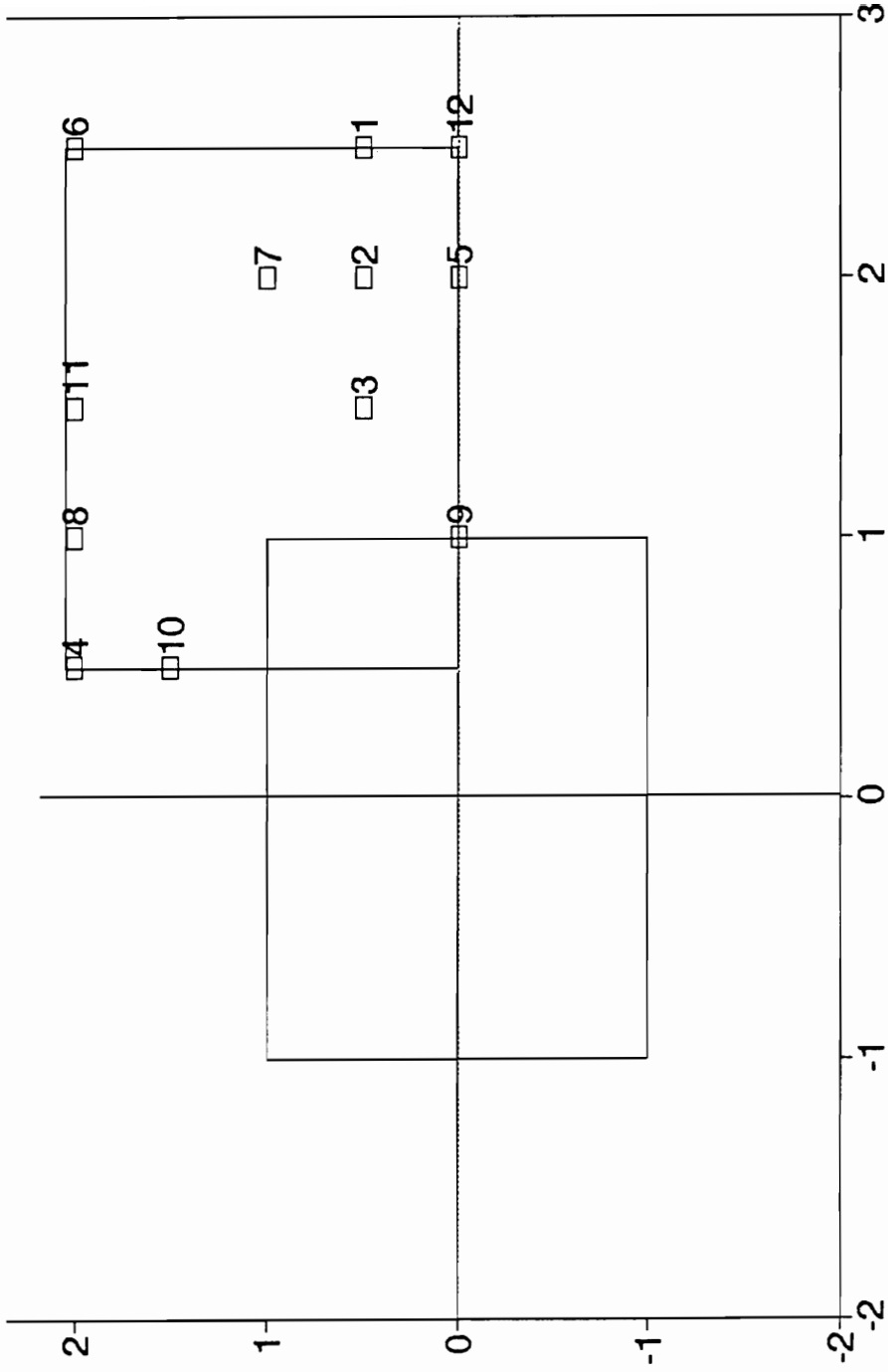


Fig. V.4.11. Twelve Points Augmented Using L.S. & BIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.2$

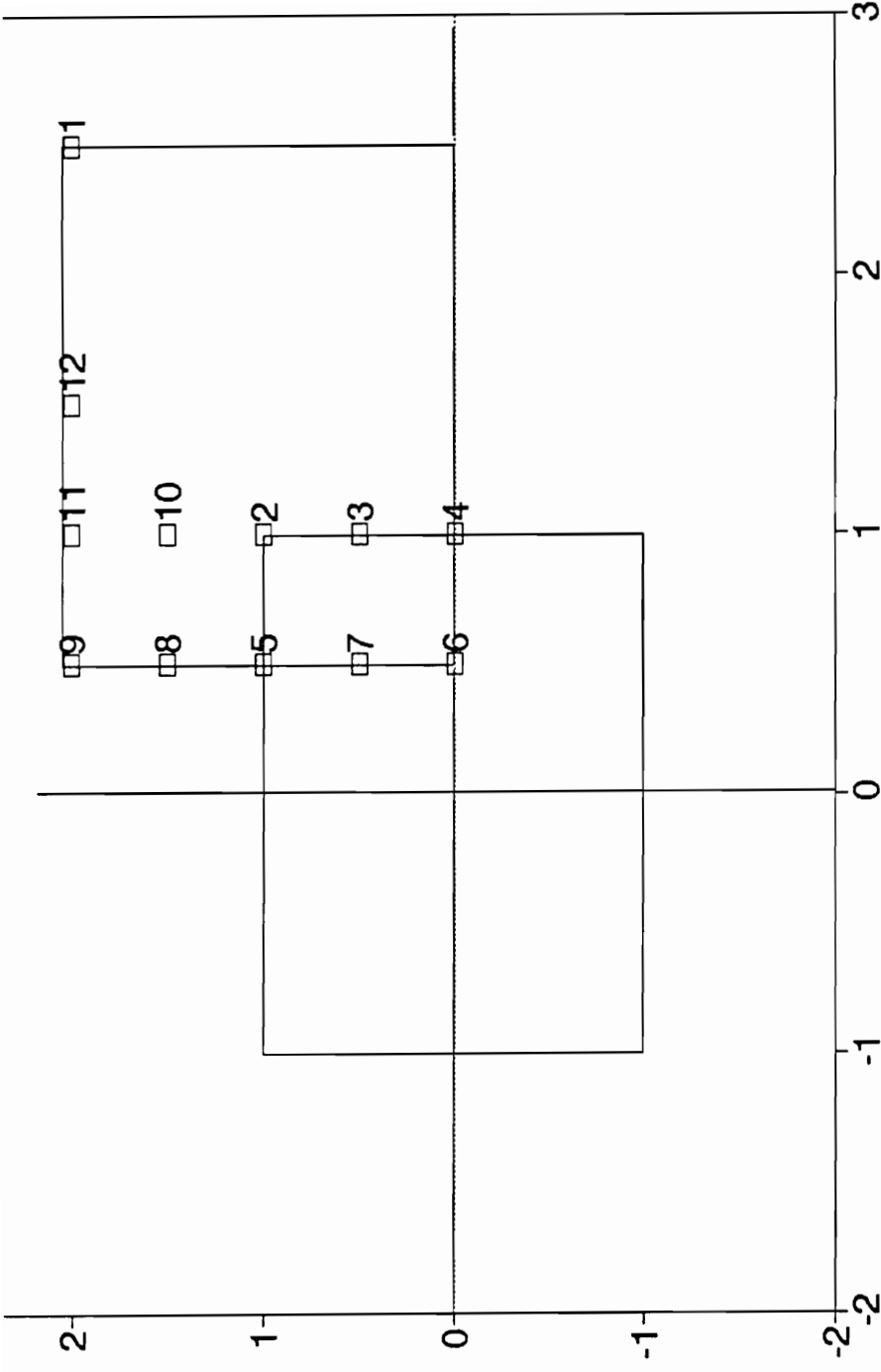


Fig. V.4.12. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with CT=1 & Sigma=0.2

**Table V.4.8. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 44.86 located at (1, 2) with  $\sigma = 0.2$  and  $CT = 1$ .

METHOD	$\overline{\text{dis.}}$	$s_{dis}$	$\overline{\text{max.}}$	$s_{max}$
HATLINK	0.64	0.22	40.83	0.117
least squares	0.50	0.00	41.19	0.092
kernel	0.65	1.00	41.27	0.062

**Table V.4.9. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 0.2$  and  $CT = 1$ .

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	696.7	647.9	2552.3	74.3	491.2	7.65
1	592.2	132.4	442.2	9.52	481.7	8.21
2	276.4	4.07	397.9	6.07	482.1	8.50
3	275.7	3.34	396.2	19.8	458.3	6.21
4	244.6	4.57	334.7	8.53	459.6	5.86
5	240.0	4.49	311.5	2.39	450.3	6.25
6	219.3	3.84	322.0	2.77	449.6	5.88
7	221.8	4.33	324.9	6.75	441.4	6.84
8	218.8	4.17	338.7	3.20	476.9	7.37
9	222.5	4.57	333.6	3.23	499.9	8.28
10	220.4	4.97	326.9	2.73	470.8	7.39
11	213.3	4.58	318.8	2.54	479.0	7.26
12	180.4	3.25	309.3	2.14	478.1	5.90



#### V.4.E CASE 7: $CT = 1$ & $\sigma = 0.5$

Now  $\sigma$  is set to 0.5. When augmenting with integrated prediction variance, this value of  $\sigma$  worked fairly well for the HATLINK method.

Plots of the twelve points in a typical design can be made for HATLINK, least squares and kernel regressions. The plot of the HATLINK design is given in Figure V.4.13, the least squares design in Figure V.4.14 and the kernel design in Figure V.4.15. For all ten repetitions, the kernel method always consistently picked the same points for augmentation.

After twelve points are augmented, the fitted maximum and distance to the true maximum are calculated for each repetition, averaged and presented in Table V.4.10. In this case, both the distance and the fitted value were about the same for all three methods. As can be seen from these values, once again the three methods tended to perform about the same in terms of estimating and locating the maximum. For all ten repetitions performed, with the augmentation of the first two points, the ESE for HATLINK dropped significantly compared to the least squares and kernel methods. Again, it appears that HATLINK worked better in estimating the overall surface. These values are presented in Table V.4.11.

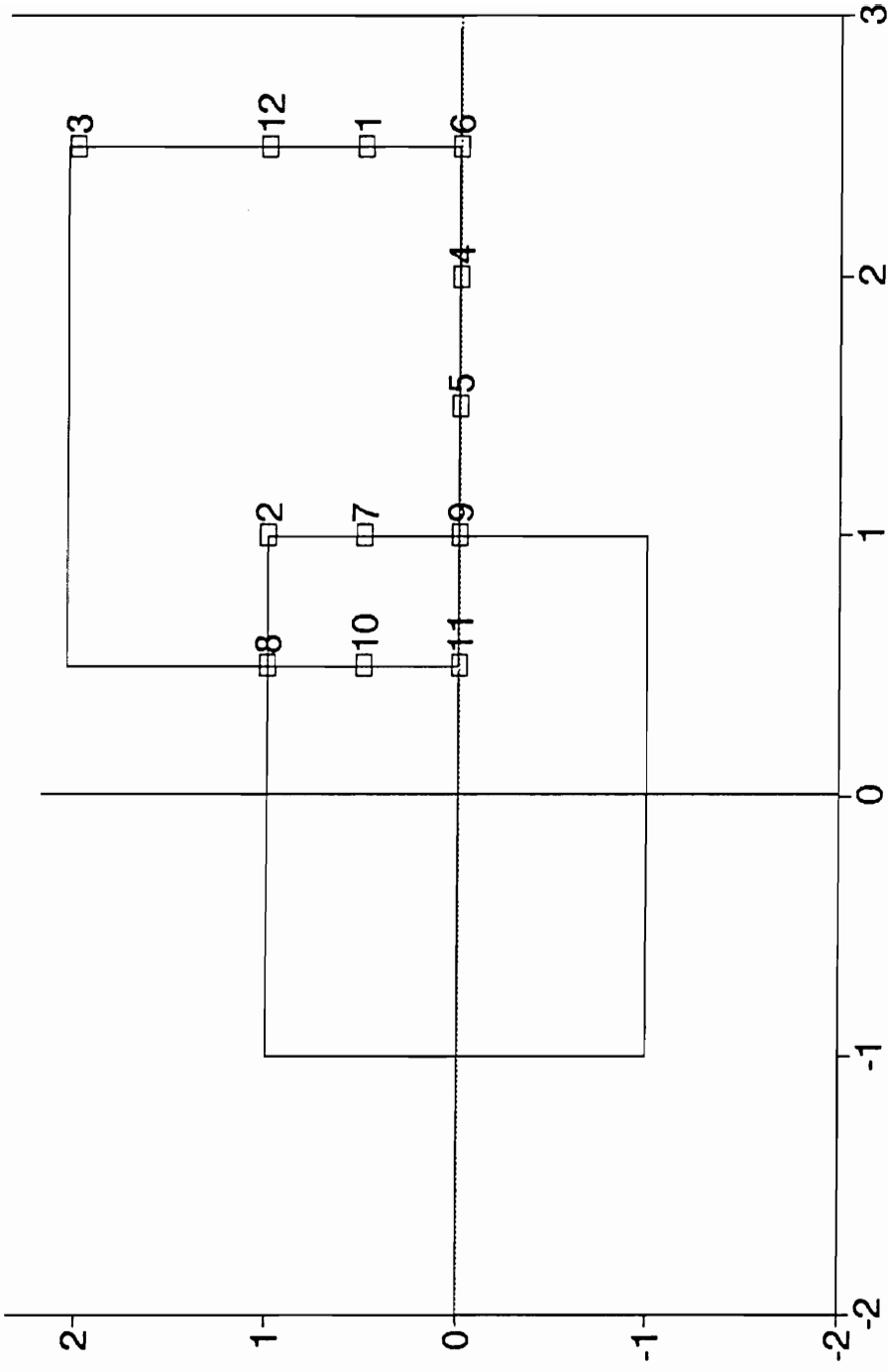


Fig. V.4.13. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with CT=1 & Sigma=0.5

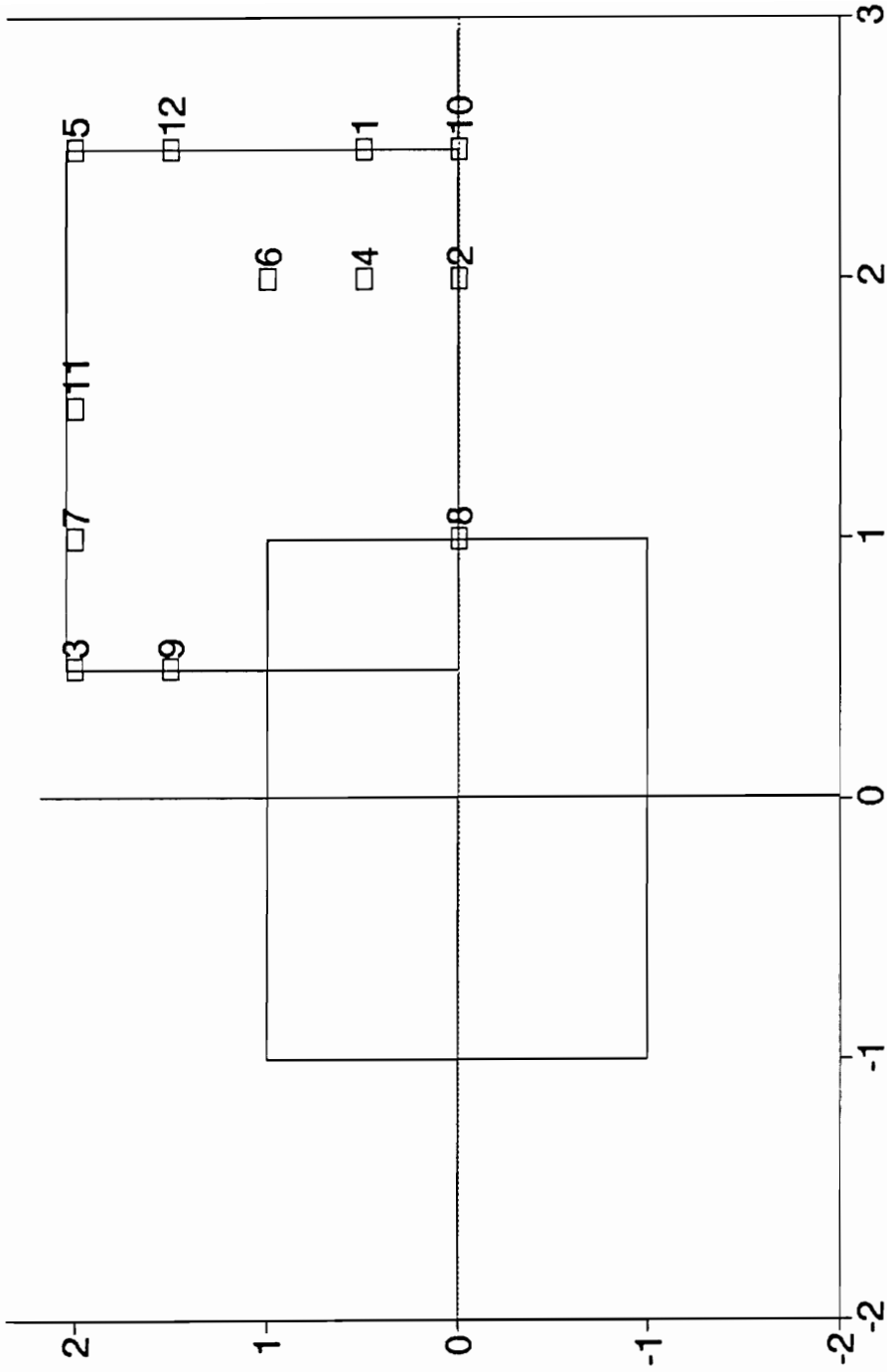


Fig. V.4.14. Twelve Points Augmented Using L.S. & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.5$

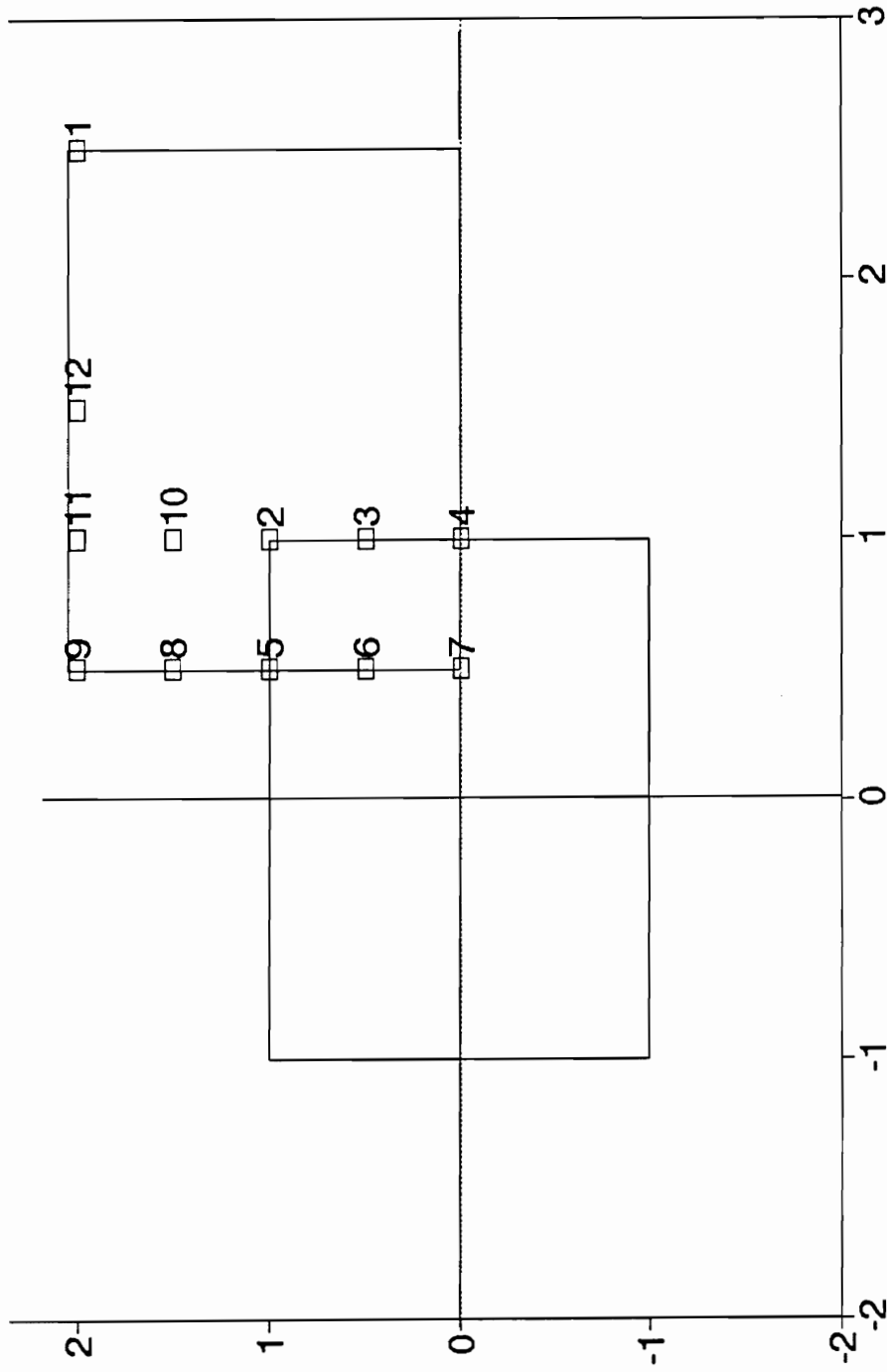


Fig. V.4.15. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=0.5$

**Table V.4.10. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 44.86 located at (1, 2) with  $\sigma = 0.5$  and  $CT = 1$ .

METHOD	$\overline{dis.}$	$s_{dis}$	$\overline{max.}$	$s_{max}$
HATLINK	0.85	0.64	41.85	0.73
least squares	0.50	0.00	41.37	0.31
kernel	0.87	1.09	41.40	0.15

**Table V.4.11. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 0.5$  and  $CT = 1$ .

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	1561.3	1146.4	2569.4	187.9	486.8	19.0
1	710.1	517.5	429.2	22.7	481.4	20.7
2	270.9	20.7	400.9	22.5	482.4	21.7
3	273.7	19.7	355.0	35.0	459.6	16.1
4	245.1	8.62	318.1	13.9	461.6	14.8
5	237.1	9.74	309.9	7.06	456.4	16.5
6	225.8	9.75	320.4	7.27	441.5	23.1
7	222.0	11.7	332.3	12.3	443.6	17.0
8	224.4	10.7	335.7	7.29	477.9	18.5
9	225.0	9.82	331.2	8.56	498.8	19.9
10	219.8	13.5	322.4	7.87	472.7	19.0
11	211.4	17.4	313.1	5.87	478.5	18.1
12	191.5	18.6	307.1	4.41	475.5	15.0

#### V.4.F CASE 8: $CT = 1$ & $\sigma = 1$

This case, with  $\sigma = 1$ , was the last case considered when performing ten repetitions. Preliminary simulations indicated that increasing  $\sigma$  even further did not affect the behavior of the HATLINK method. Remember that the smaller values of  $\sigma$  were studied in more detail because of the trouble observed when initially augmenting points by the integrated prediction variance method.

Plots of the twelve points for a typical design can be made for HATLINK, least squares and kernel regressions. The plot of the HATLINK design is given in Figure V.4.16, the least squares design in Figure V.4.17 and the kernel design in Figure V.4.18.

After twelve points are augmented, the averaged fitted maximum and average distance to the true maximum are calculated and summarized in Table V.4.12. In this case, both the distance and the fitted value were about the same for all three methods. As can be seen from these values, in terms of estimating and locating the maximum, once again, the three methods tended to perform about the same. With the augmentation of the first point, the ESE values for HATLINK always dropped significantly compared to the least squares and kernel methods. Again, it appears that HATLINK worked better in estimating the overall surface. These values are presented in Table V.4.13.

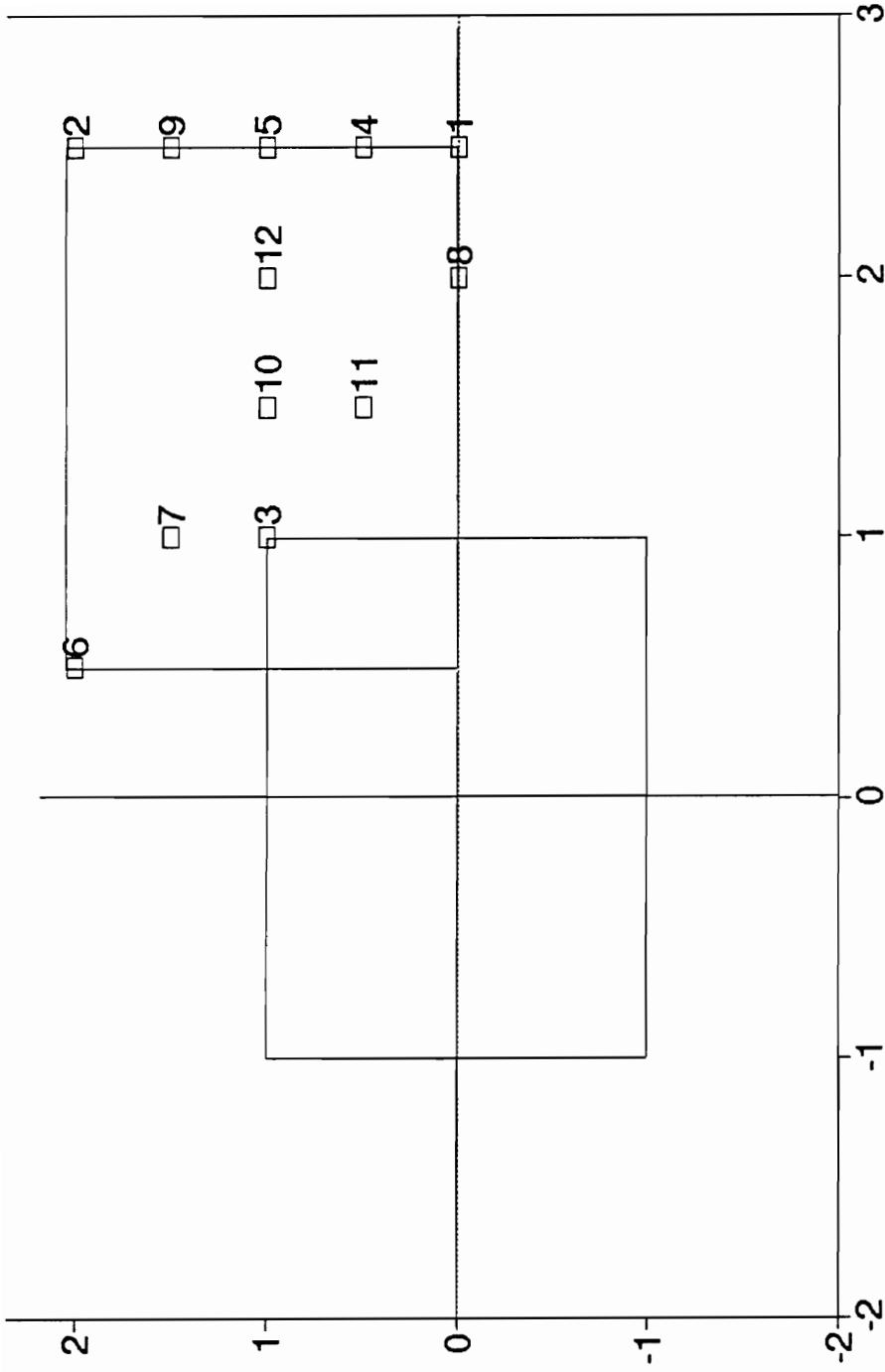


Fig. V.4.16. Twelve Points Augmented Using HATLINK & BIIV, Typical Repetition with  $CT=1$  &  $\text{Sigma}=1$

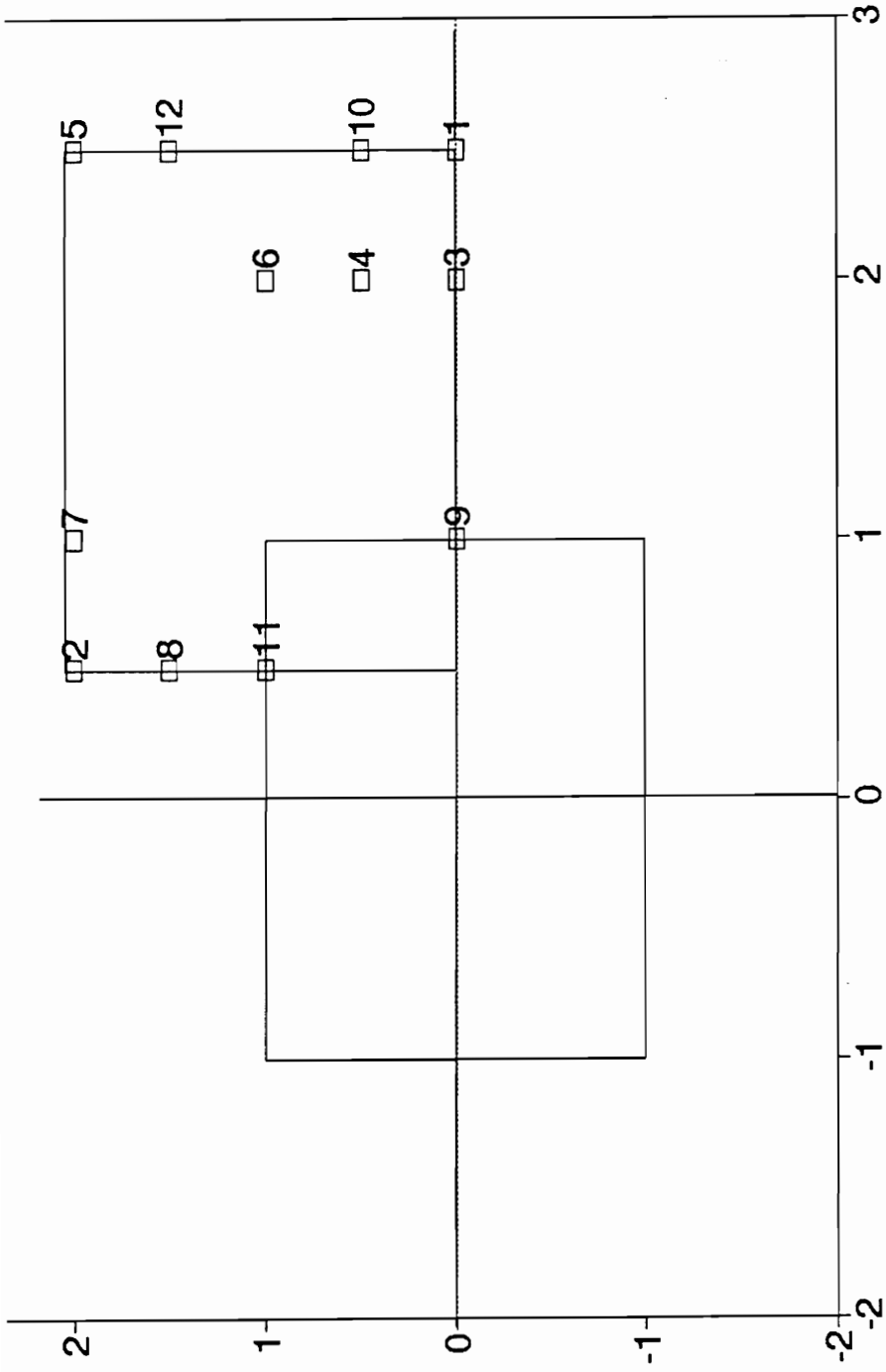


Fig. V.4.17. Twelve Points Augmented Using L.S. & BIV, Typical Repetition with  $CT=1$  &  $\Sigma=1$



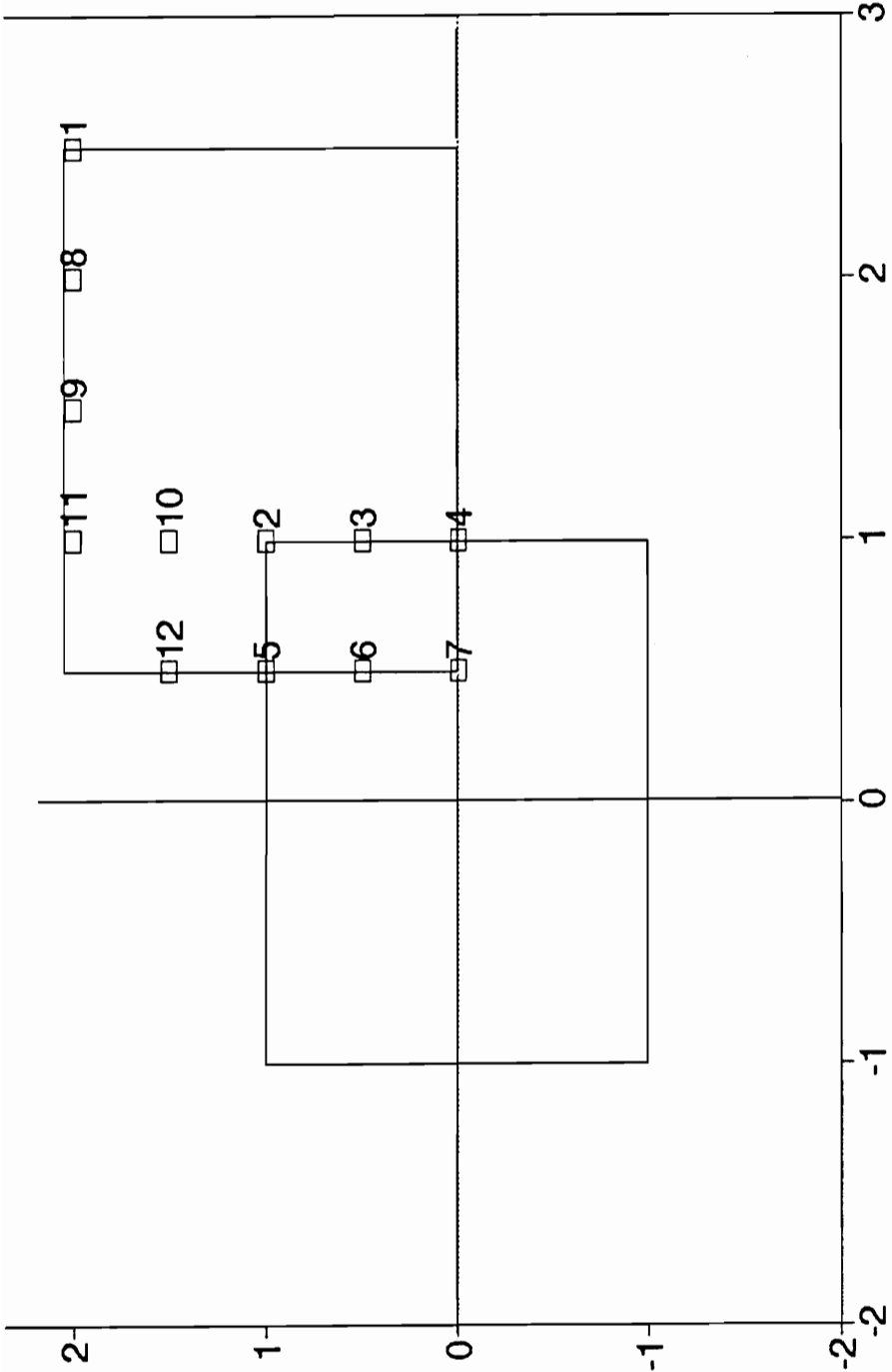


Fig. V.4.18. Twelve Points Augmented Using KERNEL & BIIV, Typical Repetition with  $CT=1$  &  $Sigma=1$

**Table V.4.12. Average Distance Between the Fitted and True Maxima, the Average Fitted Maximum and the Corresponding Standard Deviations.** These values are calculated for the ten repetitions after twelve points have been augmented. The true maximum is 44.86 located at (1, 2) with  $\sigma = 1$  and CT = 1.

METHOD	$\overline{\text{dis.}}$	$s_{dis}$	$\overline{\text{max.}}$	$s_{max}$
HATLINK	0.68	0.67	41.23	0.82
least squares	0.45	0.16	41.50	0.55
kernel	0.65	1.00	41.72	0.43

**Table V.4.13. Average ESE values after Augmentation and the Corresponding Standard Deviations of ESE Values.** These values are calculated for each point augmented for all three methods with  $\sigma = 1$  and CT = 1.

NUM	HATLINK	$s_{hat}$	LEAST SQ.	$s_{ls}$	KERNEL	$s_{ker}$
0	2533.2	472.4	2608.4	383.1	483.5	37.4
1	248.0	30.8	425.9	42.0	481.9	42.4
2	247.7	30.0	399.3	34.7	483.6	44.8
3	251.1	35.0	347.5	38.7	458.3	33.6
4	245.7	19.5	323.8	16.4	463.4	30.9
5	234.9	20.7	316.8	13.1	465.0	30.1
6	237.7	17.5	325.2	13.4	467.0	36.4
7	230.8	23.9	340.1	19.6	464.5	40.7
8	229.9	22.0	336.5	14.4	487.8	46.6
9	217.7	27.3	331.2	15.5	491.9	35.5
10	222.7	31.1	322.6	14.8	474.2	28.3
11	215.9	24.2	313.7	9.92	476.2	17.6
12	212.8	23.4	312.8	10.7	460.9	22.7

## V.5 SUMMARY OF RESULTS FROM INITIAL SIMULATIONS AND REPETITIONS

This chapter explored the BIIV method of HATLINK augmentation in order to determine the optimal technique for generating the new design, additional methods are discussed further in Appendix 3. To estimate the mixing parameter, several different methods were proposed in Chapter III and detailed further in Appendix 3. It was concluded that PRESS\* and BIIV yielded the best results in general, although some potential problems were mentioned. The effect of batch size was also studied in Section V.3. It was concluded that with the augmentation of enough batches of points, the effect of batch size appears to be minimized.

The eight cases involving repetitions studied in Section V.4 are typical cases and all had the same basic outcome. As stated above, the cases with the smaller  $\sigma$  values were considered because of the perceived trouble with the HATLINK method when augmenting points by integrated prediction variance. The HATLINK method, augmenting with BIIV, always tended to outperform both the least squares and kernel methods. This behavior was consistently seen for all ten repetitions in each case studied. For the cases where  $CT = 0.1$ , HATLINK and least squares were usually the same. With more model misspecification, after the first point or two augmented, HATLINK always did better than both least squares and kernel methods, at least in terms of the ESE values. The performance of the three methods in terms of locating and estimating the fitted maximum tended to be approximately the same, although the specific values varied somewhat from case to case.

These cases illustrate the consistency of the proposed method of augmentation from run to run, indicating that the procedure is not overly variable when the random errors change. These runs also indicate that while the augmented design may vary depending upon the random errors, the design should not vary too much. The fitted maximum and its location, after

augmenting by the desired number of points, should not change too much either. It further appears that using BIIV to augment points removes some of the trouble with smaller values of  $\sigma$  observed when augmenting points using only integrated prediction variance. This is illustrated in more detail in Appendix 3.

These cases illustrate the comparative stability of the HATLINK method, despite its stochastic nature. All of the cases studied here further illustrate the desirability of using the proposed HATLINK method of augmentation.

## CHAPTER VI

# THE HATLINK PROCEDURE IN MULTIPLE DIMENSIONS

### VI.1 INTRODUCTION

The HATLINK method proposed in this dissertation, using the PRESS\* criterion for selecting  $\lambda$  and BIIV for augmenting points is very promising. To use the method as proposed, recall that both old and new regions of interest must be determined as well as an initial point in the new region and the desired batch size. The user must also propose a parametric model in the new region of interest. In this stage of investigation of HATLINK, several two dimensional examples will be studied with the intention of observing the behavior of the HATLINK, parametric, and nonparametric methods of augmentation using BIIV in examples where the augmentation procedure is carried out completely. In such a case, not only is one augmenting with a set number of points in some specified batch size, but after augmentation the location of the maximum of the fitted surface will be determined and the overall fit considered. In the two dimensional examples, plots of both the fitted surfaces and contour plots are presented.

In the previous chapters, a set number of points were augmented by HATLINK, least squares, and kernel regression, using some specified batch size. The overall fits as measured by the ESE values were then compared, and one method was determined to be superior. It is hoped that the results given by the HATLINK fitted surface outperforms those of both least squares

and kernel regressions. For the examples in two dimensions, plots of the true surface, HATLINK, least squares and kernel fitted surfaces are presented for selected quantities of augmented points. These examples illustrate how the HATLINK method can draw from both the least squares and kernel regressions to produce a better fit than either technique alone is able to produce.

In response surface methodology, after points are augmented in the new region, not only is the fitted maximum important, but it is also desirable to determine if there is a range of operating conditions that would result in approximately the maximum value. Many times, the experimenter desires to know this range of operating values for greater flexibility in setting the design variables to achieve response values close to the true maximum. For this reason, the performance of the overall fitted surface may also be important to the researcher.

## **VI.2 FIRST EXAMPLE IN TWO DIMENSIONS**

This example takes the augmentation example initially presented in Section IV.3 and used in the previous chapters and proceeds to find the fitted maxima and present plots as described in the introduction. Both old and new regions of interest as well as an initial point in the new region of interest are the same as before. Using the same equation (IV.3.1) with  $\sigma = 1$  and  $CT = 0.7$  for the true model, twelve points will be augmented using a batch size of one. The true maximum is 42.71 and is located at (1.25, 1.60). This location is compared to the individual fitted maxima by finding the distance between the two points. After each point is augmented, the fitted maximum for the HATLINK, least squares and kernel surfaces and the distance between the fitted and true values are calculated. Unlike the previous chapter where ten repetitions were studied, only one run is performed and the plots and graphs generated will be

examined to graphically illustrate the influence of both least squares and kernel regressions upon the HATLINK method.

### VI.2.1 NUMERICAL COMPARISONS OF FIRST EXAMPLE

Augmentation of twelve points yielded the following results which are presented as follows: HATLINK in Table VI.2.1, a second order least squares model in Table VI.2.2, and kernel regression in Table VI.2.3. For the HATLINK and least squares methods, plots of the first six points augmented were presented in Figure IV.3.1 and Figure IV.3.2. Similar plots can be made for both the extra six points augmented and for the points augmented by the kernel method. Notice the ESE values, the fitted maxima and the distances to the true maximum. In terms of finding the true maximum, all three methods performed about equally well, although the kernel method did slightly better, but not enough to be very meaningful.

The distances are the same also, although the locations differed slightly. After the twelfth point was augmented, HATLINK determined the location of the maximum to be at (1.0, 1.0), while both the least squares and kernel methods placed the location at (1.5, 1.0). When computing and comparing distances, a grid of points is established and the fitted maximum is the location resulting in the largest prediction value. Thus, depending upon the fineness of the grid, there might possibly be some nonzero minimum distance between the fitted and true maxima. In such cases, it may be impossible for the fitted surface to attain the location of the true maximum value. In this example, all three methods appear to perform approximately the same in terms of minimizing the distance between the fitted and true maxima.

Notice the behavior of the ESE values in the tables. After twelve points are augmented,

HATLINK has a better overall fit than either the least squares or kernel methods. Even though all three techniques performed approximately equally well in terms of finding the fitted maximum and its location in relation to the true maximum, the HATLINK surface would be the best one to use in terms of overall performance. This could be important if one desires to find the optimal range of operating conditions as described earlier. Pictures and contour plots illustrate that the HATLINK method does indeed outperform both least squares and kernel regressions. Comparison of plots from the three methods reveal that, as conjectured, HATLINK draws from both least squares and kernel methods to produce a better overall fit.



**Table VI.2.1. Twelve Points Augmented Using HATLINK and BIIV for Example 1.** Using  $\sigma = 1$  and  $CT = 0.7$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAM</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	—	—	.507	0.00	1615.71	—	—
1	2.5000	0.0000	.525	1.00	118.39	39.09	1.13
2	2.5000	2.0000	.517	1.00	120.38	39.10	1.13
3	1.0000	1.0000	.519	1.00	110.29	39.79	.65
4	2.5000	0.5000	.515	.937	139.43	39.79	.47
5	2.5000	1.0000	.583	.625	127.55	39.94	.47
6	0.5000	2.0000	.591	.875	129.85	39.69	.65
7	0.5000	1.5000	.583	.875	132.85	39.65	.65
8	2.0000	0.5000	.564	.875	123.78	39.29	.65
9	2.0000	0.0000	.563	.937	121.62	39.32	.65
10	1.5000	0.5000	.563	.937	124.55	39.14	.65
11	2.5000	1.5000	.584	.937	121.15	39.04	.65
12	1.5000	1.0000	.590	.937	115.28	39.13	.65

**Table VI.2.2. Twelve Points Augmented Using Least Squares and BIIV for Example 1.** Using  $\sigma = 1$  and  $CT = 0.7$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	—	—	1615.71	—	—
1	2.5000	0.0000	188.73	39.20	1.33
2	2.5000	0.5000	175.52	40.21	.85
3	0.5000	2.0000	183.86	38.49	.65
4	2.0000	0.5000	191.56	38.16	.65
5	2.5000	2.0000	159.08	38.73	.27
6	2.0000	1.0000	157.84	38.84	.65
7	1.0000	2.0000	158.74	39.21	.27
8	0.5000	1.5000	159.22	38.95	.27
9	1.0000	0.0000	155.94	39.19	.65
10	2.0000	0.0000	150.30	39.19	.27
11	0.5000	1.0000	149.02	39.33	.27
12	1.5000	2.0000	148.69	39.10	.65

**Table VI.2.3. Twelve Points Augmented Using Kernel and BIIV for Example 1.** Using  $\sigma = 1$  and  $CT = 0.7$ , this lists the points augmented, the bandwidth values selected and the ESE values. The bandwidths and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	—	—	.507	179.31	—	—
1	2.5000	2.0000	.503	177.98	39.16	1.13
2	1.0000	1.0000	.506	173.39	39.79	.65
3	1.0000	0.5000	.519	169.45	39.74	.65
4	1.0000	0.0000	.523	178.84	39.75	.65
5	0.5000	1.0000	.512	183.27	39.75	.65
6	2.0000	2.0000	.518	223.27	39.69	.65
7	1.5000	2.0000	.566	224.69	39.64	1.13
8	1.0000	1.5000	.529	210.46	40.39	.27
9	1.0000	2.0000	.507	206.13	40.64	.85
10	0.5000	1.5000	.527	188.63	39.95	.27
11	0.5000	2.0000	.538	180.63	39.84	.65
12	0.5000	0.5000	.531	178.29	39.83	.65

## VI.2.2 COMPARISON OF PLOTS OF FITTED SURFACES FOR THE FIRST EXAMPLE

The first plot presented is the three dimensional plot of the true response surface. Next, for each of the HATLINK, least squares, and kernel methods, three plots of the fit are presented. The first of these plots is the fitted surface with the initial data in the new region using the first estimate of the bandwidth and lambda parameters. The second plot is the estimated curve after six points have been augmented and the bandwidth and lambda parameters have been updated. The third plot is similar to the second plot except that twelve points have been augmented. All ten plots are presented in Figure VI.2.1 through Figure VI.2.10.

As expected, based upon the ESE values all three methods perform quite poorly when only the initial data is available. This is seen by comparing Figures VI.2.2 – VI.2.4 to the true surface represented by Figure VI.2.1. While the plots for both HATLINK and kernel regression indicate some curvature, they do not capture much of the behavior of the true response surface. The plot from the least squares method, while bearing very little resemblance to the HATLINK and kernel fitted surfaces, also looks quite poor. This plot also captures very little of the behavior of the true response surface. Notice that while the true function attains a maximum of about 43, while the HATLINK and kernel fitted surfaces attains one around 38, the least squares fitted maximum is approximately 55 and is located very far from the true value. This further indicates that in this case, least squares is not estimating the function even as well as kernel regression. As expected, with no additional augmented points all three methods have substantial difficulty estimating the true surface.

When six points have been augmented with batch size of one, noticeable changes occur in both the HATLINK and least squares surfaces. The kernel surface has changed very little from the initial fit, but this is not unexpected, given the relatively constant ESE values

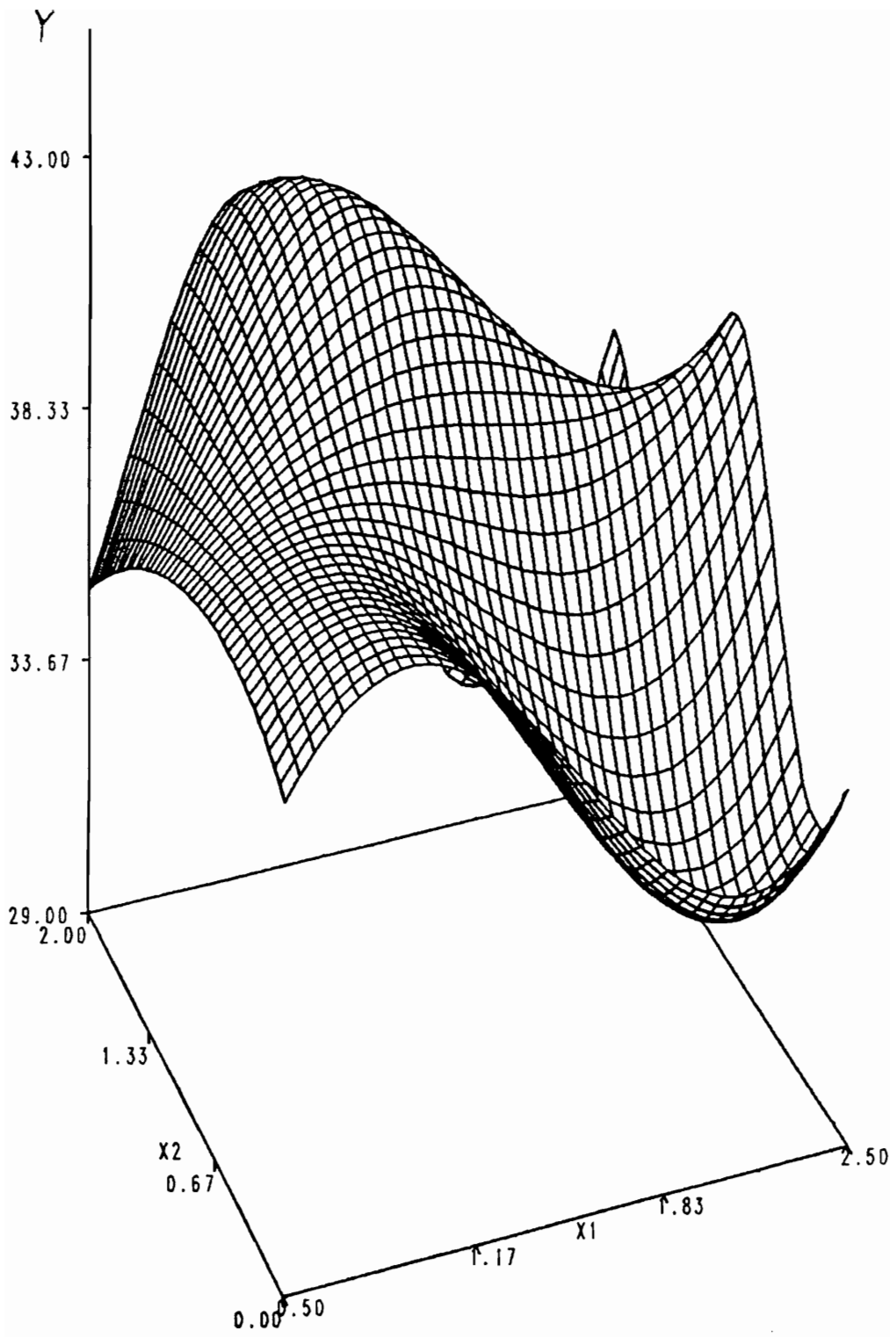


FIG. VI.2.1. TRUE RESPONSE SURFACE IN NEW REGION  
126

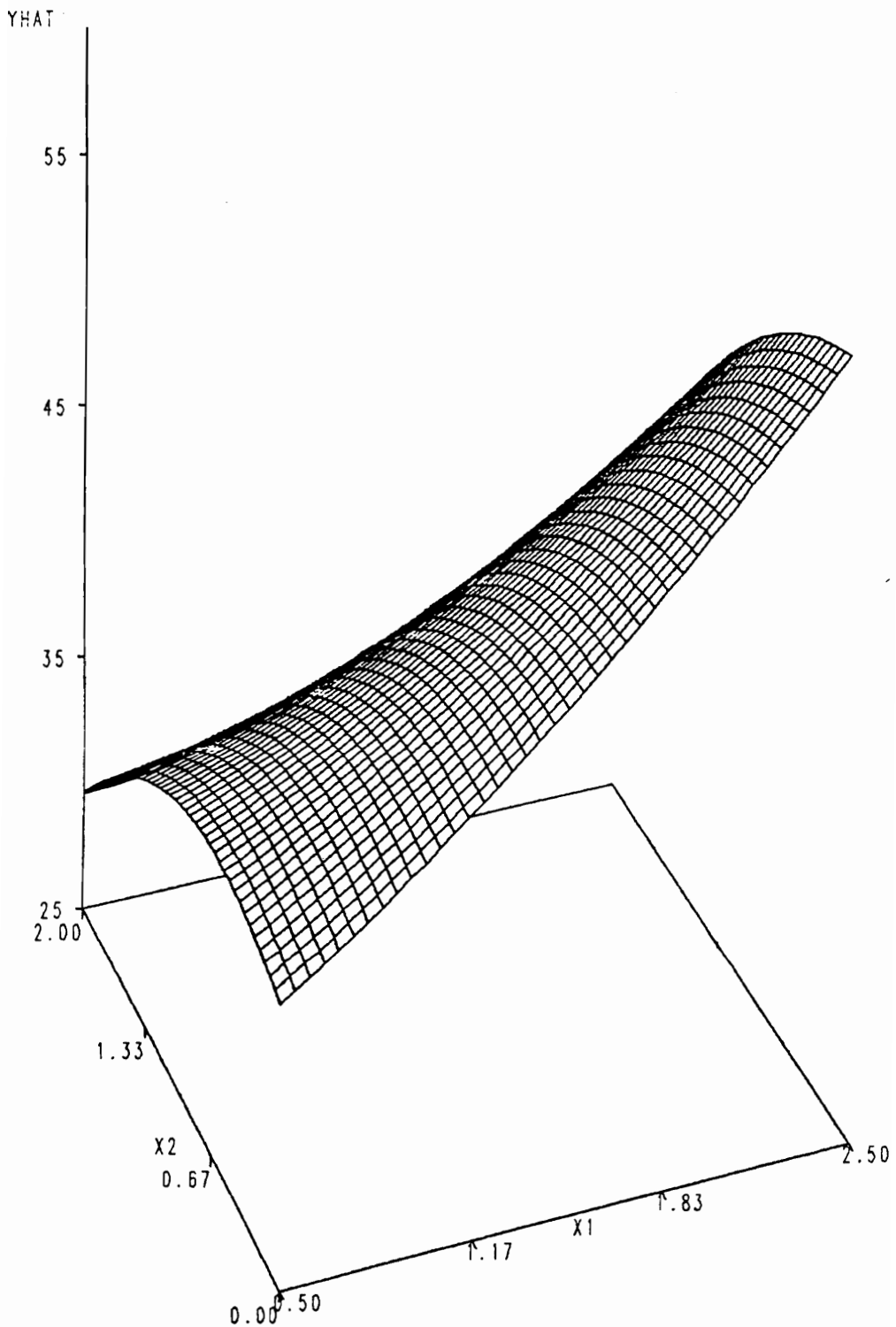


FIG. VI.2.2. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 INITIAL DATA FOR EXAMPLE 1.  $H=.507$  &  $LAM=0$

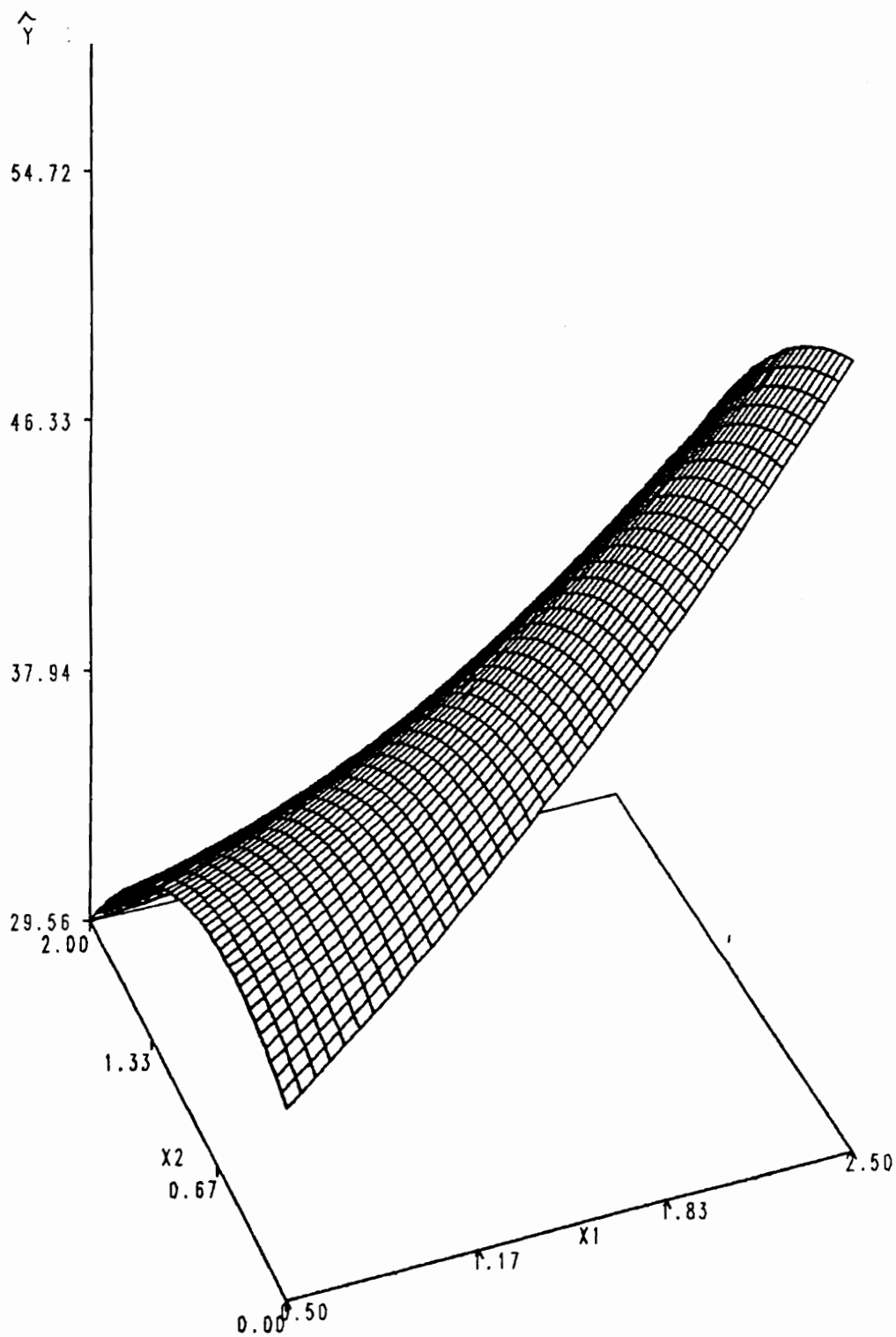


FIG. VI.2.3. FITTED 2ND ORDER L.S. MODEL  
INITIAL DATA FOR EXAMPLE 1

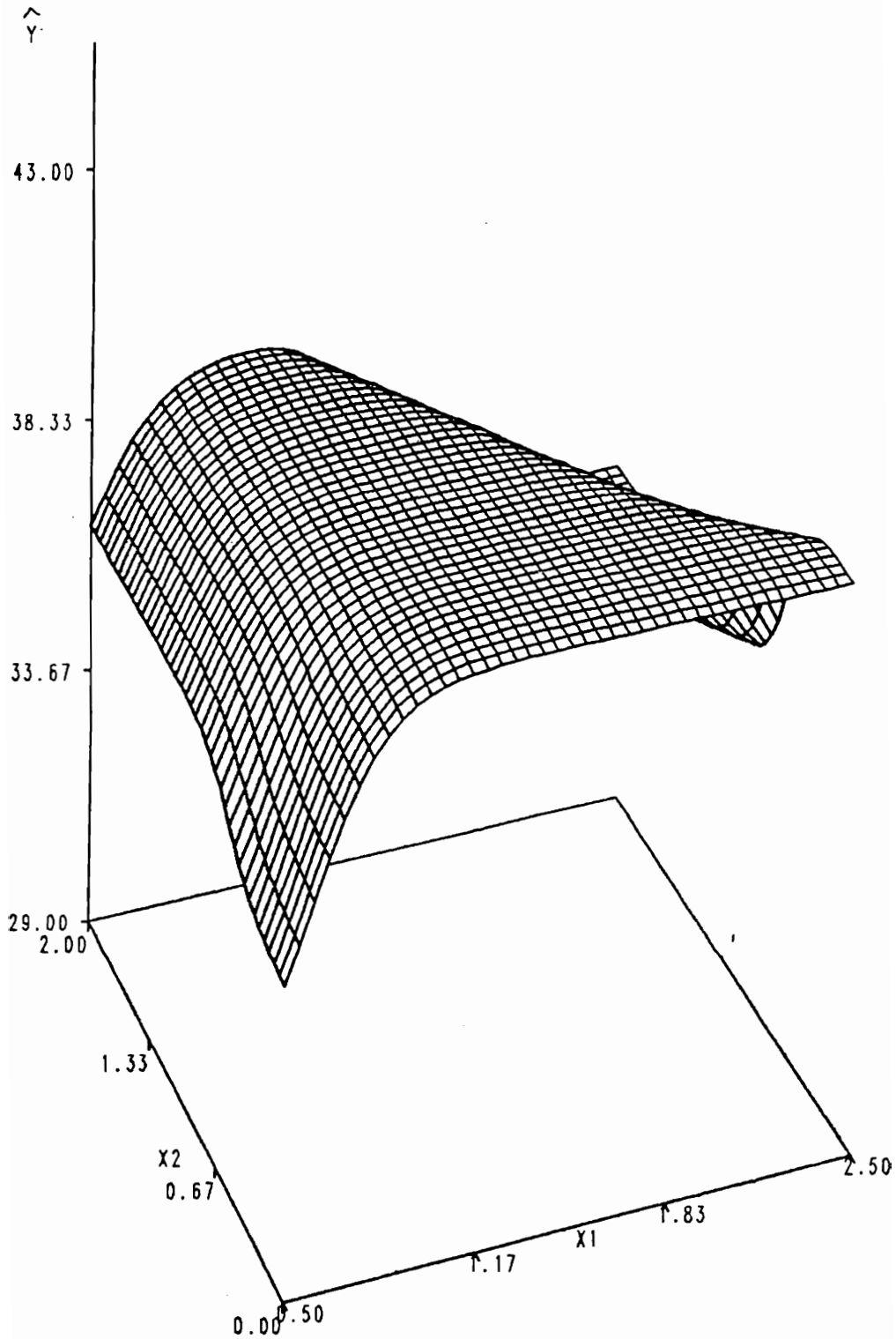


FIG. VI.2.4. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
INITIAL DATA FOR EXAMPLE 1.  $H=.507$



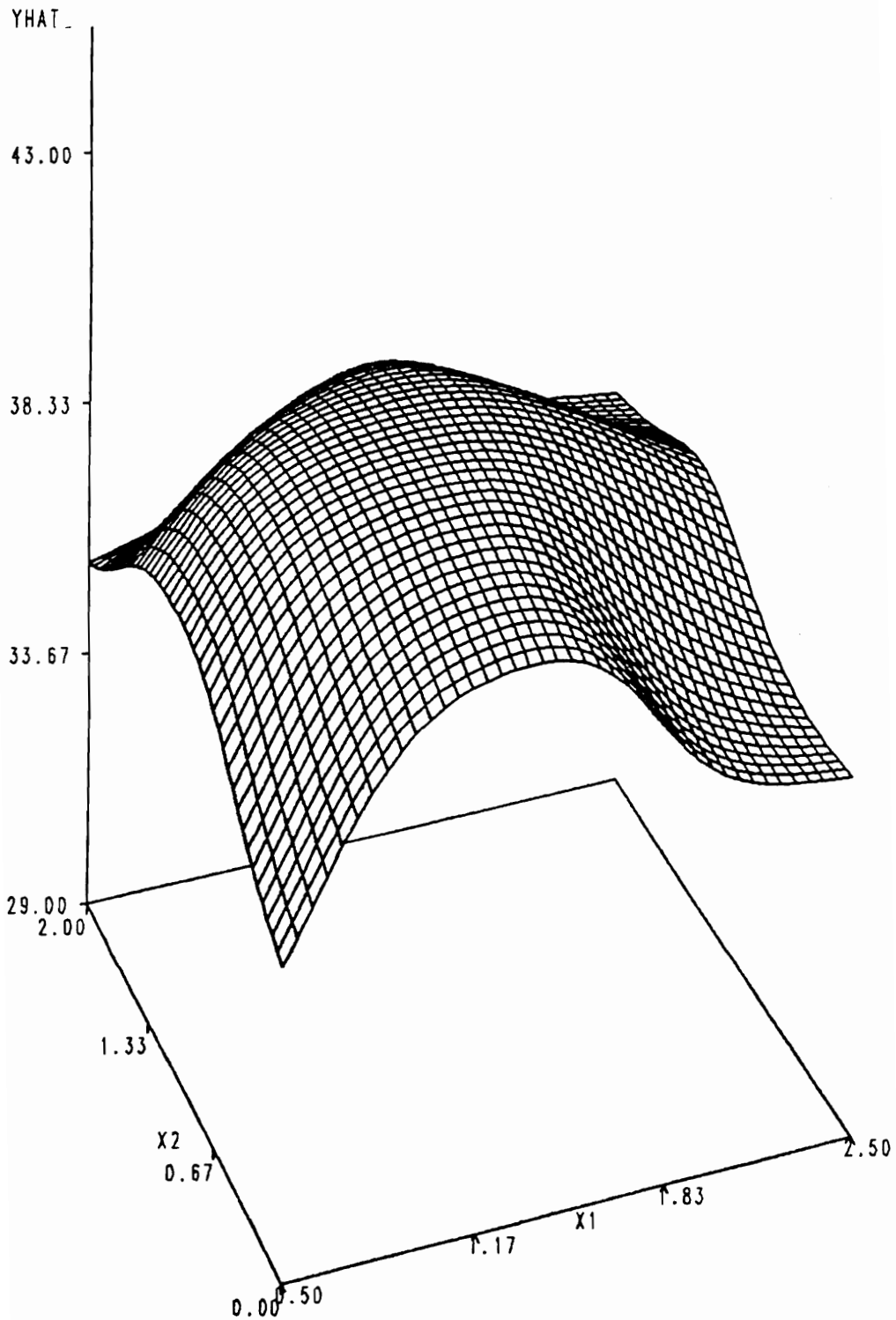


FIG. VI.2.5. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 1.  $H=.591$  &  $LAM=.875$

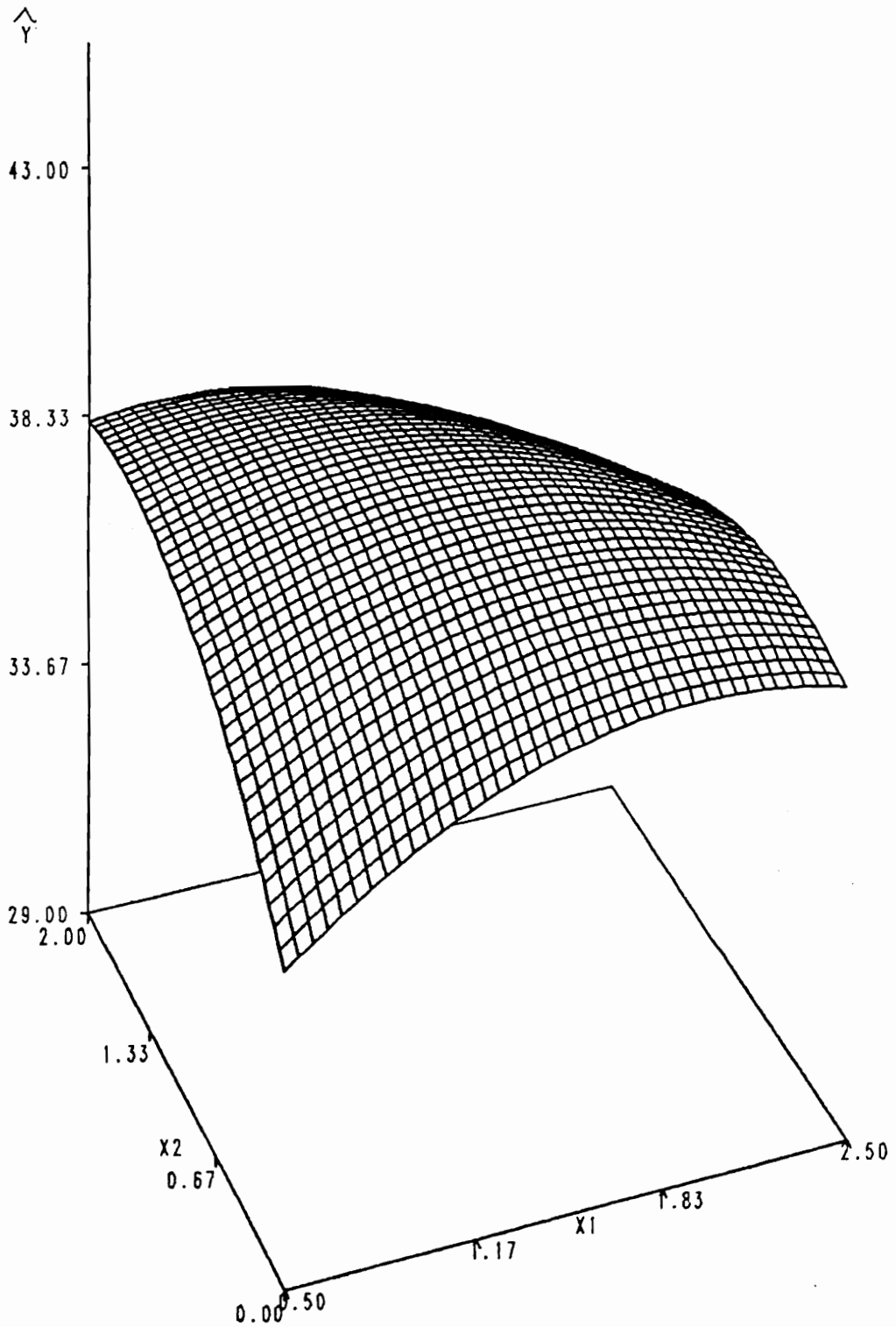


FIG. VI.2.6. FITTED 2ND ORDER L.S. MODEL  
SIX POINTS AUGMENTED FOR EXAMPLE 1

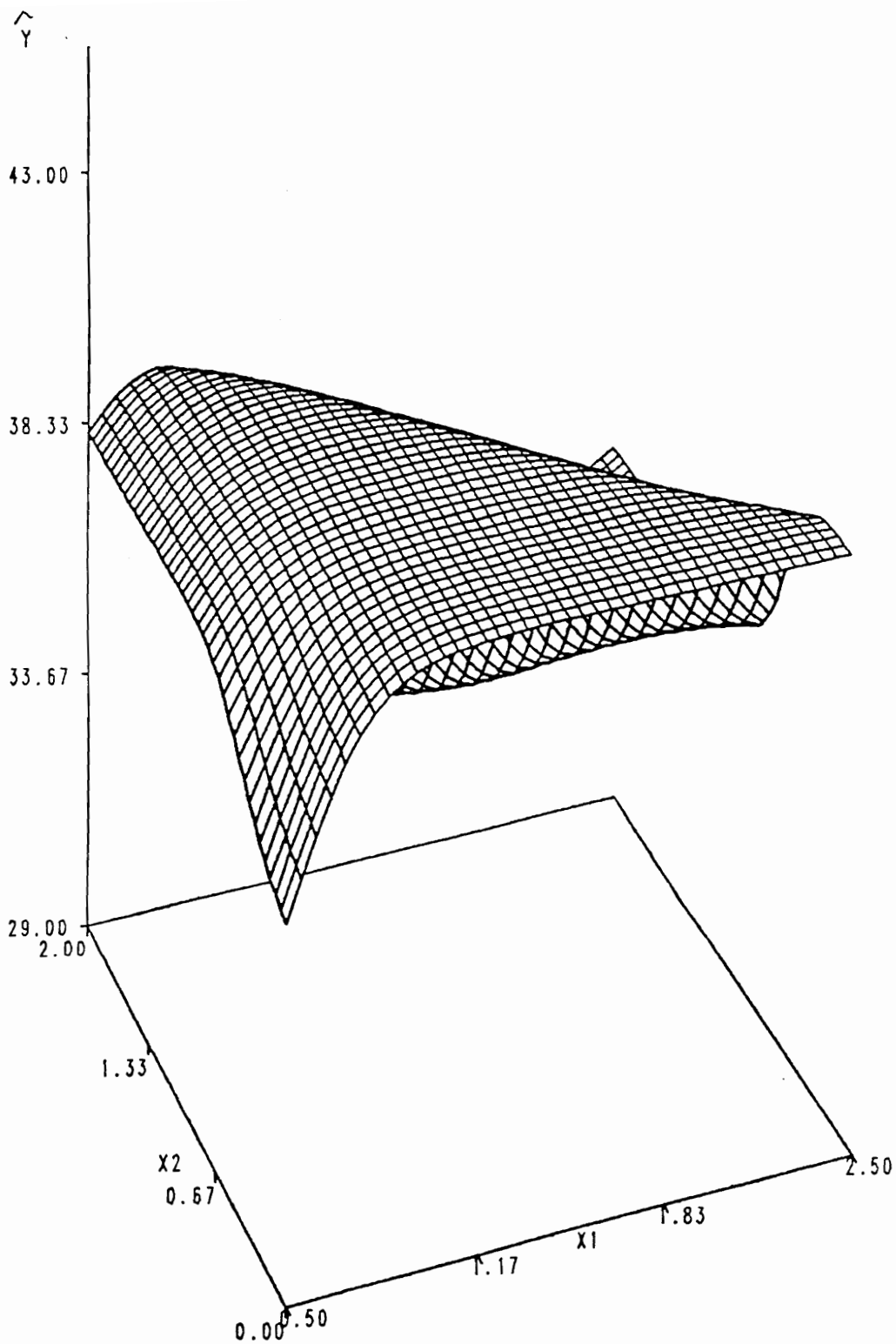


FIG. VI.2.7. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 1.  $H=.518$

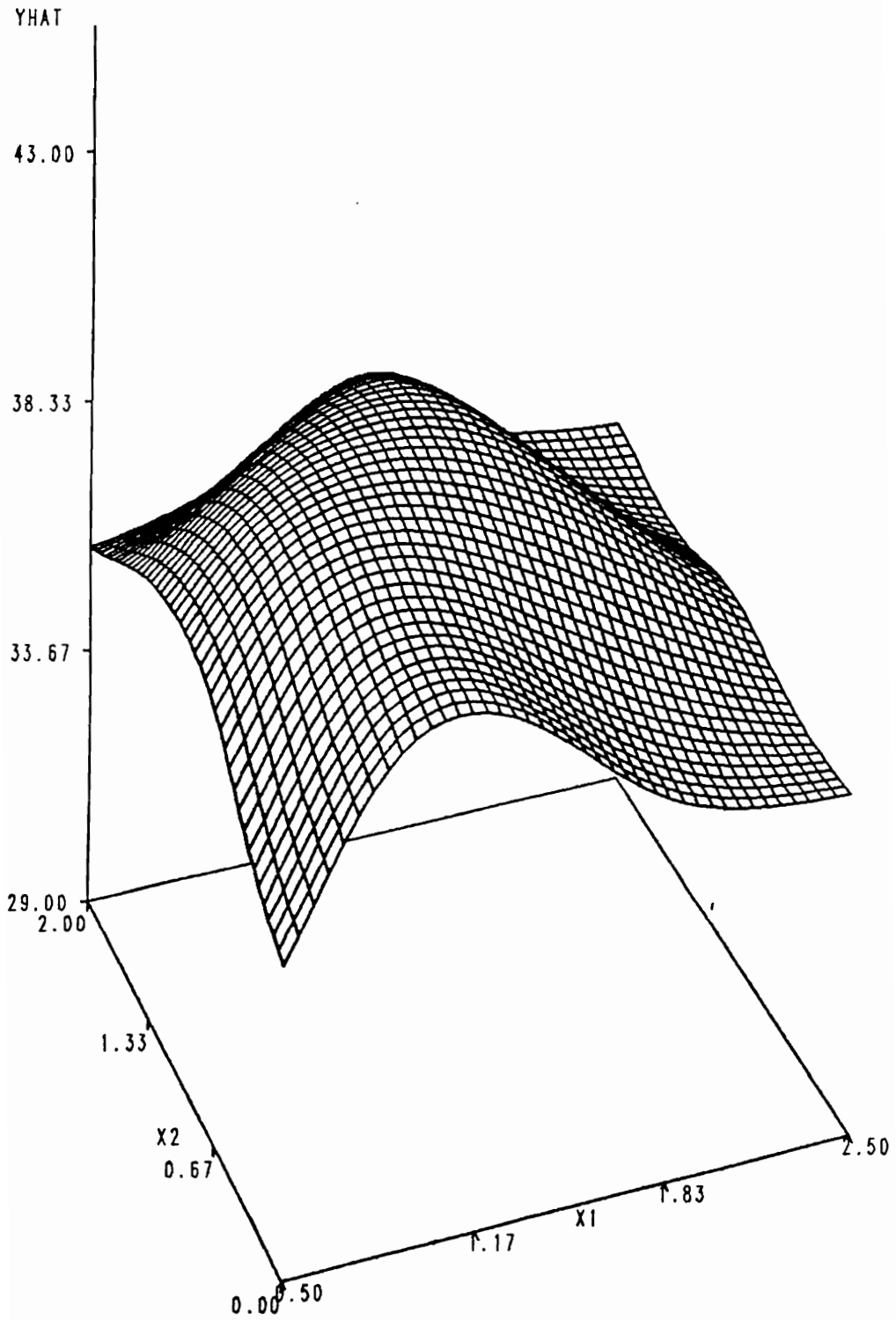


FIG. VI.2.8. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1.  $H=.590$  &  $LAM=.937$

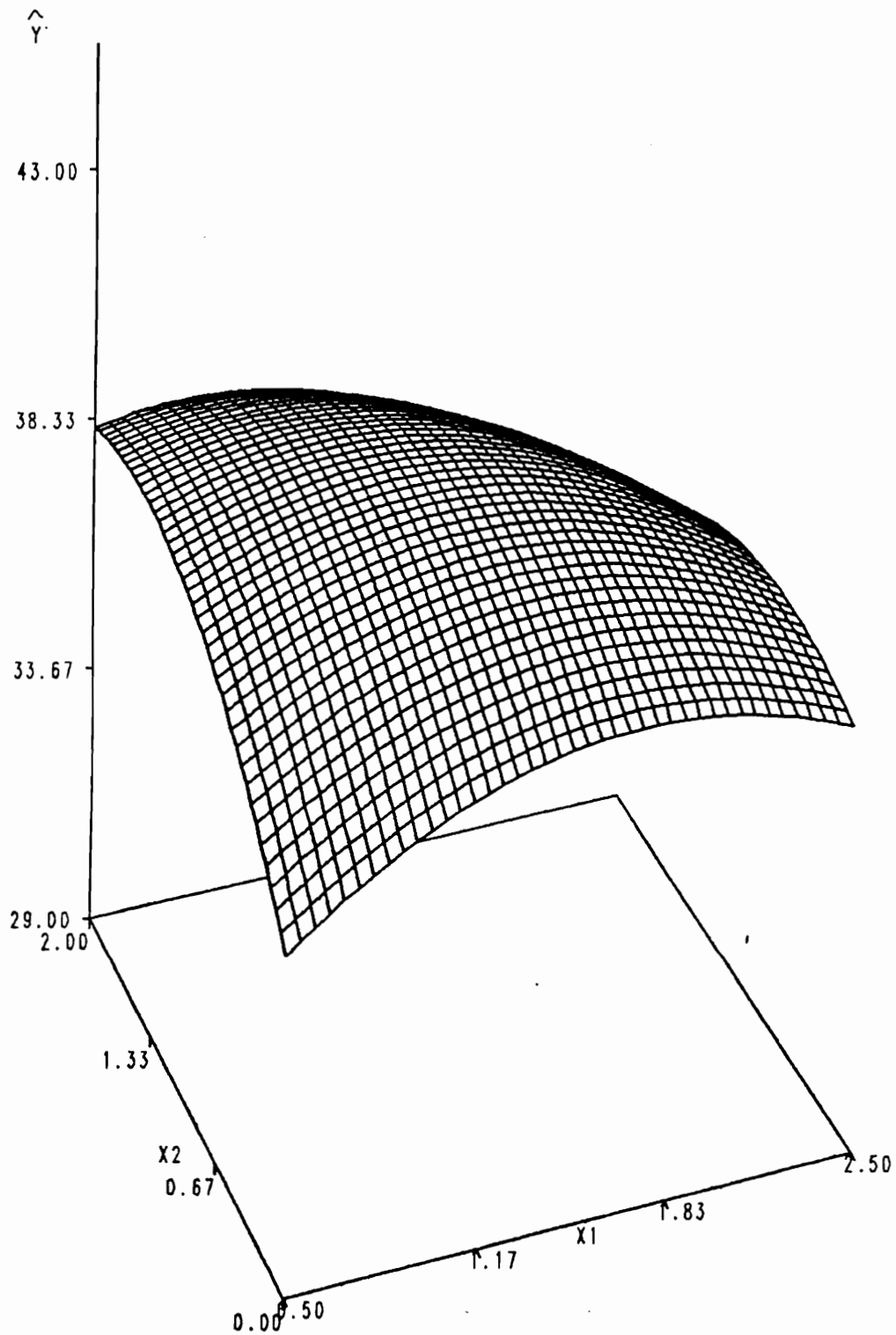


FIG. VI.2.9. FITTED 2ND ORDER L.S. MODEL  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1

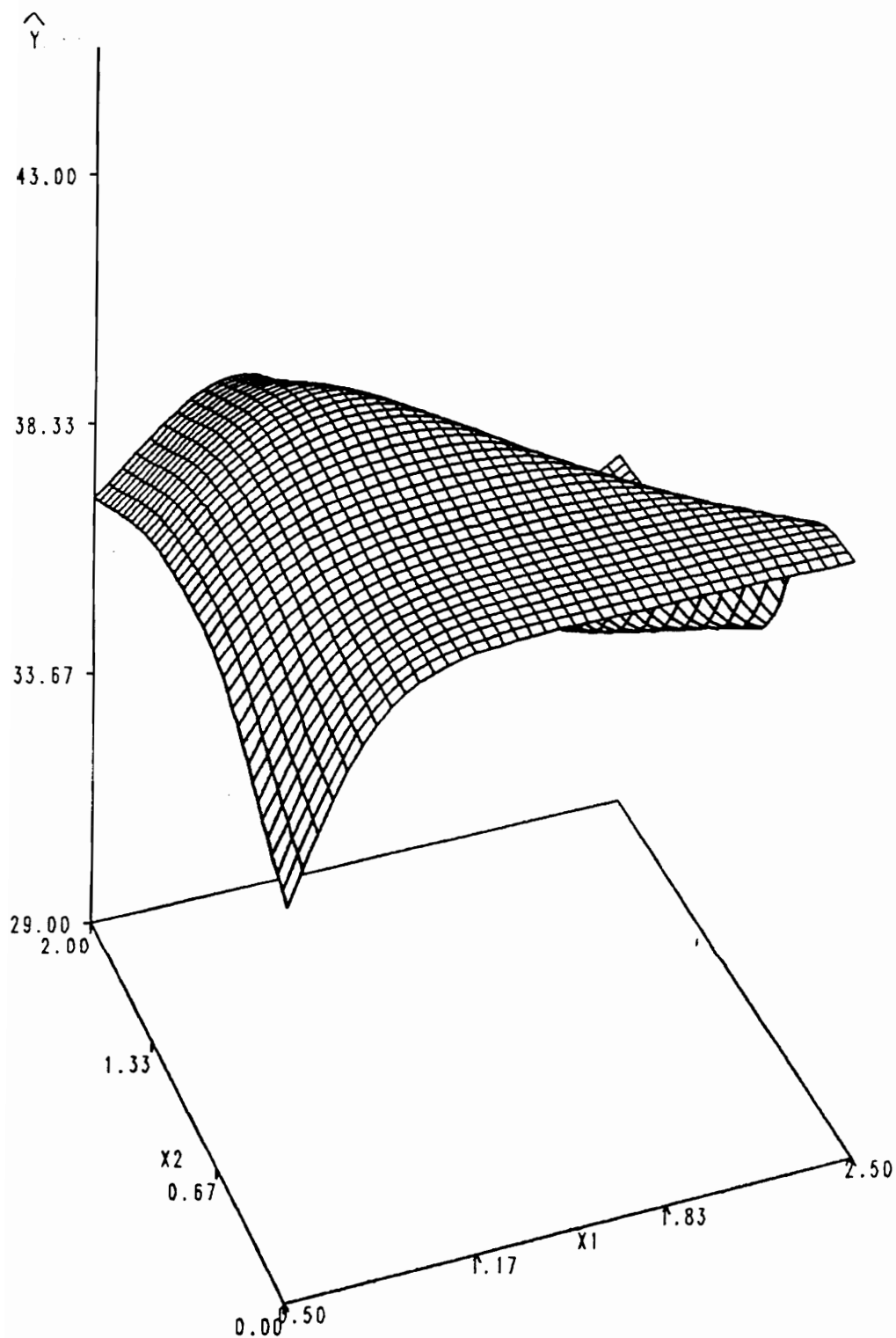


FIG. VI.2.10. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1.  $H=.531$

generated when augmenting with the six points. The least squares fit has changed substantially from its initial shape. It still does not capture the true response surface very well, but at least it does not attain as large a maximum value as before. Now the fitted maximum is closer to 38 and the other two methods.

The most interesting surface is that generated by the HATLINK technique. With the augmentation of six points, the HATLINK surface is now starting to detect some features of the true curve, indicating that it should attain some kind of minimum around the point (2.5, 0). The kernel surface was not locating this minimum, although least squares did appear to drop a little in this region. The true surface also attains another minimum in the region around (2, 2) and HATLINK visually appears to be detecting this value, which is obviously contributed by the kernel portion. This surface is also starting to emphasize the "ridge" that occurs approximately where  $x_2 = 1.6$ . By comparing all four plots, one can see that when compared to least squares and kernel regressions, the HATLINK surface comes closest to capturing the overall behavior of the true equation, and is drawing on both least squares and kernel regressions to attain a better fit than either method alone is capable of doing. Such a behavior is not unexpected, and is the reason why the HATLINK method was originally proposed.

It was speculated that the HATLINK surface would improve with additional points augmented. Now, for this example, consider what happens when twelve points are augmented. Once again, the kernel surface does not appear to have changed very much from the fit generated by either the initial data or the first set of six points augmented. Again, note the consistency of the ESE values for the kernel method. Judging from the ESE values for least squares, the fitted surface is capturing a little bit more of the true curve, although careful comparisons of the fit with six and twelve points augmented reveal that the most substantial difference appears to be around the point (2.5, 2).

HATLINK has continued to significantly improve upon how well it captures the behavior of the true surface. The main peak around the true maximum is becoming more prominent, as is the “ridge” that occurs approximately where  $x_2 = 1.6$ , and the minima occurring at all four corners is becoming a little more pronounced in the HATLINK surface. Also, the behavior along the edges of the new region is starting to more closely resemble the true curve. Notice that even after the twelve points have been augmented, neither least squares nor kernel regressions capture these behaviors as well as HATLINK. By comparing the three types of regression, once again one can see contributions from both types of regression to the HATLINK surface.

The largest fitted values and corresponding distances to the true maximum for the three methods are approximately equal. From the plots of the fitted surfaces after augmenting by twelve points, one can see that HATLINK captures the behavior of the underlying function best. This can be very important to the researcher, since many times one is interested in finding both the maximum value and some range of operating conditions. Thus HATLINK is recommended over either least squares or kernel regressions since it appears to have the capability of drawing on both types of regression to produce a better overall fit than either technique alone can produce.

### **VI.2.3 COMPARISON OF CONTOUR PLOTS FOR THE FIRST EXAMPLE**

In the previous section, plots of the various fitted surfaces were compared for the three different methods. Although some sections of the graphs were hard to see, it was clear that the HATLINK method was producing a fit superior to either that from least squares or kernel regressions. As expected, contour plots for the same conditions reveal a similar story. With a

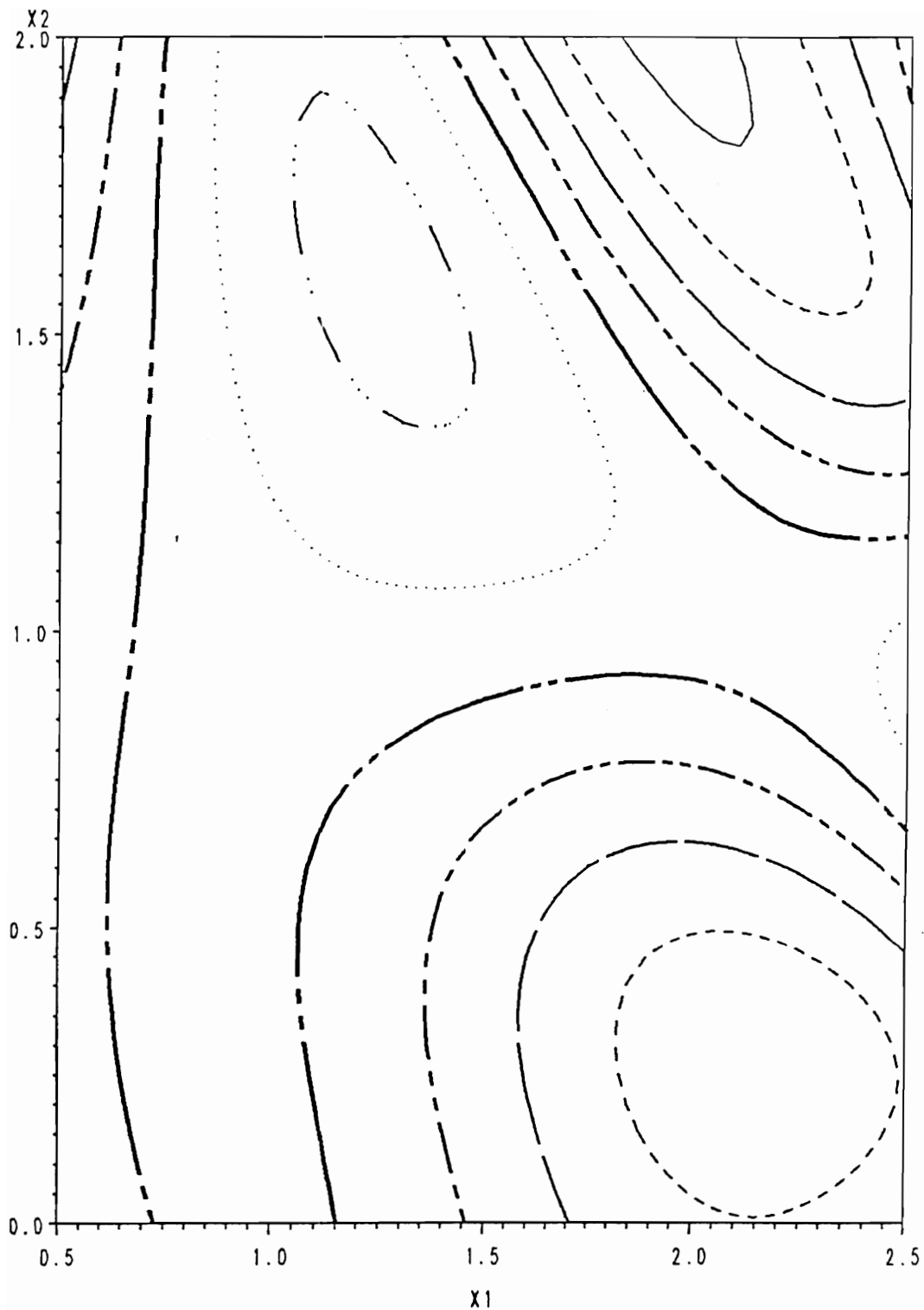


contour plot of the true equation seen in Figure VI.2.11, one can see more clearly the maximum at the point (1.25, 1.6) and the two minima around the points (2, 2) and (2.2, 0.25). The contour plots, Figure VI.2.11 to Figure VI.2.20, give one a different perspective on how well the various fitted surfaces are approximating the true surface.

As expected from Section VI.2.2, all three methods perform quite poorly with only the initial data. No method is doing very well here as seen by comparing Figure VI.2.11 – VI.2.14 to Figure VI.2.11.

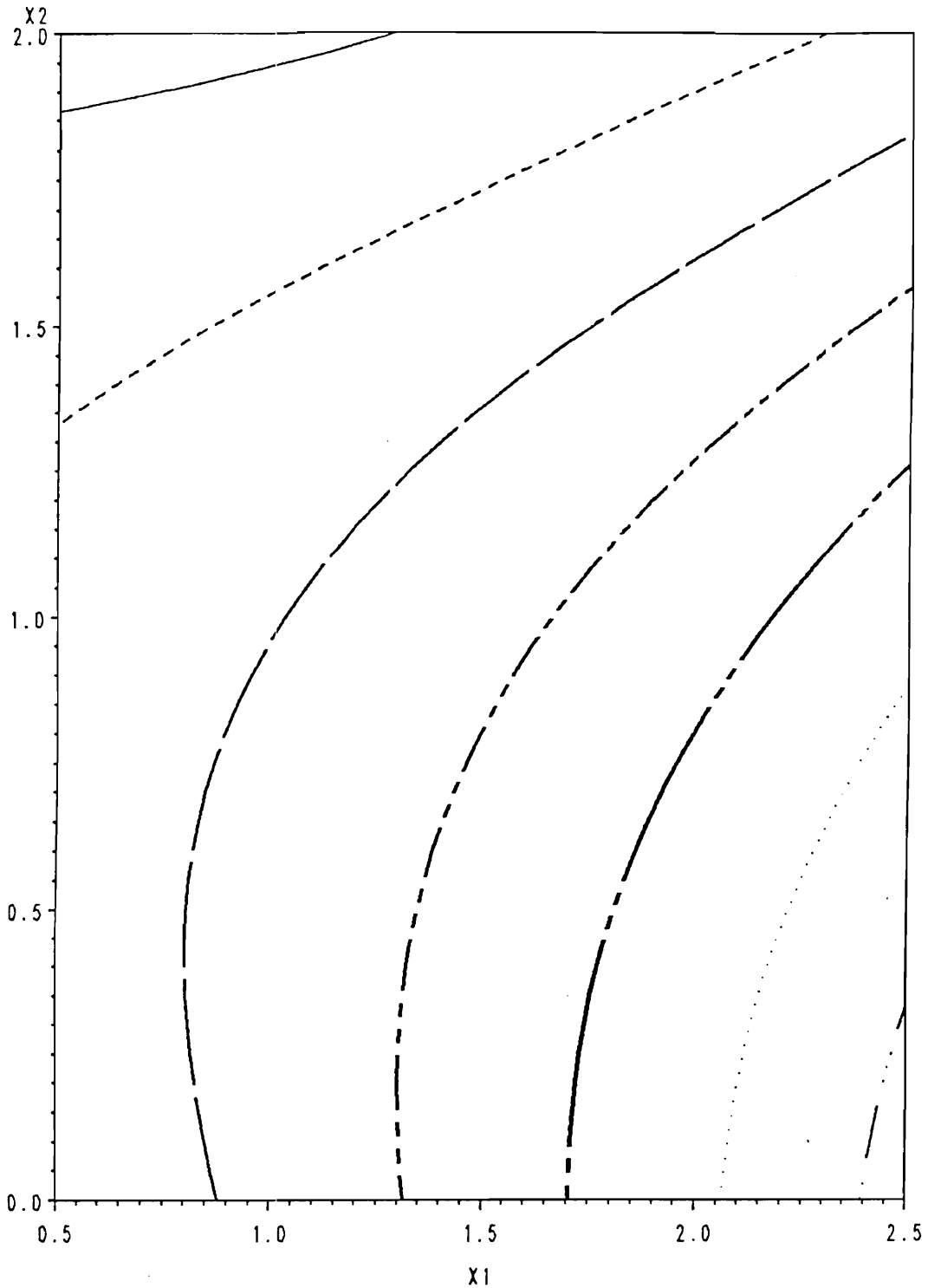
With the augmentation of six points, things begin to change radically. As expected, the kernel surface is not capturing very well the locations of the maximum and the various minima. The fit from least squares is capturing the maximum, but nothing pertaining to the two smallest values. HATLINK is not only finding the location with the largest value, but is also starting to detect the two minima occurring where  $x_1$  is approximately 2. These points are seen in Figures VI.2.15 – VI.2.17.

After augmenting with twelve points, the contour plots change yet again, see Figures VI.2.18 – VI.2.20. Comparison of those from least squares with six and twelve points augmented reveal very little difference. Notice the difference in the kernel contour plots. With twelve points augmented, not only is the maximum becoming more noticeable, but the minimum at the point (2, 2) is starting to appear. These features also appear in the HATLINK contour plot. One very interesting aspect to notice is the behavior of the various plots in the region around the second minimum at (2.2, 0.25). The HATLINK surface is detecting this somewhat in the sense that the fit is decreasing in this area. Compare this to the contour plots for least squares and kernel regressions. In this case, least squares is indicating a steady decline from the maximum in the fitted response values, with no indication of a distinct minimum occurring. The kernel contour plot is indicating a very flat surface with no changes in either



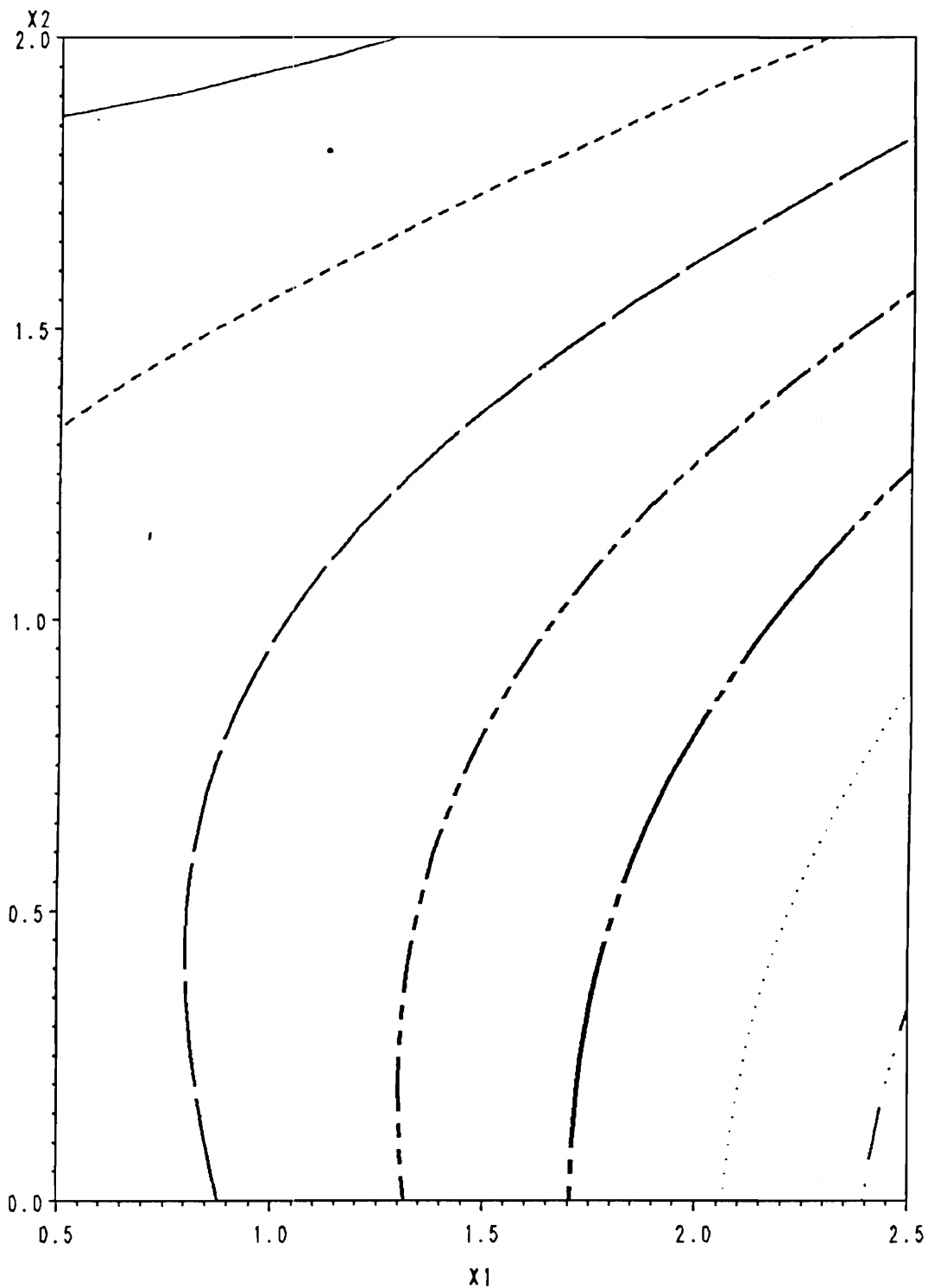
32.10    
 33.77    
 35.45    
 37.12  
 38.80    
 40.48    
 42.15

FIG. VI.2.11. TRUE RESPONSE SURFACE FOR EXAMPLE 1



YHAT:    ——— 30.82    - - - - 34.59    ——— 38.36    - - - - 42.14  
          - - - - 45.91    ······ 49.69    ······ 53.46

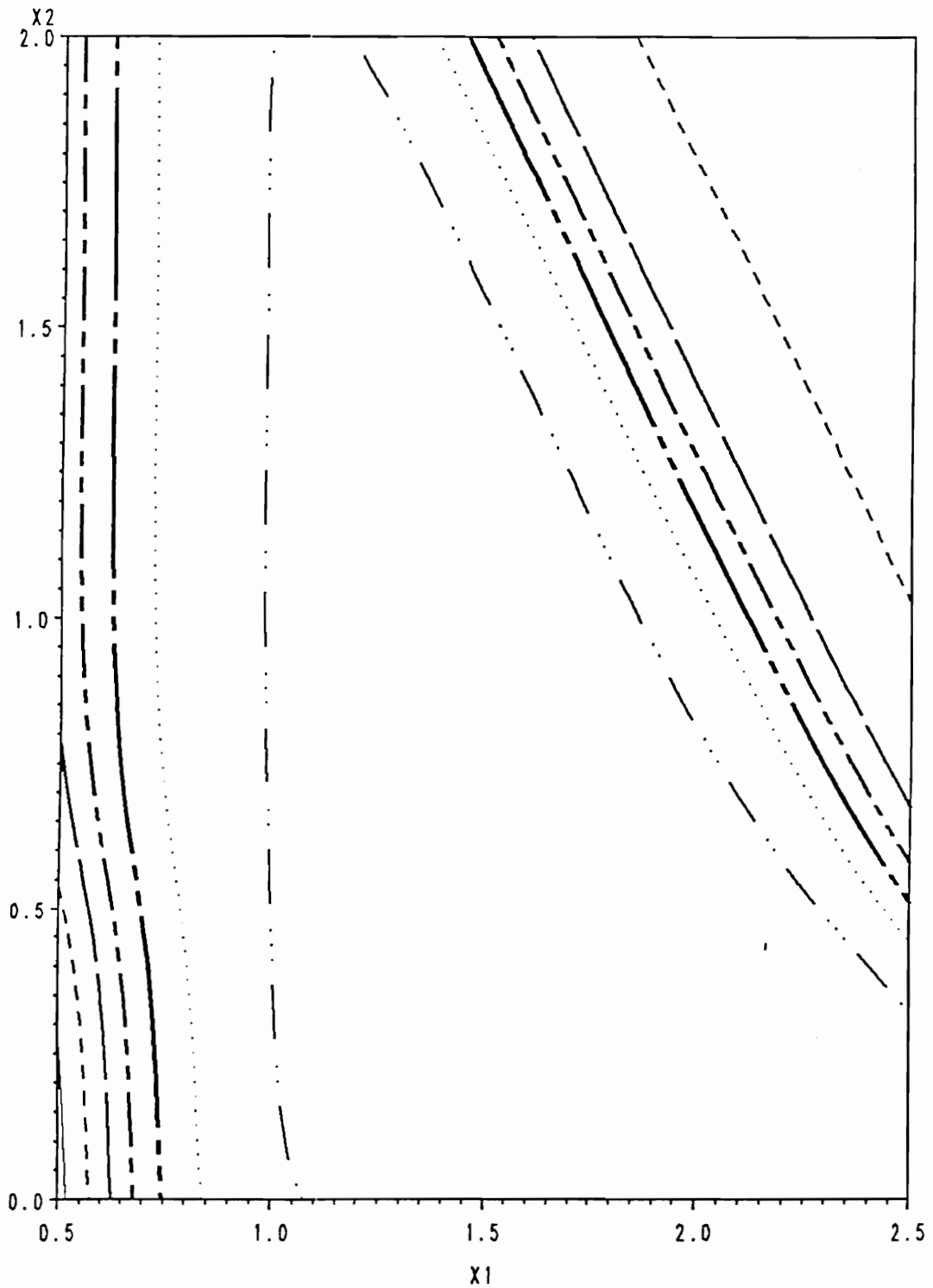
FIG. VI.2.12. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 INITIAL DATA FOR EXAMPLE 1. H=.507 & LAM=0



$\hat{Y}$

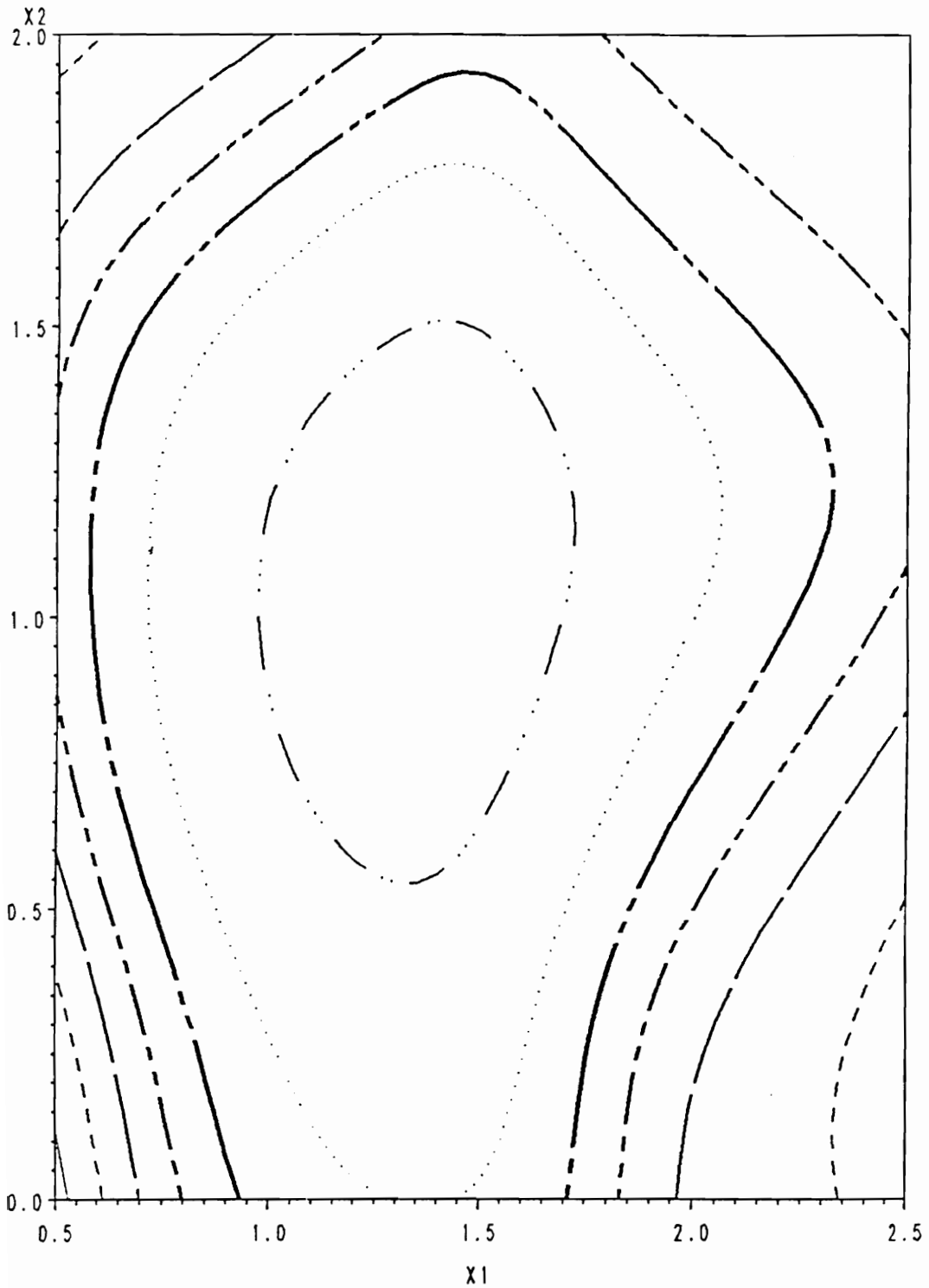
—	30.82	- - -	34.59	—	38.36	- - -	42.14
- - -	45.91	.....	49.69	..	53.46		

FIG. VI.2.13. FITTED 2ND ORDER L.S. MODEL  
INITIAL DATA FOR EXAMPLE 1



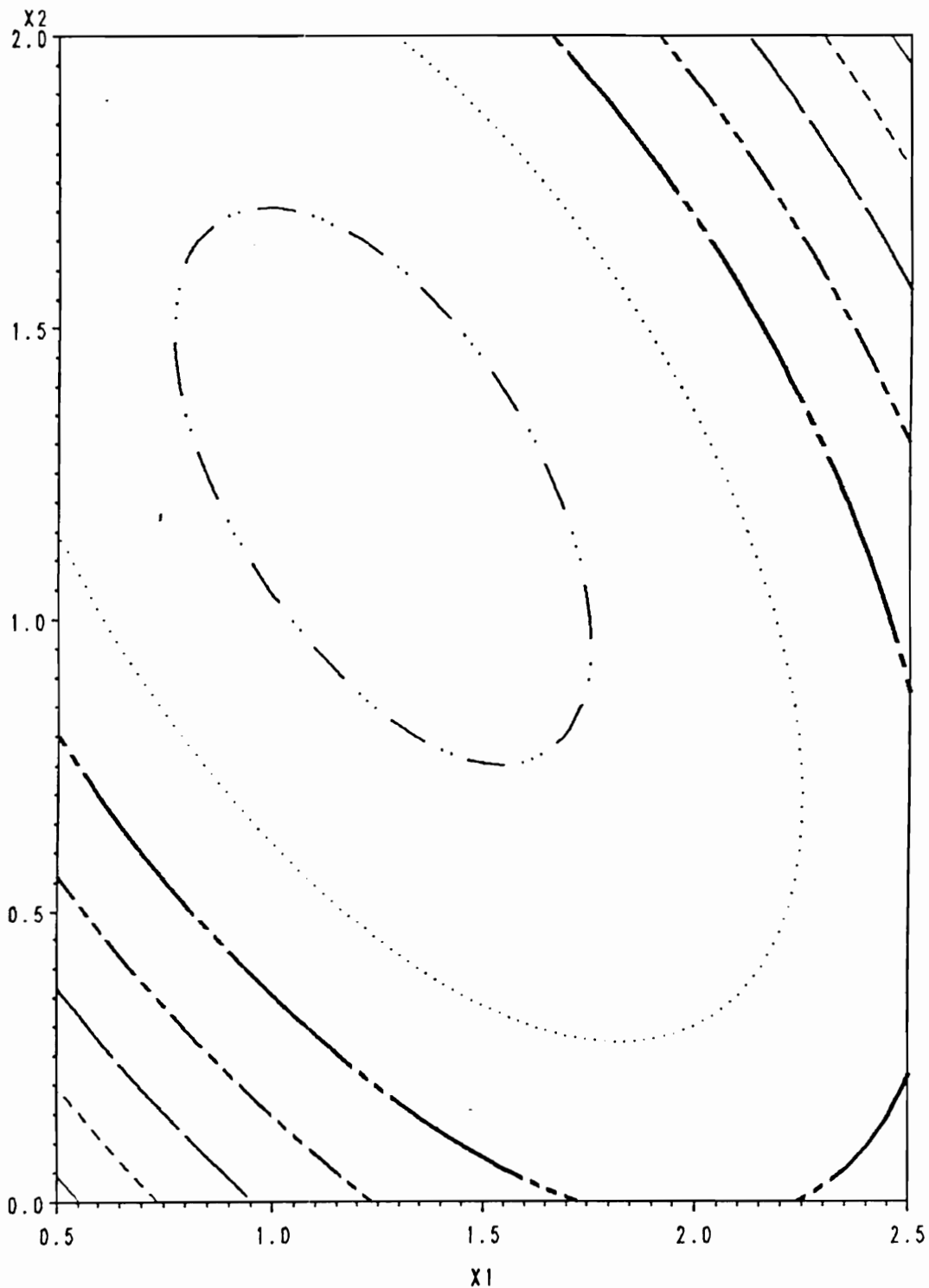
$\hat{Y}$ .      — 34.64      - - - 35.36      - · - 36.08      - - - - 36.80  
          - - - 37.52      ····· 38.24      · · - 38.96

FIG. VI.2.14. FITTED KERNEL SURFACE WITH 2ND ORDER L.S. .  
 INITIAL DATA FOR EXAMPLE 1.  $H=.507$



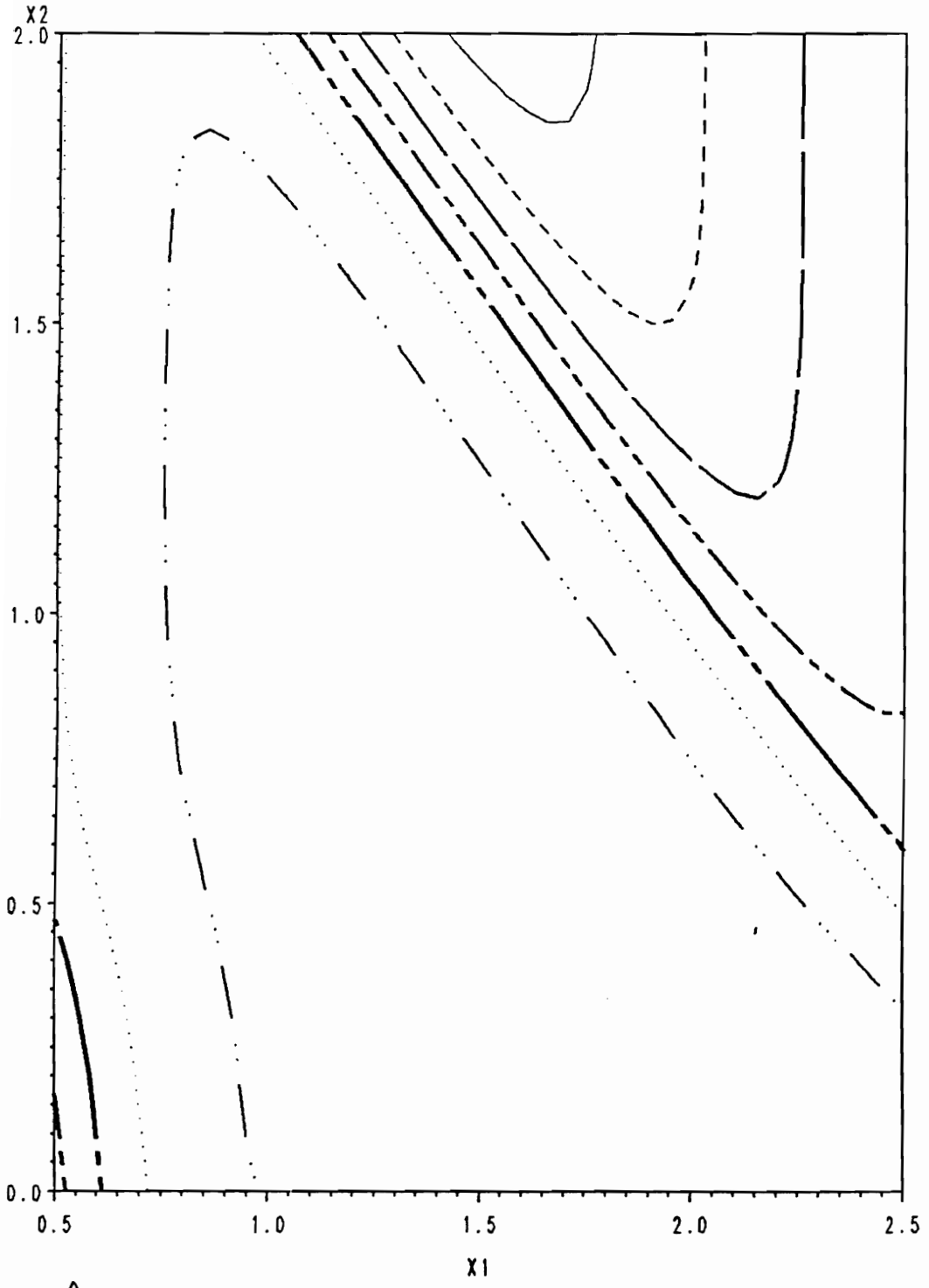
YHAT:    ——— 34.70    - - - - 35.48    ——— 36.27    - - - - 37.05  
          - - - - 37.84    ······ 38.63    - · - · 39.41

FIG. VI.2.15. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 1. H=.591 & LAM=.875



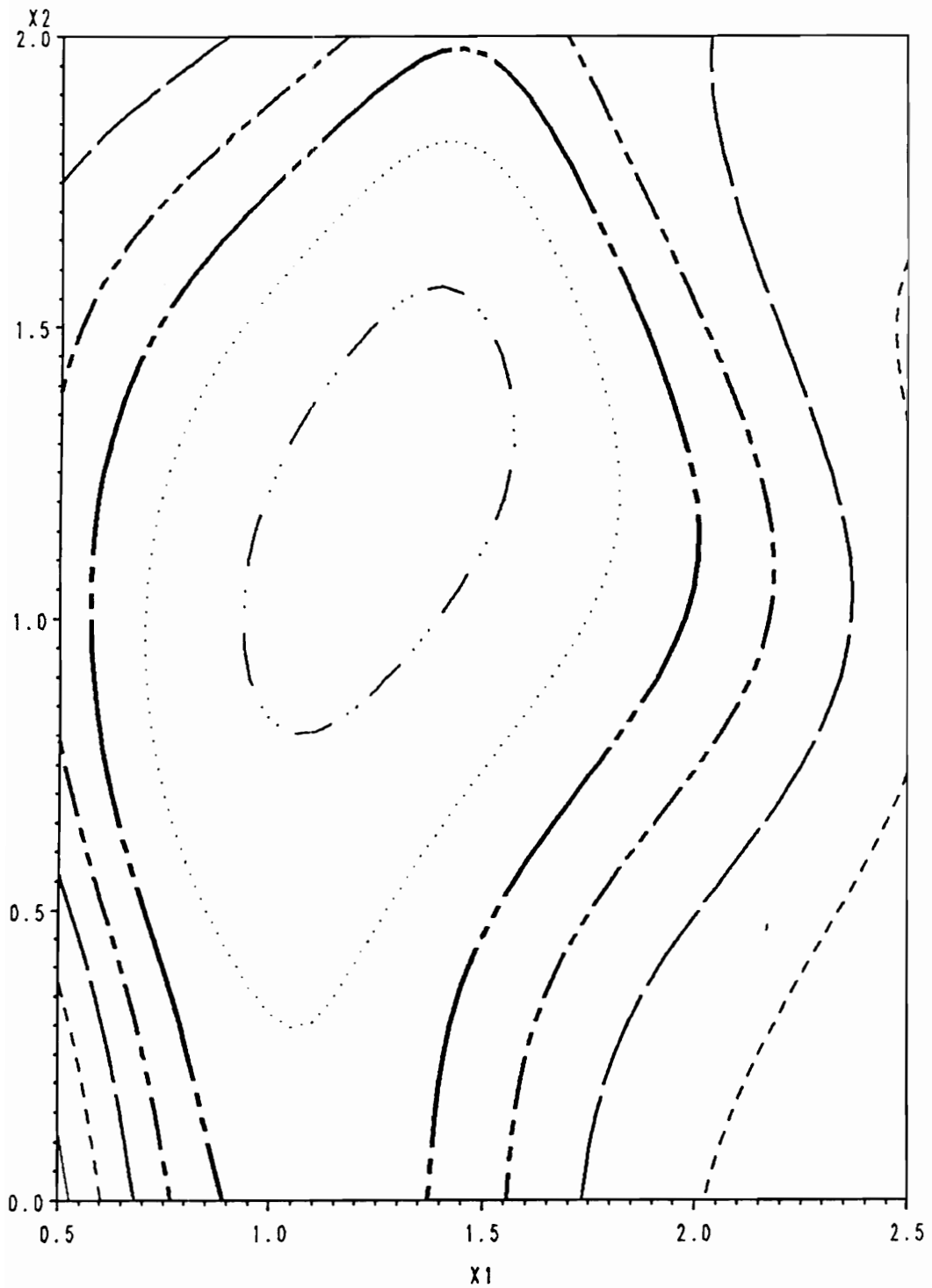
$\hat{Y}$     ——— 34.69    - - - 35.35    ——— 36.02    - - - 36.69  
          - - - 37.35    ····· 38.02    ····· 38.69

FIG. VI.2.16. FITTED 2ND ORDER L.S. MODEL  
 SIX POINTS AUGMENTED FOR EXAMPLE 1



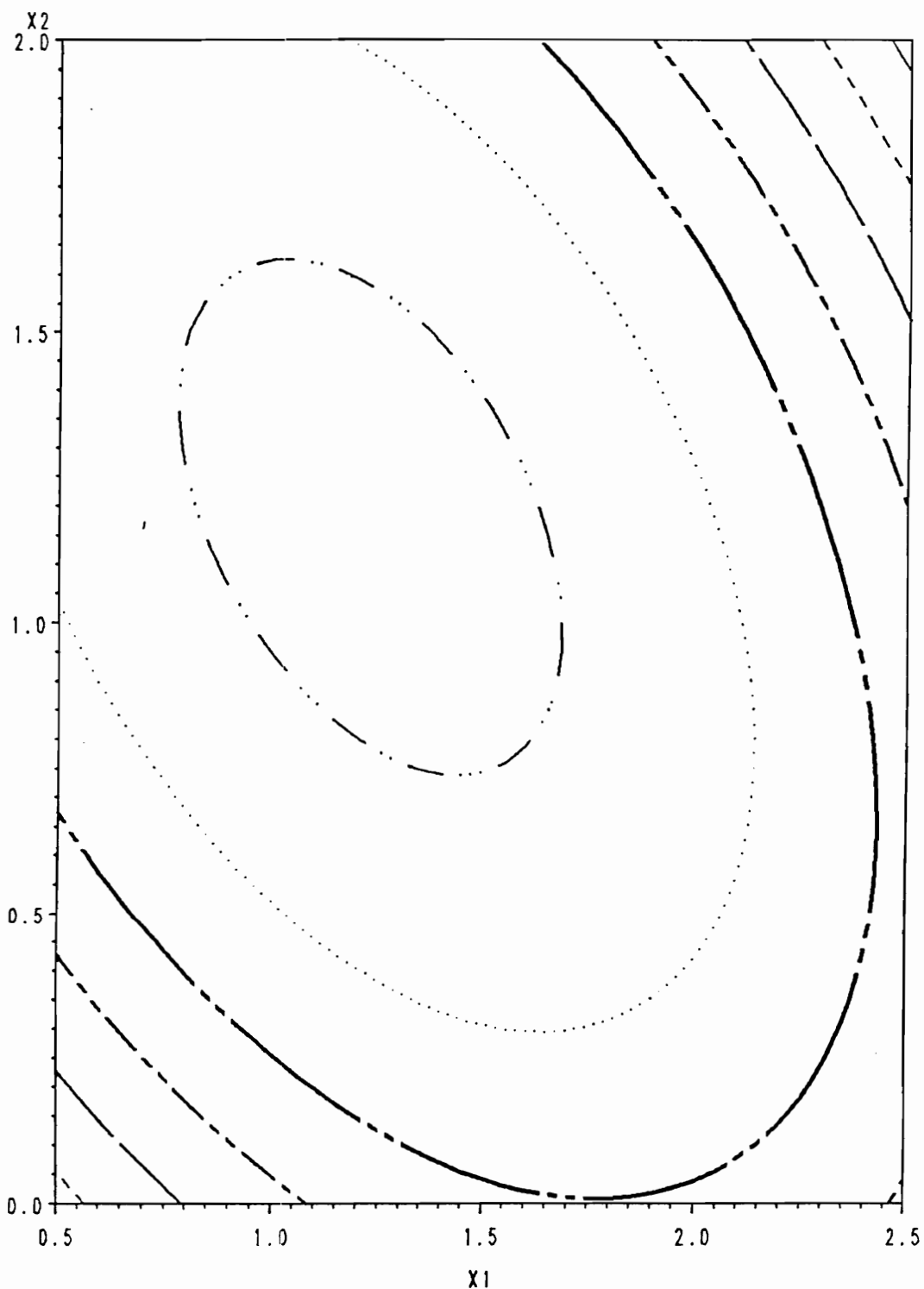
$\hat{Y}$  : ——— 32.44    - - - - 33.59    ——— 34.74    - - - - 35.90  
 - - - - 37.05    ····· 38.20    ····· 39.35  
 FIG. VI. 2.17. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 1.  $H=.518$





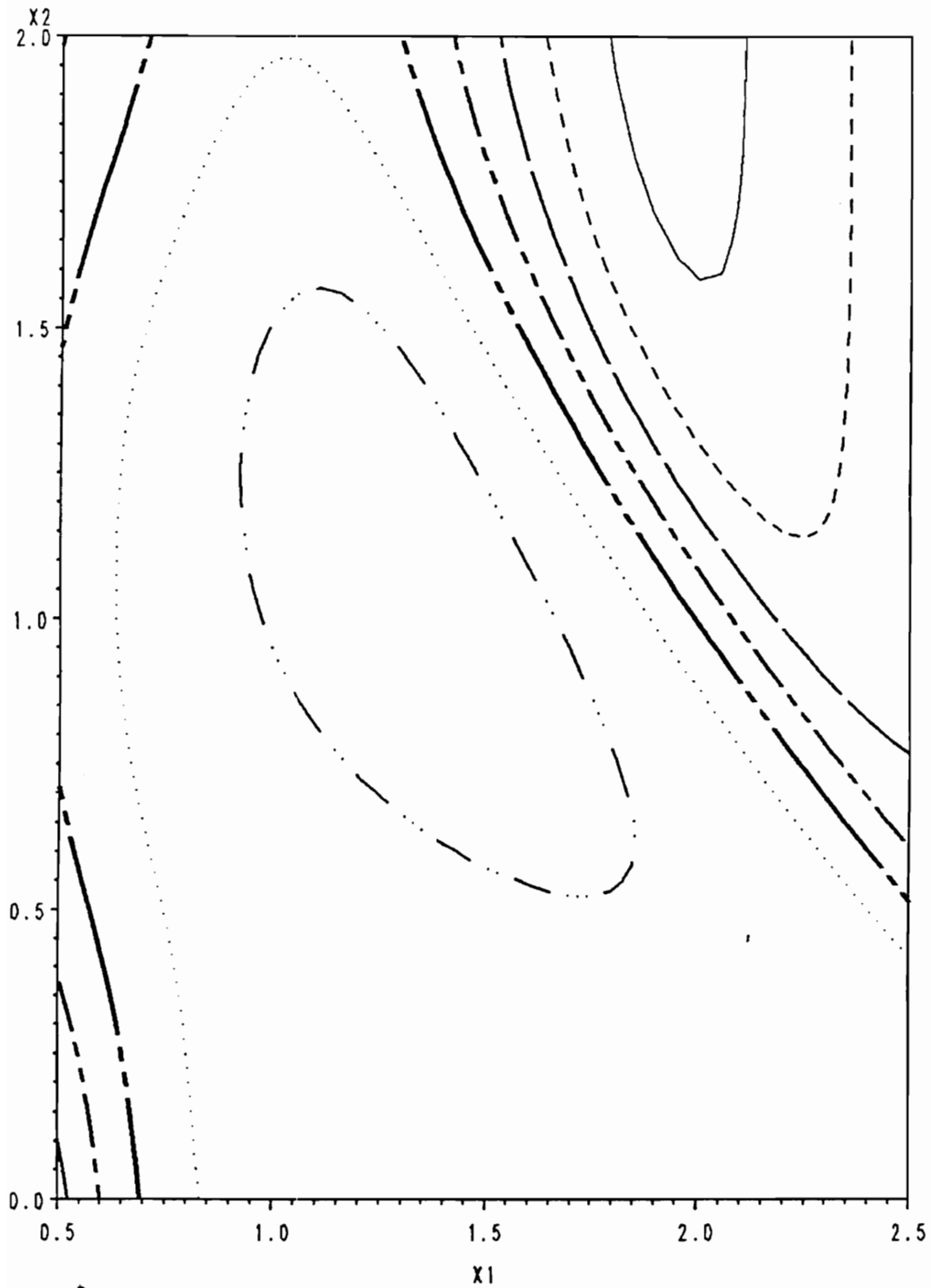
YHAT    ——— 34.64    - - - - 35.37    ——— 36.11    - - - - 36.84  
          - - - - 37.57    ······ 38.31    ······ 39.04

FIG. VI.2.18. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1.  $H=.590$  &  $LAM=.937$



$\hat{Y}$  ——— 34.38    - - - - 35.14    ——— 35.90    - - - - 36.66  
 - - - - 37.42    ..... 38.18    - - - - 38.94

FIG. VI.2.19. FITTED 2ND ORDER L.S. MODEL  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1



$\hat{Y}$     ——— 34.28    - - - - 35.18    ——— 36.07    - - - - 36.97  
          - - - - 37.86    ······ 38.76    ······ 39.65

FIG. VI.2.20. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 1.  $H=.531$

direction. In this case, it appears that HATLINK has been able to detect behavior that does not directly appear in either the least squares or kernel regression cases. This is further evidence of the superiority of the HATLINK method over both least squares and kernel regressions.

### VI.3 SECOND EXAMPLE IN TWO DIMENSIONS

For the first example, the misspecification in the model involved combinations of sine functions. It was shown that HATLINK provided improved estimation of the response surface than either least squares or kernel regressions. One might suspect that only combinations of trigonometric functions work for the model misspecification. The second example will have a different type of misspecification to illustrate the general nature of the HATLINK method. A similar procedure will be followed, except in this case the misspecification in the model will involve exponential functions. In this example, both old and new regions of interest, as well as the initial point in the new region of interest will be the same as in the previous example. Instead of using equation (IV.3.1), the true model is now assumed to be

$$Y = x_1^2 + x_1 - 3(x_2 - 1.05)^2 + 35 +$$

$$1 \left\{ \exp \left[ -x_1^2 + 4x_1 - (x_2 + .4)^2 + 4(x_2 + .4) \right] - 4 \right\} + \epsilon, \quad (\text{VI.3.1})$$

where  $\epsilon \sim N(0,1)$ . Plots of (VI.3.1) in the old and new region of interest are included as Figure VI.3.1 and Figure VI.3.2. As in the previous example, twelve points will be augmented using a batch size of one. For (VI.3.1) the true maximum is 94.83 and is located at (2.05, 1.55). This location is compared to the individual largest fitted value by finding the distance between the two points. After each point is augmented, the fitted maximum for the HATLINK, least squares

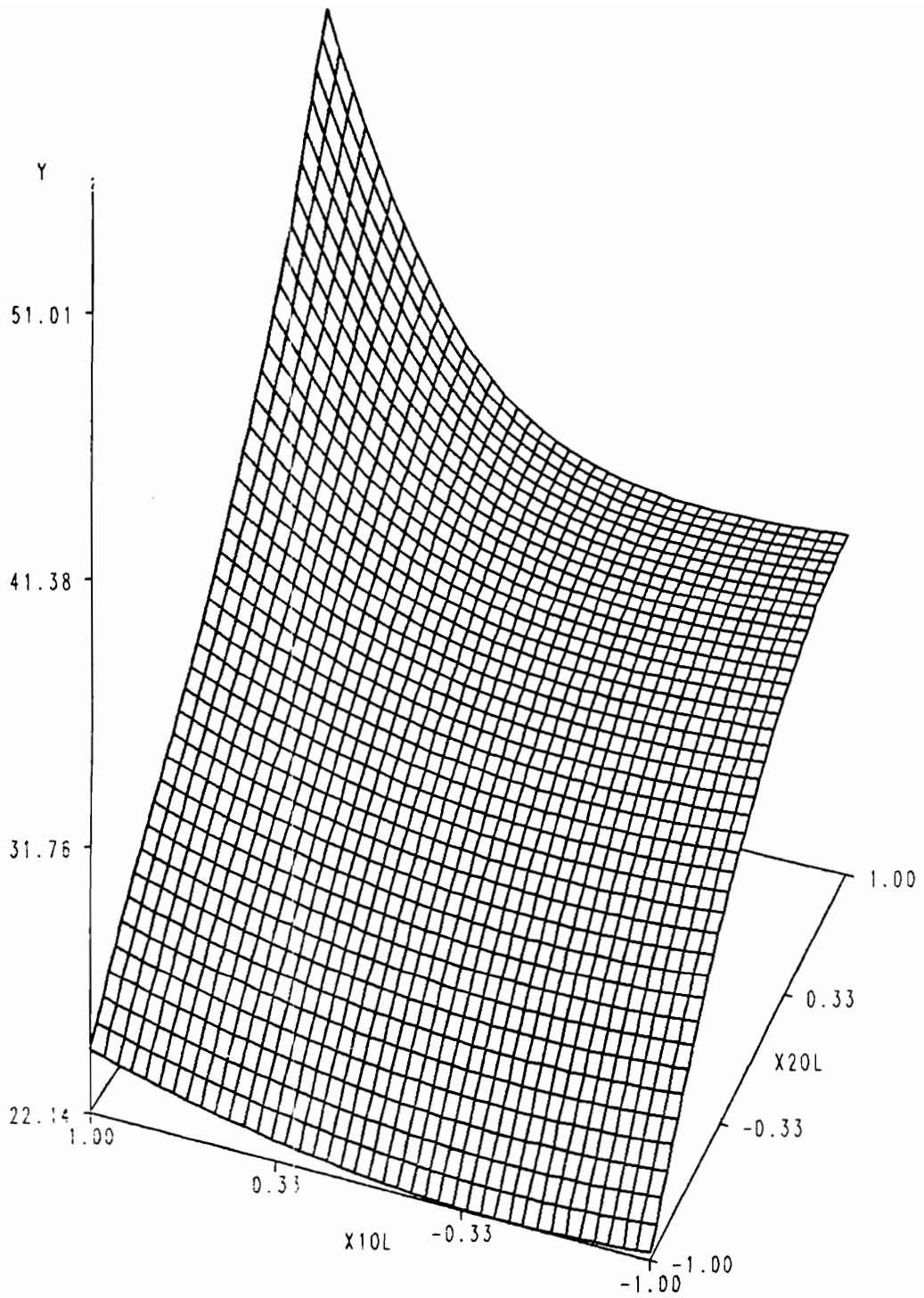


FIG. VI.3.1 TRUE MODEL IN OLD REGION FOR EXAMPLE 2  
WITH  $CT=1$

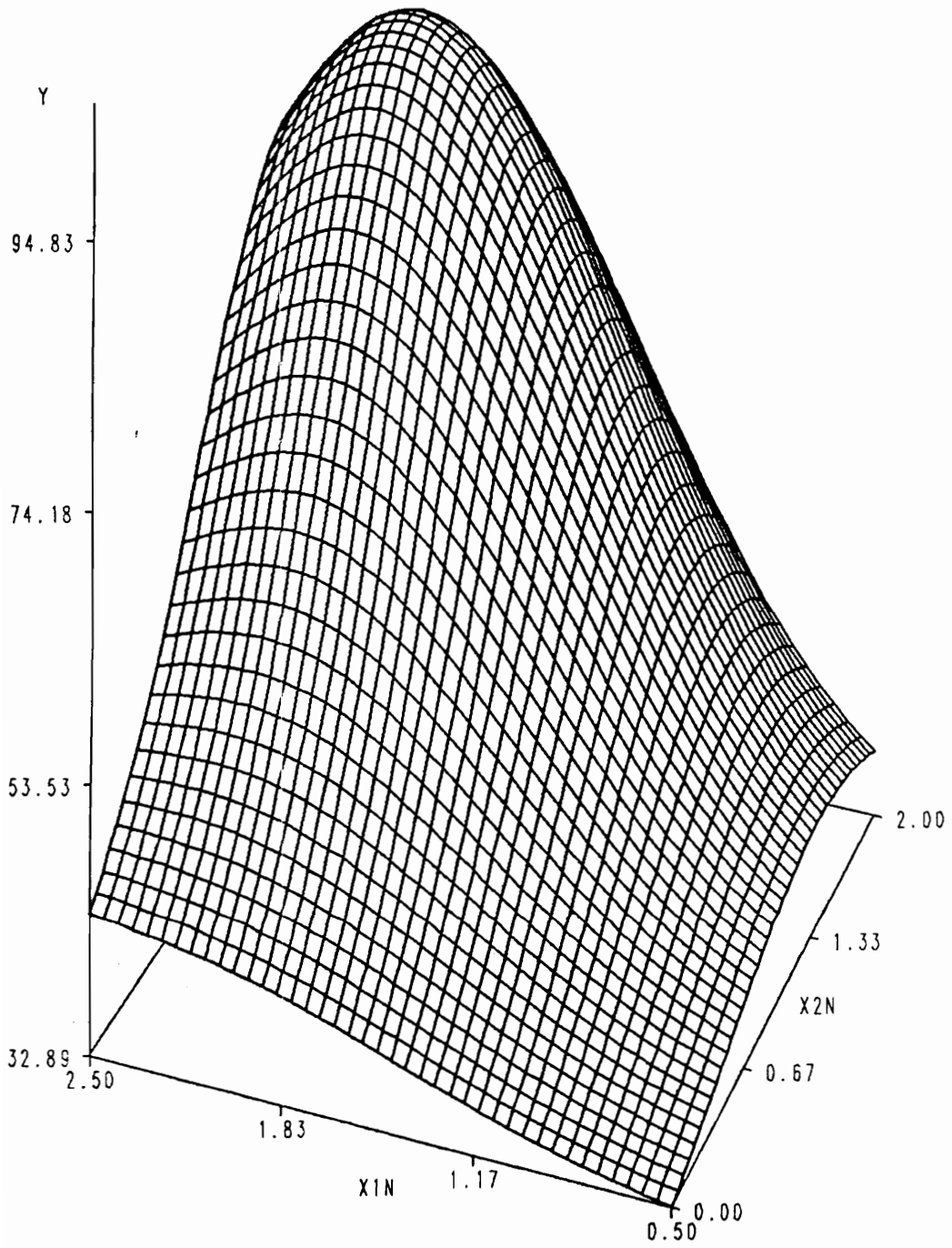


FIG. VI.3.2. TRUE MODEL IN NEW REGION FOR EXAMPLE 2  
WITH  $CT=1$

and kernel surfaces and the distance between this point and the true value will be calculated. As in the previous case, plots of the fitted surface and contour plots for the initial data, six points augmented and twelve points augmented will be generated and briefly compared. These plots will be presented for the true surface, HATLINK, least squares, and kernel regressions.

Augmentation of twelve points yielded the results presented as follows: HATLINK in Table VI.3.1, a second order least squares model in Table VI.3.2, and kernel regression in Table VI.3.3. Plots of the design points are presented in Figure VI.3.3 through Figure VI.3.5 for HATLINK, least squares and kernel regressions respectively. As in the first example, notice the ESE values, the fitted maxima, and the distances to the true location of the maximum. In terms of finding what the true maximum should be for this example, the HATLINK method performed best, estimating it somewhat better than the least squares method. Kernel regression performed the worst of the three methods in this example.

HATLINK also placed the largest fitted value closest to the true location of the maximum. The locations and distances for both least squares and kernel regression were the same. After the twelfth point was augmented, HATLINK determined that the greatest predicted response was located at (2.0, 1.5). Both the least squares and kernel methods placed it at (2.5, 2.0). As in the first example, the same considerations on using grids of points to locate the fitted maximum should be kept in mind.

Notice the behavior of the ESE values. After twelve points are augmented, HATLINK has a much better overall fit than either the least squares or kernel methods, thus indicating that the HATLINK surface would be superior in overall performance. Again, this could be important in terms of finding the optimal range of operating conditions. Pictures and contour plots illustrate that the HATLINK method outperforms both least squares and kernel methods. Comparison of plots from the three methods reveal that, as conjectured, HATLINK draws from

both least squares and kernel methods to produce a better overall fit.

As in Section VI.3, plots of the true response surface, the fitted surfaces from HATLINK, least squares, and kernel regression with the initial data, six points augmented, and twelve points augmented are included as Figure VI.3.6 through Figure VI.3.15. Analysis of these plots in a fashion similar to that performed in Section VI.3 reveal that after twelve points are augmented, the HATLINK surface is most closely resembling the true surface and contains some elements from both least squares and kernel regressions.

Contour plots of the true surface and the nine fitted surfaces reveal the superiority of the HATLINK procedure. These plots are Figures VI.3.16 through VI.3.25. Just as in the first example in Section VI.4, these contour plots reveal that HATLINK produces a better overall fitted surface. Both the contour plots and the fitted surfaces reveal that HATLINK is detecting behavior that does not directly appear in either the least squares or kernel regression cases. For future research, other examples could also be performed.



**Table VI.3.1. Twelve Points Augmented Using IIATLINK and BIIV for Example 2.** Using  $\sigma = 1$  and  $CT = 1$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>LAM</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	—	—	.641	1.00	3221.83	—	—
1	2.5000	2.0000	.480	.937	2656.27	76.21	.64
2	2.5000	0.0000	.529	.998	2744.27	76.19	.64
3	1.0000	1.0000	.568	.996	2808.67	76.19	.64
4	2.5000	0.5000	.586	.750	3015.73	76.21	.64
5	2.0000	1.5000	.363	1.00	1249.07	94.47	.07
6	2.5000	1.0000	.509	1.00	682.62	88.47	.07
7	1.5000	1.0000	.589	1.00	676.21	83.05	.07
8	2.0000	1.0000	.400	1.00	762.05	91.67	.71
9	2.0000	2.0000	.393	1.00	645.52	89.63	.07
10	1.0000	0.0000	.387	1.00	604.15	89.88	.07
11	1.0000	0.5000	.393	1.00	575.05	89.61	.07
12	1.5000	0.5000	.401	1.00	545.57	89.25	.07

**Table VI.3.2. Twelve Points Augmented Using Least Squares and BIIV for Example 2.** Using  $\sigma = 1$  and  $CT = 1$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	—	—	3429.06	—	—
1	2.5000	0.0000	3665.40	74.76	.64
2	2.5000	0.5000	3712.57	74.64	.64
3	0.5000	2.0000	4102.40	71.72	.64
4	2.0000	2.0000	2785.91	79.68	.64
5	1.5000	2.0000	2599.85	81.45	.64
6	2.5000	1.0000	2307.71	83.24	.64
7	2.5000	2.0000	2459.32	81.60	.64
8	2.0000	1.5000	1929.88	85.39	.64
9	2.0000	1.0000	1773.63	86.53	.64
10	1.5000	1.5000	1707.67	87.78	.64
11	1.5000	1.0000	1700.30	87.98	.64
12	2.0000	0.5000	1699.88	87.98	.64

**Table VI.3.3. Twelve Points Augmented Using Kernel and BIIV for Example 2.** Using  $\sigma = 1$  and  $CT = 1$ , this lists the points augmented, the bandwidth values selected and the ESE values. The bandwidths and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS is the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>BAND</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	–	–	.641	3221.83	–	–
1	2.5000	2.0000	.480	2697.62	76.19	.64
2	1.0000	1.0000	.491	2674.73	76.19	.64
3	1.0000	0.5000	.321	2891.66	76.19	.64
4	1.0000	0.0000	.350	2841.57	76.19	.64
5	0.5000	1.0000	.319	2850.09	76.19	.64
6	0.5000	0.5000	.371	2777.37	76.19	.64
7	0.5000	0.0000	.352	2793.75	76.19	.64
8	0.5000	1.5000	.337	2941.85	76.19	.64
9	0.5000	2.0000	.351	3116.70	76.19	.64
10	1.0000	1.5000	.342	2746.99	76.19	.64
11	1.0000	2.0000	.388	2961.46	76.19	.64
12	1.5000	2.0000	.605	2754.17	75.98	.64

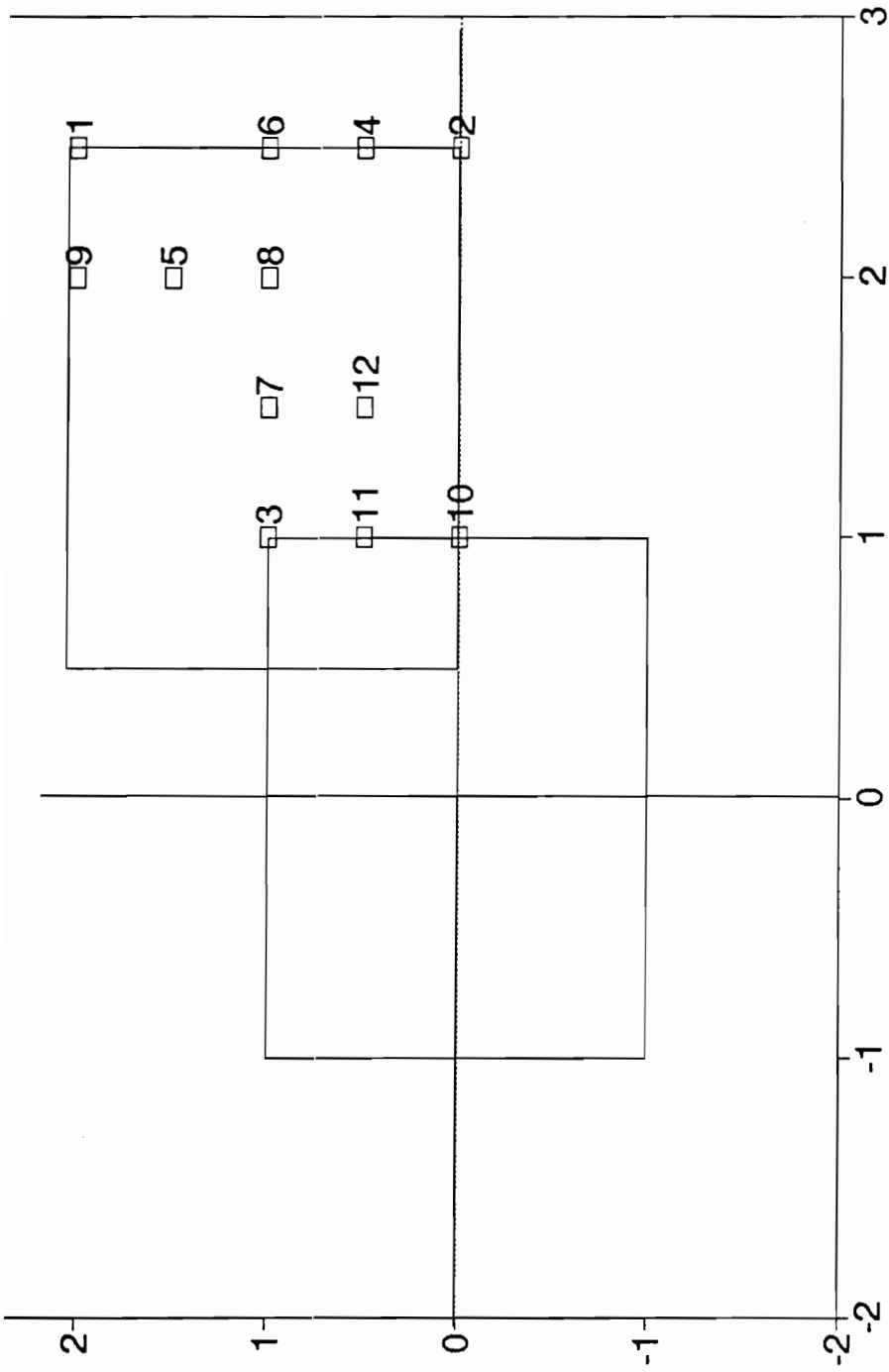


Fig. VI.3.3. Twelve Points Augmented Using HATLINK & BIIV for Example 2 with  $CT = 1$  &  $\text{Sigma} = 1$

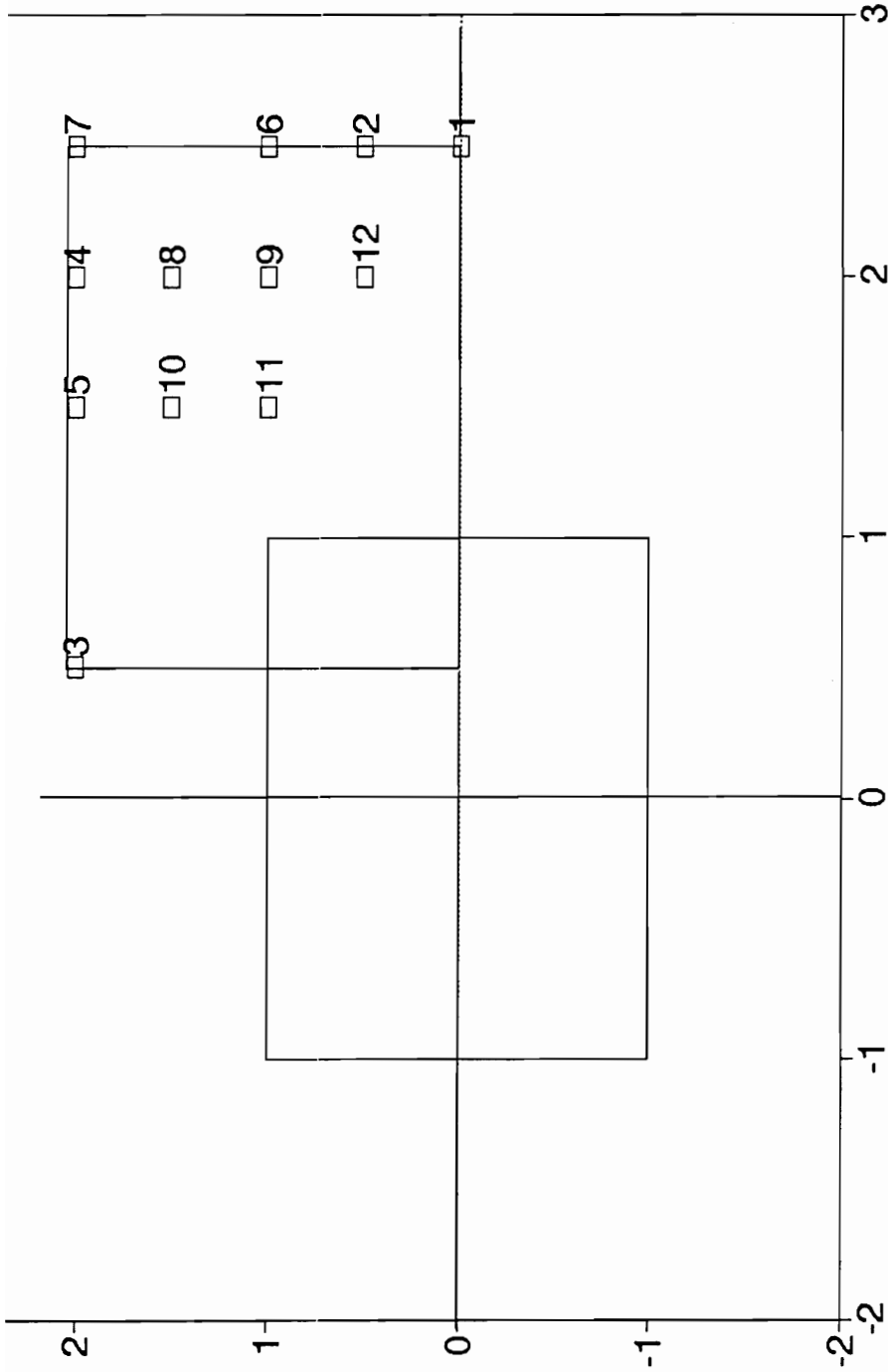


Fig. VI.3.4. Twelve Points Augmented Using L.S. & BIV for Example 2 with  $CT = 1$  &  $\Sigma = 1$

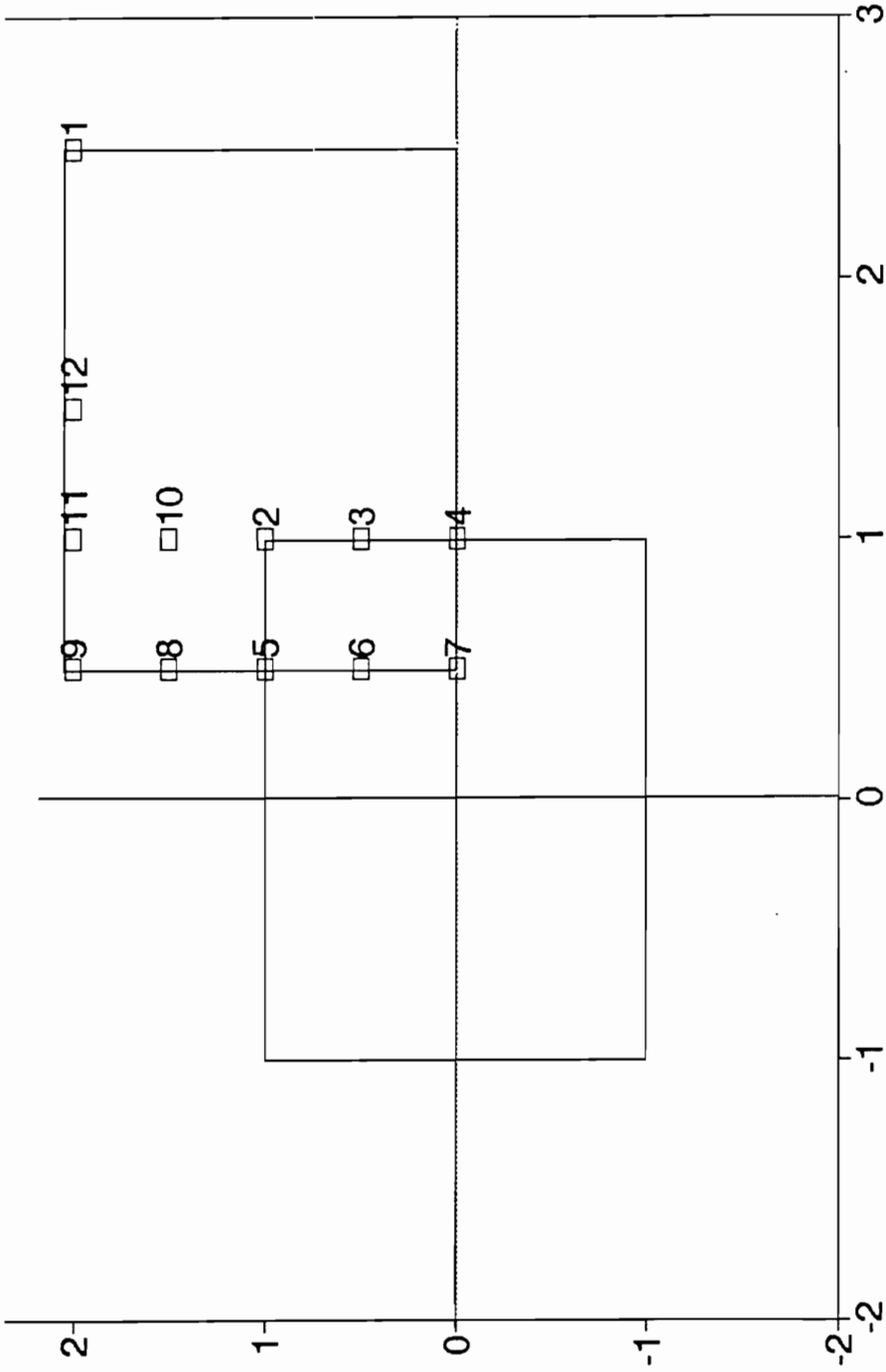


Fig. VI.3.5. Twelve Points Augmented Using KERNEL & BIV for Example 2  
with  $CT = 1$  &  $\Sigma = 1$

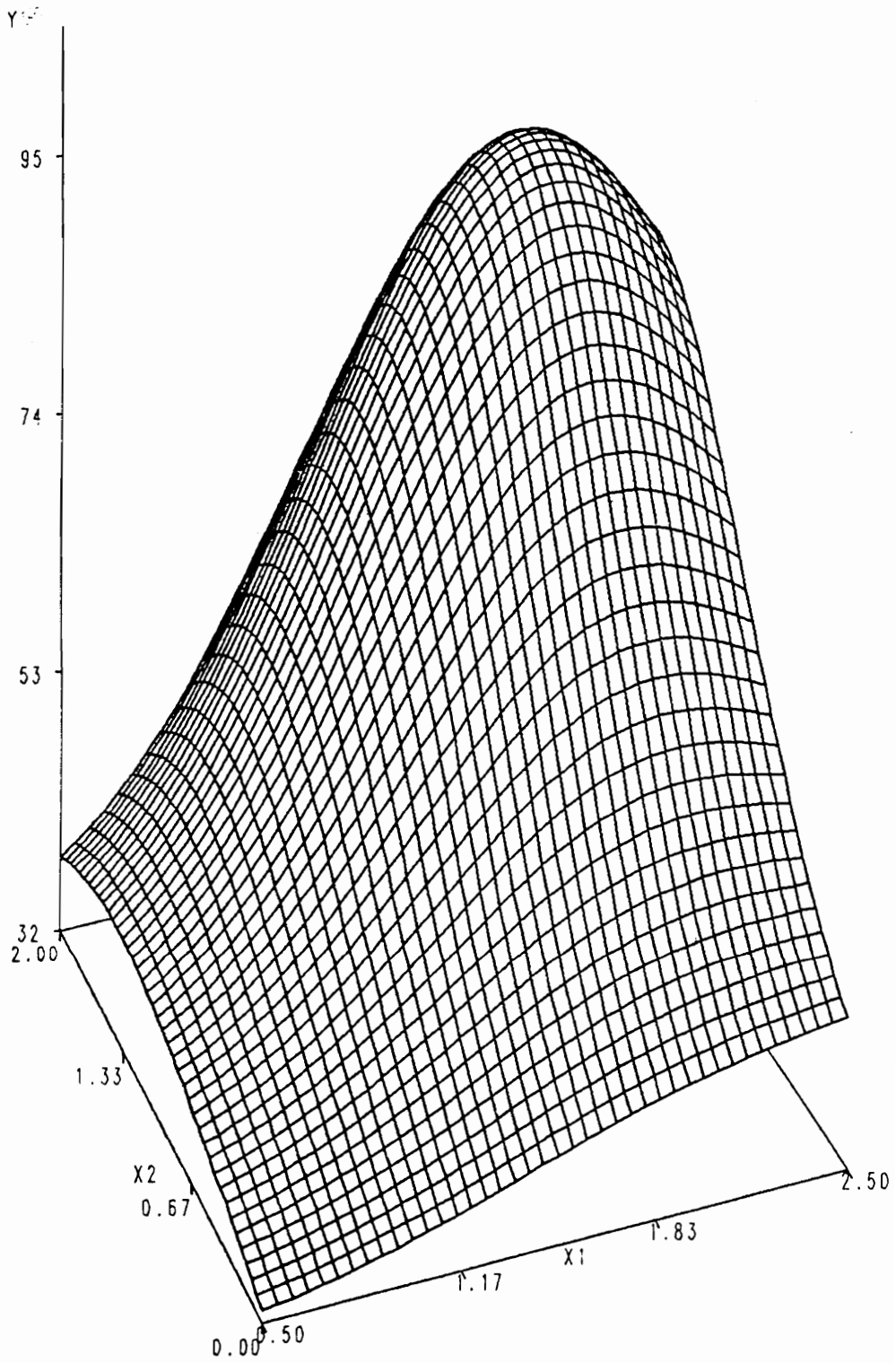


FIG. VI.3.6. TRUE RESPONSE SURFACE FOR EXAMPLE 2

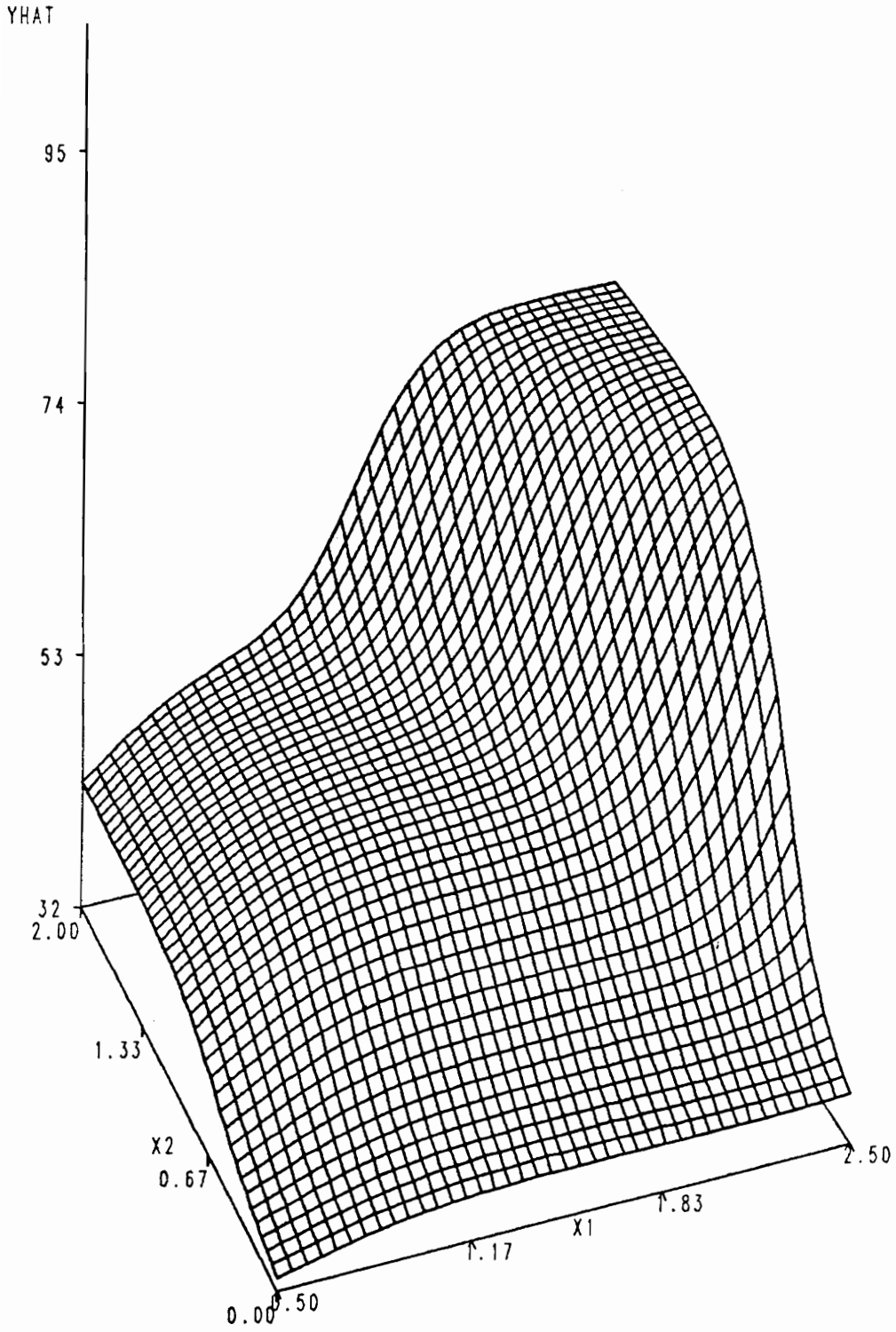


FIG. VI.3.7. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 INITIAL DATA FOR EXAMPLE 2.  $H=.641$  &  $LAM=1$



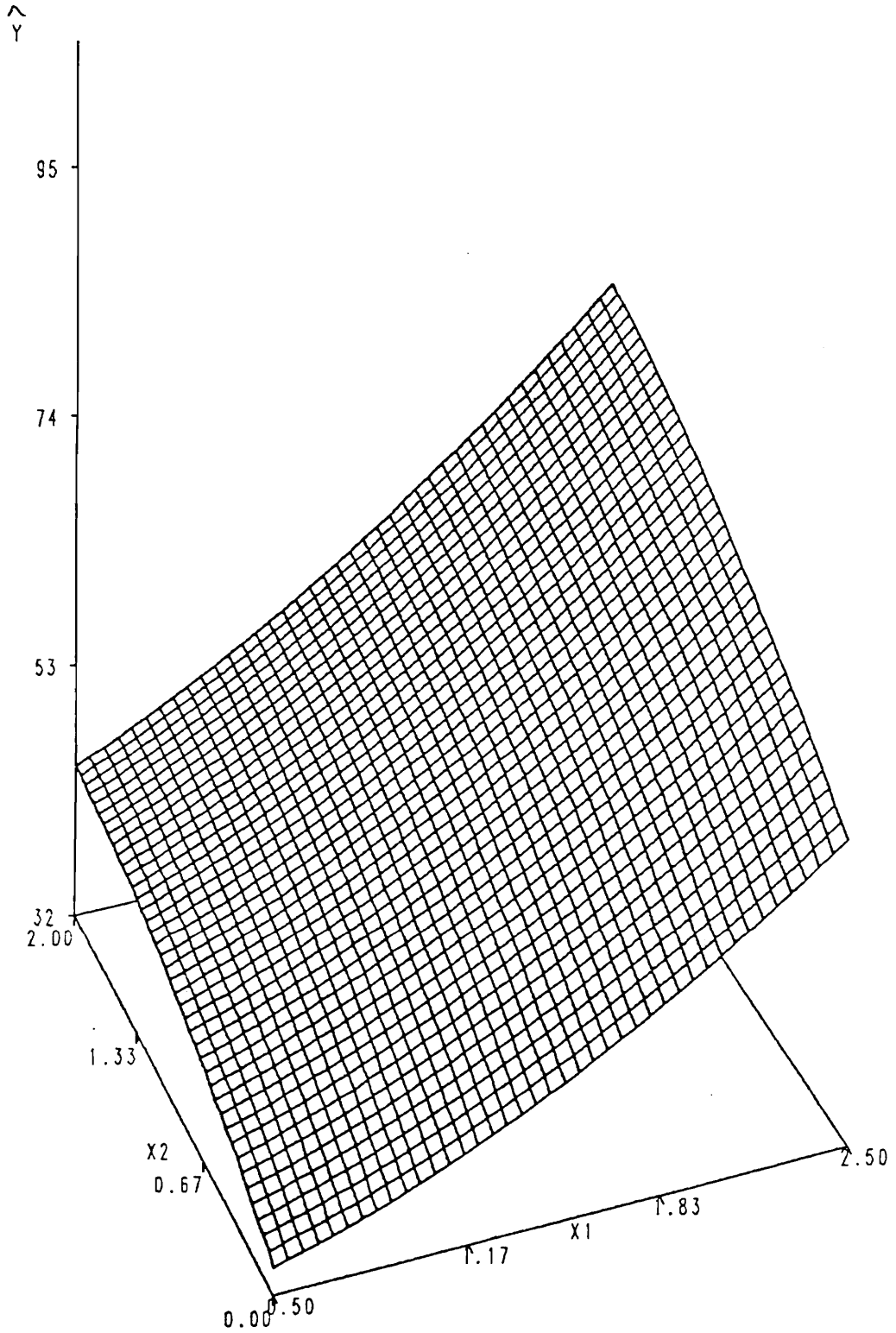


FIG. VI.3.8. FITTED 2ND ORDER L.S. MODEL  
INITIAL DATA FOR EXAMPLE 2

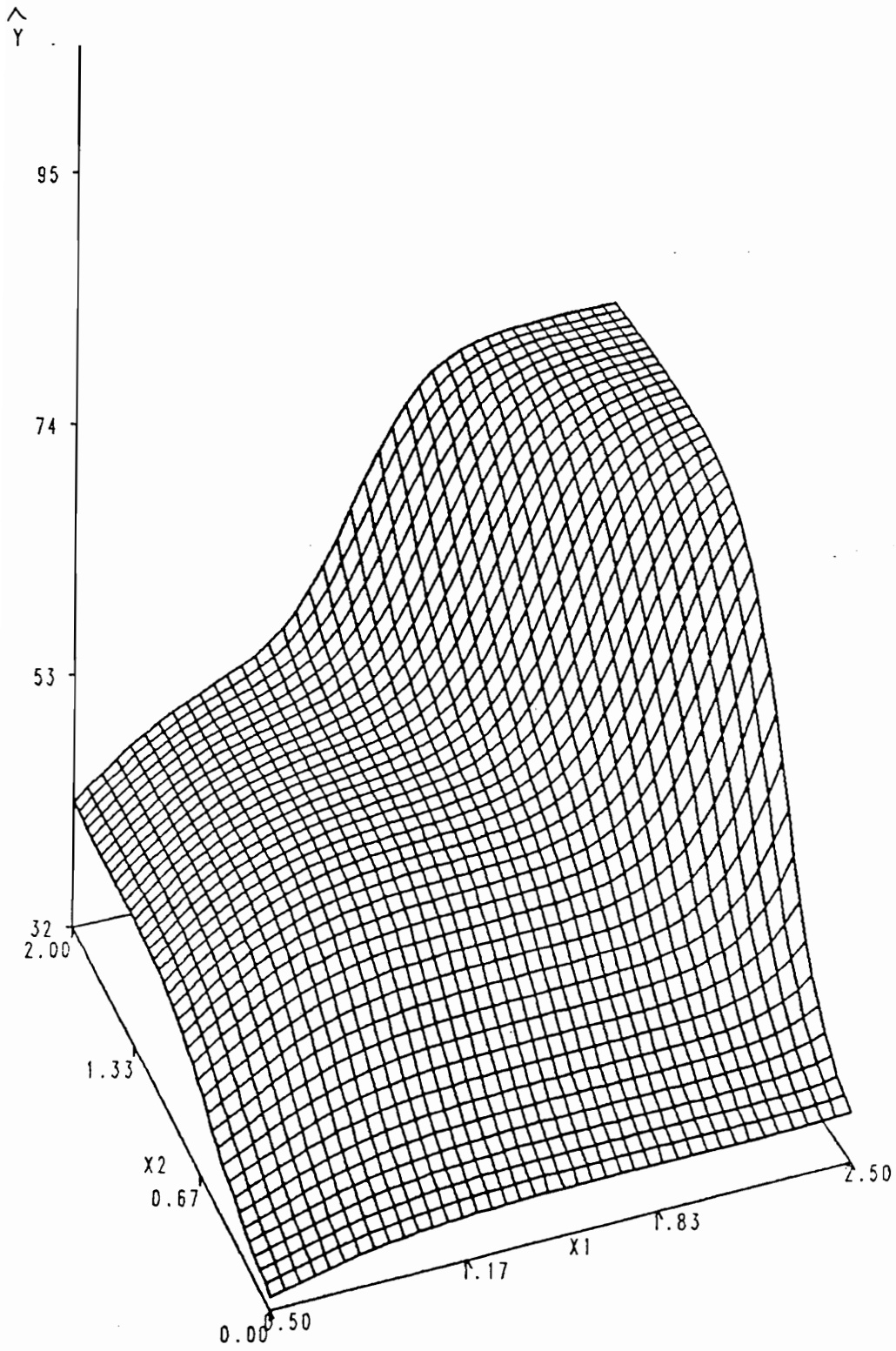


FIG. VI.3.9. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 INITIAL DATA FOR EXAMPLE 2.  $H=.641$

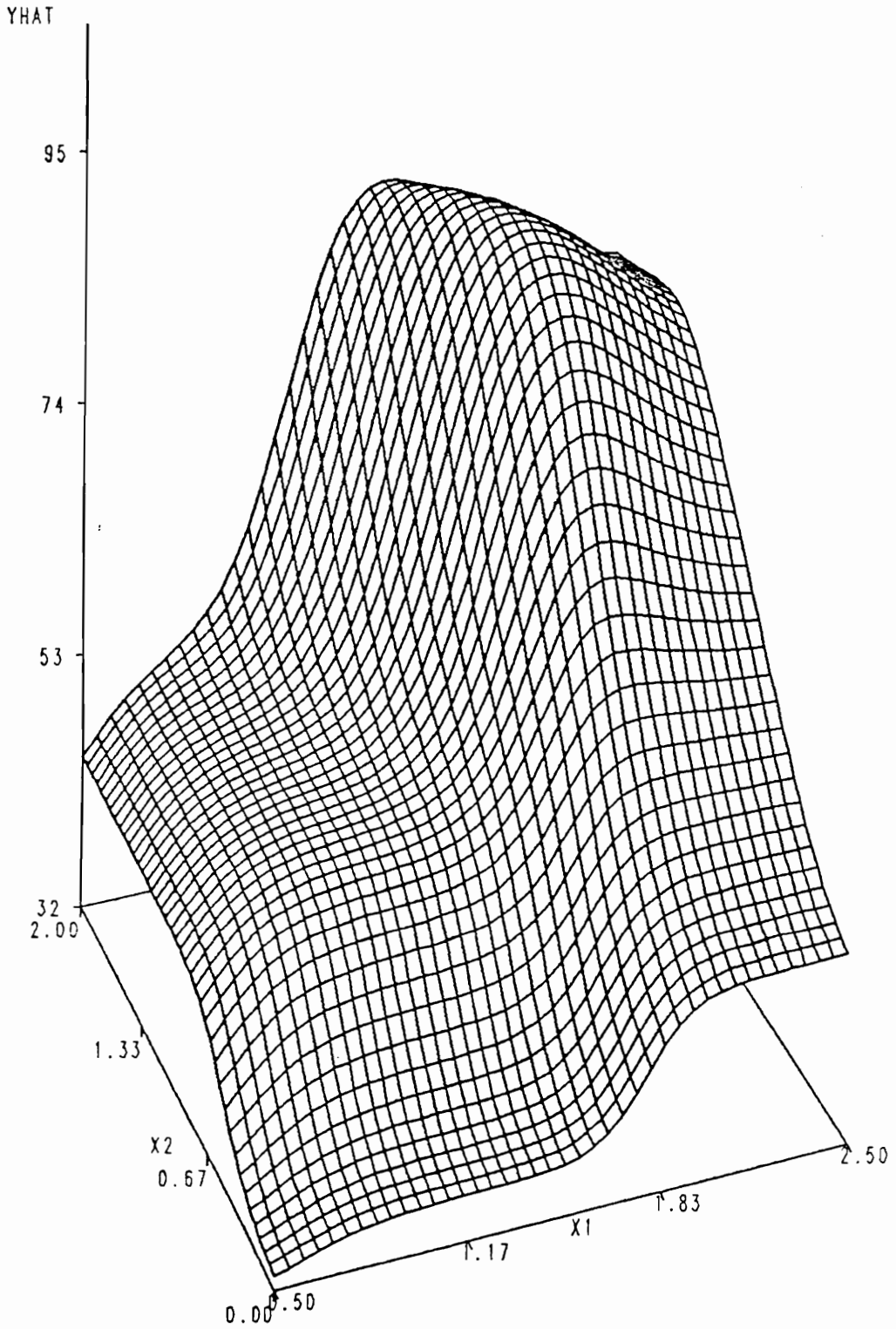


FIG. VI.3.10. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 2.  $H=.509$  &  $LAM=1$

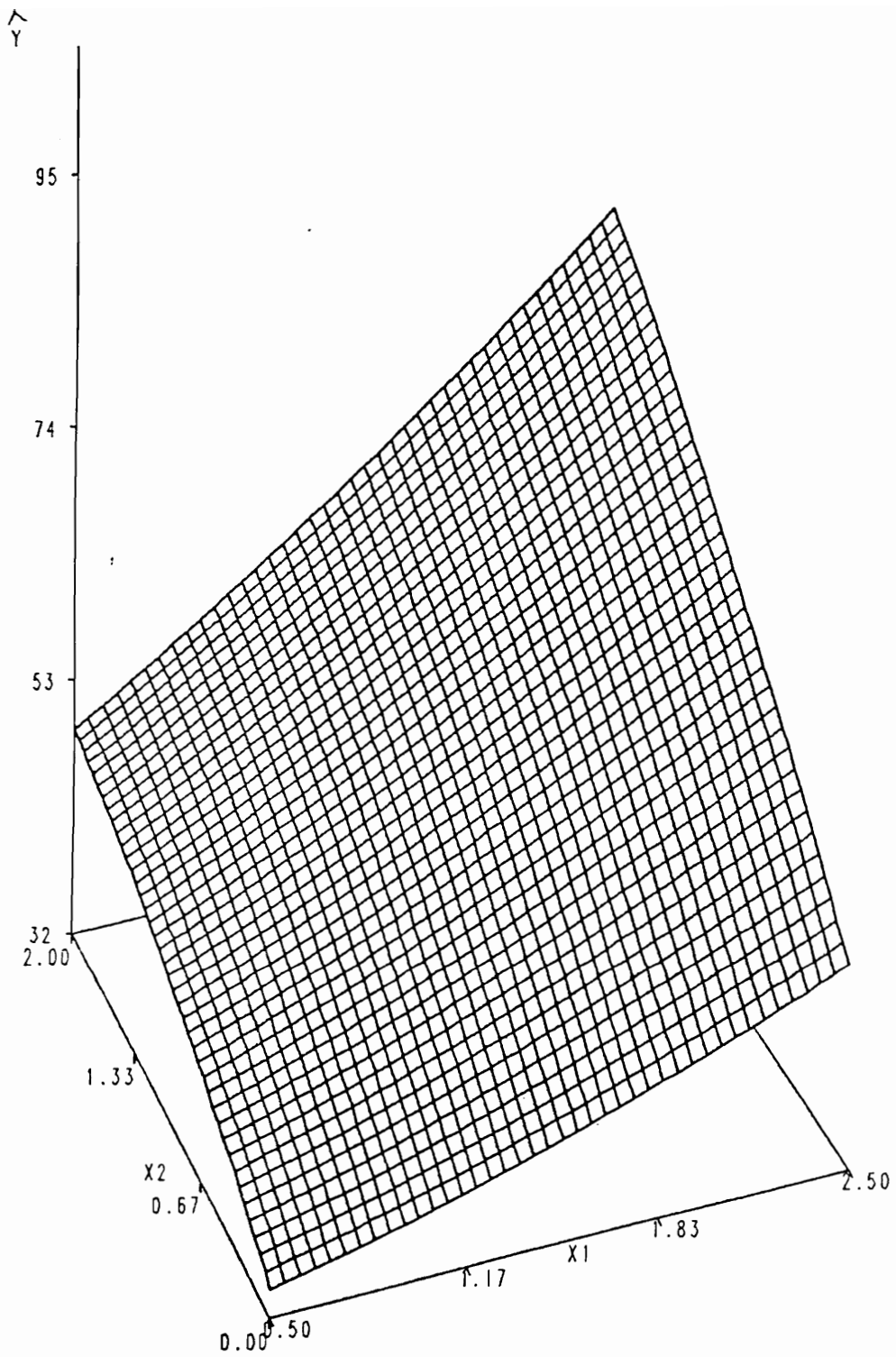


FIG. VI.3.11. FITTED 2ND ORDER L.S. MODEL  
SIX POINTS AUGMENTED FOR EXAMPLE 2

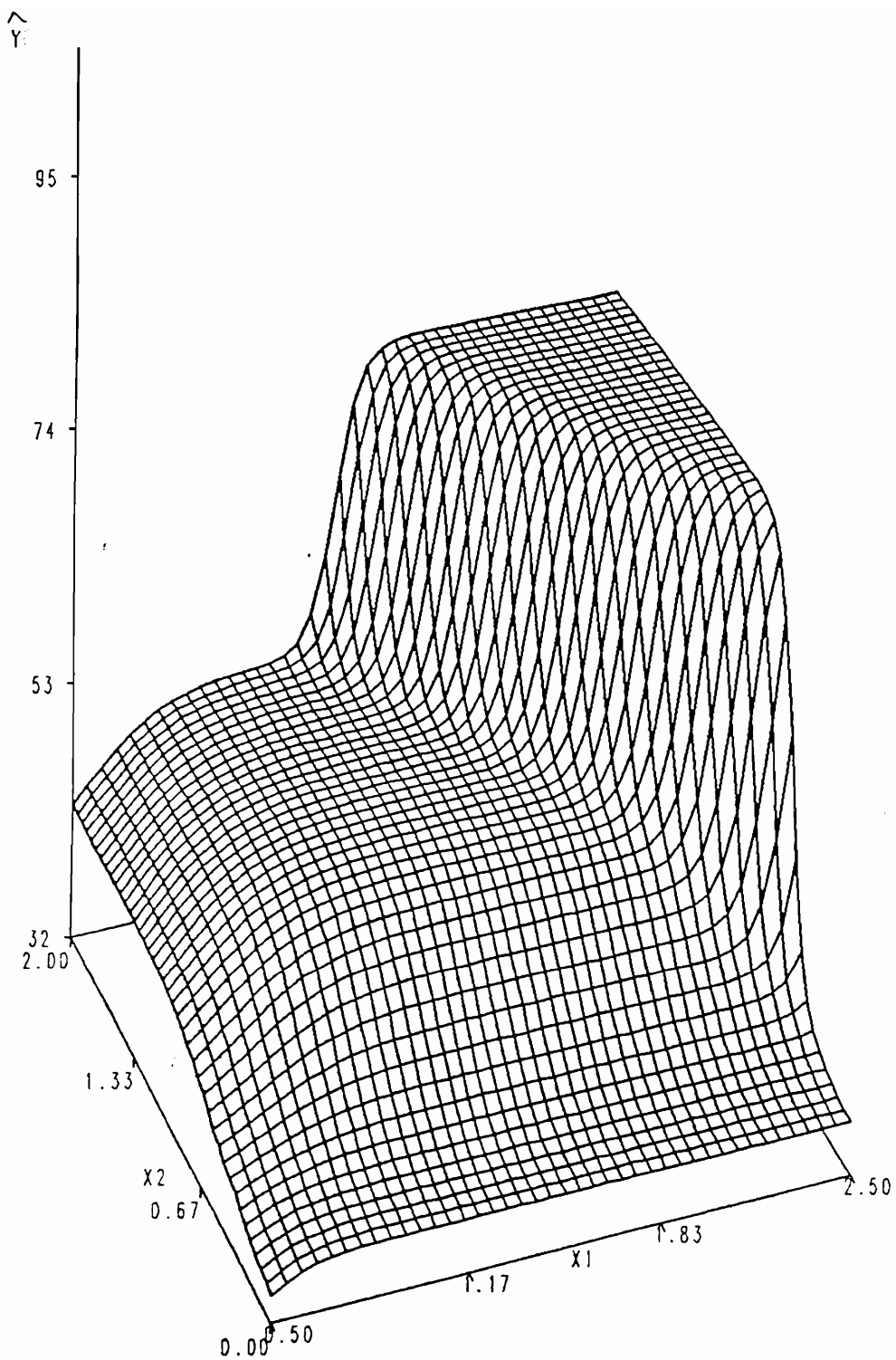


FIG. VI.3.12. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
SIX POINTS AUGMENTED FOR EXAMPLE 2.  $H=.371$

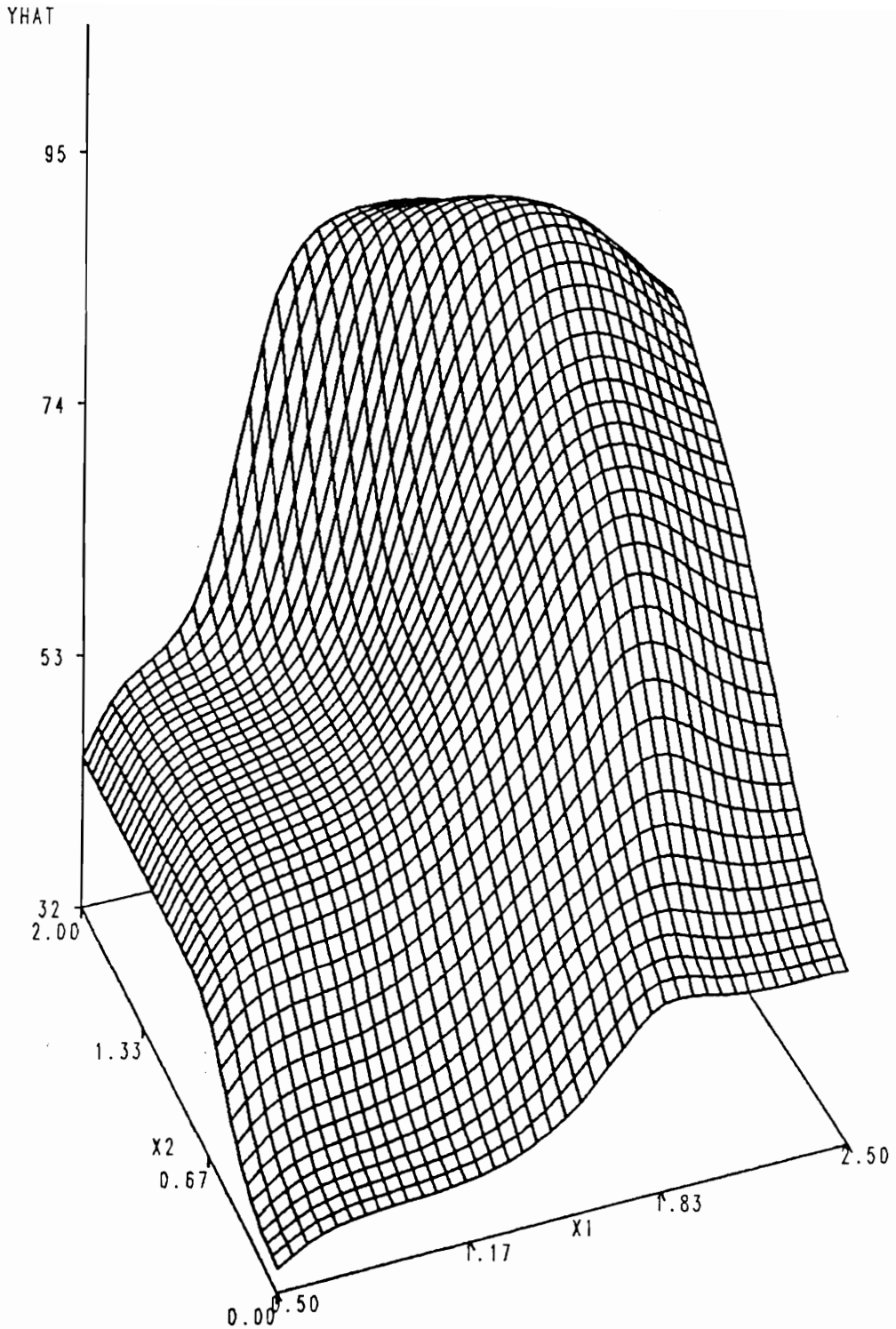


FIG. VI.3.13. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2.  $H=.401$  &  $LAM=1$

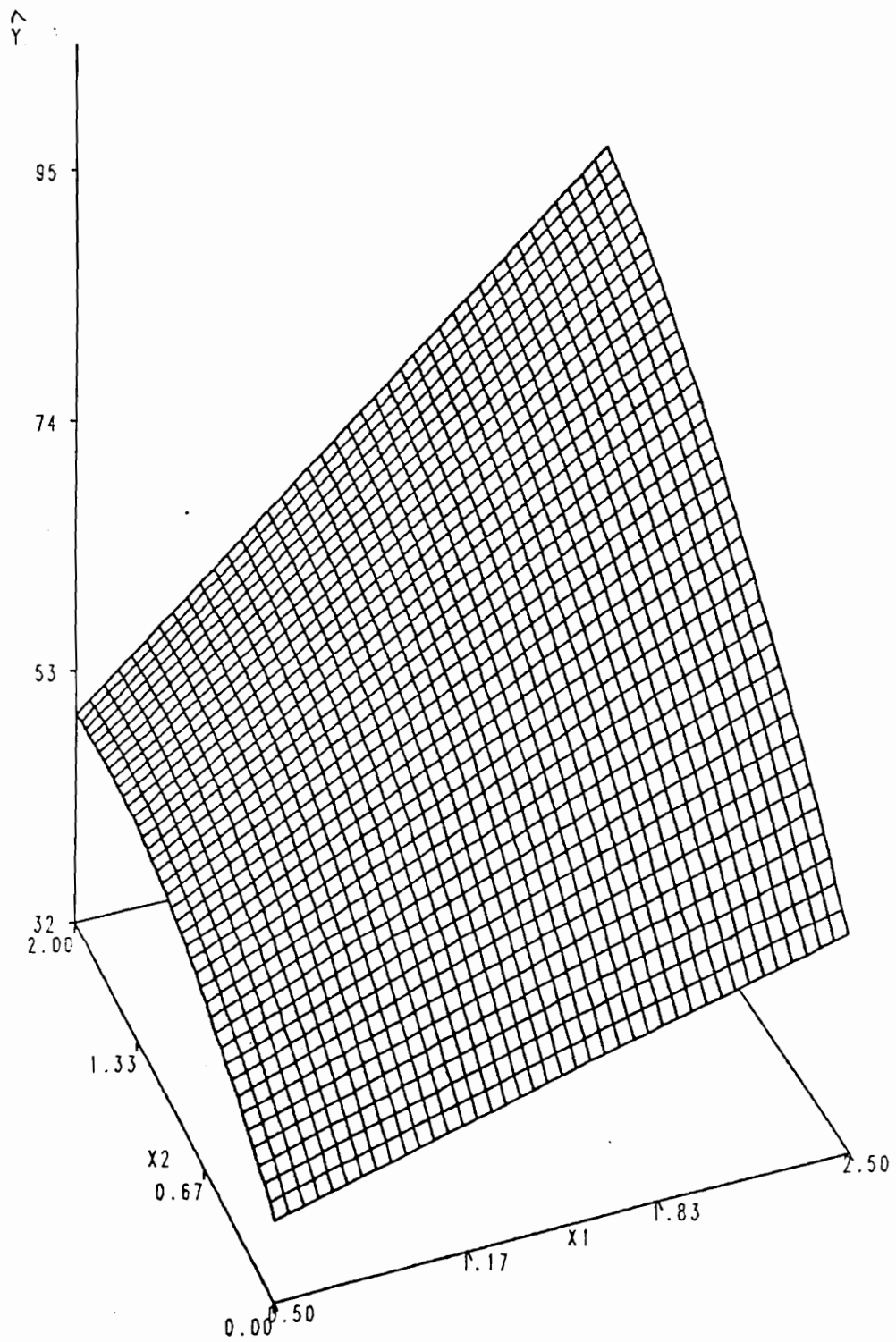


FIG. VI.3.14. FITTED 2ND ORDER L.S. MODEL  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2  
 168

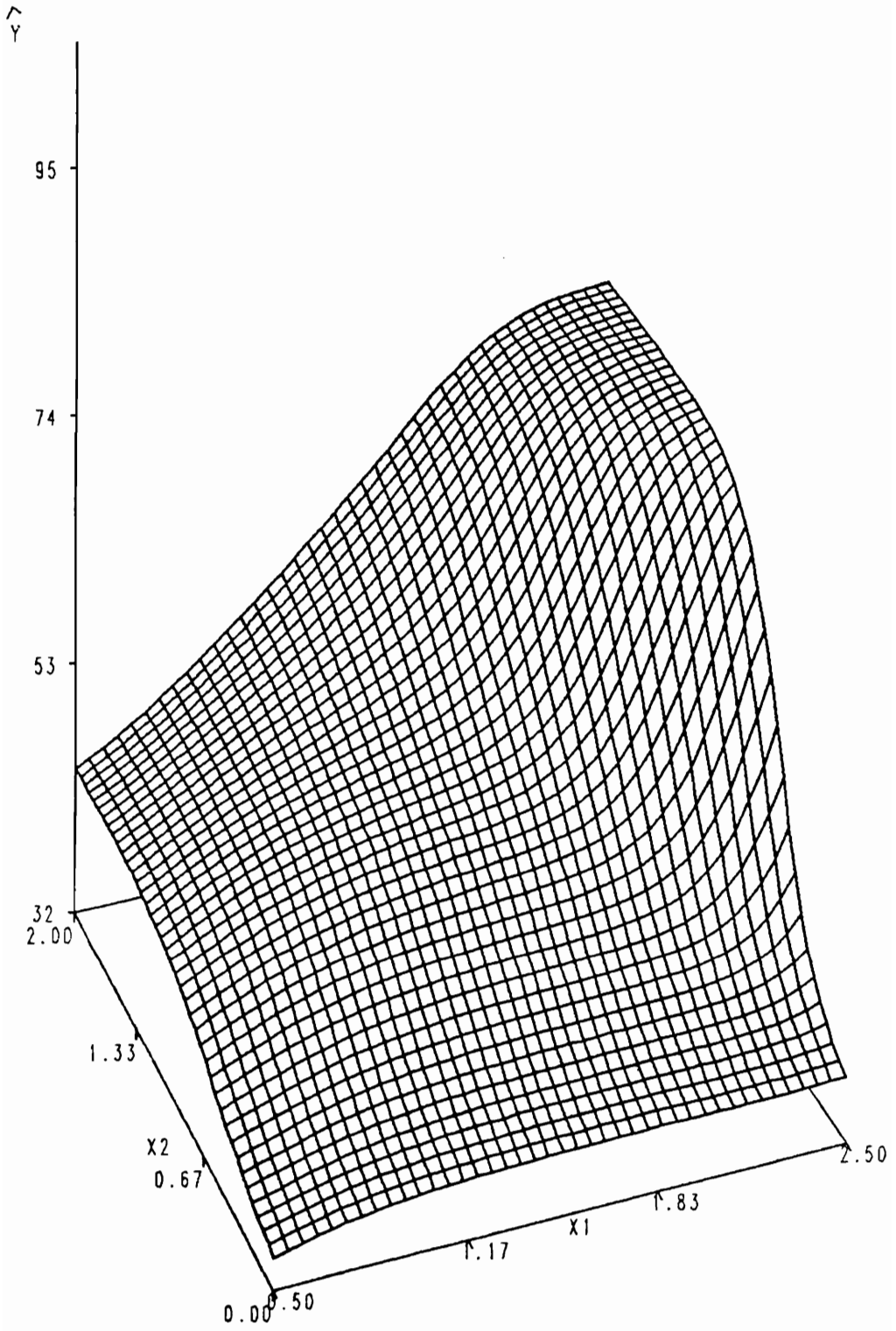
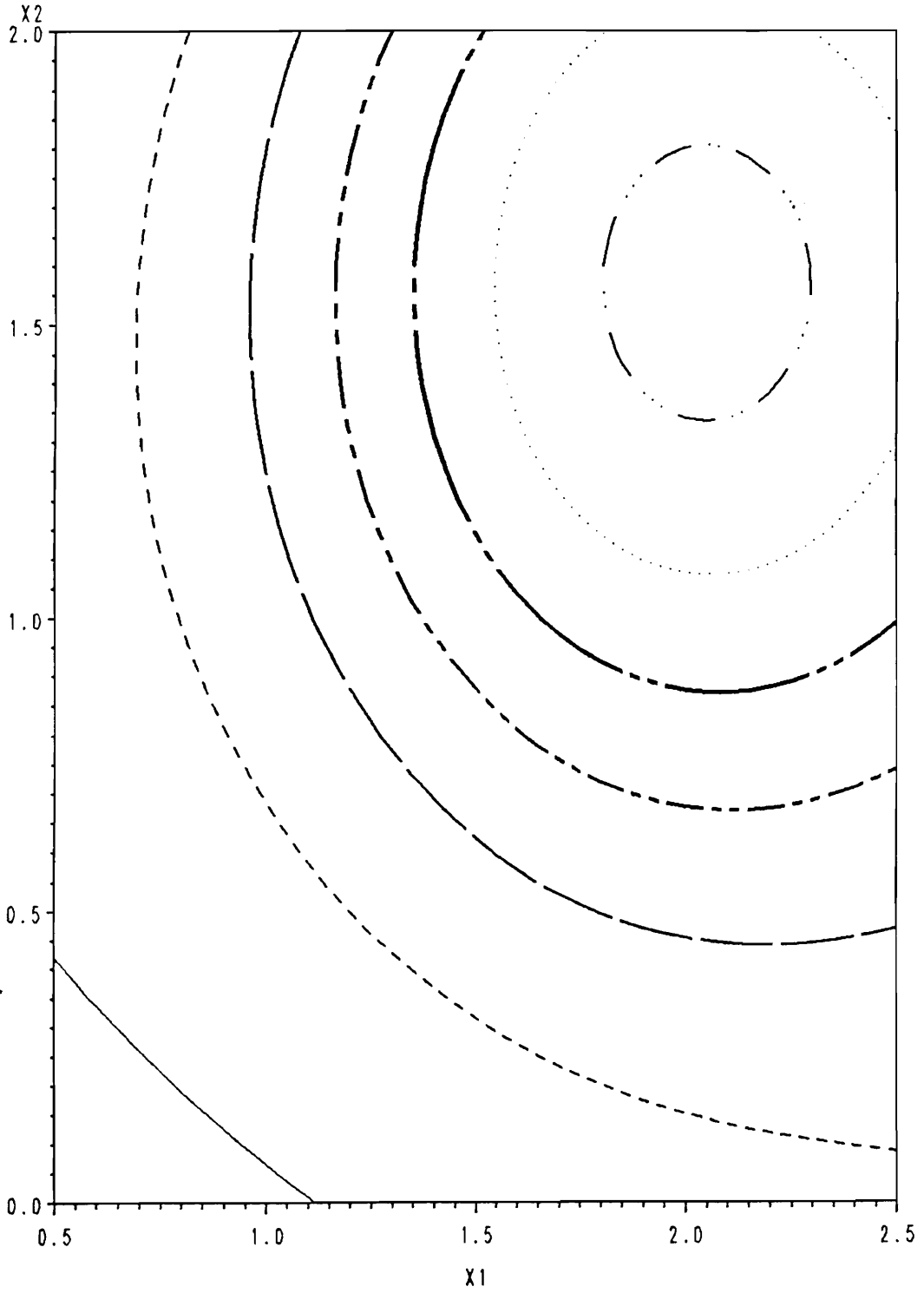
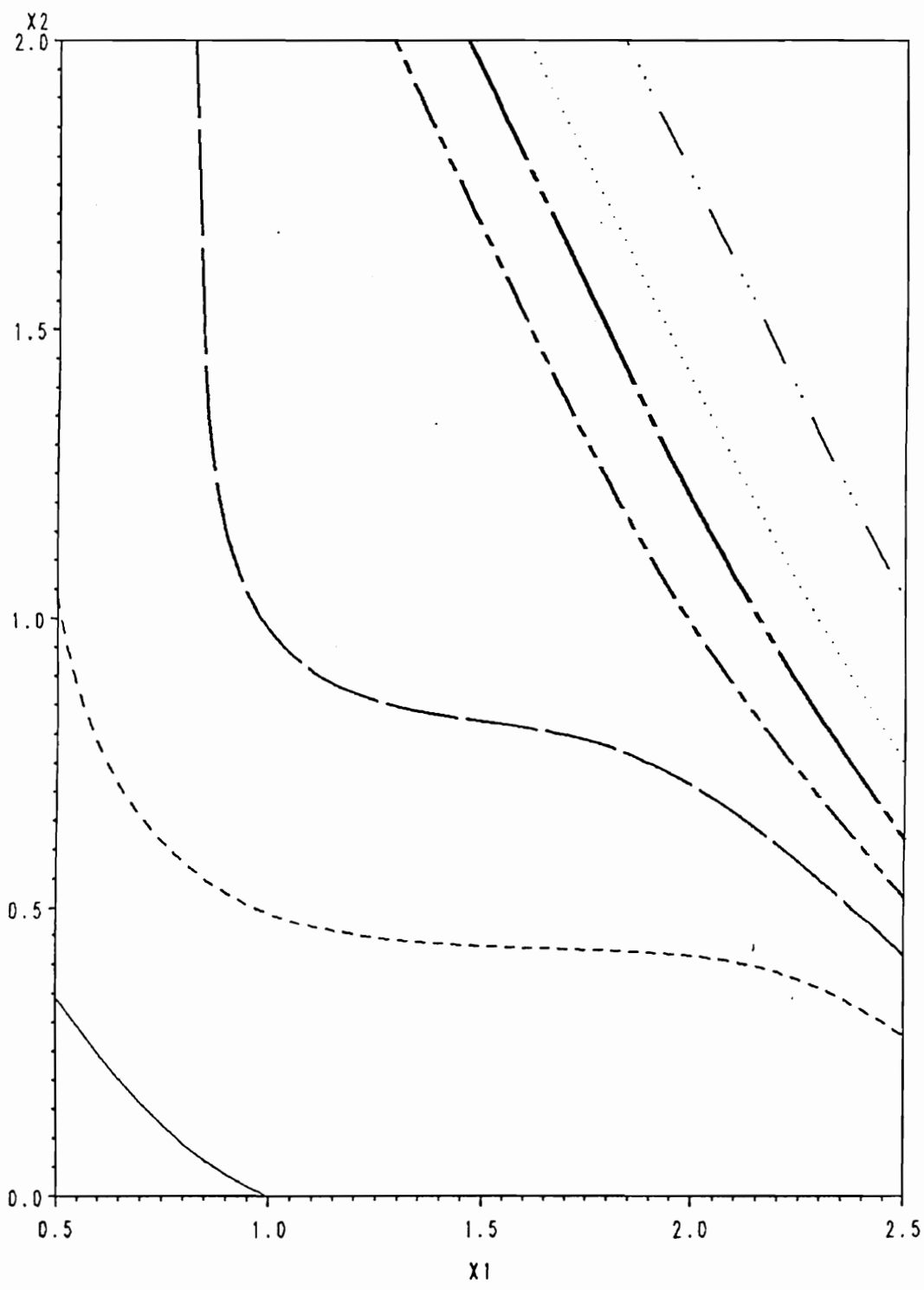


FIG. VI.3.15. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2.  $H=.605$



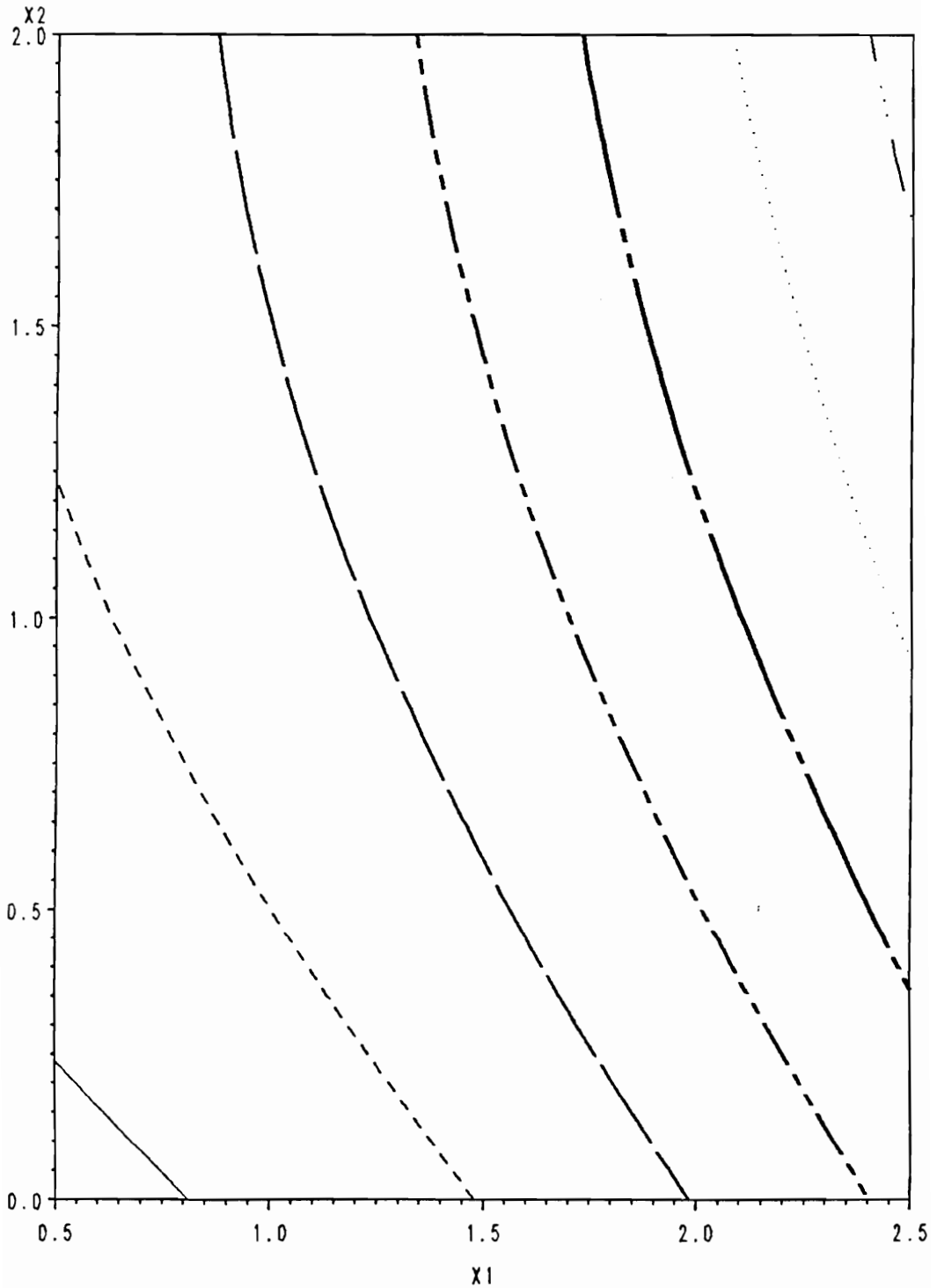


Y      ——— 35.98      - - - - 45.28      ——— 54.57      - - - - 63.86  
          - - - - 73.15      ····· 82.44      ····· 91.73  
 FIG. VI.3.16. TRUE RESPONSE SURFACE FOR EXAMPLE 2



YHAT.    ——— 35.10    - - - - 41.38    ——— 47.67    - - - - 53.95  
          - - - - 60.23    ······ 66.51    ······ 72.79

FIG. VI.3.17. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 INITIAL DATA FOR EXAMPLE 2. H=.641 & LAM=1



$\hat{Y}$

— 36.19	- - - 42.45	— 48.72	- - - 54.98
- - - 61.25	..... 67.51	- · - 73.78	

FIG. VI.3.18. FITTED 2ND ORDER L.S. MODEL  
INITIAL DATA FOR EXAMPLE 2

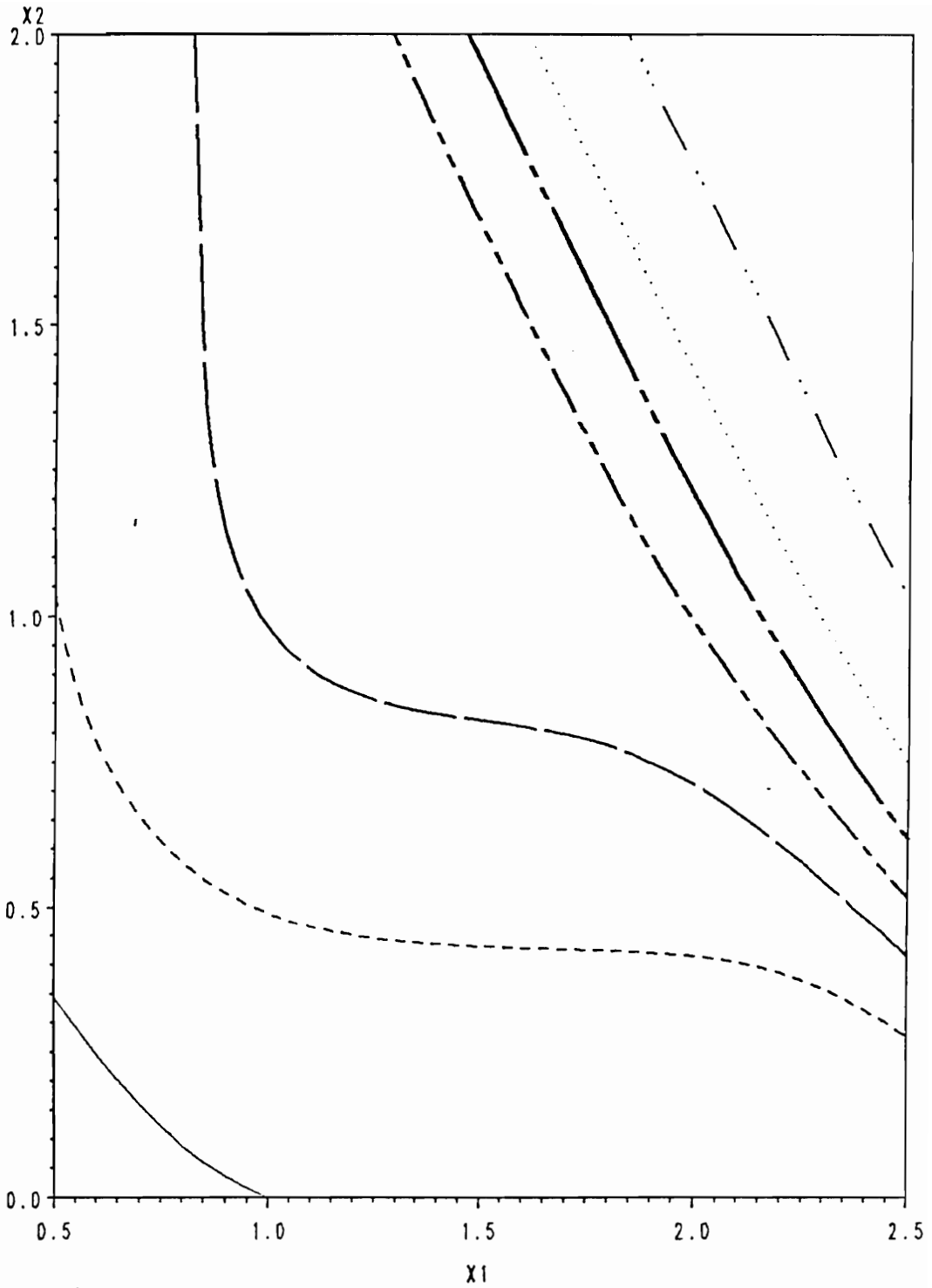
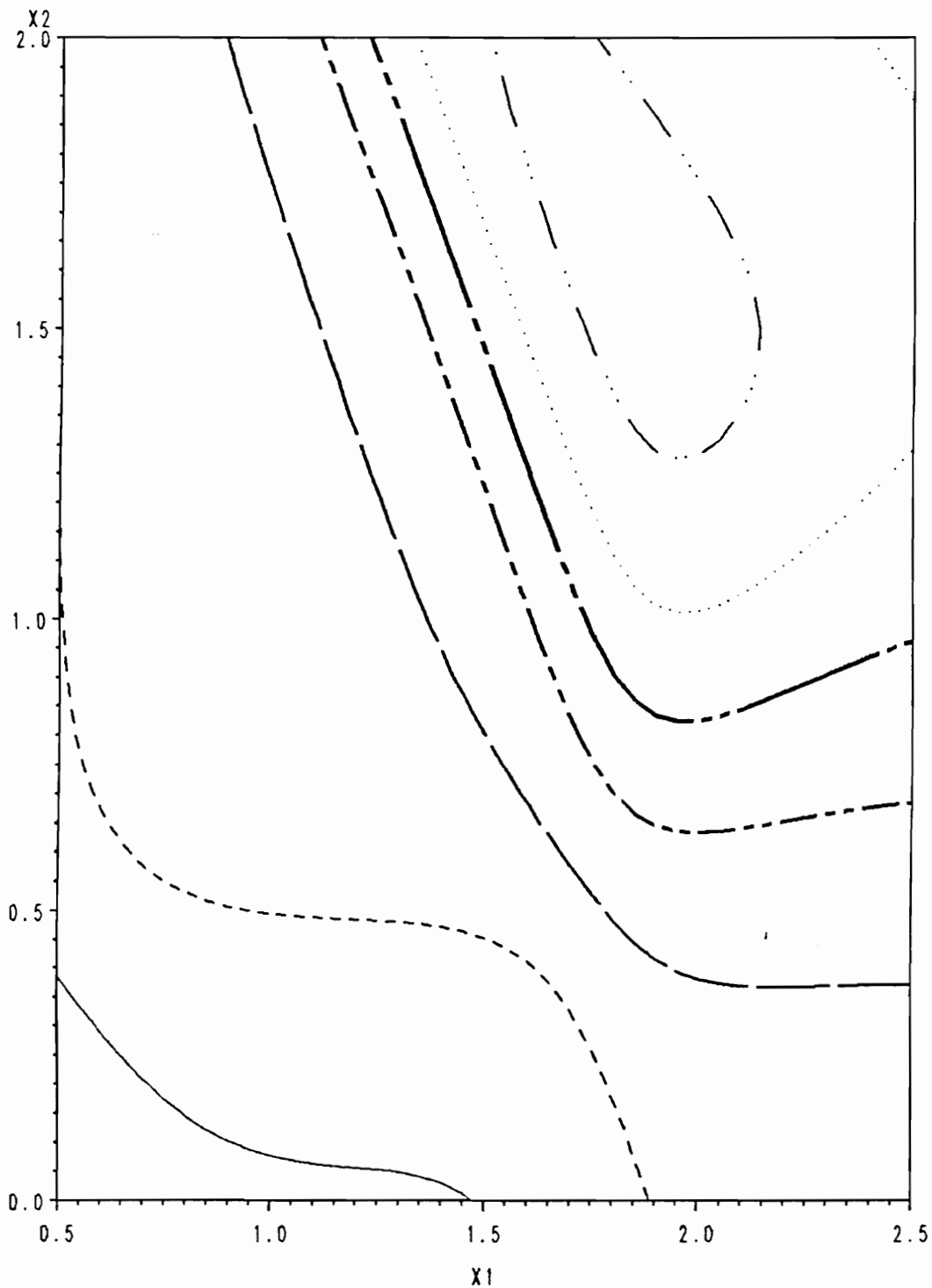


FIG. VI.3.19. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
INITIAL DATA FOR EXAMPLE 2. H=.641



YHAT    ——— 35.88    - - - - 44.31    - · - · 52.74    - - - - 61.17  
          - - - - 69.60    ······ 78.04    · · · · 86.47

FIG. VI.3.20. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 SIX POINTS AUGMENTED FOR EXAMPLE 2. H=.509 & LAM=1

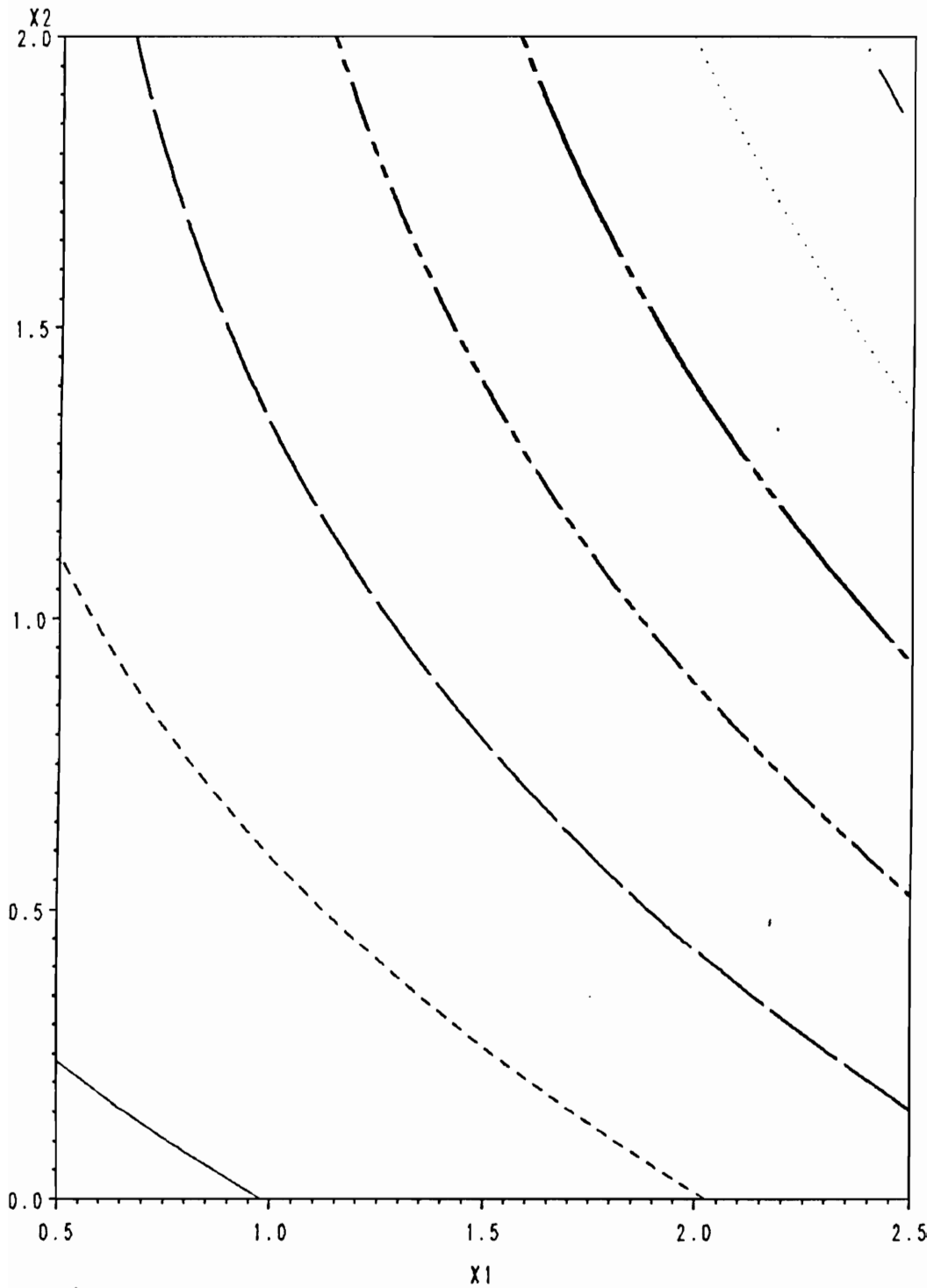
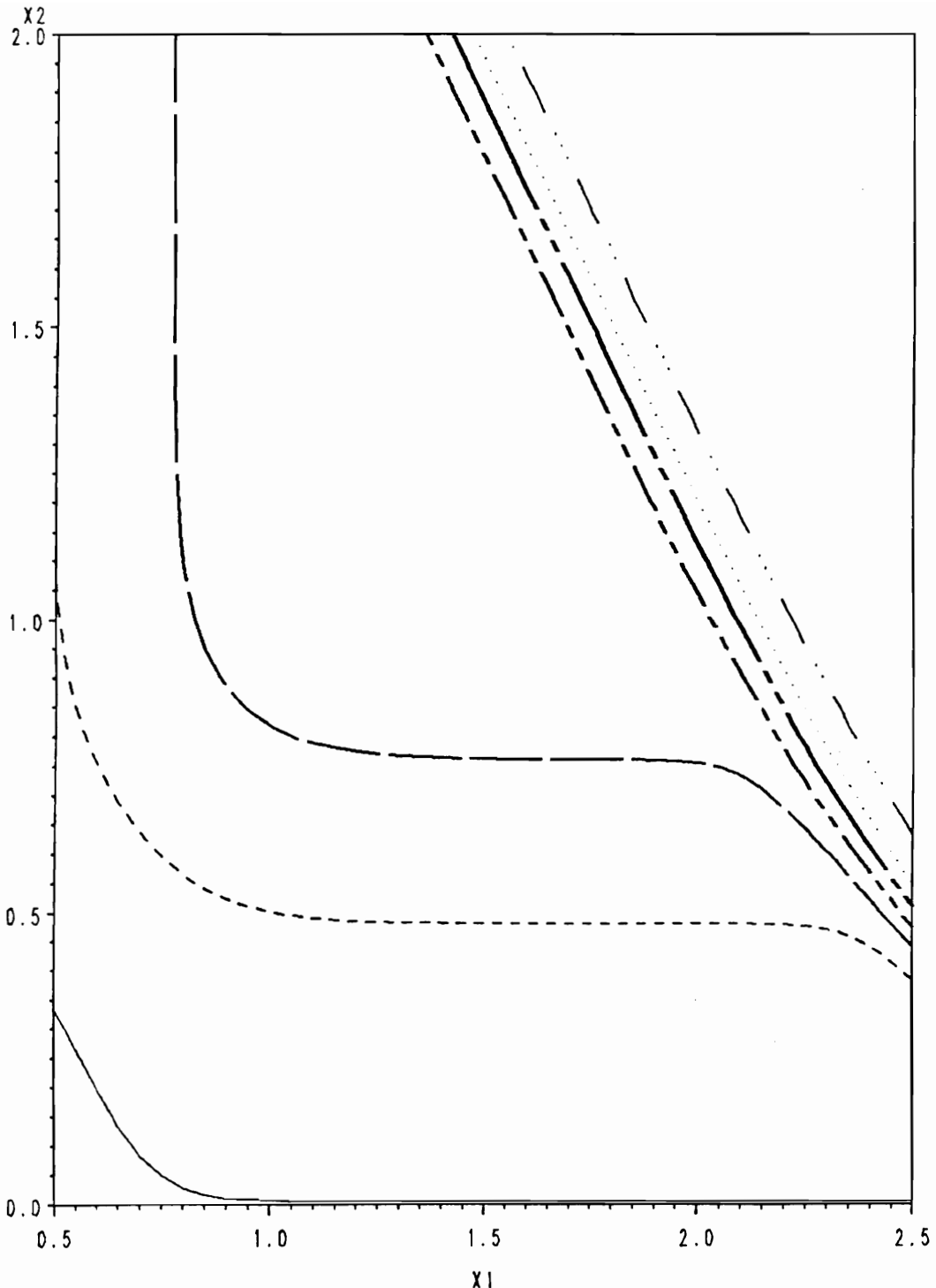
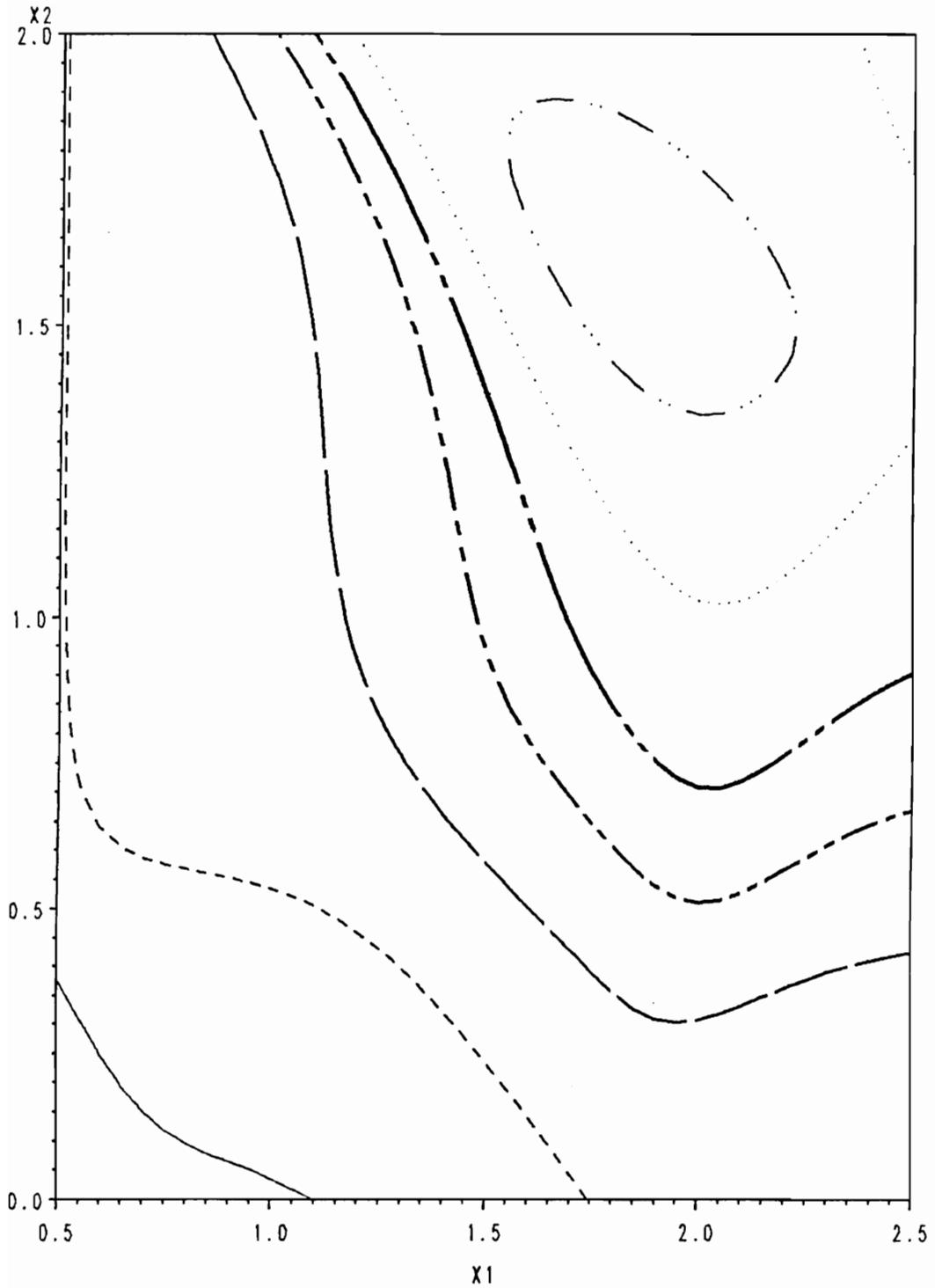


FIG. VI.3.21. FITTED 2ND ORDER L.S. MODEL  
SIX POINTS AUGMENTED FOR EXAMPLE 2



—	36.15	- - - -	42.47	—	48.80	- - - -	55.12
- - - -	61.44	· · · · ·	67.76	· · · · ·	74.08		

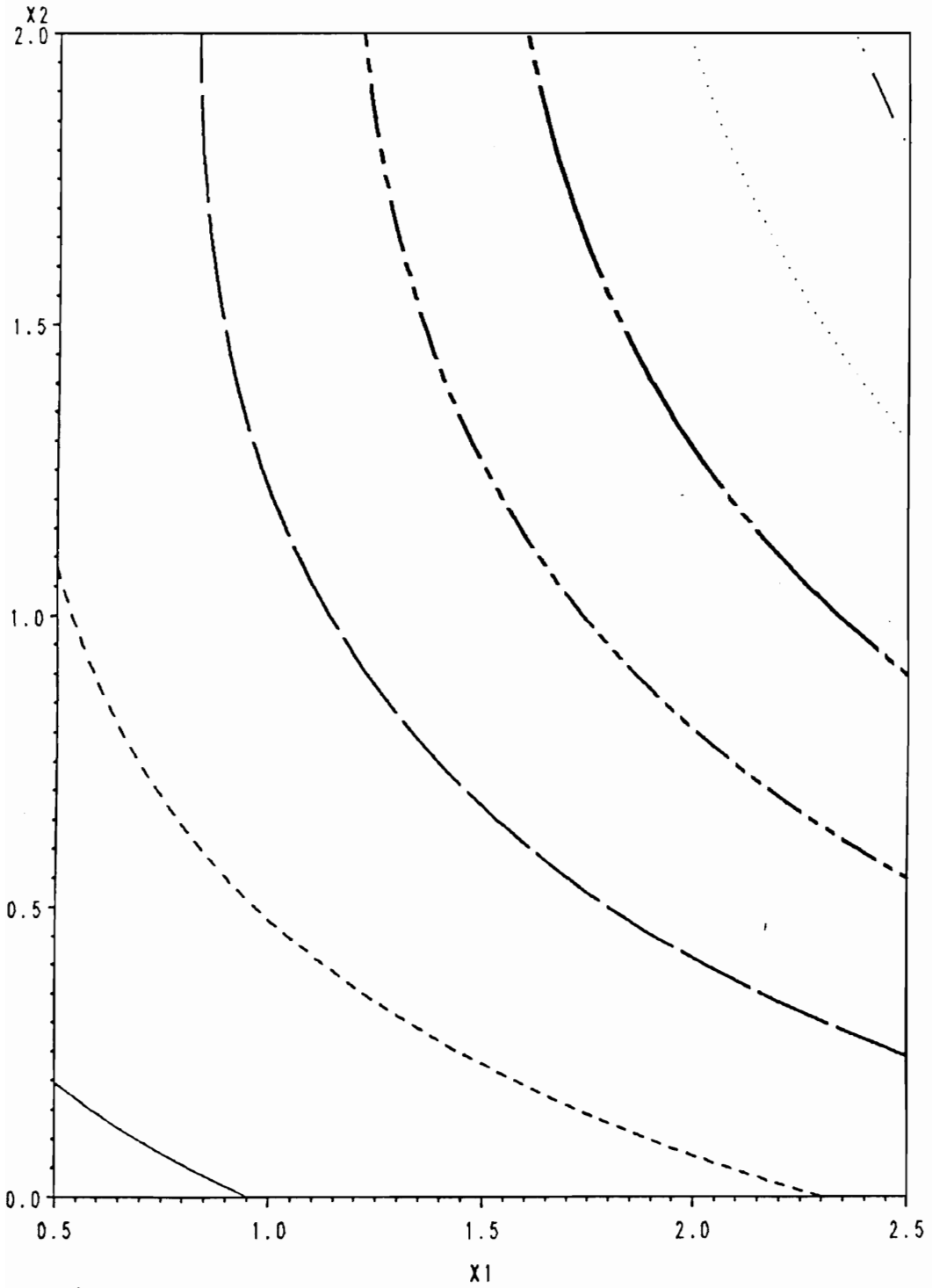
FIG. VI.3.22. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
SIX POINTS AUGMENTED FOR EXAMPLE 2. H=.371



YHAT    ——— 36.59    - - - - 44.96    — · — · 53.33    - - - - 61.70  
          - - - - 70.07    · · · · · 78.44    · · · · · 86.81

FIG. VI.3.23. FITTED HATLINK SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2. H=.401 & LAM=1





— 40.74    - - - - 48.20    — 55.66    - - - - 63.12  
 - - - - 70.57    ······ 78.03    ······ 85.49

FIG. VI.3.24. FITTED 2ND ORDER L.S. MODEL  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2

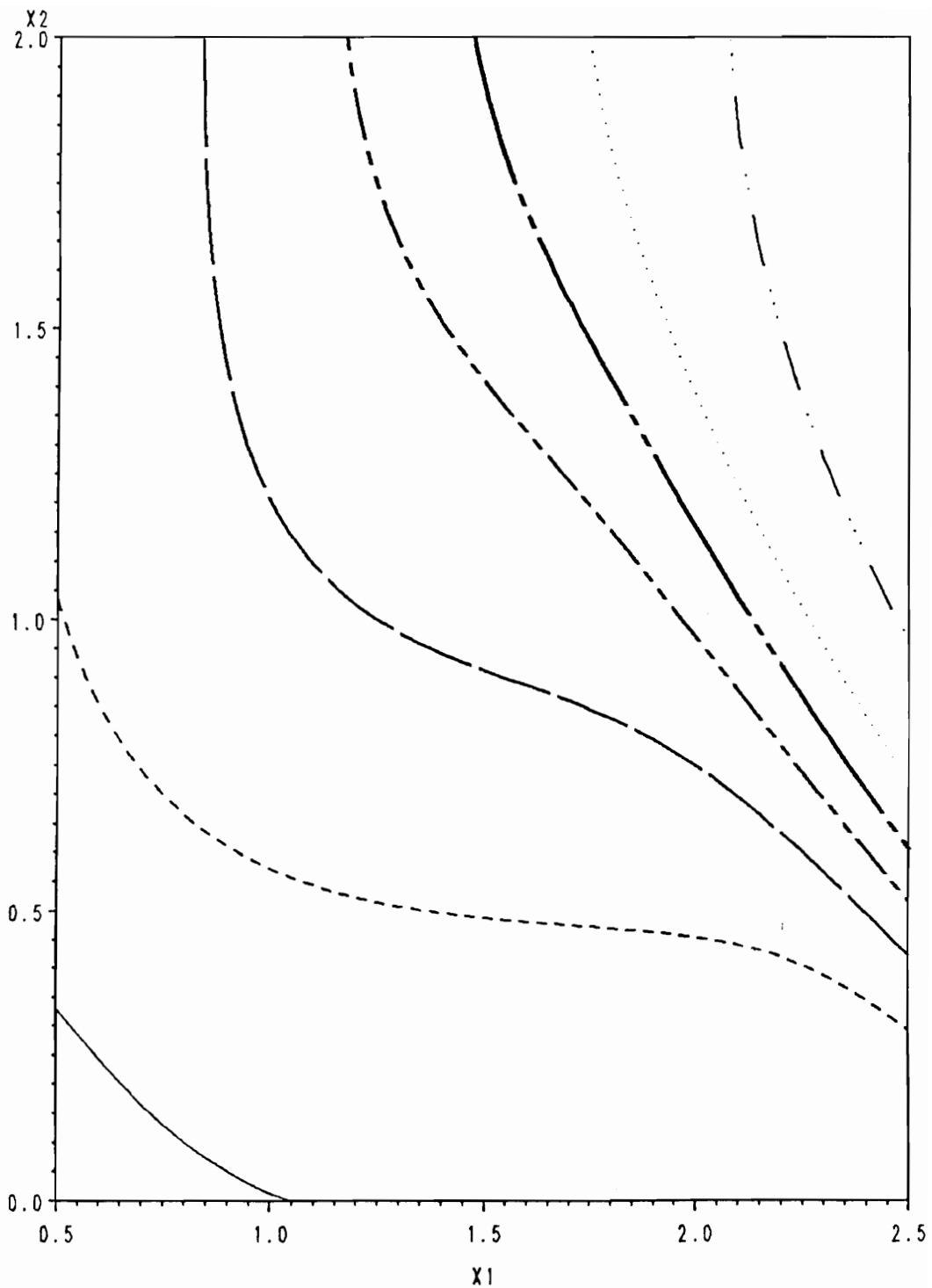


FIG. VI.3.25. FITTED KERNEL SURFACE WITH 2ND ORDER L.S.  
 TWELVE POINTS AUGMENTED FOR EXAMPLE 2.  $H=.605$

## VI.4 INTRODUCTION TO THREE DIMENSIONS

So far, all examples have been in two dimensions. When employing RSM techniques, many times the user is interested in three or more variables. The next section presents a three dimensional example of the HATLINK method of augmentation, although higher generalizations are possible. When using kernel regression for the nonparametric model, the effect of dimensionality upon kernel regression must be kept in mind when moving to higher values. It is a well known phenomena that as the dimension increases, kernel regression requires more points to adequately model the surface. For this work, the largest number of variables studied will be three. In moving to even higher dimensions, perhaps one might desire to start initially with more than one point in the new region of interest. Also, there may either be kernel functions other than the normal kernel function used in this work or other forms of nonparametric regression that would perform better in higher dimensions. These are areas for further research.

The HATLINK method of augmentation in higher dimensions may be easily extended from the two-dimensional program, available from the author, and Stroud's (1971) book on integral approximations. Using this book, an appropriate integral approximation for the desired dimension may be selected among many available choices. Besides altering the integral approximation, one must also revise the program to reflect the additional variables. The two-dimensional program was written to ease implementation when generalizing to higher dimensions.

Recall the basic assumptions of the HATLINK method on the behavior both in the old region of interest and in the overlapping area between the old and new regions. For simulations in two dimensions, along with various statistical tests, it is easy to generate pictures of the behavior of the underlying function in both the old and new regions of interest. Once points are

augmented, pictures and contour plots can be drawn of the fitted surfaces in the new region. Obviously these pictures are much harder to produce when dealing with more than two dimensions! This implies that statistical tests must be used exclusively to evaluate the fitted surfaces, for example some type of lack of fit test, and the behavior in the old region of interest to insure that the new region of interest is appropriate.

## VI.5 EXAMPLE IN THREE DIMENSIONS

This section contains an example in three dimensions of the HATLINK procedure. (The computer program for the three dimensional case is available from the author.) Just as in the previous chapter, this example also will illustrate the HATLINK augmentation procedure completely, except for the plots of the fitted surfaces and the contour plots. Ten points will be augmented with a batch size of one and the fitted maxima of the various surfaces for the initial data with five points augmented, and with ten points augmented will be located.

The effect of the extra dimension upon the number of candidate points must be kept in mind. In all of the two dimensional cases considered in this work, a grid size of 0.5 was used for each variable. This meant that for each dimension, the variable had five possible values and thus there were 25 total distinct candidate points considered. Using the same fineness of grid in the three dimensional case implies that there are now  $5^3$  or 125 candidate points! This tends to slow down the augmentation procedure somewhat, hence, for this example, only ten points were augmented instead of the twelve points augmented in the examples in the previous chapter.

Changing the grid size can also affect the magnitude of the ESE values which can be calculated for these examples. Recall the definition of the ESE value in (IV.2.1), which was the sum over the grid points of the squared difference between the true and fitted values. Altering

the grid size also varies the number of points and hence can also change the values calculated in (IV.2.1). Thus, using the same fineness of grid and with no scaling parameter of any sort, one would expect the ESE values for a three dimensional case to be larger than those for the two dimensional case. This behavior can be observed by comparing the ESE values for the examples in the previous chapters to the example presented in this chapter.

Instead of the example in two design variables presented originally in chapter two, suppose it is desired to now study the influence of temperature ( $X_1$ ), epoxy additive ( $X_2$ ), and baking temperature ( $X_3$ ) upon the strength of a particular plastic. The user still wishes to maximize the strength. Three levels of the variables were used in the experimental design. After centering and scaling, the design matrix,  $D$  and the corresponding vector,  $\underline{Y}$ , of observations are given in Table VI.5.1.

For the experiment described above, suppose that after applying steepest ascent and further considerations, the new region of interest was determined to be the cube with “front” corners at (2.5, 0, 0), (2.5, 0, 2), (2.5, 2, 0) and (2.5, 2, 2) and “back” corners at (0.5, 0, 0), (0.5, 0, 2), (0.5, 2, 0) and (0.5, 2, 2). Further, because some curvature of the response is expected in this new region, a second order model will be proposed by the user. It is desired to augment with new points in this region using the BIIV and select  $\lambda$  using the PRESS\* criteria. Because of kernel regression, one experimental run at (2.5, 2, 2) is performed and augmented to the original data set in Table VI.5.1 and a response value is obtained. This point was chosen because, based on distance, it is the farthest point from the old region of interest. Now one is ready to augment points using the BIIV criterion.

Suppose it is desired to augment by ten points using sequential design augmentation with a batch size of one and the HATLINK method. In this case, response values will be immediately observed, allowing one to re-evaluate the bandwidth and  $\lambda$  values after augmenting

with each point. The new data set used is the original data in Table VI.5.1 plus the response at (2.5, 2, 2). For comparison purposes, the true model is assumed known (but not to the researcher) to be

$$Y = x_1^2 + x_1 - 3(x_2 - 1)^2 - .5(x_3 - 2.5)^2 + 35 + .4 \left\{ 5 \left[ \sin \left( \frac{\pi * x_1 * x_2}{1.6} \right) \right] - 5 \left[ \sin \left( \frac{\pi * x_1 * (x_3 - 1.5)}{1.5} \right) \right] \right\} + \epsilon, \quad (\text{VI.5.2})$$

where  $\epsilon \sim N(0,1)$ .

For (VI.5.2) the true maximum is 45.92 and is located at (2.5, 1.50, 1.20). After each point is augmented, the fitted maxima for the HATLINK, least squares and kernel surfaces and the distances between these maxima and the true maximum will be calculated.

Augmentation of ten points yielded the results presented as follows: HATLINK in Table VI.5.2, a second order least squares model in Table VI.5.3, and kernel regression in Table VI.5.4. By drawing cubic regions, plots of the augmented points similar to those presented in Section IV.3 could be created, although this will not be performed here. The important things to notice are the ESE values, the fitted maxima and the distances to the true maximum. The HATLINK and least squares methods performed about equally well in finding what the true maximum should be, and both methods outperformed kernel by a small amount. In this case, the kernel method did slightly better, but not enough to be meaningful.

Notice that the distances between the fitted and true maximum for least squares regression is slightly less than the distance for the HATLINK method. In this case, the proposed method is doing better than least squares in terms of estimating the maximum, but it is not doing quite as well at estimating the location. Both techniques outperformed kernel regression in

Table VI.5.1. Design Matrix and Response Vector for Initial Region of Interest – Three Variables.

$D =$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ -1 & -1 & 1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$Y =$	$\begin{pmatrix} 39.2891 \\ 38.3100 \\ 31.4340 \\ 30.7222 \\ 16.6750 \\ 16.0754 \\ 28.8650 \\ 32.9284 \\ 21.6024 \\ 23.1364 \\ 22.0087 \\ 22.1389 \\ 28.6437 \\ 28.9848 \\ 20.5150 \\ 20.9731 \\ 30.1289 \\ 33.5513 \\ 21.1491 \\ 18.9562 \\ 28.5036 \\ 32.3872 \\ 28.5435 \\ 30.6702 \\ 30.4563 \\ 30.5255 \\ 27.4074 \\ 26.2458 \\ 30.3584 \\ 29.7175 \\ 27.9672 \\ 29.6356 \end{pmatrix}$
-------	--	-------	--

terms of locating the maximum. Again, keep in mind the effect of both the dimensionality and the fineness of the grid in computing and comparing distances.

After ten points are augmented, HATLINK has a better overall fit than either the least squares or kernel methods, and would be the best one to use in terms of both overall performance and finding the optimal range of operating conditions as described above. Note that initially HATLINK estimated  $\lambda$  to be one. This is obviously not optimal as seen by the ESE values. But recall that there is very little information in the new region of interest at this stage. It was speculated that HATLINK would have some trouble estimating  $\lambda$  values at the beginning. This is the reason why updating the parameters with a batch of size of one was expected to perform the best. Based upon the work in two dimensions, it is suspected that pictures and contour plots, if one could draw such plots, would illustrate that the HATLINK method outperforms both least squares and kernel methods. Just as in the two dimensional examples considered in previous chapters, the proposed technique appears to be drawing from both least squares and kernel methods to produce a better overall fit.



**Table VI.5.2. Ten Points Augmented Using IIATLINK and BIIV.** Using  $\sigma = 1$  and  $CT = 0.4$ , this lists the points augmented, the bandwidth and lambda values selected and the ESE values. The bandwidths, lambdas and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS refers to the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>X3AUG</u>	<u>BAND</u>	<u>LAM</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	–	–	–	0.657	1.000	2024.5	–	–
1	1.0000	1.0000	1.0000	0.663	1.000	1793.3	39.00	1.658
2	2.0000	2.0000	2.0000	0.668	0.504	1179.5	39.87	1.000
3	0.5000	2.0000	1.5000	0.707	0.454	1316.6	39.67	1.000
4	1.0000	2.0000	0.5000	0.696	0.436	1321.8	39.51	1.000
5	1.0000	0.5000	2.0000	0.710	0.454	1368.0	39.60	1.000
6	2.5000	1.5000	2.0000	0.717	0.350	1114.9	41.29	1.000
7	2.5000	0.0000	2.0000	0.719	0.473	835.8	42.12	1.118
8	2.5000	0.0000	1.0000	0.716	0.634	525.5	42.26	1.118
9	2.5000	0.0000	0.0000	0.732	0.635	484.1	42.44	0.707
10	2.5000	2.0000	0.0000	0.748	0.592	492.2	43.53	0.707

**Table VI.5.3. Ten Points Augmented Using Least Squares and BIIV.** Using  $\sigma = 1$  and  $CT = 0.4$ , this lists the points augmented and the ESE values. The ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS refers to the distance between the location of the fitted and true maxima.

<u>NUM.</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>X3AUG</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. design	–	–	–	1218.0	–	–
1	0.5000	2.0000	2.0000	1489.7	41.47	1.000
2	1.0000	2.0000	0.5000	1427.6	40.88	1.000
3	2.0000	0.0000	2.0000	1223.4	41.79	1.000
4	2.5000	1.0000	2.0000	1250.1	41.69	1.000
5	2.5000	0.0000	1.0000	672.0	43.32	0.500
6	2.5000	1.5000	0.0000	819.4	44.67	0.000
7	2.5000	1.5000	1.5000	799.0	44.48	0.000
8	1.5000	0.5000	0.0000	792.1	44.35	0.000
9	1.0000	0.5000	2.0000	677.0	44.08	0.500
10	2.0000	2.0000	0.0000	646.9	43.96	0.500

**Table VI.5.4. Ten Points Augmented Using Kernel and BHIV.** Using  $\sigma = 1$  and  $CT = 0.4$ , this lists the points augmented, the bandwidth values selected and the ESE values. The bandwidths and ESE values are calculated after the given point was augmented. MAX refers to the maximum fitted value and DIS refers to the distance between the location of the fitted and true maxima.

<u>NUM</u>	<u>X1AUG</u>	<u>X2AUG</u>	<u>X3AUG</u>	<u>BAND</u>	<u>ESE</u>	<u>MAX</u>	<u>DIS</u>
org. des.	–	–	–	0.657	2024.5	–	–
1	1.0000	1.0000	1.0000	0.663	1793.3	39.00	1.658
2	2.0000	2.0000	2.0000	0.668	1747.2	38.94	1.500
3	0.5000	1.0000	1.0000	0.680	1688.8	38.84	1.000
4	2.5000	2.0000	2.0000	0.672	1649.4	38.86	0.707
5	1.0000	0.5000	0.5000	0.666	1322.6	38.85	1.000
6	1.0000	0.5000	1.0000	0.651	1187.8	39.06	1.224
7	1.0000	0.0000	0.5000	0.629	1230.3	39.05	1.732
8	1.5000	2.0000	2.0000	0.620	1082.1	38.94	1.732
9	1.0000	1.0000	1.5000	0.623	1154.7	39.42	1.732
10	1.5000	1.5000	2.0000	0.611	1186.7	39.04	2.061

## VI.6 CONCLUSIONS

The examples in this chapter illustrate several aspects about the proposed augmentation method. Most importantly, they indicate why the HATLINK method as proposed in this work makes a significant contribution to augmentation techniques. As seen in the examples, the new method can often produce a better overall fit than currently used techniques of augmentation. This is especially important in the case of small amounts of model misspecification that would not be detected by customary techniques. While sometimes the HATLINK surface does not perform noticeably better than least squares or kernel regressions in finding the value or location of the maximum (as seen in the first example), it may produce a better overall fit. It has been demonstrated how the new technique is able to make use of contributions from both the parametric model through least squares and the nonparametric model through kernel regression to produce a better result.

The examples presented in this chapter also demonstrate the complete augmentation procedure. This procedure includes establishing both old and new regions of interest, as well as selecting an initial point in the new region of interest before implementation of the selected augmentation procedure. In this chapter, the same new region of interest and initial point was used for the two dimensional examples. This was to illustrate that the HATLINK procedure works in the same region with different types of misspecification in the true model

While only one example of the HATLINK method in three dimensions has been presented here, other examples could have been included. In the example presented in this chapter, the new regression technique appeared to outperform both the least squares and kernel methods of augmentation in terms of both estimating what the maximum should be and the ESE values. In terms of the distance between the true and fitted maximum, HATLINK did not

perform quite as well as the other two methods, although the distances were not too different. The HATLINK method should produce fitted surfaces that are usually better than both least squares and kernel regressions, or at least no worse.

# CHAPTER VII

## SUMMARY & AREAS FOR FURTHER RESEARCH

### VII.1 SUMMARY OF THE PERFORMANCE OF AUGMENTATION WITH HATLINK

Current methods of design augmentation use model-based approaches and assume there is minimal model misspecification. Using HATLINK to augment design points has several desirable features. It is more robust to model misspecification than the model-based approaches. The value of the mixing parameter appears to be indicative, in some sense, of the degree of model misspecification present in the user's prescribed model. The HATLINK method should be a reasonable compromise between the least squares and kernel methods.

In Chapter III, several ways of selecting  $\lambda$  were stated. After extensive study, the PRESS\* method was determined to be superior over a broad range of conditions. If the user is confident that the specified model is a good approximation to the true underlying model, then  $C_p3(\lambda)$  is recommended. A PRESS-type method was also used to estimate the bandwidth parameter.

Two main methods of augmentation, integrated prediction variance and BIIV, were studied. The simulation results presented in Chapter IV indicated that the BIIV method was superior to integrated prediction variance. Recall that BIIV attempts to incorporate a measure of bias and should perform better when there is more model misspecification. The results from

the repetitions performed in Chapter V reflect the relatively consistent results from the HATLINK method of augmentation. Recall that this consistency was important because of the stochastic nature of the proposed method of augmentation.

The effect of batch size was also studied in Chapter V. The batch size determines the frequency with which the bandwidth and  $\lambda$  parameters are updated. Thus using a batch size of one should yield the most immediate gain in terms of ESE values when augmenting with points since the two parameters are updated the most frequently. In the examples presented,  $h$  and  $\lambda$  both appear to converge separately to single numbers. With the augmentation of enough points, approximately two to four batches, the effect of batch size appears to diminish in the sense that the fits as measured by the ESE values are approximately the same.

Chapter VI examined two different families of model misspecification, one misspecified due to sinusoidal terms and the other due to exponential terms. For each of the families examined, a similar pattern of prediction performance was observed for HATLINK, least squares and kernel regressions. In all cases, the quality of fit as measured by the ESE values was superior for HATLINK after one or two points were augmented. As conjectured at the beginning, when the user's postulated model was a close approximation to the true underlying function, then the HATLINK method tended to select  $\lambda$  values emphasizing least squares regression. In such situations, the performance of kernel regression was the poorest of the regression techniques examined. When the degree of model misspecification was more extreme, least squares did not perform as well as HATLINK, which emphasized kernel regression over least squares regression. The results presented in Chapter VII further illustrate the desirability of the proposed HATLINK method of augmentation.

Most examples considered were two-dimensional in nature. The proposed method is easily extended to higher dimensions. Chapter VI also discussed the changes necessary for

moving to three dimensions and presented an example of HATLINK with three variables. One obvious disadvantage to higher dimensions is the complexity involved in presenting results in graphical form.

The HATLINK method as described is recommended for situations where the user has proposed what appears to be a reasonable parametric model, but suspects there might be some misspecification involved. The type and degree of model misspecification present is not so severe that customary lack of fit tests would detect the problem. Both old and new regions of interest must be specified as well as a design in the old region of interest and an initial point augmented in the new region. The new region of interest is expected to contain the maximum value and there is at least a small overlap between the old and new regions. The batch size and number of points to be augmented must be determined by the user. Under these circumstances, HATLINK not only appears to draw on both least squares and kernel regressions when fitting surfaces, but also detects behavior not found in either regression technique.

## **VII.2 AREAS FOR FURTHER RESEARCH**

In the proposed method of augmentation, several areas for potential improvement exist. One main area is the selection of the mixing parameter. As indicated in Section A3.1, sometimes the  $\lambda$  values selected by any of the proposed criteria do not perform equally well over a broad range of conditions such as model misspecification. Perhaps some other unknown criterion for selecting  $\lambda$  would achieve better results.

A second area for improvement lies with the augmentation criterion. As stated in Section A3.2, in many cases HATLINK yielded the lowest ESE values but failed to outperform either of the other two regression techniques when locating the maximum value. Several versions



of augmentation with BIIV were proposed and briefly examined in Appendix 3, although none appeared to perform better than the original BIIV. There may be some alteration to BIIV or some improvement to estimating the bias that would improve all components of the augmentation procedure. The BIIV method was used extensively because it appeared to work the best, but there may be some completely different augmentation criterion that would perform better in terms of both the overall fit and in locating the maximum fitted value.

As initially proposed, HATLINK is a combination of parametric and nonparametric techniques, although only the parametric least squares regression and the nonparametric kernel regression were studied. Other parametric and nonparametric methods could be examined in the future. Several nonparametric techniques were briefly described in Section III.3, for example local linear regression and nonparametric spline regression. The use of alternatives to least squares and kernel regressions might lead to improvements in the HATLINK method.

If kernel regression is used in the proposed augmentation method, any improvement to the kernel component should yield better results for HATLINK. This is especially true when the user's model is misspecified. Here, only a normal kernel function was used in order to keep computational details at a reasonable level, other kernel functions may exist that would perform better, especially in higher dimensions. Only a Euclidean measure was used to define the distance  $\|\underline{X}_i - \underline{X}_j\|$ . Perhaps a different way of defining distance would be more beneficial.

A global bandwidth, the only type studied here, uses the same value for all the regressors and thus is easier to use when programming. Furthermore, a global bandwidth does not consider cases where the response varies rapidly in one regressor, but slowly in another. In such a situation, it would be more desirable to use a narrow bandwidth for one variable and a wider bandwidth for the other variable. A variable bandwidth approach would be more desirable, although harder to implement.

One problem is measuring the goodness of fit for the HATLINK method if the true equation is not known. In simulation studies, calculating the ESE values as described is fairly easy, but is obviously harder when the true equation is unknown. Diagnostics for judging the goodness of fit in a general situation would be desirable.

Another variation of the HATLINK augmentation procedure is to start with more than one point in the new region. Because of the distance-measuring nature of kernel regression, one point augmented initially in the new region may not be enough for the HATLINK method to adequately estimate the surface. Starting with several distinct points in the new region might enable HATLINK to better fit the surface. A reasonable question to ask is if there is some optimal number and placement of points for a given situation. This would be especially important if there is a significant amount of model misspecification in the new region of interest.

In response surface methodology, it often occurs that the user has more than three variables, the highest number studied here. The proposed technique can be fairly easily extended to more variables, and the behavior of HATLINK should be studied in these cases. In dealing with higher dimensions, one should keep in mind the effect of dimensionality upon kernel regression. With enough variables, this will become a significant problem. Improvements in the nonparametric component that minimize the effect of dimensionality should yield better results, especially for multiple variables.

As can be seen, the HATLINK method of augmentation proposed here appears to work for the cases studied, although several changes are possible which may yield improvements. The HATLINK method using BIIV performs as expected and appears to be a “model-robust” method for augmenting design points in some new region of interest.

## CHAPTER VIII

### BIBLIOGRAPHY

Allen, D. A. (1974) "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction," *Technometrics*, 16, 125-127.

Benedetti, J. K. (1975) "Kernel Estimation of Regression Functions," *Proceedings of the Computer Science and Statistics Eighth Annual Symposium on the Interface*, 405-408.

Butler, G. A. (1975) "Heuristic Regression for Large Commercial Problems," *Proceedings of the Computer Science and Statistics Eighth Annual Symposium on the Interface*, 398-404.

Cheng, K. F. and Lin, P. E. (1981) "Nonparametric Estimation Function," *Z. Wahr, und Verw. Gebiete*, 57, 223-233.

Cleveland, W. S. (1979) "Robust Locally Weighted Regression and Smoothing Scatterplots," *J. Amer. Statistical Assoc.*, 74, 828-836.

Davies, Owen L. (1956) *The Design and Analysis of Industrial experiments*, Hafner Publishing Co., New York, New York.

Dwight, Herbert B. (1961) *Tables of Integrals and Other Mathematical Data*, Macmillan Publishing Co., Inc., New York, New York.

Dykstra, O., Jr. (1965) "The Orthogonalization of Undesigned Experiments," *Technometrics*, 8, 279-290.

Dykstra, O., Jr. (1971) "The Augmentation of Experimental Data to Maximize  $|X'X|$ ," *Technometrics*, 13, 682-688.

Einsporn, R. L. (1987) *HATLINK: A Link Between Least Squares Regression and Nonparametric Curve Estimation*, unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Evans, J. W. (1979) "Computer Augmentation of Experimental Designs to Maximize  $|X'X|$ ," *Technometrics*, 21, 321-330.

- Federov, V. V. (1972) *Theory of Optimal Experiments*, translated and edited by W. J. Studden and E. M. Klimko, Academic Press, New York, New York.
- Gasser, T., and Müller, H. G. (1979) "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.), 23-68 Springer-Verlag, Heidelberg.
- Gaylor, D. W. and Merrill, J. A. (1968) "Augmenting Existing Data in Multiple Regression," *Technometrics*, 10, 73-81.
- Geisser, S. (1975) "The Predictive Sample Reuse Method with Applications," *J. Amer. Stat. Assoc.*, 70, 320-328.
- Georgiev, A. A. , and Greblicki, W. (1986) "Nonparametric Function Recovering from Noisy Data," *J. Stat. Planning and Inference*, 13, 1-14.
- Giovannitti-Jensen, A. and Myers, R. H. (1989) "Graphical Assessment of the Prediction Capability of Reponse Surface Designs," *Technometrics*, 14, 767-779.
- Hebble T. L. and Mitchell, T. J. (1972) "Repairing Response Surface Designs," *Technometrics*, 14, 767-779.
- Hussey, J., Myers, R. H. and Houck, E. (1987) "Correlation Simulation Experiments in First Order Response Surface Designs", *Operations Research*, 35, 744-758.
- Kiefer, J. (1959), "Optimum Experimental Designs" *Journal of the Royal Statistical Society*, 21, 272-319.
- Kiefer, J. and Wolfowitz, J. (1960) "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.
- Kiefer, J. and Wolfowitz, J. (1959) "Optimum Designs in Regression Problems", *Annals of Mathematical Statistics*, 30, 271-294.
- Mallows, C. L. (1973) "Some Comments on  $C_p$ ," *Technometrics*, 15, 661-675.
- Müller, H. G. (1987) "Weighted Local Regression and Kernel Methods for Nonparametric Curve Fitting," *J. Amer. Statistical Assoc.* , 82, 371-386.
- Müller, H. G., and Stadtmüller, U. (1985) In Discussion of Silverman, *J. Royal Stat. Soc. B*, 47, 39.
- Müller, H. G., and Stadtmüller, U. (1987) "Variable Bandwidth Kernel Estimators of Regression Curves," *Annals of Statistics*, 15, 182-201.
- Myers, R. H. (1976) *Response Surface Methodology*, distributed by Edwards Brothers, Inc., Ann Arbor, Michigan.
- \_\_\_\_\_ (1990) *Classical and Modern Regression with Applications second edition*, Duxbury, Boston.

- Nadaraya, E. A. (1964) "On Estimating Regression", *Theory of Prob. and its Applications*, 9, 141-142.
- Priestley, M. B., and Chao, M. T. (1972) "Nonparametric Function Fitting," *J. Royal Stat. Soc. B.*, 34, 384-392.
- Rice, J. (1984) "Bandwidth Choice for Nonparametric Regression," *Annals of Statistics*, 12, 1215-1230.
- Rozum, M. (1990) *Effective Design Augmentation for Prediction*, unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- SAS Institute, Inc. (1988) *SAS/GRAPH User's Guide, Release 6.03 Edition*. Cary, NC, SAS Institute Inc.
- SAS Institute, Inc. (1985) *SAS/IML User's Guide, Version 5 Edition*. Cary, NC, SAS Institute Inc.
- Silverman, B. W. (1985) "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting (with discussion)," *J. Royal Stat. Assoc. B.*, 47, 1-53.
- Stone, M. (1974) "Cross-validatory Choice and Assessment of Statistical Predictions," (with Discussion), *J. Royal Stat. Soc. B.*, 36, 111-147.
- Stroud, A. H. (1971) *Approximate Calculation of Multiple Integrals*, Englewood Cliffs, New Jersey: Prentice-Hall.
- User Services Dept., Va. Tech Computing Center (1990) *Accessing Formac (Formula Manipulation Compiler) Under MVS*, Doc. No. MT03, Va. Tech Computing Center, Blacksburg, Va.
- Wahba, G., G.H. Golub, and C.G. Heath (1979) "Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics*, 21, 215-223.
- Watson, G. S. (1964) "Smooth Regression Analysis," *Sankhya Ser. A*, 26, 359-372.
- Welch, W. J. (1982) "Branch-and-Bound Search for Experimental Designs Based on D-optimality and Other Criteria," *Technometrics*, 26, 71-130.
- Wong, W. H. (1983) "On the Consistency of Cross-validation in Kernel Nonparametric Regression," *Annals of Statistics*, 11, 1136-1141.
- Wynn, H. P. (1970) "The Sequential Generation of D-Optimum Experimental Designs," *The Annals of Mathematical Statistics*, 41, 1655-1664.

## APPENDIX 1

### ANALYTICAL CONSIDERATIONS — FURTHER DETAILS

Recall that in Section IV.9 it is desired to augment with the point  $x_0 \in (0, 2)$  to minimize

$$\int_0^2 \frac{\text{var} [\hat{y}(\underline{x})]}{\sigma^2} dx. \quad (\text{A1.1})$$

The details of simplifying and evaluating this integral are considered in this appendix.

The design matrix for the univariate example presented in Section IV.9 is

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & x_0 \end{bmatrix} \quad \text{with } \underline{x}_j' = (1, x_j).$$

Using (IV.9.2) for the kernel function, the numerator in (A1.1) becomes

$$\sum_{j=0}^4 \lambda \left[ h_{x_j}^{ker} + (1 - \lambda) h_{x_j}^{ls} \right]^2,$$

where

$$h_{x_j}^{ker} = \frac{K\left(\frac{x - x_j}{h}\right)}{\sum_{j=0}^4 K\left(\frac{x - x_j}{h}\right)} = \frac{1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2}{\sum_{j=0}^4 \left[1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2\right]} \quad (A1.2)$$

and

$$h_{x_j}^{ls} = \underline{x}' (X' X)^{-1} \underline{x}_j . \quad (A1.3)$$

Subsequently, (A1.1) becomes

$$\int_0^2 \sum_{j=0}^4 \left\{ \lambda \left( \frac{1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2}{\sum_{j=0}^4 \left[1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2\right]} \right) + (1 - \lambda) \underline{x}' (X' X)^{-1} \underline{x}_j \right\}^2 dx. \quad (A1.4)$$

This needs to be simplified! Using a different notation to emphasize the integral as a function of  $x$ , let

$$KER(x) = \frac{1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2}{\sum_{j=0}^4 \left[1 - \frac{1}{9}\left(\frac{x - x_j}{h}\right)^2\right]} \quad (A1.5)$$

and

$$LS(x) = \underline{x}' (X' X)^{-1} \underline{x}_j \quad (A1.6)$$

After performing the sum in the denominator and simplifying, (A1.5) reduces to

$$\text{KER}(x) = \frac{9h^2 - x^2 + 2xx_j - x^2}{45h^2 - 6 - 5x^2 + 4x + 2xx_0 - x_0^2}.$$

Likewise, carrying out the matrix multiplications in (A1.6) yields

$$\text{LS}(x) = \frac{6 + x_0^2 - 2x - xx_0 - 2x_j - x_0x_j + 5xx_j}{4x_0^2 - 4x_0 + 26}.$$

Substituting in  $\text{KER}(x)$  and  $\text{LS}(x)$ , then squaring and simplifying (A1.4) yields the integrals

$$\int_0^2 \lambda^2 \sum_{j=0}^4 \text{KER}(x)^2 dx + \int_0^2 2\lambda(1-\lambda) \sum_{j=0}^4 \{\text{KER}(x)\}\{\text{LS}(x)\}dx + \int_0^2 (1-\lambda)^2 \sum_{j=0}^4 \text{LS}(x)^2 dx. \quad (\text{A1.7})$$

These integrals can now be evaluated, using tables of integrals and FORMAC, a formula manipulating computer package. The integrands consist of a polynomial in  $x$  divided by a quadratic in  $x$  raised to a power. The individual pieces are of the form

$$\int \frac{x^n}{X^m} dx, \quad (\text{A1.8})$$

where  $n$  is some positive integer,  $X = (ax^2 + bx + c)$ , and  $m$  is some positive integer. If  $m \neq 2n - 1$ , (A1.7) can be evaluated. When  $m = 2n - 1$ , the integrand was zero.

Responses corresponding to each of the original four points were simulated, and the bandwidth and lambda values were computed using the PRESS\* criterion for the bandwidth and the  $C_p3$  criterion for the lambda. The results were  $h = .86066$  and  $\lambda = .43158$ . These values were used in evaluating (A1.7).

Due to the messy nature of the algebra involved, FORMAC was used to perform the



sums and the algebra for the individual integrals in (A1.7). Using tables of integration as an aid, FORMAC was also used to evaluate the integrals at the limits of integration and simplify somewhat the answers. After all this algebra is performed, the evaluated integrals are given below.

The first integral in (A1.7) reduces to

$$\int_0^2 \lambda^2 \sum_{j=0}^4 \text{KER}(x)^2 dx =$$

$$-A \left\{ \left( \frac{C-20}{-B(G-2C)} \right) + \frac{10 \ln \left( - \left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} + A \left\{ \frac{C}{-BE} + \frac{10 \ln \left( - \left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} +$$

$$D \left\{ \frac{2(E-C)}{-B(G-2C)} + \frac{C \ln \left( - \left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} + -D \left\{ \frac{2}{-B} + \frac{C \ln \left( - \left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} +$$

$$(F - .5588 E) \left\{ \frac{C}{-5B} + \frac{2E \ln \left( - \left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} +$$

$$-F \left\{ \frac{\frac{1}{5} EC + 2(10E - C^2)}{-B(G-2C)} + \frac{2E \ln \left( - \left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} +$$

$$.9313 \left\{ \frac{8}{5(G-2C)} + \frac{3}{5} E \left\{ \frac{\frac{1}{5} (EC + 2(10E - C^2))}{B(G-2C)} + \frac{2E \ln \left( - \left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{-B\sqrt{B}} \right\} \right\}.$$

The second integral in (A1.7) is

$$\begin{aligned}
 & \int_0^2 2\lambda(1-\lambda) \sum_{j=0}^4 \{KER(x)\}\{LS(x)\} dx = \\
 & \frac{H}{10} \left\{ \frac{C \ln \left( -\left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{\sqrt{B}} - \ln(-E) \right\} + -I \frac{\ln \left( -\left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{\sqrt{B}} + \\
 & \frac{H}{50J} \left\{ C \ln(-E) + (10E - C^2) \times \left\{ \frac{\ln \left( -\left[ \frac{C-\sqrt{B}}{C+\sqrt{B}} \right] \right)}{\sqrt{B}} \right\} \right\} + \\
 & -\frac{1}{J} \left\{ \frac{H}{10} \left( \frac{C \ln \left( -\left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{\sqrt{B}} - \ln(-G+2C) \right) - I \frac{\ln \left( -\left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{\sqrt{B}} + \right. \\
 & \left. \frac{H}{50} \left( C \ln(-G+2C) + (10E - C^2) \times \left\{ \frac{\ln \left( -\left[ \frac{C-20-\sqrt{B}}{C-20+\sqrt{B}} \right] \right)}{\sqrt{B}} \right\} + \frac{2}{5} \right) \right\}.
 \end{aligned}$$

Finally, the third integral in (A1.7) becomes

$$\int_0^2 (1-\lambda)^2 \sum_{j=0}^4 LS(x)^2 dx = \frac{2K - 2L + \frac{8}{3}M}{J^2}.$$

For these integrals, the letters are defined as:

$$A = .18626 x_0^4 - 2.4835 x_0^2 + 29.8427$$

$$B = -(16 x_0^2 - 16 x_0 - 562.6607)$$

$$C = 2 x_0 + 4$$

$$D = .74505 x_0^3 - 4.9669 x_0 - 3.9735$$

$$E = x_0^2 - 27.3331$$

$$F = 1.11756 x_0^2 - 5.7119$$

$$G = x_0^2 - 7.3331$$

$$H = 1.9626 x_0^3 - .98127 x_0^2 - 2.9438 x_0 + 13.7378$$

$$I = .98127 x_0^3 + 7.1959 x_0^2 - 9.15847 x_0 + 75.23049$$

$$J = 4 x_0^2 - 4 x_0 + 26$$

$$K = 1.2924 x_0^4 - 1.2924 x_0^3 + 16.1551 x_0^2 - 7.75443 x_0 + 50.4038$$

$$L = 2.5848 x_0^3 + 2.5848 x_0^2 + 11.63165 x_0 + 33.60254$$

$$M = 6.46203 x_0^2 - 6.46203 x_0 + 42.0032.$$

As detailed in Section IV.9, the results indicate that the proposed method is reasonable, at least in a simple case. The results using the BIIV criterion for augmentation instead of the integrated prediction variance criterion should be similar, since the question is how well do the approximation formulas work when performing sequential design augmentation using HATLINK.

## **APPENDIX 2**

### **NUMERICAL CONSIDERATIONS**

#### **A.2.1 NUMERICAL CONSIDERATIONS**

As stated in Section IV.8, several problems arise in practice when carrying either (IV.1.1) or (IV.1.2) out. One problem is evaluating the integral in (IV.1.1) and (IV.1.2). Many formulas for approximating integrals exist, see for example Stroud (1971). All methods for approximating integrals require the integrand to be relatively smooth. Sometimes for certain parameters, the HATLINK prediction variance graphs are fairly rough and thus the approximation methods would not be recommended. Fortunately, for the bandwidth and mixing parameter values chosen by Einsporn's methods, plots of the HATLINK prediction variance reveal that the integrand in (IV.1.1) and (IV.1.2) is quite well-behaved.

#### **A.2.2 EXAMPLES OF THE PREDICTION VARIANCE PLOT WITH DIFFERENT $h$ & $\lambda$**

Suppose the new region of interest, the initial point in the new region, and the design are the same as in the example in Section IV.3. Before using the various techniques to approximate the integral in (IV.1.1) and (IV.1.2), the smoothness of the prediction variance plot must be checked.

As an example of the smoothness of the prediction variance plot and using a different data set for the design, the bandwidth is calculated to be 2.15 and  $\lambda$  is calculated to be 0.52. The graph of the prediction variance plot for these values of  $h$  and  $\lambda$  is presented in Figure A2.2.1. As can be seen, the plot is fairly smooth and well-behaved with little to indicate trouble in using approximation techniques. Contrast Figure A2.2.1 to the plot of the prediction variance with  $h$  and  $\lambda$  arbitrarily chosen as  $h = 12.5$  and  $\lambda = 0.90$  in Figure A2.2.2. The bandwidth is so large that kernel regression is averaging over almost all the data points for each predicted value. It has removed a large part of the variability of the original curve. A third example, with  $h = 0.20$  and  $\lambda = 0.90$  and plotted in Figure A2.2.3, illustrates some of the possible troubles if the bandwidth is too small. Note the sharp changes apparent in the plot, integral approximation techniques will not work very well in this case. For this case, kernel regression is finding prediction values based only upon the closest data points. This tends to yield a fairly rough and poorly-behaved surface for approximation techniques.

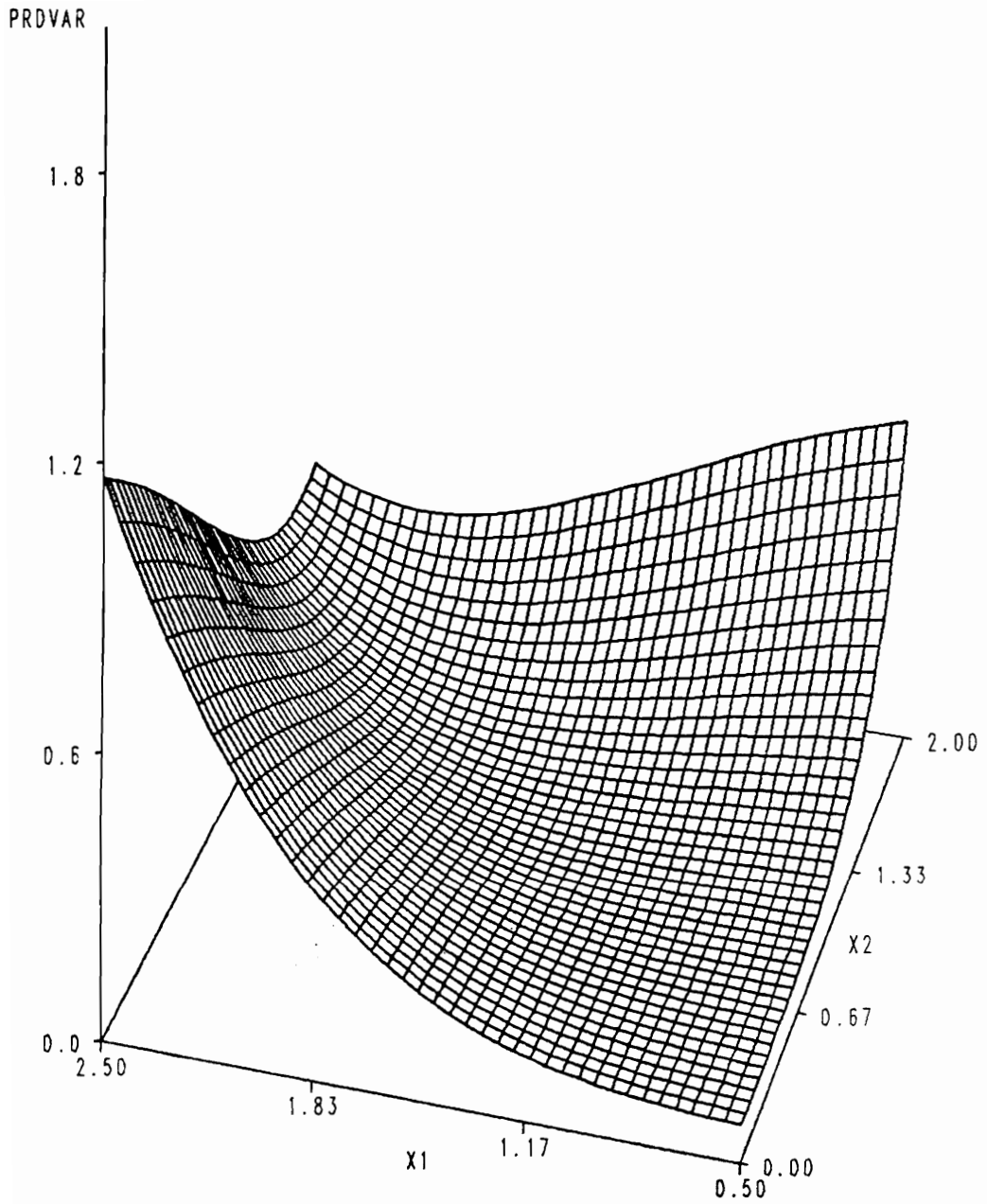


FIG. A2.2.1. PREDICTION VARIANCE PLOT FOR THE HATLINK FIT  
 BANDWIDTH IS 2.15 & LAMBDA IS 0.52 FROM DATA.

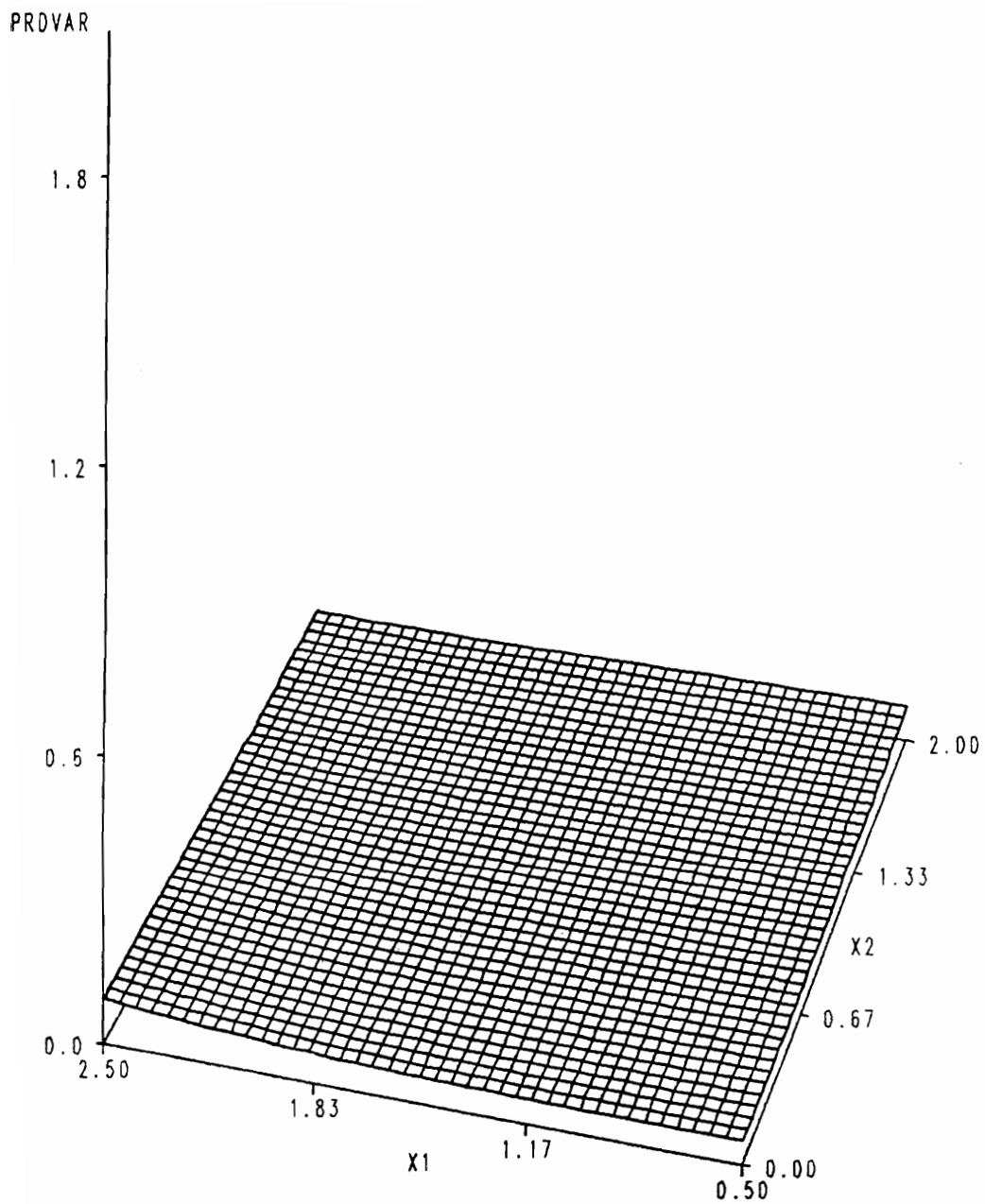


FIG. A2.2.2. PREDICTION VARIANCE PLOT FOR THE HATLINK FIT  
 BANDWIDTH IS SET TO BE 12.5 & LAMBDA IS 0.9.

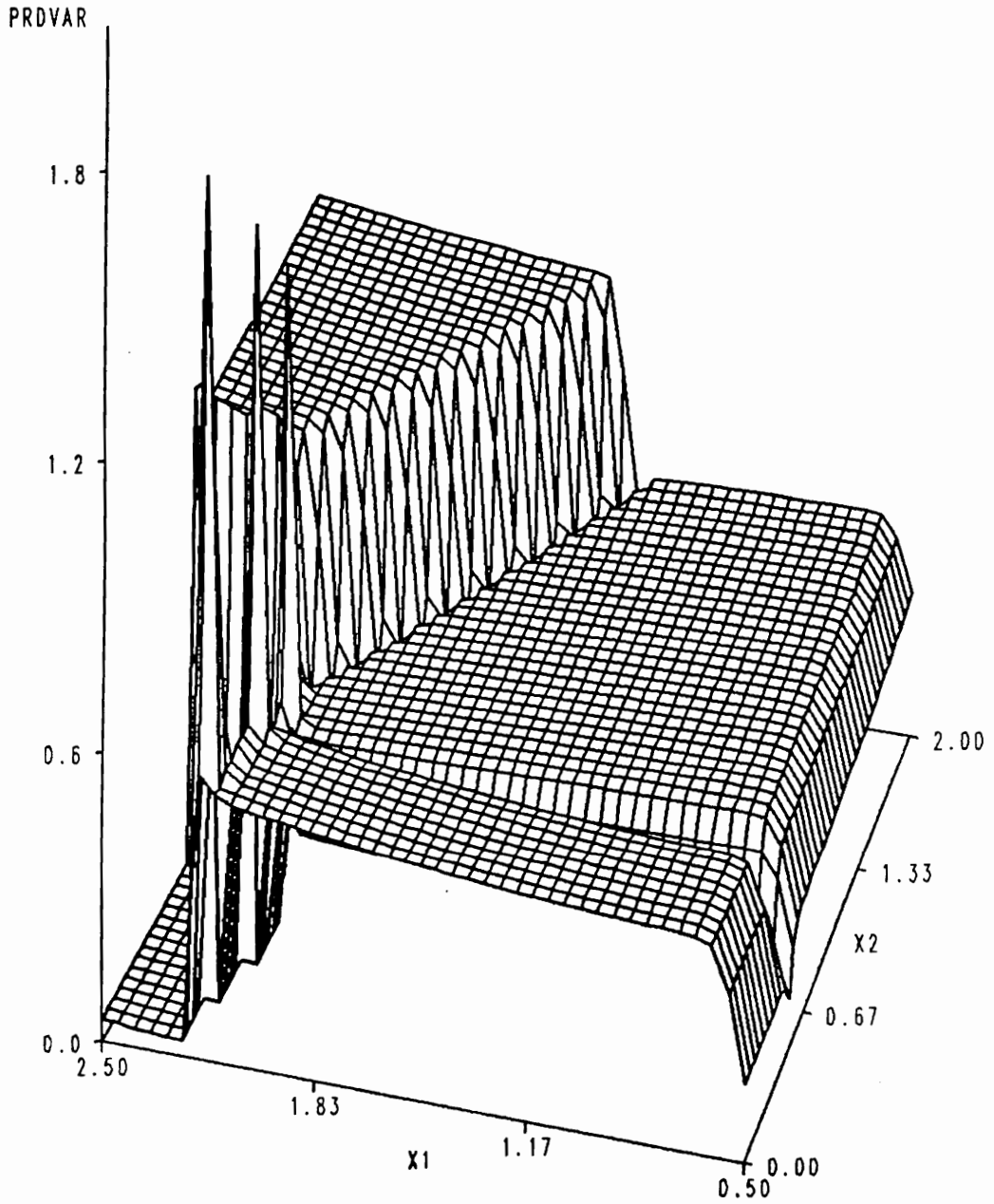


FIG. A2.2.3. PREDICTION VARIANCE PLOT FOR THE HATLINK FIT  
 BANDWIDTH IS SET TO BE 0.20 & LAMBDA IS 0.9.



### A.2.3 APPROXIMATING INTEGRALS

As mentioned in Section IV.8, the most straightforward method for evaluating the integral contained in (IV.1.1) and (IV.1.2) is a Simpson's Rule approach for higher order dimensions, which may require the function to be evaluated at many points to attain a relatively high degree of accuracy. If the function is hard or difficult to evaluate, it may become very time consuming! In preliminary work for two dimensions, 16 grid points were not sufficient but 64 or more seemed to suffice. The large grid resulted in a 4-fold increase in computer time over that required for the smaller grid.

A more efficient method of numerically evaluating integrals used here to evaluate integrals incorporates Gaussian quadrature. This involves approximating the multiple integral by a weighted sum of points evaluated with the integrand, that is

$$\int \dots \int_{R_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \simeq \sum_{i=1}^n B_i f(\nu_{i,1}, \dots, \nu_{i,n}) \quad (\text{A2.3.1})$$

where  $R_n$  is a given region in  $n$ -dimension Euclidian space  $E_n$ . The  $\underline{\nu}_i = (\nu_{i,1}, \dots, \nu_{i,n})$  are the points or nodes of the formula and lie in  $E_n$ . The  $B_i$  are constants not depending on  $f(x_1, \dots, x_n)$  called the coefficients of the formula. They act as weights on the points  $\underline{\nu}_i$ . The formulas are given for standard regions, say a circle centered at zero with radius of one or a cube centered at zero with sides of length one. Any spherical or cuboidal region can be translated and/or transformed into a standard region, and thus the formulas given in Stroud (1971) may be applied to these regions with the appropriate changes. Some of these formulas may be used to approximate (IV.1.1) and (IV.1.2) with fewer points but the same accuracy as Simpson's Rule.

## APPENDIX 3

### ADDITIONAL CONSIDERATIONS

#### A3.1 AUGMENTATION USING INTEGRATED PREDICTION VARIANCE

In preliminary studies, one main augmentation criterion considered was the integrated prediction variance criterion of (IV.1.1). Both the PRESS\* and  $C_p$  methods of selecting  $\lambda$  were considered at this stage of investigation. As stated before, the PRESS\* and  $C_{p3}$  methods of selecting  $\lambda$  appeared to perform better than any other criterion considered. The integrated prediction variance criterion appeared to suffer sometimes under certain circumstances, leading to the BIIV method of (IV.1.2).

In all simulations performed at this stage, a basic procedure was consistently used. Always using the same seed to generate the random errors and for the specified values of both  $\sigma$  and CT, twenty five points were augmented sequentially, observing responses after augmenting with each point using both HATLINK and least squares regression. This corresponds to using a batch size of one. Recall that in the case of HATLINK such a procedure allows the user, after augmenting with each point, to update both the bandwidth and the mixing parameters. After augmenting with a point, the goodness of fit was measured by the ESE criterion. The next sections summarize the results of using (IV.1.1) and the various methods of selecting  $\lambda$ .

### A3.1.A IPV & $C_{P3}$ METHODS FOR SELECTING $\lambda$

Preliminary work suggested that using the  $C_{P3}$  method for selecting  $\lambda$  would yield good estimates of  $\lambda$  over the range of values of interest in this study. This section deals with the performances of both HATLINK and least squares regression in the situation described in Section IV.3 when the mixing parameter is estimated by the  $C_{P3}$  method and integrated prediction variance is used as the method of augmentation. The augmentation is performed as detailed in Section A3.1.

Comparison of ESE values from the HATLINK and least squares methods reveal under what conditions HATLINK is producing a better fitted surface than least squares regression. For a given situation, the “winning” method is the one that yields the smaller ESE value. When  $\sigma$  is small, the HATLINK method is not doing as well initially as the least squares method. In the examples studied, regardless of the degree of model misspecification, as  $\sigma$  increases HATLINK outperforms least squares or, in the nearly correctly specified model, HATLINK at least does not appear to overly harm the user. Regardless of the initial conditions, eventually HATLINK yielded a better fit as measured by the ESE criterion than least squares regression.

By comparing the ESE values for specific values of CT, several features become apparent. Consider first the case when  $CT = 0.1$ . As illustrated in Figure A3.1.1, this case had minimal model misspecification. Thus, the  $C_{P3}$  criterion would be expected to produce a value of  $\lambda$  relatively close to zero. But this did not always immediately happen when augmenting points, especially for the smaller values of  $\sigma$ .

For  $CT = 0.1$  and regardless of the  $\sigma$  values, eventually HATLINK augmented enough points to detect to some extent the small amount of misspecification that exists in the true model. The behavior of  $\lambda$  when augmenting with the first few points was sometimes unexpected,

particularly when  $\sigma < 0.5$ . Contrary to what might have been anticipated, the  $\lambda$  values are fairly high. Intuitively this is not unexpected with  $CT = 0.1$ . The fact that these values are indeed too large is illustrated by the greater ESE values in comparison to those from least squares. As long as the  $\lambda$  value over-emphasized kernel regression, HATLINK was outperformed by the least squares method as measured by the ESE fit. This suggests that there are difficulties associated with the  $C_p3$  method of selecting  $\lambda$ , and further that these problems appear to somehow be variance based. When  $\sigma \geq 0.5$ , these problems are no longer apparent. In fact, for these higher values of  $\sigma$ , HATLINK appears to perform about as well as least squares regression. In terms of the quality of fit, the user is not being hurt too much when augmenting design points by the HATLINK method.

When  $CT = 1$ , the  $\lambda$  values selected by the  $C_p3$  criterion are likely to be close to 1, although here it is harder to have an intuitive feel for the appropriate range of  $\lambda$ . In this case, a comparison of  $\lambda$  values revealed numbers roughly between 0.6 and 0.9. These results did not look too far removed from the intuitive answer. But comparing the ESE values for the HATLINK and least squares methods reveals that for  $\sigma < 0.3$ , the least squares fit is much better than the fit for HATLINK, at least for the first 17 or so points augmented. If  $\sigma \geq 0.3$ , then HATLINK does a better job of fitting the surface than least squares does as measured by ESE. In these cases, HATLINK outperforms least squares within the first or second point augmented. This suggests again that there is possibly a variance problem with the selection of  $\lambda$  by the  $C_p3$  criterion. It appears that the mixing parameter values should possibly be closer to 1. Regardless, under some circumstances the user appears to be achieving a better fit, as measured by the ESE criterion, when augmenting design points by the HATLINK method as opposed to augmentation by least squares.

As indicated above, the method proposed does not appear to work very well when the

variance is too small. The trouble seems to be in the  $C_P3$  criterion used for the selection of  $\lambda$ . If the variance is fairly small, the  $C_P3$  method appears to not select small enough  $\lambda$  values when  $CT = 0.1$ , and not select large enough mixing parameters when  $CT = 1$ . This suggests that perhaps either a different method of choosing  $\lambda$  or a different method of augmenting with points would perform better over the range of model misspecification studied.

### A3.1.B IPV & OTHER $C_P$ CRITERIA FOR SELECTING $\lambda$

As described in Section III.7, Einsporn (1987) proposed four different ways of selecting  $\lambda$  based upon a  $C_P$  type of criterion. All four methods of choosing the mixing parameter were studied, as described in Section V.3, in the hopes that at least one of them would yield better estimates of  $\lambda$  over the broad range of model misspecification represented by  $CT = 0.1$  and  $CT = 1$ . For the different values of  $\sigma$  and  $CT$  as described above, examples were performed using the integrated prediction variance and the four different ways of selecting  $\lambda$ . As will be seen, some versions performed quite differently than others. The results for the four  $C_P$  criteria are summarized in this section.

For the four different techniques of estimating  $C_P$  described in section III.7, the key quantity involved was  $\sigma^2$ . Consequently, these versions of  $C_P$  differed mainly in the methods of estimating  $\sigma^2$  and error degrees of freedom. Note that  $C_{P1}$  and  $C_{P2}$  used  $s_{ker}^2$  to approximate  $\sigma^2$ , while  $C_{P3}$  used  $s_{ols}^2$  for  $\sigma^2$ . Instead of estimates of  $\sigma^2$ ,  $C_{P4}$  used sums of mean squared errors for HATLINK and least squares. In using these different versions of  $C_P$ , an interesting pattern quickly developed.

$C_{P1}$  and  $C_{P2}$  both used the kernel estimate of variance and both tended to select large  $\lambda$  values, regardless of the degree of model misspecification. Not unexpectedly, when  $CT = 1$ ,

both methods tended to select  $\lambda$  very close to one. What is surprising was the range of mixing parameters selected by these two methods when  $CT = 0.1$ . In this case, it would be expected to choose  $\lambda$  values relatively close to zero. Instead, numbers around 0.6 to 0.7 were selected. Comparison of the ESE values for HATLINK and least squares indicate that these are poor choices for  $\lambda$ , implying that these two methods are inflating the value of  $\lambda$  when there is minimal model misspecification. Clearly, these two versions of approximating  $C_P$  are not suitable over a broad range of  $CT$  values. They appear to be best suited for a situation where the user suspects there is some trouble with the specified model and desires to weight kernel regression over least squares regression.

$C_{P4}$  had the opposite problem, that is, it tended to underinflate the value of  $\lambda$  if there was more model misspecification. When  $CT = 0.1$ , this method yielded  $\lambda$  values fairly close to zero, as expected. But with  $CT = 1$ , it tended to estimate  $\lambda$  close to .5 and, as measured by ESE values, did not produce a very good fit. Obviously  $C_{P4}$  should not be used if the user is concerned very much about problems in the prescribed model.

For the three methods  $C_{P1}$ ,  $C_{P2}$ , and  $C_{P4}$ , there appears to be a relationship between the method of estimating  $\sigma^2$  and the value of  $CT$  for which a particular method worked well. Both  $C_{P1}$  and  $C_{P2}$  used kernel regression to estimate  $\sigma^2$ , and both methods also estimated  $\lambda$  better when  $CT = 1$ . Such a value of  $CT$  would imply that there was a lot of model misspecification and that nonparametric methods might do better than parametric methods. Conversely,  $C_{P4}$  used least squares to obtain SSE (instead of  $\sigma^2$  as in the other methods) and estimated  $\lambda$  better when  $CT = 0.1$ . This value of  $CT$  would imply that least squares would not be correct. Thus it appears that these  $C_P$  methods are tending to estimate  $\lambda$  in the direction of the type of estimate used, that is towards either least squares regression or kernel regression.

As stated in section A3.1.A, even the  $C_{P3}$  method had some problems. In this case, for

smaller variances it tended to underinflate  $\lambda$  when  $CT = 1$ . Furthermore, it tended to overinflate  $\lambda$  when  $CT = 0.1$ . Despite these problems, compared to the other three  $C_P$  versions  $C_{P3}$  appears to be the best method for estimating  $\lambda$  over a broad range of model misspecification.

The main trouble with all four versions of  $C_P$  appears to be estimating the quantity  $(s^2 - \hat{\sigma}^2)$  in such a way that the resulting  $\lambda$  value is not consistently forced towards either extreme. All new versions of  $C_P$  considered also suffered from the same problem. At this point, the PRESS\* criterion of (III.6.1) was considered to avoid the problem of estimating  $(s^2 - \hat{\sigma}^2)$ .

### A3.1.C IPV & PRESS\* CRITERIA FOR SELECTING $\lambda$

Of the two PRESS type methods proposed by Einsporn (1987), it was noted that the method of (III.5.1) tended to yield a  $\lambda$  that was too large. This was confirmed by the simulations performed using (III.5.1) to estimate the mixing parameter. PRESS tended to select a  $\lambda$  that was always at least about 0.7, regardless of the value of either  $\sigma$  or  $CT$ .

The PRESS\* method of (III.6.1) in general performed much better than both the PRESS method of (III.5.1) and the  $C_{P3}$  method of (III.7.9) when augmenting by minimizing integrated prediction variance. It yielded  $\lambda$  values very close to zero for  $CT = 0.1$ , occasionally producing  $\lambda$  values up to 0.5, especially for  $\sigma = 0.1$ . As long as the  $\lambda$  value was close to zero, both HATLINK and least squares tended to behave in a very similar fashion. According to the ESE fits, using HATLINK in these situations does not hurt the user very much. For the cases when  $CT = 1$ , the  $\lambda$  values were much closer to one than the values produced by  $C_{P3}$ .

Based upon the fits as measured by the ESE criterion, there still seems to be some trouble with the smaller variances. Again, HATLINK is not immediately outperforming least squares. For the case when  $CT = 1$ , it often required fifteen or more points before HATLINK

consistently outperformed least squares. It is interesting to note that for the larger values of  $\sigma$  (i. e.  $\sigma \geq 0.5$ ) HATLINK fits based upon  $\lambda$  selected by PRESS\* not only are better than least squares, they are also better than HATLINK fits based upon  $C_P3$ .

Based upon the results in sections A3.1.A, A3.1.B, and A3.1.C, when the degree of model misspecification is unknown, the best way of selecting  $\lambda$  is by the PRESS\* criterion. This method appears to yield the best performance over a broad range of CT values. If the user has reason to suspect that there is a fair amount of misspecification but the specified parametric model must be used, then  $C_{P1}$  or  $C_{P2}$  might work better than PRESS\*. Unless specified otherwise, for the remainder of this study all  $\lambda$  values will be found using the PRESS\* method.

#### **A3.1.D GENERALIZED CROSS VALIDATION FOR SELECTING $\lambda$**

One technique for selecting  $\lambda$  not considered by Einsporn (1987) or others is a method based upon generalized cross validation. Recall that ridge regression is a biased estimation technique used in linear regression to combat multicollinearity. In the presence of multicollinearity, least squares regression yields unbiased estimates with no upper bound upon the corresponding variances, while biased estimation techniques can sometimes yield a significant reduction in the variances with a small increase in the bias. This is achieved through a parameter  $k$ , commonly called the shrinkage parameter. If  $k$  is chosen successfully, the reduction in variance is of greater magnitude than the corresponding increase in the bias.

Several methods of selecting  $k$  are considered in Myers (1990). One commonly used method is the PRESS criterion, similar to that described by III.2.1. A second approach, also based upon a prediction approach, is the generalized cross validation or GCV given by



$$\text{GCV} = \frac{\sum_{i=1}^n e_{i,k}^2}{\left\{n - [1 + \text{tr}(\mathbf{H}_k)]\right\}^2} \quad (\text{A3.1.1})$$

where  $e_{i,k}^2$  is the  $i$ th residual for ridge regression with shrinkage parameter  $k$ . Also,  $\text{tr}(\mathbf{H}_k)$  is the trace of the ridge hat matrix with the shrinkage parameter  $k$ . The constant term represented by the intercept is not included in the hat matrix, hence the addition of the value of one is to compensate for the lack of the intercept. The value of  $k$  is selected that yields the largest reduction in variance with the smallest increase in bias. See Wahba et al. (1979) for further details concerning the nature of GCV. This paper indicated that GCV can be considered a generalized type of PRESS criterion.

One could think of the mixing parameter  $\lambda$  as a shrinkage parameter in some respects. A PRESS type of criterion is already being considered for selecting  $\lambda$ . Some bias could be occurring if the model is misspecified, and so perhaps some criterion for selecting  $\lambda$  that can incorporate bias to some extent would perform better than either the  $C_P$  or PRESS type of criteria already considered. As an alternative, it was proposed to try a variation of the GCV method in which  $\lambda$  is chosen to minimize

$$\text{GCV}^* = \frac{\sum_{i=1}^n e_{i,\lambda}^2}{\left\{n - [\text{tr}(\mathbf{H}_\lambda)]\right\}^2} \quad (\text{A3.1.2})$$

where  $e_{i,\lambda}^2$  is the  $i$ th residual of the HATLINK regression fit. In this case, there is no value of one added to the trace of the HATLINK hat matrix, since the HATLINK hat matrix includes the intercept term.

This method is similar in many respects to the PRESS method of III.2.1. As such, it was hoped that this method would perform better in selecting  $\lambda$  than the four versions of  $C_P$  that required estimating  $s^2$  and  $\sigma^2$ . Unfortunately, preliminary results using integrated prediction variance as the augmentation criterion indicated that the GCV\* method worked no better than the PRESS\* criterion, although it did appear to work better than some of the versions of  $C_P$  proposed in Chapter III.

In most cases examined, the GCV\* criterion never worked significantly better than the PRESS\* criterion, although sometimes the two techniques for evaluating  $\lambda$  worked about equally well. For the cases of smaller variance, (A3.1.2) also had trouble selecting appropriate values of  $\lambda$ . When the model misspecification was small, it selected inflated  $\lambda$  values. When augmenting with points, it persisted in this behavior much longer than the PRESS\* criterion did for the corresponding  $\sigma$  and CT values. When CT = 1, the  $\lambda$  value chosen was fairly close to one. But just like the PRESS\* criterion, when the variance was small, it required fifteen or more points augmented before HATLINK outperformed least squares as based upon the ESE fits.

In an attempt to help the GCV\* method as proposed in (A3.1.2) to perform better, several versions of (A3.1.2) were also considered, including multiplication by constants, using  $\text{tr}(\mathbf{H}_\lambda' \mathbf{H}_\lambda)$  (since  $\mathbf{H}_\lambda$  is not idempotent), and variations on the residuals in the numerator. Nothing considered appeared to overcome the observed problems. On the basis of these results, the GCV\* criterion and its variations were not considered any further.

### **A3.1.E SUMMARY OF AUGMENTATION USING IPV**

For every combination of  $\sigma$  and CT mentioned above, only one simulation was performed at this stage of the study. Furthermore, these simulations are only for one true model

and one seed value when generating the random errors. Admittedly, broad sweeping conclusions about the validity and usefulness of augmenting design points using the HATLINK method can not be made for every setting, although several things have become apparent.

Most importantly, under certain circumstances the proposed method of augmenting design points does appear to outperform the least squares method. If the user's prescribed model is indeed somewhat misspecified, HATLINK can produce a better fit than one from least squares regression. Conversely, if the prescribed model is close to the true model, the user will not be overly harmed in terms of the quality of fit. It only requires a little more computer time. As suspected, HATLINK is capable of detecting if the user's model has been incorrectly selected. Empirical evidence suggests that if the degree of misspecification is too large, then neither HATLINK nor least squares does a very good job of estimating the response surface. This appears to correspond to cases where it could be detected by conventional methods such as lack of fit tests. Thus, HATLINK appears to be useful in the case where the model is somewhat misspecified, but conventional methods would be unable to detect this situation.

Further investigations into the behavior of the augmentation method using integrated prediction variance and the selection of  $\lambda$  by PRESS\* revealed a bias problem, especially when CT was moderate to large in value. A breakdown of the ESE values into component parts demonstrated that in most cases bias contributed over 60% to the ESE values studied. This indicates that it is a larger problem than had been considered. Obviously, augmentation by minimizing the integrated prediction variance does not consider bias in any fashion, thus indicating a need to attempt to incorporate this in some manner into the augmentation procedure.

The obvious way to achieve this goal is to augment by a point to minimize the integrated prediction mean square error (IMSE). That is, select the point  $\underline{x}_0$  such that

$$\min_{\underline{x}_0 \in R_K} \left\{ \int [\text{bias}(\hat{\underline{x}})]^2 d\underline{x} + \int \text{var}(\hat{\underline{x}}) d\underline{x} \right\} \quad (\text{A3.1.3})$$

is achieved, where these integrals are calculated assuming the point  $\underline{x}_0$  has been augmented. Two difficulties arise with such a method of augmentation. First, the bias at each point  $\underline{x}$  in the region  $R_K$  must be estimated, which can be performed in a similar fashion as that used in calculating BIIV. The other problem is that due to the nature of estimating bias, the point  $\underline{x}_0$  must have a corresponding  $y_0$ . This can be based upon the current data, or approximated in some other fashion. In simulation studies, this method appeared to work better than augmentation by integrated prediction variance, but not quite as well as BIIV in most circumstances. Due to the estimation of  $y_0$  and the second integral, the IMSE method proposed in (A3.1.3) requires extra effort, but does not appear to work sufficiently better than BIIV to consider it further.

### A3.2 VARIATIONS TO THE BIIV METHOD OF AUGMENTATION

Recall that in RSM, it is desired to ultimately locate the maximum estimated response value and its corresponding location. Using the PRESS\* criterion for selecting  $\lambda$  and BIIV to augment points, the maximum estimated response value and its location was calculated for both the HATLINK and least squares methods. Both techniques were then compared to the true value and its corresponding location. It was hoped that HATLINK would perform better than least squares regression in either locating and/or estimating the maximum value. Unfortunately, for the cases considered HATLINK seems to only slightly outperform least squares in the location and value of the largest response. Plotting twelve augmented values indicated that HATLINK never placed any points close to the location of the true maximum. Due to the

nature of kernel regression, this lack of data close to the maximum makes it harder to estimate its value.

Several variations to the bias component of BIIV were considered in the hopes that this scarcity of points near the maximum could be overcome. Variations of BIIV were examined to attempt to give more weight to values closer to the maximum when augmenting with points. The overall fit as measured by the ESE values may not be as good, but it was hoped that this would be compensated by a better estimate of either the maximum value or its location. For each variation considered, one example was run with  $CT = 0.7$  and  $\sigma = 1$ . Here, the maximum value of response of 42.71 occurs at (1.25, 1.6). Any method showing promise at this stage would be investigated further.

Recall that the bias part of BIIV is estimated by  $-\left[\hat{Y}_{ker}(\underline{x}_0) - \hat{Y}_{hat}(\underline{x}_0)\right]^2$ . Each method proposed below replaces this part of BIIV without the negative, that is each replaces the estimate of squared bias in BIIV. The sixteen methods considered could be grouped into four distinct types. To ease notation here,  $\hat{Y}_{ker}(\underline{x}_0) = \hat{Y}_{ker}$  and likewise for the HATLINK fitted value at the point  $\underline{x}_0$ .

The first type consists of the predicted value from kernel regression raised to some power divided by the difference between the kernel and the HATLINK predicted values raised to some possibly different power. Where required, absolute values are taken on the difference to ensure that the resulting number is positive. These methods were considered because it was thought that possibly more weight should be given to places where the kernel predicted value was larger. By calculating BIIV in such a manner, given two locations with approximately equal estimated bias, i.e. locations where  $\hat{Y}_{ker} - \hat{Y}_{hat}$  are approximately the same, the location closer to the true maximum would hopefully be selected. Some methods were designed to keep the units the same as those for the bias squared part of the original BIIV. Three variations of this

type were:

$$\text{Meth1} = \frac{(\hat{Y}_{ker})^2}{(\hat{Y}_{ker} - \hat{Y}_{hat})^2} \quad (\text{A3.2.1})$$

$$\text{Meth2} = \frac{(\hat{Y}_{ker})}{(\hat{Y}_{ker} - \hat{Y}_{hat})^2} \quad (\text{A3.2.2})$$

$$\text{Meth3} = \left[ \frac{(\hat{Y}_{ker})^3}{|\hat{Y}_{ker} - \hat{Y}_{hat}|} \right]. \quad (\text{A3.2.3})$$

The second group of variations consisted of the negative of the difference between the kernel and the HATLINK predicted values, with absolute values where required, raised to some power divided by the kernel predicted value raised to some possibly different power. These methods are similar, in a sense, to those in the first group. To understand the reasoning behind Meth4 through Meth10, first ignore the negative sign associated with these seven methods and consider replacing the estimate of bias in BIIV with any of these methods. With the negative sign from the BIIV method, the contribution of the selected method is being maximized. Assuming two competing locations have approximately equal estimated bias, such methods would be locating the point further away from the location where the kernel predicted value is the largest. Thus the extra negative sign is required to minimize the selected method and hopefully select the point closer to the true maximum. Initially, this group looked more promising so there were seven different variations considered. Again, some were designed to keep the units the same as those in BIIV. The seven different variations were:

$$\text{Meth4} = - \left[ \frac{(\hat{Y}_{ker} - \hat{Y}_{hat})^2}{(\hat{Y}_{ker})} \right] \quad (\text{A3.2.4})$$

$$\text{Meth5} = - \left[ \frac{(\hat{Y}_{ker} - \hat{Y}_{hat})^2}{(\hat{Y}_{ker})^2} \right] \quad (\text{A3.2.5})$$

$$\text{Meth6} = - \left[ \frac{(\hat{Y}_{ker} - \hat{Y}_{hat})^4}{(\hat{Y}_{ker})^2} \right] \quad (\text{A3.2.6})$$

$$\text{Meth7} = - \left[ \frac{|\hat{Y}_{ker} - \hat{Y}_{hat}|^3}{(\hat{Y}_{ker})} \right] \quad (\text{A3.2.7})$$

$$\text{Meth8} = - \left[ \frac{(\hat{Y}_{ker} - \hat{Y}_{hat})^2}{2 * (\hat{Y}_{ker})^2} \right] \quad (\text{A3.2.8})$$

$$\text{Meth9} = - \left[ \frac{2 * (\hat{Y}_{ker} - \hat{Y}_{hat})^4}{(\hat{Y}_{ker})^2} \right] \quad (\text{A3.2.9})$$

$$\text{Meth10} = - \left[ \frac{(\hat{Y}_{ker} - \hat{Y}_{hat})^6}{(\hat{Y}_{ker})^3} \right] \quad (\text{A3.2.10})$$

The next group consists of the kernel predicted value raised to some power times the squared difference between the HATLINK and kernel predicted values. Of the variations studied in this group, all behaved almost exactly like the original part of BIIV. Consequently, only the following two versions were considered:

$$\text{Meth11} = \left(\hat{Y}_{ker}\right)\left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^2 \quad (\text{A3.2.11})$$

$$\text{Meth12} = \left(\hat{Y}_{ker}\right)^2 \left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^2. \quad (\text{A3.2.12})$$

The last group considered was of the form of the negative of some positive data-dependent scaling factor multiplied by the difference in the kernel and HATLINK predicted values raised to some power. Several different scaling factors were considered. One scaling factor was the maximum of the ratio of the kernel predicted value over the HATLINK predicted value and its inverse, that is

$$\text{maxyrat} = \text{maximum} \left( \frac{\hat{Y}_{ker}}{\hat{Y}_{hat}}, \frac{\hat{Y}_{hat}}{\hat{Y}_{ker}} \right).$$

Four methods were considered here:

$$\text{Meth13} = - \left[ \frac{\hat{Y}_{hat} * \left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^2}{\hat{Y}_{ker}} \right] \quad (\text{A3.2.13})$$

$$\text{Meth14} = - \left[ \frac{\hat{Y}_{ker} * \left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^2}{\hat{Y}_{hat}} \right] \quad (\text{A3.2.14})$$

$$\text{Meth15} = - \left[ \text{maxyrat} * \left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^2 \right] \quad (\text{A3.2.15})$$

$$\text{Meth16} = - \left[ \frac{\text{maxyrat} * \left(\hat{Y}_{ker} - \hat{Y}_{hat}\right)^4}{\left(\hat{Y}_{ker}\right)^2} \right]. \quad (\text{A3.2.16})$$

The results are summarized in Table A3.2.1.



**Table A3.2.1. Results for Variations to the Bias Part of BIIV.** In the table is the number of times the HATLINK ESE was lower than the least squares ESE out of 12 points augmented, the maximum estimated value after the 12 points were augmented, and the corresponding distance between the location of the true maximum of 42.71 at (1.25, 1.6) and the location of the estimated maximum for HATLINK, least squares and kernel regression. Spaces separate the four main groups and the original BIIV method.

Method used	# times	<u>HATLINK</u>		<u>Least squ.</u>		<u>Kernel</u>		Comments
	HATLINK wins	max	dis.	max	dis.	max.	dis	
BIIV	12	39.13	.65	39.10	.65	39.83	.65	
Meth1	2	41.77	2.03	48.36	2.03	39.83	.65	$\lambda=0$ most times
Meth2	2	41.77	2.03	48.36	2.03	39.83	.65	$\lambda=0$ most times
Meth3	5	41.77	2.03	48.36	2.03	39.83	.65	narrow wins
Meth4	9	41.56	1.66	40.13	2.03	39.83	.65	$\lambda=.75$ many times
Meth5	12	39.63	.27	39.06	.65	39.83	.65	$\lambda> .6$
Meth6	8	39.78	.65	38.23	.27	39.83	.65	$\lambda$ jumps some
Meth7	8	46.31	2.03	50.20	2.03	39.83	.65	$\lambda$ too small
Meth8	7	40.41	1.66	38.91	.65	39.83	.65	$\lambda=.75$ 10 times
Meth9	9	40.47	.65	38.58	2.03	39.83	.65	$\lambda<.75$ always
Meth10	7	38.98	.65	39.83	.65	39.83	.65	
Meth11	11	39.21	.65	39.10	.65	39.83	.65	$\lambda=1$ most times
Meth12	11	39.21	.65	39.10	.65	39.83	.65	$\lambda=1$ most times
Meth13	0	51.66	2.03	51.66	2.03	39.83	.65	$\lambda=0$ all times
Meth14	1	51.66	2.03	51.66	2.03	39.83	.65	$\lambda \neq 0$ 1 time
Meth15	1	51.66	2.03	51.66	2.03	39.83	.65	$\lambda \neq 0$ 1 time
Meth16	8	39.78	.65	38.23	.27	39.83	.65	

Ideally, at least one method would clearly work better for HATLINK in terms of ESE values, the value of the estimated maximum and the distance between the location of the estimated maximum and the location of the true maximum. As indicated by Table A3.2.1, no method worked particularly better than the original BIIV method, although Meth5 appeared to do the best overall.

For the first group of methods, the maximum response estimated using HATLINK was closer to the true maximum than the maximum obtained using least squares. But notice the distances to the true maximum. Both HATLINK and least squares regression were equally far away, about as far as possible for the region of interest. Furthermore, in terms of the ESE values HATLINK rarely performed better than least squares, even though based upon the degree of model misspecification HATLINK should outperform least squares. The main problem with these types of methods is that the selected  $\lambda$  value is too close to zero. This group was not considered further because of the poor ESE values for HATLINK and the distance between the locations of the true and fitted maxima.

The second group initially seemed to show more promise than the first group. These methods appeared to estimate more reasonable  $\lambda$  values for the degree of model misspecification considered. Also, the ESE values indicated that HATLINK was usually fitting the surface better than least squares regression. Unfortunately, the largest HATLINK fitted value is usually not located much closer to the location of the true maximum than the least squares fitted value. The one example with a distinctive “win” for HATLINK in this regard had a large distance to the true maximum (although least squares regression did also). Closer inspection of these values reveal that rather than increasing the HATLINK maximum so that it is closer to the true value, the least squares maximum is a little further away than it was for the original BIIV method.

As stated above, in terms of ESE values, the value of the maximum and the distance

between fitted and true locations, if one method must be selected, Meth5 appears to do the best overall. Based upon one run, Meth5 yielded a slightly better HATLINK maximum with a slightly smaller distance to the true as compared to the corresponding HATLINK values for BIIV. Furthermore, the least squares values for Meth5 were not quite as good as those for BIIV. Because of the stochastic nature of the augmentation method, using the same initial seed 10 repetitions of both the BIIV method and Meth5 were performed. This was done to investigate further the slightly better HATLINK fitted maximum of Meth5. If this behavior is consistently noticed, perhaps Meth5 would be better to use than BIIV. Comparison of the results of the 10 runs revealed that, after augmenting with 12 points, BIIV had an average HATLINK maximum of 39.82 while Meth5 had an average of 39.56. Based upon these 10 repetitions, Meth5 appears to be no better than BIIV on average and so nothing much was gained here.

The third group yielded almost identical results to those for BIIV. On this basis, no further variations of this type were considered.

The fourth group yielded almost identical results for HATLINK and least squares. The trouble here was that the  $\lambda$  value was zero far too many times. Only Meth16 differed from that behavior, and it behaved almost exactly like the original BIIV method did.

As indicated above, of the sixteen new variations on estimating the bias squared part of BIIV, no new method appeared to perform well enough to be considered further. Ideally, one method would perform better in terms of some ESE values, the fitted maximum and its distance to the true value. Perhaps other variations not considered could achieve all these goals. No further methods or alterations to the BIIV method will be considered here. Unless noted otherwise, the original BIIV method of (IV.1.2) will be used to augment points.

### A3.3 CONCLUSIONS

This appendix explored alternatives to the BIIV method of HATLINK augmentation. To estimate the mixing parameter, several different methods were proposed in Chapter IV. Using integrated prediction variance as the augmentation criterion, this appendix first examined the alternative ways of estimating  $\lambda$  and summarized the results, concluding that PRESS\* appeared to be superior. Also included were competing augmentation criteria, for example GCV and variations to the BIIV method. It was concluded that PRESS\* and BIIV yielded the best results in general.

## VITA

Sindee S. Sutherland was born on August 11, 1957 in Ames, Iowa to Stuart and Jean Sutherland. She graduated from Robinson High School, Fairfax, Virginia, in 1975 and entered Virginia Polytechnic Institute and State University where, after a few diversions, she earned a B.S. degree in physics in 1982. She then entered the Mathematics Department at Virginia Polytechnic Institute and State University and received a M.S. degree in 1984. She taught mathematics for several years at Radford University, Virginia Polytechnic Institute and State University, and New River Community College. In 1987, she entered the Ph.D. program in statistics at Virginia Polytechnic Institute, where she was employed as a graduate assistant. She has accepted the position of Senior Statistician at Sterling-Winthrop. The author is happily married to Keith N. Selander, whom she met as a student at Virginia Polytechnic Institute and State University.

A handwritten signature in black ink that reads "Sindee Sutherland". The signature is written in a cursive style with a large, sweeping flourish at the end.