Center for Quality and Productivity Improvement
University of Wisconsin - Madison
610 Walnut Street
Madison, WI 53705

Report No. 29

ANALYSIS OF INCOMPLETE DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C.F.J. Wu

April 1988

This research was supported by National Science Foundation Grant DMS-8420968 and was aided by access to the Mathematics Research Center and Department of Statistics research computers at the University of Wisconsin-Madison. Completed November 1987.

CENTER FOR QUALITY AND PRODUCTIVITY IMPROVEMENT
UNIVERSITY OF WISCONSIN - MADISON

Report No. 29

ANALYSIS OF INCOMPLETE DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C. F. J. Wu

April 1988

PRACTICAL SIGNIFICANCE

While incomplete data (censored or interval censored) may be easier or less costly to collect, they contain less information. More importantly they are harder to analyze. Existing methodology is inadequate for analyzing such data from highly fractionated experiments because it is computationally complicated and often not feasible to use. In this paper, we propose an iterative method which provides a simple and flexible way to consider many models simultaneously. The method's simplicity makes it easy to implement with existing software, results in computational savings, and promotes experimenter involvement. We demonstrate the procedure by reanalyzing data from three real experiments.

Keywords: quality improvement; censored and interval censored data; analysis of marginal means

---

This research was supported by National Science Foundation Grant DMS-8420968 and was aided by access to the Mathematics Research Center and Department of Statistics research computers at the University of Wisconsin-Madison. Completed November 1987.

ANALYSIS OF INCOMPLETE DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C. F. J. Wu

1. Introduction

Industrial experiments typically investigate a large number of factors in a small number of runs. Two of the experimental objective are: 1) determine the important factors that affect a critical characteristic of a product or process and 2) determine levels of the important factors that lead to improved critical characteristics. Often incomplete data are observed. This is true, for example, when a product's lifetime is the characteristic of interest. The ever increasing reliability of today's products often make it necessary to limit the duration of the experiment even when accelerated life testing is employed. Also, it may not be possible to record a unit's failure by a monitoring device, so that periodic inspection is required until the unit fails or the experiment ends. These two common types of incomplete data are right censored and interval censored, respectively.

Although less costly to collect, not only do incomplete data contain less information than complete data, but more importantly, they are harder to analyze. Existing methodology for analyzing such data is inadequate in the industrial experimental context. Standard methodology as will be described in section 2 is computationally complicated, and therefore does not promote experimenter involvement. Also, Taguchi's (1980) "minute analysis" method for analyzing interval censored data is known to be statistically incorrect (Fung 1986). We propose in this paper an iterative procedure which overcomes these drawbacks.

The iterative procedure is motivated by the fact that complete normal data are easy to analyze. This fact suggests imputing the incomplete data and treating them as complete and using a transformation of the data to achieve near normality. Then, the pseudo-complete data are analyzed using standard methods resulting in the selection of a tentative model. Based on the current selected model, the incomplete data are again imputed. This cycle of

The Center for Quality and Productivity Improvement cares about your reactions to our Reports. Please send comments (general or specific) to: Report Feedback, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705. All replies will be forwarded to the authors.

transformation, imputation, and **model selection** continues until the selected model stops changing. Because of the exploratory nature of this procedure, several models may be selected. Diagnostic checking is then done to assess the adequacy of the model(s). Based on the final model selected, one more imputation yields pseudo-complete data which are used in simple graphical plots to determine the best levels of the important factors. In some situations these plots are not applicable because of the properties of the design, but predicted responses from the final selected model can be used instead.

In section 2, the algorithm for the proposed procedure is given. In sections 3 through 5, the iterative procedure is applied to data from three real experiments. In section 3, the mixed categorical-continuous window size data of Phadke et al. (1983) from an L_{18} design are reanalyzed as left censored data. The router bit life data of Phadke (1986) from an L_{32} design are reanalyzed as right censored data in section 4. The interval censored heat exchanger data of Specht (1986) from an L_{12} design are reanalyzed in section 5. Section 6 concludes with a discussion of the procedure's pros and cons as well as its use in combination with existing methods.

2. The Proposed Procedure

Because industrial experiments often deal with quantities which take only positive values, we will consider only such responses here. One important example is a product's lifetime. A convenient way to model positive responses is to use a power transformation to transform them to near normality. Then, the convenient properties of normally distributed responses can be exploited. We take this approach because of the flexibility and simplicity it offers. We assume that the response y after some transformation, h(y), follows a linear model with normal error:

$$h(y) = \mu + \varepsilon \text{ with } \varepsilon \sim N(0, \sigma^2) . \qquad (2.1)$$

Incomplete data can be represented by the interval (a, b). For left censored data at b, $a = -\infty$. Similarly for right censored data at a, $b = \infty$.

The objective is to find $\mu = X\beta$, where β is the vector of *important* factor effects. These include main effects and interaction effects. When factors have three or more levels, β may contain linear effects, quadratic effects, etc. and their interactions.

One obvious approach for identifying the important factors and factor effects is to *specify* a comprehensive model and fit the model by maximum likelihood estimation. There are several problems with this approach, however. Typically for industrial experiments, the number of potentially important factor effects (e.g., main effects and interactions of different orders) is much larger than the number of runs, so that the maximum likelihood estimates would not exist for the comprehensive model. That is, the number of parameters is limited to the number of runs. Even if models with a smaller number of parameters are entertained, the number of such models will be too large to be computationally feasible. For example, in the router bit life experiment of section 4, there are nine factors and $\binom{9}{2} = 36$ two-factor interactions of interest. If all the models with nine main effects and three two-factor interactions were entertained, then $\binom{36}{3} = 7140$ separate MLE calculations would need to be performed. As can be seen from Table 2.1, other situations with more factors or more interactions in the model lead to an enormous number of models that need to be considered.

We briefly discuss two methods which potentially can reduce the computational effort mentioned above. Lawless and Singhal (1980) proposed an efficient algorithm for finding good fitting submodels of a full model using the maximum likelihood criterion. However, this method is usually not applicable for the situation considered here, since the maximum likehood estimates often does not exist for a comprehensive full model. Krall et al. (1975) proposed a forward selection procedure using the maximum likelihood criterion. While the problem with the existence of MLEs is lessened, the method still requires much more computation than the method proposed in this paper. If the forward selection procedure is generalized to a stepwise selection procedure, then the amount of computation is substantially increased.

Table 2.1: Number of Possible Models
For Given Number of Factors and Interactions

No. of Interactions	No. of Factors 5	6	7	8	9	10
2	45	105	210	378	630	990
3	120	455	1330	3276	7140	14190
4	210	1365	5985	20475	58905	148995
5	252	3003	20349	98280	376992	1221759

The primary difficulty with several of the approaches above is that one cannot start by specifying a comprehensive model. Taking advantage of the wealth of standard techniques for analyzing complete normal data, we employ transformation and imputation to obtain "complete normal" data, which we then use to *select* (rather than specify) a model. The procedure involves a cycle of transformation, imputation, and model selection. The algorithm for the iterative procedure consists of the following:

A. Model Selection Phase

1. Transformation
2. Initial Model Selection
3. Model Fitting
4. Imputation
5. Model Selection

Repeat steps 3 through 5 until model selection termination.

Begin at step 1 with a different transformation.

B. Model Assessment Phase

Repeat A and B until adequate model(s) are found.

The details for the algorithm are provided next.

(A1) **Transformation**. Any transformation may be used. We use the convenient Box-Cox transformation family (Box and Cox 1964) because it covers a broad range of distributions and is simple to implement. Thus, the transformation $h(y) = y^{(\lambda)}$, where

$$y^{(\lambda)} = (y^{\lambda} - 1) / \lambda \quad \lambda \neq 0$$

$$\log y \qquad \lambda = 0 \ .$$

In this step we fix λ. We do not formally handle transformation, but suggest trying several values of λ. If there are replicates, Box et al. (1978) provide a method for choosing λ.

(A2) **Initial Model Selection.** The experimenter chooses $\mu = X_0\beta_0$ (model 0) which includes main effects and interactions thought to be potentially important. We may also use the final selected models from other transformations in hopes of finding a simpler model in the current transformation.

(A3) **Model Fitting.** Fit the current model, $\mu = X_i\beta_i$ (model i), using the maximum likelihood criterion. The contribution of incomplete observation (a, b) to the likelihood is $\Phi(z_b) - \Phi(z_a)$, where $z_w = (w^{(\lambda)} - \mu) / \sigma$. The contribution of complete observation y is $\phi(z_y) \,|\, \partial y^{(\lambda)} / \partial y \,|$. Note that $\phi(z)$ and $\Phi(z)$ are the standard normal probability density function and cumulative distribution function (cdf), respectively. For censored data, an iterative least squares algorithm (Schmee and Hahn 1979) or an EM algorithm (Aitken 1981) can be used. For interval censored data, the Powell (1964) conjugate directions algorithm can be used. Professor W. Q. Meeker, Jr. graciously supplied his FORTRAN code for the latter algorithm which easily handles censored data as well. For the examples we encountered with censored data, adequate starting estimates were obtained by calculating least squares estimates by treating the censored data as complete. Similarly for interval censored data, least square estimates provided adequate starting estimates when the censored data were treated as complete and interval censored data were replaced by the interval midpoints.

(A4) **Imputation.** Impute the incomplete data by their conditional expectation:

$$E(y^{(\lambda)} \mid y \in (a, b)) = X_i\beta_i + \sigma\,(\phi(z_a) - \phi(z_b)) / (\Phi(z_b) - \Phi(z_a)) \ , \qquad (2.2)$$

$$\text{where } z_w = (w^{(\lambda)} - X_i\beta_i) / \sigma .$$

Here we use the conditional expectation as a typical value since we are interested in identifying location effects.

(A5) **Model Selection.** Apply a variety of standard techniques for analyzing complete normal data to the complete and pseudo-complete data from (A4). For example, analysis of variance (ANOVA), regression (including stepwise and subset selection procedures), normal or half-normal plots (Daniel 1959), Bayes plots (Box and Meyer 1986), and likelihood ratio tests. Stop when the current model selected is the same as the previous model, $X_i = X_{i-1}$.

(B) **Model Assessment.** Judge the final model selected by model simplicity, scientific meaningfulness, distributional adequacy, and structural model adequacy. Perform analysis of residuals using normal probability plots and plots of residuals versus predicted values to assess the model by the last two criteria. The residuals from complete data are defined to be $(y^{(\lambda)} - X\hat{\beta}) / \hat{\sigma}$.

The idea of exploiting the simplicity of complete data to solve an incomplete data problem is not new. See for example the EM algorithm (Dempster et al. 1977 and its references), the Schmee-Hahn (1979) Algorithm (an iterative least squares algorithm for handling censored regression data) and the data augmentation algorithm (Tanner and Wong 1987a, 1987b). However, its use in model selection is new, especially in the context of highly fractionated experiments.

Before proceeding to the application of our procedure to data from several real experiments, we will elaborate on model selection, model assessment, and factor level selection. For model selection, we will consider models with main effects and two-factor interactions. Regarding possible models to entertain, the 2^{k-p} fractional factorial designs should be distinguished from other designs such as those which have factors with more than two levels, mixed level designs, and Plackett-Burman designs.

For the 2^{k-p} designs, we can only entertain interactions that are not confounded with main effects. This is because when there is confounding in the 2^{k-p} design, the confounding is complete. Since only orthogonal effects are considered, we can use ANOVA, normal or half-normal probability plots, or Bayes plots to identify the important factors and effects. Therefore those interactions which the experimenter thinks may be important and which are accommodated for in the design can be considered. However, knowledge of the properties of the design can lead to the consideration of more interactions. This we call *design exploitation* and will be demonstrated in Example 2.

For the other designs, there may be partial confounding between main effects and two factor interactions. These designs may initially allow consideration of only main effects. However when only a few factors are important, interactions between them can be entertained as well. Thus, there is the potential for identifying important interaction effects. We select models entertaining these interactions by using procedures such as stepwise or subset selection regression. In other words, when the effects are not orthogonal, we can use these methods to select models as will be demonstrated in Example 1. They are also useful in handling data from experiments with unequal replication.

Regarding model assessment, distribution adequacy can be assessed by normal probability plots of the residuals. The structural adequacy can be assessed by looking for patterns in the plot of residuals versus predicted values. This raises the question of what to do when there are interval censored or censored observations. In this work, we consider the residuals from the uncensored observations. For interval censored data, we plot the predicted values and their corresponding interval observations. For distributional assessment, the Turnbull algorithm (Turnbull 1976), which calculates the empirical cdf proposed by Peto (1973), may be useful. Then the empirical cdf can be compared with the normal cdf. Finally, scientific feedback is necessary to assess the meaningfulness of the selected model.

To select the factor levels which lead to improved critical characteristics, we begin by imputing the incomplete data once more using the final model selected. Then, the analysis of marginal means (ANOMM), a simple graphical method, can be applied to these pseudo-complete data. Simply plot the means for each level of a factor when there are no interactions. Then choose the level whose mean is best. When there are interaction effects, plot the means for each combination of the factors involved in the interaction. Some care must be taken to check the properties of the design, so that the marginal mean plots are meaningful. This is especially so when there is more than one important interaction effect. Also, the validity of the ANOMM depends on the optimizing criterion. These two aspects will be discussed in Example 1. See Wu et al. (1987) for related work on the validity of ANOMM.

Another way to choose the factor levels is to use the predicted responses from the final model selected. This can be done by calculating the predicted responses for all combinations of factor levels in the model. Then, we choose the combination with the best predicted response. Note that ANOMM is a special case of this approach when there are no interactions.

Confirmatory experiments are an important aspect of experimental strategy. These are performed to evaluate whether the experimental objectives have been accomplished. They may be designed to discriminate between several choices of factor level combinations. Because the proposed procedure may suggest several models, additional runs may be required to discriminate between the different models. See Box et al. (1978) and Wu et al. (1987) for strategies on performing subsequent experiments.

In the next three sections, we demonstrate the proposed procedure using data from real experiments.

3. Example 1: Reanalysis of the Window Size Data by Phadke et al. (1983)

Consider the post-etch window size data from an experiment studying semi-conductor circuit fabrication (Phadke et al. 1983). An L_{18} $(2^1 \times 3^7)$ design was used to study the importance of nine factors (A-I) on a window forming process. The objective was to identify the important factors and their optimal levels so that the process would form window sizes close to the target of 3 μm. The L_{18} design (see Appendix 1) is a mixed-level design. A list of factors with their corresponding number of levels is given in Table 3.1.

Table 3.1: Factors and Number of Levels
for the Window Size Experiment

Label	Factor	Number of Levels
A	Mask Dimension	2
B	Viscosity	2
C	Spin Speed	3
D	Bake Temperature	2
E	Bake Time	3
F	Aperture	3
G	Exposure	3
H	Developing Time	3
I	Plasma Etch Time	3

The two-level factors B and D are incorporated into one three-level factor denoted by BD, where B_1D_1, B_2D_1, and B_1D_2 correspond to its three levels.

The data from Table III of Phadke et al. (1983) are mixed categorical-continuous. Either a window was open or not (WNO). If the window was open, then its size was measured. Note that many of the windows were not open (48%). The WNO data can be treated as left censored data by defining a censoring point L, $0 < L < \infty$. We assume that a window size smaller than L is not measurable; we choose L = 1.5 which is smaller than the smallest observed window size. The data are then modeled using (2.1) with $h(y) = y^{(0)}$, the log transformation, which has the desirable property that the predicted response in the original metric is always positive. Then the probability of a window not open is $P(WNO) = \Phi(z)$, where $z = (L^{(0)} - X\beta)/\sigma$. The WNO data are imputed by their conditional expectation using equation (2.2) which reduces to:

$$E(y^{(0)} \mid y < L) = X\beta - \sigma\phi(z)/\Phi(z) . \tag{3.1}$$

For purposes of demonstration, it will be assumed that there are 10 replicates for each of the 18 runs.

We begin the iterative scheme by entertaining a model with only main effects. This is close to the maximum number of parameters that the design can accommodate. The three-level factors contribute a linear and quadratic effect denoted by l and q, respectively. Then the initial model (model 0) consists of 15 covariates, an intercept, and a scale parameter. The maximum likelihood estimates (MLE's) for the 17 parameters are obtained and displayed in Table 3.2.

Table 3.2: Parameter Estimates for Model 0
for Window Size Experiment

σ	.26	C_l	.87	G_l	-.81
intercept	-.17	C_q	-.22	G_q	-.28
A	.45	E_l	.36	H_l	.49
BD_l	-.34	E_q	-.10	H_q	.13
BD_q	.30	F_l	.36	I_l	.14
		F_q	.14	I_q	-.28

Using (3.1), the WNO data are imputed yielding pseudo-complete data. To identify the important main effects, consider the mean squares in Table 3.3. Note that the linear and quadratic effects of factor BD correspond to D_1 vs D_2 and B_1 vs B_2 (given no D effect), respectively.

Table 3.3: Mean Squares for
Window Size Data from Model 0

mean	5.04	E_l	15.65	H_l	29.38
A	36.60	E_q	3.83	H_q	5.98
BD_l(D)	0.14	F_l	15.21	I_l	2.38
BD_q(B)	46.03	F_q	6.94	I_q	29.23
C_l	90.66	G_l	78.48	error	0.05
C_q	17.61	G_q	27.60		

Note the small mean squared error caused by the imputation of the large number of WNO data. Since there is a problem with choosing a reasonable measure of error, we do not formally use ANOVA. However, the relatively important factor effects can still be identified. Thus, tentatively C_l, B, G_l, and A are identified. Although the original analysis in Phadke et al. (1983) did not consider interactions, we will consider the six two-factor interactions between these four effects. The original 15 main effects and these six two-factor interactions are considered simultaneously by using stepwise regression, since the interactions are not orthogonal to all the main effects. Note that the same pseudo-complete data from Model 0 is still being used, here. Table 3.4 contains the output from the MINITAB stepwise regression procedure (Ryan et al. 1981) for the first eight effects. When an effect first enters the model, the table displays the t-statistic for adding the factor to the model. In subsequent steps, the t-statistics for retaining those effects already in the model are displayed.

Table 3.4: Stepwise Regression t-Statistics
for Window Size Data from Model 0

STEP	1	2	3	4	5	6	7	8
C_l	7.07	8.19	9.97	11.77	14.23	17.65	21.83	29.76
BG_l		-7.87	-9.58	-11.31	-13.67	-16.96	-20.97	-28.59
G_l			-9.28	-10.95	-13.24	-16.42	-20.31	-27.69
B				-8.39	-10.14	-12.58	-15.56	-21.20
A					9.04	11.22	13.87	18.91
G_q						-9.74	-12.05	-16.42
C_q							-9.62	-13.12
F_l								12.19
S	1.35	1.16	0.96	0.81	0.67	0.54	0.44	0.32
R^2	21.93	42.17	61.16	72.29	81.15	87.82	92.09	95.76

Table 3.4 suggests a model with the effects, C_l, B, BG_l, G_l, and A, which account for 81% of the variability. Also, a second model including G_q and C_q can be entertained. Interestingly, all six interactions considered are orthogonal to all the four corresponding main effects, C_l, B, G_l, and A. They are not all orthogonal to the remaining 11 main effects, however. So potentially many interactions are considered although for this example, only one seems to be significant. This demonstrates how easily this iterative procedure can entertain many effects.

Let us entertain the model (model 1) containing C_l, B, BG_l, G_l, A. The MLE's for model 1 are displayed in Table 3.5.

Table 3.5: Parameter Estimates for Model 1
for Window Size Experiment

σ	.32	C_l	.36
intercept	.49	G_l	-.24
A	.17	BG_l	-.31
B	-.35		

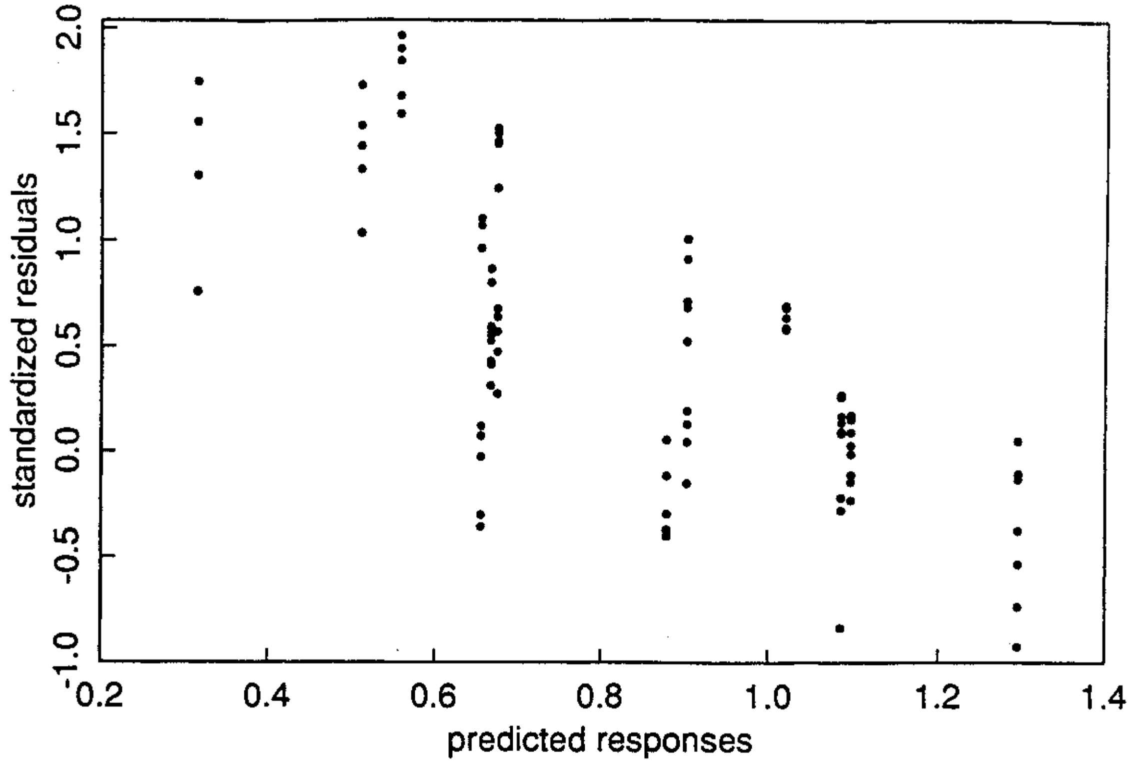
The WNO data are imputed using (3.1). A table of mean squares for main effects is similar to Table 3.3 but not included here. A stepwise regression (similar to Table 3.4 but not included here) is run considering the same 21 effects. We observe that the first five effects are the same as before, so that we terminate the model selection phase. Note that if BG_l had not been included for consideration, then H_l would have been the next effect chosen after C_l, B, G_l, and A. Here, scientific knowledge should guide which effect might be more reasonable. Further experimentation may be required to resolve this ambiguity.

We next move to the model assessment phase. We must consider the adequacy of the structural model and the transformation to normality. Figures 3.1 display a plot of residuals for the open windows against predicted values and a normal probability plot of the same residuals, respectively. Figure 3.1a reveals a definite pattern where smaller window sizes are underestimated and larger window sizes are slightly overestimated. However, the normal probability plot approximately follows a straight line, so that the assumption of normality appears to be satisfied.

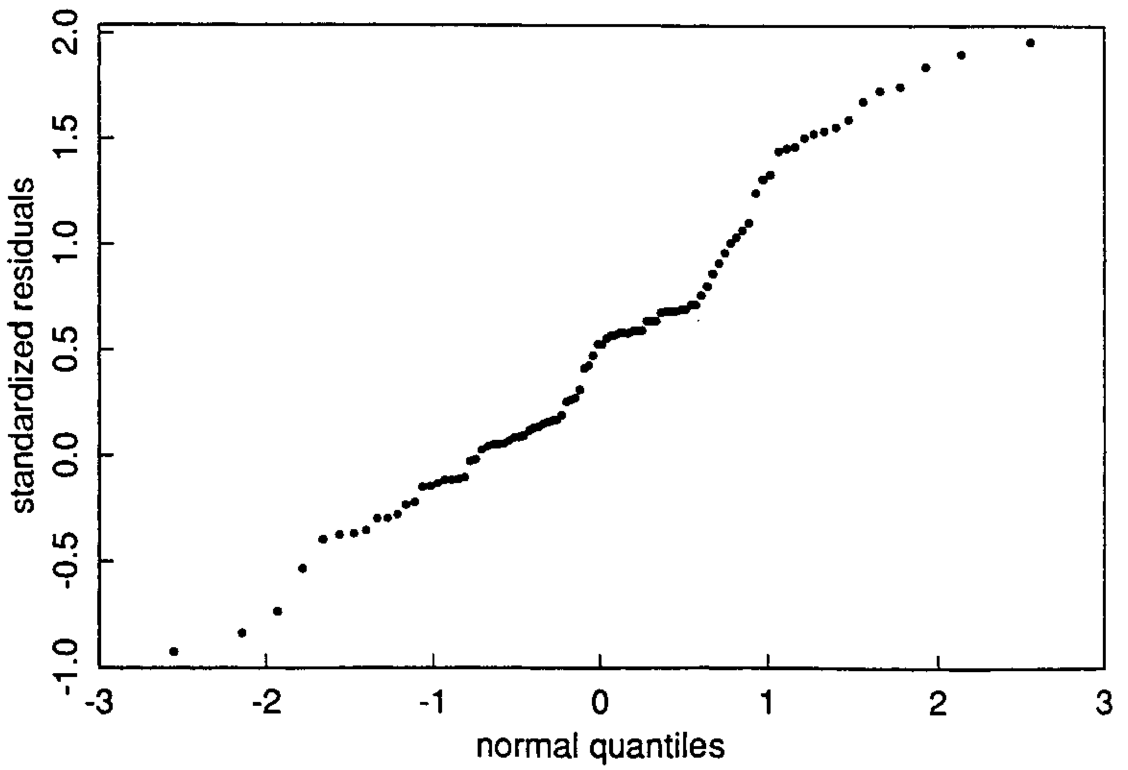
Because Figure 3.1a suggests that the structural model is not satisfactory, we entertain a new model (model 2) by adding C_q and G_q to model 1. This was the other model suggested previously, but model 1 was considered first because of its simplicity. Table 3.6 presents the estimates for model 2. The stepwise regression output (not reported here) suggests that we terminate with model 2. Figures 3.2 suggests that model 2 is more adequate than model 1. Furthermore, the log likelihoods for models 1 and 2 are -157.12 and -143.29, respectively.

Figures 3.1: Model 1 for Window Size Data

A. Uncensored Residuals versus Predicted Responses

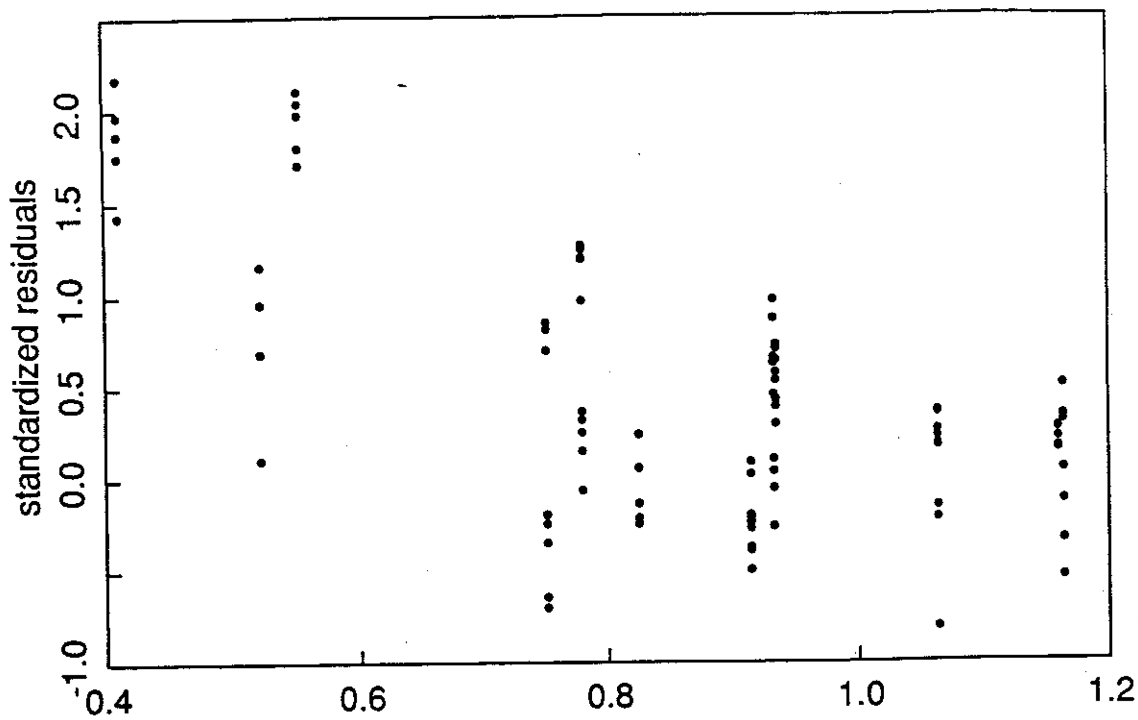


B. Normal Probability Plot of Uncensored Residuals



Figures 3.2: Model 2 for Window Size Data

A. Uncensored Residuals versus Predicted Responses



B. Normal Probability Plot of Uncensored Residuals

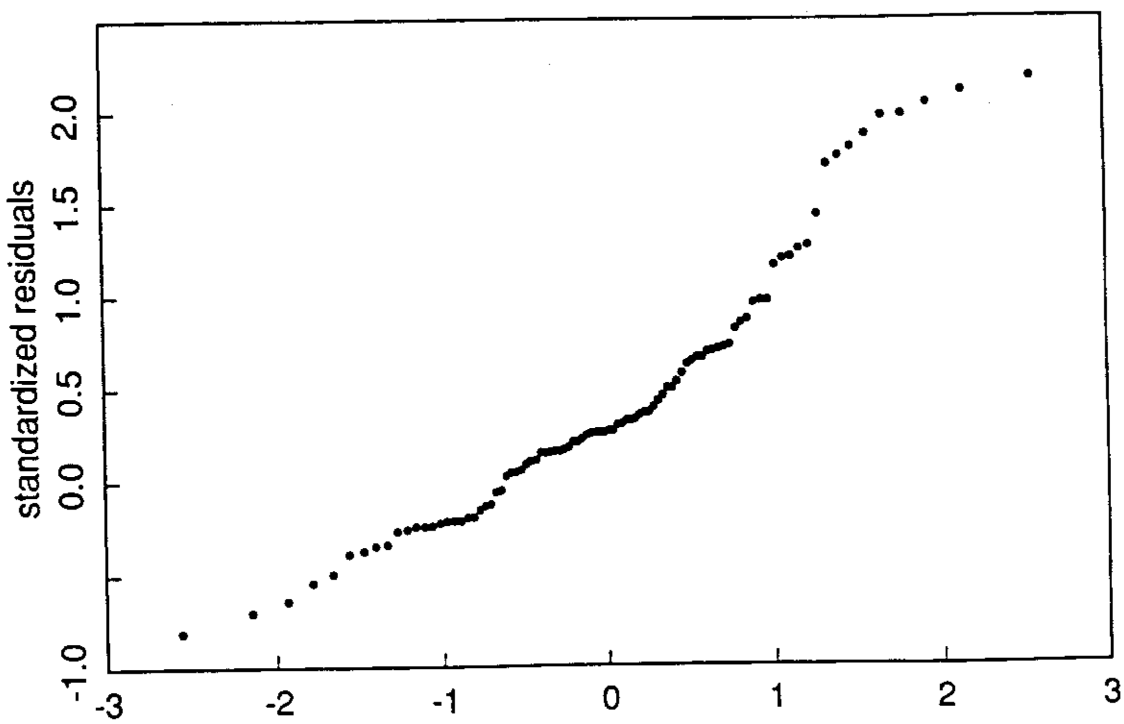


Table 3.6: Parameter Estimates for Model 2
for Window Size Experiment

σ	.30	C_l	.40
intercept	.41	C_q	-.05
A	.19	G_l	-.33
B	-.39	G_q	-.10
		BG_l	-.41

We now consider factor level selection. As mentioned in section 1, plots of the marginal means or predicted responses from the final model selected offer two ways to choose the best levels. However there are two problems with using the first method for this example: 1) the criterion of being on target and 2) properties of L_{18} for plots of marginal interaction means.

Regarding the first point, plots of marginal means are applicable for uni-directional criteria such as minimizing wear or maximizing lifetime. Consider the means at the different levels of a particular factor. Since these means average over all the levels of the other factors, they indicate the relative effect between the different levels of the particular factor. Thus, when these uni-directional criteria are used, we choose the level with the best mean. When the criterion is bi-directional such as the on target criterion, these relative effects are not helpful in determining the best factor level; the marginal mean plots are not applicable here. Thus, the factors levels obtained by selecting the level whose mean is closest to target need not lead to the best combination. However, choosing the combination of factor levels whose predicted response is closest to the target does achieve this.

Regarding the second point, suppose we have a criterion where the analysis of marginal means is applicable and the experiment uses an L_{18} design. Then, clearly if there are no interactions, the plots of marginal means provide a simple graphical method to choose the optimum levels. However, consider when there is an interaction effect such as the BG interaction identified in the current analysis. If we plot the marginal means for each of the six combinations (B has two levels, G has three levels), we need to check whether factors A and C are balanced over all six combinations. That is, if a level of one of these factors appears in one combination, it must appear in all combinations. This balancing condition is satisfied if the design is of strength 3. However, L_{18} is a design of strength 2. The use of the BD factor to accommodate B and D is an additional complication. From Table 3.7 it is clear that the requirement above does not hold. Although, Factor A is balanced, Factor C, the most important factor, is not. Therefore, plotting interaction marginal means for BG is meaningless. Similarly, plotting marginal means for factor C is meaningless. This is a drawback of the L_{18} design.

Table 3.7: L_{18} Design for Factors B, G, A, C

B	G	A	C
1	1	1	1
		2	3
		1	3
		2	1
1	2	1	2
		2	1
		1	3
		2	3
1	3	1	3
		2	2
		1	1
		2	1
2	1	1	3
		2	1
2	2	1	1
		2	2
2	3	1	2
		2	3

We next consider factor level selection using the predicted responses from the final model selected. There are 36 combinations of factors A, B, C, and G. We calculate $\mu = X\hat{\beta}$ using $\hat{\beta}$ from Table 3.6. For a particular combination, vector X is (1, A, B, C_l, C_q, G_l, G_q), where A and B take on -1 or 1, and (C_l, C_q) and (G_l, G_q) take on (−1, 1), (0, −2), or (1, 1). From calculating the predicted values for each of the 36 combinations, the combination with predicted value closest to 1.10 (=log 3) is $A_2C_3B_1G_2$. The predicted value is 1.13 and the probability of observing a window not open is .008 ($=\Phi([\log(1.5) - 1.13]/.30)$). From the design in Appendix 1, run number 18 happens to have the optimal factor level combination. Among the runs with no WNO data, run 18 has the second smallest root mean squared error or mean absolute deviation; run number 17 ($A_2C_2B_1G_1$) is slightly better.

4. Example 2: Router Bit Life Data

Consider the router bit life data from Phadke (1986). 8x4 inch printed wiring boards are cut from an 18x24 inch panel by a routing process. When the router bit becomes dull, it produces boards with rough edges which requires an extra cleaning process. Also, frequently changing the router bits is expensive. A router bit fails when it begins producing rough edges (determined by evidence of an excessive amount of dust), where router bit life is measured in (x100) inches of cut in the x-y plane. An L_{32} design was used to study the importance of nine factors (A-I), seven two-level factors and two four-level factors. Table 4.1 presents the factors and number of levels. See Appendix 2 for the L_{32} design and data.

Table 4.1: Factors and Number of Levels
for the Router Bit Life Experiment

Label	Factor	Number of Levels
A	Suction	2
B	X-Y Feed	2
C	In-Feed	2
D	Bit Type	4
E	Spindle Position	4
F	Suction Foot	2
G	Stacking Height	2
H	Slot Depth	2
I	Speed	2

The experiment was stopped after 17(x100) inches. Eight of the 32 router bits did not fail when the experiment was stopped so that some of the data are right censored. Actually, the router bits were inspected every 100 inches, so that the data are interval censored. For purposes of demonstration, we will use the midpoints of the intervals for the interval censored data. The experimenters were interested in the relative importance of the nine main effects and four two-factor interactions, BI, CI, GI, and BG. The experimental objective was to select the optimal factor levels which maximize router bit life.

Because the data are lifetimes, it is reasonable to consider some transformation of the data; the lognormal distribution is widely used for lifetime data. In fact, the log transformation seems to be the appropriate metric here for reasons to be discussed later. Only pertinent details will be given. Thus, the data are modeled using (2.1) with $h(y) = y^{(0)} = \log y$. The right censored observations are imputed using equation (2.2) which reduces to:

$$E(y^{(0)} \mid y > R) = X\beta + \sigma\phi(z)/(1 - \Phi(z)) , \tag{4.1}$$

$$\text{where } z = (R^{(0)} - X\beta) / \sigma \text{ and } R = 17.$$

Since the L_{32} design was used to accommodate nine main effects and four two-factor interactions, we take this to be the initial model (model 0). Since there are two four-level factors, the initial model contains 17 effects, an intercept, and a scale parameter.

We fit the initial model using maximum likelihood estimation and use (4.1) to impute the censored data. To identify which effects are important, consider the mean squares in Table 4.2 based on the pseudo-complete data.

Table 4.2: Mean Squares from
Router Bit Life Data for Model 0

mean	32.80	E	2.87	BG	.09
A	.41	F	7.14	BI	.05
B	7.49	G	16.78	CI	.78
C	.64	H	.02	GI	8.27
D	9.58	I	9.23	error	.89

Clearly, G, D, I, GI, B, and F are important. Factor E was thought to be unimportant a priori but was included because of a constraint imposed by the router machine; Table 4.2 confirms this.

One might propose these six effects for the next tentative model. However, knowledge of the structure of the L_{32} design and further analysis of the same pseudo-complete data, leads us to consider additional effects. First consider the alias structure of the design. Although the four-level factors prevent the design from being a 2^{k-p} factional factorial design, we can still find its alias structure by associating the effects in the L_{32} design with the 31 effects in a 2^5 full factorial design. Let 1-5 denote the generating columns. Then the factor effects in the initial model correspond to:

$$A = 2,\ B = 3,\ C = -23,\ D = (234, -25, 345),\ E = (4, 5, -45),\ F = -35,$$

$$G = -2345,\ H = -24,\ I = 1,\ BG = 245,\ BI = -13,\ CI = 123, \text{ and } GI = 12345.$$

From this we deduce that D is partially aliased with (AG, BH, CF) and E with (AH, BF, CG). Also AB, AC, BC, FG, FH, and GH are completely aliased with main effects because C=−BA and H=−FG. Out of the 21 two-factor interactions between the seven two-level factors, AF, CH, AI, FI, and HI are orthogonal to all the effects in the initial model. Thus, assuming factor E to be unimportant, we can consider these five effects as well as AH, BF, and CG. Table 4.3 displays the mean squares for these eight additional effects which is quite revealing.

Table 4.3: Additional Mean Squares from
Router Bit Life Data for Model 0

AF	5.52	CH	.02
BF	3.45	AI	.43
CG	5.00	FI	1.90
AH	.20	HI	1.48

Table 4.3 suggests that we should also consider AF and CG. Therefore the next tentative model (model 1) under consideration contains G, D, I, GI, B, F, AF, and CG. A half-normal plot in Figure 4.1 suggests the same model.

The MLE's are obtained and displayed in Table 4.4.

Table 4.4: Parameter Estimates for Model 1
for Router Bit Life Experiment

σ	.65	F	-.46
intercept	1.55	G	-.74
B	-.61	I	.57
D(2)	-1.74	AF	.52
D(3)	-1.03	CG	-.52
D(4)	.96	GI	.54

The table of mean squares based on the pseudo-complete data from model 1 are shown in Table 4.5, indicating that we can stop the iterative procedure. A half-normal plot in Figure 4.2 suggests the same. Note a better separation of the significant and insignificant effects in Figure 4.2. We terminate the model selection phase, so that model 1 is the final model.

Figure 4.1: Half-Normal Plot for Model 0 Router Bit Life Data
lambda = 0

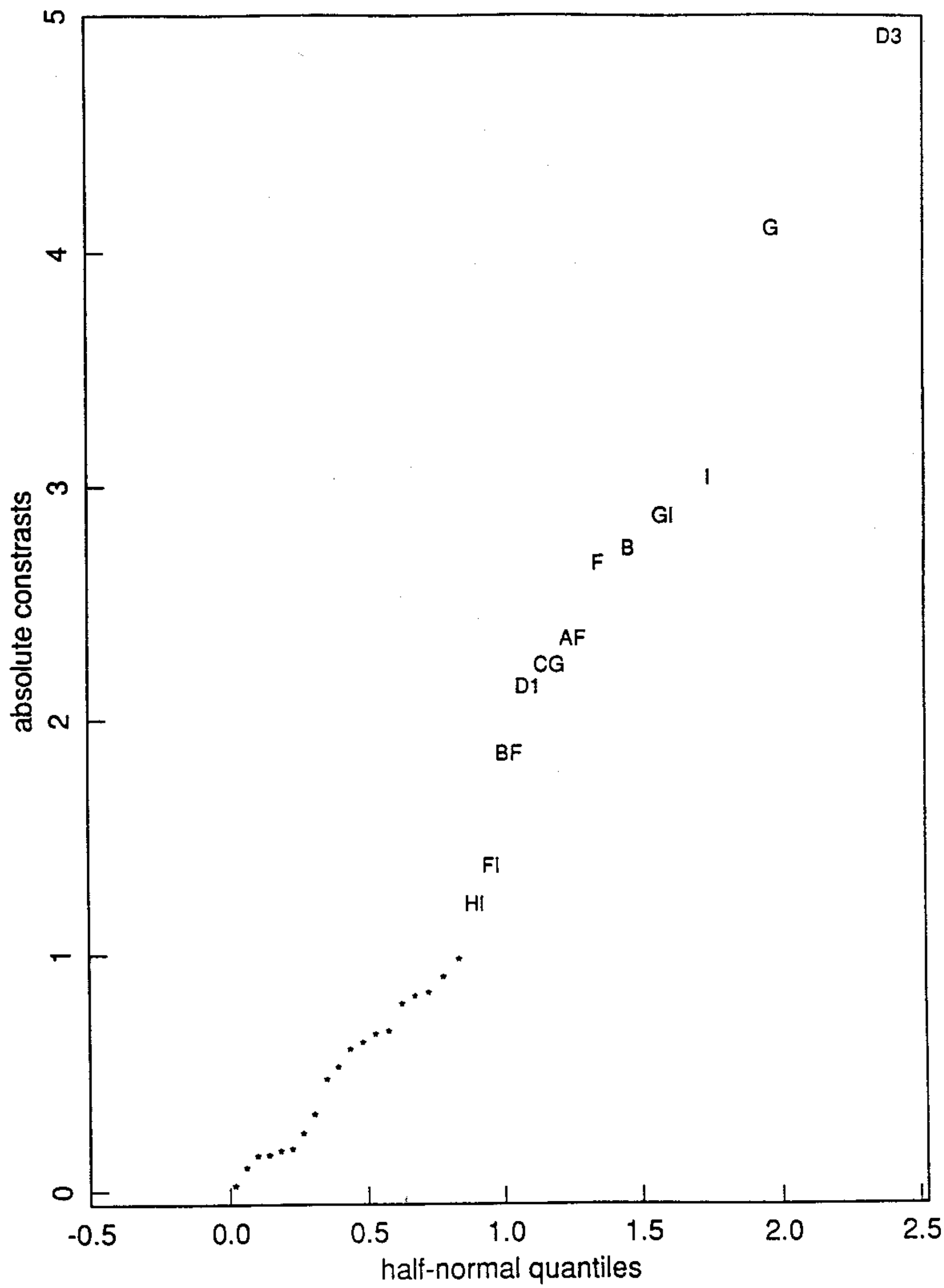


Figure 4.2: Half-Normal Plot for Model 1 Router Bit Life Data

lambda = 0

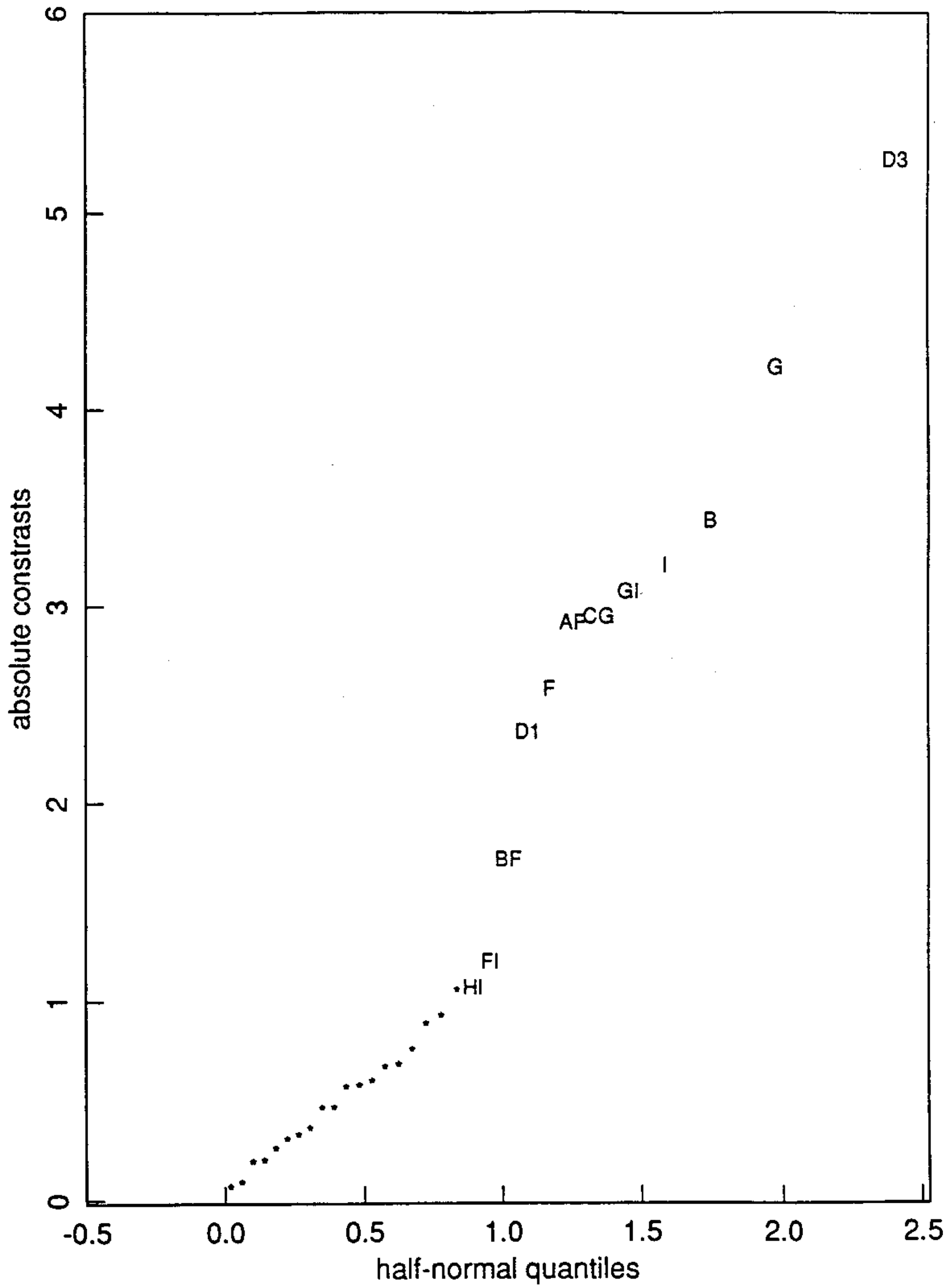


Table 4.5: Mean Squares from
Router Bit Life Data for Model 1

mean	38.48	I	10.31	CH	.09
A	.42	AF	8.54	CI	.19
B	11.83	AH	.06	FI	1.45
C	.20	AI	.54	GI	9.48
D	11.14	BF	2.96	HI	1.15
F	6.67	BG	.00	error	.42
G	17.76	BI	.03		
H	.33	CG	8.72		

Suppose that the design had not been exploited, so that model 1 did not contain the AF and CG effects. If model 1 without AF and CG was fit and the mean squares for the eight additional effects in Table 4.3 had been calculated at this point, then AF and CG would have again been identified. Thus the iterative procedure would have stopped after one more iteration obtaining the same final model as was obtained when the design was exploited at an earlier stage. This suggests that the iterative procedure does terminate with the same model even when different intermediate models are entertained.

Consider the adequacy of the distributional assumption. Figures 4.3 displays a normal probability plot of the uncensored residuals which suggests that there is also some departure from normality. A model with BF added to model 1 was considered, but there was no improvement in the normal probability plot.

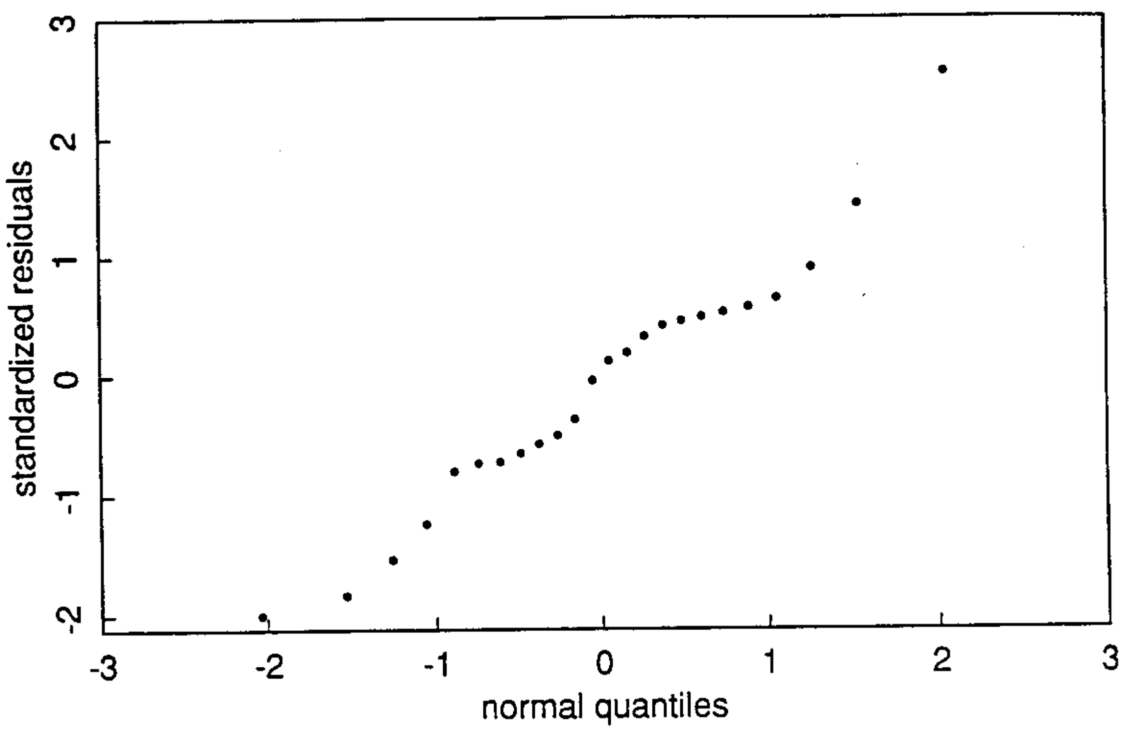
We now choose the optimal factor levels. There are 256 combinations of factors A, B, C, D, F, G, and I. By calculating the predicted response, $\hat{\mu} = X\hat{\beta}$, using $\hat{\beta}$ from Table 4.4, the combination $A_2B_1C_1D_4F_1G_1I_2$ gives the maximum predicted lifetime. Since the starting combination is $A_2B_1C_1D_4F_2G_2I_2$, the recommendation is to change F and G.

From the discussion of the analysis of marginal means (ANOMM) in Example 1, its appropriateness depends on the optimizing criterion and the design. First, the criterion of maximizing lifetime is appropriate. However, it can be verified that the combinations of CG are not balanced over the combinations of GI. Although, the ANOMM does not apply, the marginal mean plots in Figure 4.4 identify the same factor levels. This suggests that more study is required to understand why ANOMM is appropriate even though the balancing condition is violated.

Figure 4.3: Model 1 for Router Bit Life Data

lambda = 0

Normal Probability Plot of Uncensored Residuals



The use of other transformations will be discussed next. We analyzed the router bit life data for $\lambda = -1, -.5, 0, .5, 1$. Table 4.6 displays the final models selected with their corresponding log likelihoods. The order of effects is based on the magnitude of the mean squares from the pseudo-complete data from the final model.

Figures 4.4: Marginal Mean Plots for Model 1
Router Bit Life Data lamda = 0

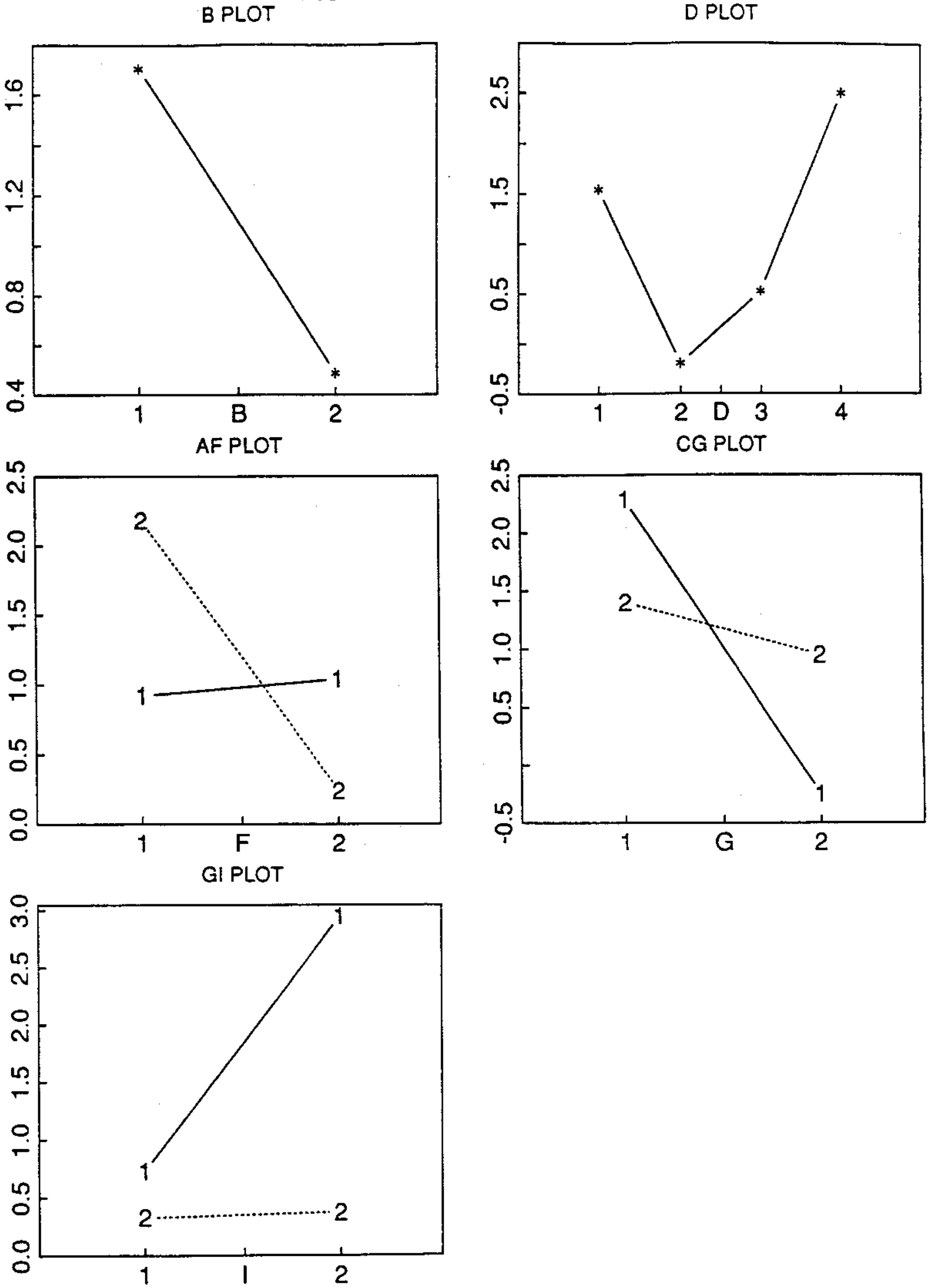


Table 4.6 Final Models for Different Transformations
with Corresponding Log Likelihoods
for the Router Bit Life Experiment
(order of effects based on mean squares)

		λ		
-1	-.5	0	.5	1
G	G	G	G	G
BF	D	B	I	I
D	B	D	GI	GI
B	BF	I	B	AF
F	F	GI	AF	F
A	I	CG	F	D
	GI	AF	D	B
	AF	F	CG	CG
	CG		HI	FI
			FI	HI
-47.70	-27.23	-29.36	-31.39	-47.89

From Table 4.6, FI and HI decrease in importance as λ decreases while A and BF increase in importance. The final model using the log transformation is simple with comparable likelihood. Although the final model with $\lambda = -1$ is the simplest, the departure from normality was substantial.

Analysis of the router bit data has revealed some interesting properties of the iterative procedure. Knowledge of the design structure can be exploited which can lead to consideration of additional effects not suggested by the experimenter. The procedure can entertain these additional effects in a flexible manner. Furthermore, the procedure was shown to terminate with the same model although different tentative intermediate models are considered.

5. Example 3: Heat Exchanger Data

Specht (1986) reports on an experiment using a 12 run Plackett-Burman design to study the importance of 10 two-level factors (A–H, J, K) on the lifetime of a heat exchanger. The unit fails when it develops a tube wall crack with lifetime measured in (x100) cycles. Each run was checked after cycles 42, 56.5, 71, 82, 93.5, 105, and 116 and stopped after 128 cycles. Note that the intervals are of uneven length. The design and data appear in Appendix 3. The goal was to find the important factors and factor settings to optimize the lifetime of the heat exchanger. We will report on the results for the reciprocal transformation. The results from other transformations will be discussed later.

The initial model (model 0) entertained is the main effects model with the 10 factors. The model is fit using maximum likelihood and the interval censored data are then imputed using (2.2). This yields pseudo-complete data which we can first analyze to determine the important main effects. By calculating the contrasts for the 10 main effects and the one error effect (denoted by e in the plots), we can identify the important effects by using half-normal plots. From Figure 5.1, only factor E appears to be important. Note that the error contrast is larger than the other 9 factor contrasts with contrast B, K, and C being larger than the remaining contrasts.

Assuming that no other main effect is important we can entertain potentially important interactions between factor E and the other nine factors. From Figure 5.2, the half-normal plot of E and these interactions suggests that EG and EH are important as well. Therefore, the next model (model 1) we entertain contains E, EG, and EH. Fitting this model yields the MLE's given in Table 5.1.

Table 5.1: Parameter Estimates for Model 1
for Heat Exchanger Data Experiment

σ	.000102
intercept	.984675
E	-.004252
EG	.002305
EH	-.001927

Based on the imputed data using model 1, the half-normal plots in Figures 5.3 and 5.4 confirm that only the factor E main effect and interactions EG and EH are important. Note the better separation between the significant and insignificant effects in Figures 5.4. Thus, we terminate the model selection phase.

Recall that in Figure 5.1, contrasts e, B, K, and C were somewhat larger than the six other contrasts. Table 5.2 presents the alias structure of the 11 contrasts in terms of the 10 main effects, EG, and EH.

Figure 5.1: Half-Normal Plot for Model 0 Heat Exchanger Data

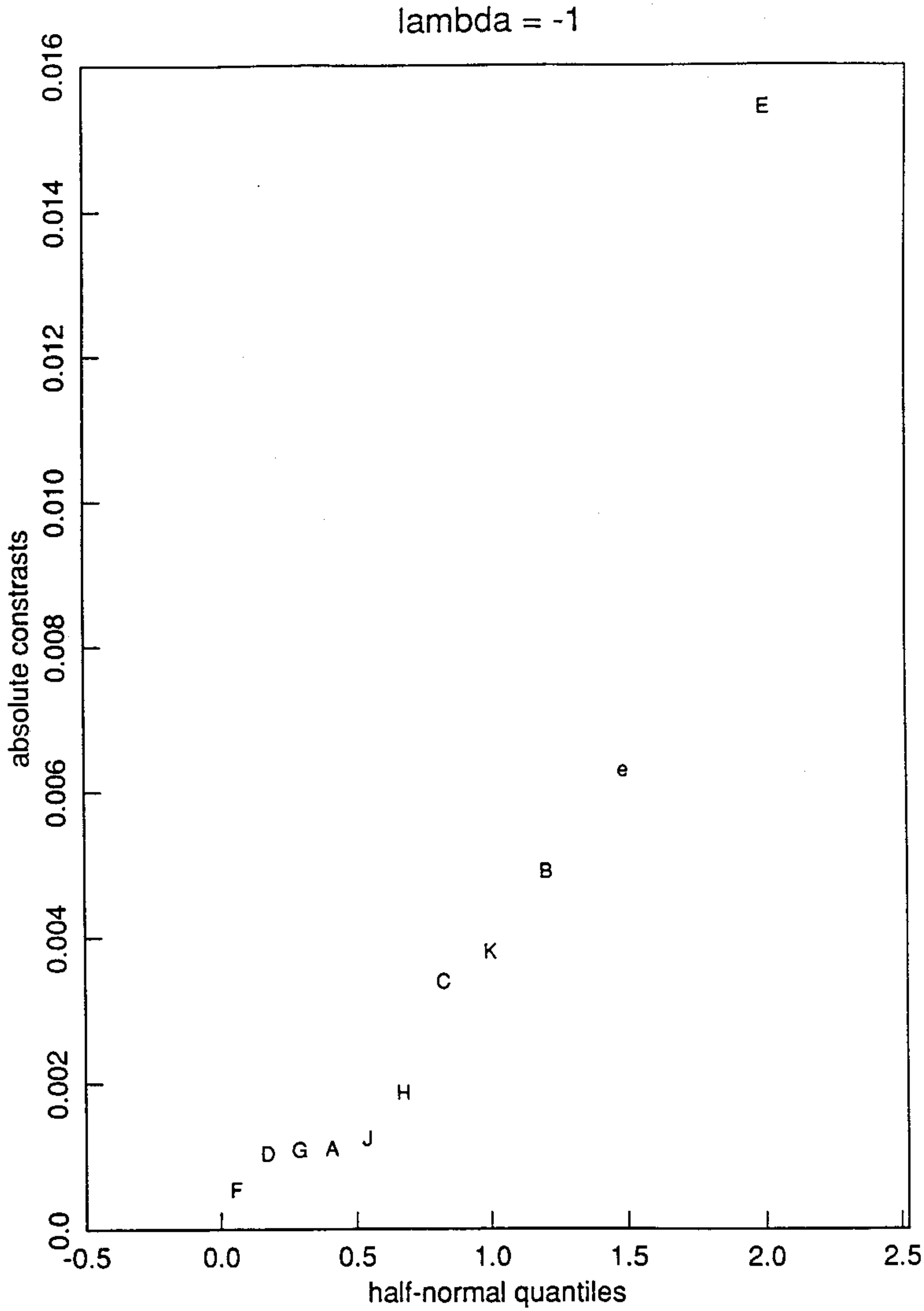


Figure 5.2: Half-Normal Plot for Model 0 Heat Exchanger Data

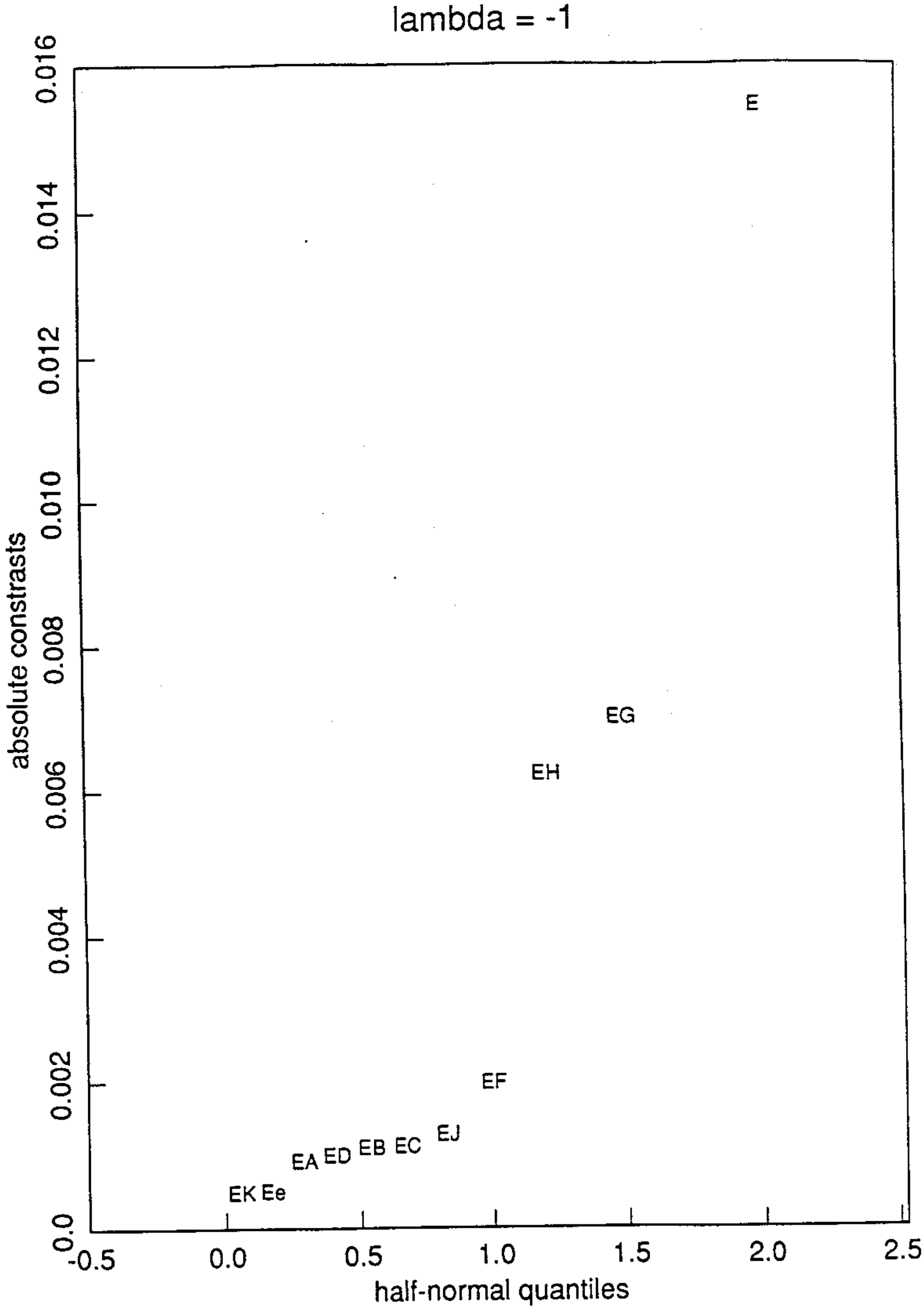


Figure 5.3: Half-Normal Plot for Model 0 Heat Exchanger Data

lambda = -1

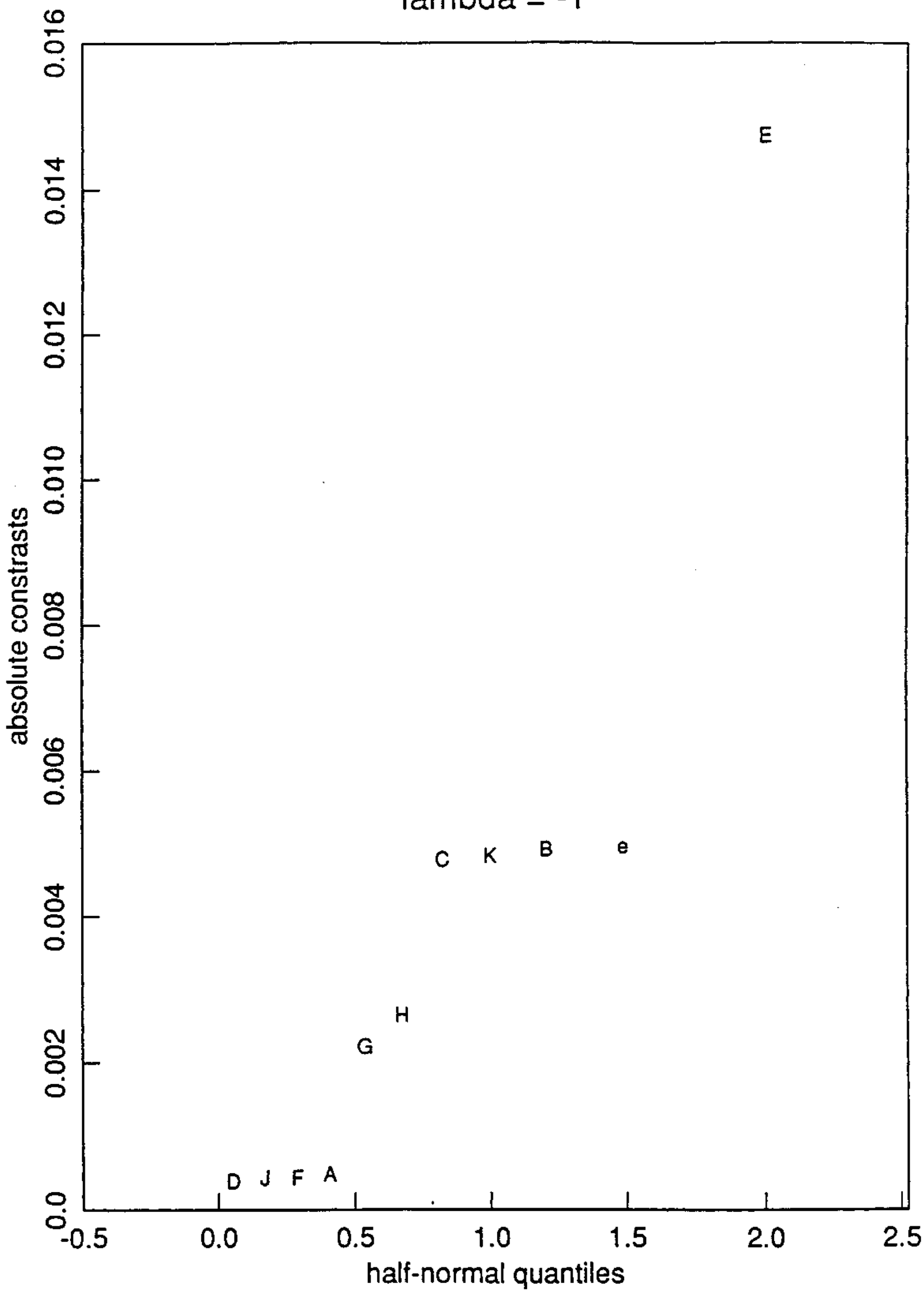


Figure 5.4: Half-Normal Plot for Model 0 Heat Exchanger Data

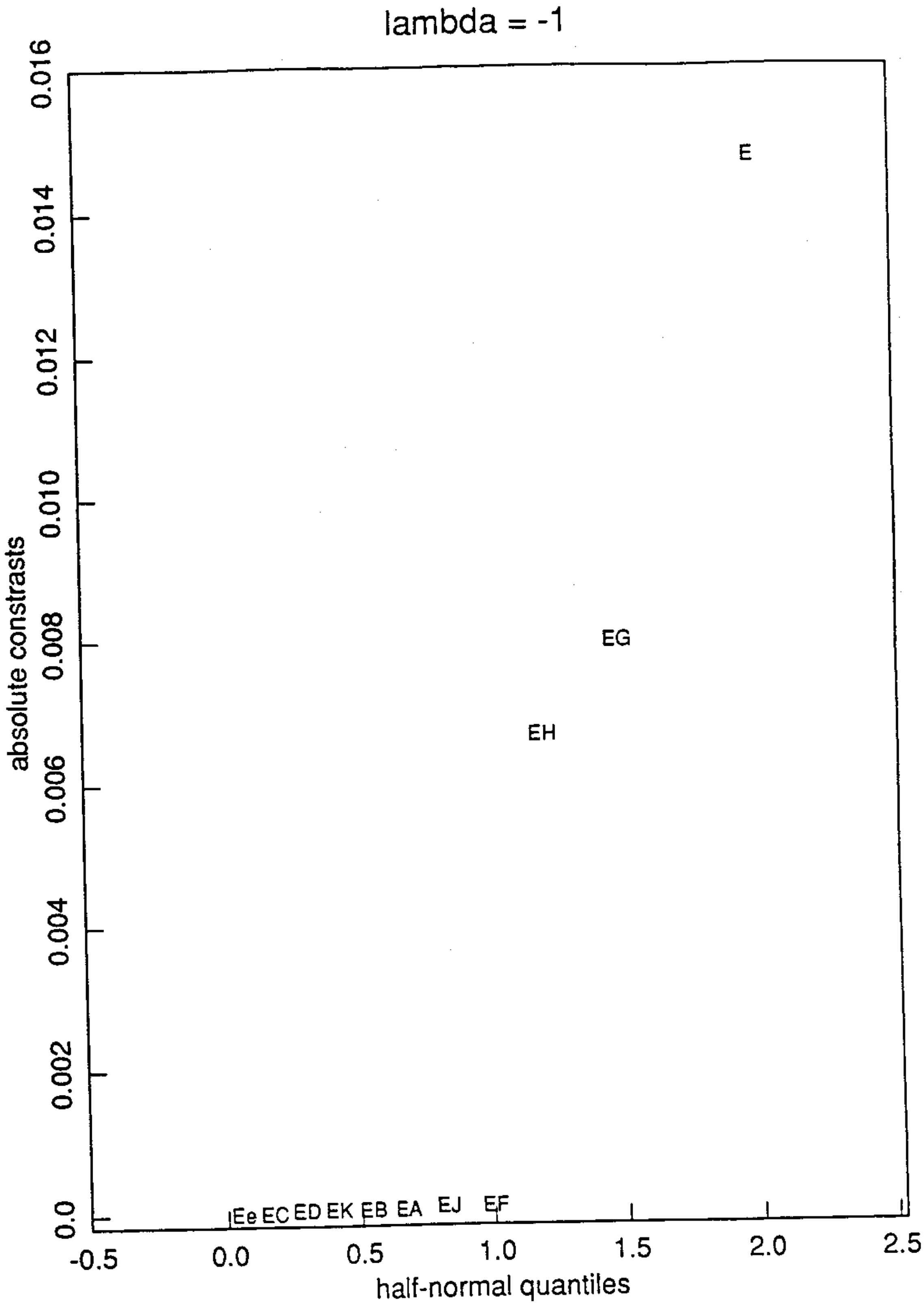


Table 5.2: Alias Structure for L_{12} Design
in the 10 main effects, EG, and EH.

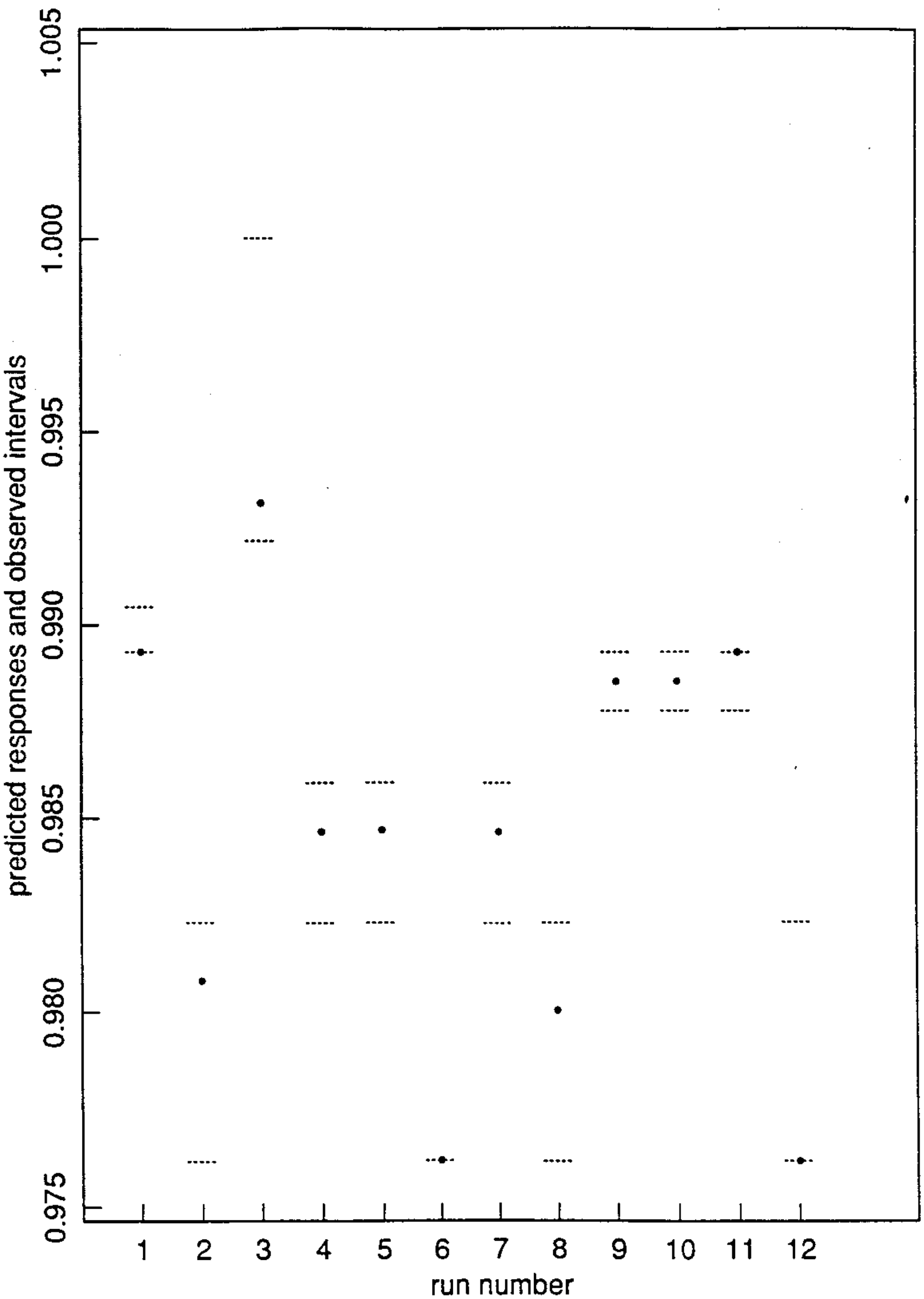
column	alias string
1	F - 1/3 EG -1/3 EH
2	B - 1/3 EG +1/3 EH
3	A - 1/3 EG -1/3 EH
4	C + 1/3 EG -1/3 EH
5	D - 1/3 EG -1/3 EH
6	E
7	G - 1/3 EG -1/3 EH
8	H - 1/3 EG -1/3 EH
9	J + 1/3 EG +1/3 EH
10	K + 1/3 EG -1/3 EH
11	e - 1/3 EG +1/3 EH

If only main effect E and interactions EG and EH are important, Table 5.2 suggests that contrasts e, B, K, and C would be larger than the remaining six contrasts. This is because the coefficients have opposite sign as do the effects EG and EH. Based on the estimates of EG and EH from Table 5.1, $1/3(EG+EH) = 0.000126$ and $1/3(EG-EH) = 0.001410$; the difference is more than ten-fold. Furthermore, e and B should be negative and K and C should be positive which can be verified by inspection of the contrasts. This further suggests that E, EG, and EH are important. Note that this is not the only explanation since the L_{12} design has a very complicated alias structure. From Draper and Stoneman (1966), each main effect is partially confounded with the 36 two factor interactions among the other nine factors. The error contrast is partially confounded with the 45 two factor interactions among the ten factors. Again, additional experiments and scientific feedback are necessary to resolve this ambiguity.

It is not clear how to do model checking with interval censored data, especially with such a small sample, except to inspect a plot of predicted values versus the observed intervals for the 12 runs. From Figure 5.5, almost all the predicted values do fall within the observed intervals.

The transformations $\lambda = 0, 1$ were also considered. The conclusions are the same except that EG and EH appear to be somewhat less important. This demonstrates that the right transformation increases the sensitivity to detect real effects (Fung 1986). A comparison of the likelihoods using $\lambda = -1, 0, 1$ for the model with E, EG, and EH are -2.77, -7.36, and -12.56, respectively. Therefore, on the basis of the likelihood criterion the reciprocal transformation is better.

Figure 5.5 Predicted Responses
for Model 1 Heat Exchanger Data



Furthermore, plots of the predicted values and observed intervals for $\lambda = 0, 1$ are not as good. Note that we also entertained a model with only factor E and a model with E and EG. Both led to the same final model above (model 1) providing further evidence that the iterative procedure terminates with the same model although different intermediate models are entertained.

We consider next the choice of the best levels for factors E, G, and H. Since maximizing lifetime is equivalent to maximizing $y^{(\lambda)}$, from Table 5.1 it can be easily seen that $E_1G_1H_2$ yields the largest predicted lifetime (146.18). Note that the lifetime for run 3 was the largest (i.e., the only right censored observation) and had these very factor levels. From Figure 5.6, the analysis of marginal means gives the same conclusions although EG and EH do not satisfy the balancing condition.

In the original analysis using Taguchi's minute analysis (Taguchi and Wu 1980), only the main effect E was detected. Note that $EG - EH \sim E$, so that incorrect settings of G and H could vitiate the effect of the optimal setting of E.

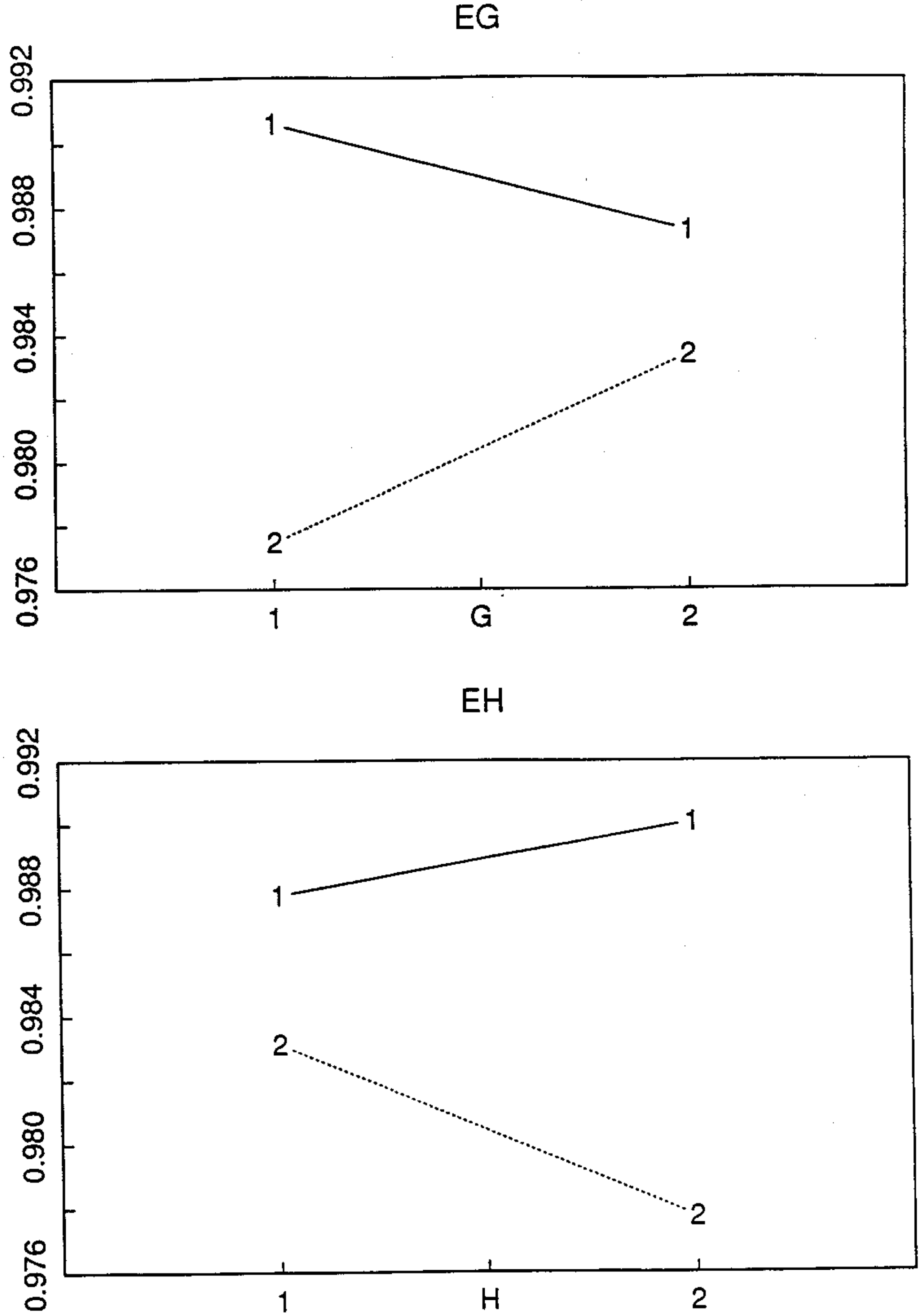
6. Discussion

The motivation for the proposed procedure is that complete normal data are much simpler to analyze than incomplete data. Via transformation and imputation, we obtain pseudo-complete normal data that can be analyzed to identify the important factors and effects. Indeed, the procedure's simplicity results in computational savings; the examples required at most two iterations. More importantly, the procedure's simplicity allows the use of standard methods such as ANOVA and graphical techniques which promotes experimenter involvement in the process of selecting the model. The procedure is flexible because it can entertain many models simultaneously when ANOVA is applied to the pseudo-complete data. A great advantage of ANOVA is its ability to consider many models simultaneously by identifying the significant effects. Also, the procedure exploits the knowledge of the design structure which leads to consideration of additional effects not originally considered by the experimenter. Finally, the procedure can be used in combination with more sophisticated methods which provides a quick yet comprehensive analysis strategy for incomplete data. More will be said about this shortly, but first we will briefly discuss some criticisms of using the procedure by itself.

One criticism it that the imputation ignores the variability of the incomplete data which might lead to the incorrect choice of factors and factor levels. Note that the validity of the model

Figures 5.6 Marginal Mean Plots Model 1

Heat Exchanger Data



is addressed in model assessment phase. Also, the procedure might miss other good fitting models since it is not as exhaustive as using the stepwise MLE procedure. When the proposed procedure is used in combination with more sophisticated methods, these criticisms disappear. The proposed procedure can be viewed as a quick way to identify a good starting model for these other procedures. Then, both adding or deleting effects in the model as in a stepwise procedure will ensure that other promising models are not missed. The Lawless and Singhal (1980) algorithm would be appropriate to use here. Much computation has been avoided by having a good starting model. Also, the analysis can be supplemented by using a Weibull or gamma distribution which is especially attractive in reliability experiments. Hence, the proposed procedure can be used alone or viewed as a complement rather than a competitor to these more sophisticated procedures.

The appropriateness of the analysis of marginal means was discussed in the context of the examples considered. Its validity depends on the criterion being optimized, the properties of the experimental design, and the final response model. It must be realized that there are assumptions made when using this graphical technique and that indiscriminate use of this method should be avoided.

Incomplete data are much easier and less costly to obtain. However, they are much harder to analyze. Numerous reports on industrial experiments using incomplete data is evidence that this is a practical and important problem. We have proposed a procedure that easily handles incomplete data using a combination of standard and easily accessible techniques. This work suggests that our procedure is a useful addition to the industrial statistician's toolbox.

Acknowledgements

We thank Bill Meeker for providing extensive comments on an earlier version. We also thank George Box, Norman Draper, Conrad Fung, Tim Kramer, and Martin Tanner for helpful discussions. This research was supported by National Science Foundation Grant DMS-8420968 and was aided by access to the the Mathematics Research Center and Department of Statistics research computers at the University of Wisconsin-Madison.

References

Aitken, M. (1981), "A Note on the Regression Analysis of Censored Data," *Technometrics*, 23, 161-163.

Box, G. and Cox, D. R. (1964), "An Analysis of Transformations," (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 26, 211-252.

Box, G., Hunter, J. S. and Hunter W. G. (1978), *Statistics for Experimenters*, New York: John Wiley.

Box, G. and Meyer, D. (1986), "An Analysis for Unreplicated Fractional Factorial Two Level Experiments, " *Technometrics*, 28, 11-28.

Daniel, C. (1959), "Use of Half-Normal Plots in Interpreting Factorial Two Level Experiments," *Technometrics*, 1, 311-341.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1-38.

Draper, N. R. and Stoneman, D. M. (1966), "Alias Relationship for Two-Level Plackett and Burman Designs," Technical Report No. 96, Department of Statistics, University of Wisconsin-Madison.

Fung, C. (1986), "Statistical Topics in Off-Line Quality Control," unpublished Ph.D. thesis, University of Wisconsin-Madison.

Krall, J. M., Uthoff, V. A. and Harley, J. B. (1975), "A Step-Up Procedure for Selecting Variables Associated with Survival," *Biometrics*, 31, 49-57.

Lawless, J. F. and Singhal, K. (1978), "Efficient Screening of Nonnormal Regression Models," *Biometrics,* 34, 318-327.

Peto, R. (1973), "Experimental Survival Curves for Interval-Censored Data," *Applied Statistics,* 22, 86-91.

Phadke, M. S. (1986), "Design Optimization Case Studies," *AT&T Technical Journal,* 65, 51-68.

Phadke, M. S., Kackar, R. N., Speeney, D. V. and Grieco, M. H. (1983), "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design Optimization," *The Bell System Technical Journal,* 1, 1273-1309.

Powell, M. J. D. (1964), "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," *Comp. J.,* 7,155-162.

Ryan, Jr., T. A, Joiner, B. L. and Ryan, B. F. (1981), *Minitab Reference Manual.*

Schmee, J. and Hahn, G. (1979), "A Simple Method for Regression Analysis with Censored Data," *Technometrics,* 21, 417-432.

Specht, N. (1985), "Heat Exchanger Product Design via Taguchi Methods," in *Third Symposium on Taguchi Methods,* Romulus, MI: American Supplier Institute, Inc., 302-318.

Taguchi, G. and Wu, Y. (1980), *Introduction to Off-Line Quality Control.* Nagoya, Japan: Central Japan Quality Control Association.

Tanner, M. A. and Wong, W. H. (1987a), "An Application of Imputation to an Estimation Problem in Grouped Lifetime Analysis," *Technometrics,* 29, 23-32.

Tanner, M. A. and Wong, W. H. (1987b), "The Calculation of Posterior Distributions by Data Augmentation," (with discussion), *Journal of the American Statistical Society,* 82, (in press).

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society,* Ser. B, 38, 290-295.

Wu, C. F. J., Mao, S. S. and Ma, F. S. (1987), "An Investigation of OA-Based Methods for Parameter Design Optimization," Report No. 24, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Appendix 1: Design for Window Size Experiment

run	factor							
	A	BD	C	E	F	G	H	I
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

Appendix 2: Design and Data for Router Bit Life Experiment

	factor									
run	A	B	C	D	E	F	G	H	I	DATA
1	1	1	1	1	1	1	1	1	1	3.5
2	1	1	1	2	2	2	2	1	1	0.5
3	1	1	1	3	4	1	2	2	1	0.5
4	1	1	1	4	3	2	1	2	1	17*
5	1	2	2	3	1	2	2	1	1	0.5
6	1	2	2	4	2	1	1	1	1	2.5
7	1	2	2	1	4	2	1	2	1	0.5
8	1	2	2	2	3	1	2	2	1	0.5
9	2	1	2	4	1	1	2	2	1	17*
10	2	1	2	3	2	2	1	2	1	2.5
11	2	1	2	2	4	1	1	1	1	0.5
12	2	1	2	1	3	2	2	1	1	3.5
13	2	2	1	2	1	2	1	2	1	0.5
14	2	2	1	1	2	1	2	2	1	2.5
15	2	2	1	4	4	2	2	1	1	0.5
16	2	2	1	3	3	1	1	1	1	3.5
17	1	1	1	1	1	1	1	1	2	17*
18	1	1	1	2	2	2	2	1	2	0.5
19	1	1	1	3	4	1	2	2	2	0.5
20	1	1	1	4	3	2	1	2	2	17*
21	1	2	2	3	1	2	2	1	2	0.5
22	1	2	2	4	2	1	1	1	2	17*
23	1	2	2	1	4	2	1	2	2	14.5
24	1	2	2	2	3	1	2	2	2	0.5
25	2	1	2	4	1	1	2	2	2	17*
26	2	1	2	3	2	2	1	2	2	3.5
27	2	1	2	2	4	1	1	1	2	17*
28	2	1	2	1	3	2	2	1	2	3.5
29	2	2	1	2	1	2	1	2	2	0.5
30	2	2	1	1	2	1	2	2	2	3.5
31	2	2	1	4	4	2	2	1	2	0.5
32	2	2	1	3	3	1	1	1	2	17*

* right censored observation

Appendix 3: Design and Data for Heat Exchanger Experiment

	factor										
run	F	B	A	C	D	E	G	H	J	K	DATA
1	1	1	1	1	1	1	1	1	1	1	(93.5, 105)
2	1	1	1	1	1	2	2	2	2	2	(42, 56.5)
3	1	1	2	2	2	1	1	2	2	2	(128, ∞)
4	1	2	1	2	2	2	2	1	1	2	(56.5, 71)
5	1	2	2	1	2	1	2	1	2	1	(56.5, 71)
6	1	2	2	2	1	2	1	2	1	1	(0, 42)
7	2	1	2	2	1	2	2	1	2	1	(56.5, 71)
8	2	1	2	1	2	2	1	1	1	2	(42, 56.5)
9	2	1	1	2	2	1	2	2	1	1	(82, 93.5)
10	2	2	2	1	1	1	2	2	1	2	(82, 93.5)
11	2	2	1	2	1	1	1	1	2	2	(82, 93.5)
12	2	2	1	1	2	2	1	2	2	1	(42, 56.5)