ANALYSIS OF CENSORED DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C. F. J. Wu

IIQP Research Report
RR-90-06

July 1990

ANALYSIS OF CENSORED DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C. F. J. Wu

IIQP Research Report
RR-90-06

July 1990

ANALYSIS OF CENSORED DATA FROM HIGHLY FRACTIONATED EXPERIMENTS

M. Hamada and C.F.J. Wu

Department of Statistics and Actuarial Science
University of Waterloo

ABSTRACT

While censored (including interval censored) data may be easier or less costly to collect than complete data, they contain less information and are harder to analyze. Existing methodology is inadequate for analyzing such data from highly fractionated experiments. 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 two real experiments. A simulation study also demonstrates its superiority over some existing methods including Taguchi's minute accumulating analysis.

Key Words: quality improvement, interval censored data, minute accumulating analysis

1. Introduction

In industrial experiments, data are often censored. For example, the ever increasing reliability of today's products often makes it necessary to limit the duration of a lifetesting experiment. Also, it may not be possible to use a monitoring device to record a unit's failure time so that periodic inspection is required until the unit fails or the experiment ends. These two common types of censored data are known as right censored and interval censored data, respectively.

While less costly to collect, censored data contain less information than complete data and are harder to analyze. This is especially so for the highly fractionated factorial experiments commonly used in industry to study a large number of factors in a small number of runs. In this context, existing methodology is inadequate for analyzing censored data to 1) determine the important factors that affect a quality characteristic of a product or process and 2) to choose levels of these factors that lead to improvement. As described in Section 2, standard methodology for censored data is computationally complicated and often infeasible to use. Alternative methods such as a quick and dirty method, the Hahn-Morgan-Schmee (1981) procedure, and Taguchi's (1987) minute accumulating analysis have more serious deficiencies which are presented in this section.

In Section 3, we propose an iterative procedure which overcomes these drawbacks. The procedure is motivated by the fact that complete normal data are easy to analyze. This suggests imputing the censored data and treating them as complete after a suitable transformation of the data to achieve near normality. Standard methods can then be used informally to select a tentative model based on the *pseudo-complete* data, i.e., the combined complete and imputed data. Next the current model is fitted and then the censored data are imputed again. This cycle of **fitting**, **imputation** and **model selection** continues until the selected model stops changing. Because of the exploratory nature of this procedure, several models may be identified. Diagnostic checking and a formal analysis of the final model(s) can then be done to assess their adequacy. While the model selection step lacks a rigorous theoretical justification and therefore should be used with caution, nevertheless, it simply and quickly identifies useful models. For the examples presented in this article, the final models chosen were confirmed by a more "rigorous" analysis in this assessment phase. Section 3 discusses these points in more detail.

In Sections 4 and 5, we use the proposed procedure to reanalyze data from two real experiments, the right censored router bit life data of Phadke (1986) and the interval censored heat exchanger life data of Specht (1986). In Section 6, a simulation study demonstrates the proposed method's superiority over some existing procedures. Section 7 concludes with a discussion of the advantages and disadvantages of the proposed procedure.

2. Some Existing Methods

We review some existing methods for handling censored data and note some of their limitations.

2.1. Fitting Comprehensive Models and Their Submodels

One obvious approach for identifying the important factors and factor effects is to *specify* a model and fit it by maximum likelihood estimation (MLE). Then the standard errors of the estimates can be used to judge the effects' importance. Note that while unequal variances induced by censoring are accounted for, there are several problems with this approach. First, the MLEs may not exist. An important example is when all the observations at a factor's high level are right censored and are uncensored at its low level. While the effect of such a factor is large and provides a great opportunity for quality improvement, its MLE does not exist. Hamada and Tse (1989) showed that the nonexistence problem is prevalent for saturated and nearly saturated models with heavy censoring. (A more detailed discussion is given in Section 7.) Second, the computational cost can be quite high because of many possible models to be fitted. Typically for industrial experiments, the number of potentially important factor effects can be much larger than the number of runs and therefore, rules out the use of a comprehensive model. Even if smaller models are entertained, the number of such models can still be very large. 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 only models with nine main effects and three two-factor interactions were entertained, then $\binom{36}{3} = 7140$ separate MLE calculations would be required.

There are two methods which can potentially reduce the amount of computation. Krall, Uthoff and Harley (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 that proposed in this article. If the forward selection procedure is generalized to a stepwise selection procedure, the amount of computation increases substantially. Lawless and Singhal (1980) proposed an efficient algorithm for finding good fitting submodels of a full model using the maximum likelihood criterion. This method encounters computational difficulties in the industrial context since the MLEs for the full model usually do not exist.

2.2. A Quick and Dirty Method

A quick and dirty (QD) method used in practice treats the censoring times as actual failure times and then analyzes them by standard methods for complete data. See Phadke (1986) for an example. Although simple, ignoring the censoring information can lead to wrong decisions. This occurs because the unobserved failure and censoring times may differ greatly depending on the particular factor level combination. The simulation study in Section 6 shows that this method can perform quite badly.

2.3. Taguchi's Minute Accumulating Analysis Method

Taguchi's (1987) minute accumulating analysis (MAA) is a method for interval censored data. MAA can also be used to analyze data with right censoring since the last group corresponds to right censored data and the uncensored data can be viewed as interval censored data with small intervals. For each observation, MAA generates a series of binary data in the following way: a "one" is assigned to all groups preceding the group in which the unit failed; the remaining groups are assigned a "zero". Thus, a time factor is created whose levels correspond to these groups. Then MAA performs an ANOVA on this generated binary data, treating them as if they came from a split-plot experiment (Taguchi 1987); the main-plot factors are those factors studied in the experiment and the sub-plot factor is the created time factor.

Fung (1986) has questioned MAA's validity by focusing on the generated dependent binary data and their corresponding huge number of degrees of freedom in the ANOVA table. Here, we focus on the main-plot analysis which, except for the last group, is approximately an ANOVA on the lifetime data. Viewed in this way, MAA like the QD method treats censoring times as actual failure times, and thus has the same deficiencies. Note that the incorrect choice of the degrees of freedom in MAA does not affect the relative importance of the factorial effects. On the other hand, treating the censoring times as actual failure times has more serious consequences. In particular, the relative importance of the effects can be reversed as shown in Section 6. See Hamada (1989b) for a detailed study of MAA's properties.

2.4. The Hahn-Morgan-Schmee Method

The Hahn-Morgan-Schmee (1981) (HMS) method is an iterative model selection method based on an iterative least squares procedure described in Schmee and Hahn (1979). The HMS method consists of a two-step loop: imputation followed by model selection. Initially, censoring times are treated as actual failure times and a model is chosen using regression based on least squares estimates (LSEs). That is, HMS' initial model is the same as that chosen by the QD method. In successive steps, the censoring times are replaced by conditional expectations based on the current chosen model's LSEs. The imputed lifetimes are then treated as actual failure times and the next model is chosen using regression. The procedure stops when the model chosen and the estimates of the effects in the model stop changing. Although simple to implement, the HMS method can perform poorly as shown in Section 6 since the final model can be unduly influenced by the initial model choice. Other problems with this method are also discussed there.

Next we propose a procedure that overcomes the drawbacks of methods which directly use the censoring information like those in Section 2.1 and yet retains the simplicity and flexibility of those in Sections 2.2-2.4.

3. The Proposed Procedure

Because industrial experiments often deal with positive valued quantities, e.g., failure times, we consider only such responses here. 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 its *flexibility* and *simplicity*.

Assume that the response y after some transformation, h(y), follows a linear model with normal error:

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

Censored 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 $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of *important* factor effects and $\mathbf{X}$ is the corresponding matrix of explanatory variables. These include main effects, interactions, linear and quadratic effects (for quantitative factors) and contrasts between groups of levels (for qualitative factors).

The primary difficulty with the standard methods for censored data presented in Section 2.1 is that one cannot start by *specifying* a comprehensive model or even a smaller saturated or nearly saturated model. However, by transforming and imputing to obtain "complete normal" data, the wealth of standard techniques for complete normal data is then available to *select* (rather than specify) a model. This is the basis of the model selection phase of the proposed procedure which involves a cycle of fitting, imputation, and model selection. The remaining two phases assess the model(s) chosen and recommend levels for the important factors.

The proposed procedure consists of the following:

A. Model Selection Phase

1. Initial Model Specification
2. Model Fitting
3. Imputation

4. Model Selection

Repeat steps 2 through 4 until model selection termination.

B. Model Assessment Phase

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

C. Factor Level Recommendation

Details for the procedure are provided next.

(A1) **Initial Model Specification.** The experimenter chooses $\boldsymbol{\mu} = \mathbf{X}_0\boldsymbol{\beta}_0$ (Model 0) which includes main effects and interactions thought to be potentially important. For highly fractionated designs, one may be restricted to a main effects model.

(A2) **Model Fitting.** Fit the current model, $\boldsymbol{\mu} = \mathbf{X}_i\boldsymbol{\beta}_i$ (Model i), using the maximum likelihood criterion. The contribution of a censored observation (a, b) to the likelihood is $\Phi(z_b)-\Phi(z_a)$, where $z_w = (h(w) - \mu)/\sigma$ and h is the transformation in (2.1). The contribution of a complete observation y is $\phi(z_y)\,|\partial h(y)/\partial y|$, where $\phi(z)$ and $\Phi(z)$ are the standard normal probability density function and cumulative distribution function (cdf), respectively. The maximum likelihood estimates can be calculated by easily available procedures such as the EM algorithm (Aitken 1981), the conjugate directions algorithm (Powell 1964), or the Newton-Raphson method.

(A3) **Imputation.** Impute the censored data by their conditional expectation:

$$E(h(y)\,|\,y \in (a,\, b)) = x_i\boldsymbol{\beta}_i + \sigma\,(\phi(z_a) - \phi(z_b))\,/\,(\Phi(z_b) - \Phi(z_a))\ , \tag{2.2}$$

where $z_w = (h(y) - x_i\boldsymbol{\beta}_i)/\sigma$. Since we are interested in identifying location effects, we use the conditional expectation as a typical value. We refer to the combined complete and imputed data as *pseudo-complete* data.

(A4) **Model Selection.** Informally apply a standard technique to the pseudo-complete data from (A3) to select a model. Stop when the current model selected is the same as the previous model, i.e., $\mathbf{X}_i\boldsymbol{\beta}_i = \mathbf{X}_{i-1}\boldsymbol{\beta}_{i-1}$.

(B) **Model Assessment.** Verify the final model chosen by performing a formal analysis. Judge it by simplicity, structural adequacy, and scientific meaningfulness. Analyze the residuals to assess the distributional adequacy.

(C) **Factor Level Recommendation.** To select factor levels which lead to improved lifetime, use the predicted responses from the final model selected. After calculating the predicted responses for all combinations of factor levels, choose the combination with the best predicted response.

The proposed method has some similarities to the HMS method, the overlap being steps A3 and A4. A crucial difference is step A2 where the MLEs are obtained for the initial model specified in step A1 and are updated for the current model selected in step A4. Unlike the HMS method, the proposed procedure directly incorporates the censoring information in the fitting step.

Next we elaborate on the selection of models based on the pseudo-complete data. Typically, one model is chosen at each iteration. If several models were equally plausible, the procedure could be applied separately to each of them. The beauty of using pseudo-complete data is that many models are simultaneously entertained so that each possible model does not have to be evaluated individually as required by the formal MLE procedures of Section 2.1. Since the problem of model selection amounts to choosing the cutoff between the significant and insignificant effects, there is a hierarchy of models; that is, if there are several plausible models, they are nested within each other. To carry out the above, one approach is to choose the largest model. If some of the parameter estimates are negligible in the subsequent fitting step, then the current model is effectively reduced to a smaller one. These two aspects, the *hierarchy* and *simultaneous consideration* of models account for the procedure's simplicity.

For 2^{k-p} fractional factorial designs, we can only entertain interactions that are not aliased with main effects. The relative size of suitably standardized effects can suggest which effects are important. Alternatively, half-normal probability plots (Daniel 1959) can be used informally to identify the important main effects and interactions. Besides the interactions which the experimenter thinks may be important, knowledge of the properties of the design can lead to the consideration of additional interactions. This we call *design exploitation* and demonstrate it in Example 1.

For 3^{k-p} fractional factorial, mixed level, and Plackett-Burman designs, there is partial aliasing between main effects and two factor interactions. These designs may initially allow consideration of only main effects. However if only a few factors are important, then some two-factor interactions may be entertained as well. That is, the same pseudo-complete data could be used to consider models containing these interactions by informally using procedures like stepwise or subset selection regression. This strategy is used in Example 2.

A word of caution in using the standard methods in step A4 is necessary since it is not theoretically justifiable. The usual assumptions of uncorrelated effects with equal variances underlying normal probability plotting do not hold because of the censoring and imputation. However, one can still rank and informally select effects based on the relative magnitudes of the standardized effects. Although a "standard" method works as well as the formal MLE method in the two examples and performs satisfactorily in the simulation study, the user should be aware of this potential pitfall.

Regarding model assessment, comparison of the final model MLEs with their respective standard errors provides a quick check of model adequacy. Note that the standard errors are easily calculated after the MLEs are obtained. This quick check confirmed the final models chosen by the proposed procedure in the examples. Distributional and structural adequacy might also be assessed by a normal probability plot of residuals and a plot of residuals versus predicted values, respectively. These techniques are discussed in Lawless (1982), but have apparently not been studied for censored data in the industrial context. Even if a few observations are censored, the standard methods of analyzing only the residuals from the uncensored observations may still not apply, since the censored data were used to obtain the estimates. More serious questions arise regarding assessment with interval censored observations. For distributional assessment, the cdf of the error distribution can be estimated using the Turnbull (1976) algorithm and compared with the normal cdf. For such few data as in

Example 2, this and perhaps any method appear to be useful for detecting only gross departures from normality. For a crude assessment of structural model adequacy, we plotted the predicted value for each experimental run to see if it fell within the corresponding observed interval.

Regarding transformation, the power transformation (Box and Cox 1964) $h(y) = y^{(\lambda)}$, where $y^{(\lambda)} = (y^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$ and $\log y$ for $\lambda = 0$, provides a convenient family of transformations and is simple to implement. Here we do not handle transformation in a formal way, but suggest trying several values of λ and applying the proposed procedure. Also, subject matter expertise should guide the choice of an appropriate transformation. Note that the other methods in Section 2 except for Taguchi's MAA can incorporate transformation.

Confirmatory experiments are an important aspect of experimental strategy. These are performed to evaluate whether the experimental objectives have been accomplished. They may also 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 choose between them. See Box, Hunter, and Hunter (1978) and Wu, Mao and Ma (1990) for strategies on performing subsequent experiments.

In the next two sections, we demonstrate the proposed procedure by reanalyzing data from two real experiments.

4. Example 1: Router Bit Life Data

Phadke (1986) reported on an experiment to improve router bit life for a routing process that cuts 8x4 inch printed wiring boards from an 18x24 inch panel. 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. Failure is 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. A 32 run design was used to study the nine factors given in Table 1. The design and data appear in Table 2.

Table 1: Factors and Number of Levels
for the Router Bit 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 had not failed when the experiment was stopped so that the corresponding data are censoring times. Actually, the router bits were inspected every 100 inches giving interval censored data. For purposes of demonstration, however, we ignore this and use the midpoints of the intervals as actual failure times. 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 factor levels which improve router bit life.

We illustrate the proposed method for the commonly used lognormal regression model. Thus, the data are modeled using (2.1) with $h(y) = y^{(0)} = \log y$. We also analyzed this data for λ = -1, -.5, .5, and 1. Only details for the log transformation are reported here since it gave a simple model with a comparable large likelihood. The right censored observations are imputed using equation (2.2) which reduces to:

$$E(\log y \mid y > R) = x\boldsymbol{\beta} + \sigma\phi(z)/(1-\Phi(z)) \; , \qquad (4.1)$$

where $z = (\log R - x\boldsymbol{\beta})/\sigma$ and $R = 17$.

Since the 32 run design was used to accommodate nine main effects and four two-factor interactions, we take this to be the initial model in step A1. Thus, the initial model (Model 0) contains 17 effects, an intercept and a scale parameter.

We fit the initial model using maximum likelihood estimation (step A2) and use (4.1) to impute the censored data (step A3). Using this pseudo-complete data, we compute in Table 3 the least squares estimates (LSEs) of the effects for Model 0 using regression from which we tentatively identify which effects are important (step A4). Some explanation is required for the two qualitative factors D and E. Each is represented as three effects whose corresponding contrasts are orthogonal to those for the remaining effects. Based on the design structure as discussed below, D's three effects are aliased with (AG, BH, CF) and E's with (AH, BF, CG). A priori, factor E was not thought to be important so that if an effect is significant, it is attributed to the corresponding interaction. On the other hand, a significant D effect is interpreted as a difference between the router bit types.

Table 2: Design and Data for Router Bit 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

From Table 3, D, G, I, GI, B, F, BF and CG appear relatively important. Interestingly, BF and CG were not considered in the original analysis because the design's structure was not taken into account. That is, knowing the design's structure allows us to entertain additional effects. Here, using what we call *design exploitation,* we can in fact consider eight additional effects.

The design's structure can be seen by associating the effects in the 32 run 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 aliased with (AG, BH, CF) and E with (AH, BF, CG) as claimed above. Also AB, AC, BC, FG, FH, and GH are completely aliased with main effects because C = –BA and H = –FG. Moreover, we can also entertain the remaining two-factor interactions among the seven two-level factors, AF, CH, AI, FI, and HI. Table 4 displays the estimates of these five additional effects (using the same Model 0 pseudo-complete data) which is quite revealing. It suggests that we should also consider AF. Therefore the next tentative model (Model 1) contains G, D, I, GI, B, F, AF, CG and BF. An informal reading of the half-normal plot of the 31 effects in Figure 1 supports the same model. (An asterisk is plotted for the 19 smallest effects and D1 and D2 denote two of D's three effects.)

Table 3: LSEs from Model 0 Pseudo-Complete Router Bit Data

A	.113	E1(AH)	-.079	I	.537
B	-.484	E2(BF)	.329	BG	.053
C	.142	E3(CG)	-.395	BI	-.039
D1(AG)	.868	F	-.472	CI	-.156
D2(BH)	-.379	G	-.724	GI	.508
D3(CF)	-.023	H	.023		

Table 4: Additional LSEs from Model 0
Pseudo-Complete Router Bit Data

AF	.415	FI	.244
AI	.115	HI	.215
CI	-.156		

*** Figure 1 about here

Next Model 1 is fit (step A2) using the original data whose MLEs appear in Table 5. For the model, we use the indicator variables for levels 2-4 denoted by D(2), D(3) and D(4) (i.e., the relative effects to level 1) since they directly provide information for comparing the different levels of D.

Table 5: MLEs for Model 1 Based on
the Original Router Bit Data

σ	.52	G	-.77
intercept	1.48	I	.56
B	-.56	AF	.51
D(2)	-1.70	BF	.39
D(3)	-.93	CG	-.50
D(4)	.98	GI	.53
F	.39		

D(x) is an indicator variable for level x.

We then impute the censored data using these MLEs (step A3). Based on the resulting pseudo-complete data, we compute the LSEs using regression (step A4) which indicate that the model selection phase can be stopped (see Table 6). A half-normal plot in Figure 2 suggests the same. Note the better separation between the significant and insignificant effects in Figure 2 as compared with Figure 1. A quick comparison of the MLEs with their corresponding standard errors confirms Model 1. Also, residual plots not shown here reveal nothing unusual.

Table 6: LSEs from Model 1 Pseudo-Complete Router Bit Data

A	.114	H	.102	CG	-.522
B	-.608	I	.568	CH	-.054
C	.078	AF	.516	CI	-.078
D1(AG)	.931	AH	-.042	FI	.213
D2(BH)	-.418	AI	.130	GI	.544
D3(CF)	-.060	BF	.304	HI	.189
F	-.457	BG	-.008		
G	-.745	BI	.032		

*** Figure 2 about here

For recommended factor levels, there are 256 combinations of factors A, B, C, D, F, G, and I. By calculating the predicted response, $\mu = x\hat{\boldsymbol{\beta}}$, using $\hat{\boldsymbol{\beta}}$ from Table 5, 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. Note that the same recommendation is obtained using the main effect and interaction mean plots in Figure 3. These plots are produced by first imputing the censored data once more using the final model. Then, for a factor with no interaction such as D, the means at each level using the pseudo-complete data are plotted. For factors with interaction effects, the means for each combination of the factors involved in the interaction are plotted. Finally, choose the level whose mean is best, e.g., the one with the longest lifetime.

*** Figure 3 about here

In the original analysis, Phadke (1986) used the quick and dirty method (see Section 2). Here it appears that it gave similar results for the initial model of 9 main effects and 4 two-factor interactions. However, we entertained 8 additional two-factor interactions of which three appear to be important, AF, BF and CG. Thus factors A and C are important through their interaction with other factors which affects the choice of factors levels. Because the recommended levels for A and C were the same as the starting ones, there is no difference in our recommendations and those of Phadke (1986). However, if cost had been a supplementary criterion and the alternative levels were cheaper, the QD approach could have chosen the wrong levels for A and C.

5. Example 2: Heat Exchanger Life Data

Specht (1986) reports on an experiment using a 12 run Plackett-Burman design to study how 10 factors (A–H, J, K) affect a heat exchanger's reliability. A 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; the design and interval censored data appear in Table 7. Note that the same final model was identified no matter what transformation was used for λ in (-1, 1). We present the results for the reciprocal transformation (λ = -1) since it gave the largest likelihood.

Table 7: 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	$(128, \infty)$
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)

The initial model (Model 0) is the main effects model with the 10 factors (step A1). The MLEs are obtained (step A2) and the interval censored data are then imputed using (2.2) (step A3). Next, the resulting pseudo-complete data are first analyzed to determine the important main effects (step A4). By calculating the contrasts for the 10 main effects and the one error effect (denoted by e in the figures), we can identify the important effects. From Figure 4, which shows the relative size of the effects, only factor E appears to be important.

*** Figure 4 about here

Assuming that no other main effect is important, we can entertain potentially important interactions between factor E and the other nine factors using the same pseudo-complete data from the initial model. From Figure 5, the size of these interactions relative to E suggests that EG and EH are important as well. Therefore, the next model (Model 1) contains E, EG, and EH. Fitting this model to the original interval censored data (step A2) yields the MLEs given in Table 8.

*** Figure 5 about here

Table 8: MLEs for Model 1 Based on the Original Heat Exchanger Data

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

Using the pseudo-complete data from Model 1 (step A3), we repeat the same calculations as for Model 0. The graphical displays of the effects in Figures 6 and 7 confirm that only the factor E main effect and interactions EG and EH are important (step A4). Also, note the increased separation between the significant and insignificant effects in Figure 7 as compared with Figure 5. Thus, we terminate the model selection phase.

*** Figures 6 and 7 about here

Table 9 presents the alias structure of the 11 contrasts in terms of the 10 main effects, EG, and EH. If only main effect E and interactions EG and EH are important, then these alias strings can be used to explain why contrasts e, B, K, and C are larger than the remaining six contrasts in Figures 4 and 6; only these four contrasts have coefficients of EG and EH with opposite signs. Based on EG = .002305 and EH = -.001927 from Table 8, (EG+EH)/3 = 0.000126 and (EG−EH)/3 = 0.001410; the difference is more than ten-fold. Moreover, we performed a more extensive search using stepwise regression for all 10 main effects and 45 two-factor interactions on both the Model 0 and Model 1 pseudo-complete data and still obtained this same model. A quick comparison of the MLEs with their corresponding standard errors also confirmed Model 1.

Table 9: Alias Structure for 12 Run 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

In Figure 8, we plot the predicted values versus the observed intervals for the 12 runs; almost all the predicted values fall within the observed intervals, suggesting that the chosen model is reasonable.

*** Figure 8 about here

We consider next the recommendation of levels for factors E, G, and H. Since maximizing lifetime is equivalent to maximizing $y^{(\lambda)}$, then it can be easily seen from Table 8 that $E_1G_1H_2$ yields the largest predicted lifetime (146.18). This is supported by the fact that run 3 in Table 7 is the only run with these same factor levels and indeed has the longest lifetime.

In the original analysis using Taguchi's minute accumulating analysis, only the main effect E was detected. The flexibility of our method makes it possible to easily entertain potentially important interactions which minute accumulating analysis cannot. It led to finding two important interactions, EG and EH. Note that $EH-EG =$ -.004232 and $E =$ -.004252 are nearly the same, so that an incorrect choice of G and H levels could vitiate the effect of the recommended level of E.

6. A Simulation Study

In this section, we study the performance of the quick and dirty (QD) and Hahn-Morgan-Schmee (HMS) methods described in Section 2 with that of the proposed method (HW). Recall from Section 2 that the main-plot analysis of Taguchi's MAA is approximately equivalent to the QD method and therefore was not included in the study. For a specified model, we used simulation to compare (i) how well each method identifies the correct model and (ii) how well it preserves the order of factor importance.

The simulation was performed for a 16 run fractional factorial experiment in six factors (A-F) defined by 5=123 and 6=234. A lognormal regression model was used with $\log y = x\boldsymbol{\beta} + \sigma\epsilon$, ϵ is N(0, 1), and $x\boldsymbol{\beta}$ consists of five real effects (A, B, C, D, AB) with β=(5, 2, 4, 1, -3); the remaining 10 effects are 0. A censoring time of 2 on the log scale was used and 500 simulations were performed for $\sigma = .5$ and 1.

To compare the performance of the three methods, we needed to adopt a formal rule for model selection. For convenience, we used the half-normal plot although it lacks theoretical justification for censored data (see Section 3). As this is a graphical method which requires visual judgement to identify points above the line through the insignificant effects, we used a more formal version for the study by specifying the following cutoff rule based on R^2. First, we assumed that the eight smallest effects were insignificant and fit a line through them noting its R^2 value. Then we added the remaining seven points one at a time, fitting a line each time, and looked for an R^2 drop of 0.1 or more (from the last fitted line). At the first such drop, the currently added effect and the remaining larger effects were identified as significant. If no R^2 drop of 0.1 or more was found, the cutoff was taken to be where the largest drop occurred. Another rule based on prediction was tried but the details are omitted since both rules gave similar results. In order to calibrate the results, we included the results based on no censoring which are denoted by U (for uncensored) in Table 10. That is, the actual failure times observed (before any censoring) at each of the 16 runs were analyzed.

An initial model with six main effects was specified for the HW method, whose results are denoted by HW in Table 10. A variation of the HW method starts with the same model as the HMS method. That is, treat censoring times as actual lifetimes and use the half-normal plot to choose the initial model. Other steps in the HW method remain unchanged. These results denoted by HW* help us understand the factors that contribute to HW's superior performance.

Results for the simulation are given for $\sigma = .5$ and 1. For each method, the first row presents the number of cases (out of 500) for which the k largest effects (k = 1 to 5) on the final plot are correctly ordered according to the true model. The second row gives the number of these cases for which all k effects were declared significant.

A summary of Table 10 follows:

(1) The results for the U method show that the R^2 rule performs quite well. For larger σ, smaller effects are missed as expected.

(2) The methods can be ranked in the following order: HW > HW* >> HMS > QD. The poorer results for these methods (compared with the U method) reflect both the loss of information from censoring as well as deficiences in these methods.

(3) The results for the QD method show that it seriously reverses the order of factor importance and demonstrates the danger of ignoring the censoring information. Also, the QD method detected spurious effects often (results not reported here).

(4) Two factors explain the HMS method's poor performance. First, it starts with the model chosen by the QD method, but this is not its only deficiency, since the HW* variation started with the same model and did much better. The main difference between HMS and HW* is that HW* (and HW) directly uses the censoring information to calculate the MLEs for the current model (in the fitting step A2). Because estimates of wrong effects in the current model tend to be small, the influence of these wrong effects in the next round of imputation is thereby reduced. On the other hand, HMS fails to use the censoring information directly so that its final model is influenced too much by the initial model. The relative disadvantage of the HMS method is lessened for larger σ because larger variation brings about more uncertainty in the imputation and therefore reduces the influence of the initial model.

(5) The results of HW* and HW show that choice of initial model can make a difference. This suggests starting with several different models and checking to see if the same final model is chosen.

(6) Since the main-plot analysis of MAA is approximately equivalent to the QD method, these results suggest that MAA can also perform poorly.

Table 10: Simulation Results for Method Comparison
Number of Cases with the k Largest Effects
Correctly Ordered and Detected Using R^2 Rule

σ	method	k				
		1	2	3	4	5
.5	U	500	500	500	500	500
		500	500	500	500	498
	QD	500	1	1	1	0
		500	1	1	1	0
	HMS	490	12	12	12	0
		490	12	12	12	0
	HW*	500	390	390	390	340
		500	390	390	390	123
	HW	500	497	485	485	485
		500	497	485	485	479
1.	U	499	499	499	499	487
		499	499	499	493	349
	QD	500	45	43	38	0
		500	45	43	38	0
	HMS	492	104	102	100	5
		492	104	102	100	1
	HW*	481	375	368	364	289
		481	375	368	363	85
	HW	487	449	370	361	306
		487	449	370	361	227

7. Discussion

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, Laird and Rubin 1977 and its references), the Schmee-Hahn (1979) algorithm and the data augmentation algorithm (Tanner and Wong 1987a, 1987b). However, in these papers estimation for a specified model rather than model selection was the goal. The proposed procedure obtains pseudo-complete normal data via transformation and imputation and informally analyzes them to identify the important factor effects. Indeed, the procedure has the attractive feature of simplicity which results in computational savings (the examples required at most two iterations) and encourages experimenter involvement in the model selection process. Moreover, it is flexible because it entertains many models simultaneously. The procedure also 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 censored data. We will discuss this shortly, but first we mention some criticisms of using the procedure by itself.

We have already pointed out that the use of standard methods on the pseudo-complete data lacks theoretical justification. Namely, the imputation step ignores the variability of the censored data, which might lead to incorrect choices of important effects. Multiple imputations of the censored data using random or systematic sampling could be tried. If these imputed values all lead to the same final model, the effect of variability of the censored data on the selected model is negligible. Since it is difficult to study these issues theoretically, we resorted to simulation to study the empirical performance of the proposed procedure in selecting appropriate models. As summarized in Section 6, it performed quite well. Furthermore, the model assessment step of the procedure provides another built-in check on the validity of the selected models.

A limitation of the method is its reliance on the existence of the MLEs; it cannot start if the MLEs for the initial (often main-effects) model do not exist. The method can still be used if an initial model whose MLEs do exist can be found. If existence problems are encountered at the beginning for a main-effects model, the data may contain too little information due to very heavy censoring. However, this problem is minimized by our strategy of building the model up rather than starting with a comprehensive model. For the two examples presented here as well as for another real example not reported, we did not encounter these problems. Note that the HMS method is not immune from this problem since the iterative least squares estimates will diverge when the MLEs do not exist.

How does one tell if he has encountered an estimability problem? Hamada and Tse (1989) conclude that because of the complicated structure in industrial experiments, it is generally difficult to tell whether the MLEs exist just by inspecting the pattern of the runs with all censored observations. However, the existence of the MLEs can be checked by a standard linear programming algorithm. Alternatively, an estimability problem may be indicated by the behavior of the iterations of the algorithm which finds the MLEs. Some things to look for are (i) some of the estimates are large and continue to increase in successive iterations, (ii) many iterations are required since the likelihood surface is flat in the neighborhood of the unbounded MLEs and (iii) the information matrix, used in the Newton-Raphson procedure, is not positive definite.

Two approaches which either handle or avoid the estimability problem altogether are being studied. A likelihood approach looks at likelihood ratio statistics for different models. While this presents no theoretical problem, there are some computational problems that need to be overcome. Also, the validity of chi-square approximation when the MLEs do not exist for some components needs to be studied. A Bayesian approach may be taken which eliminates the estimability problem since an appropriate prior leads to finite posterior estimates. The relative importance of factor effects is not expected to be very sensitive to the choice of prior.

The proposed procedure can also be viewed as a quick way to identify a good starting model for more sophisticated procedures. Adding to or deleting effects from the model as in a stepwise procedure ensures that other promising models are not missed. The analysis can also be supplemented by using a Weibull or gamma model which is especially attractive in reliability studies. Hence, the proposed procedure can be used alone or viewed as a complement to these more sophisticated procedures.

We conclude this article by noting that the usefulness of the methods considered here depends on the information contained in the data which is affected by the degree of censoring (Hamada 1989a, 1990 and cited references). For very light censoring where little if any information is lost, even the quick and dirty method works. For more censoring, the quick and dirty method breaks down but the HMS method continues to work. For moderate censoring, the proposed method outperforms the HMS and quick and dirty methods as demonstrated by the simulation. For very heavy censoring where there is little or no information in the data, no method is expected to work. Current research is looking at characterizing this classification in more precise terms.

Acknowledgements

We thank Bill Meeker and Vijay Nair for extensive comments on earlier versions, Bill Meeker for providing some useful software and Conrad Fung for helpful discussions. This paper has benefitted from comments and suggestions from the Editor, Associate Editor, and referees to whom we are grateful. Earlier research was supported by National Science Foundation Grant DMS-8420968 and was aided by access to the University of Wisconsin-Madison Department of Statistics research computer. Subsequent research was supported by General Motors of Canada Limited, the Manufacturing Research Company of Ontario, and the Natural Sciences and Engineering Research Council of Canada.

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, NY: John Wiley and Sons, Inc.

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.

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

Hahn, G. J., Morgan, C. B., and Schmee, J. (1981), "The Analysis of a Fractional Factorial Experiment with Censored Data Using Iterative Least Squares," *Technometrics,* 23, 33-36.

Hamada, M. (1989a), "The Costs of Using Incomplete Response Data for the Exponential Regression Model," *Communications in Statistics -- Theory and Methods,* 18, 1691-1714.

Hamada, M. (1989b), "An Explanation and Criticism of Minute Accumulating Analysis: Taguchi's Method for Censored Data", University of Waterloo Institute for Improvement in Quality and Productivity Research Report No. 89-11.

Hamada, M. (1990), "The Costs of Using Incomplete Exponential Data," *Journal of Statistical Planning and Inference,* to appear.

Hamada, M. and Tse, S. K. (1989), "When Do Estimates Exist from Censored Data in Industrial Experiments?," University of Waterloo Institute for Improvement in Quality and Productivity Research Report No. 89-03.

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. (1982), *Statistical Models and Methods for Lifetime Data,* New York, NY: John Wiley and Sons, Inc.

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

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

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

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. (1987), *System of Experimental Design.* White Plains, NY:

Unipub/Kraus International Publications.

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 Association,* 82, 528-540.

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. (1990), "SEL: A Search Method Based on Orthogonal Arrays," in *Design of Experiments with Applications to Engineering and Physical Sciences* (S. Ghosh, Ed.). New York, NY: Marcel Dekker, Inc.

Figure 1: Half-Normal Plot for Model 0

Pseudo-Complete Router Bit Data

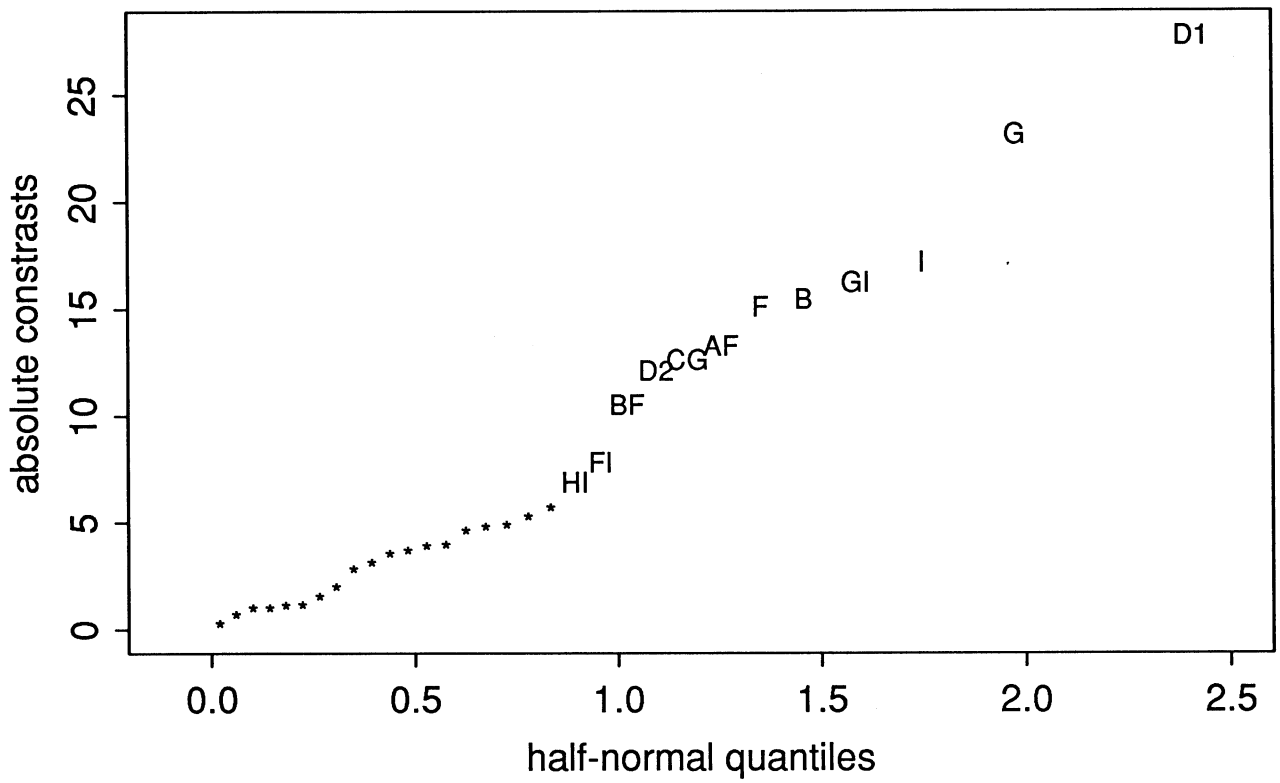
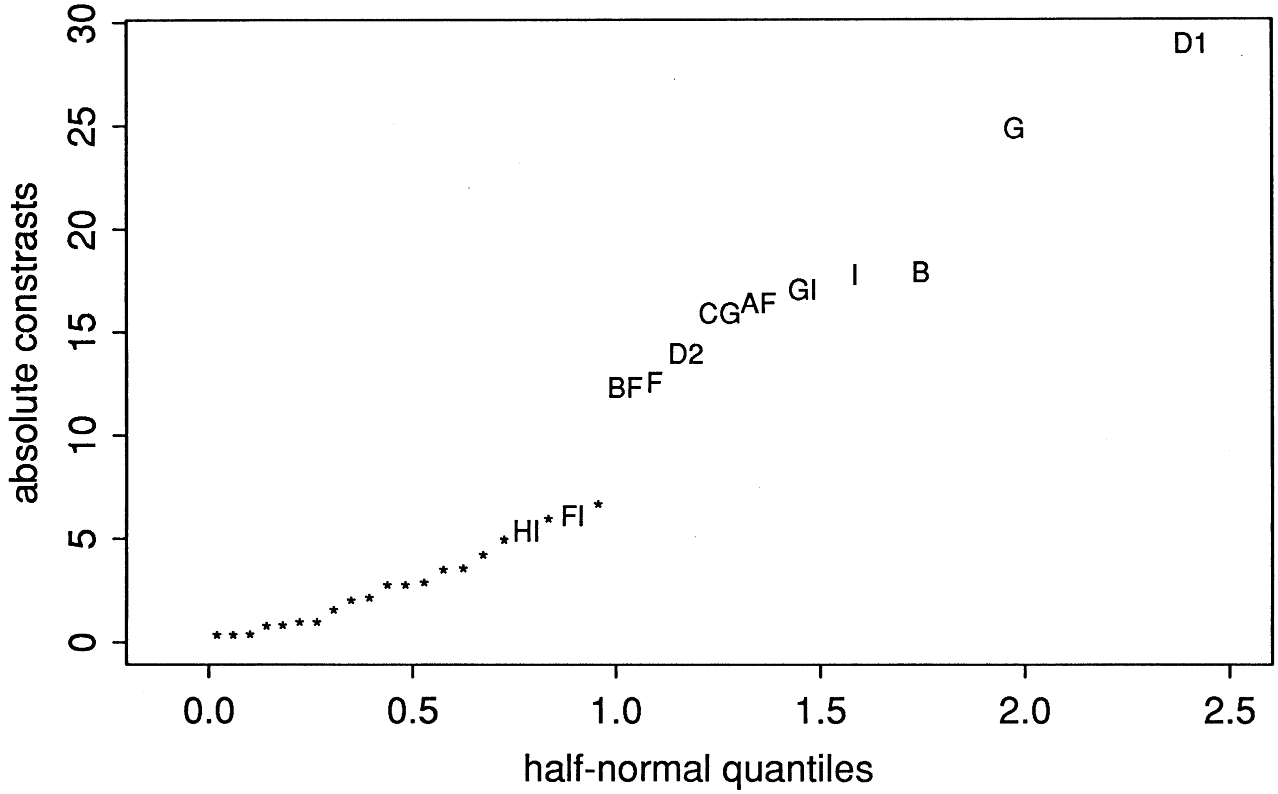


Figure 2: Half-Normal Plot for Model 1

Pseudo-Complete Router Bit Data



Figures 3: Marginal Mean Plots for Model 1

Pseudo-Complete Router Bit Data

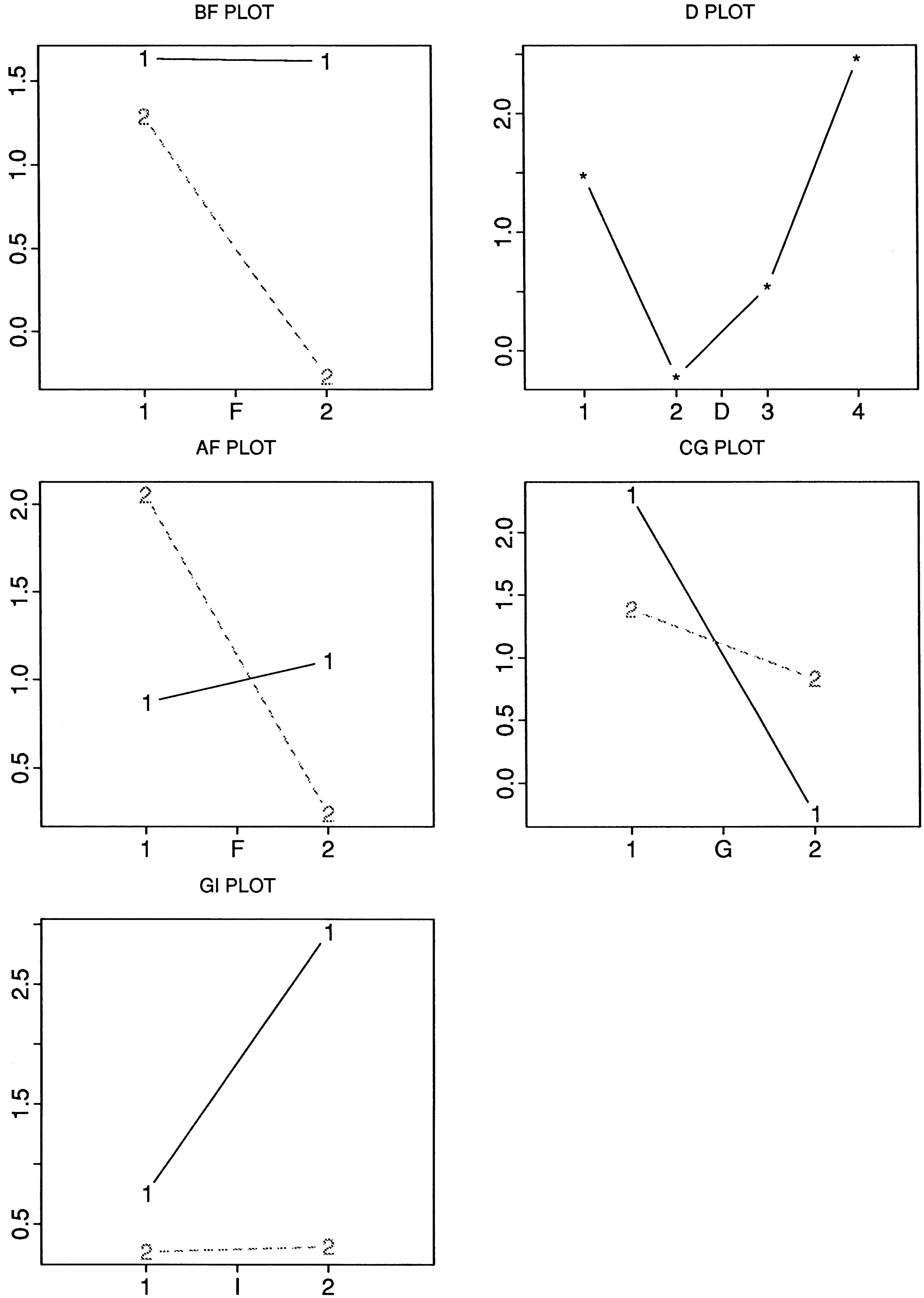


Figure 4: Effect Plot for Model 0

Pseudo-Complete Heat Exchanger Data

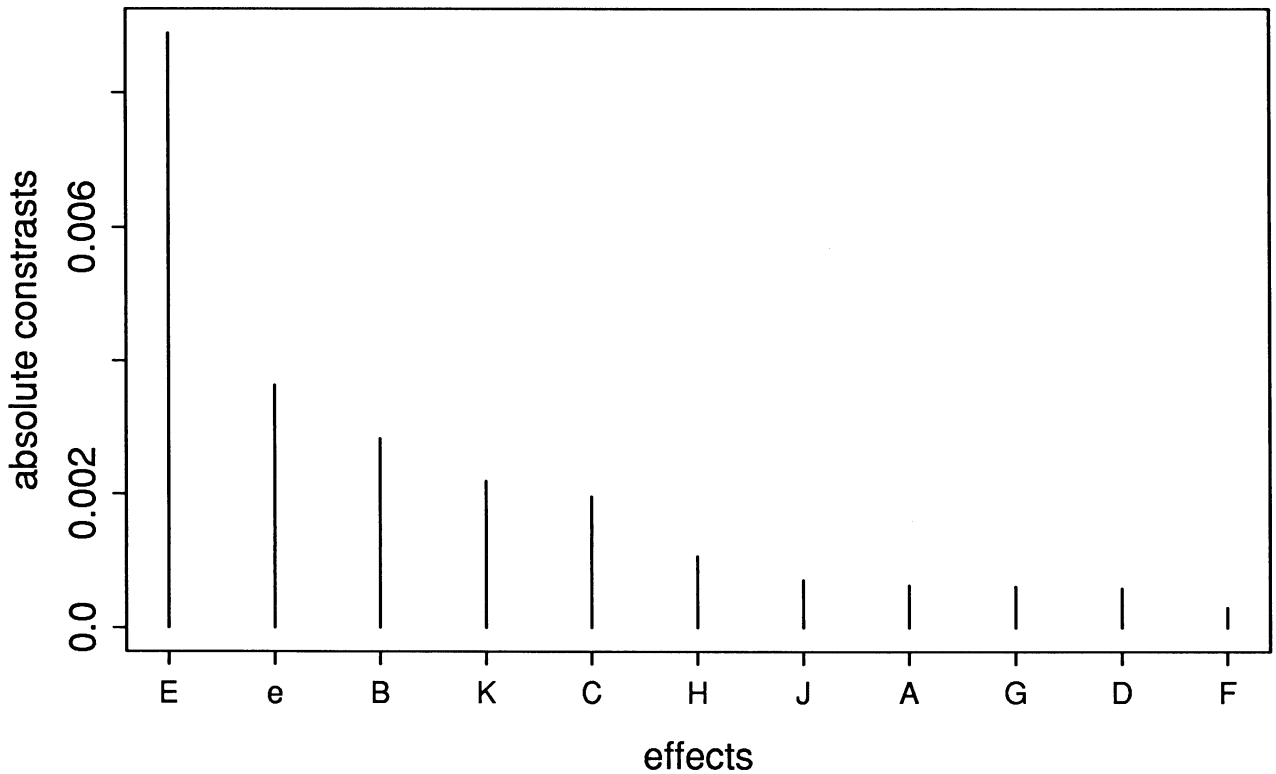


Figure 5: Additional Effect Plot for Model 0

Pseudo-Complete Heat Exchanger Data

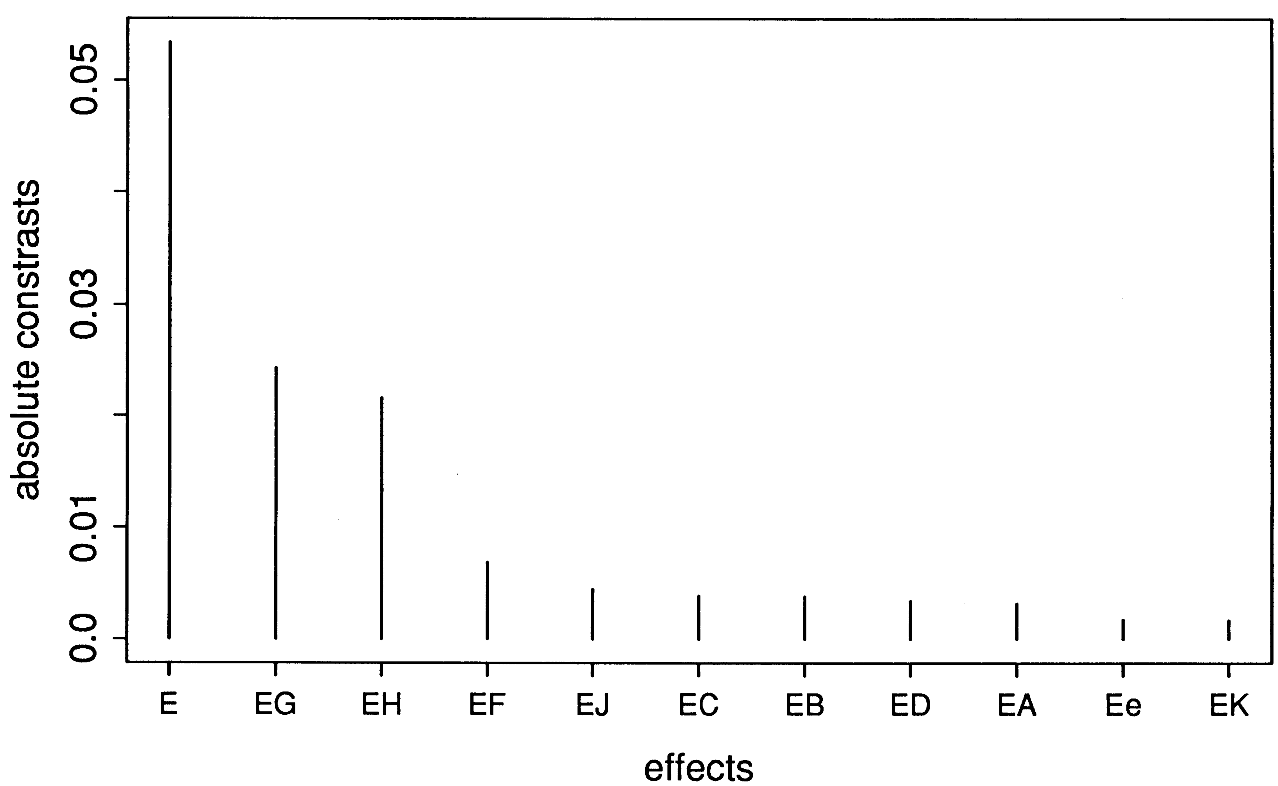


Figure 6: Effect Plot for Model 1

Pseudo-Complete Heat Exchanger Data

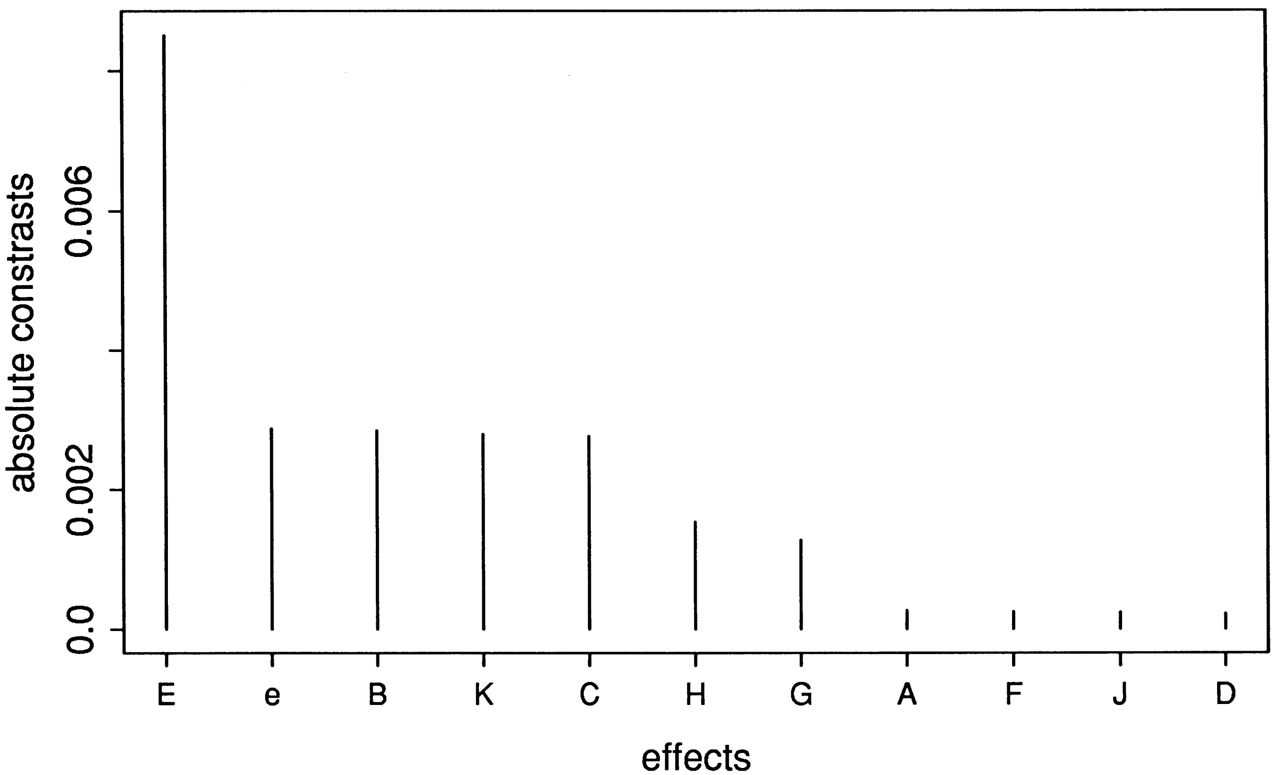


Figure 7: Additional Effect Plot for Model 1

Pseudo-Complete Heat Exchanger Data

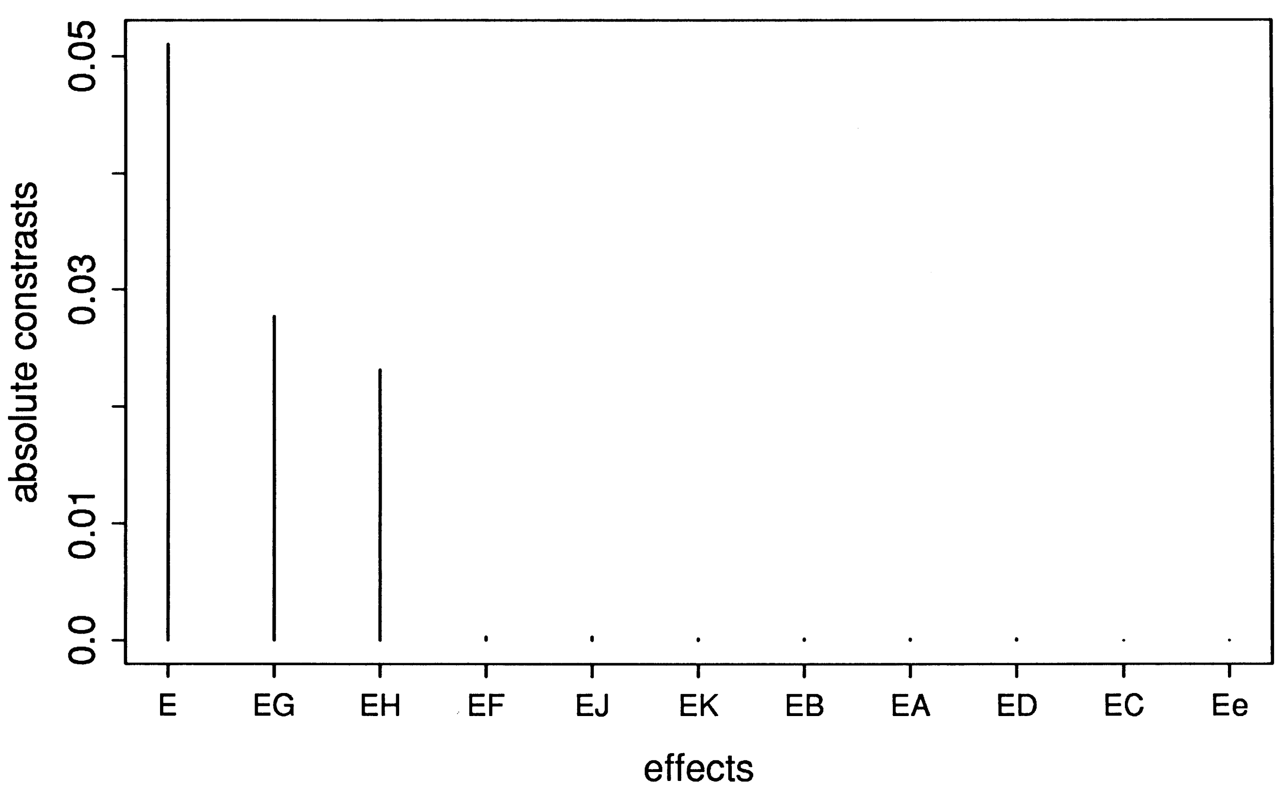


Figure 8 Predicted Responses
for Model 1 Heat Exchanger Data