WHEN DO ESTIMATES EXIST FROM CENSORED DATA IN INDUSTRIAL EXPERIMENTS

M.S. Hamada
S.K. Tse

IIQP Research Report
RR-89-03

March 1989

WHEN DO ESTIMATES EXIST FROM CENSORED DATA IN INDUSTRIAL EXPERIMENTS?

M.S. Hamada

University of Waterloo

S.K. Tse

University of New Hampshire

ABSTRACT

Often interval-censored data are observed in designed experiments for improving reliability because of cost and time constraints. Associated with this type of data are estimability problems because of the paucity of data collected in typical industrial experiments. We review the equivalence of the estimability problem with a linear programming problem and then characterize situations where these estimability problems occur by exploiting the structure in designed experiments.

The main conclusion is that it is difficult to tell just by looking at the data whether the estimates exist or not. Some surprising situations where the estimates exist and some where they do not are presented. In practice, because of the potential danger of using meaningless results, we recommend using a linear programming algorithm to check the estimability conditions. We propose a simple alternative problem that can be solved directly by standard linear programming software. These results apply to popular reliability regression models including the Weibull, lognormal, exponential, and gamma models, as well as a well-known model for binary data.

1 Introduction

In the search to improve quality through experimentation, one critical quality characteristic is a product's reliability. The experimenter's objective is to identify which factors among many are the important ones that affect the product's reliability. Because typical industrial experiments are small, estimability problems ensue unless one entertains models whose number of parameters are constrained by the number of runs (factor level combinations) in the experimental design. Moreover, in lifetesting experiments, censored data are often collected because of cost and time constraints. While data are obtained faster and cheaper, one can encounter estimability problems when fitting even those models mentioned previously. The purpose of this paper is to investigate and characterize these estimability problems for typical industrial experiments.

Different types of censored data arise in lifetesting experiments. Limiting the duration of an experiment yields right-censored observations. If the failure of a unit occurs before the limit, then the lifetime is exactly known; otherwise it is right-censored. Periodic inspection leads to left-censored or interval-censored data, where failures before the first inspection yield left-censored observations. Periodic inspection combined with limited duration yields all three types of censored data. In this paper, we will concentrate on the first scenario, limiting the experiment's duration. Application of the results to the other scenarios will be discussed.

When the data are censored, they can be analyzed by calculating the maximum likelihood estimates (MLE) for some popular reliability models such as the lognormal, Weibull, and exponential regression models. Hamada and Wu (1988) propose a strategy and methodology for analyzing interval-censored data in the industrial setting which also depends on calculating MLEs. In the industrial setting, there can be estimability problems because of the paucity of data. That is, for certain data configurations, the MLEs may not exist for every component of the unknown parameter vector. Necessary and sufficient conditions for the existence of the MLEs for these popular models are given in Silvapulle and Burridge (1986) and Hamada and Tse (1988). They show that the problem of the existence of the MLEs reduces to solving a linear programming problem. For simple linear regression, Hamada and Tse (1988) describe how the linear programming problem can be reduced to checking a few data configurations. Our initial motivation was to study whether the structure in designed experiments could be exploited in characterizing where estimability problems arise.

In practice, the experimenter uses a software package to calculate the MLEs. There can be problems when the stopping rule of the optimization algorithm is based on the increase of the likelihood function in successive iterative steps. Even when the MLEs of some parameters do not exist, there may be no indication of anything going wrong. While some of the estimates should diverge in theory, the stopping criteria may be met first since the likelihood becomes flat as the estimates diverge. Thus, there is a potential danger of making decisions based on meaningless results. Our results suggest that it is hard to tell just by looking at the data whether the MLEs exist or not, except for a few simple designs. However, as the design gets more complicated, it is much harder to characterize the estimability problem.

First, we review the equivalence of the estimability problem with a linear programming problem in Section 2. We discuss the special structure of designed experiments and how it might help to simplify the linear programming problem. In Section 3, we show the enormity of the characterization problem because there can be so many data configurations (i.e., the patterns of censored and complete observations) to consider. For example, there are over 65,000 per model for the L_{16} design. In Section 4, we characterize the estimability problem for the L_4 and L_8 designs. A table is presented giving the number of configurations for which the MLEs do not exist. For the L_8 design, we give a few simple rules which cover all possible main-effect and 2 factor interaction (f.i.) models. For larger designs, the total number of models and possible data configurations becomes unmanageable. We present some rules which include using results from smaller designs and smaller models to reduce the number of data configurations that need to be considered.

In Section 5, we propose a simple alternative linear programming problem which can be solved directly by a standard linear programming algorithm. We recommend its use which can easily be incorporated as a subroutine in existing software. In Section 6, we conclude with a summary and discussion. We also discuss how these results can be applied to a well-known model for binary data. An interesting question is what additional experimentation is needed to guarantee the existence of the MLEs. We observe that one of the reduction rules suggests a simple way to do this. Finally, we discuss the issue of what to do when the MLEs do not exist by commenting on some recent and ongoing work.

2 The Equivalent Linear Programming Problem

In this section, we give some necessary notation and review how the question of the existence of the MLEs for some popular reliability regression models reduces to solving a linear programming problem.

By modeling the log lifetime, the lognormal, Weibull, and exponential regression models fall into the following framework. Consider the model for n observations, $y_i = x_i\beta + \sigma\epsilon_i$ $(1 \leq i \leq n)$, where β, the regression parameters, and x_i, the covariates, are p dimensional vectors. The ϵ_i are independent and identically distributed with known density. For the lognormal model, ϵ is Gaussian whereas for the Weibull and exponential models, ϵ is the standard extreme value distribution. Note that σ equals one for the exponential model. Assume that for $0 \leq r_1 \leq r_2 \leq r \leq n$, the observations y_i are (i) $-\infty = a_i < y_i < b_i < \infty$ for $1 \leq i \leq r_1$ (left-censored), (ii) $-\infty < a_i < y_i < b_i = \infty$ for $r_1 + 1 \leq i \leq r_2$ (right-censored), (iii) $-\infty < a_i < y_i < b_i < \infty$ for $r_2 + 1 \leq i \leq r$ (interval-censored), and (iv) known exactly for $r + 1 \leq i \leq n$.

Provided there is at least one exactly known observation, Silvapulle and Burridge (1986) and Hamada and Tse (1988) show that the necessary and sufficient conditions for the existence of the MLEs for these models are the same. The MLEs exist if and only if there does not exist a non-zero $e \in R^p$ for which: i) $x_i e \leq 0$ for $1 \leq i \leq r_1$; ii) $x_i e \geq 0$ for $r_1 + 1 \leq i \leq r_2$; and iii) $x_i e = 0$ for $r_2 + 1 \leq i \leq n$. Thus, the question of the MLEs' existence reduces to solving a linear programming problem.

For designed experiments, in contrast with the general regression setup, there are a finite number of covariate combinations, the columns of the design matrix are orthogonal, and the entries in the design matrix for the 2 level designs are either -1 or 1. Note that this structure implies that only a finite number of data configurations need to be investigated. Our work was motivated by the question of whether this structure in designed experiments could be exploited to simplify the necessary and sufficient conditions.

Next we set up some notation to describe data configurations for designed experiments. Suppose the design has n runs. Each run is classified as L, R, or EI: classify the run as L(R) if all observations are left-(right-)censored; otherwise, classify it as EI (exactly known or interval-censored). Then, regardless of how many replications are taken, the necessary and sufficient conditions above simplify to: there does not exist a non-zero $e \in R^p$ for which $xe \leq (\geq) 0$ for an L(R) run and $xe = 0$ for an EI run, where x is the appropriate row from the regression design matrix for the model being fitted. Thus, we have a linear programming problem with n constraints in p variables.

Our focus will be on experiments with limited duration so that a run is either classified as R or EI. We will explore the relation between the number of parameters, runs, and EI runs. For instance, it is not always true that existence is guaranteed if the number of EI runs exceeds the number of parameters. In fact, the following examples demonstrate that it is difficult to tell whether the MLEs exist or not just by looking at the number of EI runs.

Example 1: Consider the data from an L_{16} design in 9 factors as shown in Table 1. Note that 12 out of 16 are EI runs and there are 10 parameters (including the intercept). We fit an exponential regression model ($f(y) = \theta exp\{-\theta y\}$, where $\theta = exp\{x\beta\}$) to the data using ISMOD (Lawless and Singhal 1987a, 1987b). The optimizer went through 7 iterations yielding the estimates and standard errors given in Table 2. The ISMOD output did not indicate a problem; in the next section we will see that the MLEs do not exist for this data configuration. As a caveat, one would generally be suspicious of standard errors based on asymptotic variance formulas for such a small sample.

This example shows that although the MLEs do not exist, the optimization program can terminate since the likelihood becomes flat as the estimates diverge. While many iterations of the optimization routine can signal problems, the defaults in a software package may preclude this possibility. Unless the stopping criteria are suitably chosen, our concern is that practitioners are not aware of this estimability problem and can make decisions based on meaningless results. Note that SAS (1985) did indicate a potential problem for this example by noting that the negative hessian used in the LIFEREG procedure was not positive definite.

Example 2: Consider the data configuration from an L_{16} design in 8 factors as displayed in Table 3. Note that only 2 out of 16 are EI runs. An exponential regression model is fitted to the data using ISMOD. Although there are 9 parameters (including the intercept), the MLEs exist.

3 Enormity of the Characterization Problem

In this section we will show the enormity of characterizing situations where the MLEs do not exist. Despite the finite number of data configurations for a given design and model, the number of data configurations to consider can still be enormous. For the 2

Table 1: Data for Example 1

			Design Matrix								
Run	Data	Type	A	B	C	D	E	F	G	H	I
1	2.0	R	1	1	1	1	1	1	1	1	1
2	0.5	EI	1	1	1	-1	1	-1	-1	-1	-1
3	0.6	EI	1	1	-1	1	-1	-1	-1	1	-1
4	2.0	R	1	1	-1	-1	-1	1	1	-1	1
5	0.7	EI	1	-1	1	1	-1	-1	1	-1	-1
6	2.0	R	1	-1	1	-1	-1	1	-1	1	1
7	2.0	R	1	-1	-1	1	1	1	-1	-1	1
8	0.8	EI	1	-1	-1	-1	1	-1	1	1	-1
9	0.9	EI	-1	1	1	1	-1	1	-1	-1	-1
10	1.0	EI	-1	1	1	-1	-1	-1	1	1	1
11	1.2	EI	-1	1	-1	1	1	-1	1	-1	1
12	1.3	EI	-1	1	-1	-1	1	1	-1	1	-1
13	1.4	EI	-1	-1	1	1	1	-1	-1	1	1
14	1.5	EI	-1	-1	1	-1	1	1	1	-1	-1
15	1.6	EI	-1	-1	-1	1	-1	1	1	1	-1
16	1.7	EI	-1	-1	-1	-1	-1	-1	-1	-1	1

Table 2: Estimates, Standard Errors from Example 1

Parameter	Estimate	Standard Error
INT	$-1.95+00$	$3.92+00$
A	$-1.66+00$	$3.92+00$
B	$1.67+00$	$3.06-01$
C	$7.43-02$	$3.06-01$
D	$2.17-02$	$3.06-01$
E	$5.06-03$	$3.06-01$
F	$-1.99+00$	$3.92+00$
G	$-1.94-02$	$3.06-01$
H	$-4.04-03$	$3.06-01$
I	$-1.99+00$	$3.92+00$

Table 3: Data Configuration for Example 2

Run	Type	Design Matrix							
1	EI	1	1	1	1	1	1	1	1
2	R	1	1	1	-1	1	-1	-1	-1
3	R	1	1	-1	1	-1	-1	-1	1
4	R	1	1	-1	-1	-1	1	1	-1
5	R	1	-1	1	1	-1	-1	1	-1
6	R	1	-1	1	-1	-1	1	-1	1
7	R	1	-1	-1	1	1	1	-1	-1
8	R	1	-1	-1	-1	1	-1	1	1
9	R	-1	1	1	1	-1	1	-1	-1
10	R	-1	1	1	-1	-1	-1	1	1
11	R	-1	1	-1	1	1	-1	1	-1
12	R	-1	1	-1	-1	1	1	-1	1
13	R	-1	-1	1	1	1	-1	-1	1
14	R	-1	-1	1	-1	1	1	1	-1
15	R	-1	-1	-1	1	-1	1	1	1
16	EI	-1	-1	-1	-1	-1	-1	-1	-1

level designs in n runs, there are 2^n configurations; for the L_8, L_{12} (Plackett and Burman 1946), and L_{16} designs, there are 256, 4096, and 65,536 data configurations for each model, respectively. So as the run size increases, the number of data configurations increases dramatically. Furthermore, the number of different models increases, so that total number of configurations for larger designs becomes prohibitive. In spite of this large number of data configurations, we can use the structure in the designs to reduce the number of configurations that need to be checked by a linear programming algorithm. Some reduction rules will be given in the next section.

In the following we will make reference to EI(R) sets. The EI(R) set for a particular data configuration contains the run numbers of the EI(R) runs with #EI(#R) denoting the number of runs in the set.

The results of Silvapulle and Burridge (1986) and Hamada and Tse (1988) suggest a geometric approach for verifying the conditions: if there does not exist a p-dimensional hyperplane which passes through all the EI runs, then the MLEs exist. If such a hyperplane exists and all the R runs fall on one side, then the MLEs do not exist; otherwise, the MLEs exist. For example, for the 2^3 design and three factor main effects model, we can represent the runs as corners of a cube. For a given corner, there are three other corners joined to it by edges of the cube. Suppose that EI runs are only at these three adjoining corners. Then, a plane passing through the three corners can easily be visualized. Since one R run will be on one side of the plane and four R runs will be on the other side, the MLEs exist. This idea of finding separating hyperplanes is simple to use for up to three factors, but is impossible to visualize for any more. Thus, this approach is not so useful for the industrial experimental setting where many more factors are investigated.

There are two configurations which are easy to check, however. First, if two runs in the EI set have design matrix rows with opposite signs, then the MLEs exist. Here, it is impossible to have all the R runs on the same side of the hyperplane which passes through the pair of runs. We will refer to this pair as the opposite sign pair. This result has implications for a strategy of additional experimentation which will be discussed in Section 6. The opposite sign pair result explains why the MLEs exist for Example 2. Second, if all the runs in the design with the same level of a factor are in the R set, then the MLEs do not exist. Here, a hyperplane can be fit through the EI set with the entire R set on the same side of the hyperplane. We will refer to this configuration as complete separation.

In the next section, we use the two easily checked configurations and some other rules which we develop to characterize the estimability problem of the L_4 and L_8 designs.

4 Rules for Reduction and Some Results for the L_4, L_8, and Larger Designs

The results for the L_4 design are based on the two easily checked configurations and are presented first. As the run size of the design increases, the complexity of the estimability problem increases requiring the use of reduction rules which are presented in the following section on results for the L_8 design.

Recall that the EI(R) set for a particular data configuration contains the run numbers of the EI(R) runs with #EI(#R) denoting the number of runs in the set. The EI(R) lists are simply lists of EI(R) sets. The notation mEI(mR) denotes lists of m size sets. Finally, let 2_c^{a-b} denote both model and design for a 2 level factors and c 2 f.i. based on a 2^{-b} fraction of a full factorial design.

4.1 L_4 Design Results

The L_4 design matrix is given in Table 4. The 2^2 and $2^{3-1}(=2_1^2)$ designs are obtained by using the first two and three columns, respectively.

The results for 2^2 are: (1) The MLEs exist for all 1R cases. (2) Of the 6 2R cases, 4

Table 4: Design Matrix for L_4

	Design Matrix		
Run	1	2	12
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

have complete separation (MLEs do not exist) and 2 are opposite sign pairs (MLEs exist). (3) All of 3R cases have complete separation. For $2^{3-1}(=2^2_1)$, all runs must be EI for the MLEs to exist.

These results can be summarized as follows: for 2^2, if 12, 13, 24, or 34 are contained in the R set, then the MLEs do not exist; for $2^{3-1}(=2^2_1)$, if the R set is contained in 1234, then the MLEs do not exist.

4.2 L_8 Design Results

The L_8 design matrix is given in Table 5 with columns used for different models displayed in Table 6. Note that for each given model, there are 256 data configurations to consider. In the following, we present some rules which can be used to greatly reduce the number of configurations that need to be checked.

1. For every data configuration, there are isomorphic configurations. That is, two configurations are isomorphic if they can be made the same by renaming and rearranging the design matrix rows and columns. For the L_8 design, some of the models with 2 f.i. are identical or isomorphic to main effects models with a larger number of factors. We need to consider only one additional model; although there are 3 additional models, 2 of these are isomorphic to main effects models. For example, the design matrices of the following designs are identical: $2^3_2 = 2^{5-2}$, $2^3_3 = 2^{6-3}$, $2^{4-1}_3 = 2^{7-4}$,

$2_1^{5-2} = (2_2^{4-1}$ or $2^{6-2})$, and $2_2^{5-2} = 2_1^{6-2} = 2^{7-4}$. The simple summary for the L_8 design below demonstrates two aspects of isomorphism. First, the following designs are isomorphic by interchanging the appropriate rows as given: $2^{5-2} \approx 2_1^{4-1}$ ($5 \leftrightarrow 7$, $6 \leftrightarrow 8$) and $2^{6-3} \approx 2_2^{4-1}$ ($1 \leftrightarrow 2$, $5 \leftrightarrow 6$). Second, the summary for 2^3 gives 6 data patterns which are isomorphic.

2. If the MLEs do not exist for a given R set, then the MLEs do not exist for larger R sets containing it.

3. If the MLEs exist for a given EI set, then the MLEs exist for larger EI sets containing it.

4. If the MLEs exist for a given EI set and model, then the MLEs exist for that EI set and any submodel.

5. If the MLEs do not exist for a given R set and model, then the MLEs do not exist for that R set and any supermodel.

6. Results for a smaller design can be applied to a larger design provided that the larger design is a replicate of the smaller design. A simple example is that complete separation for the L_2 design implies that the MLE's do not exist for complete separation for any larger 2 level design. Another example is that the MLEs' nonexistence for 1R sets for the 2^{3-1} design implies their nonexistence for some 2R sets for the 2_1^{4-1} design. By collapsing one of the factors not involved in the interaction, the larger 8 run design reduces to two replicates of the smaller 4 run design. The 1R result for 2^{3-1} also implies that the MLEs will not exist for some 4R sets for the L_{16} design. This explains why the MLEs do not exist for Example 1 since $16 = 1234 = 9$. Note that knowledge of the design's alias structure is necessary to study this data pattern, so that even this pattern would be hard to look for in practice.

Table 5: Design Matrix for L_8

	Design Matrix						
Run	1	2	3	123	12	13	23
1	1	1	1	1	1	1	1
2	1	1	-1	-1	1	-1	-1
3	1	-1	1	-1	-1	-1	1
4	1	-1	-1	1	-1	1	-1
5	-1	1	1	-1	-1	1	-1
6	-1	1	-1	1	-1	-1	1
7	-1	-1	1	1	1	-1	-1
8	-1	-1	-1	-1	1	1	1

Table 6: Columns Used for L_8 Models

	Columns Used						
Model	1	2	3	12	13	23	123
2^3	x	x	x				
2^{4-1}	x	x	x				x
2^{5-2}	x	x	x	x	x		
2^{6-3}	x	x	x	x	x	x	
2^{7-4}	x	x	x	x	x	x	x
2_1^3	x	x	x	x			
2_1^{4-1}	x	x	x	x			x
2_2^{4-1}	x	x	x	x	x		x

Thus, application of these rules eliminates the need to check every data configuration. We use these rules extensively in obtaining results for the L_8 designs. The number of cases for which the MLEs do not exist for each R set size are presented in Table 7. These numbers may also have a superscript and subscript. The subscript refers to the number of EI sets for which the MLEs exist and do not contain smaller EI sets which insure existence. The superscript refers to the number of R sets for which the MLEs do not exist and do not contain smaller R sets which insure nonexistence. Therefore, if the number does not have a superscript then all these R sets contain smaller R sets for which the MLEs do not exist. Similarly, if the number does not have a subscript, then all these EI sets (complement of the R set) contain smaller EI sets for which the MLEs exist. Hence, if we have these EI and R lists, then Rules 2 and 3 can be used to characterize the existence problem. For example, for the 2^{4-1} design in Table 7, only 4 2EI sets and 16 4R sets are required to characterize all 256 data configurations. Furthermore, for a given #R, Rule 5 can be used as we move across the table from left to right. For example, 6 of the 16 4R cases for the 2^{4-1} design we already know from the 2^3 design results. Similarly, Rule 4 can be used as we move across the table from right to left. Note that the 2^{4-1} design is not contained in the 2^{5-2} and 2^{6-3} designs so that Rule 5 cannot be used to apply the 2^{4-1} design results to the latter designs. Similar restrictions apply to Rule 4.

Next, we give a simple summary for all L_8 models. This summary can be implemented in a small computer program, thus eliminating the need for linear programming software. However, as the run size of the design increases, a simple summary is no longer possible as indicated in the next section.

- For 2^3, if the R set contains 1234, 5678, 1256, 3478, 1357, or 2468, then the MLEs do not exist.
- For 2^{4-1}, if the R set is contained in 123678, 234567, 134568, or 124578, then the MLEs exist.

Table 7: Characterization of MLEs' Non-Existence for L_8

		Design							
#R	Total	2^3	2^{4-1}	2^{5-2}	2^{6-3}	2^{7-4}	2_1^3	2_1^{4-1}	2_2^{4-1}
1	8	0	0	0	0	8^8	0	0	0
2	28	0	0	8^8	16^{16}	28	4^4	8^8	16^{16}
3	56	0	0	40	48	56	24	40	48
4	70	6^6	16^{16}	66_4	68_2	70	56_{14}	66_4	68_2
5	56	24	32	56	56	56	56	56	56
6	28	24_4	24_4	28	28	28	28	28	28
7	8	8	8	8	8	8	8	8	8

- For 2^{5-2}, if the R set is contained in 1467, 2358, 1368, or 2457, then the MLEs exist.
- For 2^{6-3}, if the R set is contained in 1467 or 2358, then the MLEs exist.
- For 2^{7-4}, if the R set contained in 12345678, then the MLEs do not exist.
- For 2_1^3, if the R set contains 12, 34, 56, or 78 or complete separation (1234, 5678, 1256, 3478, 1357, 2468), then the MLEs do not exist.
- For 2_1^{4-1}, if the R set is contained in 1458, 1368, 2457, or 2367, then the MLEs exist.
- For 2_2^{4-1}, if the R set is contained in 1368 or 2457, then the MLE s exist.

Finally, we make some comments about the results presented in Table 7. First, note the dramatic increase in the MLEs' nonexistence in moving from 2^{4-1} to 2^{5-2}. Also notice the following surprising cases where the MLEs exist for large size R sets and do not exist for small size R sets: for the former where $p \geq \#EI$, see (4R, $L_8(2^{5-2})$), (6R, $L_8(2^3)$, $L_8(2^{4-1})$); for the latter where $p \leq \#EI$, see (2R, $L_8(2^{5-2})$).

4.3 Larger Designs Results

For larger designs, we can use the rules developed in the previous section to reduce the number of data configurations that need to be studied. However, the complexity quickly increases so that an exhaustive study becomes prohibitive. We studied the L_{12} (Plackett and Burman 1946) and L_{16} designs. The results for these designs exhibit similar patterns as in the L_8 design and thus are not presented here. There are also more surprising cases where the MLEs exist for large size R sets and do not exist for small size R sets. In studying the L_{12} design, we cannot use the results from smaller designs since the L_{12} design is not a replicate of them. While results from the L_4 and L_8 designs can be used in studying the L_{16} design, many configurations not covered by these results have to be investigated. This increase in complexity makes a complete study of larger designs such as the L_{32} and L_{64} designs prohibitive.

While such information as displayed in Table 7 conveys the potential estimability problems one is faced with, it does not answer the question of whether the MLEs exist for a particular data set. In the next section, we propose a simple alternative linear programming problem that can be solved directly by standard linear programming software so that the existence of the MLEs for a particular data configuration can be checked.

5 A Simple Alternative Linear Programming Problem

The main conclusion from Section 4 is that it is not easy to tell from looking at the pattern of complete and censored observations whether the MLEs exist or not. Therefore, we recommend using a linear programming algorithm to verify the MLEs' existence. This could be easily added as a front end to the optimization program which calculates the MLEs. Because standard linear programming algorithms cannot handle this linear programming problem directly, we propose a simple alternative linear programming problem. Silvapulle and Burridge (1986) propose an alternative problem based on reducing the size of the problem, but give up the simplicity of using a standard linear programming algorithm directly. For the industrial context, reducing the problem is unnecessary since the run size for designed experiments is typically small. As was explained earlier, even if several replicates are taken, the problem is still solved in terms of the runs. For the alternative problem we will propose, the worst problem for a L_{32} design would be 63 constraints in 63 variables which is still a small problem by linear programming standards.

The linear programming problem as stated cannot be solved directly by standard linear programming algorithms since the zero vector is always a solution. We propose the following simple restatement of the linear programming problem. Suppose n is the design run size and #R is the number of R runs. We change each inequality associated with a R run into an equality by adding a new slack variable s. Since this new variable has to be negative, we introduce the corresponding constraint, $s \leq 0$. Thus, the alternative linear programming problem is to minimize the sum of all the new slack variables given all the constraints consisting of n equalities plus #R new inequalities (due to the new negative slack variables). There are two solutions to this alternative problem: either the zero vector is the only solution or the problem is unbounded from below. The MLEs exist for the former case, but not for the latter case. This alternative problem can be handled directly by a Phase 1-Phase 2 algorithm (Best and Ritter 1985).

If one is not convinced that a linear programming algorithm is needed to check for estimability for the scenario of limited experimental duration, then one should be convinced for the more general case of interval-censored data. To give an idea of the added complexity when left-censored data can also be observed, for one replicate of a L_{16} design, there are now 3^{16} configurations; for m replicates, there are $(3^{16})^m$ configurations. However, these cases can still be handled by the linear programming algorithm discussed above. In particular, when there are left-censored data and provided that there is at least one exactly known observation, the following simple changes are needed in the alternative linear programming problem. For $xe \leq 0$ associated with left-censored data, add a new positive slack variable to make it an equality. The corresponding inequality for the new slack variable is $-s \leq 0$ and to the objective function, add the negative of the slack variable. Note that the condition for left-censored data means that one is searching for a hyperplane which puts all the L runs on one side and all the R runs on the other. When none of the observations are exactly known, the conditions are more complicated for the lognormal, Weibull, and gamma regression models, but still can easily be put in this form. See Silvapulle and Burridge (1986) for these conditions.

6 Summary and Discussion

This paper has demonstrated that potential estimability problems that one encounters when fitting models to censored data from industrial experiments can be extensive. The main conclusion is that it is difficult to tell just by looking at the data pattern whether the estimates exist or not. We are convinced of this by some surprising cases where the MLEs exist for large R sets but do not exist for small R sets. While the estimability problem can be summarized by a few rules for the L_4 and L_8 designs, for larger designs the number of models and data configurations that need to be studied becomes prohibitive. This is in spite of the use of several reduction rules that were given. Therefore, for a particular data configuration we recommend using a linear programming algorithm to check if there is an estimability problem. We proposed a simple alternative problem which can be solved directly by standard software.

These results apply to the exponential regression model and to the normal, Weibull, gamma, lognormal, and loggamma regression models provided there is at least one exactly known run. These results also apply to the logistic regression model for analyzing binary data. The estimability conditions are the same as that given in Section 2 (Silvapulle 1984).

Since 0s or 1s are observed for binary data, classify a run as L(R) if all the data are 0s(1s); otherwise classify as EI. Note that all the results in Section 4 apply directly provided that there are no L runs (runs with all 0s). This is a realistic situation in practice since one is interested in finding conditions where 100% good parts are produced. That is, it is unlikely for management to allow the process to be run at combinations yielding 100% bad parts.

The results also show that larger designs, while more costly, provide better protection against estimability problems. What can be done when the MLEs do not exist? An interesting question is how many additional runs are needed to guarantee the MLEs' existence. The results for the opposite sign pair suggest a simple strategy. If we perform one additional run so that there are a pair of EI runs with opposite signs, then the MLEs exist. Thus, for one EI run, run the opposite combination until EI data are observed.

Finally, we discuss what to do when the MLEs do not exist or more generally how to analyze censored data from industrial experiments by commenting on some recent and ongoing work. Hamada and Wu (1988) propose an iterative scheme of model fitting, imputation of censored observations, and model selection. While their procedure does depend on the existence of MLEs, most estimability problems are avoided by building up the model rather than starting with a comprehensive model. Another approach is using the likelihood ratio based procedure of Lawless and Singhal (1978) which looks for factors with significant likelihood drops. While the nonexistence of MLEs does not present a theoretical impediment in this approach, it does cause a computational one. The original motivation of their work was to provide a flexible methodology for analyzing medical data where estimability problems are a rarity. Current work focuses on how to apply this approach to the industrial setting.

7 Acknowledgements

We thank M. Best for helpful discussions and J. Lawless, R. J. MacKay, and C. F. J. Wu for insightful comments on a earlier version. M. Hamada was supported by a research grant from the Manufacturing Research Corporation of Ontario and a grant from the National Science and Engineering Research Council.

References

[1] Best, M. J., Ritter, K. (1985), *Linear Programming Active Set Analysis and Computer Programs*, Englewood Cliffs: Prentice-Hall Inc.

[2] Hamada, M. and Tse, S. K. (1988), "A Note on the Existence of Maximum Likelihood Estimates in Linear Regression Models Using Interval-Censored Data," *Journal of the Royal Statistical Society*, Ser. B, 50, 293–296.

[3] Hamada, M. and Wu, C. F. J. (1988), "Analysis of Incomplete Data from Highly Fractionated Experiments," Report No. 29, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

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

[5] Lawless, J. F. and Singhal, K. (1987a), "ISMOD: an All-Subsets Regression Program for Generalized Linear Model. I. Statistical and Computation Background ," *Computer Methods and Programs in Biomedicine*, 24, 117–124.

[6] Lawless, J. F. and Singhal, K. (1987b), "ISMOD: an All-Subsets Regression Program for Generalized Linear Model. II. Program Guide and Examples," *Computer Methods and Programs in Biomedicine*, 24, 125–134.

[7] Plackett, R. L. and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments ," *Biometrika*, 33, 305–325.

[8] SAS (1985), *SAS User's Guide: Statistics, Version 5 Edition*, Cary: SAS Institute Inc.

[9] Silvapulle, M. (1981), "On the Existence of Maximum Likelihood Estimators for the Binomial Response Model," *Journal of the Royal Statistical Society*, Ser. B, 43, 310–313.

[10] Silvapulle, M. J. and Burridge, J. (1984), "Existence of Maximum Likelihood Estimates in Regression Models for Grouped and Ungrouped Data," *Journal of the Royal Statistical Society*, Ser. B, 48, 100–106.