8-2005

Design, Analysis, and Applications of Failure Amplification Experiments

Oksoun Yee

University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Oksoun Yee entitled "Design, Analysis, and Applications of Failure Amplification Experiments." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Business Administration.

Robert W. Mee, Major Professor

We have read this dissertation and recommend its acceptance:

Mary G. Leitnaker, Ramon V. Leon, Russell L. Zaretzki, Melissa R. Bowers

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation by Oksoun Yee entitled "Design, Analysis, and Applications of Failure Amplification Experiments." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Business Administration.

Robert W. Mee
Major Professor

We have read this dissertation and
recommend its acceptance:

Mary G. Leitnaker

Ramon V. Leon

Russell L. Zaretzki

Melissa R. Bowers

Accepted for the Council:

Anne Mayhew
Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

DESIGN, ANALYSIS, AND APPLICATIONS
OF FAILURE AMPLIFICATION EXPERIMENTS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Oksoun Yee
August 2005

Dedication

This dissertation is dedicated to my parents, Yee Sanir and Lim Sunok, for their unending love and support.

Acknowledgements

This work would not have been possible without the support of many people. Many thanks to my advisor, Dr. Robert Mee, for his guidance and support. I feel privileged to have been a student of such a great teacher.

I would also like to thank the other members of my dissertation committee. Thanks to Dr. Mary Leitnaker for making it possible for me to work with the Huhtamaki company and for her great advice. I want to thank Dr. Ramon Leon for introducing me to the exciting world of statistical applications in industry through my internships at the General Electric Company and his support. I would also like to thank Dr. Russell Zaretzki and Dr. Melissa Bowers for their help and work as committee members.

Thanks to Mark Bond and Mark Pettigrew from the Huhtamaki company for providing me an opportunity of consulting experience and their collaboration.

Abstract

The main focus of this study is related to the Failure Amplification Method (FAMe) proposed by Joseph and Wu (2004). They suggested the use of an "amplification factor" to increase the information from experiments with a binary response variable. In addition to the amplification factor having a known effect, Joseph and Wu recommended that, for convenience of experimentation, this factor be taken as an easy to change, split unit factor. In such cases, the analysis ought to take into account the possibility of both whole unit and split unit error variation. I present such an analysis here, where the Bayesian approach not only permits proper accounting of the error structure, but also facilitates the subsequent optimization step.

FAMe can also be extended to categorical data with more than two categories. I helped design an experiment that was conducted at Huhtamaki Consumer Packaging West Inc., Los Angeles, CA, where the response variable was an ordinal variable characterizing the quality of the Tri Web Taco Bell Disk seal. An amplification factor – speed of the production line - was a whole-unit factor that was hard to change. Therefore an application of FAMe to ordinal data is presented here as well.

It is crucial to plan an experiment carefully, particularly with categorical responses. Levels of the split-unit factor can be chosen sequentially or set in advance. In the case of the sequential design, a rule for choosing a split-unit factor level will affect consistency and bias of the parameter estimates. Theory-based sequential rules often are impractical in real life situations. Properties of sequential ad hoc designs are studied and compared to fixed designs using complete enumeration and simulation techniques.

Key words: *Binary Response, Ordinal Response, Generalized Linear Model, Mixed Model, Model Selection, Split Unit, Sequential Design, Optimization.*

Table of Contents

List of Tables

List of Figures

CHAPTER I

Introduction

When an experimenter wishes to study the effect of certain factors on a response where only attribute data are available, traditional methods of experimentation can be inefficient. Consider an experiment with a paper feeder where the probability of two types of defects - misfeeds and multifeeds - is very small. Very large sample sizes are needed in order to discriminate between the small differences in probability of defects and to find the corresponding optimum factor levels. A novel engineering-statistical approach - failure amplification method (FAMe) - was recently articulated by Joseph and Wu (2004). It allows parameter estimation and optimization with categorical responses in situations where some amplification of defects is available. The basic idea of FAMe, as stated by Joseph and Wu, is "to select a factor with a known effect on the response and use it to amplify the failure probability so as to maximize the information in the experiment."

FAMe was motivated by the operating window (OW) method proposed by Don Clausing (1994, 2004). This method is designed to assist engineers at the development stage of the process in order to improve reliability and robustness of engineering systems. The operating window metric is a range of threshold values of the operating window factor. A threshold value setting is associated with a certain probability of failure. These probabilities are set in advance and depend on an application. For the sake of simplicity, we will use 50% threshold values henceforth. The factor stack force of a paper feeder has a known effect on both misfeeds and multifeeds - small stack force will increase the probability of misfeed, and large force will induce multifeeds. Therefore stack force is chosen as an operating window factor. Other control factors such as wrap angle, belt tension and width of feed belt might affect the failure rates. At each control factor combination the threshold values of the stack force, l and u, are determined, where l (u) is the force at which there is a 50% chance of a misfeed (multifeed). Joseph and Wu (2002) provide a comprehensive modeling and optimization strategy for the OW method. At each control factor combination a signal-to-noise ratio (or some kind of performance measure) is calculated and used for further analysis. The main drawback of this approach is the loss of information when estimating the upper and lower threshold values. Joseph and Wu (2002) emphasized the need of incorporating complete data for a more informative analysis. FAMe is an extension of the operation window method that utilizes the raw data. In FAMe, the OW factor is called an amplification factor.

Another field that is closely related to FAMe is accelerated life testing (ALT). The main difference between the two methods is the response - FAMe handles categorical data, and ALT - continuous. Also FAMe is less applicable to time-related failures because it deals with sudden failures, failures that occur instantaneously at time zero. On the contrary, in the ALT experiments the response is observed over some period of time or a similar characteristic. Both FAMe and ALT require acceleration or amplification in order to induce failures, and an adequate model for extrapolation. A description of ALT and issues associated with it can be found in Nelson (1982, 1990), Meeker and Escobar (1993, 1998, 2004) and Meeker and Hamada (1995).

Recently there have been some advances in the use of design of experiments for reliability applications: Condra (2001), Hamada (1995a&b), Hamada and Wu (1995). In this literature, the response is mostly time to failure with either exact failure times or censored data, or amount of degradation. With censored data and fractional factorial designs, unique maximum likelihood estimates might not exist. Bayesian estimation provides an attractive alternative in such situations. The same problem of estimability of parameters may be encountered with a categorical response. Chipman and Hamada (1996) show that the difficulties of infinite maximum likelihood estimates can be avoided with Bayesian estimation. Another important advantage is that Bayesian estimates are more informative since they account for uncertainty of the parameter estimates in the model and facilitate obtaining meaningful summaries of the quantities of interest. This property can be utilized in the optimization step by checking sensitivity of the quantity of interest (e.g., expected loss) to the optimal factor levels.

Joseph and Wu (2004) based their choice of an amplification factor in part on the ease of experimentation with this factor. In their implementation of FAMe experimentation, an investigation of different levels of the amplification factor was performed at each combination of the control factors to be studied. That is, to facilitate data collection, the amplification variable was always taken as a split unit factor. However, Joseph and Wu (2004) failed to take the whole unit / split unit error structure into account in their analysis. I provide an alternative analysis to properly account for both whole-unit and split-unit variance components.

Generalized linear models (GLM) are widely used for analysis of binary data (McCullagh & Nelder 1989). With the presence of the random effects, traditional likelihood methods for fitting generalized linear mixed models (GLMM) pose a computational challenge in high-dimensional problems because of the numerical integration. There exist computational methods that overcome this problem, such as a Gibbs sampling approach (Zeger and Karim 1991), penalized quasilikelihood (Breslow and Clayton 1993), empirical Bayes (Stiratelli et al 1984), etc. Recently SAS Inc. released a production version of GLIMMIX that allows fitting GLMM's. The free Bayesian estimation software WinBUGS (Lunn et al, 2000) also allows fitting GLMM's and is available from www.mrc-bsu.cam.ac.uk/bugs. MLn and Stata can also be used for fitting GLMM's.

FAMe is a comprehensive approach to categorical response optimization. It consists of several stages: design, modeling, analysis, and optimization. A method of analysis and optimization of the FAMe experiments is illustrated in the next chapter. This is demonstrated on two examples from Joseph and Wu (2004). We will see how design issues affect the model fit and inference about the parameters. These issues are addressed in Chapter III. An extension of FAMe to ordinal data is presented in Chapter IV for an experiment that I helped plan and conduct.

CHAPTER II

Analysis of Split Unit Failure Amplification Experiments

A Brief Overview of Generalized Linear Mixed Models for Binary Data

In this chapter analysis and optimization for the printed circuit board and paper feeder examples from Joseph and Wu (JW, 2004) is presented. Both examples have a binary response with a split-unit structure of the designs. Therefore a brief overview of GLMM is needed.

A mixed model refers to a model with both fixed and random effects. Among other applications, such models are useful to describe data from experiments with restrictions on randomization, for example, randomized block and split-unit designs. Observations sharing the same experimental unit share a common value of a random effect and so are positively correlated.

Assume that the response vector Y follows a *Binomial(n, p)* distribution. A general form of a conditional GLMM with one random factor can defined as follows (McCulloch and Searle, 2001, Chapter 8):

$$\begin{aligned} E(Y/n \mid u) &= \mu \\ g(\mu) &= \alpha + \boldsymbol{X}\beta + \boldsymbol{Z}u\,, \end{aligned} \tag{1}$$

where μ is a conditional mean of Y/n, $g(\cdot)$ is a link function, α is the intercept, β is a vector of fixed effects and u is a vector of identically distributed random effects with variance σ_u^2. $\boldsymbol{X}$ and $\boldsymbol{Z}$ correspond to the model matrices of the fixed and random effects, respectively. In split-unit experimentation, σ_u^2 corresponds to variation due to whole units.

The most common link functions are the logit, probit and complementary log-log functions:

- $logit(\mu) = \log\dfrac{\mu}{1-\mu}$

- $probit(\mu) = \Phi^{-1}(\mu)$, where Φ is the standard normal cumulative density function
- $cloglog(\mu) = \log(-\log(1-\mu))$.

These link functions provide similar fit when $0.1 \le \mu \le 0.9$; the primary differences are in the tails of the distributions.

The conditional model in (1) can be fit either in SAS with Proc GLIMMIX procedure or with the Bayesian software WinBUGS. By default, GLIMMIX estimation is based on pseudo-likelihood techniques - see Wolfinger and O'Connell (1993) and Breslow and Clayton (1993). The GLIMMIX procedure can only fit models with normal random effects, i.e. $u \sim N(0,\sigma_u^2)$. WinBUGS estimates the parameters by applying a Monte Carlo method, the Gibbs sampler (Zeger and Karim 1991). The Bayesian method of estimation does not have a restriction on the distribution of random effects. However, convergence properties are better with conjugate priors. For the binomial data, beta and normal distributions belong to the class of conjugate priors.

Another form of a GLMM is a marginal, or unconditional model. A detailed description and differences between the conditional and marginal specification of the model can be found in McCulloch and Searle (2001, Chapter 8) and Dobson (2001, Chapter 11). In the marginal model the expected value of μ is computed by integrating with respect to the probability distribution for the random effects u:

$$E_u(\mu) = E_u[g^{-1}(\alpha + \boldsymbol{X}\beta + \boldsymbol{Z}u)].$$

Estimation of the unconditional GLMM can be done via GLIMMIX as well as GENMOD procedures in SAS. One of the available methods for fitting the marginal model is generalized estimating equation (GEE) method of Liang and Zeger (1986).

Further discussion of conditional and marginal GLMM's is given in Robinson et al (2004). They denote the conditional model as a batch-specific model, or random-effects GLMM, and the marginal model as a population-averaged model, or covariance-pattern GLMM. They give an example of a split-unit industrial experiment from film manufacturing and illustrate the implications of fitting conditional and unconditional GLMM. The main difference between the conditional and marginal specifications is that the former models random effects together with the fixed effects and the latter models only fixed effects and specifies a covariance matrix for the response.

Printed Circuit Board Example

A detailed description of the printed circuit board (PCB) example can be found in Maruthi and Joseph (1999) and JW(2004). There are two types of conflicting defects in the circuits – opens and shorts. One candidate for the amplification factor was exposure energy, with high levels leading to shorts and low levels – to opens. However, it was inconvenient to use exposure energy as an amplification factor due to budget constraints, a slow measurement process and production issues. Therefore it was decided to use line width (C_1) and spacing between a pair of conductors (C_2) as amplification factors. The levels for both C_1 and C_2 were 3, 4, 5, 6, 7 mil, where $C_2 = 10 - C_1$. Five pairs of conductors with the levels of C_1 and C_2 used in the experiment are shown in Figure 1. (All figures and tables are located in the Appendix.) The opportunity for opens increases from left to right due to the decreasing line width, whereas the opportunity for shorts decreases due to the wider distance between the pairs. The specifications for the levels of these factors are not in the control of a manufacturer and are dictated by the customer; hence they characterize complexity of PCB production. JW (2004) label such applications as the 'complexity factor' amplification method. In the normal production only 5, 6 and 7 mil were used for both line width and spacing.

In addition to complexity factors C_1 and C_2, there were eight control factors. These are listed in Table 1 with their levels. An 18-run orthogonal array (L_{18}) was used as the whole unit design for the eight control factors, crossed with a 5^{2-1} design for the split unit factors (C_1, C_2), with $C_2 = 10 - C_1$. At each treatment combination of control factors $\boldsymbol{X}$ and (C_1, C_2), the number of defects was recorded from a test pattern of 160 single conductors, or 80 pairs. We may assume that the data on opens and shorts follow *Binomial* (160, $p_1(X, C_1)$) and *Binomial* (80, $p_2(X, C_2)$), respectively. The data and L_{18} are given in Table 2.

Joseph and Wu (2004) fit the following fixed effects models to the PCB data using forward variable selection based on the Akaike information criterion (AIC):

$$\log\log\frac{1}{1-p_1} = 10.72 - .73x_{5l} - .33x_{2l} - .27x_{1l}x_{5q} - 2.768\log x_6 - 5.06\log C_1,$$

$$\log\log\frac{1}{1-p_2} = -6.66 + .48x_{1l} + .20x_{4l} - .15x_{1l}x_{5q} + 4.70\log x_6 - 7.664\log C_2,$$

where the p_1 and p_2 correspond to probabilities of failure for opens and shorts, respectively. The linear contrasts x_{il} are coded as (-1, 0, 1) and (-1, 1) for three-level and two-level factors, respectively. The quadratic contrast x_{iq} is coded as (1, -2, 1) for the three-level factor. The levels of x_6 are on the original scale - 14, 17 and 20. Note that their models are not hierarchical and they include up to third-order effects. The exposure energy and complexity factors are on the log scale. The reason for transforming amplification and complexity factors is the following. JW (2004) assume that the number of opens in a conductor and the number of shorts between a pair of conductors follow the Poisson distribution with means $\lambda_1(\boldsymbol{X}, C_1)$ and $\lambda_2(\boldsymbol{X}, C_2)$. For a Poisson random variable with mean λ, a probability that the number of defects is greater than zero is 1-exp(-λ). Then $p_1(\boldsymbol{X}, C_1) = 1 - \exp(-\lambda_1(\boldsymbol{X}, C_1))$, $p_2(\boldsymbol{X}, C_2) = 1 - \exp(-\lambda_2(\boldsymbol{X}, C_2))$, where JW assume the following models for the Poisson means:

$$\lambda_1(\boldsymbol{X}, C_1) = \frac{\lambda_1(x_1, x_2, x_3, x_4, x_5, x_7, x_8)}{C_1^{\alpha_1} x_6^{\gamma_1}},$$

$$\lambda_2(\boldsymbol{X}, C_2) = \frac{\lambda_2(x_1, x_2, x_3, x_4, x_5, x_7, x_8)\, x_6^{\gamma_2}}{C_1^{\alpha_2}},$$

where α_1, α_2, γ_1 and γ_2 are some positive constants. From the formulas above it can be derived that the probabilities of defects p_1 and p_2 follow a *cloglog* link:

$$\log\log\frac{1}{1-p_1} = \log\lambda_1(x_1, x_2, x_3, x_4, x_5, x_7, x_8) - \gamma_1\log(x_6) - \alpha_1\log C_1,$$

$$\log\log\frac{1}{1-p_2} = \log\lambda_2(x_1, x_2, x_3, x_4, x_5, x_7, x_8) + \gamma_2\log(x_6) - \alpha_2\log C_2.$$

JW factor out exposure factor x_6 in order to emphasize its reverse effect on opens and shorts and for optimization purposes. The loss is assumed to be proportional to the expected number of defects:

$$L = \omega_1 \lambda_1(\boldsymbol{X}, C_1) + \omega_2 \lambda_2(\boldsymbol{X}, C_2).$$

Next they minimize the expected loss, where expectation is taken over production levels of the complexity factors – 5, 6 and 7 mil. Thus the expected loss is

$$\begin{aligned} EL &= \omega_1 E_{C_1}\left[\lambda_1(\boldsymbol{X}, C_1)\right] + \omega_2 E_{C_2}\left[\lambda_2(\boldsymbol{X}, C_2)\right] \\ &= \frac{\omega_1 \lambda_1(x_1, x_2, x_3, x_4, x_5, x_7, x_8)}{x_6^{\gamma_1}} E\left(\frac{1}{C_1^{\alpha_1}}\right) + \omega_2 \lambda_2(x_1, x_2, x_3, x_4, x_5, x_7, x_8) x_6^{\gamma_2} E\left(\frac{1}{C_2^{\alpha_2}}\right). \end{aligned}$$

A two-step procedure is utilized for minimizing *EL*. First, JW find a performance measure independent of adjustment [PerMIA, Leon et al (1987), Leon and Wu (1992)] and minimize it with respect to the control factors excluding x_6:

$$\begin{aligned} &PM(x_1, x_2, x_3, x_4, x_5, x_7, x_8) \\ &\quad = \frac{1}{\gamma_1} \log \lambda_1(x_1, x_2, x_3, x_4, x_5, x_7, x_8) + \frac{1}{\gamma_2} \log \lambda_2(x_1, x_2, x_3, x_4, x_5, x_7, x_8). \end{aligned}$$

The second step involves finding the setting of x_6 that would result in the smallest expected loss:

$$x_6^* = \left[\frac{\gamma_1 \omega_1 E(1/C_1^{\alpha_1}) \lambda_1(x_1, x_2, x_3, x_4, x_5, x_7, x_8)}{\gamma_2 \omega_2 E(1/C_2^{\alpha_2}) \lambda_2(x_1, x_2, x_3, x_4, x_5, x_7, x_8)}\right]^{1/(\gamma_1+\gamma_2)}.$$

The above optimization procedure depends on the specific form of the model for its derivation of the optimal factor levels. We will take a different approach that is simpler conceptually and easy to implement with complete enumeration techniques.

Since JW's models do not properly account for the correlation structure of the data, I will reanalyze the data with both SAS and WinBUGS and compare the results. Failure to recognize the split-unit nature of the data will make the standard errors of the whole-unit effect estimates smaller and will most likely lead to overfitting. Bayesian models in WinBUGS were fit for a single long chain of 10,000 MCMC updates discarding the first 1,000 samples and storing every 5th value of the chain. The random whole unit effects u were sampled from a normal distribution N(0, σ_u^2), where $\sigma_u = \frac{1}{\sqrt{\tau}}$ (with the hyperparameter τ denoting the precision) and $\tau \sim Gamma(v_1, v_2)$. The random effects were assumed to follow the normal distribution in order to facilitate comparison of pseudo-likelihood estimation in GLIMMIX and Bayesian estimation in WinBUGS (GLIMMIX does not fit models with non-normal random effects). The parameters of the hyperparameter τ were taken to be $v_1 = v_2 = 0.001$, i.e. $E(\tau) = v_1/v_2 = 1$ and $Var(\tau) = v_1/(v_2)^2 = 1000$. This assumes a vague prior for the whole-unit error. At each simulation of the Markov chain the sum of the 18 random effects was constrained to be zero.

The fitted conditional models are given below in (2). Fixed effects were chosen according to the ad hoc variable selection procedure which will be discussed later in this chapter. The results from fitting GLLM's for opens using marginal models with SAS, a conditional model with SAS, and a Bayesian model with WinBUGS are shown in Tables 3, 4 and 5. Covariance pattern analysis estimates of the fixed effects have smaller standard errors. The exchangeable working correlation parameter was estimated to be -0.08, which is a measure of correlation within each whole unit. This result is doubtful since we do not expect a negative correlation within the same run. Note that the estimates from mixed models in SAS and WinBUGS closely agree, as well as 95% confidence intervals and posterior intervals. The estimate of the variance of the random effects σ_u^2 was 0.1109 (standard error = 0.0752) and 0.1396 (standard deviation = 0.1083) from SAS and WinBUGS, respectively. The Bayesian estimate of between run variation is slightly larger. Note that all of the 95% posterior intervals for the random effects encompass zero. Since the models fit in GLIMMIX and WinBUGS are essentially the same, we can use GLIMMIX for model selection and WinBUGS for optimization.

Similarly a model was obtained for shorts and the final models fitted in WinBUGS are as follows:

$$\begin{aligned}\log\log\frac{1}{1-\hat{p}_1} &= -5.034 - 0.446x_{2l} + 0.429x_{3l} + 0.572x_{4l} - 0.877x_{5l} - 0.476x_{6l} \\ &\quad - 0.458x_{7l} + 0.441x_{6q} - 3.447\log C_1 + u_1,\end{aligned} \tag{2}$$

$$\log\log\frac{1}{1-\hat{p}_2} = -5.464 + 0.665x_{1l} + 0.396x_{4l} + 1.335x_{6l} - 5.213\log C_2 + u_2,$$

where x_{il} and x_{iq} denote scaled linear and quadratic contrasts, respectively. For a three-level factor, $x_{il} = (-1, 0, 1)/\sqrt{2}$ and $x_{iq} = (1, -2, 1)/\sqrt{6}$. The log C_i contrasts correspond to log C_i = (ln C_i – 1.57)/0.68, where 1.57 is the mean of (ln 3,…, ln 7) and 0.68 = $\sqrt{\sum_i (\ln C_i - 1.57)^2}$. The codes of model fitting and optimization are given in the Appendix.

Once the model is identified, optimization can be done more easily utilizing the Bayesian estimates. The optimization stage requires choice of a loss function. A catalog of loss functions for nonnegative variables can be found in Joseph (2004). Once this choice is made, the loss is averaged over the noise factors, if any, and the expected loss is minimized with respect to control factors. With the complexity factor amplification method, expectation is taken over the noise and complexity factors, since they can not be controlled by a manufacturer.

In JW's procedure, optimization was based on an appropriate PerMIA, which is dependent on both the loss function and the assumed models for the two types of defects. We consider a different approach where there is no need to base the model choice on simplicity of theoretical form for optimization.

When there is no preference for the form of the loss function, it can be taken to be proportional to the probability of failures, since the latter are incorporated in the models through the link function. The loss for the PCB data is evaluated at each treatment combination:

$$L[i, C_1, C_2] = \omega_1\hat{p}_1[i, C_1] + \omega_2\hat{p}_2[i, C_2], \quad i = 1,\dots, \tag{3}$$

where ω_1 and ω_2 are the penalties associated with the two types of defects, and index i refers to a combination of control factor levels. The line width and spacing of C_i = 3 and 4 mils were not used in the actual production and were introduced in the experiment simply for amplification. Therefore the loss was averaged over the production appropriate levels of the complexity factors (5, 6 and 7 mils) assuming a uniform discrete distribution for C_1 and C_2:

$$EL[i] = E_{C_1,C_2}\{L[i,C_1,C_2]\} = \omega_1 E_{C_1}\hat{p}_1[i,C_1] + \omega_2 E_{C_2}\hat{p}_2[i,C_2], \quad i = 1,\ldots \ . \tag{4}$$

Note that the models in (2) suggest that $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\} = \{1, 3, 1, 1, 3, ?, 3\}$ is preferred, assuming equal ω_i's. For x_6, some trade-off is required since $\hat{p}_1$ is minimized at $x_6 = 2.3$ and $\hat{p}_2$ is minimized at the lowest level of x_6. However, it provides no information concerning sensitivity of the expected loss to the levels of the control factors and random effects. For this, we need more than point estimates.

WinBUGS will allow fitting two models for the two defects simultaneously and estimating the expected loss at each run of the Markov chain, thus taking into account variation of the parameter estimates. The optimum settings can be found by computing the expected loss at all possible treatment combinations. Since there were one 2-level and six 3-level factors for our model, we need $2 \cdot 3^6 = 1458$ treatment combinations (t.c.) for a full factorial grid. Two more t.c.'s were added that correspond to the settings similar to JW's recommendations and they were compared with our optimal settings in Table 6.

The expected loss statistics are shown in Table 6 sorted by the 97.5% column. The weights ω_i were set to 1, so *EL* is simply the sum of the expected probabilities for opens and shorts, respectively, averaging over production levels 5, 6 and 7 of the complexity factors. Since these defects are not mutually exclusive, *EL* can exceed 1. The last column indicates the criterion that resulted in the lowest expected loss. For example, treatment combination 510 has the lowest mean and smallest 97.5 percentile of the expected loss. The best setting under the main effects models contemplated earlier $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\} = \{1, 3, 1, 1, 3, ?, 3\}$ with $x_6 = 1$ has the smallest median and 2.5 percentile. This setting differs from the smallest mean only in x_6. Since the best settings in the first six rows of Table 6 are not appreciably different, an optimal setting can be chosen based on other considerations such as cost. The last eight rows of Table 6 correspond to the worst *EL*. The 97.5 percentile for the worst setting *EL*[946] is almost 50 times bigger than *EL*[510]. JW (2004) suggested $\{x_1, x_2, x_4, x_5, x_6\} = \{1, 3, 1, 2.34, 1.57\}$. *EL*[1459] has $\{x_3, x_7\} = \{1, 3\}$ in addition to the settings above and *EL*[1460] has nominal settings for $\{x_3, x_7\}$ - $\{2, 2\}$. JW optimal settings with the levels of x_3 and x_7 set to minimize the loss under our assumed models are not drastically different from our recommended settings.

Summary statistics for the expected loss provide a useful tool for assessing sensitivity of the factor levels. The predicted values of the probabilities p_1 and p_2 are marginal since they are averaged over the random effects. Variation of the random effects will increase variance of the predicted probabilities and hence the expected loss. In the PCB example we can infer that the process is fairly insensitive with the following range of the control factor settings $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$: $\{1, 2\text{-}3, 1\text{-}2, 1\text{-}2, 2\text{-}3, 1\text{-}2, 2\text{-}3\}$.

Next we contrast and compare our fitted models with those of JW. Bayesian measures of model fit and residual diagnostics may also be obtained for frequentist estimates in order to assess the goodness of fit. The Bayes p-value (Gelman et al, 2004, p.162) for a test statistic *T(y)* is defined as follows:

$$\text{Bayes p-value} = \Pr\left(\mathrm{T}(\mathrm{y}^{\mathrm{rep}}, \theta) \geq T(y, \theta) \mid y\right),$$

where the probability is taken over the posterior distribution of the parameter vector θ and the posterior predictive distribution of the replicated data. The test statistic *T(y)* can be defined in a variety of ways depending on the goals of an experimenter. We have chosen *T(y)* as a measure of how far *y* deviates from the predicted value:

$$T(y) = \sqrt{\sum_i \frac{(y_i - \hat{y}_i)^2}{N}},$$

where the summation is over the sample observations and y_i is the number of defects in the i$^{\text{th}}$ sample. The raw residuals $(y_i - \hat{y}_i)$ are not standardized by the corresponding estimate of the standard deviation of the binomial data $\sqrt{n\,\hat{p}_i(1-\hat{p}_i)}$ since there exist predicted probabilities very close to 0 and 1, or exactly 1 to the working precision. Standardizing the residuals would cause infinite values of *T(y)*.

Small values of *T(y)* indicate that the model fits the data well, while small values of $T(y^{rep})$ indicate that the data generated under the assumed model are close to the predicted values. The p-value is the proportion for which $T(y^{rep}) \geq T(y)$. A p-value close to 0 or 1 implies inappropriateness of the model – the observed data would not likely be seen under the assumed model. Plots of the $T(y, \theta)$ versus $T(y^{rep}, \theta)$ are shown in Figures 2-4.

As can be seen in the Figure 2 plots, 1000 points from MCMC simulation are spread evenly around a 45° line and the p-values are not extreme, indicating no apparent lack of fit. Figure 3 shows the scatterplots of *T(y)* versus $T(y^{rep})$ for the model without the random effects. The average residual sum of squares with the actual data is larger than the average residual sum of squares with the replicated data, meaning that some of the variation in the data is unaccounted for and the model fit is not appropriate. Similarly in Figure 4 the scatterplots are obtained for the GLM models fitted by JW and points are even further from the 45° line indicating lack of fit.

We can also use Deviance Information Criterion (DIC, Spiegelhalter et al, 2002) to compare the models. DIC can be viewed as a Bayesian analogue to AIC and is defined as a "plug-in" estimate of fit, plus the effective number of parameters p_D:

$$DIC = D(\bar{\theta}) + 2p_D$$

where θ is a vector of parameters, $D(\theta)$ is the Bayesian deviance, $D(\theta) = -2\log p(Y \mid \theta) + 2\log f(Y)$ and $p_D = \overline{D(\theta)} - D(\bar{\theta})$. $p(Y \mid \theta)$ is the conditional likelihood of *Y*, and the bar operation corresponds to the posterior mean. For members of exponential family with $E(Y) = \mu(\theta)$, $f(Y) = p(Y \mid \mu(\theta) = Y)$. The rule of thumb contemplated by Burnham and Anderson (1998) for AIC suggests that models within 1-2 of the "best" deserve consideration, and 3-7 considerably less support. According to Spiegelhalter et al. (2002), this rule works reasonably well for DIC.

Table 7 lists DIC values for the mixed models in (1), models without the random effects and JW's fixed effects models. The first two sets of models are superior to the models fitted by JW, particularly for the opens. Comparison of the first two sets of models once again suggests that the mixed model for both opens and shorts is more appropriate.

Our models and analysis differ from JW models in the following aspects:

- Our mixed models properly account for the split-unit structure of the data.
- Our models are hierarchical with only main effects and one quadratic term. JW models are not hierarchical and include third-order terms that are aliased with lower-order effects in the orthogonal array. Joseph and Wu based their variable selection

procedure on AIC values, which can lead to overfitting when ignoring the split-unit error. Our method of variable selection is given in more detail later.

- JW's PerMIA approach required certain model assumptions, including, e.g., that there were no interactions between x_6 and $\{x_1\text{-}x_5, x_7, x_8, C_i\}$. If such interaction had been needed, this would have substantially complicated their derivation, while the Bayesian calculation of the expected loss would have been no more complex.
- We tried different link functions (cloglog, logit and probit) and observed that this choice did not appreciably affect variable selection and optimum settings, but DIC was the smallest for the complementary log-log link function.
- The optimization step can be integrated with estimation of the model parameters for the two defects and the best settings can be chosen according to different criteria – mean, median, standard deviation, percentiles of the expected loss etc. It is also easy to see the degree to which *EL* is sensitive to the optimum settings.

Paper Feeder Example

JW's (2004) paper feeder example is an illustration of control factor amplification. Stack force *M* was chosen as an amplification factor with lesser force leading to misfeeds and greater force – to multifeeds. In addition to *M*, there were eight control factors and one two-level noise factor *N*, amount of paper. A complete description of this example is also in JW (2004). An orthogonal array in 18 runs was used with sequential choice of the levels of *M* within each run. The L_{18} from the PCB experiment in Table 2 was modified by changing level 3 to level 1 in x_4. Control and noise factor levels are given in Table 8. At each treatment combination a paper feeder was fed paper five times and the number of times a misfeed or multifeed occurred was recorded. The data are exhibited in Table 9. Two rows correspond to each of the 18 runs, with the first row indicating the level of the stack force *M* and the second row – number of failures out of 5 tries. The experiment appears to have been run as a split-split unit with the noise factor as a split unit factor and *M* as a split-split unit factor. For the sake of simplicity, we ignore the possible split-split unit structure of the experiment and treat it as a split-unit experiment with both M and N being split-unit factors. This simplification can be justified by the observation that changing the level of N (the amount of paper) should induce little error.

It was assumed that misfeeds and multifeeds follow *Binomial*(5, $p_1(\boldsymbol{X}, M, N)$) and *Binomial*(5, $p_2(\boldsymbol{X}, M, N)$) distributions, respectively. Mixed models were fit to the paper feeder data, with the probit link found to provide the best fit:

$$\begin{aligned}\Phi^{-1}(\hat{p}_1) = &-1.65 - 1.93x_{1l} + 2.03x_{2l} + 0.05x_{3l} + 0.52x_{4l} - 3.65x_{6l} + 0.19x_{7l} \\ &+ 0.90x_{8l} + 4.42x_{4l}x_{6l} + 2.74x_{7q} \\ &+ (-4.37 + 1.70x_{2l} - 0.28N)\ (\log M - 2.86) \\ &+ (-0.18 + 0.54x_{2l} + 0.37x_{3l} - 0.43x_{8l})\,N + u_1,\end{aligned} \tag{5}$$

$$\begin{aligned}\Phi^{-1}(\hat{p}_2) = &-0.77 - 0.88x_{2l} + 0.004x_{3l} + 0.90x_{6l} + (1.56 + 0.51x_{2l} + 0.53x_{3l})\ (\log M - 3.75) \\ &+ (-0.07 - 0.21x_{3l})\,N + u_2.\end{aligned}$$

The prior for the whole unit random errors u_i was the same as with the PCB data.

The posterior summaries of the coefficients are given in Table 10. The subscript M corresponds to the centered effect of log M. The coefficient for $N*M$ interaction was mostly negative for the misfeeds; therefore it was included in the model. The two-factor interaction $x_{4l}x_{6l}$ had the largest standard deviation. However, omitting this term has an enormous effect on the expected loss of the optimum factor settings (97.5 percentile of EL becomes more than 10 times larger).

We make several summary observations about the model fit:

- From Table 11 we can see that the total DIC value is the smallest for the mixed models. For misfeeds our mixed model was clearly preferred over JW's fixed effects model. For multifeeds, JW's nonhierarchical model has the smallest DIC value; however, if one adds terms required to make the JW model hierarchical, its DIC increases to 451.9, which is comparable to 451.6 for our model.
- An examination of the misfeed data reveals that 71% of the data were collected at the extremes of the distribution with either 0 or 5 misfeeds (see Table 12) and only 29% in the middle. The opposite is true for the multifeed data – 27% of the data were collected at the extremes and 73% - in the middle. This preponderance of extreme

outcomes for misfeeds results in bias and poor precision for the parameter estimates (see column 3 of Table 10).

- Plots of the average residual sums of squares (RSS) for the mixed and JW models are shown in Figures 5 and 6. Based on these plots, we cannot ascertain superiority of our models over JW models. Clearly, both models for misfeeds have poor fit, and *T(y)* values are less variable for their fixed effects models due to smaller standard errors of the parameter estimates. Both models do not provide an adequate fit with a significant Bayes p-value of zero. The average RSS for the actual data is higher than the average RSS for the bootstrap data, meaning that the actual data are not likely to be observed under these models. Unfortunately, we were not able to find a better model for misfeeds due to the deficiencies of the data set. If an experimenter does not observe at least two counts that are not 0 or n for a certain treatment combination of the control array, then the degree of freedom associated with this t.c. is lost. (A Bayesian analysis is able to extract some limited information.) For example, run 15 has only one observed count not 0 or n for both levels of the noise factor N. There were also four other runs with only one count not in the extremes for level 2 of the noise factor – runs 1, 6, 10 and 18. For multifeed data, only runs 12 and 13 with $N = 2$ had this pattern.
- Due to the small sample size and sparseness of the data, both methods of estimation in GLIMMIX (pseudo-likelihood based and GEE based) did not converge with the probit link even for a main effects model with misfeed data. They did converge for simple models with the logit link though. The probit model in WinBUGS converged, but was sensitive to the initial values of the parameter estimates. Even though we were able to obtain parameter estimates for misfeeds with Bayesian estimation, we observed significant lack of fit for the model.

As with the previous example, we take the sum of probabilities of the two types of defects as our loss function. The expected loss function was estimated at $2^2 3^5 = 972$ control factor combinations over the grid of M changing from 5 to 20 with a step of 2.5. The higher levels of M were considered prior to this step and it was verified that they do not produce the smallest expected loss. The loss was averaged over the noise factor *N*. Several comments can be made about the optimization step of the paper feeder example:

- The best settings from our analysis of the paper feeder data are listed in the first row of Table 13. These settings have the smallest mean, as well as 97.5 percentile of the posterior distribution of the expected loss. The median of *EL* is not a reliable criterion in this example due to the severe skewness of the posterior distribution (second row of Table 13). The mean of the expected loss at JW optimum settings (row 3) is 16 times bigger than the expected loss with our recommended levels. A 97.5 percentile criterion appears to be the best because it insures small values of *EL* on average even under more pessimistic parameter vectors.
- A histogram of 972·7 = 6804 values of 97.5 percentile of the posterior distribution of EL is displayed in Figure 7. The first bin from 0 to 0.1 contains 13 treatment combinations with $\{x_1, x_2, x_3, x_4, x_6, x_7, x_8, M\} = \{2, 3, \text{any}, 1, 3, 2, \text{any}, 10\text{-}15\}$. The worst 25 values of EL ranging from 1.6 to 1.7 have $\{x_1, x_2, x_3, x_4, x_6, x_7, x_8, M\} = \{1, 1, 1, 2, 3, 1 \text{ or } 3, \text{any}, 7.5\text{-}17.5\}$. It seems that EL is more sensitive to control factor settings than to the levels of the amplification factor in the region under consideration.
- The fact that EL takes values greater than 1 tells us that either the models do not fit well or it is possible to have both types of defects simultaneously. There are a few instances in the data when misfeeds and multifeeds occur together, for example run 7, *N* = high, *M* = 35.
- Given the data, there is enormous variability with posterior distributions of parameters, and hence of the expected loss. Thus, if we run MCMC again, results may vary. This is a deficiency primarily due to the sample size of five and the choice of the amplification factor levels, especially for misfeeds, producing less informative data.

Our conclusions from the analysis of the paper feeder data are considerably different from those of JW, as reflected by the very poor performance of their recommended optimum under our chosen model. A follow-up experiment is required to verify the results. A bigger sample size within each treatment combination would improve precision of the parameter estimates and should have been used since it would have added negligible extra cost. Design and sample size issues will be addressed in the next chapter.

Model Selection with FAME

FAMe as presented by JW (2004) can be described with the following features:

1. Two conflicting types of failure modes
2. Opportunity to induce more failures by changing the levels of an amplification factor
3. Categorical response data with small probability of failure at nominal levels of the factors
4. Split-unit structure with the amplification, complexity and/or noise factors as split-unit factor(s)
5. Low resolution fractional factorial design for the control factors
6. Little or no replication of control factor treatment combinations

Categorical data requires the use of GLM's, and the presence of random effects due to the split-unit structure requires fitting GLMM's. Features 4-6 are not essential to FAMe, but will make experimentation more efficient and economical. If the set of control factors $\boldsymbol{X}$ is a mixed level orthogonal array or nonregular array, an additional issue of complex aliasing of the effects arises.

Even without the split-unit structure of the data, variable selection with unreplicated fractional factorial designs (FFD) of low resolution and aliasing of effects is not a trivial task. George and McCulloch (1993) propose a stochastic search variable selection (SSVS) procedure for a multiple regression problem. Chipman, Hamada and Wu (1997) adopted SSVS to FFD's with complex aliasing structure and showed how the effect heredity principle can easily be incorporated with this procedure. The problem becomes even more challenging with categorical response and sparse data. Other methods for variable selection exist, but we were not able to find literature suitable for our case – mixed models with competition for explaining variation in the whole units as due to the whole unit error versus due to the fixed effects that are estimable for the orthogonal array. Hence we consider the following ad hoc method for overcoming these difficulties. A general description of this method is given below and it is illustrated on the paper feeder example following the description. With a small number of potential effects and the absence of aliasing of effects, variable selection can be done easily in GLIMMIX by specifying competing models manually. Automatic variable selection procedures such as forward, backward and stepwise are not available in GLIMMIX.

Our ad hoc method begins by fitting mixed models with split-unit effects (for example, N, M, their interaction and the random effects in the paper feeder example), without the whole-unit effects ($\boldsymbol{X}$).

$$\begin{aligned} g_1(\hat{p}_1 \mid u_1) &= f_1^1(N,M) + \boldsymbol{Z}u_1, \\ g_2(\hat{p}_2 \mid u_2) &= f_2^1(N,M) + \boldsymbol{Z}u_2, \end{aligned} \tag{6}$$

where $g_i(\cdot)$ is a link function. Note that the link function does not have to be the same for type I and type II defects. This step allows the predicted u_i values to retain differences due to fixed effects as well as whole unit error. The next step is to choose a set of active fixed factors by fitting a model for the posterior means (or medians) of the random coefficients u_i with $\boldsymbol{X}$ as predictors:

$$\begin{aligned} \hat{u}_1 &= f_1^2(\boldsymbol{X}), \\ \hat{u}_2 &= f_2^2(\boldsymbol{X}). \end{aligned} \tag{7}$$

This procedure is straightforward with orthogonal arrays of strength 4 or higher. However, when the number of degrees of freedom is only slightly larger than the number of factors, and main effects are aliased with two-factor interactions, this becomes a demanding task, unless interactions are assumed away. To circumvent the problem, an iterative stepwise variable selection procedure according to method I in Wu and Hamada (2000, p.356) may be utilized. This method is supposed to work well “when there are only a few significant interactions that are partially aliased with the main effects”.

The final step in the analysis of the FAMe data is to fit mixed models with the whole-unit and split-unit effects identified previously. Their interactions can also be entertained with a forward selection procedure.

$$
\begin{aligned}
g_1(\hat{p}_1 \mid u_1^{'}) &= f_1^1(N,M) + f_1^2(\boldsymbol{X}) + f_1^3(\boldsymbol{X},N,M) + \boldsymbol{Z}u_1^{'}, \\
g_2(\hat{p}_2 \mid u_2^{'}) &= f_2^1(N,M) + f_2^2(\boldsymbol{X}) + f_2^3(\boldsymbol{X},N,M) + \boldsymbol{Z}u_2^{'}.
\end{aligned} \tag{8}
$$

In conclusion, low resolution FFD's with complex aliasing are not recommended unless there is a strong belief that only main effects are active. The advantage of the above method is apparent in an experiment where each whole-unit corresponds to a single control factor combination, as in the case with both examples from JW (2004). The control factors do not change within each whole unit; therefore the variation due to the whole units is captured by the random effects. The use of the random effect as a response allows moving variable selection for the whole-unit factors from the GLMM realm for categorical data to a linear continuous realm. Under the assumption of normal random effects, standard methods of variable selection may be utilized.

I will now demonstrate the above procedure on the paper feeder example. Mixed models with split-unit effects in (6) were fit to the data in WinBUGS with probit links, log-linear transformation to *M*, linear effect of *N* and their interaction. The posterior summaries of the coefficients are given in Table 14. Neither *N* nor its interaction with *M* appeared to be important for multifeeds; hence it was dropped from the model. As before, the subscript *M* corresponds to the effect of log *M*. The coefficient for $N{\cdot}M$ interaction was mostly negative for the misfeeds, and so it was left in the model. Note that quite a few 95% posterior intervals for the coefficients of the random effects are either entirely negative or positive. This is an illustration of competition for explaining variation in the whole units between random effects and fixed effects from L_{18}. Later we will see how inclusion of fixed effects from the orthogonal array shrinks the random effects toward zero, since estimation of fixed and prediction of random effects are not independent of one another.

The results of the iterative stepwise variable selection procedure are summarized in Table 15. For the misfeeds, x_1, x_2, x_6, x_4x_6 and $x_7{\cdot}x_7$ were identified as useful with an R^2 of 0.87. The main effects x_4 and x_7 were not significant, but will be kept in a subsequent analysis in order to preserve the hierarchy of the model. For the multifeeds, only two main effects were identified with an R^2 of 0.66. This estimation was done via JMP.

The final step in the analysis of the paper feeder data is to identify $\boldsymbol{X} \cdot N$ and $\boldsymbol{X} \cdot M$ interactions by utilizing a forward selection procedure, which was performed manually in WinBUGS. Only interactions of N and M with the main effects were considered. For the misfeeds, $M \cdot x_2$, $N \cdot x_2$, $N \cdot x_3$ and $N \cdot x_8$ appeared to be useful in addition to the whole-unit and split-unit effects, and $M \cdot x_2$, $M \cdot x_3$, $N \cdot x_3$ – for the multifeeds. The final model was represented earlier in equation (5) and Table 10.

We followed the same model selection method with the PCB data. When using a stepwise variable selection for opens according to Wu and Hamada (2000), a complicated model with the following whole-unit effects was found: x_3, x_4, x_2x_4, x_2x_6, and x_3x_4, where x_2 is a three-level categorical factor with two degrees of freedom. Note that this model does not include x_5, which was the largest effect in the model by JW(2004). A simpler model with main effects and one quadratic term has slightly lower generalized χ^2 / df statistic – 1.47 for the latter model in (2) and 1.49 for the model with 3 two-factor interactions. Evidently, the method of Wu and Hamada does not work well in this case. It is quite possible that there are several significant two-factor interaction for opens. However, we are not able to differentiate between the two models due to the partial aliasing and therefore a parsimonious model was preferred.

I have shown that Bayesian analysis of FAMe data has definite advantages and is straightforwardly accomplished with free software WinBUGS. The examples shown in this chapter were more difficult due to the low resolution FFD for the control factors with no replication. In the absence of the above features, model selection can be done simply by fitting a set of relevant effects with SAS GLIMMIX or WinBUGS and selecting the most prominent ones.

CHAPTER III

Properties of Ad Hoc Sequential Designs with Small Sample Sizes

Introduction

It is known that the maximum likelihood estimators (MLE's) of the generalized linear models are generally biased, especially with small sample sizes *n*, even when a fixed design is used. As stated by McCullaugh and Nelder (1989, Chapter 15), "in large samples the bias of maximum-likelihood estimators is $O(n^{-1})$, and hence is negligible compared with standard errors. For samples of more modest size, or for problems in which the number of parameters is appreciable compared with *n*, the bias may not be entirely negligible." Since the Failure Amplification Method involves estimation from GLM's and GLMM's from both fixed and sequential designs, it is important to study the potential bias.

The aim of this study is to assess an increase in bias when a sequential design is used. It is also of interest to investigate when a sequential design is more appropriate than a fixed one due to decreased variance. In the PCB example from Joseph and Wu (2004) the levels of the complexity factors (line width and spacing) were fixed in advance, while in the paper feeder example the levels of the amplification factor (stack force) were presumably chosen sequentially. It was not practical to choose the levels of line width and spacing sequentially since the panels with 160 conductors were produced under the combination of the complexity and control factor levels. In a more general setting, we assume that the levels of the split-unit factor can be chosen sequentially and explore the conditions under which a sequential design is preferred.

There exists extensive literature for estimating parameters of interest on quantal response curves sequentially. When the response is binary, these types of experiments are called sensitivity experiments. A general statistical model with one factor can be formulated as follows:

$$g[p(x)] = \beta_0 + \beta_1 x\,,$$

where $g(\cdot)$ is a link function and p is the probability of failure for a Binomial response Y, Y ~ Bin(n, p).

Voelkel (1999) lists the most common objectives for sensitivity experiments as follows:

1. Estimate the setting L_p of x that corresponds to a user-specified probability p_x: $P(x \leq L_p) = F(L_p) = p_x$. The goal of this type of experiment is to estimate L_p with minimum variance. The most frequently encountered values of L_p are $L_{0.50}$ and $L_{0.10}$.
2. Obtain a good estimate of the quantal response curve in general. The levels of x are chosen in order to minimize the variance-covariance matrix of the estimates for (β_0, β_1). One such criterion is the D-optimality criterion.
3. Estimate the slope parameter β_1 of the quantal response curve. The goal is to minimize $Var(\hat{\beta}_1)$.

In FAMe experiments, several factors are investigated and it is essential to learn about the effects of the control factors as well as the split-unit factor. A split-unit factor experiment at each control factor combination may be viewed as a sensitivity experiment. An experimenter would need a good estimate of the quantal response curve in general at each control factor combination, since the location parameter of each split-unit experiment would affect precision of the control factor effects.

Voelkel (1999) also gives a review of the most common sequential methods such as the Robbins-Munro procedure (1951), Wu's sequential method (1985), Dixon and Mood's up-and-down method (1948), as well as Bayesian methods. With the exception of the up-and-down rule, the above methods utilize updating the parameter estimates by fitting a model after observing a Bernoulli response (n=1) at level x_i and choosing the next level x_{i+1} based on these updated estimates. The up-and-down rule starts with an initial value x_1 and the subsequent trials are made at a lower or upper level depending on the previous response until r trials are completed. Let d denote a change in the absolute value of x between two consecutive trials. This differential amount is fixed in advance. The up-and-down method was primarily designed to estimate the level of x at which 50% of observations fail ($L_{0.50}$). This method works best if the starting value x_1 is reasonably close to $L_{0.50}$ and if $d\beta_1$ is chosen properly, e.g., $d\beta_1 \in [0.5, 2.0]$ for the probit link. However, the up-and-down rule has poor precision for the slope of the quantal response curve, as well as percentiles L_p for small p. Furthermore, "measures of reliability may very well be misleading if the sample size is less than forty or fifty" as stated by Dixon and Mood (1948).

The remaining sequential methods mentioned above require updating of the parameter estimates after each trial and the distributional form of a model is assumed to be known in advance, for example, logit, probit or complementary log-log models. These methods are not practical with the use of designed experiments when there are many factors of interest and there is only limited time available for experimentation. For example, a fractional factorial design with N runs and r levels of the split-unit factor at each run would involve $N \cdot (r-1)$ updates of the assumed model and a reasonable guess of the location and slope parameters at each run.

An alternative to a sequential choice of factor levels with two-level fractional factorial designs is inverse binomial sampling as described by Bisgaard and Gertsbakh (2000). They provide a methodology for determining the number of defective units to detect a given change in the probability of producing a defective unit with fixed levels of Type I and Type II errors. They assume that the center of the design space is positioned at the optimal factor levels and the variability arises only due to variability of the process parameters around the nominal values. Their method can be applied to ongoing production processes where instantaneous testing of the product is possible. They argue that inverse binomial sampling is advantageous over the fixed sample size since it ensures only a certain number of defects is produced. However, their method is not practical when only a limited time is available for experimentation and the defect rate is small. For example, with 16 runs, 5% and 10% for Type I and Type II errors, respectively, 4% probability of failure at the center of the design and a change of 1% in the probability of failure when an active factor changes from -1 to 1, the required number of defectives is 4. Therefore the expected sample size at each control factor combination is 100 and the total sample size is 1600.

In light of the above discussion, an experimenter might consider an ad hoc sequential design that does not involve intermediate model fitting. Next a limited review of the bias of the fixed designs is presented, followed by the sequential ad hoc design study.

Bias of a Fixed Design with Small Sample Sizes

As already mentioned, with small sample sizes, the maximum likelihood estimates (MLE's) of the generalized linear models may be substantially biased. There exist several methods for approximating this bias. Cordeiro and McCullagh (1991) give an overview of these methods and provide general formulae for first-order approximation of biases of the maximum likelihood estimators for distributions from the exponential family, as well as an approximate formula for the bias of the parameter estimates in logistic models. In case of models with canonical link, the first-order asymptotic bias can be calculated as follows (McCullagh and Nelder, 1989, Chapter 15). Let $Y_1, \ldots Y_r$ be the set of r independent observations from *Binomial*(n, p_i), $i = 1,\ldots,r$ and $\boldsymbol{X}$ be a $r\times$(m+1) model matrix with m covariates. Then the logit model is

$$\log \frac{p_i}{1-p_i} = \sum_j x_{ij}\beta_j \; ; \qquad i=1,\ldots r.$$

The log likelihood is

$$l(\boldsymbol{p};\boldsymbol{Y}) = \sum_{i=1}^{l} \left[Y_i \log \frac{p_i}{1-p_i} + n\log(1-p_i) \right].$$

The Fisher information matrix for parameter vector $\boldsymbol{\beta}$ becomes

$$-E\frac{\partial^2 l}{\partial\beta_t \partial\beta_s} = \sum_i n \frac{(dp_i/dg_i)^2}{p_i(1-p_i)} x_{it}x_{is} = \{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\}_{ts} \;,$$

where $\boldsymbol{W}$ is a diagonal matrix of weights. Finally, the first-order asymptotic bias B_f is

$$B_f = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\xi \;,$$

where $\xi_i = -\frac{1}{2}Q_{ii}k_{3i}/k_{2i}$, $\boldsymbol{Q} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T$, and k_2 and k_3 are the second and third cumulants, respectively. For the logit link $\boldsymbol{W}$ reduces to $\boldsymbol{W} = diag\{np_i(1-p_i)\}$, $k_{2i} = np_i(1-p_i)$, $k_{3i} = np_i(1-p_i)(1-2p_i)$ and so $\xi_i = Q_{ii}(p_i - 0.5)$.

Under conditions of approximate quadratic balance (Q_{ii} = constant), a very rough approximation of the bias for small $|\boldsymbol{\beta}|$ is $B_a = (m+1)\boldsymbol{\beta}/n_{\bullet}$, where $n_{\bullet}$ is the total sample size. The approximate bias vector B_a and the parameter vector $\boldsymbol{\beta}$ are approximately co-linear. We will now examine empirically the bias and variance of $\boldsymbol{\beta}$ for a fixed design with one two-level factor with $n = 10$.

Fixed Two-Level Designs with One Factor

We will only consider logit models without the whole-unit error in this study in order to keep the discussion simple. A plot of the logistic curve with $\beta_0 = 3$ and $\beta_1 = 2$ is shown in Figure 8. The probability of failure is 50% when $x_c = -\beta_0/\beta_1$. A D-optimal design for the logistic model is to place an equal number of trials at two points around the $L_{0.50}$: $-\frac{\beta_0}{\beta_1} \pm \frac{1.5434}{\beta_1}$ (Atkinson and Donev, 1996, p. 293). The corresponding probabilities of failure at these two design points are 0.176 and 0.824, respectively.

The problem with fixed designs is that an experimenter will need to guess reasonable values of $\boldsymbol{\beta}$ prior to conducting the experiment. If these estimates are not close to the actual values, the design will be poor. Even if the guess values of $\boldsymbol{\beta}$ are correct, the D-optimal design might not be the best one with small sample sizes, which will be illustrated later in this chapter. The choice of levels of x is crucial with sensitivity experiments. If the levels are in the tails of a quantal response curve, the observed counts will most likely be 0 or n for a binomial response when n is small, resulting in non-convergence of the fitting algorithm for certain models. For example, if $\boldsymbol{Y} = (0, n)^T$ for some levels of x, MLE's do not exist. The existence of the unique MLE estimates is highly sensitive to the center of the design x_c, the distance from the center, d, and the sample size n. As mentioned earlier, the optimum $x_c = -\beta_0/\beta_1$, d=1.5434/β_1 and the equal number of samples is placed at $x_c \pm d$. The percent bias and probability that unique MLE's exist depend solely on the sample size n, $d\beta_1$ and $F(x_c)$. Bias, variance and MSE were calculated by completely enumerating all the possibilities of observed counts at $x_c \pm d$ with n = 10. There are 11^2 = 121 possible combinations of the observed counts vector $\boldsymbol{Y} = (Y_1, Y_2)^T$ for a two-level design. When at least one of the Y_i is 0 or n, unique MLE's do not exist (40 cases). Each combination of $\boldsymbol{Y}$ has a probability $p_{\boldsymbol{Y}}$ associated with it:

$$p_{\boldsymbol{Y}} = \prod_{i=1}^{r} \binom{n}{Y_i} p_i^{Y_i} (1 - p_i)^{n - Y_i}.$$

The exact bias is $B(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = \sum_{\boldsymbol{Y}} p_{\boldsymbol{Y}} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$. (Obviously $\hat{\boldsymbol{\beta}}$ depends on $\boldsymbol{Y}$. For simplicity of notation, this dependency is not shown explicitly.) The variance is $Var(\hat{\boldsymbol{\beta}}) = \sum_{\boldsymbol{Y}} p_{\boldsymbol{Y}} \left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)^2$ and the mean squared error $MSE(\hat{\boldsymbol{\beta}}) = \left[B(\hat{\boldsymbol{\beta}})\right]^2 + Var(\hat{\boldsymbol{\beta}})$. When $\hat{\boldsymbol{\beta}}$ does not exist for certain combinations of $\boldsymbol{Y}$, the above quantities are computed conditionally on uniqueness of the MLE's. Denote the sum of $p_{\boldsymbol{Y}}$ for which $\hat{\boldsymbol{\beta}}$ exist uniquely as Pr(MLE).

The results for β_0 = 3, β_1 =2, n=10, $F(x_c) = F(-\beta_0/\beta_1)$ = 0.50 and $d\beta_1$ from 0.1 to 1.9 are shown in Table 16. The probability that unique MLE's exist decreases when $d\beta_1$ increases. The near optimal design corresponds to a row with $d\beta_1$ = 1.5 with only 75% chance of the unique parameter estimates. The asymptotic variance is a square of the standard errors of the parameter estimates weighted by probability of the observed counts at fixed levels of x. The last column in Table 16 is the percentage of the first-order bias B_f.

When a two-level design is centered, a condition of approximate quadratic balance is satisfied and the approximate bias is $B_a = 2\boldsymbol{\beta}/2n = \boldsymbol{\beta}/n$. Hence the relative bias is $B_a/\boldsymbol{\beta} = 1/n$ for one-factor two-level fixed centered design, or 10% with n = 10. When Pr(MLE) is 1, bias represents the true bias and we can see that all three biases (bias

B, the first-order bias B_f and an approximation of bias B_a) are the same – 10%. When Pr(MLE) is not 1, bias, variance, MSE and asymptotic variance are conditional on existence of unique MLE's and therefore are biased themselves. Excluding the extreme cases has an effect of shrinking the true bias and variance towards zero. For example, out of the 40 outcomes when MLE's do not exist, the most likely outcome is $\boldsymbol{Y} = (1, 10)^T$ with a probability of 0.092. Had the estimate of the slope existed, it would have been very large with large variance. This explains negative bias when $d\beta_1 \geq 1.3$. Note that % bias is the same for the intercept and the slope.

If the levels of x are poorly centered, the probability of the existence of MLE estimates is even lower. For example, with the same parameter vector, sample size, $d\beta_1 = 0.5$ and $F(x_c) = F(-\beta_0/\beta_1 + 0.5) = 0.73$, Pr(MLE) = 0.86 compared to 0.98 when $F(x_c) = 0.50$.

With a sequential design, one would hope to experiment with more than just two levels. Otherwise, there is no real advantage in using a sequential design. Next we will compare fixed and sequential designs in the context of designed experiments.

Comparison of Fixed and Ad Hoc Sequential Rules with Designed Experiments and Small *n*

I will now describe an ad hoc sequential rule that one may use with experiments similar to the paper feeder example. As in that example, we will use sample size of $n = 5$ at each treatment combination. The number of levels for the split-unit factor at each control factor combination is restricted to four. In cases such as this with small sample sizes and a small number of levels, it is possible to enumerate all the possibilities. The number of all combinations of $\boldsymbol{Y}$ is $(n+1)^r$, where r is the number of levels. One such rule with $n = 5$ and $r = 4$ is shown in Table 17. The design starts at an initial level x_1 and the next sample is taken at $(x_1 + d)$ or $(x_1 - d)$ depending on the outcome Y_1. In choosing to increase or decrease the level, we assume only that we know the sign of the slope of the sequential factor. Without loss of generality, we can take it to be positive. For example, the first row of Table 17 corresponds to the following rule. If $Y_1 = 0$, we take the next sample at $x_2 = x_1 + d$. If Y_2 is 0 again, we increase the step and take the next sample at $x_3 = x_2 + 2d$. Finally, if $Y_3 = 0$, $x_4 = x_3 + 2d$. Without knowledge of the sign of the slope, after getting $Y_1 = 0$ an unbiased rule would be to choose the next level $x_2 = x_1 + d$ or $x_2 = x_1 - d$ randomly. Note that the spacing among the four levels is not necessarily a fixed differential amount d. It can be a fraction, as well as a multiple of d. The main idea of this sequential design is to obtain observed counts both in the middle and close to the tails of the distribution. More data in the middle of the probability distribution will provide more precise estimation of β_0/β_1. More data toward the tails of the distribution will provide more precise estimation of the slope β_1.

This rule is applied to a sequential factor M crossed with a control array $\boldsymbol{D}$. In other words, at each factor level combination of the control array we choose four levels of the sequential factor according to the rule in Table 17. The starting point of the design is the same for each run - x_1. Suppose we have three control factors and one sequential factor. Then $\boldsymbol{D}$ can be chosen as a full factorial design in eight runs. Denote the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma)^{\mathrm{T}}$, where γ is the slope of the sequential factor. That is, we assume a first order model.

With eight runs, sample size of five, and four levels for a sequential factor, a complete enumeration of all possible outcomes would provide $(n+1)^{r \cdot \mathrm{N}} = 7.96 \cdot 10^{24}$ possibilities. Therefore a simulation was performed with 10,000 random draws of the $\boldsymbol{Y}$.

In order to compare the sequential rule from Table 17 with a fixed design, we need to construct such a design. If the main effects are different from zero, a D-optimal design would place the design points at different levels of M from run to run. For this to be efficient with small sample sizes, accurate guess values of the control and sequential factor effects are necessary. Unfortunately, such knowledge is rarely available and therefore some protection against varying $F(x_1)$ is needed. Here we assume that we do not have such knowledge about the control factor effects and the levels of the split-unit factor are restricted to four. A proposed design is $\boldsymbol{D} \times [\hat{L}_{0.50} - d \quad \hat{L}_{0.50} - d/3 \quad \hat{L}_{0.50} + d/3 \quad \hat{L}_{0.50} + d\,]$, where $\hat{L}_{0.50}$ is a guess value of the $L_{0.50} = -\beta_0/\gamma$. Table 18 compares different values of $d\gamma$ for $\boldsymbol{\beta}^{\mathrm{T}} = (-3, 1, -2, 0.3, 2)$, $n = 5$

and $F(x_1) = 0.50$. The percentage of the bias vector B was nearly the same for all the main effects, and only the largest % bias is shown. The same is true for the estimate of the first order bias $\hat{B}_f$. Variances of $\hat{\beta}_1$ and $\hat{\beta}_3$ were not affected significantly by $d\gamma$ and were not included in the table. The existence of unique MLE's is rather insensitive to $d\gamma$ and is not a major concern as opposed to the example in Table 16. The percent of bias varied from 6.8% to 9.9% and $\hat{B}_f$ provided a good estimate of B. The approximate bias B_a was $\frac{5}{160} \times 100\% = 3.125\%$ and it is clearly underestimating the true bias. Note that variances of two largest effects β_2 and γ varied greatly with $d\gamma$ in opposite directions, i.e. Var[$\hat{\beta}_2$] (Var[$\hat{\gamma}$]) increased (decreased) with larger $d\gamma$. They become close to one another when $d\gamma = 3$. Asymptotic variance is also close to the actual variance of $\hat{\beta}$.

The same simulation was done for a design with the choice of levels of the sequential factor according to Table 17. The results are shown in Table 19. The same value of d does not result in the same levels of the split-unit factor M in fixed and sequential designs. In the fixed design, the data are collected on $[\hat{L}_{0.50} - d \quad \hat{L}_{0.50} + d]$ for M. In a sequential design, the spread of the levels of M depends on the observed counts. Here the probability of existence of unique MLE's [Pr(MLE)] is 1 for all rows. The percentage of the true bias B was similar for the common intercept β_0 and γ, as well as for the control factors. The latter is almost double $\%\hat{B}_f$. Variances of $\hat{\beta}_2$ and $\hat{\gamma}$ are similar when $d\gamma = 1$. The fixed design with $d\gamma = 3$ and the sequential design with $d\gamma = 1$ are almost identical with the difference of 2% in the maximum of the bias vector.

Fixed and sequential designs with various $\boldsymbol{\beta}$ are contrasted in Table 20. The value of $d\gamma$ was chosen such that the standard errors of β_2 and γ are comparable. The fixed design was extremely sensitive to the value of $d\gamma$. For example, for the third design with $\boldsymbol{\beta} = (-3, 3,-3, 0.3, 3)$ probability of unique MLE's was only 0.69 with $d\gamma = 3$. Larger $|\boldsymbol{\beta}|$ causes lack of convergence more often: when $\boldsymbol{\beta} = (-3, 5,-5, 0.3, 6)$ and $d\gamma = 5$, P(MLE)=0.13. Bias and variance reported for the last row in Table 20 for the fixed design are not meaningful since unique MLE's existed only 40% of the time. The sequential designs were quite robust to the choice of $d\gamma$ and they performed better with the presence of strong effects in the model.

Next we will consider a saturated model with 3 df for main effects, 3 df for two-factor interactions and 1 df for three-factor interaction for a full factorial design in three factors with N = 8 runs, $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}, \gamma)$. In general one would expect that the true model is simpler and does not involve high-order interactions. For the purpose of model selection though, often the first step is to fit a saturated model with all degrees of freedom from the control factor array. The results for $\boldsymbol{\beta}^T = (0, 1, -1, 0.5, 1.2, -0.7, 0.2, -0.1, 3)$, $d\gamma = 2$ and $F(x_1) = 0.95$ are shown in Table 21. $\hat{B}_f$ is a good approximation of the true bias except for the three-factor interaction.

Maximum likelihood estimates for a saturated model will exist only when we can estimate a location parameter from each run. For example, in the paper feeder example run 15 for misfeed data (Table 9) had only one observed count not 0 or n, and the location parameter cannot be estimated from this run. In such cases, Bayesian analysis is an alternative since it will provide a finite parameter estimate. When the number of parameters to be estimated is less than N, one may still be able to estimate the model given all the data, even with some individually non-informative treatment combinations.

In conclusion, the advantages and disadvantages of the sequential designs with GLM's are as follows:

- The main advantage of sequential design is that we can continue collecting the data until satisfactory results are obtained, time and resources permitting. General advice is to obtain at least two samples with observed counts not 0 or n.
- Sequential designs are robust to misspecification of the guess values of the parameter vector $\boldsymbol{\beta}$. In fact, even with no prior information about $\boldsymbol{\beta}$, it is still possible to obtain meaningful data sequentially. When using a fixed design, an experimenter might be left with completely non-informative data.
- Bias from a sequential design is generally higher than bias from a fixed design.

I have only considered logistic models without split-unit effects in this study. It would be useful to examine properties of sequential designs with other link functions and with the presence of the whole-unit effects. Simulation studies of GLMM's in WinBUGS would require calling WinBUGS from other programs (SAS, R, etc.) since the analysis of each simulation itself involves iterative solution. This would require significant computing resources and can be an area for further research.

CHAPTER IV

Composite Disc Experiment

Introduction

In their discussion of Joseph and Wu (2004), Leitnaker and Mee (2004) list two examples with a categorical response variable and the existence of a possible amplification factor. One of the examples involves composite disc production at Huhtamaki Consumer Packaging West Inc., Los Angeles, CA. This division of Huhtamaki manufactures packaging to quick service restaurants and beverage vendors, institutional caterers, airline caterers, etc. The composite disc is a component of the Quickspread® container that restaurants use to dispense sauces. The main consumer of this product is McDonalds (MD). Recently Huhtamaki have launched a new product – composite discs to be used in sour cream containers for Taco Bell (TB). The design of these discs is similar to the MD discs, but there are important differences. Figure 9 shows a completed TB disc that has three layers of laminated material: PET (transparent film), bleached paperboard and a triangular blue tab stock material. Figure 10 shows a disc with a blue tab partially split from the paperboard and PET. The only differences between TB and MD discs are the shape of the tab (triangular versus rectangular, respectively) and the position of the seal. The blue tabstock material is sealed to the paperboard for the MD discs and to the film for TB discs. The same machine is used for production of both types of discs.

The filling of a container flows through a valve cut in the center of the disc which is attached to the bottom of a container. One of the characteristics of the finished product that is critical to the customers of Huhtamaki is seal integrity. If a container is not sealed properly, it can cause leakage of sauce or cream into the shipping container.

The company was predominantly interested in whether it was possible to increase the speed of a production line without sacrificing the quality of the product. Currently they run two shifts with a machine operating at 70 feet per minute (FPM). If they were to increase production volume with this speed, they would have to run a third shift and hire and train additional operators. At 70 FPM, the tab sealing process for MD disks is meeting specifications and a very low percent of defects is detected. They have experimented with MD discs and found that indeed the bond degrades with increased speed. An experiment was conducted under the supervision of a project manager for his Six Sigma Black Belt Certification Project (Pettigrew, 2003). At 110 FPM, they experienced problems with the delivery end of the press. Due to the high speed, the discs were not stacked properly and were difficult to collect. A final recommendation was to increase the speed to 90 FPM and invest in improving the delivery end of the press.

Only one type of defect was encountered with MD discs – a weak seal. When the company started the production of the TB discs, they received complaints from their customers regarding a new type of defect – a very tight seal. If the bond of the seal is too strong, a blue tabstock material would not tear off completely and could potentially contaminate the food product. They have also experienced weak seals with the TB disks. Therefore TB production appears to be a good candidate for FAMe experimentation with two types of defects and speed as an amplification factor.

Design of the Taco Bell Disc Experiment

I helped design and conduct an experiment at the manufacturing facility of Huhtamaki under the supervision of the project manager Mark Bond. Due to the differences in design of TB and MD discs and limited experience with the TB product, only a modest amount of information was available about the factors that affect the sealing process. The Triweb Design Team identified eight control factors to be potentially influential. An experiment was planned with speed at three levels and the other seven factors at two levels. Factor levels and notation are given in Table 22.

A blue tab stock material type was known to affect the variability of MD discs. An operator of the TB production line was not certain whether the issue of variability of the tabstock changing from splice to splice was applicable to TB discs production since this material laminates to the film, not paper. I suggested including this factor with two levels. A “bad” splice of the blue tabstock paper was known to cause problems with MD discs. This factor is a “noise” factor and cannot be controlled during the actual production. However, for the purposes of experimentation it was controlled as a factor.

In order to construct a design with one 3-level and seven 2-level factors, a 2_{IV}^{9-4} design from Montgomery (2001, p. 671) was used as a starting point. The design generators were F = BCDE, G = ACDE, H = ABDE, J = ABCE. Thirty-one degrees of freedom for this design are distributed among 9 main effects, 21 two-factor interactions and 1 three-factor interaction. A and B factors were used to construct a 3-level factor speed (S): (-1, -1) factor combination in (A, B) corresponds to low speed, (-1, 1) and (1, -1) - to medium speed and (1, 1) – to high speed. The speed was a difficult-to-change factor. Therefore the experiment was run with eight whole units each containing four split units. A, B and CD were used to generate the whole units. The whole and split unit contrasts with aliasing are shown in Table 23.

As can be seen from the factor relation diagram in Figures 11-13, there were four production lanes from which a finished product was collected. Sixty-six samples were taken from each of the 32×4 treatment combinations and tested by three different measurement techniques, which will be described in the next section.

Measurement Process

During the preparation stage of the experiment we decided which characteristics of the sealing process to measure. A continuous response was used in the experiment with MD disks. A special machine was available at the end of the production line for measuring the force required to separate a tab from a paperboard – Imada peel-off tester (Figure 14). As described by Pettigrew (2003), “this device clamps the disc in a vise, grips the folded tab with an effecter and applies a steady pull at 12 inches per minute to remove the tab from the paperboard. A software program captures the amount of force required to peel the tab at preset intervals and averages them to deliver a single result.”

There are 360 measurements of peel force available for each disc. The program can be set to provide summary statistics such as mean, median, standard deviation etc.

A lower specification limit for the minimum peel force of the MD discs was set at 0.4. There was no need to establish an upper specification limit since they have never received any complaints about seals that are too tight. A distribution of the minimum peel force for MD discs with the machine running at 70 FPM is approximately normal with the mean of 0.87 and standard deviation of 0.068. A three-sigma lower limit is 0.666, well above the lower spec limit. With the TB discs, the opportunity for a strong bond is much greater since the tab laminates to the film and the temperature of the film and a surlyn poly layer on the tab surface can be quite high resulting in a very sticky seal. The Triweb team was not certain whether the same lower spec limit applies to the TB discs and it was required to develop upper spec limit because of the complaints with tight seals. Indeed, there was a problem on both sides of the peel strength distribution with minimum peel strength ranging from 0 to 2.4 (as will be seen later).

The degree to which continuous measurements of seal integrity correlate with actual sealing characteristics was not well known for the TB discs. Therefore we believed that it was necessary to create a categorical response variable that would allow us to ascertain the relationship between continuous and categorical responses. Based on the previous experience, an ordinal measure of the quality of the seal was developed. It is a composite measure of visual characteristics of the seal and the amount of tab paper left after the tab is removed. This categorization was assumed to correlate with the amount of peel force required to remove the tab. The seven categories are defined below:

- Category 1 – a very tight seal, with the seal area completely covered by the tab paper
- Category 2 – a tight seal, with some of the tab paper inside the seal area
- Category 3 – a tight seal, with little paper on the sealing edge and/or outside the sealing area
- Category 4 – a perfect seal with a consistent pattern and all the tab paper removed

- Category 5 – a consistent seal with a weak pull (little force was required to remove the tab), weak seal
- Category 6 – a seal with an inconsistent pattern, weak seal
- Category 7 – almost no seal or no seal at all, very weak seal

One sleeve of approximately 300 discs was collected at each treatment combination. There were 32×4 = 128 sleeves stored after the completion of the experiment. The measurement process required a substantial amount of time and it was not feasible to perform it during the actual experiment. Randomization of the discs from each sleeve would require taking all 300 discs out of the sleeve and picking random samples, which was considered to be an unnecessary task due to the absence of autocorrelation. Sixty-six samples were taken from the top of each sleeve and measurements were obtained in the following order:

1. Six samples were tested with the Imada peel tester. It takes approximately one minute to obtain a complete profile data on each sample. Summary statistics such as mean, min, max, variance and standard deviation were also recorded.
2. Forty samples were tested according to the categorical scale above with a slow peel speed.
3. Twenty samples were tested according to the categorical scale above with a fast peel speed.

The rationale for testing samples with slow and fast peel speed was that the company did not have control over the way customers remove the tab and it was desirable to make the process robust to this type of user variation. A smaller sample size for fast peeling was deemed to be suitable due to the increased probability of not removing the tab paper completely.

In the next two sections the analysis of both continuous and categorical responses will be presented and the data issues will be discussed.

Analysis of the Continuous Response

Analysis of the Mean Peel Strength

The minimum peel force was determined to be an adequate measurement of seal integrity of the MD discs owing to the process exhibiting only one type of defect – a weak seal. Since there are issues with seal integrity on both extremes of the peel strength, mean peel force appears to be more appropriate for the TB discs. Residual variation for the 32×4×6 means was only 2.7%, after accounting for variability due to 32 runs and lanes nested within runs (see Table 24). Hence we can average the six continuous measurements and proceed with the analysis. Similarly for power transformation of the mean standard deviation residual variation was 6.7% (Table 25). Transformations were applied to both responses to alleviate unequal variance problems. Note also that variation due to the lane effect is 4.5% and 5.9% of the total variation for the transformed mean and standard deviation, respectively, and is negligible compared to variation due to the fixed effects combined with random block effects from the orthogonal 32-run array. The mean of the means and standard deviations of the 6 measurements are shown in Table 26 together with the levels of the control factors and the ordinal data that will be discussed later.

The next step is to choose a set of active fixed effects taking into account the split-unit structure of the data. Statistical software such as JMP and SAS does not allow estimating fixed and random effects simultaneously when the random effects are aliased with the whole-unit effects and there are no degrees of freedom for whole-unit error. SAS procedure GLIMMIX gives a warning: "Mixed model has saturated mean and profiled variance. Fit does not proceed." Hence I will use a procedure based on Lenth's PSE method.

A model with all 31 factorial effects was fit with an R^2 of 96.2% and 127-31 = 96 degrees of freedom for the error term (variation due to lanes). RMSE for the above model was 0.106789. Speed, CD, EF, EG, FG, CEH and CGH are the seven whole-unit effects. Only Speed and CGH are partially aliased with each other. Hence CGH was replaced with the (CGH-Speed) column, making a set of orthogonal whole-unit contrasts.

Table 27 lists seven whole-unit and twenty four split-unit naïve and corrected t-ratios. When a model has error degrees of freedom, JMP calculates Lenth's PSE from the t-ratios rather than from the parameter estimates themselves. Calculated in this way, Lenth's PSE reported by JMP is actually the ratio of Lenth's estimate for σ versus the RMSE. For instance, from the seven whole-unit t-ratios in Table 27, we obtain $\text{PSE} = 1.5 \cdot (3.82 + 3.26)/2 = 5.31$. Thus, the whole-unit error mean square is much larger than the naïve mean square error. The correct t-ratios are obtained by dividing by 5.31.

We could compute Lenth's PSE for the split-unit effects in an analogous manner if they were uncorrelated. (This would be 3.55.) Instead we use JMP's calculation of the PSE from orthogonalized estimates from the model with just 24 split-unit effects $\left(\text{PSE} = \dfrac{\hat{\sigma}}{RMSE} = \dfrac{0.3632}{0.2858} = 1.27\right)$. This ratio must be multiplied by $\dfrac{0.2858}{0.1068}$ to account for the different RMSE used for the naïve t-ratios in Table 27. Thus, for split-unit effects, we divide the naïve t-ratios by 3.40 [= 1.27(0.2858/0.1068)].

In order to evaluate significance of the effects, we can assess p-values based on the critical values from Ye and Hamada (2000). The corrected t-ratios are compared to simulation-based critical values for the individual error rate. I reported p-values for the largest effects. Three effects are significant at $\alpha = 0.05$ – speed, material type J, and die pressure D. I will also include CD interaction because its p-value is only slightly higher than 0.05. In order to make a model hierarchical, the effect C needs to be included as well. Hence my final model for the square root of the mean (averaging across 6 measurements) of the mean response contains speed, blue tab material type (J), top preheat (C), die pressure (D) and CD interaction. The R^2 for this model is 86.6% with RMSE of 0.179 and the mean response of 1.04. The residual by predicted plot (Figure 15) indicates the residuals do not seem to follow random pattern at the left-hand side of the plot. The red points above the zero line are observations with lanes 2, 3 and 4. The actual response values for these points were higher than predicted by the model and therefore they ought to be closer to the middle values of strength where the target is. The blue points in the lower left quadrant of the plot correspond to the runs where the actual mean peel strength was zero or very close to zero (runs 13, 14 and 21). The lack of fit with the blue points is not a concern since we are not interested in running the process with such low peel strength. In order to investigate inadequacy of the fit with the red points, I will refit a model with the lane main effect added.

JMP output for the above model is shown on Figure 16. The R^2 increased slightly to 87.6%, and the means for the four lanes show an increasing trend from 0.97 to 1.10. The p-values in the parameter estimates section of the output are not correct, since the standard errors for the split-unit effects C, D and J and whole-unit Speed and CD are underestimated in this analysis. However, the p-value for the lane effect is correct, since it uses an estimate of within lane variation. The lane effect is significant at $\alpha = 0.05$. When I performed the variable selection procedure using Lenth's PSE method for each lane separately, the same set of effects manifested themselves and the parameter estimates for the reduced models were virtually the same, except for the speed effect. This effect was decreasing from lane 1 to 4, with the values in the range of -0.34 to -0.23. However, speed by lane interaction proved to be unimportant with the data from all four lanes (p-value = 0.27).

The parameter estimates are not affected by the invalid standard errors and are correct. For the general linear model $Y = X\beta + \varepsilon$, where ε is distributed as $N(0,\Sigma)$, the uniformly minimum variance unbiased (UMVU) estimator of β is given by the ordinary least squares (OLS) estimator $(X'X)^{-1}X'Y$ if and only if there exists a q×q nonsingular matrix F such that $\Sigma X = XF$ [Theorem 6.8.1, Graybill (1976)] where q is the number of parameters in β. Consider a mixed linear model $Y = X\beta + ZU + \varepsilon$, where X is a n×q design matrix of the fixed effects, Z is a n×m design matrix of the random effects, $U \sim N(0,G)$, $\varepsilon \sim N(0,R)$. Assuming that $G = \sigma_u^2 I_{m\times m}$ and $R = \sigma_\varepsilon^2 I_{n\times n}$, $\Sigma X = (ZGZ' + R)X = (\sigma_u^2 ZZ' + \sigma_\varepsilon^2 I)X = \sigma_\varepsilon^2(\frac{\sigma_u^2}{\sigma_\varepsilon^2}ZZ'X + X) = XF$. Let $F = F_1 + F_2$. Then $F_2 = I$ and $\sigma_u^2 ZZ'X = XF_1$. Hence $X'\sigma_u^2 ZZ'X = X'XF_1$ and $F_1 = \sigma_u^2(X'X)^{-1}X'ZZ'X$. As long as the design matrix X is nonsingular, F_1 is also nonsingular. Therefore F is nonsingular and the OLS estimate of β is also a UMVU estimator.

We can see that the largest effects are speed and blue tab paper type, both with a negative estimate. Increasing the speed will weaken the bond, as well as switching from "good" (J = -1) to "bad" (J = 1) type of paper. The response is nominal-the-best type, though at this point the target value is not available to us. The optimal setting of the factors will be discussed after we examine the relationship between continuous and ordinal responses.

From table 23 we can see that CD interaction is aliased with HJ. It is unfortunate that the only significant interaction happened to be aliased with another two-factor interaction. The project manager favors interpreting the marginally significant contrast as a CD interaction based on his expert opinion. However, in subsequent experiments it would be advantageous to make sure that both interactions are estimable.

Analysis of the Within Standard Deviation of the Peel Strength

The same steps as with the analysis of the peel strength were followed for the analysis of the within piece standard deviation. Final results are shown in the output from JMP in Figure 17. Somewhat surprisingly, the same set of effects proved to be active for the square root of the mean of each set of the standard deviations. The R^2 for the model with speed, C, D, J and CD interaction was 85.7%. The lane effect was not significant (p-value = 0.43). Even the signs of the parameter estimates were the same, meaning that pieces with stronger bonds have more within variability.

Categorical Data Analysis

Data Screening

Recall that seven categories were created which were believed to represent the strength of the bond with a perfect category in the middle. Ordinal data are encountered in industrial application frequently and often they are an alternative to unobtainable or expensive continuous measurements. This was not the case with the TB experiment. Continuous data were as easy to obtain as categorical since the Imada peel tester was already available. Actually it took longer to obtain categorical data since there were

$(40+20) \times 32 \times 4 = 7{,}680$ tabs to be removed. A single technician accomplished this enormous task.

The analysis of the categorical data with slow peel (columns 14-24 in Table 26) is presented next. A close examination of the data reveals that for the majority of runs, samples from different lanes reside in the adjacent categories, except for several outlying runs. For instance, in run 16 samples from lanes 1 and 2 were predominantly in category 6, while samples from lanes 3 and 4 were on the other side of the scale. It is unexpected that samples produced within a short period of time under the same conditions would be so disparate. While collecting the ordinal data, the technician found that an additional category is needed. Originally the category 6 was defined as “a seal with an inconsistent pattern, weak seal”. The technician discovered that samples with an inconsistent seal pattern could differ in terms of the amount of pull required to remove the tab. Therefore we need to redefine the categories and split the original category 6 into two categories. Fortunately, the technician took notes while collecting the data and was able to split the data in category 6.

The description of the redefined categories and their correspondence with the original categories is presented in Table 28. The original categories 1-4 were left unchanged, the category 6 was split into categories 5 and 7, and the original categories 5 and 7 became 6 and 8, respectively. The new categories are ordered by the amount of pull required to remove the tab, from the strongest to the weakest. However, the categorical response in this study is on a two-dimensional scale, which is illustrated in Figure 18. The amount of pull is on a horizontal axis and consistency of the seal – on a vertical axis. Vertical arrows represent variation in consistency of the seal; presumably strong seals are more consistent, but this postulation was not verified for categories with seals completely or partially covered by the tab paper. Therefore we do not have enough information to order the categories by consistency of the seal. Ordering by the amount of pull is somewhat subjective because originally categories were defined primarily based on the visual characteristics of the discs which were thought to correlate with pull strength. The underlying assumptions with pull strength ordering are that seals with more paper left after removing tabs have stronger pull (categories 1-3), a consistent seal with a good pull (perfect category 4) is stronger than a slightly inconsistent seal with a good pull (category 5) and that an inconsistent seal in category 7 involves less pull than a consistent seal in category 6. It appears from the discussion above that the combined categories 1-3, 4-5, 6-7 and 8 have less ambiguity as far as pull strength, but it would be beneficial to analyze both cases and compare the results. The data on slow peel with 8 redefined categories is given in Table 29. The number of samples in each category was as follows: 478, 1404, 675, 758, 334, 189, 464 and 818 for categories 1 to 8, respectively. Note that category 6 (C6) is the least populated one and often the data have a gap between categories 4 and 6, i.e., part of the data falls into C4 and C6 without any samples in C5.

Proportional Odds Model

One of the goals of this experiment is to identify the relationship between the ordinal and continuous responses. Let *Y* be an observed ordinal response and Y^* - an underlying unobservable (latent) continuous variable corresponding to *Y*. Then *Y* falls into category k if $\alpha_{k-1} < Y^* \le \alpha_k$, where α_k denote cutpoints, k = 1,…,K. Typically, $\alpha_0 = -\infty$ and $\alpha_k = +\infty$, while $\alpha_1,...\alpha_{k-1}$ are parameters to be estimated. A general form of a cumulative link model (Agresti, 2002) is of the form:

$$P(Y \le k \mid X) = P(Y^* \le \alpha_k \mid X) = g(\alpha_k - \beta' X),$$

where g(·) is a link function. If the latent continuous response is modeled as $Y^* = \beta' X + \varepsilon$, then normality of the error term implies a probit link for cumulative probabilities. If ε follows a logistic distribution, the cumulative logit model, or proportional odds model results.

Fitting ordinal logistic regression in the context of design of experiments with many factors presents numerous challenges. The data in Table 29 are sparse with a majority of the cells being empty. This problem is common for contingency tables with many variables and categories. Agresti (2002, p. 395) remarks that “although empty cells and sparse tables need not affect parameter estimates of interest, they can cause sampling distributions of goodness-of-fit statistics to be far from chi-squared.” In this study we will not rely on the Pearson chi-squared and likelihood-ratio chi-squared statistics to assess goodness-of-fit but apply a modification of ad hoc variable selection procedure from Chapter II.

The structure of the design in this experiment differs from JW examples in the following aspect: each whole unit is a block consisting of 4 runs while in JW examples each whole unit was a single control factor combination from L_{18}. Therefore the following procedure is deemed appropriate. First I fit a mixed effects model with the lane effect omitting any effects from the 32-run design:

$$\log\frac{P(Y \le k \mid X)}{1 - P(Y \le k \mid X)} = \alpha_k - 1.27\,L1 - 0.03\,L2 + 0.79\,L3 + u_i$$

where k = 1,…7, i = 1,…32 and *L1-L3* correspond to the dummy variables for the lane effect. The random effects u_i correspond to random run effects and contain combined variation from the whole-unit and split-unit errors. Treating the fitted random effects u_i as a response variable in the next step would allow using the same variable selection method based on Lenth's PSE as in the continuous case. The effects chosen under the above procedure are the same as with the mean peel strength – Speed, C, D, J and CD. Interestingly, correlation between the mean peel strength and the fitted random run effects u_i was 84%. The final mixed model was fit using both GLIMMIX and WinBUGS. The results were practically the same, and only the latter are reported since we will need Bayesian analysis for optimization purposes. Details on parameter estimates and posterior intervals are given in Table 30.

$$\log\frac{P(Y \le k \mid X)}{1 - P(Y \le k \mid X)} = \alpha_k - 3.06\,S + 0.32\,C + 0.67\,D - 3.51\,J + 0.80\,CD - 1.03\,L1 - 0.03\,L2 + 0.63\,L3 + b_j,$$

where, k = 1,…7, and b_j correspond to eight random block effects.

The cumulative probabilities were parameterized as $P(Y \le k \mid X) = P(Y^* \le \alpha_k \mid X) = g(\alpha_k + \beta' X)$. The sign of β has the opposite meaning with this parameterization, i.e., if the elements of β are positive, Y tends to be smaller at higher values of X. The negative effect estimate for speed suggests that the cumulative probability of Y increases as speed increases, i.e. weaker bonds are more likely with higher speed. Similarly weaker bonds are more likely with "bad" tab paper and less likely with high levels of C and D. The negative effect estimate for lane 1 means that on average lane 1 produces weaker bonds. The effect estimate for lane 4 is positive, and is equal to (0.93+0.04-0.57) = 0.40.

Optimization

A loss function for the TB experiment is not known at present and the data are being collected to evaluate customer dissatisfaction with the product. I propose an ad hoc loss function in order to show the optimal factor settings. I consider a loss function of the form $E(L) = \frac{1}{4}\sum_{i=1}^{4}\sum_{k=1}^{8}\omega_k p_{ik}$, where index i refers to lanes 1 through 4 and ω_k are the penalties associated with the probability of belonging to category k. Table 31 lists one set of possible $\omega_{k,}$ where positive values increase the loss function and negative values correspond to categories with the amount of pull close to the target.

Even though the type of tab paper (J) is a noise factor, its effect is the largest and it is not evident how to make a process robust to J. The project manager at Huhtamaki is working on resolving this issue in collaboration with the supplier of the tab paper and they plan to experiment with this factor by taking rolls that the supplier will produce with a varied composition of chemicals.

The expected loss statistics sorted by S, J and C levels are shown in Table 32. The loss is minimized with low speed, bad material and {C, D} = {-1, 1} (run 8). The next best combination corresponds to high speed, good material and {C, D} = {1, -1} (run 16). There were 24 distinct treatment combinations for a full factorial design in S, C, D and J. I have shown all 32 runs in order to see how the expected loss is affected by the random block effect. Variation due to this effect is quite large. For example, run 26 and 18 have the same levels of S, C, D and J. The mean expected loss is quite different though, -2.22 for run compared to 1.62 for run 18.

The above recommendations are of limited usefulness to the Triweb crew since they do not have control over the tab material factor and whether a new roll of material would behave as “good” or “bad”. Ideally they would like to have no or little variability in the incoming material, set speed at the high level, adjust the control factors to some fixed levels and run the machine without having to worry about quality of the product. Unfortunately, this goal is not achievable with the current variability of supplier’s material.

Correlation between the continuous and ordinal response

A bivariate analysis of the square root of the mean peel strength versus the average sample category score is shown in Figure 19. The average sample category score for each of the 128 factor-lane combinations refers to $\frac{1}{40}\sum_{k=1}^{8} k \cdot m_k$, where m_k represents the number of samples in category k. A simple linear model has an R^2 of 85% and the regression plot indicates that there is a considerable amount of variation in mean peel strength in the neighborhood of category 4. Another way of modeling a relationship between the continuous and ordinal response is to discretize the average sample category score to the nearest category and perform an analysis of means (Figure 20). The perfect category has the mean of 0.968^2 and a 95% confidence interval is (0.884^2, 1.052^2). This range is too narrow, since it simply estimates the mean and so does not account for the variation of individual disks. In order to set the specification limits for the mean peel strength, we can take the 95% lower and upper quantiles of the data with the average score of 4 – (0.886^2, 1.050^2), or (0.785, 1.103). This range of values overlaps with all but the extreme categories, and the issue is more serious at the weaker end of the peel strength distribution. It appears that a lower specification limit for the TB discs needs to be much higher than the lower specification limit of the MD discs of 0.4. However, the problem appears to be more complex with TB discs. That is, the problem with defective units cannot me managed by simply imposing tighter specification limits on peel strength.

CHAPTER V

Conclusions

An integrated approach to the design, analysis and optimization of generalized mixed linear models with FAMe experiments is presented in this dissertation. This analysis and optimization of FAMe experiments differs from JW's proposed approach in the following aspects:

- These models properly account for the split-unit structure of the data with the use of generalized linear mixed models.
- A model choice does not depend on a subsequent optimization step and is more flexible. Different link functions and interactions between an amplification and control factors may be considered.
- Variable selection is performed via an ad hoc procedure described in Chapter II. The advantage of this method is clear when the number of parameters under consideration is comparable to the available degrees of freedom, which is common in designed experiments. An intermediate step in our variable selection procedure with normal random effects as a response variable allows application of standard variable selection techniques with a linear normal response.
- Bayesian analysis in WinBUGS permits simultaneous model specification of two types of defects. Thus variation of the parameter estimates can be incorporated in quantities of interest, such as the expected loss function. Sensitivity of the optimum factor levels to the uncertainty of the parameter estimates may be assessed as well.

Data collection is crucial with FAMe experiments as with any categorical data. I have explored some of the design issues with an ad hoc sequential design with small n in Chapter III. There I have only considered fixed effects logistic models. Through this limited comparison of small sequential and fixed designs, it is evident that sequential designs that do not require estimation of parameters between runs can outperform fixed designs of similar size.

Spiess and Hamerle (2000) compare three different estimation techniques for the correlated binary response with a probit link with respect to small sample properties. These three methods (marginal ML estimation using Gauss-Hermite quadrature, GEE, and 'mean and covariance structure analysis' approach) do not include pseudo-likelihood or MCMC methods used for the analysis of FAMe examples. They study convergence, bias and efficiency of the estimation approaches via simulation. Similar study for other link functions and estimation methods can be an area for further research.

In the final chapter, I describe my experience in designing a failure amplification experiment with noise factors, and analyzing the resulting categorical and continuous data. This practical example illustrates the complexities of experimentation and data analysis.

LIST OF REFERENCES

AGRESTI, A. (2002). *Categorical Data Analysis.* John Wiley & Sons.

ATKINSON, A.C and DONEV, A.N. (1996) *Optimum Experimental Designs.* Clarendon Press, Oxford.

BISGAARD, S. and GERTBAKH, I. (2000). “2^{k-p} Experiments with Binary Responses: Inverse Binomial Sampling”. *Journal of Quality Technology* 32, pp. 148-156.

BRESLOW, N.E. and CLAYTON, D.G. (1993). “Approximate Inference in Generalized Linear Mixed Models.” *Journal of the American Statistical Association* 88, pp. 9-25.

BURNHAM, K.P. and ANDERSON, D.R.. (1998). *Model Selection and Inference.* Springer, New York.

CHIPMAN, H. and HAMADA, M. (1996). "Bayesian Analysis of Ordered Categorical Data from Industrial Experiments." *Technometrics* 38, pp. 1-10.

CHIPMAN, H.; HAMADA, M.; and WU, C.F.J. (1997). "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing." *Technometrics*. 39, pp. 372-381.

CLAUSING, D. (1994). *Total Quality Development: A step-by-step Guide to World Class Concurrent Engineering.* ASME Press, New York, NY.

CLAUSING, D. (2004). "Operating Window: An Engineering Measure for Robustness." *Technometrics* 46, pp. 25-29.

CONDRA, L.W. (2001). *Reliability Improvement with Design of Experiments.* Marcel Dekker, Inc.

CORDEIRO, G.M. and MCCULLAGH, P. (1991). “Bias Correction in Generalized Linear Models”. *Journal of the Royal Statistical Society* B 53, pp.629-643.

DIXON, J.W. and MOOD, A.M. (1948). “A Method for Obtaining and Analyzing Sensitivity Data”. *Journal of the American Statistical Association* 43, pp. 109-126.

DOBSON, A.J. (2001). *An Introduction to Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

GELMAN, A.; CARLIN, J.B.; STERN, H.S. and RUBIN, D.B. (2004). *Bayesian Data Analysis.* Chapman & Hall.

GEORGE, E.I. and MCCULLOCH, R.E. (1993). "Variable Selection Via Gibbs Sampling", *Journal of American Statistical Association* 88, pp. 881-889.

GRAYBILL, F.A. (1976). *Theory and Application of the Linear Model.* Duxbury Press

HAMADA, M. (1995a). "Using Statistically Designed Experiments to Improve Reliability and to Achieve Robust Reliability." *IEEE Transactions on Reliability* 44, pp. 206-215.

HAMADA, M. (1995b). "Analysis of Experiments for Reliability Improvements and Robust Reliability." Chapter 9 in *Recent Advances in Life-Tesing on Reliability*, R-44, pp. 206-215.

HAMADA, M. and WU, C.F.J. (1995). "Analysis of Censored Data from Fractionated Experiments: a Bayesian Approach." *Journal of the American Statistical Association* 90, pp. 467-477.

JOSEPH, V.R. (2004). "Quality Loss Functions for Nonnegative Variables and Their Applications." *Journal of Quality Technology* 36, pp. 129-138.

JOSEPH, V.R. and WU, C.F.J. (2002). “Operating Window Experiments: A Novel Approach to Quality Improvement.” *Journal of Quality Technology* 34, pp. 345-354.

JOSEPH, V.R. and WU, C.F.J. (2004). “Failure Amplification Method: An Information Maximization Approach to Categorical Response Optimization." *Technometrics* 46, pp. 1-12.

LEITNAKER, M.G. and MEE, R.W. (2004). Discussion of “Failure Amplification Method: An Information Maximization Approach to Categorical Response Optimization." *Technometrics* 46, pp. 12-14.

LEON, R.V.; SHOEMAKER, A.C.; and KACHKER, R.N. (1987). "Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-Noise Ratios" (with discussion). *Technometrics* 29, pp. 253-285.

LEON, R.V. and WU, C.F.J. (1992). "A Theory of Performance Measures in Parameter Design." *Statistica Sinica* 2, pp. 335-358.

LIANG, K.Y. and ZEGER, S.L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models”. *Biometrika* 73, pp.13-22.

LUNN D.J.; THOMAS A.; BEST N.; and SPIEGELHALTER D. (2000). "WinBUGS - A Bayesian Modeling Framework: Concepts, Structure, and Extensibility." *Statistics and Computing* 10, pp. 325-337.

MARITHI, B.N. and JOSEPH, V.R. (1999). "Improving the Yield of Printed Circuit Boards Using Design of Experiments." *Quality Engineering* 12, pp. 259-265.

MCCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models.* Chapman & Hall.

MCCULLOCH, C.E. and SEARLE, S.R. (2001). *Generalized, Linear, and Mixed Models.* John Wiley & Sons

MEEKER, W.Q. and ESCOBAR, L.A. (1993). "A review of Recent Research and Current Issues in Accelerated Testing." *International Statistical Review* 61, pp. 147-168.

MEEKER, W.Q. and ESCOBAR, L.A. (1998). *Statistical Methods for Reliability Data*, New York, Wiley.

MEEKER, W.Q. and ESCOBAR, L.A. (2004). "Reliability: The Other Dimension of Quality." *Quality Technology and Quantitative Management* 1, pp. 1-25.

MEEKER, W.Q. and HAMADA, M. (1995). "Statistical Tools for the Rapid Development and Evaluation of High-reliability Products." *IEEE Transactions on Reliability* 44, pp. 187-198.

MONTGOMERY, D.C. (2001). *Design and Analysis of Experiments.* John Wiley & Sons, Inc.

NELSON, W. (1982). *Applied Life Data Analysis.* John Wiley, NY.

NELSON, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* John Wiley & Sons.

PETTIGREW, M. J. (2003). "Composite Disc Speed Enhancement." Technical Report for Six Sigma Black Belt Certification. University of Tennessee, Knoxville, TN.

ROBBINS, H. and MUNRO, S. (1951). "A Stochastic Approximation Method". *Annals of Mathematical Statistics* 22, pp. 400-407.

ROBINSON, T.J.; MYERS, R.H.; AND MONTGOMERY, D.C. (2004). "Analysis Considerations in Industrial Split-Plot Experiments with Non-Normal Responses". *Journal of Quality Technology* 36, pp. 180-192.

SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P.; and LINDE, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society* B 64, pp. 583-639.

SPIESS, M. and HAMERLE, A. (2002). "A Comparison of Different Methods for the Estimation of Regression Models with Correlated Binary Response." *Computational Statistics & Data Analysis* 33, pp. 439-455.

STIRATELLI, R.; LAIRD, N.; AND WARE, J.H. (1984). "Random-effects Models for Serial Observations with Binary Response." *Biometrics* 40, pp. 961-971.

VOELKEL, J.G. (1999). "Sequential Experimental Designs for Sensitivity Experiments". Technical Report 1999-1, Center for Quality and Applied Statistics, Rochester Institute of Technology.

WOLFINGER, R. and O'CONNELL, M. (1993) "Generalized linear mixed models: a pseudo-likelihood approach." *Journal of Statistical Computation and Simulation* 48, pp. 233-243.

WU, C.F.J. (1985). "Efficient Sequential Designs with Binary Data". *Journal of the American Statistical Association* 80, pp. 974-984

WU, C.F.J. and HAMADA, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization.* John Wiley and Sons.

YE, K.Q. and HAMADA, M. (2000). "Critical values of the Lenth Method for Unreplicated Factorial Designs." *Journal of Quality Technology* 32, pp. 57-66.

ZEGER, S.L. and KARIM, M.R. (1991). "Generalized Linear Models with Random Effects – A Gibbs Sampling Approach." *Journal of American Statistical Association* 86, pp. 79-86.

APPENDICES

Appendix A: WinBUGS code for PCB data

MODEL

```
model {
        for (i in 1:N) {
        x1l[i] <- (2*x1[i]-3)/sqrt(2)
        x21l[i] <- x21[i]/sqrt(2)
        x22l[i] <- x22[i]/sqrt(2)
        x2l[i] <- (x2[i]-2)/sqrt(2)
        x3l[i] <- (x3[i]-2)/sqrt(2)
        x4l[i] <- (x4[i]-2)/sqrt(2)
        x5l[i] <- (x5[i]-2)/sqrt(2)
        x5q[i] <- (3*(x5[i]-2)*(x5[i]-2)-2)/sqrt(6)
        x6l[i] <- (x6[i]-2)/sqrt(2)
        x6q[i] <- (3*(x6[i]-2)*(x6[i]-2)-2)/sqrt(6)
        x7l[i] <- (x7[i]-2)/sqrt(2)
        x8l[i] <- (x8[i]-2)/sqrt(2)
                }
for (j in 1:K) {
        C[j] <- (log(M[j])-1.5664)/0.6765
                }
for (i in 1:N) {
        b1[i] ~ dnorm(0, tau1)
        b10[i] <- b1[i] - mean(b1[])
        b2[i] ~ dnorm(0, tau2)
        b20[i] <- b2[i] - mean(b2[])
        for (j in 1:K) {
# Opens
        S1[i,j] ~ dbin(p1[i,j], 160)
        cloglog(p1[i,j]) <- o[1] + o[2]*x2l[i] + o[3]*x3l[i] + o[4]*x4l[i]
                         + o[5]*x5l[i] + o[6]*x6l[i] + o[7]*x7l[i]
                        + o[8]*C[j]  + oq[1]*x6q[i]  + b10[i]
        Orep[i,j] ~ dbin(p1[i,j],160)
        or2[i,j] <- pow((S1[i,j]-160*p1[i,j]),2)
        or2rep[i,j] <- pow((Orep[i,j]-160*p1[i,j]),2)
# Shorts
        S2[i,j] ~ dbin(p2[i,j], 80)
        cloglog(p2[i,j]) <- s[1] + s[2]*x1l[i] + s[3]*x4l[i]+ s[4]*x6l[i]
                        + s[5]*C[j]  + b20[i]
        Srep[i,j] ~ dbin(p2[i,j],80)
        sr2[i,j] <- pow((S2[i,j]-80*p2[i,j]),2)
        sr2rep[i,j] <- pow((Srep[i,j]-80*p2[i,j]),2)
                        }
        orr[i] <- mean(or2[i,])
        orrep[i] <- mean(or2rep[i,])
```

```
		srr[i] <- mean(sr2[i,])
		srrep[i] <- mean(sr2rep[i,])
				}
# Optimization
		for (i in 1:1460) {
			x1lo[i] <- FF7[i,1]/sqrt(2)
			x2lo[i] <- FF7[i,2]/sqrt(2)
			x3lo[i] <- FF7[i,3]/sqrt(2)
			x4lo[i] <- FF7[i,4]/sqrt(2)
			x5lo[i] <- FF7[i,5]/sqrt(2)
			x6lo[i] <- FF7[i,6]/sqrt(2)
			x7lo[i] <- FF7[i,7]/sqrt(2)
			x6qo[i] <- (3*pow(FF7[i,6],2)-2)/sqrt(6)
					}
		for (i in 1:1460) {
		bb1[i] ~ dnorm(0, tau1)
		bb2[i] ~ dnorm(0, tau2)
		for (j in 1:K) {
			cloglog(pp1[i,j]) <- o[1] + o[2]*x2lo[i] + o[3]*x3lo[i] + o[4]*x4lo[i]
					+ o[5]*x5lo[i] + o[6]*x6lo[i] + o[7]*x7lo[i]
					+ o[8]*C[j]  + oq[1]*x6qo[i]  + bb1[i]
			cloglog(pp2[i,j]) <- s[1] + s[2]*x1lo[i] + s[3]*x4lo[i] + s[4]*x6lo[i]
					 + s[5]*C[j]  + bb2[i]
				}
		EL[i] <- (pp1[i,3]+pp1[i,4]+pp1[i,5]+pp2[i,3]+pp2[i,4]+pp2[i,5])/3
					}
# Model checking
		Tyo <- sqrt(mean(orr[]))
		Tyrepo <- sqrt(mean(orrep[]))
		Tys <- sqrt(mean(srr[]))
		Tyreps <- sqrt(mean(srrep[]))
# Priors
		for (i in 1:8) { o[i] ~ dnorm(0, 1.0E-6)}
		for (i in 1:1) { oq[i] ~ dnorm(0, 1.0E-6)}
		for (i in 1:5) { s[i] ~ dnorm(0, 1.0E-6)}
		tau1 ~ dgamma(0.001, 0.001)
		sigma1 <-1 / sqrt(tau1)
		tau2 ~ dgamma(0.001, 0.001)
		sigma2 <- 1 / sqrt(tau2)
}
```

DATA

```
list(
M=c(3,4,5,6,7),
x1=c(1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2),
x21=c(1,1,1,0,0,0,-1,-1,-1,1,1,1,0,0,0,-1,-1,-1),
 x22=c(0,0,0,1,1,1,-1,-1,-1,0,0,0,1,1,1,-1,-1,-1),
x2 = c(1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3),
x3=c(1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3),
x4=c(1,2,3,1,2,3,2,3,1,3,1,2,2,3,1,3,1,2),
x5=c(1,2,3,2,3,1,1,2,3,3,1,2,3,1,2,2,3,1),
x6=c(1,2,3,2,3,1,3,1,2,2,3,1,1,2,3,3,1,2),
x7=c(1,2,3,3,1,2,2,3,1,2,3,1,3,1,2,1,2,3),
x8=c(1,2,3,3,1,2,3,1,2,1,2,3,2,3,1,2,3,1),

N=18,
K=5,
S1 = structure(
        .Data = c(
                33,7,4,0,1,
                7,9,1,0,0,
                ...
                12,2,0,0,1),
        .Dim = c(18, 5)),
S2 = structure(
        .Data = c(
                1,0,0,0,0,
                ...
                7,2,0,0,0),
                .Dim = c(18, 5)),
FF7=structure(
        .Data = c(
        -1,-1,-1,-1,-1,-1,-1,
        ...
        1,1,1,1,1,1,1,
        -1,1,-1, -1,0.34,-0.43,1,
        -1,1,0,-1,0.34,-0.43,0),
        .Dim = c(1460, 7)))
```

INITIALS

```
list(
o =c( 0,0,0,0,0,0,0,0),
oq=c(0),
tau1 = 1,
b1 = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
s =c( 0,0,0,0,0),
tau2 =1,
b2 = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
Orep=structure(
	.Data = c(
	0,0,0,0,0,
	...
	0,0,0,0,0),
	.Dim = c(18, 5)),
Srep=structure(
	.Data = c(
	0,0,0,0,0,
	...
	0,0,0,0,0),
	.Dim = c(18, 5)),
bb1=c(0,0,...,0),
bb2=c(0,0,...,0)
)
```

Appendix B: Tables

Table 1. Factors and levels for the PCB experiment.

Control factors	Notation	Levels		
		1	2	3
Preheat	X_1	No	Yes	-
Surface preparation	X_2	Scrub	Pumice	Chemical
Lamination speed	X_3	1.2 mpm	1.5 mpm	1.8 mpm
Lamination pressure	X_4	20 psi	40 psi	60 psi
Lamination temperature	X_5	95 ^{0}C	105 ^{0}C	115 ^{0}C
Exposure energy	X_6 (m)	14	17	20
Developer speed	X_7	3 fpm	4 fpm	5 fpm
ORP	X_8	500	530	560

Table 2. OA(18, $2^1 \times 3^7$) and data from PCB experiment.

									Opens					Shorts				
									Line width C_1					Spacing C_2				
Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	3	4	5	6	7	3	4	5	6	7
1	1	1	1	1	1	1	1	1	1	0	0	0	0	33	7	4	0	1
2	1	1	2	2	2	2	2	2	4	1	0	0	0	7	9	1	0	0
3	1	1	3	3	3	3	3	3	19	2	0	0	0	14	3	1	0	0
4	1	2	1	1	2	2	3	3	9	0	0	0	0	2	0	2	0	0
5	1	2	2	2	3	3	1	1	22	1	1	1	0	7	1	2	1	0
6	1	2	3	3	1	1	2	2	8	0	0	0	0	78	30	7	1	1
7	1	3	1	2	1	3	2	3	19	1	0	0	0	9	1	3	0	0
8	1	3	2	3	2	1	3	1	4	0	1	0	0	7	0	1	0	1
9	1	3	3	1	3	2	1	2	7	0	0	0	0	4	3	0	0	0
10	2	1	1	3	3	2	2	1	22	1	0	0	1	6	0	0	0	0
11	2	1	2	1	1	3	3	2	34	2	2	0	0	13	2	0	0	0
12	2	1	3	2	2	1	1	3	13	4	1	0	0	34	5	0	1	3
13	2	2	1	2	3	1	3	2	7	0	1	0	0	8	3	0	0	0
14	2	2	2	3	1	2	1	3	25	1	0	0	0	25	8	0	2	1
15	2	2	3	1	2	3	2	1	41	1	0	0	1	7	0	0	0	0
16	2	3	1	3	2	3	1	2	45	9	5	0	1	10	6	0	0	0
17	2	3	2	1	3	1	2	3	3	0	0	0	0	8	0	0	0	0
18	2	3	3	2	1	2	3	1	7	2	0	0	0	12	2	0	0	1

Table 3. Covariance pattern analysis for opens using GEE.

Parameter	Estimate	Standard Error	Lower	Upper	Z	Pr > |Z|
Intercept	-5.040	0.112	-5.259	-4.820	-45.04	<.0001
x_{2l}	-0.445	0.134	-0.707	-0.183	-3.33	0.0009
x_{3l}	0.436	0.137	0.167	0.705	3.18	0.0015
x_{4l}	0.689	0.121	0.451	0.927	5.68	<.0001
x_{5l}	-1.045	0.147	-1.333	-0.758	-7.12	<.0001
x_{6l}	-0.498	0.136	-0.766	-0.231	-3.65	0.0003
x_{7l}	-0.387	0.147	-0.675	-0.100	-2.64	0.0083
x_{6q}	0.495	0.100	0.299	0.691	4.95	<.0001
$\ln C_1$	-3.420	0.120	-3.654	-3.185	-28.56	<.0001

Table 4. Conditional GLMM analysis for opens.

Parameter	Estimate	Standard Error	95% C.I.		t-value	Pr > |t|
			Lower	Upper		
(a) Fixed effects estimates						
Intercept	-5.014	0.140	-5.326	-4.703	-35.84	<.0001
x_{2l}	-0.441	0.177	-0.794	-0.088	-2.49	0.015
x_{3l}	0.426	0.178	0.071	0.780	2.39	0.0193
x_{4l}	0.561	0.179	0.204	0.918	3.13	0.0025
x_{5l}	-0.856	0.177	-1.208	-0.504	-4.85	<.0001
x_{6l}	-0.477	0.173	-0.821	-0.133	-2.76	0.0073
x_{7l}	-0.458	0.177	-0.811	-0.105	-2.59	0.0117
x_{6q}	0.431	0.181	0.069	0.792	2.38	0.0202
$\ln C_1$	-3.436	0.173	-3.781	-3.092	-19.88	<.0001
(b) Random effect predictions						
Run1	0.068	0.279	-0.488	0.625	0.24	0.8075
Run2	0.090	0.246	-0.400	0.580	0.37	0.7146
Run3	0.165	0.286	-0.405	0.736	0.58	0.5652
Run4	0.003	0.293	-0.581	0.587	0.01	0.9926
Run5	0.056	0.265	-0.473	0.585	0.21	0.8327
Run6	0.406	0.248	-0.088	0.900	1.64	0.1057
Run7	0.001	0.268	-0.533	0.534	0.00	0.9977
Run8	-0.446	0.268	-0.981	0.089	-1.66	0.1010
Run9	0.102	0.298	-0.492	0.696	0.34	0.7332
Run10	-0.171	0.287	-0.744	0.402	-0.6	0.5537
Run11	-0.032	0.280	-0.591	0.527	-0.11	0.9097
Run12	-0.295	0.249	-0.792	0.201	-1.19	0.2398
Run13	0.201	0.279	-0.356	0.757	0.72	0.4746
Run14	-0.063	0.257	-0.575	0.448	-0.25	0.8055
Run15	-0.253	0.269	-0.789	0.283	-0.94	0.3492
Run16	0.063	0.278	-0.491	0.616	0.23	0.8214
Run17	0.066	0.282	-0.497	0.629	0.23	0.8152
Run18	0.040	0.278	-0.515	0.595	0.14	0.8873

Table 5. Bayesian analysis for opens.

Parameter	Mean	Standard Deviation	Posterior Interval	
			2.50%	97.50%
(a) Fixed effect estimates				
Intercept	-5.034	0.117	-5.266	-4.809
x_{2l}	-0.446	0.187	-0.816	-0.069
x_{3l}	0.429	0.188	0.053	0.802
x_{4l}	0.572	0.193	0.175	0.942
x_{5l}	-0.877	0.194	-1.246	-0.479
x_{6l}	-0.476	0.186	-0.835	-0.093
x_{7l}	-0.458	0.195	-0.852	-0.064
x_{6q}	0.441	0.195	0.049	0.822
$\ln C_1$	-3.447	0.175	-3.794	-3.113
(b) Random effect predictions				
Run1	0.074	0.298	-0.515	0.687
Run2	0.097	0.246	-0.369	0.616
Run3	0.163	0.303	-0.408	0.810
Run4	-0.005	0.301	-0.614	0.585
Run5	0.052	0.269	-0.487	0.597
Run6	0.400	0.266	-0.068	0.964
Run7	-0.006	0.275	-0.558	0.535
Run8	-0.437	0.299	-1.098	0.081
Run9	0.097	0.311	-0.505	0.746
Run10	-0.173	0.304	-0.826	0.390
Run11	-0.034	0.285	-0.622	0.528
Run12	-0.274	0.256	-0.800	0.212
Run13	0.193	0.293	-0.342	0.816
Run14	-0.055	0.262	-0.578	0.466
Run15	-0.255	0.289	-0.888	0.246
Run16	0.058	0.293	-0.514	0.651
Run17	0.064	0.288	-0.507	0.656
Run18	0.041	0.291	-0.519	0.636

Table 6. Expected loss statistics and factor levels for the PCB data.

Expected Loss	x1	x2	x3	x4	x5	x6	x7	mean	sd	2.50%	median	97.50%	Minimum Value
EL[510]	1	3	1	1	3	2	3	0.00102	0.00042	0.00045	0.00094	0.00198	mean, 97.5%
EL[509]	1	3	1	1	3	2	2	0.00114	0.00045	0.00052	0.00106	0.00218	
EL[591]	1	3	2	1	3	2	3	0.00113	0.00047	0.00051	0.00105	0.00219	
EL[267]	1	2	1	1	3	2	3	0.00114	0.00045	0.00051	0.00106	0.00222	
EL[1459]	1	3	1	2	2.34	1.57	3	0.00112	0.00051	0.00051	0.00102	0.00235	JW settings
EL[507]	1	3	1	1	3	1	3	0.00104	0.00057	0.00040	0.00092	0.00243	median, 2.5%
…	…	…	…	…	…	…	…	…	…	…	…	…	
EL[1460]	1	3	2	2	2.34	1.57	2	0.00164	0.00081	0.00071	0.00148	0.00346	JW settings
…	…	…	…	…	…	…	…	…	…	…	…	…	
EL[460]	1	2	3	3	1	1	1	0.02774	0.01630	0.00831	0.02427	0.06778	
EL[218]	1	1	3	3	1	1	2	0.02769	0.01669	0.00832	0.02415	0.06883	
EL[1189]	2	2	3	3	1	1	1	0.02869	0.01661	0.00921	0.02516	0.06976	
EL[947]	2	1	3	3	1	1	2	0.02851	0.01745	0.00904	0.02483	0.07089	
EL[136]	1	1	2	3	1	1	1	0.02852	0.01753	0.00849	0.02478	0.07129	
EL[865]	2	1	2	3	1	1	1	0.02906	0.01685	0.00917	0.02534	0.07144	
EL[217]	1	1	3	3	1	1	1	0.03823	0.02317	0.01126	0.03336	0.09491	
EL[946]	2	1	3	3	1	1	1	0.03905	0.02356	0.01217	0.03378	0.09844	

Table 7. DIC values for mixed and fixed models for the PCB data. Models without the random effects have the same set of fixed effects as mixed models. JW models have a different set of fixed effects.

	Response	Dbar	Dhat	pD	DIC
Mixed models					
	Opens	274.996	259.693	15.303	290.299
	Shorts	208.549	195.751	12.798	221.347
	Total	483.545	455.444	28.101	511.646
Models w/o random effects					
	Opens	297.600	288.626	8.974	306.574
	Shorts	223.134	218.124	5.010	228.144
	Total	520.735	506.751	13.984	534.719
JW models					
	Opens	367.891	362.174	5.717	373.607
	Shorts	216.495	210.700	5.795	222.290
	Total	584.385	572.874	11.512	595.897

Table 8. Factors and levels for the paper feeder experiment.

Control factors	Notation	Levels		
		1	2	3
Feed belt material	X_1	Type A	Type B	
Speed	X_2	288 mm/s	240 mm/s	192 mm/s
Drop height	X_3	3 mm	2 mm	1 mm
Center roll	X_4	Absent	Present	-
Belt width	X_5	10 mm	20 mm	30 mm
Tray guidance angle	X_6	0	14	28
Tip angle	X_7	0	3.5	7
Turf	X_8	None	1 sheet	2 sheets
Noise factor				
Stack quantity	N	High	Low	-

Table 9. Data from the paper feeder experiment.

		Misfeed												Multifeed											
Run		N_1						N_2						N_1						N_2					
1	*M*	20	40	42.5	45	50	60	20	30	40	50	60	70	80	82.5	85	90	120	160	60	62.5	65	70	80	90
	# failures	5	5	1	0	0	0	5	0	0	0	0	0	0	2	2	2	2	2	0	1	1	3	2	3
2	*M*	0	10	15	20	30	40	0	10	15	20	40	60	30	35	40	50	60		30	40	60	70	75	80
	# failures	5	3	0	0	0	0	5	3	0	0	0	0	0	1	3	3	3		0	1	1	1	2	2
3	*M*	0	10	15	20	25		0	10	15	20	30	40	20	25	30	40			20	30	35	40	50	
	# failures	5	5	1	1	0		5	3	2	0	0	0	0	2	2	2			0	1	1	3	3	
4	*M*	20	25	30	40	60		0	20	25	30	40		50	60	65	70	80		40	50	55	60		
	# failures	5	3	1	0	0		5	5	1	0	0		0	1	2	2	2		0	0	2	2		
5	*M*	20	25	30	40	50		20	25	30	40	50		30	40	45	50	60		40	50	55	60		
	# failures	4	1	0	0	0		4	1	0	0	0		0	1	3	3	3		0	0	2	2		
6	*M*	10	15	20	30	40		10	15	20	30	40		30	40	45	50			30	40	50	55	60	
	# failures	4	2	1	0	0		3	0	0	0	0		0	1	2	3			0	1	2	2	3	
7	*M*	10	20	30	35	40		10	20	25	30	40		20	30	35	40	50		20	30	40	60	70	80
	# failures	5	4	2	1	0		5	3	0	0	0		0	1	2	2	3		0	1	1	1	2	2
8	*M*	15	20	30	35	40		20	30	35	40	60	70	70	80	100	110	120		60	70	75	80	100	
	# failures	3	2	2	3	0		5	2	4	1	1	0	0	1	1	2	2		0	1	2	2	2	
9	*M*	10	15	20	30	40		10	15	20	25	30	40	40	60	65	70	80		40	50	55	60	70	
	# failures	5	4	1	0	0		5	5	5	4	0	0	0	1	1	2	3		0	0	1	2	3	
10	*M*	0	5	10	15	20		0	5	10	15	20		5	10	15	20	30		0	5	15	20	30	
	# failures	5	1	0	0	0		5	0	0	0	0		0	1	1	3	3		0	1	0	2	2	
11	*M*	0	5	10	15	20		0	5	10	15	20		5	10	15	20	30		0	5	10	15	20	30
	# failures	5	2	0	0	0		5	1	0	0	0		0	1	1	4	3		0	1	1	0	2	2
12	*M*	0	10	15	20			0	10	15	20			30	40	50	55	60		40	45	50			
	# failures	5	4	0	0			5	4	0	0			0	1	1	2	5		0	0	1			
13	*M*	0	10	15	20			0	10	15	20			30	40	80	85	90	100	55	60				
	# failures	5	5	1	0			5	4	2	0			0	1	1	4	3	2	0	2				
14	*M*	10	20	25	30	35	40	10	20	25	30	35	40	20	30	35	40	45		20	25	30	35	40	50
	# failures	5	3	2	2	0	0	5	4	0	0	0	0	0	0	0	2	2		0	0	1	1	4	3
15	*M*	0	5	10	15			0	5	10	15	20		20	30	35	40	50		10	15	20	30		
	# failures	1	0	0	0			4	0	0	0	0		0	1	3	2	4		0	1	4	4		
16	*M*	5	10	20	30	35		5	10	20	30	40		20	30	35	40	50	60	30	40	50	55	60	
	# failures	5	1	0	0	0		5	1	0	0	0		0	1	0	2	3	5	0	1	1	2	2	
17	*M*	10	20	30	40	45	50	10	20	25	30	40		80	90	95	100			100	105	110			
	# failures	5	4	5	2	0	0	5	3	0	0	0		0	1	1	1			0	1	1			
18	*M*	10	15	20	30			10	20	30	35	40		60	65	70	80	90	120	60	70	75	80	90	
	# failures	5	5	1	0			5	5	5	0	0		0	1	2	2	2	3	0	1	2	2	3	

Table 10. Posterior summaries for the final models with the paper feeder data.

	Coefficient	Mean	St.Dev.	2.50%	median	97.50%
Misfeeds						
	α_0	-1.651	0.230	-2.121	-1.645	-1.213
	α_1	-1.926	0.739	-3.783	-1.901	-0.530
	α_2	2.027	0.732	0.526	2.030	3.391
	α_3	0.049	0.827	-1.584	0.020	1.773
	α_4	0.517	0.669	-0.795	0.539	1.882
	α_{46}	4.415	1.515	1.390	4.428	7.336
	α_6	-3.654	0.868	-5.268	-3.633	-1.922
	α_7	0.190	0.793	-1.342	0.196	1.881
	α_{7q}	2.735	0.834	1.131	2.717	4.496
	α_8	0.901	0.858	-0.770	0.923	2.452
	M	-4.366	0.379	-5.106	-4.349	-3.693
	α_{2M}	1.695	0.558	0.669	1.673	2.848
	NM	-0.280	0.117	-0.505	-0.280	-0.050
	N	-0.176	0.077	-0.327	-0.175	-0.031
	$\alpha_2 N$	0.542	0.145	0.266	0.542	0.831
	$\alpha_3 N$	0.366	0.128	0.118	0.365	0.618
	$\alpha_8 N$	-0.425	0.120	-0.657	-0.427	-0.192
	σ_1	1.720	0.541	0.981	1.616	3.083
	$u_1[1]$	0.642	1.542	-2.468	0.634	3.728
	$u_1[2]$	0.608	0.965	-1.438	0.628	2.548
	$u_1[3]$	-1.502	1.482	-4.635	-1.467	1.413
	$u_1[4]$	0.089	1.234	-2.338	0.097	2.693
	$u_1[5]$	0.896	1.124	-1.348	0.886	3.169
	$u_1[6]$	0.122	1.180	-2.369	0.171	2.435
	$u_1[7]$	0.944	1.338	-1.626	0.927	3.688
	$u_1[8]$	-0.902	1.290	-3.479	-0.910	1.679
	$u_1[9]$	-1.165	1.268	-3.705	-1.101	1.138
	$u_1[10]$	-1.118	1.250	-3.796	-1.067	1.231
	$u_1[11]$	0.738	1.284	-1.860	0.749	3.320
	$u_1[12]$	-0.512	1.492	-3.719	-0.481	2.407
	$u_1[13]$	-0.224	1.084	-2.376	-0.242	2.027
	$u_1[14]$	1.765	1.036	-0.202	1.732	3.958
	$u_1[15]$	-0.505	1.385	-3.385	-0.479	2.169
	$u_1[16]$	-1.691	1.216	-4.059	-1.724	0.881
	$u_1[17]$	0.037	1.674	-3.184	-0.010	3.241
	$u_1[18]$	1.779	1.291	-0.716	1.728	4.462

Table 10 continued. Multifeeds.

	Coefficient	Mean	St.Dev.	2.50%	median	97.50%
Multifeeds						
	α_0	-0.770	0.068	-0.905	-0.769	-0.637
	α_2	-0.875	0.242	-1.386	-0.864	-0.432
	α_3	0.004	0.229	-0.468	0.007	0.444
	α_6	0.903	0.241	0.444	0.899	1.391
	M	1.558	0.189	1.191	1.555	1.926
	$\alpha_2 M$	0.509	0.239	0.057	0.510	0.984
	$\alpha_3 M$	0.530	0.254	0.044	0.525	1.033
	N	-0.072	0.048	-0.166	-0.072	0.026
	$\alpha_3 N$	-0.211	0.085	-0.379	-0.210	-0.044
	σ_2	0.494	0.134	0.277	0.480	0.804
	$u_2[1]$	-0.185	0.312	-0.822	-0.177	0.436
	$u_2[2]$	-0.484	0.233	-0.964	-0.473	-0.043
	$u_2[3]$	-0.467	0.309	-1.098	-0.461	0.113
	$u_2[4]$	-0.273	0.234	-0.752	-0.268	0.173
	$u_2[5]$	-0.473	0.240	-0.970	-0.465	-0.025
	$u_2[6]$	0.682	0.293	0.134	0.670	1.298
	$u_2[7]$	0.129	0.301	-0.462	0.126	0.730
	$u_2[8]$	0.013	0.273	-0.519	0.012	0.557
	$u_2[9]$	-0.028	0.264	-0.555	-0.023	0.499
	$u_2[10]$	0.298	0.321	-0.334	0.298	0.933
	$u_2[11]$	0.178	0.286	-0.374	0.172	0.759
	$u_2[12]$	-0.017	0.306	-0.620	-0.014	0.589
	$u_2[13]$	0.287	0.277	-0.238	0.276	0.862
	$u_2[14]$	0.321	0.195	-0.065	0.322	0.709
	$u_2[15]$	0.711	0.301	0.157	0.696	1.334
	$u_2[16]$	0.202	0.297	-0.369	0.195	0.809
	$u_2[17]$	-0.488	0.300	-1.109	-0.478	0.066
	$u_2[18]$	-0.406	0.279	-0.983	-0.397	0.121

Table 11. DIC values for mixed and fixed models for the paper feeder data. Models without the random effects have the same set of fixed effects as mixed models. JW models have a different set of fixed effects.

	Response	Dbar	Dhat	pD	DIC
Mixed models					
	Misfeeds	276.2	251.5	24.8	301.0
	Multifeeds	430.5	409.5	21.1	451.6
	Total	706.8	660.9	45.8	752.6
Models w/o random effects					
	Misfeeds	401.4	384.5	16.9	418.2
	Multifeeds	477.0	467.8	9.2	486.2
	Total	878.3	852.3	26.1	904.4
JW models					
	Misfeeds	334.8	323.9	10.9	345.7
	Multifeeds	437.9	427.9	10.0	447.9
	Total	772.7	751.8	20.9	793.6

Table 12. Frequency of defects for the paper feeder data.

Number of defects	% data	
	Misfeed	Multifeed
0	0.50	0.26
1	0.10	0.27
2	0.06	0.28
3	0.05	0.14
4	0.07	0.03
5	0.21	0.01

Table 13. Expected loss statistics and factor levels for the paper feeder data.

Order	mean	sd	2.5%	median	97.50%	x1	x2	x3	x4	x6	x7	x8	M	Minimum value
EL[886]	0.010	0.064	0.000	0.00041	0.0532	2	3	2	1	3	2	1	10	Mean, 97.5%
EL[940]	0.023	0.111	0.000	0.00003	0.2987	2	3	3	1	3	2	1	7.5	Median
EL[697]	0.160	0.223	0.003	0.08355	0.9807	2	2	1	2	3	2	1	12.5	JW

Table 14. Posterior summaries for mixed models with split-unit effects in (6) with the paper feeder data.

	Coefficient	Mean	Std.Dev.	2.5%	Median	97.5%
Misfeeds						
	$\alpha_{1,0}$	-0.962	0.084	-1.137	-0.962	-0.796
	$\alpha_{1,M}$	-3.115	0.223	-3.558	-3.115	-2.683
	$\alpha_{1,MN}$	-0.159	0.094	-0.339	-0.160	0.024
	$\alpha_{1,N}$	-0.059	0.064	-0.188	-0.059	0.067
	σ_1	2.761	0.557	1.890	2.689	4.043
	$u_1[1]$	2.663	0.306	2.069	2.656	3.270
	$u_1[2]$	-0.856	0.307	-1.464	-0.850	-0.293
	$u_1[3]$	0.005	0.255	-0.503	0.006	0.504
	$u_1[4]$	1.945	0.264	1.433	1.945	2.466
	$u_1[5]$	1.458	0.269	0.925	1.458	1.983
	$u_1[6]$	-0.231	0.264	-0.755	-0.227	0.283
	$u_1[7]$	1.647	0.270	1.132	1.645	2.177
	$u_1[8]$	2.521	0.258	2.031	2.514	3.050
	$u_1[9]$	1.696	0.242	1.240	1.691	2.176
	$u_1[10]$	-4.098	0.525	-5.156	-4.089	-3.093
	$u_1[11]$	-3.439	0.444	-4.345	-3.438	-2.600
	$u_1[12]$	-0.555	0.291	-1.140	-0.550	0.009
	$u_1[13]$	0.008	0.273	-0.535	0.016	0.525
	$u_1[14]$	1.533	0.250	1.051	1.537	2.026
	$u_1[15]$	-7.089	0.657	-8.416	-7.083	-5.846
	$u_1[16]$	-1.358	0.354	-2.044	-1.351	-0.684
	$u_1[17]$	2.086	0.270	1.558	2.081	2.630
	$u_1[18]$	2.063	0.273	1.531	2.062	2.597
Multifeeds						
	$\alpha_{2,0}$	-0.656	0.050	-0.755	-0.656	-0.558
	$\alpha_{2,\mathrm{M}}$	1.178	0.170	0.845	1.177	1.511
	σ_2	0.702	0.172	0.427	0.682	1.093
	$u_2[1]$	-0.559	0.195	-0.957	-0.553	-0.183
	$u_2[2]$	-0.054	0.171	-0.391	-0.053	0.274
	$u_2[3]$	0.453	0.201	0.063	0.452	0.843
	$u_2[4]$	-0.396	0.197	-0.795	-0.387	-0.020
	$u_2[5]$	0.019	0.183	-0.340	0.021	0.385
	$u_2[6]$	0.093	0.187	-0.276	0.089	0.457
	$u_2[7]$	0.053	0.178	-0.301	0.054	0.416
	$u_2[8]$	-0.774	0.215	-1.196	-0.769	-0.359
	$u_2[9]$	-0.362	0.191	-0.730	-0.361	0.015
	$u_2[10]$	1.138	0.274	0.592	1.140	1.661
	$u_2[11]$	1.239	0.280	0.673	1.237	1.787
	$u_2[12]$	-0.130	0.206	-0.545	-0.125	0.273
	$u_2[13]$	-0.297	0.208	-0.716	-0.297	0.107
	$u_2[14]$	0.178	0.188	-0.191	0.173	0.555
	$u_2[15]$	0.949	0.217	0.522	0.952	1.384
	$u_2[16]$	0.139	0.176	-0.207	0.142	0.475
	$u_2[17]$	-1.222	0.279	-1.775	-1.219	-0.674
	$u_2[18]$	-0.467	0.197	-0.859	-0.464	-0.090

Table 15. Fixed effects tests with the posterior means of the random effects as a response variable for the paper feeder data.

	Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Misfeeds	X_1	1	1	12.19	8.08	0.017
	X_2	1	1	11.30	7.50	0.021
	X_4	1	1	1.16	0.77	0.401
	X_6	1	1	32.49	21.55	0.001
	X_4*X_6	1	1	18.57	12.32	0.006
	X_7	1	1	0.99	0.66	0.437
	X_7*X_7	1	1	18.19	12.07	0.006
Multifeeds						
	X_2	1	1	1.85	11.60	0.004
	X_6	1	1	2.74	17.14	0.001

Table 16. Bias, variance and MSE for the fixed two-level design with β_0 = 3, β_1 =2, n=10, $x_c = -\beta_0/\beta_1$ and $d\beta_1$ from 0.1 to 1.9. The approximate bias B_a = 10% for all rows.

$d\beta_1$	Pr(MLE)	Bias*, %	Variance*		MSE*		Asymptotic Variance *		B_f, %
			β_0	β_1	β_0	β_1	β_0	β_1	
0.1	1.00	10%	225.45	100.09	225.54	100.13	207.30	92.03	10%
0.3	0.99	10%	25.47	11.21	25.56	11.25	23.91	10.52	10%
0.5	0.98	9%	9.42	4.07	9.49	4.10	9.24	3.99	10%
0.7	0.96	7%	4.94	2.08	4.99	2.10	5.19	2.19	11%
0.9	0.94	5%	3.04	1.24	3.07	1.25	3.52	1.44	11%
1.1	0.89	3%	2.04	0.80	2.05	0.80	2.67	1.05	12%
1.3	0.83	-1%	1.44	0.54	1.44	0.54	2.17	0.81	13%
1.5	0.75	-4%	1.04	0.37	1.06	0.38	1.85	0.66	14%
1.7	0.66	-8%	0.77	0.26	0.83	0.29	1.62	0.55	16%
1.9	0.57	-12%	0.58	0.18	0.71	0.24	1.46	0.46	17%

* - conditional on existence of MLE's

Table 17. An ad hoc rule for a sequential design (with positive slope) and $n = 5$ at each of four levels of x.

Y_1		x_2	Y_2		x_3	Y_3		x_4
0		x1+d	0		x2+2d	0		x3+2d
						1		x3+2d
						2		x3+2d
						3		x3+d
						4		x3+d
						5		x3-d
			1		x2+d	0		x3+2d
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			2		x2+d	0		x3+2d
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			3		x2+d	0		x3+d
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			4		x2+d	0		x3+d
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			5		x2-d/2	0		x3+d/4
						1		x3
						2		x3
						3		x3
						4		x3
						5		x3-d/4

Table 17. Continued.

Y_1		x_2	Y_2		x_3	Y_3		x_4
1		x1+d	0		x2+2d	0		x3+2d
						1		x3+2d
						2		x3+2d
						3		x3+d
						4		x3+d
						5		(x2+x3)/2
			1		x2+d	0		x3+2d
						1		x3+2d
						2		x3+2d
						3		x3+d
						4		x3+d
						5		x2
			2		x2+d	0		x3+2d
						1		x3+2d
						2		x3+2d
						3		x3+d
						4		x3+d
						5		x2
			3		x2+d	0		x3
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			4		x2+d	0		x3
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		x2
			5		(x1+x2)/2	0		(x2+x3)/2
						1		(x2+x3)/2
						2		(x2+x3)/2
						3		(x1+x3)/2
						4		(x1+x3)/2
						5		(x1+x3)/2

Table 17. Continued.

Y_1		x_2	Y_2		x_3	Y_3		x_4
2		x1+d	0		x2+d	0		x3+2d
						1		x3+d
						2		x3+d
						3		x3+d
						4		x3+d
						5		(x2+x3)/2
			1		x2+d	0		x3+d
						1		x3+d
						2		x3+d
						3		x1
						4		x1
						5		x2
			2		x2+d	0		x3
						1		x3+d
						2		x3+d
						3		x1
						4		x1
						5		x1
			3		x1-d	0		x2+d
						1		x3+d
						2		x3+d
						3		x1
						4		x1
						5		x1
			4		x1-d	0		x2+d
						1		x3-d
						2		x3-d
						3		x3
						4		x3
						5		x3
			5		x1-d	0		x1
						1		x3-d
						2		x3-d
						3		x3
						4		x3
						5		x3

Table 17. Continued.

Y_1		x_2	Y_2		x_3	Y_3		x_4
3		x1-d	0		x1+d	0		x1
						1		x3
						2		x3
						3		x1
						4		x1
						5		(x1+x2)/2
			1		x1+d	0		x3
						1		x3
						2		x3
						3		x3+d
						4		x3+d
						5		x2-d
			2		x2-d	0		x1+d
						1		x1+d
						2		x1+d
						3		x3-d
						4		x3-d
						5		x3
			3		x2-d	0		x1+d
						1		x3-d
						2		x3-d
						3		x3-d
						4		x3-d
						5		x3
			4		x2-d	0		x2
						1		x2
						2		x2
						3		x3-d
						4		x3-d
						5		x3
			5		x2-d	0		x2
						1		x2
						2		x2
						3		x3-d
						4		x3-d
						5		x3-2d

Table 17. Continued.

Y_1		x_2	Y_2		x_3	Y_3		x_4
4		x1-d	0		x2+.5d	0		(x1+x3)/2
						1		x1+d/2
						2		x1+d/2
						3		(x2+x3)/2
						4		(x2+x3)/2
						5		
			1		x2	0		(x1+x2)/2
						1		x2-d
						2		x2-d
						3		x2-d
						4		x2-d
						5		x2-2d
			2		x2-d	0		x2
						1		x3-d
						2		x3-d
						3		x3-d
						4		x3-d
						5		x3-2d
			3		x2-d	0		x1
						1		x3-d
						2		x3-d
						3		x3-d
						4		x3-d
						5		x3-2d
			4		x2-2d	0		x3+d
						1		x3-d
						2		x3-d
						3		x3-2d
						4		x3-2d
						5		x3-2d
			5		x2-2d	0		x2-d
						1		x2-d
						2		x2-d
						3		x3-d
						4		x3-d
						5		x3-2d

Table 17. Continued.

Y_1		x_2	Y_2		x_3	Y_3		x_4
5		x1-d	0		x1-d/2	0		x1-d/4
						1		x1-d/4
						2		x1-d/4
						3		(x2+x3)/2
						4		(x2+x3)/2
						5		(x2+x3)/2
			1		x1-d/2	0		x2
						1		(x1+x3)/2
						2		(x1+x3)/2
						3		x2-d
						4		x2-d
						5		(x2+x3)/2
			2		x2-d	0		x1-d/2
						1		x2
						2		x2
						3		x2
						4		x2
						5		x2
			3		x2-d	0		x3+d/2
						1		x3-d
						2		x3-d
						3		x3-d
						4		x3-d
						5		x3-2d
			4		x2-2d	0		x3+d
						1		x3-d
						2		x3-d
						3		x3-2d
						4		x3-2d
						5		x3-3d
			5		x2-2d	0		x3+d
						1		x3+d
						2		x3+d
						3		x3-2d
						4		x3-2d
						5		x3-3d

Table 18. Fixed design with 4 levels and $\boldsymbol{\beta}^T$ = (-3, 1, -2, 0.3, 2). n = 5, F(x_1) = 0.50, and B_a = 3.125%.

$d\gamma$	Pr(MLE)	Max(B)*, %	Variance*		MSE*		Asymptotic Variance *		Max($\hat{B}_f$), %
			β_2	γ	β_2	γ	β_2	γ	
0.3	0.9859	8.4	0.11	4.91	0.12	4.93	0.11	4.55	9.1
0.5	0.9859	6.8	0.11	1.85	0.13	1.86	0.11	1.70	8.9
1	0.9948	7.4	0.12	0.54	0.14	0.55	0.11	0.50	8.3
1.5	0.9987	7.1	0.13	0.30	0.15	0.32	0.12	0.27	7.6
2	0.9997	7.1	0.14	0.22	0.15	0.24	0.12	0.20	7.2
2.5	0.9999	7.5	0.16	0.19	0.18	0.22	0.14	0.17	7.5
3	0.9996	7.9	0.19	0.18	0.21	0.20	0.16	0.15	8.1
3.5	0.9997	8.6	0.21	0.17	0.24	0.20	0.19	0.15	8.7
4	0.9994	9.9	0.24	0.18	0.28	0.21	0.21	0.15	9.5
4.5	0.9984	9.8	0.26	0.17	0.30	0.21	0.24	0.15	10.6

* - conditional on convergence

Table 19. Sequential design with 4 levels and $\boldsymbol{\beta}^T$ = (-3, 1, -2, 0.3, 2). n = 5, F(x_1) = 0.50, P(MLE) = 1, and B_a = 3.125% for all rows .

$d\gamma$	Max(B), %		Variance		MSE		Asymptotic Variance		Max($\hat{B}_f$), %
	β_0, γ	$\beta_1, \beta_2, \beta_3$	β_2	γ	β_2	γ	β_2	γ	
0.3	22.2	11.2	0.20	1.00	0.25	1.20	0.18	0.98	6.7
0.5	14.2	10.2	0.17	0.38	0.22	0.46	0.17	0.36	5.8
1	10.1	9.9	0.17	0.16	0.21	0.20	0.15	0.15	5.7
1.5	9.5	10.8	0.18	0.13	0.22	0.17	0.16	0.12	6.0
2	9.7	10.4	0.19	0.14	0.23	0.17	0.17	0.12	6.5
2.5	10.4	12.0	0.22	0.15	0.27	0.19	0.18	0.13	7.2
3	11.7	13.8	0.23	0.17	0.29	0.22	0.20	0.14	7.6
3.5	11.8	13.0	0.26	0.18	0.32	0.23	0.22	0.15	8.1
4	12.0	13.8	0.26	0.18	0.32	0.23	0.23	0.15	8.6
4.5	12.7	14.9	0.26	0.18	0.33	0.24	0.23	0.15	8.8

Table 20. Comparison of fixed and sequential designs. B, $\hat{B}_f$ and Variance are conditional on convergence.

$\boldsymbol{\beta}$	**Fixed**						**Sequential**					
	P(MLE)	max B,%	max $\hat{B}_f$, %	Variance β_1	Variance γ	$d\gamma$	P(MLE)	max B,%	max $\hat{B}_f$, %	Variance β_1	Variance γ	$d\gamma$
(-3,1,-1,0.3,2)	1	7.1	7.1	0.086	0.131	3	1	12.8	5.5	0.095	0.151	2
(-3,2,-2,0.3,2)	0.9991	9	9.1	0.194	0.203	3	1	13.7	6.8	0.206	0.19	1
(-3,3,-3,0.3,3)	0.9644	11.6	15.1	0.514	0.542	5	0.9999	15.2	9	0.594	0.523	2
(-3,4,-1,0.3,4)	0.9682	10.8	13.8	0.895	0.859	5	0.9997	10	8.4	0.67	0.606	2.5
(-3,5,-5,0.3,6)	0.4047	6.6	18.5	0.462	0.586	9	0.9994	18.4	12.8	2.227	2.835	2.5

Table 21. Sequential design for a saturated model $\boldsymbol{\beta}^T = (0, 1, -1, 0.5, 1.2, -0.7, 0.2, -0.1, 3)$. $n = 5$.

	β_0	β_1	β_2	β_3	β_{12}	β_{13}	β_{23}	β_{123}	γ
$\boldsymbol{\beta}$	0	1	-1	0.5	1.2	-0.7	0.2	-0.1	3
B	0.02	0.15	-0.16	0.08	0.18	-0.10	0.03	-0.02	0.43
B, %	-	0.15	0.16	0.16	0.15	0.14	0.16	0.22	0.14
$\hat{B}_f$, %	-	0.13	0.13	0.13	0.13	0.13	0.14	0.15	0.13
Var	0.07	0.11	0.12	0.08	0.13	0.10	0.08	0.07	0.37
MSE	0.08	0.13	0.14	0.09	0.17	0.11	0.08	0.07	0.55
Asymp. Var	0.07	0.10	0.10	0.07	0.11	0.08	0.07	0.07	0.31

Table 22. Factors and levels for the TB discs experiment.

Control factors	Notation	Levels		
		-1	0	1
Speed	S	70 FPM	90 FPM	110 FPM
Top preheat	C	155 ^oF	-	175 ^oF
Tokuden die pressure	D	200 psi	-	325 psi
Tokuden die temperature	E	435 ^oF	-	455 ^oF
Corona treater for paperboard (CTT)	F	3 kW	-	18 kW
Corona treater for tab stock (CTP)	G	0 kW	-	6 kW
Bottom preheat	H	150 ^oF	-	170 ^oF
Blue tab stock material type	J	Good	-	Bad

Table 23. Aliasing of the 31 whole and split unit effects for the TB experiment.

	Effect	Aliasing
Whole unit	S	= 0.5(A+B)
	CD	= HJ
	EF	
	EG	
	FG	= S^2
	CEH	
	CGH	= B
Split unit	C	= DHJ
	D	= CHJ
	E	
	F	
	G	
	H	
	J	
	SC	= 0.5(FH+GH)
	SD	= 0.5(FJ+GJ)
	SE	
	SF	= SG
	SH	= 0.5(CF+CG)
	SJ	= 0.5(DF+DG)
	CE	
	CG	
	CH	= DJ
	CJ	= DH
	DE	
	DG	
	EH	
	EJ	
	FJ	
	GH	
	CDF	

Table 24. Variance components for the square root of the mean peel strength. R^2 = 0.97.

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Run&Random	34.267407	0.2250047	0.0592687	0.1425379	0.4074661	92.772
Lane[Run]&Random	1.6698299	0.0109643	0.0018665	0.0080608	0.0157855	4.521
Residual		0.0065661				2.707
Total		0.2425351				100.000

Table 25. Variance components for the ¼ power transformation of the mean within piece standard deviation. R^2 = 0.93.

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Run&Random	13.053969	0.0612208	0.0162303	0.0386819	0.1113564	87.360
Lane[Run]&Random	0.8886931	0.0041678	0.0007661	0.002996	0.0061955	5.947
Residual		0.0046898				6.692
Total		0.0700785				100.000

Table 26. Continuous and ordinal data for the TB experiment with control factor levels and run order. Runs 1 - 6.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel							Fast Peel						
													C1	C2	C3	C4	C5	C6	C7	C1	C2	C3	C4	C5	C6	C7
1	1	1	1.32	0.39	0	1	-1	1	-1	1	1	-1	0	15	25	0	0	0	0	0	10	10	0	0	0	0
1	2	1	1.51	0.45	0	1	-1	1	-1	1	1	-1	2	25	13	0	0	0	0	1	17	2	0	0	0	0
1	3	1	1.46	0.44	0	1	-1	1	-1	1	1	-1	0	20	20	0	0	0	0	0	17	3	0	0	0	0
1	4	1	1.58	0.38	0	1	-1	1	-1	1	1	-1	0	20	20	0	0	0	0	0	14	6	0	0	0	0
2	1	1	0.27	0.04	0	-1	1	1	-1	1	-1	1	0	0	0	0	0	5	35	0	0	0	0	0	7	13
2	2	1	0.31	0.05	0	-1	1	1	-1	1	-1	1	0	0	0	0	0	6	34	0	0	1	0	0	18	1
2	3	1	0.32	0.07	0	-1	1	1	-1	1	-1	1	0	0	0	0	0	5	35	0	0	0	0	0	9	11
2	4	1	0.34	0.13	0	-1	1	1	-1	1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	11	9
3	1	1	1.82	0.58	0	-1	1	-1	1	-1	1	-1	5	29	6	0	0	0	0	8	12	0	0	0	0	0
3	2	1	2.18	0.87	0	-1	1	-1	1	-1	1	-1	4	31	5	0	0	0	0	1	18	1	0	0	0	0
3	3	1	2.29	0.83	0	-1	1	-1	1	-1	1	-1	8	29	3	0	0	0	0	7	13	0	0	0	0	0
3	4	1	2.01	0.50	0	-1	1	-1	1	-1	1	-1	2	34	4	0	0	0	0	0	20	0	0	0	0	0
4	1	1	0.24	0.04	0	1	-1	-1	1	-1	-1	1	0	0	0	0	0	8	32	0	0	1	0	0	6	13
4	2	1	0.37	0.09	0	1	-1	-1	1	-1	-1	1	0	0	0	0	3	19	18	0	0	0	0	0	18	2
4	3	1	0.31	0.05	0	1	-1	-1	1	-1	-1	1	0	0	0	8	2	20	10	0	0	0	2	0	12	6
4	4	1	0.36	0.08	0	1	-1	-1	1	-1	-1	1	0	0	0	2	1	14	23	0	0	0	0	0	14	6
5	1	2	3.27	0.99	-1	-1	1	1	1	1	1	-1	31	9	0	0	0	0	0	19	1	0	0	0	0	0
5	2	2	3.17	0.75	-1	-1	1	1	1	1	1	-1	30	10	0	0	0	0	0	18	2	0	0	0	0	0
5	3	2	3.07	0.65	-1	-1	1	1	1	1	1	-1	39	1	0	0	0	0	0	20	0	0	0	0	0	0
5	4	2	2.63	0.46	-1	-1	1	1	1	1	1	-1	36	4	0	0	0	0	0	18	2	0	0	0	0	0
6	1	2	0.43	0.20	-1	1	-1	1	1	1	-1	1	0	0	0	29	11	0	0	0	0	0	8	5	7	0
6	2	2	0.52	0.10	-1	1	-1	1	1	1	-1	1	0	0	0	34	6	0	0	0	0	1	6	12	1	0
6	3	2	0.48	0.10	-1	1	-1	1	1	1	-1	1	0	0	0	29	11	0	0	0	1	1	6	3	8	1
6	4	2	0.48	0.09	-1	1	-1	1	1	1	-1	1	0	1	14	16	9	0	0	0	5	1	0	1	13	0

Table 26 continued. Runs 7 -12.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel							Fast Peel						
													C1	C2	C3	C4	C5	C6	C7	C1	C2	C3	C4	C5	C6	C7
7	1	2	1.76	0.54	-1	1	-1	-1	-1	-1	1	-1	5	34	1	0	0	0	0	4	16	0	0	0	0	0
7	2	2	2.13	0.57	-1	1	-1	-1	-1	-1	1	-1	2	26	12	0	0	0	0	4	16	0	0	0	0	0
7	3	2	2.07	0.46	-1	1	-1	-1	-1	-1	1	-1	9	16	15	0	0	0	0	1	18	1	0	0	0	0
7	4	2	1.85	0.43	-1	1	-1	-1	-1	-1	1	-1	8	32	0	0	0	0	0	9	11	0	0	0	0	0
8	1	2	0.67	0.18	-1	-1	1	-1	-1	-1	-1	1	0	0	0	33	7	0	0	0	0	8	4	0	8	0
8	2	2	0.70	0.16	-1	-1	1	-1	-1	-1	-1	1	0	0	1	34	5	0	0	0	1	4	12	0	3	0
8	3	2	0.73	0.15	-1	-1	1	-1	-1	-1	-1	1	0	0	11	17	11	1	0	0	0	10	1	2	7	0
8	4	2	0.63	0.19	-1	-1	1	-1	-1	-1	-1	1	2	3	23	5	3	4	0	0	1	6	0	0	13	0
9	1	3	0.44	0.06	0	1	-1	-1	-1	1	-1	1	0	0	0	0	1	25	14	0	0	0	0	0	10	10
9	2	3	0.51	0.11	0	1	-1	-1	-1	1	-1	1	0	0	0	4	1	35	0	0	0	0	0	0	19	1
9	3	3	0.77	0.15	0	1	-1	-1	-1	1	-1	1	0	0	0	17	19	4	0	0	0	1	1	0	17	1
9	4	3	0.74	0.19	0	1	-1	-1	-1	1	-1	1	0	0	0	9	3	28	0	0	0	2	0	0	12	6
10	1	3	0.36	0.04	0	-1	1	1	1	-1	-1	1	0	0	0	0	0	19	21	0	0	0	0	0	7	13
10	2	3	0.47	0.08	0	-1	1	1	1	-1	-1	1	0	0	0	14	4	22	0	0	0	1	1	1	15	2
10	3	3	0.65	0.13	0	-1	1	1	1	-1	-1	1	0	0	1	11	10	18	0	0	2	1	2	0	15	0
10	4	3	0.64	0.17	0	-1	1	1	1	-1	-1	1	0	0	7	0	0	33	0	0	1	11	0	0	8	0
11	1	3	1.41	0.34	0	-1	1	-1	-1	1	1	-1	0	7	20	13	0	0	0	0	20	0	0	0	0	0
11	2	3	1.53	0.36	0	-1	1	-1	-1	1	1	-1	0	7	32	1	0	0	0	0	19	1	0	0	0	0
11	3	3	1.61	0.42	0	-1	1	-1	-1	1	1	-1	0	11	21	8	0	0	0	0	16	4	0	0	0	0
11	4	3	1.75	0.50	0	-1	1	-1	-1	1	1	-1	0	23	17	0	0	0	0	7	8	5	0	0	0	0
12	1	3	1.34	0.34	0	1	-1	1	1	-1	1	-1	0	2	10	28	0	0	0	0	16	4	0	0	0	0
12	2	3	1.65	0.46	0	1	-1	1	1	-1	1	-1	0	33	7	0	0	0	0	1	19	0	0	0	0	0
12	3	3	1.79	0.51	0	1	-1	1	1	-1	1	-1	1	20	19	0	0	0	0	1	19	0	0	0	0	0
12	4	3	2.13	0.46	0	1	-1	1	1	-1	1	-1	1	14	25	0	0	0	0	0	20	0	0	0	0	0

Table 26 continued. Runs 13 -18.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel							Fast Peel						
													C1	C2	C3	C4	C5	C6	C7	C1	C2	C3	C4	C5	C6	C7
13	1	4	0.00	0.00	1	-1	1	-1	1	1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
13	2	4	0.00	0.00	1	-1	1	-1	1	1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
13	3	4	0.00	0.00	1	-1	1	-1	1	1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
13	4	4	0.00	0.00	1	-1	1	-1	1	1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
14	1	4	0.00	0.00	1	1	-1	1	-1	-1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
14	2	4	0.00	0.00	1	1	-1	1	-1	-1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
14	3	4	0.00	0.00	1	1	-1	1	-1	-1	-1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
14	4	4	0.24	0.07	1	1	-1	1	-1	-1	-1	1	0	0	0	0	0	6	34	0	0	0	0	0	0	20
15	1	4	0.96	0.33	1	-1	1	1	-1	-1	1	-1	0	1	6	0	0	31	2	0	0	19	0	0	1	0
15	2	4	1.44	0.34	1	-1	1	1	-1	-1	1	-1	1	31	3	3	0	2	0	0	19	1	0	0	0	0
15	3	4	1.58	0.38	1	-1	1	1	-1	-1	1	-1	0	36	4	0	0	0	0	0	19	1	0	0	0	0
15	4	4	1.49	0.41	1	-1	1	1	-1	-1	1	-1	1	34	4	0	0	1	0	0	3	17	0	0	0	0
16	1	4	0.71	0.23	1	1	-1	-1	1	1	1	-1	0	0	0	1	0	39	0	0	0	1	0	0	19	0
16	2	4	1.22	0.34	1	1	-1	-1	1	1	1	-1	0	5	3	14	0	18	0	0	7	12	0	0	1	0
16	3	4	1.53	0.45	1	1	-1	-1	1	1	1	-1	0	27	13	0	0	0	0	0	20	0	0	0	0	0
16	4	4	1.63	0.47	1	1	-1	-1	1	1	1	-1	0	20	10	10	0	0	0	0	20	0	0	0	0	0
17	1	5	1.76	0.47	0	1	1	1	-1	1	-1	-1	1	36	3	0	0	0	0	1	19	0	0	0	0	0
17	2	5	2.13	0.57	0	1	1	1	-1	1	-1	-1	6	34	0	0	0	0	0	5	15	0	0	0	0	0
17	3	5	2.01	0.76	0	1	1	1	-1	1	-1	-1	4	36	0	0	0	0	0	10	10	0	0	0	0	0
17	4	5	2.58	0.71	0	1	1	1	-1	1	-1	-1	6	34	0	0	0	0	0	11	9	0	0	0	0	0
18	1	5	0.53	0.13	0	1	1	-1	1	-1	1	1	0	0	0	1	0	39	0	0	0	0	0	0	20	0
18	2	5	0.58	0.11	0	1	1	-1	1	-1	1	1	0	0	0	20	3	17	0	0	0	0	1	8	11	0
18	3	5	0.63	0.09	0	1	1	-1	1	-1	1	1	0	0	0	18	4	18	0	0	0	3	1	5	11	0
18	4	5	0.82	0.18	0	1	1	-1	1	-1	1	1	0	0	2	0	0	38	0	0	2	8	0	0	10	0

Table 26 continued. Runs 19 -24.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel							Fast Peel						
													C1	C2	C3	C4	C5	C6	C7	C1	C2	C3	C4	C5	C6	C7
19	1	5	0.06	0.02	0	-1	-1	1	-1	1	1	1	0	0	0	0	0	4	36	0	0	0	0	0	1	19
19	2	5	0.34	0.04	0	-1	-1	1	-1	1	1	1	0	0	0	0	0	32	8	0	0	0	0	0	9	11
19	3	5	0.63	0.15	0	-1	-1	1	-1	1	1	1	0	0	0	3	0	37	0	0	0	2	0	0	17	1
19	4	5	0.80	0.19	0	-1	-1	1	-1	1	1	1	0	0	0	1	0	39	0	0	0	2	0	0	18	0
20	1	5	1.09	0.32	0	-1	-1	-1	1	-1	-1	-1	0	0	0	31	0	9	0	0	0	2	11	0	7	0
20	2	5	1.61	0.51	0	-1	-1	-1	1	-1	-1	-1	0	12	23	4	0	1	0	1	13	6	0	0	0	0
20	3	5	1.59	0.41	0	-1	-1	-1	1	-1	-1	-1	0	29	11	0	0	0	0	0	19	1	0	0	0	0
20	4	5	1.75	0.47	0	-1	-1	-1	1	-1	-1	-1	0	7	31	2	0	0	0	0	20	0	0	0	0	0
21	1	6	0.00	0.00	1	-1	-1	-1	-1	-1	1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
21	2	6	0.00	0.00	1	-1	-1	-1	-1	-1	1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
21	3	6	0.00	0.00	1	-1	-1	-1	-1	-1	1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
21	4	6	0.00	0.00	1	-1	-1	-1	-1	-1	1	1	0	0	0	0	0	0	40	0	0	0	0	0	0	20
22	1	6	1.47	0.42	1	-1	-1	1	1	1	-1	-1	0	0	16	23	0	1	0	0	14	6	0	0	0	0
22	2	6	0.94	0.27	1	-1	-1	1	1	1	-1	-1	0	0	1	30	0	9	0	0	2	2	10	0	6	0
22	3	6	1.68	0.49	1	-1	-1	1	1	1	-1	-1	11	20	9	0	0	0	0	0	8	12	0	0	0	0
22	4	6	1.72	0.47	1	-1	-1	1	1	1	-1	-1	0	32	8	0	0	0	0	0	19	1	0	0	0	0
23	1	6	1.73	0.51	1	1	1	-1	-1	-1	-1	-1	0	19	21	0	0	0	0	0	19	1	0	0	0	0
23	2	6	2.14	0.51	1	1	1	-1	-1	-1	-1	-1	0	22	18	0	0	0	0	0	20	0	0	0	0	0
23	3	6	2.11	0.57	1	1	1	-1	-1	-1	-1	-1	0	38	2	0	0	0	0	1	19	0	0	0	0	0
23	4	6	2.14	0.48	1	1	1	-1	-1	-1	-1	-1	0	40	0	0	0	0	0	1	18	1	0	0	0	0
24	1	6	0.63	0.16	1	1	1	1	1	1	1	1	0	0	0	0	0	6	34	0	0	0	0	0	7	13
24	2	6	0.97	0.26	1	1	1	1	1	1	1	1	0	0	0	12	7	19	2	0	0	2	5	3	10	0
24	3	6	0.95	0.23	1	1	1	1	1	1	1	1	0	0	1	9	27	3	0	0	0	1	7	7	5	0
24	4	6	0.82	0.18	1	1	1	1	1	1	1	1	0	0	0	13	27	0	0	0	0	1	0	0	18	1

Table 26 continued. Runs 25 -30.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel							Fast Peel						
													C1	C2	C3	C4	C5	C6	C7	C1	C2	C3	C4	C5	C6	C7
25	1	7	2.68	0.95	0	1	1	1	1	-1	-1	-1	0	40	0	0	0	0	0	8	12	0	0	0	0	0
25	2	7	2.80	0.45	0	1	1	1	1	-1	-1	-1	5	27	8	0	0	0	0	14	6	0	0	0	0	0
25	3	7	2.68	0.55	0	1	1	1	1	-1	-1	-1	4	33	3	0	0	0	0	17	3	0	0	0	0	0
25	4	7	3.04	0.66	0	1	1	1	1	-1	-1	-1	4	34	2	0	0	0	0	16	4	0	0	0	0	0
26	1	7	0.94	0.26	0	1	1	-1	-1	1	1	1	0	0	1	0	0	39	0	0	0	4	0	0	16	0
26	2	7	1.04	0.24	0	1	1	-1	-1	1	1	1	0	0	1	26	0	13	0	0	0	5	12	0	3	0
26	3	7	0.94	0.21	0	1	1	-1	-1	1	1	1	0	0	1	19	5	15	0	0	0	4	13	0	3	0
26	4	7	1.04	0.23	0	1	1	-1	-1	1	1	1	0	0	2	1	0	37	0	0	3	4	0	0	13	0
27	1	7	2.00	0.54	0	-1	-1	-1	-1	1	-1	-1	0	40	0	0	0	0	0	2	18	0	0	0	0	0
27	2	7	1.61	0.41	0	-1	-1	-1	-1	1	-1	-1	1	34	5	0	0	0	0	3	17	0	0	0	0	0
27	3	7	1.98	0.72	0	-1	-1	-1	-1	1	-1	-1	0	26	14	0	0	0	0	2	16	2	0	0	0	0
27	4	7	1.70	0.50	0	-1	-1	-1	-1	1	-1	-1	0	18	20	2	0	0	0	0	19	1	0	0	0	0
28	1	7	0.60	0.14	0	-1	-1	1	1	-1	1	1	0	0	0	5	1	34	0	0	0	1	3	0	16	0
28	2	7	0.71	0.15	0	-1	-1	1	1	-1	1	1	0	0	0	37	3	0	0	0	0	0	18	0	2	0
28	3	7	0.74	0.14	0	-1	-1	1	1	-1	1	1	0	0	0	30	5	5	0	0	0	1	17	0	2	0
28	4	7	0.96	0.24	0	-1	-1	1	1	-1	1	1	0	0	2	19	0	19	0	0	0	4	4	0	12	0
29	1	8	1.13	0.34	-1	-1	-1	-1	1	1	1	1	0	0	9	30	0	1	0	0	1	17	1	0	1	0
29	2	8	1.30	0.29	-1	-1	-1	-1	1	1	1	1	0	0	3	37	0	0	0	0	0	9	8	0	3	0
29	3	8	1.29	0.31	-1	-1	-1	-1	1	1	1	1	0	0	17	23	0	0	0	0	5	12	3	0	0	0
29	4	8	1.30	0.33	-1	-1	-1	-1	1	1	1	1	7	8	10	15	0	0	0	2	8	10	0	0	0	0
30	1	8	1.46	0.41	-1	1	1	1	-1	-1	1	1	2	13	20	5	0	0	0	1	17	2	0	0	0	0
30	2	8	1.29	0.31	-1	1	1	1	-1	-1	1	1	3	22	15	0	0	0	0	5	13	2	0	0	0	0
30	3	8	1.34	0.32	-1	1	1	1	-1	-1	1	1	13	21	5	1	0	0	0	4	15	1	0	0	0	0
30	4	8	1.66	0.33	-1	1	1	1	-1	-1	1	1	6	24	10	0	0	0	0	7	11	2	0	0	0	0

Table 26 continued. Runs 31 -32.

Run	Lane	Block	Mean (mean)	Mean (StD)	Speed	C	D	E	F	G	H	J	Slow Peel C1	C2	C3	C4	C5	C6	C7	Fast Peel C1	C2	C3	C4	C5	C6	C7
31	1	8	3.18	0.72	-1	-1	-1	1	-1	-1	-1	-1	17	22	1	0	0	0	0	8	12	0	0	0	0	0
31	2	8	3.00	0.94	-1	-1	-1	1	-1	-1	-1	-1	14	23	3	0	0	0	0	18	2	0	0	0	0	0
31	3	8	2.78	0.70	-1	-1	-1	1	-1	-1	-1	-1	31	9	0	0	0	0	0	19	1	0	0	0	0	0
31	4	8	2.39	0.55	-1	-1	-1	1	-1	-1	-1	-1	19	21	0	0	0	0	0	19	1	0	0	0	0	0
32	1	8	3.64	1.04	-1	1	1	-1	1	1	-1	-1	24	14	2	0	0	0	0	15	5	0	0	0	0	0
32	2	8	3.69	1.06	-1	1	1	-1	1	1	-1	-1	36	3	1	0	0	0	0	17	3	0	0	0	0	0
32	3	8	3.21	0.93	-1	1	1	-1	1	1	-1	-1	39	1	0	0	0	0	0	20	0	0	0	0	0	0
32	4	8	3.41	1.02	-1	1	1	-1	1	1	-1	-1	39	1	0	0	0	0	0	19	1	0	0	0	0	0

Table 27. t-ratios of the factorial effects for the square root of the mean of the mean for the TB disc experiment.

Effect type	Term	Naïve t-ratio	Corrected t-ratio	p-value
whole-unit	Speed	-21.44	-4.04	.01<p<.02
whole-unit	C*D	11.58	2.18	.05<p<.06
whole-unit	C*E*H	3.97	0.75	.30<p<.40
whole-unit	E*G	-3.82	-0.72	
whole-unit	E*F	3.26	0.61	
whole-unit	CGH-Speed	2.26	0.43	
whole-unit	F*G	-1.20	-0.23	
split-unit	J[-1]	38.80	11.41	p<.001
split-unit	D	9.17	2.70	.02<p<.03
split-unit	C	5.74	1.69	.09<p<.10
split-unit	C*E	-5.67	-1.67	
split-unit	Speed*C	5.66	1.66	
split-unit	F	3.99	1.17	
split-unit	H	3.39	1.00	
split-unit	E	3.26	0.96	
split-unit	C*J[-1]	-3.13	-0.92	
split-unit	Speed*E	2.86	0.84	
split-unit	E*H	2.79	0.82	
split-unit	C*G	2.58	0.76	
split-unit	Speed*D	2.37	0.70	
split-unit	C*H	-2.28	-0.67	
split-unit	Speed*G	2.27	0.67	
split-unit	D*E	2.18	0.64	
split-unit	Speed*J[-1]	2.08	0.61	
split-unit	F*J[-1]	1.82	0.53	
split-unit	D*G	-1.79	-0.53	
split-unit	Speed*H	-1.62	-0.48	
split-unit	G	1.55	0.46	
split-unit	G*H	-1.29	-0.38	
split-unit	E*J[-1]	-0.85	-0.25	
split-unit	C*D*F	-0.49	-0.14	

Table 28. Redefined versus original categories.

Redefined categories		Original categories
#	Description	#
1	very tight seal, with seal area completely covered by tab paper	1
2	tight seal, with some of tab paper inside seal area	2
3	tight seal, with little paper on sealing edge and/or outside the sealing area	3
4	perfect seal with consistent pattern and all tab paper removed	4
5	seal with inconsistent pattern and good pull	6
6	consistent seal with weak pull	5
7	seal with inconsistent pattern and weak pull	6
8	almost no seal or no seal at all, very weak seal	7

Table 29. Ordinal response data with 8 redefined categories (slow peel).

Run	Lane	Block	Speed	C1	C2	C3	C4	C5	C6	C7	C8
1	1	1	0	0	15	25	0	0	0	0	0
1	2	1	0	2	25	13	0	0	0	0	0
1	3	1	0	0	20	20	0	0	0	0	0
1	4	1	0	0	20	20	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	5	35
2	2	1	0	0	0	0	0	0	0	6	34
2	3	1	0	0	0	0	0	0	0	5	35
2	4	1	0	0	0	0	0	0	0	0	40
3	1	1	0	5	29	6	0	0	0	0	0
3	2	1	0	4	31	5	0	0	0	0	0
3	3	1	0	8	29	3	0	0	0	0	0
3	4	1	0	2	34	4	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	8	32
4	2	1	0	0	0	0	0	0	3	19	18
4	3	1	0	0	0	0	8	0	2	20	10
4	4	1	0	0	0	0	2	0	1	14	23
5	1	2	-1	31	9	0	0	0	0	0	0
5	2	2	-1	30	10	0	0	0	0	0	0
5	3	2	-1	39	1	0	0	0	0	0	0
5	4	2	-1	36	4	0	0	0	0	0	0
6	1	2	-1	0	0	0	29	0	11	0	0
6	2	2	-1	0	0	0	34	0	6	0	0
6	3	2	-1	0	0	0	29	0	11	0	0
6	4	2	-1	0	1	14	16	0	9	0	0
7	1	2	-1	5	34	1	0	0	0	0	0
7	2	2	-1	2	26	12	0	0	0	0	0
7	3	2	-1	9	16	15	0	0	0	0	0
7	4	2	-1	8	32	0	0	0	0	0	0
8	1	2	-1	0	0	0	33	0	7	0	0
8	2	2	-1	0	0	1	34	0	5	0	0
8	3	2	-1	0	0	11	17	0	11	1	0
8	4	2	-1	2	3	23	5	4	3	0	0
9	1	3	0	0	0	0	0	0	1	25	14
9	2	3	0	0	0	0	4	2	1	33	0
9	3	3	0	0	0	0	17	0	19	4	0
9	4	3	0	0	0	0	9	5	3	23	0
10	1	3	0	0	0	0	0	12	0	7	21
10	2	3	0	0	0	0	14	7	4	15	0
10	3	3	0	0	0	1	11	13	10	5	0
10	4	3	0	0	0	7	0	18	0	15	0
11	1	3	0	0	7	20	13	0	0	0	0
11	2	3	0	0	7	32	1	0	0	0	0
11	3	3	0	0	11	21	8	0	0	0	0
11	4	3	0	0	23	17	0	0	0	0	0
12	1	3	0	0	2	10	28	0	0	0	0
12	2	3	0	0	33	7	0	0	0	0	0
12	3	3	0	1	20	19	0	0	0	0	0
12	4	3	0	1	14	25	0	0	0	0	0

Table 29. Continued.

Run	Lane	Block	Speed	C1	C2	C3	C4	C5	C6	C7	C8
13	1	4	1	0	0	0	0	0	0	0	40
13	2	4	1	0	0	0	0	0	0	0	40
13	3	4	1	0	0	0	0	0	0	0	40
13	4	4	1	0	0	0	0	0	0	0	40
14	1	4	1	0	0	0	0	0	0	0	40
14	2	4	1	0	0	0	0	0	0	0	40
14	3	4	1	0	0	0	0	0	0	0	40
14	4	4	1	0	0	0	0	0	0	6	34
15	1	4	1	0	1	6	0	31	0	0	2
15	2	4	1	1	31	3	3	2	0	0	0
15	3	4	1	0	36	4	0	0	0	0	0
15	4	4	1	1	34	4	0	1	0	0	0
16	1	4	1	0	0	0	1	39	0	0	0
16	2	4	1	0	5	3	14	18	0	0	0
16	3	4	1	0	27	13	0	0	0	0	0
16	4	4	1	0	20	10	10	0	0	0	0
17	1	5	0	1	36	3	0	0	0	0	0
17	2	5	0	6	34	0	0	0	0	0	0
17	3	5	0	4	36	0	0	0	0	0	0
17	4	5	0	6	34	0	0	0	0	0	0
18	1	5	0	0	0	0	1	0	0	39	0
18	2	5	0	0	0	0	20	0	3	17	0
18	3	5	0	0	0	0	18	0	4	18	0
18	4	5	0	0	0	2	0	0	0	38	0
19	1	5	0	0	0	0	0	0	0	4	36
19	2	5	0	0	0	0	0	0	0	32	8
19	3	5	0	0	0	0	3	0	0	37	0
19	4	5	0	0	0	0	1	0	0	39	0
20	1	5	0	0	0	0	31	9	0	0	0
20	2	5	0	0	12	23	4	1	0	0	0
20	3	5	0	0	29	11	0	0	0	0	0
20	4	5	0	0	7	30	3	0	0	0	0
21	1	6	1	0	0	0	0	0	0	0	40
21	2	6	1	0	0	0	0	0	0	0	40
21	3	6	1	0	0	0	0	0	0	0	40
21	4	6	1	0	0	0	0	0	0	0	40
22	1	6	1	0	0	16	23	1	0	0	0
22	2	6	1	0	0	1	30	9	0	0	0
22	3	6	1	11	20	9	0	0	0	0	0
22	4	6	1	0	32	8	0	0	0	0	0
23	1	6	1	0	19	21	0	0	0	0	0
23	2	6	1	0	22	18	0	0	0	0	0
23	3	6	1	0	38	2	0	0	0	0	0
23	4	6	1	0	40	0	0	0	0	0	0
24	1	6	1	0	0	0	0	0	0	6	34
24	2	6	1	0	0	0	12	0	7	19	2
24	3	6	1	0	0	1	9	0	27	3	0
24	4	6	1	0	0	0	13	0	27	0	0

Table 29. Continued.

Run	Lane	Block	Speed	C1	C2	C3	C4	C5	C6	C7	C8
25	1	7	0	0	40	0	0	0	0	0	0
25	2	7	0	5	27	8	0	0	0	0	0
25	3	7	0	4	33	3	0	0	0	0	0
25	4	7	0	4	34	2	0	0	0	0	0
26	1	7	0	0	0	1	0	39	0	0	0
26	2	7	0	0	0	1	26	13	0	0	0
26	3	7	0	0	0	1	19	15	5	0	0
26	4	7	0	0	0	2	1	37	0	0	0
27	1	7	0	0	40	0	0	0	0	0	0
27	2	7	0	1	34	5	0	0	0	0	0
27	3	7	0	0	26	14	0	0	0	0	0
27	4	7	0	0	18	20	2	0	0	0	0
28	1	7	0	0	0	0	5	34	1	0	0
28	2	7	0	0	0	0	37	0	3	0	0
28	3	7	0	0	0	0	30	5	5	0	0
28	4	7	0	0	0	2	19	19	0	0	0
29	1	8	-1	0	0	9	30	0	0	1	0
29	2	8	-1	0	0	3	37	0	0	0	0
29	3	8	-1	0	0	17	23	0	0	0	0
29	4	8	-1	7	8	10	15	0	0	0	0
30	1	8	-1	2	13	20	5	0	0	0	0
30	2	8	-1	3	22	15	0	0	0	0	0
30	3	8	-1	13	21	5	1	0	0	0	0
30	4	8	-1	6	24	10	0	0	0	0	0
31	1	8	-1	17	22	1	0	0	0	0	0
31	2	8	-1	14	23	3	0	0	0	0	0
31	3	8	-1	30	10	0	0	0	0	0	0
31	4	8	-1	19	21	0	0	0	0	0	0
32	1	8	-1	24	14	2	0	0	0	0	0
32	2	8	-1	36	3	1	0	0	0	0	0
32	3	8	-1	39	1	0	0	0	0	0	0
32	4	8	-1	39	1	0	0	0	0	0	0

Table 30. Bayesian analysis of TB ordinal data.

Parameter	Mean	Standard Deviation	2.50%	97.50%
(a) Fixed effect estimates				
α_1	-6.369	0.114	-6.598	-6.145
α_2	-2.009	0.056	-2.120	-1.900
α_3	-0.246	0.053	-0.350	-0.144
α_4	2.054	0.063	1.931	2.178
α_5	3.049	0.071	2.912	3.189
α_6	3.697	0.077	3.546	3.850
α_7	5.254	0.092	5.074	5.434
S	-3.058	0.502	-4.099	-2.025
C	0.316	0.029	0.260	0.372
D	0.668	0.029	0.610	0.725
J	-3.511	0.058	-3.623	-3.396
CD	0.796	0.324	0.153	1.379
L1	-1.032	0.050	-1.128	-0.940
L2	-0.033	0.050	-0.130	0.063
L3	0.634	0.050	0.532	0.726
(b) Random effect predictions				
block 1	-0.720	0.330	-1.387	-0.120
block 2	0.313	0.566	-0.803	1.417
block 3	-0.094	0.329	-0.758	0.505
block 4	0.560	0.628	-0.711	1.781
block 5	-1.629	0.330	-2.235	-0.964
block 6	0.589	0.565	-0.509	1.710
block 7	0.276	0.331	-0.329	0.941
block 8	0.705	0.628	-0.510	1.986
σ_b	1.039	0.326	0.592	1.872

Table 31. Penalties associated with each category.

Category	Description	Penalty
C1	very tight seal, with seal area completely covered by tab paper	5
C2	tight seal, with some of tab paper inside seal area	3
C3	tight seal, with little paper on sealing edge and/or outside the sealing area	-3
C4	perfect seal with consistent pattern and all tab paper removed	-5
C5	seal with inconsistent pattern and good pull	-3
C6	consistent seal with weak pull	3
C7	seal with inconsistent pattern and weak pull	5
C8	almost no seal or no seal at all, very weak seal	10

Table 32. Expected loss statistics for TB experiment.

Run	S	C	D	J	mean	sd	2.50%	median	97.50%
8	-1	-1	1	1	-2.74	0.07	-2.88	-2.74	-2.60
16	1	1	-1	-1	-2.67	0.07	-2.82	-2.68	-2.53
29	-1	-1	-1	1	-2.58	0.09	-2.74	-2.58	-2.40
6	-1	1	-1	1	-2.39	0.11	-2.60	-2.39	-2.16
26	0	1	1	1	-2.22	0.12	-2.44	-2.22	-1.99
15	1	-1	1	-1	-2.17	0.12	-2.39	-2.17	-1.93
22	1	-1	-1	-1	-1.85	0.12	-2.09	-1.85	-1.59
1	0	1	-1	-1	-0.74	0.14	-1.01	-0.74	-0.47
20	0	-1	-1	-1	-0.66	0.14	-0.94	-0.66	-0.38
30	-1	1	1	1	-0.13	0.17	-0.46	-0.13	0.20
12	0	1	-1	-1	0.22	0.14	-0.05	0.22	0.49
3	0	-1	1	-1	0.35	0.15	0.06	0.35	0.63
23	1	1	1	-1	1.02	0.13	0.77	1.02	1.26
11	0	-1	1	-1	1.22	0.12	0.98	1.22	1.46
18	0	1	1	1	1.62	0.24	1.16	1.62	2.11
28	0	-1	-1	1	1.79	0.26	1.29	1.79	2.30
27	0	-1	-1	-1	1.94	0.10	1.73	1.95	2.14
17	0	1	1	-1	2.01	0.10	1.82	2.01	2.19
25	0	1	1	-1	3.44	0.06	3.33	3.44	3.55
10	0	-1	1	1	3.45	0.24	2.97	3.45	3.91
7	-1	1	-1	-1	3.54	0.06	3.43	3.54	3.66
24	1	1	1	1	3.84	0.23	3.39	3.84	4.30
5	-1	-1	1	-1	3.95	0.05	3.84	3.95	4.05
31	-1	-1	-1	-1	4.27	0.05	4.17	4.27	4.37
32	-1	1	1	-1	4.84	0.02	4.81	4.84	4.87
2	0	-1	1	1	5.03	0.24	4.56	5.03	5.49
9	0	1	-1	1	5.23	0.21	4.81	5.23	5.65
19	0	-1	-1	1	6.50	0.18	6.13	6.50	6.86
4	0	1	-1	1	6.60	0.18	6.23	6.60	6.96
21	1	-1	-1	1	7.97	0.14	7.68	7.97	8.25
13	1	-1	1	1	8.36	0.13	8.09	8.36	8.61
14	1	1	-1	1	9.07	0.09	8.90	9.08	9.24

Appendix C: Figures

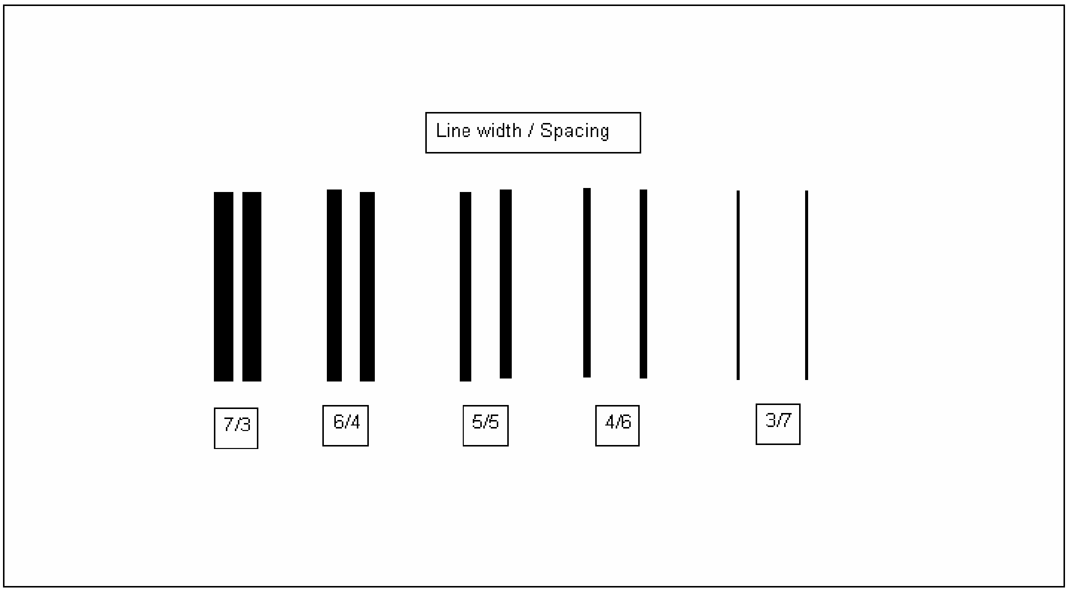


Figure 1. Line width and spacing for 5 pairs of conductors in PCB experiment.

a) Opens. A p-value = 0.30 b) Shorts. A p-value = 0.24.

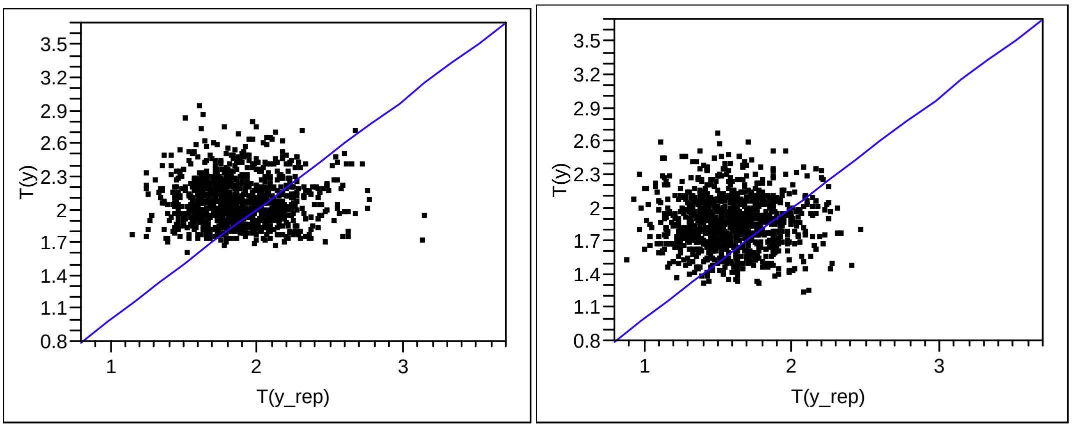


Figure 2. Realized versus posterior predictive distributions for the test quantity T(y) for the mixed PCB models (GLMM).

a) Opens. A p-value = 0.016 b) Shorts. A p-value = 0.032.

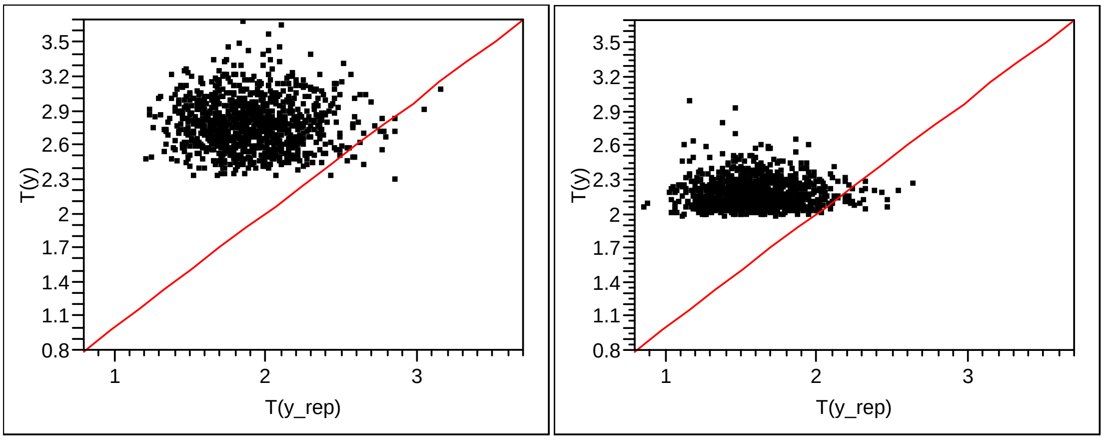


Figure 3. Realized versus posterior predictive distributions for the test quantity *T(y)* for the PCB models without the random effects (GLM).

a) Opens. A p-value = 0 b) Shorts. A p-value = 0.044

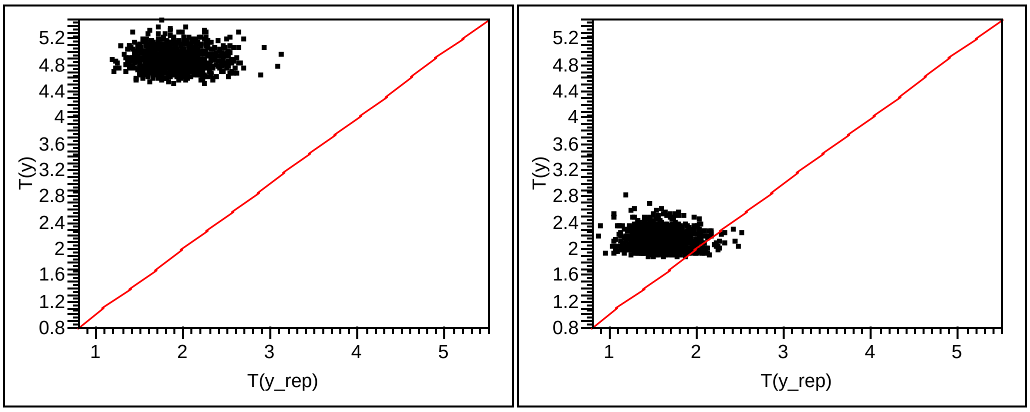


Figure 4. Realized versus posterior predictive distributions for the test quantity T(y) for the PCB models according to JW.

a) Misfeeds. A p-value = 0. b) Multifeeds. A p-value = 0.38.

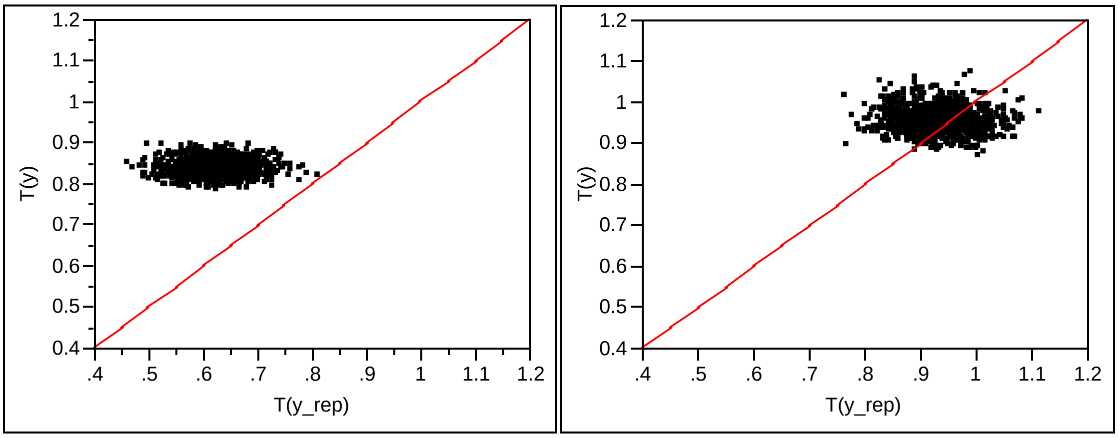


Figure 5. Realized versus posterior predictive distributions for the test quantity *T(y)* for the mixed models in the paper feeder example.

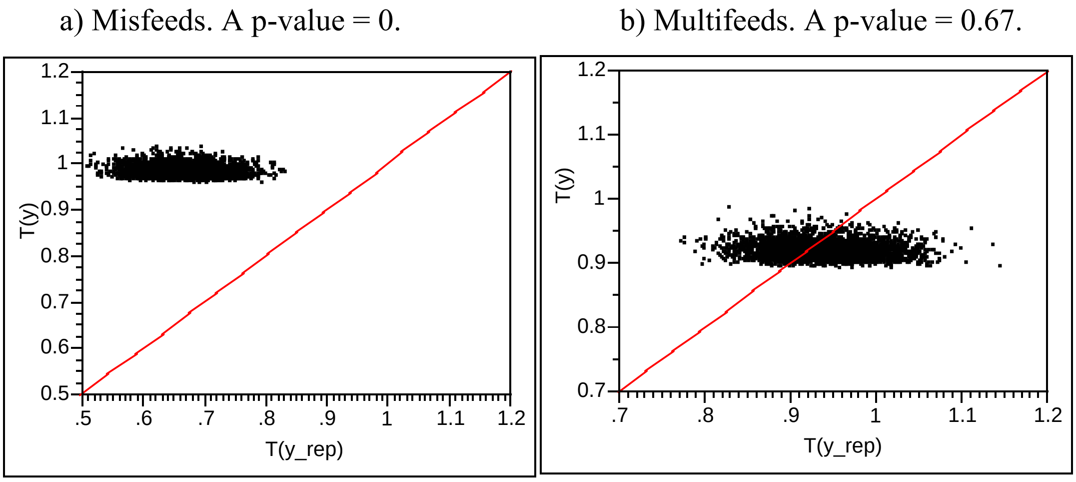


Figure 6. Realized versus posterior predictive distributions for the test quantity *T(y)* for JW models in the paper feeder example.

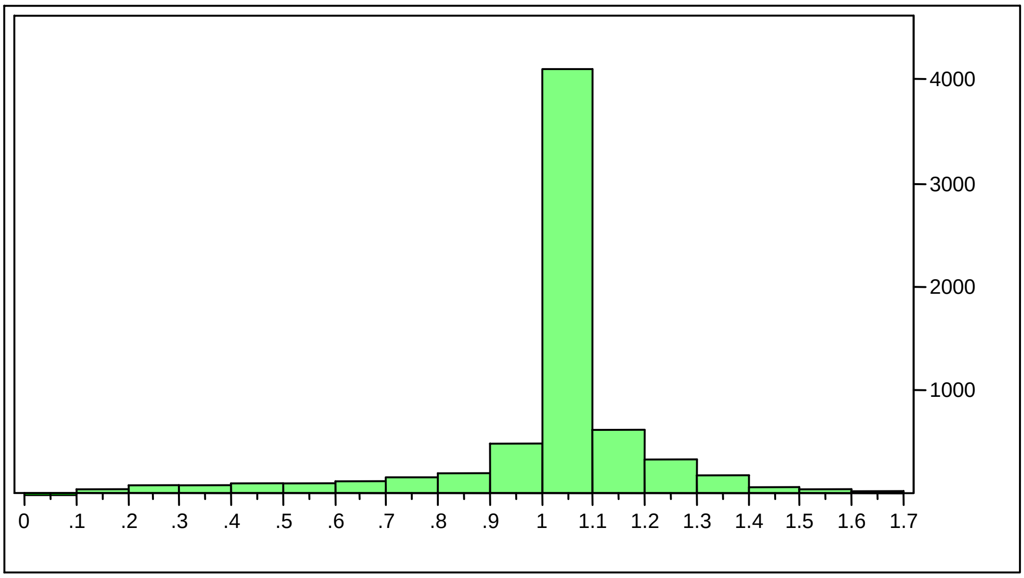


Figure 7. Histogram of the 97.5 percentile of the expected loss for the paper feeder example.

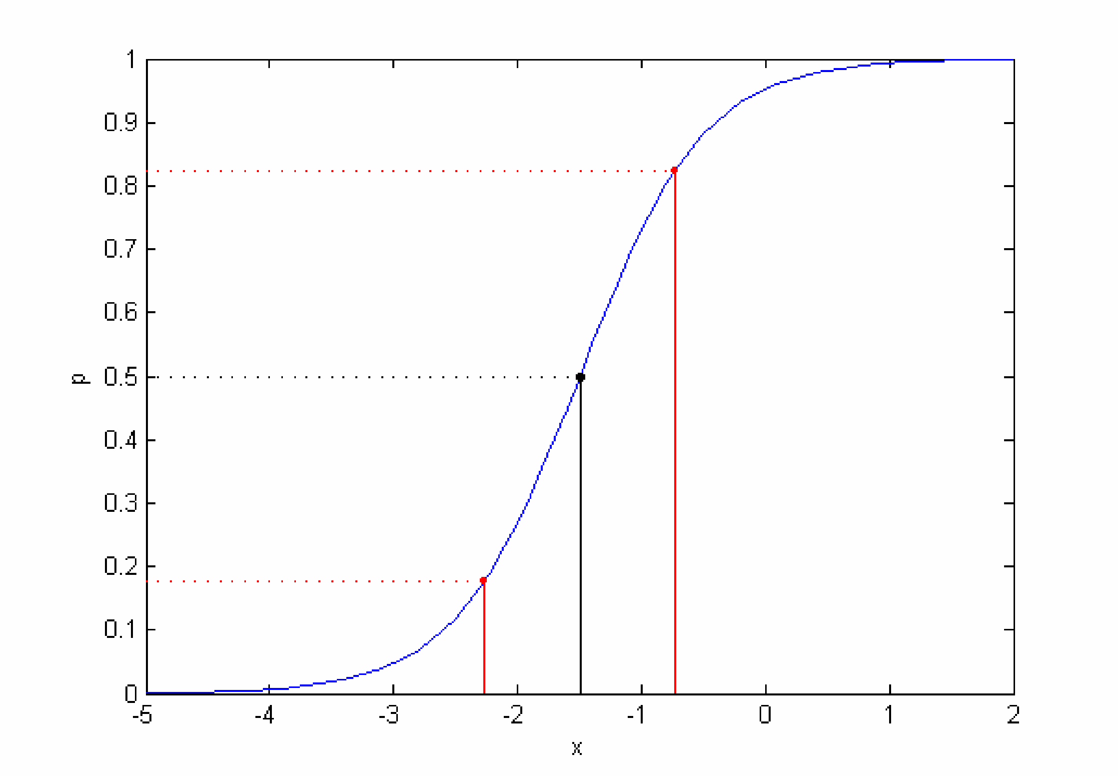


Figure 8. Logistic curve with $\beta_0 = 3$ and $\beta_1 = 2$. Red lines correspond to the D-optimal design levels of x.

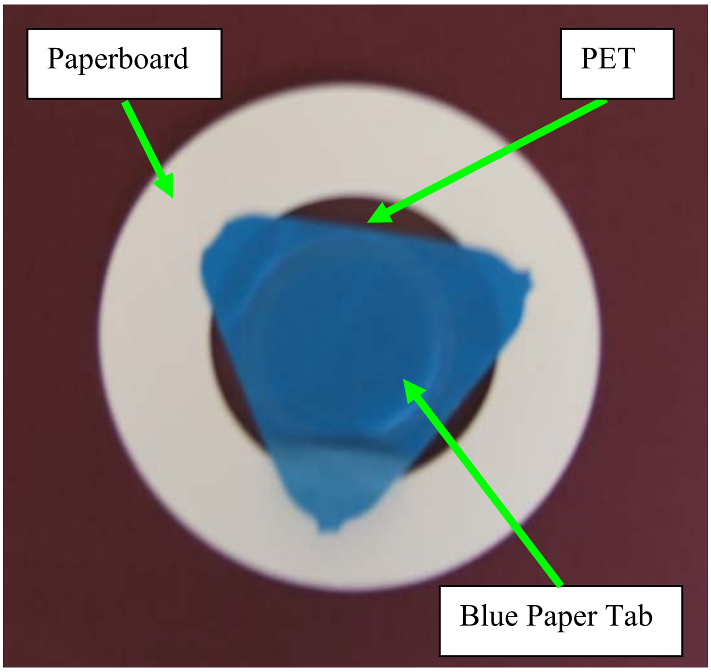


Figure 9. Completed Taco Bell disc

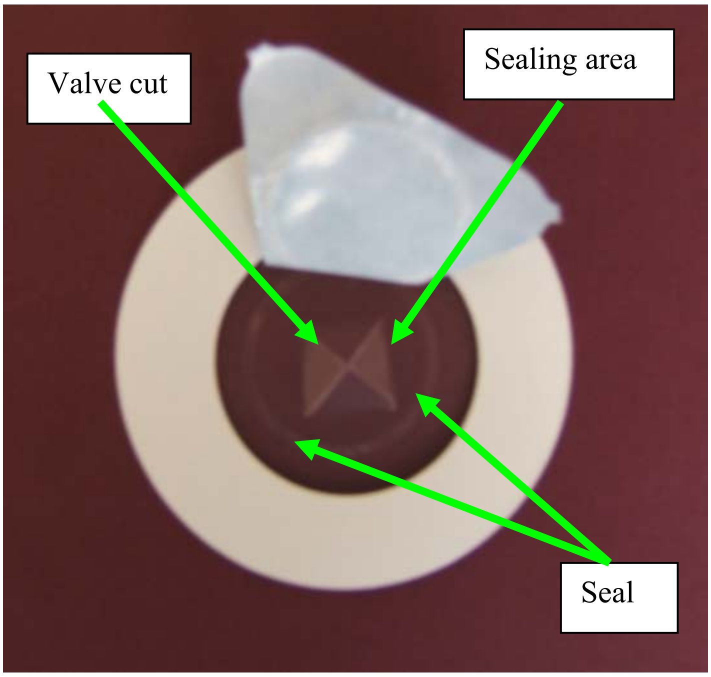


Figure 10. Taco Bell Disc with area under the tab.

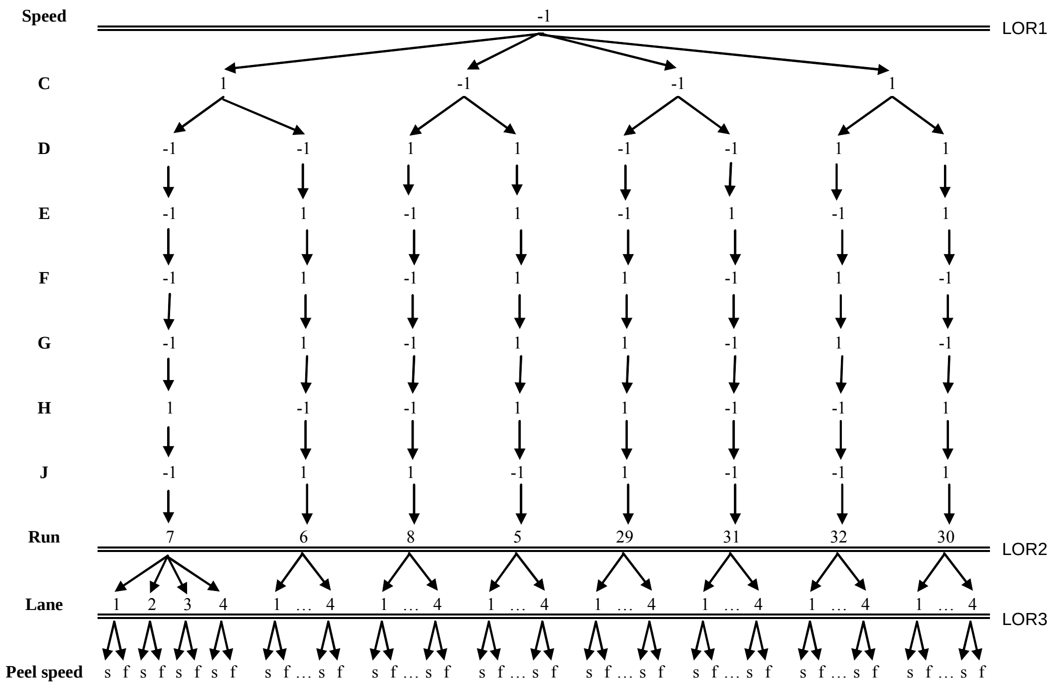


Figure 11. Factor relation diagram for TB experiment with low level of speed.

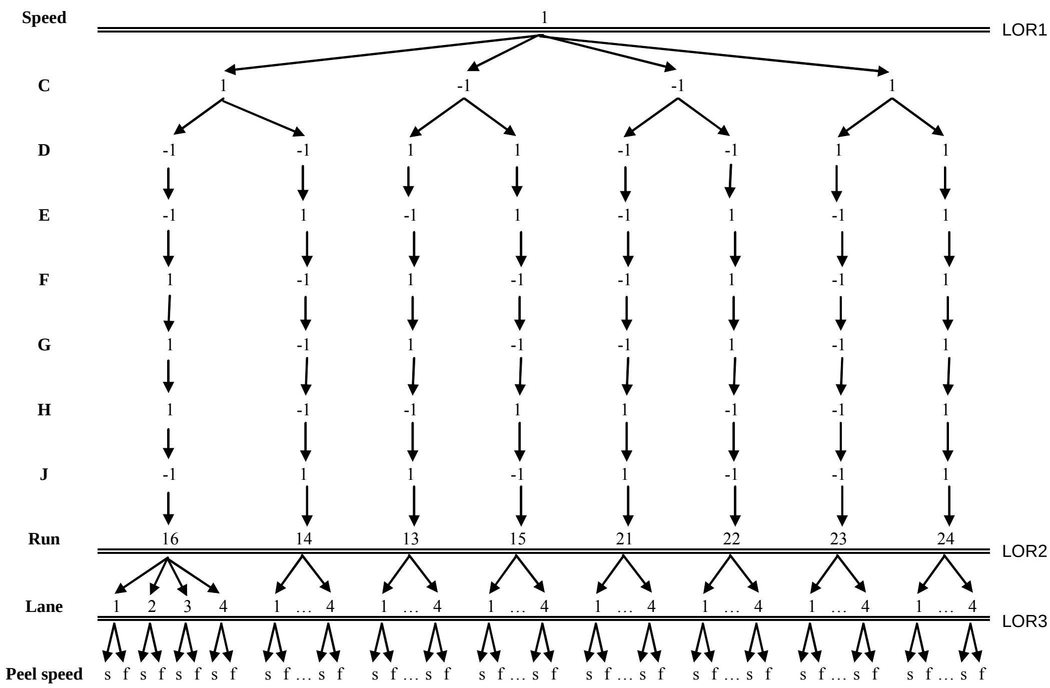


Figure 12. Factor relation diagram for TB experiment with high level of speed.

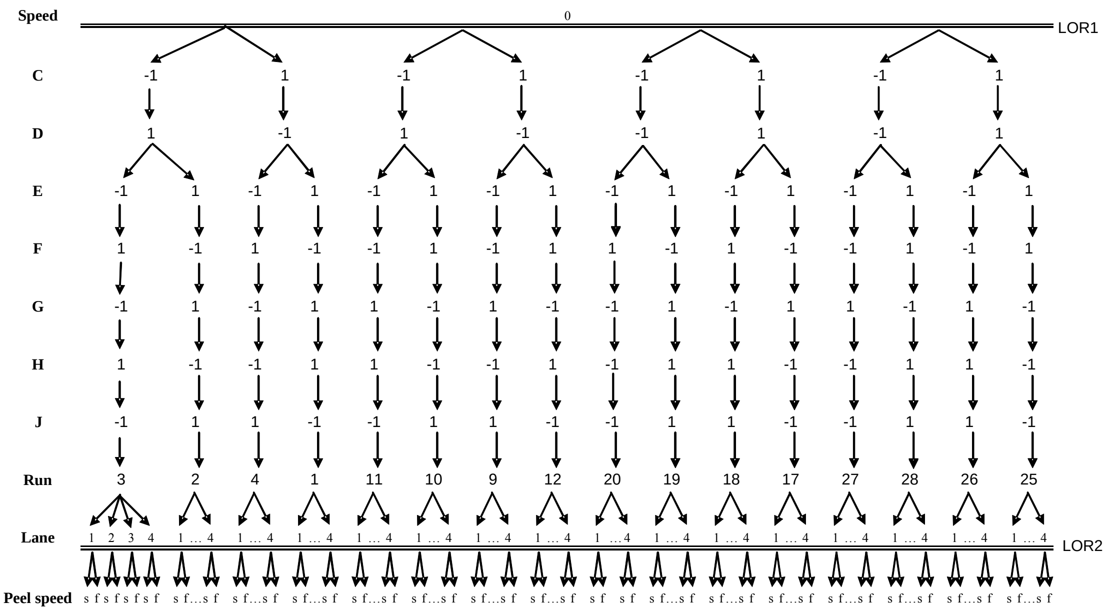


Figure 13. Factor relation diagram for TB experiment with medium level of speed.

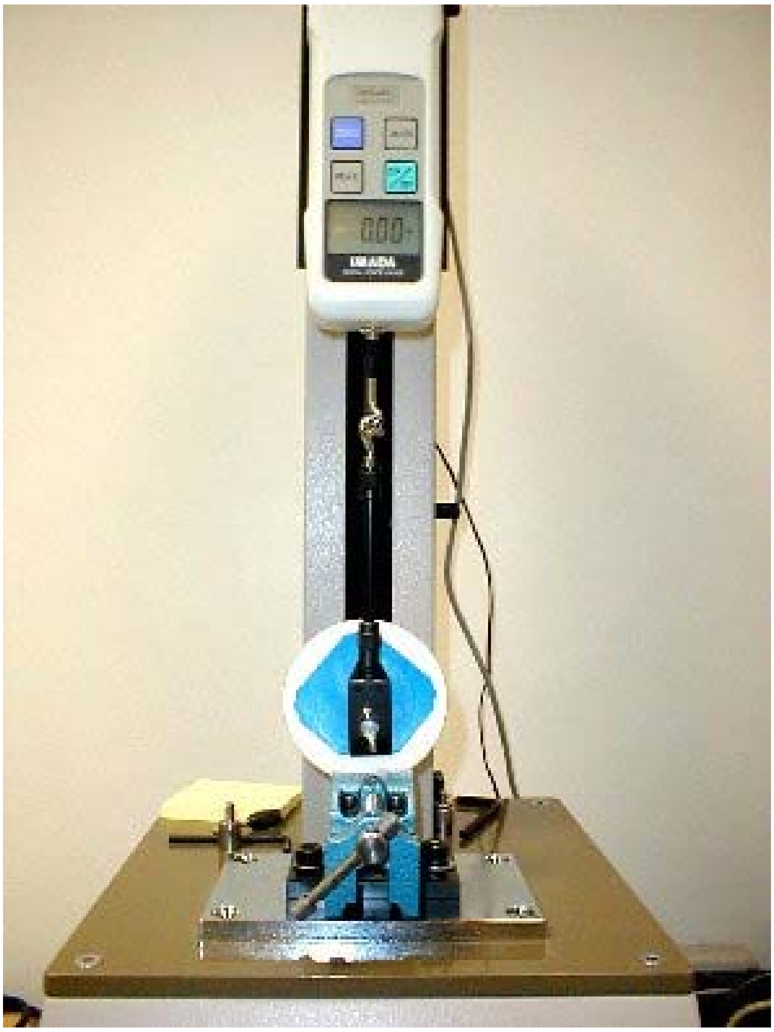


Figure 14. Imada Peel Tester

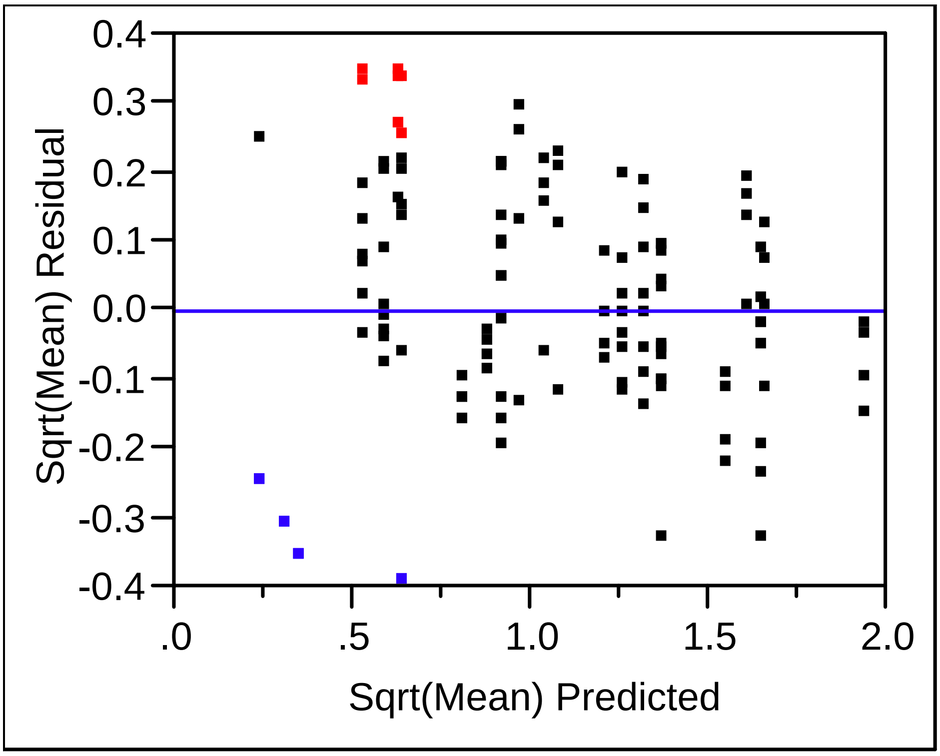


Figure 15. Residual by predicted plot for the square root of the mean of means response.

Summary of Fit

RSquare	0.875514
RSquare Adj	0.867145
Root Mean Square Error	0.174724
Mean of Response	1.040036
Observations (or Sum Wgts)	128

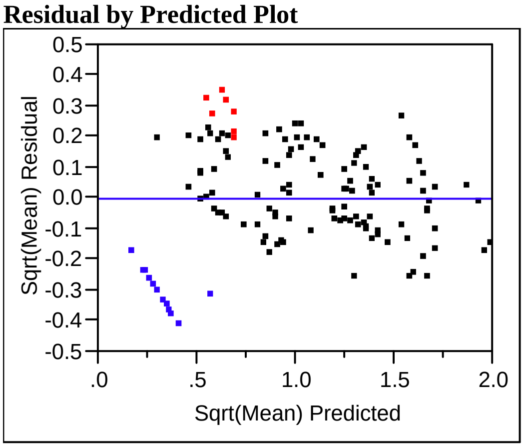
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	25.550180	3.19377	104.6165
Error	119	3.632876	0.03053	Prob > F
C. Total	127	29.183055		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.0400357	0.015444	67.34	<.0001
Speed	-0.286247	0.02184	-13.11	<.0001
C	0.0541485	0.015444	3.51	0.0006
D	0.0866005	0.015444	5.61	<.0001
J	-0.3661946	0.015444	23.71	<.0001
C*D	0.1093234	0.015444	7.08	<.0001
Lane[1]	-0.071501	0.026749	-2.67	0.0086
Lane[2]	-0.005305	0.026749	-0.20	0.8431
Lane[3]	0.0217959	0.026749	0.81	0.4168

Residual by Predicted Plot



Effect Details
Lane
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	0.9685350	0.03088708	0.96854
2	1.0347303	0.03088708	1.03473
3	1.0618316	0.03088708	1.06183
4	1.0950457	0.03088708	1.09505

Figure 16. Analysis of the square root of the mean(mean) with Speed, C, D, J, CD and lane effects.

Summary of Fit

RSquare	0.857157
RSquare Adj	0.851303
Root Mean Square Error	0.099979
Mean of Response	0.526661
Observations (or Sum Wgts)	128

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	7.3178493	1.46357	146.4171
Error	122	1.2194992	0.01000	Prob > F
C. Total	127	8.5373485		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5266613	0.008837	59.60	<.0001
Speed	-0.138649	0.012497	-11.09	<.0001
C	0.0248045	0.008837	2.81	0.0058
D	0.0407884	0.008837	4.62	<.0001
J	-0.2051652	0.008837	23.22	<.0001
C*D	0.0564553	0.008837	6.39	<.0001

Residual by Predicted Plot

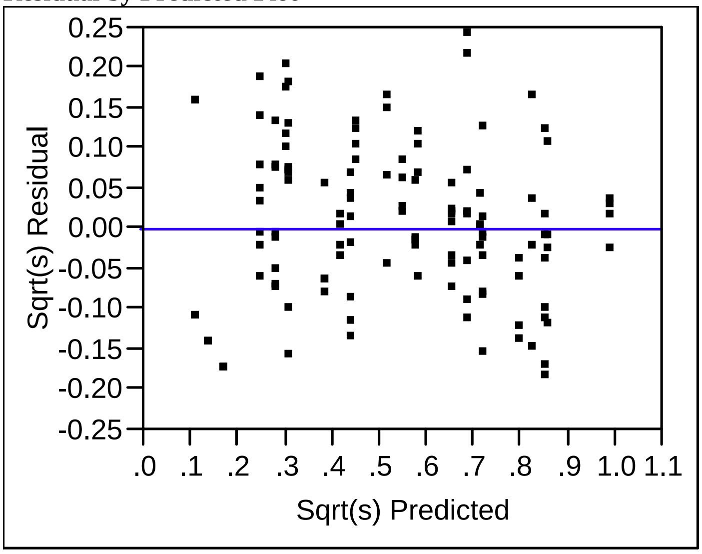
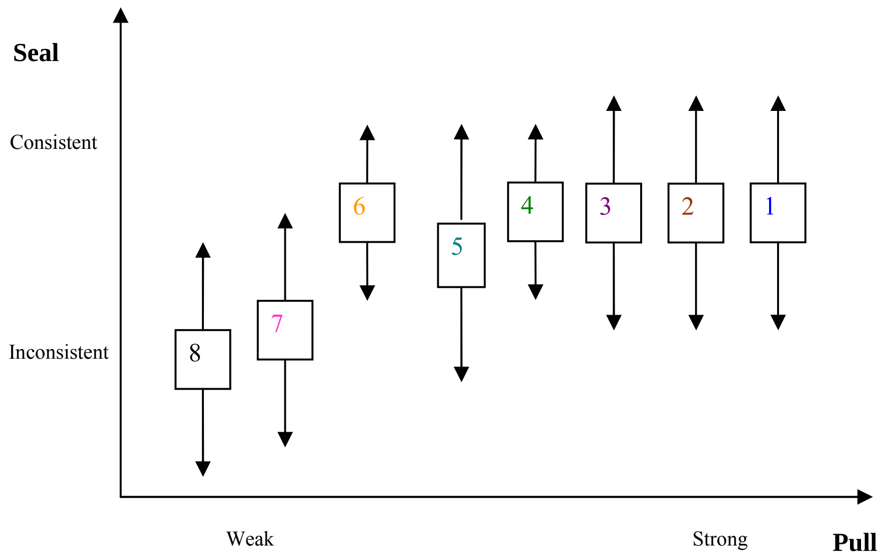


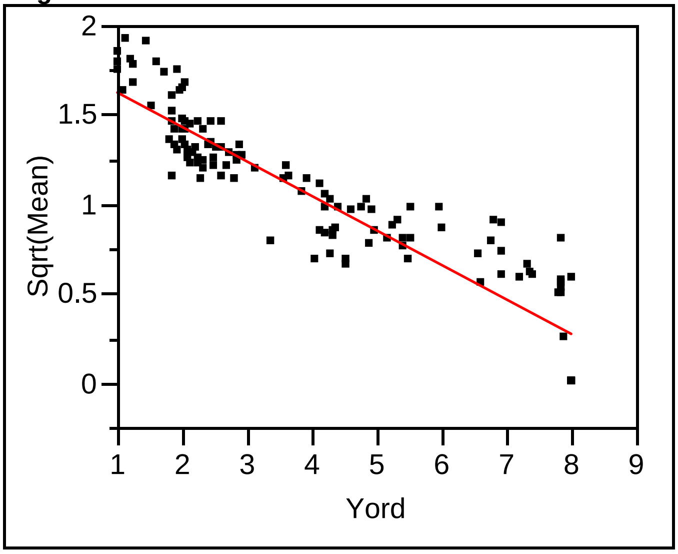
Figure 17. Analysis of the square root of the mean of the standard deviation with Speed, C, D, J and CD effects.



Categories:
1 – completely covered seal area
2 – paper inside seal
3 – paper on sealing edge and/or outside
4 – perfect seal
5 – inconsistent seal with good pull
6 – good looking seal, but has weak pull
7 – inconsistent seal, weak pull
8 – almost no seal or no seal at all

Figure 18. Redefined categories on a two-dimensional scale: consistency of the seal and amount of pull.

Response Sqrt(Mean)
Regression Plot



Summary of Fit

RSquare	0.850808
RSquare Adj	0.849624
Root Mean Square Error	0.185888
Mean of Response	1.040036
Observations (or Sum Wgts)	128

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	24.829184	24.8292	718.5507
Error	126	4.353871	0.0346	Prob > F
C. Total	127	29.183055		<.0001

Parameter Estimates

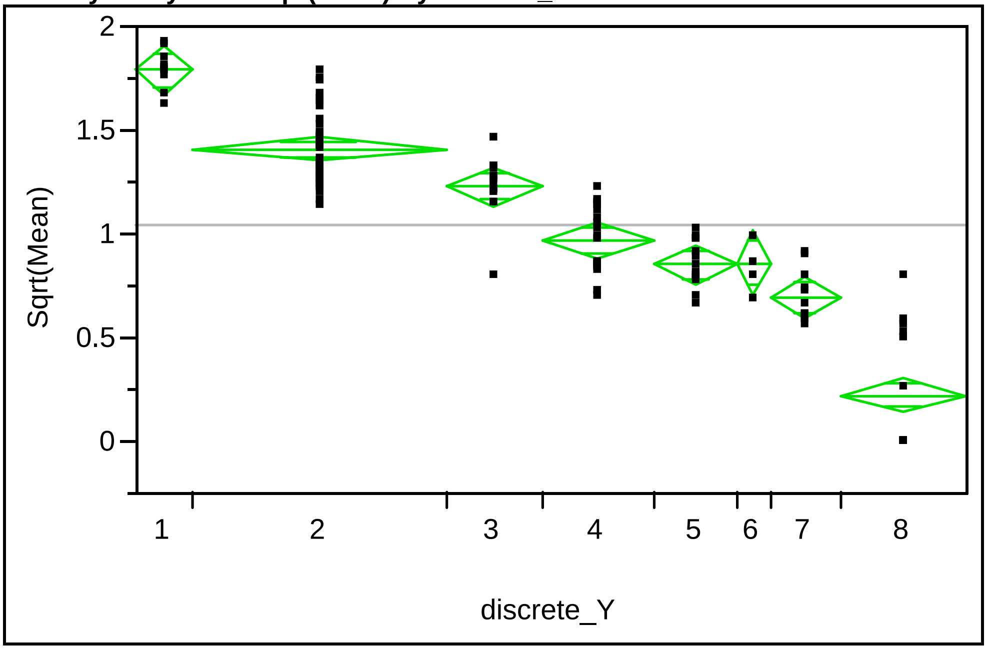
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.8284475	0.03369	54.27	<.0001
Yord	-0.192782	0.007192	-26.81	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Yord	1	1	24.829184	718.5507	<.0001

Figure 19. Relationship between the ordinal and continuous responses.

Oneway Analysis of Sqrt(Mean) By discrete_Y



Oneway Anova
Summary of Fit

Rsquare	0.87436
Adj Rsquare	0.867032
Root Mean Square Error	0.174799
Mean of Response	1.040036
Observations (or Sum Wgts)	128

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
discrete_Y	7	25.516510	3.64522	119.3019	<.0001
Error	120	3.666545	0.03055		
C. Total	127	29.183055			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	9	1.78887	0.05827	1.6735	1.9042
2	39	1.40798	0.02799	1.3526	1.4634
3	15	1.22637	0.04513	1.1370	1.3157
4	17	0.96796	0.04239	0.8840	1.0519
5	13	0.85101	0.04848	0.7550	0.9470
6	5	0.85977	0.07817	0.7050	1.0145
7	11	0.69416	0.05270	0.5898	0.7985
8	19	0.22447	0.04010	0.1451	0.3039

Std Error uses a pooled estimate of error variance

Figure 20. Relationship between the continuous response and the actual ordinal *Y* rounded to the nearest category.

VITA

Oksoun Yee was born in Yuzhno-Sakhalinsk, Russia. She graduated from Novosibirsk State University, Russia with a Bachelor of Science in Mathematics and a Master of Science in Mathematics. She earned a Master of Science in Statistics and Doctorate in Business Administration with a concentration in Statistics from the University of Tennessee, Knoxville.

Oksoun was awarded the Graduate Student Excellence Award from the Department of Statistics, University of Tennessee, in 2002. She is currently employed by Schering-Plough Corporation as a research statistician.