Statistica Sinica **8**(1998), 67-85

NOISE FACTORS, DISPERSION EFFECTS, AND ROBUST DESIGN

David M. Steinberg and Dizza Bursztyn

Tel-Aviv University

Abstract: There has been great interest recently in the use of designed experiments to improve quality by reducing the variation of industrial products. A major stimulus has been Taguchi's robust design scheme, in which experiments are used to detect factors that affect process variation. We study here one of Taguchi's novel ideas, the use of noise factors to represent varying conditions in the manufacturing or use environment. We show that the use of noise factors can dramatically increase power for detecting factors with dispersion effects, provided the noise factors are explicitly modeled in the subsequent analysis.

Key words and phrases: Interactions, off-line QC, parameter design, product array experiments, quality improvement, Taguchi methods, transmitted variation.

1. Introduction

Planned experiments have become a major tool for quality improvement during the last decade, stimulated in large part by the quality engineering ideas of Genichi Taguchi (Kackar (1985), Phadke (1989), Nair (1992)). A major goal of Taguchi's quality improvement experiments is to determine which *design factors* (i.e., controllable process parameters) have *dispersion effects* (i.e., affect variability) and thereby to find settings of the design factors that will minimize variability. *Robust design* refers to quality engineering activities aimed at achieving that goal. One of the novel ideas in Taguchi's work on industrial experiments is the use of *noise factors*, which are impossible or too expensive to control during product manufacture or use but can be set at fixed levels in an experiment and varied jointly with design factors.

Two examples will help clarify the aims and methods of robust design experiments. Pignatiello and Ramberg (1985) described a study to identify process conditions that would consistently produce leaf springs for motor vehicles with a free height of 8 inches. Four design factors were studied: furnace temperature, heating time, transfer time (from the furnace to a press that forms the camber of the spring), and hold down time in the press. The final production step called for submersing the spring assembly in a hot oil quench. The process engineers suspected that the oil temperature influenced spring length. This temperature was difficult to control in regular production but could be monitored for inclusion in the experiment as a noise variable, with the goal of identifying settings of design factors that would minimize its effect on spring length.

Engel (1992) described a robust design experiment on an injection molding process. The goal of this experiment was to find process settings that would consistently obtain a target shrinkage value. Seven process parameters were included as design factors. Three factors, percent regrind, moisture content, and ambient temperature, were included as noise factors. These factors could be controlled for the experiment, but would not be controlled in regular production.

This article will focus on two questions: (i) should noise factors be included in an experiment? (ii) if they are included, how should the experimental data be analyzed? We show that the use of noise factors can substantially improve the ability to detect dispersion effects, but only if the noise factors are explicitly modeled. Our study was motivated in part by Gunter (1988) (see also Carroll and Ruppert (1988)), who noted that power to detect dispersion effects can be quite low when conventional replication is used. We show that the use of product array experiments can provide valuable information on dispersion effects with much smaller samples. Among practitioners, a popular method of analyzing robust design experiments with noise factors is to score each design factor combination by a "signal-to-noise (SN) ratio", as advocated by Taguchi (1987). Our results show that the SN analysis fails to exploit the potential of noise factors.

The paper is organized as follows. In Section 2, we present the most commonly used experimental plans for noise factors and some simple models that relate noise factors to dispersion effects. In Section 3 we present a study of statistical power in a simple setting. The power study proves the value of using noise factors in robust design experiments. In Section 4 we discuss extensions of the results to more complex and realistic settings. We illustrate the ideas in Section 5 by re-analyzing the data from the leaf spring experiment. We conclude with discussion and summary in Section 6.

2. Designs, Models and Analyses

Most robust design experiments, following the work of Taguchi (1987), have used product arrays for joint study of design and noise factors. In these designs, separate arrays are chosen for the design and the noise factors and then each combination in the design factor array is paired with each combination in the noise factor array, producing a matrix of data. The analysis recommended by Taguchi involves two steps: (i) summarize the data for each design factor combination, then (ii) study how the summary measure depends on the design factors. To test for dispersion effects of the design factors, Taguchi typically uses the logarithm of the coefficient of variation as a summary statistic.

Recent statistical research on robust design experiments has favored a "response model" analysis, which includes effects for both noise factors and design factors. This idea is implicit in Easterling (1985) and has since been espoused by Lorenzen and Villalobos (1990), Shoemaker, Tsui and Wu (1991), Steinberg and Bursztyn (1994), Tsui (1996), Welch, Yu, Kang and Sacks (1990), and many of the discussants in Nair (1992).

The simplest response model of interest is one that includes main effects for all the factors together with design factor by noise factor interactions. The model can be written as follows. Let Y denote the quality characteristic of interest, $X_1, \ldots, X_k$ the design factors and $N_1, \ldots, N_t$ the noise factors. Then

$$\mathrm{E}\{Y(X_1, \ldots, X_k, N_1, \ldots, N_t)\} = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{u=1}^{t} \alpha_u N_u + \sum_{i=1}^{k} \sum_{u=1}^{t} \alpha_{iu} X_i N_u. \quad (1)$$

We will assume that all the observations are independent and have common variance σ^2. This assumption is discussed briefly in Section 6.

In actual production, we can set the levels of the design factors, but the noise factors will vary according to their natural distributions. We assume in equation (1) that the noise factors have been coded so that N_u is the deviation of the noise factor from its expectation. Then, following Myers et al. (1992), if the design factors are set to $X_1, \ldots, X_k$, the process mean will be

$$\beta_0 + \sum_{i=1}^{k} \beta_i X_i \quad (2)$$

and the process variance will be

$$\sigma^2 + \mathrm{Var}_N \Big\{ \sum_{u=1}^{t} [\alpha_u + \sum_{i=1}^{k} \alpha_{iu} X_i] N_u \Big\} = \sigma^2 + \alpha(X)' V \alpha(X), \quad (3)$$

where $\alpha_u(X) = \alpha_u + \sum_{i=1}^{k} \alpha_{iu} X_i$ is the effect of N_u when the design factors are set at X, and V is the random variation covariance matrix of the noise factors. This matrix can be estimated from data external to the experiment, using observed process data for noise factors that reflect variable production conditions or field data for environmental factors.

The design factor by noise factor interactions link the design factors to the process variance. If X_i interacts with N_u, then we can attempt to choose a level of X_i for which the effect of N_u is close to 0. Shoemaker et al. (1991) showed that Taguchi's product array designs enable estimation of all these interactions.

The following simple analysis is suggested:

1. Perform a standard factorial analysis of the $k+t$ factor experiment to estimate all main effects and design factor by noise factor interactions. A normal (or half-normal) probability plot of all the orthogonal contrasts is useful for screening out the important effects.
2. Minimize the process variance of Y by choosing levels of the design factors so that each of the sums $\alpha_u + \sum_{i=1}^{k} \alpha_{iu} X_i$ is close to 0.
3. Adjust the process mean to a target level, if that is one of the goals of the experiment, using design factors that have strong main effects but minimal interactions with noise factors.

The end result is a process that is on-target and has minimal transmitted variability. The key feature of this procedure is step 2, where design factor by noise factor interactions are used to choose levels of the design factors that neutralize the effects of the noise factors and thereby minimize transmitted variation. The idea of exploiting these interactions was also discussed by Lorenzen and Villalobos (1990), Myers et al. (1992), Shoemaker et al. (1991), and Shoemaker and Tsui (1993). See also sections 2, 3 and 6 of Nair (1992).

The above "algorithm" for analysing robust design experiments must, of course, be complemented by common sense, process knowledge and good data analysis. For example, it may be desirable, as suggested by Box (1988), to transform the raw data before conducting the analysis. The choice of some design factor levels will entail trade-offs between the process mean and the process variance. An advantage of the response model approach is that it makes these trade-offs explicit and easy to compute.

3. Noise Factors and Power to Detect Dispersion Effects

3.1. Testing for dispersion effects

Including noise factors in an experiment may be costly and difficult. It will often be much simpler to assess dispersion from "conventional" replicate observations at design factor settings. Is the use of noise factors worth the trouble? If they are included, should their effects be modeled explicitly or is the row-summary analysis advocated by Taguchi (1987) preferable? In this section, we focus on the ability to detect dispersion effects in the simple setting of a two-level factorial design with just one control factor X and one noise factor N. The results for this simple case are useful in analyzing more realistic experiments with many factors, which we discuss in Section 4.

For now, we assume that

$$Y(X,N) = \beta_0 + \beta_X X + \beta_N N + \beta_{XN} XN + \epsilon, \tag{4}$$

where $\epsilon \sim N(0, \sigma^2)$. Recall from Section 2 that X has a dispersion effect if it interacts with N. We compare several design and analysis strategies in terms of their power for level α tests of the hypotheses

$$H_0 : \beta_{XN} = 0$$
$$H_1 : \beta_{XN} \neq 0.$$

Although we do not recommend a rigid testing approach to data analysis, the power comparisons provide an excellent basis for comparing the strategies.

An important decision in planning the experiment is whether or not to include the noise factor N. If so, we assume that a 2^2 design will be used with $m/2$ observations in each cell. If not, we assume that m observations are taken at each of two levels of X and that N varies at random according to a $N(0, \sigma_N^2)$ distribution. At the analysis stage, one must decide whether to use row-summary statistics or to explicitly model the noise factor effects. Each of the following sections examines a different design and analysis strategy. An important quantity in describing the results is $\delta = \beta_{XN}\sigma_N/\sigma$.

3.2. Noise factor not included in the design

When the noise factor is not included in the experiment, one can test for a dispersion effect using the ratio of the sample variances of the m observations at each level of X, $F = s^2(+)/s^2(-)$. For example, the analysis of Nair and Pregibon (1988) reduces to an F-test in our simple setting. The use of Taguchi's signal-to-noise ratio for nominal-is-best type problems is essentially equivalent to comparing the sample variances of the *logged* data (Box (1988)).

The F-test is two-sided because either level of X might have the lower variance. Thus we reject H_0 if $F > F_c$ or $F < 1/F_c$, where the critical value F_c is the upper $\alpha/2$ quantile of the $F(m-1, m-1)$ distribution. The power of the F-test depends on the true process variances at each level of X, which are

$$\sigma^2(-) = \sigma^2 + \sigma_N^2(\beta_N - \beta_{XN})^2 \qquad \text{and} \qquad \sigma^2(+) = \sigma^2 + \sigma_N^2(\beta_N + \beta_{XN})^2 \quad (5)$$

when $X = -1$ and when $X = 1$, respectively. Then $(\sigma^2(-)/\sigma^2(+))F$ has an $F(m-1, m-1)$ distribution and the power of the F-test is

$$\begin{aligned}
\text{Power}(\beta_{XN}, \beta_N) &= \Pr\{F > F_c\} + \Pr\{F < 1/F_c\} \\
&= \Pr\{(\sigma^2(-)/\sigma^2(+))F > (\sigma^2(-)/\sigma^2(+))F_c\} \\
&\quad + \Pr\{(\sigma^2(-)/\sigma^2(+))F < \sigma^2(-)/(\sigma^2(+)F_c)\} \\
&= \Pr\{F(m-1, m-1) > (\sigma^2(-)/\sigma^2(+))F_c\} \\
&\quad + \Pr\{F(m-1, m-1) < \sigma^2(-)/(\sigma^2(+)F_c)\} \\
&= \Pr\{F(m-1, m-1) > (\sigma^2(-)/\sigma^2(+))F_c\} \\
&\quad + \Pr\{F(m-1, m-1) > (\sigma^2(+)/\sigma^2(-))F_c\}. \qquad (6)
\end{aligned}$$

The last equality follows from the fact that if W has an F distribution with equal degrees of freedom in numerator and denominator, then $1/W$ has the same distribution.

The power of the F-test depends on β_N as well as the other parameters. Figure 1 illustrates the effect of β_N on the power of the F-test when $\delta = 1$ and $m = 8$. The variance ratio is then

$$\frac{\sigma^2(-)}{\sigma^2(+)} = \frac{1 + ((\beta_N/\beta_{XN}) - 1)^2}{1 + ((\beta_N/\beta_{XN}) + 1)^2}.$$

The power is highest when β_N and β_{XN} are approximately equal, so that the noise variable makes almost no contribution to $\sigma^2(-)$ and the variance ratio is far from 1. However, moderate reductions in process variance are difficult to detect, as noted by Gunter (1988).

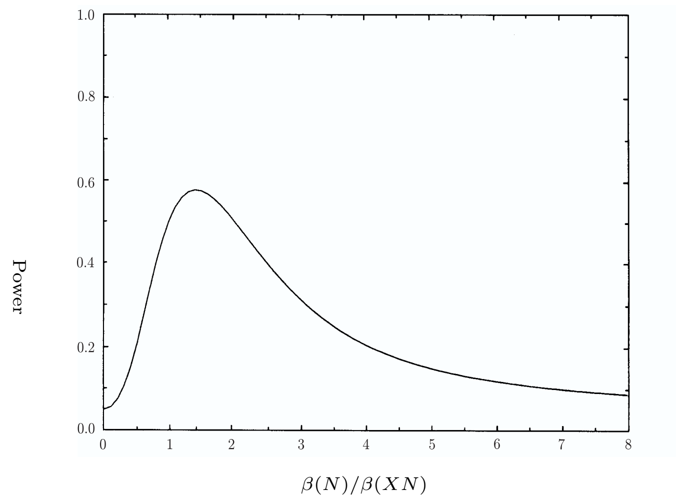

Figure 1. The power of the F-test when the noise factor is not controlled and $\delta = 1$. When β_N/β_{XN} is not close to 1, the F-test has almost no ability to detect the dispersion effect of X.

In the comparisons that follow, we exaggerate the efficiency of the F-test by maximizing the power over β_N. The relevant bound is given in the following Lemma, which we prove in the appendix.

Lemma. *When the noise factor is not controlled,*

$$\text{Power}(\beta_{XN}, \beta_N) \leq \Pr\{F(m-1, m-1) > rF_c\} + \Pr\{F(m-1, m-1) > (1/r)F_c\}, \tag{7}$$

where

$$r = \frac{(1+\delta^2)^{1/2} - 1}{(1+\delta^2)^{1/2} + 1} \tag{8}$$

is the minimal value of the variance ratio $\sigma^2(-)/\sigma^2(+)$ *for fixed* β_{XN}.

3.3. Noise factor included in the design with response model analysis

Suppose that the noise factor has been included in the experiment at levels $\pm\gamma\sigma_N$. A standard factorial analysis computes the interaction term between X and N as

$$XN = \bar{Y}(XN = 1) - \bar{Y}(XN = -1). \tag{9}$$

The regression coefficient $\hat{\beta}_{XN}$ is related to XN, for our scaling convention, by

$$\hat{\beta}_{XN} = XN/(2\gamma\sigma_N). \tag{10}$$

The obvious way to test H_0 is via the t-ratio,

$$t = \sqrt{(m/2)}XN/s, \tag{11}$$

where s^2 is the standard unbiased estimator of σ^2. The level α t-test rejects H_0 if $|t| > t_{2m-4,1-\alpha/2}$, the upper $\alpha/2$ quantile of the t distribution with $2m-4$ degrees of freedom. The t-test can be used only if $m > 2$; otherwise there are no residual degrees of freedom to estimate σ^2. The power is

$$\text{Power}(\beta_{XN}) = \Pr\{|t| > t_{2m-4,1-\alpha/2}\} \tag{12}$$

and can be computed from the fact that t has a non-central t distribution with $2m-4$ degrees of freedom and non-centrality parameter $\gamma\delta\sqrt{2m}$.

3.4. Noise factor included in the design with row summary analysis

We now consider what happens when the noise factor has been included in the experiment but the analysis is based on row summaries, as in Section 3.2. Again the natural test statistic is the variance ratio $F = s^2(+)/s^2(-)$, perhaps after appropriate transformation of the original data.

When N has been controlled, the F-test is no longer valid! Under model (4), the sample variance at each level of X includes a "between groups" sum of squares component that reflects the different means at the two levels of N. Thus $s^2(+)$ and $s^2(-)$ have non-central χ^2 distributions with $m-1$ degrees of freedom and non-centrality parameters

$$\lambda(+) = m(\beta_N + \beta_{XN})^2\gamma^2\sigma_N^2/(2\sigma^2) \tag{13a}$$

and

$$\lambda(-) = m(\beta_N - \beta_{XN})^2\gamma^2\sigma_N^2/(2\sigma^2), \tag{13b}$$

respectively, and the variance ratio has a doubly non-central F distribution. Analyzing the data as though the noise factors had not been included may lead not only to reduced power, but to incorrect rejection probabilities. If, for example, H_0 is true and most of the process variance is contributed by the noise factor, the distribution of the variance ratio will be heavily concentrated about 1 and the true level of significance will be much smaller than α. A test for H_0 based on the true null distribution is impractical, since the rejection region depends on β_N.

Although the F-test is invalid when N has been controlled, such row summary analyses are commonly used by practitioners (see many articles in the proceedings of the American Supplier Institute's symposia) and have been applied extensively in statistical journals (Engel (1992), Ghosh and Duh (1992), Nair and Pregibon (1988), Tuck, Lewis and Cottrell (1993), Vining and Myers (1990) and Rosenbaum (1994)). Thus we think it is relevant to consider the ability of this analysis to detect dispersion effects. Let F^* denote a random variable with the appropriate doubly non-central F distribution and let F_c be the upper $\alpha/2$ quantile of the central $F(m-1, m-1)$ distribution Then

$$\text{Power}(\beta_{XN}, \beta_N) = \Pr\{F^* > F_c\} + \Pr\{F^* < 1/F_c\}. \tag{14}$$

Using the technique in Scheffé (1959) to approximate the doubly non-central F leads to

$$\text{Power}(\beta_{XN}, \beta_N) \approx \Pr\{F(m_1, m_2) > aF_c\} + \Pr\{F(m_1, m_2) < a/F_c\}, \tag{15}$$

where $a = (m-1+\lambda(-)^2)/(m-1+\lambda(+)^2)$ and the degrees of freedom are

$$m_1 = \frac{[m-1+\lambda(+)^2]^2}{m-1+2\lambda(+)^2} \qquad \text{and} \qquad m_2 = \frac{[m-1+\lambda(-)^2]^2}{m-1+2\lambda(-)^2}.$$

For our comparisons we again exaggerate the efficiency of this test by maximizing the power over β_N. We have no analytic bound here, so the maximization is numerical. Inspection of the results shows that the best value of β_N here almost eliminates the effect of the noise factor at one of the levels of the design factor and is similar to that derived for the case when the noise factor is not controlled.

3.5. Power comparison

Figures 2 and 3 graph the power as a function of $\delta = \beta_{XN}\sigma_N/\sigma$ for $m = 4$ and $m = 8$. For the t-test, the power is graphed for $\gamma = 0.5$ and for $\gamma = 1$. For the F-test with N controlled, the power is graphed only for $\gamma = 1$. For both F-tests, we have graphed the upper bounds to the power.

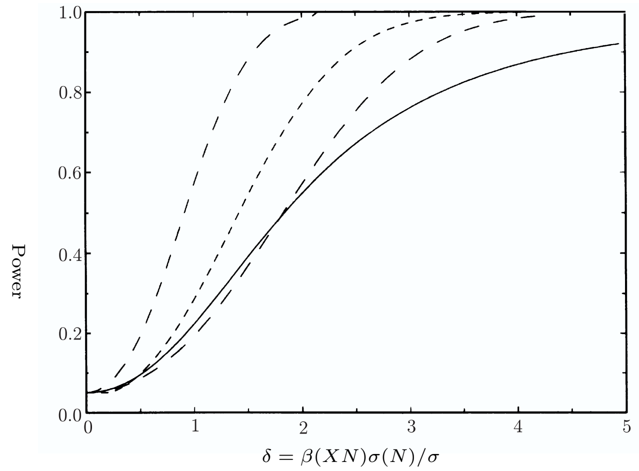

Figure 2. The power of the dispersion tests with 4 observations at each setting of X. The two long dashed lines are the t-tests – the higher line corresponds to $\gamma = 1.0$ and the lower line to $\gamma = 0.5$; the solid line is the F-test when N is not controlled; the short dashed line is the F-test when N is controlled with $\gamma = 1.0$. For both F-tests, the graph shows maximal power over β_N for given δ.

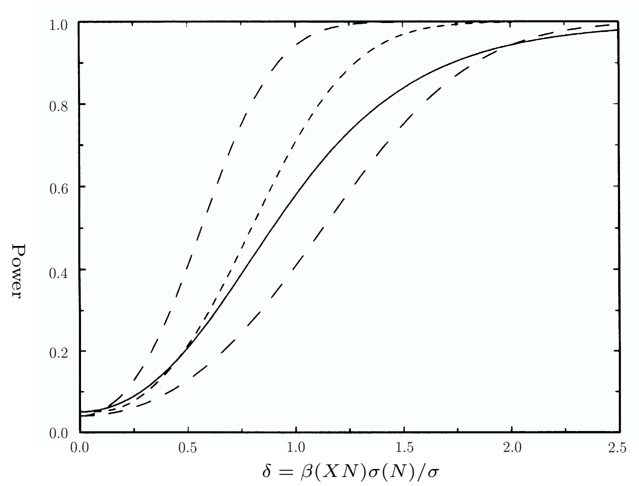

Figure 3. The power of the dispersion tests with 8 observations at each setting of X. The two long dashed lines are the t-tests – the higher line corresponds to $\gamma = 1.0$ and the lower line to $\gamma = 0.5$; the solid line is the F-test when N is not controlled; the short dashed line is the F-test when N is controlled with $\gamma = 1.0$. For both F-tests, the graph shows maximal power over β_N for given δ.

The striking fact is that the t-test has much greater power for detecting whether X has a dispersion effect than do either of the F-tests. Even when the noise factor is controlled at $\pm 0.5\sigma_N$, the power of the t-test is roughly equal to the *best attainable* power of the F-tests. From Figure 1, we know that the power of the F-test can fall well below the maximum.

The power comparison answers the two questions that we posed in Section 1: (i) Including noise factors in a robust design experiment can significantly increase the power to detect dispersion effects and can thus overcome the difficulties foreseen by Gunter (1988) and by Carroll and Ruppert (1988); (ii) It is essential that the analysis explicitly model the noise factors when they have been controlled in the experiment, rather than computing summary measures for each design factor combination.

Bérubé and Nair (1998) reached similar conclusions about the usefulness of noise factors, but from a slightly different perspective. They showed that the ability to estimate dispersion effects is an increasing function of the percentage of variance in the response that can be attributed to noise factors. Thus the experiment will be most informative if it includes noise factors with large effects on the response. These results complement our conclusions that using noise factors can improve power to detect dispersion effects.

3.6. Choice of noise factor levels

If the noise factor is controlled, how should one choose its experimental levels? Setting the levels at $\pm\gamma\sigma_N$, the non-centrality parameter for the power calculation is proportional to $\gamma\sqrt{(m/2)}$. Thus doubling γ has roughly the same effect on power as quadrupling the sample size, which suggests that γ should be chosen as large as possible. A drawback to choosing γ large is that equation (4) may cease to be an adequate model for the data (see Tribus and Szonyi (1989)). Equation (12) for calculating power will then be incorrect.

We recommend $\gamma = 1.5$ as a good default value, with $\gamma = 1$ a minimal choice. Three reasons guide our advice: (1) The most important noise factors will often have an effect on the response that, if not perfectly linear, is at least monotonic. (2) For noise factors with a monotonic effect, large values of γ will be effective for identifying levels of the control factors at which the noise factor has a small effect. (3) The potential gains in power are substantial.

4. Dispersion Effects in Multi-Factor Experiments

In this section we consider the effect of both design and analysis on the ability to detect dispersion effects in the common practical setting of experiments with many factors and no replication. We assume that equation (1) provides a reasonable model when both design factors and noise factors are controlled in the experiment and that a product array design with m observations in each row has been used to facilitate the use of row summary analyses. We assume that the noise factors vary randomly if they are not included.

We again conclude that there is much to be gained from including noise factors in the experiment, but only if they are explicitly modeled in the analysis. Row summary analyses typically have low power, and, in certain circumstances, can be totally misleading, missing relevant dispersion effects, identifying nonexistent effects, and advocating sub-optimal design factor settings. In the following subsections, we review the results for full factorial analyses, develop some theory for studying row summary analyses, and then compare the methods in several specific settings.

4.1. Full factorial analysis with noise factors

When noise factors are included in a robust design experiment, we can conduct a standard analysis, focusing on main effects and design factor by noise factor interactions. A normal, or half-normal, probability plot can be used to identify the important contrasts. The power for detecting a design factor by noise factor interaction, and hence a dispersion effect, is again given by equation (12), with appropriate adjustments to reflect the sample size and the error degrees of freedom.

4.2. Row summary analyses

A number of different row summary statistics might be used to look for dispersion effects. We will focus on comparing row variances. Our conclusions will be relevant to the analysis proposed by Nair and Pregibon (1988), which is based on modeling the logged variances, and to Taguchi's signal-to-noise ratio for reducing variance about a target value.

It is useful to decompose each row variance into components for each noise factor and for residual variation. Denote the m observations in row i by $y_{i,j}$, the row average by $\bar{y}_i$ and the design factor combination by X^i. The row variance is

$$s_i^2 = \sum (y_{i,j} - \bar{y}_i)^2/(m-1) = \Big(\sum_{u=1}^{t} SSN_{u,i} + SSE_i\Big)/(m-1), \qquad (16)$$

where $SSN_{u,i}$ is the sum of squares for the uth noise factor and SSE_i is the residual sum of squares for the data in row i.

If the noise factors were controlled in the experiment, this is a standard ANOVA decomposition and

$$SSN_{u,i} = m[\bar{y}_i(N_u = 1) - \bar{y}_i(N_u = -1)]^2/4. \qquad (17)$$

If N_u was set at $\pm\gamma_u\sigma_u$, where $\mathrm{Var}\{N_u\} = \sigma_u^2$ under natural variation, then $\alpha_u(X^i)$ is the effect of N_u for row i and

$$\mathrm{E}\{s_i^2\} = \sigma^2 + \frac{m}{m-1}\sum[\alpha_u(X^i)]^2\gamma_u^2\sigma_u^2. \tag{18}$$

If the noise factors were not controlled, we can think of equation (16) as decomposing the row variance into noise factor and residual components via the regression model

$$Y_{i,j} = \mu(X^i) + \sum \alpha_u(X^i)N_{u(i,j)} + \epsilon_{i,j}, \tag{19}$$

where $N_{u(i,j)}$ is the value of the noise factor N_u for the jth observation in row i. In this case,

$$\mathrm{E}\{s_i^2\} = \sigma^2 + \frac{m}{m-1}\alpha(X^i)'V\alpha(X^i), \tag{20}$$

where V is the covariance matrix of the noise factors under random variation.

Thus in both cases the sum of squares component for the noise factors increases as a function of the distance of $\alpha_u(X^i)$ from the origin and similar conclusions will hold for the ability to detect dispersion effects.

4.3. Several design factors interact with one noise factor

Suppose there is just one active noise factor, N_u, which interacts with the design factors $X_1,\ldots,X_r$. The effect of N_u for a setting X of the design factors will be

$$\alpha_u(X) = \alpha_u + \sum_{i=1}^{r}\alpha_{iu}X_i.$$

The full factorial analysis will detect the interactions if they are large relative to the error standard deviation.

The sensitivity of the row variance comparison will depend on the *relative* sizes of the variances, as in Section 3. If α_u is much larger than the interactions, then all the rows will have similar variances and analyzing them will have low power to detect dispersion effects.

Now consider what happens if $\alpha_u \approx 0$. If just one design factor, say X_1, interacts with N_u, then $\alpha_u(X)$ will have similar magnitude, but opposite sign, depending on the setting of X_1. The row variances will all be similar and no dispersion effects will be found. However, if X_1 is a continuous factor, the effect of N_u can be neutralized by setting X_1 to an intermediate level (see Lorenzen and Villalobos (1990)). If several design factors interact with N_u, there may be many settings of the design factors at which the interactions cancel one another and thus minimize transmitted variation. Estimating the interaction coefficients easily allows these settings to be identified. The row variances can differ substantially, since at other settings, the interactions may amplify one another, resulting in large variances. However, no single factor adjustment will make the slope of N_u closer to 0, so an analysis of row variances will not show any main effects for the design factors. Instead, the analysis will detect *interactions* among design factors. Moreover, if the design factor plan has resolution III, such an interaction may be incorrectly interpreted as the main effect of a third factor which has no dispersion effect at all. (See Steinberg and Bursztyn (1994) for an example.)

The most favorable situation for the row variance analysis occurs when α_u is large enough so that $\alpha_u(X)$ has constant sign, but is near 0 for at least some of the experimental settings of the design factors. This mirrors the analysis in Section 3 that found highest power for the $F-$test when $\beta_N \approx \beta_{XN}$. The variance ratios will be most extreme if one of the design factors has a dominant interaction rather than if all the factors have moderate interactions. But that brings us back to the case studied in Section 3, where power is much lower than with the response model analysis.

4.4. Interactions with several noise factors

From equations (18) and (20), the expected sum of squares for row i has a noise factor component and a residual component. The row summary analysis will be sensitive to a dispersion effect of a design factor X only if the noise factor component dominates the sum of squares and is close to 0 for one experimental level of X. If just one of the noise factors dominates the sum, we are again back in the setting of Section 3, with low power for the row summary analyses. If several noise factors are important, it may be that setting X near its high level neutralizes some of the noise factors but setting X near its low level is better for neutralizing others. In this case, an analysis of the row variances is unlikely to detect any dispersion effect for X. In fact, the situation may be much more promising and knowledge of such an interaction pattern can be quite useful for reducing process variance. If several design factors interact with the noise factors, it will often be possible to find joint settings of the design factors that neutralize all the noise factors. Alternatively, the setting of X might be chosen to neutralize some of the noise factors with the others controlled by measures to reduce their natural variation such as improving the production environment or tightening tolerances. The response model analysis effectively exposes all these possibilities.

5. Example

In this section we illustrate the ideas of the previous sections by analyzing the experiment described by Pignatiello and Ramberg (1985), whose goal was to identify process conditions that would consistently produce leaf springs with a free height of 8 inches. The experimental plan was a 2^{4-1} factorial in the design factors crossed with the high and low levels of the noise factor, the oil quench temperature. Three replicates were observed at each setting in the design.

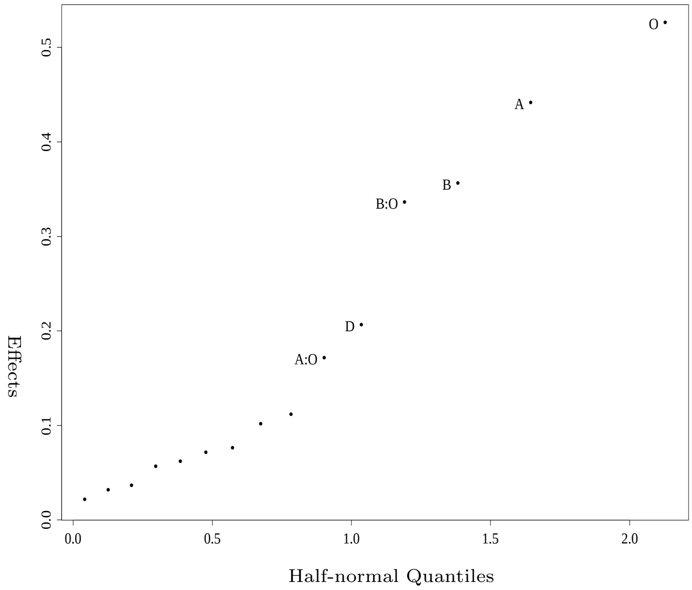

Figure 4. Half-normal probability plot of the factor effects in the leaf spring experiment.

Table 1. The leaf spring experiment – major effects and interactions on free spring height. The four design factors are Furnace Temperature (A), Heating Time (B), Transfer Time (C), and Hold Down Time (D). The noise factor is Oil Quench Temperature (O).

Main Effects	
O	-0.26
A	0.22
B	-0.18
D	0.10
C	-0.03

Interactions	
BO	0.165
AO	0.085

The response model approach treats the leaf spring experiment as a 2^{5-1} fractional factorial replicated three times. Figure 4 shows a half-normal probability plot of the 15 effects and Table 1 lists the main effects and the large interactions. The oil quench temperature (O) has a large main effect, -0.26, confirming that uncontrolled production variation may cause substantial variation in free height. Two design factors have large interactions with oil quench temperature, heating time (B) and furnace temperature (A). The effect of the oil quench temperature can be neutralized by setting these factors at their high levels. The estimated effect of the oil quench temperature is then $-0.26 + 0.165 + 0.085 = -.01$.

Pignatiello and Ramberg (1985) presented an analysis of these data based on the signal-to-noise ratio $SN = -20\log(s/\bar{y})$ that has been proposed by Taguchi (1987) for experiments whose goal is to minimize variance about a target value. In this analysis, SN is computed from the average $\bar{y}$ and the standard deviation s of the 6 observations at each of the eight design factor combinations. The results are summarized in Table 2. The SN analysis indicates that heating time (B) has a dispersion effect and should be set at its high level. However, interpretation of any additional effects is complicated by the large interactions. The next largest main effect is due to transfer time (C), but is smaller in magnitude than an effect which may be due to the BC interaction. If we assume that the BC interaction really is responsible for that large effect, then the analysis suggests setting transfer time to its low level. The dispersion effect of the furnace temperature is not detected in this analysis.

Table 2. The leaf spring experiment – major effects and interactions on the signal-to-noise ratio. The four design factors are labeled A-D.

Main Effects	
B	9.27
C	−4.57
D	2.94
A	−0.34

Interactions	
$AD + BC$	−5.19
$AC + BD$	3.45
$AB + CD$	−2.30

The analysis of summary measures from the leaf spring experiment misses the simple solution found by the response model analysis for neutralizing the effect of the oil quench temperature. Moreover, the failure occurs in a setting that is ideally conducive to the row summary analysis. As noted in the last section, this analysis should be most efficient when the effect of the noise factor is always in the same direction, but is close to 0 for some combinations of the design factors. In the leaf spring experiment, the effect of the oil quench temperature at the eight design combinations ranges from -0.54 to 0.01. The variations in the effect of the oil temperature are effectively explained by the settings of the furnace temperature and the heating time, but the SN analysis fails to identify the dispersion effects.

6. Discussion

Robust design experiments with controlled noise factors can be an extremely valuable tool for quality improvement. In this section we discuss some additional practical issues that add further weight to our recommendations.

We have assumed throughout that the residual variance is constant. What happens if the residual variance depends on the design factors? This situation might occur if some important noise factors could not be controlled in the experiment. Non-constant residual variance will be reflected in row summary measures of variation and has been advanced by Taguchi (1987), Taguchi (1992), and Phadke (1992) as one of the reasons for using row summary analyses. This feature strikes us as small compensation for the great loss in power to detect dispersion effects directly related to the noise factors. Moreover, two critical issues are ignored. First, the row summary analyses will still suffer from the lack of power noted by Gunter (1988) – differences in residual variance will be almost impossible to detect unless they are extremely large. Second, differences in residual variation can also be studied from the response model analysis by examining residuals (Asscher (1995), Bergman and Hynén (1997)), an approach that we think is much more direct and effective than analyzing row summaries.

It has been argued that the product array designs proposed by Taguchi (1987) for robust design experiments are impractical because they require too many runs (see the discussion in Nair (1992)). For identifying dispersion effects, our results show that good designs must enable estimation of design factor by noise factor interactions. To estimate all these interactions, product arrays constructed from Plackett-Burman designs have minimal sample sizes and thus can be an effective choice. Engineering considerations may suggest that some interactions are much more likely than others, in which case smaller designs can often be found (see Shoemaker et al. (1991) and Welch et al. (1990)).

Finally, it is important to consider robust design experiments in the larger context of quality improvement. Often the empirical analysis of these experiments will indicate settings of design factors that provide a higher quality process. However, the knowledge gained from a thorough analysis of the experiment may lead to improvements well beyond those immediate empirical confines. We believe that one of the greatest benefits of including noise factors in the design and performing a response model analysis is the specific information that is obtained on which noise factors really affect variation and which design factors, if any, can be used to neutralize their effects (see also Shoemaker et al. (1991) and Tsui (1996)). The knowledge that a noise factor has a large effect but does not interact with any of the design factors can stimulate engineers to propose other design factors that are good candidates for interactions or to devise measures to control its natural variation. Knowing that a noise factor has a small effect may enable engineers to scale back costly programs for its control. Row summary analyses, focusing only on immediate empirical gains, ignore the detailed information that can be obtained from relating the results to the noise factors and leave many valuable stones unturned.

We are convinced that there are great advantages to including noise factors in robust design experiments. However, the common practice of analyzing row summaries sacrifices many of those advantages and can lead to incorrect conclusions and poor process settings. We have shown that direct modeling of noise factor effects and their interactions with design factors greatly improves the ability to detect dispersion effects, facilitates determination of design factor settings that minimize variation, and provides valuable information that can be used to achieve further improvements in quality.

Acknowledgements

The work of D. M. Steinberg was carried out in part while visiting the Center for Quality and Productivity Improvement and the Department of Statistics at the University of Wisconsin. The visit to CQPI was aided by a grant from the Alfred P. Sloan Foundation. We would like to thank our colleagues at CQPI for their helpful comments.

Appendix – Maximal Power for the F-Test

Given the model assumptions in Section 3, we now prove the lemma stated there.

Lemma. *Suppose the noise factor is not controlled, m observations are taken at each level of the control factor, and an F-test is used to compare the sample variances. For fixed β_{XN}, the power of the test satisfies the inequality*

$$\text{Power}(\beta_{XN}, \beta_N) \leq \Pr\{F(m-1, m-1) > rF_c\} + \Pr\{F(m-1, m-1) < r/F_c\},$$

where F_c is the upper $\alpha/2$ quantile of the $F(m-1, m-1)$ distribution, $r = ((1+\delta^2)^{1/2} - 1)/((1+\delta^2)^{1/2} + 1)$, and $\delta = \beta_{XN}\sigma_N/\sigma$.

Proof. From equation (6), the power depends on β_{XN} and β_N through the true variance ratio, $R = \sigma^2(-)/\sigma^2(+)$. The power at R is equal to the power at $1/R$, so we can assume without loss of generality that $R \leq 1$. Let $\nu = m - 1$. We then have

$$\begin{aligned}\text{Power}(\beta_{XN}, \beta_N) &= \Pr\{F(\nu,\nu) > RF_c\} + \Pr\{F(\nu,\nu) < R/F_c\} \\ &= 1 - G(RF_c) + G(R/F_c),\end{aligned}$$

where G denotes the distribution function of $F(\nu, \nu)$. First, we will show that the power is monotone decreasing for $R < 1$. Differentiating the power with respect to R gives

$$-F_c g(RF_c) + g(R/F_c)/F_c, \tag{21}$$

where g, the density function for $F(\nu,\nu)$, is

$$g(u)=\frac{\Gamma(\nu)u^{(\nu-2)/2}}{[\Gamma(\nu/2)]^2(1+u)^\nu}. \tag{22}$$

Using (22) in (21), we find that the derivative of the power function is proportional to

$$R^{(\nu-2)/2}[(F_c+R)^{-\nu}-(1+F_cR)^{-\nu}]. \tag{23}$$

The sign of the derivative is thus the same as the sign of $(1+RF_c)-(F_c+R)=(1-R)(1-F_c)$. Since $F_c>1$, the derivative is negative when $R<1$, proving the monotonicity.

The power of the F test is therefore maximized when R is minimal. For fixed β_{XN}, simple calculus shows that r is the minimal value of R.

References

Asscher, J. (1995). Design and analysis of robust design experiments with two components of variance. Unpublished Ph. D. dissertation, The Technion – Israel Institute of Technology.

Bergman, B. and Hynén, A. (1997). Testing for dispersion effects from unreplicated designs. *Technometrics* **39**, 191-198.

Bérubé, J. and Nair, V. N. (1998). Exploiting the inherent structure in robust parameter design experiments. *Statist. Sinica* **8**, 43-66.

Box, G. E. P. (1988). Signal-to-noise ratios, performance criteria, and transformations. *Technometrics* **30**, 1-40 (with discussion).

Carroll, R. J. and Ruppert, D. (1988). Discussion of "Signal-to-noise ratios, performance criteria, and transformations", by G. E. P. Box. *Technometrics* **30**, 30-31.

Easterling, R. G. (1985). Discussion of "Off-line quality control, parameter design, and the Taguchi method", by R. N. Kackar. *J. Qual. Tech.* **17**, 191-192.

Engel, J. (1992). Modeling variation in industrial experiments. *Appl. Statist.* **41**, 579-593.

Ghosh, S. and Duh, Y.-J. (1992). Determination of optimum experimental conditions using dispersion main effects and interactions of factors in replicated factorial experiments. *J. Appl. Statist.* **19**, 367-378.

Gunter, B. (1988). Discussion of "Signal-to-noise ratios, performance criteria, and transformations", by G. E. P. Box. *Technometrics* **30**, 32-35.

Kackar, R. N. (1985). Off-line quality control, parameter design, and the Taguchi method. *J. Qual. Tech.* **17**, 176-209 (with discussion).

Lorenzen, T. J. and Villalobos, M. A. (1990). Understanding robust design, loss functions, and signal to noise ratios. General Motors Research Laboratories Report GMR-7118.

Myers, R. H., Khuri, A. I. and Vining, G. (1992). Response surface alternatives to the Taguchi robust parameter design approach. *Amer. Statist.* **46**, 131-139.

Nair, V. N., Editor, (1992). Taguchi's parameter design: A panel discussion. *Technometrics* **34**, 127-161.

Nair, V. N. and Pregibon, D. (1988). Analyzing dispersion effects from replicated factorial experiments. *Technometrics* **30**, 247-257.

Phadke, M. S. (1989). *Quality Engineering Using Robust Design.* Prentice Hall, New York.

Phadke, M. S. (1992). Discussion in "Taguchi's parameter design: A panel discussion", edited by V. N. Nair. *Technometrics* **34**, 127-161.

Pignatiello, J. J. and Ramberg, J. S. (1985). Discussion of "Off-line quality control, parameter design, and the Taguchi method", by R. N. Kackar. *J. Qual. Tech.* **17**, 198-206.

Rosenbaum, P. R. (1994). Dispersion effects from fractional factorials in Taguchi's method of quality design. *J. Roy. Statist. Soc.* Ser. *B* **56**, 641-652.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley, New York.

Shoemaker, A. C. and Tsui, K.-L. (1993). Response model analysis for robust design experiments. *Comm. Statist. Ser. A* **22**, 1037-1064.

Shoemaker, A. C., Tsui, K.-L. and Wu, C. F. J. (1991). Economical experimentation methods for robust design. *Technometrics* **33**, 415-427.

Steinberg, D. M. and Bursztyn, D. (1994). Dispersion effects in robust design experiments with noise factors. *J. Qual. Tech.* **26**, 12-20.

Taguchi, G. (1987). *System of Experimental Design*, Vol. 1 & 2. UNIPUB, White Plains, NY.

Taguchi, S. (1992). Discussion in "Taguchi's parameter design: A panel discussion", edited by V. N. Nair. *Technometrics* **34**, 127-161.

Tribus, M. and Szonyi, G. (1989). An alternative view of the Taguchi approach. *Qual. Prog.* May, 46-52.

Tsui, K.-L. (1996). A critical look at Taguchi's modelling approach for robust design. *J. Appl. Statist.* **23**, 81-95.

Tuck, M. G., Lewis, S. M. and Cottrell, J. I. L. (1993). Response surface methodology and Taguchi: A quality improvement study from the milling industry. *Appl. Statist.* **42**, 671-681.

Vining, G. G. and Myers, R. H. (1990). Combining Taguchi and response surface philosophies: A dual response approach. *J. Qual. Tech.* **22**, 38-45.

Welch, W. J., Yu, T. K., Kang, S. M. and Sacks, J. (1990). Computer experiments for quality control by parameter design. *J. Qual. Tech.* **22**, 15-22.

Department of Statistics and Operations Research, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Ramat Aviv 69978, Israel.

E-mail: dms@math.tau.ac.il

E-mail: dizza@math.tau.ac.il

(Received November 1995; accepted October 1996)