Center for Quality and Productivity Improvement
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

Report No. 60

An Analysis of Taguchi's Method of Confirmatory Trials

Søren Bisgaard and Neil Diamond

October 1990

The Center for Quality and Productivity Improvement cares about your reactions to our reports. Please direct comments (general or specific) to: Report Editor, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705; (608) 263–2520. All comments will be forwarded to the author(s).

CQPI
UNIVERSITY OF WISCONSIN
Report No. 60
Report Series
in Quality and Productivity

An Analysis of Taguchi's Method of Confirmatory Trials

Søren Bisgaard
Center for Quality and Productivity Improvement *and* Department of Industrial Engineering
University of Wisconsin
Madison, Wisconsin

Neil Diamond
Footscray Institute of Technology
Victoria, Australia

ABSTRACT

Taguchi has suggested a new method of confirmatory trials that is intended to test for possible interactions effects. The method is being promoted by many recent authors. A careful analysis of this method, however, shows that confirmatory trials are ineffective in detecting the presence of interaction effects. In fact in many cases, it is shown that the probability of detecting an interaction effect decreases rather than increases as the size of the interaction effect increases. Thus, we advice practitioners not to adopt this new method as a test for interactions.

KEYWORDS: *Fractional factorials, interaction effects, confirmatory trials, power of the test, prediction interval, alias matrix*

This work was sponsored by the National Science Foundation Grant DDM–8808138 and Grant ECD–872 1545.

CQPI Report No. 60, October 1990

An Analysis of Taguchi's Method of Confirmatory Trials

Søren Bisgaard and Neil Diamond

Taguchi has suggested a new method of confirmatory trials that is intended to test for possible interactions effects. The method is being promoted by many recent authors. A careful analysis of this method, however, shows that confirmatory trials are ineffective in detecting the presence of interaction effects. In fact in many cases, it is shown that the probability of detecting an interaction effect decreases rather than increases as the size of the interaction effect increases. Thus, we advice practitioners not to adopt this new method as a test for interactions.

1. INTRODUCTION

Dr. Taguchi has stimulated among engineers a much welcomed interest in statistics and specifically in experimental design. A large and ever increasing number of manufacturing and design engineers are today using factorial designs in their daily work. This positive development seems in large part to be due to Taguchi's many practical examples of the use of experimental design in the discrete parts manufacturing industries.

The introduction of Taguchi Methods has, however, also been surrounded by some controversy. This controversy seldom concerns the engineering aspects of these methods which are widely acknowledged as very good. Instead the critique is almost exclusively focussed on some of the statistical methods employed by Dr. Taguchi to achieve his engineering goals. This conclusion is more thoroughly elaborated upon in Box, Bisgaard, and Fung (1988).

One of the primary reasons for criticizing Taguchi's use of statistics is his unusual attitude toward interaction effects in fractional factorial experiments. In particular, his method of provisionally ignoring interaction effects and then checking later for their possible existence via the use of *confirmatory trials* is novel and has been a topic of much discussion. In this article we will explore the merit of this method which is now in wide use among engineers.

In a recent article Taguchi discusses the key ideas of his philosophy. He writes (Taguchi, 1988, p. 27):

> The greatest general difference between Taguchi methods and traditional design of experiments is the belief that it is useless and counterproductive to assign interactions, whether or not there are interactions. Orthogonal arrays should be used to help us to discover experimental failures when interactions exist.

At a more technical level Taguchi explains earlier in the same article (Taguchi, 1988, p. 25):

> In order to prove that there are no interactions, one assigns only main effects to the orthogonal array and finds the optimum conditions. One makes the assumption that there is no interaction and estimates the process average. If there is little interaction, the estimated value of the process average and its confirmatory experiment value will match well. If the effect of interactions is large, these two values will differ greatly...

In the past few years it seems to have been taken for granted that it is possible to detect interaction effects by this approach. Several authors have even praised Taguchi's method as a real innovation. However, based on the analysis presented in this paper we conclude that his approach is ineffective in detecting the presence of interaction effects. In fact in many cases, we will show that the probability of detecting an interaction effect decreases instead of increases as the size of the interaction effect increases.

2. SOME PRELIMINARY LEAST SQUARES THEORY

In response surface work it is common to distinguish between the model we "hope" is true and the model we "fear" might be true. See for example Box and Draper (1987). The model we "hope" is true might be a linear main-effects-only model. However, the model we "fear" could include various interaction effects. In matrix notation we may write the main-effects-only model as

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon \tag{1}$$

and the more elaborate model including possible interaction effects as

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon . \tag{2}$$

In both cases we assume that $E\{\varepsilon\} = 0$ and $V\{\varepsilon\} = \mathbf{I}_n\sigma^2$.

If we analyzed the data **y** as if (1) is correct, the least squares estimates of the main effects are

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} \ . \tag{3}$$

If model (1) is correct, (3) provides unbiased estimates of β_1. However, if instead model (2) is true, the estimates given by (3) will be biased. In fact $E\{\hat{\beta}_1\} = \beta_1 + \mathbf{A}\beta_2$ where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$. The matrix **A** is also known as the *alias* or the *bias matrix.*

Based on the assumption that model (1) is correct we might proceed to compute an Analysis of Variance table as given in Table 1.

Table 1
Analysis of Variance Table under the assumption that Model (1) is correct.

Source	Sum of Squares	df's	Expected Sums of Squares
Main Effects	$\hat{\beta}_1\mathbf{X}_1'\mathbf{y}$	ν_1	$\beta_1'\mathbf{X}_1'\mathbf{X}_1\beta_1 + \nu_1\sigma^2$
Residuals	$\mathbf{y}'\mathbf{y} - \hat{\beta}_1\mathbf{X}_1'\mathbf{y}$	$n-\nu_1$	$(n-\nu_1)\sigma^2$
Total	$\mathbf{y}'\mathbf{y}$	n	

Now suppose instead Model (2) is correct. In that case the expected sums of squares are changed as given in Table 2 below. The derivations of the expected sums of squares are supplied in Appendix as Result A1.

Table 2
Analysis of Variance Table set up as if Model (1) is correct when in fact Model (2) is.

Source	Sum of Squares	df's	Expected Sums of Squares
Main Effects	$\hat{\beta}_1\mathbf{X}_1'\mathbf{y}$	ν_1	$(\beta_1 + \beta_2\mathbf{A})'\mathbf{X}_1'\mathbf{X}_1(\beta_1 + \beta_2\mathbf{A}) + \nu_1\sigma^2$
Residuals	$\mathbf{y}'\mathbf{y} - \hat{\beta}_1\mathbf{X}_1'\mathbf{y}$	$n-\nu_1$	$\beta_2'(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1\mathbf{A})\beta_2 + (n-\nu_1)\sigma^2$
Total	$\mathbf{y}'\mathbf{y}$	n	

Note in particular in Table 2 that the residual sum of squares is inflated by the interaction effects β_2. Thus, it will be difficult to know what "large" means in the context of a large deviation between an observed and a predicted value of a confirmatory trial. In Section 4 we will study this problem in more detail by way of two special cases that will show some of the possible consequences. However, let us here continue with the general theory.

3. THE POWER OF TAGUCHI'S TEST

Suppose model (1) is correct. Then the best prediction (in the least squares sense) of a new observation y_{n+1} at a design point given by the row vector $\mathbf{x}_{01}$ is $\hat{z}_{n+1} = \mathbf{x}_{01}\hat{\beta}_1$. It is readily seen that

$$E\{\hat{z}_{n+1}\} = E\{\mathbf{x}_{01}\hat{\beta}_1\} = \mathbf{x}_{01}\beta_1$$

and $V\{\hat{z}_{n+1}\} = \sigma^2 \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'$.

Now suppose we conduct q confirmatory trials at $\mathbf{x}_{01}$ yielding the observations $y_{n+i}, i = 1,\ldots,q$ then $E\{y_{n+i}\} = \mathbf{x}_{01}\beta_1$ and $V\{y_{n+i}\} = \sigma^2$. Denote by $\bar{y}_q$ the average of the q confirmatory trials. Consequently

$$E\{\bar{y}_q - \hat{z}_{n+1}\} = 0 \text{ and}$$

$$V\{\bar{y}_q - \hat{z}_{n+1}\} = \sigma^2(1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}').$$

Adding the assumption that $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1,\ldots,n+q$ we have that

$$\frac{\bar{y}_q - \hat{z}_{n+1}}{\sigma\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}} \sim N(0,1) \quad (4)$$

If we substitute $s^2 = (\mathbf{y}'\mathbf{y} - \hat{\beta}_1'\mathbf{X}_1'\mathbf{y})/(n - v_1)$ for σ^2 then

$$\frac{\bar{y}_q - \hat{z}_{n+1}}{s\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}} \sim t(n - v_1) \quad (5)$$

where $t(n\text{-}v_1)$ denote the *central* Student's t distribution. Using (5) we can construct a $(1-\alpha)$ 100% *prediction* interval C such that $Pr\{\bar{y}_q \in C\} = 1 - \alpha$ where

$$C = [\mathbf{x}_{01}\hat{\beta}_1 - \Delta; \mathbf{x}_{01}\hat{\beta}_1 + \Delta] \text{ and } \Delta = t(n - v_1)_{1-\alpha/2}\, s\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}.$$

Taguchi's test for interaction effects is based on observing whether or not an average of q confirmatory trial falls outside the interval C. Thus the *power* of his test is $p(\theta) = Pr\{\bar{y}_q \notin C | \theta\}$ where θ is some vector of parameters indexing the possible departures from the model (1).

The prediction interval C is computed on the assumption that model (1) is correct. However, let us now evaluate the consequence of some of the terms of the vector β_2 being active and thus model (2) being more appropriate. In that case both the expectation of the predicted value and the expected value of the observed confirmatory trials are biased from what we obtained above based on model (1). In fact,

$$E\{\hat{z}_{n+1}\} = E\{\mathbf{x}_{01}\hat{\beta}_1\} = \mathbf{x}_{01}\beta_1 + \mathbf{x}_{01}\mathbf{A}\beta_2 \text{ and } E\{\bar{y}_q\} = \mathbf{x}_{01}\beta_1 + \mathbf{x}_{02}\beta_2$$

where $\mathbf{x}_{02}$ denote the coordinates of the design point $\mathbf{x}_{01}$ but in terms of the additional columns of the design matrix coming from model (2). Therefore

$$\begin{aligned} E\{\bar{y}_q - \hat{z}_{n+1}\} &= \mathbf{x}_{01}\beta_1 + \mathbf{x}_{02}\beta_2 - \mathbf{x}_{01}\beta_1 - \mathbf{x}_{01}\mathbf{A}\beta_2 \\ &= \mathbf{x}_{02}\beta_2 - \mathbf{x}_{01}\mathbf{A}\beta_2 \end{aligned}$$

The variance of the prediction, $V\{\hat{z}_{n+1}\} = \sigma^2 \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'$ remains the same as before. Therefore

$$V\{\bar{y}_q - \hat{z}_{n+1}\} = \sigma^2(1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}')$$

and hence

$$\frac{\bar{y}_q - \hat{z}_{n+1}}{\sigma\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}} \sim N(\delta, 1)$$

where the (normalized) bias is

$$\delta = \delta(\beta_2) = \frac{(\mathbf{x}_{02} - \mathbf{x}_{01}\mathbf{A})\beta_2}{\sigma\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}}. \tag{6}$$

From Table 2 we have that

$$E\{s^2\} = \beta_2'(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1\mathbf{A})\beta_2/(n - \nu_1) + \sigma^2.$$

Thus the error variance estimate might also be biased. Consequently $(n\text{-}\nu_1)s^2/\sigma^2$ is no longer distributed as a central $\chi^2(n\text{-}\nu_1)$ but as a *non-central* $\chi^2(n\text{-}\nu_1, \lambda)$ distribution with non-centrality parameter

$$\lambda = \lambda(\beta_2) = \beta_2'(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1\mathbf{A})\beta_2/\sigma^2 \tag{7}$$

Hence the t-ratio used for computing the prediction interval is no longer distributed as a standard Student's t-distribution but as a *doubly non-central* $t(n\text{-}\nu_1, \delta, \lambda)$, see Robbins (1948). See also Appendix Result A2. Consequently the power of Taguchi's test for H_0: $\beta_2 = 0$ versus H_1: $\beta_2 \neq 0$ is

$$\begin{aligned} Pr\{\bar{y}_q \notin C|\beta_2\} &= 1 - Pr\left\{ t(n-\nu_1)_{\alpha/2} \leq \frac{\bar{y}_q - \mathbf{x}_{01}\hat{\beta}_1}{s\sqrt{1/q + \mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'}} \leq t(n-\nu_1)_{1-\alpha/2} \right\} \\ &= 1 - \int_{t \in C^*} f(t|\delta(\beta_2), \lambda(\beta_2))dt \end{aligned} \tag{8}$$

where $C^* = [t(n-\nu_1)_{\alpha/2}; t(n-\nu_1)_{1-\alpha/2}]$ i.e. the interval spanned by the upper and lower $\alpha/2$ quantiles of the *central* Student's t distribution, and $f(t \mid \delta, \lambda)$ denotes the density function of the *doubly non-central* t distribution.

4. EXAMPLES

In this section we will present two examples that each show some of the draw backs of Taguchi's confirmatory trial method. In particular we will show the power of Taguchi's test for interactions. The problems that are illustrated with these two examples are representative of the problems that we have found with the confirmatory trial methods for other larger experiments.

EXAMPLE 1. A 2^{5-2} FRACTIONAL FACTORIAL EXPERIMENT

A typical yet simple experiment used for screening is the 2^{5-2} fractional factorial design with generators $D = AB$ and $E = AC$. If we think the interaction effects are inert then the full first order model is

$$y = \beta_0 + x_A\beta_A + x_B\beta_B + x_C\beta_C + x_D\beta_D + x_E\beta_E + \varepsilon \quad . \tag{9}$$

But if in fact all the possible two factor interactions are active then the "correct" model is

$$
\begin{aligned}
y = {} & \beta_0 + x_A\beta_A + x_B\beta_B + x_C\beta_C + x_D\beta_D + x_E\beta_E \\
& + x_A x_B \beta_{AB} + x_A x_C \beta_{AC} + x_A x_D \beta_{AD} + x_A x_E \beta_{AE} \\
& + x_B x_C \beta_{BC} + x_B x_D \beta_{BD} + x_B x_E \beta_{BE} \\
& + x_C x_D \beta_{CD} + x_C x_E \beta_{CE} + x_D x_E \beta_{DE} + \varepsilon
\end{aligned} \tag{10}
$$

In matrix notation we can write (9) and (10) as (1) and (2) respectively where, $\beta_1' = (\beta_0, \beta_A, \beta_B, \beta_C, \beta_D, \beta_E)$, $\beta_2' = (\beta_{AB}, \beta_{AC}, \beta_{AD}, \beta_{AE}, \beta_{BC}, \beta_{BD}, \beta_{BE}, \beta_{CD}, \beta_{CE}, \beta_{DE})$,

$$
\mathbf{X}_1 = \begin{bmatrix}
1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

and

$$
\mathbf{X}_2 = \begin{bmatrix}
1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\
-1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}.
$$

Computing the alias matrix we find that the five main effects are confounded as given below:

$$
\beta_1 + \mathbf{A}\beta_2 = \begin{bmatrix}
\beta_0 \\
\beta_A + \beta_{BD} + \beta_{CE} \\
\beta_B + \beta_{AD} \\
\beta_C + \beta_{AE} \\
\beta_D + \beta_{AB} \\
\beta_E + \beta_{AC}
\end{bmatrix}.
$$

The expected mean squares for the main effects for model (9) when in fact (10) is more appropriate is given in the Analysis of Variance table below.

Table 3
Analysis of Variance Table set up as if model (9) is correct when in fact model (10) is.

Source	Sum of Squares	df's	Expected Sums of Squares
Average	$\bar{y}^2$	1	$\mu^2+\sigma^2$
A	$\hat{\beta}_A^2$	1	$(\beta_A+\beta_{BD}+\beta_{CE})^2+\sigma^2$
B	$\hat{\beta}_B^2$	1	$(\beta_B+\beta_{AD})^2+\sigma^2$
C	$\hat{\beta}_C^2$	1	$(\beta_C+\beta_{AE})^2+\sigma^2$
D	$\hat{\beta}_D^2$	1	$(\beta_D+\beta_{AB})^2+\sigma^2$
E	$\hat{\beta}_E^2$	1	$(\beta_E+\beta_{AC})^2+\sigma^2$
Residuals	$\mathbf{y}'\mathbf{y}-\hat{\beta}_1\mathbf{X}_1'\mathbf{y}$	2	$8(\beta_{DE}^2+2\beta_{BC}\beta_{DE}+\beta_{CD}^2+2\beta_{BE}\beta_{CD}+\beta_{BE}^2+\beta_{BC}^2)+2\sigma^2$
Total	$\mathbf{y}'\mathbf{y}$	8	

This Analysis of Variance table vividly shows the problem with Taguchi's confirmatory trial method. From it we can see that both the expected sums of squares of the main effects and the error variance estimate potentially are inflated by two-factor interaction effects. Thus either the prediction will be biased, or the confidence band will be inflated, or both. Which of these possibilities occurs will depend on which interaction terms turn out to be significant. In particular note that the interactions $\beta_{BD}, \beta_{CE}, \beta_{AD}, \beta_{AE}, \beta_{AB}$, and β_{AC} bias the main effect estimates but do not inflate the error variance. On the other hand the interactions $\beta_{BC}, \beta_{DE}, \beta_{BE}$, and β_{CD} inflate the error variance estimate but do not bias the main effect estimates.

Confirmatory trials need not be restricted to the design points of the particular fraction run. In fact it is explicitly part of the Taguchi Method that the "optimal" solution should be sought among any of the possible points in the corresponding full factorial design. Thus for computing the power of Taguchi's test note that for all design points of the full 2^5 factorial design

$$\mathbf{x}_{02}-\mathbf{x}_{01}\mathbf{A}=(x_{AB}-x_D, x_{AC}-x_E, x_{AD}-x_B, x_{AE}-x_C, x_{BC}, x_{BD}-x_A, x_{BE}, x_{CD}, x_{CE}-x_A, x_{DE}) \ .$$

As an example of what might happen, suppose for simplicity that only β_{AB} is active. In that case $\lambda=0$ and

$$\begin{aligned}\delta &= (\mathbf{x}_{02}-\mathbf{x}_{01}\mathbf{A})\beta_2 \Big/ \sigma\sqrt{1/q+\mathbf{x}_{01}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_{01}'} \\ &= (x_{AB}-x_D)\beta_{AB} / \sigma\sqrt{1/q+3/4} \\ &= \begin{cases} 2\beta_{AB}/\sigma\sqrt{1/q+3/4} & \text{if } x_{AB}=-x_D=1 \\ -2\beta_{AB}/\sigma\sqrt{1/q+3/4} & \text{if } x_{AB}=-x_D=-1 \\ 0 & \text{if } x_{AB}x_D=1 \end{cases} \ .\end{aligned}$$

Thus $2^4=16$ of the $2^5=32$ design points have the property that $x_{AB}x_D=x_Ax_Bx_D=1$ and hence for 50% of the predictions $\lambda=\delta=0$. Consequently the power of Taguchi's test (8) remains *constant* and equal to $100\alpha\%$ no matter how large the interaction effect is. At the remaining 50% of the design points $\lambda=0$ but $\delta\neq 0$. Thus for only half of the points is the power of the test a function of the interaction effect. Figure 1 shows the power as functions of the standardized interaction term β_{AB}/σ for $\alpha = 0.05$ when only one confirmatory trial is conducted. Specifically (*i*) shows the the situation where $\lambda = \delta = 0$, and (*ii*) the case when $\lambda = 0$, $\delta \neq 0$.

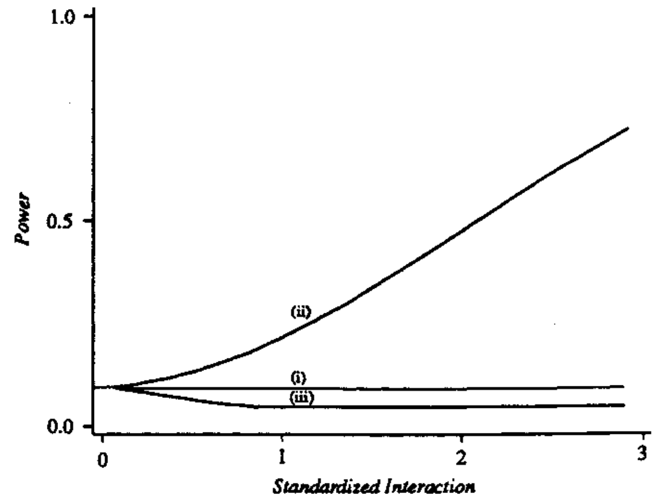

Figure 1. *Power of Taguchi's confirmatory trial versus the standardized interaction effect for the 2^{5-2} design with one confirmatory trial where (i) $\lambda = \delta = 0$, (ii) $\lambda = 0, \delta \neq 0$ and (iii) $\lambda \neq 0, \delta \neq 0$.*

Consider now instead what will happen if one of the interaction effects that are active only bias the the error estimate. As an example suppose only β_{BC} is active. In that case

$$\delta = x_{BC}\beta_{BC} / \sigma\sqrt{1/q + 3/4} \text{ and}$$

$$\lambda = 8\beta_{BC}^2 / \sigma^2 .$$

The power function for this case is shown in Figure 1 (*iii*) for $q = 1$. We see that the power *decreases* as a function of an increasing interaction effect. Figure 2 shows that for $q = \infty$ the situation is qualitatively the same so increasing the sample size will not improve the situation.

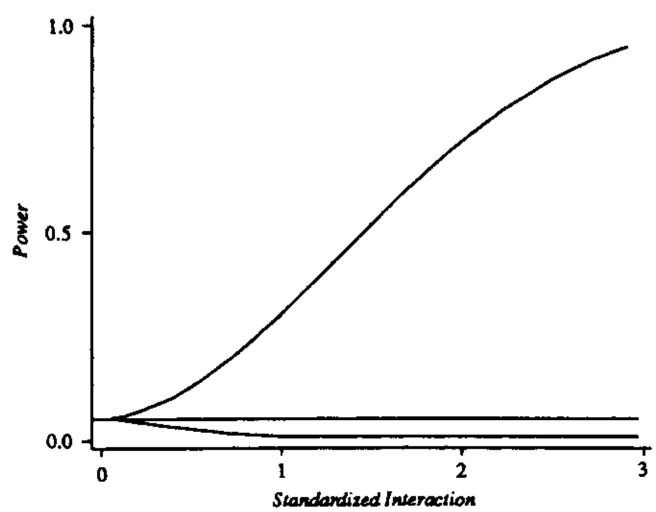

Figure 2. *Power versus the standardized interaction effect for the 2^{5-2} design with ∞ confirmatory trials.*

Most of Taguchi's applications of fractional factorial designs are screening experiments. When conducting screening experiments we are in a state of relative ignorance and thus we are at the mercy of Mother Nature as to which interaction effects turn out to be active. Thus the discussion above shows that it is not likely that confirmatory trials will be of much help in deciding whether interaction effects are present.

EXAMPLE 2. A 3^2 FULL FACTORIAL EXPERIMENT

Many of Taguchi's applications of the confirmatory trial techniques are in the context of three level factorial experiments. For those experiments he finds the "optimum" by maximizing the marginals. This approach is equivalent to fitting a quadratic model without the interaction terms and subsequently finding the "optimum" based on this model. For the 3^2 design his approach would be equivalent to using the following model:

$$y = \beta_0 + x_A\beta_A + x_B\beta_B + x_A^2\beta_{AA} + x_B^2\beta_{BB} + \varepsilon \quad . \tag{11}$$

A full second order response surface model including the interaction effect would on the other hand be written as

$$y = \beta_0 + x_A\beta_A + x_B\beta_B + x_A^2\beta_{AA} + x_B^2\beta_{BB} + x_A x_B\beta_{AB} + \varepsilon \tag{12}$$

These two models can be written in the form of (1) and (2). Doing that we have that

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 \\ 1 & 0 & -1 & -2 & 1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 0 & 1 & -2 \\ 1 & 0 & 0 & -2 & -2 \\ 1 & 1 & 0 & 1 & -2 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 0 & 1 & -2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

$\beta_1' = (\beta_0, \beta_A, \beta_B, \beta_{AA}, \beta_{BB})$ and $\beta_2 = (\beta_{AB})$. The alias matrix is then $\mathbf{A}' = (0, 0, 0, 0, 0)$.

The expected sum of squares computed as if model (11) is correct when in fact model (12) is correct is

$$(\beta_1 + \mathbf{A}\beta_2)'\mathbf{X}_1'\mathbf{X}_1(\beta_1 + \mathbf{A}\beta_2) + \nu_1\sigma^2 =$$

$$9\beta_0^2 + 6\beta_A^2 + 6\beta_B^2 + 18\beta_{AA}^2 + 18\beta_{BB}^2 + \nu_1\sigma^2 \quad .$$

Similarly we find the expected value of the "residual" sum of squares to be

$$\beta_2'(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1\mathbf{A})\beta_2 + (n - \nu_1)\sigma^2 = 4\beta_{AB}^2 + 4\sigma^2 \quad .$$

Consequently $\lambda = 4\beta_{AB}^2 / \sigma^2$ and $\delta = \pm\beta_{AB} / \sigma\sqrt{1/q + x_{01}(X_1'X_1)x_{01}'}$ where $\mathbf{x}_{02} = (x_{AB})$ is the coordinate of the interaction column at the corresponding design point of the factorial design. As indicated in Figure 3, for five out of the nine design points $\mathbf{x}_{AB} = 0$. For those points $\delta = 0$ and the power of Taguchi's test is monotonically decreasing and is therefore less than $\alpha = 0.05$ for all values of the interaction effect.

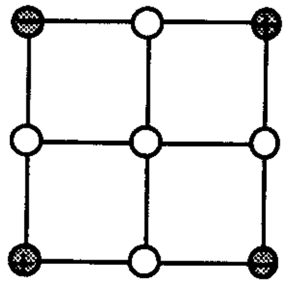

Figure 3. *The nine design points of a 3^2 experiment. The open circles indicate the design point where* $x_{AB} = 0$ *and the shaded circles are the points where* $x_{AB} = \pm 1$.

For the remaining four design points $x_{AB} = \pm 1$. The power functions for this situation with one confirmatory trial are shown in Figure 4. Specifically (*i*) shows the power for any of the five points where $\delta = 0$ and (*ii*) shows the power for the remaining points where $\delta = \pm\beta_{AB} / \sigma\sqrt{1/q + x_{01}(X_1'X_1)x_{01}'}$. As indicated in Figure 5 for $q = \infty$ the situation does not improve substantially by increasing the sample size.

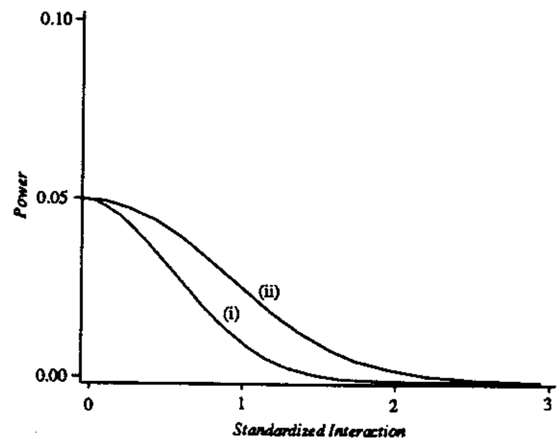

Figure 4. *Power versus the standardized interaction effect for the 3^2 design with one confirmatory trial where (i) δ = 0 and (ii) δ ≠ 0.*

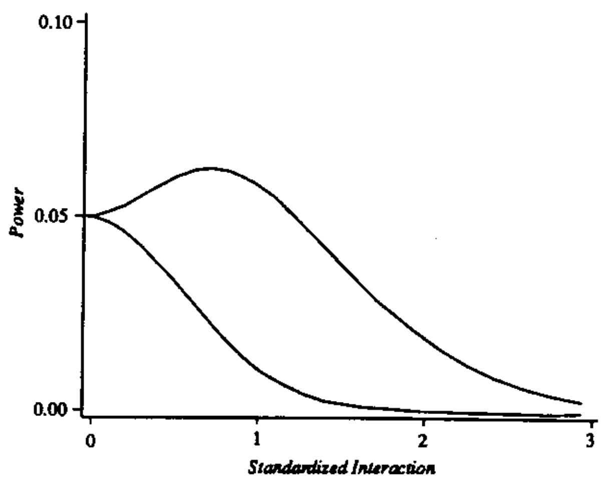

Figure 5. *Power versus the standardized interaction effect for the 3^2 design with ∞ confirmatory trials.*

Power function calculations do not always illustrate the performance of a test in the most intuitive way. A more dramatic display of the problems with Taguchi's confirmatory trial method can be obtained by looking at the prediction interval itself as a function of a possible interaction effect. For our 3^2 experiment the *expected values* of the standardized prediction limits for the false model (11) are given by

$$\frac{\bar{y}_q - \hat{z}_{n+1}}{\sigma} \in \pm t(9-5)_{1-\alpha/2}\sqrt{\left(1/q + 5/9\right)\left(1+[\beta_{AB}/\sigma]^2\right)}$$

and the *expected value* of the standardized prediction limits for the true model (12) are

$$\frac{\bar{y}_q - \hat{z}_{n+1}}{\sigma} \in \beta_{AB}\, x_A x_B/\sigma \pm t(9-6)_{1-\alpha/2}\sqrt{\left(1/q + 6/9\right)}\ .$$

Figure 6 (*i*) shows how the bias of the error term increasingly inflates the prediction interval as a function of an increasing standardized interaction effect assuming the false linear-effects-only model. Thus as the graph shows it will be rather unlikely to observe a confirmatory trial outside the limits since the prediction interval limits increase as a function of the interaction effect. For comparison we have overlaid the prediction interval based on the correct model shown as (*ii*). Here the width of the interval stays constant but the location changes.

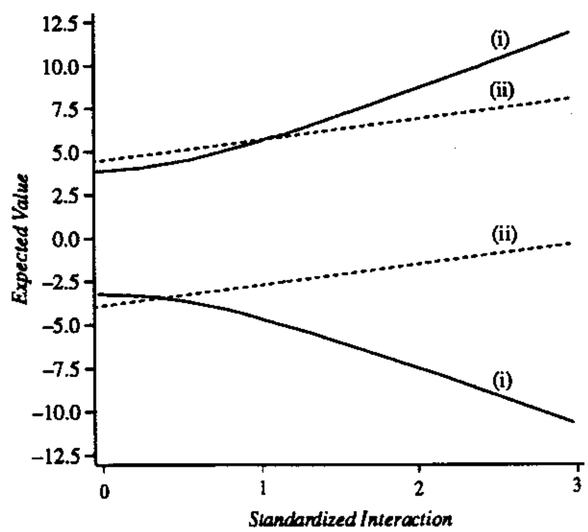

Figure 6. *Expected value of the standardized prediction limits for the 3^2 design with one confirmatory trial for (i) the false model and (ii) the true model.*

5. CONCLUSION

We have in the background research work for this article investigated the behavior of Taguchi's confirmatory trial method for several other more complex experiments often used by Dr. Taguchi. They all essentially reveal the same type of problems as those discussed in connection with the relatively simple experiments provided in Section 4. Thus based on the general theory of Section 2 and 3 and the examples in Section 4 we conclude that the performance of Taguchi's confirmatory trial method is problematic. In fact the performance may be extremely poor. We therefore recommend that the confirmatory trial technique advocated by Taguchi should not be relied on as a test for possible interaction effects. However, if the intention with the confirmatory trial is merely to verify the outcome of chosen factor combination as opposed to testing for possible interaction effects as stated by Taguchi in the quote above, then there is of course little reason to criticize the confirmatory trial practice.

APPENDIX

In this appendix we provide proofs and explanations for some of the results used in this paper.

Result A1. If $E\{\mathbf{y}\} = \mu = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$,

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}\ ,$$

$$\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2,$$

rank $\{\mathbf{X}_1'\mathbf{X}_1\} = \nu_1$ and $V\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$ then

$$E\{\hat{\beta}_1'\mathbf{X}_1'\mathbf{y}\} = (\beta_1 + \mathbf{A}\beta_2)'\mathbf{X}_1'\mathbf{X}_1(\beta_1 + \mathbf{A}\beta_2) + \nu_1\sigma^2$$

and

$$E\{\mathbf{y}'\mathbf{y} - \hat{\beta}_1'\mathbf{X}_1'\mathbf{y}\} = \beta_2'\mathbf{X}_2'(\mathbf{X}_2 - \mathbf{X}_1'\mathbf{A})\beta_2 + (n - \nu_1)\sigma^2\ .$$

Proof:

$$\begin{aligned} E\{\hat{\beta}_1'\mathbf{X}_1'\mathbf{y}\} &= E\{\mathbf{y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}\} \\ &= \mu'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mu + tr\{\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{I}_n\sigma^2\} \\ &= (\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2)'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) + \nu_1\sigma^2 \\ &= [(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_1\beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2]'\mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) + \nu_1\sigma^2 \\ &= (\beta_1 + \mathbf{A}\beta_2)'\mathbf{X}_1'\mathbf{X}_1(\beta_1 + \mathbf{A}\beta_2) + \nu_1\sigma^2 \end{aligned}$$

Moreover,

$$\begin{aligned} E\{\mathbf{y}'\mathbf{y}\} &= tr\{\mathbf{I}_n\sigma^2\} + \mu'\mu \\ &= n\sigma^2 + \beta_1'\mathbf{X}_1'\mathbf{X}_1\beta_1 + \beta_1'\mathbf{X}_1'\mathbf{X}_2\beta_2 + \beta_2'\mathbf{X}_2'\mathbf{X}_1\beta_1 + \beta_2'\mathbf{X}_2'\mathbf{X}_2\beta_2\ . \end{aligned}$$

By subtraction we then get

$$E\{\mathbf{y}'\mathbf{y} - \beta_1'\mathbf{X}_1'\mathbf{y}\} = \beta_2'\mathbf{X}_2'(\mathbf{X}_2 - \mathbf{X}_1\mathbf{A})\beta_2 + (n - \nu_1)\sigma^2\ .$$

Result A2. If $Z \sim \mathrm{N}(0,1)$, δ is a fixed constant, and $Y \sim$ *Non – Central Chi*^{2} with f degrees of freedom and non centrality parameter λ then

$$T' = \frac{Z + \delta}{\sqrt{Y / f}}$$

is distributed as a *doubly non-central t-distribution*. See Robbins (1948) and Krishnan (1968). For the power computations in this paper we have used the series expansion of the doubly non-central t provided by Bulgren and Amos (1968).

REFERENCES

Box, G., Bisgaard, S. and Fung, C.. (1988) "An Explanation and Critique of Taguchi's Contributions to Quality Engineering," *Quality and Reliability Engineering International*, Vol. 4, 123–131

Box, G. E. P., and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley and Sons.

Bulgren, W. G. and Amos, D. E. (1968), "A Note on Representations of the Doubly Non-Central *t* Distribution", *Journal of the American Statistical Association*, 63, 1013-1019.

Krishnan, M. (1968), "Series Representations of the Doubly Noncentral *t*-Distribution," *Journal of the American Statistical Association*, Vol. 63, pp. 1004–1012.

Robbins, M.(1948),"The Distribution of Student's *t* When the Population Means are Unequal," *Ann. Math. Statist.* 19:406-410 .

Taguchi , G. (1988), "The Development of Quality Engineering," *The ASI Journal*, Vol. 1, Number 1, Fall.