Handling Uncertainty in Analysis of Robust Design Experiments

Hugh Chipman
Graduate School of Business, University of Chicago,
1101 E. 58th Street, Chicago, IL, 60637
Email: hugh.chipman@gsb.uchicago.edu

July 1996

Abstract

In the analysis of robust design experiments, a model is typically fit to experimental data, and then used to select levels of control variables that desensitize the response to uncontrollable variation. Usually, model uncertainty (and sometimes parameter uncertainty) is not formally accounted for in the optimization process. This can lead to unrealistic improvements, and perhaps even sub-optimal performance. This paper considers the use of Bayesian methods in the fitting of models and their subsequent optimization. Through the use of these methods, reliable assessments of uncertainty may be incorporated into the analysis of such data.

Key Words: Bayesian Methods, Parameter and Model Uncertainty, Performance Measure, Model Selection, Response Model

Introduction

The purpose of a robust design experiment is to adjust the parameters of a product or process so that it is insensitive to uncontrollable variation. The performance of products and processes is sensitive to variables which are controllable under regular production (control variables) and those which are expensive or impossible to control under production (noise variables). The idea of robust design experiments is to adjust the control variables to make the process/product robust to the uncontrollable noise variables.

An experiment is carried out in which control and noise variables are set at fixed levels. The two basic stages are (I) using the data to model the response and (II) using this model to select optimal settings of the control variables. Both stages have challenges. In (I), some terms in the model may be related to other terms, complicating selection of a model. Examples of this would include interactions not being allowed to enter unless corresponding main effects are active, or the grouping of dummy variables associated with a multilevel categorical predictor. In (II), the optima selected can depend heavily on which model is selected. In fact, if a poor model is selected, the "optimal" settings chosen could increase rather than decrease variation.

In this paper, I illustrate the application of Bayesian methods to both stages. In (I), the model selection methodology of Box and Meyer (1993) is used, in conjunction with priors for variable selection with related predictors (Chipman, 1996). The resultant posterior distributions are used in (II) to assess the influence of parameter and model uncertainty on the optimization process. This makes it possible to protect against choosing a poor model which would lead to sub-optimal settings. The assessment of model uncertainty is similar in spirit to Draper (1995), which also discusses this issue from a Bayesian perspective. DuMouchel and Jones (1994) present recent work on Bayesian model selection in an industrial setting, but from a design standpoint. Another important distinction is that they do not place probabilities on models.

The following illustrative example is used throughout the paper. The data are from an experiment on the manufacture of leaf springs for trucks, and were originally analyzed by Pignatiello and Ramberg (1985). There are four control variables (B, C, D, E) and one noise variable (O). The descriptions and levels of the variables are given in Table 1. The response (Y) is the free height of the spring. The experiment is a cross-product matrix of the type suggested by Taguchi. It consists of a 2^{4-1} control variable matrix crossed with a 2^1 noise variable matrix, with three replicates observed at

Table 1: Variables studied in the experiment and their levels

Letter	Factor	Levels Low	High
B	High heat temperature (° F)	1840	1880
C	Heating time (seconds)	25	23
D	Transfer time (seconds)	12	10
E	Hold down time (seconds)	2	3
O	Quench oil temperature (° F)	130-150	150-170

Table 2: Variable settings and corresponding free height data.

B	C	D	E	O^-			O^+		
−	−	−	−	7.78	7.50	7.78	7.25	7.81	7.12
+	−	−	+	8.15	7.88	8.18	7.88	7.88	7.44
−	+	−	+	7.50	7.50	7.56	7.56	7.50	7.50
+	+	−	−	7.59	7.63	7.56	7.75	7.75	7.56
−	−	+	+	7.94	7.32	8.00	7.44	7.88	7.44
+	−	+	−	7.69	7.56	8.09	7.69	8.06	7.62
−	+	+	−	7.56	7.18	7.62	7.18	7.44	7.25
+	+	+	+	7.56	7.81	7.81	7.50	7.69	7.59

each of the 16 combinations. The data are given in Table 2. The goal of the experiment is to have the response near 8.0 inches with as little variation as possible. By choosing appropriate levels of $B - E$, it is hoped that the target value can be attained, and a minimum of variation in O transmitted to the response. In this article, the error terms are assumed to be i.i.d. If the replicates were not randomized (as is probably the case), a hierarchical model could be used.

Two classes of methods for the analysis of robust design data are the response model approach (Welch et. al. 1990) and the summary measure approach. The former models the response directly and treats the noise variables as fixed for the purpose of model building (I). In (II), regular production conditions are simulated by assuming a distribution for the noise variables, which induces extra variation in Y. By choosing levels of control variables, the variation transmitted to Y may be reduced. The response model approach is also discussed in Shoemaker, Tsui and Wu (1991), Myers, Khuri, and Vining (1992), and Steinberg and Bursztyn (1994). The summary measure approach analyzes a summary measure calculated using noise replicates and does not explicitly include noise variables in the model. This approach relies on a cross-product array formed by combining an array of control variables with an array of noise variables. Each summary measure is calculated for a row of the control matrix. Common summary measures include the signal to noise ratio, as suggested by Taguchi, or the sample mean and standard deviation. Myers and Vining (1990) discuss response surface methods in the joint analysis of mean and standard deviations. Nelder and Lee (1991) develop methods based on Generalized Linear Models for similar analysis of location and dispersion. A good overview of the summary measure and response model approaches is given in Nair (1992).

The method discussed in this article is not intended as a replacement of or competitor to the methods just mentioned. Rather, it should be viewed as an "add-on" for assessment of uncertainty which is applicable to any of these methods.

The article is arranged as follows: A Bayesian method for model selection is reviewed, and then applied in assessing uncertainty in response models. These ideas are illustrated with the leaf spring example. The paper concludes with extensions and other related issues.

Model Selection and Uncertainty

A set of possible linear models, $M_1, \ldots M_m$ are considered. Each M_i specifies the columns of a predictor matrix X_i and corresponding coefficient vector β_i. Model i has the form

$$Y = X_i\beta_i + \epsilon, \tag{1}$$

where ϵ is a vector of i.i.d. $\mathrm{N}(0, \sigma^2)$ random variables. X_i is assumed to have t_i columns ($0 \le t_i \le p$) in addition to an intercept column, which is assumed to be present in all models. Each column will be referred to as a *term* to distinguish it from the variables used to form it. In the response model approach, X_i could include control and noise main effect terms and interaction terms. In the summary measure approach, it would not contain terms involving noise variables.

Following Box and Meyer (1993), each term has a prior probability π of being in the model. Assuming these terms to be independent yields a prior on model M_i as:

$$\Pr(M_i) = \pi^{t_i}(1-\pi)^{p-t_i} \propto \left(\frac{\pi}{1-\pi}\right)^{t_i}. \tag{2}$$

Conditional on model M_i, all elements of β_i except the intercept (β_{i0}) are assigned independent $N(0, \gamma^2\sigma^2)$ priors. An uninformative prior is put on β_{i0} and $\log(\sigma)$, so that $f(\beta_{i0}, \sigma) \propto 1/\sigma$. The parameters π and γ then define the prior. π may be chosen by considering the expected number of active terms. Box and Meyer (1993, 1992) suggest choosing γ to minimize probability on the null model.

The posterior probability of model $M_i, i = 0, \ldots, 2^p$ is given by

$$\Pr(M_i|Y) = C\left(\frac{\pi}{1-\pi}\right)^{t_i} \gamma^{-t_i} \frac{|X_0'X_0|^{1/2}}{|\Gamma_i + X_i'X_i|^{1/2}} \left(\frac{S(\hat{\beta}_i) + \hat{\beta}_i'\Gamma_i\hat{\beta}_i}{S(\hat{\beta}_0)}\right)^{-(n-1)/2} \tag{3}$$

where

$$\Gamma_i = \frac{1}{\gamma^2}\begin{bmatrix} 0 & 0 \\ 0 & I_i \end{bmatrix},$$

$$\hat{\beta}_i = (\Gamma_i + X_i'X_i)^{-1}X_i'Y,$$

$$S(\hat{\beta}_i) = (Y - X_i\hat{\beta}_i)'(Y - X_i\hat{\beta}_i),$$

X_0 corresponds to the intercept-only model, and C is a normalizing constant. There is a slight difference from Box and Meyer's (1993) form of the equation, since interactions are treated here as distinct terms. The posterior on models captures model uncertainty, and the posterior on β_i, σ (given in the appendix) captures parameter uncertainty for a given model.

In small to medium sized problems (such as the example considered here), the number of possible models is small enough that all probabilities may be evaluated and the normalizing constant calculated exactly. In larger problems, posterior probabilities on the model space may be determined using stochastic search techniques. George (1995) provides a review of this literature.

More flexible priors on the model space could be used. For example, investigators might be uncomfortable with models that contain interactions but no corresponding main effects. Chipman (1996) develops priors in which inclusion or exclusion of terms depends on other terms in the model. For example, a control by noise interaction (say CN) would be less likely to be included if neither C or N were included. In the response model approach, the large number of control by noise interactions makes this prior attractive because it allows more control in specifying which models are sensible. For such priors, (3) would be the same, except the $(\pi/(1-\pi))^{t_i}$ term would be replaced by a different prior probability.

Model Uncertainty Applied to the Robust Design Problem

Regression model (1) is used in both the response model and summary measure approaches. They differ in what Y is used, and which terms are included in the predictor matrix X. Assessment of parameter and model uncertainty can be applied to either approach. Once the model is fit, one or more performance measures are to be optimized. The posterior on models and parameters leads to a posterior on the performance measures. This posterior will give a more realistic comparison of performance under different settings of the control variables.

The application of this idea will be demonstrated for the leaf spring data using the response model approach. The following terms are considered: B, C, D, E, BC, BD, CD, O, and seven interactions between O and the other terms given. Some two-way interactions (like $BC = DE$) are aliased with each other. Pignatellio and Ramberg (1985) elect to identify these as as BC, BD, CD. Since all three appear inactive, this is not an important issue. All other aliased terms are of different orders - in these cases, the lower order term is assumed to be active. Only six terms (B, C, E, O, BO, CO) appear important, as indicated by the half normal plot of the estimated effects in Figure 1. It is unclear whether E and BO are active, as they have smaller effects than the other four terms. The p-values for E and BO are in the range (0.01, 0.03) while the other four terms have p-values below 0.0005. All other p-values are greater than 0.15.

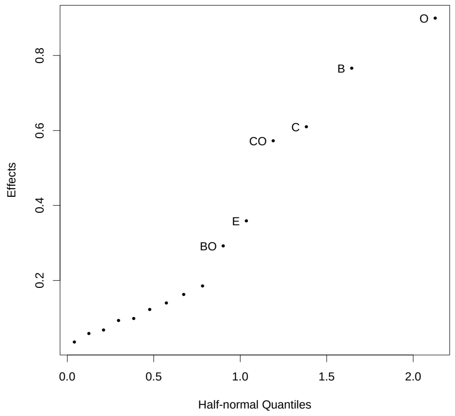
Suppose first that parameter and model uncertainty are ignored, and the following model is fit using the data:

$$Y = 7.636 + 0.111B - 0.088C + 0.052E - 0.130O + 0.042BO + 0.083CO + \epsilon, \tag{4}$$

where $\mathrm{Var}(\epsilon)$ is estimated by $\hat{\sigma}^2 = 0.126^2$. Now assume that O is random with $E(O) = 0$ and $V(O) = \sigma_O^2$. Conditional on knowing β and σ (here $\hat{\beta}, \hat{\sigma}$ are assumed to be the true values),

$$\begin{aligned} \mu_Y &= E(Y) \\ &= 7.636 + 0.111B - 0.088C + 0.052E \end{aligned} \tag{5}$$

Figure 1: Half normal plot of estimated effects for the leaf spring data



and

$$\begin{aligned}\sigma_Y^2 &= V(Y) \\ &= V(-0.130O + 0.042BO + 0.083CO) + \sigma^2 \\ &= (-0.130 + 0.042B + 0.083C)^2\sigma_O^2 + \sigma^2\end{aligned} \tag{6}$$

These are the two performance measures used in the response model approach. In this case, it is desired to have μ_Y near 8.0 and minimize σ_Y. The bracketed term in (6) means that choice of B and C can reduce the variation in Y. Control by noise interaction terms provide the means for adjusting σ_Y.

One problem is that $\hat{\beta}$ and $\hat{\sigma}$ are estimates that have uncertainty associated with them. This uncertainty is transmitted to μ_Y and σ_Y via (5) and (6). Optimization should account for this uncertainty. This can be achieved by use of a posterior distribution for β, σ. A similar approach is explored from a non-Bayesian perspective by Easterling (in preparation) and Myers and Montgomery (1995, p. 496 – 498).

There is another problem. Often the model is uncertain. In this example, terms E and BO may or may not be active. Inclusion or exclusion of these terms should change the form of (5) and (6) and consequently the effectiveness of B for adjusting σ_Y and of E for adjusting μ_Y. This type of problem has been noted by Shoemaker, Tsui, and Wu (1991), who observe that improper selection of the model could lead to settings that increase

Table 3: The four most probable models. Negligible mass is associated with other models.

B	C	E	O	BO	CO	Prob
•	•	•	•	•	•	0.29
•	•		•	•	•	0.12
•	•	•	•		•	0.37
•	•		•		•	0.21

rather than decrease variation. Specifically, selection of control by noise interactions has great influence on adjustment of σ_Y. To protect against this, model uncertainty (3) can be used to combine performance measures across different models. The distributions of μ_Y and σ_Y will be mixtures of several continuous distributions. Each component of the mixture will be formed from expressions like (5) and (6), with mixture probabilities given by (3). This Bayesian model mixing does not have an obvious frequentist analog. Certainly an unweighted average of performance measures over promising models is less attractive, since some models may be more likely than others.

In the leaf spring example, the first step in incorporating uncertainty is to calculate $\Pr(M_i)$ for all models M_i. Since nine of the 15 candidate terms are clearly inactive, there are effectively $2^6 = 64$ models to consider rather than $2^{15} = 32768$. The prior parameter γ is set at 0.7 to minimize the posterior probability on an intercept-only model. Choice of $\pi = 0.25$ corresponds to the prior belief that $0.25 \times 15 = 3.75$ terms are expected to be active. The four models given in Table 3 have virtually all the posterior mass. Because these models have active main effects that correspond to active interactions, the independence prior (2) on models is used.

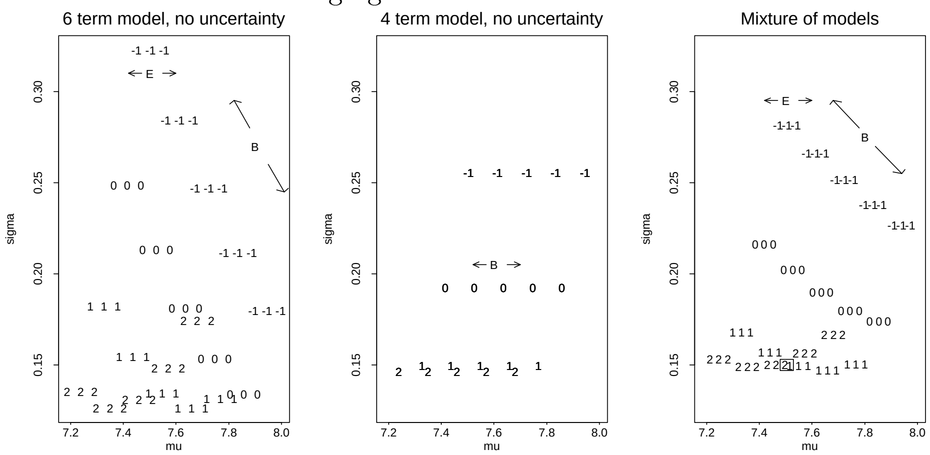
If O is viewed as random, then optimal levels of B, C, and (for two models) E are sought. Initial B, C, E values on a 5^3 grid on $\{-2, -1, 0, 1, 2\}$ were considered. To simplify presentation, this has been refined to 60 (B, C, E) values given by $B \in \{-2, -1, 0, 1, 2\} \times C \in \{-1, 0, 1, 2\} \times E \in \{-1, 0, 1\}$. Some extreme grid points (e.g. 2, -2) are retained to achieve performance near the target. If O is assumed equally likely to take values -1 and $+1$, then $\mathrm{Var}(O) = 1$.

Optimization in three situations is considered:

1. *six term model:* model (4) with terms B, C, E, O, BO, CO is used in conjunction with point estimates to estimate μ_Y and σ_Y.

2. *four term model:* a reduced form of model (4), with terms B, C, O, CO

Figure 2: Estimates of performance measures μ_Y and σ_Y at 60 different settings of B, C, E. Point estimates and single models are used in the two left plots. Posterior expectations over four models are used in the right plot. Plotting symbols indicate levels of the variable C, and labeled arrows indicate the effect of changing B and E.



is used with point estimates in the same fashion.

3. *mixture model:* the posterior distributions of β_i, σ are mixed over the four models in Table 3, and expected values of μ_Y and σ_Y are calculated.

For each of these three cases, estimates of μ_Y, σ_Y are obtained at each of the 60 (B, C, E) combinations. The first two cases use point estimates, which are plugged into equations of the form of (5) and (6), although the terms included depend on the model. Monte Carlo simulation with 10,000 draws is used to calculate the posterior for the mixture model. Details of this calculation are given in the appendix.

The estimated performance measures are graphically displayed in Figure 2. Each graph plots (μ_Y, σ_Y) at the 60 different settings of (B, C, E). Plotting symbols indicate the level of C corresponding to each point. The effects of changing B and E are indicated by labeled arrows. For example, changing E results in horizontal shifts in the six term model and the mixture model, and has no effect in the four term model. The desired target region of each plot is the lower right, which corresponds to a mean of 8.0 and low variance.

The two left plots indicate that model choice has serious implications for optimization. In the six term model, B, C, and E adjust μ_Y and B, C adjust σ_Y. In the four term model, E no longer has any effect, and B only adjusts μ_Y. By using a posterior expectation over models to estimate (μ_Y, σ_Y), this instability is reduced.

The mixture plot may be thought of a weighted average of four plots (one for each model in Table 3, including the two given in Figure 2). Relative to the smallest model considered (here the four term model), the mixture of models offers more adjustment variables. When compared to the largest model (here six terms), this averaging affects the optimization by diluting (or shrinking) the effect of some variables on μ_Y and σ_Y. In comparing the six term model to the mixture, B is still adjusts σ_Y, but has less influence. The effect of C as a noise adjustment variable remains unchanged. This is because C is included in all models, while B is included only in some. A similar pattern is observable for E. This dilution of effects means that attaining the optimum may be more difficult than if the largest model is used, as is the case here.

Note that the effect of E is changed less than that of B. This is because $\Pr(E \text{ active}) = 0.66$ and $\Pr(B \text{ active}) = 0.41$. There is more certainty that E is active, so its effectiveness as an adjustment variable is reduced less than that of B.

To illustrate another point, suppose that the target value for μ_Y was 7.5 rather than 8.0 and model and parameter uncertainty are accounted for by using the right plot. Two settings (indicated by a box in the plot) appear promising. They have the values $(B, C, E) = (0, 1, -1)$ (setting 1) and $(0, 2, 1)$ (setting 2). Based on the expected values given in Figure 2, there is little to distinguish the two settings. The spread of the posteriors is different, with

$$\text{Var}(\mu_Y) = 0.00156, \text{Var}(\sigma_Y) = 0.000100, \text{ setting } 1$$
$$\text{Var}(\mu_Y) = 0.00271, \text{Var}(\sigma_Y) = 0.000192, \text{ setting } 2.$$

Setting 1 is less risky, since is has the same expected performance measures, but less uncertainty than setting 2. The spread of both performance measures is affected by the change of C from 1 to 2, a change that amplifies the uncertainty in β_C and β_{CO}. Choosing this level of C parallels the idea of choosing control variables to minimize transmitted variation.

To summarize, there are two consequences of posterior averaging: (1) when there is uncertainty about whether terms are active, the effect of adjusting these terms is reduced but not removed. Instability due to choice of a specific model is reduced. (2) Estimates of uncertainty in performance measures may be utilized to choose between apparently comparable settings.

Discussion

There are two ways in which this methodology may be used. First, it can be used as a criterion to be optimized, as illustrated here. Second, it may be used as a diagnostic to assess already chosen settings of control variables taking into account model and parameter uncertainty. This could be accomplished by identification of one or more optimal settings without using the Bayesian approach. Posteriors would then be used to incorporate uncertainty into the evaluation of the optimal settings.

The same posterior calculations and assessment can be applied to other performance measures. Conditional on the parameters and model, one or more performance measure can be defined. These performance measures are functions of the unknown parameters, and the model selected. The posterior distribution of the performance measure can be calculated provided the posterior distribution on parameters and models is available. For example, one could calculate a sample average and variance for the six values in each row of Table 2, and fit two linear models relating each of these to the control variables. Given β and σ for each model, the performance measures can be calculated for a given setting of control variables. The posteriors of β and σ may be used to determine the posterior of each performance measure at this setting. The impact of the uncertainty assessment will be similar to the example presented here: weaker effects will have their influence reduced, and uncertainty in the performance measure may be assessed.

The performance measure may not always be a simple linear function. In the response model approach, one could calculate the probability that Y falls within a range of interest, and then try to maximize this measure. In such situations, a simulation-based calculation approach is attractive, since calculation of the posterior of this nonlinear performance measure is straightforward.

Appendix

The posterior distribution for β_i, σ given model M_i is

$$\Pr(\beta_i|\sigma, M_i)\Pr(\sigma|M_i).$$

The individual components are

$$\beta_i|\sigma, M_i \sim N(\hat{\beta}_i, [\Gamma + X'X]^{-1}\sigma^2),$$
$$\sigma|M_i \sim IG(n - t_i - 1, [S(\hat{\beta}_i) + \hat{\beta}_i'\Gamma_i\hat{\beta}_i]/(n - t_i - 1)).$$

This result may be found in many Bayesian references, including Zellner (1971).

These posteriors are used in the Monte Carlo calculation of the posterior distribution of μ_Y and σ_Y. In the leaf spring example, draws of β, σ were made from the four models given in Table 3, in proportion to the posterior probabilities of these models. For example, 2900 values of (β_i, σ) were drawn from the posterior for the model with B, C, E, O, BO, and CO. For each (β_i, σ) pair, a corresponding μ_Y, σ_Y were calculated. The 10000 resultant values of μ_Y, σ_Y represent draws from the posterior, and incorporate both model and parameter uncertainty.

Acknowledgments

I would like to thank R. G. Easterling, M. S. Hamada, R. J. MacKay, and two referees for comments and helpful suggestions. This research was supported by the University of Chicago Graduate School of Business.

References

Box, G. E. P. and Meyer, R. D. (1992) "Finding the Active Factors in Fractionated Screening Experiments". University of Wisconsin-Madison, Center for Quality and Productivity Improvement, Report #80.

Box, G. E. P. and Meyer, R. D. (1993). "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, 25, 94–105.

Chipman, H. (1996). "Bayesian Variable Selection with Related Predictors", *Canadian Journal of Statistics* 24, 17–36.

Draper, D. (1995). "Assessment and Propagation of Model Uncertainty", *Journal of the Royal Statistical Society, series B* 57, 45–97.

Easterling, R. G. (in preparation)."Virtual Prediction of Process Capability", Sandia National Laboratories.

George, E. I. (1995). "Bayesian Model Selection", preliminary draft for *Encyclopedia of Statistical Sciences.*

Myers, R. H. and Montgomery (1995). *Response Surface Methodology.* John Wiley and Sons, New York, NY.

Myers, R. H., Khuri, A. I. and Vining, G. (1992). "Response Surface Alternatives to the Taguchi Robust Parameter Design". *The American Statistician* 46, 131–139.

Nair, V. N. (Editor) (1992). "Taguchi's Parameter Design: A Panel Discussion". *Technometrics* 34, 128–161.

Nelder, J. A. and Lee, Y. (1991). "Generalized Linear Models for the Analysis of Taguchi-Type Experiments". *Applied Stochastic Models and Data Analysis* 7, 107–120.

Pignatiello, J. J. and Ramberg, J. S. (1985). Discussion of "Off-Line Quality Control, Parameter Design, and the Taguchi Method" by R. N. Kackar, *Journal of Quality Technology*, 17, 198–206.

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". *Journal of Quality Technology* 26, 12–20.

Vining, G. and Myers, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach". *Journal of Quality Technology* 22, 38–45.

Welch, W. J., Yu, T. K., Kang, S. M. and Sacks, J. (1990). "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, 15–22.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* Robert E. Krieger Publishing Company, Malabar, FL.