Bootstrap Confidence Regions for Optimal Operating Conditions in Response Surface Methodology

Roger D. Gibb[*]   I-Li Lu[†]

Walter H. Carter Jr[‡]

[*]The Procter & Gamble Company, gibb.rd@pg.com

[†]The Boeing Company, i-li.lu@boeing.com

[‡]Virginia Commonwealth University, whcarter@vcu.edu

Bootstrap Confidence Regions for Optimal Operating Conditions in Response Surface Methodology

Roger D. Gibb, I-Li Lu, and Walter H. Carter Jr

Abstract

This article concerns the application of bootstrap methodology to construct a likelihood-based confidence region for operating conditions associated with the maximum of a response surface constrained to a specified region. Unlike classical methods based on the stationary point, proper interpretation of this confidence region does not depend on unknown model parameters. In addition, the methodology does not require the assumption of normally distributed errors. The approach is demonstrated for concave-down and saddle system cases in two dimensions. Simulation studies were performed to assess the coverage probability of these regions.

AMS 2000 subj Classification: 62F25, 62F40, 62F30, 62J05.

Key words: Stationary point; Kernel density estimator; Boundary kernel.

1 Introduction

One of the reasons for using a response surface analysis is to determine the operating conditions which, without loss of generality, maximize a response y. Development of confidence regions for these operating conditions has been considered by several investigators. The purpose of this report is to demonstrate that likelihood-based bootstrap confidence region methodology is a powerful alternative which does not suffer from some of the limitations associated with existing approaches.

Over the experimental region it is assumed that the relationship of y and the regressors $x_1, x_2, \ldots, x_k$ can be expressed

$$y = g(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon, \tag{1}$$

where g is an unknown continuous, differentiable function and ε is a random source of variability not accounted for in g. Often $g(\boldsymbol{x}, \boldsymbol{\theta})$ can be adequately approximated with a second-order polynomial, i.e.

$$g(\boldsymbol{x}) \approx \beta_0 + \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}, \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)$ and

$$\boldsymbol{B} = \begin{pmatrix} \beta_{11} & \beta_{12}/2 & \ldots & \beta_{1k}/2 \\ \beta_{12}/2 & \beta_{22} & \ldots & \beta_{2k}/2 \\ \vdots & & \ddots & \vdots \\ \beta_{1k}/2 & \beta_{2k}/2 & \ldots & \beta_{kk}/2 \end{pmatrix}. \tag{3}$$

Box and Hunter (1954) constructed a confidence region for the stationary point of a second-order response surface. The stationary point is given by

$$\boldsymbol{x}_{sp} = -\frac{\boldsymbol{B}^{-1}\boldsymbol{\beta}}{2}, \tag{4}$$

the solution to the system of equations $\partial y/\partial \boldsymbol{x} = \boldsymbol{0}$. The interpretation and relevance of the stationary point depends on the nature of the response surface which is determined by $\boldsymbol{B}$. Consider the following cases: 1) the eigenvalues of $\boldsymbol{B}$ are mixed in sign, 2) the eigenvalues of $\boldsymbol{B}$ are all positive and 3) the eigenvalues of $\boldsymbol{B}$ are all negative.

In cases 1 and 2 the stationary point is a saddle point and the location of minimum response, respectively, not the location of maximum response and, therefore, is not of interest for the purpose of this report. In case 3 the stationary point is the location of maximum response. In practice, the model parameters, including the elements of $\boldsymbol{B}$, are unknown but can be estimated from the data. Since there is uncertainty associated with any estimator of $\boldsymbol{B}$ there is also uncertainty in assessing the nature of the true response surface and, therefore, the interpretation of Box and Hunter's confidence region.

Another issue of practical importance is the data are observed over a treatment space of finite dimensions, i.e. the experimental region, and one is generally unwilling to extrapolate to areas outside this region. Peterson (1992) utilized a transformation technique to develop a confidence region for the stationary point constrained to a specified region. However, as the approach is founded on the stationary point, the methodology suffers from the same interpretation difficulty as Box and Hunter's confidence region.

In cases 1 and 2 the maximum response is undefined unless interest is restricted to a subset of the treatment space. Ridge analysis was developed by Hoerl (1959) and refined by Draper (1963) to estimate the stationary point subject to the constraint $\boldsymbol{x}'\boldsymbol{x} = r^2$. Under their approach the investigator must specify a value of the Lagrangian multiplier μ which, if chosen greater than the largest eigenvalue of $\hat{\boldsymbol{B}}$, facilitates determination of the location of the predicted constrained maximum. Stablein, Carter and Wampler (1983) constructed a confidence region for the constrained

stationary point conditional on the investigator's choice of μ. However, proper interpretation of their confidence region depends on whether the choice of μ is greater than the largest eigenvalue of $\boldsymbol{B}$. Since the eigenvalues of $\boldsymbol{B}$ can only be estimated, there is uncertainty involved in the confidence region's interpretation.

Let $\boldsymbol{x}_{cm}$ be the operating conditions associated with maximum $g(\boldsymbol{x}, \boldsymbol{\theta})$ subject to the constraint that $\boldsymbol{x}$ is within the experimental region. Denote $\hat{\boldsymbol{x}}_{cm}$ and $\hat{\boldsymbol{\theta}}$ as estimators of $\boldsymbol{x}_{cm}$ and $\boldsymbol{\theta}$ respectively, then in practice, $\hat{\boldsymbol{x}}_{cm}$ can be calculated from $g(\boldsymbol{x}; \hat{\boldsymbol{\theta}})$ using a numerical optimization algorithm, such as the Nelder-Mead (1965) simplex. A favorable property of $\boldsymbol{x}_{cm}$ is its interpretation does not depend on unknown model parameters, unlike the stationary point or constrained stationary point. A drawback, however, is that in some instances $\hat{\boldsymbol{x}}_{cm}$ is not a consistent estimator. Consider the case where $g(\boldsymbol{x}, \boldsymbol{\theta})$ is continuous and $\varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$. If the model is correct, $\hat{\boldsymbol{\theta}}$ is a least-squares estimator and $\boldsymbol{x}_{cm}$ is unique, then $\hat{\boldsymbol{x}}_{cm}$ can be shown to be consistent (Kendall and Stuart, 1979, Chapter 18). To demonstrate an instance where $\boldsymbol{x}_{cm}$ is not unique and, therefore, $\hat{\boldsymbol{x}}_{cm}$ is not consistent, consider the $k = 2$ second-order response surface where $\beta_1 = \beta_2 = 0$ and $\beta_{11} = \beta_{22} > 0$. If the experimental region is the two dimensional direct product of the interval $[-a, a]$, i.e. $[-a, a]^2$, then the maximum response within the experimental region occurs at $(-a, -a)$, $(a, -a)$, $(-a, a)$ and (a, a). A formal assessment of whether $\boldsymbol{x}_{cm}$ is unique could be made with tests of the appropriate hypotheses. For example, the likelihood of the symmetric model described earlier could be investigated by testing $\mathrm{H}_0 \colon \beta_1 = \beta_2 = 0, \beta_{11} = \beta_{22}$. If there is insufficient evidence to reject H_0 further investigation is warranted.

Construction of a confidence region for $\boldsymbol{x}_{cm}$ using classical methods would require knowledge of the sampling distribution of $\hat{\boldsymbol{x}}_{cm}$. As this distribution is unknown it is reasonable to explore the use of bootstrap confidence region methods that do not require its mathematical derivation. Unlike the confidence region methods described

earlier, the bootstrap approach does not require normality assumptions on the model errors.

2 Likelihood-Based Bootstrap Confidence Regions

Hall (1987, 1992) describes three methods for constructing likelihood-based bootstrap confidence regions for a k-variate parameter vector $\boldsymbol{\xi}$ given a sample of size n, namely, the percentile-t method, the ordinary percentile (hybrid) method, and the percentile method. In regards to which of these methods is appropriate when $\boldsymbol{\xi} \equiv \boldsymbol{x}_{cm}$, there are several factors that merit consideration. First, the percentile-t method requires an accurate estimate for $\boldsymbol{V}$, the asymptotic covariance matrix of $n^{\frac{1}{2}}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})$, where $\boldsymbol{V}$ is assumed to be positive definite. The analytic expression for $\hat{\boldsymbol{V}}$, an estimate of $\boldsymbol{V}$, is unknown and accurate estimates using resampling techniques are unlikely to obtain in the small sample designed experiment setting. Second, and perhaps more importantly, only the percentile method preserves the range of $\boldsymbol{\xi}$ in all cases. In particularly, for $\boldsymbol{\xi} \equiv \boldsymbol{x}_{cm}$, only confidence regions under the percentile method are guaranteed to be bounded by the experimental region. Therefore, for the purposes of this report, attention is focused exclusively on the percentile method to construct likelihood-based bootstrap confidence regions for $\boldsymbol{x}_{cm}$.

2.1 Percentile Method Algorithm

An adaptation of Hall's (1987) likelihood-based bootstrap confidence region algorithm when $\boldsymbol{\xi} = \boldsymbol{x}_{cm}$ is given below, followed by several notes regarding its implementation.

1. For the response surface model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, determine $\hat{\boldsymbol{\theta}}$ using the method of least-squares.

2. Standardize the residual vector $\hat{\boldsymbol{\varepsilon}}$ with the elementwise operation

$$\hat{\boldsymbol{\varepsilon}}_s = \hat{\boldsymbol{\varepsilon}}/\sqrt{\text{diag}(I - X(X'X)^{-1}X')}.$$

3. Compute b bootstrap estimates of $\boldsymbol{x}_{cm}$ by following steps 3(a) - 3(d) a total of b times (see implementation notes for convenient choices of b).

   (a) Generate a bootstrap sample of $\hat{\boldsymbol{\varepsilon}}_s$, denoted by $\hat{\boldsymbol{\varepsilon}}_s^*$.

   (b) Calculate the corresponding bootstrap sample of the data with $\boldsymbol{y}^* = \boldsymbol{X}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\varepsilon}}_s^*$.

   (c) For the response surface model $\boldsymbol{y}^* = \boldsymbol{X}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$, determine $\hat{\boldsymbol{\theta}}^*$ using the method of least-squares.

   (d) Based on $\hat{\boldsymbol{\theta}}^*$, calculate $\hat{\boldsymbol{x}}_{cm}^*$.

4. Fit a nonparametric density $\hat{f}$ to the $\hat{\boldsymbol{x}}_{cm,i}^*$, $i = 1, 2,\ldots, b$.

5. The contour on $\hat{f}$ of smallest content that captures $100(1 - \alpha)\%$ of the b estimates $\hat{\boldsymbol{x}}_{cm}^*$ is a $100(1 - \alpha)\%$ likelihood-based confidence region for $\boldsymbol{x}_{cm}$ using the percentile method.

2.2   Implementation Notes

Steps $1 - 3$ are an application of bootstrapping residuals to generate bootstrap estimates of $\boldsymbol{x}_{cm}$. Bootstrapping residuals was proposed by Efron (1979) and is considered appropriate when $\boldsymbol{X}$ is fixed and the model is correct with exchangeable errors. Standardization of the residual vector in step 2 ensures that the variance of each $\hat{\varepsilon}_{s,i}$ matches that of the unobservable ε_i. A simple bootstrap sample is generated by drawing a sample of size n with replacement from the elements of $\hat{\boldsymbol{\varepsilon}}_s$. The balanced bootstrap, proposed by Davison et al. (1986), was used in the examples and

simulation studies of section 2.4 to ensure that $\overline{\hat{\varepsilon}}^* = \overline{\hat{\varepsilon}}$. In step 4 the conditional distribution of $\hat{\boldsymbol{x}}_{cm}$ is approximated nonparametrically and used as the basis for selecting the confidence region boundary.

In practice, the level contour on $\hat{f}$ of smallest content that captures $100(1-\alpha)\%$ of the b estimates $\hat{\boldsymbol{x}}_{cm}^*$ was determined as follows: Choose b such that $(1-\alpha)b$ is a non-negative integer and assume that $\hat{f}(\hat{\boldsymbol{x}}_{cm,i}^*)$, $i = 1, 2, \ldots, b$, are unique. Let $\hat{f}_\alpha$ be the value of $\hat{f}$ corresponding to the level contour of smallest content on $\hat{f}(\boldsymbol{x})$ that captures exactly $(1-\alpha)b$ of the $\hat{\boldsymbol{x}}_{cm,i}^*$, $i = 1, 2, \ldots, b$. To calculate $\hat{f}_\alpha$ consider that, by definition, $\hat{f}(\hat{\boldsymbol{x}}_{cm,i}^*) \geq \hat{f}_\alpha$ for all $\hat{\boldsymbol{x}}_{cm,i}^*$ captured by the contour and $\hat{f}(\hat{\boldsymbol{x}}_{cm,i}^*) < \hat{f}_\alpha$ for every other $\hat{\boldsymbol{x}}_{cm,i}^*$. Therefore, $\hat{f}_\alpha$ is the $(1-\alpha)b^{th}$ element of the set with elements $\hat{f}(\hat{\boldsymbol{x}}_{cm,i}^*)$, $i = 1, 2, \ldots, b$, sorted in descending order. In the event that all $\hat{f}(\hat{\boldsymbol{x}}_{cm,i}^*)$, $i = 1, 2, \ldots, b$, are not unique, the contour $\hat{f}_\alpha(\boldsymbol{x})$ may capture more than $(1-\alpha)b$ of the $\hat{\boldsymbol{x}}_{cm,i}^*$, $i = 1, 2, \ldots, b$, in which case the confidence region is conservative. After identifying $\hat{f}_\alpha$ in this manner the confidence region boundary was graphically displayed using the GCONTOUR procedure of SAS®.

2.3   Application to Second-Order Response Surfaces

Likelihood-based confidence regions were constructed for $\boldsymbol{x}_{cm}$, where $g(\boldsymbol{x}, \boldsymbol{\theta})$ was a second-order polynomial ($k = 2$) for the following cases: 1) concave down with $\boldsymbol{x}_{cm}$ located inside the experimental region and 2) saddle system with $\boldsymbol{x}_{cm}$ located, necessarily, on a boundary of the experimental region. The experimental region was defined as the two dimensional direct product of the interval $[-1.4, 1.4]$. Model parameters and coordinates of $\boldsymbol{x}_{cm}$ for both models are summarized in table 1.

As indicated in section 2.1, a nonparametric estimate of the sampling distribution of $\hat{\boldsymbol{x}}_{cm}$ conditional on b bootstrap estimates of $\boldsymbol{x}_{cm}$ serves as the basis for constructing

the confidence region boundary. A product kernel density estimator for this purpose is given by

$$\hat{f}(\boldsymbol{x}) = \frac{1}{bh_1h_2} \sum_{i=1}^{b} \left\{ \prod_{j=1}^{2} K\left(\frac{x_i - \hat{x}^*_{cm,i,j}}{h_j} \right) \right\}, \tag{5}$$

where K is a univariate kernel and h_i is the bandwidth for the ith coordinate (see, e.g., Silverman, 1986, Scott, 1992). Gasser and Müller (1979) showed that if K is symmetric the bias of $\hat{f}(\boldsymbol{x})$ can be inflated near the boundaries. Müller (1988) showed that uniform bias over the entire support of $f(\boldsymbol{x})$ can be achieved if an asymmetric boundary corrected kernel is used in (5). Gasser and Müller (1979) and Jones (1993), among others, describe univariate boundary kernels designed for estimating densities with a single boundary. Densities with support over an interval may be estimated using two such kernels, one at each boundary, but this is appropriate only if the bandwidth is small relative to the interval length. Hart and Wehrly's (1992) linear boundary kernel was implemented in (5), as this kernel is specifically designed to estimate densities with support on a interval, where the bandwidth need not be small relative the interval length. The biweight density was used as a basis for constructing this boundary kernel.

Two methods of multivariate bandwidth selection were explored: Normal rule-of-thumb (see, e.g., Scott, 1992) and the plug-in approach of Wand and Jones (1994, pp 107-113). The sample standard deviation in each coordinate of $\hat{\boldsymbol{x}}^*_{cm}$ was used to estimate the scale of the conditional distribution of $\hat{\boldsymbol{x}}_{cm}$, as some measure of scale is required by both bandwidth selectors. The possibility exists for all bootstrap estimates for $\boldsymbol{x}_{cm}$ to be located on the experimental region boundary, a situation that occured almost exclusively with the saddle system. To ensure nonzero scale estimates, the standard deviation for the ith coordinate was calculated from a modified dataset

given by

$$z_i = \begin{cases} \boldsymbol{x}^*_{i,cm} - \text{sign}(\boldsymbol{x}^*_{i,cm})\delta & \text{if} \quad \text{abs}(\boldsymbol{x}^*_{i,cm}) = 1.4 \\ \boldsymbol{x}^*_{i,cm} & \text{otherwise,} \end{cases} \tag{6}$$

where δ is a random $U(0, 0.05)$ deviate. The effect of using the modified dataset to estimate standard deviation was negligible in the concave-down system, as most bootstrap estimates did not fall on a boundary. For the saddle system, using the modified dataset prevented nonzero bandwidth estimates and, thus, permitted use of the product kernel density estimator.

2.4   Simulation Study Results

It is of interest to compare the bootstrap confidence region performance under both methods of bandwidth selection, i.e. Normal rule-of-thumb and the plug-in approach. Data for 500 experiments were simulated by adding a $N(0, 3^2)$ deviate to the 'true' response at operating conditions associated with a $k = 2$ rotatable central-composite design (5 center runs, $n = 13$). Confidence regions using both bandwidth selectors were constructed using 2000 bootstrap samples in each simulated experiment. Bandwidths under the plug-in method were slightly larger, on average, than those of the Normal rule-of-thumb (see table 2). This result is not unexpected, as simulation studies of Wand and Jones (1994) indicate a tendency for the multivariate plug-in method to oversmooth in some cases.

A contour of the true concave-down response surface and a representative confidence region for $\boldsymbol{x}_{cm}$ with $b = 2000$ are provided in figures 1 and 2, respectively. Bootstrap estimates for $\boldsymbol{x}_{cm}$ are indicated with points in figure 2, with a solid line identifying the confidence region boundary. Note that the confidence region boundary is within the experimental region, as required. Had the hybrid or percentile-t methods been implemented this would not have been true in all cases.

The simulated experiments were arranged into five groups of 100 and the coverage probability calculated in each group for confidence coefficients ranging from $90-100\%$. The mean coverage probability, with error bars placed at one standard error from the mean in either direction, is plotted in figure 3. Note that there is little difference in coverage probability under the two approaches, a result expected from the close similarity in bandwidths seen in table 2. One approach for potentially improving the coverage probability is explored in the discussion section.

A contour plot of the true saddle system and a representative confidence region for $\boldsymbol{x}_{cm}$ are provided in figures 4 and 5, respectively. Note that in this case $\boldsymbol{x}_{cm}$ is located on the boundary of the experimental region. Since all bootstrap estimates for $\boldsymbol{x}_{cm}$ are also on the boundary, visual identification of the confidence region is difficult. In the figure, two small triangles identify the location of lower and upper boundaries in the x_2 coordinate. In the x_1 coordinate the confidence region extends slightly away from the experimental region boundary.

Figure 6 summarizes the coverage probability under both bandwidth selectors for the saddle system. As with the concave-down response surface, the coverage probability is statistically indistinguishable under the two approaches. The coverage probability is closer to the nominal level than for the concave-down response surface.

Calculation of the Normal rule-of-thumb bandwidths requires less time and computational resources than bandwidths under Wand and Jones' plug-in method. Since the coverage probability was essentially the same under both methods, the Normal rule-of-thumb method was implemented in the simulation studies that follow.

A simulation study was conducted to assess whether increasing the number of bootstrap samples would improve the coverage probability. Confidence regions were constructed from the same simulated data described earlier using $b = 4000$ and $b = 6000$. In neither the saddle system nor concave-down case was the coverage probability

significantly effected by increasing b.

To explore the effect of sample size on coverage probability, confidence regions for $\boldsymbol{x}_{cm}$ were constructed under both second-order response surfaces for $n = 13$, $n = 26$ and $n = 208$, corresponding to 1, 2 and 16 experimental replications of a rotatable central-composite design with 5 center runs. The results for the concave-down and saddle system with $b = 2000$ are summarized in figures 7 and 8, respectively. In both cases the coverage probability approaches the nominal level with increasing sample size.

3   Discussion

Existing confidence regions for operating conditions associated with the maximum of a response surface, either unconstrained or constrained to a specified region, are based on the stationary point. These approaches are only applicable to second-order models under the assumption of normally distributed errors. Also, the interpretation of these confidence regions can be ambiguous, since this requires assessment of the unknown elements of the $\boldsymbol{B}$ matrix.

In contrast, the interpretation of a likelihood-based bootstrap confidence region for $\boldsymbol{x}_{cm}$ does not depend on the nature of the response surface, assuming $\boldsymbol{x}_{cm}$ is unique. For example, if $g(\boldsymbol{x}, \boldsymbol{\theta})$ is a second-order polynomial the confidence region interpretation is the same whether $g(\boldsymbol{x}, \boldsymbol{\theta})$ is concave-down, concave-up or a saddle system. In addition, the approach is not restricted to second-order models nor does it require assumptions on the model error distribution, except for exchangeability of the errors. In principle, the methodology is also applicable to models where $g(\boldsymbol{x}, \boldsymbol{\theta})$ is nonlinear in $\boldsymbol{\theta}$, though Hjorth (1994, pp 190) indicates that in non-linear regression applications a direct analogue of standardized residuals is generally not available.

Simulation results from section 2.4 (see figure 7) indicate the bootstrap confidence region coverage probability was less than nominal under the concave-down system ($n = 13$). Coverage probability was higher for the saddle system ($n = 13$), but still below the nominal level (see figure 8). A possible reason that higher coverage probability was observed for the saddle system is the estimated conditional distribution of $\hat{\boldsymbol{x}}_{cm}$ is essentially restricted to one dimension in this case, thereby reducing the 'curse of dimensionality'. Evidence that the coverage probability in both systems converges to the nominal level with increasing sample size, an indication of the asymptotic accuracy of the methodology, is apparent in figures 7 and 8.

Loh (1987, 1991) proposed a bootstrap calibration technique to improve the accuracy of confidence sets. In brief, the approach consists of estimating the confidence coefficient associated with the desired coverage probability. For example, to achieve a coverage probability of 90% in the concave-down response surface case ($n = 13$), it is apparent in figure 7 that a confidence coefficient of approximately 99.5% is required. It is also apparent that in this case realization of coverage probabilities greater than $\approx 91\%$ are not feasible with this approach, unless the sample size is increased. For the saddle system ($n = 13$) a confidence coefficient of approximately 95% is needed to achieve a coverage probability of 90% (see figure 8).

The accuracy of bootstrap confidence regions for $\hat{\boldsymbol{x}}_{cm}$ using the percentile method is largely determined by how accurately the conditional distribution of $\hat{\boldsymbol{x}}_{cm}$, f, is estimated. Under the kernel method implemented in section 2, choice of bandwidth has a critical impact on this accuracy. Hall (1987) indicated that bandwidths which are optimal in a global sense, such as minimization of the mean integrated squared error of $\hat{f}$, can produce confidence regions that are "unduly bumpy". He noted that this can be attributed to the fact that the ratio of the variance to the squared bias of $\hat{f}$ is greater in the tails than in the central region of the distribution. Therefore,

the Normal rule-of-thumb and plug-in bandwidth selectors were investigated for their tendency to avoid undersmoothing. The $k = 2$ concave-down and saddle system cases explored in section 2.4 are examples where these methods provide reasonable bandwidth estimates. Jones (1993) followed a similar approach when he used a plug-in bandwidth selector in conjunction with a univariate boundary kernel.

There are situations, however, in which bandwidth selection is more challenging. For example, if f is bimodal, use of the standard deviation to estimate the scale of f can result in considerable oversmoothing. Janssen et al. (1995) developed more robust scale estimators for such cases. Consider, also, a situation where all bootstrap estimates for $\boldsymbol{x}_{cm}$ are equally distributed on the two boundaries $\mathcal{A} = \{\boldsymbol{x} \colon x_1 = 1.4, x_2 \in (-1.4, 1.4)\}$ and $\mathcal{B} = \{\boldsymbol{x} \colon x_1 \in (-1.4, 1.4), x_2 = 1.4\}$. On boundary $\mathcal{A}$ bandwidth h_1 should be much smaller than h_2. However, on boundary $\mathcal{B}$ the opposite is true. In this case, a variable kernel density estimator would seem more appropriate than the fixed bandwidth kernel estimator implemented in section 2.4, as this would allow for different levels of smoothing depending on the location of bootstrap estimates for $\boldsymbol{x}_{cm}$.

Application of the bootstrap confidence region methodology was restricted to the $k = 2$ case where the experimental region was rectangular in shape. The bootstrap and kernel density estimation techniques described in section 2 are not limited to two dimensions. Wand and Jones' plug-in bandwidth selector can also be extended to higher dimensions, though with greater computational expense. In light of the close comparison of coverage probability under the Normal rule-of-thumb and plug-in methods observed in section 2.4, rule-of-thumb methods may be adequate in many higher-dimensional applications. Graphical presentation of confidence regions for $k > 3$ is a challenging issue, though this is not unique to our application. Staniswalis, Messer and Finston (1993) describe a multivariate boundary kernel designed for regions of

arbitrary shape which obtains the correct order of bias over the entire region (see, also, Scott, 1992, pp 155).

Throughout this article discussion has focused on maximizing a single response. However, researchers in many areas of application are faced with the problem of simultaneously improving multiple responses that depend on a common set of controllable variables. Since a single operating condition is rarely optimal for all responses, compromise must be incorporated into the estimation procedure. Desirability optimization methodology (Harrington, 1965, Derringer and Suich, 1980) addresses this problem and has been proven effective in a wide range of applications involving continuous responses. In brief, the approach consists of estimating the operating conditions that maximize D, a simultaneous measure of the desirability of all the responses. An arguable shortcoming of the methodology is it does not provide for estimation of the variability of these operating conditions. Bootstrap confidence region methodology is defined in sufficiently general terms to allow for construction of a confidence region for the operating condition that maximizes D constrained to the experimental region, under the assumption that this parameter is unique.

Of practical interest is the computational resources and time required to construct a likelihood-based bootstrap confidence region. The programs that generated the results of section 2.4 were written in SAS$^{®}$ and run on a 300 megahertz Pentium$^{®}$ II desktop computer with 128 megabytes of RAM. Approximately 2 minutes is required to generate a single confidence region with $b = 2000$, of which about 25 seconds are used for steps $1 - 3$ with the balance of the time taken in steps 4 and 5. Wand (1994) describes methods of accelerating computations required for multivariate bandwidth optimization and kernel density estimation. These methods were not implemented due to technical limitations of SAS$^{®}$. It is expected that steps 4 and 5 would require considerably less time under their optimized procedures.

References

Box, G.E.P., and Hunter, J.S. (1954), "A Confidence Region for the Solution of a Set of Simultaneous Equations With An Application to Experimental Design," *Biometrika*, 41, 190-199.

Davison, A.C., Hinkley, D.V., and Schechtman, E. (1986), "Efficient Bootstrap Simulation," *Biometrika*, 73, 555-566.

Derringer, G., and Suich, R. (1980), "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, 12, 214-219.

Draper, N.R. (1963), "Ridge Analysis for Response Surfaces," *Technometrics*, 5, 469-479.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1-26.

Gasser, T., and Müller, H.-G (1979), "Kernel Estimation of Regression Functions," In *Smoothing Techniques for Curve Estimation* (eds. T. Gasser and M. Rosenblatt), Heidelberg: Springer-Verlag, pp 23-69.

Hall, P. (1987), "On the Bootstrap and Likelihood-Based Confidence Regions," *Biometrika*, 74, 481-493.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Harrington, E.C. (1965), "The Desirability Function," *Industrial Quality Control*, 21, 494-498.

Hart, J.D. and Wehrly, T.E. (1992), "Kernel Regression When the Boundary Region is Large, With An Application to Testing the Adequacy of Polynomial Models," *The Journal of the American Statistical Assocication*, 87, 1018-1024.

Hoerl, A.E. (1959), "Optimum Solution of Many Variables Equations," *Chemical Engineering Progress*, 55, 67-78.

Hjorth, J.S.U. (1994), *Computer Intensive Statistical Methods, Validation Model Selection and Bootstrap*, London: Chapman & Hall.

Janssen, P., Marron, J.S., Veraverbeke, N., and Sarle, W. (1995), "Scale Measures for Bandwidth Selection," *The Journal of Nonparametric Statistics*, 5, 359-380.

Jones, M.C. (1993), "Simple Boundary Correction for Kernel Density Estimation," *Statistics and Computing*, 3, 135-146.

Kendall, M. and Stuart, A. (1979), *The Advanced Theory of Statistics, Volume II: Inference and Relationship,* 4th edition, New York: Macmillan.

Loh, W.-Y. (1987), "Calibrating Confidence Coefficients," *The Journal of the American Statistical Association*, 82, 155-162.

Loh, W.-Y. (1991), "Bootstrap Calibration for Confidence Interval Construction and Selection," *Statistica Sinica*, 1, 479-495.

Müller, H.G. (1988). *Lecture Notes in Mathematics: Nonparametric Regression Analysis of Longitudinal Data.* New York: Springer-Verlag, 46.

Nelder, J.A., and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308-313.

Peterson, J.J. (1992), "Confidence Regions for Constrained Response Surface Optima," Presented at the August, 1992 Joint Statistical Meetings, Atlanta, GA.

Scott, D.W. (1992), *Multivariate Density Estimation: Theory, Practice and Visualization*, New York: Wiley.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Stablein, D.M., Carter, W.H., and Wampler, G.L. (1983), "Confidence Regions for Constrained Optima in Response-Surface Experiments," *Biometrics*, 39, 759-763.

Staniswalis, J.G., Messer, K., and Finston, D.R. (1993), "Kernel Estimators for Multivariate Regression," *Journal of Nonparametric Statistics* **3**, 103-121.

Wand, M.P. (1994), "Fast Computation of Multivariate Kernel Estimators," *Journal of Computational and Graphical Statistics*, 3, 433-445.

Wand, M.P., and Jones M.C. (1994), "Multivariate Plug-In Bandwidth Selection," *Computational Statistics Quarterly*, 9, 97-116.

Table 1: True second-order model parameters and associated $\boldsymbol{x}_{cm}$.

Response Surface	β_0	β_1	β_2	β_{12}	β_{11}	β_{22}	$\boldsymbol{x}_{cm}$
concave down	86.850	5.242	4.778	-0.775	-2.781	-2.524	$(0.828, 0.819)$
saddle	90.259	-6.425	1.244	-0.775	2.781	-2.524	$(-1.4, 0.462)$

Table 2: Mean and standard error (in parenthesis) of bandwidths estimated from 500 simulated response surfaces using the Normal rule-of-thumb and plug-in methods.

Bandwidth selector	Concave-down		Saddle system	
	h_1	h_2	h_1	h_2
Normal rule-of-thumb	0.196	0.213	0.011	0.261
	(0.0033)	(0.0037)	(6.9×10^{-6})	(0.0049)
Plug-in method	0.214	0.233	0.013	0.313
	(0.0034)	(0.0038)	(13×10^{-6})	(0.0060)
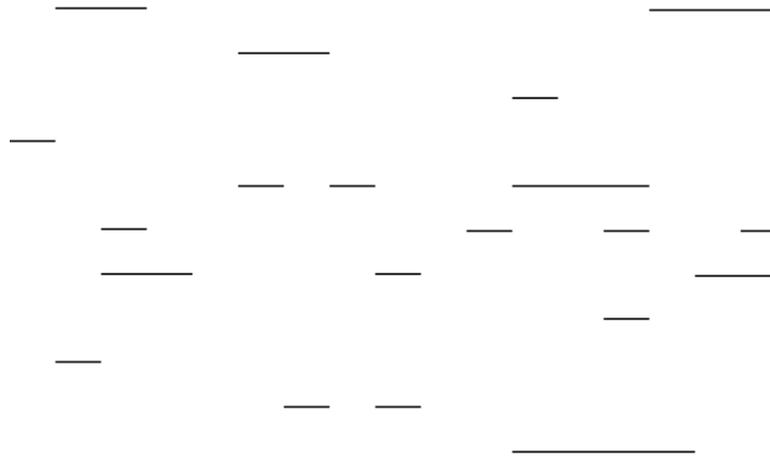
Figure 1: Contour plot of the true concave-down response surface, where $\times$ identifies $\boldsymbol{x}_{cm}$.
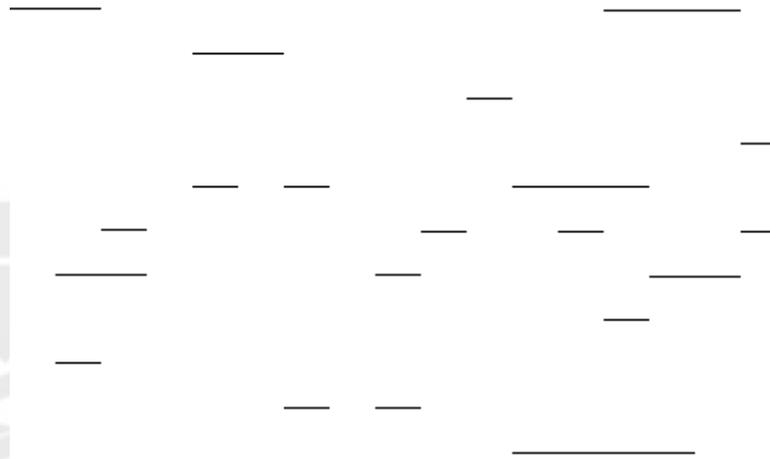

Figure 2: A 90% confidence region for $\boldsymbol{x}_{cm}$ where the true response surface is concave-down and $\times$ identifies $\boldsymbol{x}_{cm}$.
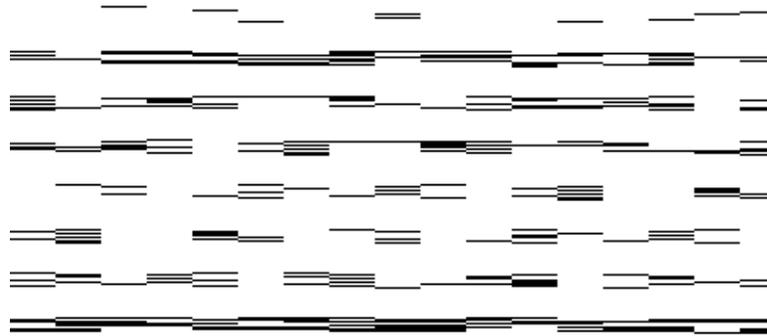
Figure 3: Comparison of coverage probability under the Normal rule-of-thumb and plug-in bandwidth selectors. The true response surface is concave-down.
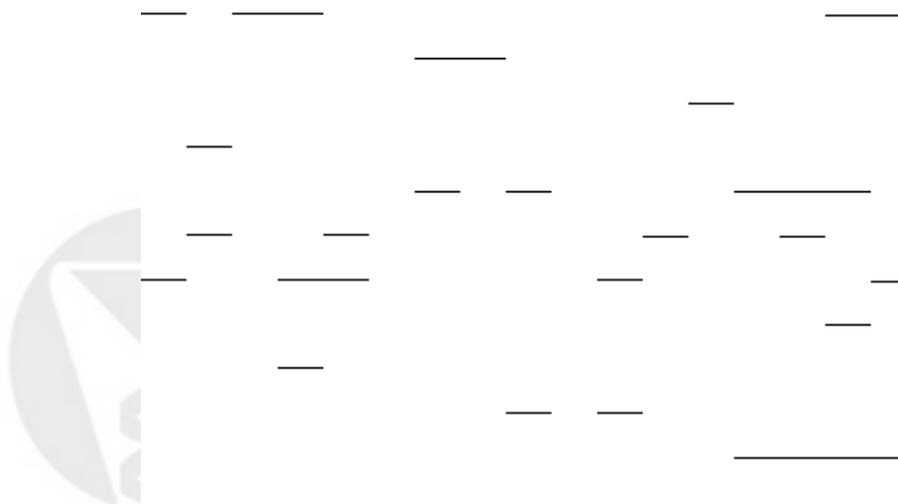

Figure 4: Contour plot of the true saddle system response surface, where $\times$ identifies $\boldsymbol{x}_{cm}$.
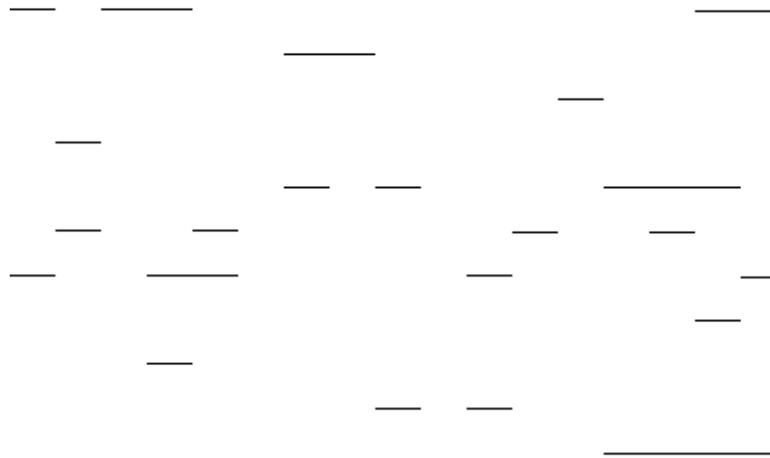
Figure 5: A 90% confidence region for $\boldsymbol{x}_{cm}$ where the true response surface is saddle system and $\times$ identifies $\boldsymbol{x}_{cm}$.
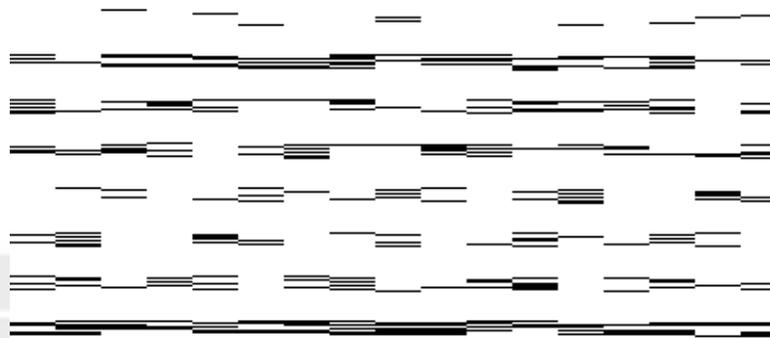

Figure 6: Comparison of coverage probability under the Normal rule-of-thumb and plug-in bandwidth selectors. The true response surface is a saddle system.

Figure 7: Comparison of coverage probability for three different sample sizes where the response surface is concave-down and $b = 2000$ with Normal-rule-of-thumb bandwidths.
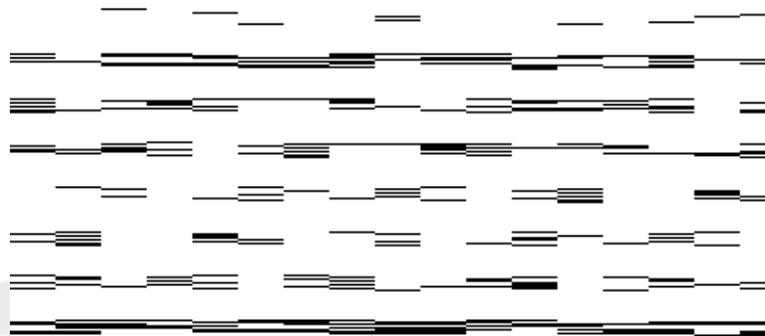

Figure 8: Comparison of coverage probability for three different sample sizes where the response surface is a saddle system and $b = 2000$ with Normal-rule-of-thumb bandwidths.