Research Article

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1058

Published online 10 August 2009 in Wiley InterScience

Bayesian Analysis for Weighted Mean-squared Error in Dual Response Surface Optimization

In-Jun Jeong,[a]*† Kwang-Jae Kim[b] and Dennis K. J. Lin[c]

Dual response surface optimization considers the mean and the variation simultaneously. The minimization of mean-squared error (MSE) is an effective approach in dual response surface optimization. Weighted MSE (WMSE) is formed by imposing the relative weights, ($\lambda, 1-\lambda$), on the squared bias and variance components of MSE. To date, a few methods have been proposed for determining λ. The resulting λ from these methods is either a single value or an interval. This paper aims at developing a systematic method to choose a λ value when an interval of λ is given. Specifically, this paper proposes a Bayesian approach to construct a probability distribution of λ. Once the distribution of λ is constructed, the expected value of λ can be used to form WMSE.

Keywords: Bayesian analysis; interval of weight; weighted mean-squared error

1. Introduction

Response surface methodology consists of a group of techniques used in the empirical study between a response and a number of input variables. For detailed descriptions on various response surface techniques, see, for example, Box and Draper[1] and Myers and Montgomery[2]. The conventional response surface methodology focuses on the mean of the response, assuming that the variance of the response is constant. However, the constant variance assumption may not be valid in practice.

The dual response surface approach has received a great deal of attention for its attempt to tackle the non-equal variance problem (Vining and Myers[3]). Suppose that there is a response y that is determined by k input variables, coded $x_1, x_2, \ldots, x_k$. The dual response surface approach first fits models for the mean ($\hat{\omega}_\mu$) and the standard deviation ($\hat{\omega}_\sigma$) as separate responses. A quadratic (second-order) polynomial form is typically used for the model building:

$$\begin{aligned}\hat{\omega}_\mu &= \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2 + \sum\sum_{i<j}^{k} \hat{\beta}_{ij} x_i x_j \\ \hat{\omega}_\sigma &= \hat{\gamma}_0 + \sum_{i=1}^{k} \hat{\gamma}_i x_i + \sum_{i=1}^{k} \hat{\gamma}_{ii} x_i^2 + \sum\sum_{i<j}^{k} \hat{\gamma}_{ij} x_i x_j\end{aligned} \tag{1}$$

Various methods have been proposed for the optimization of the dual response surfaces (Del Castillo and Montgomery[4], Lin and Tu[5], Copeland and Nelson[6], Del Castillo *et al.*[7], Kim and Lin[8]). Tang and Xu[9] and Köksoy and Doganaksoy[10] provide a good review on these existing methods.

Lin and Tu[5] propose an effective approach based on the mean-squared error (MSE) minimization:

$$\underset{\mathbf{x}}{\text{Minimize}}\ \ \text{MSE} = (\hat{\omega}_\mu - T)^2 + \hat{\omega}_\sigma^2 \tag{2}$$

where T is the target value for the mean and $\mathbf{x} = (x_1, \ldots, x_k)$ is a vector of the input variables. It should be noted that Equation (2) is the MSE minimization model for a nominal-the-best (NTB)-type response. The goal of the model is to have the mean response at some target value with a minimum variation. MSE consists of two terms: the squared bias ($(\hat{\omega}_\mu - T)^2$) and the variance ($\hat{\omega}_\sigma^2$). As noted in Lin and Tu[5], a natural extension of MSE, viz., weighted MSE (WMSE) is formed by imposing the relative weights on the squared bias and variance terms:

$$\text{WMSE} = \lambda(\hat{\omega}_\mu - T)^2 + (1-\lambda)\hat{\omega}_\sigma^2 \tag{3}$$

where λ is a weighting factor ($0 \le \lambda \le 1$).

[a] Technology Business Division, Electronics and Telecommunications Research Institute, Daejeon 305-700, Republic of Korea
[b] Division of Mechanical and Industrial Engineering, Pohang University of Science and Technology, Pohang, Kyungbuk 790-784, Republic of Korea
[c] Department of Statistics, The Pennsylvania State University, University Park, PA 16802, U.S.A.
* Correspondence to: In-Jun Jeong, Technology Business Division, Electronics and Telecommunications Research Institute, Daejeon 305-700, Republic of Korea.
† E-mail: ijjeong@etri.re.kr

One reason to introduce the weighting to the MSE criterion is that one of the squared bias and variance could be more important than the other. As an example, in the manufacturing process of tin plates, a kind of steel product, reducing the variation of hardness is critical to quality. This is because the final products (e.g. beverage can) made of the plates with a high variability are easy to crack, while the bias associated with the mean value does not significantly affect the likelihood of crack formation (Kim *et al.*[11]). In this case, the variance is much more critical than the squared bias. In order to solve such a problem appropriately, the relative importance of the squared bias and variance needs to be incorporated into the MSE criterion through the weighting.

Another reason is that the MSE criterion does not consider the relative scale (or magnitude) of the squared bias and variance. As mentioned in Kim and Lin[8], the minimization of MSE can be driven mainly by one of the two components whichever has a larger scale, although a small amount of the other might have a significant impact on the overall quality. In order to consider the impact of the squared bias and variance in a balanced manner, the scale effect should be adjusted through a proper weight.

An appropriate value of λ should be determined to make the WMSE criterion effective. To date, a few methods have been proposed for determining λ. Ding *et al.*[12] propose a data-driven method to determine λ. It first generates the ideal point in the mean-standard deviation $(\hat{\omega}_\mu, \hat{\omega}_\sigma)$ space, which consists of the marginally optimal values of $\hat{\omega}_\mu$ and $\hat{\omega}_\sigma$ obtained by individually minimizing $(\hat{\omega}_\mu - T)^2$ and $\hat{\omega}_\sigma^2$, respectively. Then, it finds an efficient curve through minimizing WMSE for all possible λ values between 0 and 1. It has been shown that any optimal point must lie on such an efficient curve (corresponding to a specific value of λ). Finally, it identifies the point on the efficient curve that is closest to the ideal point and the corresponding λ value. The obtained λ value is considered as an optimal choice.

Jeong *et al.*[13] propose a preference-based method to determine λ. The basic premise of this method is that λ should be congruent with a decision maker (DM)'s preference structure. The squared bias and variance values resulting from a process setting of the input variables are considered to form a vector. The DM expresses his/her preference structure by providing the rankings of such alternative vectors in a pairwise manner. Then, the method finds the λ value(s) congruent with the given rankings of the vectors.

The motivation of the data-driven method and the preference-based method is that the DM encounters difficulties in directly specifying a value of λ, which is a mathematical representation of the relative importance and/or scale of the squared bias and variance. They focus on finding a value of λ that is considered reasonable by the DM rather than having this as a user-specified characteristic, and therefore facilitating the process through which he/she chooses a value of λ systematically.

Unlike the Ding *et al.*[12]'s method, which always generates a single value of λ, the Jeong *et al.*[13]'s method provides either a single value or an interval of λ, depending on the consistency of the preference structure expressed by the DM. In the single value case, the generated value of λ can be directly used to combine the squared bias and variance in WMSE. In the interval case, a specific value of λ in the interval needs to be chosen to form WMSE for practical use. The choice of λ should be made carefully, particularly when the optimal setting is sensitive with respect to the chosen λ value.

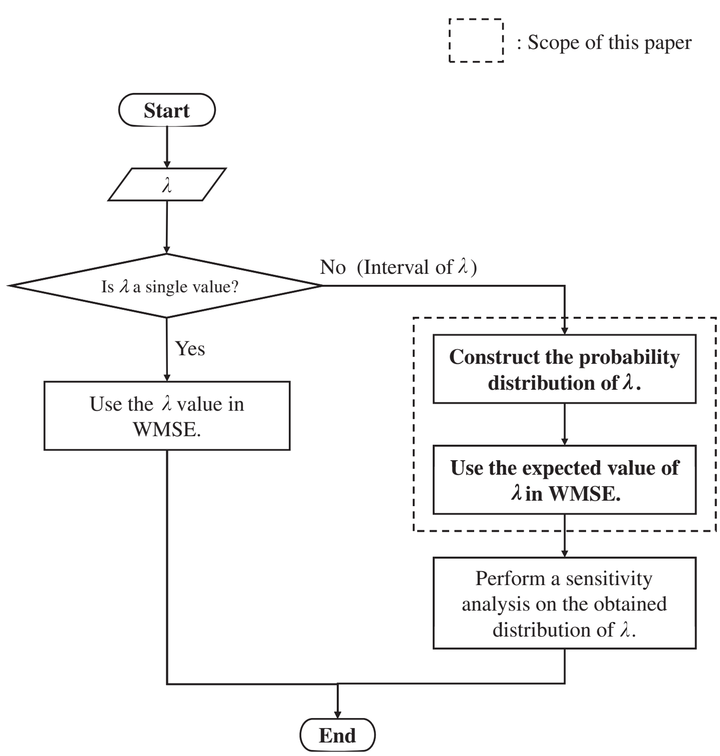

Figure 1. The context and scope of this paper

A simple method to help to choose a λ value is to narrow the interval, as mentioned in Jeong *et al.*[13]. The interval can be narrowed by using additional pairwise rankings of new alternative vectors. The new alternative vectors can be predicted by using the regression models in Equation (1) at various interpolated design points. If the interval is not sufficiently narrow, mild extrapolations could be employed. This method would require a large number of iterations with the DM before reaching a sufficiently narrow interval.

This paper aims at developing a systematic method to choose a λ value when an interval of λ is given. Specifically, this paper proposes to construct a probability distribution of λ, from which a meaningful choice of λ can be made. A Bayesian approach is used for this purpose. The lower and upper bounds of the λ interval are employed as the sample information in the Bayesian analysis. Once the (posterior) distribution of λ is constructed, the expected value of λ can be used to form WMSE. The Bayesian approach is able to utilize the prior knowledge to construct the distribution of λ. In addition, it can update the distribution of λ with the knowledge obtained through the sample information. The more the knowledge accumulates, the more the degree of belief in the value of λ increases.

Note that the proposed Bayesian approach is applicable in the single value case as well as in the interval case. However, it is particularly important to handle the case where λ is obtained as an interval but a single λ value in the interval should be chosen. Therefore, this study will focus on the interval case. The context and scope of this paper are shown in Figure 1. The preference-based method is reviewed in the next section, and then the proposed Bayesian scheme is presented. The popular 'printing process' problem is used for a detailed illustration on how to apply the proposed scheme, followed by a sensitivity analysis on various prior assumptions.

2. Review of the preference-based method

Consider m alternative vectors consisting of the squared bias and variance values. Let $\mathbf{z}^i=(z_1^i,\ z_2^i)^{\mathrm{T}}(i=1,2,\ldots,m)$, where z_1^i and z_2^i are the squared bias and variance components, respectively. The WMSE value associated with $\mathbf{z}^i$ is expressed as

$$\mathrm{WMSE}^i=\boldsymbol{\lambda}\mathbf{z}^i$$

where $\boldsymbol{\lambda}=(\lambda,1-\lambda)$. The basic premise of the preference-based method is that the pairwise ranking of $\mathbf{z}^i$'s given by the DM should be consistent with the corresponding WMSE values. For example, if the DM prefers $\mathbf{z}^i$ to $\mathbf{z}^j$ (denoted as $\mathbf{z}^i\succ\mathbf{z}^j$), $\boldsymbol{\lambda}\mathbf{z}^i$ should be less than $\boldsymbol{\lambda}\mathbf{z}^j$ ($\boldsymbol{\lambda}\mathbf{z}^i<\boldsymbol{\lambda}\mathbf{z}^j$).

The elicitation process to obtain the pairwise rankings is basically the DM's preference articulation process (Jeong *et al.*[13]). Given $\mathbf{z}^i$ and $\mathbf{z}^j$, the DM is asked to make a judgment as to which alternative is preferred to the other. In order to make such a judgment, the DM would compare their squared bias and variance values and evaluate the tradeoffs in terms of the overall quality. From a practical viewpoint, the DM would think of their implications on various quality indices, such as process capability, proportion of rejects, quality loss, and the like. This judgment requires knowledge about the process/product under study. Here, it should be noted that including the MSE value for judgment is improper, because the pairwise rankings based on MSE would lead to the unweighting (i.e. $\lambda=0.5$).

The preference-based method first extracts the pairwise rankings of $\mathbf{z}^1,\mathbf{z}^2,\ldots,\mathbf{z}^m$ from the DM. The pairwise ranking has three types: (i) $\mathbf{z}^i$ is preferred to $\mathbf{z}^j$ ($\mathbf{z}^i\succ\mathbf{z}^j$); (ii) $\mathbf{z}^i$ is less preferred to $\mathbf{z}^j$ ($\mathbf{z}^i\prec\mathbf{z}^j$); and (iii) the comparison is to be held back ($\mathbf{z}^i?\mathbf{z}^j$). It is worth mentioning that the DM may wish to refer to the unsquared bias and standard deviation for the pairwise comparison, because it is easier to think in terms of actual units rather than squared units. In addition, the unsquared bias tells one whether the bias is positive or negative, about which the DM may have a concern. The resulting pairwise rankings are summarized as an index set of ordered pairs:

$$I_P=\{(i,j)|\mathbf{z}^i\succ\mathbf{z}^j\}$$

where the subscript P denotes *Preferred*. The set I_P is associated with the cases (i) and (ii). Note that (ii) can be easily transformed to (i) by changing the position of $\mathbf{z}^i$ and $\mathbf{z}^j$.

The extracted pairwise rankings are transformed to inequalities between the corresponding WMSE values:

$$\boldsymbol{\lambda}\mathbf{z}^i<\boldsymbol{\lambda}\mathbf{z}^j\quad\forall(i,j)\in I_P \tag{4}$$

By solving each inequality in Equation (4) for $\lambda\in[0,1]$, the individual set of λ's for the (i,j) pair ($\Lambda_{(i,j)}$) is obtained as an interval. The set of λ's satisfying all the inequalities (Λ) is obtained by intersecting $\Lambda_{(i,j)}$'s, $\forall(i,j)$. If $\Lambda\neq\emptyset$, the algorithm ends, and the current solution Λ is the final set of λ's. The set Λ, obtained as an interval, is the set of λ's that are congruent with all the current pairwise rankings. Otherwise, there exists an inconsistency among the pairwise rankings. The preference-based method attempts to remove such an inconsistency through a revision process. (For detailed information on the revision process, see Jeong *et al.*[13]—in particular, Steps 3 and 4 of the method.) If the inconsistency cannot be removed nevertheless, a single value of λ, which minimizes the degree of inconsistency, is obtained instead.

3. Proposed Bayesian scheme

The Bayesian analysis combines the prior probability distribution on a parameter (θ), which is unobservable, and the sample information (x), which is observable, to construct the posterior probability distribution of θ given x. In our work, λ is regarded as a parameter, and the lower and upper bounds of the λ interval (say, $\underline{\lambda}$ and $\overline{\lambda}$) serve as the sample information. Let $\pi(\lambda)$ be the prior probability density function (pdf) of λ, and $f(\underline{\lambda}, \overline{\lambda}|\lambda)$ be the joint likelihood function of $\underline{\lambda}$ and $\overline{\lambda}$ given λ. Then, the posterior pdf of λ given $\underline{\lambda}$ and $\overline{\lambda}$ is obtained as

$$p(\lambda|\underline{\lambda}, \overline{\lambda}) = \frac{\pi(\lambda) f(\underline{\lambda}, \overline{\lambda}|\lambda)}{\int_\lambda \pi(\lambda) f(\underline{\lambda}, \overline{\lambda}|\lambda)\,\mathrm{d}\lambda} \tag{5}$$

In Equation (5), $\pi(\lambda)$ needs to be determined based on the prior knowledge. Various methods have been proposed for this purpose (Berger[14]). The likelihood function $f(\underline{\lambda}, \overline{\lambda}|\lambda)$ can be determined based on theoretical assumptions and available empirical information on $\underline{\lambda}$, $\overline{\lambda}$, and λ. In this study, the likelihood function is derived from the lower and upper bounds of $\Lambda_{(i,j)}$'s in conjunction with order statistics. (The derivation of the likelihood function will be described in more detail in Section 4.1.) Once $p(\lambda|\underline{\lambda}, \overline{\lambda})$ is constructed, the expected value of λ given $\underline{\lambda}$ and $\overline{\lambda}$ is obtained as

$$E[\lambda|\underline{\lambda}, \overline{\lambda}] = \int_\lambda \lambda \cdot p(\lambda|\underline{\lambda}, \overline{\lambda})\,\mathrm{d}\lambda$$

This expected value of λ can then be used in Equation (3), instead of λ, to form WMSE. In addition to the posterior expectation, a sensitivity analysis can be performed by using the lower and upper percentiles (e.g. the 5th and 95th percentiles) of the obtained distribution of λ (Figure 1).

4. Illustrative example

The 'printing process' problem, which was originally discussed in Box and Draper[1] (p. 247), is used here to illustrate the proposed Bayesian scheme. The purpose of the problem is to improve the quality of a printing process (y) by controlling speed (x_1), pressure (x_2), and distance (x_3). The experiment was conducted in a 3^3 factorial design with three replicates. Table I shows the data set, along with the unsquared bias and standard deviation values as well as the squared bias and variance values. In the problem, the fitted response surfaces for the mean and the standard deviation are as follows (Vining and Myers[3]):

$$\hat{\omega}_\mu = 327.6 + 177.0x_1 + 109.4x_2 + 131.5x_3 + 32.0x_1^2 - 22.4x_2^2 - 29.1x_3^2 + 66.0x_1x_2 + 75.5x_1x_3 + 43.6x_2x_3 \quad (T = 500)$$

$$\hat{\omega}_\sigma = 34.9 + 11.5x_1 + 15.3x_2 + 29.2x_3 + 4.2x_1^2 - 1.3x_2^2 + 16.8x_3^2 + 7.7x_1x_2 + 5.1x_1x_3 + 14.1x_2x_3$$

For this problem, Jeong *et al.*[13] obtained an interval of λ using the preference-based method. As previously mentioned, the squared bias and variance values in Table I, resulting from the 27 experimental conditions of the input variables, might be used to form the alternative vectors. However, only seven vectors (denoted as $\mathbf{z}^1, \ldots, \mathbf{z}^7$ in Table I), which passed a screening test, were finally used. The method extracted the initial pairwise rankings of $\mathbf{z}^1, \ldots, \mathbf{z}^7$ from the DM and transformed them to the inequalities between the corresponding WMSE values. The initial pairwise rankings are given in Table II. By solving the inequalities for $\lambda \in [0, 1]$, $\Lambda_{(i,j)}$'s and Λ were obtained. But, the set Λ was null (i.e. $\Lambda = \emptyset$). Then, the method removed inconsistent pairwise rankings through the revision process. Specifically, it changed $\mathbf{z}^3 \prec \mathbf{z}^5$ into $\mathbf{z}^3 \succ \mathbf{z}^5$, and deleted $\mathbf{z}^4 \succ \mathbf{z}^6$. From the revised pairwise rankings, $\Lambda_{(i,j)}$'s and Λ were obtained again. Then, the set Λ was obtained as (0.064, 0.246). The intermediate and final results of the example are summarized in Table III. In this paper, the closed interval [0.064, 0.246] is used for computational convenience. The interval [0.064, 0.246] is the set of λ values satisfying all the inequalities (i.e. all the pairwise rankings) in Table III. That is, any value of λ in this interval is congruent with the DM's stated preferences and thus may be used in WMSE. However, at this point, it is not clear which value is most desirable to form WMSE.

Figure 2 shows the optimal settings (Figure 2(a)), the corresponding squared bias and variance values (Figure 2(b)), and the unsquared bias and standard deviation values (Figure 2(c)) for various λ values within the obtained interval. The optimal settings (x_2^* and x_3^* in particular) and the squared bias (unsquared bias) and variance (standard deviation) values do not show a stable pattern within the interval. In order to make a meaningful choice of λ in the interval, we construct a probability distribution of λ using the proposed Bayesian scheme, assuming the given interval serves as the sample information in the Bayesian analysis.

4.1. Construction of the posterior pdf of λ given $\underline{\lambda}$ and $\overline{\lambda}$

To construct the posterior pdf of λ given $\underline{\lambda}$ and $\overline{\lambda}$ (i.e. $p(\lambda|\underline{\lambda}, \overline{\lambda})$), the prior pdf $\pi(\lambda)$ and the joint likelihood function $f(\underline{\lambda}, \overline{\lambda}|\lambda)$ are needed. In this example, we obtain $f(\underline{\lambda}, \overline{\lambda}|\lambda)$ as the product of the likelihood function of $\underline{\lambda}$ given λ (say, $f_1(\underline{\lambda}|\lambda)$) and the likelihood function of $\overline{\lambda}$ given $\underline{\lambda}$ and λ (say, $f_2(\overline{\lambda}|\underline{\lambda}, \lambda)$). Then, Equation (5) can be transformed into

$$p(\lambda|\underline{\lambda}, \overline{\lambda}) = \frac{\pi(\lambda) f_1(\underline{\lambda}|\lambda) f_2(\overline{\lambda}|\underline{\lambda}, \lambda)}{\int_\lambda \pi(\lambda) f_1(\underline{\lambda}|\lambda) f_2(\overline{\lambda}|\underline{\lambda}, \lambda)\,\mathrm{d}\lambda}$$

Now, $\pi(\lambda)$, $f_1(\underline{\lambda}|\lambda)$, and $f_2(\overline{\lambda}|\underline{\lambda}, \lambda)$ are to be determined.

Table I. The printing process study data

i	x_1	x_2	x_3	y_1^i	y_2^i	y_3^i	$\bar{y}^i$	$(\bar{y}^i - T)$ (unsquared bias)	s^i (standard deviation)	$(\bar{y}^i - T)^2$ (squared bias)	$(s^i)^2$ (variance)	Code
1	−1	−1	−1	34	10	28	24.00	−476.00	12.49	226 576.00	156.00	
2	0	−1	−1	115	116	130	120.33	−379.67	8.39	144 146.78	70.33	
3	1	−1	−1	192	186	263	213.67	−286.33	42.83	81 986.78	1834.33	
4	−1	0	−1	82	88	88	86.00	−414.00	3.46	171 396.00	12.00	
5	0	0	−1	44	178	188	136.67	−363.33	80.41	132 011.11	6465.33	
6	1	0	−1	322	350	350	340.67	−159.33	16.17	25 387.11	261.33	$\mathbf{z}^1$
7	−1	1	−1	141	110	86	112.33	−387.67	27.57	150 285.44	760.33	
8	0	1	−1	259	251	259	256.33	−243.67	4.62	59 373.44	21.33	
9	1	1	−1	290	280	245	271.67	−228.33	23.63	52 136.11	558.33	
10	−1	−1	0	81	81	81	81.00	−419.00	0.00	175 561.00	0.00	
11	0	−1	0	90	122	93	101.67	−398.33	17.67	158 669.44	312.33	
12	1	−1	0	319	376	376	357.00	−143.00	32.91	20 449.00	1083.00	$\mathbf{z}^2$
13	−1	0	0	180	180	154	171.33	−328.67	15.01	108 021.78	225.33	
14	0	0	0	372	372	372	372.00	−128.00	0.00	16 384.00	0.00	$\mathbf{z}^3$
15	1	0	0	541	568	396	501.67	1.67	92.50	2.78	8556.33	$\mathbf{z}^4$
16	−1	1	0	288	192	312	264.00	−236.00	63.50	55 696.00	4032.00	
17	0	1	0	432	336	513	427.00	−73.00	88.61	5329.00	7851.00	$\mathbf{z}^5$
18	1	1	0	713	725	754	730.67	230.67	21.08	53 207.11	444.33	
19	−1	−1	1	364	99	199	220.67	−279.33	133.82	78 027.11	17908.33	
20	0	−1	1	232	221	266	239.67	−260.33	23.46	67 773.44	550.33	
21	1	−1	1	408	415	443	422.00	−78.00	18.52	6084.00	343.00	$\mathbf{z}^6$
22	−1	0	1	182	233	182	199.00	−301.00	29.44	90 601.00	867.00	
23	0	0	1	507	515	434	485.33	−14.67	44.64	215.11	1992.33	$\mathbf{z}^7$
24	1	0	1	846	535	640	673.67	173.67	158.21	30 160.11	25 030.33	
25	−1	1	1	236	126	168	176.67	−323.33	55.51	104 544.44	3081.33	
26	0	1	1	660	440	403	501.00	1.00	138.94	1.00	19303.00	
27	1	1	1	878	991	1161	1010.00	510.00	142.45	260 100.00	20 293.00	

$T = 500$.

Table II. The Initial pairwise rankings on the illustrative example

	$\mathbf{z}^1$	$\mathbf{z}^2$	$\mathbf{z}^3$	$\mathbf{z}^4$	$\mathbf{z}^5$	$\mathbf{z}^6$	$\mathbf{z}^7$
$\mathbf{z}^1$							
$\mathbf{z}^2$	?						
$\mathbf{z}^3$	*	*					
$\mathbf{z}^4$	$\prec$	$\prec$	$\prec$				
$\mathbf{z}^5$	$\prec$	$\prec$	$\succ$	?			
$\mathbf{z}^6$	$\succ$	*	$\succ$	$\prec$	$\succ$		
$\mathbf{z}^7$	$\succ$	$\succ$	?	$\succ$	*	?	

*The blank cell indicates that the pairwise comparison is not made due to a dominance relationship between the corresponding vectors.

First, $\pi(\lambda)$ should be determined based on, if any, the prior knowledge on λ as mentioned in Section 3. However, we have no prior knowledge in this problem. Therefore, we assume that any weighting of the squared bias or variance is equally valued. That is, $\pi(\lambda)$ is determined as the uniform pdf on $[0, 1]$:

$$\pi(\lambda) = 1, \quad 0 \leq \lambda \leq 1$$

This pdf is called a 'noninformative' prior pdf in the Bayesian literature. The use of a noninformative prior pdf is a common practice when no prior knowledge is available. The underlying idea behind adopting the noninformative prior pdf is that the likelihood functions play a dominant role in the construction of the posterior pdf (Raghunathan[15]). A thorough sensitivity analysis will be conducted in a later section to assess the impact of the choice of the prior pdf.

Table III. The intermediate and final results on the illustrative example

(i, j)	Pairwise rankings: Initial	Pairwise rankings: Revised	Inequalities	$\Lambda_{(i,j)}$
(1, 2)	$\mathbf{z}^1 ? \mathbf{z}^2$	$\mathbf{z}^1 ? \mathbf{z}^2$	—	—
(1, 4)	$\mathbf{z}^1 \succ \mathbf{z}^4$	$\mathbf{z}^1 \succ \mathbf{z}^4$	$\lambda\mathbf{z}^1 < \lambda\mathbf{z}^4$	[0, **0.246**)
(1, 5)	$\mathbf{z}^1 \succ \mathbf{z}^5$	$\mathbf{z}^1 \succ \mathbf{z}^5$	$\lambda\mathbf{z}^1 < \lambda\mathbf{z}^5$	[0, 0.275)
(1, 6)	$\mathbf{z}^1 \prec \mathbf{z}^6$	$\mathbf{z}^1 \prec \mathbf{z}^6$	$\lambda\mathbf{z}^1 > \lambda\mathbf{z}^6$	(0.004, 1]
(1, 7)	$\mathbf{z}^1 \prec \mathbf{z}^7$	$\mathbf{z}^1 \prec \mathbf{z}^7$	$\lambda\mathbf{z}^1 > \lambda\mathbf{z}^7$	(**0.064**, 1]
(2, 4)	$\mathbf{z}^2 \succ \mathbf{z}^4$	$\mathbf{z}^2 \succ \mathbf{z}^4$	$\lambda\mathbf{z}^2 < \lambda\mathbf{z}^4$	[0, 0.268)
(2, 5)	$\mathbf{z}^2 \succ \mathbf{z}^5$	$\mathbf{z}^2 \succ \mathbf{z}^5$	$\lambda\mathbf{z}^2 < \lambda\mathbf{z}^5$	[0, 0.309)
(2, 7)	$\mathbf{z}^2 \prec \mathbf{z}^7$	$\mathbf{z}^2 \prec \mathbf{z}^7$	$\lambda\mathbf{z}^2 > \lambda\mathbf{z}^7$	(0.043, 1]
(3, 4)	$\mathbf{z}^3 \succ \mathbf{z}^4$	$\mathbf{z}^3 \succ \mathbf{z}^4$	$\lambda\mathbf{z}^3 < \lambda\mathbf{z}^4$	[0, 0.343)
(3, 5)	$\mathbf{z}^3 \prec \mathbf{z}^5$	$\mathbf{z}^3 \succ \mathbf{z}^5$	$\lambda\mathbf{z}^3 < \lambda\mathbf{z}^5$	[0, 0.415)
(3, 6)	$\mathbf{z}^3 \prec \mathbf{z}^6$	$\mathbf{z}^3 \prec \mathbf{z}^6$	$\lambda\mathbf{z}^3 > \lambda\mathbf{z}^6$	(0.032, 1]
(3, 7)	$\mathbf{z}^3 ? \mathbf{z}^7$	$\mathbf{z}^3 ? \mathbf{z}^7$	—	—
(4, 5)	$\mathbf{z}^4 ? \mathbf{z}^5$	$\mathbf{z}^4 ? \mathbf{z}^5$	—	—
(4, 6)	$\mathbf{z}^4 \succ \mathbf{z}^6$	$\mathbf{z}^4 ? \mathbf{z}^6$	—	—
(4, 7)	$\mathbf{z}^4 \prec \mathbf{z}^7$	$\mathbf{z}^4 \prec \mathbf{z}^7$	$\lambda\mathbf{z}^4 > \lambda\mathbf{z}^7$	[0, 0.969)
(5, 6)	$\mathbf{z}^5 \prec \mathbf{z}^6$	$\mathbf{z}^5 \prec \mathbf{z}^6$	$\lambda\mathbf{z}^5 > \lambda\mathbf{z}^6$	[0, 0.909)
(6, 7)	$\mathbf{z}^6 ? \mathbf{z}^7$	$\mathbf{z}^6 ? \mathbf{z}^7$	—	—
				$\Lambda = ($**0.064**, **0.246**$)$

Next, $f_1(\underline{\lambda}|\lambda)$ and $f_2(\bar{\lambda}|\underline{\lambda}, \lambda)$ are determined using the lower and upper bounds of $\Lambda_{(i,j)}$'s as the sample information. In the preference-based method, the set Λ is obtained by intersecting $\Lambda_{(i,j)}$'s. That is, the maximum among the lower bounds and minimum among the upper bounds of $\Lambda_{(i,j)}$'s become the lower and upper bounds of Λ (i.e. $\underline{\lambda}$ and $\bar{\lambda}$), respectively. Then, a specific value between $\underline{\lambda}$ and $\bar{\lambda}$ is chosen as λ. Here, the lower bounds (or upper bounds) of $\Lambda_{(i,j)}$'s can be viewed as a sample from a continuous distribution on $[0, \lambda]$ (or $[\lambda, 1]$) when λ is known.

The individual sets $\Lambda_{(i,j)}$'s are based upon I_P, which is derived from the sample squared biases and variances, and the DM's opinions. In an experiment, the value of a response variable depends on the level of input variables (factors). The input levels are purposely designed by an experimental objective. Therefore, it cannot be assured that the response values (resulting from the designed input levels) and corresponding squared biases and variances are random samples. Moreover, it is not known exactly how the DM might construct I_P. Therefore, under current circumstances, we cannot conclude that the lower bounds (or upper bounds) of $\Lambda_{(i,j)}$'s are a 'random' sample. However, for the purpose of facilitating a theoretic development, they will be treated as a random sample.

Suppose that $\underline{\lambda}_1, \underline{\lambda}_2, \ldots, \underline{\lambda}_n$ are a sample of lower bounds randomly selected from the uniform distribution on $[0, \lambda]$. Here a uniform distribution is assumed as a typical Bayesian practice when the relevant information is vague. Let $\underline{\lambda}_{(1)} \le \underline{\lambda}_{(1)} \le \cdots \le \underline{\lambda}_{(n)}$ be the ordered values. Here, the nth-order statistic $\underline{\lambda}_{(n)}$ is considered as $\underline{\lambda}$. The pdf of $\underline{\lambda}$ becomes $f_1(\underline{\lambda}|\lambda)$:

$$f_1(\underline{\lambda}|\lambda) = g(\underline{\lambda}|\lambda, n) = n(G(\underline{\lambda}|\lambda))^{n-1} g(\underline{\lambda}|\lambda) = n\left(\frac{\underline{\lambda}}{\lambda}\right)^{n-1}\left(\frac{1}{\lambda}\right) = \frac{n\underline{\lambda}^{n-1}}{\lambda^n}, \quad 0 \le \underline{\lambda} \le \lambda \tag{6}$$

where $g(\underline{\lambda}|\lambda)$ and $G(\underline{\lambda}|\lambda)$ are the uniform pdf and cumulative density function (cdf) of $\underline{\lambda}$ given λ, respectively:

$$g(\underline{\lambda}|\lambda) = \left(\frac{1}{\lambda}\right) \quad \text{and} \quad G(\underline{\lambda}|\lambda) = \left(\frac{\underline{\lambda}}{\lambda}\right), \quad 0 \le \underline{\lambda} \le \lambda$$

Likewise, suppose that $\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_n$ are a sample of upper bounds randomly selected from the uniform distribution on $[\lambda, 1]$. Then, let $\bar{\lambda}_{(1)} \le \bar{\lambda}_{(1)} \le \cdots \le \bar{\lambda}_{(n)}$ be the ordered values. Here, the first-order statistic $\bar{\lambda}_{(1)}$ is $\bar{\lambda}$. The pdf of $\bar{\lambda}$ becomes $f_2(\bar{\lambda}|\underline{\lambda}, \lambda)$:

$$f_2(\bar{\lambda}|\underline{\lambda}, \lambda) = h(\bar{\lambda}|\underline{\lambda}, \lambda, n) = n(1 - H(\bar{\lambda}|\underline{\lambda}, \lambda))^{n-1} h(\bar{\lambda}|\underline{\lambda}, \lambda) = n\left(1 - \frac{\bar{\lambda} - \lambda}{1 - \lambda}\right)^{n-1}\left(\frac{1}{1 - \lambda}\right) = \frac{n(1 - \bar{\lambda})^{n-1}}{(1 - \lambda)^n}, \quad \underline{\lambda} \le \lambda \le \bar{\lambda} \le 1 \tag{7}$$

where $h(\bar{\lambda}|\underline{\lambda}, \lambda)$ and $H(\bar{\lambda}|\underline{\lambda}, \lambda)$ are the uniform pdf and cdf of $\bar{\lambda}$ given $\underline{\lambda}$ and λ, respectively:

$$h(\bar{\lambda}|\underline{\lambda}, \lambda) = \left(\frac{1}{1 - \lambda}\right) \quad \text{and} \quad H(\bar{\lambda}|\underline{\lambda}, \lambda) = \left(\frac{\bar{\lambda} - \lambda}{1 - \lambda}\right), \quad \underline{\lambda} \le \lambda \le \bar{\lambda} \le 1$$

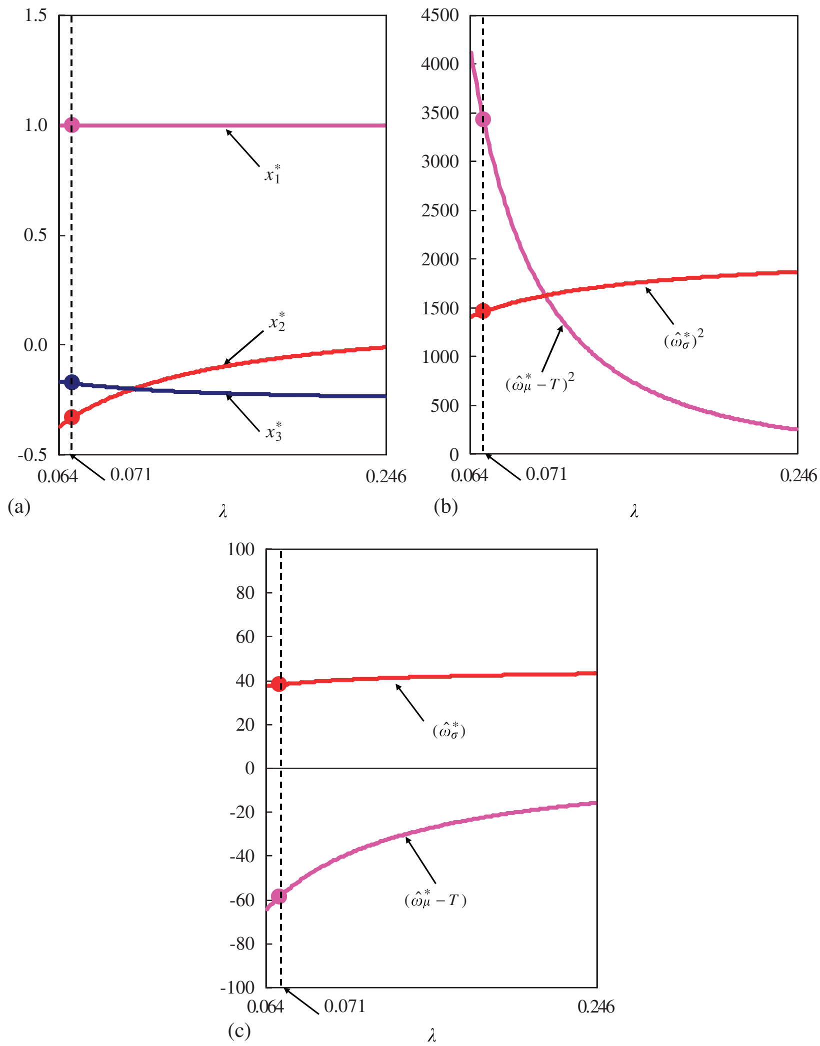

Figure 2. The optimal settings, the corresponding squared bias and variance values, and the unsquared bias and standard deviation values for $\lambda \in [0.064, 0.246]$: (a) the optimal settings versus λ; (b) the squared bias and variance values versus λ; and (c) the unsquared bias and standard deviation values versus λ. This figure is available in colour online at www.interscience.wiley.com/journal/qre

For detailed descriptions on the distribution of order statistics, see, for example, David and Nagaraja[16], Rohatgi[17], and Hogg *et al.*[18].

The right-hand side of $f_2(\overline{\lambda}|\underline{\lambda}, \lambda)$ in Equation (7) does not include $\underline{\lambda}$. The lower and upper bounds $\underline{\lambda}$ and $\overline{\lambda}$ are not independent in general because of the basic relationship between them, namely, $\overline{\lambda} > \underline{\lambda}$. However, once λ is given, the information on $\underline{\lambda}$ is essentially redundant in $f_2(\overline{\lambda}|\underline{\lambda}, \lambda)$ since $\overline{\lambda} \geq \lambda \geq \underline{\lambda}$. Thus, $\underline{\lambda}$ does not influence $\overline{\lambda}$ when λ is given. In a way, they are 'conditionally' independent.

Then, $p(\lambda|\underline{\lambda}, \overline{\lambda})$ is constructed as

$$p(\lambda|\underline{\lambda},\overline{\lambda}) = p(\lambda|\underline{\lambda},\overline{\lambda},n) = \frac{\left(\dfrac{n\underline{\lambda}^{n-1}}{\lambda^n}\right)\left\{\dfrac{n(1-\overline{\lambda})^{n-1}}{(1-\lambda)^n}\right\}}{\displaystyle\int_{\underline{\lambda}}^{\overline{\lambda}}\left(\dfrac{n\underline{\lambda}^{n-1}}{\lambda^n}\right)\left\{\dfrac{n(1-\overline{\lambda})^{n-1}}{(1-\lambda)^n}\right\}\mathrm{d}\lambda} = \frac{\dfrac{1}{\{\lambda(1-\lambda)\}^n}}{\displaystyle\int_{\underline{\lambda}}^{\overline{\lambda}}\dfrac{1}{\{\lambda(1-\lambda)\}^n}\mathrm{d}\lambda} = \frac{1}{k}\frac{1}{\{\lambda(1-\lambda)\}^n}, \quad \underline{\lambda} \leq \lambda \leq \overline{\lambda}$$

where

$$k = \int_{\underline{\lambda}}^{\overline{\lambda}} \frac{1}{\{\lambda(1-\lambda)\}^n}\mathrm{d}\lambda, \quad 0 < \underline{\lambda} \leq \lambda \leq \overline{\lambda} < 1$$

In the example, $\underline{\lambda}$ and $\overline{\lambda}$ were obtained as 0.064 and 0.246, respectively. They were made from the 12 lower and upper bounds of $\Lambda_{(i,j)}$'s given in Table III. Then, with $\underline{\lambda} = 0.064$, $\overline{\lambda} = 0.246$, and $n = 12$, $p(\lambda|\underline{\lambda} = 0.064, \overline{\lambda} = 0.246, n = 12)$ is obtained as

$$p(\lambda|\underline{\lambda} = 0.064, \overline{\lambda} = 0.246, n = 12) = \frac{1}{k}\frac{1}{\{\lambda(1-\lambda)\}^{12}}, \quad 0.064 \leq \lambda \leq 0.246$$

where

$$k=\int_{0.064}^{0.246}\frac{1}{\{\lambda(1-\lambda)\}^{12}}\,\mathrm{d}\lambda$$

Figure 3 plots $\pi(\lambda)$ and $p(\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12)$ versus λ. As shown in this figure, $p(\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12)$ drastically decreases as λ moves from 0.064 to 0.246. This form of $p(\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12)$ can be attributed to the assumed likelihood functions given in Equations (6) and (7). In $f_1(\underline{\lambda}|\lambda)$, the smaller the given λ value is, the larger the pdf value is. In $f_2(\overline{\lambda}|\underline{\lambda},\lambda)$, the larger the given λ value is, the larger the pdf value is. Such properties of $f_1(\underline{\lambda}|\lambda)$ and $f_2(\overline{\lambda}|\underline{\lambda},\lambda)$ collectively make the unique form of $p(\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12)$.

The results discussed above are based upon the assumption that $\underline{\lambda}_1,\underline{\lambda}_2,\ldots,\underline{\lambda}_n$ and $\overline{\lambda}_1,\overline{\lambda}_2,\ldots,\overline{\lambda}_n$ follow a uniform distribution. It should be noted that other distributions can be used as appropriate. A beta distribution would be a good alternative in the sense that it can produce various distribution shapes by changing its parameters. However, in the case of a beta distribution, the posterior pdf, $p(\lambda|\underline{\lambda},\overline{\lambda})$, is not obtained in a closed form, but in a numerical manner.

4.2. *Expected value of λ*

The expected value of λ given $\underline{\lambda}=0.064$, $\overline{\lambda}=0.246$, and $n=12$ is obtained as

$$E[\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12]=\int_{0.064}^{0.246}\lambda\cdot p(\lambda|\underline{\lambda}=0.064,\overline{\lambda}=0.246,n=12)\,\mathrm{d}\lambda=\int_{0.064}^{0.246}\lambda\cdot\left(\frac{1}{k}\right)\left[\frac{1}{\{\lambda(1-\lambda)\}^{12}}\right]\mathrm{d}\lambda=0.071$$

Using the obtained expected value of λ, the WMSE criterion can be posed as the following optimization problem:

$$\begin{aligned}&\underset{\mathbf{x}}{\text{minimize}} && \text{WMSE}=(0.071)(\hat{\omega}_\mu-T)^2+(1-0.071)\hat{\omega}_\sigma^2 && \text{(8-1)}\\ &\text{subject to} && \hat{\omega}_\sigma\geq 0 && \text{(8-2)}\\ & && -1\leq x_i\leq 1,\quad i=1,2,3 && \text{(8-3)}\end{aligned}\qquad(8)$$

It should be noted that Equation (8-2) represents a constraint that the standard deviation response should be non-negative. Expressing $\log\hat{\omega}_\sigma$ as a linear model instead of $\hat{\omega}_\sigma$ in Equation (1) can also ensure the positiveness of $\hat{\omega}_\sigma$. This standard log-linear model has been used in Chiao and Hamada[19]. The optimal setting resulting from Equation (8) is $(x_1^*,x_2^*,x_3^*)=(1.000,-0.331,-0.175)$. The corresponding squared bias (unsquared bias) and variance (standard deviation) values are

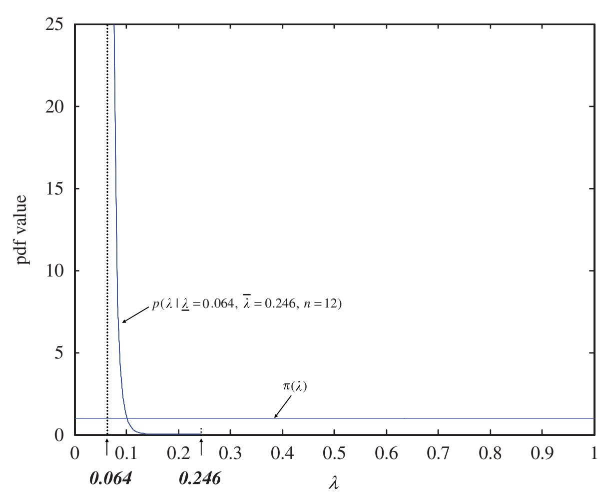

Figure 3. The uniform prior pdf and the corresponding posterior pdf. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Table IV. Comparison with other λ values

Method	λ	$(x_1^*,\ x_2^*,\ x_3^*)$	$(\hat{\omega}_\mu^*-T)^2$	$(\hat{\omega}_\mu^*-T)$	$(\hat{\omega}_\sigma^*)^2$	$(\hat{\omega}_\sigma^*)$
Lin and Tu[5]	0.5	(1.000, 0.074, −0.252)	28.20	−5.31	1976.87	44.46
Ding *et al.*[11]	0.6	(1.000, 0.089, −0.255)	12.67	−3.56	1995.50	44.67
Proposed scheme	0.071	(1.000, −0.331, −0.175)	3424.59	−58.52	1457.10	38.17

$(\hat{\omega}^*_\mu - T)^2 = 3424.59$ $((\hat{\omega}^*_\mu - T) = -58.52)$ and $(\hat{\omega}^*_\sigma)^2 = 1457.10$ $(\hat{\omega}^*_\sigma = 38.17)$, respectively. If an arbitrary value of λ between 0.064 and 0.246 were used to form WMSE, the optimal setting would be different as manifested by the changes in x^*_2 and x^*_3 as shown in Figure 2. Consequently, $(\hat{\omega}^*_\mu - T)^2$ could have changed from 4117.79 to 248.38 (from -64.17 to -15.76 for $(\hat{\omega}^*_\mu - T)$) and $(\hat{\omega}^*_\sigma)^2$ from 1407.00 to 1867.71 (from 37.51 to 43.22 for $\hat{\omega}^*_\sigma$).

For comparison purposes, we solved the optimization model of Equation (8) for $\lambda = 0.5$ and $\lambda = 0.6$. These λ values are the ones that were obtained in Lin and Tu[5] and Ding *et al.*[12], respectively. The results are summarized in Table IV, along with that from $\lambda = 0.071$. As expected, the optimal setting, and the corresponding squared bias (unsquared bias) and variance (standard deviation) values show important differences depending upon the value of λ. The value of 0.071 is considered as a meaningful choice in the sense that it is the consequence of utilizing all the relevant information available, namely, the prior knowledge on λ and the interval of λ congruent with the DM's preference structure. Note that, when $\lambda = 0.071$, the variance (standard deviation) value is significantly reduced, although the squared bias (unsquared bias) value is much sacrificed. In the next section, a sensitivity analysis will be performed for various prior distributions.

5. Sensitivity analysis

In the previous section, the uniform pdf was used as $\pi(\lambda)$. In this section, we investigate how sensitively $p(\lambda|\underline{\lambda}, \overline{\lambda})$ changes when $\pi(\lambda)$ varies. We employ a beta pdf as $\pi(\lambda)$ with parameters $\alpha > 0$ and $\beta > 0$. The beta pdf of λ is defined as

$$\pi(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1} (1-\lambda)^{\beta-1}, \quad 0 \le \lambda \le 1 \tag{9}$$

where $\Gamma(\alpha)$ is the gamma function. The beta distribution has a random variable space $[0, 1]$, which coincides with the range of λ. Thus, the beta distribution is suitable for the prior probability distribution on λ. The likelihood functions and the sample information are the same as those in the Illustrative Example section.

5.1. Construction of the posterior pdfs

To consider a variety of beta pdfs, we employ 10 combinations of α and β, which represent five and two different values for μ and σ, respectively. The values of μ are 0.03, 0.064, 0.155, 0.246, and 0.5. The value 0.03 is about half of the lower bound, 0.064 the lower bound itself, 0.155 the center between the lower and upper bounds, 0.246 the upper bound itself, and 0.5 about twice larger than the upper bound. The values of σ are 0.1 and 0.02, which represent relatively high and low variability levels, respectively. The 10 beta pdfs are summarized in Table V.

Combining $\pi(\lambda)$, $f_1(\underline{\lambda}|\lambda)$, and $f_2(\overline{\lambda}|\underline{\lambda}, \lambda)$, $p(\lambda|\underline{\lambda}, \overline{\lambda})$ is constructed as

$$p(\lambda|\underline{\lambda}, \overline{\lambda}) = p(\lambda|\underline{\lambda}, \overline{\lambda}, n)$$
$$= \frac{\left\{ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1}(1-\lambda)^{\beta-1} \right\} \left(\frac{n\underline{\lambda}^{n-1}}{\lambda^n} \right) \left\{ \frac{n(1-\overline{\lambda})^{n-1}}{(1-\lambda)^n} \right\}}{\int_{\underline{\lambda}}^{\overline{\lambda}} \left\{ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1}(1-\lambda)^{\beta-1} \right\} \left(\frac{n\underline{\lambda}^{n-1}}{\lambda^n} \right) \left\{ \frac{n(1-\overline{\lambda})^{n-1}}{(1-\lambda)^n} \right\} d\lambda}$$

Table V. *$E[\lambda|\underline{\lambda} = 0.064,\ \overline{\lambda} = 0.246,\ n = 12]$'s and $E[L]$'s from the 10 beta prior pdfs*

	$\pi(\lambda)$				
	$(\alpha,\ \beta)$	$\Leftrightarrow$	$(\mu,\ \sigma)^*$	$E[\lambda\|\underline{\lambda},\ \overline{\lambda},\ n]$	$E[L]$
Case 1	(0.057, 1.853)		(0.03, 0.1)	0.072	0.083
Case 2	(0.319, 4.671)		(0.064, 0.1)	0.071	0.084
Case 3	(1.875, 10.22)		(0.155, 0.1)	0.071	0.084
Case 4	(4.317, 13.23)		(0.246, 0.1)	0.074	0.082
Case 5	(12.00, 12.00)		(0.5, 0.1)	0.138	0.047
Case 6	(2.153, 69.60)		(0.03, 0.02)	0.069	0.086
Case 7	(9.521, 139.2)		(0.064, 0.02)	0.069	0.086
Case 8	(50.60, 275.8)		(0.155, 0.02)	0.128	0.050
Case 9	(113.8, 348.9)		(0.246, 0.02)	0.224	0.072
Case 10	(312.0, 312.0)		(0.5, 0.02)	0.246	0.091

$^*\mu = \alpha/(\alpha+\beta)$ and $\sigma^2 = \alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$.

$$= \frac{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} n^2\{\underline{\lambda}(1-\bar{\lambda})\}^{n-1} \dfrac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{\{\lambda(1-\lambda)\}^n}}{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} n^2\{\underline{\lambda}(1-\bar{\lambda})\}^{n-1} \displaystyle\int_{\underline{\lambda}}^{\bar{\lambda}} \dfrac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{\{\lambda(1-\lambda)\}^n}\,\mathrm{d}\lambda}$$

$$= \frac{1}{k}\frac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{\{\lambda(1-\lambda)\}^n}, \quad \underline{\lambda}\le\lambda\le\bar{\lambda}$$

where

$$k=\int_{\underline{\lambda}}^{\bar{\lambda}} \frac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{\{\lambda(1-\lambda)\}^n}\,\mathrm{d}\lambda, \quad 0<\underline{\lambda}\le\lambda\le\bar{\lambda}<1$$

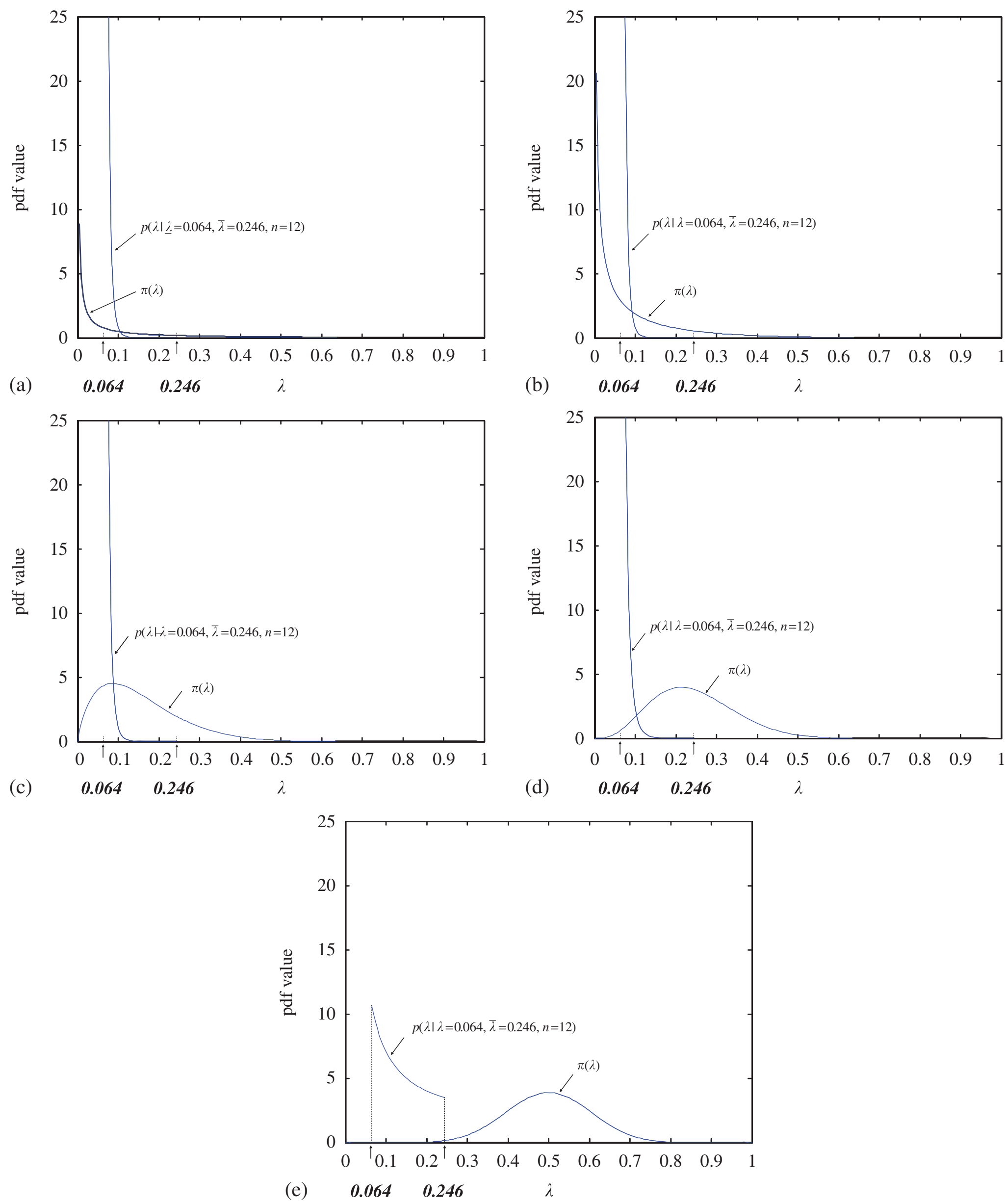

Figure 4. The beta prior pdfs and the corresponding posterior pdfs: (a) Case 1: $\alpha=0.057$, $\beta=1.853$ ($\mu=0.03, \sigma=0.1$); (b) Case 2: $\alpha=0.319$, $\beta=4.671$ ($\mu=0.064, \sigma=0.1$); (c) Case 3: $\alpha=1.875$, $\beta=10.22$ ($\mu=0.155, \sigma=0.1$); (d) Case 4: $\alpha=4.317$, $\beta=13.23$ ($\mu=0.246, \sigma=0.1$); (e) Case 5: $\alpha=12.00$, $\beta=12.00$ ($\mu=0.5, \sigma=0.1$); (f) Case 6: $\alpha=2.153$, $\beta=69.60$ ($\mu=0.03, \sigma=0.02$); (g) Case 7: $\alpha=9.521$, $\beta=139.2$ ($\mu=0.064, \sigma=0.02$); (h) Case 8: $\alpha=50.60$, $\beta=275.8$ ($\mu=0.155, \sigma=0.02$); (i) Case 9: $\alpha=113.8$, $\beta=348.9$ ($\mu=0.246, \sigma=0.02$); and (j) Case 10: $\alpha=312.0$, $\beta=312.0$ ($\mu=0.5, \sigma=0.02$). This figure is available in colour online at www.interscience.wiley.com/journal/qre

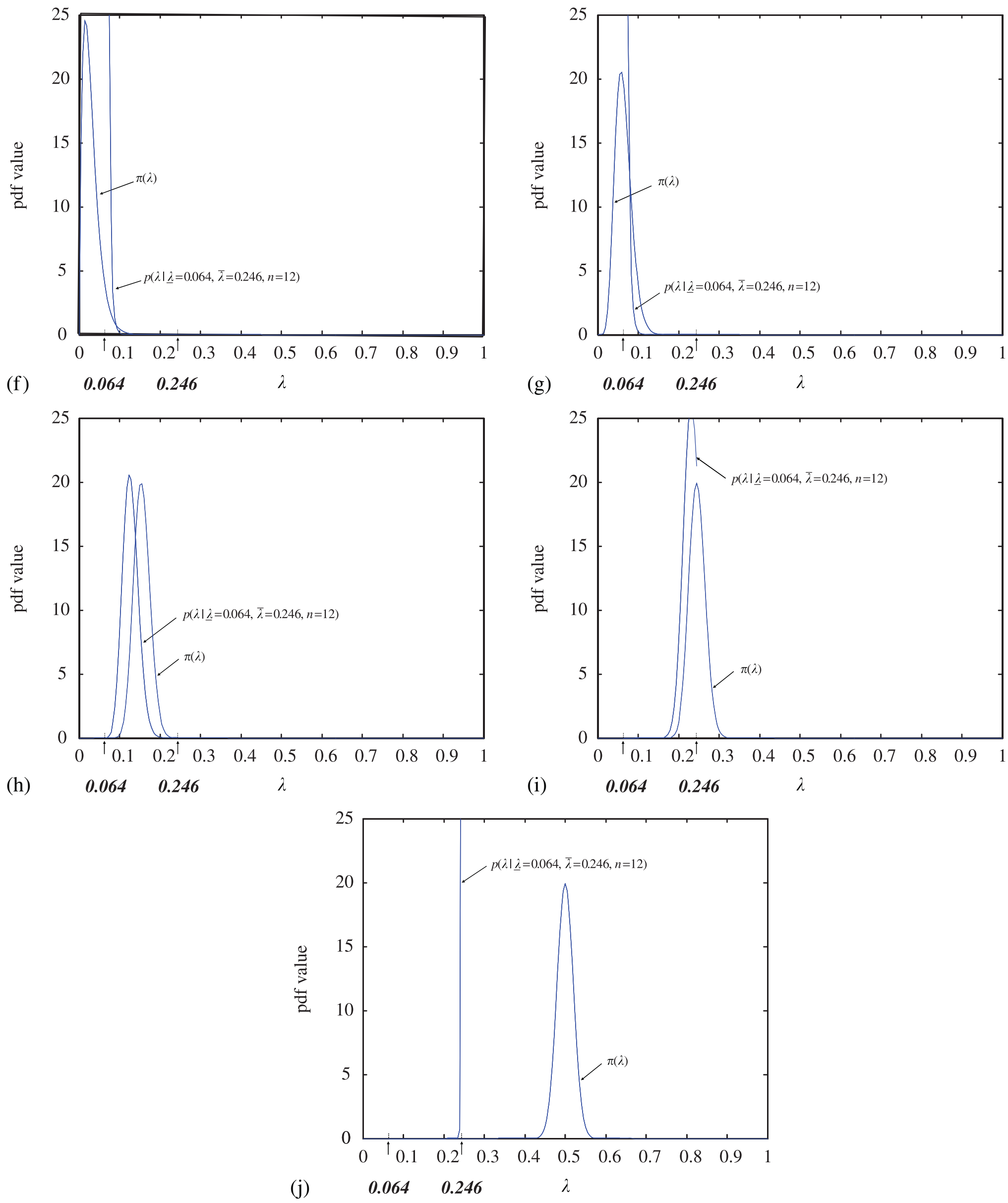

Figure 4. *Continued*

Then, with $\underline{\lambda}=0.064$, $\overline{\lambda}=0.246$, and $n=12$, $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$ is obtained. Figure 4 plots $\pi(\lambda)$ and $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$ versus λ for the 10 cases. Figure 5 shows the superimposed $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$'s for the 10 cases.

As shown in Figures 4 and 5, when the variability of $\pi(\lambda)$ is relatively high, the mean of $\pi(\lambda)$ has little effect on $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$. When the variability is relatively low, on the other hand, the mean seems to have a significant effect. When $\sigma=0.1$ (Figures 4(a) to 4(e)), the form of $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$ stays unchanged regardless of μ except for Case 5. Moreover, this form is almost the same as that constructed from the noninformative prior pdf in Figure 3. However, when $\sigma=0.02$ (Figures 4(f) to 4(j)), the form of $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$ changes sensitively with respect to μ. The pdf value of $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$ increases with λ when μ is 0.5 (Case 10), while it decreases when μ is 0.03 and 0.064 (Cases 6 and 7). As expected, $\pi(\lambda)$ with a lower variability has a stronger impact on $p(\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12)$.

Here, as the reviewer suggested, some additional advice on the appropriate choice of $\pi(\lambda)$ would be helpful. When there is no prior knowledge about λ, any value of λ in [0, 1] is equally desirable. In this case, the two parameters of $\pi(\lambda)$ in Equation (9) should be set at $\alpha=\beta=1$ (Type I), with which the beta distribution reduces to the uniform distribution on [0, 1]. If both weight values are known to be equal *a priori*, the value of λ is desired to be set at 0.5 and $\alpha=\beta>1$ (Type II). The specific value of α (or β) depends on the variability of λ. As the variability gets lower, the value of α (or β) gets larger. If the weight of the squared bias is greater than that of the variance, the value of λ is set at the value over 0.5 and $\alpha>\beta$ (Type III). In the opposite case, it is set at

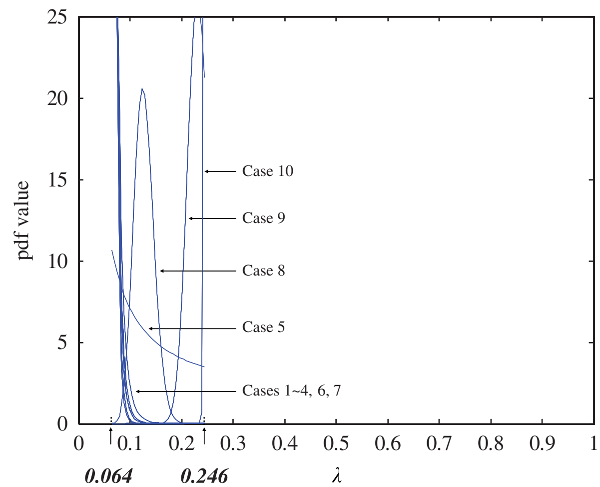

Figure 5. The posterior pdfs from the 10 beta prior pdfs. This figure is available in colour online at www.interscience.wiley.com/journal/qre

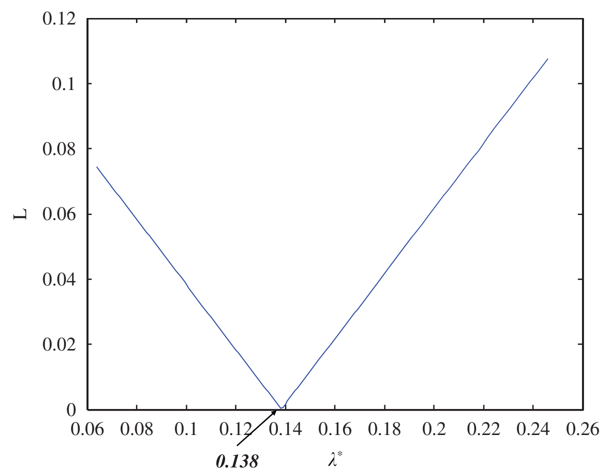

Figure 6. The pattern of L's for various values of λ^* within [0.064, 0.246] and $E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]=0.138$ (Case 5). This figure is available in colour online at www.interscience.wiley.com/journal/qre

the value under 0.5 and $\alpha<\beta$ (Type IV). For both cases, the specific values of α and β depend on the (conjectured) distribution shape of λ.

Among the 11 prior pdfs used in Sections 4 and 5, the uniform pdf corresponds to Type I (Figure 3), Cases 5 and 10 of the beta pdfs to Type II (Figures 4(e) and 4(j)), and Cases 1–4 and 6–9 to Type IV (Figures 4(a)–4(d) and Figures 4(f)–4(i)). In essence, the goodness of the choice of $\pi(\lambda)$ should be evaluated by how accurately the chosen $\pi(\lambda)$ displays the prior knowledge about λ (Iversen[20]).

5.2. The goodness of the expected value of λ

Using the obtained posterior pdfs, $E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]$'s are computed. They are summarized in Table V. Then, we check how good the obtained $E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]$ is as an estimate of the true λ (say, λ^*). The value of λ^* is unknown, but it would be reasonable to assume that λ^* exists within [0.064, 0.246]. In this work, the 'absolute error loss' is employed as a measure of the goodness (Berger[14], pp. 62–63). The absolute error loss (L) is defined as

$$L=|\lambda^*-E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]|$$

As an example, if $\lambda^*=0.155$ and $E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]=0.138$, which is the expected value of λ in Case 5, then $L=0.017$ $(=|0.155-0.138|)$. However, λ^* is unknown, and assumed to lie in [0.064, 0.246]. Thus, L can also change with respect to λ^*. Figure 6 plots the pattern of L's for various values of λ^* within [0.064, 0.246]. Here, suppose λ^* follows a uniform distribution on [0.064, 0.246]. Then, the 'expected absolute error loss' can be obtained as

$$E[L]=\int_{0.064}^{0.246}|\lambda^*-E[\lambda|\underline{\lambda}=0.064, \bar{\lambda}=0.246, n=12]|\cdot\left(\frac{1}{0.246-0.064}\right)\mathrm{d}\lambda^*=\int_{0.064}^{0.246}|\lambda^*-0.138|\cdot\left(\frac{1}{0.246-0.064}\right)\mathrm{d}\lambda^*=0.047$$

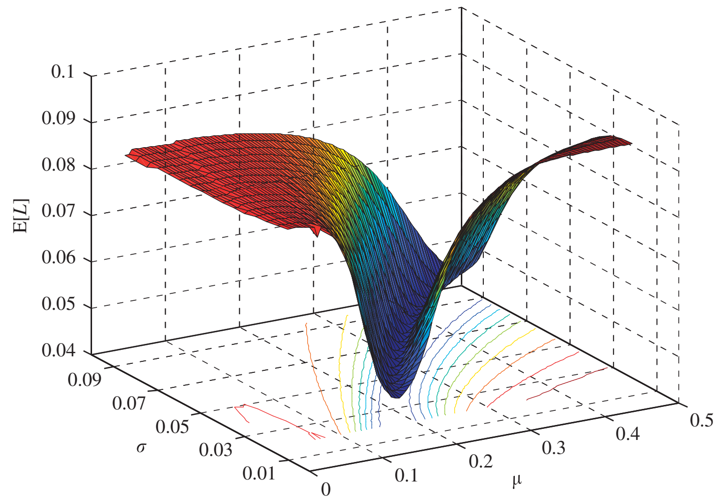

Figure 7. A three-dimensional plot of the $E[L]$, μ, and σ. This figure is available in colour online at www.interscience.wiley.com/journal/qre

For $E[\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12]$'s in Table V, $E[L]$'s are obtained under the uniform distribution assumption on λ^*, and shown in the rightmost column in Table V. In this table, $E[L]$ decreases as $E[\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12]$ gets close to 0.138. Specifically, $E[L]$ is minimized when $E[\lambda|\underline{\lambda}=0.064, \overline{\lambda}=0.246, n=12]=0.155$. Figure 7 shows a three-dimensional plot of the $E[L]$, μ, and σ. This figure shows $E[L]$'s obtained from $\pi(\lambda)$'s with various pairs of μ and σ in $[0.03, 0.5]$ and $[0.02, 0.1]$, respectively, including the original 10 cases in Table V.

6. Conclusion

Dual response surface optimization considers the mean and the variation simultaneously. The minimization of MSE is an effective approach in dual response surface optimization. WMSE is formed by imposing the relative weights, $(\lambda, 1-\lambda)$, on the squared bias and variance terms of MSE. To date, quite a few methods have been proposed for determining λ. The resulting λ from these methods is either a single value or an interval. In the interval case, a value of λ in the interval needs to be chosen to form WMSE.

This paper has developed a systematic method to choose a λ value when an interval of λ is given. Specifically, this paper has proposed a Bayesian scheme to construct a distribution of λ, from which a meaningful choice of λ can be made. The lower and upper bounds of the λ interval are employed as the sample information in the Bayesian analysis. Once the posterior distribution of λ is constructed, the expected value of λ can then be used to form WMSE. The Bayesian approach has an advantage in that it can utilize the prior knowledge to construct the distribution of λ. In addition, it can update the distribution of λ with the knowledge obtained through the sample information. The more the knowledge accumulates, the more the degree of belief in the value of λ increases.

The proposed Bayesian scheme has been illustrated through an example with a noninformative prior distribution. Furthermore, a sensitivity analysis has been conducted for a variety of beta prior distributions.

References

1. Box GEP, Draper NR. *Empirical Model Building and Response Surfaces*. Wiley: New York, 1987.
2. Myers RH, Montgomery DC. *Response Surface Methodology* (2nd edn). Wiley: New York, 2002.
3. Vining GG, Myers RH. Combining Taguchi and response surface philosophies a dual response approach. *Journal of Quality Technology* 1990; **22**(1):38–45.
4. Del Castillo E, Montgomery DC. A nonlinear programming solution to the dual response problem. *Journal of Quality Technology* 1993; **25**(3):199–204.
5. Lin D, Tu W. Dual response surface optimization. *Journal of Quality Technology* 1995; **27**(1):34–39.
6. Copeland KA, Nelson PR. Dual response optimization via direct function minimization. *Journal of Quality Technology* 1996; **28**(3):331–336.
7. Del Castillo E, Fan SK, Semple J. The computation of global optima in dual response systems. *Journal of Quality Technology* 1997; **29**(3):347–353.
8. Kim K, Lin D. Dual response surface optimization: A fuzzy modeling approach. *Journal of Quality Technology* 1998; **30**(1):1–10.
9. Tang LC, Xu K. A unified approach for dual response surface optimization. *Journal of Quality Technology* 2002; **34**(4):437–447.
10. Köksoy O, Doganaksoy N. Joint optimization of mean and standard deviation using response surface methods. *Journal of Quality Technology* 2003; **35**(3):239–252.
11. Kim K, Jun C, Cho H, Hong Y, Kim Y, Kim C, Youm S, Lee S, Lee J, Chang B, Jeong B, Jeong I. Development of a six-sigma improvement model: Reducing the variation of tin plates' hardness. *Technical Report*, POSCO, Republic of Korea, 2000.
12. Ding R, Lin D, Wei D. Dual response surface optimization: A weighted MSE approach. *Quality Engineering* 2004; **16**(3):377–385. DOI: 10.1081/QEN-120027940.

13. Jeong I, Kim K, Chang S. Optimal weighting of bias and variance in dual response surface optimization. *Journal of Quality Technology* 2005; **37**(3):236–247.
14. Berger JO. *Statistical Decision Theory and Bayesian Analysis* (2nd edn). Springer: New York, 1985.
15. Raghunathan TE. Bayesian analysis of quality level using simulation methods. *Journal of Quality Technology* 2000; **32**(2):172–182.
16. David HA, Nagaraja HN. *Order Statistics* (3rd edn). Wiley: New York, 2003.
17. Rohatgi VK. *An Introduction to Probability Theory and Mathematical Statistics*. Wiley-Interscience: New York, 1976.
18. Hogg R, Craig A, McKean J. *Introduction to Mathematical Statistics* (6th edn). Prentice-Hall: Upper Saddle River, 2005.
19. Chiao C, Hamada M. Analyzing experiments with correlated multiple responses. *Journal of Quality Technology* 2001; **33**(4):451–465.
20. Iversen GR. *Bayesian Statistical Inference*. SAGE Publications: Newbury Park, 1984.

Authors' biographies

In-Jun Jeong is a Senior Researcher in the Technology Business Division at the Electronics and Telecommunications Research Institute, Republic of Korea. He earned his BSc, MSc, and PhD in Industrial Engineering from Pohang University of Science and Technology, Republic of Korea. His research interests include quality engineering and management, technology transfer and commercialization, and technology evaluation. He has published research articles in various journals including *Journal of Quality Technology*, *Quality and Reliability Engineering International*, *European Journal of Operational Research*, *Computers and Operations Research*, *Expert Systems with Applications*, *International Journal of Industrial Engineering*, *International Journal of Reliability*, *Quality and Safety Engineering*, and the like.

Kwang-Jae Kim is a Professor in the Department of Industrial and Management Engineering at Pohang University of Science and Technology, Korea. He earned his BSc in Industrial Engineering in 1984 from Seoul National University, Korea, his MSc in Industrial Engineering in 1986 from Korea Advanced Institute of Science and Technology, Korea, and his PhD in Management Science in 1993 from Purdue University. His research interests include quality assurance in product and process design, new product/service development, and service engineering. He has published over 50 research papers in various refereed journals. He is a member of ASQ, IIE, and INFORMS.

Dr Dennis K. J. Lin is a University Distinguished Professor of Supply Chain and Statistics at Penn State University. His research interests include quality engineering, industrial statistics, data mining and response surface. He has published over 150 papers in a wide variety of journals. He currently serves as co-editor for *Applied Stochastic Models for Business and Industry*, and associate editor for various journals: *Technometrics*, *Statistica Sinica*, *Journal of Quality Technology*, *Journal of Data Science*, *Quality Technology & Quality Management*, *Journal of Statistics and Its Applications*, and *Journal of Statistical Theory and Practice*. Dr Lin is an elected fellow of ASA and ASQ, an elected member of ISI, a lifetime member of ICSA, a fellow of RSS, and a Chang-Jiang Scholar of China at Renmin University. He is also the recipient of the 2004 Faculty Scholar Medal Award at Penn State University.