Backfitting Estimation of a Response Surface Model

Jhoanne Marsh C. Gatpatan
University of the Philippines in the Visayas
cyann_mars@yahoo.com

The backfitting algorithm is used in estimating a response surface model with covariates from a data generated through a central composite design. Backfitting takes advantage of the orthogonality generated by the central composite design on the design matrix. The simulation study shows that backfitting yield estimates and predictive ability of the model comparable to that from ordinary least squares when the response surface bias is minimal. Ordinary least squares estimates generally fails when the response surface bias is large, while backfitting exhibits robustness and still produces reasonable estimates and predictive ability of the fitted model. Orthogonality facilitates the viability of assumptions in an additive model where backfitting is an optimal estimation algorithm.

Keywords: *backfitting, response surface model, second order model, central composite design*

1. Introduction

Response Surface Methodology (RSM) is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes (Myers and Montgomery 2002). Response surface designs are powerful class of experimental designs, possibly identifying a closer link between empirical and theoretical response surfaces than the traditional experimental designs (Inouye 2001). The applications of RSM have been very helpful and extensively used in the industry (e.g., food manufacturing, drug manufacturing, etc.) where the relationship between the response and the design factors are approximated by some low-order polynomials. Aslan (2008) used RSM in a central composite design (CCD) for modeling and optimization of influence of operating variables on performance of multigravity separator for chromite concentration. Similarly, Li et al. (2008) used RSM for the design optimization of integrated circuits. On the other hand, Hsu (1995) also used RSM to study the effect of processing variables on the quality of frozen surimi. RSM is useful not only in product design but also in systems modeling in biology and ecology. Inouye (2001) used simulated data in comparing the ability of six different response surface designs to estimate correctly the competition coefficients of simple competition models under a variety of conditions. There are also applications in the physical sciences, biological sciences, as well as social sciences (see for example Khuri and Cornell 1987).

As an empirical strategy, RSM aims to choose designs that are simple and results are easily interpretable. The optimal design that is desirable in RSM is usually evaluated based on two very important characteristics: orthogonality and rotatability. Orthogonal designs result to uncorrelated terms in the model, hence parameter estimates are uncorrelated as well (Khuri and Cornell 1987). A first-order orthogonal design is ideal in least squares estimation since the design matrix X'X is diagonal and hence, interpretation of the regression coefficients is simplified because the effects are additive and not conditional on other terms in the model. According to Myers and Montgomery (2002), the orthogonality of the design matrix X implies that the roles of two variables can be assessed independently of each other. Rotatable designs, on the other hand, yield the same prediction variance $V(\hat{y}(x))$ at all points x that are equidistant from the center point. The desirable feature is that quality of prediction is invariant to any rotation of the coordinate axes in the space of predictor variables (Khuri and Cornell 1987). As noted by Myers and Montgomery (2002), a rotatable design will not require knowledge of the exact location of the optimum point (for as long as it is within the design space) before the experiment is conducted.The CCD results to design points that possess both the properties of orthogonality and rotatability.

In recent modeling paradigms, the additive model was introduced where the response variable is expressed as a sum of functions of predictors. The form of the function is not necessarily specified, but the usual linear smoother may also be considered. Since each component of the model has an additive effect to the response, it facilitates the piecewise estimation of each component. The effect of one component is estimated by regressing the response on the particular variable, the next variable is regressed on the residuals from the first, and so on, until all components are accounted. This estimation method called backfitting yields many computational advantages (see Hastie and Tibshirani 1990 for details). Given an additive model, the backfitting algorithm examines the effect of each variable/component individually, hence unlike the maximum likelihood estimation, it is not flooded with parameters being estimated thereby affecting the speed of convergence. In the case of least squares estimation in every stage, the design matrix will have a "better configuration" since it is already low-dimensional and this avoids the ill-conditioning problem.

This ensures the stability of the resulting least squares estimators in the backfitting algorithm.

In RSM, there are generally very few design points that are used to generate responses since cost of experimentation is often prohibitive. With very few points, the configuration of the design matrix can be prone to ill-conditioning, hence stability of the parameter estimates might be threatened. Since a CCD generates orthogonal points, this can easily satisfy the additivity of the components in an additive model, hence backfitting can potentially mediate in the balance between cost (fewer design points) and stability of parameter estimates.

In some response surface experiments, multicollinearity can be present among the design factors (especially since there are only very few design points) and this can have serious effects on the stability of estimated parameters and on the usefulness of the estimated model if ordinary least squares is used in fitting the model (see Carley et al. 2004 for detailed discussion). Multicollinearity is also easily observed in experiments with high dimensions, i.e., too many design factors. In the presence of multicollinearity, the fitted model may still be useful despite the imprecision of the estimated regression coefficients, but predictions of new observations requiring extrapolation beyond the experimental region can yield poor results.

This paper proposes to use backfitting in estimating a response surface model with covariates in a CCD. Through a simulation study, the accuracy of the parameter estimates will be assessed along with the predictive ability of the estimated response surface.

We would like to provide an alternative estimation procedure to ordinary least squares when the postulated model does not fit the data well, e.g., when there is too much bias in the approximate response surface function. Backfitting can 'optimize' the extraction of signals from the response and hence, minimizes the remaining noise that is unaccounted for. The method can also handle contamination and possibly misspecification that may occur during experimentation and during the response surface estimation. Robustness of the method is feasible because of the orthogonality of the design matrix generated from the Central Composite Design.

2. Alternative Estimation of the Response Surface

As Khuri and Cornell (1987) defined it, a central composite design consists of: (1) a complete (or fraction of a) 2^k factorial design, where the factor levels are coded -1, +1 values. This is called the factorial portion of the design; (2) n_0 center points ($n_0 \geq 1$); (3) two axial points on the axis of each design variable at a distance of α from the design center. This is called the axial portion of the design. Myers and Montgomery (2002) gave the roles of these three components of the design. A variance-optimal design for a first order model or first-order with two factor interaction model is what the factorial points represent. On the other hand, information about the existence of a curvature in the system is what the center point runs provide. Moreover, axial points contribute to the efficient estimation of the pure quadratic terms if a curvature exists in the system.

Among the designs commonly used in response surface modeling, CCD is popular since it implies two desirable properties: othogonality and rotatability. Orthogonality facilitates the attainment of many optimal characteristics of least squares method. Draper and Smith (1998) discusses the implications of orthogonality of columns of X. Suppose the data matrix is partitioned into $X=(X_1,X_2)$ and if $X_1X_2=0$, then in the model $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, $\hat{\beta}_1 = (X_1{}'X_1)^{-1}X_1Y$, and $\hat{\beta}_2 = (X_2{}'X_2)^{-1}X_2Y$ and $SS(\hat{\beta}) = SS(\hat{\beta}_1) + SS(\hat{\beta}_2)$. This implies that the contribution of X_1 and X_2 in explaining Y may be separately assessed without affecting the other block of independent variables. The model is necessarily additive.

Benefitting from the orthogonality of the design matrix from the CCD, the response surface model is additive. Thus, the backfitting algorithm that performs best in an additive model can simplify the estimation of the parameters of the response surface model. Opsomer (2000) shows that if covariates are independent, then the asymptotic bias of the linear smoother obtained through a backfitting algorithm becomes very small. This implies that if for instance, the response surface is approximated by the second order-model, the order bias of the approximation is minimal when the model is estimated through linear smoothing in a backfitting algorithm. Butler (2006) further noted the minimum response surface bias in CCD for any number of center points.

Given a data from a CCD, the following backfitting algorithm in a linear smoother is proposed to estimate the response surface model with covariates.

Step 1: Postulate an additive model of the form

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k + \delta_1Z_1 + \delta_2Z_2 + \ldots + \delta_jZ_j + \varepsilon \quad (1)$$

or a second-order model of the form

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{11}X_1^2 + \ldots + \beta_{kk}X_k^2 + \beta_{12}X_{12} + \ldots + \beta_{kj}X_{kj} \quad (2)$$
$$+ \delta_1Z_1 + \delta_2Z_2 + .. + \delta_jZ_j + \varepsilon$$

Step 2: Estimate a regression model with the first factor X_1 using ordinary least squares. Compute the residual $e_1 = y - \hat{\beta}_0 - \hat{\beta}_1x_1$ from the estimated regression coefficients.

Step 3: Regress the residual e_1 with X_2. Compute the remaining residual left by taking out the effect of X_2. Continue iteration until the effect of the last factor is accounted and the remaining residuals can be attributed to the covariates.

Step 4: The covariate with the highest correlation with the response variable will be regressed with the residual left in Step 3. Continue with the estimation of the covariate effect until all covariates are included in the model, the last set of residuals are attributed to pure error.

Aside from taking advantage of the additivity of the response surface model, the backfitting algorithm will also facilitate estimation of the model when there are many factors and covariates involved since CCD is expected to require few design points only.

3. Simulation Study

A simulation study is presented to illustrate the relative advantage of the proposed estimation procedure for a response surface model. The data generation process is expected to mimic a response surface setting that enables fitting of the first- or second-order models.

The 'true' response functions assumed are given by the following:

Model 1: $y = 1 - 0.1x_1 + 0.01x_2 - 0.01z_1 - 0.01z_2 + k\varepsilon$ (3)

Model 2: $y = 1 - 0.1x_1 + 0.001x_2 - 0.01z_1 - 0.01z_2 + 0.001w + k\varepsilon$ (4)

where x_1 and x_2 are the factors, z_1 and z_2 are covariates, w is another variable that also affects the response variable but will be assumed later as not accessible to the experimenter, k is a constant to be used in inducing contamination that controls the correlation of the factors with the response variable, and ε is the error term.

These models can generate scenarios of ideal response surface modeling situations as well as those with misspecifications. We simulated two types of misspecifications. Data coming from model (3) illustrates misspecification in terms of assumption on errors while data from model (4) will show the misspecification through variable omission when w is left out during the estimation and pretend that it is part of the error. The error terms are used in simulating the misspecification reflected as violation of the usual regression assumption, e.g., normal distribution.

For the 'true' model, two error term distributions (normal and poisson) are considered. The normal distribution cases are N(0,1), N(0,100), and N(100,100), while the poisson distributions are Po(10), and Po(100). The normal distribution for simulated error terms is important because in ordinary least squares regression, this will ensure the validity of the Gauss-Markov theorem. The case N(0,1) illustrates most ideal scenario, N(0,100) for highly heteroskedastic errors, possibly due to outlying random shocks. The last distribution of error terms represents the case where there is misspecification in the assumptions related to the error distribution. Po(10) represents a skewed discrete error with a smaller mean and variance while Po(100) illustrates a scenario of error terms coming from a skewed population with larger mean and variance. When the response variable is count, a skewed distribution may best characterize the behavior of the error term. The choice of the distribution of error terms having varying variances and means will also illustrate the presence of contamination in the experimentation. In controlled experiments, responses and experimental errors are expected to exhibit smaller variances within the same setting. Variances become larger when certain experimental conditions are not properly controlled. Scenarios of different levels of contamination constant *k* (high, moderate, and low) are also included in the simulation by multiplying it to the error terms. The contamination constant *k* controls the correlation of the response variable and the factors. Lower values of *k* mean higher correlation between the response and the predictor variables. Low values of *k* illustrate the case that data coming from models (3) and (4) can potentially lead to good model fit while higher levels of *k*, can lead to either moderate to poor model fit. Higher coefficient of determination (R^2) implies that ordinary least squares (OLS) is optimal by the virtue of Gauss–Markov Theorem under some regularity conditions on the error distribution. Conversely, low R^2 means poor model fit, perhaps due to misspecification, or inappropriate model, which leads to failure of OLS estimates. In RSM which is associated with high data generation cost, a more robust procedure is needed.

In all the scenarios, two factors, X_1 and X_2, are simulated based on the pre-specified range of values $6<X_1<11$ and $100<X_2<300$. These values are chosen to induce variation into the predictor variables to facilitate estimation of the response surface. The first variable has a smaller range compared to the second to imitate certain characteristics of factors that are commonly used. The choice of having a shorter range of values for X_1 illustrates the instance where short range of values for one factor easily causes multicollinearity to manifest.

There are four possible combinations of distributions from which these factors are simulated from: two continuous probability distributions, two discrete probability distributions, a combination of a discrete and a continuous probability distribution, i.e., 1) N(9,1) , N(200, 961) ; 2) Po(8), Po(200); 3) N(9,1), Po(200); and 4) Po(8), N(200,961). Normal distribution is considered for the continuous case to illustrate the scenario that the factor comes from a symmetric continuous distribution. On the other hand, a skewed discrete population is illustrated by considering the poisson distribution usually for count data.

The covariates Z_1 and Z_2 are generated from N(0,1) and N (0,100), respectively. The covariates are simulated to capture instances where uncontrollable factors are present in the experiment.

A population of 1,000 observations is simulated from each of the 120 simulation scenarios, see Table 1. Twenty datasets are generated for each of these populations. The CCD points (there are 9 points of these including the axial point, 2^k points, and a center point) are added within each set of the Xs to come up with 1000 observations. These CCD points are computed based on the means and standard deviations of X_1

and X_2, respectively, using the transformation $x^* = \frac{X - \overline{X}}{sd}$, where x* is the coded value of the factor levels and X is the natural value of the factor. The remaining 991 observations are generated from the corresponding sets of combinations of distributions mentioned above.

From these 1,000 observations, random samples of size 9, using simple random sampling without replacement (SRSWOR) is also drawn. Only nine sample points are chosen to facilitate comparison with CCD points.

The parameters of the true models (coefficients) summarized in Table 1 are arbitrarily chosen. The response variable y is computed using Models 1 and 2. Ordinary least squares regression is used to estimate the Model 1. Model 2 is also estimated with ordinary least squares but the variable w is intentionally deleted to simulate a case where a factor is missed out so that response surface bias is introduced. The estimation is also done using all 1,000 observations. For each of the datasets, we pick out 1) CCD points, and 2) random samples obtained from simple random sampling without replacement. Then OLS regression to the CCD points and random sample points are also implemented. Backfitting was used in estimating the response surface using the CCD points and to the random samples.

Table 1. Simulation scenarios

Model 1														
N(0,1)			N(0,100)			N(100,100)			Po(10)			Po(100)		
H	M	L	H	M	L	H	M	L	H	M	L	H	M	L
NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN
PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP
Model 2														
N(0,1)			N(0,100)			N(100,100)			Po(10)			Po(100)		
H	M	L	H	M	L	H	M	L	H	M	L	H	M	L
NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN	PN
PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP

Note: H-High values of k; M-Moderate values of k; L – Low values of k
NN - X_1~N(9,1), X_2~N(200,961), Z_1~N(0,1), Z_2~N(0,100)
NP - X_1~N(9,1), X_2~Po(200), Z_1~N(0,1), Z_2~N(0,100)
PN - X_1~Po(8), X_2~N(200,961), Z_1~N(0,1), Z_2~N(0,100)
PP - X_1~Po(8), X_2~Po(200), Z_1~N(0,1), Z_2~N(0,100)

The mean absolute percentage error (MAPE), which is the average of the absolute value of the quotient of the residual and the response variable multiplied by 100%, i.e. MAPE = average | [(y- $\hat{y}$)/y] *100%| , is used to evaluate the predictive ability of the estimated models obtained from backfitting. The MAPE values obtained using the backfitting of CCD points and SRS points are compared to that of the

MAPE values of the original population which used OLS in estimation. The relative advantage of backfitting will already manifest once it yield comparable performance in predicting responses to that of the actual population. Note that the proposed estimation method tend to set aside the effect of other factors while estimating others, hence, it may not be attractive at the start, but as the number of experimental factors increases, it will manifest the computational advantage.

3.1 Predictive ability

The MAPE computed from the 20 datasets simulated from Model 1 with low and high values of contamination constant *k* and having normally distributed error terms are summarized in Tables 2 to 4. High and low values of k illustrate how the estimation procedure will perform compared to OLS when the model fits the data well/poorly. In Table 2 where there is no misspecification and the error term behaves as N(0,1), the OLS provides MAPE very close to the MAPE obtained when the entire population is used in estimating the parameters. OLS is clearly optimal since all the requirements for the Gauss-Markov Theorem to hold is present. Models with parameters estimated from backfitting yield inferior predictive ability compared to OLS. However, even for the case where the error term is N(0,1), the backfitting algorithm shows slight advantage over OLS when model fit is poor (contaminated errors). The varying combination of distributions where the factors are generated from does not necessarily influence predictive ability of the model whether using OLS or backfitting or whether there is a good or poor model fit to the data.

Table 2. Average MAPE of datasets simulated from model 1 with error term N (0,1)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.999	0.998	0.999	0.999	**R^2**	0.11	0.041	0.1749	0.0995
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		0.39	0.38	0.38	0.37		229.80	251.31	142.82	176.42
OLS_{CCD}		0.53	0.64	0.55	0.55		242.85	431.18	155.76	215.67
BF_{CCD}		15.47	7.15	15.75	6.34		268.90	405.25	166.45	207.42
OLS_{SRS}		0.59	0.62	0.56	0.53		241.60	297.02	182.61	206.72
BF_{SRS}		15.09	7.90	15.05	6.88		275.06	269.87	187.14	181.43

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population data
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

In Table 3, as the variance of the error distribution increases, the predictive ability of the model estimated from backfitting changes. Even with good model fit, models estimated using backfitting yield comparable MAPE to those estimated using OLS. When model fit worsen (still with large error variance), the advantage of backfitting becomes clearer. The MAPE produced by models estimated using backfitting are comparable to the MAPE when the entire population data is used. Those MAPE coming from OLS–estimated models are way higher than the MAPE when the model is estimated using the entire population data.

Table 3. Average MAPE of datasets simulated from model 1 with error term N(0,100)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.9211	0.801	0.951	0.916	**R^2**	0.0053	0.00421	0.00658	0.00489
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		3.94	3.86	3.80	3.71		230.97	252.82	389.13	633.14
OLS_{CCD}		5.37	6.39	5.54	5.56		735.00	780.64	727.00	932.45
BF_{CCD}		16.57	9.29	16.80	8.22		619.36	631.82	628.89	688.06
OLS_{SRS}		5.78	5.91	6.04	5.73		792.27	713.67	945.66	1351.67
BF_{SRS}		16.54	8.76	20.28	8.89		602.22	591.71	1020.33	686.88

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

In Table 4, where aside from large error variance, mean is also nonzero, the predictive ability of OLS and backfitting are comparable to the predictive ability of a model estimated using the entire population data. This observation is valid whether there is good or poor model fit.

With poisson error distribution with lower mean, the backfitting estimates yield models that predict with similar reliability to OLS. The advantage of OLS over backfitting for models with good fit is no longer that remarkable. And as the model fit deteriorates, backfitting becomes more accurate over OLS in predicting responses. Similar is also true even when the mean (variance) of the poisson distribution increases further (see Table 6 for details). In Table 7, while the error term is N(0,1), a predictor is intentionally missed out in the estimation to simulate a scenario where there is omitted variable. For the case where there is a good model fit, OLS is only slightly advantageous over backfitting. When model fit declines, backfitting exhibits advantages over OLS. Similar is true when the variance of the error distribution increases (see Table 8 for details) or when aside from large variance, mean of error distribution is also nonzero (see Table 9 for details).

Table 4. Average MAPE of datasets simulated from model 1 with error term N(100,100)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.921	0.801	0.951	0.916	**R^2**	0.0053	0.00421	0.0066	0.0049
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		2.62	2.60	2.54	2.51		7.99	7.99	7.92	7.92
OLS_{CCD}		3.55	4.29	3.68	3.75		10.72	13.17	11.42	11.76
BF_{CCD}		10.90	6.23	11.03	5.54		10.03	10.88	10.38	10.57
OLS_{SRS}		4.09	4.69	4.09	3.58		10.91	12.34	11.24	11.79
BF_{SRS}		11.73	6.30	11.12	6.01		9.69	9.91	10.19	10.52

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 5. Average MAPE of datasets simulated from model 1 with error term Po(10)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.99	0.98	0.99	0.99	**R^2**	0.02	0.007	0.025	0.016
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		1.17	1.16	1.13	1.11		23.52	23.68	23.04	23.04
OLS_{CCD}		1.75	1.77	1.68	1.51		30.70	35.20	33.26	31.22
BF_{CCD}		14.19	6.87	13.28	6.52		28.48	32.86	30.40	28.62
OLS_{SRS}		1.90	1.76	1.66	1.65		31.60	35.60	36.49	31.55
BF_{SRS}		14.73	7.21	15.05	7.09		30.83	29.66	31.12	28.97

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 6. Average MAPE of datasets simulated from model 1 with error term Po(100)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.92	0.801	0.951	0.92	**R^2**	0.005	0.004	0.007	0.005
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		2.62	2.60	2.54	2.51		8.00	7.99	7.92	7.92
OLS_{CCD}		3.55	4.30	3.68	3.76		10.78	13.18	11.41	11.78
BF_{CCD}		10.90	6.22	11.03	5.55		10.02	10.88	10.37	10.59
OLS_{SRS}		4.27	4.84	3.83	4.16		11.95	11.81	11.00	10.74
BF_{SRS}		11.08	6.23	10.58	5.93		10.22	10.55	9.72	9.48

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 7. Average MAPE of datasets simulated from model 2 with error term N (0,1)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.95	0.99	0.989	0.997	**R^2**	0.022	0.022	0.0853	0.0834
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		6.05	3.38	31.41	5.27		414.79	334.08	14804.54	516.10
OLS_{CCD}		8.71	4.48	42.86	8.01		720.37	661.85	4614.44	775.82
BF_{CCD}		10.00	5.94	41.59	14.55		684.90	556.59	13343.80	720.41
OLS_{SRS}		9.04	5.17	29.59	9.55		778.91	593.15	5140.77	1174.39
BF_{SRS}		12.11	6.48	113.71	16.78		562.08	565.26	6569.49	897.44

Note:NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 8. Average MAPE of datasets simulated from model 2 with error term N(0,100)

		Low k					High k			
Method	**k**	0.001	0.001	0.001	0.001	**k**	1.00	1.00	1.00	1.00
	R^2	0.95	0.99	0.99	1	**R^2**	0.0038	0.004	0.004	0.005
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		6.05	3.38	31.41	5.27		181.11	137.50	144.79	134.73
OLS_{CCD}		8.71	4.48	42.86	8.01		690.97	581.91	688.22	560.83
BF_{CCD}		10.00	5.94	41.59	14.55		586.34	463.18	543.12	502.59
OLS_{SRS}		9.02	4.93	45.91	8.93		975.46	717.75	813.93	665.67
BF_{SRS}		12.05	6.16	47.17	12.10		1021.92	547.97	682.24	550.21

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 9. Average MAPE of datasets simulated from model 2 with error term N(100,100)

		Low k					High k			
Method	**k**	0.001	0.001	0.001	0.001	**k**	1.00	1.00	1.00	1.00
	R^2	0.95	0.99	0.989	0.997	**R^2**	0.004	0.004	0.004	0.005
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		4.70	2.49	8.46	4.56		8.07	8.11	8.08	8.12
OLS_{CCD}		6.77	3.26	15.18	6.55		11.48	11.79	11.04	10.63
BF_{CCD}		7.73	4.34	16.87	9.98		10.30	10.27	9.80	9.98
OLS_{SRS}		6.65	3.75	15.80	7.20		11.58	10.92	10.99	11.93
BF_{SRS}		9.10	5.96	27.85	8.82		10.11	10.28	10.01	9.96

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

In Tables 10 and 11, in addition to a missed factor, the errors are also coming from a skewed distribution (poisson). Again, when the model fits the data well, the predictive ability of models with parameters estimated using backfitting is comparable to those estimated using OLS. And as fit of the model deteriorates, backfitting becomes more advantageous (better predictive ability).

Table 10. Average MAPE of datasets simulated from model 2 with error term Po (10)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.92	0.94	0.979	0.99	**R^2**	0.01	0.01	0.01	0.01
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		6.37	4.96	21.75	10.39		28.41	28.68	28.48	28.13
OLS_{CCD}		9.52	7.83	28.74	23.73		37.51	42.83	38.67	37.76
BF_{CCD}		11.60	8.10	105.92	26.27		34.85	39.01	35.53	37.46
OLS_{SRS}		11.56	6.48	36.74	19.24		37.36	47.09	43.01	42.31
BF_{SRS}		12.97	6.69	52.08	22.65		33.20	37.36	39.19	33.17

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 11. Average MAPE of datasets simulated from model 2 with error term Po (100)

		Low k					High k			
Method	**k**	0.001	0.001	0.001	0.001	**k**	1.00	1.00	1.00	1.00
	R^2	0.95	0.99	0.989	0.997	**R^2**	0.004	0.004	0.004	0.005
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{POPN}		4.70	2.49	12.98	5.11		8.07	8.11	8.08	8.13
OLS_{CCD}		6.77	3.26	24.61	7.25		11.48	11.85	11.06	10.65
BF_{CCD}		7.73	4.34	26.53	10.70		10.32	10.27	9.82	10.00
OLS_{SRS}		7.39	3.81	20.27	6.36		11.04	12.13	12.22	13.46
BF_{SRS}		9.07	4.42	31.12	13.10		10.25	10.69	10.38	11.43

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSPOPN - OLS is used in estimation using the actual simulated population
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

It is clear that when backfitting is used in estimating parameters of a response surface model, the fitted model exhibits robustness to various nuisances associated with experimentation and model fitting. Contamination may occur during the experimentation stage to generate data that will be used in fitting the response surface model. Furthermore, some important predictors may be missed out in the desire to limit the number of factors to be included in the experiment. As a result, a model that poorly fits the data may be achieved. This will provide a poor approximation of the true response surface. If instead, backfitting is used in estimation, even with such problems with data-generation and model fitting, we can still expect a model that approximates the true response surface fairly well.

3.2 Parameter estimates

We summarize in Tables 12 to 14 the average percent difference between the parameter estimates from the 20 datasets using Model 1 with normal error terms and the true parameter values. When the model fits the data well, regardless of mean and variance values of the error distribution, backfitting provides estimates that are too far from the actual parameters compared to the estimates obtained using OLS. The optimality of OLS is again exhibited as it produced estimates that are very close to the true parameter values, see Table 12 for details. However, as the model fit deteriorates, regardless of the error distribution parameters, backfitting with both CCD points and SRS points are more advantageous over OLS in terms of parameter estimates. Backfitting was able to estimate parameters and are relatively closer to the actual parameters of the population than the estimates obtained using OLS. In the case where the error terms have large variance, the estimates obtained using backfitting with SRS points, given that a factor was generated from poisson distribution, the other, from normal distribution, are relatively far from the true parameter values. This can be explained by the kind of samples obtained in simple random sampling selection. Unlike the CCD which generated data that are predetermined or fixed before experimentation, SRS generates data that may vary as caused by sampling error. It is for this reason, that in addition to the desirable characteristic of CCD (orthogonality), backfitting is proposed as an estimation algorithm for a more robust approximation of the response surface.

Table 12. Average percent difference of parameter estimates from model 1 with error term N(0,1)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.9992	0.9975	0.9995	0.9991	**R^2**	0.10997	0.04099	0.17487	0.09947
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}		8.84	17.01	10.34	10.84		799.59	2301.73	1805.58	2392.49
BF_{CCD}		366.63	222.10	383.65	190.21		779.53	1225.95	1396.59	1604.71
OLS_{SRS}		9.63	13.71	9.26	8.29		1218.88	927.71	2308.28	2532.44
BF_{SRS}		291.69	209.60	263.75	162.00		964.82	838.42	1548.84	1810.07

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 13. Average percent difference of parameter estimates from model 1 with error term N(0,100)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.921	0.8013	0.951	0.916	**R^2**	0.0053	0.0042	0.00658	0.0049
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}		89.81	183.23	98.39	109.88		4396.60	6050.12	23235.92	3641.40
BF_{CCD}		392.99	321.08	415.56	221.21		3781.39	2555.76	17658.38	2799.82
OLS_{SRS}		107.44	90991.62	97.47	94.42		5781.15	5037.46	10951.15	4147.73
BF_{SRS}		309.71	271.47	354.90	171.30		3931.35	2979.18	16957.24	2027.08

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 14. Average percent difference of parameter estimates from model 1 with Error term N(100,100)

		Low k					High k			
Method	**k**	0.01	0.01	0.01	0.01	**k**	1.00	1.00	1.00	1.00
	R^2	0.921	0.801	0.951	0.916	**R^2**	0.0053	0.00421	0.00658	0.0049
		Combination of X_1 and X_2					**Combination of X_1 and X_2**			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}		85.29	177.16	95.74	104.41		3933.09	5539.17	21828.61	3097.86
BF_{CCD}		372.15	301.08	395.25	202.18		3366.91	2206.82	17032.69	2688.39
OLS_{SRS}		129.52	229.79	100.00	74.60		3620.56	3645.03	11788.54	3998.69
BF_{SRS}		359.72	246.48	227.53	202.20		2661.64	1621.96	21044.95	2984.56

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

When the error distribution is poisson and the model fits the data well, with lower mean/variance, backfitting yields inferior estimates over those obtained using OLS. However, backfitting yield better estimates than OLS when the model fits the data poorly. In Table 15, the varying combination of distributions where the factors were generated from, does not necessarily influence the quality of estimates whether using OLS nor backfitting. Noticeably, when model fit is not good, and among the four combinations of distributions that factors generated come from the combination of factors 1 and 2 coming from poisson distribution illustrates the optimality of the backfitting estimation.

Table 15. Average percent difference of parameter estimates from model 1 with error term Po(10)

	Low k					High k			
Method k	0.01	0.01	0.01	0.01	k	1.00	1.00	1.00	1.00
R^2	0.9915	0.97531	0.99482	0.99092	R^2	0.01595	0.00694	0.0252	0.01592
	Combination of X_1 and X_2					Combination of X_1 and X_2			
	NN	NP	PN	PP		NN	NP	PN	PP
OLS_{CCD}	32.27	41.77	33.64	28.88		2254.86	3172.71	4125.48	1099.66
BF_{CCD}	329.24	186.83	267.91	203.82		1949.41	2677.98	3412.17	669.85
OLS_{SRS}	40.78	39.39	34.30	39.84		2010.85	3091.18	4803.45	987.43
BF_{SRS}	311.01	168.40	308.20	173.28		1582.22	2678.05	2564.43	622.90

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

In Table 16, good model fit again exhibited the advantage of OLS over backfitting. In some extreme cases (high variance), the estimates from backfitting can be comparable to the estimates obtained using OLS. Evidences of backfitting yielding better estimates over OLS is clearly seen when model fit deteriorates. Except for a missed out variable, when the model fit is good and under the most ideal setting N(0,1), OLS and backfitting produced comparable estimates of the parameters. This is also true for the other cases of normal error distribution but with large variance, see Tables 18-19 for details. The advantages of backfitting over OLS manifests when the model fits the data poorly, see Tables 17-19.

Table 16. Average percent difference of parameter estimates from model 1 with error term Po(100)

	Low k					High k			
Method k	0.01	0.01	0.01	0.01	k	1.00	1.00	1.00	1.00
R^2	0.92101	0.80118	0.95107	0.91603	R^2	0.00528	0.004215	0.006625	0.004875
	Combination of X_1 and X_2					Combination of X_1 and X_2			
	NN	NP	PN	PP		NN	NP	PN	PP
OLS_{CCD}	85.28	178.38	96.04	105.02		7704.86	3460150.13	4083.97	3782.31
BF_{CCD}	372.37	303.43	396.97	202.78		6980.75	3895.80	3134.18	3341.43
OLS_{SRS}	122.78	152.74	113.25	148.71		11720.58	1327091.84	7438.18	2355.14
BF_{SRS}	278.12	162.22	252.34	217.32		2623.10	6641.87	5508.30	1608.18

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 17. Average percent difference of parameter estimates from model 2 with error term N(0,1)

	Low k					High k			
Method k	0.01	0.01	0.01	0.01	**k**	1	1	1	1
R^2	0.95154	0.985385	0.9886	0.996665	R^2	0.022245	0.02176	0.0853	0.083395
	Combination of X_1 and X_2					Combination of X_1 and X_2			
	NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}	33.84	22.30	46.31	24.17		4130.28	1452.54	1283.38	2383.22
BF_{CCD}	42.26	44.21	71.13	73.30		1405.77	700.92	693.28	2869.53
OLS_{SRS}	42.20	28.52	37.16	30.86		2223.29	1069.18	1183.91	4821.04
BF_{SRS}	57.01	49.27	70.15	79.09		1071.87	704.52	832.88	3082.58

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using Central Composite Design points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 18. Average percent difference of parameter estimates from model 2 with error term N(0,100)

		Low k					High k			
Method	**k**	0.001	0.001	0.001	0.001	**k**	1.00	1.00	1.00	1.00
	R^2	0.95154	0.985385	0.9886	0.996665	R^2	0.003845	0.003845	0.00438	0.00487
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}		33.84	22.30	46.31	24.17		5970.30	5834.56	250430.61	3382.12
BF_{CCD}		42.26	44.21	71.13	73.30		4464.05	3094.52	229988.12	2022.77
OLS_{SRS}		35.91	28.93	52.15	30.78		5806.98	11260.40	587003.62	3528.47
BF_{SRS}		63.25	45.97	73.66	55.17		2706.18	8910.14	582469.05	2019.23

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

With poisson error distribution and in addition, there is a predictor that is missed out, Tables 20-21 show the quality of the estimates obtained from backfitting are comparable to the quality of estimates obtained from OLS. The estimates of backfitting, on the other hand, are better than the estimates of OLS when the model does not fit the data well.

Table 19. Average percent difference of parameter estimates from model 2 with error term N(100,100)

Method		Low k					High k			
	k	0.001	0.001	0.001	0.001	k	1.00	1.00	1.00	1.00
	R^2	0.95154	0.9854	0.9886	0.9967	R^2	0.003845	0.00385	0.00438	0.00487
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	NP	PN	PP		NN	NP	PN	PP
OLS_{CCD}		33.68	22.21	46.22	24.07		4373.67	4961.82	128156.76	3069.41
BF_{CCD}		41.96	43.95	70.87	73.04		3128.60	2675.00	51341.24	1967.24
OLS_{SRS}		37.80	25.04	33.55	36.14		4697.38	5058.30	138150.07	9788.76
BF_{SRS}		69.15	56.62	97.80	59.52		3550.28	3304.00	7074.23	5553.65

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

Table 20. Average percent difference of parameter estimates from model 2 with error term Po(10)

Method		Low k					High k			
	k	0.01	0.01	0.01	0.01	k	1.00	1.00	1.00	1.00
	R^2	0.9173	0.9444	0.97923	0.98687	R^2	0.00718	0.00531	0.01086	0.01237
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	NP	PN	PP		NN	NP	PN	PP
OLS_{CCD}		58.73	54.51	54.38	47.77		2741.20	41217.70	2331.33	1141.52
BF_{CCD}		95.43	59.61	100.92	93.71		1524.67	36497.88	1344.01	883.94
OLS_{SRS}		102.46	39.53	68.61	59.50		2439.04	27277.03	3879.13	2570.62
BF_{SRS}		118.52	46.07	89.73	58.20		1147.31	24498.41	2783.32	842.17

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

The different scenarios illustrated that when there is contamination (e.g., error distribution, missed out factor) in the data and in the process of response surface estimation, backfitting provides estimates that are relatively robust to such contaminations. The use of CCD to generate data in fitting the response surface is necessary to ensure orthogonality and subsequently, additivity of the response surface that will facilitate the implementation of the backfitting algorithm.

Table 21. Average percent difference of parameter estimates from model 2 with error term Po(100)

Method		Low k					High k			
	k	0.001	0.001	0.001	0.001	k	1.00	1.00	1.00	1.00
	R^2	0.95154	0.98538	0.988595	0.99666	R^2	0.003815	0.00386	0.004335	0.004885
		Combination of X_1 and X_2					Combination of X_1 and X_2			
		NN	**NP**	**PN**	**PP**		**NN**	**NP**	**PN**	**PP**
OLS_{CCD}		33.61	22.30	46.16	23.99		7217.73	6374.97	5405.40	2746.09
BF_{CCD}		41.96	43.98	70.86	73.03		5506.84	2485.53	2874.34	1733.80
OLS_{SRS}		49.37	31.78	47.54	21.94		11667.60	5226.17	12279.07	6660.86
BF_{SRS}		63.95	42.36	103.57	59.91		7075.23	1953.64	9483.85	6362.91

Note: NN - X1~N(9,1), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
NP - X1~N(9,1), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
PN - X1~Po(8), X2~N(200,961), Z1~N(0,1), Z2~N(0,100)
PP - X1~Po(8), X2~Po(200), Z1~N(0,1), Z2~N(0,100)
OLSCCD - OLS is used in estimation using CCD points
OLSSRS - OLS is used in estimation using simple random sample points
BFCCD- Backfitting is used in estimation using CCD points
BFSRS- Backfitting is used in estimation using simple random sample points

4. Conclusions

The simulation study shows that when the model fits fairly well to the data, backfitting is comparable to ordinary least squares in estimating a response surface model. Backfitting, as an iterative procedure, works well under cases where severe misspecification in the model is present, possibly caused by contamination in the course of experimentation. These problems like omission of variables, error assumption, or experimental contamination, all leading towards failure of OLS to approximate the response surface. Backfitting is robust to various nuisances in experimentation that usually generates data used in fitting a response surface. When the fitted response surface is prone to bias due to high dimensionality of operating variables, backfitting can intervene in the estimation, provided that a design leading to an orthogonal design matrix is used.

REFERENCES

Aslan, N. (2008), Application of Response Surface Methodology and Central Composite Rotatable Design for Modeling and Optimization of a Multigravity Separator for Chromite Concentration, *Powder Technology* 185: 80-86.

Butler, N. (2006), On the Minimum bias response surface designs of Box and Draper, *Journal of Statistical Planning and Inference* 136: 3221-3230.

Carley, K., Kamneva, N. and Reminga, J. (2004), Response Surface Methodology CASOS Report, CMU-ISRI- 04-136.

Draper, N. and Smith, H. (1998), *Applied Regression Analysis* 3rd ed., New York: John Wiley and Sons, pp. 165 – 167.

Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Model*, London: St. Edmundsbury Press.

Inouye, B. (2001), Response Surface Experimental Designs for Investigating Interspecific Competition, *Ecology* 82: 2696-2706.

Khuri, A. and Cornell, J. (1987), *Response Surfaces Designs and Analyses,* New York: Marcel Dekker Inc.

Hsu, S. (1995), Optimization of the Surimi Processing System with a Central Composite Design Method, *Journal of Food Engineering* 24: 101-111.

Li, Y., Li, Y. and Yu, S. (2008), Design Optimization of a Current Mirror Amplifier Integrated Circuit using a Computational Statistics Technique, *Mathematics and Computers in Simulation* 79:1165-1177.

Myers, R. and Montgomery, D. (2002), *Response Surface Methodology Process and Product Optimization using Designed Experiments,* Canada: John Wiley & Sons Inc.

Opsomer, J. (2000), Asymptotic Properties of Backfitting Estimators, *Journal of Multivariate Analysis* 73: 166-179.