Robust Parameter Design: A Semi-Parametric Approach

Stephanie M. Pickle
Department of Statistics, Virginia Polytechnic Institute & State University

Timothy J. Robinson
Department of Statistics, University of Wyoming

Jeffrey B. Birch
Department of Statistics, Virginia Polytechnic Institute & State University

Christine M. Anderson-Cook
Statistical Sciences Group, Los Alamos National Laboratory

Abstract

Parameter design or robust parameter design (RPD) is an engineering methodology intended as a cost-effective approach for improving the quality of products and processes. The goal of parameter design is to choose the levels of the control variables that optimize a defined quality characteristic. An essential component of robust parameter design involves the assumption of well estimated models for the process mean and variance. Traditionally, the modeling of the mean and variance has been done parametrically. It is often the case, particularly when modeling the variance, that nonparametric techniques are more appropriate due to the nature of the curvature in the underlying function. Most response surface experiments involve sparse data. In sparse data situations with unusual curvature in the underlying function, nonparametric techniques often result in estimates with problematic variation whereas their parametric counterparts may result in estimates with problematic bias. We propose the use of semi-parametric modeling within the robust design setting, combining parametric and nonparametric functions to improve the quality of both mean and variance model estimation. The proposed method will be illustrated with an example and simulations.

KEY WORDS: Response surface; Nonparametric regression; Model robust regression; Genetic algorithm

1 Introduction

In the mid 1980's, Japanese quality consultant Genichi Taguchi popularized a cost-efficient approach to quality improvement known as robust parameter design (RPD). Taguchi postulated that there are two types of factors which operate on a process: control factors and noise factors. Control factors are variables whose levels remain unchanged in the process once they are set, whereas the levels of the noise factors change randomly within the process and cause unwanted variation in the response, y. The goal of robust parameter design is to determine levels of the control factors which cause the response to be robust to changes in the levels of the noise variables. A popular design for studying both the impact of control factors and noise factors on a process is the crossed array. A $2^2 \times 2^2$ crossed array is shown in Figure 1. Variation in the process which results from uncontrollable fluctuations in the noise factors can be summarized by taking the sample variance of the points in the noise factor space at each of the control factor settings (4 points in the $2^2 \times 2^2$ example). The process can be made robust to the variation associated with the noise factors by choosing the factor combination of the control factors corresponding to the smallest sample variance. It is often the case that the levels of the noise factors are unobservable not only in the process but also in a controlled experimental setting. In these situations, replications at the control factor settings provide the researcher with an idea of process variability and the approach to robust design is the same; namely, to choose the factor combination in the control factor space which corresponds to the smallest sample variance. It is these types of situations which will be the focus of this manuscript.

Instead of using only the sample variances for describing the underlying process variance, Vining and Myers (1990) introduced a dual model response surface approach to RPD in which it is assumed that both the mean and variance can be described by separate parametric regression models. Optimal control factor settings are then found using constrained optimization (constrained estimated mean and minimized process variance with respect to control factor settings).

If one or both models are misspecified by the researcher, the estimates may be highly biased and, consequently, the optimal control factor settings may be misspecified. Vining and Bohn (1998) [henceforth referred to as VB] point out that traditional parametric models are often inadequate, particularly when modeling the variance, and suggest the use of nonparametric techniques for modeling the variance. Unfortunately, in sparse data situations, which are typically the case with response surface experiments, nonparametric techniques often result in highly variable estimates. To overcome the pitfalls associated with each method, we propose the use of semi-parametric modeling within the robust design setting.

By using a semi-parametric technique for modeling, we can combine parametric and nonparametric functions to improve the quality of both the mean and variance models. The resulting semi-parametric estimates will have smaller bias and variance. Furthermore, these hybrid estimates will result in a better understanding of the process at hand. In turn, the optimal factor settings are

less likely to be misspecified. We will illustrate the use and benefit of our proposed method with an application to the Box and Draper (1987) printing ink study and simulations. We also suggest the use of a more flexible and efficient optimization routine, the genetic algorithm, for determining optimal control factor settings.

2 An Overview of Parametric and Nonparametric Approaches

2.1 Parametric Approach

Given the data from a crossed array, there are a number of potential approaches to directly modeling the mean and variance as a function of the control factors. A general approach is to assume that the underlying functional forms for the mean and variance models can be expressed parametrically. Assuming a d point design with n_i replicates at each location ($i = 1, 2, \ldots, d$), the point estimators of the process mean and variance, $\bar{y}_i$ and s_i^2, respectively, form the data for the dual response system. Since the purpose of this article is to demonstrate the utility of a hybrid approach (combining a parametric and nonparametric approach to modeling) for robust design, we will consider an "off the shelf" model for the mean. An "off the shelf" model for the process mean is linear in the model parameters and can be written as:

$$\text{Means Model:} \qquad \bar{y}_i = \mathbf{x}_i'\boldsymbol{\beta} + g^{1/2}\left(\mathbf{x}_i^{*\prime};\boldsymbol{\gamma}\right)\varepsilon_i, \tag{1}$$

where $\mathbf{x}_i'$ and $\mathbf{x}_i^{*\prime}$ are $1 \times k$ and $1 \times l$ vectors of means model and variance model regressors, respectively, expanded to model form, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $k \times 1$ and $m \times 1$ vectors of mean and variance model parameters, respectively, g is the underlying variance function, and ε_i denotes the random error for the mean function. The ε_i are assumed to be uncorrelated with mean zero and variance of one. Note that the model terms for the i^{th} observation in the means model are denoted by $\mathbf{x}_i'$ while the model terms for the variance model are denoted by $\mathbf{x}_i^{*\prime}$. This allows for the fact that the process mean and variance may not depend on the same set of regressors.

Similar to the modeling of the mean, various modeling strategies have been utilized for estimating the underlying variance function. Bartlett and Kendall (1946) demonstrated that if the errors are normal about the mean model and if the design points are replicated, the variance can be modeled via a log-linear model with the d sample variances utilized for the responses. A great deal of work has also been done using generalized linear models for estimating the variance function. Although not an exhaustive list, the reader is referred to Box and Meyer (1986), Aitkin (1987), Grego (1993), Lee and Nelder (2003), and Myers, Brenneman, and Myers (2005). As mentioned previously, since the purpose of this manuscript is to demonstrate the utility of a hybrid approach to modeling, we choose an "off the shelf" approach to variance modeling. The

log-linear model proposed by Bartlett and Kendall (1946) is a popular one [see Vining and Myers (1990) and Myers and Montgomery (2002)] and is written explicitly as:

$$\text{Variance Model:} \qquad \ln(s_i^2) = g^*(\mathbf{x}_i^*) + \eta_i = \mathbf{x}_i^{*\prime}\boldsymbol{\gamma} + \eta_i, \tag{2}$$

where η_i denotes the model error term whose expectation is assumed to be zero and whose variance is assumed constant across the d design points.

Assuming the model forms for the mean and variance given in (1) and (2), the model parameters are estimated using the following estimated weighted least squares (EWLS) algorithm:

Step 1: Fit the variance model, $\ln(s_i^2) = \mathbf{x}_i^{*\prime}\boldsymbol{\gamma} + \eta_i$, via ordinary least squares (OLS), obtaining $\hat{\boldsymbol{\gamma}}^{(OLS)} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^*\mathbf{y}^*$ where $\mathbf{y}^*$ is the $d \times 1$ vector of log transformed sample variances.

Step 2: Use $\hat{\sigma}_i^2 = \exp\left(\mathbf{x}_i^{*\prime}\hat{\boldsymbol{\gamma}}^{(OLS)}\right)$ as the estimated variances to compute the $d \times d$ estimated variance-covariance matrix for the means model, $\hat{\mathbf{V}} = diag\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_d^2\right)$.

Step 3: Use $\hat{\mathbf{V}}^{-1}$ as the estimated weight matrix to fit the means model, yielding $\hat{\boldsymbol{\beta}}^{(EWLS)} = \left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\bar{\mathbf{y}}$ where $\bar{\mathbf{y}}$ denotes the $d \times 1$ vector of sample averages.

The algorithm above yields the following estimates of the process mean and variance functions:

$$\text{Estimated Process Mean:} \qquad \widehat{E[y_i]}^{(EWLS)} = \mathbf{x}_i'\hat{\boldsymbol{\beta}}^{(EWLS)} \tag{3}$$

$$\text{Estimated Process Variance:} \qquad \widehat{Var[y_i]}^{(OLS)} = \exp\left(\mathbf{x}_i^{*\prime}\hat{\boldsymbol{\gamma}}^{(OLS)}\right). \tag{4}$$

Once estimates of the mean and variance have been calculated, the goal becomes finding the operating conditions for the control factors such that the mean is as close as possible to the target while maintaining minimum process variance.

Any control factor which influences the expression in (4) is known as a *dispersion* factor. Any control factor that does not influence the expression in (4) but does influence the expression in (3) is known as an *adjustment* factor. When both dispersion and adjustment factors are present, the robust design problem can be approached in a two-step fashion. Specifically, levels of the dispersion factors are chosen so as to minimize the estimated process variance in (4), and then the levels of the adjustment factors are chosen so as to bring the estimated process mean in (3) to a desired level. If only dispersion factors are present and these factors also influence the process mean, the researcher is left with finding the levels of the control factors that yield a desirable trade-off between low variance and a deviation from the targeted mean. This is often

accomplished via minimization of an objective function such as the squared error loss (SEL):

$$\text{SEL} = E\left[y(\mathbf{x}) - T\right]^2 = \left\{E\left[y(\mathbf{x})\right] - T\right\}^2 + Var\left[y(\mathbf{x})\right], \qquad (5)$$

where T denotes the target value for the process mean. Minimization can be accomplished via non-linear programming using a method such as the generalized reduce gradient or the Nelder-Mead simplex algorithm. The squared error loss approach is also useful when adjustment factors are present but are not strong enough to bring the mean to the targeted value. Note that the determined set of optimal operating conditions is highly dependent on quality estimation of both the mean and variance functions. Misspecification of the forms of either the mean or variance models can have serious implications in process optimization.

2.2 Nonparametric Approach

Situations may arise in which the user cannot explicitly state parametric forms for the dual model. In these situations, parametric specifications may result in serious bias of the estimated mean and/or variance. To prevent the bias induced by parametric model misspecification, VB and Anderson-Cook and Prewitt (2005) [henceforth referred to as AP] suggest the use of nonparametric regression for estimating the process mean and variance. Expressing the dual model where the mean and variance functions (h and g^*, respectively) are assumed to have unknown but smooth forms we have:

$$\text{Means Model:} \qquad \bar{y}_i = h\left(\mathbf{x}_i'\right) + g^{1/2}\left(\mathbf{x}_i^{*\prime}\right)\varepsilon_i$$

$$\text{Variance Model:} \qquad \ln(s_i^2) = g^*\left(\mathbf{x}_i^{*\prime}\right) + \eta_i.$$

Similar to parametric regression, estimators are linear combinations of the response values $\bar{y}_i$ and $\ln(s_i^2)$; however, the weighting schemes in some nonparametric regression methods assign more weight to observations closest to the point of prediction, $\mathbf{x}_0$. The nonparametric fits are more flexible than the parametric fits as they are not confined to the user's specified form. This enables the nonparametric approach to more adequately fit processes whose underlying models have more complicated forms than those expressed by the linear models in (1) and (2).

Several fitting techniques have been proposed in the nonparametric regression literature such as kernel regression [see for example Nadaraya (1964), Watson (1964), Priestley and Chao (1972), and Gasser and Müller (1984)], local polynomial models [see for example Fan and Gijbels (1996)], spline-based smoothers, and series-based smoothers [see for example Ruppert, Wand, and Carrol (2003)]. VB first applied nonparametric smoothing in the RPD setting by using the Gasser-Müller estimator for the dual response problem. AP continued with this idea by using the Nadaraya-Watson estimator and local polynomial regression, the method used in this research. Local polynomial regression (LPR) is a popular class of nonparametric smoothing methods and

is particularly appealing in response surface applications due to its robustness to biased estimates at the boundary of the design space. LPR is essentially a weighted least squares (WLS) problem where the weights are given by a kernel function. The polynomial form of the local polynomial fit can be of order one or greater and we focus on degree $p = 1$ (local linear regression (LLR)) in this article.

For the multiple regressor case, at point $\mathbf{x}_0 = (x_{01}, x_{02}, \ldots, x_{0k})$ where prediction is desired, we define the kernel function as:

$$\mathbf{K}(\mathbf{x}_0, \tilde{\mathbf{x}}_i) = \frac{1}{b^k} \prod_{j=1}^{k} K\left(\frac{x_{0j} - \tilde{x}_{ij}}{b}\right), \tag{6}$$

where $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$, $K\left(\frac{x_{0j}-\tilde{x}_{ij}}{b}\right)$ is a univariate kernel function, and b is the bandwidth. Note that when estimating both the mean and variance nonparametrically, a different kernel function may be used for the mean than for the variance since the regressors effecting the mean do not necessarily effect the variance. The choice of kernel function is not crucial to the performance of the estimator (Simonoff, 1996). Thus, for convenience, we will use the simplified Gaussian kernel, $K(u) = e^{-u^2}$.

The smoothness of the estimated function is controlled by the bandwidth, b. Since the coding of variables in response surface designs typically involves centering and scaling, the units are comparable in all directions. Thus, it is reasonable to use the same bandwidth, b, in all dimensions as expressed in (6). The choice of bandwidth is critical and the literature is rich with bandwidth selection methods [see for example Härdle (1990), Härdle et al. (2004)]. Typically the bandwidth is chosen to minimize some optimality criteria such as MSE. Mays, Birch, and Starnes (2001) [henceforth referred to as MBS] introduce a penalized cross-validation technique, PRESS**, for choosing an appropriate bandwidth. The approach chooses the bandwidth as the value b that minimizes PRESS**, defined as:

$$\text{PRESS}^{**} = \frac{\text{PRESS}}{d - trace\left(H^{(LLR)}\right) + (d - (k+1))\,\frac{SSE_{max} - SSE_b}{SSE_{max}}},$$

where SSE_{max} is the largest error sum of squares over all possible bandwidth values, SSE_b is the error sum of squares associated with a particular bandwidth value b, k is the number of regressors, and the prediction error sum of squares, PRESS, is given by:

$$\text{PRESS} = \sum_{i=1}^{d} \left(y_i - \hat{y}_{i,-i}\right)^2,$$

where $\hat{y}_{i,-i}$ denotes the estimated response obtained by leaving out the i^{th} observation when estimating at location $\mathbf{x}_i$. The LLR smoother matrix, $\mathbf{H}^{(LLR)}$,

is defined as:

$$\mathbf{H}^{(LLR)} = \begin{bmatrix} \mathbf{h}_1^{(LLR)\prime} \\ \mathbf{h}_2^{(LLR)\prime} \\ \vdots \\ \mathbf{h}_d^{(LLR)\prime} \end{bmatrix}$$

where $\mathbf{h}_i'$ is defined below. MBS show that PRESS** performs well by guarding against very small and very large bandwidths.

The nonparametric estimate of the dual model is found by first estimating the underlying variance function and then, using the estimated variances as weights, an estimated weighted local linear regression (EWLLR) fit is found for the mean. For more information regarding weighted LLR, the reader is referred to Lin and Carroll (2000). Expressions for the fits are provided below:

Estimated Process Mean:
$$\begin{aligned} \widehat{E[y_0]}^{(EWLLR)} &= \mathbf{x}_0' \hat{\boldsymbol{\beta}}^{(EWLLR)} \\ &= \mathbf{x}_0' \left(\mathbf{X}'\mathbf{W}_0\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{W}_0\bar{\mathbf{y}} \\ &= \mathbf{h}_0^{(EWLLR)\prime}\bar{\mathbf{y}} \end{aligned} \tag{7}$$

Estimated Process Variance:
$$\begin{aligned} \widehat{Var[y_0]}^{(LLR)} &= \exp\left(\mathbf{x}_0^{*\prime}\hat{\boldsymbol{\gamma}}^{(LLR)}\right) \\ &= \exp\left[\mathbf{x}_0^{*\prime}\left(\mathbf{X}^{*\prime}\mathbf{W}_0^*\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{W}_0^*\mathbf{y}^*\right] \\ &= \exp\left(\mathbf{h}_0^{(LLR)\prime}\mathbf{y}^*\right). \end{aligned} \tag{8}$$

Regarding notation for the means fit, $\mathbf{h}_0^{(EWLLR)\prime} = \mathbf{x}_0'(\mathbf{X}'\mathbf{W}_0\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_0\bar{\mathbf{y}}$, $\mathbf{W}_0 = \left\langle \sqrt{h_{0i}^{(KER)}} \right\rangle \hat{\mathbf{V}}^{-1} \left\langle \sqrt{h_{0i}^{(KER)}} \right\rangle$ where $\left\langle \sqrt{h_{0i}^{(KER)}} \right\rangle$ is the diagonal matrix containing the square roots of the kernel weights associated with $\mathbf{x}_0$, $\left\langle \sqrt{h_{0i}^{(KER)}} \right\rangle = diag\left(\sqrt{h_{01}^{(KER)}}, \sqrt{h_{02}^{(KER)}}, \ldots, \sqrt{h_{0d}^{(KER)}}\right)$ with $h_{0i}^{(KER)} = \frac{\mathbf{K}(\mathbf{x}_0, \tilde{\mathbf{x}}_i)}{\sum_{i=1}^{d} \mathbf{K}(\mathbf{x}_0, \tilde{\mathbf{x}}_i)}$, and $\hat{\mathbf{V}}$ is the estimated variance-covariance matrix, $\hat{\mathbf{V}} = diag\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_d^2\right)$. Regarding notation for the variance fit, $\mathbf{h}_0^{(LLR)\prime} = \mathbf{x}_0^{*\prime}(\mathbf{X}^{*\prime}\mathbf{W}_0^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{W}_0^*\mathbf{y}^*$ and $\mathbf{W}_0^*$ is the diagonal matrix containing the kernel weights associated with $\mathbf{x}_0^*$. Under the assumption of normality of $\bar{\mathbf{y}}$ and $\mathbf{y}^*$, the estimates of $E[y_0]$ and $Var[y_0]$ given by (7) and (8) are the local maximum likelihood estimates of Fan, Heckman, and Wand (1995).

Similar to the parametric approach to robust design, once estimates of the mean and variance functions have been calculated, a squared error loss approach will be used for process optimization. Unfortunately, most of the analytic optimization methods suggested for the parametric approach are based on gradient techniques which require continuous functions with derivatives for the estimated mean and variance functions. Since the mean and variance estimates from nonparametric methods do not result in closed form expressions, these optimization

routines are no longer applicable. VB utilize a simplex search based on the AMOEBA algorithm (Vetterling et al., 1992) which does not require the calculation of derivatives; however, simplex methods tend to be time consuming and often find local, as opposed to global optima [for details, see Haupt and Haupt (2004)]. Therefore we advocate the use of genetic algorithms for optimization.

The genetic algorithm (GA), originally developed by Holland (1975), has become a popular optimization technique. It is especially useful for optimizing functions that do not have known parametric forms, as it does not require derivatives to find the optimal solutions. Instead, the GA is based on the principles of genetics and uses evolutionary concepts such as selection, crossover, and mutation to find the optimal solutions. Furthermore, GA uses an intelligent, sequential search strategy which enables the user to find global, not local, solutions more efficiently (Goldberg, 1989). Thus, we will use the GA for process optimization.

2.3 Parametric vs. Nonparametric

Parametric and nonparametric approaches to modeling each possess positive and negative attributes. The parametric method is superior if the true, underlying functions can be adequately expressed parametrically and if the user correctly specifies the parametric forms. However, if either of the models is misspecified, the estimates may be highly biased and optimal control factor settings may be miscalculated. On the other hand, if the user has no idea about the true form of the underlying functions, nonparametric methods offer a nice alternative. Nonparametric methods can provide superior fits by capturing structure in the data that a misspecified parametric model cannot. However, nonparametric methods were originally developed for situations with large sample sizes whereas a main underpinning of RSM is the use of cost-efficient experimental designs (i.e., small sample sizes). In small sample settings, nonparametric fitting techniques may fit irregularities in the data too closely thereby creating estimated mean and variance functions that are highly variable. Consequently, optimization may be based on non-reproducible idiosyncrasies in the data. MBS introduce methods which are essentially hybrids of the parametric and nonparametric methods. These semi-parametric approaches produce estimated functions which are characterized by lower bias than parametric approaches and lower variance than nonparametric approaches. The details of this hybrid approach appear in the next section.

3 A Semi-Parametric Approach

When used individually, both parametric and nonparametric regression methods have shortcomings. In this section we present the use of semi-parametric techniques for overcoming some of these drawbacks, especially in situations where the user has partial knowledge of the underlying model or the data contains important "bumps" that parametric models cannot capture. The semi-parametric

estimates proposed combine individual parametric and nonparametric fits via appropriately chosen mixing parameters. We detail the methodologies below.

3.1 Model Robust Regression 1 (MRR1)

Einsporn and Birch (1993) proposed a semi-parametric method for modeling the mean response for assumed constant error variance. Their technique, model robust regression 1 (MRR1), combines parametric and nonparametric fits to the raw data in a convex combination via a mixing parameter, λ. For instance, if $\hat{\mathbf{y}}^{(OLS)}$ denotes the vector of ordinary least squares estimates of the mean and $\hat{\mathbf{y}}^{(LLR)}$ denotes the vector of local linear regression estimates of the mean, then the MRR1 estimated mean responses are obtained as:

$$\hat{\mathbf{y}}^{(MRR1)} = \lambda \hat{\mathbf{y}}^{(LLR)} + (1-\lambda)\hat{\mathbf{y}}^{(OLS)},$$

where $\lambda \in [0,1]$. For cases where the user's specified parametric model is correct, the optimal value of the mixing parameter, λ, is 0 and as the amount of model misspecification increases, λ increases from 0 to 1.

Similar to the choice of bandwidth in LLR, the choice of mixing parameter, λ, involves a bias-variance trade-off. MBS derive the following data driven expression for the asymptotically optimal value of the mixing parameter, λ, for MRR1:

$$\hat{\lambda}_{opt}^{(MRR1)} = \frac{\left\langle \hat{\mathbf{y}}_{-i}^{(LLR)} - \hat{\mathbf{y}}_{-i}^{(OLS)}, \mathbf{y} - \hat{\mathbf{y}}^{(OLS)} \right\rangle}{\left\| \hat{\mathbf{y}}^{(LLR)} - \hat{\mathbf{y}}^{(OLS)} \right\|^2}, \tag{9}$$

where the i^{th} observations of $\hat{\mathbf{y}}_{-i}^{(LLR)}$ and $\hat{\mathbf{y}}_{-i}^{(OLS)}$ are $\hat{y}_{i,-i}^{(LLR)}$ and $\hat{y}_{i,-i}^{(OLS)}$, respectively. The values $\hat{y}_{i,-i}^{(LLR)}$ and $\hat{y}_{i,-i}^{(OLS)}$ denote the LLR and OLS estimates, respectively, obtained by leaving out the i^{th} observation when estimating at $\mathbf{x}_i$. The notation $\langle\ \rangle$ represents the inner product and $\|\ \|$ represents the standard L_2 (Euclidean) norm.

MRR1 produces a smooth estimate that captures important anomalies in the data, which parametric methods are incapable of modeling. By containing a parametric portion for the overall fit, MRR1 brings stability to the overall fit and eliminates over-fitting the data, a problem associated with nonparametric regression. Thus, MRR1 estimates often have smaller bias and variance than their individual parametric and nonparametric counterparts, especially for small sample sizes. It should be noted, however, if there are locations in the data where both the parametric and nonparametric estimates are too high or too low, then the MRR1 estimates will also be too high or too low as the method has no means to correct for the error.

3.2 Model Robust Regression 2 (MRR2)

Model robust regression 2 (MRR2) was introduced by Mays, Birch, and Einsporn (2000) as an improvement to the MRR1 approach for estimating the mean

with constant variance. Similar to MRR1, MRR2 combines a parametric fit and a nonparametric fit via a mixing parameter; however, the parametric fit is to the raw data (as in MRR1) while the nonparametric fit is to the residuals from the parametric fit. The vector of residuals ($\mathbf{r}$) represents the structure in the data which is not captured by the user specified parametric model. The vector of residuals is fit nonparametrically via LLR resulting in the following vector of smoothed residuals:

$$\hat{\mathbf{r}} = \mathbf{H}_r^{(LLR)}\mathbf{r},$$

where $\mathbf{H}_r^{(LLR)}$ is computed similarly to the LLR smoother matrix in (7) but with the "response" variable being the residuals from the OLS fit to the raw data. The MRR2 estimates are then obtained by adding a portion of the LLR smoothed residuals back to the original parametric fit, yielding:

$$\hat{\mathbf{y}}^{(MRR2)} = \hat{\mathbf{y}}^{(OLS)} + \lambda\hat{\mathbf{r}},$$

where $\lambda \in [0, 1]$. The size of the mixing parameter for MRR2 does not necessarily represent the amount of model misspecification as it does in MRR1. Instead, the value of λ indicates the amount of correction needed from the residual fit. Notice that $\lambda = 1$ actually represents only a 50% contribution from the residual fit as the coefficient for the parametric fit is always one. Similar to MRR1, MBS derive a data driven expression for the asymptotically optimal mixing parameter. The expression is given as:

$$\hat{\lambda}_{opt}^{(MRR2)} = \frac{\left\langle \hat{\mathbf{r}}, \mathbf{y} - \hat{\mathbf{y}}^{(OLS)} \right\rangle}{\|\hat{\mathbf{r}}\|^2}. \tag{10}$$

MBS show that both MRR methods perform as well or better than the individual parametric and nonparametric methods for varying degrees of model misspecification both in an asymptotic sense as well as small sample settings.

3.3 Dual Model Robust Regression (DMRR)

Robinson and Birch (2002) extend the MRR techniques to models with non-constant error variance. Robinson and Birch (2002) consider the unreplicated design case and therefore use a residual-based variance estimate. Similar to the results of MBS, Robinson and Birch (2002) show that dual model robust regression (DMRR) is not only asymptotically superior to its parametric and nonparametric counterparts but that DMRR also performs better in small sample settings. Consequently, we propose DMRR as a natural tool for RSM and the RPD problem. Although originally developed for the regression case with little or no replication, we demonstrate here its extension to replicated designs in which the variance may be modeled directly.

Assuming that the process mean and variance functions can be expressed as functions with two components, a user supplied parametric component and a

"lack of fit" component, the dual model can be written as:

$$\text{Means Model:} \qquad \bar{y}_i = h\left(\mathbf{x}_i'; \boldsymbol{\beta}\right) + f\left(\mathbf{x}_i\right) + g^{1/2}\left(\mathbf{x}_i^{*\prime}; \boldsymbol{\gamma}\right)\varepsilon_i,$$

$$\text{Variance Model:} \qquad \ln(s_i^2) = g^*\left(\mathbf{x}_i^{*\prime}; \boldsymbol{\gamma}\right) + l\left(\mathbf{x}_i^*\right) + \eta_i.$$

Regarding notation, $h\left(\mathbf{x}_i'; \boldsymbol{\beta}\right)$ and $g^*\left(\mathbf{x}_i^{*\prime}; \boldsymbol{\gamma}\right)$ denote the user specified parametric forms for the mean and variance functions [for purposes of discussion in this manuscript we assume $h\left(\mathbf{x}_i'; \boldsymbol{\beta}\right) = \mathbf{x}_i'\boldsymbol{\beta}$ and $g^*\left(\mathbf{x}_i^{*\prime}; \boldsymbol{\gamma}\right) = \mathbf{x}_i^{*\prime}\boldsymbol{\gamma}$]. The "lack of fit" components for the mean and variance functions, $f\left(\mathbf{x}_i\right)$ and $l\left(\mathbf{x}_i^*\right)$, respectively, represent the portions of the mean and variance functions which cannot be captured parametrically. The only assumptions placed on f and l is that they are reasonably smooth functions.

As mentioned earlier, MBS demonstrate that MRR2 performs better than MRR1 overall. However, for variance modeling, there is nothing to guarantee positive fits in the MRR2 approach. Hence for the dual modeling problem, we suggest the use of MRR1 for estimation of the variance and MRR2 for estimation of the means model. Thus, the following algorithm can be used to find the DMRR estimates:

Step 1: Fit the variance model, $\ln(s_i^2) = \mathbf{x}_i^{*\prime}\boldsymbol{\gamma} + l\left(\mathbf{x}_i^*\right) + \eta_i$, via model robust regression 1 (MRR1). MRR1 yields the variance model robust regression (VMRR) estimates:

$$\begin{aligned}
\widehat{Var[\mathbf{y}]}^{(VMRR)} &= \exp\left[\lambda_\sigma \hat{\mathbf{y}}^{*(LLR)} + (1-\lambda_\sigma)\,\hat{\mathbf{y}}^{*(OLS)}\right] \\
&= \exp\left[\lambda_\sigma \mathbf{H}_\sigma^{(LLR)}\mathbf{y}^* + (1-\lambda_\sigma)\,\mathbf{H}_\sigma^{(OLS)}\mathbf{y}^*\right] \\
&= \exp\left[\lambda_\sigma \mathbf{H}_\sigma^{(VMRR)}\mathbf{y}^*\right],
\end{aligned}$$

where $\lambda_\sigma \in [0,1]$ is the variance model mixing parameter and $\mathbf{H}_\sigma^{(VMRR)}$ is the smoother matrix for the VMRR fit to the log transformed sample variances.

Step 2: Use $\hat{\sigma}_i^2 = \exp\left(\mathbf{h}_{i,\sigma}^{(VMRR)\prime}\mathbf{y}^*\right)$ as the estimated variances to compute the estimated variance-covariance matrix for the means model, $\hat{\mathbf{V}} = diag\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_d^2\right)$, where $\mathbf{h}_{i,\sigma}^{(VMRR)}$ is the i^{th} row of $\mathbf{H}_\sigma^{(VMRR)}$.

Step 3: Use $\hat{\mathbf{V}}^{-1}$ as the estimated weight matrix to obtain the parametric estimate of the means model via estimated weighted least squares (EWLS). EWLS yields $\hat{\boldsymbol{\beta}}^{(EWLS)} = \left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\bar{\mathbf{y}}$ and $\widehat{E[y_i]}^{(EWLS)} = \mathbf{x}_i'\hat{\boldsymbol{\beta}}^{(EWLS)} = \mathbf{H}_\mu^{(EWLS)}\bar{\mathbf{y}}$.

Step 4: Form the residuals from the EWLS fits found in Step 3, $\mathbf{r} = \bar{\mathbf{y}} - \widehat{E[\mathbf{y}]}^{(EWLS)}$, and perform local linear regression on the residuals. LLR

yields $\hat{\mathbf{r}} = \mathbf{H}_r^{(LLR)}\mathbf{r}$, where $\mathbf{H}_r^{(LLR)}$ is the smoother matrix for the LLR fit to the residuals from the EWLS fit to the means model.

Step 5: Obtain the means model robust regression (MMRR) estimates via MRR2 as:

$$\begin{aligned}\widehat{E[\mathbf{y}]}^{(MMRR)} &= \widehat{E[\mathbf{y}]}^{(EWLS)} + \lambda_\mu \hat{\mathbf{r}} \\ &= \left[\mathbf{H}_\mu^{(EWLS)} + \lambda_\mu \mathbf{H}_r^{(LLR)}\left(\mathbf{I} - \mathbf{H}_\mu^{(EWLS)}\right)\bar{\mathbf{y}}\right], \\ &= \mathbf{H}_\mu^{(MMRR)}\bar{\mathbf{y}}\end{aligned}$$

where $\lambda_\mu \in [0, 1]$ is the means model mixing parameter.

For the nonparametric estimates, the bandwidths, b_μ and b_σ, will be chosen as the values that minimize PRESS**. The mixing parameters, λ_σ and λ_μ, for the variance and mean fits, respectively, will be chosen via the asymptotically optimal expressions for MRR1 and MRR2 as given in (9) and (10), respectively, with $\hat{\mathbf{y}}^{(OLS)}$ replaced with $\hat{\mathbf{y}}^{(EWLS)}$. Similar to the parametric and nonparametric approaches, once estimates of the mean and variance functions have been calculated, a squared error loss approach will be used for process optimization. Furthermore, as in the nonparametric approach, the genetic algorithm will be used for optimization since the estimates of the mean and variance functions do not take on closed form expressions.

4 The Printing Ink Example

The Box and Draper (1987) printing ink study has been analyzed throughout the RPD literature [see for example Vining and Myers (1990) and VB]. The purpose of the study was to examine the effect of three factors, speed (x_1), pressure (x_2), and distance (x_3), on a printing machine's ability to apply ink to package labels. The experiment used a 3^2 complete factorial with three replicates at each design point. The goal of the study was to find an optimal location where the process variance is minimized and the process mean achieves a target value of 500. Table 1 provides the results of the experiment. Note that two locations ($i = 10$ and 14) have a sample standard deviation of zero. Thus, to accommodate the log transformation for the variance model, we will replace the observed sample variances, s_i^2, with $s_i^2 + 1$. For our purposes, we will assume that the user has specified a first-order model for the log transformed variance model and a second-order model for the mean.

The nonparametric and semi-parametric approaches involve the choice of an appropriate global bandwidth for the kernel function. Using PRESS**, we obtain a bandwidth of 0.63 for the variance model and 0.52 for the means model in the nonparametric approach. These bandwidths meet the recommendations of AP. A bandwidth of 0.51 was chosen for the nonparametric smooth of the EWLS residuals in the semi-parametric fit to the mean.

Furthermore, the semi-parametric approach involves the choice of appropriate mixing parameters, λ_σ and λ_μ. For the variance model, the asymptotically optimal data driven mixing parameter is found to be 0.6812. This value suggests that there is a moderate amount of variance model misspecification; that is, there are some trends in the data that the parametric model cannot adequately capture. The asymptotically optimal data driven mixing parameter for the means model is found to be 1.0. Therefore, the addition of the entire nonparametric residual fit provides necessary correction to the parametric means model.

Using the genetic algorithm with the squared error loss (SEL) objective function as given in (5), we obtain the optimal factor settings displayed in Table 2. Notice that all three approaches suggest an x_1 value of 1.000. Also the nonparametric and semi-parametric methods both recommend an x_2 value of 1.000. The most dramatic difference in the solutions is seen in the x_3 coordinate. While all three methods suggest negative values for the factor, the actual settings for x_3 vary greatly. Coincidentally, x_3 is the most significant factor in the parametric variance model.

In Table 2, we also see that the optimal factor settings for the semiparametric approach ($x_1 = 1.000, x_2 = 0.543, x_3 = -0.382$) yield an estimated process mean of 497.6278, which is closest to the target value of 500, and an estimated process variance of 1019.523, the lowest among the three methods. As a result, the semi-parametric approach also results in the lowest estimated SEL value. Thus, we would conclude that this new method performs better than its parametric and nonparametric counterparts. Ultimately, the only way to determine which approach gives the best optimization results is to perform a confirmatory experiment. Unfortunately, we cannot do so for this example, but we can use simulations to compare the three approaches in general.

5 Simulations

In the printing ink example, the semi-parametric fit was observed to be superior to its parametric and nonparametric counterparts in terms of SEL. In this section, we compare the three methods more generally in terms of fit via a simulation study. The performance of the semi-parametric approach will be compared to the parametric and nonparametric approaches in four scenarios: the researcher correctly specifies the forms of both the underlying mean and variance functions, the researcher correctly specifies the form of the underlying variance function but misspecifies the means model, the researcher correctly specifies the form of the means model but misspecifies the variance model, and the researcher incorrectly specifies the forms of both the underlying mean and variance functions. For each scenario, Monte Carlo simulations will be used to generate 500 data sets, each of which are based on the following underlying dual

model:

$$\begin{aligned} y_i &= 20 - 10x_{1i} - 25x_{2i} - 15x_{1i}x_{2i} + 20x_{1i}^2 + 50x_{2i}^2 \\ &\quad +\gamma_\mu \left[10\sin\left(4\pi x_{1i}\right) + 10\cos\left(4\pi x_{2i}\right) + 10\sin\left(4\pi x_{1i}x_{2i}\right)\right] \\ &\quad +g^{1/2}\left(\mathbf{x}_i^*\right)\varepsilon_i \end{aligned} \tag{11}$$

$$\ln\left(\sigma_i^2\right) = g^*\left(\mathbf{x}_i^*\right) = 1.5 - x_{1i} + 1.5x_{2i} + \gamma_\sigma\left[-4x_{1i}x_{2i} + 2x_{1i}^2 + x_{2i}^2\right], \tag{12}$$

where $\varepsilon_i \sim \mathrm{N}(0,1)$, γ_μ represents the means model misspecification parameter, and γ_σ represents the variance model misspecification parameter. As in the printing ink example, we assume a full second-order model is specified by the user for the mean and a first-order model specification for the log transformed sample variances. In all scenarios, factors x_1 and x_2 have four levels with values are taken to be 0, 1/3, 2/3, and 1 for each factor. The data will be generated as if a 4^2 complete factorial experiment was run with three replicates at each design point for a total of 16 design points and 48 experimental runs. As the values of γ_μ and γ_σ increase, the amount of misspecification increases in the means and variance model, respectively. Five degrees of means model misspecification will be studied (γ_μ = 0.00, 0.25, 0.50, 0.75, and 1.00), and five degrees of variance model misspecification will be studied (γ_σ = 0.00, 0.25, 0.50, 0.75, and 1.00). The R language is used for computations and simulations.

Figure 2 shows the response surface for the true underlying means model when $\gamma_\mu = 0.00$, and the response surfaces of the mean function for the varying degrees of model misspecification (γ_μ = 0.25, 0.50, 0.75, and 1.00) appear in Figures 3 through 6, respectively. Notice that as γ_μ increases, the curvature of the mean surface becomes much more pronounced. Figure 7 shows the response surface for the true underlying variance model when $\gamma_\sigma = 0.00$, and the response surfaces of the variance function for the varying degrees of model misspecification (γ_σ = 0.25, 0.50, 0.75, and 1.00) appear in Figures 8 through 11, respectively. Again notice that as γ_σ increases, the curvature of the variance surface becomes much more pronounced.

Comparisons will be based on the Monte Carlo simulated integrated mean squared error for the mean and variance estimates given by:

$$SIMSEM = \frac{\sum asem}{500}, \qquad asem = \frac{\sum\left(E[y_i] - \hat{y}_i\right)^2}{1600},$$

$$SIMSEV = \frac{\sum asev}{500}, \qquad asev = \frac{\sum\left(\sigma_i^2 - \hat{\sigma}_i^2\right)^2}{1600},$$

where 'asem' and 'asev' denote the average squared error for the mean and variance functions, respectively, across 1600 $\mathbf{x}_0$ locations (based on a 40×40 uniform grid of points in the regressor space) for each of the 500 simulated data sets and $E[y_i]$ and σ_i^2 are the true underlying mean and variance models, respectively.

5.1 Means and Variance Models Correctly Specified

For the scenario in which the researcher correctly specifies the forms of both the underlying mean and variance functions, we would expect the parametric approach to be superior for both models. In this scenario, the means model is given by (11) with $\gamma_\mu = 0.00$ and the variance model is given by (12) with $\gamma_\sigma = 0.00$. The first row of Table 3 and the first row of Table 4 provide a comparison of the three approaches based on the SIMSE values for this scenario. Regarding the estimated mean, Table 3 shows that the parametric approach performs best as it yields an SIMSEM value of 0.4335. The semi-parametric approach is a close second with a value of 0.6151, whereas the nonparametric fit is much worse with a SIMSEM value of 7.9748. As for the estimated variance, Table 4 shows that the parametric approach performs best with a SIMSEV value of 17.9190. Note when $\gamma_\mu = \gamma_\sigma = 0.00$, one would expect the parametric to do best. It is interesting, however, that the semi-parametric method is a close competitor to the parametric approach in this scenario.

5.2 Variance Model Correctly Specified, Means Model Misspecified

Rows 6, 11, 16, and 21 of Table 3 provide the SIMSEM values for the scenario in which the researcher correctly specifies the variance model but misspecifies the mean model (i.e., $\gamma_\sigma = 0.00$ and $\gamma_\mu > 0.00$). The semi-parametric approach outperforms its parametric and nonparametric counterparts for moderate levels of mean misspecification (i.e., $\gamma_\mu \leq 0.50$) and competes closely with the nonparametric approach for more extreme levels of misspecification (i.e., $\gamma_\mu \geq 0.75$). Note that as γ_μ increases from zero, the parametric fit to the mean becomes much less competitive. The semi-parametric fit to the mean is at least highly competitive or superior to the other methods for the entire range of γ_μ.

5.3 Means Model Correctly Specified, Variance Model Misspecified

Rows 2 through 5 of Table 3 show the SIMSEM values for the scenario in which the researcher correctly specifies the means model (i.e., $\gamma_\mu = 0.00$) but misspecifies the form of the variance (i.e., $\gamma_\sigma > 0.00$). Even though the estimated variances are incorporated in the weighted estimation of the mean, variance misspecification has little impact on the estimated mean. Table 3 shows that the SIMSEM values consistently increase for all three approaches as the variance misspecification increases. Nonetheless, the parametric and semi-parametric approaches still perform best when the means model is correctly specified, whereas the nonparametric fit is consistently subpar. As for the estimated variance, rows 2 through 5 of Table 4 provide the SIMSEV values for each of the methods. The semi-parametric approach performs best when there is moderate to large variance misspecification (i.e., $\gamma_\sigma > 0.50$) and is competitive for lower degrees of misspecification.

5.4 Means and Variance Models Misspecified

The remainder of the results in Table 3 provide a comparison for the scenario in which the researcher misspecifies the forms of both the underlying mean and variance functions (i.e., $\gamma_\mu > 0.00$ and $\gamma_\sigma > 0.00$). Again, it appears that variance misspecification has little impact of mean estimation. The biggest impact on mean estimation is the user's specification of the underlying means model. Overall, the semi-parametric approach performs best if there is small to moderate mean misspecification and/or moderate to large variance misspecification. It is important to note that even when the semi-parametric approach is not best in terms of SIMSEM or SIMSEV, its fit is highly competitive with the superior approach. While the parametric and nonparametric methods are best in some situations (parametric SIMSEM and SIMSEV values best when $\gamma_\mu = \gamma_\sigma = 0.00$ and nonparametric SIMSEM values best when $\gamma_\mu \geq 0.75$), these methods are noticeably inferior when there are small to moderate levels of misspecification.

6 Conclusions

The dual model response surface approach to RPD has been shown to work well when the variance of response is not constant over the experimental region and can be successfully modeled using regression methods. One drawback, however, is that optimization depends too heavily on the assumption of well estimated models for the process mean and variance, and it is often the case that user specified parametric models are not flexible enough to adequately model the process mean and variance. VB and AP suggest the use of nonparametric smoothing when the user is unable to specify the explicit forms for the mean and/or variance functions. However, in small sample settings, which are customary for response surface experiments, the nonparametric approach often produces estimates that are highly variable. Consequently, we suggest a semi-parametric approach, which combines the user's specified parametric model with a nonparametric fit, to provide better estimates of both the process mean and variance.

Using the Box and Draper (1987) printing ink data to compare the parametric, nonparametric, and semi-parametric approaches, we find that the semi-parametric approach performs best in terms of SEL. The optimization based on the semi-parametric approach recommends control factor settings which result in the estimated mean being closer to target as well as the smallest estimate of process variance. As previously mentioned, confirmatory experiments are necessary to prove which approach performs best for this example; however, the semi-parametric approach produces better results than its parametric and nonparametric counterparts previously seen in the literature.

To compare the three approaches more generally, a simulation study was conducted. Variance model misspecification was observed to have little impact on the quality of the estimated mean. If the user correctly specifies the mean and variance models, the parametric approach is best with the semi-

parametric method a close second. The nonparametric method, on the other hand, is vastly inferior in terms of SIMSEM. The nonparametric method, while best for large degrees of mean misspecification, is only slightly better than the proposed semi-parametric approach. When the mean is misspecified, the parametric method is clearly inferior. For small to moderate mean misspecification, the semi-parametric method is always superior. Since, in practice, one never knows if the forms of the underlying models have been correctly specified, we advocate a method that performs consistently well over all degrees of potential misspecification. The semi-parametric method is the only one which consistently performs well. The proposed semi-parametric approaches for mean and variance modeling are easy to implement and the R code is available upon request from the authors.

References

Aitkin, M., 1987. Modeling variance heterogeneity in Normal regression using GLIM. Applied Statistics 36, 332-339.

Anderson-Cook, C.M., Prewitt, K., 2005. Some guidelines for using nonparametric methods for modeling data from response surface designs. Journal of Modern Applied Statistical Methods 4, 106-119.

Bartlett, M.S., Kendall, D.G., 1946. The statistical analysis of variance heterogeneity and the logarithmic transformation. Journal of the Royal Statistical Society Series B 8, 128-138.

Box, G.E.P., Draper, N.R., 1987. Empirical Model Building and Response Surface. Wiley, New York.

Box, G.E.P., Meyer, R.D., 1986. Dispersion effects from fractional designs. Technometrics 28, 19-27.

Einsporn, R., Birch, J.B., 1993. Model robust regression: using nonparametric regression to improve parametric regression analyses. Technical Report 93-5. Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA.

Fan, J., Heckman, N.E., Wand, M.P., 1995. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. Journal of the American Statistical Association 90, 141-150.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and Its Applications. Chapman & Hall, London.

Gasser, T., Müller, H.G., 1984. Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171-185.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.

Grego, J.M., 1993. Generalized linear models and process variation. Journal of Quality Technology 25, 288-295.

Härdle, W., 1990. Applied Non-Parametric Regression. Cambridge University Press, Cambridge.

Härdle, W., Müller, M., Sperlichm, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer, Berlin.

Haupt, R.L., Haupt, S.E., 2004. Practical Genetic Algorithms. Wiley, New York.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Lee, Y., Nelder, J.A., 2003. Robust design via generalized linear models. Journal of Quality Technology 35, 2-12.

Lin, X., Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association 95, 520-534.

Mays, J., Birch, J.B., Einsporn, R., 2000. An overview of model-robust regression. Journal of Statistical Computation and Simulation 66, 79-100.

Mays, J., Birch, J.B., Starnes, B.A., 2001. Model robust regression: combining parametric, nonparametric, and semiparametric methods. Journal of Nonparametric Statistics 13, 245-277.

Myers, R.H., Montgomery, D.C., 2002. Response Surface Methodology: Process and Product Optimization Using Design Experiments, 2nd ed. Wiley, New York.

Myers, W.R., Brenneman, W.A., Myers, R.H., 2005. A dual-response approach to robust parameter design for a generalized linear model. Journal of Quality Technology 37, 130-138.

Nadaraya, E., 1964. On estimating regression. Theory of Probability and Its Applications 9, 141-142.

Priestley, M.B., Chao, M.T., 1972. Non-parametric function fitting. Journal of the Royal Statistical Society Series B 34, 385-392.

Robinson, T.J., Birch, J.B., 2002. Dual model robust regression: robust to model misspecification. Technical Report 02-2. Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press, Cambridge.

Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer-Verlag, New York.

Vetterling, W.T., Teukolsky, S.A., Press, W.H., Flannery, B.P., 1992. Numerical Recipes: Example Book (Fortran), 2nd ed. Cambridge University Press, Cambridge.

Vining, G.G., Bohn, L.L., 1998. Response surfaces for the mean and variance using a nonparametric approach. Journal of Quality Technology 30, 282-291.

Vining, G.G., Myers, R.H., 1990. Combining Taguchi and response surface philosophies: a dual response approach. Journal of Quality Technology 22, 38-45.

Watson, G., 1964. Smoothing regression analysis. Sankhya Series A 26, 359-372.

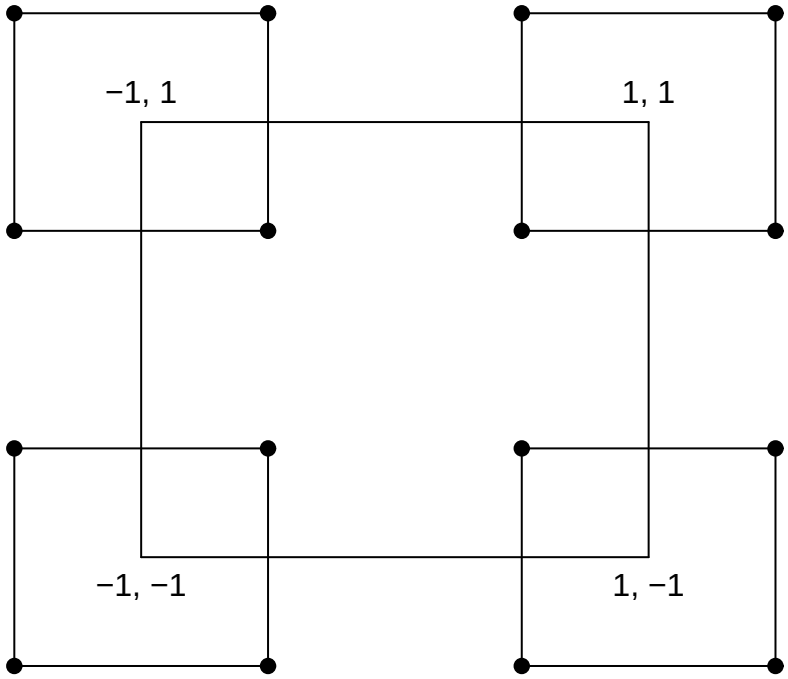

Figure 1: The $2^2 \times 2^2$ crossed array.

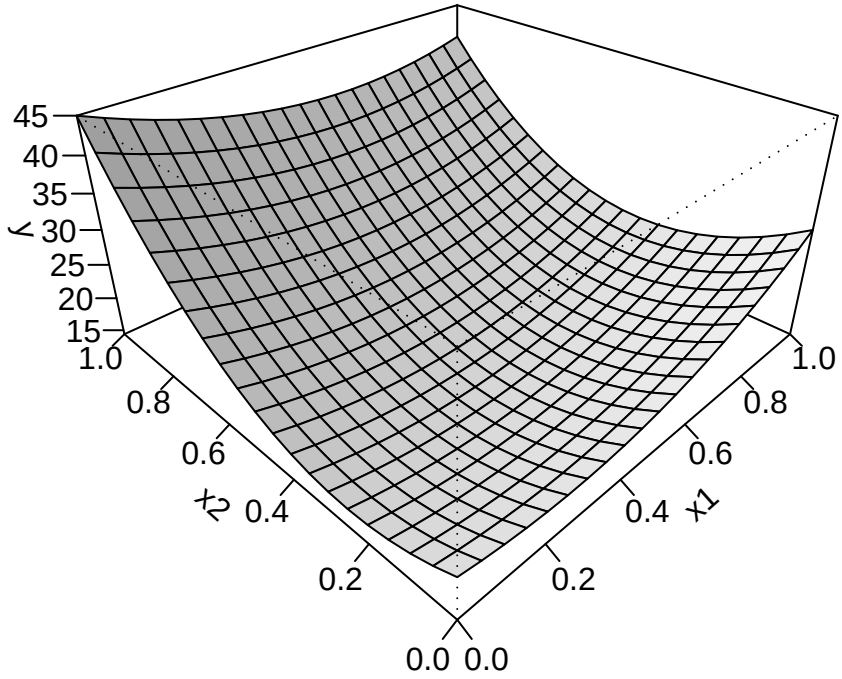

Figure 2: Response surface for the true underlying means model when $\gamma_\mu = 0.00$.

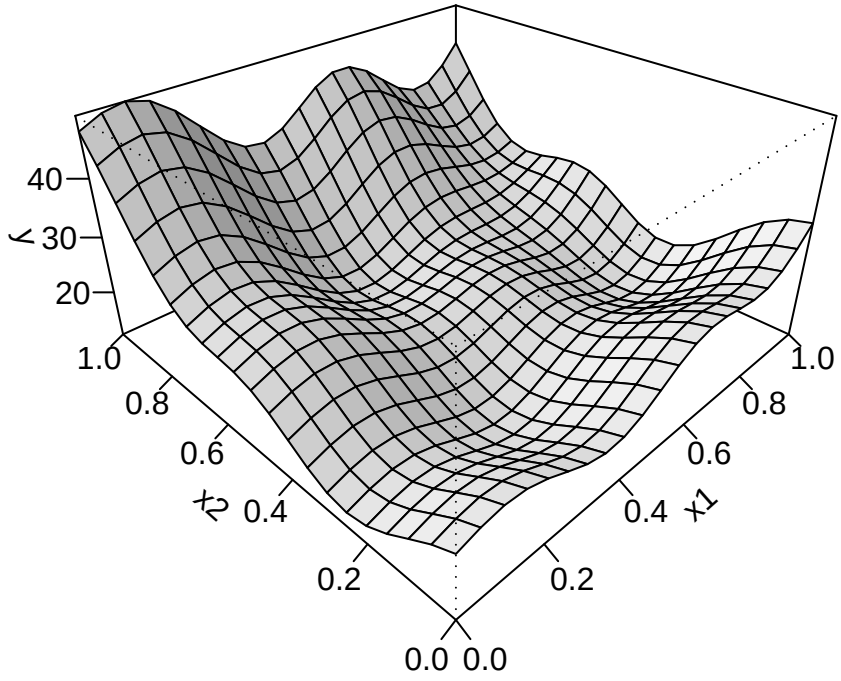

Figure 3: Response surface for the means model when $\gamma_\mu = 0.25$.

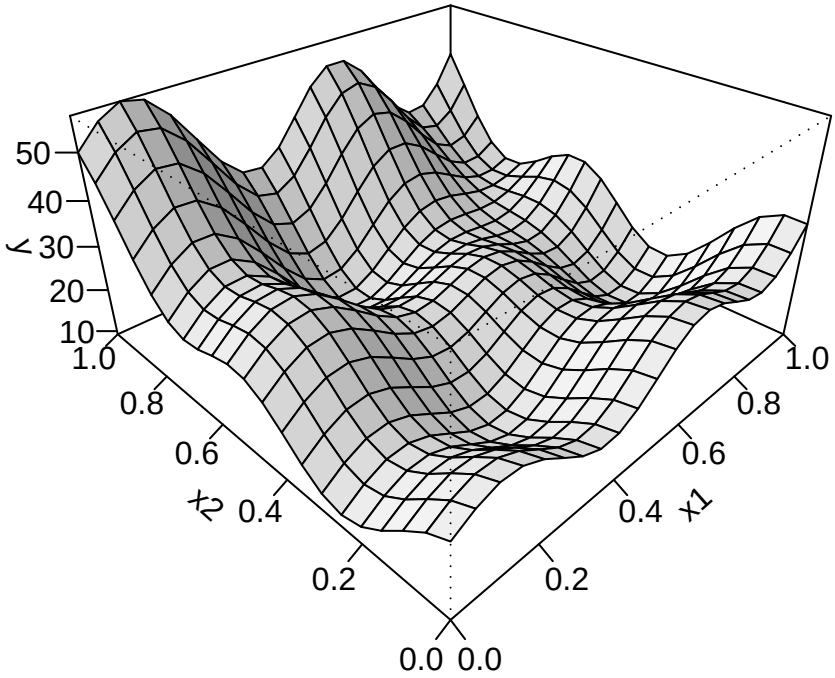

Figure 4: Response surface for the means model when $\gamma_\mu = 0.50$.

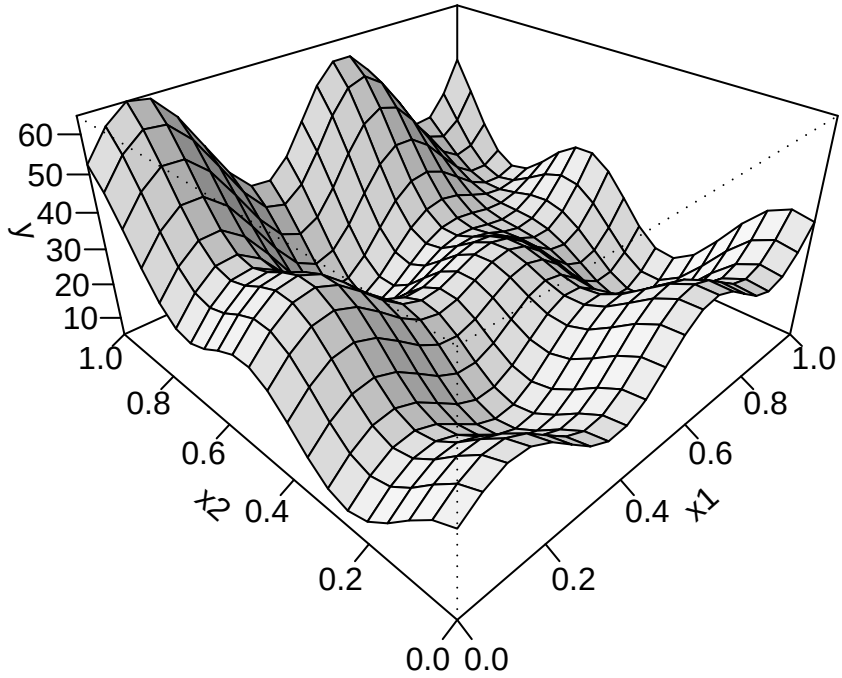

Figure 5: Response surface for the means model when $\gamma_\mu = 0.75$.

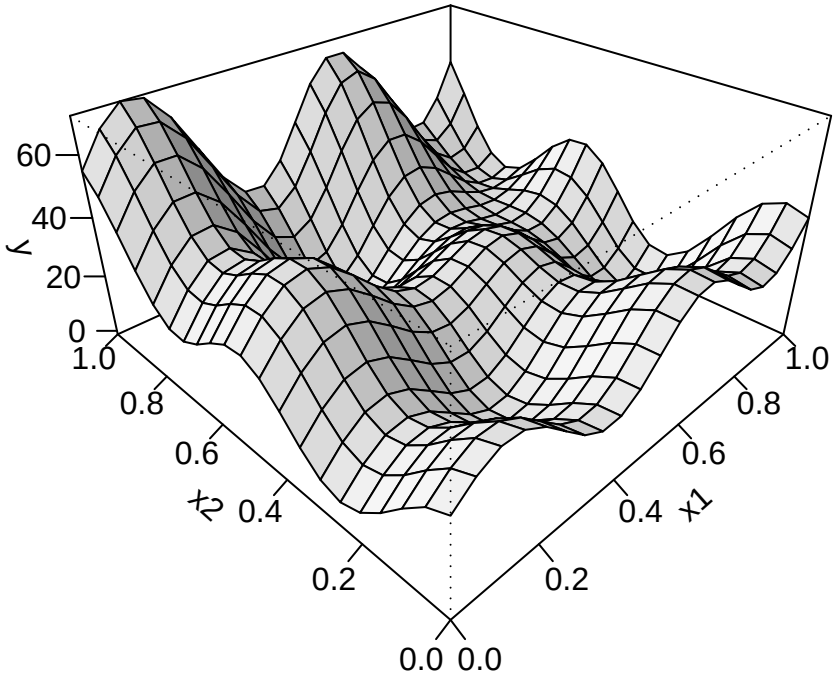

Figure 6: Response surface for the means model when $\gamma_\mu = 1.00$.

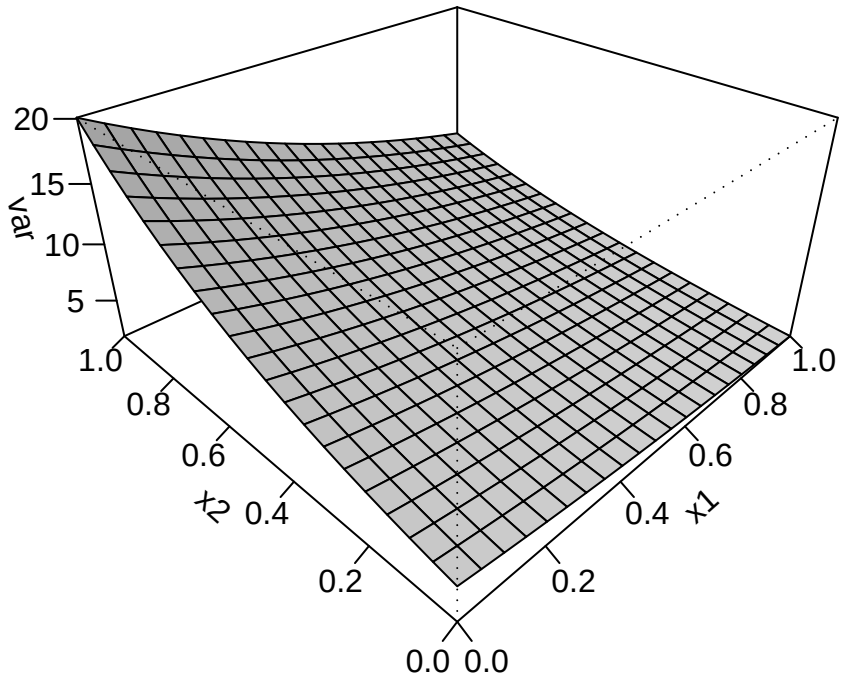

Figure 7: Response surface for the true underlying variance model when $\gamma_\sigma = 0.00$.

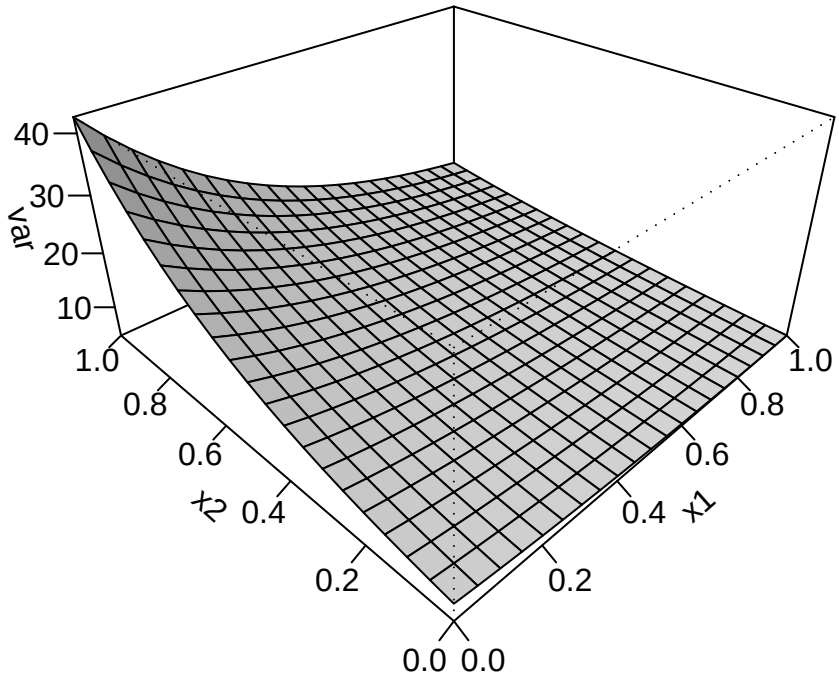

Figure 8: Response surface for the variance model when $\gamma_\sigma = 0.25$.

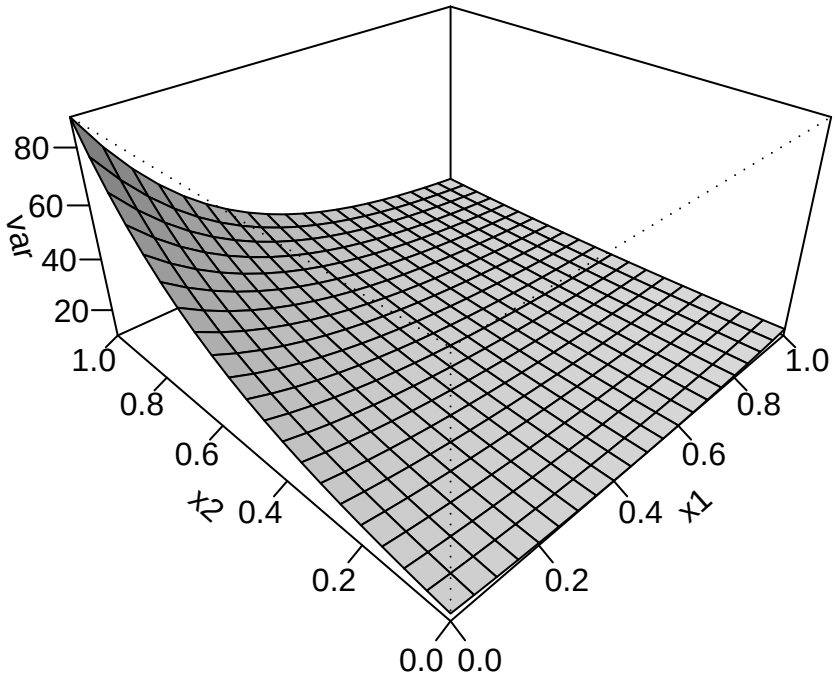

Figure 9: Response surface for the variance model when $\gamma_\sigma = 0.50$.

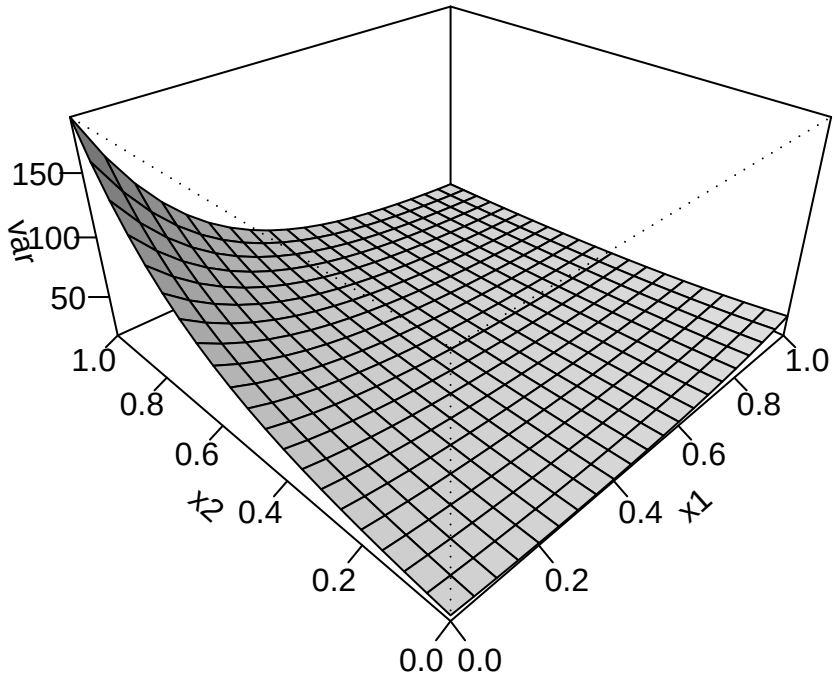

Figure 10: Response surface for the variance model when $\gamma_\sigma = 0.75$.

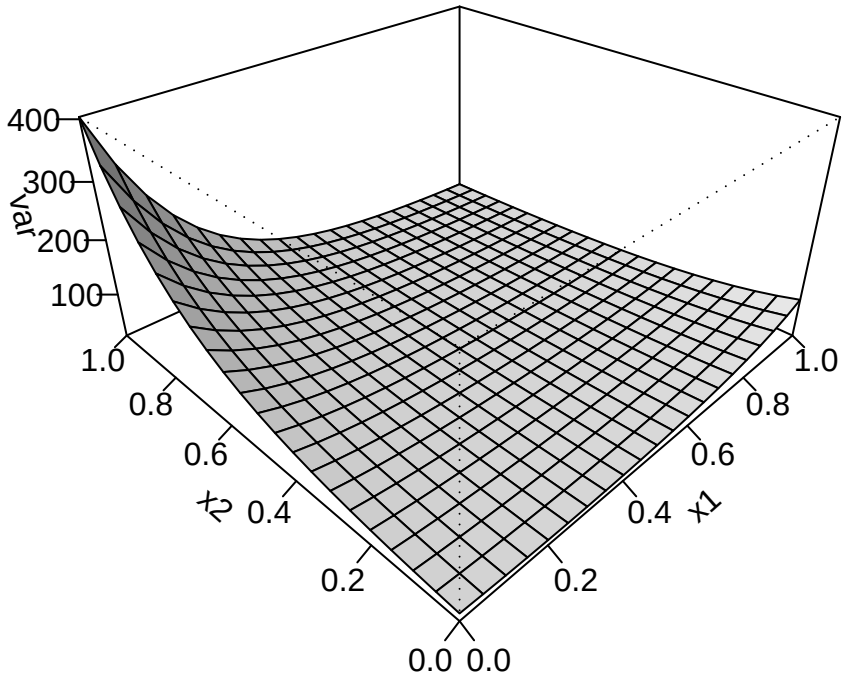

Figure 11: Response surface for the variance model when $\gamma_\sigma = 1.00$.

i	x_{1i}	x_{2i}	x_{3i}	y_{1i}	y_{2i}	y_{3i}	$\bar{y}_i$	s_i
1	-1	-1	-1	34	10	28	24.00	12.49
2	0	-1	-1	115	116	130	120.33	8.39
3	1	-1	-1	192	186	263	213.67	42.83
4	-1	0	-1	82	88	88	86.00	3.46
5	0	0	-1	44	178	188	136.67	80.41
6	1	0	-1	322	350	350	340.67	16.17
7	-1	1	-1	141	110	86	112.33	27.57
8	0	1	-1	259	251	259	256.33	4.62
9	1	1	-1	290	280	245	271.67	23.63
10	-1	-1	0	81	81	81	81.00	0.00
11	0	-1	0	90	122	93	101.67	17.67
12	1	-1	0	319	376	376	357.00	32.91
13	-1	0	0	180	180	154	171.33	15.01
14	0	0	0	372	372	372	372.00	0.00
15	1	0	0	541	568	396	501.67	92.50
16	-1	1	0	288	192	312	264.00	63.50
17	0	1	0	432	336	513	427.00	88.61
18	1	1	0	713	725	754	730.67	21.08
19	-1	-1	1	364	99	199	220.67	133.82
20	0	-1	1	232	221	266	239.67	23.46
21	1	-1	1	408	415	443	422.00	18.52
22	-1	0	1	182	233	182	199.00	29.44
23	0	0	1	507	515	434	485.33	44.64
24	1	0	1	846	535	640	673.67	158.21
25	-1	1	1	236	126	168	176.67	55.51
26	0	1	1	660	440	403	501.00	138.94
27	1	1	1	878	991	1161	1010.00	142.45

Table 1: The printing ink data.

Approach	x_1	x_2	x_3	$\widehat{E[y_i]}$	$\widehat{Var[y_i]}$	$\widehat{SEL}$
Parametric	1.000	0.358	-0.112	497.619	1723.693	1729.363
Nonparametric	1.000	1.000	-0.352	496.866	1088.455	1098.276
Semi-parametric	1.000	1.000	-0.522	**497.629**	**1019.523**	**1025.150**

Table 2: Comparison of the recommended optimal factor settings for the printing ink data. Best values in bold.

γ_μ	γ_σ	Parametric	Nonparametric	Semi-parametric
	0.00	**0.4335**	7.9748	0.6151
	0.25	**0.4503**	8.2321	0.6520
0.00	0.50	**0.4806**	8.4893	0.7148
	0.75	**0.5285**	8.6231	0.8200
	1.00	**0.6001**	8.6842	0.9801
	0.00	11.7193	16.2605	**11.1885**
	0.25	11.6521	16.3836	**10.9981**
0.25	0.50	11.6099	16.6595	**10.8863**
	0.75	11.5983	17.1650	**10.8326**
	1.00	11.9108	18.0681	**11.1520**
	0.00	44.5181	42.9345	**40.9599**
	0.25	44.4003	42.9855	**40.4008**
0.50	0.50	44.8819	43.4888	**40.1894**
	0.75	44.8342	43.8464	**39.9063**
	1.00	44.8238	45.4837	**39.8345**
	0.00	98.8521	**87.5581**	90.1062
	0.25	98.6950	**87.5757**	89.9734
0.75	0.50	98.5082	**87.7151**	87.9934
	0.75	98.3134	88.2083	**87.1736**
	1.00	98.1409	90.1727	**86.6895**
	0.00	174.7138	**150.1532**	158.6082
	0.25	174.5361	**150.1094**	156.7779
1.00	0.50	174.2771	**150.2388**	155.0822
	0.75	175.6504	**151.2052**	154.1700
	1.00	174.4280	153.0754	**152.1491**

Table 3: Simulated integrated mean squared error values for the means model (SIMSEM) for 500 Monte Carlo runs. Best values in bold.

γ_σ	Parametric	Nonparametric	Semi-parametric
0.00	**17.9190**	19.8901	19.5246
0.25	**17.7432**	19.8656	19.3408
0.50	20.0876	21.0396	**19.9436**
0.75	25.6025	23.6397	**21.6610**
1.00	35.5426	28.2999	**25.6706**

Table 4: Simulated integrated mean squared error values for the variance model (SIMSEV) for 500 Monte Carlo runs. Best values in bold.