ESTIMATION OF NONLINEAR SIMULATION METAMODELS USING CONTROL VARIATES

M. ISABEL REIS SANTOS AND ACÁCIO M. O. PORTA NOVA

Departmento de Matemática, Instituto Superior Técnico, Portugal

Departmento de Engenharia e Gestão, Instituto Superior Técnico, Portugal

The method of control variates has been intensively used for reducing the variance of estimated (linear) regression metamodels in simulation experiments. In contrast to previous studies, this paper presents a procedure for applying multiple control variates when the objective is to estimate and validate a *nonlinear* regression metamodel for a single response, in terms of selected decision variables. This procedure includes robust statistical regression techniques for estimation and validation. Assuming joint normality of the response and controls, confidence intervals and hypothesis tests for the metamodel parameters are obtained. Finally, results for measuring the efficiency of the use of control variates are discussed.

(SIMULATION; MODELING; NONLINEAR METAMODELS; VARIANCE REDUCTION)

1. Introduction

Computer simulation models are commonly used for estimating and validating metamodels. A simulation metamodel is a mathematical relationship between the input (input parameters or design variables) and the output (response) of the computer simulation model; see Barton (1992). If this auxiliary model is an accurate representation of the simulation model, it can be very useful for prediction and sensitivity analysis, since it uses fewer computer resources, when compared with the more time consuming and expensive simulation program. In order to improve the efficiency of metamodel estimation, it is common to use the control variates technique. This technique is one of the most widely used variance reduction methods, since it is not very difficult to implement, it is a general method and it does not alter the underlying stochastic process.

Many authors have studied the method of control variables in the context of linear metamodel estimation; see, for example, Nozari, Arnold and Pegden (1984), Porta Nova and Wilson (1989) and Shih and Song (1995). In particular, the polynomial form of the general linear regression model has been extensively analyzed. However, polynomials are unable to produce a global fit to curves of arbitrary shape. Moreover, in real-life systems nonlinearity is common and approximation using polynomials becomes unrealistic. Consequently, in these problems, polynomials often do not provide good fits—e.g., in problems involving queueing systems (Friedman and Friedman (1985)). An alternative that provides better and more realistic global fits is the use of statistical *nonlinear* regression techniques (Santos and Porta Nova (1999)). In this paper, we apply the method of control variates to the estimation and validation of a nonlinear metamodel of a single simulation response, expressed in terms of multiple inputs. In fact, the procedure presented here is a generalization of the work of Nozari, Arnold and Pedgen (1984) to nonlinear simulation metamodels.

This paper is organized as follows. In Section 2, we formulate the nonlinear metamodeling problems, both without and with control variates. In Section 3, we obtain some distribution-free results. In Section 4, we discuss the metamodel estimation problem under a joint normality assumption. The minimum variance ratio and the loss factor are also obtained. In Section 5, the methodology described in this paper is illustrated using a simple $M/M/1$ queueing system. Finally, Section 6 is reserved for conclusions.

2. Nonlinear Metamodel Estimation

2.1 *Nonlinear Metamodels*

Consider an experimental design consisting of n different design points, defined by the d decision variables $\{X_{il} : i = 1, \ldots, \mathsf{n}; l = 1, \ldots, \mathsf{d}\}$. For each design point, r independent replications of the simulation model are carried out and the experiment yields $\{(Y_{ij}, \mathbf{C}_{ij}) : i = 1, \ldots, \mathsf{n}, j = 1, \ldots, \mathsf{r}\}$, where Y_{ij} is the relevant system response and $\mathbf{C}_{ij}$ is a vector of q concomitant control variables, with a known mean. Without loss of generality, we assume $\mathcal{E}[\mathbf{C}_{ij}] = \mathbf{0}$, with $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$. Suppose that the simulation model (computer program) can be represented by the metamodel

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \epsilon_{ij}, \tag{1}$$

for $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$, where $\epsilon_{ij} \sim \mathrm{NID}\,(0, \sigma^2)$, with $\sigma > 0$, and $\boldsymbol{\theta}$ is an $\mathsf{m} \times 1$ vector of unknown parameters. Under mild regularity conditions, every nonlinear control variable scheme behaves asymptotically like a linear control variable scheme; see Glynn and Whitt (1989) and Loh (1994). As a result, we only consider linear schemes involving control variables. Thus, we assume that the error $\epsilon_{ij} = Y_{ij} - f(\mathbf{X}_{i.}, \boldsymbol{\theta})$, in problem (1), has a linear regression on the control vector $\mathbf{C}_{ij}$, with an unknown $\mathsf{q} \times 1$ vector of control coefficients $\boldsymbol{\delta}$ and an error ε_{ij}. This way, the simulation model can also be represented by the replicated simulation metamodel

$$Y_{ij} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \mathbf{C}_{ij}\boldsymbol{\delta} + \varepsilon_{ij}, \tag{2}$$

with $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$.

Let $\mathbf{Z}$ be the following random matrix:

$$\mathbf{Z} = \begin{bmatrix} Y_{11} & C_{111} & \ldots & C_{11\mathsf{q}} \\ \vdots & \vdots & & \vdots \\ Y_{\mathsf{n}1} & C_{\mathsf{n}11} & \ldots & C_{\mathsf{n}1\mathsf{q}} \\ \vdots & \vdots & & \vdots \\ Y_{1\mathsf{r}} & C_{1\mathsf{r}1} & \ldots & C_{1\mathsf{r}\mathsf{q}} \\ \vdots & \vdots & & \vdots \\ Y_{\mathsf{nr}} & C_{\mathsf{nr}1} & \ldots & C_{\mathsf{nrq}} \end{bmatrix} = [\mathbf{YC}]. \tag{3}$$

Assume that the row vector $\mathbf{Z}_{l.}$ is continuous and has the same probability density function for all $l = 1, \ldots, \mathsf{N} = \mathsf{nr}$, such that the following dispersion matrix exists:

$$\mathbf{\Sigma} = \mathcal{D}[\mathbf{Z}_{l.}] = \begin{bmatrix} \sigma^2 & \boldsymbol{\sigma}_{YC} \\ \boldsymbol{\sigma}_{CY} & \mathbf{\Sigma}_C \end{bmatrix}, \tag{4}$$

where $\sigma^2 = \mathrm{Var}[Y_{ij}]$. As a consequence, $\mathbf{\Sigma}_C = \mathcal{D}[\mathbf{C}_{ij}]$ is nonsingular and also positive definite with probability one, for all replications of all experimental points (Porta Nova and Wilson (1989)). The covariance vector between Y_{ij} and $\mathbf{C}_{ij}$, denoted by $\boldsymbol{\sigma}_{YC} = \mathcal{C}[Y_{ij}, \mathbf{C}_{ij}]$, is assumed to be constant for all $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$, with $\boldsymbol{\sigma}_{CY} = \boldsymbol{\sigma}_{YC}^T$.

In order to simplify the estimation procedure, instead of problems (1) and (2), we consider respectively the equivalent least squares problems, in which the individual observations, at each design point, are replaced by their averages across runs:

$$\bar{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\epsilon}_{i.}, \qquad i = 1, 2, \ldots, \mathsf{n}, \tag{5}$$

with $\bar{\epsilon}_{i.} \sim \mathrm{NID}(0, \sigma_{\bar{Y}}^2)$, $\sigma_{\bar{Y}}^2 = \mathrm{Var}[\bar{Y}_{i.}] = \sigma^2/\mathsf{r}$ and

$$\bar{Y}_{i.} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\mathbf{C}}_{i.}\boldsymbol{\delta} + \bar{\varepsilon}_{i.}, \qquad i = 1, 2, \ldots, \mathsf{n}, \tag{6}$$

where $\bar{\mathbf{C}}_{i.} = (\bar{C}_{i.1}, \ldots, \bar{C}_{i.\mathsf{q}})$, with $\bar{C}_{i.k} = \frac{1}{\mathsf{r}}\sum_{j=1}^{\mathsf{r}} C_{ijk}$.

2.2 *Objectives*

In this paper, two kinds of results are exposed:

(i) Assuming that the metamodel (2) is valid, we obtain the approximated minimum variance ratio, the nonlinear least squares estimator $\hat{\boldsymbol{\delta}}$ (for the true vector of control coefficients $\boldsymbol{\delta}$), and the corresponding controlled nonlinear least squares estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ (to estimate the true vector of metamodel coefficients $\boldsymbol{\theta}$);

(ii) Assuming that the response and the control variables have a joint multivariate normal distribution, we derive the approximated loss factor and we construct asymptotic confidence regions for $\boldsymbol{\theta}$. We also propose procedures for testing hypotheses about the metamodel parameters.

3. General Results on Metamodel Estimation

In this section, we present results on metamodel estimation using control variables that do not depend on the assumption of joint normality between the response and the control variables.

3.1 *Minimum Variance Ratio*

If the Jacobian matrix $\mathbf{F}$ of $\mathbf{f} = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}^*), \ldots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}^*)^T$ has full column rank m, then we apply result (12.21) of Seber and Wild (1989) to problem (5), obtaining the following asymptotic ordinary nonlinear least squares estimator of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^* + \left(\mathbf{F}^T\mathbf{F}\right)^{-1} \mathbf{F}^T \left[\bar{\mathbf{Y}} - \mathbf{f}\right], \tag{7}$$

where $\boldsymbol{\theta}^*$ is the exact value of $\boldsymbol{\theta}$ and $\bar{Y} = (\bar{\mathbf{Y}}_{1.}, \ldots, \bar{Y}_{\mathsf{n}.})^T$ (in order to simplify the notation, we use $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*)$ and $\mathbf{f} = \mathbf{f}(\boldsymbol{\theta}^*)$). The mean and the covariance of this estimator are obtained applying (12.23) of Seber and Wild (1989) to problem (5),

$$\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*, \qquad \mathcal{D}[\hat{\boldsymbol{\theta}}] = \frac{\sigma^2}{\mathsf{r}}(\mathbf{F}^T\mathbf{F})^{-1}. \tag{8}$$

When control variables are observed, and for a fixed vector of control coefficients $\boldsymbol{\phi}$, the least squares estimator of $\boldsymbol{\theta}$ is given approximately by

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} - \mathbf{f}], \tag{9}$$

where

$$\bar{\mathbf{C}} = \begin{bmatrix} \bar{C}_{1.1} & \ldots & \bar{C}_{1.\mathsf{q}} \\ \vdots & & \vdots \\ \bar{C}_{\mathsf{n}.1} & \ldots & \bar{C}_{\mathsf{n}.\mathsf{q}} \end{bmatrix}.$$

This estimator is obtained representing problem (6) in the form $\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi} = f(\mathbf{X}_{i.}, \boldsymbol{\theta}) + \bar{\varepsilon}_{i.}$ and then determining the ordinary nonlinear least squares estimator as in (7) (considering a fixed $\boldsymbol{\phi}$).

The approximation (9) is equivalent to

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T[\bar{\mathbf{Y}} - \mathbf{f}] - (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\bar{\mathbf{C}}\boldsymbol{\phi},$$

and, when control variables are not used, the least squares estimator is given by (7), so

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \hat{\boldsymbol{\theta}} - (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\bar{\mathbf{C}}\boldsymbol{\phi},$$

that is, $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \neq \hat{\boldsymbol{\theta}}$, or in other words, observing control variables with known means results in a different estimator of $\boldsymbol{\theta}$. As a consequence, if the random matrix has a probability density function, then using $\mathcal{E}[\bar{\mathbf{C}}] = \mathbf{0}$ and $\mathcal{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}^*$, we obtain

$$\mathcal{E}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \boldsymbol{\theta}^*. \tag{10}$$

To obtain the dispersion matrix $\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})]$, it is useful to write the estimator (9) in the form $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) \approx \mathbf{A}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}] + \mathrm{b}$, where $\mathbf{A} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T$ and $\mathrm{b} = \boldsymbol{\theta}^* - (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{f}$. Thus, we can write

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})] = \mathbf{A}\mathcal{D}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi}]\mathbf{A}^T, \tag{11}$$

where $\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\phi} = (\bar{Y}_{1.} - \bar{\mathbf{C}}_{1.}\boldsymbol{\phi}, \ldots, \bar{Y}_{\mathsf{n}.} - \bar{\mathbf{C}}_{\mathsf{n}.}\boldsymbol{\phi})^T$. Since this work is in the context of the method of independent replications, we have

$$\mathrm{Cov}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi}, \bar{Y}_{i'.} - \bar{\mathbf{C}}_{i'.}\boldsymbol{\phi}] = 0, \qquad i \neq i'. \tag{12}$$

Moreover,

$$\mathrm{Var}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\phi}] = \sigma^2_{\bar{Y}} + \boldsymbol{\phi}^T\boldsymbol{\Sigma}_{\bar{C}}\boldsymbol{\phi} - 2\boldsymbol{\phi}^T\boldsymbol{\sigma}_{\bar{C}\bar{Y}}. \tag{13}$$

The vector of control coefficients that minimizes this variance is given by $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1}_{\bar{C}}\boldsymbol{\sigma}_{\bar{C}\bar{Y}}$; see (8) and (9) of Lavenberg and Welch (1981). But $\boldsymbol{\Sigma}_{\bar{C}} = \boldsymbol{\Sigma}_C/\mathsf{r}$ and $\boldsymbol{\sigma}_{\bar{C}\bar{Y}} = \boldsymbol{\sigma}_{CY}/\mathsf{r}$, and as a result

$$\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1}_C\boldsymbol{\sigma}_{CY}. \tag{14}$$

Substituting (14) in (13), we obtain

$$\mathrm{Var}[\bar{Y}_{i.} - \bar{\mathbf{C}}_{i.}\boldsymbol{\delta}] = \frac{1}{\mathsf{r}}\tau^2, \tag{15}$$

where

$$\tau^2 = \sigma^2 - \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}^{-1}_C\boldsymbol{\sigma}_{CY}. \tag{16}$$

(12), (15) and (16) imply that $\mathcal{D}[\bar{\mathbf{Y}} - \bar{\mathbf{C}}\boldsymbol{\delta}] = \frac{1}{\mathsf{r}}(\sigma^2 - \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}^{-1}_C\boldsymbol{\sigma}_{CY})\mathbf{I}_{\mathsf{n}}$. Substituting this dispersion matrix in (11) and since $\mathbf{A} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T$, we have

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})] = \frac{1}{\mathsf{r}}(\sigma^2 - \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}^{-1}_C\boldsymbol{\sigma}_{CY})(\mathbf{F}^T\mathbf{F})^{-1}. \tag{17}$$

Using (8) and (17), we conclude that the maximum reduction in variance that is possible to obtain, with the use of control variables, is given approximately by the *minimum variance ratio*

$$\eta(\boldsymbol{\delta}) = \frac{|\mathcal{D}[\hat{\boldsymbol{\theta}}(\boldsymbol{\delta})]|}{|\mathcal{D}[\hat{\boldsymbol{\theta}}]|} \approx 1 - \rho^2_{YC}, \tag{18}$$

where $\rho^2_{YC} = \boldsymbol{\sigma}_{YC}\boldsymbol{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY}/\sigma^2$ is the multiple correlation coefficient between Y_{ij} and $\mathbf{C}_{ij}$, $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$. In general, as in the linear case, $\boldsymbol{\delta}$ is unknown, so it must be estimated and, as a consequence, the variance will increase. We will use a loss factor to quantify the percentage increase in variance when $\boldsymbol{\delta}$ must be estimated.

3.2 *Controlled Nonlinear Least Squares Estimator*

In order to obtain estimators for the unknown true parameters $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, we resort to the method of nonlinear least squares. Given appropriate regularity conditions (Seber and Wild (1989)), then for large N, the least squares estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ in (6) satisfy, approximately:

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) \approx \boldsymbol{\theta}^* + (\mathbf{F}^T\mathbf{F})^{-1}[\bar{\mathbf{Y}} - \mathbf{f} - \mathbf{C}\hat{\boldsymbol{\delta}}], \tag{19}$$

$$\hat{\boldsymbol{\delta}} \approx (\mathbf{C}^T\mathbf{P}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{P}[\bar{\mathbf{Y}} - \mathbf{f}], \tag{20}$$

where

$$\mathbf{P} = \mathbf{I}_{\mathsf{n}} - \mathbf{F}(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T. \tag{21}$$

These results are obtained as follows. The Taylor's series expansion of $f_i = f(\mathbf{X}_{i.}, \boldsymbol{\theta})$ about the point $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ yields $\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$. As a result, (6) becomes

$$\bar{\mathbf{Y}} - \mathbf{f}(\boldsymbol{\theta}^*) \approx \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \mathbf{C}\boldsymbol{\delta} + \bar{\varepsilon}. \tag{22}$$

Applying (6) and (8) of Searle (1971), page 342, to this (linearized) problem (22), we can write

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^* \approx (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T(\bar{Y} - \mathbf{f}) - (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{C}\hat{\boldsymbol{\delta}},$$

As a consequence, we obtain (19). Finally, (20) and (21) are obtained using (10) and (11) of Searle (1971), page 342.

3.3 *Validation Procedure*

After estimating the metamodel parameters, we must test the ability of the estimated metamodel to approximate the simulation model response (i.e., to ascertain if the estimated metamodel adequately fits the simulation data). To test the adequacy of the metamodel (6), we propose the following F-test for lack of fit (see Seber and Wild (1989), page 32):

$$F = \frac{(SSE - SSPE)/(\mathsf{n} - \mathsf{m} - \mathsf{q})}{SSPE/(\mathsf{N} - \mathsf{n})},$$

where, in this situation, $SSE = \sum_{i=1}^{\mathsf{n}}\sum_{j=1}^{\mathsf{r}}[Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) - \mathbf{C}_{ij}\hat{\boldsymbol{\delta}}]^2$ is the usual residual sum of squares and $SSPE = \sum_{i=1}^{\mathsf{n}}\sum_{j=1}^{\mathsf{r}}[Y_{ij} - \bar{Y}_{i.}]^2$ is the pure error sum of squares. When the metamodel is valid and if there exists a parameterization for which it can be adequately approximated by a linear model, then F is roughly distributed as an $F_{\mathsf{n}-\mathsf{m}-\mathsf{q},\mathsf{n}(\mathsf{r}-1)}$ distribution.

4. Results for Normal Nonlinear Metamodels

In order to simplify the presentation of the results, when the response and the control variables have a joint normal distribution, we will introduce some additional notation and hypotheses. Consider the models (5) and (6). Suppose that, for the l-th design point,

$$\mathbf{Z}_{l.} \sim \mathrm{N}_{\mathsf{q}+1}\left(\left(\mu_l, \mathbf{0}^T\right), \mathbf{\Sigma}\right), \tag{23}$$

where $\mathbf{\Sigma}$ is given by (4). As a result, the $\mathsf{N} \times (\mathsf{q}+1)$ random matrix, defined by (3), has a multivariate normal distribution

$$\mathbf{Z} \sim \mathrm{N}_{\mathsf{N},\mathsf{q}+1}\left(\boldsymbol{\mu}_Z, \Xi, \mathbf{\Sigma}\right), \tag{24}$$

with unknown $\mathbf{\Sigma}$ and $\mathcal{E}[\mathbf{Z}] = \boldsymbol{\mu}_Z = (\boldsymbol{\mu}_Y, \mathbf{0})$, where

$$\boldsymbol{\mu}_Y = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \dots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}))^T.$$

The dispersion matrix between the ith and kth rows is $\mathcal{D}[Z_{i.}, Z_{k.}] = \Xi_{ik}\mathbf{\Sigma}$ for $1 \leq i, k \leq \mathsf{N}$ and the dispersion matrix between the jth and lth columns is $\mathcal{D}[Z_{j.}, Z_{.l}] = \mathbf{\Sigma}_{jl}\Xi$ for $1 \leq j, l \leq \mathsf{q}+1$. Suppose also that Ξ and $\mathbf{\Sigma}$ are positive definite. Moreover, the rows of $\mathbf{Z}$ are mutually independent, since they correspond to independent executions of the simulation program. As a consequence, we consider $\Xi = \mathbf{I}_{\mathsf{N}}$ in the following development. If the $\mathsf{q} \times \mathsf{q}$ matrix $\mathbf{\Sigma}_C$ is positive definite, then the conditional distribution of $\mathbf{Y}$ given $\mathbf{C}$ is given by

$$\mathbf{Y}|\mathbf{C} \sim \mathrm{N}_{\mathsf{N}}\left(\boldsymbol{\mu}_{Y.C}, \tau^2\mathbf{I}_{\mathsf{N}}\right), \tag{25}$$

with τ^2 given by (16) and

$$\boldsymbol{\mu}_{Y.C} = \boldsymbol{\mu}_Y + \mathbf{C}\mathbf{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY}; \tag{26}$$

see Theorem 17.2-g), of Arnold (1981). As a consequence, conditioning on $\mathbf{C}$, we conclude that the correct metamodel is (2). In the following development, we will see that the asymptotic nonlinear least squares estimators are unbiased, both conditionally and unconditionally. Moreover, the approximated confidence region, for the true metamodel parameter vector, centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ will also be obtained.

4.1 *Distribution of the Controlled Estimator*

If $\mathbf{Z} \sim \mathrm{N}_{\mathsf{N},\mathsf{q}+1}(\boldsymbol{\mu}_Z, \mathbf{I}_{\mathsf{N}}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, then we will use the fact that the conditional distribution of $\mathbf{Y}$ given $\mathbf{C}$ is normal and given by (25). Conditioning on $\mathbf{C}$, we see that the correct nonlinear metamodel for a normal response is

$$\mathbf{Y} = \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) + \mathbf{C}\boldsymbol{\delta} + \mathbf{e}, \tag{27}$$

where $\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) = (f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{1.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}), \ldots, f(\mathbf{X}_{\mathsf{n}.}, \boldsymbol{\theta}))^T$ (a vector with N components) and $\mathbf{e} = (\varepsilon_{11}, \ldots, \varepsilon_{1\mathsf{r}}, \ldots, \varepsilon_{\mathsf{n}1}, \ldots, \varepsilon_{\mathsf{nr}})^T$. Using the Taylor's expansion in a neighborhood of $(\mathbf{X}, \boldsymbol{\theta}^*)$, we obtain

$$\tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}) \approx \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*) + \tilde{\mathbf{F}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*), \tag{28}$$

where $\tilde{\mathbf{F}}$ is the Jacobian matrix of $\tilde{\mathbf{f}}$, calculated at $(\mathbf{X}, \boldsymbol{\theta}^*)$. As a result, (27) can be rewritten as

$$\mathbf{G} \approx \tilde{\mathbf{F}}\boldsymbol{\lambda} + \mathbf{C}\boldsymbol{\delta} + \mathbf{e} \tag{29}$$

where $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$ and $\boldsymbol{\lambda} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$. Applying Searle (1971), page 342, to the problem (29), we obtain $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{G} - (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{C}\hat{\boldsymbol{\delta}}$, where $\hat{\boldsymbol{\delta}} = (\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{G}$ and $\tilde{\mathbf{P}} = \mathbf{I}_{\mathsf{N}} - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T$. That is,

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) \approx \mathbf{B}\mathbf{G} \qquad \text{with} \qquad \mathbf{B} = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T[\mathbf{I}_{\mathsf{N}} - \mathbf{C}(\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{P}}]. \tag{30}$$

Since (25) is verified, we have approximately $\mathbf{G}|\mathbf{C} \sim \mathrm{N}_{\mathsf{N}}(\boldsymbol{\mu}_{G.C}, \tau^2\mathbf{I}_{\mathsf{N}})$, where $\boldsymbol{\mu}_{G.C} = \boldsymbol{\mu}_{Y.C} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$. Using this result and (30), we can apply Theorem 17.2-d) of Arnold (1981) and so $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_{\mathsf{m}}(\mathbf{B}\boldsymbol{\mu}_{G.C}, \tau^2\mathbf{B}\mathbf{B}^T)$. Suppose that $\tilde{\mathbf{F}}$ has rank m, then $\tilde{\mathbf{P}}$ is an orthogonal projection of $\mathbb{R}^{\mathsf{N}}$ into $\mathcal{R}(\tilde{\mathbf{F}})^{\perp}$; see Seber (1989). Since $\tilde{\mathbf{P}}$ is an orthogonal projection, then $(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\tilde{\mathbf{P}} = \tilde{\mathbf{P}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} = 0$ and, as a consequence, $\mathbf{B}\mathbf{B}^T = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}[\mathbf{I}_{\mathsf{m}} + \tilde{\mathbf{F}}^T\mathbf{C}(\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}]$. Applying Nozari (1982), page 121, to the linearized problem (29), we obtain $\mathcal{E}[\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \boldsymbol{\lambda}$. As a result, $\mathbf{B}\boldsymbol{\mu}_{G.C} = \boldsymbol{\lambda}$ and we have $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_{\mathsf{m}}(\boldsymbol{\lambda}, \tau^2\mathbf{B}\mathbf{B}^T)$. But $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}}) = \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*$ and $\boldsymbol{\lambda} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$, so $\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \boldsymbol{\theta}$ (given $\mathbf{C}$, $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ is an unbiased estimator of $\boldsymbol{\theta}$), $\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \tau^2\mathbf{B}\mathbf{B}^T$ and

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C} \sim \mathrm{N}_{\mathsf{m}}\left(\boldsymbol{\theta}, \tau^2\mathbf{B}\mathbf{B}^T\right). \tag{31}$$

In order to construct a confidence region for $\boldsymbol{\theta}$ centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, we will present an estimator of τ^2. Applying Theorem 3.3 of Porta Nova (1985), to the problem (29), we obtain

$$\hat{\mathbf{e}}^T\hat{\mathbf{e}}|\mathbf{C} \sim \mathrm{W}_1(\mathsf{N} - \mathsf{m} - \mathsf{q}, \tau^2, 0),$$

that is, $\hat{\mathbf{e}}^T\hat{\mathbf{e}}|\mathbf{C} \sim \tau^2\chi^2_{\mathsf{N}-\mathsf{m}-\mathsf{q}}$, where $\hat{\mathbf{e}} = \mathbf{G} - \tilde{\mathbf{F}}\hat{\boldsymbol{\lambda}} - \mathbf{C}\hat{\boldsymbol{\delta}}$; see Theorem 17.6-b) of Arnold (1981). But $\mathbf{G} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \boldsymbol{\theta}^*)$ and $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$, then $\hat{\mathbf{e}} = \mathbf{Y} - \tilde{\mathbf{f}}(\mathbf{X}, \hat{\boldsymbol{\theta}}) - \mathbf{C}\hat{\boldsymbol{\delta}}$. As a result, given $\mathbf{C}$, an unbiased estimator for τ^2 is given by

$$\hat{\tau}^2|\mathbf{C} = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{\mathsf{N}-\mathsf{m}-\mathsf{q}} = \frac{1}{\mathsf{N}-\mathsf{m}-\mathsf{q}}\sum_{i=1}^{\mathsf{n}}\sum_{j=1}^{\mathsf{r}}[Y_{ij} - f(\mathbf{X}_{i.}, \hat{\boldsymbol{\theta}}) - \mathbf{C}_{ij}\hat{\boldsymbol{\delta}}]^2, \tag{32}$$

where

$$\hat{\tau}^2|\mathbf{C} \sim (\mathsf{N}-\mathsf{m}-\mathsf{q})^{-1}\tau^2\chi^2_{\mathsf{N}-\mathsf{m}-\mathsf{q}}. \tag{33}$$

The unbiasedness results from a property of the χ^2 distribution: $\mathrm{E}[\hat{\tau}^2|\mathbf{C}] = \mathrm{E}[\hat{\mathbf{e}}^T\hat{\mathbf{e}}]/(\mathsf{N}-\mathsf{m}-\mathsf{q}) = (\mathsf{N}-\mathsf{m}-\mathsf{q})\tau^2/(\mathsf{N}-\mathsf{m}-\mathsf{q}) = \tau^2$.

4.2 *Asymptotic Variance Ratio and Loss Factor*

Just like Venkatraman and Wilson (1986) and Porta Nova and Wilson (1989), and in contrast to Nozari, Arnold and Pedgen (1984) and Rubinstein and Marcus (1985), we consider that the adequate generalization of the performance measures (variance ratio and loss factor), introduced by Lavenberg, Moeller and Welch (1982), is based on the unconditioned dispersion matrix of the controlled coefficients in metamodel (2). As a consequence, we will now obtain the unconditioned variance of $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$.

Since (31) is verified, we have

$$\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \boldsymbol{\theta} \qquad \text{and} \qquad \mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] = \tau^2\mathbf{B}\mathbf{B}^T.$$

So,

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] = \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] + \mathcal{D}[\mathcal{E}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] = \mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \tag{34}$$

and

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}[\mathbf{I}_{\mathsf{m}} + \tilde{\mathbf{F}}^T\mathbf{C}(\mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C})^{-1}\mathbf{C}^T\tilde{\mathbf{F}}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}].$$

Let $\mathbf{U} = (\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\mathbf{C}$ and $\mathbf{V} = \mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C}$. Then, we can write

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}] \approx \tau^2(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} + \tau^2\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T. \tag{35}$$

Since $\mathbf{C} \sim \mathrm{N}_{\mathsf{N},\mathsf{q}}(\mathbf{0}, \mathbf{I}_{\mathsf{N}}, \boldsymbol{\Sigma}_C)$ and $\tilde{\mathbf{P}}$ is an orthogonal projection of $\mathbb{R}^{\mathsf{N}}$ into a space of dimension $\mathsf{N}-\mathsf{m}$ ($\mathcal{R}(\mathbf{F})^{\perp}$), we apply Theorem 17.7 of Arnold (1981) to obtain $\mathbf{V} = \mathbf{C}^T\tilde{\mathbf{P}}\mathbf{C} \sim \mathrm{W}_{\mathsf{q}}\,(\mathsf{N}-\mathsf{m}, \boldsymbol{\Sigma}_C)$.

Using the fact that $(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^T\tilde{\mathbf{P}} = \mathbf{0}$ and, since $\tilde{\mathbf{P}}$ is positive definite, we apply Theorem 17.7-b.2) of Arnold (1981) and we can conclude that $\mathbf{U}$ and $\mathbf{V}$ are independent. Taking the expected value in (35), we obtain

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} + \tau^2\mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T]$$

Using the results in the Appendix of Nozari, Arnold and Pedgen (1984), we have

$$\mathcal{E}[\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T] = \frac{\mathsf{q}}{\mathsf{N}-\mathsf{m}-\mathsf{q}-1}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}.$$

Combining the two previous results, we obtain

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2\frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{N}-\mathsf{m}-\mathsf{q}-1}(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1}.$$

But $\tilde{\mathbf{F}}^T = [\mathbf{F}^T \ldots \mathbf{F}^T]$, so $(\tilde{\mathbf{F}}^T\tilde{\mathbf{F}})^{-1} = \mathbf{F}^T\mathbf{F}/\mathsf{r}$ and, as a result,

$$\mathcal{E}[\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})|\mathbf{C}]] \approx \tau^2\frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{r}(\mathsf{N}-\mathsf{m}-\mathsf{q}-1)}(\mathbf{F}^T\mathbf{F})^{-1},$$

that is, from (34), we have the following asymptotic result

$$\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})] \approx \tau^2(\mathbf{F}^T\mathbf{F})^{-1}\frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{r}(\mathsf{N}-\mathsf{m}-\mathsf{q}-1)}.$$

This approximation, in conjunction with (8) and (16), allows us to obtain the following approximated generalized variance ratio:

$$\eta(\hat{\boldsymbol{\delta}}) = \frac{\left|\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]\right|}{\left|\mathcal{D}[\hat{\boldsymbol{\theta}}]\right|} \approx \frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{N}-\mathsf{m}-\mathsf{q}-1}\frac{\tau^2}{\sigma^2} = \frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{N}-\mathsf{m}-\mathsf{q}-1}(1-\rho^2_{YC}). \tag{36}$$

Comparing this with the minimum variance ratio (18), we observe a degradation of the maximum variance reduction, namely the loss factor:

$$\mathrm{LF}(\hat{\boldsymbol{\delta}}) = \frac{\mathsf{N}-\mathsf{m}-1}{\mathsf{N}-\mathsf{m}-\mathsf{q}-1}. \tag{37}$$

4.3 *Asymptotic Confidence Regions for the Metamodel Coefficients*

The objective of this section is to determine a confidence rectangle consisting of m confidence intervals for $\boldsymbol{\theta}_j$, $j = 1, \ldots, \mathsf{m}$. In fact, it is simpler to represent graphically and to explain the meaning of a confidence rectangle of this type, when compared with the more common approximated confidence ellipsoid.

Since (31) is verified, then using Theorem 3.10 of Arnold, we can ensure that conditioning on $\mathbf{C}$:

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)}{\tau^2}|\mathbf{C} \sim \chi^2_{\mathsf{m}} \tag{38}$$

On the other hand, (33) can be rewritten as

$$(\mathsf{N}-\mathsf{m}-\mathsf{q})\frac{\hat{\tau}^2}{\tau^2}|\mathbf{C} \sim \chi^2_{\mathsf{N}-\mathsf{m}-\mathsf{q}}. \tag{39}$$

Combining (38) and (39), we obtain

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)}{\mathsf{m}\hat{\tau}^2}|\mathbf{C} \sim \mathrm{F}_{\mathsf{m},\mathsf{N}-\mathsf{m}-\mathsf{q}}.$$

Given $\mathbf{C}$, an asymptotic confidence region for $\boldsymbol{\theta}$, with conditional coverage probability of at least $1-\alpha$, is given by

$$\left\{\boldsymbol{\theta}^* : \frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}^*)}{\mathsf{m}\hat{\tau}^2} \leq \mathrm{F}_{\mathsf{m},\mathsf{N}-\mathsf{m}-\mathsf{q};1-\alpha}\right\}. \tag{40}$$

Since this confidence region has conditional coverage of at least $1-\alpha$, it has also unconditional coverage of at least $1-\alpha$.

Let $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ be the i-th component of the vector $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ and let $\hat{\tau}^2[\mathbf{B}\mathbf{B}^T]_{ii}$ be the corresponding variance estimator (the i-th diagonal element of $\hat{\tau}^2\mathbf{B}\mathbf{B}^T$). Since $\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})$ is conditionally independent of $\hat{\tau}^2[\mathbf{B}\mathbf{B}^T]_{ii}$ given $\mathbf{C}$, the results (31) and (33) imply that

$$\frac{\hat{\boldsymbol{\theta}}_i(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}_i^*}{\hat{\tau}\sqrt{\left[\mathbf{B}\mathbf{B}^T\right]_{ii}}}|\mathbf{C} \sim t_{\mathsf{N}-\mathsf{m}-\mathsf{q}},$$

where $t_{\mathsf{N}-\mathsf{m}-\mathsf{q}}$ represents the Student t-distribution with $\mathsf{N}-\mathsf{m}-\mathsf{q}$ degrees of freedom. As a consequence, using the Bonferroni method, a confidence rectangle for $\boldsymbol{\theta}$ with conditional coverage probability of at least $1-\alpha$ has the form

$$\hat{\boldsymbol{\theta}}_k(\hat{\boldsymbol{\delta}}) \pm t_{\mathsf{N}-\mathsf{m}-\mathsf{q};1-\alpha/(2p)}\hat{\tau}\left[\mathbf{B}\mathbf{B}^T\right]_{kk}^{1/2}, \quad k=1,\ldots,p, \tag{41}$$

where $1 \leq p \leq \mathsf{m}$. This confidence region has conditional coverage probability of at least $1-\alpha$, so it also has unconditional coverage probability of at least $1-\alpha$.

4.4 *Hypothesis Testing on the Metamodel Coefficients*

Suppose that we want to test the hypothesis $\mathrm{H}_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ *versus* $\mathrm{H}_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Since (31) is verified and using Theorem 3.10 of Arnold (1981), we observe that, conditioning on $\mathbf{C}$:

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}_0)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}_0)}{\tau^2} \sim \chi^2_{\mathsf{m}}, \tag{42}$$

if H_0 is true. Combining (42) and (39) and following an identical procedure to the one considered in 4.3, we reject H_0, with confidence level $100(1-\alpha)\%$, if

$$\frac{(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}_0)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})-\boldsymbol{\theta}_0)}{\mathsf{m}\hat{\tau}^2} > \mathrm{F}_{\mathsf{m},\mathsf{n}-\mathsf{m}-\mathsf{q};1-\alpha}. \tag{43}$$

5. Numerical Results for Queueing Application

We illustrate our methodology using a simple $M/M/1$ queueing system. We assume that customers arrive according to a Poisson process with a constant expected arrival rate, λ, and that service times follow an exponential distribution with a constant expected service time, $1/\mu \equiv 1$. The performance measure of interest is the average waiting time in the queue. The objective is to express this response as a function of the queue utilization factor, $\rho = \lambda/\mu$ (a single decision variable). The available concomitant output variables, that can be used as controls, are the average service time and the average inter-arrival time. In this experiment, after some minor adjustments, twelve ($\mathsf{n} = 12$) different values for ρ were considered: $\{\rho_i : i = 1, 12\} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.7, 0.75, 0.85, 0.9, 0.95\}$. There were $\mathsf{r} = 20$ replications for each of the $\mathsf{n} = 12$ design points. Different replications used the same value for the independent variable ρ_i, but different pseudo-random number seeds. Each of these 20 replications started with an empty and idle system (no customers waiting). At each design point, we ran Welch's procedure (Welch (1983)) to determine the length of each simulation run and the initial data deletion. For example, at the design point $\rho = 0.95$, we ignored 3500 observations from the beginning of each run and we used only the remaining 36500 observations (approximately 85% of the number of observations in the run), while considering a Welch window of 10000; see Table 1. We then compared the dispersion diagram based on the collected data (Figure 1) with commonly

Table 1: Initial data deletion.

	Observations		Welch's
ρ_i	Deleted	In run	window
0.10,0.20,0.30	500	3500	1000
0.40	1000	7000	1000
0.50	1500	10000	1000
0.55,0.60,0.70,0.75	1500	10000	4000
0.85	2000	20000	8000
0.90	2500	20000	10000
0.95	3500	40000	10000

available theoretical curves. To relate the average waiting time in the queue with the utilization factor, we chose the hyperbolic metamodel $Y_{ij} = \frac{\theta_1 X_i}{1+\theta_2 X_i} + \epsilon_{ij}$ (where Y_{ij} is the average queue waiting time during the j-th run at experimental point i), with $\epsilon_{ij} \sim \mathrm{N}\left(0, \sigma_i^2\right)$ ($i = 1, \ldots, 12$, $j = 1, \ldots, 20$), and σ_i^2 varies with i. For stabilizing the variance, we took logarithms on both sides of the above

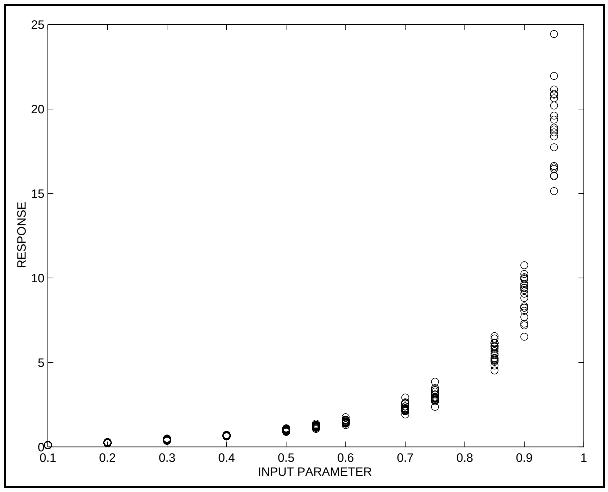

Figure 1: Dispersion diagram of the $M/M/1$ queue.

expression, obtaining

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1+\theta_2 X_i} + \upsilon_{ij}, \quad i = 1, \ldots, 12, j = 1, \ldots, 20, \tag{44}$$

with $\upsilon_{ij} = \log(1 + \epsilon_{ij}/\mathrm{E}[Y_{ij}])$. $\mathrm{E}[\upsilon_{ij}] \approx 0$, because ϵ_{ij} is small when compared with $\mathrm{E}[Y_{ij}]$ and $\mathrm{Var}[\upsilon_{ij}]$ is approximately constant for all $i = 1, \ldots, \mathsf{n}$ and $j = 1, \ldots, \mathsf{r}$. We measure the variance heterogeneity using the quantity

$$het = \max_{i=1,\ldots,12} \widehat{\mathrm{Var}}[\upsilon_{ij}] / \min_{i=1,\ldots,12} \widehat{\mathrm{Var}}[\upsilon_{ij}],$$

with

$$\widehat{\mathrm{Var}}[\upsilon_{ij}] = \left[\frac{1}{\mathsf{r}-1} \sum_{j=1}^{\mathsf{r}} \left(\log Y_{ij} - \frac{1}{\mathsf{r}} \sum_{j=1}^{\mathsf{r}} \log Y_{ij} \right)^2 \right]^{1/2};$$

see Kleijnen (1992). We obtained a het value approximately constant and equal to 1. For improving the efficiency of metamodel estimation, we chose the following control variates

$$C_{kij} = \frac{t_{kij} - \mu_{ki}}{\varsigma_{ki}}, \quad k = 1, 2, \quad i = 1, \ldots, \mathsf{n}, \quad \mathsf{j} = 1, \ldots, \mathsf{r},$$

where t_{1ij} is the average inter-arrival time and t_{2ij} is the average service time. Both were sampled from exponential distributions with known means and variances: $\mathrm{E}[t_{1ij}] = \mu_{1i} = 1/\rho_i$, $\mathrm{Var}[t_{1ij}] = \varsigma_{1i}^2 = 1/\rho_i^2$, $\mathrm{E}[t_{2ij}] = \mu_{2i} = 1/\mu^2 = 1$ and $\mathrm{Var}[t_{2ij}] = \varsigma_{2i}^2 = 1/\mu^2 = 1$. The hypothesized controlled problem is

$$\log Y_{ij} = \log \frac{\theta_1 X_i}{1+\theta_2 X_i} + \delta_1 C_{1ij} + \delta_2 C_{2ij} + \nu_{ij}, \;\; i = 1, \ldots, 12, \;\; j = 1, \ldots, 20.$$

5.1 *Estimation and Validation*

We obtained the least squares estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ using the Levenberg-Marquart method, implemented in MATLAB, with a termination tolerance of 10^{-6} and maximum number of function evaluations equal to 600; see Table 2.

Table 2: Estimated Metamodel Coefficients.

Metamodel Coefficients	Direct Estimator $\hat{\boldsymbol{\theta}}$	Controlled Estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$
θ_1	0.9982	1.0001
θ_2	-0.9992	-0.9991

After estimation, the validation of the controlled metamodel was carried out. Since $\mathrm{F}_{\mathsf{n}-\mathsf{m}-\mathsf{q},\mathsf{N}-\mathsf{n};\alpha} = \mathrm{F}_{8,228;0.05} \approx 1.95$, with $\mathsf{N} = 240$, based on the F-test, we do not reject the metamodel with control variables; see Table 3. As a consequence, based on this validation procedure, we do not reject the controlled metamodel.

Table 3: Testing for lack-of-fit.

Source	d.f.	Sum of Squares	Mean of Squares	F
Lack-of-fit	8	0.008036	0.001004	0.119
Pure Error	228	1.923	0.008435	

5.2 *Confidence Regions and Hypothesis Testing*

Since $\mathrm{F}_{\mathsf{m},\mathsf{N}-\mathsf{m}-\mathsf{q};\alpha} = \mathrm{F}_{2,236;0.05} = 3.034$ and $\hat{\tau}^2 = 8.184 \times 10^{-3}$, the 95% approximated confidence ellipsoid for $\boldsymbol{\theta}$, centered at $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$, is given by (40):

$$\left\{\boldsymbol{\theta}^* : (\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*)^T(\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}^*) \leq 4.966 \times 10^{-2}\right\},$$

where

$$\mathbf{B}\mathbf{B}^T = \begin{bmatrix} 3.2891 \times 10^{-2} & 1.8495 \times 10^{-3} \\ 1.8495 \times 10^{-3} & 2.2099 \times 10^{-4} \end{bmatrix}.$$

The corresponding approximated confidence rectangle for $\boldsymbol{\theta}$, with coverage probability of at least $1 - \alpha = 0.95$, from (41), is given in Table 4. In the construction of this confidence rectangle, we used $t_{\mathsf{N}-\mathsf{m}-\mathsf{q};1-\alpha/(2\mathsf{m})} = t_{236;0.9875} \approx 2.256$. We tested the hypothesis $\mathrm{H}_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0 = (1.0, -1.0)^T$ *versus*

Table 4: Approximated Bonferroni 95% Confidence Intervals.

Metamodel Coefficients	Controlled Estimator
θ_1	$1.0001188 \pm 3.35 \times 10^{-3}$
θ_2	$-0.9991349 \pm 2.74 \times 10^{-4}$

$\mathrm{H}_1 : \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0 = (1.0, -1.0)^T$, with a confidence level of 0.95%, using $\mathrm{F}_{\mathsf{m},\mathsf{N}-\mathsf{m}-\mathsf{q};1-\alpha} = \mathrm{F}_{2,236;0.95} = 3.034$. We obtained $(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0)^T(\mathsf{m}\hat{\tau}^2\mathbf{B}\mathbf{B}^T)^{-1}(\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}}) - \boldsymbol{\theta}_0) = 0.3848$ (43). Therefore, we do not reject the null hypothesis H_0.

5.3 *Experimental Variance Ratio and Loss Factor*

In order to estimate the variance ratios and loss factors, we adapted the procedure described in Porta Nova and Wilson (1989) to our situation. Thus, we performed a meta-experiment with $\mathsf{K} = 30$ independent replications of the basic experiment, consisting of twelve design points ($\rho = 0.1, 0.2, etc.$) and twenty independent replications for each design point. For each kth replication of the basic experiment ($k = 1, \ldots, \mathsf{K}$), we calculate the direct estimator $\hat{\boldsymbol{\theta}}^k$ and the control-variate estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})^k$ for the metamodel coefficient vector $\boldsymbol{\theta}$. From the random sample $\{\hat{\boldsymbol{\theta}}^k : 1 \leq k \leq \mathsf{K}\}$ we compute a unbiased estimator for $\mathcal{D}[\hat{\boldsymbol{\theta}}]$ as follow:

$$\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}] = \frac{1}{\mathsf{K}} \sum_{k=1}^{\mathsf{K}} \left(\hat{\boldsymbol{\theta}}^k - \bar{\hat{\boldsymbol{\theta}}}\right)\left(\hat{\boldsymbol{\theta}}^k - \bar{\hat{\boldsymbol{\theta}}}\right)^T, \quad \text{where} \quad \bar{\hat{\boldsymbol{\theta}}} = \frac{1}{\mathsf{K}} \sum_{k=1}^{\mathsf{K}} \hat{\boldsymbol{\theta}}^k;$$

and similarly from the random sample $\{\hat{\boldsymbol{\theta}}^k : 1 \leq k \leq \mathsf{K}\}$ we compute an unbiased estimator $\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]$ of $\mathcal{D}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]$. Based on this estimators, we compute the following estimator of the variance ratio (36), wich we call the *observed variance ratio*:

$$\hat{\eta}(\hat{\boldsymbol{\delta}}) = \frac{\left|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})]\right|}{\left|\hat{\mathcal{D}}[\hat{\boldsymbol{\theta}}]\right|}. \tag{45}$$

In order to obtain the minimum variance ratio (18), at each design point we compute unbiased estimators for the variance of Y, σ^2, covariance vector between Y and $\mathbf{C}$, $\boldsymbol{\sigma}_{YC}$, and despersion matrix of $\mathbf{C}$, $\boldsymbol{\Sigma}_C$. For example, if Y_i^k denotes the mean response observed at the ith design point on the kth independent replication of the basic experiment, then the variance of Y is estimated by

$$(\hat{\sigma}^2)^i = \frac{1}{\mathsf{K}} \sum_{k=1}^{\mathsf{K}} \left(Y_i^k - \bar{Y}\right)^2,$$

where $\bar{Y} = \sum_{j=1}^{\mathsf{K}} Y_i^j/\mathsf{K}$. As a result, a pooled estimator of σ^2 based on all n experimental points is given by: $\bar{\hat{\sigma}}^2 = \sum_{i=1}^{\mathsf{n}} (\hat{\sigma}^2)^i/\mathsf{n}$. The other estimators are obtained in a similar way: $\bar{\hat{\boldsymbol{\sigma}}}_{YC} = \sum_{i=1}^{\mathsf{n}} \hat{\boldsymbol{\sigma}}_{YC}^i/\mathsf{n}$, where

$$\hat{\boldsymbol{\Sigma}}_C^i = \frac{1}{\mathsf{K}} \sum_{k=1}^{\mathsf{K}} (\bar{\mathbf{C}}_{i.}^k - \bar{\bar{\mathbf{C}}}_{i.})^T (\bar{\mathbf{C}}_{i.}^k - \bar{\bar{\mathbf{C}}}_{i.}),$$

with $\bar{\bar{\mathbf{C}}}_{i.} = \sum_{k=1}^{\mathsf{K}} \bar{\mathbf{C}}_{i.}^k/\mathsf{K}$; and $\bar{\hat{\boldsymbol{\sigma}}}_{YC} = \sum_{i=1}^{\mathsf{n}} \hat{\boldsymbol{\sigma}}_{YC}^i/\mathsf{n}$, where

$$\hat{\boldsymbol{\sigma}}_{YC}^i = \frac{1}{\mathsf{K}} \sum_{k=1}^{\mathsf{K}} (\bar{Y}_{i.}^k - \bar{\bar{Y}}_{i.})(\bar{\mathbf{C}}_{i.}^k - \bar{\bar{\mathbf{C}}}_{i.}).$$

Using this numerical values, we calculate the following estimator of (18) that we call the *estimated minimum variance ratio*:

$$\hat{\eta}(\boldsymbol{\delta}) = 1 - \frac{\bar{\hat{\boldsymbol{\sigma}}}_{YC} \bar{\hat{\boldsymbol{\Sigma}}}_C^{-1} \bar{\hat{\boldsymbol{\sigma}}}_{YC}^T}{\bar{\hat{\sigma}}^2}. \tag{46}$$

Multiplying (46) by the loss factor (37), we obtain the *predicted variance ratio*

$$\ddot{\eta} = \hat{\eta}(\boldsymbol{\delta}) \mathrm{LF}(\hat{\boldsymbol{\delta}}).$$

The *observed loss factor* is given by

$$\widehat{\mathrm{LF}}(\hat{\boldsymbol{\delta}}) = \frac{\hat{\eta}(\hat{\boldsymbol{\delta}})}{\hat{\eta}(\boldsymbol{\delta})}.$$

The predicted variance ratio can be compared with the observed variance ratio and the observed loss factor can be compared with the theoretical loss factor. The numerical results are reported in Table 5.

Table 5: Estimation of the Variance Ratios and Loss Factors.

	Estimated Variance Ratios			Loss Factor		
		Actual				
Metamodel Coefficient	Minimum $\hat{\eta}(\boldsymbol{\delta})$	Predicted $\ddot{\eta}(\hat{\boldsymbol{\delta}})$	Observed $\hat{\eta}(\hat{\boldsymbol{\delta}})$	True $\mathrm{LF}(\hat{\boldsymbol{\delta}})$	Estimated $\widehat{\mathrm{LF}}(\hat{\boldsymbol{\delta}})$	% Error
$\boldsymbol{\theta}$	0.61	0.62	0.99	1.01	1.20	60.6%

In the example described here, the maximum percentage reduction in generalized variance that can be achieved using control variates is approximately $100[1 - \hat{\eta}(\boldsymbol{\delta})]\% = 39\%$. Since we do not know $\boldsymbol{\delta}$, it must be estimated. Nevertheless the resulting estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$ has a smaller variance, when compared with the estimator $\hat{\boldsymbol{\theta}}$ without control variates. The numerical results presented here are in agreement with the theoretical results developed in this article. The error obtained for the loss factor is similar to the value obtained by Porta Nova and Wilson (1989), 57.7%. These authors analyzed a queueing network simulation in the context of a linear metamodel estimation.

6. Conclusions

The main objective of this paper is to establish some important results on the use of multiple control variates for improving the precision of nonlinear regression metamodel estimation. This technique can be useful in many situations where it is possible to identify effective concomitant control variables. Since nonlinear regression models are better, than linear models, in capturing the shape of arbitrary mathematical functions, we emphasize the importance of using valid nonlinear metamodels in simulation studies. Also, nonlinear metamodels allow us to characterize the precision of the fit by the use of confidence intervals and they are more robust than linear models when extrapolating from the actual experimental domain.

For experimental designs with a sufficiently large number of experimental points and under certain regularity conditions, the efficiency of metamodel estimation can be improved using the method of control variables. However, whether a regression metamodel is used in the simulation context or not, it must be validated. The validation can be made using, for example, the lack of fit F-test present in the statistical literature on nonlinear regression models. In our experimental study, we observed a marked sensitivity of the variance ratio $\eta(\hat{\boldsymbol{\delta}})$ and the loss factor $\mathrm{LF}(\hat{\boldsymbol{\delta}})$ with respect to the validity of the assumed controlled problem (2). As a consequence, it is imperative to resort to statistical validation techniques, like the above mentioned F-test, in order to verify the capability of the controlled metamodel in representing the simulation model.

References

Anderson, T. W. 1984. An Introduction to Multivariate Statistical Analysiss, John Wiley and Sons, New York, USA.

Arnold, S. F. 1981. The Theory of Linear Models and Multivariate Analysis, John Wiley and Sons, New York, USA.

Barton, R. R. 1992. Metamodels for Simulation Input-Output Relations. In Proceedings of Winter Simulation Conference, pp. 289-299.

Friedman, L. W., and Friedman, H. H. 1985. Validating the simulation metamodel: Some practical approaches. Simulation, 45 (3) 144-146.

Glynn, P. W., and Whitt, W. 1989. Indirect Estimation Via $L = \lambda W$, Operations Research, 37 (1) 82-103.

Loh, W. W. 1994. On the Method of Control Variates, Unpublished Ph.D. dissertation, Department of Operations Research, Stanford University, Stanford CA. (Downloadable from website http://www-or.stanford.edu/ simlib/theses.html).

Kleijnen, J. P. C. 1992. Regression Metamodels for Simulation with Common Random Numbers: Comparison of Validation Tests and Confidence Intervals. Management Science, 38 (8) 1164-1184.

Lavenberg, S. S., and Welch, P. D. 1981. A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations. Management Science, 27 (3) 322-335.

Lavenberg, S. S., Moeller, T. L., and Welch, P. D. 1982. Statistical Results on Control Variables with Application to Queueing Network Simulation. Operations Research, 30, 182-202.

Nozari, A. 1982. On Simulation Experiments for Multipopulation Problems. Unpublished Ph.D. dissertation, Pennsylvania State University.

Nozari, A., Arnold, S. F., and Pegden, C. D. 1984. Control Variates for Multipopulation Simulation Experiments. IIE Transactions, 16, 159-169.

Porta Nova, A. M. 1985. A Generalized Approach to Variance Reduction in Discrete-Event Simulation Using Control Variables. Unpublished Ph.D. dissertation. University of Texas.

Porta Nova, A. M., and Wilson, J. R. 1989. Estimation of Multiresponse Simulation Metamodels Using Control Variates. Management Science, 35 (11) 1316-1333.

Rubinstein, R. Y., and Marcus, R. 1985. Efficiency of Multivariate Control Variates in Monte Carlo Simulation, Operations Research, 33 (3) 661-677.

Santos, M. I., and Porta Nova, A. M. 1999. The Main Issues in Nonlinear Simulation Metamodel Estimation. In Proceedings of Winter Simulation Conference, pp. 502-509.

Searle, S. R. 1971. Linear Models, John Wiley and Sons, New York, USA.

Seber, G. A. F., and Wild, C. J. 1989. *Nonlinear Regression*. John Wiley and Sons, New York, NY, USA.

Shih, N. H., and Song, W. T. 1995. Estimating Simulation Metamodels Using Correlated Control Variates. Communications in Statistics B - Simulation and Computation, 24 (1) 79-90.

Venkatraman, S., and Wilson, J. 1986. The Efficiency of Control Variates in Multiresponse Simulation. Operations Research Letters, 5 (1) 37-42.

Welch, P. D. 1983. The statistical analysis of simulation results. In S. S. Lavenberg, editor, Computer Performance Modeling Handbook, pages 268-328. Academic Press, New York, NY, USA.