World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:9, No:12, 2015

Selection of Designs in Ordinal Regression Models under Linear Predictor Misspecification

Ishapathik Das

Abstract—The purpose of this article is to find a method of comparing designs for ordinal regression models using quantile dispersion graphs in the presence of linear predictor misspecification. The true relationship between response variable and the corresponding control variables are usually unknown. Experimenter assumes certain form of the linear predictor of the ordinal regression models. The assumed form of the linear predictor may not be correct always. Thus, the maximum likelihood estimates (MLE) of the unknown parameters of the model may be biased due to misspecification of the linear predictor. In this article, the uncertainty in the linear predictor is represented by an unknown function. An algorithm is provided to estimate the unknown function at the design points where observations are available. The unknown function is estimated at all points in the design region using multivariate parametric kriging. The comparison of the designs are based on a scalar valued function of the mean squared error of prediction (MSEP) matrix, which incorporates both variance and bias of the prediction caused by the misspecification in the linear predictor. The designs are compared using quantile dispersion graphs approach. The graphs also visually depict the robustness of the designs on the changes in the parameter values. Numerical examples are presented to illustrate the proposed methodology.

Keywords—Model misspecification, multivariate kriging, multivariate logistic link, ordinal response models, quantile dispersion graphs.

I. Introduction

THIS article presents a method of selecting robust designs for ordinal response models under model misspecification. Several authors discussed about the model misspecification and its effect on design selection for linear and generalized linear models (with a single response) till date; however, there is a lack in works for the situation where the responses are ordinal. In case of generalized linear models (GLMs), model misspecification may occur when the assumptions regarding the form of the linear predictor and/or the link function is wrong. For addressing the uncertainty in the linear predictor, we assume an unknown function and an iterative method is provided for estimating the unknown function at the design points of the model. Then, the multivariate kriging method is used to estimate the function at all points of the design region using the estimates at design points as a training data sets. The designs are compared using the quantile dispersion graphs approach on the basis of a scalar valued function of the mean squared error of prediction. These graphs allow design comparison over the entire experimental region. Also they provide assessment of the designs' sensitivity to changes in the parameter values.

I. Das is with the Department of Statistics, Kumaun University, SSJ Campus, Almora, 263601, India (phone: +919634590436; e-mail: ishapathik@gmail.com).

Some of the references for selecting optimal designs in multivariate GLMs are [1]–[3]. Reference [1] provides optimal designs for bivariate logistic models while [3] discussed designs for multinomial responses. However, neither of these papers investigate the effect of model misspecification on design selection for ordinal responses or multivariate GLMs.

Here, we provide a method of comparing designs in ordinal response models using QDGs under model misspecification. The comparison criterion is the mean squared error of prediction (MSEP) of the designs which takes into account the variance and the bias in the parameter estimates due to model misspecification. The QDGs technique allows a comparison of the designs on the entire experimental designs and not just at a single point like A, D or G optimal designs. QDGs also allow one to study the robustness of the designs to parameter changes.

The rest of the article is organised as follows: Section II discusses the multivariate generalized models and ordinal regression models. In Section III, the problem of model misspecification in ordinal response models and the expression for the MSEP under misspecification are given. The comparison criterion and the QDGs approach is explained in Section IV. In Section V, numerical examples illustrating the proposed, methodology are discussed followed by concluding remarks in Section VI.

II. Multivariate Generalized Linear Models

In multivariate GLMs (generalized linear models), we assume that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are independent q dimensional response variables such that each $\mathbf{y}_i$ belongs to an exponential family, which has the form

$$s(\mathbf{y}_i|\boldsymbol{\theta}_i, \phi, \omega_i) = \exp\left\{\frac{[\mathbf{y}_i'\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)]}{\phi}\omega_i + c(\mathbf{y}_i, \phi, \omega_i)\right\}. \quad (1)$$

where b(.) and c(.) are known function and ϕ is dispersion parameter possibly unknown. Here, mean response $\boldsymbol{\mu}_i = \boldsymbol{\mu}(\boldsymbol{\theta}_i) = E(\mathbf{y}_i|\boldsymbol{\theta}_i) = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{iq}]^T$ and linear predictor $\boldsymbol{\eta}_i = [\eta_{i1}, \eta_{i2}, \ldots, \eta_{iq}]^T$ are q dimensional vectors. The mean response $\boldsymbol{\mu}_i$ is related to the linear predictor $\boldsymbol{\eta}_i = \mathbf{Z}_i\boldsymbol{\beta}$ by link function $\mathbf{g}$ as

$$\boldsymbol{\eta}_i = \mathbf{g}(\boldsymbol{\mu}_i), \quad (2)$$

where

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}_i) & \mathbf{O}_{1\times p_2} & \ldots & \mathbf{O}_{1\times p_q} \\ \mathbf{O}_{1\times p_1} & \mathbf{f}_2(\mathbf{x}_i) & \ldots & \mathbf{O}_{1\times p_q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{1\times p_1} & \mathbf{O}_{1\times p_2} & \ldots & \mathbf{f}_q(\mathbf{x}_i) \end{bmatrix}, \quad (3)$$

$\mathbf{f}_j(\mathbf{x})$ is known vector function of $\mathbf{x}$, $\mathbf{O}_{1\times p_j}$ is matrix of order $1 \times p_j$ with all elements zero, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_q^T]^T$ is $p \times 1$ vector of unknown parameters such that each $\boldsymbol{\beta}_j = [\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp_j}]^T$ is $p_j \times 1$ vector of unknown parameters for jth response with $p = \sum_{j=1}^{q} p_j$. For simplicity we denote $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]^T$ such that first p_1 elements of $\boldsymbol{\beta}$ come from $\boldsymbol{\beta}_1$, next p_2 elements of $\boldsymbol{\beta}$ come from $\boldsymbol{\beta}_2$ and so on. It is usually assumed that inverse of $\mathbf{g}$ exists and it is denoted as $\mathbf{h}$. For natural link function we have $\boldsymbol{\theta}_i = \boldsymbol{\eta}_i$. In general, we can relate $\boldsymbol{\theta}_i$ to linear predictor $\boldsymbol{\eta}_i$ using $\boldsymbol{\mu}_i = \boldsymbol{\mu}(\boldsymbol{\theta}_i)$ by function $\mathbf{u}$ which is defined as

$$\boldsymbol{\theta}_i = \mathbf{u}(\boldsymbol{\eta}_i) = \mathbf{u}(\mathbf{Z}_i\boldsymbol{\beta}) = \boldsymbol{\mu}^{-1}(\mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta})) \quad \text{[4, pp. 347]} \tag{4}$$

For example, if $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are independent random variables such that each $\mathbf{y}_i$ is multinomial distributed having $(q+1)$ categories with parameter $(\boldsymbol{\pi}_i, n_i)$ with the density given by

$$P(\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iq})) = \frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iq}!(n_i - y_{i1} - y_{i2} - \ldots - y_{iq})!} \times \pi_{i1}^{y_{i1}}\pi_{i2}^{y_{i2}}\ldots\pi_{iq}^{y_{iq}} \times (1 - \pi_{i1} - \pi_{i2}\ldots - \pi_{iq})^{(n_i - y_{i1} - y_{i2} - \ldots - y_{iq})}$$

Then the density of $\bar{\mathbf{y}}_i = \mathbf{y}_i/n_i$ has the form

$$s(\bar{\mathbf{y}}_i|\boldsymbol{\theta}_i, \phi, \omega_i) = \exp\left\{\frac{[\bar{\mathbf{y}}_i'\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)]}{\phi}\omega_i + c(\mathbf{y}_i, \phi, \omega_i)\right\}, \tag{5}$$

where $\boldsymbol{\theta}_i = \left[\log(\frac{\pi_{i1}}{1-\pi_{i1}-\pi_{i2}-\ldots-\pi_{iq}}), \ldots, \log(\frac{\pi_{iq}}{1-\pi_{i1}-\pi_{i2}-\ldots-\pi_{iq}})\right]^T$, $b(\boldsymbol{\theta}_i) = -\log(1 - \pi_{i1} - \pi_{i2} - \ldots - \pi_{iq})$, $c(\mathbf{y}_i, \phi, \omega_i) = \log\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iq}!(n_i - y_{i1} - y_{i2} - \ldots - y_{iq})!}\right)$, $\omega_i = n_i$ and $\phi = 1$. Here $\boldsymbol{\mu}_i = E(\bar{\mathbf{y}}_i) = \boldsymbol{\pi}_i$ for $i = 1, 2, \ldots, n$.

The logistic link function is given by

$$\boldsymbol{\eta}_i = \mathbf{g}(\boldsymbol{\pi}_i) = \boldsymbol{\theta}_i$$

Here, $\boldsymbol{\theta}_i = \boldsymbol{\eta}_i$ and hence $\mathbf{u}$ is identity function.

A. Ordinal Response Models

Suppose $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are independent q dimensional ordinal responses so that each $\mathbf{y}_i$ is multinomial distributed having $(q+1)$ ordered categories with parameter $(\boldsymbol{\pi}_i, n_i)$. The simple cumulative model [5, pp. 243] for ordinal responses is given by

$$\gamma_{ij} = \sum_{r=1}^{j} \boldsymbol{\pi}_{ir} = P(\mathbf{y}_i \le j) = F(\theta_{0j} + \mathbf{x}_i^T\boldsymbol{\theta}) \quad \text{for } j = 1, 2, \ldots, q\ , \tag{6}$$

where F^{-1} is link function, $\theta_{01}, \theta_{02}, \ldots, \theta_{0q}, \boldsymbol{\theta}$ are parameters and $\mathbf{x}_i$ is design vector. The general form of the multivariate GLMs for the categorical responses is given by $\mathbf{g}(\boldsymbol{\mu}_i) = \mathbf{g}(\boldsymbol{\pi}_i) = \mathbf{Z}_i\boldsymbol{\beta}$ or $\boldsymbol{\mu}_i = \boldsymbol{\pi}_i = \mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta})$ [5, pp. 261]. The equivalent form for the ordinal responses is given by

$$\begin{aligned}\boldsymbol{\mu}_i &= \boldsymbol{\pi}_i = \begin{pmatrix}\pi_{i1}\\ \pi_{i2}\\ \vdots\\ \pi_{iq}\end{pmatrix}\\ &= \mathbf{h}\left\{\begin{pmatrix}1 & \ldots & 0 & \mathbf{x}_i^T\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \ldots & 1 & \mathbf{x}_i^T\end{pmatrix}\begin{pmatrix}\theta_{01}\\ \vdots\\ \theta_{0q}\\ \boldsymbol{\theta}\end{pmatrix}\right\}\\ &= \mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta}),\end{aligned} \tag{7}$$

where

$$\mathbf{Z}_i = \begin{pmatrix}1 & \ldots & 0 & \mathbf{x}_i^T\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \ldots & 1 & \mathbf{x}_i^T\end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix}\theta_{01}\\ \vdots\\ \theta_{0q}\\ \boldsymbol{\theta}\end{pmatrix} \text{ and } \mathbf{h} = \begin{pmatrix}h_1\\ h_2\\ \vdots\\ h_q\end{pmatrix}$$

which is given by

$$h_j(\eta_{i1}, \ldots, \eta_{iq}) = F(\eta_{ij}) - F(\eta_{ij-1}),\quad \text{where } \eta_{ij} = \theta_{0j} + \mathbf{x}_i^T\boldsymbol{\theta}, \text{ for } j = 1, 2, \ldots, q. \tag{8}$$

From (7), and (8), we should have $\eta_{i1} \le \eta_{i2} \ldots \le \eta_{iq}$ and hence $\theta_{01} \le \theta_{02} \le \ldots \le \theta_{0q}$.

III. Model Misspecification

Experimenter usually assumes the linear predictor of the form

$$\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\boldsymbol{\beta}, \tag{9}$$

Hence, the estimated mean response is

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\pi}(\mathbf{x}) = \mathbf{h}(\boldsymbol{\eta}(\mathbf{x})), \tag{10}$$

However, the correct form the linear predictor may be different. Suppose, the correct form of the linear predictor is

$$\boldsymbol{\eta}_T(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\boldsymbol{\beta} + \mathbf{f}(\mathbf{x}) \tag{11}$$

Then, the estimated mean response using the correct form of linear predictor is

$$\boldsymbol{\mu}_T(\mathbf{x}) = \boldsymbol{\pi}_T(\mathbf{x}) = \mathbf{h}(\boldsymbol{\eta}_T(\mathbf{x})) = \mathbf{h}(\mathbf{Z}(\mathbf{x})\boldsymbol{\beta} + \mathbf{f}(\mathbf{x})), \tag{12}$$

Reference [6] provides an approximate bias and variance of $\hat{\boldsymbol{\beta}}$ for univariate generalized linear models when the linear predictor is misspecified. We generalize it for ordinal response models.

Let $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_n]^T$ and $\boldsymbol{\mu}_T = [\mu_{T,1}, \mu_{T,2}, \ldots, \mu_{T,n}]^T$ with $\boldsymbol{\mu}_i = \mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta}_0)$ and $\boldsymbol{\mu}_{T,i} = \mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta} + \mathbf{f}(\mathbf{x}_i))$, where $\boldsymbol{\beta}_0$ is the model parameter vector. Let us denote $\mathbf{W}_i(\boldsymbol{\beta}) = \mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i^T(\boldsymbol{\beta})$ and $\mathbf{W}_{T,i}(\boldsymbol{\beta}) = \mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_{T,i}\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i^T(\boldsymbol{\beta})$, where $\mathbf{D}_i(\boldsymbol{\beta}) = \frac{d\mathbf{h}(\boldsymbol{\eta}_i)}{d\boldsymbol{\eta}}$ is the derivative of $\mathbf{h}(\boldsymbol{\eta})$ evaluated at $\boldsymbol{\eta}_i = \mathbf{Z}_i\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_{T,i}$ are the variance of $\mathbf{y}_i$ under assumed model and true model respectively. Finally, let $\mathbf{P}$ be the $nq \times nq$ diagonal matrix having each diagonal element $\frac{1}{n}$, $\mathbf{D} = diag(\mathbf{D}_i(\boldsymbol{\beta}_0))$, $\boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_i)$, $\mathbf{W} = diag(\mathbf{W}_i(\boldsymbol{\beta}_0))$, $\mathbf{W}_T = diag(\mathbf{W}_{T,i}(\boldsymbol{\beta}_0))$ and $\mathbf{X} = [\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_n^T]^T$. Then

the approximate bias and variance of $\hat{\boldsymbol{\beta}}$ are given by

$$Bias(\hat{\boldsymbol{\beta}}) = \mathbf{H}_n^{-1}\mathbf{b}, \quad (13)$$

$$Var(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}, \quad (14)$$

where $\mathbf{b} = \mathbf{X}^T\mathbf{P}\mathbf{D}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu})$, $\tilde{\mathbf{H}}_n = \mathbf{X}^T\mathbf{P}\mathbf{W}_T\mathbf{X}$ and $\mathbf{H}_n = \mathbf{X}^T\mathbf{P}\mathbf{W}\mathbf{X} - \mathbf{R}$ with

$$\mathbf{R} = \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{q}\mathbf{Z}_i^T\mathbf{U}_{ir}(\boldsymbol{\beta}_0)\mathbf{Z}_i(y_{ir} - \mu_{ir}(\boldsymbol{\beta}_0)),$$

where $\mathbf{U}_{ir}(\boldsymbol{\beta}_0) = \frac{\mathbf{d}^2\mathbf{u}_r(\mathbf{Z}_i\boldsymbol{\beta}_0)}{d\boldsymbol{\eta}d\boldsymbol{\eta}^T}$.

Now, using the above expressions of bias and variance of $\hat{\boldsymbol{\beta}}$, the mean squared error (MSE) of $\hat{\boldsymbol{\mu}}(\mathbf{x})$ is given by

$$\begin{aligned} &MSE(\hat{\boldsymbol{\mu}}(\mathbf{x})) = \\ &\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T}\mathbf{Z}(\mathbf{x})Var(\hat{\boldsymbol{\beta}}) \times \\ &\mathbf{Z}^T(\mathbf{x})\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T}^T \\ +\ &\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T}[\mathbf{Z}(\mathbf{x})Bias(\hat{\boldsymbol{\beta}}) - \mathbf{f}(\mathbf{x})] \times \\ &[\mathbf{Z}(\mathbf{x})Bias(\hat{\boldsymbol{\beta}}) - \mathbf{f}(\mathbf{x})]^T\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\eta_T}^T \end{aligned} \quad (15)$$

The derivation of the above expressions for $Bias(\hat{\boldsymbol{\beta}})$, $Var(\hat{\boldsymbol{\beta}})$ and $MSE(\hat{\boldsymbol{\mu}}(\mathbf{x}))$ are given in Appendix A and B.

IV. Comparison Criteria for Designs

Here, we want to compare designs in ordinal regression models under linear predictor misspecification. The MLEs of the unknown parameters are biased due to model misspecification. So, we compare designs on the basis of MSEP of the estimated mean response as it takes into account both the variance and bias of the parameter estimates. In the univariate response case MSEP is a scaler. So, in univariate case, if MSEP of a design D_1 is smaller than that of design D_2, then we say that design D_1 has better prediction capabilities than design D_2. However, for multivariate case, MSEP is a matrix. So, we have to consider a scaler valued function of MSEP matrix for such comparison. Several scaler valued functions of a matrix such as determinant, largest eigen value, etc exist in literature. For our numerical example, we consider the largest eigen value of MSEP, denoted as EMSEP, for comparing the designs.

From (15), it is observed that, MSEP at a point $\mathbf{x}$ depends on unknown parameter vector $\boldsymbol{\beta}$, and unknown function $\mathbf{f}$. Hence, EMSEP is a function of $\boldsymbol{\beta}$, $\mathbf{f}$, and $\mathbf{x}$. For, evaluating EMSEP at a point $\mathbf{x}$, we need to have values of $\boldsymbol{\beta}$, and $\mathbf{f}(\mathbf{x})$. We use the MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ for evaluating EMSEP at $\mathbf{x}$. In the next section, we describe an algorithm for estimating $\mathbf{f}(\mathbf{x})$.

A. Algorithm for Estimating $\mathbf{f}$

We first generalize the estimation technique of [7] for the multinomial response model case. This method can be used if an initial data set is available.

Let $\mathbf{d(x)}$ be the difference between the estimated and the true response at a design point $\mathbf{x}$,

$$\mathbf{d(x)} = \boldsymbol{\pi}_T(\mathbf{x}) - \boldsymbol{\pi}(\mathbf{x}) = \mathbf{h}[\boldsymbol{\eta}(\mathbf{x}) + \mathbf{f(x)}] - \mathbf{h}[\boldsymbol{\eta}(\mathbf{x})]. \quad (16)$$

Using first order Taylor series expansion around $\boldsymbol{\eta}(\mathbf{x})$, we have

$$\mathbf{d(x)} = \left[\frac{\partial\mathbf{h}\{\boldsymbol{\eta}(\mathbf{x})\}}{\partial\boldsymbol{\eta}(\mathbf{x})}\right]\mathbf{f(x)}. \quad (17)$$

Thus, the estimate of $\mathbf{f}$ is given by,

$$\hat{\mathbf{f}}(\mathbf{x}) = \left[\frac{\partial\mathbf{h}\{\boldsymbol{\eta}(\mathbf{x})\}}{\partial\boldsymbol{\eta}(\mathbf{x})}\right]^{-1}_{\hat{\boldsymbol{\eta}}(\mathbf{x})}\hat{\mathbf{d}}(\mathbf{x}), \quad (18)$$

where $\hat{\mathbf{d}}(\mathbf{x}) = \bar{\mathbf{y}}(\mathbf{x}) - \hat{\boldsymbol{\pi}}(\mathbf{x}) = \bar{\mathbf{y}}(\mathbf{x}) - \mathbf{h}[\mathbf{Z}(\mathbf{x})\hat{\boldsymbol{\beta}}]$, and $\hat{\boldsymbol{\beta}}$ is the MLEs of $\boldsymbol{\beta}$ using assumed form of linear predictor $\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\boldsymbol{\beta}$.

By estimating $\mathbf{f}$ by the above method, we expect to have better fit using true form of linear predictor over assumed form of linear predictor. That means the deviance using true form of linear predictor should be less than that of the fitted model using assumed form of linear predictor. However, it is observed that, as $\mathbf{f}$ is estimated using first order Taylor series approximation, the deviance is increased or not decreased significantly using true form of linear predictor for some data sets. So, when the approximation is not close enough then we don't get desired results. Here, we propose an algorithm which ensure that we will get better fit using true form of linear predictor over assumed form of linear predictor of the model. In this algorithm, the function $\mathbf{f}(\mathbf{x})$ is estimated using an iterative method. Let us denote for $l \geq 1$, $\boldsymbol{\eta}^{(l)}(\mathbf{x})$, $\mathbf{d}^{(l)}(\mathbf{x})$, and $\mathbf{f}^{(l)}(\mathbf{x})$ for the estimates of $\boldsymbol{\eta}(\mathbf{x})$, $\mathbf{d}(\mathbf{x})$, and $\mathbf{f}(\mathbf{x})$ respectively at lth iteration. Also, let $Dev(l)$ be the deviance of the fitted model using $\boldsymbol{\eta}^{(l)}(\mathbf{x})$. For $l = 1$, $\boldsymbol{\eta}^{(1)}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the MLEs of $\boldsymbol{\beta}$ using assumed form of linear predictor $\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\boldsymbol{\beta}$, $\mathbf{d}^{(1)}(\mathbf{x}) = \bar{\mathbf{y}}(\mathbf{x}) - \mathbf{h}[\boldsymbol{\eta}^{(l)}(\mathbf{x})]$, and $\mathbf{f}^{(1)}(\mathbf{x}) = \left[\frac{\partial\mathbf{h}\{\boldsymbol{\eta}(\mathbf{x})\}}{\partial\boldsymbol{\eta}(\mathbf{x})}\right]^{-1}_{\boldsymbol{\eta}^{(1)}(\mathbf{x})}\mathbf{d}^{(1)}(\mathbf{x})$. Now for $l \geq 1$, the algorithm goes as follows

$$\begin{aligned} \boldsymbol{\eta}^{(l+1)}(\mathbf{x}) &= \boldsymbol{\eta}^{(l)}(\mathbf{x}) + r\mathbf{f}^{(l)}(\mathbf{x}), \\ \mathbf{d}^{(l+1)}(\mathbf{x}) &= \bar{\mathbf{y}}(\mathbf{x}) - \mathbf{h}[\boldsymbol{\eta}^{(l+1)}(\mathbf{x})], \text{ and} \\ \mathbf{f}^{(l+1)}(\mathbf{x}) &= \left[\frac{\partial\mathbf{h}\{\boldsymbol{\eta}(\mathbf{x})\}}{\partial\boldsymbol{\eta}(\mathbf{x})}\right]^{-1}_{\boldsymbol{\eta}^{(l+1)}(\mathbf{x})}\mathbf{d}^{(l+1)}(\mathbf{x}). \end{aligned} \quad (19)$$

The above iteration is continued till $Dev(l)$ is close to zero, i.e., $Dev(l) < \epsilon$ for some chosen $\epsilon > 0$. Here, $r \in (0, 1]$ is a correction factor suitably chosen so that the iteration is conversed. In several example, it is observed that smaller values of r ensure the convergence of the algorithm. However, it increases the number of iteration. So, we have to choose r suitably so that, the iteration is conversed as well as it reduces the computation time.

Note the above described method can be used to estimate $\mathbf{f}$ only at the design points. However, we need to compute EMSEP at any point in the region $\mathcal{R}$, thus an estimate of $\mathbf{f}$ at any point in the design region is necessary. Using multivariate kriging ([8], [9]) details in Appendix C, we estimate $\mathbf{f}$ at any point in the design region. The estimates of $\mathbf{f}$ at the design points obtained by the above algorithm act as a training data

set for kriging.

B. Quantile Dispersion Graphs for Comparing Designs

As we mentioned in Section IV, it is observed that the EMSEP values are functions of $\mathbf{f}(\mathbf{x})$, $\boldsymbol{\beta}$ and $\mathbf{x}$. So, we may denote the EMSEP values corresponding to design D as a function $\tau_D[\mathbf{x}, \boldsymbol{\beta}, \mathbf{f}(\mathbf{x})]$. By using the method described in Section IV-A, we can get the estimates $\hat{\mathbf{f}}(\mathbf{x})$ of $\mathbf{f}(\mathbf{x})$, and the dependence on $\mathbf{f}$ can be removed by using the estimates $\hat{\mathbf{f}}(\mathbf{x})$. Here, it is assumed that $\boldsymbol{\beta}$ lies in the parameter space $\mathcal{C}$. If an initial data set is available, we use the $100(1-\alpha)\%$ confidence region of $\boldsymbol{\beta}$ ([10]),

$$\mathcal{C} = \{\boldsymbol{\gamma} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})'[\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}) \leq \chi^2_{\alpha,p}\}. \tag{20}$$

as the parameter space $\mathcal{C}$. To compare over the entire experimental region $\mathcal{R}$, we partition $\mathcal{R}$ into several concentric regions $\mathcal{R}_\nu$ by reducing its boundary by a shrinkage factor ν. The values of ν are so chosen that the concentric regions cover the entire region $\mathcal{R}$. For a design D and a fixed $\boldsymbol{\beta} \in \mathcal{C}$ the quantiles of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$ are computed for $\mathbf{x} \in \mathcal{R}_\nu$. The $\mathfrak{p}$th quantile of design D is denoted by $Q_D(\mathfrak{p}, \boldsymbol{\beta}, \nu)$.

To address the dependency on $\boldsymbol{\beta}$, a subset of values of $\boldsymbol{\beta}$ is selected from $\mathcal{C}$ and denoted by C, $Q_D(\mathfrak{p}, \boldsymbol{\beta}, \nu)$ values are then found for $\boldsymbol{\beta} \in C$. For a fixed $\mathfrak{p}$ and ν the minimum and maximum quantiles,

$$Q_D^{\min}(\mathfrak{p}, \nu) = min_{\boldsymbol{\beta} \in C}\{Q_D(\mathfrak{p}, \boldsymbol{\beta}, \nu)\},\ \mathfrak{p} \in [0, 1], \tag{21}$$
$$Q_D^{\max}(\mathfrak{p}, \nu) = max_{\boldsymbol{\beta} \in C}\{Q_D(\mathfrak{p}, \boldsymbol{\beta}, \nu)\},\ \mathfrak{p} \in [0, 1]. \tag{22}$$

are computed over the values of $\boldsymbol{\beta}$ in C. By plotting $Q_D^{min}(\mathfrak{p}, \nu)$ and $Q_D^{max}(\mathfrak{p}, \nu)$ against the probabilities $\mathfrak{p} \in [0, 1]$, we get the quantile dispersion graphs (QDGs) for design D over the region $\mathcal{R}_\nu$.

Small and close values of $Q_D^{\min}(\mathfrak{p}, \nu)$ and $Q_D^{\max}(\mathfrak{p}, \nu)$ are desirable. Small values of the minimum and maximum quantiles indicate that the design D has good prediction capability in the presence of linear predictor misspecification. While, close values of $Q_D^{\min}(\mathfrak{p}, \nu)$ and $Q_D^{\max}(\mathfrak{p}, \nu)$ imply that the design is robust to changes in the parameter values $\boldsymbol{\beta}$.

V. Examples

The examples is based on simulated data where $q = 3$ (four categories) and there are two covariates x_1 and x_2. For generating the data set, the true linear predictor is

$$\begin{aligned}
\eta_{T,1}(\mathbf{x}) &= -4.2 + 2.5x_1 - 3.7x_2 + f_{T,1}(\mathbf{x}),\\
\eta_{T,2}(\mathbf{x}) &= -3.1 + 2.5x_1 - 3.7x_2 + f_{T,2}(\mathbf{x}),\\
\eta_{T,3}(\mathbf{x}) &= -1.5 + 2.5x_1 - 3.7x_2 + f_{T,3}(\mathbf{x}),
\end{aligned}$$

where the function $\mathbf{f}_T = [f_{T,1}, f_{T,2}, f_{T,3}]'$ is defined by

$$\begin{aligned}
f_{T,1}(\mathbf{x}) &= -|5\pi \sin(2\pi x_1)\cos(x_2)|,\\
f_{T,2}(\mathbf{x}) &= |\pi \sin(2\pi x_1)\cos(x_2)|,\\
f_{T,3}(\mathbf{x}) &= 6|\pi \sin(2\pi x_1)\cos(x_2)|,
\end{aligned}$$

and the true link function is the multivariate logistic link function. The responses are taken at the design points of design D_1, a 5×7 factorial design and there are $n_i = 6$ experimental units at each run. Designs and the corresponding responses are given in Table I.

Suppose the experimenter assumed the following model,

$$\begin{aligned}
\eta_1(\mathbf{x}) &= \theta_{01} + \theta_1 x_1 + \theta_2 x_2,\\
\eta_2(\mathbf{x}) &= \theta_{02} + \theta_1 x_1 + \theta_2 x_2\\
\eta_3(\mathbf{x}) &= \theta_{03} + \theta_1 x_1 + \theta_2 x_2
\end{aligned} \tag{23}$$

and the multivariate logistic link function. The parameter estimates are shown in Table II. The resultant deviance is 77.1884 with 30 degrees of freedom (p value$<$ 0.0001), indicating a lack of fit. A possible cause of the large deviance may be due to a misspecified linear predictor. Suppose we misspecified the linear function by a function $\mathbf{f}$ which is unknown, then the true $\boldsymbol{\eta}$ can be written as,

$$\begin{aligned}
\eta_1(\mathbf{x}) &= \theta_{01} + \theta_1 x_1 + \theta_2 x_2 + f_1(\mathbf{x}),\\
\eta_2(\mathbf{x}) &= \theta_{02} + \theta_1 x_1 + \theta_2 x_2 + f_2(\mathbf{x}),\\
\eta_3(\mathbf{x}) &= \theta_{03} + \theta_1 x_1 + \theta_2 x_2 + f_3(\mathbf{x}),
\end{aligned} \tag{24}$$

where $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})]'$.

Our interest here is to study the effect of the misspecification in the linear predictor on design selection. Suppose we compare the performance of design D_1 under misspecification with two other designs D_2 (uniform shell design) and D_3 (central composite design). Using QDGs we will study the effect of model misspecification on the prediction capabilities of the three designs and choose the one which is most robust under the current misspecification. The design points of D_2 and D_3 are listed in Table I, note they have the same number of experimental runs as the original design D_1.

The prediction capabilities of the three designs are compared on the basis of their EMSEP values. Note that EMSEP depends on the unknown function $\mathbf{f}(\mathbf{x})$, the parameter vector $\boldsymbol{\beta}$ and the design points $\mathbf{x}$.

To compare the designs on the entire experimental region, $\mathcal{R}$ is divided into several concentric regions, $\mathcal{R}_\nu$, given by

$$\begin{aligned}
\mathcal{R}_\nu &= \{\mathbf{x} : lb_1 \leq x_1 \leq ub_1, lb_2 \leq x_2 \leq ub_2\},\\
&\quad 0.5 < \nu \leq 1,
\end{aligned} \tag{25}$$

with $lb_i = a_i + (1-\nu)(b_i - a_i)$ and $ub_i = b_i - (1-\nu)(b_i - a_i)$. Here, a_i and b_i are the upper and lower bounds of x_i for $i = 1, 2$. For this example we choose $a_1 = -1$, $a_2 = -1$, $b_1 = 1$ and $b_2 = 1$. From the boundary of each region $\mathcal{R}_\nu$, 1000 points are selected.

For estimation of $\mathbf{f}(\mathbf{x})$ at the selected points in $\mathcal{R}$ we proceed in the following way. First we use the method described in Section IV-A to compute $\hat{\mathbf{f}}(\mathbf{x})$ at the design points of D_1. Then, using $\hat{\mathbf{f}}(\mathbf{x})$ at the design points as a training data, multivariate kriging is applied to find $\hat{\mathbf{f}}(\mathbf{x})$ at all points in the experimental region $\mathcal{R}$, where, $\mathcal{R} = \{\mathbf{x} : -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$.

For kriging, we fit the intercept only model with unknown parameter vector $\boldsymbol{\varpi}_0$,

$$\mathbf{f(x)} = \boldsymbol{\varpi}_0 + \boldsymbol{\Pi}(\mathbf{x}), \tag{26}$$

where $\boldsymbol{\Pi}(\mathbf{x})$ is a stationary Gaussian process with zero mean, unknown variance and correlation function. The form of the correlation function is assumed to be the generalized

TABLE I
DESIGNS D_1 (5×7 FACTORIAL), D_2 (UNIFORM SHELL DESIGN) AND D_3 (CENTRAL COMPOSITE DESIGN) AND THE RESPONSES FOR THE EXAMPLE; THERE ARE $n_i = 6$ EXPERIMENTAL UNITS AT EACH RUN

D_1		D_2		D_3		Responses			
x_1	x_2	x_1	x_2	x_1	x_2	y_{i1}	y_{i2}	y_{i3}	y_{i4}
-1.0	-1.00	-1.0	0.00	-1	-1	0	0	4	2
-1.0	-0.67	-1.0	0.00	-1	-1	0	1	1	4
-1.0	-0.33	-1.0	0.00	-1	-1	0	0	0	6
-1.0	0.00	-1.0	0.00	-1	-1	0	0	1	5
-1.0	0.33	-1.0	0.00	1	-1	0	0	0	6
-1.0	0.67	-0.5	-0.87	1	-1	0	0	0	6
-1.0	1.00	-0.5	-0.87	1	-1	0	0	0	6
-0.5	-1.00	-0.5	-0.87	1	-1	2	2	1	1
-0.5	-0.67	-0.5	-0.87	-1	1	1	1	3	1
-0.5	-0.33	-0.5	-0.87	-1	1	1	0	3	2
-0.5	0.00	-0.5	0.87	-1	1	2	0	2	2
-0.5	0.33	-0.5	0.87	-1	1	0	0	0	6
-0.5	0.67	-0.5	0.87	1	1	0	0	0	6
-0.5	1.00	-0.5	0.87	1	1	0	0	0	6
0.0	-1.00	-0.5	0.87	1	1	2	2	1	1
0.0	-0.67	0.0	0.00	1	1	0	1	3	2
0.0	-0.33	0.0	0.00	1	0	3	2	1	0
0.0	0.00	0.0	0.00	1	0	1	1	1	3
0.0	0.33	0.0	0.00	1	0	1	3	0	2
0.0	0.67	0.0	0.00	1	0	0	0	1	5
0.0	1.00	0.5	-0.87	-1	0	0	0	0	6
0.5	-1.00	0.5	-0.87	-1	0	5	1	0	0
0.5	-0.67	0.5	-0.87	-1	0	4	2	0	0
0.5	-0.33	0.5	-0.87	-1	0	2	3	0	1
0.5	0.00	0.5	-0.87	0	1	1	1	3	1
0.5	0.33	0.5	0.87	0	1	2	1	2	1
0.5	0.67	0.5	0.87	0	1	0	0	3	3
0.5	1.00	0.5	0.87	0	1	0	0	1	5
1.0	-1.00	0.5	0.87	0	-1	6	0	0	0
1.0	-0.67	0.5	0.87	0	-1	6	0	0	0
1.0	-0.33	1.0	0.00	0	-1	5	1	0	0
1.0	0.00	1.0	0.00	0	-1	4	2	0	0
1.0	0.33	1.0	0.00	0	0	3	1	1	1
1.0	0.67	1.0	0.00	0	0	0	1	2	3
1.0	1.00	1.0	0.00	0	0	0	0	0	6

exponential correlation function given in [8] with unknown parameter vector $\boldsymbol{\Theta}$. The value of the log likelihood function is used for selecting the appropriate model (26). The MLEs of the parameters in the kriging model are given in Table III.

For investigating the dependency of τ_D on the unknown parameter vector $\boldsymbol{\beta}$, 1000 values of $\boldsymbol{\beta}$ are randomly selected from the $100(1-\alpha)\%$ confidence region of $\boldsymbol{\beta}$ (20) to form C. For a fixed $\boldsymbol{\beta} \in C$, $\mathbf{x} \in \mathcal{R}_\nu, \nu \in (0.5, 1]$ and corresponding $\hat{\mathbf{f}}(\mathbf{x})$, the quantiles of τ_D (Q_D) are computed for design D. For each value of $\boldsymbol{\beta} \in C$ the procedure of calculating Q_D is repeated and finally the minimum and maximum quantiles, $Q_D^{\min}(\mathfrak{p}, \nu)$ and $Q_D^{\max}(\mathfrak{p}, \nu)$ for $\mathfrak{p} = 0(0.5)1$ are found. By plotting $Q_D^{\min}$ and $Q_D^{\max}$ against $\mathfrak{p}$, we get the QDGs for design D. The QDGs for designs D_1, D_2, and D_3 are given in Fig. 1 for $\nu = 0.7, 0.8, 0.9, 1.0$.

From Fig. 1, we observe that minimum quantiles of all the three designs are very close to each other for all values of p and ν. For $\nu = 0.8, 0.9, 1.0$, the maximum quantiles of designs D_1 and D_2 are smaller than those of design D_3 for

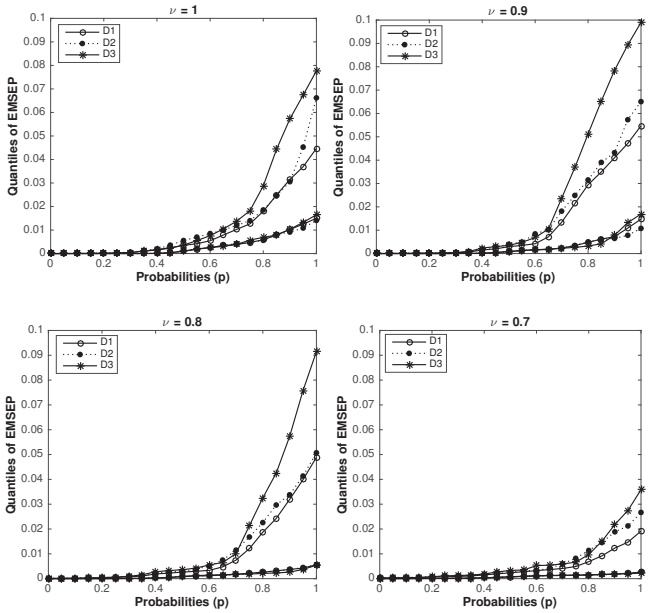

Fig. 1 Quantile dispersion graphs for designs D_1, D_2, and D_3

$\mathfrak{p} > 0.5$. Hence, away from the center of the region, designs D_1 and D_2 have better prediction capabilities than the design D_3 under the misspecification of the linear predictor. However, the maximum quantiles of designs D_1 and D_2 are close to each other implying that designs D_1 and D_2 has comparable prediction capabilities near the boundary of the region. As we move towards the center of the region, the maximum quantiles of designs D_2 and D_3 move close to each other, while the maximum quantiles of the design D_1 are smaller than those of designs D_2 and D_3. Hence, near the center of the experimental region, designs D_2 and D_3 have comparable performances and the design D_1 has better prediction capabilities than designs D_2 and D_3. Another point to note is that the differences between Q_D^{min} and Q_D^{max} values of design D_1 are smaller than those of designs D_2 and D_3, implying that design D_1 is more robust to the changes in the regression parameter vector $\boldsymbol{\beta}$ as compared to D_2 and D_3 throughout the experimental region.

TABLE II
MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS IN THE MODEL (23) FOR THE EXAMPLE

Parameter	Estimate
θ_{01}	-3.7481
θ_{02}	-2.5556
θ_{03}	-1.1148
θ_1	2.5332
θ_2	-3.3203
Deviance=77.1884, DF=30.	

VI. CONCLUDING REMARKS

In this article, we use the QDGs approach to compare designs for ordinal response models when the form of the linear predictor is uncertain. The QDGs approach compares the designs over the entire experimental region not just at a

TABLE III
MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS IN THE MODEL (26) FOR EXAMPLE 1

Parameter	Estimate
ϖ_{01}	-0.1816
ϖ_{02}	-0.1926
ϖ_{03}	0.4125
σ_1^2	0.2672
σ_2^2	0.3341
σ_3^2	0.3731
Θ_1	1.4359
Θ_2	1.8946
Θ_3	0.4125
log likelihood	-44.0035

point like a single value criterion, such as D efficiency. We use an unknown function $\mathbf{f}$ for addressing the misspecification of linear predictor. Multivariate parametric kriging is used to estimate $\mathbf{f}$ at any point in the design region. The proposed method assesses how the prediction capability of the designs change as points are selected from various locations in the experimental region and the robustness of the designs to the changes in the unknown model parameters.

APPENDIX A
DERIVATION OF FORMULA 13 AND 14

The log likelihood function of the responses for multivariate GLMs [4, pp. 105] is given by

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\mu}_i), \tag{27}$$

where $l_i(\boldsymbol{\mu}_i) = \frac{\mathbf{y}_i'\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)}{\phi}\omega_i$ for $i = 1, 2, \ldots, n$. Using the link $\boldsymbol{\mu}_i = \mathbf{h}(\mathbf{Z}_i\boldsymbol{\beta})$, we have

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{Z}_i^T \mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] \tag{28}$$

[4, pp. 105]

and

$$-\frac{d^2 l(\boldsymbol{\beta})}{d\boldsymbol{\beta} d\boldsymbol{\beta}^T} = \sum_{i=1}^{n} \mathbf{Z}_i^T \mathbf{W}_i(\boldsymbol{\beta})\mathbf{Z}_i - \sum_{i=1}^{n}\sum_{r=1}^{q} \mathbf{Z}_i^T \mathbf{U}_{ir}(\boldsymbol{\beta})\mathbf{Z}_i(y_{ir} - \mu_{ir}(\boldsymbol{\beta})), \tag{29}$$

[4, pp. 348]

where $\mathbf{U}_{ir}(\boldsymbol{\beta}) = \frac{\mathbf{d}^2\mathbf{u}_r(\mathbf{Z}_i\boldsymbol{\beta})}{d\boldsymbol{\eta} d\boldsymbol{\eta}^T}$. Now,

$$\begin{aligned} & E(\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}) = \\ & \frac{1}{\sqrt{n}} E\left(\sum_{i=1}^{n} \mathbf{Z}_i^T \mathbf{D}_i(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)]\right) \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{Z}_i^T \mathbf{D}_i(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_{T,i} - \boldsymbol{\mu}_i) \\ & = \sqrt{n}\sum_{i=1}^{n} \frac{1}{n}\mathbf{Z}_i^T \mathbf{D}_i(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_{T,i} - \boldsymbol{\mu}_i) \\ & = \sqrt{n}\mathbf{X}^T\mathbf{P}\mathbf{D}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}) = \sqrt{n}\mathbf{b} \end{aligned} \tag{30}$$

and

$$\begin{aligned} & Var(\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}) \\ = & \sum_{i=1}^{n} \frac{1}{n}\mathbf{Z}_i^T \mathbf{D}_i(\boldsymbol{\beta}_0)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_{T,i}\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i^T(\boldsymbol{\beta}_0)\mathbf{Z}_i \\ = & \mathbf{X}^T\mathbf{P}\mathbf{W}_T\mathbf{X} \end{aligned} \tag{31}$$

By central limit theorem for independent not identical distributed random variables we get $\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}}$ has a multivariate normal distribution with mean $\sqrt{n}\mathbf{b}$ and variance $\mathbf{X}^T\mathbf{P}\mathbf{W}_T\mathbf{X} = \tilde{\mathbf{H}_n}$. Also, from (29) we have

$$\begin{aligned} & -\frac{1}{n}\frac{d^2 l(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta} d\boldsymbol{\beta}^T} = \sum_{i=1}^{n} \frac{1}{n}\mathbf{Z}_i^T \mathbf{W}_i(\boldsymbol{\beta}_0)\mathbf{Z}_i - \\ & \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{q} \mathbf{Z}_i^T \mathbf{U}_{ir}(\boldsymbol{\beta}_0)\mathbf{Z}_i(y_{ir} - \mu_{ir}(\boldsymbol{\beta}_0)) \\ & = \mathbf{X}^T\mathbf{P}\mathbf{W}\mathbf{X} - \mathbf{R} \end{aligned} \tag{32}$$

Now, expanding $\frac{dl(\boldsymbol{\beta})}{d\beta_j}$ around $\boldsymbol{\beta}_0$, we get

$$\begin{aligned} & \frac{dl(\boldsymbol{\beta})}{d\beta_j} = \frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} + \sum_{k=1}^{p}(\beta_k - \beta_{0,k})\frac{d^2 l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k} \\ & + \frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{p}(\beta_k - \beta_{0,k})(\beta_l - \beta_{0,l})\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l}. \end{aligned} \tag{33}$$

Here, β_j and $\beta_{0,j}$ are denoted for the jth term of the vector $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ respectively and $\boldsymbol{\beta}_*$ is the point lying on the line segment joining by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. Now, putting $\hat{\boldsymbol{\beta}}$ (MLE of $\boldsymbol{\beta}$) in (33) and using $\frac{dl(\hat{\boldsymbol{\beta}})}{d\boldsymbol{\beta}} = 0$ we get,

$$\begin{aligned} 0 &= \frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} + \sum_{k=1}^{p}(\hat{\beta}_k - \beta_{0,k})\frac{d^2 l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k} \\ & + \frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{p}(\hat{\beta}_k - \beta_{0,k})(\hat{\beta}_l - \beta_{0,l})\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l} \\ \Rightarrow -\frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} &= \sum_{k=1}^{p}(\hat{\beta}_k - \beta_{0,k})\frac{d^2 l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k} \\ & + \frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{p}(\hat{\beta}_k - \beta_{0,k})(\hat{\beta}_l - \beta_{0,l})\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l} \\ \Rightarrow -\frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} &= \sum_{k=1}^{p}(\hat{\beta}_k - \beta_{0,k})\left[\frac{d^2 l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k}\right. \\ & \left. + \frac{1}{2}\sum_{l=1}^{p}(\hat{\beta}_l - \beta_{0,l})\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j \beta_k d\beta_l}\right] \\ \Rightarrow -\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} &= \sqrt{n}\sum_{k=1}^{p}(\hat{\beta}_k - \beta_{0,k})\left[\frac{1}{n}\frac{d^2 l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k}\right. \\ & \left. + \frac{1}{2n}\sum_{l=1}^{p}(\hat{\beta}_l - \beta_{0,l})\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l}\right] \end{aligned} \tag{34}$$

Now, $\frac{d^3 l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l}$ is bounded when the distribution of response variables follow multinomial distribution. Also, by consistency of $\hat{\boldsymbol{\beta}}$, we have $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to 0$ as $n \to \infty$. Hence, $\frac{1}{2n}\sum_{l=1}^{p}(\hat{\beta}_l -$

$\beta_{0,l})\frac{d^3l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l} \to 0$ as $n \to \infty$. So,

$$\left[\frac{1}{n}\frac{d^2l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k} + \frac{1}{2n}\sum_{l=1}^{p}(\hat{\beta}_l - \beta_{0,l})\frac{d^3l(\boldsymbol{\beta}_*)}{d\beta_j d\beta_k d\beta_l}\right] \to \frac{1}{n}\frac{d^2l(\boldsymbol{\beta}_0)}{d\beta_j d\beta_k} = -H_{jk} \quad (35)$$

Where H_{jk} is the element of the matrix $\mathbf{H}_n = -\frac{1}{n}\frac{d^2l(\boldsymbol{\beta}_0)}{d\boldsymbol{\beta}d\boldsymbol{\beta}^T} = \mathbf{X}^T\mathbf{P}\mathbf{W}\mathbf{X} - \mathbf{R}$ (by (32)). So, the limit distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is the solution of

$$\sum_{k=1}^{p} H_{jk}\sqrt{n}(\hat{\beta}_k - \beta_{0,k}) = \frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_0)}{d\beta_j} \quad (36)$$

This gives the limit distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is the limit distribution of $\mathbf{H}_n^{-1}\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_o)}{d\boldsymbol{\beta}}$. We have shown that $\frac{1}{\sqrt{n}}\frac{dl(\boldsymbol{\beta}_o)}{d\boldsymbol{\beta}}$ follows asymptotically multivariate normal distribution with mean $\sqrt{n}\mathbf{b}$ and variance $\mathbf{X}^T\mathbf{P}\mathbf{W}_T\mathbf{X} = \tilde{\mathbf{H}}_n$. Hence $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has multivariate normal distribution with mean $\sqrt{n}\mathbf{H}_n^{-1}\mathbf{b}$ and variance $\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}$. Hence $Bias(\hat{\boldsymbol{\beta}}) = \mathbf{H}_n^{-1}\mathbf{b}$ and $Var(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}$.

Appendix B
Derivation of formula 15

The mean squared error of $\hat{\boldsymbol{\mu}}(\mathbf{x})$ is given by

$$MSE(\hat{\boldsymbol{\mu}}(\mathbf{x})) = Var(\hat{\boldsymbol{\mu}}(\mathbf{x})) + Bias(\hat{\boldsymbol{\mu}}(\mathbf{x})(Bias\hat{\boldsymbol{\mu}}(\mathbf{x}))^T \quad (37)$$

Since $\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}(\hat{\boldsymbol{\eta}}(\mathbf{x}))$, expanding $\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}(\hat{\boldsymbol{\eta}}(\mathbf{x}))$ around $\boldsymbol{\eta}_T(\mathbf{x})$ and approximating up to first order derivative of the Taylor series, we have,

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}(\hat{\boldsymbol{\eta}}(\mathbf{x})) = \mathbf{h}(\boldsymbol{\eta}_T(\mathbf{x})) + \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T}(\hat{\boldsymbol{\eta}}(\mathbf{x}) - \boldsymbol{\eta}_T(\mathbf{x})), \quad (38)$$

Now, taking expectation on both side of (38), we get

$$\begin{aligned} E(\hat{\boldsymbol{\mu}}(\mathbf{x})) &= E[\mathbf{h}(\boldsymbol{\eta}_T(\mathbf{x})] \\ &+ \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T}(E(\hat{\boldsymbol{\eta}}(\mathbf{x})) - \boldsymbol{\eta}_T(\mathbf{x})) \\ \Rightarrow E(\hat{\boldsymbol{\mu}}(\mathbf{x})) &= \mathbf{h}(\boldsymbol{\eta}_T(\mathbf{x})) \\ &+ \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} Bias[\hat{\boldsymbol{\eta}}(\mathbf{x})] \end{aligned} \quad (39)$$

From (38), we get

$$\begin{aligned} Var(\hat{\boldsymbol{\mu}}(\mathbf{x})) &= Var(\mathbf{h}(\hat{\boldsymbol{\eta}}(\mathbf{x}))) \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} Var(\hat{\boldsymbol{\eta}}(\mathbf{x}))\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]^T_{\boldsymbol{\eta}_T} \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} \mathbf{Z}(\mathbf{x})Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T(\mathbf{x})\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]^T_{\boldsymbol{\eta}_T} \\ &\quad (\text{Since } \hat{\boldsymbol{\eta}}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\hat{\boldsymbol{\beta}}) \end{aligned} \quad (40)$$

Now,

$$\begin{aligned} Bias(\hat{\boldsymbol{\mu}}(\mathbf{x})) &= E[\hat{\boldsymbol{\mu}}(\mathbf{x})] - \boldsymbol{\mu}_T(\mathbf{x}) \\ &= E[\hat{\boldsymbol{\mu}}(\mathbf{x})] - \mathbf{h}(\boldsymbol{\eta}_T(\mathbf{x})) \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} Bias[\hat{\boldsymbol{\eta}}(\mathbf{x})] \text{ (by (39))} \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} [E(\hat{\boldsymbol{\eta}}(\mathbf{x})) - \boldsymbol{\eta}_T(\mathbf{x})] \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} [\mathbf{Z}(\mathbf{x})E(\hat{\boldsymbol{\beta}}) - \mathbf{Z}(\mathbf{x})\boldsymbol{\beta} - \mathbf{f}(\mathbf{x})] \\ &= \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(x))}{d\boldsymbol{\eta}(x)}\right]_{\boldsymbol{\eta}_T} [\mathbf{Z}(\mathbf{x})Bias(\hat{\boldsymbol{\beta}}) - \mathbf{f}(\mathbf{x})] \end{aligned} \quad (41)$$

Hence, from (37), (40) and (41) we have,

$$\begin{aligned} & MSE(\hat{\boldsymbol{\mu}}(\mathbf{x})) \\ = & \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} \mathbf{Z}(\mathbf{x})Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T(\mathbf{x})\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]^T_{\boldsymbol{\eta}_T} \\ & + \left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]_{\boldsymbol{\eta}_T} [\mathbf{Z}(\mathbf{x})Bias(\hat{\boldsymbol{\beta}}) - \mathbf{f}(\mathbf{x})] \times \\ & [\mathbf{Z}(\mathbf{x})Bias(\hat{\boldsymbol{\beta}}) - \mathbf{f}(\mathbf{x})]^T\left[\frac{d\mathbf{h}(\boldsymbol{\eta}(\mathbf{x}))}{d\boldsymbol{\eta}(\mathbf{x})}\right]^T_{\boldsymbol{\eta}_T} \end{aligned} \quad (42)$$

Appendix C
Parametrical Empirical kriging for multiple outputs

Let us consider the regression model where for $\mathbf{x} \in S \subset \mathbb{R}^d$, we have multiple outputs $\Upsilon_1(.), \Upsilon_2(.), \ldots, \Upsilon_m(.)$ and the model is given by [9] (page 102, equation (4.2.1))

$$\Upsilon_i(\mathbf{x}) = \boldsymbol{\Lambda}_i'(\mathbf{x})\boldsymbol{\delta}_i + G_i(\mathbf{x}), \quad (43)$$

where for each $i = 1, 2, \ldots, m$, the $G_i(.)$ has multivariate normal distribution having mean zero, unknown variance and unknown correlation function. Here, $\boldsymbol{\Lambda}_i(.)$ is a $p_i \times 1$ known vector function and $\boldsymbol{\delta}_i$ is a $p_i \times 1$ vector of unknown parameters. For simplicity let us denote $\boldsymbol{\delta} = [\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \ldots, \boldsymbol{\delta}_m']'$ as the $p \times 1$ unknown parameters with $p = \sum_{i=1}^{m} p_i$. For the above model we assume that the covariance between $G_i(\mathbf{x}_1)$ and $G_i(\mathbf{x}_2)$ depends only on $(\mathbf{x}_1 - \mathbf{x}_2)$ and it is given by $Cov\{G_i(\mathbf{x}_1), G_i(\mathbf{x}_2)\} = \sigma_i^2\rho_i(\mathbf{x}_1 - \mathbf{x}_2)$, where $\rho_i(\mathbf{x}_1 - \mathbf{x}_2)$ is the correlation between $\{G_i(\mathbf{x}_1)\}$ and $\{G_i(\mathbf{x}_2)\}$. Also, we assume joint covariance structure of the $\{G_i(.)\}$ given by $Cov\{G_i(\mathbf{x}_1), G_j(\mathbf{x}_2)\} = \sigma_i\sigma_j\rho_{ij}(\mathbf{x}_1 - \mathbf{x}_2)$, where $\rho_{ij}(.)$ is the cross correlation function of $Z_i(.)$ and $Z_j(.)$ for $i \neq j$. Reference [9] discusses various types of correlation functions for $\rho_i(.)$ and $\rho_{ij}(.)$. In our numerical example we use the Generalized exponential correlation function for ρ_i and ρ_{ij} given below. Suppose we need to find $\rho_i(\mathbf{x}_1 - \mathbf{x}_2)$ and $\rho_{ij}(\mathbf{x}_1 - \mathbf{x}_2)$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two realization of $\mathbf{x}$ and let $\boldsymbol{\varepsilon} = \mathbf{x}_1 - \mathbf{x}_2 \in \mathbb{R}^d$. Then $\rho_i(\boldsymbol{\varepsilon})$ and $\rho_{ij}(\boldsymbol{\varepsilon})$ have the form $\zeta(\boldsymbol{\varepsilon}|\boldsymbol{\Theta})$, where $\zeta(\boldsymbol{\varepsilon}|\boldsymbol{\Theta})$ is given by

$$\zeta(\boldsymbol{\varepsilon}|\boldsymbol{\Theta}) = \prod_{k=1}^{d}\exp(-\Theta_k|\varepsilon_k|^{\Theta_{d+1}}),\ 0 < \Theta_{d+1} \leq 2.$$

Suppose, we want to predict $\Upsilon_1(.)$ at new design point $\mathbf{x}_0$. Then the estimate of $\Upsilon_1(\mathbf{x}_0)$, $\hat{\Upsilon}_1(\mathbf{x}_0)$ is given by [9, pp. 107,

equation (4.2.12)]

$$\hat{\Upsilon}_1(\mathbf{x}_0) = \mathbf{\Lambda}_0'\hat{\boldsymbol{\delta}} + \mathbf{r}_0'\mathbf{\Sigma}^{-1}(\mathbf{\Upsilon}^N - \mathbf{F}\hat{\boldsymbol{\delta}}), \tag{44}$$

where $\hat{\boldsymbol{\delta}}$ is the generalized least square estimate of $\boldsymbol{\delta}$ using $N = \sum_{i=1}^{m} n_i$ observations $\mathbf{\Upsilon}^N = [(\mathbf{\Upsilon}_1^{n_1})', (\mathbf{\Upsilon}_2^{n_2})', \ldots, (\mathbf{\Upsilon}_m^{n_m})']'$ with n_i observations for the ith response variable given by $\mathbf{\Upsilon}_i^{n_i} = [\Upsilon_i(\mathbf{x}_1^i), \ldots, \Upsilon_i(\mathbf{x}_{n_i}^i)]'$ at the design points $\mathbf{x}_1^i, \mathbf{x}_2^i, \ldots, \mathbf{x}_{n_i}^i$. Here

$$F = \begin{pmatrix} \mathbf{\Lambda}_1^T(x_0) & \ldots & \mathbf{O}_{1\times p_m} \\ \mathbf{F}_1 & \ldots & \mathbf{O}_{n_1\times p_m} \\ \vdots & \ddots & \vdots \\ \mathbf{O}_{n_m\times p_1} & \ldots & \mathbf{F}_m \end{pmatrix}, \tag{45}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{R}_1 & \tau_2\mathbf{R}_{12} & \ldots & \tau_m\mathbf{R}_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_m\mathbf{R}_{1m}^T & \tau_m\mathbf{R}_{2m}^T & \ldots & \tau_m^2\mathbf{R}_m \end{pmatrix}, \tag{46}$$

$\mathbf{\Lambda}_0' = [\mathbf{\Lambda}_1'(\mathbf{x}_0), \mathbf{O}_{1\times(p-p_1)}]$, and $\mathbf{r}_0' = [\mathbf{r}_{11}', \tau_2\mathbf{r}_{12}', \ldots, \tau_m\mathbf{r}_{1m}']$, where

- $\tau_i = \sigma_i/\sigma_1,\ 2 \le i \le m$,
- $\mathbf{\Lambda}_1(x_0)$ is the vector of regressor at the design point $\mathbf{x}_0$ for $\Upsilon_1(.)$,
- $\mathbf{F}_i = (\mathbf{\Lambda}_i'(\mathbf{x}_l^i))$ is the $n_i \times p_i$ matrix of regressor for ith response,
- $\mathbf{R}_i$ is the matrix of order $n_i \times n_i$ for the correlations between the elements of $\mathbf{\Upsilon}_i^{n_i}$,
- $\mathbf{r}_{1i}$ is the $n_i \times 1$ vector of correlations between $\Upsilon_1(x_0)$ and $\mathbf{\Upsilon}_i^{n_i}$, and
- $\mathbf{R}_{ij}$ is the matrix of order $n_i \times n_j$ for the correlations between $\mathbf{\Upsilon}_i^{n_i}$ and $\mathbf{\Upsilon}_j^{n_j}$.

The unknown parameter vector $\mathbf{\Theta}$ for the correlation function is estimated using maximum likelihood estimation.

REFERENCES

[1] M. A. Heise and R. H. Myers, "Optimal designs for bivariate logistic regression," *Biometrics*, vol. 52, pp. 613–624, 1996.
[2] S. S. Zocchi and A. C. Atkinson, "Optimum experimental designs for multinomial logistics models," *Biometrics*, vol. 55, pp. 437–443, 1999.
[3] S. Mukhopadhyay and A. I. Khuri, "Comparison of designs for multivariate generalized linear models," *Journal of Statistical Planning and Inference*, vol. 138, pp. 169–183, 2008.
[4] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. New York: Springer, 2001.
[5] G. Tutz, *Regression for categorical data*. Cambridge University Press, 2011, vol. 34.
[6] A. J. Adewale and X. Xu, "Robust designs for generalized linear models with possible overdispersion and misspecified link functions," *Computational Statistics and Data Analysis*, vol. 54, pp. 875–890, 2010.
[7] A. J. Adewale and D. P. Wiens, "Robust designs for misspecified logistic models," *Journal of Statistical Planning and Inference*, vol. 139, pp. 3–15, 2009.
[8] S. N. Lophaven, H. B. Nielsen, and J. Sondergaard, "A matlab kriging toolbox," Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, Tech. Rep., 2002.
[9] T. J. Santner, B. J. Williams, and W. Notz, *The Design And Analysis of Computer Experiments*. Springer-Verlag, 2003.
[10] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, vol. 54, pp. 426–482, 1943.