D-Optimal Design for a Seemingly Unrelated Linear Model

Moudar Soumaya[1,2*], Rainer Schwabe[2]

1. Department of Statistics, Faculty of Sciences, Al-Baath University, Homs, P.O. Box 77, Syrian Arab Republic
2. Institute for Mathematical Stochastics, Otto-von-Guericke University, PF 4120, 39016 Magdeburg, Germany

Abstract: In applications often more than one dependent variable is observed in each experimental unit. In some of these situations the explanatory variables may be adjusted separately for the components in these models. For example, if one is interested in both pharmacokinetics and pharmacodynamics, the time points need not be identical for the measurements of the two quantities within one subject. As the observations will be correlated within one unit, the data may be described by a multivariate model, which has the structure of a seemingly unrelated regression.

Keywords: multivariate linear model, seemingly unrelated regression, D-optimal design, product type design.

1 Introduction

The model of seemingly unrelated regression (SUR) has been introduced by Zellner (1962) and since then various types of it are playing an important role in many areas of science. The determination of D-optimal designs for such models, which is the aim of the present work, can be based on a multivariate equivalence theorem by Fedorov (1972) and is related to results by Kurotschka and Schwabe (1996), where design problems for multivariate experiments are reduced to their univariate counter-parts, and uses techniques concerning product type designs derived by Schwabe (1996). By this tools for SUR models, in which an intercept is included in each component, it can be shown that the D-optimal design can be generated as the product of marginal designs, which are D-optimal in the univariate marginal models for the components.

The paper is organized as follows: In the second section we specify the model and characterize optimal designs in the third section. In section 4 the results are illustrated by means of a simple example. Finally, section 5 contains some conclusions.

2 Model specification

The model contains m-dimensional multivariate observations for n individuals. The components of the multivariate observations can be heterogeneous, which means that the response can be described by different regression functions and different experimental settings, which may be chosen from different experimental regions. Then the observation of the jth component of individual i can be described by

$$Y_{ij} = \mathbf{f}_j(x_{ij})^\top \boldsymbol{\beta}_j + \varepsilon_{ij} = \sum_{l=1}^{p_j} f_{jl}(x_{ij})\beta_{jl} + \varepsilon_{ij}\,, \tag{1}$$

where $\mathbf{f}_j = (f_{j1}, ..., f_{jp_j})^\top$ are the known regression functions and $\boldsymbol{\beta}_j = (\beta_{j1}, ..., \beta_{jp_j})^\top$ the unknown parameter vectors for the jth component and the experimental setting x_{ij} may be chosen from an experimental region $\mathcal{X}_j$.

Denote by $\mathbf{Y}_i = (Y_{i1}, ..., Y_{im})^\top$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im})^\top$ the multivariate vectors of observations and error terms, respectively, for individual i and correspondingly the block diagonal multivariate

*Corresponding author: moudar.soumaya@st.ovgu.de

regression function

$$\mathbf{f}(\mathbf{x}) = \operatorname{diag}\left(\mathbf{f}_j(x_j)\right)_{j=1,\dots,m} = \begin{pmatrix} \mathbf{f}_1(x_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{f}_m(x_m) \end{pmatrix}$$

for the multivariate experimental setting $\mathbf{x} = (x_1, ..., x_m)^\top \in \mathcal{X} = \times_{j=1}^n \mathcal{X}_j$.

Then the individual observation vector ca be written as

$$\mathbf{Y}_i = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i\,, \tag{2}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_m^\top)^\top$ is the complete stacked parameter vector for all components. For the error vectors $\boldsymbol{\varepsilon}_i$ it is assumed that they have zero mean, that they are uncorrelated across the individuals and that they have a common positive definite covariance matrix $\operatorname{Cov}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$ within the individuals.

Finally, denote by $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_n^\top)^\top$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, ..., \boldsymbol{\varepsilon}_n^\top)^\top$ the stacked vectors of all observations and all error terms, respectively. Then we can write the complete observation vector as

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \varepsilon\,, \tag{3}$$

where $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), ..., \mathbf{f}(\mathbf{x}_n))^\top$ is the complete experiment design matrix. The complete observational error $\boldsymbol{\varepsilon}$ then has the covariance matrix $\mathbf{V} = \operatorname{Cov}(\boldsymbol{\varepsilon}) = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and "$\otimes$" denotes the Kronecker product.

If we assume the covariance matrix $\boldsymbol{\Sigma}$ and, hence $\mathbf{V}$ known, we can estimate the parameter $\boldsymbol{\beta}$ efficiently by the Gauss-Markov estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{GM}} = (\mathbf{F}^T\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{V}^{-1}\mathbf{Y} \tag{4}$$

with its covariance matrix equal to the inverse of the corresponding information matrix

$$\mathbf{M} = \mathbf{F}^T\mathbf{V}^{-1}\mathbf{F} = \sum_{i=1}^n \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top \tag{5}$$

which is the sum of the individual information

Remark 2.1 *The univariate marginal models of the components have the following form*

$$\mathbf{Y}^{(j)} = \mathbf{F}^{(j)}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}^{(j)}\,, \tag{6}$$

where $\mathbf{Y}^{(j)} = (Y_{1j}, ..., Y_{nj})^\top$ *and* $\boldsymbol{\varepsilon}^{(j)} = (\varepsilon_{1j}, ..., \varepsilon_{nj})^\top$ *are the vectors of observations and errors for the* j*th component, respectively, and* $\mathbf{F}^{(j)} = (\mathbf{f}_j(x_{1j}), ..., \mathbf{f}_j(x_{nj}))^\top$ *is the design matrix for the* j*th marginal model. The corresponding error terms are uncorrelated and homoscedastic,* $Cov(\boldsymbol{\varepsilon}^{(j)}) = \sigma_j^2\mathbf{I}_n$, *where* $\sigma_j^2 = \sigma_{jj}$ *is the* j*th diagonal entry of* $\boldsymbol{\Sigma}$.

3 Optimal designs

We can define an experimental design $\xi = \begin{pmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_k \\ w_1 & \dots & w_k \end{pmatrix}$ by the set of all different experimental settings $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$, $i = 1, ..., k$, with the corresponding relative frequencies $w_i = \frac{n_i}{n}$, where n_i is the number of replications at $\mathbf{x}_i$, Then the corresponding standardized information matrix can be obtained as

$$\mathbf{M}(\xi) = \sum_{i=1}^k w_i \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top. \tag{7}$$

For analytical purposes we consider approximate designs, for which the weights $w_i \geq 0$ need not be multiples of $\frac{1}{n}$, but only have to satisfy $\sum_{i=1}^{k} w_i = 1$. As information matrix are not necessarily comparable, we have to take some real-valued criterion function of the information matrix. In this paper we will adopt the most popular criterion of D-optimality, which aims at maximizing the determinant of the information matrix. This is equivalent to minimizing the volume of the confidence ellipsoid in the case of Gaussian noise: A design ξ^* is said to be D-optimal, if $\det \mathbf{M}(\xi^*) \geq \det \mathbf{M}(\xi)$ for all other competing designs ξ.

A useful tool for checking the performance of a given candidate design is a multivariate version of the equivalence theorem for D-optimality (see Fedorov (1972), theorem 5.2.1):

Theorem 3.1 *The approximate design ξ^* is D-optimal in the multivariate linear model if and only if*

$$\text{trace}\left(\Sigma^{-1}\mathbf{f}(\mathbf{x})^\top \mathbf{M}(\xi^*)^{-1}\mathbf{f}(\mathbf{x})\right) \leq p\,, \tag{8}$$

for all $x \in \mathcal{X}$, where $p = \sum_{j=1}^{m} p_j$ is the number of parameters in the model.

The quantity $\varphi(\mathbf{x};\xi) = \text{trace}\,(\Sigma^{-1}\mathbf{f}(\mathbf{x})^\top(\mathbf{M}(\xi))^{-1}\mathbf{f}(\mathbf{x}))$ will often be called the sensitivity function of the design ξ, which shows which experimental settings are "most informative". In particular, for an optimal design ξ^* the sensitivity function attains its maximum p at the support points.

The quality of a competing design ξ can be measured in terms of its D-efficiency

$$\text{eff}_D(\xi) = \left(\frac{\det \mathbf{M}(\xi)}{\det \mathbf{M}(\xi^*)}\right)^{1/p} \tag{9}$$

compared to the D-optimal design ξ^*. The efficiency states, how much less observations are required, when the optimal design ξ^* is used instead of ξ.

To obtain a complete characterization of the D-optimal designs we have to require that all marginal models related to the components contain an intercept, $f_{j1}(\mathbf{x}) \equiv 1$, say. Then the following general result holds.

Theorem 3.2 *Let ξ_j^* be D-optimal for the jth marginal component (6) on the marginal design region $\mathcal{X}_j$ with an intercept included, $j = 1, ..., m$, then the product type design*

$$\xi^* = \otimes_{j=1}^{m} \xi_j^* \tag{10}$$

is D-optimal for the SUR model (3) on the design region $\mathcal{X} = \times_{j=1}^{m} \mathcal{X}_j$.
The sensitivity function φ does not depend on $\mathbf{\Sigma}$.

The proof is based on an application of the equivalence theorem 3.1 after an orthogonalization with respect to the constant regression functions f_{j1}. Theorem 3.2 may fail to hold, if the regression functions of the marginal components do not contain an intercept.

4 Example: Bivariate straight line regression

To illustrate the results we consider the SUR model with simple straight line regression models for the components,

$$Y_{ij} = \beta_{j0} + \beta_{j1}x_{ij} + \varepsilon_{ij}\,. \tag{11}$$

on the unit interval $\mathcal{X}_1 = \mathcal{X}_2 = [0,1]$ as experimental regions. Then it is well-known that the D-optimal designs for the marginal models $\xi_1^* = \xi_2^* = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$ assign equal weights to each

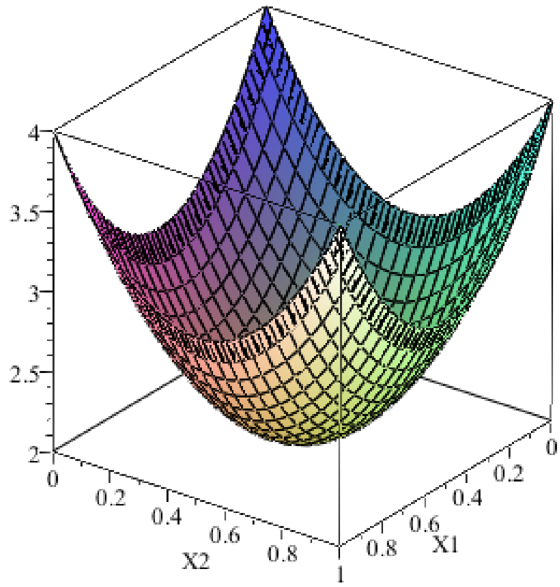

Figure 1: Sensitivity function φ for the D-optimal design $\xi_1^* \otimes \xi_2^*$

of the endpoint of the interval. By Theorem 3.2 the product type design

$$\xi^* = \xi_1^* \otimes \xi_2^* = \begin{pmatrix} (1,1) & (0,0) & (1,0) & (0,1) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

is D-optimal for the SUR model (11) on $\mathcal{X} = [0,1]^2$.

The corresponding sensitivity function

$$\varphi(\mathbf{x}; \xi^*) = \text{trace}(\mathbf{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top \mathbf{M}(\xi^*)^{-1}\mathbf{f}(\mathbf{x})) = 4 - 4x_1 + 4x_1^2 - 4x_2 + 4x_2^2 \tag{12}$$

is plotted in figure 1. It can be easily seen that the sensitivity function is independent of $\mathbf{\Sigma}$ and satisfies the condition $\varphi(\mathbf{x}; \xi^*) \leq p = 4$ for all $\mathbf{x} \in \mathcal{X}$.

An obvious alternative would be a multivariate linear regression design

$$\xi_0 = \begin{pmatrix} (1,1) & (0,0) \\ 1/2 & 1/2 \end{pmatrix},$$

where $x_1 = x_2$ is required and the corresponding marginals of ξ_0 are optimal in the marginal models. While the statistical analysis would simplify for such a design, as the Gauss-Markov estimator reduces to ordinary least squares for any $\mathbf{\Sigma}$, the D-efficiency

$$\text{eff}_D(\xi) = (1 - \varrho^2)^{1/4}$$

compared to the D-optimal design ξ^* depends heavily on the correlation $\varrho = \sigma_{12}/(\sigma_1\sigma_2)$ and tends to zero as $|\varrho|$ tends to one. The corresponding behavior is depicted in figure 2.

5 Conclusions

While the data analysis is well developed for SUR models, there seemed to be no results available on design optimization in such models so far. To fill this gap we establish that under certain regularity conditions D-optimal designs for seemingly unrelated regression and related multivariate linear models can be generated as products of the D-optimal designs for the corresponding

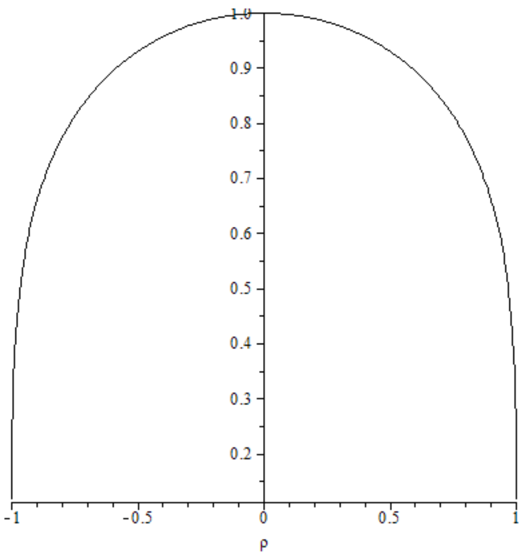

Figure 2: D-efficiency of the multivariate linear regression design ξ_0

univariate models of the single components. This construction turns out to yield optimal designs independent of the covariance structure of the components. Thus design optimization for SUR models can be reduced to univariate problems, for which the theory is well developed. In the special case that the components share the same model structure it might be tempting to simplify the design by letting the experimental settings equal across all components within each unit. Then the observations would result in a MANOVA or multivariate regression model and the analysis would be essentially facilitated. However, an example shows that the efficiency of such MANOVA designs may substantially decrease, if the correlation between the components increases.

References

[1] FEDOROV V.V. (1972): Theory of Optimal Experiments. *Academic Press, New York.*

[2] KUROTSCHKA, V.G. AND SCHWABE, R.(1996): The Reduction of Design Problems for Multivariate Experiments to Univariate Possibilities and their Limitations . *In: E. Brunner and M. Denker (Eds.), Research Developments in Probability and Statistics. Festschrift in Honor of Madan L. Puri on the Occasion of his 65th Birthday. VSP, Utrecht 193–204.*

[3] SCHWABE, R. (1996): Optimum Designs for Multi-factor Models.*Springer, New York.*

[4] ZELLNER, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias. *Journal of the American Statistical Association, 57, 348-368.*