Journal of Multivariate Analysis 142 (2015) 48–56

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Optimal design for multivariate observations in seemingly unrelated linear models

Moudar Soumaya [a,*], Norbert Gaffke [b], Rainer Schwabe [b]

[a] Process Systems Engineering, Otto-von-Guericke University Magdeburg, PF 4120, 39016 Magdeburg, Germany

[b] Institute for Mathematical Stochastics, Otto-von-Guericke University Magdeburg, PF 4120, 39016 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 13 February 2015
Available online 14 August 2015

AMS subject classifications:
62K05
62H12

Keywords:
Multivariate linear model
Seemingly unrelated regression
Optimal design
Product type design

ABSTRACT

The concept of seemingly unrelated models is used for multivariate observations when the components of the multivariate dependent variable are governed by mutually different sets of explanatory variables and the only relation between the components is given by a fixed covariance within the observational units. A multivariate weighted least squares estimator is employed which takes the within units covariance matrix into account. In an experimental setup, where the settings of the explanatory variables may be chosen freely by an experimenter, it might be thus tempting to choose the same settings for all components to end up with a multivariate regression model, in which the ordinary and the least squares estimators coincide. However, we will show that under quite natural conditions the optimal choice of the settings will be a product type design which is generated from the optimal counterparts in the univariate models of the single components. This result holds even when the univariate models may change from component to component. For practical applications the full factorial product type designs may be replaced by fractional factorials or orthogonal arrays without loss of efficiency.

1. Introduction

In an experiment often more than one dependent variable is observed for each observational unit. Sometimes for these dependent components the explanatory variables may be adjusted separately. For example, one might be interested in some processes over time like pharmacokinetics and pharmacodynamics, where observations can be made at the same subjects, but where the time points need not be identical for the measurements of the different quantities within one subject. As typically observations are correlated within units, the data are properly described by a multivariate model with separate sets of explanatory variables.

Such models have been introduced by Zellner [8] in econometrics and have been called seemingly unrelated regression (SUR) models, because the corresponding univariate models for the components do not seem to have anything in common at a first glance. However, it has been pointed out that the correlation between the variables could be employed to transfer useful information from one component to another. Since its introduction various modifications have been considered in observational studies, and the corresponding statistical analysis has been well developed.

In experimental situations these seemingly unrelated regression models have been used less frequently, and, to the best knowledge of the authors, no explicit result is available for the construction of optimal designs in such experiments. There

* Corresponding author.
E-mail address: moudar.soumaya@gmail.com (M. Soumaya).

http://dx.doi.org/10.1016/j.jmva.2015.07.011

is a certain belief that it is sufficient to choose optimal marginal designs for the single components. However, we will show in this paper that additionally the joint distribution of the marginal designs plays an important role, because information may be transferred between the components. In particular, we will establish that product type designs are optimal if all univariate models related to the components contain a constant term (intercept or grand mean). The optimal designs for the multivariate model will be constructed as products of their corresponding counterparts which are optimal in the univariate models of the components. These full factorial product type designs may be replaced by fractional factorials or by suitable orthogonal arrays, which are also optimal, because only the two-dimensional marginals of the designs are involved in the optimization. We will also show that these designs outperform competitors which do not share the factorial structure.

Proofs will be based on Fedorov's [1] multivariate versions of the equivalence theorems for optimal designs. Some techniques concerning product type designs are adopted from Schwabe [5], and there seems to be a relation between univariate additive models and the seemingly unrelated linear models treated here. However, in contrast to the univariate additive case the optimality of product type designs is not restricted to the commonly used *D*-criterion, but carries over to linear criteria and, in particular, to the *A*- and *IMSE*-criteria.

The paper is organized as follows: in the second section we specify the model and collect some relevant issues of optimal designs in the third section. In Section 4 we present the optimality of product type designs and illustrate their performance by an example in the bivariate case in Section 5. Section 6 contains some discussion of the results. Technical proofs are deferred to Appendix.

2. Model specification

We consider multivariate linear models in which m-dimensional observations $\mathbf{Y}_i$ depend on some explanatory variables for n experimental units $i = 1, \ldots, n$. The components (variables) Y_{ij} of the multivariate observations $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^\top$ are assumed to be seemingly unrelated. This means that the settings x_{ij} of the explanatory variables may differ across the components. More generally, we even allow for different univariate linear models for the components, i.e. different explanatory variables, different regression functions and different experimental regions. This model approach covers and generalizes the concept of seemingly unrelated regression (SUR) by Zellner [8] and may also contain components of the analysis of variance type or with both qualitative and quantitative factors of influence.

For each component j the observation Y_{ij} of unit i is specified by a linear model

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

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

The combined observational vector $\mathbf{Y}_i$ can then be written as a multivariate linear model

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

where $\mathbf{f}$ is the block diagonal multivariate regression function

$$\mathbf{f}(\mathbf{x}) = \operatorname{diag}(\mathbf{f}_j(x_j))_{j=1,\ldots,m} = \begin{pmatrix} \mathbf{f}_1(x_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{f}_m(x_m) \end{pmatrix} \tag{3}$$

for the multivariate experimental setting $\mathbf{x} = (x_1, \ldots, x_m)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ is the stacked parameter vector of dimension $p = \sum_{j=1}^m p_j$ for all components and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})^\top$ the multivariate vector of error terms for unit i. We assume additionally that the components x_{ij} of the multivariate experimental setting $\mathbf{x}$ may be chosen independently for each component resulting in a rectangular form of the experimental region, $\mathbf{x} \in \mathcal{X} = \times_{j=1}^m \mathcal{X}_j$. To assure estimability it is further assumed that the components of $\mathbf{f}_j$ are linearly independent functions on $\mathcal{X}_j$ for each $j = 1, \ldots, m$, which implies that the components of $\mathbf{f}$ are linearly independent functions on $\mathcal{X}$.

To complete the model the $\boldsymbol{\varepsilon}_i$ are assumed to be zero mean error vectors with homogeneous non-singular covariance matrix $\boldsymbol{\Sigma} = \operatorname{Cov}(\boldsymbol{\varepsilon}_i)$ which are uncorrelated across the units. Hence, the observational vectors $\mathbf{Y}_i$ inherit the covariance structure from the error terms, $\operatorname{Cov}(\mathbf{Y}_i) = \boldsymbol{\Sigma}$, and the covariance structure does not depend on the experimental settings $\mathbf{x}_i$.

Typically observations will be collected in a data matrix with individual observational vectors $\mathbf{Y}_i$ as rows. Following Zellner [8] the observational vector of the whole experiment is usually obtained by vectorization of the data matrix, i.e. by stacking the columns on top of each other which represent the components. This emphasizes the relation of the multivariate model to its univariate components.

In contrast to this common approach we will vectorize the transposed data matrix by stacking the individual observation $\mathbf{Y}_i$ on top of each other: for the whole experiment denote the stacked vectors of all observations and all error terms