

# Bayesian $D$ -optimal designs for error-in-variables models

*to the memory of Kathryn Chaloner*

Maria Konstantinou and Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

## Abstract

Bayesian optimality criteria provide a robust design strategy to parameter misspecification. We develop an approximate design theory for Bayesian  $D$ -optimality for non-linear regression models with covariates subject to measurement errors. Both maximum likelihood and least squares estimation are studied and explicit characterisations of the Bayesian  $D$ -optimal saturated designs for the Michaelis-Menten, Emax and exponential regression models are provided. Several data examples are considered for the case of no preference for specific parameter values, where Bayesian  $D$ -optimal saturated designs are calculated using the uniform prior and compared to several other designs, including the corresponding locally  $D$ -optimal designs, which are often used in practice.

Keywords: error-in-variables models, classical errors, Bayesian optimal designs,  $D$ -optimality

AMS Subject Classification: 62K05

## 1 Introduction

Locally optimal designs, as termed by Chernoff [1953], depend on the model parameters when the model generating the data is nonlinear. In many cases these parameters are unknown at the design stage and therefore, a best guess of the parameter values is required for the locally optimal designs to be used in practice. This approach however, can result in inefficient designs

if the parameters are misspecified. Hence there is the need to overcome this dependence and construct robust designs that estimate the model parameters with high precision and thus perform well even when there is imperfect knowledge of the true parameter values.

In many practical situations some information about the parameter values, such as a range of plausible values, can be provided by the experimenter. Based on such an uncertainty space a robust design strategy is that of Bayesian optimal designs introduced by Pronzato and Walter [1985], Chaloner [1989] and Chaloner and Larntz [1989]. Bayesian optimality incorporates the parameter uncertainty in the formulation of the optimality criteria through a prior distribution on the parameter space and the proposed criteria are based on classical optimality criteria (see, for example, Chaloner [1993] and Chaloner and Verdinelli [1995] for more details). Therefore, many of the well established results of classical design theory can be directly extended to the Bayesian framework. The construction of Bayesian optimal designs for several regression models has been studied by many authors such as Chaloner and Larntz [1992], Dette and Neugebauer [1997], Han and Chaloner [2003], Dette et al. [2007] and Burghaus and Dette [2014].

In this paper we investigate Bayesian optimal designs for a class of error-in-variables models, that is, of regression models where one or more of the covariates involved cannot be observed directly. The relationship between the true (unobserved) and observed covariates is described by the error model and according to its structure a distinction is made between the classical and Berkson errors. For a detailed review see, for example, Fuller [1987] and Carroll et al. [1995]. Our focus is on classical errors which include the sampling and instrument recording errors frequently arising in practice.

Despite of their importance, the literature on optimal designs for error-in-variables models with classical errors is rather scarce [see Keeler and Reilly [1992] and Dovi et al. [1993] for early references]. Recently, Konstantinou and Dette [2015] develop an approximate optimal design theory for local optimality criteria in error-in-variables models with classical errors and provide analytical results on locally  $D$ -optimal designs for some commonly used nonlinear models when these are subject to the classical error structure. This paper extends their work and provides the corresponding approximate design theory for Bayesian optimality. We thus obtain designs which are optimal for parameter estimation and robust over the specified parameter space.

In Section 2 we introduce the approximate design problem in the context of error-in-variables models subject to classical errors and present the limiting properties of the maximum likelihood and least squares estimators. The approximate design theory for Bayesian optimality is then provided in Section 3 along with the general equivalence theorem and a sufficient condition for Bayesian  $D$ -optimality for maximum likelihood and least squares estimation, respectively. In Section 4 we provide analytical characterisations of Bayesian  $D$ -optimal saturated designs for the Michaelis-Menten, Emax and exponential regression models when these are subject to classical errors. Finally, in Section 5 we consider the case of a uniform prior on the parameter space. Via several data examples, we establish the superiority of the resulting Bayesian  $D$ -

optimal designs by comparing them to the corresponding locally  $D$ -optimal designs, explicitly defined in Konstantinou and Dette [2015], as well as to other designs frequently used in practice.

## 2 Approximate designs and parameter estimation

We assume that the observations are generated by a nonlinear model and consider a repeated observations set-up under which a total of  $r_i$  ( $i = 1, \dots, n$ ) measurements are taken at each of the fixed experimental conditions  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . We further assume that one is unable to observe the true covariate values  $\mathbf{x}_i$ 's,  $i = 1, \dots, n$ , directly due to measurement errors such as sampling and instrumental error. Therefore, a classical error model specifying the conditional distribution of the observed given the true (unobserved) covariates is considered. Throughout this paper we assume classical additive errors, that is,

$$\begin{aligned} Y_{ij} &= m(\mathbf{x}_i, \boldsymbol{\theta}) + \eta_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, r_i \\ \mathbf{X}_{ij} &= \mathbf{x}_i + \boldsymbol{\varepsilon}_{ij}, \end{aligned} \tag{2.1}$$

where  $\boldsymbol{\theta} = (\theta_0, \dots, \theta_p)^T$  is the vector of unknown model parameters,  $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})^T \in \mathcal{X}$  is the vector of true covariates with  $\mathcal{X} \subset \mathbb{R}^q$  denoting the design space and  $\mathbf{X}_{ij}$  denotes the observed vector of the  $j$ th repeated measurement at the  $i$ th experimental condition. Furthermore, the vectors  $(\eta_{ij}, \boldsymbol{\varepsilon}_{ij})^T$  of response errors  $\eta_{ij}$  and covariate errors  $\boldsymbol{\varepsilon}_{ij}$  are assumed to be independent and identically normally distributed with mean  $\mathbf{0}$  and variance covariance matrix  $\Sigma_{\eta\varepsilon}$  being positive definite and the regression function  $m(\mathbf{x}, \boldsymbol{\theta})$  is continuous and twice differentiable with respect to both  $\mathbf{x}$  and  $\boldsymbol{\theta}$ .

We consider approximate designs in the sense of Kiefer [1974] which are defined as probability measures on the design space  $\mathcal{X}$  with finite support. Using the limiting relation

$$\lim_{r \rightarrow \infty} \frac{r_i}{r} = \omega_i > 0, \quad i = 1, \dots, n,$$

where  $r = \sum_{i=1}^n r_i$  denotes the total sample size, an approximate design is of the form

$$\xi = \left\{ \begin{array}{ccc} \mathbf{x}_1 & \dots & \mathbf{x}_n \\ \omega_1 & \dots & \omega_n \end{array} \right\}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^n \omega_i = 1,$$

where the  $\mathbf{x}_i$ 's and  $\omega_i$ 's are called support points and weights of the design, respectively. The goal of the experiment is to estimate the parameters of the underlying model in (2.1) involving the true covariates. In an error-in-variables models set-up however, the true covariate values are unobservable and thus unknown. Therefore, an approximate design provides the experimenter with target values for the true covariates  $\mathbf{x}_i$ ,  $i = 1, \dots, n$  which he would then try to achieve through the observed covariate values  $\mathbf{X}_{ij}$ ,  $i = 1, \dots, n; j = 1, \dots, r_i$ .

Following the methodology in Fuller [1987], Konstantinou and Dette [2015] derived the asymptotic properties of the maximum likelihood and least squares estimators for the parameter vector denoted by  $\hat{\boldsymbol{\theta}}_{ML}$  and  $\hat{\boldsymbol{\theta}}_{LS}$  respectively. In particular, under assumptions of regularity,

$$\sqrt{r}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_{\text{true}}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, M_{ML}^{-1}(\xi, \boldsymbol{\theta})),$$

and

$$\sqrt{r}(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}_{\text{true}}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, M_{LS}^{-1}(\xi, \boldsymbol{\theta})),$$

where  $\xrightarrow{\mathcal{L}}$  denotes convergence in distribution and the information matrices  $M_{ML}(\xi, \boldsymbol{\theta})$  and  $M_{LS}(\xi, \boldsymbol{\theta})$  are given by

$$M_{ML}(\xi, \boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{1}{\sigma_1(\mathbf{x}, \boldsymbol{\theta})} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T d\xi(\mathbf{x}) \quad (2.2)$$

$$M_{LS}(\xi, \boldsymbol{\theta}) = D_0(\xi, \boldsymbol{\theta}) D_1^{-1}(\xi, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta}), \quad (2.3)$$

with

$$D_k(\xi, \boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{[\sigma_1(\mathbf{x}, \boldsymbol{\theta})]^k}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T d\xi(\mathbf{x}), \quad k = 0, 1, \quad (2.4)$$

$$\sigma_k(\mathbf{x}, \boldsymbol{\theta}) = \left( 1, \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}} \right)^T \right) (\Sigma_{\eta^\varepsilon})^k \left( 1, \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}} \right)^T \right)^T, \quad k = 0, 1. \quad (2.5)$$

### 3 Bayesian $D$ -optimal saturated designs

A locally optimal design maximises an appropriate concave functional of the information matrix, here  $M_{ML}(\xi, \boldsymbol{\theta})$  or  $M_{LS}(\xi, \boldsymbol{\theta})$ , called an optimality criterion. In general, locally optimal designs depend on the unknown parameter vector  $\boldsymbol{\theta}$  which must be specified for their implementation. The Bayesian approach on the other hand, takes into account any prior information available for  $\boldsymbol{\theta}$  leading to more robust optimality criteria.

We consider the construction of Bayesian  $D$ -optimal designs introduced by Pronzato and Walter [1985] and Chaloner and Larntz [1989]. Let  $\boldsymbol{\theta} \in \Theta$ , where  $\Theta \subset \mathbb{R}^{p+1}$  and also let  $\pi$  denote a prior distribution on the parameter space  $\Theta$ . A design  $\xi_\pi^*$  is called Bayesian  $D$ -optimal with respect to the prior  $\pi$  for models of the form (2.1) if it maximises the function

$$\Phi_\pi(\xi) = \int_{\Theta} \log |M(\xi, \boldsymbol{\theta})| \pi(d\boldsymbol{\theta}), \quad (3.1)$$

where the information matrix  $M(\xi, \boldsymbol{\theta})$  is that corresponding to maximum likelihood or least squares estimation, given in equations (2.2) and (2.3) respectively, according to the preferable estimation method for the parameter vector.

In the case of maximum likelihood estimation the criterion (3.1) is concave with respect to the design  $\xi$ . Hence using Theorem 3.3 in Dette et al. [2007], the general equivalence theorem for

characterising and checking Bayesian  $D$ -optimality of a candidate design for models of the form (2.1) is given below.

**Theorem 3.1.** *A design  $\xi_\pi^*$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$  for maximum likelihood estimation in model (2.1) if and only if the inequality*

$$\int_{\Theta} d_{ML}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) := \int_{\Theta} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{M_{ML}^{-1}(\xi_\pi^*, \boldsymbol{\theta})}{\sigma_1(\mathbf{x}, \boldsymbol{\theta})} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \pi(d\boldsymbol{\theta}) \leq p + 1,$$

*holds for all  $\mathbf{x} \in \mathcal{X}$ . Furthermore, the maximum is achieved at the support points of  $\xi_\pi^*$ .*

On the other hand, when the vector of model parameters  $\boldsymbol{\theta}$  is estimated via least squares the mapping  $\xi \rightarrow M_{LS}(\xi, \boldsymbol{\theta})$  and thus the optimality criterion (3.1) is not concave. However, the following theorem provides a necessary condition for Bayesian  $D$ -optimality. That is, a design that does not satisfy this condition cannot be Bayesian  $D$ -optimal.

**Theorem 3.2.** *If the design  $\xi_\pi^*$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$  for least squares estimation in model (2.1), then the inequality*

$$\int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) := \int_{\Theta} 2d_0(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) - \sigma_1(\mathbf{x}, \boldsymbol{\theta})d_1(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) \leq p + 1,$$

*holds for all  $\mathbf{x} \in \mathcal{X}$ , where*

$$d_k(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) = \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{D_k^{-1}(\xi_\pi^*, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \pi(d\boldsymbol{\theta}), \quad k = 0, 1.$$

*Furthermore, the maximum is achieved at the support points of  $\xi_\pi^*$ .*

For purposes of comparison with the corresponding locally  $D$ -optimal designs found in Konstantinou and Det [2015], in what follows we study saturated designs. These are designs that have the same number of support points as the dimension  $p + 1$  of the parameter vector  $\boldsymbol{\theta}$ . Lemma 3.1 shows that regardless of the estimation method the Bayesian  $D$ -optimal saturated design is equally weighted.

**Lemma 3.1.** *The Bayesian  $D$ -optimal saturated design with respect to the prior  $\pi$  for maximum likelihood or least squares estimation in model (2.1), puts equal weights at its support points.*

**Remark 3.1.** If we also take into account uncertainty on the response and covariate errors which are assumed to be  $(\eta_{ij}, \boldsymbol{\varepsilon}_{ij})^T \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_{\eta\varepsilon})$  ( $i = 1, \dots, n; j = 1, \dots, r_i$ ), the Bayesian  $D$ -optimality criterion becomes

$$\Phi_{\pi_1, \pi_2}(\xi) = \int_{\Sigma} \int_{\Theta} \log |M(\xi, \boldsymbol{\theta})| \pi_1(d\boldsymbol{\theta}) \pi_2(d\Sigma_{\eta\varepsilon}), \quad (3.2)$$

where  $\pi_1$  is a prior distribution on the parameter space  $\Theta$  and  $\pi_2$  is a prior distribution on the space of positive definite covariance matrices  $\Sigma$ . Then for a Bayesian  $D$ -optimal design with respect to the priors  $\pi_1$  and  $\pi_2$  maximising (3.2), Lemma 3.1 still holds. Furthermore, the statements of Theorems 3.1 and 3.2 are also true, where the integration has to be performed also with respect to the prior  $\pi_2$ .

## 4 Application to specific nonlinear models

In this section we specify the underlying regression function  $m(\mathbf{x}_i, \boldsymbol{\theta})$  and in particular, we consider three nonlinear models widely used in applications for the modelling of the dose-response relationship. Namely, we consider the Michaelis-Menten and Emax models specified by

$$m_1(x, \boldsymbol{\theta}) = \frac{\theta_1 x}{(\theta_2 + x)}, \quad (x \in [0, x_u] \subset \mathbb{R}_0^+), \quad (4.1)$$

and

$$m_2(x, \boldsymbol{\theta}) = \theta_0 + \frac{\theta_1 x}{\theta_2 + x}, \quad (x \in [0, x_u] \subset \mathbb{R}_0^+), \quad (4.2)$$

respectively, and also the three-parameter exponential regression model given by

$$m_3(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 e^{-\theta_2 x} \quad (x \in [0, x_u] \subset \mathbb{R}_0^+). \quad (4.3)$$

In the Michaelis-Menten model the parameter  $\theta_1 > 0$  is the maximum achievable response and  $\theta_2 > 0$  is the dose  $x$  where the response is half-maximal. Similarly in the Emax model,  $\theta_1$  and  $\theta_2$  are the asymptotic maximum increase of the response and the dose producing half of the asymptotic maximum effect respectively and  $\theta_0 \geq 0$  is the placebo effect, that is, the response at dose  $x = 0$ . Finally, for the three-parameter exponential regression model the parameter  $\theta_1 > 0$  is involved in the placebo effect along with  $\theta_0 \geq 0$  and  $\theta_2 \in \mathbb{R} \setminus \{0\}$  describes the rate of the dose effect. The construction of optimal designs in the case of no measurement error in the covariates has been discussed by Dette et al. [2010], Rasch [1990] and Han and Chaloner [2003] for the Michaelis-Menten, the Emax and the exponential regression model, respectively, among others.

For the sake of simplicity we further assume that for these univariate models the response and covariate errors are uncorrelated. Therefore, the variance-covariance matrix of the vector measurement errors given in (2.1) and the  $\sigma$ -functions defined in (2.5) become

$$\Sigma_{\eta\varepsilon} = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}, \quad \sigma_0(x, \boldsymbol{\theta}) = 1 + \left( \frac{\partial m(x, \boldsymbol{\theta})}{\partial x} \right)^2, \quad \sigma_1(x, \boldsymbol{\theta}) = \sigma_\eta^2 + \left( \frac{\partial m(x, \boldsymbol{\theta})}{\partial x} \right)^2 \sigma_\varepsilon^2.$$

In the following two theorems we derive the Bayesian  $D$ -optimal saturated designs for maximum likelihood estimation in each of the aforementioned nonlinear models with measurement errors as in (2.1). Using these analytical characterisations the design problem is reduced to finding the solution of an equation in one variable and therefore, the numerical effort for design search reduces substantially.

**Theorem 4.1.** *The Bayesian  $D$ -optimal saturated design with respect to a prior  $\pi$  for maximum likelihood estimation in the Michaelis-Menten model (4.1) with measurement errors as in (2.1)*

is equally supported at points  $x_1^*$  and  $x_u$ , whereas for the Emax model (4.2) with measurement errors as in (2.1) it is equally supported at points 0,  $x_1^*$  and  $x_u$ . The non-trivial support point  $x_1^* \in (0, x_u)$  is a solution of the equation

$$\int_{\Theta} \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} \pi(d\boldsymbol{\theta}) = 0, \quad (4.4)$$

in the interval  $(0, x_u)$ , where  $\varrho_{\varepsilon\eta}^2 = \sigma_{\varepsilon}^2 / \sigma_{\eta}^2$ .

**Theorem 4.2.** *The Bayesian  $D$ -optimal saturated design with respect to a prior  $\pi$  for maximum likelihood estimation in the exponential regression model (4.3) with measurement errors as in (2.1) is equally supported at points 0,  $x_1^*$  and  $x_u$ . The non-trivial support point  $x_1^* \in (0, x_u)$  is a solution of the equation*

$$\int_{\Theta} \frac{1 - e^{\theta_2 x_u} + \theta_2 x_u e^{\theta_2 x_1}}{x_1 - x_u + x_u e^{\theta_2 x_1} - x_1 e^{\theta_2 x_u}} - \frac{\theta_2 e^{2\theta_2 x_1}}{e^{2\theta_2 x_1} + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} \pi(d\boldsymbol{\theta}) = 0, \quad (4.5)$$

in the interval  $(0, x_u)$ , where  $\varrho_{\varepsilon\eta}^2 = \sigma_{\varepsilon}^2 / \sigma_{\eta}^2$ .

The corresponding analytical results for Bayesian  $D$ -optimal saturated designs for least squares estimation are given below. We note that in the case of Theorem 4.4 the two non-trivial support points of the design can only be evaluated using numerical optimisation.

**Theorem 4.3.** *The Bayesian  $D$ -optimal saturated design with respect to a prior  $\pi$  for least squares estimation in the Michaelis-Menten model (4.1) with measurement errors as in (2.1) puts equal masses at points  $x_1^*$  and  $x_u$ , whereas for the Emax model (4.2) with measurement errors as in (2.1) it puts equal masses at points 0,  $x_1^*$  and  $x_u$ . The non-trivial support point  $x_1^* \in (0, x_u)$  is a solution of the equation*

$$\int_{\Theta} \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} + \frac{2\theta_1^2 \theta_2^2}{(\theta_2 + x_1)[(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2]} \pi(d\boldsymbol{\theta}) = 0, \quad (4.6)$$

in the interval  $(0, x_u)$  and  $\varrho_{\varepsilon\eta}^2 = \sigma_{\varepsilon}^2 / \sigma_{\eta}^2$ .

**Theorem 4.4.** *The Bayesian  $D$ -optimal saturated design with respect to a prior  $\pi$  for least squares estimation in the exponential regression model (4.3) with measurement errors as in (2.1) is always supported at the larger end-point  $x_u$  of the design space.*

**Remark 4.1.** With the assumption of uncorelated response and covariate errors, the prior  $\pi_2$  on the space of matrices  $\boldsymbol{\Sigma}$  reduces to a prior on the space of the corresponding error variances  $\sigma_{\eta}^2$  and  $\sigma_{\varepsilon}^2$ , which is a subset of  $\mathbb{R}^2$ . Using the Bayesian  $D$ -optimality criterion (3.2) instead of (3.1), does not affect the characteristics of the Bayesian  $D$ -optimal saturated designs with respect to  $\pi_1$  and  $\pi_2$  which remain the same as described in Theorems 4.1-4.4. In each case,



the value of the non-trivial support point does change however, as it is a solution of a different equation. For example, in the case of Theorem 4.1, the analogous of equation (4.4) is

$$\int_{\mathbb{R}^+} \int_{\Theta} \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} \pi_1(d\boldsymbol{\theta}) \pi_2(d\varrho_{\varepsilon\eta}^2) = 0,$$

where here  $\pi_2$  is a prior on the ratio  $\varrho_{\varepsilon\eta}^2 = \sigma_{\varepsilon}^2 / \sigma_{\eta}^2$  which is induced by the given prior on the space of error variances.

## 5 Data example

The theoretical results of the previous section are now illustrated via several data examples. In what follows, a uniform prior distribution is used on the parameter space corresponding to the case of no preference for specific parameter values. Under this concept there is no need for the experimenter to specify a prior, thereby avoiding a step that is often difficult in practice. A number  $\nu$  of equally spaced values are taken from each of the parameter's uncertainty intervals and the resulting prior points on the entire parameter space are equally likely to be observed. We note that when the “true” parameter values are not specified (Tables 1 and 4) the efficiency of a design  $\xi$  is calculated via

$$\text{eff}(\xi) = \frac{\exp\{\frac{1}{p+1}\Phi_{\pi}(\xi)\}}{\exp\{\frac{1}{p+1}\Phi_{\pi}(\xi_{Bay}^*)\}} = \exp\left\{\frac{1}{p+1}[\Phi_{\pi}(\xi) - \Phi_{\pi}(\xi_{Bay}^*)]\right\}, \quad (5.1)$$

where  $\xi_{Bay}^*$  is the Bayesian  $D$ -optimal saturated design for the prior  $\pi$  and a specific error-ratio value  $\varrho_{\varepsilon\eta}^2$ . Finally, when nominal values for the parameters are considered for the calculations (Tables 2 and 3), we use the  $D$ -efficiency defined for a design  $\xi$  as

$$\text{eff}_D(\xi) = \left( \frac{\det\{M(\xi, \boldsymbol{\theta})\}}{\det\{M(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta})\}} \right)^{1/p+1}, \quad (5.2)$$

where  $p + 1$  is the number of model parameters and  $\xi_{\boldsymbol{\theta}}^*$  is the locally  $D$ -optimal design for errors-in-variables models with classical errors using the parameter values vector  $\boldsymbol{\theta}$ , explicitly defined in Konstantinou and Dette [2015].

We begin with an investigation of how the Bayesian  $D$ -optimal saturated designs change in the presence of error in the covariates. For this purpose we consider an example discussed in Mihara et al. [2000]. These authors model the velocity of a biochemical reaction (CSD-plus pyrovate) with respect to the concentration of a substrate (L-cysteine sulfinat) via the Michaelis-Menten model. The design space in this example is  $\mathcal{X} = [0, 80]$  and the obtained parameter estimates are  $(\theta_1, \theta_2) = (16, 3.5)$ . However, in their set-up Mihara et al. [2000] do not take into account possible errors in the measurement of the substrate concentration. Such



errors are the result of instrument recording errors which, as mentioned in the introduction, correspond to the classical error structure.

To study the case where a parameter space is provided by the experimenter, we use the estimates stated above as a starting point for the choice of an uncertainty space. In particular, we consider the parameter space  $\Theta = [8, 24] \times [1.75, 5.25]$  which corresponds to the  $\pm 50\%$  region around the point of parameter estimates. Using Theorems 4.1 and 4.3 we find the Bayesian  $D$ -optimal two-point designs for maximum likelihood and least squares estimation respectively in the Michaelis-Menten model (4.1) with classical errors as in (2.1). Table 1 presents the support points of these designs for various values of the error-variances ratio  $\varrho_{\varepsilon\eta}^2 = \sigma_\varepsilon^2/\sigma_\eta^2$ . For these calculations two uniform priors are considered, using  $\nu = 5$  and  $\nu = 11$  equally-spaced values from each of the parameter space intervals  $[8, 24]$  and  $[1.75, 5.25]$ . Note that the point  $(\theta_1, \theta_2) = (16, 3.5)$  of the parameter estimates is included in the resulting prior points for both priors. We also calculate the efficiencies (5.1) of the Bayesian designs assuming no error in the covariates, that is,  $\varrho_{\varepsilon\eta}^2 = 0$  which are also given in Table 1. Using a uniform prior with  $\nu = 11$  this design is equally supported at points 3.06 and 80 for maximum likelihood estimation and at points 5.82 and 80 when the parameters are estimated via least squares.

As for the case of locally  $D$ -optimal designs discussed in Konstantinou and Dette [2015], taking the error in the covariate into account results in the non-trivial support point of the Bayesian  $D$ -optimal design to move further away from  $x = 0$  compared to its value when  $\varrho_{\varepsilon\eta}^2$  is assumed to be zero. As the  $\varrho_{\varepsilon\eta}^2$ -value becomes larger, the value of the non-trivial support point increases further. The choice of estimation method also seems to have an effect on the Bayesian  $D$ -optimal design with the non-trivial support point of the design for least squares estimation always being larger. Furthermore, the Bayesian  $D$ -optimal design assuming no error in the covariate has efficiency less than 90% for some error-variances ratio values  $\varrho_{\varepsilon\eta}^2 > 0$ . Most importantly, even if the covariate error variance  $\sigma_\varepsilon^2$  is small but equal to the response error variance  $\sigma_\eta^2$  (hence  $\varrho_{\varepsilon\eta}^2 = 1$ ), the efficiency of the Bayesian design ignoring the covariate error is 82.44% if the parameters are estimated via maximum likelihood. Therefore, the usual strategy of ignoring the covariate error if it is believed to be small can result in inefficient designs. We finally note that the Bayesian design using  $\nu = 11$  turned out to have efficiencies of approximately 100% when compared with the corresponding Bayesian design for  $\nu = 101$ . Hence for the rest of this section the uniform prior with  $\nu = 11$  is used.

We now assess the robustness of Bayesian  $D$ -optimal saturated designs against misspecifications of the model parameters. Konstantinou and Dette [2015] use data discussed in Frisillo and Stewart [1980] on how the wave velocity of ultrasonic signals relates to the percent gas-brine saturation. In this example the use of error-in-variables models is justified by the possible false measurement of the intensity of an X-ray beam, which is then used to determine the percentage of gas-brine saturation (see Frisillo and Stewart [1980] for more details). Konstantinou and Dette [2015] fit the exponential regression model (4.3) to the data and obtain the parameter estimates

Table 1: *Left part: Support points of the Bayesian  $D$ -optimal two-point designs using a uniform prior with  $\nu = 5$  or  $\nu = 11$  values for each parameter. Right part: Efficiencies (%) of the Bayesian  $D$ -optimal two-point design ( $\nu = 11$ ) for  $\varrho_{\varepsilon\eta}^2 = 0$ .*

		Support points		
$\varrho_{\varepsilon\eta}^2$	Estimation Method	$\nu = 5$	$\nu = 11$	Efficiencies (%)
4/1	MLE	(8.02,80)	(8.12,80)	62.92
	LSE	(9.14,80)	(9.21,80)	84.68
2/1	MLE	(6.79,80)	(6.86,80)	72.96
	LSE	(8.14,80)	(8.19,80)	91.48
1/1	MLE	(5.77,80)	(5.82,80)	82.44
	LSE	(7.36,80)	(7.40,80)	95.97
1/2	MLE	(4.94,80)	(4.99,80)	90.11
	LSE	(6.78,80)	(6.82,80)	98.38
1/4	MLE	(4.30,80)	(4.34,80)	95.26
	LSE	(6.37,80)	(6.42,80)	99.44

$(\theta_0, \theta_1, \theta_2) = (1210, 66.07, 0.0696)$ . Then using these estimates and assuming that  $\varrho_{\varepsilon\eta}^2 = 1$ , they find the locally  $D$ -optimal designs on the design space  $\mathcal{X} = [0, 35]$  for model (4.3) with errors as in (2.1), which are given by

$$\xi_{loc}^{MLE} = \begin{pmatrix} 0 & 17.23 & 35 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad \xi_{loc}^{LSE} = \begin{pmatrix} 1.26 & 21.54 & 35 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad (5.3)$$

for maximum likelihood and least squares estimation respectively.

For the construction of the corresponding Bayesian  $D$ -optimal designs we consider the parameter space  $\Theta = \{1210\} \times [33, 100] \times [0.01, 0.3]$ . The estimate for  $\theta_0$  is used since this parameter does not affect the design and the interval used for  $\theta_1$  corresponds roughly to the  $\pm 50\%$  interval around its estimate. The estimate for the parameter  $\theta_2 = 0.0696$  suggests a very small dose effect. Hence the corresponding uncertainty interval is chosen such that larger rates of dose effect

are considered. Theorem 4.2 provides a complete analytical characterisation of the Bayesian  $D$ -optimal three-point design for maximum likelihood estimation. In the case of least squares estimation Theorem 4.4 only provides us with the larger support point and the weights of the design and thus the other two support points were found numerically using the Particle Swarm Optimisation (PSO) algorithm (see, for example, Clerc [2006]). Assuming that  $\varrho_{\varepsilon\eta}^2 = 1$ , the Bayesian  $D$ -optimal three-point designs for maximum likelihood and least squares estimation in model (4.3) with errors as in (2.1) are given by

$$\xi_{Bay}^{MLE} = \begin{pmatrix} 0 & 11.59 & 35 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad \xi_{Bay}^{LSE} = \begin{pmatrix} 6.79 & 16.33 & 35 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}. \quad (5.4)$$

The following two tables present the efficiencies of the locally and Bayesian  $D$ -optimal designs, given in (5.3) and (5.4) respectively, along with the efficiencies of the uniform design  $\xi_{uni}$  allocating equal weights at points 0, 17.5 and 35, for the four end-points of the parameter space. We thus examine the efficiencies (5.2) of these designs when the “true” parameter values are equal to either one of the extreme values of their corresponding uncertainty interval. Table 2 presents the efficiencies for the case of maximum likelihood estimation whereas the corresponding results for least squares estimation are given in Table 3.

Table 2: *Efficiencies (%) of the locally and Bayesian  $D$ -optimal three-point designs for maximum likelihood estimation assuming  $\varrho_{\varepsilon\eta}^2 = 1$  and of the uniform design on  $\mathcal{X} = [0, 35]$ .*

	Efficiencies (%)		
$(\theta_1, \theta_2)$	$\xi_{loc}^{MLE}$	$\xi_{Bay}^{MLE}$	$\xi_{uni}$
(33, 0.01)	99.91	94.25	99.82
(33, 0.3)	31.57	73.09	30.20
(100, 0.01)	100	93.09	99.97
(100, 0.3)	49.30	96.45	47.23
Average	70.20	89.22	69.31

We observe that for both estimation methods the efficiencies of the designs fluctuate, with the Bayesian  $D$ -optimal design always being more robust. Although in the case of least squares estimation the Bayesian design has efficiencies below 90%, it has the largest average efficiency.

Table 3: *Efficiencies (%) of the locally and Bayesian D-optimal three-point designs for least squares estimation assuming  $\varrho_{\varepsilon\eta}^2 = 1$  and of the uniform design on  $\mathcal{X} = [0, 35]$ .*

	Efficiencies (%)		
$(\theta_1, \theta_2)$	$\xi_{loc}^{LSE}$	$\xi_{Bay}^{LSE}$	$\xi_{uni}$
(33, 0.01)	88.61	59.16	99.86
(33, 0.3)	15.17	58.82	24.40
(100, 0.01)	90.94	61.03	99.99
(100, 0.3)	15.39	75.17	24.27
Average	52.53	63.55	62.13

This is due to the fact that the choice of a uniform prior for the construction of the Bayesian  $D$ -optimal designs results in the average, over the parameter space, of the values for the determinant of the information matrix to be maximised. Therefore, when there is no preference for specific parameter values, the use of the Bayesian design is a consistently more efficient choice avoiding the risk of having an extremely inefficient design if the parameters are misspecified.

Throughout this paper the response and covariate errors are assumed to be known at least up to the value of the error-variances ratio  $\varrho_{\varepsilon\eta}^2$ . However,  $\varrho_{\varepsilon\eta}^2$  does affect the optimal choice of the Bayesian design (see Table 1). We thus conclude this section with a robustness assessment of the Bayesian  $D$ -optimal designs with respect to misspecification of the  $\varrho_{\varepsilon\eta}^2$ -value. In particular, we use equation (5.1) to calculate the efficiencies of the Bayesian  $D$ -optimal designs for  $\varrho_{\varepsilon\eta}^2 = 1$ , given in (5.4), and of the uniform design three-point design  $\xi_{uni}$  on  $\mathcal{X} = [0, 35]$  when the Bayesian  $D$ -optimal design  $\xi_{Bay}^*$  corresponds to various other  $\varrho_{\varepsilon\eta}^2$ -values. The results are given in Table 4

It is evident that under both estimation methods the Bayesian  $D$ -optimal design assuming equal response and covariate errors is extremely robust and efficient against misspecification of the error-variances ratio value  $\varrho_{\varepsilon\eta}^2$ . On the contrary the “off the shelf” uniform design commonly used in practice has efficiencies well below 90%.

Table 4: *Efficiencies (%) of the Bayesian D-optimal three-point designs assuming  $\varrho_{\varepsilon\eta}^2 = 1$  and of the uniform three-point designs on  $\mathcal{X} = [0, 35]$ .*

		Efficiencies	
$\varrho_{\varepsilon\eta}^2$	Estimation Method	$\xi_{Bay}$	$\xi_{uni}$
4/1	MLE	97.48	91.51
	LSE	97.66	74.02
2/1	MLE	99.32	86.93
	LSE	99.37	75.13
1/1	MLE	100	81.77
	LSE	100	76.08
1/2	MLE	99.30	76.43
	LSE	99.28	76.99
1/4	MLE	97.34	71.35
	LSE	96.95	78.04

**Acknowledgements** This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG).

## References

- I. Burghaus and H. Dette. Optimal designs for nonlinear regression models with respect to non-informative priors. *Journal of Statistical Planning and Inference*, 154:12–25, 2014.
- R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models*. Taylor & Francis, 1995.
- K. Chaloner. Bayesian design for estimating the turning point of a quadratic regression. *Comm. Statist., Theory Methods*, 18(4):1385–1400, 1989.
- K. Chaloner. A note on optimal bayesian design for nonlinear problems. *Journal of Statistical*

- Planning and Inference*, 37(2):229 – 235, 1993.
- K. Chaloner and K. Larntz. Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21(2):191 – 208, 1989.
- K. Chaloner and K. Larntz. Bayesian design for accelerated life testing. *Journal of Statistical Planning and Inference*, 33(2):245 – 259, 1992.
- K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.
- H. Chernoff. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24(4):586–602, 1953.
- M. Clerc. *Particle Swarm Optimization*. Iste Publishing Company, London, 2006.
- H. Dette and H. M. Neugebauer. Bayesian d-optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, 60:331–349, 1997.
- H. Dette, L. M. Haines, and L. A. Imhof. Maximin and bayesian optimal designs for regression models. *Statistica Sinica*, 17:463–480, 2007.
- H. Dette, C. Kiss, M. Bevanda, and F. Bretz. Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, 97(2):513–518, 2010.
- V. G. Dovi, A. P. Reverberi, and L. Maga. Optimal design of sequential experiments for error-in-variables models. *Computers & Chemical Engineering*, 17(1):111 – 115, 1993.
- A. L. Frisillo and T. J. Stewart. Effect of partial gas/brine saturation on ultrasonic absorption in sandstone. *Journal of Geophysical Research*, 85(B10):5209–5211, 1980.
- W. A. Fuller. *Measurement Error Models*. Wiley, New York, 1987.
- C. Han and K. Chaloner. D-and c-optimal designs for exponential regression models used in pharmacokinetics and viral dynamics. *Journal of Statistical Planning and Inference*, 115: 585–601, 2003.
- S. Keeler and P. Reilly. The design of experiments when there are errors in all the variables. *Canadian Journal of Chemical Engineering*, 70:774–778, 1992.
- J. Kiefer. General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 2(5):849–879, 1974.
- M. Konstantinou and H. Dette. Locally optimal designs for errors-in-variables models. *Biometrika*, 102(4):951–958, 2015.
- H. Mihara, T. Kurihara, T. Yoshimura, and N. Esaki. Kinetic and mutational studies of three NifS homologs from *escherichia coli*: Mechanistic difference between L-Cysteine Desulfurase and L-Selenocysteine Lyase Reactions. *Journal of Biochemistry*, 127(4):559–567, 2000.
- L. Pronzato and E. Walter. Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75(1):103–120, 1985.
- D. Rasch. Optimum experimental design in nonlinear regression. *Comm. Statist. Theory Methods*, 19(12):4789–4806, 1990.

# Appendix

## Proof of of Theorem 3.2

*Proof.* Let  $\xi_\pi^*$  be a Bayesian  $D$ -optimal design with respect to the prior  $\pi$  for least squares estimation in any functional model of the form (2.1). For any other design  $\xi$  and  $a \in [0, 1]$  also let  $\xi_a = (1 - a)\xi_\pi^* + a\xi$ . Then the Frechet derivative of the criterion function  $\Phi_\pi(\xi)$  at  $\xi_\pi^*$  in the direction of  $\xi - \xi_\pi^*$  is

$$\begin{aligned} \frac{d}{da} \Phi_\pi(\xi_a) \big|_{a=0} &= \int_{\Theta} \frac{d}{da} (\log |M_{LS}(\xi_a, \boldsymbol{\theta})|) \big|_{a=0} \pi(d\boldsymbol{\theta}) \\ &= \int_{\Theta} \frac{d}{da} (2 \log |D_0(\xi_a, \boldsymbol{\theta})|) \big|_{a=0} - \frac{d}{da} (\log |D_1(\xi_a, \boldsymbol{\theta})|) \big|_{a=0} \pi(d\boldsymbol{\theta}) \\ &= \int_{\Theta} 2 \text{tr} \{ D_0^{-1}(\xi_\pi^*, \boldsymbol{\theta}) [D_0(\xi, \boldsymbol{\theta}) - D_0(\xi_\pi^*, \boldsymbol{\theta})] \} - \text{tr} \{ D_1^{-1}(\xi_\pi^*, \boldsymbol{\theta}) [D_1(\xi, \boldsymbol{\theta}) - D_1(\xi_\pi^*, \boldsymbol{\theta})] \} \pi(d\boldsymbol{\theta}) \\ &= \int_{\Theta} 2 \text{tr} \{ D_0^{-1}(\xi_\pi^*, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta}) \} - \text{tr} \{ D_1^{-1}(\xi_\pi^*, \boldsymbol{\theta}) D_1(\xi, \boldsymbol{\theta}) \} \pi(d\boldsymbol{\theta}) - (p + 1). \end{aligned}$$

Now using Dirac measures  $\delta_{\mathbf{x}}$  with weight 1 at the support points  $\mathbf{x} \in \mathcal{X}$  of the design  $\xi$  we have that

$$\begin{aligned} \text{tr} \{ D_k^{-1}(\xi_\pi^*, \boldsymbol{\theta}) D_k(\delta_{\mathbf{x}}, \boldsymbol{\theta}) \} &= \text{tr} \left\{ D_k^{-1}(\xi_\pi^*, \boldsymbol{\theta}) \frac{(\sigma_1(\mathbf{x}, \boldsymbol{\theta}))^k}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \\ &= \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{(\sigma_1(\mathbf{x}, \boldsymbol{\theta}))^k}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} D_k^{-1}(\xi_\pi^*, \boldsymbol{\theta}) \left( \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &= (\sigma_1(\mathbf{x}, \boldsymbol{\theta}))^k d_k(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}), \quad k = 0, 1. \end{aligned}$$

Since  $\xi_\pi^*$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$ ,  $d\Phi_\pi(\xi_a)/da \big|_{a=0}$  is non-positive for all designs  $\xi$ , and the inequality  $\int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) \leq p + 1$  for all  $\mathbf{x} \in \mathcal{X}$  follows.

Now let us assume for the Bayesian  $D$ -optimal design  $\xi_\pi^*$  that  $\max_{\mathbf{x} \in \mathcal{X}} \int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) < p + 1$ . This yields that

$$\int_{\mathcal{X}} \int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) d\xi_\pi^*(\mathbf{x}) < (p + 1) \int_{\mathcal{X}} d\xi_\pi^*(\mathbf{x}) = p + 1.$$

On the other hand, it follows from the definition of the functions  $d_0$  and  $d_1$  and a straightforward calculation that for any design  $\xi$

$$\int_{\mathcal{X}} \int_{\Theta} d_0(\mathbf{x}, \xi, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) d\xi(\mathbf{x}) = \int_{\mathcal{X}} \int_{\Theta} \sigma_1(\mathbf{x}, \xi, \boldsymbol{\theta}) d_1(\mathbf{x}, \xi, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) d\xi(\mathbf{x}) = \int_{\Theta} p + 1 \pi(d\boldsymbol{\theta}) = p + 1,$$

which yields  $\int_{\mathcal{X}} \int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) d\xi_\pi^*(\mathbf{x}) = p + 1$  and contradicts our initial assumption. Hence  $\max_{\mathbf{x} \in \mathcal{X}} \int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) = p + 1$  and from  $\int_{\mathcal{X}} \int_{\Theta} d_{LS}(\mathbf{x}, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) d\xi_\pi^*(\mathbf{x}) = p + 1$  it follows that this maximum is attained at each support point of the Bayesian  $D$ -optimal design  $\xi_\pi^*$ .

□



### Proof of Lemma 3.1

*Proof.* Let  $\xi = \{\mathbf{x}_0, \dots, \mathbf{x}_p; \omega_0, \dots, \omega_p\}$  be any saturated design. Also let  $X$  be the  $(p+1) \times (p+1)$  matrix with  $i$ th row given by  $\left(\frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T$ ,  $i = 0, \dots, p$ ,  $W = \text{diag}(\omega_0, \dots, \omega_p)$  and  $L_k = \text{diag}(\sigma_k(\mathbf{x}_0, \boldsymbol{\theta}), \dots, \sigma_k(\mathbf{x}_p, \boldsymbol{\theta}))$  for  $k = 0, 1$ . Under this notation the determinants of the information matrices  $M_{ML}(\xi, \boldsymbol{\theta})$  and  $M_{LS}(\xi, \boldsymbol{\theta})$ , given in (2.2) and (2.3) respectively, become

$$\begin{aligned} \det \{M_{ML}(\xi, \boldsymbol{\theta})\} &= [\det X]^2 [\det L_1]^{-1} [\det W] \\ \det \{M_{LS}(\xi, \boldsymbol{\theta})\} &= \det \{D_0(\xi, \boldsymbol{\theta}) D_1^{-1}(\xi, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta})\} = [\det X]^2 [\det L_0]^{-1} [\det L_1]^{-1} [\det W]. \end{aligned}$$

Hence the criterion (3.1) becomes

$$\int_{\Theta} \log |W| + 2 \log |X| - \log |L_1| \pi(d\boldsymbol{\theta}),$$

and

$$\int_{\Theta} \log |W| + 2 \log |X| - \log |L_0| - \log |L_1| \pi(d\boldsymbol{\theta}),$$

for maximum likelihood and least squares estimation respectively. Maximising any of the above expressions with respect to the weights gives  $\omega_i = 1/(p+1)$ , for all  $i = 0, \dots, p$  since  $\det W = \omega_0 \omega_1 \dots \omega_p$ , which proves the assertion.  $\square$

### Proof of Theorem 4.1

*Proof.* From Lemma 3.1 it follows that a Bayesian  $D$ -optimal saturated design on  $\mathcal{X} = [0, x_u]$  for maximum likelihood estimation in the Michaelis-Menten model with measurement errors as in (2.1), puts equal weights at its support points. For any design  $\xi_{\pi} = \{x_1, x_2; 0.5, 0.5\}$  ( $0 < x_1 < x_2 \leq x_u$ ), the criterion defined in (3.1) for the information matrix given in (2.2) becomes

$$\begin{aligned} \Phi_{\pi}(\xi_{\pi}) &= \int_{\Theta} 2 \log \theta_1 + 2 \log x_1 + 2 \log x_2 + 2 \log(x_2 - x_1) - \log 4 \\ &\quad - \log(\sigma_{\eta}^2(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_{\varepsilon}^2) - \log(\sigma_{\eta}^2(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\varepsilon}^2) \pi(d\boldsymbol{\theta}). \end{aligned}$$

It is easy to check that for fixed  $x_1$ , this is increasing with  $x_2$  and therefore maximized for  $x_2^* = x_u$ . The smaller support point of the optimal design is found by solving  $\partial \Phi_{\pi}(\xi_{\pi}) / \partial x_1 |_{x_2=x_u} = 0$  for  $x_1 \in (0, x_u)$ . This is equivalent to solving

$$\int_{\Theta} \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2\sigma_{\eta}^2(\theta_2 + x_1)^3}{\sigma_{\eta}^2(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_{\varepsilon}^2} \pi(d\boldsymbol{\theta}) = 0, \quad (.5)$$

for  $x_1 \in (0, x_u)$ .

Similarly using Lemma 3.1, a Bayesian  $D$ -optimal saturated design on  $\mathcal{X} = [0, x_u]$  for maximum likelihood estimation in the Emax model with measurement errors as in (2.1), is equally

weighted. Thus for the three-point design  $\xi_\pi = \{x_0, x_1, x_2; 1/3, 1/3, 1/3\}$  ( $0 < x_0 < x_1 < x_2 \leq x_u$ ) we have that

$$\begin{aligned} \Phi_\pi(\xi_\pi) = \int_{\Theta} & 2 \log \theta_1 + 4 \log \theta_2 + 2 \log(x_0 - x_1) + 2 \log(x_0 - x_2) + 2 \log(x_1 - x_2) - \log 27 \\ & - \log(\sigma_\eta^2(\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) \pi(d\theta). \end{aligned}$$

The criterion above is decreasing with  $x_0$  and increasing with  $x_2$  and therefore, it is maximised at  $x_0^* = 0$  and  $x_2^* = x_u$ . The equation  $\partial \Phi_\pi(\xi_\pi) / \partial x_1 |_{x_0=0; x_2=x_u} = 0$  gives again equation (.5).  $\square$

## Proof of Theorem 4.2

*Proof.* Following similar arguments as in the proof of Theorem 4.1, for an equally weighted three-point design  $\xi_\pi = \{x_0, x_1, x_2; 1/3, 1/3, 1/3\}$  ( $0 \leq x_0 < x_1 < x_2 \leq x_u$ ), the criterion defined in (3.1) using the information matrix given in (2.2) for the Emax model (4.3), becomes

$$\begin{aligned} \Phi_\pi(\xi_\pi) = \int_{\Theta} & 2 \log \theta_1 + 2 \log(e^{\theta_2 x_2}(x_0 - x_1) + e^{\theta_2 x_0}(x_1 - x_2) + e^{\theta_2 x_1}(x_2 - x_0)) - \log 27 \\ & - \log(\sigma_\eta^2 e^{2\theta_2 x_0} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2 e^{2\theta_2 x_1} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2 e^{2\theta_2 x_2} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) \pi(d\theta). \end{aligned}$$

It is easy to check that  $e^{\theta_2 x_2}(x_0 - x_1) + e^{\theta_2 x_0}(x_1 - x_2) + e^{\theta_2 x_1}(x_2 - x_0)$  is increasing with  $x_0$ , decreasing with  $x_2$  and negative for all  $x_0 < x_1 < x_2$ . Therefore, the criterion is decreasing with  $x_0$  and increasing with  $x_2$  and thus it is maximized at  $x_0^* = 0$  and  $x_2^* = x_u$ . The middle support point of the Bayesian  $D$ -optimal design is found by solving  $\partial \Phi_\pi(\xi_\pi) / \partial x_1 |_{x_0=0; x_2=x_u} = 0$  for  $x_1 \in (0, x_u)$  which is equivalent to solving

$$\int_{\Theta} \frac{1 - e^{\theta_2 x_u} + \theta_2 x_u e^{\theta_2 x_1}}{x_1 - x_u + x_u e^{\theta_2 x_1} - x_1 e^{\theta_2 x_u}} - \frac{\theta_2 e^{2\theta_2 x_1}}{e^{2\theta_2 x_1} + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} \pi(d\theta) = 0,$$

for  $x_1 \in (0, x_u)$ , where  $\varrho_{\varepsilon\eta}^2 = \sigma_\varepsilon^2 / \sigma_\eta^2$ .  $\square$

## Proof of Theorem 4.3

*Proof.* From Lemma 3.1, it follows that a Bayesian  $D$ -optimal saturated design for least squares estimation assigns equal weights at its support points. For the Michaelis-Menten model with measurement errors as in (2.1) and a two-point equally weighted design  $\xi_\pi = \{x_1, x_2; 0.5, 0.5\}$  ( $0 < x_1 < x_2 \leq x_u$ ), the criterion defined in (3.1) for the information matrix given in (2.3) becomes

$$\begin{aligned} \Phi_\pi(\xi_\pi) = \int_{\Theta} & 2 \log \theta_1 + 2 \log(x_2 - x_1) + 4 \log(\theta_2 + x_1) + 4 \log(\theta_2 + x_2) - \log 4 \\ & - \log(\sigma_\eta^2(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) \\ & - \log((\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2) - \log((\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2) \pi(d\theta). \end{aligned}$$

For fixed  $x_1$ , the above expression is increasing with  $x_2$  and therefore maximized at  $x_2^* = x_u$ . The smaller support point of the optimal design is found by solving  $\partial\Phi_\pi(\xi_\pi)/\partial x_1|_{x_2=x_u} = 0$  for  $x_1 \in (0, x_u)$ , which is equivalent to solving

$$\int_{\Theta} \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_{\varepsilon\eta}^2} + \frac{2\theta_1^2 \theta_2^2}{(\theta_2 + x_1)[(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2]} \pi(d\boldsymbol{\theta}) = 0, \quad (.6)$$

for  $x_1 \in (0, x_u)$ .

In the case of the Emax model with errors as in (2.1), the criterion for a three-point equally weighted design  $\xi_\pi = \{x_0, x_1, x_2; 1/3, 1/3, 1/3\}$  ( $0 \leq x_0 < x_1 < x_2 \leq x_u$ ) becomes

$$\begin{aligned} \Phi_\pi(\xi_\pi) = \int_{\Theta} & 2\log \theta_1 + 4\log \theta_2 + 2\log(x_2 - x_1) + 2\log(x_2 - x_0) + 2\log(x_1 - x_0) + 4\log(\theta_2 + x_0) \\ & + 4\log(\theta_2 + x_1) + 4\log(\theta_2 + x_2) - \log 4 - \log(\sigma_\eta^2(\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) \\ & - \log(\sigma_\eta^2(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log((\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2) \\ & - \log((\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2) - \log((\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2) \pi(d\boldsymbol{\theta}). \end{aligned}$$

It is easy to check, following similar arguments as before, that the above expression is decreasing with  $x_0$ , increasing with  $x_2$  and thus maximized at  $x_0^* = 0$ ,  $x_2^* = x_u$  and  $x_1^*$  is the solution of the same equation as for the Michaelis-Menten model, that is, equation (.6).  $\square$

## Proof of Theorem 4.4

*Proof.* For the three-parameter exponential regression model with measurement errors as in (2.1) and a three-point equally weighted design  $\xi_\pi = \{x_0, x_1, x_2; 1/3, 1/3, 1/3\}$  ( $0 \leq x_0 < x_1 < x_2 \leq x_u$ ), the criterion defined in (3.1) for the information matrix given in (2.3) becomes

$$\begin{aligned} \Phi_\pi(\xi_\pi) = \int_{\Theta} & 2\log \theta_1 + 2\log(e^{\theta_2 x_2}(x_0 - x_1) + e^{\theta_2 x_0}(x_1 - x_2) + e^{\theta_2 x_1}(x_2 - x_0)) - \log 27 \\ & - \log(\sigma_\eta^2 e^{2\theta_2 x_0} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2 e^{2\theta_2 x_1} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) - \log(\sigma_\eta^2 e^{2\theta_2 x_2} + \theta_1^2 \theta_2^2 \sigma_\varepsilon^2) \\ & - \log(1 + \theta_1^2 \theta_2^2 e^{-2\theta_2 x_0}) - \log(1 + \theta_1^2 \theta_2^2 e^{-2\theta_2 x_1}) - \log \log(1 + \theta_1^2 \theta_2^2 e^{-2\theta_2 x_2}) \pi(d\boldsymbol{\theta}). \end{aligned}$$

Following the proof of Theorem 4.2, the criterion is increasing with  $x_2$  and thus it is maximized at  $x_2^* = x_u$ .  $\square$