

Locally optimal designs for errors-in-variables models

Maria Konstantinou, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

maria.konstantinou@ruhr-uni-bochum.de

holger.dette@ruhr-uni-bochum.de

Abstract

This paper considers the construction of optimal designs for nonlinear regression models when there are measurement errors in the predictor. Corresponding (approximate) design theory is developed for maximum likelihood and least squares estimation, where the latter leads to non-concave optimisation problems. For the Michaelis-Menten, EMAX and exponential regression model D -optimal designs can be found explicitly and compared with the corresponding designs derived under the assumption of no measurement error in concrete applications.

AMS Subject Classification: 62K05

Keywords and Phrases: error-in-variable model, optimal design, nonlinear regression, D -optimality

1 Introduction

Errors-in-variables models, also referred to as *measurement error models*, differ from the classical regression models in that one or more of the independent variables involved are subject to measurement error. It goes without saying that in applied work covariates are often measured imprecisely with the source of error varying, for example, from simple instrument or human error to systematic error in questionnaires or error in measuring long-term quantities such as food intake [Carroll et al. (1995)]. Overlooking this error will result in biased parameter estimates [Carroll et al. (1995)] and therefore in inaccurate conclusions to be drawn from the experiment. In view of the importance of measurement

error models, mainly due to their wide range of applications, analysis strategies for errors-in-variables models received much attention leading to a considerably large literature on the subject.

For illustration purposes let us consider a nonlinear regression model involving only one independent variable given by

$$Y_i = m(x_i, \boldsymbol{\theta}) + \eta_i, \quad i = 1, \dots, n,$$

and assume that one cannot observe x_i directly but instead observes

$$X_i = x_i + \epsilon_i, \quad i = 1, \dots, n,$$

where $\boldsymbol{\theta}$ is the vector of unknown model parameters and η_i and ϵ_i are the response and covariate errors respectively. The observed variable X is called the *manifest variable* whereas the true unobserved variable x is called the *latent variable*. Depending on the error structure, measurement error models are classified accordingly and different analysis methods are used. The model is said to be a *structural* or a *functional* model when the true value of the x 's are assumed to be random variables or a sequence of unknown fixed constants respectively. In both cases, the measurement error is independent of the latent variable and is referred to as *classical error*. The third basic measurement error model is the *Berkson error model* under which the observed variable X is fixed and controlled by the experimenter. Measurement error now enters via the setting of the value of the covariate and is independent of the manifest variable. The Berkson error model is used in situations where the experimenter sets target values $X_i, i = 1, \dots, n$, for the covariate but the true values $x_i, i = 1, \dots, n$, used in the experimenter differ from the ones originally designed. For a detailed review of the characteristics and estimation results see, for example, Bard (1974), Fuller (1987), Carroll et al. (1995) and Buonaccorsi (2010).

Based on the analysis methods readily available for measurement error models, the issue of efficient planning of experiments involving such models follows naturally. Due to the broad applicability of errors-in-variables models finding efficient experimental designs is of great interest as these designs minimise both the cost and duration of experiments while at the same time maximize the precision of their conclusions. Despite its importance, the literature on optimal designs for models with error-in-variables is rather scarce. One of the first authors that examined the effects of measurement error on the planning of experiments involving errors-in-variables models is Box (1963). He considers response surface models with Berkson type errors and explores the robustness of factorial and fractional factorial designs. Following Box (1963), Draper and Beggs (1971) and Vuchkov

and Boyadjieva (1983) propose the use of the sum of squared differences of the responses and the maximum element of the information matrix respectively, as a measure of robustness against Berkson type errors. Furthermore, Tang and Bacon-Shone (1992) study the construction of Bayesian optimal designs for the Berkson type probit model.

More recent results on Berkson error models are given in Pronzato (2002) and Donev (2004). Besides the initial assumption of errors in setting the independent variable values, Pronzato (2002) further assumes that the true values, and therefore the actual design used, is known at the end of the experiment. He uses the expected value of the determinant of the information matrix as an extension to the D -optimality criterion. Donev (2004) considers both the cases of unknown and known true values at the end of the study. Under the former scenario he proposes the use of the sum of the variances of the estimated responses as a measure of robustness whereas in the latter case he argues that minimising the expected value of the inverse information matrix is a more appropriate criterion than that proposed by Pronzato (2002).

In contrast to the above literature, little research has been done on how to plan experiments involving a structural or a functional model. To the best of our knowledge there are only two papers considering this type of design problem. Keeler and Reilly (1992) use the classical D -optimality criterion to construct sequential designs for functional models. The measurement errors of the independent variables enter via the maximum likelihood estimation and hence via the information matrix. Structural models are considered in Dovi et al. (1993). Measurement error is again accounted for in the estimation stage however, optimal designs are constructed using the expectation of the D -optimality criterion with respect to the fluctuations of the design variables assumed to be random.

In the present paper we focus on the classical error structure for which measurement error occurs due to recording devices error, administration error, etc. [Carroll et al. (1995)]. We feel this is an assumption arising more frequently in applications compared to the Berkson error structure which assumes, for example, zero measuring device error. Furthermore, we focus on functional rather than structural models because the latter require further assumption to be made on the distribution of the latent variable and even in this case the estimation procedures for functional modelling are robust to misspecifications of the distribution of the independent variable [Carroll et al. (1995)]. In Section 2 we introduce a nonlinear functional model involving a vector of covariates. The limiting properties of both the maximum likelihood and least squares estimators for the vector of unknown parameters are established and the approximate optimal design problem in the context of nonlinear models subject to the functional error structure is introduced. Then in Section 3 we focus on the construction of locally optimal designs with respect

to Kiefer’s Φ_ℓ -optimality criteria for such functional models. We state the general equivalence theorem for maximum likelihood estimation and find a necessary condition for Φ_ℓ -optimality under least squares estimation. In Section 4 we illustrate our approach through an application to the D -optimality criterion for the case of only one explanatory variable and uncorrelated measurement errors. Extensions of the well-known result on the equality of the weights of locally D -optimal saturated designs are also given under both estimation methods. We provide analytical characterisations of locally D -optimal design for three widely used nonlinear models under the functional error structure considered with maximum likelihood and least squares estimation being considered separately. Three real examples are considered in Section 5 for each of which we compare the locally D -optimal designs arising for maximum likelihood and least squares estimation in the Michaelis-Menten error-in-variables model with the D -optimal design for the model under the assumption of no measurement error. Finally, proofs of all results are presented in an Appendix and an Online Supplement, which contains further details.

We conclude this introduction by mentioning once again that the focus of this paper is on locally optimal designs for functional errors-in-variables models, since our goal is to explore the differences which may appear in the optimal designs if the error in the predictor is ignored. These designs require some preliminary information regarding the unknown parameters and there are numerous situations where such information is available, for example, in Phase II dose-finding trials [see Dette et al. (2008)]. Moreover, if such knowledge is not available the theory developed in this paper can be further extended using similar arguments as in Chaloner and Verdinelli (1995) and Dette (1997) to obtain optimal designs with respect to robust optimality criteria.

2 Estimation in functional models

Throughout this paper we consider nonlinear functional models of the form

$$\begin{aligned} Y_{ij} &= m(\mathbf{x}_i, \boldsymbol{\theta}) + \eta_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, r_i \\ \eta_{ij} &\sim N(0, \sigma_\eta^2), \end{aligned} \tag{1}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)^T$ is the vector of unknown model parameters, n is the number of distinct experimental conditions, r_i , $i = 1, \dots, n$, is the number of repeated measurements taken at the i th observation point and η_{ij} , $i = 1, \dots, n$, $j = 1, \dots, r_i$, is the measurement error in the response Y_{ij} . The vector of unknown true values of the independent variables $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iq})^T$, with q not necessarily equal to p , is also subject to measurement

error. In particular, we assume that we observe

$$\begin{aligned}\mathbf{X}_{ij} &= \mathbf{x}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, r_i \\ \boldsymbol{\epsilon}_{ij} &\sim N(0, \sigma_\epsilon^2),\end{aligned}\tag{2}$$

with $\mathbf{x}_i \in \mathcal{X}$ being fixed where $\mathcal{X} \subset \mathbb{R}^q$ denote the design space. We further suppose that the vectors of measurement errors are independent and identically distributed, that is, $(\eta_{ij}, \boldsymbol{\epsilon}_{ij})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\epsilon})$, where

$$\boldsymbol{\Sigma}_{\eta\epsilon} = \begin{pmatrix} \sigma_\eta^2 & \sigma_{\eta, \epsilon_1} & \dots & \sigma_{\eta, \epsilon_q} \\ \sigma_{\epsilon_1, \eta} & \sigma_{\epsilon_1}^2 & \dots & \sigma_{\epsilon_1, \epsilon_q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\epsilon_q, \eta} & \sigma_{\epsilon_q, \epsilon_1} & \dots & \sigma_{\epsilon_q}^2 \end{pmatrix},$$

is a nonsingular matrix. For the sake of brevity we focus on the assumption of a normal distribution for the errors in model (1) and (2). Under other distributional assumptions similar results can be obtained using the methodology introduced in this paper.

2.1 Maximum likelihood estimation

We note that since the vector of measurement errors is normally distributed the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{ML}$ coincides with the weighted least squares estimator with weight matrix equal to $\boldsymbol{\Sigma}_{\eta\epsilon}^{-1}$ and is defined as the first component of $(\boldsymbol{\theta}, \mathbf{x})$ that minimises

$$\sum_{i=1}^n \sum_{j=1}^{r_i} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i) \boldsymbol{\Sigma}_{\eta\epsilon}^{-1} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i)^T.$$

In the rest of the article we only referred to maximum likelihood estimation but nevertheless all our results for this estimation method are also valid for weighted least squares estimation.

Following Fuller (1987), we impose the following assumptions:

- (a) The function $m(\mathbf{x}_i, \boldsymbol{\theta})$ is continuous and has continuous first and second order derivatives with respect to both \mathbf{x}_i and $\boldsymbol{\theta}$.
- (b) For every $\zeta > 0$ there exists a constant $\delta_\zeta > 0$ such that

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{r_i} \inf_{Y_{ij}=m(\mathbf{x}_i, \boldsymbol{\theta})} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i) \boldsymbol{\Sigma}_{\eta\epsilon}^{-1} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i)^T > \delta_\zeta,$$

for all $\boldsymbol{\theta}$ satisfying $|\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{true}}| > \zeta$, where $\boldsymbol{\theta}_{\text{true}}$ denotes the “true” vector of unknown parameters.

The limiting properties of the estimator $\hat{\boldsymbol{\theta}}_{ML}$ as the sample sizes increase are derived under the assumption

$$\lim_{r_i \rightarrow \infty} \frac{r_i}{r} = \omega_i > 0, \quad i = 1, \dots, n, \quad (3)$$

as $r_i \rightarrow \infty$, where $r = \sum_{i=1}^n r_i$. Throughout this paper we collect the information provided by the different experimental conditions and the limiting relation above in the matrix

$$\xi = \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ \omega_1 & \cdots & \omega_n \end{array} \right\}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^n \omega_i = 1. \quad (4)$$

Note that ξ defines a probability measure on the design space \mathcal{X} , and following Kiefer (1974) we call any probability measure of the form (4) an *approximate design* on $\mathcal{X} (\subset \mathbb{R}^q)$. By a slight modification of Theorem 3.2.1 in Fuller (1987) we have that

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

where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution and the inverse of the asymptotic covariance matrix is given by

$$\begin{aligned} M_{ML}(\xi, \boldsymbol{\theta}) &= \int_{\mathcal{X}} \frac{1}{\sigma_1(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) d\xi(\mathbf{x}) \\ &= \sum_{i=1}^n \frac{\omega_i}{\sigma_1(\mathbf{x}_i, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right), \end{aligned} \quad (5)$$

with

$$\sigma_1(\mathbf{x}_i, \boldsymbol{\theta}) = \left(1, \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \mathbf{x}_i} \right) \boldsymbol{\Sigma}_{\eta\epsilon} \left(1, \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \mathbf{x}_i} \right)^T. \quad (6)$$

We conclude this section noting that Fuller (1987) examines a more general set-up where the model cannot be written as in (1). On the other hand, he fixes the sample sizes r_i , $i = 1, \dots, n$, to 1 and considers homoscedastic error variances converging to zero.

2.2 Least squares estimation

The (ordinary) least squares estimator $\hat{\boldsymbol{\theta}}_{LS}$ is defined as the first component of $(\boldsymbol{\theta}, \mathbf{x})$ minimizing the expression

$$\sum_{i=1}^n \sum_{j=1}^{r_i} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i) (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i)^T.$$

Given assumptions (a) and

(b1) For every $\zeta > 0$ there exists $\delta_\zeta > 0$ such that

$$Q'_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{r_i} \inf_{Y_{ij}=m(\mathbf{x}_i, \boldsymbol{\theta})} (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i) (Y_{ij} - m(\mathbf{x}_i, \boldsymbol{\theta}), \mathbf{X}_{ij} - \mathbf{x}_i)^T > \delta_\zeta,$$

for all $\boldsymbol{\theta}$ satisfying $|\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{true}}| > \zeta$.

and following the methodology in the proof of Theorem 3.2.1 in Fuller (1987) it can be shown that

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

where the inverse of the asymptotic covariance matrix is now given by

$$M_{LS}(\xi, \boldsymbol{\theta}) = D_0(\xi, \boldsymbol{\theta}) D_1^{-1}(\xi, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta}). \quad (7)$$

Here the matrices $D_0(\xi, \boldsymbol{\theta})$ and $D_1(\xi, \boldsymbol{\theta})$ are defined as

$$D_0(\xi, \boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{1}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) d\xi(\mathbf{x}), \quad (8)$$

$$D_1(\xi, \boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{\sigma_1(\mathbf{x}, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) d\xi(\mathbf{x}), \quad (9)$$

where the function $\sigma_1(\mathbf{x}, \boldsymbol{\theta})$ is given in (6). and

$$\sigma_0(\mathbf{x}_i, \boldsymbol{\theta}) = \left(1, \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \mathbf{x}_i} \right) \left(1, \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \mathbf{x}_i} \right)^T. \quad (10)$$

In the literature on optimal design, the matrices $M_{ML}(\xi, \boldsymbol{\theta})$ and $M_{LS}(\xi, \boldsymbol{\theta})$ are called *information matrices* of the design ξ [see Pukelsheim (2006)].

3 Optimal Designs

Following Kiefer (1974) we consider the construction of optimal approximate designs of the form (4) that are defined as probability measures with finite support on the design space $\mathcal{X}(\subset \mathbb{R}^q)$. This means that the support points \mathbf{x}_i , $i = 1, \dots, n$ define the distinct experimental points where observations are to be taken and the weights ω_i represent the relative proportion of observations to be taken at the corresponding support point. If an approximate design is given and r observations can be taken, a rounding procedure is applied to obtain integers r_i ($i = 1, \dots, n$) from the not necessarily integer valued quantities $\omega_i r$ ($i = 1, \dots, n$) [see Pukelsheim and Rieder (1992)]. In this case the relation (3) obviously holds and by the discussion in Section 2.1 and 2.2 the precision of the maximum likelihood and least squares estimators can be measured by the inverses of the information matrices $M_{ML}(\xi, \boldsymbol{\theta})$ and $M_{LS}(\xi, \boldsymbol{\theta})$, respectively. Therefore, optimal designs are constructed by maximizing a function, say Φ , of the information matrix with respect to the design ξ which is called *optimality criterion*. This function differs according to the combination of model parameters we are interested in estimating and there are numerous criteria which can be used to discriminate between competing designs [see, for example, Pukelsheim (2006) or Atkinson et al. (2007)]. In the following we concentrate on Kiefer's Φ_ℓ criteria, which are defined by

$$\Phi_\ell(M(\xi, \boldsymbol{\theta})) = \left(\text{tr} \left\{ \frac{1}{p+1} M^\ell(\xi, \boldsymbol{\theta}) \right\} \right)^{1/\ell}, \quad \ell \in [-\infty, 1) \quad (11)$$

and contain the famous D -, A - and E -criterion ($\ell = 0$, $\ell = -1$, $\ell = -\infty$) as special cases. A very useful tool for the characterisation of optimal designs and for checking the optimality of a candidate design is the general equivalence theorem [see, for example, Silvey (1980) or Atkinson et al. (2007)]. In the case of maximum likelihood estimation in the functional models under consideration, the general equivalence theorem for Φ_ℓ -optimality is presented below.

Theorem 1. *If $\ell \in (-\infty, 1)$, a design $\xi_{\boldsymbol{\theta}}^*$ is locally Φ_ℓ -optimal for maximum likelihood estimation in any functional model of the form (1) and (2) if and only if the inequality*

$$d_{ML}(\mathbf{x}, \xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) := \text{tr} \left\{ M_{ML}^{\ell-1}(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) \frac{\left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)}{\sigma_1(\mathbf{x}, \boldsymbol{\theta})} \right\} \leq \text{tr} \{ M_{ML}^\ell(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) \},$$

holds for all $\mathbf{x} \in \mathcal{X}$ with equality at the support points \mathbf{x}_i^ of $\xi_{\boldsymbol{\theta}}^*$. The matrix $M_{ML}(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta})$ and the function $\sigma_1(\mathbf{x}, \boldsymbol{\theta})$ are given in (5) and (6) respectively.*

On the other hand, when the vector of unknown parameters $\boldsymbol{\theta}$ is estimated via least squares the criterion function, $\Phi_\ell(M_{LS}(\xi, \boldsymbol{\theta}))$ where $M_{LS}(\xi, \boldsymbol{\theta})$ is given in (7), is not concave and therefore the general equivalence theorem does not hold. The following theorem provides a necessary (but not sufficient) condition for the Φ_ℓ -optimality of a candidate design, that is, a design that does not satisfy this condition cannot be Φ_ℓ -optimal for errors-in-variables models of the form (1) and (2).

Theorem 2. *If $\ell \in (-\infty, 1)$, any locally Φ_ℓ -optimal design ξ_θ^* on \mathcal{X} for least squares estimation in any functional model of the form (1) and (2) satisfies the inequality*

$$d_{LS}(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta}) := 2d_0(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta}) - \sigma_1(\mathbf{x}, \boldsymbol{\theta})d_1(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta}),$$

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

$$\begin{aligned} d_0(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta}) &= \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{D_0^{-1}(\xi_\theta^*, \boldsymbol{\theta}) M_{LS}^\ell(\xi_\theta^*, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right), \\ d_1(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta}) &= \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \frac{D_0^{-1}(\xi_\theta^*, \boldsymbol{\theta}) M_{LS}^{\ell+1}(\xi_\theta^*, \boldsymbol{\theta}) D_0^{-1}(\xi_\theta^*, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \end{aligned}$$

and $\sigma_0(\mathbf{x}, \boldsymbol{\theta})$, $\sigma_1(\mathbf{x}, \boldsymbol{\theta})$ are given in (10), (6), respectively. Furthermore, the maximum of $d_{LS}(\mathbf{x}, \xi_\theta^*, \boldsymbol{\theta})$ is achieved at the support points \mathbf{x}_i^* of ξ_θ^* .

Theorems 1 and 2 can also be extended to the case $\ell = -\infty$, but details are omitted for the sake of brevity. We refer to Pukelsheim (2006), Section 7.22, for the analogue of Theorem 1. The extension of Theorem 2 to the case $\ell = -\infty$ can be derived using the argument in Sections 11.10 - 11.13 of this reference. These results can be used to find the exact number or an upper bound for the number of support points of the locally Φ_ℓ -optimal design for models as in (1) and (2). This is illustrated in the following section for the D -optimality criterion and three specific functional errors-in-variables models of interest.

4 D -optimal designs for nonlinear error-in-variables models

In this section we focus on D -optimal designs which maximize the determinant of the information matrix $M(\xi, \boldsymbol{\theta})$ with respect to the design ξ [see, for example, Atkinson et al. (2007)]. This criterion corresponds to the case $\ell = 0$ of Kiefer's Φ_ℓ -criterion (11) and D -optimal designs minimise the volume of the confidence ellipsoid for the parameter

estimators. In this case the inequalities in Theorems 1 and 2 reduce to

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

$$d_{LS}(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) := \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T [2d_0(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) - \sigma_1(\mathbf{x}, \boldsymbol{\theta})d_1(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta})] \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \leq p + 1, \quad (12)$$

where

$$d_k(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) := \frac{D_k^{-1}(\xi_{\theta}^*, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})}, \quad k = 0, 1.$$

When a design has as many support points as model parameters it is said to be a saturated design. From Lemma 5.1.3 in Silvey (1980) it follows that when the parameters of any functional model as in (1) and (2) are estimated via maximum likelihood, the locally D -optimal saturated design is equally supported.

Lemma 1. *The locally D -optimal saturated design on \mathcal{X} for maximum likelihood estimation in any functional model of the form (1) and (2), puts equal weights at its support points.*

The Lemma below shows that the same result also holds for least squares estimation.

Lemma 2. *The locally D -optimal saturated design on \mathcal{X} for least squares estimation in any functional model of the form (1) and (2) puts equal weights at its support points.*

In what follows we consider three well-known and widely used nonlinear models and investigate the construction of locally D -optimal designs when the functional error structure described in (1) and (2) is assumed. Both maximum likelihood and least squares estimation are examined separately. Our approach is illustrated through an application to the case of one explanatory variable and we further assume for the sake of simplicity that the measurement errors of the dependent and independent variables are uncorrelated. Therefore, the variance-covariance matrix of the vector of measurement errors is now given by

$$\Sigma_{\eta\epsilon} = \begin{pmatrix} \sigma_{\eta}^2 & 0 \\ 0 & \sigma_{\epsilon}^2 \end{pmatrix}, \quad (13)$$

and the functions $\sigma_0(x, \boldsymbol{\theta})$ and $\sigma_1(x, \boldsymbol{\theta})$ in (10) and (6) become

$$\sigma_0(x, \boldsymbol{\theta}) = 1 + \left(\frac{\partial m(x, \boldsymbol{\theta})}{\partial x} \right)^2 \quad (14)$$

and

$$\sigma_1(x, \boldsymbol{\theta}) = \sigma_\eta^2 + \left(\frac{\partial m(x, \boldsymbol{\theta})}{\partial x} \right)^2 \sigma_\epsilon^2. \quad (15)$$

Namely, we consider the Michaelis-Menten, the EMAX and the exponential regression model, which are widely used in various areas of scientific research, such as pharmacology, medicine and biochemistry among others and in particular, they are used for modelling the dose-response relationship.

The Michaelis-Menten model is specified by

$$m_1(x, \boldsymbol{\theta}) = \frac{\theta_1 x}{(\theta_2 + x)}, \quad x \in \mathcal{X} = [0, x_u] \subset \mathbb{R}_0^+, \quad (16)$$

where the function m_1 relates the velocity of a reaction to the concentration x of a substrate, the parameter $\theta_1 \geq 0$ corresponds to the maximum reaction velocity and $\theta_2 \geq 0$ is the half-saturation constant, that is, the concentration-value x where the velocity is half-maximal. Optimal designs for this model in the case of no measurement error have been discussed by Duggleby (1979), Dunn (1988), Rasch (1990), Dette and Wong (1999) and Dette et al. (2010b), among others.

An increasing dose-response relationship with maximum effect achieved asymptotically at large dose levels is often modeled using the EMAX model given by

$$m_2(x, \boldsymbol{\theta}) = \theta_0 + \frac{\theta_1 x}{\theta_2 + x}, \quad x \in \mathcal{X} = [0, x_u] \subset \mathbb{R}_0^+, \quad (17)$$

[see, for example, Bretz et al. (2005)]. The explanatory variable x denotes the dose, $\theta_0 \geq 0$ usually corresponds to the placebo effect, that is, at dose $x = 0$, $\theta_1 \geq 0$ is the asymptotic maximum increase above the placebo and $\theta_2 \geq 0$ is the dose producing half of the asymptotic maximum effect. Optimal designs for the EMAX model with no measurement error have been determined by Dette et al. (2008) and Dette et al. (2010a), among others.

Finally, we consider the exponential regression model

$$m_3(x, \boldsymbol{\theta}) = \theta_0 e^{-\theta_1 x}, \quad x \in \mathcal{X} = [x_l, x_u] \subset \mathbb{R}, \quad (18)$$

which is the simplest and one of the most widely used models for survival studies such as clinical trials. As before x is the dose and the function m_3 represents the time until the occurrence of a particular event of interest, for example, the cure of patients. The parameter $\theta_0 \geq 0$ denotes the placebo effect and $\theta_1 (\neq 0)$ describes the rate of the explanatory effect. Because of its importance, numerous authors have worked on the construction

of optimal designs in the case of no measurement error [see, for example, Melas (1978), Mukhopadhyaya and Haines (1995), Dette and Neugebauer (1997), Han and Chaloner (2003) and Dette et al. (2006)].

4.1 Maximum likelihood estimation

When the model parameters are estimated via maximum likelihood the information matrix of a given design ξ is defined in (5). We also note that in this case classical optimal design theory results, such as the general equivalence in Theorem 1 and Lemma 1, are applicable. We begin with a discussion on locally D -optimal designs for the Michaelis-Menten model when the explanatory variable is subject to measurement error. The following result shows that the locally D -optimal design for the Michaelis-Menten model is uniquely determined and supported at exactly two points and it is therefore a saturated design.

Theorem 3. *The locally D -optimal design on $\mathcal{X} = [0, x_u]$ for maximum likelihood estimation in the Michaelis-Menten model (16) with measurement errors as in (1) and (2) is unique and puts equal masses at the points x_1^* and x_u , where x_1^* is the unique solution of the equation*

$$\frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} = 0, \quad (19)$$

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

We observe that the larger support point of the locally D -optimal design is always fixed at the upper end-point x_u of the design space regardless of the parameter vector θ . Therefore, using Theorem 3 the design problem reduces to an optimisation problem in one variable. Note that in the case of no error in the predictor, that is $\varrho_{\varepsilon\eta}^2 = 0$, the solution of equation (19) is given by $x_1^* = \frac{\theta_2 x_u}{2\theta_2 + x_u}$, which is the result obtained by Rasch (1990) for the D -optimal design in the Michaelis-Menten model with no measurement error. Moreover, if $\varrho_{\varepsilon\eta}^2 \rightarrow \infty$ we obtain $x_1^* \rightarrow \frac{x_u}{2}$, and an application of the implicit function theorem shows that x_1^* is a strictly increasing function of the ratio $\varrho_{\varepsilon\eta}^2 = \sigma_\varepsilon^2 / \sigma_\eta^2$. This means that the “non-trivial” support point of the D -optimal design for a model with no measurement error is always smaller than the corresponding point of the D -optimal design for the Michaelis-Menten model with measurement error.

Remark 1. A careful inspection of the proof of Theorem 3 shows that for any $\ell \in [-\infty, 1)$ the locally Φ_ℓ -optimal design for maximum likelihood estimation in the Michaelis-Menten model with measurement errors is always supported at the right boundary of the design space and a further point $x_1^* \in (0, x_u)$. Similar conclusions can be drawn for all other scenarios discussed in this section.

We now consider the D -optimal design problem for the EMAX model with measurement errors of the form (1) and (2). Note that for any design ξ , the information matrix for the EMAX model and therefore the resulting D -optimal design is independent of the parameter θ_0 and thus it can be treated as a nuisance parameter.

Theorem 4. *The locally D -optimal design on $\mathcal{X} = [0, x_u]$ for maximum likelihood estimation in the EMAX model (17) with measurement errors as in (1) and (2) is unique and puts equal masses at exactly three points $0, x_1^*, x_u$ where the support point x_1^* is the unique solution of the equation (19) in the interval $(0, x_u)$.*

The above theorem shows that the locally D -optimal design for the EMAX model is always supported at the two boundaries of the design space and its middle support point is equal to the smaller support point of the locally D -optimal design for the corresponding Michaelis-Menten model.

Finally we derive an analytical characterisation of the locally D -optimal design for the exponential regression model.

Theorem 5. *The locally D -optimal design on $\mathcal{X} = [x_l, x_u]$ for maximum likelihood estimation in the exponential regression model (18) with measurement errors as in (1) and (2) is unique and has exactly two support points.*

(a) If $\theta_1 > 0$, this design is

$$\xi_{\theta}^* = \begin{Bmatrix} x_l & x_1^* \\ 0.5 & 0.5 \end{Bmatrix},$$

where $x_1^* = \min\{x_u, \tilde{x}_1\}$ with \tilde{x}_1 being the unique solution of the equation

$$(\sigma_{\eta}^2 e^{2\theta_1 x_1} + \theta_0^2 \theta_1^2 \sigma_{\epsilon}^2) - (x_1 - x_l) \theta_1 \sigma_{\eta}^2 e^{2\theta_1 x_1} = 0,$$

in the interval (x_l, ∞) .

(b) If $\theta_1 < 0$, this design is

$$\xi_{\theta}^* = \begin{Bmatrix} x_0^* & x_u \\ 0.5 & 0.5 \end{Bmatrix},$$

where $x_0^* = \max\{x_l, \tilde{x}_0\}$ with \tilde{x}_0 being the unique solution of the equation

$$-(\sigma_{\eta}^2 e^{2\theta_1 x_0} + \theta_0^2 \theta_1^2 \sigma_{\epsilon}^2) - (x_u - x_0) \theta_1 \sigma_{\eta}^2 e^{2\theta_1 x_0} = 0,$$

in the interval $(-\infty, x_u)$.

Theorem 5 provides a complete classification of the locally D -optimal design depending on the sign of the parameter θ_1 , that is, whether the dose has an increasing or a decreasing

effect on the response. One of the support points of the optimal design is always fixed at one of the end-points of the design space thus simplifying the design problem. Moreover, in the “classical” case of no measurement error, that is $\sigma_\varepsilon^2 = 0$, the design reduces to the D -optimal design obtained by Han and Chaloner (2003), which puts equal masses at the two points x_l and $\min\{x_u, x_l + \frac{1}{\theta_1}\}$ (in the $\theta_1 > 0$ case) and $\max\{x_l, x_u + \frac{1}{\theta_1}\}$ and x_u (in the $\theta_1 < 0$ case). An application of the implicit function theorem shows that the non-trivial support point of the optimal design in the model with no measurement error is always smaller (larger) than the corresponding model with measurement error when $\theta_1 > 0$ ($\theta_1 < 0$).

4.2 Least squares estimation

As mentioned in Section 3, the general equivalence theorem cannot be used for the characterisation of optimal designs for least squares estimation because the mapping $\xi \rightarrow M_{LS}(\xi, \boldsymbol{\theta})$ is in general not concave under the measurement error structure we consider. However, Theorem 2 (in the case $\ell = 0$) gives a necessary condition for a design to be D -optimal for least squares estimation (see (12)). In what follows we use this necessary condition to obtain an upper bound for the number of support points of the locally D -optimal designs for each of the nonlinear models under consideration. For purposes of comparison with the optimal designs for maximum likelihood estimation, we then investigate the construction of locally D -optimal saturated designs which according to Lemma 2 assign equal weights to each of the support points.

We begin with an explicit result on the number of support points of the locally D -optimal design for the Michaelis-Menten model when its parameters are estimated via least squares.

Lemma 3. *Under least squares estimation, the locally D -optimal design on $\mathcal{X} = [0, x_u]$ for the Michaelis-Menten model (16) with measurement errors as in (1) and (2) has at most four support points.*

We note that as in Section 4.1, when the Michaelis-Menten model is used, none of the observations can be taken at point $x = 0$. An analytical characterisation of the two-point locally D -optimal design is given below.

Theorem 6. *The locally D -optimal saturated design on $\mathcal{X} = [0, x_u]$ for least squares estimation for the Michaelis-Menten model (16) with measurement errors as in (1) and*

(2) puts equal masses at the points x_1^* and x_u , where x_1^* is a solution of the equation

$$\frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\varepsilon\eta}^2} + \frac{2\theta_1^2 \theta_2^2}{(\theta_2 + x_1)[(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2]} = 0, \quad (20)$$

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

At this point a comparison of the locally D -optimal (saturated) designs in the Michaelis-Menten model with measurement errors for least squares and maximum likelihood estimation might be of interest. Note that the D -optimal designs differ only in the defining equations (19) and (20) of the interior support point x_1^* by the term

$$\frac{2\theta_1^2 \theta_2^2}{(\theta_2 + x_1)[(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2]} \leq \frac{2\theta_1^2}{\theta_2[\theta_2^2 + \theta_1^2]}. \quad (21)$$

Consequently, if the term on the right-hand side of (21) is small, the designs for maximum likelihood and least squares estimation do not differ very much (see Section 5 for some numerical examples), while the differences become larger if this expression is increasing.

Next we consider corresponding results for the EMAX model.

Lemma 4. *Under least squares estimation, the locally D -optimal design on $\mathcal{X} = [0, x_u]$ for the EMAX model model (17) with measurement errors as in (1) and (2) has at most five support points.*

Theorem 7. *The locally D -optimal saturated design on $\mathcal{X} = [0, x_u]$ for least squares estimation in the EMAX model (17) with measurement errors as in (1) and (2) puts equal masses at the points $0, x_1^*, x_u$, where x_1^* is a solution of the equation (20) in the interval $(0, x_u)$.*

We conclude again with a brief discussion on the locally D -optimal design for the exponential regression model. The following lemma provides an upper bound for the number of support points of the optimal design. Its proof is presented in detail in the online appendix and uses the properties of extended Chebyshev systems [see Karlin and Studden (1966)].

Lemma 5. *Under least squares estimation, the locally D -optimal design on $\mathcal{X} = [x_l, x_u]$ for the exponential regression model model (18) with measurement errors as in (1) and (2) has at most four support points.*

For the exponential regression model, an analytical characterisation of the locally D -optimal saturated design such as the one presented in Theorem 5 can not be produced for

the case of least squares estimation. The design is therefore constructed by maximising

$$\det M_{LS}(\xi, \boldsymbol{\theta}) = \frac{\theta_0^2(x_1 - x_0)^2}{4 [\sigma_\eta^2 e^{2\theta_1 x_0} + \theta_0^2 \theta_1^2 \sigma_\epsilon^2] [\sigma_\eta^2 e^{2\theta_1 x_1} + \theta_0^2 \theta_1^2 \sigma_\epsilon^2]} \frac{e^{-2\theta_1 x_0} e^{-2\theta_1 x_1}}{[1 + \theta_0^2 \theta_1^2 e^{-2\theta_1 x_0}] [1 + \theta_0^2 \theta_1^2 e^{-2\theta_1 x_1}]},$$

with respect to all two-point designs ξ equally supported at points x_0 and x_1 .

5 Data examples

To illustrate our results we use several examples from the literature for the choice of parameter values and design space. In particular, we use a clinical dose finding study considered in Dette et al. (2008), data from hormone receptor assays considered in Cressie and Keightley (1981) and a study of Mihara et al. (2000), who investigated the behaviour of CSD-plus pyrovate with L-cysteine sulfinat as substrate. We determine locally D -optimal designs for maximum likelihood and saturated locally D -optimal designs for least squares estimation in the Michaelis-Menten model with measurement error.

In the first example the response is an anxiety scale score used to determine the efficacy of an anti-anxiety drug whose dose corresponds to the explanatory variable. The dose-response relationship was modelled using a homoscedastic normal model although, as mentioned in Dette et al. (2008), prior to the start of the study, various other models were considered to be good alternatives. One of the models under consideration was the EMAX model assuming that the average placebo effect is zero or in other words the Michaelis-Menten model. By preliminary experiments prior information regarding the parameter values is available, that is $(\theta_1, \theta_2) = (7/15, 25)$, and the design space is given by $\mathcal{X} = [0, x_u] = [0, 150]$. Under maximum likelihood and least squares estimation we find (saturated) locally D -optimal designs for the Michaelis-Menten model under the functional error structure described in (1) and (2) which, according to Theorems 3 and 6, are equally supported with point $x = x_u$ always being one of their support points. The left part of Table 1 presents the smaller support points of the saturated locally D -optimal design for both maximum likelihood and least squares estimation for various values of the measurement errors ratio $\varrho_{\epsilon\eta}^2 = \sigma_\epsilon^2 : \sigma_\eta^2$.

It can be observed that in this first example the D -optimal designs for the two estimation methods are nearly identical. These small differences can be explained by the discussion presented in the paragraph following Theorem 6. For the situation under consideration the quantity on the right-hand side of inequality (21) is given by $2.78 \cdot 10^{-5}$ and there is practically no difference between equations (19) and (20). Consequently, whether the model parameters will be estimated through maximum likelihood or least squares does

Table 1: *Smaller support points x_{ML}^* and x_{LS}^* of the locally D -optimal (saturated) designs for maximum likelihood and least squares estimation in the Michaelis-Menten model. Left part: the clinical trial example considered in Dette et al. (2008); $\mathcal{X} = [0, 150]$. Middle part: the hormone receptor assay data considered in Cressie and Keightley (1981); $\mathcal{X} = [0, 2000]$. Right part: the example considered in Mihara et al. (2000); $\mathcal{X} = [0, 80]$.*

| | $\theta_1 = 7/15, \theta_2 = 25$ | | $\theta_1 = 0.69, \theta_2 = 0.6$ | | $\theta_1 = 16, \theta_2 = 3.5$ | |
|-------------------------------|----------------------------------|------------|-----------------------------------|------------|---------------------------------|------------|
| $\varrho_{\varepsilon\eta}^2$ | x_{ML}^* | x_{LS}^* | x_{ML}^* | x_{LS}^* | x_{ML}^* | x_{LS}^* |
| 4:1 | 18.754 | 18.755 | 194.79 | 195.66 | 8.499 | 9.468 |
| 2:1 | 18.751 | 18.753 | 193.06 | 193.95 | 7.145 | 8.390 |
| 1:1 | 18.751 | 18.751 | 192.18 | 193.07 | 6.039 | 7.572 |
| 1:2 | 18.750 | 18.751 | 191.74 | 192.63 | 5.155 | 6.982 |
| 1:4 | 18.750 | 18.751 | 191.51 | 192.41 | 4.479 | 6.586 |

not affect the optimal choice of the D -optimal design. We finally note that the locally D -optimal design for the Michaelis-Menten model with no measurement errors puts equal masses at the points 18.75 and 150. Thus in this case there is no substantial difference between the D -optimal designs for models with and without measurement errors.

Our second example discusses the application of the Michaelis-Menten model to the hormone receptor assays discussed in Cressie and Keightley (1981). These authors provided several estimation methods for the parameters θ_1, θ_2 in the Michaelis-Menten model. For the data set considered in Table 1 of this reference the design space is given by $\mathcal{X} = [0, 2000]$ and the parameter estimates are $\theta_1 = 43.95$ and $\theta_2 = 236.53$ [see Dette and Biedermann (2003), Section 4]. The locally D -optimal designs (more precisely, the corresponding interior support point) are presented in the middle part of Table 1, whereas the locally D -optimal design for the corresponding model with no measurement error puts equal masses at the points 191.23 and 2000. Here we observe more substantial differences between the D -optimal design for models with and without measurement errors, in particular, if the noise ratio $\varrho_{\varepsilon\eta}^2$ is large.

Our final example considers the investigation of the behaviour of CSD-plus pyrovate with L-cysteine sulfinat as substrate [see Mihara et al. (2000) for details]. In this case the parameter estimates (by maximum likelihood) in the Michaelis-Menten model are given by $\theta_1 = 1, \theta_2 = 3.5$, the design space is $\mathcal{X} = [0, 80]$ and the locally D -optimal design for the model with no measurement errors puts equal masses at the points 3.218 and 80. The D -optimal design corresponding to maximum likelihood and least squares estimation in the Michaelis-Menten model with measurement errors can be found in the right part of Table 1. We observe that all designs differ substantially and therefore, in this example it

is of particular importance to take the measurement error into account for the planning of the experiment.

In all examples under consideration we observe that for fixed ratio-value $\sigma_\epsilon^2 : \sigma_\eta^2$ the smaller support point of the optimal design for least squares estimation is always greater than the corresponding x_{ML}^* of the measurement errors ratio-values. Furthermore, in both the cases of maximum likelihood and least squares estimation the value of the smaller support point increases as the covariate measurement error σ_ϵ^2 increases. On the other hand, an increase in the value of σ_η^2 has a decreasing effect on both x_{ML}^* and x_{LS}^* .

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG). The research of H. Dette reported in this publication was also partially supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix

Proof of the results in Section 3

Proof of Theorem 2: Let ξ_θ^* be a locally D -optimal design for least squares estimation of the parameters involved in a functional model as in (1) and (2). For any other design ξ and $a \in [0, 1]$ also let $\xi_a = (1 - a)\xi_\theta^* + a\xi$ and

$$\pi(\xi) = \log \Phi_\ell(M_{LS}(\xi, \boldsymbol{\theta})) = \frac{1}{\ell} [\log(p + 1) + \log \text{tr}\{M_{LS}^\ell(\xi, \boldsymbol{\theta})\}].$$

The Frechet derivative of $\pi(\xi)$ at ξ_{θ}^* in the direction of $\xi - \xi_{\theta}^*$ is given by

$$\begin{aligned}
\left. \frac{d}{da} \pi(\xi_a) \right|_{a=0} &= \left. \frac{1}{\ell \operatorname{tr}\{M_{LS}^{\ell}(\xi_a, \boldsymbol{\theta})\}} \right|_{a=0} \left. \frac{d}{da} \operatorname{tr}\{M_{LS}^{\ell}(\xi_a, \boldsymbol{\theta})\} \right|_{a=0} \\
&= \frac{1}{\operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\}} \operatorname{tr} \left\{ M_{LS}^{\ell-1}(\xi_{\theta}^*, \boldsymbol{\theta}) \left. \frac{d}{da} D_0(\xi_a, \boldsymbol{\theta}) D_1^{-1}(\xi_a, \boldsymbol{\theta}) D_0(\xi_a, \boldsymbol{\theta}) \right|_{a=0} \right\} \\
&= \frac{1}{\operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\}} [2 \operatorname{tr}\{D_1^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta})\} - \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\} \\
&\quad - \operatorname{tr}\{D_1^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\xi_{\theta}^*, \boldsymbol{\theta}) D_1^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_1(\xi, \boldsymbol{\theta})\}] \\
&= \frac{1}{\operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\}} [2 \operatorname{tr}\{D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta})\} - \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\} \\
&\quad - \operatorname{tr}\{D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell+1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_1(\xi, \boldsymbol{\theta})\}].
\end{aligned}$$

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

$$\begin{aligned}
\operatorname{tr}\{D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0(\delta_{\mathbf{x}}, \boldsymbol{\theta})\} &= \operatorname{tr} \left\{ D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta}) \frac{\left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \right\} \\
&= \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \frac{D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \\
&= d_0(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}),
\end{aligned}$$

and similarly

$$\operatorname{tr}\{D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) M_{LS}^{\ell+1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) D_1(\xi, \boldsymbol{\theta})\} = \sigma_1(\mathbf{x}, \boldsymbol{\theta}) d_1(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}).$$

Since ξ_{θ}^* is D -optimal $\left. \frac{d}{da} \pi(\xi_a) \right|_{a=0}$ is non-positive for all designs ξ , the inequality

$$d_{LS}(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) \leq \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\},$$

for all $\mathbf{x} \in \mathcal{X}$ follows.

For the proof of the equality part of Theorem 2 let us assume that

$$\max_{\mathbf{x} \in \mathcal{X}} d_{LS}(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) < \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\}.$$

Then

$$\int_{\mathcal{X}} d_{LS}(\mathbf{x}, \xi_{\theta}^*, \boldsymbol{\theta}) d\xi_{\theta}^*(\mathbf{x}) < \int_{\mathcal{X}} \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\} d\xi_{\theta}^*(\mathbf{x}) = \operatorname{tr}\{M_{LS}^{\ell}(\xi_{\theta}^*, \boldsymbol{\theta})\}.$$

Now for any design ξ

$$\begin{aligned}
\int_{\mathcal{X}} d_{LS}(\mathbf{x}, \xi, \boldsymbol{\theta}) d\xi(\mathbf{x}) &= 2tr \left\{ D_0^{-1}(\xi, \boldsymbol{\theta}) M_{LS}^\ell(\xi, \boldsymbol{\theta}) \int_{\mathcal{X}} \frac{\left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} d\xi(\mathbf{x}) \right\} \\
&\quad - tr \left\{ D_0^{-1}(\xi, \boldsymbol{\theta}) M_{LS}^{\ell+1}(\xi, \boldsymbol{\theta}) D_0^{-1}(\xi, \boldsymbol{\theta}) \int_{\mathcal{X}} \sigma_1(\mathbf{x}, \boldsymbol{\theta}) \frac{\left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)}{\sigma_0(\mathbf{x}, \boldsymbol{\theta})} d\xi(\mathbf{x}) \right\} \\
&= 2tr \left\{ D_0^{-1}(\xi, \boldsymbol{\theta}) M_{LS}^\ell(\xi, \boldsymbol{\theta}) D_0(\xi, \boldsymbol{\theta}) \right\} \\
&\quad - tr \left\{ D_0^{-1}(\xi, \boldsymbol{\theta}) M_{LS}^{\ell+1}(\xi, \boldsymbol{\theta}) D_0^{-1}(\xi, \boldsymbol{\theta}) D_1(\xi, \boldsymbol{\theta}) \right\} \\
&= tr \{ M_{LS}^\ell(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) \},
\end{aligned}$$

which contradicts our initial assumption. Hence $\max_{\mathbf{x} \in \mathcal{X}} d_{LS}(\mathbf{x}, \xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) = tr \{ M_{LS}^\ell(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) \}$ and this is attained at each support point \mathbf{x}_i^* of the locally D -optimal design $\xi_{\boldsymbol{\theta}}^*$.

Proof of the results in Section 4

Proof of Lemma 2: Let

$$\xi = \begin{Bmatrix} \mathbf{x}_0 & \dots & \mathbf{x}_p \\ \omega_0 & \dots & \omega_p \end{Bmatrix},$$

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 determinant of the information matrix $M_{LS}(\xi, \boldsymbol{\theta})$ given in (7) becomes

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

Maximising the above expression 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.

Proof of the results in Section 4.1

Here we only present the proof of Theorem 3. Similar arguments are used for the proofs of Theorems 4 and 5 which can be found in the online appendix.

Proof of Theorem 3: Let $\xi_{\boldsymbol{\theta}}^*$ be a locally D -optimal design for maximum likelihood

estimation in the Michaelis-Menten model given in (16) subject to measurement errors as in (1) and (2). Also let

$$M_{ML}^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) = \begin{pmatrix} m_1 & m_2 \\ m_2 & m_3 \end{pmatrix},$$

where $m_1, m_2, m_3 \in \mathbb{R}$. From Theorem 1 for $\ell = 0$, we obtain that a locally D -optimal design must satisfy the inequality

$$P_1(x) := m_1 x^2 (\theta_2 + x)^2 - 2m_2 \theta_1 x^2 (\theta_2 + x) + m_3 \theta_1^2 x^2 - 2 [\sigma_\eta^2 (\theta_2 + x)^4 + (\theta_1 \theta_2)^2 \sigma_\epsilon^2] \leq 0,$$

for all $x \in \mathcal{X} = [0, x_u]$ with equality at the support points of ξ_{θ}^* . Here $P_1(x)$ is a polynomial of degree 4 and therefore, it can have at most 4 roots and at most 3 turning points.

Assuming that $P_1(x)$ has exactly 4 roots, that is, the locally D -optimal design has exactly 4 support points, we have that at least the two middle roots must be turning points since $P_1(x) \leq 0, \forall x \in \mathcal{X}$. Hence $P_1(x)$ has a total of at least 5 turning points which contradicts the fact that the degree of $P_1(x)$ is 4. Now let $P_1(x)$ have 3 roots. Since $P_1(x) \leq 0, \forall x \in \mathcal{X}$, the middle root must be a turning point and the end-points 0 and x_u of the design space \mathcal{X} are also roots. However, $P_1(0) = -2 < 0$ and thus $x = 0$ cannot be a support point which contradicts the assumption of a three-point design. We therefore conclude that the locally D -optimal design for the Michaelis-Menten model is supported at exactly two-points but no observations can be taken at point $x = 0$.

From Lemma 1 it follows that a locally D -optimal design for the Michaelis-Menten model is equally supported at its two support points. It therefore remains to evaluate these points by maximising the determinant of the information matrix with respect to all designs with two support points and equal weights. Let

$$\xi_{\theta} = \left\{ \begin{array}{cc} x_1 & x_2 \\ 0.5 & 0.5 \end{array} \right\},$$

be an equally weighted two-point design. Using (5), the determinant of the information matrix of the design ξ_{θ} is given by

$$\det \{M_{ML}(\xi_{\theta}, \boldsymbol{\theta})\} = \frac{\theta_1^2 x_1^2 x_2^2 (x_2 - x_1)^2}{4 [\sigma_\eta^2 (\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_\epsilon^2] [\sigma_\eta^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_\epsilon^2]}.$$

For fixed x_1 , maximising $\det \{M_{ML}(\xi_{\theta}, \boldsymbol{\theta})\}$ is equivalent to maximising

$$2 \log x_2 + 2 \log (x_2 - x_1) - \log (\sigma_\eta^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_\epsilon^2).$$

Taking the derivative of the above expression with respect to x_2 we have that for all $x_2 \in (x_l, x_u]$

$$\begin{aligned} & \frac{2}{x_2} + \frac{2}{x_2 - x_1} - \frac{4\sigma_\eta^2(\theta_2 + x_2)^3}{\sigma_\eta^2(\theta_2 + x_2)^4 + \theta_1^2\theta_2^2\sigma_\epsilon^2} \\ &= 2 \frac{(2x_2 - x_1)\theta_1^2\theta_2^2\sigma_\epsilon^2 + \sigma_\eta^2(\theta_2 + x_2)^3 [(2x_2 - x_1)\theta_2 + x_1x_2]}{x_2(x_2 - x_1) [\sigma_\eta^2(\theta_2 + x_2)^4 + \theta_1^2\theta_2^2\sigma_\epsilon^2]} > 0. \end{aligned}$$

Hence, regardless of the value of x_1 , the determinant is increasing with x_2 and therefore maximised for $x_2^* = x_u$. It remains to determine the smaller support point x_1^* of the locally D -optimal design. This is found by solving

$$\frac{\partial \log \det \{M_{ML}(\xi_{\boldsymbol{\theta}}, \boldsymbol{\theta})\}}{\partial x_1} \Big|_{x_2=x_u} = 0 \iff \frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2\sigma_\eta^2(\theta_2 + x_1)^3}{\sigma_\eta^2(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2\sigma_\epsilon^2} = 0,$$

for $x_1 \in (0, x_u)$. Solving the above equation is equivalent to solving

$$f(x_1) := (x_u - 2x_1) [\sigma_\eta^2(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2\sigma_\epsilon^2] - 2\sigma_\eta^2(\theta_2 + x_1)^3 x_1(x_u - x_1) = 0,$$

for $x_1 \in (0, x_u)$. Now $f(0) > 0$, $f(x_u) < 0$ and it is easy to check that the function f has exactly one turning point in $(0, x_u)$ which is a minimum. Hence $f(x_1) = 0$ has a unique solution in the interval $x_1 \in (0, x_u)$, which completes the proof.

Proof of the results in Section 4.2

As in the previous section, for the sake of brevity only Lemma 3 and Theorem 6 are proven here since the proofs of the remaining results of Section 4.2 use similar arguments. These can be found in the online appendix.

Proof of Lemma 3: Let $\xi_{\boldsymbol{\theta}}^*$ be a locally D -optimal design for least squares estimation in the Michaelis-Menten model given in (16) subject to measurement errors as in (1) and (2). Also let

$$D_0^{-1}(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) = \begin{pmatrix} d_1 & d_2 \\ d_2 & d_3 \end{pmatrix} \quad \text{and} \quad D_1^{-1}(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) = \begin{pmatrix} d_4 & d_5 \\ d_6 & d_6 \end{pmatrix},$$

where $d_1, \dots, d_6 \in \mathbb{R}$. For the Michaelis-Menten model, the necessary condition for D -

optimality given in Theorem 2 for $\ell = 0$ becomes

$$\begin{aligned} & \left(\frac{x}{(\theta_2 + x)}, \frac{-\theta_1 x}{(\theta_2 + x)^2} \right) \left\{ 2D_0^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) - \left[\sigma_{\eta}^2 + \frac{\theta_1^2 \theta_2^2}{(\theta_2 + x)^4 \sigma_{\epsilon}^2} \right] D_1^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) \right\} \left(\frac{x}{(\theta_2 + x)}, \frac{-\theta_1 x}{(\theta_2 + x)^2} \right)^T \\ & \leq 2 \left[1 + \frac{\theta_1^2 \theta_2^2}{(\theta_2 + x)^4} \right], \quad \forall x \in \mathcal{X} = [0, x_u], \end{aligned}$$

with equality at the support points of ξ_{θ}^* . Multiplying through the above expression with $(\theta_2 + x)^8$ we get $P_4(x) \leq 0$ where $P_4(x)$ is a polynomial of degree 8. Therefore, $P_4(x)$ has at most 8 roots and at most 7 turning points.

Assuming that $P_4(x)$ has 8, 7, or 6 roots leads to a contradiction as these correspond to a total number of at least 13, 11 and 9 turning points respectively. Furthermore, if $P_4(x) = 0$ at exactly 5 points, then the 3 middle roots must be turning points and also the end-points of the design space are roots of the polynomial. However, $x = 0$ cannot be a root ($P_4(0) < 0$) and therefore the assumption of 5 roots is also rejected. Hence the locally D -optimal design for the Michaelis-Menten model has at most 4 support points but no observations can be taken at point $x = 0$.

Proof of Theorem 6: From Lemma 2 we know that a locally D -optimal saturated design for the Michaelis-Menten model puts equal weights at its support points. Let

$$\xi_{\theta} = \begin{Bmatrix} x_1 & x_2 \\ 0.5 & 0.5 \end{Bmatrix},$$

be a two-point equally weighted design. The determinant of the information matrix $M_{LS}(\xi_{\theta}, \boldsymbol{\theta})$ given in (7) becomes

$$\det \{M_{LS}(\xi_{\theta}, \boldsymbol{\theta})\} = \frac{\theta_1^2 x_1^2 x_2^2 (x_2 - x_1)^2}{4 [\sigma_{\eta}^2 (\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2] [\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2]} \frac{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4}{[(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2] [(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2]}.$$

For fixed x_1 , maximising the above expression is equivalent to maximising

$$2 \log x_2 + 2 \log(x_2 - x_1) - \log(\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2) + 4 \log(\theta_2 + x_2) - \log((\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2).$$

Differentiating this with respect to x_2 gives

$$\frac{2}{x_2} + \frac{2}{x_2 - x_1} - \frac{4\sigma_{\eta}^2 (\theta_2 + x_2)^3}{\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2} + \frac{4\theta_1^2 \theta_2^2}{(\theta_2 + x_2) [(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2]},$$

which is positive for all $x_2 \in (0, x_u]$ following the proof of Theorem 3. Hence for fixed x_1 , the determinant of the information matrix is increasing with x_2 and therefore maximised for $x_2^* = x_u$. The smaller support point x_1^* is found by solving

$$\left. \frac{\partial \log \det \{M_{LS}(\xi_{\theta}, \theta)\}}{\partial x_1} \right|_{x_2=x_u} = 0, \quad x_1 \in (0, x_u),$$

which is equivalent to solving

$$\frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2\sigma_{\eta}^2(\theta_2 + x_1)^3}{\sigma_{\eta}^2(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2\sigma_{\epsilon}^2} + \frac{2\theta_1^2\theta_2^2}{(\theta_2 + x_1)[(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2]} = 0,$$

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

References

- Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum Experimental Designs, with SAS (Oxford Statistical Science Series)*. Oxford University Press, USA, 2nd edition.
- Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press.
- Box, G. E. (1963). The effects of errors in the factor levels and experimental design. *Technometrics*, 5(2):247–262.
- Bretz, F., Pinheiro, J. C., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61(3):738–748.
- Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis.
- Carroll, R., Ruppert, D., and Stefanski, L. (1995). *Measurement Error in Nonlinear Models*. Taylor & Francis.
- Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.
- Cressie, N. A. C. and Keightley, D. D. (1981). Analysing data from hormone-receptor assays. *Biometrics*, 37:235–249.
- Detle, H. (1997). Designing experiments with respect to ‘standardized’ optimality criteria. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(1):97–110.
- Detle, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, 98(463):679–686.
- Detle, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.

- Dette, H., Kiss, C., Bevanda, M., and Bretz, F. (2010a). Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, 97(2):513–518.
- Dette, H., Kiss, C., and Wong, W. K. (2010b). A web-based tool for finding optimal designs for the Michaelis-Menten model and an overview. *Statistics in Biopharmaceutical Research*, 2(3):383–393.
- Dette, H., Martinez Lopez, I., Ortiz Rodriguez, I., and Pepelyshev, A. (2006). Maximin efficient design of experiment for exponential regression models. *Journal of Statistical Planning and Inference*, 136:4397–4418.
- Dette, H. and Neugebauer, H. M. (1997). Bayesian D -optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, 60:331–349.
- Dette, H. and Wong, W. K. (1999). E -optimal designs for the Michaelis-Menten model. *Statistics & Probability Letters*, 44:405–408.
- Donev, A. N. (2004). Design of experiments in the presence of errors in factor levels. *Journal of Statistical Planning and Inference*, 126(2):569 – 585.
- Dovi, V. G., Reverberi, A. P., and Maga, L. (1993). Optimal design of sequential experiments for error-in-variables models. *Computers & Chemical Engineering*, 17(1):111 – 115.
- Draper, N. R. and Beggs, W. J. (1971). Errors in the factor levels and experimental design. *The Annals of Mathematical Statistics*, 42(1):46–58.
- Duggleby, R. G. (1979). Experimental designs for estimating the kinetic parameters for enzyme-catalysed reactions. *Journal of Theoretical Biology*, 81:671–684.
- Dunn, G. (1988). Optimal designs for drug, neurotransmitter and hormone receptor assays. *Statistics in Medicine*, 7:805–815.
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- Han, C. and Chaloner, K. (2003). D -and c -optimal designs for exponential regression models used in pharmacokinetics and viral dynamics. *Journal of Statistical Planning and Inference*, 115:585–601.
- Karlin, S. and Studden, W. J. (1966). *Tchebysheff Systems: With Application in Analysis and Statistics*. Wiley, New York.
- Keeler, S. and Reilly, P. (1992). The design of experiments when there are errors in all the variables. *Canadian Journal of Chemical Engineering*, 70:774–778.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 2(5):849–879.
- Melas, V. B. (1978). Optimal designs for exponential regression. *Mathematische Operationsforschung und Statistik*, 9(1):45–59.

- Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N. (2000). 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.
- Mukhopadhyaya, S. and Haines, L. M. (1995). Bayesian D -optimal designs for the exponential growth model. *Journal of Statistical Planning and Inference*, 44(3):385–397.
- Pronzato, L. (2002). Information matrices with random regressors. application to experimental design. *Journal of Statistical Planning and Inference*, 108(1-2):189–200.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, 79:763–770.
- Rasch, D. (1990). Optimum experimental design in nonlinear regression. *Comm. Statist. Theory Methods*, 19(12):4789–4806.
- Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.
- Tang, P. K. and Bacon-Shone, J. (1992). Bayesian optimal designs for probit regression with errors-in-variables. Technical Report 31, University of Hong Kong. Dept. of Statistics.
- Vuchkov, I. N. and Boyadjieva, L. N. (1983). The robustness of experimental designs against errors in the factor levels. *Journal of Statistical Computation and Simulation*, 17(1):31–41.

Online Appendix: further technical details:

Proof of Theorem 4: Let ξ_{θ}^* be a locally D -optimal design for maximum likelihood estimation in the EMAX model given in (17) subject to measurement errors as in (1) and (2). Also let

$$M_{ML}^{-1}(\xi_{\theta}^*, \theta) = \begin{pmatrix} m_1 & m_2 & m_3 \\ m_2 & m_4 & m_5 \\ m_3 & m_5 & m_6 \end{pmatrix},$$

where $m_1, \dots, m_6 \in \mathbb{R}$. Using Theorem 1 in Section 3 for $\ell = 0$, a locally D -optimal design must satisfy the inequality

$$P_2(x) := m_1(\theta_2 + x)^4 + 2m_2x(\theta_2 + x)^3 - 2m_3\theta_1x(\theta_2 + x)^2 + m_4x^2(\theta_2 + x)^2 - 2m_5\theta_1x^2(\theta_2 + x) + m_6\theta_1^2x^2 - 3[\sigma_{\eta}^2(\theta_2 + x)^4 - 3\theta_1^2\theta_2^2\sigma_{\epsilon}^2] \leq 0,$$

for all $x \in \mathcal{X} = [0, x_u]$ with equality at the support points of ξ_{θ}^* . Similar arguments as given in the proof of Theorem 3 show that the locally D -optimal design must be supported at exactly three points including the boundary points 0 and x_u of the design space. Using Lemma 1 the locally D -optimal design for the EMAX model is found by maximising the determinant of the information matrix with respect to all three-point designs of the form

$$\xi_{\theta} = \begin{Bmatrix} 0 & x_1 & x_u \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}.$$

Using (5), the determinant of the information matrix of any design ξ_{θ} is given by

$$\det \{M_{ML}(\xi_{\theta}, \theta)\} = \frac{\theta_1^2\theta_2^4x_1^2x_u^2(x_1 - x_u)^2}{27[\sigma_{\eta}^2\theta_2^4 + \theta_1^2\theta_2^2\sigma_{\epsilon}^2][\sigma_{\eta}^2(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2\sigma_{\epsilon}^2][\sigma_{\eta}^2(\theta_2 + x_u)^4 + \theta_1^2\theta_2^2\sigma_{\epsilon}^2]}.$$

Taking the derivative of the logarithm of the above expression with respect to x_1 and equating it to zero we get

$$\frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2\sigma_{\eta}^2(\theta_2 + x_1)^3}{\sigma_{\eta}^2(\theta_2 + x_1)^4 + \theta_1^2\theta_2^2\sigma_{\epsilon}^2} = 0,$$

for $x_1 \in (0, x_u)$ which has a unique solution (see proof of Theorem 3).

Proof of Theorem 5: Here we only present the proof of part (a). Part (b) can be proven following along the same lines using symmetry arguments and is therefore omitted. Let ξ_{θ}^* be a locally D -optimal design for maximum likelihood estimation in the

exponential regression model given in (18) subject to measurement errors as in (1) and (2). Also denote by

$$M_{ML}^{-1}(\xi_{\theta}^*, \boldsymbol{\theta}) = \begin{pmatrix} m_1 & m_2 \\ m_2 & m_3 \end{pmatrix},$$

its corresponding information matrix, where $m_1, m_2, m_3 \in \mathbb{R}$. It follows from the general equivalence theorem given in Theorem 1 that a locally D -optimal design must satisfy the inequality

$$P_3(x) := m_1 - 2m_2\theta_0x + m_3\theta_0^2x^2 \leq 2(\sigma_\eta^2e^{2\theta_1x} + \theta_0^2\theta_1^2\sigma_\epsilon^2) := g_1(x),$$

for all $x \in \mathcal{X} = [x_l, x_u]$ with equality at its support points.

Equation $P_3(x) = g_1(x)$ has at most three solutions as these are the intersection points of an increasing function ($g_1(x)$) and a polynomial of degree 2 ($P_3(x)$). Suppose that this equation has exactly 3 solutions, that is, the locally D -optimal design has three support points $x_l \leq x_1 < x_2 < x_3 \leq x_u$. Then $P_3(x_k) = g_1(x_k)$, $k = 1, 2, 3$. By Cauchy's mean value theorem there exist points \tilde{x}_k , $k = 1, 2$ such that

$$x_1 < \tilde{x}_1 < x_2 < \tilde{x}_2 < x_3 \quad \text{and} \quad P_3'(\tilde{x}_k) = g_1'(\tilde{x}_k), k = 1, 2.$$

Since $P_3(x) \leq g_1(x)$ on \mathcal{X} we also have that $P_3'(x_2) = g_1'(x_2)$. By applying the mean value theorem again, there exist \hat{x}_k , $k = 1, 2$ such that

$$\tilde{x}_1 < \hat{x}_1 < x_2 < \hat{x}_2 < \tilde{x}_2 \quad \text{and} \quad P_3''(\hat{x}_k) = g_1''(\hat{x}_k), k = 1, 2.$$

Now $P_3''(x)$ is a constant and $g_1''(x) = 8\sigma_\eta^2\theta_1^2e^{2\theta_1x}$ is an injective function. Therefore, $P_3''(x) = g_1''(x)$ has at most one solution on \mathcal{X} which contradicts the assumption of three support points. Hence a locally D -optimal design for the exponential regression model has exactly two support points and following Lemma 1, these must be equally weighted.

For the exponential regression model with measurement error, the determinant of the information matrix of a design ξ_{θ} equally supported at points x_0 and x_1 becomes

$$\det \{M_{ML}(\xi_{\theta}, \boldsymbol{\theta})\} = \frac{\theta_0^2(x_1 - x_0)^2}{4[\sigma_\eta^2e^{2\theta_1x_0} + \theta_0^2\theta_1^2\sigma_\epsilon^2][\sigma_\eta^2e^{2\theta_1x_1} + \theta_0^2\theta_1^2\sigma_\epsilon^2]}.$$

By assumption $\theta_1 > 0$, it is easy to see that for fixed x_1 , $\det \{M_{ML}(\xi_{\theta}, \boldsymbol{\theta})\}$ is decreasing with x_0 and therefore, it is maximised at $x_0^* = x_l$. The larger support point x_1^* of the

locally D -optimal design is then found by solving the equation

$$\left. \frac{\partial \det \{M_{ML}(\xi_{\theta}, \theta)\}}{\partial x_1} \right|_{x_0=x_l} = 0,$$

for $x_1 \in (x_l, x_u]$. The equation above can be rewritten as

$$\sigma_{\eta}^2 + \theta_0^2 \theta_1^2 \sigma_{\epsilon}^2 e^{-2\theta_1 x_1} = \theta_1 \sigma_{\eta}^2 (x_1 - x_l),$$

and it has a unique solution in (x_l, ∞) since $\theta_1 \sigma_{\eta}^2 (x_1 - x_l)$ is strictly increasing for $x_1 \in (x_l, \infty)$, whereas the L.H.S. function is strictly decreasing with x_1 . If the solution is attained outside the design space we then take $x_1^* = x_u$.

Proof of Lemma 4: Let ξ_{θ}^* be a locally D -optimal design for least squares estimation in the EMAX model given in (17) subject to measurement errors as in (1) and (2). Under this model, the necessary condition for D -optimality given in Theorem 2 ($\ell = 0$) becomes

$$\begin{aligned} & \left(1, \frac{x}{(\theta_2 + x)}, \frac{-\theta_1 x}{(\theta_2 + x)^2} \right) \left\{ 2D_0^{-1}(\xi_{\theta}^*, \theta) - \left[\sigma_{\eta}^2 + \frac{\theta_1^2 \theta_2^2}{(\theta_2 + x)^4 \sigma_{\epsilon}^2} \right] D_1^{-1}(\xi_{\theta}^*, \theta) \right\} \left(\frac{x}{(\theta_2 + x)}, \frac{1, -\theta_1 x}{(\theta_2 + x)^2} \right)^T \\ & \leq 3 \left[1 + \frac{\theta_1^2 \theta_2^2}{(\theta_2 + x)^4} \right], \quad \forall x \in \mathcal{X} = [0, x_u], \end{aligned}$$

with equality at the support points of ξ_{θ}^* . Since both $D_0^{-1}(\xi_{\theta}^*, \theta)$ and $D_1^{-1}(\xi_{\theta}^*, \theta)$ are real-valued 3×3 matrices, multiplying through the above expression with $(\theta_2 + x)^8$ gives $P_5(x) \leq 0$, where $P_5(x)$ is a polynomial of degree 8.

Using the same arguments as in the proof of Lemma 3, the assumptions of 8,7 and 6 roots are rejected. Moreover, $P_5(0) < 0$ and so the 5-roots assumption is valid. Therefore, the locally D -optimal design for the EMAX model has at most 5 support points.

Proof of Theorem 7: Let

$$\xi_{\theta} = \left\{ \begin{array}{ccc} x_0 & x_1 & x_2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\},$$

be a three-point equally weighted design. Using (7), the determinant of the information

matrix of this design is

$$\det \{M_{LS}(\xi_{\theta}, \boldsymbol{\theta})\} = \frac{\theta_1^2 \theta_2^4 (x_2 - x_1)^2 (x_2 - x_0)^2 (x_1 - x_0)^2}{27 [\sigma_{\eta}^2 (\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2] [\sigma_{\eta}^2 (\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2] [\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2]} \frac{(\theta_2 + x_0)^4 (\theta_2 + x_1)^4 (\theta_2 + x_2)^4}{[(\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2] [(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2] [(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2]}.$$

For fixed x_0 and x_1 , maximising the above expression is equivalent to maximising

$$2 \log(x_2 - x_1) + 2 \log(x_2 - x_0) - \log(\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2) + 4 \log(\theta_2 + x_2) - \log((\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2),$$

for $x_2 \in (x_1, x_u]$. Differentiating this with respect to x_2 gives

$$\frac{2}{x_2 - x_1} + \frac{2}{x_2 - x_0} - \frac{4\sigma_{\eta}^2 (\theta_2 + x_2)^3}{\sigma_{\eta}^2 (\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2} + \frac{4\theta_1^2 \theta_2^2}{(\theta_2 + x_2) [(\theta_2 + x_2)^4 + \theta_1^2 \theta_2^2]}.$$

It is easy to show that the expression above is positive for all $0 \leq x_0 < x_1 < x_2 \leq x_l$ and therefore increasing with x_2 . Hence $\det \{M_{LS}(\xi_{\theta}, \boldsymbol{\theta})\}$ is maximised at $x_2^* = x_u$.

Similarly for fixed x_1 and x_2 , the smaller support point of the locally D -optimal saturated design is found by maximising

$$2 \log(x_2 - x_0) + 2 \log(x_1 - x_0) - \log(\sigma_{\eta}^2 (\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2) + 4 \log(\theta_2 + x_0) - \log((\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2),$$

for $x_0 \in [0, x_1)$. It can be easily shown that the derivative of the above expression given by

$$-\frac{2}{x_2 - x_0} - \frac{2}{x_1 - x_0} + \frac{4}{\theta_2 + x_0} - \frac{4\sigma_{\eta}^2 (\theta_2 + x_0)^3}{\sigma_{\eta}^2 (\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2 \sigma_{\epsilon}^2} - \frac{4(\theta_2 + x_0)^3}{(\theta_2 + x_0)^4 + \theta_1^2 \theta_2^2},$$

is negative for all $0 \leq x_0 < x_1 < x_2 \leq x_l$ and as before we conclude that $x_0^* = 0$.

The interior point x_1^* of the saturated locally D -optimal design can be found by differentiating the logarithm of $\det \{M_{LS}(\xi_{\theta}, \boldsymbol{\theta})\}$ for $x_0 = 0$ and $x_2 = x_u$ with respect to x_1 and equating the resulting expression to zero. This gives x_1^* as a solution of the equation

$$\frac{1}{x_1} - \frac{1}{x_u - x_1} - \frac{2(\theta_2 + x_1)^3}{(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2 \varrho_{\epsilon\eta}^2} + \frac{2\theta_1^2 \theta_2^2}{(\theta_2 + x_1) [(\theta_2 + x_1)^4 + \theta_1^2 \theta_2^2]} = 0,$$

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

Proof of Lemma 5: Let ξ_{θ}^* be a locally D -optimal design for least squares estimation in the exponential regression model given in (18) subject to measurement errors as in (1)

and (2). The necessary condition for D -optimality presented in Theorem 2 is given by

$$P_6(x) := e^{-2\theta_1 x} (1, -\theta_0 x) \{2D_0^{-1}(\xi_{\theta}^*, \theta) - [\sigma_{\eta}^2 + \theta_0^2 \theta_1^2 e^{-2\theta_1 x} \sigma_{\epsilon}^2] D_1^{-1}(\xi_{\theta}^*, \theta)\} (1, -\theta_0 x)^T \\ - 2 [1 + \theta_0^2 \theta_1^2 e^{-2\theta_1 x}] \leq 0, \quad \forall x \in \mathcal{X} = [x_l, x_u],$$

for the exponential regression model with equality at the support points of ξ_{θ}^* .

The function $P_6(x)$ is a linear combination of the functions $1, e^{-2\theta_1 x}, e^{-4\theta_1 x}, xe^{-2\theta_1 x}, xe^{-4\theta_1 x}, x^2 e^{-2\theta_1 x}, x^2 e^{-4\theta_1 x}$ which have Wronskian derivative equal to

$$W(1, e^{-2\theta_1 x}, \dots, x^2 e^{-4\theta_1 x}) = 1048576\theta_1^{15} e^{-18\theta_1 x} > 0, \quad \forall x \in \mathcal{X} \subset \mathbb{R}.$$

Therefore, these functions form an extended Chebyshev system and so $P_6(x)$ has at most 6 distinct zeros.

Taking the derivative of $P_6(x)$ with respect to x we get a linear combination of the functions $e^{-2\theta_1 x}, e^{-4\theta_1 x}, xe^{-2\theta_1 x}, xe^{-4\theta_1 x}, x^2 e^{-2\theta_1 x}, x^2 e^{-4\theta_1 x}$ which again have

$$W(e^{-2\theta_1 x}, \dots, x^2 e^{-4\theta_1 x}) = 2048\theta_1^9 e^{-18\theta_1 x} > 0, \quad \forall x \in \mathcal{X} \subset \mathbb{R},$$

and thus form an extended Chebyshev system. Hence $P_6(x)$ has at most 5 turning points.

The assumptions of 6 and 5 roots can be immediately rejected as these correspond to a total number of at least 9 and 7 turning points respectively. Therefore, the locally D -optimal design for the exponential regression model has at most 4 support points.