

# Designing to detect heteroscedasticity in a regression model

Alessandro Lanteri<sup>1</sup>, Samantha Leorato<sup>1</sup>, Jesús López-Fidalgo<sup>2</sup>  
and Chiara Tommasi<sup>1</sup>

<sup>1</sup>Department of Economics, Management and Quantitative Methods, University of Milan, Milano, Italy

<sup>2</sup>Institute of Data Science and Artificial Intelligence, University of Navarre, Pamplona, Spain

Address for correspondence: A. Lanteri, Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio, 7, 20122 Milano, Italy. Email: [alessandro.lanteri@unimi.it](mailto:alessandro.lanteri@unimi.it)

## Abstract

We consider the problem of designing experiments to detect the presence of a specified heteroscedasticity in Gaussian regression models. We study the relationship of the  $D_{S^*}$ - and KL-criteria with the noncentrality parameter of the asymptotic chi-squared distribution of a likelihood-based test, for local alternatives. We found that, when the heteroscedasticity depends on one parameter, the two criteria coincide asymptotically and that the  $D_1$ -criterion is proportional to the noncentrality parameter. Differently, when it depends on several parameters, the KL-optimum design converges to the design that maximizes the noncentrality parameter. Our theoretical findings are confirmed through a simulation study.

**Keywords:** asymptotic power, heteroscedasticity, likelihood-based tests, noncentrality parameter, optimal discrimination designs

## 1 Introduction

In the literature of Optimal Design of Experiments, many papers concern precise estimation of the regression coefficients or some transformation of them; classical references are [Atkinson et al. \(2007\)](#), [Silvey \(1980\)](#), and [Pázman \(1986\)](#), among others. Much less attention is deserved to the estimation of the error variance function, because frequently the error term is assumed to be homoscedastic or with a heteroscedasticity derived from a variance function known up to a proportionality constant.

The goal of this paper is to design an experiment to detect a specific kind of heteroscedasticity in a nonlinear Gaussian regression model

$$y_i = \mu(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad \varepsilon_i \sim N(0; \sigma^2 h(\mathbf{x}_i; \boldsymbol{\gamma})), \quad i = 1, \dots, n, \quad (1)$$

where  $y_i$  is a response variable with nonlinear mean function  $\mu(\mathbf{x}_i; \boldsymbol{\beta})$  and error variance  $\sigma^2 h(\mathbf{x}_i; \boldsymbol{\gamma})$ ; herein  $\boldsymbol{\beta} \in \mathbb{R}^m$  is the vector of regression coefficients,  $\sigma^2$  is an unknown constant and  $h: \mathbb{R}^p \times \mathbb{R}^s \mapsto \mathbb{R}_+$  is a continuous positive function that depends on a parameter vector  $\boldsymbol{\gamma} \in \mathbb{R}^s$ . Assume that there exists a value  $\boldsymbol{\gamma}_0$  such that  $h(\mathbf{x}_i; \boldsymbol{\gamma}_0) = 1$ ; in other words,  $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$  leads to the homoscedastic model.

In many practical problems, it may be convenient to test for the model heteroscedasticity with a likelihood-based test (log-likelihood ratio, score, or Wald test), since its asymptotic distribution is known to be a chi-squared random variable with  $s$  degrees of freedom. In particular, we consider local alternatives:

$$\begin{cases} H_0: & \boldsymbol{\gamma} = \boldsymbol{\gamma}_0, \\ H_1: & \boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \frac{\boldsymbol{\lambda}}{\sqrt{n}}, \quad \boldsymbol{\lambda} \neq \mathbf{0}, \end{cases} \quad (2)$$

with  $\lambda \in \mathbb{R}^s$ , and we aim at designing an experiment with the goal of maximizing in some sense the (asymptotic) power of a likelihood-based test. In this study, we justify the use of the  $D_s$ - and the KL-criterion (López-Fidalgo et al., 2007) to optimally design an experiment with the inferential goal of testing hypotheses (2). Some other authors who considered the inferential issue of hypotheses testing (but concerning the regression functions) are Stigler (1971), Spruill (1990), and Dette and Titoff (2009), who used the  $D_s$ -optimality and/or the  $T$ -criterion proposed by Atkinson and Fedorov (1975); see also Chapter 6 of Fedorov (1972).

Both the  $D_s$ - and KL-criteria are related, in a different way, to the noncentrality parameter of the asymptotic chi-squared distribution of a likelihood-based test. We prove that the first-order approximation of KL-criterion for discriminating between the homoscedastic and the heteroscedastic models coincides with the noncentrality parameter. Although the same result does not hold in general for the  $D_s$ -optimality, the  $D_s$ -criterion is proportional to the noncentrality parameter when  $\gamma$  is scalar. Therefore, in the scalar case, the two criteria lead to the same optimal design, asymptotically.

The most relevant result is that the sequence of the KL-optimum designs is proved to converge to the design that maximizes the noncentrality parameter. Furthermore, the analytic expression of this limiting design (at which we expect to reach the maximum asymptotic power of the test) is also provided.

The paper is organized as follows. In Section 2, with regard to problem (2), we give the expression of the noncentrality parameter for the chi-square asymptotic distribution of a likelihood-based test. We also recall how the  $D_s$ -criterion is related to the noncentrality parameter and we prove that, in this context, the  $D_s$ -optimum design can be obtained as a D-optimality solution. In Section 3, we provide the closed-form expression for the KL-optimal design, which is quite uncommon. Furthermore, we prove that the KL-criterion differs from the noncentrality parameter by a term that goes to zero uniformly; this implies that the sequence of the KL-optimum designs converges to the design maximizing the noncentrality parameter. Section 4 presents three examples where we compute the asymptotic power of the log-likelihood ratio statistic for several designs; in addition, the exact powers are computed through a simulation study. These numerical outcomes confirm the theoretical findings. All the proofs of the theoretical results stated in the main body of this manuscript can be found in the [online supplementary material](#).

## 2 $D_s$ -criterion and noncentrality parameter

Let  $\mathbf{x}$  denote an experimental point that can vary in a compact experimental domain  $\mathcal{X} \subseteq \mathbb{R}^p$ . Following Kiefer and Wolfowitz (1959), an approximate design  $\zeta = \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ \zeta(\mathbf{x}_1) & \cdots & \zeta(\mathbf{x}_k) \end{array} \right\}$  is a discrete probability measure on  $\mathcal{X}$ . Let  $\Xi$  denote the set of all possible designs. A design is said ‘optimal’ if it maximizes some concave criterion function of  $\zeta \in \Xi$ , which reflects the inferential goal. The  $D_s$ -criterion of optimality is commonly used for precise estimation of a subset of parameters. Herein, we are interested in estimating as precisely as possible the variance parameter vector  $\gamma$ . The  $D_s$ -optimum design maximizes the precision matrix of the maximum likelihood estimator of  $\gamma$ .

From model (1), the log-likelihood function for one observation at  $\mathbf{x}$  is

$$l(\gamma; \boldsymbol{\beta}, \sigma^2, \gamma, \mathbf{x}) = \left\{ -\frac{1}{2} \frac{[y - \mu(\mathbf{x}; \boldsymbol{\beta})]^2}{\sigma^2 h(\mathbf{x}; \gamma)} - \frac{1}{2} [\log(2\pi) + \log(\sigma^2) + \log(h(\mathbf{x}; \gamma))] \right\}.$$

The score vector, denoted by  $\mathbf{u} = (\mathbf{u}_\beta^T, u_{\sigma^2}, \mathbf{u}_\gamma^T)^T$  is

$$\begin{aligned} \mathbf{u}_\beta &= \frac{\partial l(\gamma; \boldsymbol{\beta}, \sigma^2, \gamma, \mathbf{x})}{\partial \boldsymbol{\beta}} = \frac{[y - \mu(\mathbf{x}; \boldsymbol{\beta})]}{\sigma^2 h(\mathbf{x}; \gamma)} \nabla \mu(\mathbf{x}; \boldsymbol{\beta}), \\ u_{\sigma^2} &= \frac{\partial l(\gamma; \boldsymbol{\beta}, \sigma^2, \gamma, \mathbf{x})}{\partial \sigma^2} = \frac{1}{2} \frac{[y - \mu(\mathbf{x}; \boldsymbol{\beta})]^2}{\sigma^4 h(\mathbf{x}; \gamma)} - \frac{1}{2\sigma^2}, \\ \mathbf{u}_\gamma &= \frac{\partial l(\gamma; \boldsymbol{\beta}, \sigma^2, \gamma, \mathbf{x})}{\partial \gamma} = \frac{1}{2} \frac{\nabla h(\mathbf{x}; \gamma) [y - \mu(\mathbf{x}; \boldsymbol{\beta})]^2}{\sigma^2 h(\mathbf{x}; \gamma)^2} - \frac{1}{2} \frac{\nabla h(\mathbf{x}; \gamma)}{h(\mathbf{x}; \gamma)}, \end{aligned}$$

where  $\nabla\mu(\mathbf{x}; \boldsymbol{\beta}) = \partial\mu(\mathbf{x}; \boldsymbol{\beta})/\partial\boldsymbol{\beta}$  and  $\nabla h(\mathbf{x}; \gamma) = \partial h(\mathbf{x}; \gamma)/\partial\gamma$ . After some algebra, we have that

$$J(\mathbf{x}; \boldsymbol{\beta}, \sigma^2, \gamma) = E[\mathbf{u}\mathbf{u}^T] = \begin{bmatrix} \frac{\nabla\mu(\mathbf{x}; \boldsymbol{\beta})\nabla\mu(\mathbf{x}; \boldsymbol{\beta})^T}{\sigma^2 h(\mathbf{x}; \gamma)} & 0 & \mathbf{0}^T \\ 0 & \frac{1}{2\sigma^4} & \frac{\nabla h(\mathbf{x}; \gamma)^T}{2\sigma^2 h(\mathbf{x}; \gamma)} \\ 0 & \frac{\nabla h(\mathbf{x}; \gamma)}{2\sigma^2 h(\mathbf{x}; \gamma)} & \frac{\nabla h(\mathbf{x}; \gamma)\nabla h(\mathbf{x}; \gamma)^T}{2h(\mathbf{x}; \gamma)^2} \end{bmatrix},$$

where the expectation is taken with respect to the Normal law.

The Fisher information matrix is

$$\mathcal{I}(\boldsymbol{\zeta}; \boldsymbol{\beta}, \sigma^2, \gamma) = \sum_{i=1}^k J(\mathbf{x}_i; \boldsymbol{\beta}, \sigma^2, \gamma) \boldsymbol{\zeta}(\mathbf{x}_i) = \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{12}^T & \mathcal{I}_{22} \end{bmatrix},$$

where

$$\mathcal{I}_{11} = \begin{bmatrix} \mathbf{M}(\boldsymbol{\zeta}; \boldsymbol{\beta}, \sigma^2, \gamma) & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}; \quad \mathbf{M}(\boldsymbol{\zeta}; \boldsymbol{\beta}, \sigma^2, \gamma) = \sum_{i=1}^k \frac{\nabla\mu(\mathbf{x}_i; \boldsymbol{\beta})\nabla\mu(\mathbf{x}_i; \boldsymbol{\beta})^T}{\sigma^2 h(\mathbf{x}_i; \gamma)} \boldsymbol{\zeta}(\mathbf{x}_i),$$

$$\mathcal{I}_{12} = \begin{bmatrix} \mathbf{0}^T \\ \sum_{i=1}^k \frac{\nabla h(\mathbf{x}_i; \gamma)^T}{2\sigma^2 h(\mathbf{x}_i; \gamma)} \boldsymbol{\zeta}(\mathbf{x}_i) \end{bmatrix}; \quad \mathcal{I}_{22} = \sum_{i=1}^k \frac{\nabla h(\mathbf{x}_i; \gamma)\nabla h(\mathbf{x}_i; \gamma)^T}{2 h(\mathbf{x}_i; \gamma)^2} \boldsymbol{\zeta}(\mathbf{x}_i).$$

The asymptotic covariance matrix of the MLE for  $\boldsymbol{\gamma}$  is  $[\mathcal{I}_{22.1}(\boldsymbol{\zeta}; \boldsymbol{\gamma})]^{-1}$ , where  $\mathcal{I}_{22.1}(\boldsymbol{\zeta}; \boldsymbol{\gamma}) = \mathcal{I}_{22} - \mathcal{I}_{12}^T \mathcal{I}_{11}^{-1} \mathcal{I}_{12}$  is the Schur complement of  $\mathcal{I}_{22}$  in  $\mathcal{I}(\boldsymbol{\zeta}; \boldsymbol{\beta}, \sigma^2, \gamma)$ . The  $D_s$ -optimum design for  $\boldsymbol{\gamma}$  minimizes the generalized variance of the MLE for  $\boldsymbol{\gamma}$ :

$$\boldsymbol{\zeta}_{D_s} = \arg \max_{\boldsymbol{\zeta}} |\mathcal{I}_{22.1}(\boldsymbol{\zeta}; \boldsymbol{\gamma})|. \tag{3}$$

The following theorem proves that the  $D_s$ -optimality for  $\boldsymbol{\gamma}$  actually is D-optimality for an ‘auxiliary’ linear regression. This result is quite useful from a practical point of view, because there is much statistical software to compute D-optimum designs, which can be exploited to determine  $\boldsymbol{\zeta}_{D_s}$ .

**Theorem 1** The  $D_s$ -optimum design for  $\boldsymbol{\gamma}$  coincides with the D-optimum design for estimating  $(\alpha_0, \boldsymbol{\alpha}^T)$ , with  $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_s)$ , in the following linear regression model:

$$y_i = \alpha_0 + \boldsymbol{\alpha}^T \nabla \log h(\mathbf{x}_i; \tilde{\boldsymbol{\gamma}}) + \varepsilon_i, \quad \varepsilon_i \sim N(0; \sigma^2), \quad i = 1, \dots, n,$$

where

$$\nabla \log h(\mathbf{x}; \boldsymbol{\gamma}) = \left( \frac{\partial \log h(\mathbf{x}; \boldsymbol{\gamma})}{\partial \gamma_1}, \dots, \frac{\partial \log h(\mathbf{x}; \boldsymbol{\gamma})}{\partial \gamma_s} \right)^T$$

and  $\tilde{\boldsymbol{\gamma}}$  is a nominal value for  $\boldsymbol{\gamma}$ .

**Corollary 1** If  $\boldsymbol{\gamma}$  is scalar, then  $s = 1$  and the  $D_1$ -optimal design for  $\boldsymbol{\gamma}$  is

$$\boldsymbol{\zeta}_{D_1} = \left\{ \begin{matrix} x_l & x_u \\ 0.5 & 0.5 \end{matrix} \right\}, \tag{4}$$

where  $x_l \in \{x : \text{*argmin}_x \nabla \log h(x, \tilde{\boldsymbol{\gamma}})\}$  and  $x_u \in \{x : \text{*argmax}_x \nabla \log h(x, \tilde{\boldsymbol{\gamma}})\}$  with  $\nabla \log h(x, \boldsymbol{\gamma}) = \partial \log h(\mathbf{x}_i; \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}$ .

When the inferential goal is testing homoscedasticity ( $\gamma = \gamma_0$ ) versus the local alternatives  $\gamma = \gamma_0 + \lambda/\sqrt{n}$ , the  $D_s$ -optimum design is commonly applied because somehow it maximizes the asymptotic power of a likelihood-based test. It is well known that, under the alternative hypotheses, the log-likelihood ratio, Wald and Score tests are asymptotically distributed as a chi-squared random variable with noncentrality parameter  $\zeta(\xi; \lambda; \gamma_0) = \lambda^T \mathcal{I}_{22.1}(\xi; \gamma_0) \lambda$ .

Therefore, the  $D_s$ -optimum design (3) with  $\gamma = \gamma_0$  maximizes 'in some sense'  $\zeta(\xi; \lambda; \gamma_0)$ , without considering any specific value of  $\lambda$ .

Replacing  $\mathcal{I}_{11}$ ,  $\mathcal{I}_{12}$ , and  $\mathcal{I}_{22}$  into the expressions of  $\mathcal{I}_{22.1}(\xi; \gamma)$  and setting  $\gamma = \gamma_0$ , we can provide the analytic expression of the noncentrality parameter:

$$\begin{aligned} \zeta(\xi; \lambda; \gamma_0) = & \frac{1}{2} \lambda^T \left( \sum_{i=1}^k \nabla h(\mathbf{x}_i; \gamma_0) \nabla h(\mathbf{x}_i; \gamma_0)^T \xi(\mathbf{x}_i) \right. \\ & \left. - \sum_{i=1}^k \nabla h(\mathbf{x}_i; \gamma_0) \xi(\mathbf{x}_i) \sum_{i=1}^k \nabla h(\mathbf{x}_i; \gamma_0)^T \xi(\mathbf{x}_i) \right) \lambda. \end{aligned} \quad (5)$$

**Remark 1** From Corollary 1, if  $\gamma$  is scalar and the inferential goal is testing hypothesis (2), then the  $D_1$ -optimal design for  $\gamma$  is equally supported at  $\mathbf{x}_l \in \operatorname{argmin}_x \nabla h(\mathbf{x}; \gamma_0)$  and  $\mathbf{x}_u \in \operatorname{argmax}_x \nabla h(\mathbf{x}; \gamma_0)$ . Furthermore, from Equation (5), it exactly maximizes the noncentrality parameter.

### 3 Noncentrality parameter as first-order approximation of the KL-criterion

The KL-criterion proposed by López-Fidalgo et al. (2007) can be applied to discriminate between heteroscedastic and homoscedastic models, under  $H_1$  and  $H_0$ , respectively. If the heteroscedastic model is completely known, i.e., if  $(\boldsymbol{\beta}_1^T, \sigma_1, \gamma_1^T)$  are the assumed known parameter values under  $H_1$ , then the KL-criterion is

$$I_{12}(\xi; \gamma_1) = \min_{\boldsymbol{\beta}, \sigma^2} \sum_{i=1}^k \left( \frac{\sigma_1^2 b(\mathbf{x}_i) + [\mu(\mathbf{x}_i; \boldsymbol{\beta}_1) - \mu(\mathbf{x}_i; \boldsymbol{\beta})]^2}{\sigma^2} - \log \frac{\sigma_1^2 b(\mathbf{x}_i)}{\sigma^2} \right) \xi(\mathbf{x}_i), \quad (6)$$

where  $b(\mathbf{x}_i) = b(\mathbf{x}_i; \gamma_1)$ . Differentiating the sum on the right-hand side of (6) with respect to  $\boldsymbol{\beta}$  and  $\sigma^2$  and setting it equal to zero the KL-criterion becomes

$$I_{12}(\xi; \gamma_1) = 1 + \log A_b - \log G_b, \quad (7)$$

where  $A_b = \sum_{i=1}^k b(\mathbf{x}_i) \xi(\mathbf{x}_i)$  and  $G_b = \prod_{i=1}^k [b(\mathbf{x}_i)]^{\xi(\mathbf{x}_i)}$  are the arithmetic and the geometric means of the values  $b(\mathbf{x}_i)$ ,  $i = 1, \dots, k$ , respectively.

For the next results we need to introduce some regularity assumptions on  $b$ . Let  $\mathcal{B}(\gamma_0, r)$  be the ball of radius  $r$  centred at  $\gamma_0$ . We assume that there exists a value  $r > 0$ , and finite constants  $M_1$  and  $M_2$  such that the function  $b(\mathbf{x}, \gamma)$  is twice continuously differentiable with respect to  $\gamma \in \mathcal{B}(\gamma_0, r)$  for all  $\mathbf{x} \in \mathcal{X}$  and that

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla b(\mathbf{x}, \gamma_0)\| & \leq M_1 < \infty; \\ \sup_{\mathbf{x} \in \mathcal{X}} \sup_{\gamma \in \mathcal{B}(\gamma_0, r)} \|b''(\mathbf{x}, \gamma)\|_2 & \leq M_2 < \infty, \end{aligned} \quad (8)$$

where  $b''(\mathbf{x}, \gamma) = (\partial^2 / \partial \gamma \partial \gamma^T) b(\mathbf{x}, \gamma)$  and  $\|A\|_2$  indicates the spectral norm of a square matrix  $A$ :  $\|A\|_2 = \sup_{\mathbf{u} \neq 0} \|A\mathbf{u}\| / \|\mathbf{u}\|$ , that coincides with the square root of the largest eigenvalue of  $A^T A$ . The next result follows from Equation (7).

**Theorem 2** Let  $\gamma_1 = \gamma_0 + \lambda/\sqrt{n}$ , and let assumption (8) hold for some constants  $M_1, M_2$  and  $r > 0$ . Then, the KL-criterion is related to the noncentrality parameter of the asymptotic distribution of a likelihood-based test by

$$I_{12}(\zeta; \gamma_1) = 1 + \frac{1}{n} \zeta(\zeta; \lambda; \gamma_0) + O\left(\frac{\|\lambda\|^3}{n^{3/2}}\right).$$

Theorem 2 essentially states that if  $n$  is large enough to make the error term negligible, the KL-optimum design maximizes the noncentrality parameter. Therefore, when  $\gamma$  is scalar (and  $n$  is sufficiently large), KL- and  $D_1$ -optimal designs should almost coincide, because the  $D_1$ -optimality is proportional to the noncentrality parameter. Differently, when  $\gamma$  is not scalar, the  $D_s$ -criterion is no longer proportional to the noncentrality parameter and therefore, KL- and  $D_s$ -optimum designs differ. In Section 4, for some values of  $\lambda$  and  $n$ , we compare the power of the log-likelihood ratio test when different designs are applied: the KL-optimum design for several specific values of  $\lambda$ , the  $D_s$ -optimal design and a uniform design, which is frequently applied in practice.

The following theorem has been very helpful to develop the simulation study, because it provides a closed form for the KL-optimum design.

**Theorem 3** Let us denote by  $\underline{h} = \inf_x b(x; \gamma_1) > 0$  and  $\bar{h} = \sup_x b(x; \gamma_1) < \infty$ . Let further  $\mathcal{X}_l = \{x : b(x; \gamma_1) = \underline{h}\}$  and  $\mathcal{X}_u = \{x : b(x; \gamma_1) = \bar{h}\}$ . A KL-optimal design is

$$\zeta_{\gamma_1}^{KL} = \left\{ \begin{array}{cc} \underline{x} & \bar{x} \\ \omega & 1 - \omega \end{array} \right\}, \quad \omega := \omega(\gamma_1) = \frac{\bar{h}}{\bar{h} - \underline{h}} - \frac{1}{\log \bar{h} - \log \underline{h}}, \quad (9)$$

where  $\underline{x} \in \mathcal{X}_l$  and  $\bar{x} \in \mathcal{X}_u$ .

If  $n\omega$  is not an integer number, then the best approximation is its integer part.

Under mild assumptions, the following theorem proves that, when  $n$  goes to infinity (and thus  $\gamma_1 \rightarrow \gamma_0$ ), the KL-optimum design tends to become equally supported at  $\underline{x}$  and  $\bar{x}$ .

**Theorem 4** Let  $\bar{h} = \sup_x b(x; \gamma_1)$  and  $\underline{h} = \inf_x b(x; \gamma_1)$ . If  $\underline{x} = \arg \inf_x b(x; \gamma_1)$  and  $\bar{x} = \arg \sup_x b(x; \gamma_1)$  do not depend on  $\gamma_1$  and  $b(x; \gamma_1)$  is such that  $\bar{h}/\underline{h} \rightarrow 1$  as  $\gamma_1 \rightarrow \gamma_0$ , then

$$\lim_{\gamma_1 \rightarrow \gamma_0} \omega(\gamma_1) = 1/2,$$

where  $\omega(\gamma_1)$  is defined by (9). In particular, for all  $\gamma_1$  such that  $\bar{h}/\underline{h} > 1$ ,

$$\omega(\gamma_1) > 1/2.$$

Let the limiting KL-optimal design be denoted by

$$\zeta_{\gamma_0}^{KL} = \left\{ \begin{array}{cc} \underline{x} & \bar{x} \\ 0.5 & 0.5 \end{array} \right\}.$$

Theorem 2 shows the asymptotic link between the KL-divergence and the noncentrality parameter of a likelihood-based test (under the alternative). The following result proves that the asymptotic expansion in Theorem 2 holds uniformly in  $\zeta$ , and this implies that, as  $n$  goes to infinity, the sequence of KL-optimal designs converges to the design which maximizes the noncentrality parameter.

**Theorem 5** Assume that (8) holds for some constants  $M_1, M_2$ , and  $r > 0$ . Also assume that  $\lambda$  belongs to a bounded set  $\Lambda : \sup_{\lambda \in \Lambda} \|\lambda\| \leq L < \infty$ . Then, the following result holds for all  $\lambda \in \Lambda$

$$\sup_{\zeta \in \Xi} |n(I_{12}(\zeta; \gamma_1) - 1) - \zeta(\zeta, \lambda, \gamma_0)| = O(n^{-1/2}).$$

From Theorems 4 and 5 we have the following result.

**Corollary 2** The design which maximizes the noncentrality parameter  $\zeta(\zeta, \lambda, \gamma_0)$  is the limiting KL-optimal design  $\xi_{\gamma_0}^{\text{KL}}$ .

## 4 Applications

In this section, we compare the performance of different designs in testing hypotheses (2) through a simulation study. We apply the log-likelihood ratio test using samples of data generated from each design and then we compare the corresponding powers. We also develop a comparison among the asymptotic powers of the test statistic.

In these applications, we consider model (1) with  $\sigma^2=1$  and three different kinds of heteroscedasticity:

- Case 1:  $h(x, \gamma) = e^{\gamma x}$ ;
- Case 2:  $h(x, \gamma) = 1 + 0.1(\gamma x + \sin(2\pi\gamma x))$ ;
- Case 3:  $h(x, \gamma) = 1 + \gamma_1 x + \gamma_2 x^2$ .

Without loss of generality, we set  $x \in [0, 1]$  to guarantee the above variance functions to be positive for  $\gamma \in \mathbb{R}_+^s$ . As mean function we consider,

$$\mu(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x, \quad \beta_0 = \beta_1 = 1$$

in Cases 1 and 2, and the EMAX model

$$\mu(x, \boldsymbol{\beta}) = \beta_0 + \frac{\beta_1 x}{\beta_2 + x}, \quad \beta_0 = \beta_1 = 1 \text{ and } \beta_2 = 0.25$$

in Case 3.

Let us note that  $D_s$ - and KL-optimal designs depend only on the variance function  $h(x, \gamma)$  and are not affected by  $\mu(x, \boldsymbol{\beta})$ , as shown in Sections 2 and 3. The computation of the likelihood-based tests, however, depends on the choice of the mean function. Therefore, this choice acts on the finite sample powers. However, this dependence should disappear as  $n$  increases. We point out that, in Case 1, we have developed simulations also for the EMAX model but the results are not reported herein because they are very similar to those given in Table 2. In the same way, in Case 3, we have considered also the linear mean function obtaining results different but consistent (in their trend) with those displayed in Table 4.

In Cases 1 and 2, function  $h(x, \gamma)$  depends on a one-dimensional parameter  $\gamma \in \mathbb{R}$ , and thus  $s = 1$ . As we can appreciate from Figure 1, in the first case  $h(x, \gamma)$  is monotonic in  $x$ , while in the second, for large enough values of  $\gamma$ , describes a nonmonotonic heteroscedasticity. Case 3, instead, concerns a two-dimensional parameter  $\gamma \in \mathbb{R}^2$ , and thus  $s = 2$ . Note that, in all the three cases, the homoscedastic model is obtained when  $\gamma_0 = 0$  or  $\gamma_0 = 0$  (Cases 1 and 2 or Case 3, respectively). Therefore, from  $H_1$  in (2),  $\gamma = \lambda/\sqrt{n}$  in the first two cases and  $\gamma = (\gamma_1, \gamma_2)^T$  in Case 3, where  $\gamma_1 = \lambda_1/\sqrt{n}$  and  $\gamma_2 = \lambda_2/\sqrt{n}$ .

To discriminate these heteroscedastic models from the homoscedastic case, we firstly compute several KL-optimal designs (for different choices of  $\lambda, \lambda_1$ , and  $\lambda_2$ ) and the  $D_s$ -optimal design (for  $\gamma = 0$  and  $\gamma = 0$ ). In addition, as a benchmark, we consider a uniform design

$$\xi_{U_s} = \begin{Bmatrix} 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{Bmatrix}.$$

Uniform designs are a common choice in practice and, in specific contexts, they provide nice theoretical guarantees (see Biedermann & Dette, 2001; Wiens, 1991, 2019).

From each heteroscedastic model (Cases 1, 2, and 3) we generate  $n$  responses from model (1), in different settings, applying the KL-optimal design, the  $D_s$ -optimal design, and the uniform design described above.

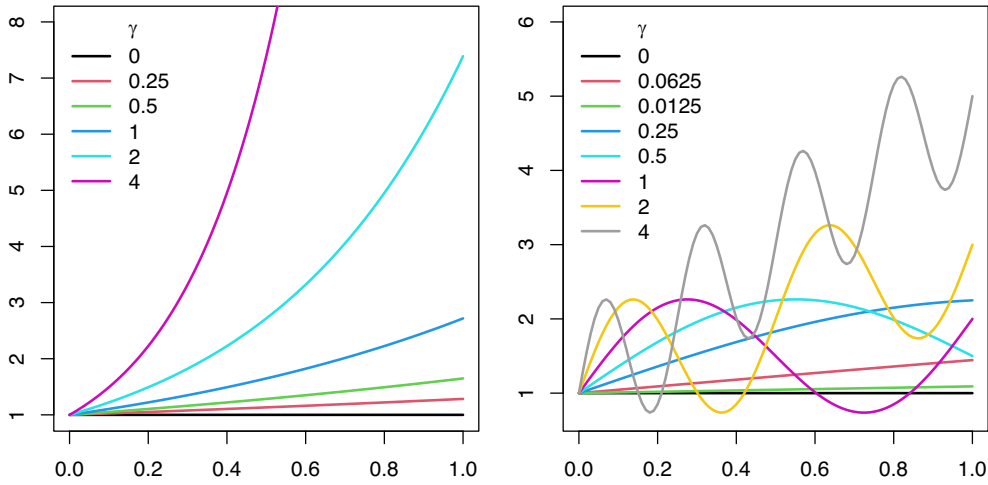


Figure 1. Case 1 (left) and Case 2 (right) variance functions  $h(x, \gamma)$ , for different values of  $\gamma$ .

Table 1. Optimal designs for Cases 1, 2, and 3

Case 1			
$\xi_{D_1} = \begin{Bmatrix} 0 & 1 \\ 0.500 & 0.500 \end{Bmatrix}$	$\xi_{0.25}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.521 & 0.479 \end{Bmatrix}$	$\xi_{0.5}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.542 & 0.458 \end{Bmatrix}$	
$\xi_{1}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.582 & 0.418 \end{Bmatrix}$	$\xi_{2}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.657 & 0.343 \end{Bmatrix}$	$\xi_{4}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.769 & 0.231 \end{Bmatrix}$	
Case 2			
$\xi_{D_1} = \begin{Bmatrix} 0 & 1 \\ 0.500 & 0.500 \end{Bmatrix}$	$\xi_{0.0625}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.504 & 0.496 \end{Bmatrix}$	$\xi_{0.125}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.507 & 0.493 \end{Bmatrix}$	$\xi_{0.25}^{KL} = \begin{Bmatrix} 0.00 & 0.55 \\ 0.510 & 0.490 \end{Bmatrix}$
$\xi_{0.5}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.510 & 0.490 \end{Bmatrix}$	$\xi_{1}^{KL} = \begin{Bmatrix} 0.72 & 0.28 \\ 0.512 & 0.488 \end{Bmatrix}$	$\xi_{2}^{KL} = \begin{Bmatrix} 0.36 & 0.64 \\ 0.520 & 0.480 \end{Bmatrix}$	$\xi_{4}^{KL} = \begin{Bmatrix} 0.18 & 0.89 \\ 0.532 & 0.468 \end{Bmatrix}$
Case 3			
$\xi_{D_2} = \begin{Bmatrix} 0 & 0.5 & 1 \\ 0.333 & 0.333 & 0.333 \end{Bmatrix}$	$\xi_{(0,0)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.500 & 0.500 \end{Bmatrix}$	$\xi_{(0.05,0.05)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.501 & 0.499 \end{Bmatrix}$	
$\xi_{(0.125,0.125)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.519 & 0.481 \end{Bmatrix}$	$\xi_{(0.125,0.25)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.526 & 0.474 \end{Bmatrix}$	$\xi_{(0.25,0.25)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.534 & 0.466 \end{Bmatrix}$	
$\xi_{(0.125,0.5)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.540 & 0.460 \end{Bmatrix}$	$\xi_{(0.25,0.5)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.546 & 0.454 \end{Bmatrix}$	$\xi_{(0.5,0.5)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.557 & 0.443 \end{Bmatrix}$	
$\xi_{(0.25,1)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.567 & 0.433 \end{Bmatrix}$	$\xi_{(0.5,1)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.575 & 0.425 \end{Bmatrix}$	$\xi_{(1,1)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.590 & 0.410 \end{Bmatrix}$	
$\xi_{(0.5,2)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.602 & 0.398 \end{Bmatrix}$	$\xi_{(1,2)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.612 & 0.388 \end{Bmatrix}$	$\xi_{(2,2)}^{KL} = \begin{Bmatrix} 0 & 1 \\ 0.629 & 0.371 \end{Bmatrix}$	

Table 1 lists all  $D_s$ - and KL-optimal designs we consider in our experiments. In the first two cases, the  $D_1$ -optimal design is supported at 0 and 1 as a consequence of Corollary 1. In Case 1, all the KL-optimal designs share the same support points, that are 0 and 1. This behaviour can be explained by Theorem 3: the monotonicity of the variance function implies that  $\underline{x}$  and  $\bar{x}$

**Table 2.** Monte Carlo actual significance level  $\hat{\alpha}$  and finite powers  $\hat{\eta}$ , and their asymptotic counterparts, obtained using different optimality criteria for Case 1, with variance function  $h(x; \gamma) = e^{\gamma x}$ , in various experimental settings

$\lambda$	$n$	$\gamma$	$\zeta_{\gamma}^{\text{KL}}$		$\zeta_{D_1}$		$\zeta_{U_S}$	
			$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$
5	25	1	0.0680	0.3777	0.0672	0.4090	0.0686	0.2362
	100	0.5	0.0555	0.4185	0.0509	0.4280	0.0548	0.2342
	400	0.25	0.0463	0.4334	0.0500	0.4203	0.0488	0.2434
	$\infty$	0	0.0500	0.7054	0.0500	0.7054	0.0500	0.4239
10	25	2	0.0674	0.8813	0.0672	0.9196	0.0672	0.6464
	100	1	0.0517	0.9304	0.0509	0.9371	0.0548	0.6937
	400	0.5	0.0517	0.9353	0.0500	0.9446	0.0488	0.7035
	$\infty$	0	0.0500	0.9988	0.0500	0.9988	0.0500	0.9424
20	25	4	0.0863	0.9962	0.0672	1.0000	0.0672	0.9886
	100	2	0.0591	1.0000	0.0509	1.0000	0.0548	0.9976
	400	1	0.0557	1.0000	0.0500	1.0000	0.0488	0.9990
	$\infty$	0	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

coincide with the extremes of the experimental domain  $[0; 1]$ . In Case 2, instead, the variance function is not monotone, for  $\gamma > 0.2755$ , and thus it does not necessarily reach its maximum and minimum at the extremes of  $[0; 1]$ . When the support of  $\zeta_{\gamma}^{\text{KL}}$  is not constant with respect to  $\gamma$ , the limit result in Theorem 4 does not necessarily apply, although it is still true that  $\omega(\gamma) > 1/2$ .

Let us note that, from Table 1 in both the scalar cases as  $\gamma = \lambda/\sqrt{n}$  goes to  $\gamma_0 = 0$ , the KL-optimal design approaches the  $D_1$ -optimal design, which is the design which maximizes the noncentrality parameter (see Remark 1); this is consistent with Equation (5) (Remark 1) and Corollary 2 that imply asymptotic equivalence of  $D_1$ - and KL-criteria. This is not true for Case 3, which concerns a multidimensional parameter  $\gamma$ ; in this case, the  $D_S$ -optimal design has three support points while the KL-optimum design is always supported at 0 and 1. Note that in our settings, for Case 3,  $\zeta_{(\gamma_1, \gamma_2)}^{\text{KL}} = \zeta_{(\gamma_2, \gamma_1)}^{\text{KL}}$ . For this reason we do not consider redundant couples in our outcomes.

In Tables 2–4, we display the experimental results for Cases 1, 2, and 3, respectively. We present Monte Carlo estimates of the actual significance level  $\hat{\alpha}$  and the finite test powers  $\hat{\eta}$  of the log-likelihood ratio (LR) test, for different optimality criteria in several experimental settings. With a slight abuse of notation, the rows corresponding to  $n = \infty$  display the nominal significance level  $\alpha$  and the asymptotic test powers. Let us note that the asymptotic powers have been analytically computed from

$$P(X_s^2[\zeta(\zeta; \lambda; \gamma_0)] \geq \chi_{s; 1-\alpha}^2),$$

where  $\chi_{s; 1-\alpha}^2$  is the quantile of a chi-squared random variable (rv) with  $s$  degrees of freedom (df),  $\alpha = 0.05$ , and  $X_s^2[\zeta(\zeta; \lambda; \gamma_0)]$  is a noncentral chi-squared rv with  $s$  df and noncentrality parameter  $\zeta(\zeta; \lambda; \gamma_0)$  given by (5). Finite sample powers are computed as the proportions in the  $M$  replications of LR-statistics larger than the asymptotic threshold  $\chi_{s; 1-\alpha}^2$ . In this paper, only tables for  $\alpha = 0.05$  are included. Different significant levels produce similar results.

The values of the asymptotic powers confirm the theoretical results described in Remark 1 and Corollary 2. From Tables 2 and 3, we can observe that the  $D_1$ -optimal design, which maximizes the noncentrality parameter for any value of  $\lambda$  in the scalar case, actually guarantees the largest asymptotic powers in all the scenarios. On the other hand, when  $\lambda$  is a vector, from Table 4 (see the rows corresponding to  $\gamma^T = (0, 0)$ ), we have that the design with the largest asymptotic power (for any choice of  $\lambda$ ) is the limiting KL-optimal design,  $\zeta_{\gamma_0}^{\text{KL}} = \zeta_0^{\text{KL}}$  and this is consistent with Corollary 2.



**Table 3.** Monte Carlo actual significance level  $\hat{\alpha}$  and finite powers  $\hat{\eta}$ , and their asymptotic counterparts, obtained using different optimality criteria for Case 2, with variance function  $h(x; \gamma) = 1 + 0.1[\gamma x + \sin(2\pi\gamma x)]$ , in various experimental settings

$\lambda$	$n$	$\gamma$	$\xi_{\gamma}^{KL}$		$\xi_{D_1}$		$\xi_{U_3}$	
			$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$
10	100	1	0.0000	0.0003	0.0000	0.0021	0.0000	0.0005
	400	0.5	0.0081	0.0305	0.0085	0.0108	0.0036	0.0120
	1,600	0.25	0.0480	0.3743	0.0482	0.3709	0.0596	0.2222
	6,400	0.125	0.0506	0.6096	0.0492	0.6159	0.0572	0.3921
	25,600	0.0625	0.0489	0.6916	0.0514	0.6943	0.0512	0.4049
	$\infty$	0	0.0500	0.9537	0.0500	0.9537	0.0500	0.7307
20	100	2	0.0000	0.0217	0.0000	0.0389	0.0000	0.0090
	400	1	0.0000	0.0899	0.0085	0.0814	0.0036	0.0739
	1,600	0.5	0.0493	0.3819	0.0482	0.0963	0.0596	0.2709
	6,400	0.25	0.0486	0.9133	0.0501	0.9125	0.0596	0.7274
	25,600	0.125	0.0518	0.9939	0.0514	0.9937	0.0541	0.8952
	$\infty$	0	0.0500	1.0000	0.0500	1.0000	0.0500	0.9993
40	100	4	0.0000	0.2236	0.0000	0.1945	0.0000	0.0635
	400	2	0.0000	0.3543	0.0085	0.2547	0.0036	0.1579
	1,600	1	0.0109	0.5399	0.0482	0.2660	0.0596	0.4365
	6,400	0.5	0.0521	0.9206	0.0501	0.2932	0.0596	0.7294
	25,600	0.25	0.0492	1.0000	0.0514	1.0000	0.0541	0.9955
	$\infty$	0	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

To analyse the finite sample performances of tests, we generated, for each design and for each scenario (choice of  $h$ ,  $\lambda$ , and  $n$ ),  $M = 10,000$  Monte Carlo replications of samples of increasing sizes. Specifically, we chose  $n = 25, 100, 400$  for Case 1 and Case 3, whereas  $n = 100, 400, 1,600, 6,400, 25,600$  for Case 2. In fact, the nonmonotonicity of the conditional variance function of Case 2 makes it quite difficult to discriminate between the two hypotheses in (2), thus a larger sample size is needed to obtain some nonzero power. Several alternatives are considered, defined by  $\lambda = \{5, 10, 20\}$  in Case 1,  $\lambda = \{10, 20, 40\}$  in Case 2, and  $\lambda_i = \{2.5, 5, 10\}$ , with  $i = 1, 2$ , in Case 3 (removing redundant couples).

**Case 1.** Besides  $\xi_{D_1}$  and  $\xi_{U_3}$ , we considered the KL-optimal designs,  $\xi_{\gamma}^{KL}$ , for  $\gamma = \lambda/\sqrt{n}$  in  $\{0.25, 0.5, 1, 2, 4\}$ .

From Table 2, which summarizes the finite sample and asymptotic results under Case 1, we may appreciate that, coherently with the asymptotic findings, the  $D_1$ -optimal design provides the highest finite power almost uniformly, with just one exception where  $\xi_{\gamma}^{KL}$  is slightly better. In addition, we can notice how, consistently with Remark 1, Corollary 2 and Theorem 4, the more  $\gamma$  decreases the smaller the difference between KL- and  $D_1$ -optimal designs.

In this example,  $h(x; \gamma) = e^{\gamma x}$  is a *well-behaved* variance function; powers increase rapidly both with  $n$  and  $\lambda$ , although for  $\lambda = 5$  it seems that a sample size of  $n = 400$  is not large enough to approach the asymptotic power. Our benchmark, the uniform design, appears to be the worst choice in all scenarios. For all criteria, the actual significance level converges to the nominal level, displayed on rows  $n = \infty$ , as  $n$  increases and it is relatively close to it already for  $n = 25$ . In particular, design  $\xi_{D_1}$  shows the fastest convergence.

**Case 2.** As previously pointed out, when the variance function is  $h(x; \gamma) = 1 + 0.1[\gamma x + \sin(2\pi\gamma x)]$ , powers of the LR-statistic are very low and tend to increase quite slowly with  $n$  and  $\lambda$ , as displayed in Table 3.

**Table 4.** Monte Carlo actual significance level  $\hat{\alpha}$  and finite powers  $\hat{\eta}$ , and their asymptotic counterparts, obtained using different optimality criteria for Case 3, with variance function  $h(x; \gamma) = 1 + \gamma_1 x + \gamma_2 x^2$  and mean function  $\mu(x, \beta) = \beta_0 + \beta_1 x / (\beta_2 + x)$ , in various experimental settings

$\lambda^T$	$n$	$\gamma^T$	$\xi_{\gamma}^{\text{KL}}$		$\xi_{D_2}$		$\xi_{U_5}$	
			$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\eta}$
(2.5, 2.5)	25	(0.5, 0.5)	0.0639	0.2300	0.1423	0.2808	0.1649	0.2973
	100	(0.25, 0.25)	0.0527	0.3016	0.1111	0.2855	0.1062	0.2682
	400	(0.125, 0.125)	0.0480	0.3448	0.0696	0.2668	0.0678	0.2031
	$\infty$	(0, 0)	0.0500	0.6028	0.0500	0.4384	0.0500	0.3400
(2.5, 5)	25	(0.5, 1)	0.0690	0.3387	0.1423	0.3656	0.1649	0.3632
	100	(0.25, 0.5)	0.0549	0.4928	0.1111	0.4466	0.1062	0.3875
	400	(0.125, 0.25)	0.0492	0.6093	0.0696	0.4360	0.0678	0.3551
	$\infty$	(0, 0)	0.0500	0.9291	0.0500	0.8038	0.0500	0.6753
(5, 5)	25	(1, 1)	0.0687	0.4530	0.1423	0.4495	0.1649	0.4318
	100	(0.5, 0.5)	0.0569	0.6743	0.1111	0.5777	0.1062	0.5011
	400	(0.25, 0.25)	0.0522	0.8125	0.0696	0.6315	0.0678	0.5135
	$\infty$	(0, 0)	0.0500	0.9965	0.0500	0.9668	0.0500	0.9028
(2.5, 10)	25	(0.5, 2)	0.0712	0.5234	0.1423	0.5240	0.1649	0.4960
	100	(0.25, 1)	0.0537	0.7974	0.1111	0.7054	0.1062	0.6286
	400	(0.125, 0.5)	0.0529	0.9283	0.0696	0.7821	0.0678	0.6738
	$\infty$	(0, 0)	0.0500	1.0000	0.0500	0.9983	0.0500	0.9876
(5, 10)	25	(1, 2)	0.0702	0.6049	0.1423	0.5849	0.1649	0.5584
	100	(0.5, 1)	0.0531	0.8793	0.1111	0.7800	0.1062	0.7039
	400	(0.25, 0.5)	0.0519	0.9739	0.0696	0.8861	0.0678	0.7896
	$\infty$	(0, 0)	0.0500	1.0000	0.05	1.0000	0.05	0.9990
(10, 10)	25	(2, 2)	0.0647	0.7235	0.1423	0.6827	0.1649	0.6234
	100	(1, 1)	0.0527	0.9624	0.1111	0.9004	0.1062	0.8289
	400	(0.5, 0.5)	0.0501	0.9976	0.0696	0.9701	0.0678	0.9264
	$\infty$	(0, 0)	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

It is worth noting that the actual significance level gets close to the nominal level only for  $n \geq 1,600$ , for all designs. This means that the asymptotic critical values used to determine the rejection region are too high, which reflects in a poor performance of the LR-test for small samples in all scenarios. In general, it seems that KL-optimal designs provide higher finite power faster (i.e., for lower  $n$ ) with respect to the  $D_1$ -optimal design. Furthermore, in our experiments, only for  $n = 25, 600$  and  $\gamma = 0.0625$  the  $D_1$ -optimal design shows its asymptotic dominance over its competitors, while in the other cases it is dominated by the KL-optimal design. Unfortunately, the finite sample power of KL-designs is strongly dependent on the knowledge of the alternative and thus it is difficult to suggest a rule of thumb for the choice of  $\gamma$  in the case of uncertainty. To this extent, we suggest the reading of [Tommasi and López-Fidalgo \(2010\)](#).

**Case 3.** [Table 4](#) displays the finite powers and actual significance level (and their asymptotic counterparts) of the LR-tests associated to the different designs for variance function  $h(x; \gamma) = 1 + \gamma_1 x + \gamma_2 x^2$  and mean function  $\mu(x, \beta) = \beta_0 + \beta_1 x / (\beta_2 + x)$ . We recall that, for this particular specification of  $h(\cdot; \cdot)$ ,  $\xi_{(\gamma_1, \gamma_2)}^{\text{KL}} = \xi_{(\gamma_2, \gamma_1)}^{\text{KL}}$ , and thus, without loss of generality, we consider only KL-optimal designs with  $\gamma_1 \leq \gamma_2$ .

The results in [Table 4](#) show that, for  $n \leq 100$ , the actual significance levels of the  $D_s$ - and the uniform designs are two to three times larger than the nominal level. We can notice that KL-optimal designs always dominate the other two designs, presenting larger finite powers and/

or smaller actual significant levels. This becomes more evident as  $n$  increases, which is in line with the asymptotic theory: for small values of  $\|\eta\|$  the KL-criterion tends to coincide with the noncentrality parameter of the asymptotic chi-squared distribution under the alternative (Theorem 2).

## 5 Conclusions

The main goal of this paper is the study of optimality criteria for detecting heteroscedasticity in a nonlinear Gaussian regression model, under local alternatives. Three tests of hypotheses are typically used for discriminating between nested rival models: the likelihood ratio, the score and the Wald tests. All of them are related to the likelihood function (thus, they are referred to likelihood-based tests) and, under local alternatives, they share asymptotically the same noncentral chi-squared distribution. On the other hand, from the experimental design perspective,  $D_s$ - and KL-optimality are the most common criteria to design experiments for discrimination. In this paper, we prove theoretically to what extent both these criteria are related to the noncentrality parameter of a likelihood-based test for discriminating heteroscedasticity versus homoscedasticity. The KL-criterion is proved to be asymptotically equivalent to the noncentrality parameter. Therefore, asymptotically the KL-optimum design guarantees the maximum power of the likelihood-based tests. The  $D_s$ -criterion instead is proportional to the noncentrality parameter whenever the variance function depends just on one parameter (even for finite samples, not only asymptotically). Therefore, in this last case, the two criteria of optimality are asymptotically equivalent.

The numerical outcomes obtained from a simulation study confirm these theoretical results.

*Conflict of interest:* The authors have no conflicts of interest to declare.

## Data availability

The code to reproduce the experiments that support the findings of this study is openly available at <https://github.com/alanteri/DDHRM>.

## Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*.

## References

- Atkinson A. C., Donev A. N., & Tobias R. D. (2007). *Optimum experimental designs, with SAS*. Oxford University Press.
- Atkinson A. C., & Fedorov V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, 62(1), 57–70. <https://doi.org/10.1093/biomet/62.1.57>
- Biedermann S., & Dette H. (2001). Optimal designs for testing the functional form of a regression via non-parametric estimation techniques. *Statistics & Probability Letters*, 52(2), 215–224. [https://doi.org/10.1016/S0167-7152\(00\)00244-3](https://doi.org/10.1016/S0167-7152(00)00244-3)
- Dette H., & Titoff S. (2009). Optimal discrimination designs. *The Annals of Statistics*, 37(4), 2056–2082. <https://doi.org/10.1214/08-AOS635>
- Fedorov V. V. (1972). *Theory of optimal experiments*. Academic Press.
- Kiefer J., & Wolfowitz J. (1959). Optimum designs in regression models. *The Annals of Mathematical Statistics*, 30(2), 271–294. <https://doi.org/10.1214/aoms/1177706252>
- López-Fidalgo J., Tommasi C., & Trandafir P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 231–242.
- Pázman A. (1986). *Foundation of optimal experimental design*. D. Reidel Publishing Company.
- Silvey S. D. (1980). *Optimal design*. Chapman and Hall.
- Spruill M. C. (1990). Good designs for testing the degree of a polynomial mean. *Sankhya, Series B*, 52(1), 67–74. <https://www.jstor.org/stable/25052622>
- Stigler S. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, 66(334), 311–318. <https://doi.org/10.1080/01621459.1971.10482260>
- Tommasi C., & López-Fidalgo J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Computational Statistics and Data Analysis*, 54(1), 143–150. <https://doi.org/10.1016/j.csda.2009.07.022>

- Wiens D. P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Statistics & Probability Letters*, 12(3), 217–221. [https://doi.org/10.1016/0167-7152\(91\)90081-2](https://doi.org/10.1016/0167-7152(91)90081-2)
- Wiens D. P. (2019). Maximin power designs in testing lack of fit. *Journal of Statistical Planning and Inference*, 199, 311–317. <https://doi.org/10.1016/j.jspi.2018.07.007>