DT-Optimum designs for model discrimination and parameter estimation

A.C. Atkinson*, Department of Statistics
London School of Economics, London WC2A 2AE, UK

February 18, 2005

Abstract

The paper introduces DT optimum designs that provide a specified balance between model discrimination and parameter estimation. An equivalence theorem is proved for the case of two models and extended to an arbitrary number of models and of combinations of parameters. A numerical example shows the power of the procedure. The relationship with designs for choosing the degree of a polynomial is established.

Keywords: D-optimum design; D_1-optimum design; Equivalence theorem; Nonlinear model; T-optimum design

1 Introduction

Optimum designs for discrimination between models often have very poor properties for estimation of the parameters in the chosen model. For example, Atkinson and Fedorov (1975) consider discrimination between two linear polynomial regression models, a constant and a quadratic. The T-optimum design for discrimination between these two models puts points at the minimum and maximum of the quadratic over the design region. This two-point design is totally uninformative for estimation of the three parameters of the quadratic model, should this be the true model. There is a long history of papers that seek to find a balance between model discrimination and parameter estimation, at least from Hill, Hunter, and Wichern (1968) to Biswas and Chaudhuri (2002) and Waterhouse, Eccleston, and Duffull (2004). Here

*e-mail: `a.c.atkinson@lse.ac.uk`

we use optimum design theory to provide designs of known properties with a specified balance between estimation and discrimination. An equivalence theorem is stated, which makes available the standard algorithms for the construction and checking of optimum designs.

T-optimum designs for discrimination between two models are recalled in §2. The better-known D-optimum designs for parameter estimation are very briefly described in §3. The main theoretical results for two models are in §4, with a numerical example in §5. The generalisation to any number of models and subsets of parameters is in §6. Linear models are specifically considered in §7 and a link made to work on determining the order of a one-factor polynomial.

2 T-optimum Designs

Atkinson and Fedorov (1975) consider designs for discrimination between two rival regression models $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$. The models may be linear or nonlinear in the parameters which are estimated by least squares. Suppose that the first model is true, so that the observations

$$y_i = \eta_t(x) + \epsilon_i = \eta_1(x, \theta_1) + \epsilon_i. \tag{1}$$

For a design ξ that puts weights w_i at k support points $x_i \in \mathcal{X}$, the lack of fit sum of squares for model 2 is made as large as possible by maximising

$$\Delta_1(\xi) = \sum_{i=1}^{k} w_i \{\eta_t(x_i) - \eta_2(x_i, \hat{\theta}_{t2})\}^2, \tag{2}$$

where

$$\sum_{i=1}^{k} w_i \{\eta_t(x_i) - \eta_2(x_i, \hat{\theta}_{t2})\}^2 = \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^{k} w_i \{\eta_t(x_i) - \eta_2(x_i, \theta_2)\}^2. \tag{3}$$

Let ξ_T^* be the design maximising (2). Under some regularity conditions, Atkinson and Fedorov (1975) prove the following theorem about T-optimum designs.

THEOREM 1.

(i) a necessary and sufficient condition for a design ξ_T^* to be T-optimum is fulfilment of the inequality

$$\psi_1(x, \xi_T^*) \leq \Delta_1(\xi_T^*) \qquad x \in \mathcal{X},$$

where

$$\psi_1(x,\xi) = \{\eta_t(x) - \eta_2(x, \hat{\theta}_{t2})\}^2;$$

(ii) the upper bound of $\psi_1(x, \xi_T^*)$ is achieved at the points of the optimum design;

(iii) for any nonoptimum design ξ, that is a design for which $\Delta_1(\xi) < \Delta_1(\xi_T^*)$,

$$\sup_{x\in\mathcal{X}} \psi_1(x,\xi) > \Delta_1(\xi_T^*).$$

The conditions include the continuity and differentiability with respect to θ_j of the models η_j. More importantly, when the design is singular for θ_2 in (3), it is necessary to extend (i) by introduction of a measure on θ_2. This measure is not needed in the example of §5.

Nesting makes it impossible to discriminate between all models. If model 1 is a special case of model 2, then $\Delta(\xi) = 0$ since, if model 1 is, as assumed, true, then model 2 must also be true. Partially nested linear models are considered in §7. In general the parameter estimate of the false model, $\hat{\theta}_{t2}$ will, as (3) and (1) show, depend on the design ξ and on θ_1. Thus T-optimum designs are often locally optimum.

3 D-optimum Designs

Suppose for the moment that the models are linear in the parameters, so that in (1) $\eta_1(x,\theta_1) = f_1^{\mathrm{T}}(x)\theta_1$, where θ_1 is $p_1 \times 1$. In matrix form this becomes $\mathrm{E}\,Y = F_1\theta_1$ where F_1 is $k \times p$. The information matrix of the least squares estimates $\hat{\theta}_1$ from the design ξ is

$$M_1(\xi) = F_1^{\mathrm{T}} W F_1,$$

where $W = \mathrm{diag}\,\{w_i\}$. D-optimum designs maximise $\log|M_1(\xi)|$. For nonlinear models the parameter sensitivities $F_1(x,\theta) = \{\partial\eta_1(x,\theta_1)/\partial\theta_{1j}\}$ and the locally D-optimum designs, introduced by Box and Lucas (1959), maximise $\log|M_1(\xi,\theta_o)| = \log|F_1(x,\theta_o)^{\mathrm{T}} W F_1(x,\theta_o)|$ for some point prior value θ_o. The original equivalence theorem for D-optimum designs is due to Kiefer and Wolfowitz (1960).

Theorem 2.

(i) a necessary and sufficient condition for a design ξ_D^* to be D-optimum is fulfilment of the inequality

$$d_1(x,\xi_D^*) \leq p_1 \qquad x \in \mathcal{X},$$

where

$$d_1(x,\xi_D^*) = f_1^{\mathrm{T}}(x) M_1^{-1}(\xi_D^*) f_1(x);$$

(ii) the upper bound of $d_1(x, \xi_D^*)$ is achieved at the points of the optimum design;

(iii) for any nonoptimum design ξ, that is a design for which $\log |M_1(\xi)| < \log |M_1(\xi_D^*)|$,

$$\sup_{x \in \mathcal{X}} d_1(x, \xi) > p_1.$$

4 DT-optimum Designs

The two theorems have the common structure of equivalence theorems for convex design problems (Fedorov 1972, Pukelsheim 1993, Fedorov and Hackl 1997). However, the two design criteria are very different in behaviour. The T-optimum criterion $\Delta(\xi)$ has the physical dimensions of a sum of squares of observations, whereas the D-optimum criterion $|M(\xi)|$ is a function solely of the regressor variables. Rescaling the y_i causes a change in the relative magnitudes of $\Delta_1(\xi)$ and $\log |M_1(\xi)|$. To combine the two criteria requires a common scale of comparison for which we use efficiencies. The T-efficiency of any design ξ relative to the T-optimum design ξ_T^* is

$$E_f^T(\xi) = \Delta_1(\xi)/\Delta_1(\xi_T^*), \tag{4}$$

whereas, the D-efficiency is

$$E_f^D(\xi) = \{|M_1(\xi)|/|M_1(\xi_D^*)|\}^{1/p_1}. \tag{5}$$

The T-efficiency (4) does not depend on the scaling of the y_i.

To obtain designs for both discrimination and parameter estimation we maximise a weighted product of the efficiencies

$$\{E_f^T(\xi)\}^{1-\kappa}\{E_f^D(\xi)\}^{\kappa} = \{\Delta_1(\xi)/\Delta_1(\xi_T^*)\}^{1-\kappa}\{|M_1(\xi)|/|M_1(\xi_D^*)|\}^{\kappa/p_1}\,(0 \le \kappa \le 1). \tag{6}$$

When $\kappa = 0$ we obtain T-optimality, with D-optimality for $\kappa = 1$. We investigate the properties of designs as κ varies. To clarify the structure of the design criterion, take logs in (6), when the right-hand side becomes

$$(1-\kappa)\log\Delta_1(\xi)+(\kappa/p_1)\log|M_1(\xi)|-(1-\kappa)\log\Delta_1(\xi_T^*)-(\kappa/p_1)\log|M_1(\xi_D^*)|. \tag{7}$$

The terms involving ξ_T^* and ξ_D^* are constants when a maximum is found over ξ, so that the criterion to be maximized is

$$\Phi_1^{(DT)}(\xi) = (1-\kappa)\log\Delta_1(\xi) + (\kappa/p_1)\log|M_1(\xi)|. \tag{8}$$

Designs maximizing (8) are called DT-optimum and are denoted ξ_{DT}^*.

THEOREM 3.

(i) a necessary and sufficient condition for a design ξ^*_{DT} to be DT-optimum is fulfilment of the inequality

$$\psi_1^{(DT)}(x, \xi^*_{DT}) \leq 1 \qquad x \in \mathcal{X},$$

where

$$\psi_1^{(DT)}(x,\xi) = (1-\kappa)\psi_1(x,\xi)/\Delta_1(\xi) + (\kappa/p_1)d_1(x,\xi)$$
$$= (1-\kappa)\{\eta_t(x) - \eta_2(x,\hat{\theta}_{t2})\}^2/\Delta_1(\xi) + (\kappa/p_1)f_1^{\mathrm{T}}(x)M_1^{-1}(\xi)f_1(x).$$

(ii) the upper bound of $\psi_1^{(DT)}(x, \xi^*_{DT})$ is achieved at the points of the optimum design;

(iii) for any nonoptimum design ξ, that is a design for which $\Phi_1^{(DT)}(\xi) < \Phi_1^{(DT)}(\xi^*_{DT})$,

$$\sup_{x \in \mathcal{X}} \psi_1^{(DT)}(x, \xi^*_{DT}) > 1.$$

PROOF

Since $0 \leq \kappa \leq 1$, $\Phi_1^{(DT)}(\xi)$ (8) is a convex combination of two design criteria, the second of which is D-optimality. The first criterion is $\log \Delta_1(\xi)$, the logarithm of that for T-optimality. As $\Delta_1(\xi) \geq 0$, $\log \Delta_1(\xi)$ is a concave function of a concave design criterion. Therefore the DT-criterion satisfies the conditions of convex optimum design theory and an equivalence theorem applies similar to Theorems 1 and 2.

To find the derivative function $\psi_1(x,\xi)$ for T-optimality Atkinson and Fedorov (1975) calculate the directional derivative $\{\partial\Delta_1(\xi)/\partial\alpha\}_{\alpha=0}$, with $\xi = (1-\alpha\xi_1) + \alpha\xi_2$. Since

$$\frac{\partial \log \Delta_1(\xi)}{\partial \alpha} = \frac{1}{\Delta_1(\xi)}\frac{\partial \Delta_1(\xi)}{\partial \alpha},$$

we obtain the first term in $\psi_1^{(DT)}(x,\xi)$. The second term is that from D-optimality and the theorem is proved.

5 A Numerical Example

To illustrate these results we return to Example 1 from Atkinson and Fedorov (1975) mentioned in §1. There are two models, a quadratic and a constant.

$$\eta_t(x) = \eta_1(x) = \theta_{1,0} + \theta_{1,1}x + \theta_{1,2}x^2, \quad \eta_2(x,\theta) = \theta_2. \qquad (9)$$

The simpler model 2 is thus nested within the quadratic. The T-optimum design provides maximum power for testing that $\theta_{1,1}$ and $\theta_{1,2}$ are both zero.

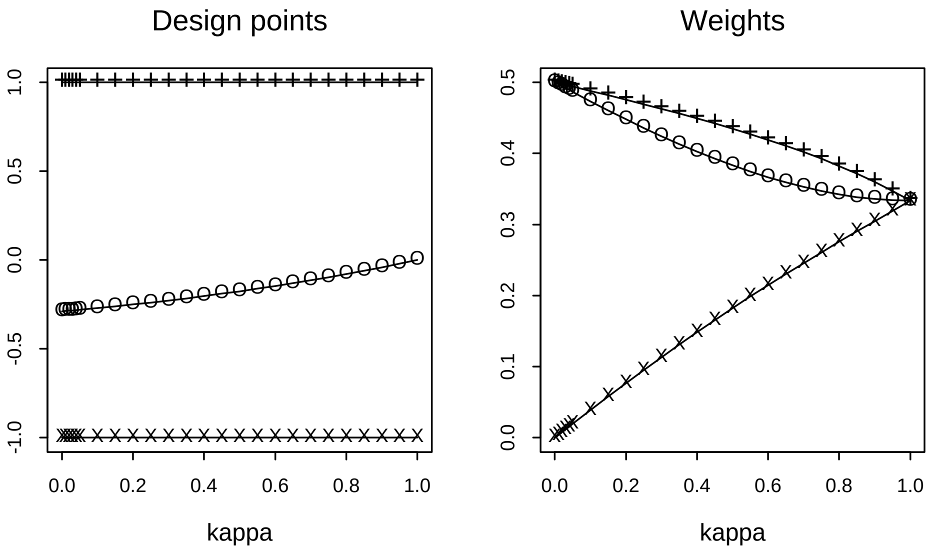

Figure 1: Structure of designs as κ varies. Left-hand panel, support points x_i; right-hand panel design weights w_i. Except for $\kappa = 0$, the T-optimum design, there are three points of support at -1, 1 and an intermediate value. For $\kappa = 1$ the D-optimum design with equal support at -1, 0 and 1 is obtained. The same plotting symbol is used in the two panels for the same value of i

For the parameter values $\theta_{1,0} = 1$, $\theta_{1,1} = 0.5$ and $\theta_{1,2} = 0.8660$, with $\mathcal{X}$ = [-1, 1] the two component optimum designs are

$$\xi_T^* = \left\{ \begin{array}{cc} -0.2887 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \; \xi_D^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}. \tag{10}$$

The two-point T-optimum design depends solely on the ratio $\theta_{1,2}/\theta_{1,1}$, here $\sqrt{3}$. It has zero efficiency for estimating the parameters of the model. On the other hand, the efficiency of the D-optimum design for testing, $E_f^T(\xi_D^*)$, is a respectable 64.46%.

To explore the properties of the proposed designs, DT-optimum designs were found for a series of values of κ between zero and one. The resulting designs are plotted in Figure 1. All designs were checked for optimality using the equivalence condition from Theorem 3. The left-hand panel shows the design points; for $\kappa > 0$ there are three at -1, 1 and an intermediate value which tends to 0 as $\kappa \to 1$. The weights in the right-hand panel change in a similarly smooth way from 0, 0.5 and 0.5 to one third at all design points when $\kappa = 1$.

The efficiencies in Figure 2 likewise change in a smooth manner. The T-efficiency decreases almost linearly as κ increases, whereas the D-efficiency increases rapidly away from zero as the design weight at $x = -1$ becomes

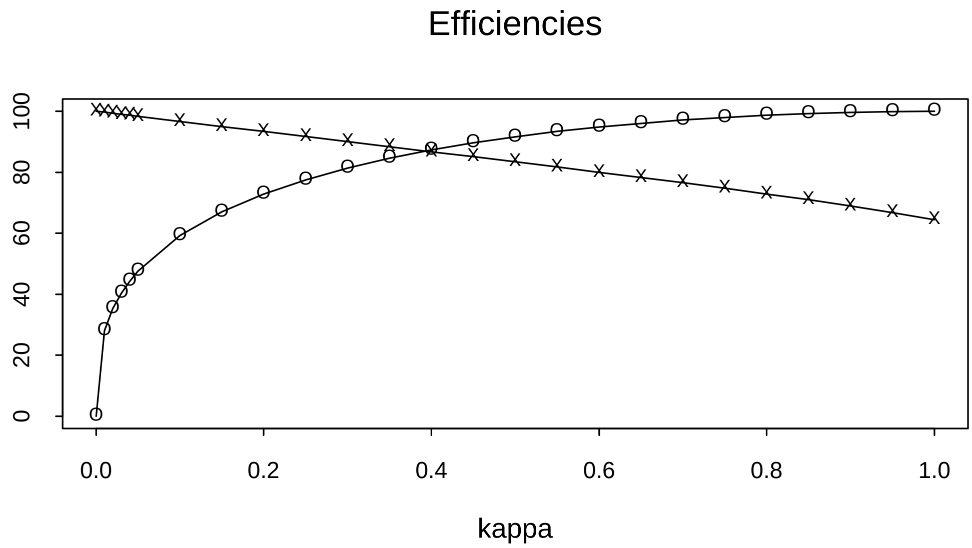

Figure 2: Efficiencies of designs as κ varies: O D-efficiency; X T-efficiency. The product of the efficiencies is a maximum at $\kappa = 0.5$

non-negligible. The product of these efficiencies is, as it must be from (8), a maximum at $\kappa = 0.5$. The T-efficiency for this value is 83.38%, with the D-efficiency equal to 91.63%. We have found a single design that is highly efficient both for parameter estimation and model testing.

6 Generalized DT-optimality

For discrimination between two separate, that is partially non-nested, models, both non-centrality parameters may be of interest as well as the two sets of model parameters.

The criterion and proof of hte equivalence theorem for DT-optimality extend straightforwardly to such generalized DT-optimum designs with c non-centrality parameters $\Delta_j(\xi)$ and m sets of model parameters p_k; in some cases only some subsets of the parameters may be of interest. The criterion (8) expands to

$$\Phi^{(GDT)}(\xi) = \sum_{j=1}^{c} a_j \log \Delta_j(\xi) - \sum_{k=1}^{m} (b_k/s_k) \log |A_k^{\mathrm{T}} M_k^{-1}(\xi) A_k|, \tag{11}$$

where the a_j and b_k are sets of non-negative coefficients reflecting the importance of the parts of the design criterion. The matrices of coefficients A_k are $p_k \times s_k$, defining the linear combinations of the p_k parameters in model k that are of interest. For D-optimality A_k is the identity matrix of dimension

p_k; for D_s-optimality $s_k < p_k$ and A_k is the $s_k \times s_k$ identity matrix adjoined with $p_k - s_k$ rows of zeroes. The negative sign for the second term on the right-hand side of (11) arises because the covariance matrix of the estimates is minimized.

The equivalence theorem states that

$$\psi^{(GDT)}(x, \xi^*_{GDT}) \leq \sum_{j=1}^{c} a_j + \sum_{k=1}^{m} b_k \qquad x \in \mathcal{X}, \qquad \text{where}$$

$$\psi^{(GDT)}(x, \xi) = \sum_{j=1}^{c} a_j \psi_j(x, \xi)/\Delta_j(\xi) + \sum_{k=1}^{m} (b_k/s_k) d_k^A(x, \xi) \quad \text{and}$$

$$d_k^A(x, \xi) = f_k^{\mathrm{T}} M_k^{-1}(\xi) A_k \{A_k^{\mathrm{T}} M_k^{-1}(\xi) A_k\}^{-1} A_k^{\mathrm{T}} M_k^{-1}(\xi) f_k.$$

The form of $d_k^A(x, \xi)$ comes from the equivalence theorem for D_A-optimality (Sibson 1974).

7 Linear Models

We now suppose explicitly that the two models are linear in the parameters

$$\text{Model 1:} \quad \mathrm{E}\, Y = F_1 \theta_1; \quad \text{Model 2:} \quad \mathrm{E}\, Y = F_2 \theta_2.$$

The non-centrality parameter when the first model is true and is not a special case of the second model is

$$\Delta_1(\xi) = \theta_1^{\mathrm{T}} M_{1(2)}(\xi) \theta_1, \qquad \text{where} \tag{12}$$

$$M_{1(2)}(\xi) = M_{11}(\xi) - M_{12}(\xi) M_{22}^{-1}(\xi) M_{21}(\xi), \quad M_{ij}(\xi) = F_i^{\mathrm{T}} W F_j.$$

If the two models have terms in common, the elements of $M_{1(2)}(\xi)$ corresponding to these elements are zero. Let the combined model with duplicate terms eliminated be

$$\mathrm{E}\, Y = F\theta = F_1 \theta_1 + \tilde{F}_2 \tilde{\theta}_2 = \tilde{F}_1 \tilde{\theta}_1 + F_2 \theta_2, \tag{13}$$

where $\tilde{F}_j \tilde{\theta}_j$ represent the complement of model not j in the combined model $F\theta$. Then the non-centrality parameter (12) can be replaced by

$$\Delta_1(\xi) = \tilde{\theta}_1^{\mathrm{T}} \tilde{M}_{1(2)}(\xi) \tilde{\theta}_1, \qquad \text{where} \tag{14}$$

$$\tilde{M}_{1(2)}(\xi) = \tilde{M}_{11}(\xi) - \tilde{M}_{12}(\xi) M_{22}^{-1}(\xi) \tilde{M}_{21}(\xi), \quad \tilde{M}_{ij}(\xi) = \tilde{F}_i^{\mathrm{T}} W \tilde{F}_j.$$

A more general discussion, for T-optimum designs, is in Atkinson and Fedorov (1975). We now consider DT-optimality.

When the two models differ by a single term, $\tilde{\theta}_1$ is scalar and

$$\log\{\Delta_1(\xi)\} = \log[\tilde{\theta}_1\{\tilde{M}_{11}(\xi) - \tilde{M}_{12}(\xi)M_{22}^{-1}(\xi)\tilde{M}_{21}(\xi)\}\tilde{\theta}_1]. \tag{15}$$

But, from a standard result in the theory of D_s-optimum designs, for example Atkinson and Donev (1992, p. 109),

$$|\tilde{M}_{11}(\xi) - \tilde{M}_{12}(\xi)M_{22}^{-1}(\xi)\tilde{M}_{21}(\xi)| = \frac{|M(\xi)|}{|M_{22}(\xi)|},$$

where $M(\xi) = F^{\mathrm{T}}WF$. Then (15) becomes

$$\log\{\Delta_1(\xi)\} = 2\log\tilde{\theta}_1 + \log|M(\xi)| - \log|M_{22}(\xi)|$$

and the criterion for DT-optimality (8) is

$$\Phi_1^{(DT)}(\xi) = (1-\kappa)\log|M(\xi)| - (1-\kappa)\log|M_{22}(\xi)| + \kappa\log|M_1(\xi)|, \tag{16}$$

a form of generalized D-optimality involving three information matrices, those of the two component models and that of the combined model with duplicated terms eliminated. The criterion does not depend on the value of $\tilde{\theta}_1$.

There is an appreciable literature on the special case of (16) in which it is desired both to estimate the $m+1$ parameters of the single-factor polynomial of order m and to test whether the term in x^m is necessary. The D_s-optimum design for the parameter β_m is often called D_1 optimum. Since the second model is a special case of the first, $F = F_1$ in (13) and the criterion (16) can be written

$$\Phi_m^{(DT)}(\xi) = \log|M_m(\xi)| - (1-\kappa)\log|M_{m-1}(\xi)|,$$

where $M_m(\xi)$ is the information matrix for the polynomial of order m. When $\kappa = 0$, D_1-optimality is obtained. The problem is complicated by the desire to find a good design for all $m \in [m-k, m+k], k \geq 1$. See, for example, Dette and Franke (2001) or Zen and Tsai (2002).

8 Extensions

It has been assumed here that the response is univariate. The extension of T-optimality to correlated multivariate observations is given by Uciński and Bogacka (2005). The combination of this criterion with the multivariate D-optimality of Draper and Hunter (1966) yields DT-optimality in the multivariate case.

Acknowledgments

I am grateful to Professor Dariusz Uciński of the University of Zielona Góra, and to Dr Barbara Bogacka, Queen Mary University of London, for continuing stimulating discussions about D- and T-optimum designs.

References

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Designs.* Oxford: Oxford University Press.

Atkinson, A. C. and V. V. Fedorov (1975). The design of experiments for discriminating between two rival models. *Biometrika 62*, 57–70.

Biswas, A. and P. Chaudhuri (2002). An efficient design for model discrimination and parameter estimation. *Biometrika 89*, 709–718.

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika 46*, 77–90.

Dette, H. and T. Franke (2001). Robust designs for polynomial regression by maximizing a minimum of D- and D_1-efficiencies. *Annals of Statistics 29*, 1024–1049.

Draper, N. R. and W. G. Hunter (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika 53*, 596 – 599.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* New York: Academic Press.

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments.* Lecture Notes in Statistics 125. New York: Springer.

Hill, W. J., W. G. Hunter, and W. D. Wichern (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics 10*, 145–160.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics 12*, 363–366.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* New York: Wiley.

Sibson, R. (1974). D_A-optimality and duality. In J. Gani, K. Sarkadi, and I. Vincze (Eds.), *Progress in Statistics, Vol.2 – Proc. 9th European Meeting of Statisticians, Budapest.* Amsterdam: North-Holland.

Uciński, D. and B. Bogacka (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society, Series B 67*, 3–18.

Waterhouse, T. H., J. A. Eccleston, and S. B. Duffull (2004). On optimal design for discrimination and estimation. In J. Antoch (Ed.), *COMPSTAT 2004: Proceedings in Computational Statistics*, pp. 1963–1970. Heidelberg: Physica-Verlag.

Zen, M.-M. and M.-H. Tsai (2002). Some criterion-robust optimal designs for the dual problem of model discrimination and parameter estimation. *Sankhyā, Series B 64*, 322–338.