International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 5 Issue 2 || February. 2017 || PP-71-77

Hellinger Optimal Criterion and $\mathcal{H}P_A$- Optimum Designs for Model Discrimination and Probability-Based Optimality

W. A. Hassanein[1], N. M. Kilany[2]
[1]*Faculty of Science, Tanta University*
[2]*Faculty of Science, Menoufia University*

ABSTRACT: *Kullback-Leibler (KL) optimality criterion has been considered in the literature for model discrimination. However, Hellinger distance has many advantages rather than KL-distance. For that reason, in this paper a new criterion based on the Hellinger distance named by Hellinger ($\mathcal{H}$) -optimality criterion is proposed to discriminate between two rival models. An equivalence theorem is proved for this criterion. Furthermore, a new compound criterion is constructed that possess both discrimination and a high probability of desired outcome properties. Discrimination between binary and Logistic GLM are suggested based on the new criteria.*

KEYWORDS: *KL-optimality; P-optimality; Compound criteria; Equivalence theorem.*

I. INTRODUCTION

The key importance in various theoretical and applied statistical inference and data processing problems is the distance (divergence) measures. They are mainly namely the $f-$ divergences and the Bergman divergences. f – divergences between probability densities are defined as:

$$I_f(p,q) = \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx$$

with f a convex function satisfying $f(1)=0, f'(1)=0, f''(1)=1$. Some of the well-known measures of f – divergences are; Kullback-Leibler divergence, Hellinger distance, $\chi 2$-divergence, Csiszár α- divergence, and Kolmogrov total variation distance.

Hellinger distance (also called Bhattacharyya distance), since it was first defined in its modern version in Bhattacharyya [4], is used to measure the similarity between two points of a parametric family. Under certain regularity conditions, its limit behavior as the difference in the parameter values goes down to 0, is closely related to Fisher information. Hellinger distance can also be used to study information properties of a parametric set in non-regular situations (e.g., when Fisher information does not exist). It promises certain advantages relative to such alternative information measures as Kullback-Leibler divergence.

Kullbach-Leibler -distance plays a major role in information theory and finds many natural applications in Bayesian parametric estimation. However, neither Kullbach-Leibler nor 2 chi-square distance measures are symmetric *Shemyakin* [8]. Hellinger metric is symmetric, non-negative and it satisfies the triangular inequality. Extra properties of Hellinger distance were reviewed in several studies, e.g., *Gibbs and Su* [5]. The advantages of Hellinger distance rather than Kullbach-Leibler -distance motivate us to propose a new optimality criterion based on Hellinger distance and unite it to form a compound criterion to achieve more provided properties.

López-Fidalgo [6] was introduced an optimal experiment criterion for discriminating between non-normal models namely KL-optimality. It is mainly based on Kullback-Leibler (KL) distance. Most of the proposals assume the normal distribution for the response and provide optimality criteria for discriminating between regression models. *Tommasi et. al.* [10] proposed a max-min approach for discriminating among competing statistical models (probability distribution families). However, designs that are optimal for model discrimination may be inadequate for parameter estimation. Hence, some compound criteria are found to yield designs that offer efficient parameter estimation and model discrimination for example, DT by Atkinson [2], DKL-optimality criterion by *Tommasi* [9], CDT by *Abd El-Monsef and Seyam* [1].

McGree and Eccleeston [7] proposed a criterion that maximizing a probability of a desired outcome named by Probability-based optimality (P-optimality) and a compound criterion that unite the D-optimality and P-optimality called by DP-optimality is also studied for generalized linear models. P-optimality is different from the criterion proposed by *Verdinelli and Kadane* [11] for a Bayesian optimal design for linear models, which attempted to maximize the information and outcome. Their criterion was motivated by impracticality of running an experiment that observes new successes, despite the ability to estimate model parameters.

The paper is organized as follows: The optimum design background is introduced in Section 2. Probability-based design criteria is presented in Section 3. A new criterion namely $\mathcal{H}$-optimality is proposed in Section 4, the properties of this criterion are discussed, and a general equivalence theorem is derived. In Section 5, new compound design, $\mathcal{HPA}$ -optimum design, is derived. An equivalence theorem for the new compound criterion is proved, which is the basis for the numerical construction and checking of optimum designs. Hellinger Discrimination between binary models with high optimum probability of success is defined in Section 6.

II. OPTIMUM DESIGN PRELIMINARIES

Consider the generalized linear models GLMs

$$E(Y) = \mu = \eta = g^{-1}(\mathrm{X}\beta)$$

which is defined by the distribution of the response, Y, a linear predictor η and two functions:

- A link function $g(.)$ that describes how the mean, $E(Y_i) = \mu_i$ depends on the linear predictor $g(\mu_i) = Y_i$.
- A variance function that describes how the variance, $Var(Y_i)$ depends on the mean
$$Var(Y_i) = \phi(V(\mu))$$

where the dispersion parameter ϕ is a constant.

In GLMs, the errors or noise ϵ_i have relaxed assumptions where it may or may not have normal distribution. GLMs are commonly used to model binary or count data. Some common link functions are used such that the identity, logit, log and probit link to induce the traditional linear regression, logistic regression, Poisson regression models.

A design ξ defines, for $i = 1, \dots, n$, the vector of experimental conditions $x_i \in \chi$ related to y_i, where χ is a compact experimental domain and the experimental weights w_i corresponding to each x_i, where $\sum_{i=1}^{n} w_i = 1$.The design space can be then expressed as $\delta = \{\xi_i \in X^n \times [0,1]^n : \sum_{i=1}^{n} w_i = 1\}$. Such designs are called approximate or continuous designs.

III. PROBABILITY-BASED OPTIMALITY: P_A-OPTIMALITY CRITERION

Every so often, experimenters wish to increase or maximize the probability of an outcome. To this aim, *McGree and Eccleston* [7] have offered a P-optimality criterion, which is defined as a criterion that maximizes a function of the probability of observing a particular outcome. The general form of P-optimality considered for the logistic GLM is the maximization of the function Φ_P defined as

$$\Phi_P(\xi) = f(\pi_i(\boldsymbol{\theta}, \xi_i)), \qquad \text{for } i = 1,2, \dots \dots, n$$

where, f $(\cdot)$ is some function of the probability of success $\pi_i(\boldsymbol{\theta}, \xi_i)$, for i=1,…,n.

Specifically, *McGree and Eccleston* [7] suggested two forms of P-optimality; one of them concerns with maximizes the minimum probability of success for a given design ξ, and the other maximizes the average probability of success of a given design. The criterion considered here is the maximization of a weighted sum of the probabilities of success. The form of this criterion is

$$\Phi_{P_A}(\xi) = \sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i, \quad \text{for } i = 1,2, \dots \dots, n \qquad (1)$$

where, $\pi_i(\boldsymbol{\theta}, \xi_i)$ is the i-th probability of success given by ξ_i and w_i is the experimental effort relating to the i-th support point. In this criterion, design weights have been included and will play a role in maximizing the probabilities.

A P_A-optimal design satisfies the following equivalence theorem, proved by *McGree and Eccleston* [7]

Theorem 1. For a P_A-optimal design, $\xi^*_{P_A}$, the following three conditions are equivalent.

(i) The design $\xi^*_{P_A}$ satisfies the inequality

$$\psi_{P_A}(x, \xi^*_{P_A}) \leq 0,$$

where,

$$\psi_{P_A}(x, \xi) = \frac{\Phi_{P_A}(x) - \Phi_{P_A}(\xi)}{\Phi_{P_A}(\xi)}$$

is the directional derivative of $\Phi_{P_A}(\xi)$ in the direction of $\delta_{\xi_x} = \xi_x - \xi$

(ii) The upper bound of $\psi_{P_A}(x, \xi^*_{P_A})$ is attained at the points of the optimum design.

(iii) For any non-optimum design ξ, that is the design for which;

$$\Phi_{P_A}(\xi) < \Phi_{P_A}(\xi^*_{P_A}),\ sup_{x\in\chi}\psi_{P_A}(x, \xi) > 0.$$

The P_A - efficiency of a design ξ relative to the optimum design $\xi^*_{P_A}$ is

$$Eff_{P_A}(\xi) = \frac{\sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i}{\sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi^*_{P_A}) w_i}$$

This efficiency is a pure number in (0, 1) which measures the goodness of a design ξ.

IV. HELLINGER OPTIMUM DESIGNS: $\mathcal{H}$-OPTIMALITY CRITERION

Hellinger distance measure is the corner stone for the new optimality criterion proposed in this paper to discriminate between two rival probability densities. Suppose that a parametric family of probability measures $\{F_\theta, \theta \in \Theta\}$ is defined on a measurable space $(\chi, \mathcal{B})$ hence, all measures from the family are absolutely continuous with respect to some σ-finite measure on $\mathcal{B}$. Let y be an observable random variable and let $f_1(y, x, \theta_1)$ and $f_2(y, x, \theta_2)$ be two rival probability density functions of y which depend on an experimental condition $x \in \chi$ and on a vector of unknown parameters, $\theta_i \in \Theta_i, i = 1,2$.

To discriminate between $f_1(y, x, \theta_1)$ and$f_2(y, x, \theta_2)$, the $\mathcal{H}$-optimality criterion will defined. If $f_1(y, x, \theta_1)$ is assumed to be completely known "true" model, then the $\mathcal{H}$-optimality criterion function $H_{21}(\xi)$ is

$$H_{21}(\xi) = \min_{\theta_2\in\Theta_2} \int_\chi \mathcal{H}(f_1, f_2, x, \theta_2)\, \xi(dx) \tag{2}$$

where,

$$\mathcal{H}(f_1, f_2, x, \theta_2) = [\int_\chi \{\sqrt{f_1(y, x; \theta_1)} - \sqrt{f_2(y, x; \theta_2)}\}^2 dy]^{\frac{1}{2}}, \quad x \in \chi \tag{3}$$

is Hellinger distance between $f_1(y, x, \theta_1)$ and the alternative model $f_2(y, x, \theta_2)$. A design denoted by $\xi^*_{H_{21}}$, which maximizes $H_{21}(\xi)$ will be called $\mathcal{H}$-optimum design. A design for which the optimization problem

$$\Omega_2(\xi) = \{\hat{\theta}_2 : \hat{\theta}_2(\xi) = arg \min_{\theta_2 \in \Theta_2} \int_{\chi} \mathcal{H}(f_1, f_2, x, \theta_2)\, \xi(dx)\} \tag{4}$$

has a unique solution, is a regular design otherwise is called a singular design.

Assuming that the optimal design is regular, Let ξ and $\tilde{\xi}$ be any two designs; then the directional derivative of H_{21} at ξ in the directional of $\delta_{\tilde{\xi}} = \tilde{\xi} - \xi$ is defined as

$$\partial H_{21}(\xi, \tilde{\xi}) = \lim_{\lambda \to 0^+} [\frac{H_{21}\{(1-\lambda)\xi + \lambda\tilde{\xi}\} - H_{21}(\xi)}{\lambda}] \tag{5}$$

Suppose that ξ is regular, then

$$\partial H_{21}(\xi, \tilde{\xi}) = \int_{\chi} \psi_{21}(x, \xi)\tilde{\xi}(dx)$$

where, $\psi_{21}(x, \xi)$ is the directional derivative of H_{21}at ξ in the direction of $\delta_{\xi_x} = \xi_x - \xi$ and ξ_x is the design which puts the whole mass at point x.

Theorem 2. Suppose that $\xi^*$is regular $\mathcal{H}$-optimum design.

(i) The design $\xi^*$is $\mathcal{H}$-optimum design if and only if $\psi_{21}(x, \xi^*) \leq 0, x \in \chi$, where, $\psi_{21}(x, \xi) = \mathcal{H}(f_1, f_2, x, \hat{\theta}_2) - \int_{\chi} \mathcal{H}(f_1, f_2, x, \hat{\theta}_2)\xi(dx)$ is the directional derivative of $H_{21}(\xi)$ in the direction of $\delta_{\xi_x} = \xi_x - \xi$ and $\hat{\theta}$ is the unique solution of the optimization problem (4).

(ii) The function $\psi_{21}(x, \xi^*)$ achieves its maximum value at the points of the optimal design support.

Proof.

Consider equations (2) and (5), let

$$h(\xi, \theta_2) = \int_{\chi} [\int_{\chi} \{\sqrt{f_1(y, x; \theta_1)} - \sqrt{f_2(y, x; \theta_2)}\}^2 dy]^{\frac{1}{2}} \xi(dx)$$

Thus, $H_{21}(\xi) = \min_{\theta_2 \in \Theta_2} \{h(\xi, \theta_2)\}$. Following *Uciński and Bagacka* [12], we obtain

$$\partial H_{21}(\xi, \tilde{\xi}) = \min_{\theta_2 \in \Theta_2(\xi)} \{\partial h(\xi, \theta_2, , \tilde{\xi})\}.$$

If $\Theta_2(\xi) = \{\hat{\theta}_2\}$, then $\partial H_{21}(\xi, \tilde{\xi}) = \partial h(\xi, \theta_2, , \tilde{\xi})$.

Let

$$r(x) = [\int_{\chi} \{\sqrt{f_1(y, x; \theta_1)} - \sqrt{f_2(y, x; \hat{\theta}_2)}\}^2 dy]^{\frac{1}{2}},$$

$$R(\xi) = g(\xi, \hat{\theta}_2) = \int_{\chi} r(x)\xi(dx)$$

where, ξ is any design. For any other design $\bar{\xi}$,

$$\partial R(\xi, \bar{\xi}) = \lim_{\lambda \to 0^+} (\frac{1}{\lambda}[\int_{\chi} r(x) \left\{(1-\lambda)\xi + \lambda\bar{\xi}\right\} - \int_{\chi} r(x)\xi(dx)\,])$$

$$= \int_{\chi} r(x)\bar{\xi}(dx) - \int_{\chi} r(x)\xi(dx)$$

Then,

$$\partial R(\xi, \bar{\xi}) = \partial H_{21}(\xi, \bar{\xi}) = \int_{\chi} \psi_{21}(x, \xi)\bar{\xi}(dx) \tag{6}$$

where,

$$\psi_{21}(x, \xi) = \mathcal{H}(f_1, f_2, x, \hat{\theta}_2) - \int_{\chi} \mathcal{H}(f_1, f_2, x, \hat{\theta}_2)\xi(dx) = \partial H_{21}(\xi, \xi_x)$$

Thus, $\psi_{21}(x,\xi)$ is the directional derivative of $H_{21}(\xi)$ in the direction of $\delta_{\xi_x} = \xi_x - \xi$.

Since, $H_{21}(\xi)$ is concave function of ξ, then the optimality of ξ^*exists if and only if

$\partial H_{21}(\xi,\xi^*) \leq 0$ for any design ξ. From equation (6), a necessary and sufficient condition for the optimality of ξ^* is that ξ^* satisfies the inequality

$$\max_{\xi}\{\int_{\chi}\psi_{21}(x,\xi^*)\xi(dx)\} \leq 0$$

Accordingly, $\psi_{21}(x,\xi^*) \leq 0\,, \quad \forall x \in \chi$

To prove that $\psi_{21}(x,\xi^*)$ attains its maximum value of 0 at all points of ξ^*, suppose the reverse, i.e suppose that there is a set $\Upsilon \subset supp(\xi^*)$ and a scalar a such that

$$\int_{\Upsilon}\psi_{21}(x,\xi^*)\xi^*(dx) \leq a < 0 \text{ and } \psi_{21}(x,\xi^*) = 0, \quad \forall x \in \chi - \Upsilon.$$

Then,

$$\int_{\chi}\psi_{21}(x,\xi^*)\xi^*(dx) \leq a < 0$$

This is a contradict the requirement

$$\int_{\chi}\psi_{21}(x,\xi^*)\xi^*(dx) = 0$$

where ξ^* is the optimal design that is obtained from (6) for $\xi^* = \bar{\xi}$.

V. $\mathcal{H}P_A$- OPTIMALITY CRITERION

In general, P_A-optimal designs have little ability or efficiency for discriminate between any two statistical models. However, a combined criterion involving both $\mathcal{H}$- and P_A-optimality should yield designs that offer true model and a high probability of observing a particular outcome. A method for forming this compound criterion, similar to that of *Atkinson* [3] for DT-optimality.

From the definition of the compound design criterion which is a weighted geometric mean of efficiencies design ξ, we defined a new compound criteria which combines $\mathcal{H}$- and P_A- optimality, weighted by a pre-defined mixing constant $0 \leq \alpha \leq 1$. This criterion will be called $\mathcal{H}P_A$- optimality. That is

$$\begin{aligned}\Phi_{\mathcal{H}P_A}(\xi) &= [Eff_{\mathcal{H}}(\xi)]^{\alpha}\left[Eff_{P_A}(\xi)\right]^{1-\alpha} \\ &= \left[\frac{\mathrm{H}_{21}(\xi)}{\mathrm{H}_{21}(\xi^*_{21})}\right]^{\alpha}\left[\frac{\sum_{i=1}^{n}\pi_i(\boldsymbol{\theta},\xi_i)w_i}{\sum_{i=1}^{n}\pi_i\left(\boldsymbol{\theta},\xi^*_{P_A}\right)w_i}\right]^{1-\alpha}\end{aligned} \tag{7}$$

when $\alpha = 0$ we obtain P_A-optimality and when $\alpha = 1$ we obtain $\mathcal{H}$ -optimality. Taking the logarithm of (7) yields,

$$log\,\Phi_{\mathcal{H}P_A}(\xi) = \alpha log\mathrm{H}_{21}(\xi) + (1-\alpha)log\sum_{i=1}^{n}\pi_i(\boldsymbol{\theta},\xi_i)w_i \tag{8}$$

Because the terms involving $\xi^*_{\mathcal{H}}$ and $\xi^*_{P_A}$ are constants, when a maximum is found over ξ, they can be ignored. A $\mathcal{H}P_A$-optimum design, $\xi^*_{\mathcal{H}P_A}$, maximizes $\log\Phi_{\mathcal{H}P_A}(\xi)$.

The equivalence theorem may be stated as follows,

Theorem 3. For $\mathcal{H}P_A$-optimal design, $\xi^*_{\mathcal{H}P_A}$, the following three conditions are equivalent.

(i) A necessary and sufficient condition for a design $\xi^*_{\mathcal{H}P_A}$ to be $\mathcal{H}P_A$-optimum is fulfillment of the inequality

$$\psi_{\mathcal{H}P_A}(x, \xi^*_{\mathcal{H}P_A}) \leq 0, \qquad x \in \chi$$

where

$$\psi_{\mathcal{H}P_A}(x, \xi) = \alpha \frac{\psi_{21}(x, \xi)}{\mathrm{H}_{21}(\xi)} + (1 - \alpha) \frac{\psi_{P_A}(x, \xi)}{\sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i}$$

is the directional derivative of the criterion function (8) at ξ in the direction of $\delta_{\xi_x} = \xi_x - \xi$.

(ii) The upper bound of $\psi_{\mathcal{H}P_A}(x, \xi^*_{\mathcal{H}P_A})$ is attained at the points of the optimum design.

(iii) For any non optimum design ξ, that is a design for which $\Phi_{\mathcal{H}P_A}(\xi) < \Phi_{\mathcal{H}P_A}(\xi^*_{\mathcal{H}P_A})$, $sup_{x \in \chi} \psi_{\mathcal{H}P_A}(x, \xi^*_{\mathcal{H}P_A}) > 0$

Proof. Since, $0 \leq \alpha \leq 1$, $log\, \Phi_{\mathcal{H}P_A}(\xi)$ given by (8) is a convex combination of logarithm of two design criteria. The first of which is $log \mathrm{H}_{21}(\xi)$ and the second criterion is $log \sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i$. As $\mathrm{H}_{21}(\xi) \geq 0$ and $\sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i \geq 0$, $log \mathrm{H}_{21}(\xi)$ and $log \sum_{i=1}^{n} \pi_i(\boldsymbol{\theta}, \xi_i) w_i$ are concave functions of concave design criteria. Consequently, the criterion $log\, \Phi_{\mathcal{H}P_A}(\xi)$ is a convex combination of two concave functions. Therefore $log\, \Phi_{\mathcal{H}P_A}(\xi)$ is concave and the $\mathcal{H}P_A$- criterion satisfies the conditions of convex optimum design theory and an equivalence theorem applies similar to Theorems 1 and 2. Furthermore, $\psi_{\mathcal{H}P_A}$ is the linear combination of the directional derivatives given by theorems 1 and 2. That is, the first term of $\psi_{\mathcal{H}P_A}$ is that from $\mathcal{H}$-otimality and the second term is from P_A-otimality. Thus, the theorem has been proved.

VI. HELLINGER DISCRIMINATION FOR BINARY MODELS WITH P_A -OPTIMUM DESIGN

Assume that the response variable y has a binomial distribution satisfies the P_A -optimum design probability of success $P(Y = 1) = \sum_{i=1}^{n} \pi_{ij}(x, \theta_j, \xi_{ij}) w_i$, $j = 1,2$ where θ_j are the parameters for the two possible models. To obtain the optimal design for discrimination between binary models which satisfy P_A-criterion, maximization of the following criterion using Hellinger distance is considered;

$$\Delta_2(\xi, \theta_1) = inf_{\theta_2 \in \Omega_2} \left(\left[\int_{\chi} \left\{ \sqrt{max \sum_{i=1}^{n} \pi_{i1}(x, \theta_1, \xi_{i1}) w_i} - \sqrt{max \sum_{i=1}^{n} \pi_{i2}(x, \theta_2, \xi_{i2}) w_i} \right\}^2 + \left\{ \sqrt{1 - max \sum_{i=1}^{n} \pi_{i1}(x, \theta_1, \xi_{i1}) w_i} - \sqrt{1 - max \sum_{i=1}^{n} \pi_{i2}(x, \theta_2, \xi_{i2}) w_i} \right\}^2 \xi(dx) \right]^{\frac{1}{2}} \right) \qquad (9)$$

The proposed optimality criterion (9) can be appropriate for GLMs according to the applicability of P_A-optimality for these models.

REFERENCES

[1]. Abd El-Monsef, M.M.E. and Seyam, M.M (2011). "*CDT-Optimum Designs for Model Discrimination, Parameter Estimation and Estimation of a Parametric Function*". Journal of Statistical Planning and Inference, 141, 639-643.

[2]. Atkinson, A. C. (2008). "*DT–Optimum Designs for Model Discrimination and Parameter Estimation*". Journal of Statistical Planning and Inference, 138, 56–64.

[3]. Atkinson, A., Donev, A. and Tobias, R. (2007). *"Optimum experimental designs, with SAS"*. Oxford university press, New York.

[4]. Bhattacharyya, A. (1943). "*On a Measure of Divergence between Statistical Populations Defined by their Probability Distributions*". Bulletin of Calcutta Mathematical Society, 35, 99-109.

[5]. Gibbs, A. L. and Su, E. W. (2002). "*On Choosing and Bounding Probability Metrics*". International Statistical Review, 70(3), 419-435.

[6]. López-Fidalgo, J., Tommasi, C., Trandafir, P. (2007). "*An Optimal Experimental Design Criterion for Discriminating Between Non-Normal Models*". Journal of the Royal Statistical Society, B, 69, 231-242.
[7]. McGree, J. M. and Eccleston, J. A. (2008). "*Probability-Based Optimal Design*". Australian and New Zealand Journal of Statistics, 50 (1), 13- 28.
[8]. Shemyakin, A. (2014). "*Hellinger Distance and Non-Informative Priors*". Bayesian Analysis, TPA, 1-18.
[9]. Tommasi, C. (2009). "*Optimal Designs for both Model Discrimination and Parameter Estimation*". Journal of Statistical Planning and Inference, 139, 4123-4132.
[10]. Tommasi, C., Martín- Martín, R., López-Fidalgo, J. (2015). "*Max-min Optimal Discriminating Designs for Several Statistical Models*". Statistical Computing. DOI 10.1007/s11222-015-9606-1.
[11]. Verdinelli, I. and Kadane, J. . B. (1992). "*Bayesian Designs for Maximizing Information and Output*". Journal of American Statistical Association, 87, 510-515.
[12]. Uciński, D. and Bagacka, B. (2005). "*T-optimum Designs for Discrimination between Multiresponse Dynamic Models*." Journal of Royal Statistical Society, B, 67, 3-18.