Optimal designs for discriminating between functional linear models

Verity Fisher and Dave Woods

Southampton Statistical Sciences Research Institute, University of Southampton, United Kingdom

{V.Fisher,D.Woods}@southampton.ac.uk

Introduction

Improvements in online measuring have facilitated an increase in the number of observations that can be taken on each experimental unit in industrial and scientific experiments [1]. It can often be assumed that the data from each run are generated by a smooth underlying function, and here we are interested in how changes to the levels of the controllable factors influence these functions.

Our motivation for this work is an experiment to study wear in a pin and disc assembly (Figure 1) for a given lubricant performed by the National Centre for Advanced Tribology Southampton. The effects of four factors were investigated in 20 runs, each defined by a different treatment, or combination of values of the factors.

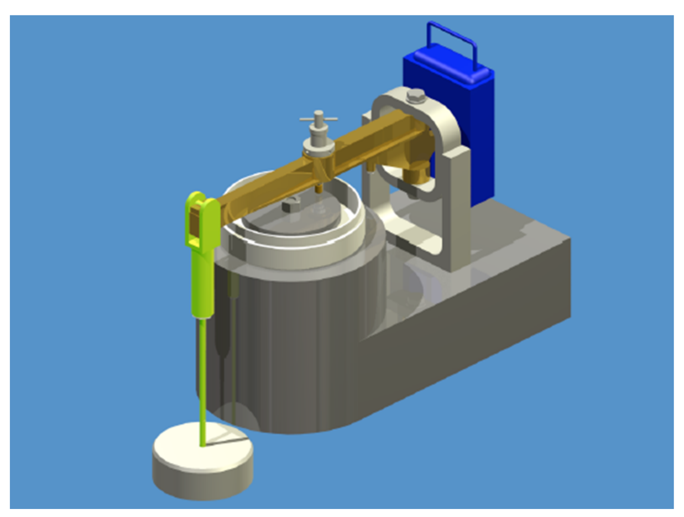

Figure 1: *Schematic of the pin and disc equipment.*

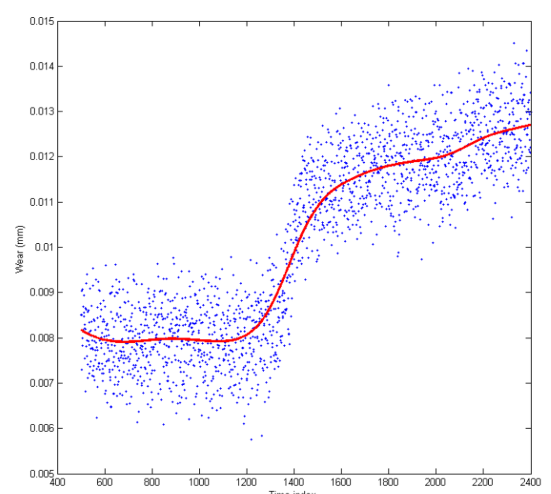

Figure 2: *Example of smoothed wear data from one run of the experiment.*

The response was the total wear of the pin and disc, measured by a linear variable displacement transformer. Figure 2 shows an example of the data produced from this experiment, which can be assumed to be a realisation of a smooth function. To understand the effects of the four treatment factors on the observed functions, we can use the data to choose between contending functional linear models. Hence, we desire to find T-optimal designs [2] which provide the most information for discriminating between functional linear models.

Functional linear models

We assume the following linear model for the functional responses from an N-run experiment:

$$\text{M1}: \qquad \boldsymbol{Y}(t) = \mathbf{X}_1\boldsymbol{\beta}_1(t) + \boldsymbol{\varepsilon}(t)\,,$$

with, for $t \in \mathcal{T} \subset \mathbb{R}$, $\boldsymbol{Y}(t) = (Y_1(t), \dots, Y_N(t))^{\mathrm{T}}$, $\mathbf{X}$ an $N \times p_1$ model matrix, $\boldsymbol{\beta}_1(t) = (\beta_{11}(t), \dots, \beta_{1p_1}(t))^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \dots, \varepsilon_N(t))$. The errors $\varepsilon_j(t)$ and $\varepsilon_j(u)$ are realisations from a Gaussian stochastic process mean zero and covariance function $\gamma(t, u)$; for $i \neq j$, $\varepsilon_i(t)$ and $\varepsilon_j(u)$ are assumed independent.

That is, the observed functions $Y_j(t)$ are assumed to be linear combinations of unknown functions $\beta_{1k}(t)$ with the addition of independent error functions $\varepsilon_j(t)$ ($j = 1, \dots, N$; $k = 1, \dots, p_1$).

We assume that the aim of the experiment is to discriminate between model M1 and a rival model

$$\text{M2}: \qquad \boldsymbol{Y}(t) = \mathbf{X}_2\boldsymbol{\beta}_2(t) + \boldsymbol{\varepsilon}(t)\,,$$

with $\mathbf{X}_2$ an alternative model matrix with corresponding vector of unknown functions $\boldsymbol{\beta}_2(t) = (\beta_{21}(t), \dots, \beta_{2p_2}(t))^{\mathrm{T}}$. To discriminate between these models, Faraway [3] suggested the test statistic

$$T = \int_t \left[\hat{\boldsymbol{Y}}_2(t) - \hat{\boldsymbol{Y}}_1(t)\right]^{\mathrm{T}} \left[\hat{\boldsymbol{Y}}_2(t) - \hat{\boldsymbol{Y}}_1(t)\right] \,\mathrm{d}t\,, \qquad (1)$$

where $\hat{\boldsymbol{Y}}_i = (\hat{Y}_1(t), \dots, \hat{Y}_N(t))^{\mathrm{T}}$ are the fitted functions from model Mi.

T-optimality

We find approximate optimal designs in f factors which are represented by a measure ξ on the design region $\mathcal{X} = [-1, 1]^f$:

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{array} \right\}, \qquad (2)$$

where $x_j \in \mathcal{X}$ are the support points with associated weights $0 < w_j \leq 1$; $\sum_{j=1}^n w_j = 1$.

Using data collected using design (2), we obtain fitted functions

$$\hat{\boldsymbol{Y}}_1(t) = \mathbf{X}_1\left(\mathbf{X}_1^{\mathrm{T}}\mathbf{W}\mathbf{X}_1\right)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{W}\boldsymbol{Y}(t) = \mathbf{H}\boldsymbol{Y}(t)\,,$$

where $\mathbf{W} = \mathrm{diag}(w_1, \dots, w_n)$ and $\mathbf{X}_i$ is now defined for the n support points.

If we assume we observe data from M2 without error, from (1) we find a T-optimal design $\xi^\star$ by maxmizing

$$\begin{aligned} \Phi(\xi) &= \int_t \left[\mathbf{X}_2\boldsymbol{\beta}_2(t) - \mathbf{H}\mathbf{X}_2\boldsymbol{\beta}_2(t)\right]^{\mathrm{T}} \mathbf{W} \left[\mathbf{X}_2\boldsymbol{\beta}_2(t) - \mathbf{H}\mathbf{X}_2\boldsymbol{\beta}_2(t)\right] \,\mathrm{d}t \\ &= \int_t \boldsymbol{\beta}_2^{\mathrm{T}}(t)\mathbf{X}_2^{\mathrm{T}}\left[\mathbf{I} - \mathbf{H}\right]^{\mathrm{T}}\mathbf{W}\left[\mathbf{I} - \mathbf{H}\right]\mathbf{X}_2\boldsymbol{\beta}_2(t)\,\mathrm{d}t\,. \end{aligned} \qquad (3)$$

Lemma: Assume M1 is nested within M2, so $p_1 < p_2$, $\mathbf{X}_2 = [\mathbf{X}_1 : \mathbf{X}_{21}]$ and $\boldsymbol{\beta}_2^{\mathrm{T}}(t) = [\boldsymbol{\beta}_1^{\mathrm{T}}(t), \boldsymbol{\beta}_{21}^{\mathrm{T}}(t)]$ with $\mathbf{X}_{21}$ an $n \times (p_2 - p_1)$ model matrix and $\boldsymbol{\beta}_{21}$ an $(p_2 - p_1)$ vector of unknown functions. Then objective function (3) is given by

$$\Phi(\xi) = \int_t \boldsymbol{\beta}_{21}^{\mathrm{T}}(t)\mathbf{X}_{21}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{H}\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_{21}\boldsymbol{\beta}_{21}(t)\,\mathrm{d}t\,,$$

and hence does not depend on the parameter vector $\boldsymbol{\beta}_1(t)$ which is common to both M1 and M2.

Proof. The proof is analogous to that for the scalar regression case [4]. □

Theorem: Assume M1 is nested in M2, as in the lemma, and $p_2 = p_1 + 1$; that is, models M1 and M2 differ by only one term. Then the T-optimal design does not depend on the unknown function $\boldsymbol{\beta}_{21}(t)$.

Proof. If $p_2 - p_1 = 1$, $\mathbf{X}_{21}$ is a $n \times 1$ vector, and $\boldsymbol{\beta}_{21}(t)$ is a single function $\beta_{21}(t)$ and hence for given t is a scalar. From the lemma,

$$\begin{aligned} \Phi(\xi) &= \int_t \beta_{21}^2(t)\mathbf{X}_{21}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{H}\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_{21}\,\mathrm{d}t \\ &= \mathbf{X}_{21}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{H}\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_{21}\int_t \beta_{21}^2(t)\,\mathrm{d}t \\ &\propto \mathbf{X}_{21}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{H}\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_{21}\,, \end{aligned} \qquad (4)$$

where the constant of proportionality does not depend on ξ. Therefore, the T-optimal design that maximises (4) does not depend on the function $\beta_{21}(t)$. □

Corollary: When M1 is nested in M2 and $p_2 = p_1 + 1$, it follows directly from (4) that the same design $\xi^\star$ is T-optimal design for both the functional linear model and the scalar linear model.

Example

We construct a T-optimal design to compare the functional linear models

$$Y(t) = \beta_0(t) + \beta_1(t)x + \epsilon(t)\,, \qquad (5)$$

and

$$Y(t) = \theta_0(t) + \theta_1(t)x + \theta_2(t)x^2 + \epsilon(t)\,. \qquad (6)$$

That is, we find an optimal design to test if (5) is appropriate given data from (6). As the models differ by only one term, from the theorem we know that $\xi^\star$ will not depend on any of the unknown functions. Maximising (4) using the Nelder-Mead algorithm, we find the optimal design

$$\xi^\star = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.25 & 0.5 & 0.25 \end{array} \right\}, \qquad (7)$$

which, from the corollary, is also T-optimal for comparing first- and second-order scaler regression models.

Simulation study

We assess the power for rejecting H_0 : "model (5) is correct" using exact T-optimal designs with N runs, obtained by rounding (7), and test statistic (1). The power is approximated through a simulation study with data generated from model (6) assuming:

- the functions $Y_j(t)$ are observed at points $t_1, \dots, t_m \in [-1, 1]$, $j = 1, \dots, N$;
- $\theta_k(t) = \alpha_{k0} + \alpha_{k1}t + \alpha_{k2}t^2$, $k = 0, 1, 2$;
- Cov $(\varepsilon_g(t_u), \varepsilon_h(t_v)) = \sigma_a^2\rho^{|u-v|} + \sigma_b^2$ for $g = h$ and $0 < \rho < 1$, and 0 otherwise.

For each of $S = 1000$ generated data sets (with $\sigma_a^2 = 0.1$, $\sigma_b^2 = 2$, $\rho = **$), we approximate (1) as

$$T \approx \sum_{j=1}^{N} \left[\hat{\boldsymbol{Y}}_2(t) - \hat{\boldsymbol{Y}}_1(t)\right]^{\mathrm{T}} \left[\hat{\boldsymbol{Y}}_2(t) - \hat{\boldsymbol{Y}}_1(t)\right]\,,$$

and calculate $\mathfrak{F} = (N - p_2)T/\text{RSS}$, where RSS is the residual sum of squares from model (6). Under H_0, $\mathfrak{F}$ follows an F distribution [5] with $\hat{\lambda}$ and $\hat{\lambda}(n - p_2)$ degrees of freedom (DoF) where $\hat{\lambda} = \mathrm{tr}(\hat{\Sigma})^2/\mathrm{tr}(\hat{\Sigma}^2)$ is the DoF adjustment factor and $\hat{\Sigma}$ is the empirical variance-covariance matrix for $\boldsymbol{Y}$, estimated from fitting (6). We approximate the power as the proportion of simulations for which H_0 is rejected.

Figure 3 displays the results of this study. Obviously, as the number of runs increases, the power of the test increases for all values of α_{22}. Usually, the difference in power for different N is smallest for larger values of α_{2j}. This is intuitive, as larger values of α_{2j} lead to larger θ_2 and hence an easier discrimination problem. We also see that larger values of α_{20} lead to higher power for all values of α_{22}; increasing α_{21} has much less effect.

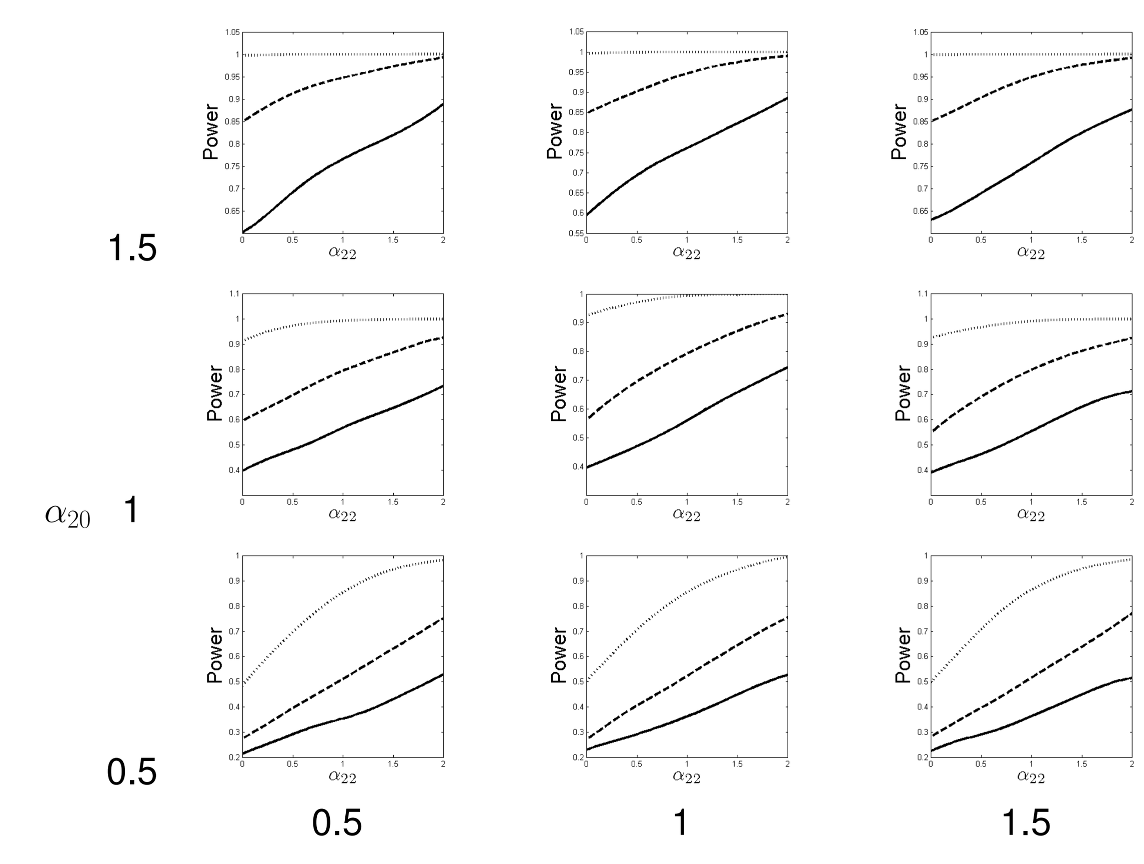

Figure 3: *Example 1: Simulated power from 1000 simulations under different parameter values, α_{20}, α_{21} and α_{22}, and different numbers, N, of runs. The three curves represent $(-)$ $N = 12$, $(--)$, $N = 24$, $(\cdots)$ $N = 72$*

References

1. Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis.* New York: Springer.
2. Atkinson, A. C. and Fedorov, V. V. (1975). Biometrika, 62, 57-70.
3. Faraway, J.J. (1997). Technometrics, 39, 254-261.
4. Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS.* Oxford: Oxford University Press. p.360
5. Shen, Q. and Faraway, J. (2004). Statistica Sinica, 14, 1239-1257.