

D-optimal Design for Polynomial Regression: Choice of Degree and Robustness

In this paper we show that for the D-optimal design, departures from the design are much less important than a departure from the model. As a consequence, we propose, based on D-optimality, a rule for choosing the regression degree.

We also study different types of departures from the model to define a new class of D-optimal designs, which is robust and more efficient than the uniform one.

1. Introduction

One of the main problems in the polynomial regression concerns a choice of the regression degree. In the literature on optimal design it is usually assumed that a statistical model is known. In practice, however, the exact degree of the polynomial is not known with certainty. Moreover, the design support may not correspond to the optimal one given by an analytical solution.

Consider the common polynomial regression model with heteroscedastic error

$$E(y|x) = f^T(x)\beta, \quad V(y|x) = \sigma^2/\lambda(x),$$

where the explanatory variable x varies in I , a real compact interval,

$$f^T(x) = (1, x, x^2, \dots, x^m), \quad \beta = (\beta_0, \beta_1, \dots, \beta_m)^T$$

and $\lambda(x) > 0$ is the efficiency function of the experiment under study.

These models are widely used when the response is curvilinear or when nonlinear relationships are approximated by a polynomial over reasonably small range of the x 's.

A design of experiment is a probability measure ξ defined on the interval I . An exact design is a probability measure ξ with finite support x_1, x_2, \dots, x_n at which uncorrelated observations are taken, such that $\xi(x_i) = n_i/N$ where n_i is the number of observations taken at x_i and $N = \sum n_i$ is the total number of observations. Otherwise, ξ is said to be an approximate design.

The information matrix of an approximate design, denoted by $M(\xi)$, is defined by the equality

$$M(\xi) = \int_I \lambda(x) f^T(x) f(x) d\xi(x),$$

or, equivalently, for an exact design by

$$M(\xi) = \sum \lambda(x_i) f^T(x_i) f(x_i) \xi(x_i).$$

In that case the covariance matrix of the generalized least squares estimates of β is $(\sigma^2/N)M^{-1}(\xi)$, and the variance of the prediction at a given point x is $\sigma^2 f^T(x) M^{-1}(\xi) f(x)$. $|M^{-1}(\xi)|$ is known as the generalised variance.

An optimal design minimizes (or maximizes) Φ_m , a specific convex (or concave) function of the information matrix. $\Phi_m(M(\xi)) = [\text{trace}(M(\xi))^{-1}]^{1/m}$ defines a class of classical criteria. For $m = 1$

we have the A-optimality. The D-optimality corresponds to $m \rightarrow 0$, the criteria based on the determinant of the information matrix. For $m \rightarrow \infty$ the E-optimality is defined through the eigenvalues of the information matrix.

The purpose of this paper is, for a certain class of efficiency functions and D-optimality, to study the influence of departures from the optimal design on the values of the criteria.

Our numerical results show that the influence on the value of D-efficiency is more important when perturbations are applied to boundary points than to others.

As a consequence of a numerical study of the behavior of the standardized determinant of the information matrix with respect to the degree of the polynomial regression for classical efficiency functions, we propose the following rule for choosing the regression degree.

For homoscedastic cases and for Jacobi's type efficiency functions the regression degree should be 'the smallest reasonable'. For Laguerre efficiency functions the regression degree should be the 'the largest reasonable'. For the Hermite function there is no general rule.

We also point out that departures from the design are much less important than a departure from the model.

Finally we define a new class of designs, *higher-order D-optimal*, which is

- (1) more efficient than uniform designs
- and
- (2) as robust as uniform ones.

So, the family of higher-order D-optimal designs can be considered as a near optimal, robust family of approximate designs.

2. Review of Classical Theory

In this article we restrict our attention to feasible designs for the polynomial regression. Feasibility corresponds to the possibility of estimating the full parameter β .

A design ξ^* , satisfying the equation

$$|M(\xi^*)| = \max_{\xi \in \mathfrak{J}} |M(\xi)|,$$

is called *D-optimal*, where \mathfrak{J} is the set of all feasible designs. D-optimal designs are well known for an important class of the efficiency functions given in Theorems 1 and 2.

Theorem 1 [Fedorov (1972)]

- a) For the polynomial regression of degree n , the D-optimal design is unique, and it is concentrated at $n+1$ equally weighted points.
- b) For the following efficiency functions:

1. $\lambda(x) = 1$, $-1 \leq x \leq 1$;
2. $\lambda(x) = (1-x)^{\alpha+1} (1+x)^{\beta+1}$, $-1 < x < 1, \alpha > -1$ and $\beta > -1$;
3. $\lambda(x) = \exp(-x)$, $0 \leq x < \infty$;
4. $\lambda(x) = x^{\alpha+1} \exp(-x)$, $0 < x < \infty, \alpha > -1$;
5. $\lambda(x) = \exp(-x^2)$, $-\infty < x < \infty$,

the design points are roots of the corresponding polynomials:

1. $(1-x^2)P'_n(x)$, where $P_n(x)$ is the n^{th} Legendre polynomial;
2. $J_{n+1}^{\alpha,\beta}(x)$, where $J_{n+1}^{\alpha,\beta}(x)$ is the $(n+1)^{\text{th}}$ Jacobi polynomial with parameters α, β ;
3. $xL_n^1(x)$, where $L_n^1(x)$ is the n^{th} Laguerre polynomial with parameter 1;
4. $L_{n+1}^\alpha(x)$, where $L_{n+1}^\alpha(x)$ is the $(n+1)^{\text{th}}$ Laguerre polynomial with parameter α ;
5. $H_{n+1}(x)$, where $H_{n+1}(x)$ is the $(n+1)^{\text{th}}$ Hermite polynomial.

The generalization of the second part of Theorem 1, proposed by [Antille (1977)], was followed by numerous authors [Huang et al. (1995)], [He et al. (1996)], [Chang and Lin (1997)], [Imhoff et al. (1998)], [Dette et al. (1999)] and others, and finally resulted in Theorem 2.

Theorem 2 [Antille, Dette, Weinberg (2001)]

For the following efficiency functions:

1. $\lambda(x) = x^{-\gamma} \exp(-\delta/x)$, $0 \leq x < \infty$, $\delta > 0$ and $\gamma > 2n$;
2. $\lambda(x) = (1+x^2)^{\alpha+1} \exp(2\beta \arctan(x))$, $-\infty < x < \infty$, $\alpha \leq -n-1$ and $\beta \in \mathbb{R}$,

the design points are roots of the corresponding polynomials:

1. $B_{n+1}^{-\gamma,\delta}(x)$, where $B_{n+1}^{-\gamma,\delta}(x)$ is the generalized Bessel polynomial with parameters γ, δ ;
2. $J_{n+1}^{\alpha+i\beta, \alpha-i\beta}(ix)$, the $(n+1)^{\text{th}}$ Jacobi polynomial with parameters $\alpha + i\beta, \alpha - i\beta$.

Figures 1 and 2 contain graphical representations of examples of optimal designs for different degrees of regression and for the efficiency functions 1 and 4 ($\alpha = 5$) of Theorem 1.

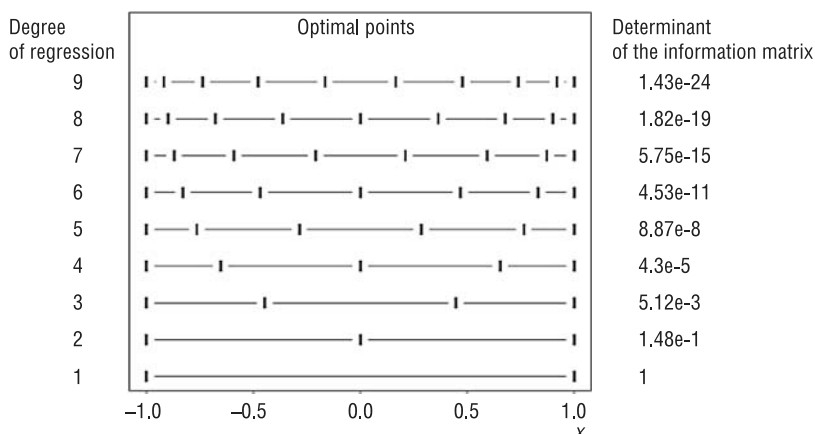


Figure 1. Optimal design for Legendre efficiency function

For more details on optimal designs we refer to the monographs [Fedorov (1972)], [Silvey (1980)], [Atkinson and Donev (1992)], and [Pukelsheim (1993)].

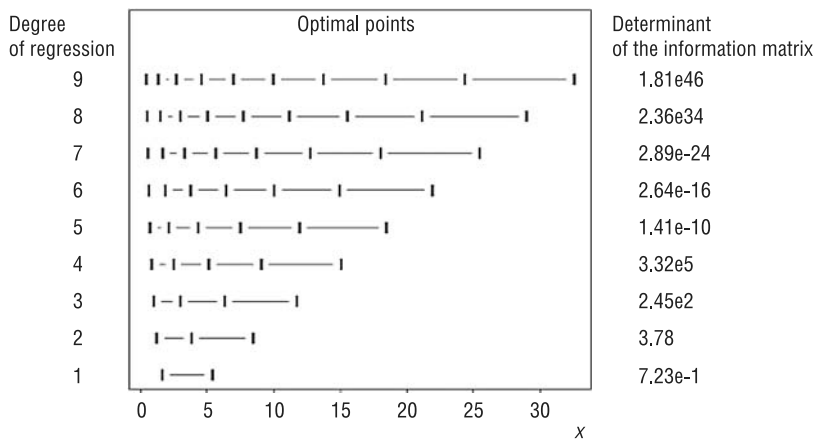


Figure 2. The optimal design for the Generalized Lagurre efficiency function ($\alpha = 5$)

3. Robustness of D-optimal Designs

3.1 Influence of Design Points

Classical theory on optimal designs for polynomial regression provides analytical solutions to optimization problems. However the design points may not correspond to the optimal ones. In this section we study the influence of different perturbations on the design points.

In order to compare designs in term of D-optimality let us define D_{eff} as the D-efficiency of a design ξ with respect to the optimal design ξ^* :

$$D_{eff} = \frac{|M(\xi)|}{|M(\xi^*)|}.$$

The standardised D-efficiency is the n^{th} root of the D-efficiency, where n is the degree of the polynomial regression.

As an example we consider the well known homoscedastic case for the regression of degree 4 with explanatory variable $x \in [-1; +1]$. Figures 3 and 4 exhibit perturbed designs and their respec-

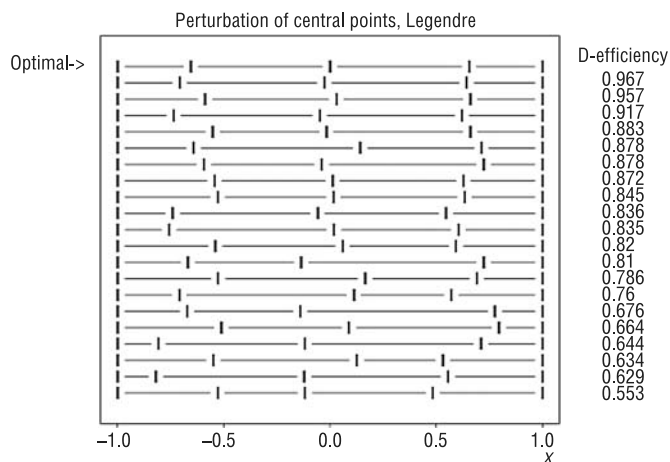


Figure 3. Perturbation of central points of the optimal design for the Legendre efficiency function

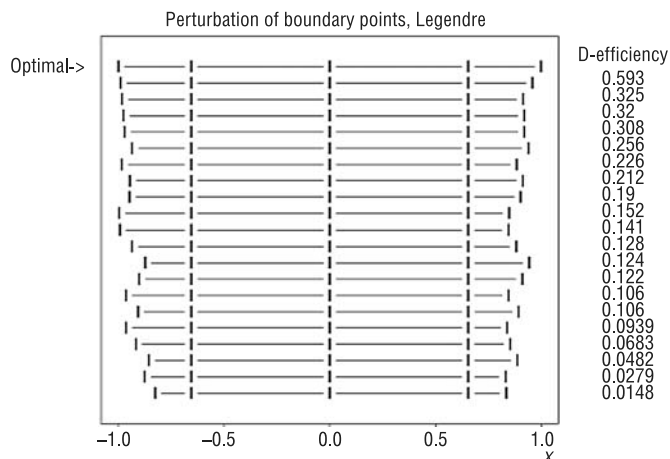


Figure 4. Perturbation of boundary points of the optimal design for the Legendre efficiency function

tive D-efficiencies. In Figure 3 perturbations are applied to central points and in Figure 4 to boundary ones.

The numerical results show that the influence on D-efficiency is more important when perturbations are applied to boundary points than to others.

Another way to illustrate this fact is presented in Figure 5.

We shift respectively x_1 towards x_2 , then x_2 towards x_3 , after that x_3 towards x_2 and finally x_2 towards x_1 and plot, in Figure 5, the D-efficiency versus the distance between two neighbouring points, where $x_1 < x_2 < x_3 < x_4 < x_5$ are the D-optimal points. Due to the symmetry of the design we do not need to consider the moves of the points x_4 and x_5 . In each case the D-optimal design corresponds to the maximal distance between two neighbouring points, with the D-efficiency equal to 1, and when the two points are close, the D-efficiency is close to 0. Looking at Figure 5 from right to left we observe that the decrease of the D-efficiency curves is more important when

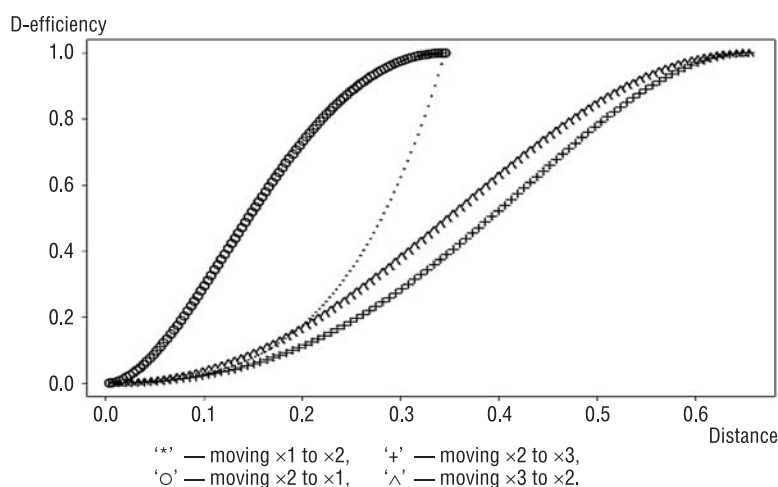


Figure 5. The D-efficiency versus (vertical axis) the distance between two neighbouring points (horizontal axis)

x_1 moves towards x_2 than when x_2 moves towards x_1 or when x_2 towards x_3 . Therefore and obviously, the conclusions are similar to those drawn in the cases of Figures 3 and 4.

3.2. Influence of the Degree of Regression

Usually the degree of a regression is supposed to be known to allow computation of optimal designs. In practice however, the exact degree of the polynomial is not known with certainty.

In this section we study the behavior of the determinant of the information matrix with respect to the degree of the polynomial regression for classical efficiency functions.

In each case we consider, for comparison,

$$|M|^{\frac{1}{m}} = \left[\left(\frac{1}{m+1} \right)^{m+1} \prod_{i=1}^{m+1} (\lambda(x_i)) \prod_{k < j} (x_k - x_j)^2 \right]^{\frac{1}{m}},$$

the standardized determinant of the information matrix. For the classical efficiency functions, D-optimal design points are roots of orthogonal polynomials. The property of these roots, which interlace for increasing degrees, implies that $\prod_{k < j} (x_k - x_j)^2$ is a decreasing function of degree if the design support is in the interval $[-1; +1]$. Hence, for efficiency functions, satisfying $|\lambda(x)| \leq 1$, $x \in [-1; +1]$, the standardised determinant of the information matrix is a decreasing function of regression degree. The opposite is true for Laguerre efficiency functions.

So for D-optimal designs generated by Legendre or Jacobi polynomials the generalised variance is an increasing function of the degree of regression. For Laguerre or generalised Laguerre the reverse is true. No general result exists for the Hermite case.

Hence, as long as D-optimality is concerned, a rule of choice of the degree can be formulated as follows:

for homoscedastic cases and for Jacobi's type efficiency functions the regression degree should be 'the smallest reasonable'; for Laguerre efficiency functions the regression degree should be the 'the largest reasonable'; for the Hermite function a case by case study has to be done.

Table 1 presents standardized determinants for D-optimal designs for various efficiency functions.

Table 1

Standardized determinants of the information matrices for D-optimal designs for degrees of regression from 1 to 9

Polynomials of Legendre, $\lambda(x) = 1$								
1.	0.385	0.172	0.081	0.0389	0.0189	0.00924	0.00454	0.00224
Polynomials of Jacobi, $\lambda(x) = (1-x)^\alpha(1+x)^\beta, \alpha = 1, \beta = 1$								
0.148	0.0716	0.035	0.0173	0.00854	0.00423	0.0021	0.00105	0.00052
Polynomials of Jacobi, $\lambda(x) = (1-x)^\alpha(1+x)^\beta, \alpha = 2, \beta = 2$								
0.0819	0.0353	0.0159	0.00739	0.00349	0.00167	0.000802	0.000388	0.000189

Polynomials of Jacobi, $\lambda(x) = (1-x)^\alpha(1+x)^\beta, \alpha = 2, \beta = 4$								
0.102	0.0326	0.0126	0.00524	0.00228	0.00102	0.000462	0.000213	0.0000996
Polynomials of Legendre, $\lambda(x) = e^{-x}$								
0.135	0.199	0.415	1.12	3.71	14.5	65.7	337	1940
Polynomials of Generalized Laguerre, $\lambda(x) = x^\alpha e^{-x}, \alpha = 1$								
0.0733	0.231	0.695	2.4	9.58	43.8	226	1300	8230
Polynomials of Generalized Laguerre, $\lambda(x) = x^\alpha e^{-x}, \alpha = 2$								
0.268	0.824	2.68	10.1	44	218	1210	7480	50800
Polynomials of Generalized Laguerre, $\lambda(x) = x^\alpha e^{-x}, \alpha = 2.5$								
0.723	1.94	6.26	24	107	546	3120	19800	138000
Polynomials of Hermite, $\lambda(x) = e^{-x^2}$								
0.184	0.158	0.161	0.187	0.241	0.337	0.506	0.811	1.37
Polynomials of Hermite (x unscaled), $\lambda(x) = ke^{-(x/2\sigma)^2}, \sigma^2 = 1, k = 10$								
36.8	14.1	13.9	18.8	30.5	55.9	113	245	568
Polynomials of Hermite (x unscaled), $\lambda(x) = ke^{-(x/2\sigma)^2}, \sigma^2 = 0.1, k = 1$								
0.0368	0.0141	0.00644	0.00335	0.00192	0.0012	0.00081	0.000581	0.00044

3.3. Departure from the Optimal Design Versus Departure from the Model

The analysis performed in this part of the study was motivated by [Huber (1975), (1981)], who pointed out that D-optimal designs are strongly non-robust in the case of rather small deviations from linearity and showed that the uniform designs have much more satisfactory behavior than the optimal ones.

First, for illustrative purposes, in Table 2 we present the values of the standardized determinants for different regression degrees for uniform designs. It can be easily seen that departure of a model affects more the value of the standardized determinant than departures of the optimal design. We also note that with increasing number of points in uniform designs the values of standardized determinants monotonically increase as well.

At the next step we introduce the concept of a family of D-optimal designs. D-optimal designs of order k for the regression degree $l < k$ are called *higher-order D-optimal designs*. For Legendre, Jacobi, and Hermite efficiency functions we compute the standardized determinants for different degrees of regression for higher-order D-optimal designs. So, for example, for the optimal design of order 9, which contains 10 points we calculate the standardized determinants for the polynomial regressions of degrees from 1 to 9. These results are presented in Tables 3, 4, and 5. Our goal would be, on the one hand, to verify the Huber's assertions and, on the other hand, to compare the performance of the higher-order D-optimal designs versus uniform ones.

Again we observe that departures from the design are much less important than the departure from the model (Tables 3, 4, and 5), the values of standardized determinants are determined more by the degree of regression than by the «quality» of the design.

Table 2

Standardized determinants of the information matrices for uniform designs
for the Legendre efficiency function

Number of points	Regression degree								
	1	2	3	4	5	6	7	8	9
	Standardized Determinants								
	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-4}$
2	2.50								
3	2.96	1.14							
4	3.12	1.40	5.16						
5	3.20	1.51	6.47	2.32					
6	3.24	1.58	7.17	2.96	1.04				
7	3.27	1.62	7.60	3.35	1.35	4.69			
8	3.28	1.64	7.87	3.60	1.55	6.12	2.10		
9	3.29	1.66	8.06	3.78	1.70	7.13	2.77	0.941	
10	3.30	1.67	8.20	3.90	1.79	7.87	3.26	1.25	4.21

Comparing Tables 2 and 3 for Legendre efficiency functions, we observe that, in difference from the uniform designs, the values of standardized determinants for higher order D-optimal designs monotonically decrease with increasing number of points. But this decrease is slow and monotonic; the values for higher-order D-optimal designs remain always superior to the ones for the uniform designs. The same observation is true for other efficiency functions.

Table 3

Standardized determinants for the information matrices for higher order D-optimal
designs for the Legendre efficiency function

Number of points	Regression degree								
	1	2	3	4	5	6	7	8	9
	Standardized Determinants								
	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$
2	10.00								
3	6.67	3.85							
4	6.00	3.10	1.72						
5	5.71	2.90	1.49	8.10					
6	5.56	2.80	1.42	7.24	3.89				
7	5.45	2.74	1.38	6.97	3.55	1.89			
8	5.38	2.70	1.36	6.83	3.45	1.75	9.24		
9	5.33	2.67	1.34	6.73	3.39	1.71	8.67	4.54	
10	5.29	2.65	1.33	6.66	3.35	1.68	8.47	4.30	2.24

Table 4

Standardized determinants for the information matrices for higher order D-optimal designs for Jacobi (1, 1) efficiency functions

Number of points	Regression degree								
	1	2	3	4	5	6	7	8	9
	Standardized Determinants								
	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-4}$
2	14.8								
3	9.60	7.16							
4	8.40	5.74	3.50						
5	7.84	5.33	3.02	1.73					
6	7.51	5.13	2.87	1.54	8.54				
7	7.30	5.00	2.79	1.48	7.80	4.23			
8	7.15	4.91	2.74	1.45	7.56	3.93	2.10		
9	7.03	4.85	2.70	1.43	7.43	3.82	1.97	1.05	
10	6.95	4.80	2.84	1.42	7.34	3.77	1.92	0.98	5.20

Table 5

Standardized determinants for the information matrices for higher order D-optimal designs for the Hermite efficiency function

Number of points	Regression degree								
	1	2	3	4	5	6	7	8	9
	Standardized Determinants								
	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^0$
2	18.4								
3	10.8	15.8							
4	8.00	12.0	16.1						
5	6.38	10.3	13.5	1.87					
6	5.31	8.99	12.1	1.64	2.41				
7	4.55	8.05	11.0	1.50	2.16	3.37			
8	3.98	7.31	10.1	1.40	2.02	3.08	5.07		
9	3.54	6.71	9.40	1.31	1.91	2.92	4.70	8.12	
10	3.18	6.21	8.80	1.24	1.81	2.78	4.48	7.60	1.37

This analysis shows, that for the same or close number of observations (points) higher-order D-optimal designs are

- (1) more efficient than uniform designs
- and
- (2) as robust as uniform designs.

So the family of higher-order D-optimal designs can be considered as a near optimal, robust family of approximate designs.

4. Conclusion

Analytical solutions to optimization problems in the optimal design theory do not exist in many cases. Numerical calculations for D-optimal designs in the polynomial regression pointed out the following results:

1. For Legendre, Jacobi and other non-exponential efficiency functions lower regression degrees correspond to higher D-efficiency; for Laguerre efficiency functions higher degrees correspond to higher D-efficiency and for Hermite efficiency there is no general rule.
2. For Legendre and Laguerre efficiency functions, less central is a point more important is its influence on the D-efficiency; for Hermite efficiency functions this result is not valid; for Jacobi efficiency functions there is no general rule, the results depend on its parameters.
3. We confirm the Huber's hypotheses, that departures from the design are considerably less important than departures from the model.
4. For Legendre efficiency functions we can compare uniform designs with higher-order D-optimal designs, introduced in this paper. D-optimal higher-order designs are significantly more efficient, and the family of higher-order D-optimal designs can be considered as a near optimal, robust family of approximate designs.

References

- Antille G. Plans d'expérience. Measures et plans d'expérience D-optimaux pour les polynômes de régression // *C.R. Acad. Sci. Paris*. 1977. №285.
- Antille G., Dette H., Weinberg A. A note on optimal designs in weighted polynomial regression for the classical efficiency functions // *J. Statistic. Plann. Inference*. 2003. №113.
- Chang F. C., Lin G. C. D-optimal designs for weighted polynomial regression. *J. Statistic. Plann. Inference*. 1997. №62.
- Dette H., Haines L. M., Imhoff L. Optimal designs for rational models and weighted polynomial regression. *Ann. Statist.* 1999. №27.
- Fedorov V. V. Theory of Optimal Experiments // *Academic Press*, New York, 1972.
- He Z., Studden W. J., Sun D. Optimal designs for rational models // *Ann. Statist.* 1996. №24.
- Huang M.-N. L., Chang F. C., Wong W. K. D-optimal designs for polynomial regression without an intercept. *Statist. Sinica*. 1995. №5.
- Imhoff L., Krafft O., Schaefer M. D-optimal designs for polynomial regression with weight function $\omega(x) = x / (1 + x)$ // *Statist. Sinica* 1998. №8.
- Karlin S., Studden W. J. Optimal experimental designs // *Ann. Math. Statist.* 1966. №37.
- Karlin S., Studden W. J. Tchebycheff systems: with Applications. Analysis and Statistics. Wiley, New York, 1966.
- Kiefer J. C. General equivalence theory for optimum design (approximate theory) // *Ann. Statist.* 1974. №2.
- Pukelsheim F. Optimal Design of Experiments. Wiley, New York, 1993.
- Silvey S. D. Optimal Design. Chapman and Hall, London, 1980.