ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

E5-97-254

Cs.Török*

ON THE ESTIMATION OF THE DEGREE OF REGRESSION POLYNOMIAL

Submitted to «Biometrics»

*Dept. of Mathematics, Technical University, Košice, Slovakia;
E-mail: torokcs@ccsun.tuke.sk

1997

... the caprices of small samples
have to be taken into account ...
F.Mosteller, J.W.Tukey

1 Introduction [1]

The most popular type of regression model used for handling problems that cannot be described by straight lines is the *polynomial model*, in which the dependent variable is related to functions of the powers of the independent variable.

A one-variable polynomial model is defined as follows

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_r x^r + \varepsilon, \tag{1}$$

where y represents the dependent variable, x the independent variable, $\varepsilon \approx \mathcal{N}(0, \sigma^2)$ is a random error term and $\alpha_0, \alpha_1, \ldots, \alpha_r$ are the unknown regression parameters. The highest exponent, or power, of x used in the model is known as the *degree* of the model, and it is customary for a model of degree r to include all terms with lower powers of the independent variable. In selecting a polynomial model, the goal is usually to select the polynomial model of lowest order that adequately represents the trend in the plot.

Polynomial models are primarily used as means to fit a relatively smooth curve to a set of data but the degree of polynomial required to fit a set of data is not usually known a priori.
There are several approaches to determining the *right* degree of the polynomial.

The *partial sums of squares*, that do not depend on the order in which the independent variables are listed in regression model, are used in linear regression procedures. For fitting polynomial models by linear regression procedures *sequential sums of squares* may be used [6].
It is customary to build an appropriate polynomial model by sequentially fitting equations with higher order terms until a satisfactory degree of fit has been accomplished. In other words, it is started by fitting a simple linear regression of y on x. Then a model is specified with linear and quadratic terms, to ascertain if adding the quadratic term improves the fit by significantly reducing the residual mean square. The process can be continued by adding and testing the contribution of a cubic term, then a fourth power term, and so on, until no additional terms are needed.

Another method for checking the appropriateness of a model is to plot the *residual* values. The residual plot can show a systematic pattern when the specified degree of polynomial is inadequate.

[1]The author is grateful for support by the Russian Fundamental Science Foundation under grant 95-01-01467.

Modern regression techniques solve the problem of estimating the degree of polynomial by introducing a penalization function ([1], [2], [3], [7]). If the penalization function is appropriately chosen a *consistent estimator* for the degree can be found. These techniques may be not satisfactory for small number of observations.

We present a new approach to determining the degree of regression polynomial. The main tool of this approach is the discrete projective transformation ([4], [5], [9]). One of the main features of this transformation is the fact, that the DPT decreases the degree of the polynomial by two while the derivation only by one.

The simulation result from section 2 shows the dependence of an application of penalization function on the number of observations. We introduce in section three the DPT, give the basic definitions and properties. In the next two sections we investigate the DPT of polynomials and the aspects of numerical calculation of DPT. The section six is devoted to the use of DPT in estimating the degree of regression polynomial on a real set of data. The last section consists of conclusions.

2 Penalization functions

For given pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ the regression parameters $\alpha_0, \alpha_1, \ldots, \alpha_r$ can be easily estimated by the least squares method (LSM). The main problem is to estimate the degree r of the polynomial, i.e. the number $r+1$ of parameters $\alpha_0, \alpha_1, \ldots, \alpha_r$.

The modern methods solve the problem by introducing a penalization function $q_n(k)$. Akaike was one of the first statisticians, that used penalization functions. His FPE and AIC criteria [1], [2] were proposed to estimate the degree of a stationary autoregressive process in time series analysis.

Assume that $0 \le r \le R$, where R is a given number. Denote s_k^2 an estimator of σ^2, where the model (1) takes into account $k+1$ parameters. The random behaviour of s_k^2 does not allow to determine the beginning of the asymptotically constant part of the function s_k^2, $k = 0, 1, 2, \ldots, R$. The penalization function takes into account besides s_k^2 the number of the parameters too and it is the product

$$g_n(k) = s_k^2(1 + q_n(k))$$

which is analyzed. The function $q_n(k)$ penalizes the growing number k of parameters in the model. If the number of parameters in the model is small and insufficient, $k < r$, the estimated residual variance s_k^2 is large. If we take too many parameters, $k > r$, then the penalization function causes, that the considered product $g_n(k)$ increases.

The number of parameters, which minimizes the product, is taken as an estimator of the dimension of the model. Different authors use different penalization functions. If the penalization function is appropriately chosen, e.g. [3]

$$q_n(k) = kcN^{-\beta}, \qquad c = 1,\ \beta = \frac{1}{4},$$

the estimator is consistent.

In [3] simulated data were used to investigate the dependence of the above criterion on the c, β, σ^2 and regression coefficients. Our simulation study shows that there is a dependence on the number n of observations too. We present the results from simulated data in tab. 1. We simulated a model

$$y = -x + 15x^2 - 10x^3 + x^4 - x^5 + \varepsilon,$$

with $x_1 = -0.5$, step $h = 0.1$ and pseudo random numbers $\varepsilon \approx \mathcal{N}(0, 0.5^2)$.

As we see, if e.g. $n = 24$, then the criterion, based on the minimal value of $g_n(k) = s_k^2(1 + kN^{-0.25})$ gives the degree 5 only for every third simulated sample.

Tab.1 The number of samples with corresponding tested degree by penalization func.

Degree / Sample size	1	2	3	4	5	6	7	8+	Number of simulated samples
40	0	0	0	0	96	4	0	0	100
32	0	0	0	0	94	4	1	1	100
24	0	0	0	55	33	8	2	2	100
16	0	0	78	5	5	5	2	5	100
12	0	0	49	9	10	7	8	17	100

3 DPT

The discrete projective transformation was introduced and its basic properties studied in [4], [5]. This section is devoted to the description of DPT on the base of these works.

If we choose x_0 and fix on a curve of an arbitrary continuous function $f(x)$ two different points $(x_0 + \lambda;\ f(x_0 + \lambda))$ and $(x_0 + L;\ f(x_0 + L))$, then the discrete projective transformation maps any point $P_f \equiv P(x;\ f(x))$ (not equal with the previous ones) of the curve onto the corresponding point $P_h \equiv P(\tau;\ h(\tau))$ on the curve of a new function $h(\tau)$ of simpler geometrical structure

$$\mathcal{D} : P_f \longrightarrow P_h,$$

where $\tau = x - x_0$. We will assume that throughout the paper $x_0 = 0$.

NOTATION *Let us introduce for any* $\tau \ne \lambda \ne L$, $\tau, \lambda, L \in \mathbf{R}$ *the following notations*

$$p_\lambda \equiv p_1(\tau; \lambda, L) = \frac{\tau L}{(\lambda - \tau)(\lambda - L)}, \qquad p_L \equiv p_2(\tau; \lambda, L) = \frac{\lambda \tau}{(L - \lambda)(L - \tau)}, \tag{2}$$

$$p_\tau \equiv p_3(\tau; \lambda, L) = \frac{\lambda L}{(\tau - \lambda)(\tau - L)},$$

and under the addition condition $p_3 \ne 0$ *the notations*

$$d_1(\tau; \lambda, L) = -\frac{p_1}{p_3}, \qquad d_2(\tau; \lambda, L) = -\frac{p_2}{p_3}, \qquad d_3(\tau; \lambda, L) = \frac{1}{p_3}.$$

We mention that the functions d_1, d_2, d_3 are defined only for nonzero λ and L.

The cross-ratio functions $p_i, d_i, i = \overline{1,3}$ play a key role in the definition of DPT and its inverse transformation.

DEFINITION 1 *Let $\lambda \neq L$.*

a) The DPT of an arbitrary differentiable function $f(x)$ is analytically defined as follows

$$\mathcal{D}[f(x)] \equiv f^{\triangleleft}(x) \equiv h(\tau) = (\mathbf{P}, \mathbf{F}), \quad (3)$$

where $(\cdot,\cdot)$ denotes the dot product and

$$\mathbf{P} = [p_\lambda, p_L, p_\tau], \qquad \mathbf{F} = [f(\lambda), f(L), f(\tau)]. \quad (4)$$

b) The inverse DPT is defined by

$$\mathcal{D}^{-1}[h(\tau)] \equiv f(x) = (\mathbf{D}, \mathbf{H}),$$

where

$$\mathbf{D} = [d_1, d_2, d_3], \qquad \mathbf{H} = [f(\lambda), f(L), h(x)]. \quad (5)$$

Before investigating the main features of the DPT we give the basic **properties** of the cross-ratio functions p_i and d_i, $i = \overline{1,3}$:

a) p_i and d_i are independent of scaling (a compression and stretching) i.e. for any $k \neq 0$

$$\left.\begin{array}{l} p_i(k\tau; k\lambda, kL) = p_i(\tau; \lambda, L)), \\ d_i(k\tau; k\lambda, kL) = d_i(\tau; \lambda, L)), \end{array}\right\} \quad i = 1, 2, 3;$$

b) p_i and d_i possess the property of normalization

$$\sum_{i=1}^{3} p_i = 1, \qquad \sum_{i=1}^{3} d_i = 1;$$

c) for $\tau = 0$ we have

$$\begin{array}{c} p_3 = d_3 = 1 \quad \textit{and} \\ p_1 = p_2 = d_1 = d_2 = 0; \end{array}$$

d) the graphs of p_i and d_i, $i = \overline{1,3}$, for $\lambda = -1$ and $L = 1$ are in Fig, 1,2.

Taking into account the property b) of functions $p_i, i = \overline{1,3}$, we get from (3) the

DEFINITION 2 *The DPT of any differentiable function $f(x)$ for any $\lambda \neq L$ may be defined as*

$$\mathcal{D}[f(x)] \equiv f^{\triangleleft}(x) \equiv h(\tau) = \frac{\tau L}{\lambda - L} \cdot \frac{f(\tau) - f(\lambda)}{\tau - \lambda} + \frac{\lambda \tau}{L - \lambda} \cdot \frac{f(L) - f(\tau)}{L - \tau} + f(\tau). \quad (6)$$

Formulae (3), (6) define the DPT for all τ, except of two values: $\tau = \lambda$ and $\tau = L$. From the second definition we get due to the differentiability $f(\cdot)$ the DPT at points λ and L

$$h(\lambda) = \lim_{\tau \to \lambda} h(\tau) = -u f'(\lambda) + v_f \lambda^2 + f(\lambda), \quad (7)$$

and

$$h(L) = \lim_{\tau \to L} h(\tau) = u f'(L) - v_f L^2 + f(L), \quad (8)$$

where

$$u = \frac{\lambda L}{L - \lambda} \quad \text{and} \quad v_f = \frac{f(L) - f(\lambda)}{(L - \lambda)^2}.$$

From definition 2 and the property c) of p_i we get immediately one of the basic properties of DPT:

$$\mathcal{D}[f(0)] = f(0),$$

i.e. the point $[0, f(0)]$ is an immovable point for DPT.
The further two main features of DPT are the linearicity

$$\mathcal{D}[af_1(x) + bf_2(x)] = a\mathcal{D}[f_1(x)] + b\mathcal{D}[f_2(x)] = ah_1(\tau) + bh_2(\tau)$$

where $a, b \in \mathbf{R}$ and f_1, f_2 are arbitrary differentiable functions, and the decrease of the degree of power function by 2

$$\mathcal{D}[x^n] = g(\tau, \tau^2, \ldots, \tau^{n-2}), \qquad n \geq 3.$$

The latter property we formulate as (see [4])

Lemma 1 *Let $\lambda \neq L$. Then*

a) for $n = 1, 2$

$$\mathcal{D}[x^n] = 0,$$

b) and for any natural $n \geq 3$ we have

$$\mathcal{D}[x^n] = \lambda L \tau \sum_{i=1}^{n-1} \lambda^{i-1} \sum_{k=1}^{n-i-1} L^{k-1} \tau^{n-i-k-1}. \quad (9)$$

We mention that from the property c) of p_i it follows immediately that for any constant $c \in \mathbf{R}$

$$\mathcal{D}[c] = c.$$

Hence and from the linearicity of DPT and Lemma 1 we get in particular:

$$\begin{array}{l} \mathcal{D}[a + bx] = a, \\ \mathcal{D}[a + bx + cx^2] = a \end{array}$$

Corollary 1 *Let $\Delta f = (f(\lambda) - f(0), f(L) - f(0), f(x) - f(0))$, where $f(\tau)$ is an arbitrary polynomial of degree less than 3. Then the vectors P and Δf are orthogonal*

$$(P, \Delta f) = 0.$$

4 DPT of polynomials

Let us introduce for the right hand side of (9) the notation

$$G_n(x;\lambda,L)=\lambda L\tau\sum_{i=1}^{n-1}\lambda^{i-1}\sum_{k=1}^{n-i-1}L^{k-1}\tau^{n-i-k-1}.$$

As we can see from the next lemma, $G_n \equiv G_n(x;\lambda,L)$ may be computed by recursion.

Lemma 2 *For $n \geq 3$*

$$G_n = x(G_{n-1}-\lambda G_{n-2}) + L^{n-3} + \lambda G_{n-1},$$

where $G_0 = G_1 = 0$.

For the derivation of matrix equation for the DPT of polynomials it is very important the representation of G_n by power functions x^i.

Lemma 3 *For $n \geq 3$*

$$G_n(x;\lambda,L)=\sum_{i=0}^{n-3}x^i g_{n-3-i}(\lambda,L), \qquad (10)$$

where

$$g_i(\lambda,L)=\sum_{j=0}^{i}L^{i-j}\lambda^j, \qquad j=\overline{0,n-3}. \qquad (11)$$

Remarks:

1. There is a single recursion relation for g_i:

$$\begin{aligned} g_0 &= 1, \\ g_i &= g_{i-1}\lambda + L^i, \qquad i=1,2,\ldots \end{aligned} \qquad (12).$$

2. Although the g_i, $i=0,1,\ldots$ are independent of n, the coefficients g_{n-3-i} by x^i in (10) depend on n.

Formulae for $g_i(\lambda,L)$ and $G_n(x;\lambda,L)$, $n=\overline{3,8}$ are in Tab.2 and 3 respectively.

Tab.2 Formulae for $g_i(\lambda,L)$ and $g_i(-L,L)$, $i=\overline{0,5}$

i	$g_i(\lambda,L)$	$g_i(-L,L)$
0	1	1
1	$\lambda+L$	0
2	$\lambda^2+\lambda L+L^2$	L^2
3	$\lambda^3+\lambda^2L+\lambda L^2+L^3$	0
4	$\lambda^4+\lambda^3L+\lambda^2L^2+\lambda L^3+L^4$	L^4
5	$\lambda^5+\lambda^4L+\lambda^3L^2+\lambda^2L^3+\lambda L^4+L^5$	0

Tab.3 Formulae for $G_n(x;\lambda,L)$ and $G_n(x;-L,L)$, $n=\overline{3,8}$

n	$G_n(x;\lambda,L)$	$G_n(x;-L,L)$
3	$g_0=1$	1
4	$x+g_1=x+\lambda+L$	x
5	$x^2+xg_1+g_2=x^2+x(\lambda+L)+\lambda^2+\lambda L+L^2$	L^2+x^2
6	$x^3+x^2g_1+xg_2+g_3$	xL^2+x^3
7	$x^4+x^3g_1+x^2g_2+xg_3+g_4$	$L^4+x^2L^2+x^4$
8	$x^5+x^4g_1+x^3g_2+x^2g_3+xg_4+g_5$	$xL^4+x^3L^2+x^5$

From Lemmas 1 and 3 we get the expansion of DPT x^n by power functions x^i.

Lemma 4 *Let $\lambda \neq L$. Then for any integer $n \geq 3$ we have*

$$\mathcal{D}[x^n;\lambda,L]\equiv\mathcal{D}[x^n]=\sum_{i=1}^{n-2}x^i d_{in}(\lambda,L), \qquad (13)$$

where

$$d_{in}(\lambda,L)=\lambda L g_{n-2-i}(\lambda,L), \quad i=\overline{1,n-2}, \qquad (14)$$

end $g\cdot$ is defined by (11),(12) (see also Tab.2).

Formulae for d_{in}, $i=\overline{1,6}$, $n=3,8$ and for $\mathcal{D}[x^n]$ are in Tab.4-7.

Tab.4 $d_{in}(\lambda,L)$, $i=\overline{1,5}, n=\overline{3,7}$

i \ n	3	4	5	6	7
1	λL	λLg_1	λLg_2	λLg_3	λLg_4
2		λL	λLg_1	λLg_2	λLg_3
3			λL	λLg_1	λLg_2
4				λL	λLg_1
5					λL

Tab.5 $d_{in}(-L,L)$, $i=\overline{1,5}, n=\overline{3,7}$

i \ n	3	4	5	6	7
1	$-L^2$	0	$-L^4$	0	$-L^6$
2	0	$-L^2$	0	$-L^4$	0
3	0	0	$-L^2$	0	$-L^4$
4	0	0	0	$-L^2$	0
5	0	0	0	0	$-L^2$

Tab.6 The DPT of power function $\mathcal{D}[x^n;\lambda,L]$, $\overline{3,6}$

n	$\mathcal{D}[x^n;\lambda,L]$
3	$Lx\lambda$
4	$x^2\lambda L+(L+\lambda)L\lambda x$
5	$\lambda Lx^3+(L+\lambda)L\lambda x^2+(L^2+\lambda L+\lambda^2)L\lambda x$
6	$\lambda Lx^4+(L+\lambda)L\lambda x^3+(\lambda L+L^2+\lambda^2)L\lambda x^2+(L^3+\lambda L^2+\lambda^3+\lambda^2 L)L\lambda x$

Tab.7 The DPT of power function, $\lambda = -L$

n	$\mathcal{D}[x^n; -L, L]$
3	$-L^2x$
4	$-L^2x^2$
5	$-L^2x^3 - L^4x$
6	$-L^2x^4 - L^4x^2$
7	$-L^2x^5 - L^4x^3 - xL^6$
8	$-L^2x^6 - L^4x^4 - L^6x^2$

From Lemma 4 one can easily get the

Corollary 2 *For $\lambda \neq L$ and $n \geq 3$ $\mathcal{D}[x^n]$ can be presented by the recursion formula*

$$\mathcal{D}[x^n] = x(\mathcal{D}[x^{n-1}] + \lambda L g_{n-3}).$$

From lemma (4) we can derive the DPT of polynomial of degree r

$$f_r(x) = \alpha_{r0} + \alpha_{r1}x + \alpha_{r2}x^2 + \ldots + \alpha_{rr}x^r$$

in matrix form.

Theorem 1 *Let $\lambda \neq L$. Then for any integer $r \geq 0$*

$$\mathcal{D}[f_r(x)] \equiv f_r^{\triangleleft}(x) = X_r{}^T \Delta A_r, \qquad (15)$$

where the vectors X_r and A_r are defined by

$$\begin{aligned} X_r &= (1, x, x^2, \ldots, x^r)^T, \\ A_r &= (\alpha_{r0}, \alpha_{r1}, \alpha_{r2}, \ldots, \alpha_{rr})^T, \end{aligned} \qquad (16)$$

$$\Delta = \begin{bmatrix} 1 & O_r^T \\ O_r & D \end{bmatrix}, \qquad (17)$$

the vector O_r consists of zeros and the nonzero elements d_{ij} of matrix D are defined by formula (14) (see also Tab 3,4).

The result of theorem follows from the next facts:

- the element d_{ik} is equal the coefficient by x^i in the DPT of x^k - $\mathcal{D}[x^k]$, see (13);
- since for any constant c, $\mathcal{D}[c] = c$ and $\mathcal{D}^{-1}[c] = c$, every element of the first row and first column of matrix Δ equals zero except $\Delta_{11} (\equiv d_{00}) = 1$;
- if $k = 1, 2$, then $d_{ik} = 0$ for every i, since $\mathcal{D}[x] = \mathcal{D}[x^2] = 0$;
- if $i = r-1, r$, or if $i \geq 2$, $k \geq 3$ and $k - i \leq 1$, then $d_{ik} = 0$, since DPT decreases the degree of power functions by 2.

The scheme of the matrix Δ is

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{1,3} & d_{1,4} & d_{1,5} & d_{1,6} \\ 0 & 0 & 0 & 0 & d_{2,4} & d_{2,5} & d_{2,6} \\ 0 & 0 & 0 & 0 & 0 & d_{3,5} & d_{3,6} \\ 0 & 0 & 0 & 0 & 0 & 0 & d_{4,6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda L & \lambda^2 L + \lambda L^2 & \lambda^3 L + \lambda^2 L^2 + \lambda L^3 & \lambda^4 L + \lambda^3 L^2 + \lambda^2 L^3 + \lambda L^4 \\ 0 & 0 & 0 & 0 & \lambda L & \lambda^2 L + \lambda L^2 & \lambda^3 L + \lambda^2 L^2 + \lambda L^3 \\ 0 & 0 & 0 & 0 & 0 & \lambda L & \lambda^2 L + \lambda L^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda L \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For $\lambda = -L$ and $r = 8$ we get (see also Tab.5)

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -L^2 & 0 & -L^4 & 0 & -L^6 & 0 \\ 0 & 0 & 0 & 0 & -L^2 & 0 & -L^4 & 0 & -L^6 \\ 0 & 0 & 0 & 0 & 0 & -L^2 & 0 & -L^4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -L^2 & 0 & -L^4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -L^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -L^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let us see two examples on the DPT of polynomials

$$\begin{aligned} &\mathcal{D}[f_6(x)] \equiv f_6^{\triangleleft}(x) = \alpha_0 + \\ &+x\left(\lambda L\alpha_3 + \left(\lambda^2 L + \lambda L^2\right)\alpha_4 + \left(\lambda^3 L + \lambda^2 L^2 + \lambda L^3\right)\alpha_5 + \left(\lambda^4 L + \lambda^3 L^2 + \lambda^2 L^3 + \lambda L^4\right)\alpha_6\right) + \\ &+x^2\left(\lambda L\alpha_4 + \left(\lambda^2 L + \lambda L^2\right)\alpha_5 + \left(\lambda^3 L + \lambda^2 L^2 + \lambda L^3\right)\alpha_6\right) + \\ &+x^3\left(\lambda L\alpha_5 + \left(\lambda^2 L + \lambda L^2\right)\alpha_6\right) + x^4\lambda L\alpha_6. \end{aligned}$$

This DPT in case $\lambda = -L$ is reduced to

$$\mathcal{D}[f_6(x)] \equiv f_6^{\triangleleft}(x) = \alpha_0 + x\left(-L^2\alpha_3 - L^4\alpha_5\right) + x^2\left(-L^2\alpha_4 - L^4\alpha_6\right) - x^3L^2\alpha_5 - x^4L^2\alpha_6$$

The evaluation of DPT of higher orders is done recursively. We show how is computed the k-th DPT $Y^{\triangleleft k}$ on the base of the $k-1$-th DPT.

First of all the function $f^{(k-1)\triangleleft}(x)$ must be computed analytically from the definition 1 or 2 (formulae (3), (6) respectively). The function $f^{(k-1)\triangleleft}(x)$ is needed to evaluate $f^{k\triangleleft}(\lambda)$ and $f^{k\triangleleft}(L)$, more precisely the derivation $\frac{df^{(k-1)\triangleleft}(x)}{dx}$ in λ and L.

We have

$$f^{(k-1)\triangleleft}(x) \equiv \mathcal{D}[f^{(k-1)\triangleleft}(x)] = (P, F^{k-2}), \tag{19}$$

where P is defined by (4) and

$$F^{k-2} = [y_1^{(k-2)\triangleleft}, y_n^{(k-2)\triangleleft}, f^{(k-2)\triangleleft}(x)].$$

The k-th DPT of Y is equal

$$Y^{k\triangleleft} = (f^{k\triangleleft}(\lambda), y_1^{*k\triangleleft}, y_2^{*k\triangleleft}, \ldots, y_{n-2}^{*k\triangleleft}, f^{k\triangleleft}(L))^T,$$

where

$$\begin{aligned}
f^{k\triangleleft}(\lambda) &= -u\frac{df^{(k-1)\triangleleft}(\lambda)}{dx} + v_{k-1}\lambda^2 + y_1^{(k-1)\triangleleft},\\
f^{k\triangleleft}(L) &= u\frac{df^{(k-1)\triangleleft}(L)}{dx} - v_{k-1}L^2 + y_n^{(k-1)\triangleleft},\\
v_{k-1} &= \frac{y_n^{(k-1)\triangleleft} - y_1^{(k-1)\triangleleft}}{(L-\lambda)^2},\\
Y^{*k\triangleleft} &= P_\lambda \cdot y_1^{(k-1)\triangleleft} + P_L \cdot y_n^{(k-1)\triangleleft} + P_\tau \star Y^{*(k-1)\triangleleft},
\end{aligned}$$

and

$$Y^{*(k-1)\triangleleft} = (y_2^{(k-1)\triangleleft}, y_3^{(k-1)\triangleleft}, \ldots, y_{n-1}^{(k-1)\triangleleft})^T.$$

Let us now consider the situation, when

$$f(x) = f_r(x) = \alpha_{r0} + \alpha_{r1}x + \alpha_{r2}x^2 + \ldots + \alpha_{rr}x^r. \tag{20}$$

Let

$$\mu = max(r-2k,\ 0).$$

If for any $k \geq 1$ $\mu = 0$, then $f_r^{k\triangleleft}(x) = const \equiv \alpha_{r0}$ and $Y^{k\triangleleft} = (\alpha_{r0}, \cdots, \alpha_{r0})^T$.

Let us have a look at the case when $\mu > 0$. There are two ways to compute $Y^{k\triangleleft}$.

1. We can follow the algorithm described above, however now the function $f_r^{(k-1)\triangleleft}(x)$ in (19) does not need to be computed analytically.
Let (we remind that by the assumption $\mu \geq 1$)

$$f_r^{(k-1)\triangleleft}(x) \equiv f_\nu(x) = \alpha_{\nu 0} + \alpha_{\nu 1}x + \cdots + \alpha_{\nu\nu}x^\nu,$$

where $\nu = r-2(k-1)$ (of course the coefficients $\alpha_{\nu 0}, \alpha_{\nu 1}, \cdots, \alpha_{\nu\nu}$ have to be remembered). Then

$$\frac{df_r^{(k-1)\triangleleft}(x)}{dx} \equiv \frac{f_\nu(x)}{dx} = \alpha_{\nu 1} + 2\alpha_{\nu 2}x + \cdots + \nu\alpha_{\nu\nu}x^{\nu-1}.$$

The rest of computation is carried out by analogy with the general case. Mention must be made that a convenient way to compute the coefficients of polynomial $f_r^{(k-1)\triangleleft}(x)$ is due to Lemma 5 (see the next way).

2. The second way, which is simpler than the first one however it needs more machine operations, is to use purely the result of Lemma 5.
Let

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^r \\ 1 & x_2 & x_2^2 & \ldots & x_2^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^r \end{pmatrix}.$$

Then from (15) and (18) we get that

$$\begin{aligned}
Y^{\triangleleft} &= \mathbf{X}\Delta A^r (= \mathbf{X}B^1),\\
Y^{k\triangleleft} &= \mathbf{X}\Delta B^{k-1}, \qquad k = 2, 3, \ldots
\end{aligned} \tag{21}$$

where Δ is given by (17), A^r by (16) and $B^{k-1} = \Delta^{k-1}A^r$. The process may be optimized by decreasing the corresponding dimensions of matrices in every iteration step by 2. Consequently the dimensions of matrices in (21) may be reduced to

$$\mathbf{X}_{n\times(r-2k+3)},\quad \Delta_{(r-2k+3)\times(r-2k+3)},\quad B^{k-1}_{(r-2k+3)\times 1}.$$

6 The DPT in regression analysis

The discrete projective transformation was defined in the previous section for continuous functions. From (3) it follows that the DPT of discrete points may be computed too. Let us consider the set of data in tab.8 (see [6]). Our goal is to determine from the discrete points $[x_i, y_i], i = \overline{1,21}$ the appropriate degree of the regression polynomial.

Tab.8

X	14	21	28	35	42	49	56	63	70	77	84
Y	590	910	1305	1730	2140	2725	2890	3685	3920	4325	4410
X	91	98	105	112	119	126	133	140	147	154	
Y	4485	4515	4480	4520	4545	4525	4560	4565	4626	4566	

The discrete points may be transformated by (3) after the choice of two *basic points* $[\lambda, f(\lambda)]$ and $[L, f(L)]$. The basic points may be chosen from the discrete points $[x_i, y_i]$ and their DPT's. Unfortunately this choice has an essential drawback: by computation of every DPT we lose two observations.
Here we show another approach: the basic points will be from the estimated polynomial and its discrete projective transformations.

From the theorem we immediately get the matrix equation of the k-th DPT of polynomial.

Lemma 5 *Let $\lambda \neq L$ and $r \geq 0$. Then for every $k \geq 1$*

$$\underbrace{\mathcal{D}[\mathcal{D}[\ldots\mathcal{D}[}_{k} f_r(x) \underbrace{]\ldots]]}_{k} \equiv \mathcal{D}^k[f_r(x)] \equiv f_r^{k\triangleleft}(x) = X_r{}^T \Delta^k A_r. \quad (18)$$

Examples we show only for D^k, because for every positive natural k

$$\Delta^k = \begin{pmatrix} 1 & O_r^T \\ O_r & D^k \end{pmatrix}.$$

For $\lambda = -L$ and $r = 8$ we have

$$D^2 = \begin{bmatrix} O & L^4 & 0 & 2L^6 & 0 \\ \vdots & 0 & L^4 & 0 & 2L^6 \\ \vdots & 0 & 0 & L^4 & 0 \\ \vdots & 0 & 0 & 0 & L^4 \\ O & \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

$$D^3 = \begin{bmatrix} O & -L^6 & 0 \\ \vdots & 0 & -L^6 \\ O & \ldots & \ldots \end{bmatrix} \quad \text{and} \quad D^4 = [O].$$

The DPT's of order 2 and 3 of polynomial $f_6(x)$ are:

$$\mathcal{D}^2[f_6(x)] \equiv f_6^{2\triangleleft}(x) = \alpha_0 + x\left(\lambda^2 L^2 \alpha_5 + 2\lambda L\left(\lambda^2 L + \lambda L^2\right)\alpha_6\right) + x^2\lambda^2 L^2 \alpha_6,$$

$$\mathcal{D}^3[f_6(x)] \equiv f_6^{3\triangleleft}(x) = \alpha_0.$$

5 The numerical realization of DPT

In this section we describe the numerical implementation of DPT of any differentiable function and of polynomials in particular. As we will see the evaluation of DPT of polynomials may be realized purely numerically without any analytical computation.

In numerical algorithms functions $f(x)$ are presented over an interval $[a,\ b]$ in most cases with pairs $[x_i,\ f(x_i)]$, where $f(x_i)$, $i = \overline{1,n}$ are the function values at the meshpoints x_i, $i = \overline{1,n}$. The distance between two successive meshpoints x_{i-1}, x_i, $i = \overline{2,n}$ may be constant or changeable. It is supposed that $x_1 = a$, $x_n = b$ and $x_i < x_j$, if $i < j$, $i = \overline{1, n-1}$, $j = \overline{i+1, n}$.
So let us consider two vectors

$$X = (x_1, x_2, \ldots, x_n)^T, \quad \text{and}$$
$$Y = (y_1, y_2, \ldots, y_n)^T,$$

where $y_i = f(x_i)$. Our goal is to show an effective way to compute the numerical values of the first, second, ... DPT's of $f(x)$, i.e. the vectors

$$Y^{k\triangleleft} \equiv \underbrace{\mathcal{D}[\mathcal{D}[\ldots\mathcal{D}[}_{k} Y \underbrace{]\ldots]]}_{k} = (y_1^{k\triangleleft}, y_2^{k\triangleleft}, \ldots, y_n^{k\triangleleft})^T, \quad k = 1, 2, \ldots,$$

where $y_i^{k\triangleleft} = f^{k\triangleleft}(x_i)$, $i = \overline{1,n}$ and

$$f^{k\triangleleft}(x_i) = \underbrace{\mathcal{D}[\mathcal{D}[\ldots\mathcal{D}[}_{k} f(x) \underbrace{]\ldots]]}_{k}.$$

By $f^{k\triangleleft}(x)$ is denoted the k-th DPT or the DPT of order k of the function $f(x)$.

The DPT of $f(x)$ is determined by two basic points λ, L and their values of function $f(\lambda)$, $f(L)$. Without loss of generality we take $\lambda = x_1$ and $L = x_n$.
Let us introduce three vectors

$$P_\lambda = \begin{pmatrix} p_\lambda(x_2) \\ p_\lambda(x_3) \\ \vdots \\ p_\lambda(x_{n-1}) \end{pmatrix}, \quad P_L = \begin{pmatrix} p_L(x_2) \\ p_L(x_3) \\ \vdots \\ p_L(x_{n-1}) \end{pmatrix}, \quad P_r = \begin{pmatrix} p_r(x_2) \\ p_r(x_3) \\ \vdots \\ p_r(x_{n-1}) \end{pmatrix},$$

where p_λ, p_L, and p_r are defined by (2). These vectors are used without any change in computations of DPT of any order. It should be noted that they do not contain the corresponding values in the basic points $x_1 = \lambda$ and $x_n = L$. The DPT at these points is calculated by (7), (8) and at $x_2, x_3, \ldots, x_{n-1}$ by (3) or (6).

The first DPT of Y is given by

$$Y^{\triangleleft} = (f^{\triangleleft}(\lambda), y_1^{\star\triangleleft}, y_2^{\star\triangleleft}, \ldots, y_{n-2}^{\star\triangleleft}, f^{\triangleleft}(L))^T,$$

where (see (7), (8))

$$f^{\triangleleft}(\lambda) \equiv y_1^{\triangleleft} = -uf'(\lambda) + v_0\lambda^2 + y_1$$
$$f^{\triangleleft}(L) \equiv y_n^{\triangleleft} = uf'(L) - v_0 L^2 + y_n$$
$$u = \frac{\lambda L}{L-\lambda}, \quad v_0 = \frac{y_n - y_1}{(L-\lambda)^2},$$

and $y_1^{\star\triangleleft}, y_2^{\star\triangleleft}, \ldots, y_{n-2}^{\star\triangleleft}$ are the elements of vector

$$Y^{\star\triangleleft} = P_\lambda \cdot y_1 + P_L \cdot y_n + P_r \star Y^\star,$$

where

$$Y^\star = (y_2, y_3, \ldots, y_{n-1})^T,$$

by the dot $\cdot$ is denoted the multiplication of vector by scalar and by asterisk $\star$ the multiplication of two vectors by the corresponding elements

$$P_r \star Y^\star = \begin{pmatrix} p_r(x_2)y_2 \\ \vdots \\ p_r(x_{n-1})y_{n-1} \end{pmatrix}.$$

The process of the estimation the degree of polynomial by DPT consists of two parts.

I. Computation of DPT's.

The algorithm of computation of DPT is the following:

1. we determine the *initial* polynomial function $\hat{f}_k = \hat{\alpha}_{k,0} + \hat{\alpha}_{k,1}x + \ldots + \hat{\alpha}_{k,k}x^k$ of degree k on the base of points $[x_i, y_i]$ by the least square method;
2. with the two basic points $[\lambda, \hat{f}_k(\lambda)]$ and $[L, \hat{f}_k(L)]$ we compute by (3) the first DPT of points $[x_i, y_i]$, which we denote as $[d_k^{\triangleleft}x_i, d_k^{\triangleleft}y_i]$;
3. we determine the polynomial function $\hat{f}_{k-2} = \hat{\alpha}_{k-2,0} + \hat{\alpha}_{k-2,1}x + \ldots + \hat{\alpha}_{k-2,k-2}x^{k-2}$ of degree $k-2$ by (21);
4. with the basic points $[\lambda, \hat{f}_{k-2}(\lambda)]$ and $[L, \hat{f}_{k-2}(L)]$ we compute by (3) the second DPT of points $[d_k^{\triangleleft}x_i, d_k^{\triangleleft}y_i]$, which we denote $[d_k^{2\triangleleft}x_i, d_k^{2\triangleleft}y_i]$;
5. if $k-4 \geq 1$, then the steps 3, 4 may be continued;
6. the whole process 1-5 may be repeated by higher degree k.

The sample

$$d_k^{p\triangleleft} = \{[d_1^{p\triangleleft}x_i, d_k^{p\triangleleft}y_1], [d_k^{p\triangleleft}x_2, d_k^{p\triangleleft}y_2], \ldots, [d_n^{p\triangleleft}x_i, d_k^{p\triangleleft}y_n]\}$$

is got from $\{[x_i, y_i]\}_{i=\overline{1,n}}$ by p consecutive DPT, where k is the degree of the initial tested polynomial.

The tab.9 shows the scheme of the computation of $d_k^{p\triangleleft}$ - the DPT's of initial sample.

Tab.9 The computation scheme of DPT's.

Degree of polynomial	3	4	5	6	7	8	...
					$d_7^{\triangleleft}$	$d_8^{\triangleleft}$	...
			$d_5^{\triangleleft}$	$d_6^{\triangleleft}$	$d_7^{2\triangleleft}$	$d_8^{2\triangleleft}$	...
Last non zero DPT	$d_3^{\triangleleft}$	$d_4^{\triangleleft}$	$d_5^{2\triangleleft}$	$d_6^{2\triangleleft}$	$d_7^{3\triangleleft}$	$d_8^{3\triangleleft}$	...
Zero DPT	$d_3^{2\triangleleft}$	$d_4^{2\triangleleft}$	$d_5^{3\triangleleft}$	$d_6^{3\triangleleft}$	$d_7^{4\triangleleft}$	$d_8^{4\triangleleft}$	...

Remark: If λ and L are too near x_{min} and x_{max} then the DPT of observations from the left and right hand side of the interval $[x_{min}, x_{max}]$ might be very far away from the axe x. The experiment shows that the distance $3h$-$8h$ from x_{min} and x_{max} seems to be appropriate. We carried out the computation with $\lambda = -21$ and $L = 189$.

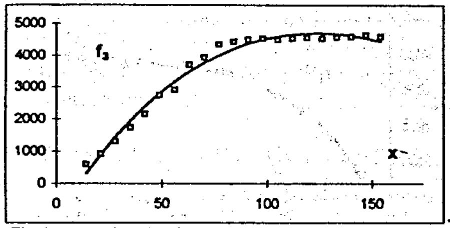

Fig.1 Approximation by polynomial of degree 3

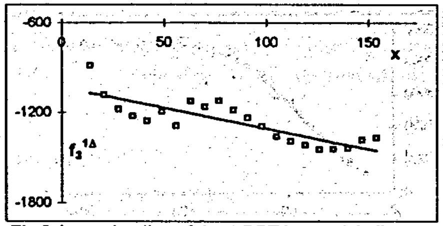

Fig.2 Approximation of the 1.DPT by a stright line

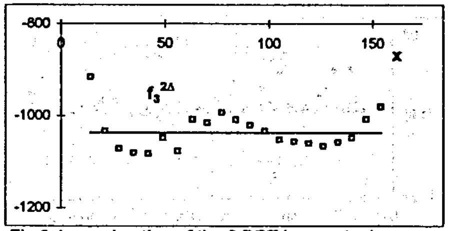

Fig.3 Approximation of the 2.DPT by constant

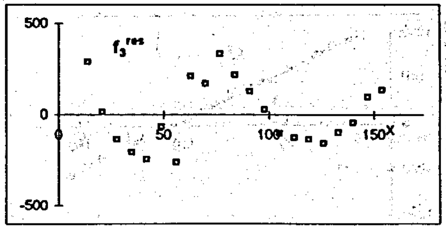

Fig.4 Residuum by polynomial of degree 3

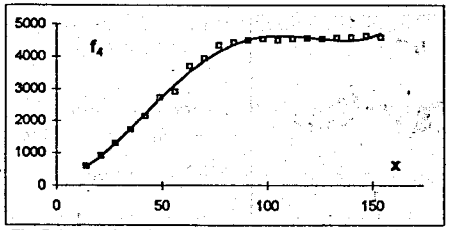

Fig.5 Approximation by polynomial of degree 4

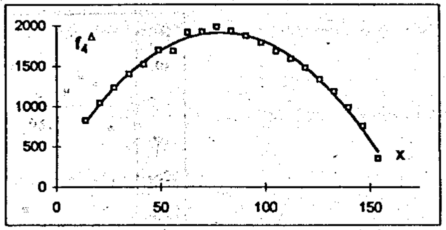

Fig.6 Approximation of the 1.DPT by a parabola

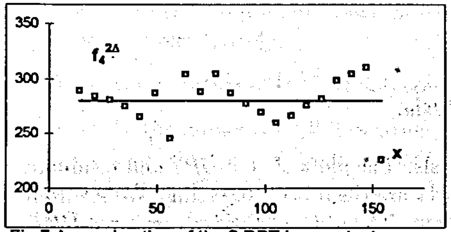

Fig.7 Approximation of the 2.DPT by constant

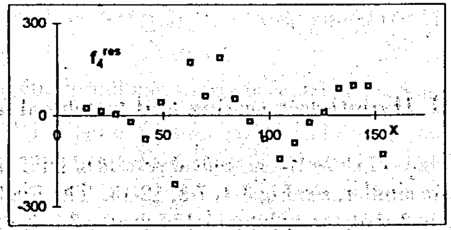

Fig.8 Residuum by polynomial of degree 4

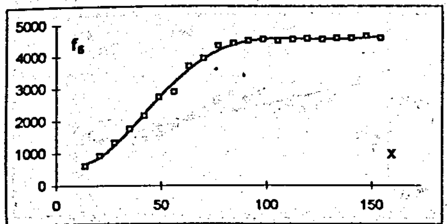

Fig.9 Approximation by polynomial of degree 5

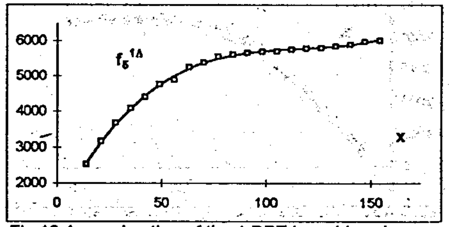

Fig.10 Approximation of the 1.DPT by cubic polynom.

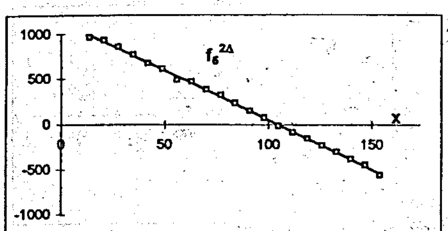

Fig.11 Approximation of the 2.DPT by a stright line

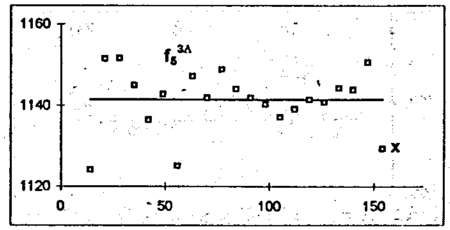

Fig.12 Approximation of the 3.DPT by constant

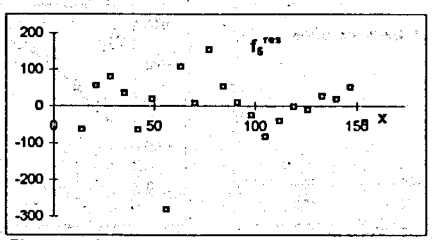

Fig.13 Residuum by polynomial of degree 5

II. Hypotheses testing and graphical analysis.

Fig.1-13 show the graphical results of DPT analysis. The plots of *zero DPT* and residuum are similar, see Fig.3-4, 7-8, 12-13. The Fig.2, 6, 11 are the most interesting. We see from Fig.2 that the polynomial of degree 3 is insufficient. The plots of the *last non zero DPT*, f_4^{Δ}, $f_5^{2\Delta}$, in Fig.6, 11 reveal no patterns, so the appropriate degree of the polynomial model by the graphical analysis may be 4 or 5.

The DPT analysis enables to use in addition to the visual criterion statistical criteria too. We can test single hypotheses, that f_4^{Δ} and $f_5^{2\Delta}$ form a parabola and a straight line respectively. The results of these statistical tests do not contradict the graphical ones. Mention must be made that the standard regression criteria give degree 5 and the penalization criterion degree 6.

The tab.10 shows the hypotheses for the last non zero DPT's.

Tab.10 Hypotheses for the last non zero DPT's.

Degree of polynomial	3	4	5	6	7	8	...
Last non zero DPT	d_3^{Δ}	d_4^{Δ}	$d_5^{2\Delta}$	$d_6^{2\Delta}$	$d_7^{3\Delta}$	$d_8^{3\Delta}$	...
Hypotheses	line	parabola	line	parabola	line	parabola	...

7 Conclusions

This work is an outcome of study on how the discrete projective transformation can be used in estimating the degree of polynomial in regression models. The proposed approach is illustrated on a real set of data.

The special structure of the DPT of polynomials allowed us to construct a *recursive matrix algorithm* for the DPT of estimated regression polynomial.

As we pointed out, the main problem of the new method is the right choice of the basic points for DPT. In this paper we chose the basic points on the estimated regression curves and their DPT's.

The main features of DPT analysis in the estimating of the degree of the regression polynomial are:

- the use of DPT leads to both graphical and statistical criteria;
- the DPT's are associated with the stages of passage from regression polynomial to the residuum;
- the simplest statistical criterion is based on testing of prime hypothesis that the last nonzero DPT represents a line or parabola, if the real degree r of polynomial is odd or even respectively;
- the computation of DPT of polynomial is simple and has a form of matrix equation;
- unlike the numerical differentiation the DPT is not a local operation and due to the basic points λ and L the influence of the distance between the observation on the evaluation of DPT of sample may be eliminated.

The generalization of the new approach to the multidimensional case is under investigation.

There are three files: **DPT.m**, **DPTexmp1.mws** and **DPTexmp2.mws** on the JINR CV server. MapleV Release4 (Windows) programs DPTexmp1.mws, DPTexmp2.mws show the use of DPT.m package of Maple procedures, that carry out the above presented methods and computations. These files can be downloaded by FTP from JINR server **cv.jinr.ru** as anonymous from directory **/pub/PC/maple/dpt**.

References

[1] Akaike H., Fitting autoregressive models for prediction. Ann. Inst. Statist. Math. 21(1969), 243-247

[2] Akaike H., Information theory and an extension of the maximum likelihood principle. Proc. 2nd Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory (1972), 267-281

[3] Anděl J., Garrido M.T., Insua A., Estimating the dimension of a linear model, Kybernetika 17 (1981) 514-525

[4] Dikoussar, N.D., Adaptive projective filters for track finding, Comput. Phys. Commun. 79 (1994) 39-51

[5] Dikoussar, N.D.: Function parametrization by using 4-point transforms, Comput. Phys. Commun. 99 (1997) 235-254

[6] R.J.Freund, R.C.Littell: SAS System for Regression, SAS Institute Inc., Cary NC USA, 1991

[7] Geweke J., Meese R., Estimating regression models of finite but unknown order, International Economic Review 22 (1981) 55-70

[8] F.Mosteller, J.W.Tukey, Data analysis and regression, Addison-Wesley, Massachusetts (Moskva 1982)

[9] Török Cs., N.D.Dikoussar, Approximation with Discrete Projective Transformation, Computers and Mathematics with Applications, accepted for publication (1997)

Received by Publishing Department
on August 15, 1997.