AN ALGORITHM AND PROGRAM FOR POLYNOMIAL REGRESSION ANALYSIS BASED ON THE LEVEL AND CONSISTENCY ESTIMATES

KONSTANTIN MOMIROVIĆ and MIRELA MATEČIĆ
Institute of Kinesiology, Computing center, Zagreb

ŽIGA KNAP
Institute of Sociology and Philosophy, Ljubljana

An algorithm is proposed and a program is written for polynomial regression analysis, under the least squares and covariance maximisation models, of some criterion variate in the space of regressors defined by second order polynomials of measures of level and consistency on a homogenous set of characteristics in Harris universal metrics.

ALGORITAM I PROGRAM ZA POLINOMIJALNU REGRESIJSKU ANALIZU NA OSNOVU PROCJENE NIVOA I KONZISTENCIJE

Predložen je algoritam i napisan program za polinomijalnu regresijsku analizu, pod modelom najmanjih kvadrata i pod modelom maksimalizacije kovarijance, neke kriterijske varijable u prostoru regresora definiranih polinomom drugog reda mjera nivoa i konzistencije izvedenih iz skupa homogenih karakteristika reparametriziranih na Harrisovu univerzalnu metriku.

KEYWORDS
Polynomial regression / level / consistency / Harris space

1. INTRODUCTION

In a recent work Hošek (1988) proved that, beside level, the inconsistency of social status can be, in a polynomial model, the significant predictor of some types, and especially of aggressive behaviour. In that work an algorithm and program proposed and written by Momirović and Erjavec (1988) was applied; the main feature of this algorithm is second order polynomial regression analysis with primary regressors defined as the first principal component and the standard deviations of deviates from the first principal components of variates transformed to standardized image form. Although transformation to image metrics has some advantages (normalization of variables and reduction of inconsistency due to error) the proposed procedure is at the same time too exotic and too sophisticated; and use of least squares criterion in full polynomial model leads to peculiar effects because of the ill defined covariance matrix among regressors.

The aim of this paper is to present a slightly modified another algorithm for the regression analysis under polynomial model defined by the level and inconsistency measures as primary regressors. Both least squares and covariance maximization model in analysis of some criterion variable are applied; regressors are defined as the first principal component and the standardized deviation about it of variates reparametrized to Harris universal metrics. This allows the full amount of inconsistency about the variable with maximal representativness for a set of variables defining social status indicators but not restricted only to that and, by covariance maximization criterion, elimination of the effects of ill deffined covariance matrix.

2. ALGORITHM

Let $E = \{ e_i ; i = 1, \ldots, n \}$ be a random sample from a population P; let $V = \{ v_j ; j = 1, \ldots, m \}$ be a set of variables with, at least, weak ordering properties, and let K be a quantitative variable with logical status of criterion predictable by V or some function(s) of V. Let $K = E \otimes K$ be standardized vector of criterion variable, and let $Z = E \otimes V$ be a standardized matrix of E described on V, so that $R = Z^T Z \frac{1}{n}$ is a simple correlation matrix of variables from V. Define $U^{-2} = \text{diag}\ R^{-1}$ so that $H = ZU^{-1}$ is a data matrix in Harris universal metric space for the variables from V on E with covariance matrix $C = H^T H = U^{-1} R U^{-1}$. Let λ be the first eigenvalue of C and X eigenvector, $X^T X = 1$ associated to λ.

The measure of level of characteristics defined by V is defined as $L = HX \lambda^{-1/2}$, and the measure of consistency of these characteristics as $S^* = (((H - LE_m^T) \circ (H - LE_m^T))^T E_n)^{1/2}$, with $\circ$ denoting Hadamard multiplication, E_m is summation vector of order m, and E_n summation vector of order n. Standardization of S^*, obtained by $S = (S^* - P S^*) D_{S^*}^{-1}$, with $P = E_n (E_n^T E_n)^{-1} E_n$ and $D_S^2 = (S^{*T} S^* - S^{*T} P S^*) \frac{1}{n}$ is then performed in order to obtain a comparable metrics for the level and consistency measures. The content of these measures can be estimated from structural vectors $F_L = UX \lambda^{1/2}$ and $F_S = Z^T S \frac{1}{n}$.

Second order polinomials of L and S are defined by the set $B^* = L, S, L^2, S^2, L \circ S$. Denote by $B = (B^* - PB^*) D_B^{-1}$, $D_B^2 = \text{diag}\,(B^{*T} B^* - B^* P B^*) \frac{1}{n}$ the standardized second order polynomials of level and consistency measures.

Let $Q = B^T K \frac{1}{n}$ be a crosscorrelation vector between so obtained regressors and criterion variables and $R_B = B^T B \frac{1}{n}$ be a correlation matrix for regressors. The least squares regression can be then defined, in the frame of general canonical model, as the solution of the problem

$$(1) \qquad BV = K^* \qquad \begin{aligned} K^T K^* \tfrac{1}{n} &= \rho = \max \\ K^{*T} K^* \tfrac{1}{n} &= 1, \end{aligned}$$

and the maximum covariance (so called stupid regression) as the solution of the problem

$$(2) \qquad BW = \tilde{K} \qquad \begin{aligned} K^T \tilde{K} \tfrac{1}{n} &= c = \max \\ W^T W &= 1. \end{aligned}$$

The well known (Dobrić, Štalec and Momirović, 1984; Dobrić, 1986) solutions of (1) and (2) are

$$V = R_B^{-1} Q \rho^{-1}$$

with $\rho^2 = Q^T R^{-1} Q$, and

$$W = Q(Q^T Q)^{-1/2};$$

with $\sigma^2 = W^T R W$ and $\eta = c\,\sigma^{-1}$, quasimultiple correlations between B^* and K, the solution of (2), optimally scaled for error, is

$$W^* = V \sigma^{-1} \eta .$$

Test of significance for ρ , the standard multiple correlation between B^* and K is $f_\rho = \rho^2/(1-\rho^2) * 5/(n-6)$ distributed as Snedecor f with 5 and n-6 degrees of freedom, and test of significance of quasimultiple correlation η is simply $f_\eta = \eta^2((n-2)/(1-\eta^2))$, distributed as Snedecor f with n-2 degrees of freedom. Tests of significance of elements of V are derived from $(1-\rho^2)\mathrm{R}_{\mathrm{B}}^{-1}(n-6)^{-1}$, their covariance matrix, and test of significance of elements W are equal to tests of significance of elements of Q, which are the simple correlations between variables from B and criterion variable K .

3. PROGRAM

The proposed algorithm is almost literally implemented in macro program THEOPHANIA, written in 4.04 B version of GENSTAT. The entire listing of this macro is presented. Samples of behaviour of THEOPHANIA can be obtained from the second author.

Macro program THEOPHANIA is stored in SRCE*GENS-MACRO. library and can be activated by 'REFERENCE' program which generates following structures:

(1) variate structures V (1...M) for original regressors
(2) variate structure K for criterion variate
(3) pointer structure VNM for regressors name

```
'MACRO'   THEOPHANIA  $

        WRITTEN BY
                K. MOMIROVIĆ AND Z. KNAP. MOSCOW AND KYEW. APRIL 1989.
        IMPLEMENTATION
                M. MATEČIĆ. ZAGREB. SEPTEMBAR 1989.
        FUNCTION
                POLYNOMIAL REGRESSION ANALYSIS UNDER THE LEAST SQUARES
                AND COVARIANCE MAXIMISATION MODEL OF SOME CRITERION
                VARIATE IN THE SPACE OF REGRESSORS DEFINED BY SECOND
                ORDER POLYNOMIALS OF MEASURES OF LEVEL AND CONSISTENCY
                ON A HOMOGENOUS SET OF CHARACTERISTICS IN HARRIS METRICS.
        DOCUMENTATION
                K. MOMIROVIĆ, M. MATEČIĆ AND Z. KNAP:
                AN ALGORITHM AND PROGRAM FOR POLYNOMIAL REGRESSION
                ANALYSIS BASED ON LEVEL AND CONSISTENCY ESTIMATES.
                METODOLOŠKI ZVESKI, 5, JUS, LJUBLJANA, 1989.
        REQUIREMENTS
                REFERENCE PROGRAM MUST TRANSFERE TO THEOPHANIA THE
                FOLLOWING STRUCTURES:
                (1) A VARIATE STRUCTURE V(1...M)$N FOR ORIGINAL REGRESSORS
                (2) A VARIATE STRUCTURE K$N FOR CRITERION VARIATE
                (3) A POINTER STRUCTURE VNM FOR REGRESSORS NAMES
''
''
        SECTION O.
        DECLARATIONS.
''
'LOCAL'   NENT,NVAR,CRIT,FREG,AS,SD,ASC,SDC,C,QT,Q,R,BETA,GAMA,FBETA,
          FGAMA,QBETA,QGAMA,RHOSQ,RHO,ETASQ,ETA,FRHO,FETA,QRHO,QETA,RELRE,
          ERRRHO,ERRETA,DF1,DF2,DF,U,SS,RELRER,RELBG,RELFS,LUD1,LUD2,
          LUD3,LUD4,LUD5,UV,F,S,SEBETA,SEQ,ERRRSQ
'SCAL'    NENT,NVAR,ASC,SDC,RHOSQ,RHO,ETASQ,ETA,FRHO,FETA,QRHO,
          QETA,RELRER,RELREE,RELBG,RELFS,ERRRHO,ERRETA,
          DF1,DF2,DF,LAM,ERRRSQ
'STAR'
'EQUA'    NENT,NVAR = N,M
'RUN'
'SCAL'    AS(1...NVAR) : SD(1...NVAR)
'VARI'    L,CON,L2,CON2,LCON,CRIT$ NENT
'SET'     FREG= L,CON,L2,CON2,LCON,CRIT
'VARI'    VRS(1...NVAR)$MENT
'STAR'
'EQUA'    VRS(1...NVAR) = V : CRIT = K
'RUN'
'DSSP'    SSS $ VRS(1...NVAR) : SS $ FREG
'DIAG'    USNVAR : UU$5 : LAMBDA$1
'POIN'    PNT = L,CON,L2,CON2,LCON,CRIT
'POIN'    RGNM = L,CON, L2,CON2,LCON
'VARI'    LUD5S NENT
'SYMM'    C$ PNT : RR,R$ RGNM : CC$ NVAR
```

```
'MATR'      QT$ 1,RGNM : Q,BETA,GAMA,FBETA,FGAMA,S,F,QBETA,QGAMA,
                         SEBETA,SEQ$ RGNM,1
'MATR'      AS $ 1, NVAR
'MATR'      HL,HCON$ VNM,1 : LUD1$ NENT,NVAR : LUD2$ NVAR,NENT
'MATR'      LUD3$ NENT,1 : LUD4$ 1, NENT
''
        SECTION 1.
        REPARAMETRIZATION OF REGRESSORS AND CRITERION.
''
'SSP'       SS S
'EQUA'      CC,AS,NENT= SSS
'EQUA'      AS(1...NVAR)= AS
'CALC'      U= CC/NENT : U=SQRT(U)
'EQUA'      SD(1...NVAR)= U
'DEVA'      U
'CALC'      CC= CORMAT(CC)
'CALC'      U= INV(CC) : U= SQRT(U)
'CALC'      VRS(1...NVAR)= VRS(1. .NVAR)-AS(1...NVAR)
'CALC'      VRS(1...NVAR)= VRS(1...NVAR)/SD(1...NVAR)
'EQUA'      SD(1...NVAR)= U
'CALC'      VRS81...NVAR)= VRS(1...NVAR)/SD(1...NVAR)
'DEVA'      SSS,CC,AS,SD
'EQUA'      CRIT= K
'CALC'      ASC= MEAN(CRIT) : SDC= SQRT(VAR(CRIT))
'CALC'      CRIT= (CRIT-ASC)/SDC
'SSP'       SSS
'EQUA'      CC,AS,NENT= SSS
'DEVA'      ASC,SDC
'CALC'      CC= CC/NENT
''
        SECTION 2.
        CALCULATION OF LEVEL AND CONSISTENCY ESTIMATES.
''
'LRV'       CC;HL,LAMBDA,DF1
'EQUA'      LAM= LAMBDA
'DEVA'      DF1
'EQUA'      LUD2= VRS(1...NVAR)$ NENT
'CALC'      LUD1= TRANS(LUD2)
'DEVA'      LUD2
'CALC'      L= PDT(LUD1;HL) : L= L/SQRT(LAM) : HL= HL/SQRT(LAM)
'CALC'      HL= PDT(U;HL)
'CALC'      L2= L*L
'CALC'      VRS(1...NVAR)= VRS(1...NVAR)-L
'CALC'      CON= VVAR(VRS(1...NVAR)) : CON=SQRT(CON)
'CALC'      CON= (CON-MEAN(CON))/(SQRT(VAR(CON)))
'EQUA'      LUD4= CON
'CALC'      LUD3= TRANS(LUD4)
'DEVA'      LUD4
'CALC'      HCON= TPDT(LUD1;LUD3)   :   HCON= HCON/NENT
'CALC'      HCON= PDT(U;HCON)
'CALC'      CON2= CON*CON
'CALC'      LCON= L*CON
'EQUA'      FREG= L,CON,L2, CON2, LCON,CRIT
```

```
'DEVA'      LUD1,LUD3,ASC,SDC,VRS(1...NVAR)
''
         SECTION 3.
         LEAST SQUARES POLYNOMIAL REGRESSION
''
'SSP'       SS
'CALC'      C= CORMAT(SS)
'EQUA'      R,QT,DF= C
'CALC'      Q= TRANS(QT)
'CALC'      RR= INV(R)
'CALC'      BETA= PDT(RR;Q)
'CALC'      RHOSQ= PDT(QT;BETA)
'CALC'      RHO= SQRT(RHOSQ)
'CALC'      DF2= NENT-5-1
'CALC'      ERRRSQ= 1.0-RHOSQ
'CALC'      RR= RR*ERRRSQ : RR=RR/DF2
'CALC'      UU= RR
'EQUA'      SEBETA= UU
'CALC'      FBETA= (BETA*BETA)/SEBETA
'CALC'      QBETA= FPROB(FBETA;1;DF2)
'CALC'      FRHO= RHOSQ/ERRRSQ
'CALC'      FRHO= FRHO*(DF2/5)
'CALC'      FRHO= FPROB(FRHO;5;D2)
'CALC'      ERRRSQ= SQRT(ERRRSQ)
'CALC'      SEBETA= SQRT(SEBETA)
'CALC'      F= Q/RHO
''
         SECTION 4.
         STUPID POLYNOMIAL REGRESSION ANALYSIS.
''
'CALC'      DF= SQRT(SUM(Q*Q))
'CALC'      ETA= DF/SQRT(DF)
'CALC'      ETASQ= ETA*ETA : ERRETA= 1.0-ETASq
'CALC'      GAMA= Q/DF : S= PDT(R;GAMA)
'CALC'      DF= TPDT(GAMA;S)
'CALC'      S= S/SQRT(DF)
'CALC'      DF1= NENT-2
'CALC'      FETA= ETASQ*(DF1/ERRETA)
'CALC'      QETA= FPROB(FETA;1;DF1)
'CALC'      ERRETA= SQRT(ERRETA)
'CALC'      FGAMA= Q*Q*(DF1/1.0-DF1)
'CALC'      QGAMA= FPROB(FGAMA;1;DF1)
'CALC'      SEQ= (1.0-Q)/SQRT(NENT)
''
         SECTION 6.
         RELATIONS BETWEEN LEAST SQUARES AND MAXIMUM COVARIANCE
         POLYNOMIAL REGRESSION ANALYSIS.
''
'CALC'      RELRER= ETA/RHO
'CALC'      RELREE= ERRRSQ/ERRETA
'CALC'      RELBG= 1.0/SQRT(SUM(BETA*BETA))
'CALC'      RELFS= 1.0/SQRT(SUM(GAMA*GAMA))
'CALC'      RELBG= RELBG*(SUM(BETA*GAMA))*RELFS
'CALC'      RELFS= 1.0/SQRT(SUM(F*F))
'CALC'      DF= 1.0/SQRT(SUM(S*S))
```

```
'CALC'      RELFS= RELFS*(SUM(F*S)*DF
''
        SECTION 7.
        PRINTED OUTPUT.
''
'PAGE'
'LINE'      10
'CAPT'      ''***************************************************
              *                 T H E O P H A N I A               *
              *                                                   *
              *  POLYNOMIAL REGRESSION ANALYSIS BASED ON LEVEL    *
              *                                                   *
              *  AND CONSISTENCY ESTIMATES OF A SET OF            *
              *  REGRESSORS IN UNIVERSAL METRIC SPACE             *
              *                                                   *
              *****************************************************
'LINE'      6
'CAPT'      '' NUMBER OF ENTITIES   (NENT) AND NUMBER OF ORIGINAL
               REGRESSORS (NVAR)''
'PRINT/S'      NENT,NVAR $ 16.0
'PAGE'
'LINE'      2
'CAPT'      '' STRUCTURE OF LEVEL (HL) AND INCONSISTENCY (HCON)
               ESTIMATES''
'PRINT/P'      HL,HCON $ 16.3
'PAGE'
'LINE'      2
'CAPT'      '' CORRELATIONS OF REGRESSORS (R) AND COVARIANCES OF
               STANDARDISED LEAST SQUARES PARTIAL REGRESSION
               COEFFICIENTS (RR)''
'PRIN'         R $ 10.3
'LINE'      6
'PRIN'         RR $ 10.3
'PAGE'
'LINE'      2
'CAPT'      '' MAIN RESULTS OF LEAST SQUARES REGRESSION ANALYSIS:
               RHOSQ = COEFFICIENT OF DETERMINATION
               RHO   = MULTIPLE CORRELATION
               ERRRSQ= STANDARD ERROR OF PREDICTION
               FRHO  = F-TEST OF HYPOTHESIS RHO=0
               QRHO  = PROBABILITY OF ERROR IN INFERENCE RHO .NE.0
               IN NEXT TABLE THE FOLLOWING CHARACTERISTICS
               OF REGRESSORS ARE PRESENTED:
               BETA  = STANDARDISED PARTIAL REGRESSION COEFFICIENTS
               SEBETA= STANDARD DEVIATIONS OF BETA COEFFICIENTS
               FBETA = TESTS OF HYPOTHESES BETA(J)=0
               QBETA = PROBABILITY OF ERROR IN INFERENCES BETA(J).NE.0
               F     = STRUCTURE OF REGRESSION FACTOR''
'PRINT/S'      RHOSQ,RHO,ERRRSQ,FRHO,QRHO $ 10.3
'LINE'      3
'PRINT/P'      BETA,SEBETA,FBETA,QBETA,F  $  10.3
'PAGE'
'LINE'      2
```

```
'CAPT'      '' MAIN RESULTS OF MAXIMUM COVARIANCE
               REGRESSION ANALYSIS:
               ETASQ = COEFFICIENT OF QUASIDETERMINATION
               ETA   = QUASIMULTIPLE COEFFICIENT OF CORRELATION
               ERRETA= STANDARD ERROR OF PREDICTION BASED ON
                       MAXIMUM COVARIANCE METHOD
               FETA  = APPROXIMATE TEST OF HYPOTHESIS ETA=0
               QETA  = PROBABILITY OF ERROR IN INFERENCE ETA .NE.0
               IN THE NEXT TABLE THE FOLLOWING CHARACTERISTICS OF
               REGRESSORS UNDER MAXIMUM COVARIANCE MODEL ARE PRESENTED:
               Q     = CORRELATIONS BETWEEN REGRESSORS AND CRITERION
               SEQ   = STANDARD DEVIATIONS OF Q
               GAMA  = STUPID REGRESSION COEFFICIENTS
               FGAMA = TESTS OF HYPOTHESES Q(J).GAMA(J) = 0
               QGAMA = PROBABILITIES OF ERROR IN INFERENCES
                       Q(J), GAMA(J) .NE.0
               S     = STRUCTURE OF REGRESSION FACTOR UNDER
                       MAXIMUM COVARIANCE MODEL.''
'PRIN/S'       ETASQ,ETA,ERRETA,FETA,QETA $ 10.3
'LINE'      3
'PRIN/P'       Q,SEQ,GAMA,FGAMA,QGAMA,S  $ 10.3
'PAGE'
'LINE'      2
'CAPT'      '' RELATIONS BETWEEN LEAST SQUARES AND MAXIMUM COVARIANCE
               REGRESSION ANALYSIS:
               RELRER = CORRELATION OF REGRESSION ESTIMATES
               RELREE = CORRELATION OF ERROR ESTIMATES
               RELBG  = CONGRUENCE OF REGRESSION COEFFICIENTS
               RELFS  = CONGRUENCE OF FACTOR STRUCTURES''
'PRIN/S'       RELRER,RELREE,RELBG,RELFS $ 10.3
'PAGE'
''
          END OF MACRO THEOPHANIA:
          AUTHORS ARE GRATEFULL TO A. HOSHEK AND S. SAKSIDA
          FOR MANY CRITICAL COMMENTS AND HELPFUL SUGGESTIONS,
          AND TO ACADEMIES OF SCIENTCIES OF SSSR AND USSR
          FOR PROVIDING EXEPTIONAL WORKING FACILITIES.
''
'ENDMACRO/LOCAL=DESTROY'
```

REFERENCES

1. Dobrić, V., J. Štalec and K. Momirović (1984):
   Note on some relationships between least squares and robust regression analysis.
   Proc. of 6th International Symposium "Computer at the University", 507:1-7

2. Dobrić, V. (1986):
   On a class of robust methods for multivariate data analysis.
   COMPSTAT 86, Physica-Verlag, Heidelberg, pp. 211-216.

3. Hošek, A. (1988):
   Effect of inconsistency of social status on some types of aggressive behaviour (in Croatian; English summary).
   Metodološki zveski, 3, pp. 119-132.

4. Momirović, K. and N. Erjavec (1988):
   IRENA. Program library SRCE*GENS-MACRO.
   University computing center, Zagreb