Generalization of regression analysis to the spatial context

SK Mishra

North-Eastern Hill University, Shillong

12. April 2004

Generalization of Regression Analysis to the Spatial Context

SK Mishra
Dept. of Economics
NEHU, Shillong, India.

1. Introduction: The conventional (linear) regression analysis assumes that the dependent variable (regressand), y, is a linear function of $X=(x_1, x_2, ..., x_m)$ such that $y = X\beta$. The (regression) parameters, $\beta = (\beta_1 \;\; \beta_2 \;\; ... \;\; \beta_m)'$, may be visualized as $\beta_j = \frac{\partial y}{\partial x_j}$; $j = 1,2,...,m$. In the population, however, y may be influenced by many other variables uncorrelated with $X=(x_1, x_2, ..., x_m)$. Hence, if we draw a sample (consisting of n individuals, $n > m$) and we describe our sample as $(y[n], \;\; X[n,m])$, no β (howsoever we choose them) will exactly satisfy the relationship $y = X\beta$. A discrepancy vector $u = (u_1 \;\; u_2 \;\; ... \;\; u_n)'$ will make up the equality relationship such that $y = X\beta + u$. Fixing the $X[n,m]$ matrix, if we draw g repeated samples, we will obtain g number of discrepancy vectors, $u^{(1)}, u^{(2)}, ..., u^{(g)}$. The conventional regression analysis assumes that $E(u_i^{(1)}, u_i^{(2)}, ..., u_i^{(g)}) = 0 \;\; \forall \;\; i = 1,2,...,n$. Here $E(.)$ is the expectation of (.). Moreover, it assumes that $E(u_i u_j') = [\sigma^2] \;\; \forall \;\; i, j = 1,2,...,n$ is a diagonal matrix with strictly positive diagonal elements all equal. Additionally, it assumes that $X[n,m]$ is non-stochastic and of full rank m. Under these (Gauss-Markov) assumptions, β is estimated by the Least Squares method, which gives us $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ and this $\hat{\beta}_{OLS}$ is the best linear unbiased estimator of the population parameter vector, β.

If we measure the variate values of y and each column of X as a (signed) deviation from their respective (arithmetic) mean values, we may obtain $\hat{\beta}_{OLS} = [V_{XX}]^{-1} V_{Xy}$ where V_{XX} is the variance-covariance matrix of X (with itself) and V_{Xy} is the vector of covariances of X *and* y. If v_{rs}, an element of the variance-covariance matrix V_{XX}, is the co-variance of x_r *and* $x_s \in X$, it is given by $v_{rs} = \frac{1}{n}\sum_{i=1}^{n}(x_{ir} - \bar{x}_r)(x_{is} - \bar{x}_s) = (1/n)\sum_{i=1}^{n} x_{ir}x_{is} - \bar{x}_r\bar{x}_s$. If $r = s$, then $v_{rs} = v_{rr} = v_{ss}$ is called the variance (of x_r *or* x_s). The covariance of x_r and y also is defined in the similar manner. This is the conventional view of variance and covariance.

The point of our concern here is that conventionally variance is visualized as the expectation of (squared) deviations of the individual variate values from the *mean value* of the variate. Similarly, covariance (of any two variates) is visualized as the expectation of the product of deviations of the variates concerned from their respective mean values. That is, the variance $v_{rr} = E(x_r - \bar{x}_r)(x_r - \bar{x}_r) = \sum_{i=1}^{n}(x_r - \bar{x}_r)(x_r - \bar{x}_r)p_i$ and, in the similar manner, covariance $v_{rs} = E(x_r - \bar{x}_r)(x_s - \bar{x}_s) = \sum_{i=1}^{n}(x_r - \bar{x}_r)(x_s - \bar{x}_s)p_i$. It is assumed that the probabilities of the occurrence of the squared deviations (as well as the product of deviations) are uniformly constant, or $p_i = 1/n \;\forall\; i = 1, 2, ..., n$. These are the bits of a commonplace knowledge in statistics.

2. Variance as the Expectation of the Product of Inter-individual Differences: Let us look at the variance (and covariance) slightly unconventionally. Covariance of x_r and x_s may be obtained as $v_{rs} = \frac{1}{2n^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir} - x_{jr})(x_{is} - x_{js})\right]$. By expanding the RHS we get

$$v_{rs} = \frac{1}{2n^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}x_{is} + x_{jr}x_{js} - x_{ir}x_{js} - x_{jr}x_{is})\right]$$

$$= \frac{1}{2n^2}\left[\sum_{i=1}^{n}(nx_{ir}x_{is} + \sum_{j=1}^{n}x_{jr}x_{js} - x_{ir}\sum_{j=1}^{n}x_{js} - x_{is}\sum_{j=1}^{n}x_{jr})\right]$$

$$= \frac{1}{2n^2}\left[n\sum_{i=1}^{n}x_{ir}x_{is} + n\sum_{j=1}^{n}x_{jr}x_{js} - \sum_{i=1}^{n}x_{ir}\sum_{j=1}^{n}x_{js} - \sum_{i=1}^{n}x_{is}\sum_{j=1}^{n}x_{jr}\right] = \left[\frac{1}{n}\sum_{i=1}^{n}x_{ir}x_{is} - \bar{x}_r\bar{x}_s\right].$$

Analogous to the expectation interpretation of arithmetic mean $= \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}x_i p_i$: $p_i = 1/n \;\forall\; i = 1, 2, ..., n$, we may reinterpret v_{is}. Denoting the joint probability of occurrence of $(x_{ir} - x_{jr})(x_{is} - x_{js})$ by p_{ij} and assigning a value of $1/(2n^2)$ to it (uniformly for all $i, j = 1, 2, ..., n$) we may consider the covariance of the variates x_r and x_s, (v_{rs}), as the expectation of $(x_{ir} - x_{jr})(x_{is} - x_{js}) = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir} - x_{jr})(x_{is} - x_{js})p_{ij}$. Further, if $r = s$, the same interpretation applies to the variance of x_r (or x_s) as well.

To think aloud, it is *not necessary* to assign the value of $1/(2n^2)$ to all p_{ij} uniformly. As it is done in the case of weighted average where we obtain $\bar{x} = \sum_{i=1}^{n}x_i w_i$: $\sum_{i=1}^{n}w_i = 1$ (and wherein $w_i = p_i$), we may assign different values to p_{ij} with the constraints that $p_{ij} \ge 0$ and $\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij} = 1$. If $p_{ij} = 1/(2n^2) \;\forall\; i, j = 1, 2, ..., n$ we obtain variance (as well as covariance) as obtained by the conventional methods (in the conventional sense). However, if p_{ij} is different than $1/(2n^2)$ we obtain differently weighted variance (as well as covariance).

There is an additional but very important point to be noted. When variance (or covariance) is computed in the conventional sense, permutation of individuals in the sample does not effect on the numerical value of variance (or covariance). This is so due to identical or location-indifferent weight, ($p_{ij} = 1/(2n^2)$), assigned to each and every inter-individual difference such as $(x_{ir} - x_{jr})(x_{is} - x_{js})$ and $(x_{ir} - x_{jr})(y_i - y_j)$. Thus, the order of the individuals in the sample is immaterial. However, when p_{ij} are not assigned identical weights throughout, the value of covariance (or variance) is not impervious to permutation (or reshuffling) of individuals in the sample. The order relationship among the individuals in the sample matters and is important. However, time series data have one-way order and matters are different. In spatial data that characterize two-way order, the matters are much more different.

3. The Spatial Context: Our day-to-day experience suggests that certain variables are *local* in their effects. The influence of such variables is limited within the boundaries of the spatial entity (district) where they are physically located. In contrast, the effects of some other variables are *percolating* or *pervasive* in nature. They permeate through the district boundaries or sometimes grossly transcend the local borders. The intensity of influence of such variable often decreases with an increase in the distance traversed, though the rate of such decay may be slow or rapid. Therefore, the value of the dependent variable observed in district i (say, y_i) may be influenced by the value of an explanatory variable x_r in the district j (say, x_{jr} $i \neq j$).

In the spatial context, therefore, contiguity (interactivity or connectedness) is very important. Any two spatial entities (or districts) are said to be contiguous (to each other) if they have a common boundary or common vertex (or both). In this sense, a spatial entity is always contiguous to itself. In the most simple case we may assign a value of unity to c_{ij} if the spatial entities i and j are contiguous, else $c_{ij} = 0$. Here $c_{ij} \in C(n,n)$, the contiguity matrix that describes the contiguity relationship among the n different spatial entities under consideration. Accordingly, $p_{ij} = c_{ij} / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$.

In the real world, 'connectedness' (interactivity or contiguity) is not a simple binary relationship that may capture the openness of the spatial entities to each other. One may discriminate among the instances of 'interactivity' or 'connectedness' arising due to common vertex and common boundary segments of different magnitudes. There can be several other criteria to measure 'interactivity' or 'connectedness.' In any case, c_{ij} may be assigned a numerical value and accordingly, $p_{ij} = c_{ij} / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$ may be obtained. Once p_{ij} have been obtained, one may compute $v_{rs} = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_{ir} - x_{jr})(x_{is} - x_{js}) p_{ij}$; $r, s = 1, 2, ..., m$ constituting the contiguity (connectedness) weighted variance-covariance matrix with regard to X and, similarly, the contiguity (connectedness) weighted co-variance vector of X *and* y.

At this juncture it is pertinent to note that p_{ij} need not be constant across the variables. It may be perfectly justified to use different p_{ij} for different variables or couplets, such as (x_r, x_s) or (x_r, y). It depends on the nature of variables, since some variables are local and others are pervasive in their effects.

For sake of discrimination now we would denote the contiguity (connectedness) weighted variance-covariance matrix of X by V^*_{XX} and similarly, the co-variance vector of X *and* y will be denoted by V^*_{Xy}. Explicitly,

$$v^*_{rs} \in V^*_{XX} = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_{ir} - x_{jr})(x_{is} - x_{js}) p_{ij} \quad for \quad p_{ij} \neq 1/(2n^2) \quad uniformly.$$

$$v^*_{x_r y} \in V^*_{Xy} = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})(y_i-y_j)p_{ij} \quad for \quad p_{ij} \neq 1/(2n^2) \;\; uniformly.$$

On the other hand, the conventional variance-covariance matrix of X will be denoted by V_{XX} and the co-variance vector of X *and* y will be denoted by V_{Xy}. Explicitly,

$$v_{rs} \in V_{XX} = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})(x_{is}-x_{js})p_{ij} \quad for \quad p_{ij} = 1/(2n^2) \;\; uniformly.$$

$$v_{x_r y} \in V_{Xy} = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ir}-x_{jr})(y_i-y_j)p_{ij} \quad for \quad p_{ij} = 1/(2n^2) \;\; uniformly.$$

4. Interactivity-weighted Regression Coefficients: Now we obtain the contiguity (connectedness or interactivity) weighted regression coefficient vector, $\hat{\beta}^* = (\hat{\beta}^*_1 \quad \hat{\beta}^*_2 \ldots \hat{\beta}^*_m)' = [V^*_{XX}]^{-1}V^*_{Xy}$. Against this, the conventional (OLS) regression coefficient vector is given by $\hat{\beta} = (\hat{\beta}_1 \quad \hat{\beta}_2 \ldots \quad \hat{\beta}_m)' = [V_{XX}]^{-1}V_{Xy}$. It may also be noted that $\hat{\beta} = \hat{\beta}^*$ if $p_{ij} = 1/(2n^2) \;\forall\; i, j = 1, 2, ..., n$. In a sense, $\hat{\beta}$ is only a special case of $\hat{\beta}^*$. In the conventional regression analysis we assume that every entity is connected to every other entity and that too equally, leading to $p_{ij} = 1/(2n^2) \;\forall\; i, j = 1, 2, ..., n$. However, in case of $\hat{\beta}^*$ we assign different values to p_{ij} depending on contiguity or the degree of connectedness. From the viewpoint of interpretation, while the conventional regression coefficients ($\hat{\beta}$) ignore interactions, connectedness or contiguity relations among the spatial the proposed regression coefficients ($\hat{\beta}^*$) incorporate these relations among the spatial entities.

When we estimate $\hat{\beta}$ or $\hat{\beta}^*$ using the variance-covariance matrices (whether V_{XX} and V_{Xy} giving $\hat{\beta}$, or V^*_{XX} and V^*_{Xy} giving $\hat{\beta}^*$), the intercept term, say β_0 remains unestimated. One may obtain the estimated value of β_0 by the well-known relationship, $\hat{\beta}_0 = \bar{y} - (\hat{\beta})'\bar{x}$ and $\hat{\beta}^*_0 = \bar{y} - (\hat{\beta}^*)'\bar{x}$.

5. A FORTRAN Computer Program to Implement the Interaction-weighted Method: We give here the codes of the computer program that may be used to compute the regression coefficients with interaction weights. The inputs to the program are : Y(N), X(N,M) and C(N,N). Here Y is the dependent variable, X are M explanatory variables; Y and X are in N observations; C(N,N) is the matrix of interaction values.

References

1. Fröberg, CE (1965). *Introduction to Numerical Analysis*, Addison-Wesley, London.
2. Krishnamurthy, EV and SK Sen (1976). *Computer-Based Numerical Algorithms*, Affiliated East-West Press, New Delhi.

```
C     MAIN PROGRAM ===============================================
      DOUBLE PRECISION X(100,10),Y(100),E(100),Y0(100),A0(10),A(10)
      DOUBLE PRECISION TMP1,TMP2,XX(10,10),XY(10),V(10,10),W(10,10)
      DOUBLE PRECISION RAND
      INTEGER *2 C(100,100)
      CHARACTER *11 FIL
C     ------------------------------------------------------------
      WRITE(*,*)'DECIDE VALUES OF N, M AND FILE'
      WRITE(*,*)'N=NO. OF OBSERVATIONS; M=NO. OF EXPLANATORY VARIABLES'
      WRITE(*,*)'FIL IS THE NAME OF FILE STORING X(N), y(N) AND C(N,N)'
C     ------------------------------------------------------------
      READ(*,*) N,M,FIL
      OPEN(7,FILE=FIL)
      DO 1 I=1,N
      READ(7,*)(C(I,J),J=1,N)
    1 CONTINUE
      DO 2 I=1,N
      READ(7,*) Y(I),(X(I,J),J=1,M)
    2 CONTINUE
      CLOSE(7)
  100 DO 99 IZ=1,2
      ICI=IZ-1
      DO 7 J=1,M
      XY(J)=0.0
      DO 8 JJ=1,M
      XX(J,JJ)=0.0
      DO 8 I=1,N
      DO 8 II=1,N
      TMP1=X(I,J)-X(II,J)
      TMP2=X(I,JJ)-X(II,JJ)
      TMP=TMP1*TMP2
      IF(ICI.EQ.1) TMP=TMP*C(I,II)
      XX(J,JJ)=XX(J,JJ)+TMP
    8 CONTINUE
      DO 7 I=1,N
      DO 7 II=1,N
      TMP1=X(I,J)-X(II,J)
      TMP2=Y(I)-Y(II)
      TMP=TMP1*TMP2
      IF(ICI.EQ.1) TMP=TMP*C(I,II)
      XY(J)=XY(J)+TMP
    7 CONTINUE
      DO 20 J=1,M
      DO 21 JJ=1,M
   21 XX(J,JJ)=XX(J,JJ)/(N**2)
   20 XY(J)=XY(J)/(N**2)
      NN=1
C     To invert XX Cayley-Hamilton method is used (see Froberg, 1964)
      CALL EIGEN(XX,M,NN,V)
      DO 9 J=1,M
      DO 9 JJ=1,M
      IF(J.NE.JJ) XX(J,JJ)=0.0
      IF((J.EQ.JJ).AND.(XX(J,JJ).GT.1.0D-99)) THEN
      XX(J,JJ)=1.0/XX(J,JJ)
      ELSE
      XX(J,JJ)=0.0
      ENDIF
```

```
    9 CONTINUE
      DO 10 J=1,M
      DO 10 JJ=1,M
      W(J,JJ)=0.0
      DO 10 I=1,M
      W(J,JJ)=W(J,JJ)+V(J,I)*XX(I,JJ)
   10 CONTINUE
      DO 11 J=1,M
      DO 11 JJ=1,M
      XX(J,JJ)=0.0
      DO 11 I=1,M
      XX(J,JJ)=XX(J,JJ)+W(J,I)*V(JJ,I)
   11 CONTINUE
      DO 12 J=1,M
      A(J)=0
      DO 12 JJ=1,M
      A(J)=A(J)+XX(J,JJ)*XY(JJ)
   12 CONTINUE
      WRITE(*,*) 'ICI= ',ICI
      WRITE(*,*)'Coefficients = ',(A(J),J=1,M)
   99 CONTINUE
      END
C     ----------------------------------------------------------------
      SUBROUTINE EIGEN(A,N,NN,V)
C     Adapted from Krisnamurthy & Sen (1976)
      DOUBLE PRECISION A(10,10),V(10,10),W(10,10),P(10)
      DOUBLE PRECISION PMAX,EPLN,TAN,SIN,COS,AI,TT,TA,TB
      DIMENSION MM(10)
C     ------------ INITIALISATION -------------------------
C      WRITE(*,*)'ENTERS EIGEN'
      DO 50 I=1,N
      DO 51 J=1,N
      V(I,J)=0.0
   51 W(I,J)=0.0
      P(I)=0.0
   50 CONTINUE
      PMAX=0
      EPLN=0
      TAN=0
      SIN=0
      COS=0
      AI=0
      TT=0
      EPLN=1.0D-310
C     ------------------------------------------------------
      IF(NN.NE.0) THEN
        DO 3 I=1,N
        DO 3 J=1,N
        V(I,J)=0.0
        IF(I.EQ.J) V(I,J)=1.0
    3   CONTINUE
      ENDIF
    2 NR=0
    5 MI=N-1
      DO 6 I=1,MI
      P(I)=0.0
      MJ=I+1
```

```
      DO 6 J=MJ,N
      IF(P(I).GT.DABS(A(I,J))) GO TO 6
        P(I)=DABS(A(I,J))
        MM(I)=J
    6 CONTINUE
    7 DO 8 I=1,MI
      IF(I.LE.1) GOTO 10
      IF(PMAX.GT.P(I)) GOTO 8
   10 PMAX=P(I)
      IP=I
      JP=MM(I)
    8 CONTINUE
C      EPLN=DABS(PMAX)*1.0D-09
      IF (PMAX.LE.EPLN) THEN
C      WRITE(*,*)'PMAX EPLN',PMAX, EPLN
C      PAUSE'CONVERGENCE CRITERION IS MET'
      GO TO 12
      ENDIF
      NR=NR+1
C     WRITE(*,*)'PMAX, EPLN',PMAX,EPLN
   13 TA=2.0*A(IP,JP)
      TB=(DABS(A(IP,IP)-A(JP,JP))+
     1DSQRT((A(IP,IP)-A(JP,JP))**2+4.0*A(IP,JP)**2))
C       WRITE(*,*) 'TA TB = ',TA,TB
         TAN=TA/TB
C       WRITE(*,*) 'TAN = ',TAN
      IF(A(IP,IP).LT.A(JP,JP)) TAN=-TAN
   14 COS=1.0/DSQRT(1.0+TAN**2)
      SIN=TAN*COS
      AI=A(IP,IP)
      A(IP,IP)=(COS**2)*(AI+TAN*(2.0*A(IP,JP)+TAN*A(JP,JP)))
      A(JP,JP)=(COS**2)*(A(JP,JP)-TAN*(2.0*A(IP,JP)-TAN*AI))
      A(IP,JP)=0.0
      IF(A(IP,IP).GE.A(JP,JP)) GO TO 15
      TT=A(IP,IP)
      A(IP,IP)=A(JP,JP)
      A(JP,JP)=TT
      IF(SIN.GE.0) GO TO 16
      TT=COS
      GO TO 17
   16 TT=-COS
   17 COS=DABS(SIN)
      SIN=TT
   15 DO 18 I=1,MI
      IF(I-IP) 19, 18, 20
   20 IF(I.EQ.JP)GO TO 18
   19 IF(MM(I).EQ.IP) GO TO 21
      IF(MM(I).NE.JP) GO TO 18
   21 K=MM(I)
      TT=A(I,K)
      A(I,K)=0.0
      MJ=I+1
      P(I)=0.0
      DO 22 J=MJ,N
      IF(P(I).GT.DABS(A(I,J))) GO TO 22
      P(I)=DABS(A(I,J))
      MM(I)=J
```

```
   22 CONTINUE
      A(I,K)=TT
C      WRITE(*,*)'IN LOOP 18'
   18 CONTINUE
      P(IP)=0.0
      P(JP)=0.0
      DO 23 I=1,N
      IF(I-IP) 24, 23, 25
   24 TT=A(I,IP)
      A(I,IP)=COS*TT+SIN*A(I,JP)
      IF(P(I).GE.DABS(A(I,IP))) GO TO 26
      P(I)=DABS(A(I,IP))
      MM(I)=IP
   26 A(I,JP)=-SIN*TT+COS*A(I,JP)
      IF(P(I).GE.DABS(A(I,JP))) GO TO 23
   30 P(I)=DABS(A(I,JP))
      MM(I)=JP
      GO TO 23
   25 IF(I.LT.JP) GO TO 27
      IF(I.GT.JP) GO TO 28
      IF(I.EQ.JP) GO TO 23
   27 TT=A(IP,I)
      A(IP,I)=COS*TT+SIN*A(I,JP)
      IF(P(IP).GE.DABS(A(IP,I))) GO TO 29
      P(IP)=DABS(A(IP,I))
C     SEE THIS IS ONE OR I
      MM(IP)=I
   29 A(I,JP)=-TT*SIN+COS*A(I,JP)
      IF(P(I).GE.DABS(A(I,JP))) GO TO 23
      GO TO 30
   28 TT=A(IP,I)
      A(IP,I)=TT*COS+SIN*A(JP,I)
      IF(P(IP).GE.DABS(A(IP,I))) GO TO 31
      P(IP)=DABS(A(IP,I))
      MM(IP)=I
   31 A(JP,I)=-TT*SIN+COS*A(JP,I)
      IF(P(JP).GE.DABS(A(JP,I))) GO TO 23
      P(JP)=DABS(A(JP,I))
      MM(JP)=I
   23 CONTINUE
      IF(NN.EQ.0) GOTO 7
      DO 32 I=1,N
      TT=V(I,IP)
      V(I,IP)=TT*COS+SIN*V(I,JP)
      V(I,JP)=-TT*SIN+COS*V(I,JP)
   32 CONTINUE
      GO TO 7
   12 RETURN
      END
```