Estimation of Production Possibility Surface from Empirical Data

SK Mishra
Dept. of Economics
NEHU, Shillong, India

1. Introduction

A production possibility surface is the frontier surface determined by the available resources in an economy and the technology applied for production. An efficient utilization of resources is presumed. The surface is in the non-negative orthant, meaning thereby that neither the resources nor the quantity produced may take on a negative value. Mathematically, the shape of the surface is ellipsoidal. In a linear programming formulation of determining the efficient resource allocation for optimal production, this surface is approximated by a polyhedron in the non-negative orthant.

It is interesting to fit a production possibility surface to empirical data on two accounts : first, it is an example of a nonlinear surface fitting that cannot appreciably be approximated by a simple linearization method, and second, that it is an example of the surface that cannot ordinarily be fitted by most of the conventional non-linear regression algorithms. These algorithms fail mostly because they approach the best fitting surface from both sides, interior and exterior of the surface, and thereby land themselves into the problem of finding square root of a negative quantity.

2. The nature of the surface

The production possibility surface is a segment of an ellipsoid with nonnegative values, all lying in the nonnegative orthant. Algebraically, it may be expressed as

$$\sum_{i=1}^{m}(x_i/a_i)^2 = K^2;\ \ x_i \geq 0;\ \ a_i > 0;\ \ K > 0;\ \ m \geq 2 \qquad \ldots (1).$$

3. Estimation of the ellipsoidal surface from sample data

Empirical data may be collected on $x_i = (x_{ji};\ \ j = 1, 2, ..., n)$ such that

$$\begin{aligned} x_1 &= (x_{11}, x_{12}, ..., x_{n1}) \\ x_2 &= (x_{12}, x_{22}, ..., x_{n2}) \\ &.\ \ .\ \ .\ \ .\ \ .\ \ ....\ \ . \\ x_m &= (x_{1m}, x_{1m}, ..., x_{nm}) \end{aligned} \qquad \ldots (2)$$

where n (n > m) is the number of observations in the sample. Based on the information in the sample, it is required to fit the ellipsoidal function as in (1) above that amounts to estimation of a_i $(i = 1, 2, ..., m)$, such that the distance between the observed points and the estimated points is minimized. The distance to be minimized may be suitably defined. Usually, one of the variables is chosen as dependent variable and the others are considered as explanatory variables. Then the vertical distance between the observed and the expected values of the dependent variable is minimized. However, one may choose to minimize the horizontal distance or the distance defined on any chosen angle to the normal of the ellipsoid at any point (see Kmenta, 1971 pp. 317-319 and Fröberg, 1965 pp. 280-281). Depending on certain considerations such as the existence of outliers in the sample, one may choose absolute distance to be minimized (Dasgupta and Mishra, 2004). To minimize the absolute distance one may use the iterative estimation procedure suggested by Abdelmalek (1974), Schlossmacher (1973) or Tylor (1974).

Unfortunately, the usual procedure in which one chooses a metric (e.g. L=1 for absolute and L= 2 for least squares or Euclidean) and defines the loss function such as

$$S=\sum_{j=1}^{n}\left|x_{jq}-\{K^2-\sum_{i=1}^{m}(x_{ji}/a_i)^2\}^{1/2}\right|^{L};\ i\neq q \qquad \ldots(3)$$

which is minimized with respect to a_i fails to operate. It is required, therefore, that estimation of the parameters of an ellipsoid is attempted differently.

In a simple bivariate case (ellipse), given n points $p_i=(x_i,y_i);\ i=1,2,...,n$ and the generic implicit equation $f(p,a)=a_1x^2+a_2xy+a_3y^2+a_4x+a_5y+a_6=0$, one may attempt to obtain a solution vector $\hat{a}$ such that it minimizes the distance between the observed and the estimated points.

4. A brief account of previous works

Statistical fitting of ellipse (or ellipsoid) to empirical data remained unattractive for a long time. However, in the last two decades or so, it attracted many researchers especially to solve the pattern recognition problems. Bookstein (1979) introduced a very general conic fitting method and showed that under a quadratic constraint on the parameters, the least squares method can be formulated as a problem of solving the rank-deficient generalized eigenvalue system. Based on Bookstein's method, Sampson (1982) developed an *iterative* method yielding better approximation. Taubin (1991) attempted to improve upon Bookstein's and Sampson's methods by formulating the least squares fitting method as the generalized eigen-system problem. Rosin (1993) and Gander et al. (1994) used conic fitting methods under linear constraints to fit an ellipse.

Kanatani (1994) showed how conic fitting methods yield estimates that are statistically biased. Pilu (1996) proposed a direct least squares fitting method specific to ellipses. He showed that the method works fast, fits an ellipse quite accurately and is also extremely robust to noise. Halir and Flusser (1998) present a numerically stable non-iterative algorithm for fitting an ellipse to a set of data points. The approach is based on a least squares method and it guarantees an ellipse-specific solution for scattered or noisy data. The optimal solution is computed directly. This leads to a simple, stable and robust fitting method, which can be easily implemented with a great efficiency. Matei and Meer (2000) proposed an improved maximum likelihood estimator for ellipse fitting based on the heteroscedastic errors-in-variables regression algorithm. The technique significantly reduces the bias of the parameter estimates present in the direct least squares method, while it is numerically more robust than renormalization, and requires less computations than minimizing the geometric distance with the Levenberg-Marquardt optimization procedure.

5. Estimation of parametric equations of the ellipsoid

In this paper we propose a new, simple iterative method for our limited purpose at hand. We have tried to estimate a_i of the ellipsoid by fitting the parametric equations to the empirical data. The parametric equations of the ellipsoid (m=2) are given as

$$x_{ji} = a_1 \sin(\theta_j);\;\; x_{j2} = a_2 \cos(\theta_j) \qquad \ldots(4)$$

where θ_j is the eccentric angle with reference to the point (x_{j1}, x_{j2}). The numerical value of $\cos(\theta_j)$ at point $(x_{j\alpha}, x_{j\beta})$ of the ellipse (ellipsoid for m=2) is given by $x_{j\beta}/R$ (where α and β are the axes along which the minor and the major of the ellipse lie respectively and R is the radius of the circle surrounding the ellipse such that R is not smaller than the major of the ellipse). At each iteration, once $\sin(\theta_j)$ and $\cos(\theta_j)$ for all j are available, a_1 and a_2 may be estimated by the method of least squares (LS) or least absolutes (LA). The value of K may suitably be normalized to unity without affecting the relationship. In successive iterations, R shrinks gradually, and finally converges to become approximately equal to the major of the ellipse. Estimation formulas (obtained by the method of LS or LA) at each iteration are given as

$$\hat{a}_1 = [\sum_{j=1}^{n}\{x_{j1}\sin(\theta_j)\}]/[\sum_{j=1}^{n}\{\sin(\theta_j)^2\}];\;\; j = 1,2,...,n$$
$$\hat{a}_2 = [\sum_{j=1}^{n}\{x_{j2}\cos(\theta_j)\}]/[\sum_{j=1}^{n}\{\cos(\theta_j)^2\}];\;\; j = 1,2,...,n \qquad \ldots(5)$$

At each iteration, with R shrinking closer to the major of the ellipse, $\hat{K}$, $\hat{a}_1$ and $\hat{a}_2$ improve such as to minimize the sum of the squared (or absolute) error.

A similar procedure may be adopted in case of an ellipsoid in three or more dimensions. For example, the equation of an ellipsoid in three dimensions is given as in (1) above for m=3. The parametric equations are given by

$$x_{j1} = a_1 \cos(\theta_j)\sin(\phi_j);\;\; x_{2j} = a_2 \sin(\theta_j)\sin(\phi_j);\;\; x_{j3} = a_3 \cos(\phi_j) \qquad \ldots(6)$$

Here also, using the method of LS or LA, the parameters may be estimated as

$$\hat{a}_1 = [\sum_{j=1}^{n}\{x_{j1}\cos(\theta_j)\sin(\phi_j)\}]/[\sum_{j=1}^{n}\{\cos(\theta_j)\sin(\phi_j)\}^2]$$
$$\hat{a}_2 = [\sum_{j=1}^{n}\{x_{j2}\sin(\theta_j)\sin(\phi_j)\}]/[\sum_{j=1}^{n}\{\sin(\theta_j)\sin(\phi_j)\}^2] \qquad \ldots(7)$$
$$\hat{a}_3 = [\sum_{j=1}^{n}\{x_{j3}\cos(\phi_j)\}]/[\sum_{j=1}^{n}\{\cos(\phi_j)\}^2]$$

6. Numerical examples

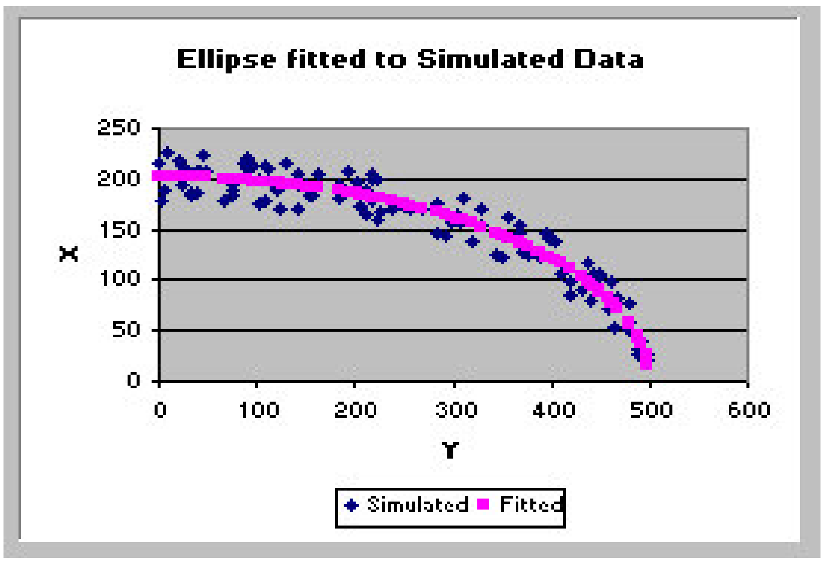
Twenty-one observations on x and y with elliptical relations $x_i = a_1[K^2 - (y_i/a_2)^2]^{1/2}$ are used here as a test example for the numerical exercise. Data were generated with the parameters specified as $K = 10.2;\; a_1 = 1;\; a_2 = 2.9$. Small random errors (u) were added to x.

Vertical distance on x has been minimized by the method of LS and LA. The results are given in table 1. The results are very close to the parameters. The estimated parameters as well as the variables by LS and LA estimators are very close since the errors have no outliers.

As an additional illustration to ellipse fitting, 100 pairs of (x, y) were generated and random errors were added to x as described in the BASIC program given below.

```
10 N=100:A=2:B=5:K=100:DIM X(N),Y(N)
20 RANDOMIZE:OPEN"O",#1,"el1":PRINT "Data stored in file el1"
30 FOR I=1 TO N:S=RND(S):Y(I)=B*K*S
40 X(I)=A*(K^2-(Y(I)/B)^2)^.5 +(RND(S)-.5)*50
50 PRINT#1, USING "#####.#####"; X(I),Y(I):NEXT I:CLOSE:END
```

With randomization seed = 7711 data were generated and the ellipse was estimated. The simulated and estimated values are plotted on the graph below.



Some experimental results of LS and LA fitting of ellipse to simulated data are presented in tables 2 and 3. The results indicate that the fits are effective and consistent.

7. Concluding remarks

Statistical fitting of certain surfaces, if tried directly with the algebraic expression thereof, may be problematic. Equations of asteroid, cycloid, hyperbola, spirals (see Mishra, 2004), etc. are the typical examples that would be difficult to fit to empirical data directly. Although one may always fit an approximating polynomial of a desired order in any kind of data, but the interpretation of the fitted polynomial may not be easy or at times, possible. Therefore, such curves (surfaces) may be fitted more conveniently if they are transformed into their respective parametric equations. The parameters θ, ϕ, etc. may be much more conveniently interpreted as they relate to simple ratios of the variables.

References

Abdelmalek, NN (1974) "On Discrete Linear L_1 Approximation of L_1 Solution of Over identified Linear Equations." *J. of Approximation Theory*, 11, 38-53.

Bookstein, FL (1979) "Fitting conic sections to scattered data." *Computer Graphics and Image Processing*, 9, 56-71.

Dasgupta, M and SK Mishra (2004) "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review." *Social Science Research Network (SSRN).* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=552502

Fröberg, CE (1965) *Introduction to Numerical Analysis*. Addison Wesley, London.

Gander, W, GH Golub and R Strebel (1994) "Least-squares Fitting of Circles and Ellipses." *BIT*, 43, 558-578.

Halir, R and Flusser, J (1998) "Numerically Stable Direct Least Squares Fitting of Ellipses." in Skala, V (ed.) *Proc. Int. Conf. in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*. 125–132.

Kanatani, K (1994) "Statistical Bias of Conic Fitting and Renormalization." *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 16(3), 320-326.

Kmenta, J (1971) *Elements of Econometrics*. Macmillan, New York.

Matei, B and P Meer (2000) "Reduction of Bias in Maximum Likelihood Ellipse Fitting." At http://citeseer.ist.psu.edu/matei00reduction.html

Mishra, SK (2004) "An Algorithm for Fitting Archimedean Spiral to Empirical Data." *SSRN* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=531542

Pilu, M (1996) "Direct Least Squares Fitting of Ellipses." Working Paper, available at http://www. portal.acm.org/citation.cfm?id=302950

Rosin, PL (1993) "A Note on the Least Squares Fitting of Ellipses." *Pattern Recognition Letters*, 14, 799-808.

Sampson, PD (1982) "Fitting Conic Sections to very Scattered Data: An Iterative Refinement of the Bookstein Algorithm." *Computer Graphics and Image Processing*, 18, 97-108.

Schlossmacher, EJ (1973) "An Alternative Technique for Absolute Deviation Curve Fitting." *J. of the American Statistical Association*, 68, 857-859.

Taubin, G (1991) "Estimation of Planar Curves, Surfaces and Non-planar Space Curves defined by Implicit Equations, with Applications to Edge and Range Image Segmentation." *IEEE PAMI*, 13(11), 1115-1138.

Tylor, LD (1974) "Estimation my Minimizing the Sum of Absolute Errors." in Zarembka, P (ed) *Frontiers of Econometrics*, Academic Press, New York.

Table 1. Test data and Estimated Ellipse (Production Possibility Frontier)								
	Test data			Min Absolute distance			Min Euclidean distance	
Sl no.	x	y		$\hat{x}_{LA}$	$\hat{e}_{LA}$		$\hat{x}_{LS}$	$\hat{e}_{LS}$
1	10.10	5.00		10.04992	0.05008		10.05684	0.04316
2	10.00	6.00		9.98496	0.01504		9.99159	0.00841
3	9.90	7.00		9.90763	-0.00763		9.91394	-0.01394
4	9.80	8.00		9.81765	-0.01765		9.82357	-0.02357
5	9.70	9.00		9.71467	-0.01467		9.72013	-0.02013
6	9.60	10.00		9.59827	0.00173		9.60321	-0.00321
7	9.50	11.00		9.46794	0.03206		9.47231	0.02769
8	9.30	12.00		9.32312	-0.02312		9.32683	-0.02683
9	9.20	13.00		9.16310	0.03690		9.16607	0.03393
10	9.00	14.00		8.98709	0.01291		8.98924	0.01076
11	8.80	15.00		8.79411	0.00589		8.79534	0.00466
12	8.60	16.00		8.58303	0.01697		8.58323	0.01677
13	8.30	17.00		8.35247	-0.05247		8.35151	-0.05151
14	8.10	18.00		8.10077	-0.00077		8.09851	0.00149
15	7.80	19.00		7.82589	-0.02589		7.82216	-0.02216
16	7.50	20.00		7.52529	-0.02529		7.51989	-0.01989
17	7.20	21.00		7.19575	0.00425		7.18842	0.01158
18	6.80	22.00		6.83307	-0.03307		6.82351	-0.02351
19	6.40	23.00		6.43167	-0.03167		6.41945	-0.01945
20	6.00	24.00		5.98375	0.01625		5.96830	0.03170
21	5.50	25.00		5.47790	0.02210		5.45838	0.04162
	Parameter K = 10.20; A = 2.90			K=10.19603; A=2.90714 Sum(abs(error)) = 0.44642			K=10.20356; A=2.89995 Sum(error^2) = 0.034997	

Table 2. Experimental Results of Ellipse fitting by LS					
Seed = 3311 u~N(0,10) Experiments=100 Parameters K=200; a_2=B=2.5	Sample Size (n)	$\hat{K}$	$\hat{\sigma}(\hat{K})$	$\hat{B}$	$\hat{\sigma}(\hat{B})$
	15	199.9196	2.7743	2.4978	0.04444
	30	200.1236	1.9314	2.5010	0.02586
	60	200.1352	1.7288	2.4985	0.02259
	100	200.0184	1.2469	2.4997	0.01692

Table 3. Experimental Results of Ellipse fitting by LA					
Seed = 3311 u~N(0,10) Experiments=100 Parameters K=200; a_2=B=2.5	Sample Size (n)	$\hat{K}$	$\hat{\sigma}(\hat{K})$	$\hat{B}$	$\hat{\sigma}(\hat{B})$
	15	200.0877	2.9247	2.4922	0.04839
	30	200.1107	1.9601	2.5014	0.02696
	60	200.0977	1.7392	2.4999	0.02328
	100	200.0063	1.2342	2.5001	0.01670

```
C     MAIN PROGRAM FOR FITTING AN ELLIPSE TO EMPIRICAL DATA
      DOUBLE PRECISION X(100),Y(100)
      DOUBLE PRECISION EPSILON
      CHARACTER *15 FI,FO
      WRITE(*,*)'INPUT FILE TO READ X AND Y DATA ?'
      READ(*,*) FI
      WRITE(*,*)'OUTPUT FILE TO STORE RESULTS'
      READ(*,*) FO
      WRITE(*,*)'NUMBER OF OBSERVATIONS ?'
      READ(*,*) N
      OPEN(7,FILE=FI)
      DO 1 I=1,N
    1 READ(7,*) X(I),Y(I)
      CLOSE(7)
      WRITE(*,*)'DISTANCE TO MINIMIZE ? ABSOLUTE(1), EUCLIDEAN(2)'
      READ(*,*) L
      OPEN(7,FILE=FO,STATUS='NEW')
      NO=7
      CALL ELLIPSE(N,X,Y,L,FO,NO)
      CLOSE(7)
      WRITE(*,*)'RESULTS STORED IN OUTPUT FILE = ',FO
      END
C     ------------------------------------------------------------
      SUBROUTINE ELLIPSE(N,X,Y,L,FO,NO)
      DOUBLE PRECISION X(100),Y(100),YC(100),XCOS(100),YSIN(100)
      DOUBLE PRECISION ANGLE(100),YEXP(100),EPSILON,DELTA
      DOUBLE PRECISION MAXX,MAXY,MINY,R,AAA,FX,SX,SXY,P1,P2,SY,SYX
      DOUBLE PRECISION K1,B1,S,SP,D
      INTEGER CHNG
      CHARACTER DST *9, FO *15
      EPSILON=1.0D-09
      MAXIT=100000
      CHNG=0
      DST='SQUARE'
      IF(L.EQ.1) DST='ABSOLUTE'
C     FIND MAX(X), MAX(Y), AND MAX(MAX(X),MAX(Y))
  240 MAXX=X(1)
      MAXY=Y(1)
      MINY=Y(1)
      DO 1 I=2,N
      IF(X(I).GT.MAXX) MAXX=X(I)
      IF(Y(I).GT.MAXY) MAXY=Y(I)
      IF(Y(I).LT.MINY) MINY=Y(I)
    1 CONTINUE
      R=MAXX
      IF(MAXX.LT.MAXY) R=MAXY
      IF((R.EQ.MAXX).AND.(CHNG.EQ.0)) THEN
      WRITE(NO,*)'MAJOR LIES ON THE X AXIS'
      ENDIF
      IF(R.EQ.MAXX) GOTO 400
      WRITE(NO,*)'MAJOR LIES ON THE Y AXIS'
      WRITE(NO,*)'Y AND X ARE NOW EXCHANGED'
      DO 2 I=1,N
      AAA=X(I)
      X(I)=Y(I)
      Y(I)=AAA
```

```
    2 CONTINUE
      CHNG=1
      GOTO 240
  400 R=(MAXX/(1.0-MINY/MAXY))*1.2
      DELTA=0.05*R
C     R IS THE RADIUS OF THE CIRCLE THAT INCLUDES THE MAJOR
C     TO ALL GIVEN VALUES OF X FIND YC ON CIRCLE WITH RADIUS R
      SP=-10000.0
      ITER=0
  460 ITER=ITER+1
      DO 3 I=1,N
      FX=R**2-X(I)**2
      IF(FX.LT.0.0) GOTO 800
      YC(I)=DSQRT(R**2-X(I)**2)
    3 CONTINUE
      DO 4 I=1,N
      ANGLE(I)=DATAN(YC(I)/X(I))
      XCOS(I)=DCOS(ANGLE(I))
      YSIN(I)=DSIN(ANGLE(I))
    4 CONTINUE
      SX=0.0
      SXY=0.0
      DO 5 I=1,N
      SX=SX+XCOS(I)**2
      SXY=SXY+XCOS(I)*X(I)
    5 CONTINUE
      P1=SXY/SX
      SY=0.0
      SYX=0.0
      DO 6 I=1,N
      SY=SY+YSIN(I)**2
      SYX=SYX+YSIN(I)*Y(I)
    6 CONTINUE
      P2=SYX/SY
      K1=P2
      B1=P1/P2
      S=0.0
      DO 7 I=1,N
      YEXP(I)=DSQRT(DABS(K1**2-(X(I)/B1)**2))
      IF(L.EQ.1) S=S+DABS(Y(I)-YEXP(I))
      IF(L.EQ.2) S=S+(Y(I)-YEXP(I))**2
    7 CONTINUE
      IF(SP.LT.0.0) GOTO 820
      IF(S.LT.SP) GOTO 820
      IF(DABS(S-SP).LT.EPSILON) GOTO 860
  800 R=R+DELTA
      DELTA=DELTA/2.0
      R=R-DELTA
      IF(ITER.LT.MAXIT) GOTO 460
  820 SP=S
      R=R-DELTA
      IF(ITER.LT.MAXIT) GOTO 460
      WRITE(NO,*)'DID NOT CONVERGE IN',MAXIT,' ITERATIONS.'
      WRITE(NO,*)'MOST APPROXIMATE VALUES ARE'
      GOTO 870
  860 WRITE(NO,*)'CONVERGENCE REACHED. FINAL VALUES ARE'
      WRITE(NO,*)'ITERATIONS PERFORMED = ',ITER
```

```
870 IF(CHNG.EQ.0.0) THEN
    WRITE(NO,*)'K=',K1,' AND B=',B1
    WRITE(NO,*)'Y, X, EXPECTED Y AND EXPECTED ERRORS ARE :'
    ENDIF
    IF(CHNG.EQ.1.0) THEN
    WRITE(NO,*)'K=',K1,' AND A=',B1
    WRITE(NO,*)'X, Y, EXPECTED X AND EXPECTED ERRORS ARE :'
    ENDIF
    DO 8 I=1,N
    D=Y(I)-YEXP(I)
    WRITE(NO,100) Y(I),X(I),YEXP(I),D
  8 CONTINUE
    WRITE(NO,*)'SUM OF ',DST,' OF DEVIATIONS =',S
    WRITE(NO,*)'-----------------------------------------------'
    WRITE(NO,*)'FORMULA FOR ELLIPSE ESTIMATED HERE IS '
    IF(CHNG.EQ.0) THEN
    WRITE(NO,*)'Y(I)=SQRT(K**2-(X(I)/B)**2)'
    WRITE(NO,*)'K=',K1,' AND B=',B1
    ENDIF
    IF(CHNG.EQ.1) THEN
    WRITE(NO,*)'X(I)=SQRT(K**2-(Y(I)/A)**2)'
    WRITE(NO,*)'K=',K1,' AND A=',B1
    ENDIF
    WRITE(NO,*)'-----------------------------------------------'
100 FORMAT(1X,4G15.8)
    RETURN
    END
```