THE QUADRATIC TRANSPORTATION PROBLEM AS A MODEL OF INTERREGIONAL MIGRATION PATTERNS

Waldo R. Tobler
University of California,
Santa Barbara

December 1982
CP-82-84

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Spatial interaction models have played an important role in two tasks in the Human Settlements and Services Area. In the Public Facilities Location Task they have been used to represent the locational behavior of establishments and households. In the Urban Change Task they have been used to describe internal migration patterns. In this paper, Waldo Tobler introduces a new spatial interaction model and outlines some of its properties. Variants of the basic model are noted and a computer listing is provided for readers wishing to explore the usefulness of the model as a descriptor of movement patterns.

Andrei Rogers
Chairman
Human Settlements
and Services Area

ABSTRACT

In the Quadratic Transportation Problem the objective is to minimize the functional

$$\sum_i \sum_j x_{ij}^2 c_{ij}$$

subject to the constraints

$$\sum_j x_{ij} = O_i \quad , \quad \sum_i x_{ij} = I_j \quad , \quad x_{ij} \geq 0$$

Here we interpret x_{ij} as the quantity of movement (migrants, commuters, trade, telephone calls, etc.) between places i and j during a given interval of time. The transport disutility or cost is labeled c_{ij} and is assumed to be known. The problem solution is $x_{ij} = (\alpha_i + \beta_j)/c_{ij}$, and the Lagrangians can be interpreted as estimates of shadow prices. Variants of the basic model are noted and competing spatial interaction models are cited. The model is tested using empirical data on the visitation of persons to a set of recreational facilities. A computer program listing is provided.

CONTENTS

THE QUADRATIC TRANSPORTATION PROBLEM AS A MODEL OF INTERREGIONAL MIGRATION PATTERNS

INTRODUCTION

Spatial interaction models purport to describe the amount of movement between sets of places. There are many such models, each with many variants, and the literature is extensive. In this short report I introduce a "new" spatial interaction model and outline some of its properties. Whether, or when (under which circumstances), this model should supersede those now in use will need to be decided by the reader. To introduce the subject a well-known model is described.

1. THE LINEAR TRANSPORTATION PROBLEM (L.T.P.)

The objective function is to

$$\text{Min:} \quad \sum_i \sum_j x_{ij} c_{ij} \qquad \begin{matrix} i = 1,\ldots,R \\ j = 1,\ldots,C \end{matrix}$$

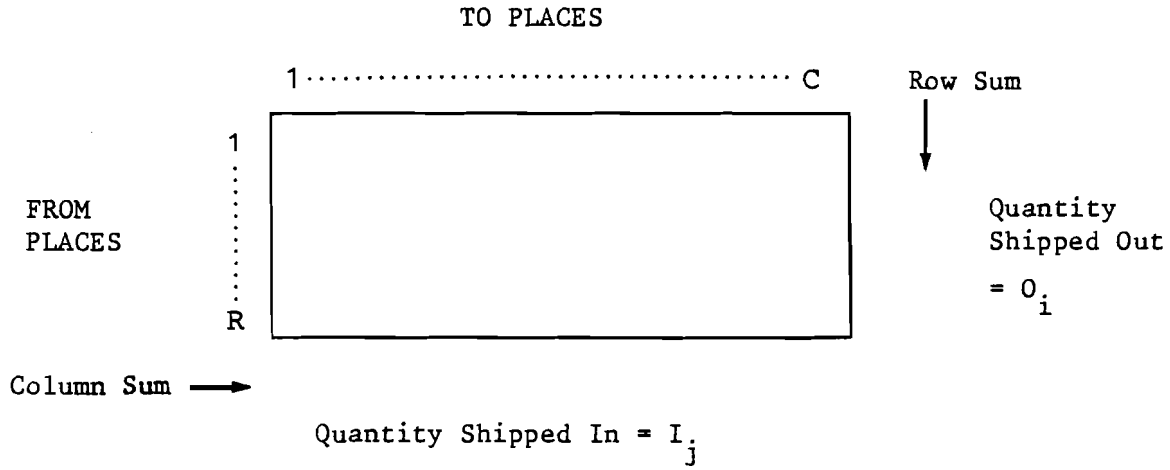
$$\text{Subject to:} \quad \sum_j x_{ij} = O_i \qquad j = 1,\ldots,C$$

$$\sum_i x_{ij} = I_j \qquad i = 1,\ldots,R$$

$$x_{ij} \geq 0$$

This problem has a dual, whose variables are normally interpreted as shadow prices, and several variants, of which the transshipment problem is worthy of citation in the present context. Further detail is not required since the L.T.P. is well known. Recall, however, that the number of non-zero x_{ij} in the solution does not exceed $C + R - 1$, and that these values will be integers if the marginal sums O_i and I_j are integers.

The L.T.P. can be laid out in the form of a rectangular table, as follows:



In addition to the known marginal totals O_i and I_j, the transport cost quantities c_{ij} are also given. The solution to the L.T.P. consists in finding the entries x_{ij} in the box to satisfy the objective function.

The important point is that virtually all of the current spatial interaction models can be cast in this same rectangular table format, and with virtually identical constraints. What distinguishes one model from another is the objective function. Several possibilities are given in the ensuing paragraphs. The context of the study should enable one to decide between various objective functions. Whether the point is to obtain a realistic description of natural events or an idealistic (normative) one will influence this decision.

2. THE QUADRATIC TRANSPORTATION PROBLEM (Q.T.P.)

2.1 The Problem

The objective function is to

$$\text{Min:} \quad \sum_i \sum_j x_{ij}^2 c_{ij} \qquad \begin{matrix} i = 1,\ldots,R \\ j = 1,\ldots,C \end{matrix}$$

subject to the previous constraints. With Lagrangians this becomes

$$\varepsilon^2 = \sum_i \sum_j x_{ij}^2 c_{ij} + \sum_i \alpha_i \left(O_i - \sum_j x_{ij} \right) + \sum_j \beta_j \left(I_j - \sum_i x_{ij} \right)$$

Setting the appropriate derivatives to zero yields

$$x_{ij} = \frac{1}{2}(\alpha_i + \beta_j) / c_{ij} \qquad c_{ij} > 0$$

and

$$\alpha_i \sum_j \frac{1}{c_{ij}} + \sum_j \frac{\beta_j}{c_{ij}} = 2\, O_i$$

$$\sum_i \frac{\alpha_i}{c_{ij}} + \beta_j \sum_i \frac{1}{c_{ij}} = 2\, I_j$$

These last two equations are easily written as a single simple matrix equation, of rank $R + C - 1$, and directly solvable. Once the Lagrangians are known the x_{ij} can be computed. This derivation does not consider the non-negativity constraint which must be taken into account by means fully discussed in Dorigo and Tobler (1983).

Properties which distinguish the solution to the Q.T.P. from that of the L.T.P. are that

a) The x_{ij} are on average smaller numbers. This is forced by the quadratic term in the objective function.

b) The number of non-zero x_{ij} will exceed $R + C - 1$, and will approach RC.

c) The x_{ij} are generally not integers.

Properties a) and b) are more in accord with empirical spatial interaction tables than are the solutions to the L.T.P. This is expected because commodity flows are rendered more reliable by a diversity of sources, urban traffic is diverted to avoid congestion, and migration patterns are rendered diffuse due to information inadequacies. Spatial allocation models which use the L.T.P. thus yield results which are less realistic than can be obtained through the use of the Q.T.P. solution. Property c) of the L.T.P. is desirable, however, and suggests investigation of an Integer Q.T.P.

2.2 Variants on the Q.T.P.

It is obvious that a Quadratic Transshipment Problem can be formulated, with solution procedures modeled on those of the linear case. This can be given an interesting interpretation. Let b_{ij} be the length of the border between regions i and j. Then the objective function

$$\text{Min:} \quad \sum_i \sum_j x_{ij}^2 / b_{ij} \qquad b_{ij} > 0$$

represents a problem in which the square of the flux across these borders is minimized. Now most regions in the domain of interest will not have common boundaries, $b_{ij} \equiv 0$. In order to satisfy the constraints it may be necessary to "transship" entities through adjoining regions. Actual routing of movements can thus be modeled.

Alternately consider objective functions as follows

2.2.1 $$\text{Min:} \quad \sum_i \sum_j x_{ij}^2 c_{ij} / O_i I_j$$

or

2.2.2 $$\text{Min:} \quad \sum_i \sum_j x_{ij}^2 c_{ij} / P_i P_j$$

The second of these yields

$$x_{ij} = \frac{1}{2}(\alpha_i + \beta_j)\frac{P_i P_j}{c_{ij}}$$

as solution, neglecting the non-negativity constraint (easily added, as before). This is recognized as a variant form of the "spatial gravity model", as widely used in Geography, Regional Science, and related fields. The variant 2.2.1 requires less data (the P_i, P_j are "populations" of the source or destination regions). Both of these variants can be interpreted as incorporating "economies of scale" into the transportation system by allowing the magnitude of the movements to influence the transportation cost structure. Further details on these Q.T.P. model variants can be found in Tobler (1983).

The spatially continuous version of the basic Q.T.P. requires minimization of the double integral

$$\int\int_{\mathscr{R}} \left[\frac{\partial\alpha}{\partial x}^2 + \frac{\partial\alpha}{\partial y}^2 + \frac{\partial\beta}{\partial x}^2 + \frac{\partial\beta}{\partial y}^2 + \lambda^2(\alpha+\beta)^2 + 2\gamma(\alpha+\beta)\right]dxdy$$

with $\frac{\partial\alpha}{\partial\eta} = \frac{\partial\beta}{\partial\eta} = 0$ as the Neumann condition on the boundary. The solution to this least squares problem is

$$\nabla^2\alpha = I - 4(\alpha + \beta)$$

$$\nabla^2\beta = O - 4(\alpha + \beta)$$

where $\alpha(x,y)$ and $\beta(x,y)$ are differentiable spatial functions and $O(x,y)$ and $I(x,y)$ are spatially continuous source and sink density functions. This is a coupled system of simultaneous linear partial differential (Helmholz) equations. Observe that this solution yields two-way flows, continuously routed from one place to another through adjacent places. Subtraction of one equation from the other yields Poisson's equation with the difference between the in and out, i.e., the amount of change at a place, as the driver. The solution of this single partial

differential equation can then be represented as a spatially varying attractivity field or potential, with flows shown as a curl-free vector field; see Tobler (1981) for examples. Addition of the two equations yields a single P.D.E. for the gross movements with similar properties.

3. OTHER SPATIAL INTERACTION MODELS

Most widely used today is the entropy model (Wilson 1967):

$$\text{Max:} \qquad -\sum_i \sum_j x_{ij} \ln x_{ij}$$

where the following condition

$$\sum_i \sum_j x_{ij} c_{ij} = D$$

is added to the previous constraints. D is a quantity which is (rather unrealistically) assumed known a priori. This model has as solution

$$x_{ij} = \alpha_i \beta_j O_i I_j \exp(-\gamma c_{ij})$$

Here the Lagrangians enter in multiplicative form, not in the simpler additive form of the Q.T.P. These two models (and some others) are applied to empirical migration data in Tobler (1983), with the Q.T.P. yielding a very slightly better fit to the data than the entropy model. In the migration case R = C and the interaction table is square, but this does not add complexity. Of course a separate analysis may be undertaken for each migrant category or age group.

It is perhaps of interest to consider an even simpler model, namely

$$\text{min:} \qquad \sum_i \sum_j x_{ij}^2$$

$$\text{subject to:} \quad \sum_j x_{ij} = O_i \ , \ \sum_j x_{ij} = I_j \ , \ x_{ij} \geq 0$$

$$\sum_i \sum_j x_{ij} c_{ij} = D$$

The constraints are identical to those used in the entropy model but a somewhat simpler objective function is specified. The solution is

$$x_{ij} = \alpha_i + \beta_j + \gamma c_{ij}$$

The solution procedure is similar to that used for the Q.T.P., and it is again necessary to be careful to not violate the non-negativity constraint. This simple linear model resembles the regression equation often used in movement studies, with origin effects, destination effects, and an impedence between the sets of places. Of course we expect γ to be negative.

Since $x_{ij} \ln x_{ij}$ is not very different from x_{ij}^2, an objective function of the form $\sum_i \sum_j x_{ij}(\ln x_{ij}) c_{ij}$ or $\sum_i \sum_j x_{ij}(\ln x_{ij}) c_{ij}/O_i I_j$ is suggested and these can also be found in the literature. The total cost constraint D is then no longer needed.

4. EMPIRICAL EXAMPLE

The data, tabulated in the Appendix, come from Cesario (1973); also see Slater (1974), Cesario (1974), and Baxter and Ewing (1979) who analyze the same table. Given is the number of people from each of ten counties who visited five parks during a single day, and the distance between these parks and the counties. It is rather absurd that one distance be used to represent the separation of a county and a park (see the map in Baxter and Ewing, 1979), but this is common in this type of modeling and the convention is accepted here.

From the data the outsums O_i and insums I_j are computed, and the objective is to re-estimate the movement quantities by filling in the body of the table. The results include the

Lagrangians as "pushes" and "pulls". They are of course estimates of the shadow prices, and are determined only up to a constant of integration as in any Neumann problem.

The root mean square errors compare with a value of circa 5.8×10^6 reported by Cesario (1974) and a value of 7.0×10^6 computed by Baxter and Ewing (1979). Cesario's model (1973, 1974) minimizes the RMSE without exactly satisfying the marginal constraints, and thus is not directly comparable to the L.T.P., Q.T.P., or entropy models. But the results suggest that the Q.T.P. solution is a plausible candidate descriptor of the events in question. It is more difficult to decide whether it is a better descriptor than the other models (except the L.T.P. which would only poorly represent the actual movements.

5. COMPUTER PROGRAM

The appended computer listing is slightly modified from an earlier version written by Dr. G. Dorigo while a post-doctoral resident at the University of California at Santa Barbara in 1980. It should be self explanatory.

DATA

Observed Movements

From County \ To Park	Big Pocono	Gouldsboro	Hickory Run	Promised Land	Tobyhanna	
Berks	46	35	333	84	69	567
Carbon	50	33	1670	71	91	1915
Lackawanna	230	6970	141	977	1917	10235
Lehigh	307	520	1458	315	387	2987
Luzerne	255	3366	4586	303	595	9105
Monroe	376	313	253	150	848	1940
Northhampton	385	1121	1263	499	981	4249
Pike	17	7	26	87	6	143
Schuylkill	63	101	1886	48	40	2138
Wayne	8	20	12	124	18	182
Total	1737	12486	11628	2658	4952	33461

Distances (miles)	Big Pocono	Gouldsboro	Hickory Run	Promised Land	Tobyhanna
Berks	95	101	89	115	96
Carbon	40	52	30	71	44
Lackawanna	45	21	46	35	29
Lehigh	47	62	57	62	62
Luzerne	55	45	25	65	49
Monroe	17	26	45	26	24
Northhampton	41	56	64	60	52
Pike	49	53	80	35	47
Schuylkill	70	77	57	85	71
Wayne	53	37	72	22	37

Source: Cesario (1973), Table 5, p. 245.

Model Results

1) $x_{ij} = (\alpha_i + \beta_j) / c_{ij}$

From County \ To Park	Big Pocono	Gouldsboro	Hickory Run	Promised Land	Tobyhanna	Pushes
Berks	0	275	292	0	0	45453
Carbon	0	717	1183	0	15	64461
Lackawanna	644	4462	1998	1163	1968	177282
Lehigh	48	1080	1143	226	470	123843
Luzerne	688	2279	4030	763	1346	194997
Monroe	0	1255	685	0	0	55139
Northhampton	357	1418	1212	440	822	148640
Pike	0	100	43	0	0	436
Schuylkill	0	762	998	67	311	107261
Wayne	0	137	45	0	0	0
Pulls	-119336	10132	6492	-95868	-63122	$\sum_i \sum_j x_{ij}^2 c_{ij}$ = 2.25 E+9

2) $x_{ij} = (\alpha_i + \beta_j) O_i I_j / c_{ij}$

From County \ To Park	Big Pocono	Gouldsboro	Hickory Run	Promised Land	Tobyhanna	Pushes
Berks	31	197	212	42	86	4.04 E-3
Carbon	111	532	890	108	275	7.28 E-4
Lackawanna	395	5026	2246	917	1652	6.71 E-5
Lehigh	207	1022	1061	263	434	1.81 E-3
Luzerne	347	2615	4562	517	1063	4.86 E-4
Monroe	191	738	418	227	365	0
Northhampton	325	1547	1293	374	709	1.68 E-3
Pike	9	53	34	21	26	1.59 E-3
Schuylkill	114	682	876	155	311	2.35 E-3
Wayne	8	73	36	34	32	7.78 E-4
Pulls	1.93 E-3	1.58 E-3	1.67 E-3	2.29 E-3	1.82 E-3	$\sum_i \sum_j x_{ij}^2 c_{ij}$ = 42.22

```
100 REM QUADRATIC TRANSPORTATION PROBLEM
102 REM W.R.TOBLER 25 SEPT 1982
105 REM ROWS, COLUMNS
110 DATA 4,5
115 REM OUT,IN-SUMS
120 DATA7,3,5,15,1,8,8,9,4
125 REM DISTANCES
130 DATA9,4,3,7,5,4,2,7,2,5
140 DATA3,7,12,4,9,9,4,3,5,1
200 READ NR,NC:REM # ROWS, #COLS
210 DIM O(NR),I(NC),I9(NC),O9(NR)
220 DIMM(NR,NC),D(NR,NC)
230 GOSUB3000
600 REM
610 DIM R(NC),E(NR),R1(NC),E1(NR)
620 FORK=1TONR:E(K)=1:E1(K)=0:NEXTK
630 FORK=1TONC:R1(K)=0:NEXTK
640 LP=99999:IR=0:OT=0:T2=2:EP=12:SH=0.5
650 REM GET TRIAL SOLUTION
655 TS=1
660 GOSUB 2000
670 OT=OT+1:T=1.0E32
680 REM CHECK SOLUTION
690 FORK=1TONR
700 IFE(K)>TTHEN720
710 T=E(K)
720 NEXTK
730 FORJ=1TONC
735 I9(J)=0
740 R(J)=R(J)+T
750 NEXTJ
760 FORK=1TONR
765 O9(K)=0
770 E(K)=E(K)-T
780 NEXTK
790 F1=1:T=0:DF=0:T3=0:S=0:S1=0
800 FORK=1TONR
810 FORJ=1TONC
820 M(K,J)=0
830 NEXTJ
840 NEXTK
850 REM
860 FORK=1TONR
870 FORJ=1TONC
890 SM=(R(J)+E(K))*D(K,J)
900 SM=SM*SH
901 REM LINES 904,905 FORCE
902 REM AN INTEGER SOLUTION
903 REM THEY MAY BE REMOVED
904 Q%=SM+SH
905 SM=Q%
906 REM PUSH,PULL & FUNCTIONAL
907 REM VALUES ALSO CHANGE
908 REM WITH INTEGER SOLUTION
910 M(K,J)=SM
920 I9(J)=I9(J)+SM
930 O9(K)=O9(K)+SM
940 IFSM>=0THEN970
950 F1=0
960 D(K,J)=0
970 NEXTJ
980 NEXTK
990 REM
1000 DF=0
1010 FORK=1TONR
1020 DF=DF+ABS(O(K)-O9(K))
1030 NEXTK
1040 FORJ=1TONC
1050 DF=DF+ABS(I(J)-I9(J))
1060 NEXTJ
1070 REM
1080 IFF1>0THEN1100
1090 GOTO650
1100 REM DONE, PRINT RESULTS
1110 PRINT!"  "
1120 PRINT!"RESULTS ARE"
1125 PRINT!"ITERATIONS=";OT;" ";IR
```

```
1130 FORK=1TONR
1135 S2=0
1140 FORJ=1TONC
1150 PRINT!K;J;M(K,J)
1152 IFD(K,J)<=0THEN1160
1154 S=M(K,J)/D(K,J)
1156 T=T+S
1158 T3=T3+M(K,J)*S
1160 S2=S2+M(K,J)
1161 S1=S1+M(K,J)
1162 NEXTJ
1165 PRINT!"ROW SUM=";S2;O(K)
1170 NEXTK
1172 PRINT!"GRAND SUM=";S1
1173 PRINT!"COLUMN SUMS="
1174 FORJ=1TONC
1175 S=0
1176 FORK=1TONR
1177 S=S+M(K,J)
1178 NEXTK
1179 PRINT!J;S;I(J)
1180 NEXTJ
1181 IFDF<1THEN1186
1182 PRINT!"DISCREPANCY DUE TO"
1183 PRINT!"FORCED INTEGER SOLUTION IS"
1184 PRINT!DF
1186 PRINT!" "
1187 PRINT!"PUSHES"
1188 FORK=1TONR
1190 PRINT!K;E(K)
1200 NEXTK
1205 PRINT!"PULLS"
1210 FORJ=1TONC
1220 PRINT!J;R(J)
1230 NEXTJ
1233 PRINT!"FUNCTIONAL VALUE FOR"
1235 PRINT!"LINEAR=";T
1237 PRINT!"QUADRATIC=";T3
1245 PRINT!" "
1246 PRINT!"DONE"
1250 END
```

```
2000 REM MAIN ITERATION FOR LAGRANGIANS
2002 REM PULLS ASSOCIATED WITH SINKS
2004 REM PUSHES ASSOCIATED WITH SOURCES
2010 FORL=1TOLP
2020 SS=TS:TS=0
2025 REM ESTIMATE PUSH (R)
2030 FORJ=1TONC
2040 T=0:S0=0
2050 FORK=1TONR
2060 S0=S0+D(K,J)
2070 T=T+E(K)*D(K,J)
2080 NEXTK
2090 R(J)=(T2*I(J)-T)/S0
2100 NEXTJ
2110 REM ESTIMATE PULL (E)
2120 FORK=1TONR
2130 T=0:S0=0
2140 FORJ=1TONC
2150 S0=S0+D(K,J)
2160 T=T+R(J)*D(K,J)
2170 NEXTJ
2180 E(K)=(T2*O(K)-T)/S0
2190 NEXTK
2200 REM NOW CHECK CONVERGENCE
2210 IR=IR+1:T=0
2230 FORK=1TONR
2240 DF=ABS(E(K)-E1(K))
2250 E1(K)=E(K)
2260 IFABS(E(K))<TSTHEN2280
2270 TS=ABS(E(K))
2280 IFDF<SSTHEN2310
2290 IFDF<TTHEN2310
2300 T=DF
2310 NEXTK
2320 FORJ=1TONC
2330 DF=ABS(R(J)-R1(J))
2340 R1(J)=R(J)
2350 IFABS(R(J))<TSTHEN2370
2360 TS=ABS(R(J))
2370 IFDF<SSTHEN2400
2380 IFDF<TTHEN2400
2390 T=DF
2400 NEXTJ
2410 IFT=0THEN2500
2440 TS=TS*10↑-EP
2450 NEXTL
2500 RETURN
```

```
3000 REM READ DATA
3005 REM OUTSUMS=ORIGINS=SOURCES=SUPPLIES
3006 SM=0
3010 FORK=1TONR
3020 READ O(K)
3025 SM=SM+O(K)
3030 NEXTK
3035 REM INSUMS=DESTINATIONS=SINKS=DEMANDS
3040 FORK=1TONC
3050 READ I(K)
3055 SM=SM-I(K)
3060 NEXTK
3062 IFSM=0THEN3066
3063 PRINT!"SUMMATION ERROR"
3065 REM DISTANCES
3066 X=1
3067 REM SET X=0 TO MODULATE
3068 REM DISTANCES BY FLOW SIZE
3070 FORK=1TONR
3080 FORJ=1TONC
3085 D(K,J)=0
3090 READ T
3100 IFT=0THEN3120
3110 D(K,J)=1/T
3112 IFX=1THEN3120
3114 D(K,J)=D(K,J)*O(K)*I(J)
3120 NEXTJ
3130 NEXTK
3140 PRINT!"ALL DATA ARE IN"
3150 RETURN
```

```
RUN


ALL DATA ARE IN

RESULTS ARE
ITERATIONS= 2   41
 1  1  0
 1  2  2
 1  3  3
 1  4  2
 1  5  0
ROW SUM= 7  7
 2  1  0
 2  2  1
 2  3  0
 2  4  2
 2  5  0
ROW SUM= 3  3
 3  1  0
 3  2  1
 3  3  1
 3  4  3
 3  5  0
ROW SUM= 5  5
 4  1  1
 4  2  3
 4  3  4
 4  4  3
 4  5  4
ROW SUM= 15  15
GRAND SUM= 30
COLUMN SUMS=
 1  1  1
 2  7  8
 3  8  8
 4  10  9
 5  4  4
DISCREPANCY DUE TO
FORCED INTEGER SOLUTION IS
 2

PUSHES
 1  14.7881588
 2  0
 3  14.2863455
 4  22.4425273
PULLS
 1 -12.4179445
 2  4.06999424
 3  2.68715162
 4  7.16224817
 5 -14.482015
FUNCTIONAL VALUE FOR
LINEAR= 120
QUADRATIC= 290

DONE
```

REFERENCES

Baxter, M., and G. Ewing (1979) Calibration of Production-Constrained Trip Distribution Models and the Effect of Intervening Opportunities. *Journal of Regional Science* 19(3):319-330.

Cesario, F. (1973) A Generalized Trip Distribution Model. *Journal of Regional Science* 13(2):233-247.

Cesario, F. (1974) More on the Generalized Trip Distribution Model. *Journal of Regional Science* 14(3):389-397.

Dorigo, G., and W. Tobler (1983) Push-Pull Migration Laws. *Annals of the Association of American Geographers*. Forthcoming (March).

Slater, P. (1974) Exploratory Analysis of Trip Distribution Data. *Journal of Regional Science* 14(3):377-388.

Tobler, W. (1981) A Model of Geographic Movement. *Geographical Analysis* 13(1):1-20.

Tobler, W. (1983) An Alternate Formulation for Spatial Interaction Modelling. *Environment and Planning A*. Forthcoming.

Wilson, A. (1967) A Statistical Theory of Spatial Distribution Models. *Transportation Research* 1:253-269.