Queen's Economics Department Working Paper No. 73

THE GRAVITY HYPOTHESIS AND TRANSPORTATION COST MINIMIZATION

John M. Hartwick
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

2-1972

THE GRAVITY HYPOTHESIS AND TRANSPORTATION COST MINIMIZATION

by

John M. Hartwick

Queen's University

DISCUSSION PAPER NO. 73

February, 1972

Abstract

The Gravity Hypothesis and Transportation Cost Minimization

John M. Hartwick
Queen's University

The Hitchcock-Koopmans transportation is linked to the gravity "model" of regional analysis with the aid of two new analytical results. Eighty transportation problems are generated by Monte Carlo methods and solved by linear programming. The transportation cost minimizing flows are compared with the flows generated by gravity methods in least squares regressions and in other non-parametric tests. Flows generated by gravity methods are indicated to be relatively poor proxies for those generated in a transportation cost minimizing system.

February, 1972

THE GRAVITY HYPOTHESIS AND TRANSPORTATION COST MINIMIZATION

John M. Hartwick*

Queen's University

1. Introduction

The gravity hypothesis, that the flows commodities or resources between two points can be approximated with the aid of a simple log linear function, has proved useful to regional economists obliged to work on problems with few data. Proxy data can be generated easily from the gravity "model".[1] Geographers have found the gravity hypothesis useful for the same reason as economists and in addition have turned to the hypothesis as a conceptual framework for organizing diverse formulations for estimating flows in space.[2] Tideman [1968] appears to be the only person who has successfully derived gravity - like functions from a simple model of the profit maximizing behaviour of a firm in geographic space. Isard [1960, p. 515] pointed out that the quality of proxies or estimates of interpoint flows generated with the gravity function was relatively poor for disaggregated industrial groups. Relatively good quality estimates were generated for aggregated industrial groups. Isard worked with observed interpoint flows as the bases for comparison with the flows generated with a gravity "model".

---

* Philip G. Hartwick handled the programming involved in the empirical tests in this paper. I am much indebted to him for his help.

In this paper, I compare flows generated from a transportation cost minimizing linear program with those generated with the gravity function. This comparison is intended to provide a test, alternate to Isard's, dealing with the quality of estimates generated with the gravity function.[3] In this case, theoretically optimal flows are used as the bases for comparison.

In Section 2 the Hitchock-Koopmans transportation problem in linear programming is related to the gravity "model". In Section 3 results on the quality of proxy flows generated with the gravity function are reported. In Section 4 special economic landscapes are considered and related to the previous results.

2. The Transportation Problem and the Gravity "Model"

Our economic landscape consists of m geographically distinct points of supply for a commodity per unit time and n geographically distinct points of demand per unit time. The demand points are separated from the supply points. The transportation problem is to determine a schedule of flows from the supply points to the demand points which minimizes the total transportation cost involved in the shipment and at least satisfies all the requirements of all demanders.

Let σ_i $(i = 1,\ldots, m)$ be the initial supplies and δ_j $(j = 1,\ldots, n)$ be the initial demands. Let τ_{ij} be the cost of transporting a unit of the commodity between supply point i and demand point j and let ξ_{ij} be the flow of the commodity between points i and j.

Formally, the transportation problem (Gale [1960, pp. 4-17]) is to determine ξ_{ij} so as to minimize

$$(2.1)\quad T = \sum_i \sum_j \tau_{ij}\, \xi_{ij}$$

subject to

$$(2.2)\qquad \sum_{i=1}^{m} \xi_{ij} \geqq \delta_j \qquad j = 1,\ldots, n$$

$$(2.3)\qquad \sum_{j=1}^{n} \xi_{ij} \leqq \sigma_i \qquad i = 1,\ldots, m$$

$$(2.4)\qquad \xi_{ij} \geqq 0 \qquad (i = 1,\ldots, m;\; j = 1,\ldots, n)$$

The dual to the transportation problem is to determine n prices π_j^1 at the demand points and m prices π_i at the supply points so as to maximize

$$(2.5)\qquad \sum_{j=1}^{n} \pi_j^1\, \delta_j - \sum_{i=1}^{n} \pi_i \sigma_i$$

subject to

(2.6) $\pi_j^1 - \pi_i \leqq \tau_{ij}$ $\quad$ $(i = 1,\ldots, n$; $j = 1,\ldots, n)$

(2.7) $\pi_j^1 \geqq 0$ $\quad$ $j = 1,\ldots, n$

and

$\pi_i \geqq 0$ $\quad$ $i = 1,\ldots, m$

It will be useful in the subsequent analysis to note that the feasibility of the transportation problem (2.2 - 2.4) can be expressed as finding elements $\xi_{ij} \geqq 0$ which satisfy the row and column constraints in the following table.

ξ_{11}	ξ_{21}	.	.	.	ξ_{m1}	$\geqq\delta_1$
						$\geqq\delta_2$
ξ_{12}	ξ_{22}	.	.	.	ξ_{m2}	$\vdots$
$\vdots$	$\vdots$				$\vdots$	
ξ_{1n}	ξ_{2n}	.	.	.	ξ_{mn}	$\geqq\delta_n$
						δ_{n+1}
$\leqq$	$\leqq$				$\leqq$	
σ_1	σ_2	.	.	.	σ_m	

Table 2.1

We shall link the transportation problem to the gravity "model" in the following two theorems. First we shall prove that if ξ_{ij} is optimal for (2.1 - 2.4) then demands will be exactly met[3], i.e. condition 2.2 will have equalities for the solution values. Intuitively, we observe that if this theorem were not true, then we could take an excess demand, leave it at any supply point and reduce the total transportation cost bill.

Secondly we shall prove that the transportation problem is feasible if and only if the total initial supplies at least equal the total initial demands. This result is well-known but the proof we develop has not been presented in the literature and firmly links the gravity "model" to the transportation problem.

Theorem 1: If vector $(\hat{\xi}_{ij})$ solves the transportation problem (i.e. 2.1 - 2.4) then $\sum_{i=1}^{m} \hat{\xi}_{ij} = \delta_j \quad (j = 1,\dots, n)$

Proof: Assume the contrary; that is assume $\sum_{i=1}^{m} \hat{\xi}_{ik} > \delta_k$ for the k th demand point.

Now $\pi_k^1 = 0$ by the equilibrium theorem [Gale 1960; th. 1.2] and $\pi_k^1 - \pi_i < \tau_{ik}$ for $(i = 1,\dots, m)$,

since $\pi_i \geq 0$ and we assume $\tau_{ij} > 0$ for all i and j.

By the equilibrium theorem again $\hat{\xi}_{ik} = 0$ for all i.

That is $\sum_{i=1}^{m} \hat{\xi}_{ik} = 0 < \delta_k$ which is a contradiction.

Hence the theorem.

The converse of this theorem is of course not true. Note the key condition is that $\frac{\partial T}{\partial \xi_{ij}} > 0$ for all i and j and that nonlinear total transportation cost functions can result in the above condition on demands.

Theorem 2: The transportation problem (2.1 - 2.4) is feasible if and only if $\sum_{i=1}^{m} \sigma_i \geqq \sum_{j=1}^{n} \delta_j$

Proof: I Assume the problem is feasible.

That is there exists $\xi_{ij} \geqq 0$ such that

(2.2) $$\sum_{i=1}^{n} \xi_{ij} \geqq \delta_j \qquad (j = 1,\ldots, n)$$

and

(2.3) $$\sum_{j=1}^{n} \xi_{ij} \leqq \sigma_i \qquad (i = 1,\ldots, m)$$

Summing (2.2) over j and (2.3) over i we get

$$\sum_{i=1}^{m} \sigma_i \geqq \sum_{i=1}^{m} \sum_{j=1}^{n} \xi_{ij} \geqq \sum_{j=1}^{n} \delta_j$$

II Assume $\sum_{i=1}^{m} \sigma_i \geqq \sum_{j=1}^{n} \delta_j$

Define $\delta_{n+1} = \sum_i \sigma_i - \sum_j \delta_j$ as in Table 2.1

Now define $K = \sum_{i=1}^{m} \sigma_i = \sum_{j=1}^{n+1} \delta_j$

Note that $\xi^1_{ij} = \frac{\sigma_i \delta_j}{K}$ $(i = 1, \ldots, m$
$j = 1, \ldots, n + 1)$

satisfies (2.2), (2.3) and (2.4) is thus a feasible solution.[4]

Observe that the feasible solution constructed in part II of theorem 2 is a gravity-like formulation where the constructed flows are not adjusted with transportation costs as they are in the usual proxy flows generated from a gravity function. The familiar gravity function is

$$x_{ij} = k \frac{\sigma_i \delta_j}{S} \quad f\ (t_{ij}) \tag{2.8}$$

where x_{ij} is the proxy flow, σ_i the flow at the point of origin, δ_j is the flow at the point of destination, t_{ij} is the cost of transporting a unit of the flow between points i and j, S is the total flow moving between points in the system, and k is a constant. Transshipping between either two points of demand or two of supply is ruled out. The function $f(t_{ij})$ is generally taken to be of the form $t_{ij}^{-\alpha}$ with α taking the value 1 or 2 _a priori_.

If x_{ij} is an observed flow, then k and α are selected in order to make the estimate, the right hand side of (2.8), approximate "closely" the value of the observed flow, the left-hand side of (2.8).

We shall refer to the value $\frac{\sigma_i \delta_j}{S}$ as the simple gravity estimate. Observed that by dividing (2.8) through by $\frac{\sigma_i \delta_j}{S}$, substituting for $f(t_{ij})$, and taking logs of both sides, we get

$$(2.9) \qquad \log \frac{x_{ij}}{\hat{x}_{ij}} = \beta + \alpha \log t_{ij}$$

where $\beta = \log k$ and $\hat{x}_{ij} = \frac{\sigma_i \delta_j}{S}$

Equation (2.9) was estimated in a least-squares regression in which x_{ij} was a flow optimizing a Hitchock-Koopmans linear programming transportation problem, t_{ij} was the unit transportation cost in moving the commodity from point i to point j, σ_i was the supply quantity at point i, δ_j was the demand quantity at point j and S was the sum of all supply quantities.

3. The Quality of Gravity Estimates

In excess of seventy-five separate regressions were run for an equation of the form of (2.9). Data were generated randomly for quantities of supplies and demands and transportation costs in an 8 by 9 Hitchock-Koopmans transportation problem.[5] The gravity estimation is considered to be of a high quality if the estimate of α is negative, statistically significantly different from zero, and stable. We do not ask that α be equal to either 1 or 2.

First we observe that of the 80 regressions run α ranged from + .0455 to - 1.703. Observe that none reached the classic gravity value of physics of - 2.00.

In Table 3.1 we note that a higher proportion were indeed of the conventionally accepted sign, namely negative. Also we observe that the Kendall rank correlation coefficient between the transport cost minimizing flows and the simple gravity flows is generally positive or of the expected sign. The rank correlation coefficient between the transport cost minimizing flows and the transportation costs is generally also of the expected sign, namely negative.

The exceptions are of course important to take note of since they indicate the fact that gravity measures are not consistently of high quality.

The stability of the estimates of α were tested by applying the Chow test to pairs of regressions of equation

Table 3.1

Summary of Empirical Tests

Number with values of	Slopes (α)	Slopes significant at .95 confidence level	Kendall Rank Correlation: x_{ij}'s & $\hat{x}_{ij}$'s	Kendall Rank Correlation significant of .95 confidence level x_{ij}'s & $\hat{x}_{ij}$'s	Kendall Rank Correlation: x_{ij}'s & t_{ij}'s	Kendall Rank Correlation Significant at .95 confidence level: x_{ij}'s & t_{ij}'s
Positive sign	12	0	30	20	13	0
negative sign	68	7	5	1	22	9
Total	80	7	35	21	35	9

(2.9) taken in a sequence. The pairs were chosen sequentially as the randomly generated problems were developed in the computer. That is the first regression was compared with the second, the second with the third and so on. The results are reported in Table 3.2.

CHOW TEST	
	No. of Slope Values
Significant at .95 Confidence Level	15
Not Significant	2
Total	17

Table 3.2

Note that we indeed have values of α which fail the Chow test or are derived in a statistical sense from a different population.

In general then I conclude that the gravity function (2.9) yields poor estimates of actual interpoint flows where the actual flows simultaneously assume a transportation cost minimizing set of values for a distribution system. The simplest way to summarize the results of the "Monte Carlo" experiments seemed to be to plot the scatter consisting of a set of pairs, each pair being comprised of a slope value α and and the corresponding intercept value β. These values were duly recorded and a remarkably regular linear pattern emerged. The pattern is contained in Figure 3.1

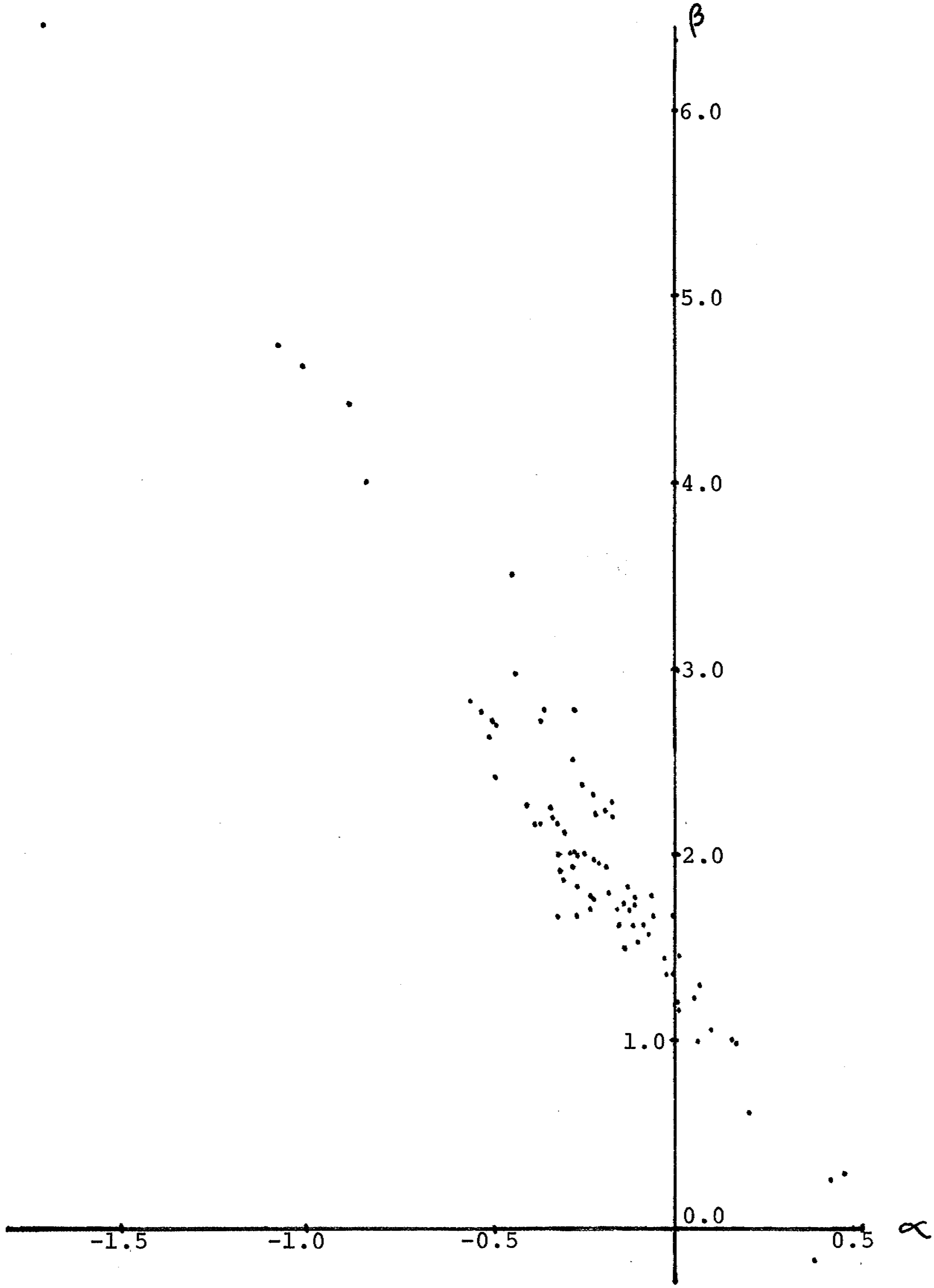


Figure 3.1

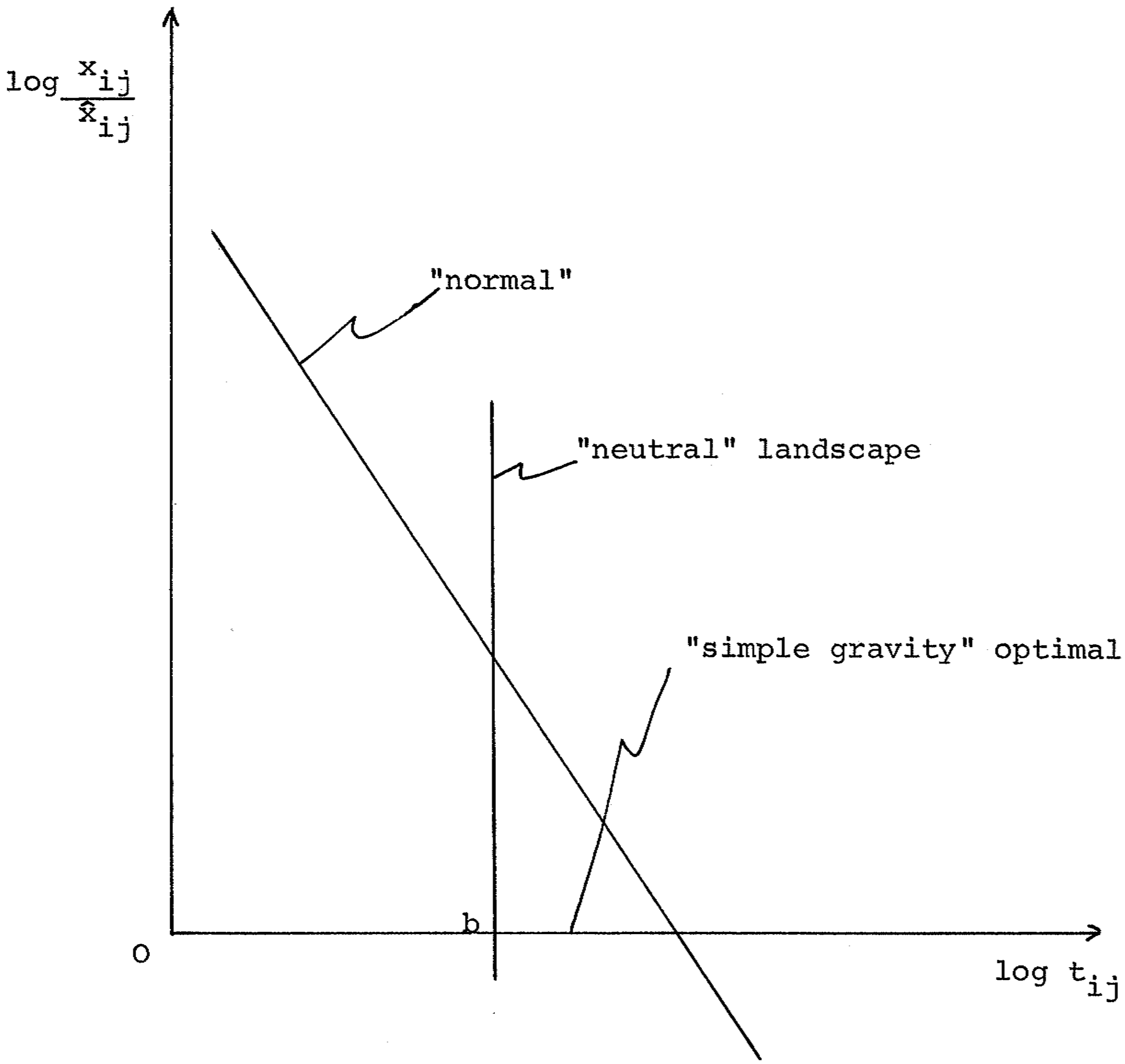
The first thing we see is that the two parameters α and k in equation (2.8) with the substitution for $f(t_{ij})$ are not independent and so an estimate of one determines the value of the other.

(2.10) $\beta = 1.4 - 3\alpha$.

Moreover the nature of the dependency of the slope value and intercept in equation (2.9) is such that all estimated gravity functions (equations of the form of (2.9)) have the common solution $(t_{ij}, \frac{x_{ij}}{\hat{x}_{ij}}) = (3, 1.4)$. In other words, the gravity function is stable in the sense expressed by equation (2.10) or equivalently in the sense that the log linear gravity function pivots about a unique point for estimation based on alternative data sets.

4. Gravity and Special Economic Landscapes

In this section, we shall examine whether it is possible to predict the nature of flows in a landscape from knowledge of transportation costs and/or particular quantities of supplies and demands. For example consider the situation in which the unit transportation costs between any two points, one of supply and one of demand, are equal. What will be the nature of the transportation cost minimizing set of flows? Will simple gravity estimates (recall the definition $\hat{x}_{ij} = \frac{\sigma_i \delta_j}{S}$ where S is the sum of all supplies) be optimal? It is easy to visualize the nature of the problems posed in the classic gravity "model" diagram illustrating equation (2.9) below as Figure 4.1.



$\log \frac{x_{ij}}{\hat{x}_{ij}}$
"normal"
"neutral" landscape
"simple gravity" optimal
b
O
$\log t_{ij}$

Figure 4.1

See

Figure 4.1

If simple gravity is also transportation cost minimizing then $x_{ij} = \hat{x}_{ij}$ and $\log \frac{x_{ij}}{\hat{x}_{ij}} = 0$ which in Figure 4.1 means all observations lie along the $\log t_{ij}$ axis.[7] The "normal" case is when all observations form an approximately "linear scatter" to the right of the $\log \frac{x_{ij}}{\hat{x}_{ij}}$ axis such that a least squares line through the scatter has a negative slope. Such a line is indicated in Figure 4.1.

If simple gravity were optimal for a system with equal transportation costs between any two points of demand and supply, then all observations would be co-incident at a point on the log t_{ij} axis. A moment's reflection will make clear that any feasible solution to the relevant transportation problem which satisfies all demands with equalities will be optimal. Recall that we proved in theorem 1 that a necessary condition for a solution to the transportation problem to be optimal was that all demands were just satisfied or met as equalities in (2.2). Theorem 2 made use of the fact that simple gravity estimates were always feasible and satisfied demands as equalities.

If we label the situation when all transportation costs are equal as a "neutral" landscape then for arbitrary t_{ij}, we see in Fibure 4.1 a hypothetical schedule satisfying a transportation cost minimization criterion. A hypothetical situation in which "simple gravity" is optimal as well as there being a "neutral" landscape, is point b in Figure 4.1. Note that the existence of a multiplicity of optimal solutions to the "neutral" landscape problem implies that the slopes and intercepts will in general not satisfy relationship (2.10). In fact since the intercept does not exist in the case of the "neutral" landscape and in addition the slope does not exist (even as infinity) when the observed line degenerates to a point such as b, we can consider the "neutral" landscape situation as a poorly-behaved exception to the relationship captured in equation (2.10).

Consider the possibility of other cases in which the simple gravity values are also transportation cost minimizing ones. Let me indicate one special situation. Let all supplies, demands, and transportation costs be integers. It is well known that the transportation cost minimizing solution to such a problem has integer values.[8] Thus a necessary condition for simple gravity values to be transportation cost minimizing in the problem specified in integers is that the simple gravity solution be in integers. Clearly, randomly generated transportation problems with integer specifications will generally not have integer simple gravity values. This can be confirmed by setting down a few examples in a form such as that in Table 2.1. Thus we can conclude that simple gravity values will in general not be transportation cost minimizing values and will not generally be observed in nature.

Footnotes:

1. See for example Leontief and Strout [1963] for the use of gravity estimates in an interregional input-output model. Lowry [1964] used gravity to generate flows of commuting trips in an urban model.

2. See for example Wilson [1970; Chapter 3].

3. It is interesting to note that both the NBER urban model and Curtis Harris' massive interregional model at the University of Maryland use the Hitchock-Koopmans transportation problem to generate shadow prices on flows at sites of origin and destination in order to generate new output levels of commodities and housing in the "next period". In a real sense, the estimates of interpoint flows generated by transportation cost minimization have superceded those generated in a gravity "model" in some recent applied work.

4. Neither Gale [1960] nor Dantzig [1963], the two well-known sources I examined, consider this result. In fact Dantzig proves a theorem in Section 15.2 ("Allocation with Surplus and Deficit") in which he explicitly assumes "the availabilities exceed the requirements but requirements must be met exactly". Theorem 1 indicates that the assumption is not required.

5. This proof of part II was presented in a class assignment by my student, Miss Diane Cummings, when she was an undergraduate at Queen's University.

6. As Samuelson [1952] indicated, transportation cost minimization, as an objective which nature is assumed to pursue, is a special case of the maximization of gross "social payoff" (gross economic rent) minus transportation costs. One might expect the flows generated from a program in which transportation costs are minimized only to approximate those generated when net "social payoff" is maximized. These considerations open up a broad area investigation, namely whether efficient economic organization in space is transportation cost minimizing, and we shall not pursue the topic.

7. When $x_{ij} = \hat{x}_{ij}$ for all i and j we can say that the solution of the transportation problem yields no information, that is

$$\sum_i \sum_j \log \frac{x_{ij}}{\hat{x}_{ij}} = 0 \text{ where } \log \frac{x_{ij}}{\hat{x}_{ij}} \text{ is a}$$

measure of information in the sense of electrical engineers. For additional comments on the use of information theoretic concepts in economics as well as for examples, see Theil [1967] and Hartwick [1971].

8. See for example Dantzig [1963; p. 305].

References

Dantzig, George B., 1963, Linear Programming and Extensions, Princeton: Princeton University Press.

Gale, David, 1960, The Theory of Linear Economic Models, New York: McGraw Hill.

Hartwick, John M., 1971, "Notes on the Isard and Chenery-Moses Interregional Input-Output Models", Journal of Regional Science, Vol. 11, No. 1, April.

Isard, Walter, 1960, Methods of Regional Analysis, Cambridge, Mass.: The MIT Press.

Leontief, W. and A. Strout, 1963, "Multiregional Input-Output Analysis" in Tibor Barna, ed., Structural Interdependence and Economic Development, New York: St. Martin's Press.

Lowry, Ira S., 1964, A Model of Metropolis, Santa Monica: Rand Corporation.

Samuelson, Paul A., 1952, "Spatial Price Equilibrium and Linear Programming", American Economic Review, 42, June.

Theil, H., 1967, Economics and Information Theory, Chicago: Rand-McNally.

Tideman, Nicholaus, 1968, "The Theoretical Efficacy of "Potential" and Transport Cost Models" Discussion Paper # 38, Program on Regional and Urban Economics, Harvard University, 1968.

Wilson, A.G., 1970, Entropy in Urban and Regional Modelling, London: Pion Limited.

Ministry of State Ministère d'État
Urban Affairs Affaires urbaines
Canada Canada

March 10, 1972

The Secretary,
Institute for Economic Research,
Department of Economics,
Queen's University,
Kingston, Ontario

Dear Ms:

Please have the enclosed manuscript made into a discussion paper.

Yours sincerely,

John M. Hartwick

John M. Hartwick

Encl.

sent to duplicating Mar 13/72

Ottawa K1A 0P6
(613) 993-1491