TN 11: A MODEL OF VISITOR FLOWS CONSIDERING A BASIC PARTICIPATION FUNCTION AND AN "ALTERNATIVE FACTOR":SIMULATION AND PARAMETER ESTIMATION

BY J. BEAMAN, H.K. CHEUNG, AND N.H. DO

ABSTRACT

This paper presents what may be called a generalization of the Cesario model. Specifically, emissiveness is considered. The goal is to show what the behavioural structures may be behind Cesario’s origin emissiveness. Functions are introduced that specify ways of defining Cesario's emissiveness. They are shown to be plausible in that they reduce to emissiveness in the way one might think of it if an origin was only served by one destination (an isolated destination). Also, the functions are constructed in such a way that people do not simply double the number of visits they make to destinations because a second destination is introduced into a system that already has one destination, even if the new destination is as attractive as the original destination and at the same distance from the given origin.

Ewing, in the review section of this chapter, comments on this formulation and relates it to other formulations in Chapter 2, particularly TN 30. Ewing also raises issues and points to alternative approaches to generalization.

Estimation problems resulting from the form of functions defining emissiveness are described. The authors report that original estimation attempts resulted in the discovery of a degeneracy problem (two or more parameters were related so variation in some defined/caused variation in others/another). The nature of this problem is described and two approaches to determining the parameter of the generalized model are presented. The one method involves determining coefficients similar to the original Cesario emissivenesses but this time corrected for availability of supply. The other approach is analogous to that explored by Cesario when he tried to explain his emissiveness values by socio-economic information on origins. Functions considered and concepts relate matters considered in TN 30.

It is not claimed that the modification to the Cesario model improves the model's overall effectiveness in explaining destination use. It is pointed out that no more total variance in origin-destination flow data is explained by this generalization than by the original model. Rather, it is suggested that by eliminating effects of supply on Cesario's emissiveness coefficients it is made possible to explain what are called the inherent emissiveness coefficients by using socio-economic characteristics. Absolute emissiveness and its relation to demographic factors are indicated to be matters that should be incorporated into a Cesario-like model when it is used to estimate destination use in an altered system.

In the way of general critique, it is pointed out that the Cesario model could be described as not being adequately successful in explaining people’s general origin-destination flow behaviour. The generalized model is noted as being equally deficient since only the internal structure of the Cesario model has been modified. Some indications are given about why both these models might be deficient so researchers can have a better feel for where the model proposed is applicable and in what directions research might be focused.

INTRODUCTION

Cesario has presented a gravity model formulation to explain observed outdoor recreation travel patterns between origins/geographic areas (origins) and destinations for overnight camping. The basic concepts behind this model are that given otherwise equivalent situations:

(1) associated with each origin is an inherent "emissiveness", E(o), which is a (relative) measure of the degree to which people in that origin will participate; (2) associated with each destination is an inherent "attractiveness", A(d), which is a (relative) measure of the degree to which people will visit that destination; and, (3) visits between any origin and destination decrease in a systematic way with the distance between them. The complete formulation of the Cesario model and procedures for estimating the E(o)'s and A(d)'s are given in CORD Study TN 4.

Cesario suggests that the emissiveness and attractiveness parameters computed using his model are functions of origin and of destination attributes, respectively. Now, without getting into the issue of whether the suggestion by Cesario meant that he thought that the emissiveness coefficients calculated were dependent only on the socio-economic characteristics of the population, one may note that it is reasonable to propose that the emissiveness of an origin depends on the amount of supply to which it is exposed. It is this relationship that is of concern in this paper as the authors proceed to generalize the Cesario model in such a way as to explicitly introduce supply related effects into Cesario's emissiveness coefficient.

There are a number of factors that can be conceptualized as influencing people's use of destinations. The ones studied here can be called the "alternative" factor, and the factor of "supply generating demand". One may consider the following example as an aid to understanding what prompts the authors to suggest that there is a need to study the likely influence (1) of alternatives and (2) of "the level of supply" on people's behaviour. Suppose an origin which is supplied by a single destination for camping *which is not at capacity at any time* is given access to another destination, with the second destination equally as attractive as the first (e.g., identical in what is offered) and adjacent to, or at an equal distance but in the opposite direction from the origin as the first destination. Naïve application of a gravity model based on use of the one destination would imply equal use for the second destination. But is it reasonable to argue that total visits ("demand" – really consumption) for the two destinations will be greater than total visits when there was only one destination. If the second destination offers the same attractive "good" (camping experience of same attractiveness) at the same price (distance, time, camping fee etc.) as the first destination, it is reasonable to argue that "consumption" should not change. Does putting out two rows of boxes of brand-X soap at one price suggest consumption will increase? Without price change or removal of a competitor, wouldn't the purchase of brand-X remain constant? Why wouldn't camping visits just divide equally between the destinations (at least after the transient effect of "checking out" the new destination wore off?

Does the original demand merely divide between the two destinations? If not and it increases, does one have a "paradoxical" situation of supply generating demand? Of course any paradox in the above example is explained away if one supposes that the alternative destination provides variety (even if attractivenesses are equal, one can go somewhere different), and this stimulates demand. Further, supply may affect people's awareness or knowledge of an activity, in this case camping. Awareness can prompt more activity, eventually resulting in participation being related to level of supply as indicated in Figure 1. Supply influencing demand for goods in this manner has been reported for reservoir recreation by Clawson and Knetsch (1966, see also Ch. 1, particularly parts authored by Knetsch and see TN 29).

When supply is limited one might see little demand even at a low cost. Given less limited supply and even some increase in price, there could be much greater demand. Clawson and Knetsch (1966) clearly consider that increasing supply can alter people's propensity to participate in an activity in a way different from just going more because supply is closer. From a dynamic and behavioural perspective, it is possible that at some point awareness, variety etc.

are such that demand "takes off" in relation to supply. Each new facility that is added can further serve to convince people that an activity is the thing to do. That is, it is the thing to do. However, because of limits on time and money, growth is limited. Figure 1 illustrates saturation being reached. At saturation, little demand expansion is to be expected when new supply is introduced.

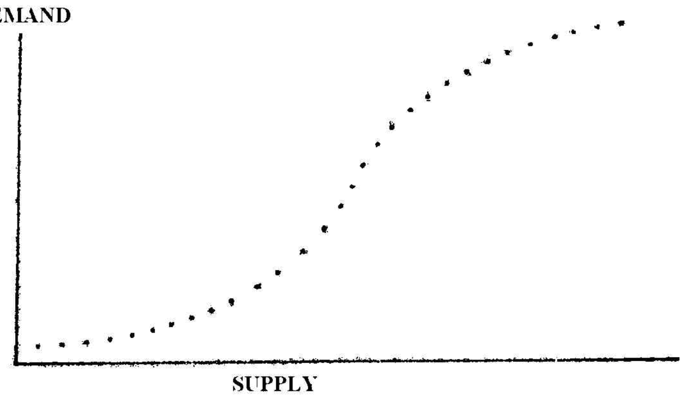
To keep this work in context, one should look at Knetsch's material in Chapter 1. There one can note that Knetsch raises some important issues regarding the difference between basing planning on maximizing consumption and on meeting demand. Expanding supply and justifying this on the basis of increased consumption is different from looking at where demand is being met and targeting some geographic equity how demand is met. As Knetsch argues, understanding behaviour is critical to meeting demand rather than reacting to consumption.

As for the s-curve, one may note that the type of "s-curve model" shown in Figure 1 is commonly used to explain the growth in sales when new products are introduced. In terms of tourism life cycle (Butler), increase may be followed by decline. Certainly, for products decline occurs when one product replaces another or a "fad" fades. Regardless, here the concern is not with a dynamic formulation. Concern is with the new equilibrium that will be reached, *at least for some time*, once new supply has been introduced into a system and people have "totally" adapted to the new configuration.

MODEL SPECIFICATION

Consider, for example, an origin supplied by two destinations. Let one destination be a new destinations equally as attractive as an original destination. Also, let the two destinations be at an equal distance in the opposite directions from the origin that generates virtually all visitors to the destinations. As described above, it is possible to think that after the second destinations was created and its use reached equilibrium, total attendance at the two destinations would not be the same as it was at the one destination before creation of the second. However, if creation of the new destination results in some people "seeing" a new product as available, one could expect an increase in total use as suggested in Figure 1. The size of the market for the product would increase. Furthermore, "consumers" of the old destination, may see going to the new destination and providing a something different (variety if nothing else). Probability of participation in the activity may thus increase for a particular segment. What is important about the example just introduced is that, in classical supply-demand terms, the new destination offers the same product at the same travel-price, etc., so one might say that consumption should not change. However, the west Texas reservoir creation situation cited in Clawson-Knetsch (1966) or arguments based on increased utility of trips when a variety of destinations is available suggest that the original use does not merely divide between the two destinations when the second destination is created. Total attendance at destinations is expected to increase: under certain conditions creating supply is expected to "create" demand resulting in increased consumption.

FIGURE 1: EQUILIBRIUM "LEVEL" OF DEMAND (PARTICIPATION) ASSOCIATED WITH SPECIFIC "LEVELS" OF SUPPLY



Deriving the model discussed here is a convenient way to expose the premises on which the SGD (supply generates "demand") model is based. Accept that for an isolated destination, the only destination serving an origin with a certain experience for which there is no substitute, visitor flows, V(o,d), are specified by Equation 1 below:

(1) VI(o,d)=F(D(o,d)),IA(d),Æ(o)) WHERE

VI(o,d) is the flow of some fairly homogenous type of visitor/visitor parties from origin, o, to an isolated destination, d;

D(o,d) is the "distance" from o to d measured in appropriate time-distance units;

F(, ,) is a function expressing the way V(o,d) varies with "distance";

IA(d) is the inherent attractiveness of the destination, d; and

Æ(o) is the “absolute” emissiveness of the origin, o.

Equation 1 is a special case of a more general equation discussed by Cesario (see TN 4). It implies that the amount of visitor flow from an origin to the only destination of a given kind that serves it for main-destination is dependent on the attractivity of the destination area (see TN 9 and TN 33 on the issue of destination area attractivity) and the distance to the destination, and all this in the context of a certain type of people/parties going to the given destination for a given purposes (see TN 33 for more general concerns).

But, one may ask, what happens when there are two destinations of differing or similar attractiveness at differing or similar distances from an origin that serve the origin in similar ways for a given type of e.g. main-destination visits (e.g. week-end recreational camping visits)? It is suggested that the total participation that will be generated when new destinations are entered into a system is not generally as great as if every destination could be treated as the only destination for the given purpose. If going to each destination was “independent” of the others existing flows computed using Equation 1 could be added. Rather it is proposed that introducing a new destination into a system will (at equilibrium) usually cause total use to rise less than the amount that the new destination would receive if it were the only destination (it replaced the existing destination), but total trips will be more than there were when only the original destination was there.

The ideas already introduced make it convenient to state:

Assumption 1: It is possible and meaningful to speak of an origin's isolated absolute emissiveness, Æ(o). [*One should note that the authors are really proposing to deal with dynamic* *change by modeling the equilibrium that arises after a transition effects related to introducing a new element into a system largely disappear*.]

In particular, one should note that since the distance functions used here are of the form $e^{-aD(o,d)}$ there is no problem with use becoming infinite as distance goes to zero (as would be the case with D(o,d) to some NEGATIVE power). Actually, one does not expect use of destinations to increase very rapidly as one goes from being one quarter of a mile from a destination to one eighth of a mile. Nor does one expect use to become "infinite" because a person lives in a destination. As well, for large distances there are problems with the "distance to a power" function (e.g. if that power is 2 or less- see Beaman and Smith TN on alternative factors) that do not arise when an exponential function is used. In this regard, behaviourally the exponential function used makes more sense than using distance raised to some negative power to explain number of trips based on distance traveled. Regardless, a distance function can be built on several shapes (see TN 1) with each one being appropriate to a different range of distance/travel time.

Assumption 2 suggests what isolated absolute emissiveness really means.

Assumption 2: The isolated absolute emissiveness of an origin to its isolated destination is a parameter that expresses the aggregate effect of a number of factors on the level of participation in an activity that the origin would generate to its isolated destination of unit attractivity at zero distance (or some other distance depending on the distance function) and it may be defined by one of the following increasingly specific equations:

(2A) Æ(o)=F(origin demographics and background e.g. vis-à-vis an activity of concern)

(2B) Æ(o)=C(1,o)POP(o)G(% age-gender distribution, % income distribution, etc.)

(2C) Æ(o)=C(2,o)(N(o)U+ΣB(i,j)n(i,j))

WHERE N(o) refers to the population of o and n(i,j) is the proportion or percent of people in o who have level j of socio-economic variable I (see Tn 6 & 12)

In Equations 2B and 2C above the constants C(1,o) and C(2,o) can be considered to reflect origin uniqueness. It is an effect that remains once total population and a population's demographics and other factors have been considered. Also the reason for having C(1,o) and C(2,o) in the above relates to problems of defining absolute scales and to identification problems in estimation, both of which are commented on subsequently. Furthermore, one can note that in some areas there is a history of water related activity, certain hunting, or other activities. A large or small C(1,o) or C(2,o) could be typical of origins in a region. A history of participation may relate to supply availability (see TN 29).

In particular, Equation 2C for which a derivation is given in TN 6 implies that Æ(o) has a simple form that involves the total potential for participation in an activity being proportional to the number of people in an origin area and dependent on the socio-economic characteristics of that population (being poor, aged, etc.). In fact, one may suggest that the age, income, etc. effects implied by Equation 2C are corrections that may be applied to any origin (see TN 12) while origin uniqueness is embodied in C(2,o). Nevertheless, demographic effects considered in TN 12 may be deceptive because correct values must be determined considering interactions. As well as gender interactions such as documented in TN 20, there may be combined effects of gender and geographic area.

The preceding assumption is important in accepting Assumption 3 which simply restates Equation l, taking into account the effect of distance to travel to a destination and destination attractivity:

Assumption 3: If an origin is served by a single destination for a given activity that has no substitutes (see TN 33) then total emissiveness of an origin is concentrated on a single destination so that actual visitor flows can be calculated if distance and the attractiveness of the destination are introduced as indicated in Equation 2:

(2) $VI(o,d)=e^{-rD(o,d)}A(d)Æ(o)=e^{-rD(o,d)}A(d)C(1,o)POP(o)G(o)$

WHERE r is a constant which may be called the impedance of distance (TN 14),

A(d) is the attractiveness of destination d,

G(o) is a correction for population distribution, and

C(1,o) is the per capita isolated destination emissiveness of o.

Now it has already been stressed that the primary concern here is generalizing Equation 2 to an expression that is appropriate when more than one destination is considered. Yet there is the need to present an expression that reduces to the isolated destination equation (Equation 2) when only an isolated destination is considered. Also, there is the problem raised earlier of not having the use of each one of several destinations be the same as if they were isolated destinations serving an origin when they all serve an origin. So what is needed is that an alternative factor of sorts be defined which expresses how origin emissiveness increases due to alternative destinations being available (on the effect of alternatives see TN 1, 3, 33 and sources cited there). In particular, it is proposed that a form of the type given in the Equation 3 group be considered.

Assumption 4: Attendance at a destination from an origin is defined by Equation 3 if it is accepted that destinations are such that it is (i) meaningful to say that they have an attractivity A(d), that is not influenced by other destinations being in the system (this is relaxed in TN 33), (2) that single purpose main destination visits are made to the destinations and (3) capacity of destinations is large enough that it does not influence use levels by affecting attractivity.

(3A) $V(o,d)=e^{-rD(o,d)} A(d)(PP(o)/(\Sigma_d P(o,d))((\Sigma_d(P(o,d)/PP(o))^{SE})^{SM}POP(o)CG(o)$

WHERE r, D(o,d), A(d) and POP(o) are as defined before,

P(o,d) is the product $A(d)\, e^{-rD(o,d)}$, and PP(o) is the maximum of the P(o,d) over all d associated with an o, and CG(o)=C(1,o)G(o).

Two alternative ways of writing Equation 3A which are more compact and of importance later are:

(3B) $V(o,d)= e^{-rD(o,d)} A(d)N(o)M(o)AIU(o)$

WHERE

AIU(o)=absolute isolated use PP(o)POP(o)CG(o) which is the flow that would occur to each destination or destinations with the highest P(o,d) if that destination were the only isolated destination serving o,

$M(o)=(\Sigma_d (P(o,d)/PP(o,d))^{SE})^{SM}$ is the supply-generated-"demand" multiplier which expresses how much the absolute isolated use is inflated due to having more than one destination near an origin, and

$N(o)=PP(o)/\Sigma_d P(o,d)$ is a prorating ratio which normalizes the total use generated, $M(o)AIU(o)$ according to the magnitude of $P(o,d)=\exp(-rD(o,d))A(d)$

From the subscripts involved in Equation 3B as indicated in Equation 3C, it is possible to see that one has a Cesario-type equation for use of a destination from different origins as determined by destination attractiveness and origin emissiveness parameters (see TN 4 and References cited there).

(3C) $V(o,d)= e^{-rD(o,d)} A(d)E(o)$

WHERE $E(o)=N(o)M(o)AIU(o)$

Though the issue is not pursued here (see TN 33), the preceding discussion can be reworded to give two further generalizations of the Cesario model. One involves introducing a destination alternative factor. Beaman (TN 9) presents empirical evidence that such a function involving the distribution of destinations around a given destination may be justified. The other generalization has to do with introducing considerations of other activities (substitutability—see TN 32 and 37).

A matter which has not been stressed yet in this paper and which is not discussed subsequently is the importance of recognizing what type of people making what kind of use of destinations a given model is appropriate for. In application it is obviously necessary to have data in *which attractiveness* so that attractiveness of a destination to fishermen is not confused with attractiveness to "recreational" campers. Also, as stressed elsewhere, weekend and holiday day-use attractiveness should not be confused with attractiveness for recreational camping (see TN 8). A destination that is attractive for a holiday may be too far away for day use or a weekend trip. Is it attractive but too far or unattractive because one does not want to drive several hours to spend one hour at a destination? Or, is it possible that both a distance cost and a “time available effect are operative? Furthermore attractiveness to day-users, non-main destination users etc. should not be confused (see e.g. user definitions in TN 30).

MODEL JUSTIFICATION

In the way of justification for the equations introduced one can note that when an isolated destination is considered one sees that from Equation 3, Equation 4C, for use of an isolated destination, is easily obtained. Since $PP(o)=P(o,d)$:

(4A) $V(o)= e^{-rD(o,d)} A(d)(PP(o)/\Sigma_d PP(o))(\Sigma_d(PP(o)/PP(o))^{SE})^{SM}POP(o)CG(o)$

WHERE the sums have one term, the one for the single origin destination pair (o,d).

(4B) $V(o)= e^{-rD(o,d)} A(d)(1)(\Sigma_d(1)^{SE})^{SM}POP(o)CG(o)$

(4C) $V(o)=e^{-r(D(o,d)}A(d) POP(o)CG(o)$

Equation 4C is the equation for the use of an isolated destination with attractivity A(d) from an origin o served only by d (see Assumption 3).

It is also easy to see that if one considers several destinations Equation 3 implies that the total flow from an origin $TF(o)=\Sigma V(o,d)$ is such that all destinations receive less than or no more flow than they would when they are isolated destinations if SE = 1 and SM=1. Specifically:

(5) $TF(o)\leq\Sigma_d VI(o,d)$ since

(6) $V(o,d)\leq VI(o,d)$

Equation 6 holds since the factor which inflates actual outflow from an origin is less than or equal to the factor that is implicit in the right side of Equation 5. The multipliers of concern satisfy the following relationship:

(6A) $M(o)=(\Sigma_d(P(o,d)/PP(o))^{SE})^{SM} \leq \Sigma_d P(o,d)/PP(o)$ for $SE \geq 1$ and $SM \leq 1$.

This is because by the definition of PP(o) and P(o,d) their ratio is less than one so that even without taking the root SM:

(6B) $M(o)=\Sigma_d (P(o,d)/PP(o))^{SE} \leq \Sigma_d P(o,d)/PP(o))$

Equality holds for SE=SM=1. Though there may be a reason to consider SE < 1 and/or SM > 1 as indicated in the review of the chapter, this possibility is not allowed here. One may wonder why the “root” SM was introduced. The power SM is actually only considered to be less than one here so that if several destinations have P(o,d)=PP(o), one does not have equal increase or decrease in flow for each of these destinations that is included in or excluded from a system. For example, consider two destinations:

(6C) $M(o)=(\Sigma_d (1)^{SE})^{SM}=2^{SM} < \Sigma_d 1=2$ if $SM \neq 1$

The doubling of emissiveness is indicated by the left side of Equation 6C being 2 if SM equals one. This illustrates the general concern. As long as all new destinations that are introduced into a system are introduced so that their P(o,d)=PP(o), each new destination receives as many visits as the destination d which originally had PP(o)=PP(o, d) *if SM equals 1*. But introducing the *root* SM changes the model so that use does not "multiply" with the number of destinations in the way just indicated.

In inductive terms, Equation 3 defines a system in which the total emissiveness of an origin, reduced for distance effects, is prorated among the various destinations according to the flow that they would receive if they were isolated destinations. But they receive a smaller total flow than would be obtained if each destination were an isolated destination (unless SE=SM=1 or for special cases if SM=1). In this context, Equation 3 is interesting to consider when examining the effect of creating a new destination.

Consider what happens when a system has one destination, d=l, at a distance D(o,d) from an origin, if this destination has an attractivity of Y, SM=1/2 and SE=2. When a second destination d=2 with attractivity Y and also at a distance D(o,d) from the origin is introduced into the system, say in the opposite direction from o as the first, Equation 4 implies that for both destinations the new use at both d=1 and 2 is:

(7) $V(,,N)=((1^{SE}+1^{SE})^{SM}/(1+1))V(,,o)=(2^{1/2}/2)V(,,o)=.707V(,,o)$

WHERE V(,,N) indicates the new system use at d=1 or 2, and V(,,o) refers to old system use at d=1.

Thus, the ratio of total attendance at the kind of destinations considered in the old system to attendance in the new system is:

(8) $2V(,,N)/V(,,o)=2(.707)=1.414$

About 40% more visitors are generated, but the same "good" is offered at the same price and consumption is up by 40% - supply is generating "demand" (participation) - and it is reasonable to suggest that the "supply has generated demand".

Related to this it is interesting to consider the cost-benefit associated with any increased satisfaction of people: are there any psychological benefits that would not have occurred had no new destination been in? And, if there are psychological benefits, how would they have differed if the new destination had been as attractive but had been located at such a distance that (P(o,d)/PP(o))=1/2 so only a much smaller supply generated demand would have occurred (12%). Clearly, there may be advantages in not subsidizing visits that do not benefit people very much.

ESTIMATION

Equation 3 is the kind of equation discussed by Cesario. In the terminology of this paper it specifies a Cesario type model:

(9) V(o,d)=F(D(o,d))A(d)E(o)

Because of the identity in form between Equation 3C and Equation 9, the attractiveness of the destinations A(d)'s are numbers which one may believe can be estimated using Cesario's approach and one may also expect that a constant, k, and the impedance of distance coefficient, r, associated with the distance function can be estimated. Caution about whether the r and A(d)'s will be "easily" or "accurately" estimated may relate (1) a bias depending on logarithmic transformations (see TN 4) and (2) heteroscedasticity considerations (see TN 19). Still, parameters can at least be estimated as Cesario has suggested.

Having read the last paragraph the reader may be asking why C(1,0) and C(2,0) appeared and now there is reference to k. The reason is that what could be called the "absolute" A()'s and E()'s which appear in Equation 9 are not what are estimated by Cesario. Because there is a problem in knowing on what "absolute" scale to measure attractiveness of a destination or origin emissiveness, these are measured on relative scales so that if AA(d) and EE(o) are the estimates obtained:

(10) A(d)=some constant times AA(d)=k(1)AA(d)

(11) E(o)=some constant times EE(o)=k(2)EE(o)

Actually, both AA(d)'s and EE(o)'s are estimated in such a way that the product of all AA(d)'s equals one and the same applies to the EE(o)'s. One cannot estimate n parameters for n parks. A constraint is necessary if one is to be able to get estimates. This is because parameters estimated in the linear form of the model, obtained by taking logarithms of the multiplicative version of the model, are to sum to zero so there is not redundancy with k=k(1)k(2). Therefore, one parameter is minus the sum of the others, e.g., $a_1=-(a_2+\cdots+a_n)$ for n parameters. This matter is discussed in texts covering dummy variable regression (e.g., where male effect is minus the female effect.

Now, the k(1) and k(2) introduced are critical in understanding other estimation problems. The constant estimated in the regression, k, is, as can be seen below, a number that reflects the values of both k(1) and k(2). The relation that shows this is:

(12) V(o,d)=F(D(o,d))k(1)AA(d)k(2)EE(o)=kF(D(o,d))AA(d)EE(o)= F(D(o,d))A(d)E(o)

Without information to define k(1) or k(2) the other component of the estimated k cannot be determined. Stated differently, one cannot break up k into its k(1) and k(2) components unless one has information allowing either k(1) or k(2) to be determined.

Given the preceding considerations the estimation procedure endorsed here for obtaining the parameters of the generalized Cesario model is essentially the same two step procedure proposed by Cesario. However, there are complications which arise because of the SE and SM parameters. Regardless, the first stage estimation is by the analysis of covariance (see TN 4), to produce estimates of k, r the A(d)'s and E(o)'s.

In the second step of estimation the emissivity of the Cesario formulation become the focus of attention. One might think that one has the information shown in Figure 2, which appears to show M times N observations to be used in estimating M + 3 parameters: the CG(o)'s, K(2), SE and SM in Equation 13 below.

(13) $EE(o)=E(o)/k(2)=(1/k(2))CG(o)POP(o)(PP(o)/\Sigma P(o,d))(\Sigma(P(o,d)/PP(o))^{SE})^{SM}$

But now it is necessary to consider whether the equation for which parameters are to be determined is identified: whether it is mathematically possible to determine unique parameter estimates. For the original Equation 13 there are M blocks of N equations each as shown in Figure 2. Thus it appears that M + 3 parameters are to be estimated from around M x N observations. But consider taking the natural logarithms of both sides of Equation 13 and moving all known terms to the left-hand side. Then one has:

(14) $DD(o)=SM(\ln(\Sigma_d(P(o,d)/PP(o))^{SE})+\ln(CG(o))+\ln(1/k(2))$

WHERE DD(o) is the sum of all known terms or one writes

(15) $DD(o)=SM(\ln(\Sigma_d(P(o,d)/PP(o))^{SE})+U(o)+KK(2)+\varepsilon(o)$

WHERE $U(o)=\ln(CG(o))$, $KK(2)=\ln(1/K(2))$ and $\varepsilon(o)$ reflects the presence of error.

FIGURE 2: A LINEAR MODEL PICTURE OF THE RELATION BETWEEN E( )'s AND THE DATA FROM WHICH THEY CAN BE ESTIMATED

	V(1,1) =	•••	+A(1)E(1)	
	•		•	
Block 1	•		•	N Equations
	•		•	
	V(1,N) =	•••	+A(N)E(1)	
	V(2,1) =	•••	+A(1)E(2)	
	•		•	
Block 2	•		•	N Equations
	•		•	
	V(2,24) =	•••	+A(N)E(2)	
	•		•	
	•		•	N(M-3) Equations
	•		•	
	V(M,1) =	•••	+A(1)E(M)	
	•		•	
Block M	•		•	N Equations
	•		•	
	V(M,N) =	•••	+A(N)E(M)	

So, when one moves to estimation of Equation 14, it can be seen that each of the M blocks in Figure 2 contains equations which could be transformed using first estimation cycle estimates to give equations like Equation 15. Each block essentially contains different observations of the same true value of EE(o) as a function of SE, SM, KK(2) and the U(o)'s. Thus one can see that E(1) or equivalently EE(1) is common to all equations in block 1 of Figure 2, E(2) is common to all equations of block 2 etc.

So, one might as well only use the M estimated E(o)ns, the EE(o)'s, as the dependent variables, in order to estimate the unknown parameters for all origins. The unknown parameters are as just indicated SE, SM, K(2), and the U(o)'s or in terms of Equation 14 they are SE, SM, K(2), and the CG(o)'s. Regardless, as already stressed there are M + 3 parameters to be estimated. Now, even though the EE(o)'s were such that Σ ln EE(o)=0 and thus have M-1 degrees of freedom, the DD(o)'s which are the dependent variable in Equation 15 do not satisfy the same linear constraint and can be considered to be M pieces of information allowing M parameters to be estimated. But it has been suggested that M+3 parameters are to be estimated. From M observation this is not possible unless something is done.

Firstly, in proceeding to obtain estimates it is necessary to recognize that if KK(2) and G(o)'s are considered in the same analysis of variance, in the way that Cesario computed a constant and origin effects, a constraint must be placed on the G(o)'s to remove a linear relation between them and the "mean term" KK(2). So consider that the constraint is applied so that the estimated U(o)'s have a sum of zero. Then there are M + 2 parameters to estimate. But applying the constraint is an indication that it is not the absolute CG(o)'s, or in logarithmic form absolute U(o)'s that are being estimated but:

(16) UU(o)=k(3)U(o)

So, also one sees that one will not get an estimate of ln(k(2)) but k(3)ln(k(2)). Still with the constraint on the U(o)'s there are two too many parameters to be estimated. So the problem is how to get estimates. One way is to eliminate the need to estimate two of the U(o)'s. One basis for doing this is to say that some origins are so similar that they are accepted as having the same U(o)'s. For example one may assume that U(i)=U(j) and U(k)=U(r) where i, j, k and r indicate at least 3 different origins (i could equal k).

Estimation in which the strategy just described was used to achieve estimates of parameters was tested using a special non-linear estimation program. This is commented on in the discussion section of the paper.

As long as only M-3 U(o)'s are introduced one can estimate the G(o) up to a multiplicative constant. Still, the ln(CG(o))=U(o) are not the ultimate object of interest. The planner or researcher ultimately wants to be able to explain U(o)'s or the corresponding CG(o)'s by observable information (e.g. socio-economic). Viewed in this way one can see why there need not be interest in estimates of the M-3 CG(o)'s but there should be interest in parameters that show how CG(o)'s for origins relate to socio-economic information about the residents of the origins.

A second alternative can in fact be pursued which is analogous to Cesario Stage II estimation. If one accepts that the U(o)'s in Equation 15 can be written as:

(17) U(o)=ln(Æ(o)/POP(o))

WHERE by Equation 2C

(18) Æ(o)=C(2,o)(UN(o)+ΣB(i,j)n(i,j))

Taking UN(o) out of the sum and using the Taylor series expansion of the natural log (1 + X) being approximately X for X << (much less than) 1, one has:

(19) U(o)=-ln(POP(o))+ln(N(o))+ln(C(2,o)U)+ln(1+Σ(B(i,j)n(i,,j))/UN(o))

(20) U(o)≡ln(U)+ln(C(2,o))+(Σ(B(i,j)n(i,j))/UN(o)

Substituting the approximation for U(o) into Equation 15 one gets the equation for which estimation may be carried out:

(21) $DD(o)=SM(\ln((\Sigma(P(o,d)/PP(o))^{SE}))+KK(2)+\ln(U)+\ln(C(2,o))+(\Sigma(B(i,j)u(i,j))/UN(o)$

Now, there may be some concern with what was done above because (Σ(B(i,j)u(i,j))/UN(o) might not be small. But, experience in working on TN 16 and on TN 12 has shown that the sum was small. However, if this is a problem there is an alternative formulation where the mean participation rate, y mean, is used instead of U and the sum of concern then is only a correction for a population difference from the average population for which the B(i,j) were estimated. The sum in this alternative formulation given by Equation 17 of TN 6 will only be large in the rare case that a population differs drastically from the population for which the B(i,j) were estimated.

Depending on whether or not the U and B(i,j) are to be treated as parameters to be estimated or known values used to arrive at an equation in which SE, SM, KK(2) and C(2,o)'s are the unknowns, there will be different matters of concern. If the B(i,j)'s are to be estimated, it is reasonable to follow the strategy that Cesario implicitly did in his emissiveness analysis of considering C(2,o)=a constant for all o. Then, one need only be concerned that M-3 of the B(i,j)'s can be estimated along with (1) a constant that has a value dependent on ln(C(2,o)), ln(U) and KK(2); (2) SE; and (3) SM. On the other hand if U and B(i,j) values are determined exogenously and are used along with known n(i,j) and N(o) to make a correction for population distribution, one has the problem discussed earlier with respect to estimating the CG(o)'s or their logs, the U(o)'s. As indicated in that discussion, identification can be achieved by specifying that C(2,i)=C(2,j) and C(2,k)=C(2,r) while the C(2,o)'s are transformed in scale by applying the estimation constraint that the sum of the estimated ln(C(2,o)) must equal zero.

RESULTS AND DISCUSSION

The model proposed here preserves the elements of the Cesario approach, but extends it to include an alternative factor and other parameters to "explain" how people respond to change in supply.

There was no question that the "Stage I" emissiveness and attractiveness parameters could be estimated. But there was enough concern about whether Stage II estimation would be successful that the estimation test alluded to earlier was carried out. A simple example estimation was run using a two destination, five origin system with sufficient constraints on the CG(o) so that the parameters in Equation 5 were identifiable. The values of the dependent variable DD(o) of Equation 6 were computed based on "correct" values for SM, SE, KK(2) and the CG(o). In practice, of course, the DD(o) would have a stochastic element but concern was not with the sensitivity of the parameters to error in DD(o) (so error was not introduced showing parameter values could be exactly recovered and thus estimation was possible). Using the predetermined DD(o) as dependent variables and an assumed initial value of SE, Equation 5 was estimated by an iterative least squares regression.

The iterative procedure used is straightforward. In the second millennium one would use a routing such as SAS’s PROC NLIN (see e.g. a version of the SAS/STAT manual by SAS Institute, Cary, NC)to do what is descried and new powerful estimation capabilities. After assuming an initial value for SE, SM an iterative process begins. Initial parameter values are corrected according to a "gradient" criterion which moves them towards the value which minimizes the sum of squared residuals. The corrected values are used in successive steps to produce new values until there is "convergence" to an acceptable minimum sum of squared residuals (e.g., changing parameters does not reduce a sum by enough to affect parameter values). After one series of iterations in a variety of different tests may be performed (e.g., starting with different initial values) to see that the minimum found is a minimum and the only reasonable solution to consider (e.g., a “local” minimum or maximum can be detected and confused with an appropriate solution). Regardless, by showing that parameters used to create/simulate a situation could be recovered, it was shown that the structure is such that the model is estimable, at least when the kinds of assumptions specified earlier can be made and parameters do not meet certain conditions. It is possible to specify parameter values for which estimation cannot be carried out (SE=SM=l). However, except for unlikely conditions, and one can expect to get estimated values except when error or structural problems with the model are such that no "reasonable" parameter values will explain dependent values.

Thus what is proposed in this paper is a modification of the Cesario estimation approach that allows one to take into account people's response to alternative destinations . In this model one explicitly recognizes that the emissiveness estimated in Cesario's first step of estimation should not be analyzed without considering the configuration of supply around an origin. The formulation is such that one has an explicit function defined which allows one to take into account the attractivity of alternative destinations and their geographic configuration, when one wants to calculate what will happen when a new destination is added to a system or an existing destination is removed. Actually a similar operation can be carried out using the model introduced in TN 1. There Cheung even presents an example of how a system is altered when a new destination is introduced. He has computed new origin destination flows based on changing his alternative factor. Incidentally Cheung's alternative factor has much in common with the alternative factor derived here and others that have been defined (see TN 30 and TN 3). However, as Ewing points out in the Review notes on this chapter there may be a big difference between using an ad hoc alternative factor and to some degree determining the alternative factor influence based on behaviour.

The reader may wish to note that originally the degeneracy problem encountered in Stage II estimation (which was described in the Estimation section of the paper) was not recognized. Bacon (a statistics professor at Queens university) worked on the estimation problem to find out why attempts to estimate 27 parameters for 26 origins always failed because matrices which were to be inverted to estimate the parameters were of rank 25 or less. Careful examination of the problem showed that one really only had 26 "Independent" pieces of information on which to estimate parameters. It is in this way that it was recognized that the “equivalent” origin approach to reducing the number of parameters, introduced earlier, could be used to get parameter estimates. In fact, it was recognized that without a way to reduce the number of parameters to be estimated, estimation would fail.

It is not being suggested that the model defined gives a better fit to data than the Cesario model. Obviously the fit to given data is determined by the first cycle of estimation. What is claimed is that it is necessary to recognize that Cesario’s emissiveness values should be corrected for supply effects to get emissiveness values that relate to destination independent origin attributes. This is a reason that Cesario's attempt to estimate the relationship of his emissiveness values to the aggregate socio-economic characteristics was not successful. Given that more variance is associated with the “alternative factor” component of Cesario's E's than with origin attributes affecting the inherent emissiveness factors, one needs to extract the destination related variance prior to trying to establish origin related influences.

It would be quite possible to take a given system of destinations and carry out a simulation exercise based on parameter estimates to find out how much of the variation in the Stage I emissiveness is related to the configuration of destinations around the origins considered, and how much is related to socio-economic characteristics of the inhabitants of these origins. Simulations along this line have shown that once size of population has been considered, variation from other socio-economic characteristics may be expected to produce no more than about 5 to 10 percent difference in the amount of participation that different origins will generate. (CG's do not vary much but C(2,o)'s may). The point is that the configuration of destinations around origins may account for a much larger variation in the amount of participation that origins generate than do origin variables (e.g., socio-economic and other variables mentioned above).

In practical terms the last paragraph indicates that a relatively wide spread of the EE(o) values is to be expected when there is high variability in the distribution of supply around various origins. The distribution of destinations is affecting the values that EE(o)'s have as well as these being influenced by N(o) and differences in the socio-economic characteristics of origins.

These authors believe that in terms of the model presented, the important research step that must be taken is not the consideration of the effect of income or other variables on EE(o)'s which is what Cesario suggested (see TN 4), but the identification of the influence of supply on EE(o)'s and then consideration of the importance of C(2,o) or C(1,o) factors of origin uniqueness in explaining behaviour. The matter of origin uniqueness relates to a “cultural” affinity to activities such as an activity being popular in one area and not in another regardless of supply. In essence, one is acknowledging that people’s background matters and that does not depend on demographics and may not even be reflected in “psychographics”.

It would be nice to be able to report that computer runs have been made using the data which Cesario used that confirm the validity of "The Generalized Cesario Model". However, this is NOT EXACTLY the case. Weighted regression runs (TN 19) to estimate more accurate emissiveness parameters than those presented in TN 4 have been processed. The E()'s obtained have been analyzed to determine if they are in accord with the formulation for an origin.

(22) E()≡K(Prorating ratio)(supply generates demand effect)(Population)

(23) $E()=K\ X(2,o)X(3,o)^{SM}POP(o)$

Actually, it was convenient to determine if the prorating ratio and "population have an exponent of one. Because, as indicated above, the supply generates demand effect has an exponent SM, this was estimated. The results were that Cesario's emissiveness values were not compatible with the generalization in this TN, but rather with the generalization in which SE is less than one, not greater than one, and in which the maximum potential to visit a destination, from PP(o), is an important variable. One model developed had a R^2 of 0.80 when population alone was introduced. When the other two variables were included R^2 was 0.95. When these other two variables (PP(o) and X(3,o) with SE < 1) were introduced, each alone, (given the other and population) had already explained a significant amount of variance. However, this model is not discussed because its definition and the rationale behind it, would be a deviation from the theme of this paper. It is planned to present these results in a separate non-CORD Study paper.

Nevertheless, the results of estimating parameters for SE=1.2 and 1.4 were invalid, because of the collinearity between X(2,o) and X(3,o). The correlation for SE=1.2 was 0.985, while the correlation for SE=1.4 was 0.955. To obtain estimates of parameters, either X(2,o) or X(3,o) was eliminated from the equation. When this was done, it was found that the coefficients

B(2) and (-B(3)) did not have values between zero and one (*modern programs allow parameter values to be constrained so some problems mentioned can be avoided*). That was to be expected because B(2) should be one and X(2,o) and X(3,o) have about the same variance and negative correlation. Therefore, 1 > B(3) > 0. This means that if B(2) is estimated, it should be TOO SMALL by the value of SM; and if SM is estimated it should be TOO SMALL by -1, because its value will include -B(2). Actually, B(2) values were in the -2 range, and B(3) values were in the +2 range. In other words, the values of B(2) and B(3) suggest that the structure of the particular Generalized Cesario Model used was not appropriate to the data. Because the two regression coefficients relating emissiveness to X(2,o) and X(3,o) were too large, it seems clear that:

1. being located in park-like land, there was not much need to use "park-influenced emissiveness"; and

2. the "configuration" of destinations around an origin reflect: a)a history of trying to meet needs, or b)little need to meet needs because of a situation such as that suggested in 1.

The kind of research suggested above has not proceeded beyond the stage just described because as work on TN 19 and TN 35 progressed it became clear that it was necessary to use weighted regression to get more accurate estimates of the EE(o)'s than originally obtained by Cesario. These were obtained late in 1974. However, getting the new estimates of the EE(o)'s, which were much more accurate than ordinary least squares estimates, only raised a new problem and that was that the Cesario model should be rejected as structurally invalid for the analysis of the data to which it was being applied. Certain obvious problems with the data used include (i) using combined data for week day, weekend days and holidays (see TN 8) and (2) problems in defining if a visitor was really a main destination visitor (see TN 30). Also there is a problem commented on subsequently in that in the Cesario data are only data on part of the destinations in a system.

In summary, recognizing the problems already cited meant that EE(o)'s and destination attractiveness estimated by Cesario might not have clear meanings. Thus further sophisticated analysis of the EE(o)'s was not justified. Ideally a new analysis will be carried out on a new large data set on a truly complete system of destinations and with good definitions and data collection practices being used in assembling these data (see TN 21 on the problems with CORD Study Park User Surveys). Then, one will be justified in more elaborate empirical analyses of the EE(o)'s than that presented earlier.

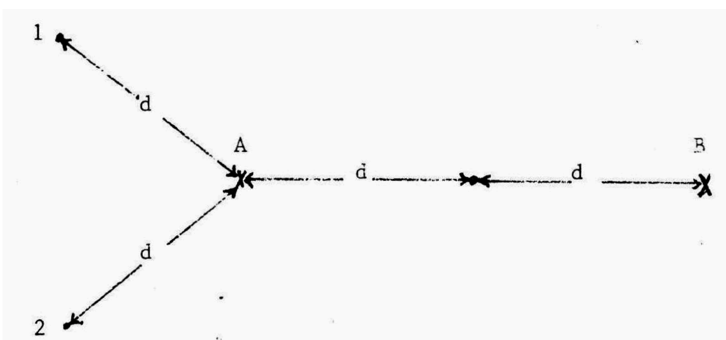
However, the proceeding discussion evades the issue of how appropriate the Cesario model for Ontario is for making estimates for use of new parks. Applying the Cesario model to estimate parameters that explain visitor flows for Ontario (TN 4) and then using the results to simulate origin-destination flows has shown that whereas the Cesario model (when used with real data) results in an R-squared in the order of .8, it should result in a much higher R-squared, as indicated by the R-squared values obtained when estimating parameters for simulated Ontario visitor flows. For the Ontario Parks on which estimation was originally done, parameter estimation on simulated data for Ontario Parks showed that the R-squared for estimation should be in excess of .99. The "generalization" suggested may explain part of the variance not being explained by the “Cesario” relation between population and emissiveness, but it does not explain any of the 20% of variance which was left unexplained after the Cesario model was applied. This obvious failure in terms of explanation points up the structural deficiency of the Cesario model (for tests of model structure see TN 19 and TN 35).

One may reasonably ask what could be wrong with a model as general as the Cesario model. A simple answer is that the model does not allow for asymmetries in behaviour. The kind of situation which might be creating problems is shown in Figure 3, where people do not go to a given destination from a certain origin because it is readily available as supply to other locations. People in origin 3 have a destination which from their perspective appears more attractive than destination A which appears attractive to origin 1 and 2. In particular the other source of problems, as suggested earlier, is almost certainly aggregation over different types of use (see TN 3, TN 8 and TN 40, etc.) Obviously if data on camping combines “resident”/continuous summer users or holiday users with weekend users different functions should be used to explain the behaviour of these people. It is not enough to simply deal with campers.

At this point it remains unknown whether introducing visit type and other considerations would result in a drastic improvement in the R^2 of the Cesario model. Such matters must be taken up to determine the appropriateness of the Cesario model to sufficiently disaggregated data.

Previous parts of this discussion section have suggested ways of testing the validity of the model proposed here. However, one of the most direct validity tests has not been mentioned. The B(i,j)'s, which can be estimated if it is assumed C(2,o) is equal to a constant for all o's, should agree with the B(i,j)'s obtained by an analysis of variance of national survey information on people's use of parks. CORD Study national survey information provides an abundance of data on parks use and even the similar Canadian Government Travel Bureau surveys provide some data. So if there is success in estimating B(i,j)'s for the Cesario model when it is applied to disaggregated data for which it is more structurally valid than it is for the Ontario data (TN 4), then there is an abundance of "independent" estimates of these B(i,j) so that comparisons can be made resulting in a test of model validity.

FIGURE 3: THE RELATION OF THREE ORIGINS TO TWO PARKS THAT ILLUSTRATES A SITUATION WHERE THE CESARIO MODEL WOULD NOT GIVE GOOD USE ESTIMATES*



Above 1, 2 and 3 are origins while A and B are destinations each with attractivity X.

*** It is possible to further generalize the Cesario Model so that the attractivity of A to people from 3 is "corrected" so that it is less attractive than B because of the "competition" effect from origins 1 and 2. This is not discussed here, and it involves major changes to Cesario's original formulation, that have nothing to do with the generalization of the model introduced here.**

Finally, only once in the preceding discussion has the importance of system definition been introduced. Yet, for Stage II analysis, defining what destinations are considered by people from a given origin for a given type of trip (week day, weekend day, weekend overnight, or “extended” holiday) is crucial. With data on any number of origins and destinations one can carry out Stage I analysis but if one does not have attractiveness for all destinations affecting use from a given origin then such factors as the following cannot be defined:

(24) $SM(\ln(\Sigma_d (PP(o)/P(o,d))^{SE}))$

(*Actually, except for simulations, one may argue that SE and SM are not system's parameters but should change over time or with changes in supply. This matter is of importance still it is not examined in this paper. A paper must end somewhere.*) No more need to be said if specifying "the" system were a simple matter of enumerating certain types of facilities. Unfortunately, the problem of substitutability should be considered for reasons suggested by Veal (**reference not located**) and also discussed in TN 10 and 33 (see also TN 32 and TN 37). Concerns about how the decision to go to a given place is made are elaborated in TN 33 where substitutability is introduced into the Cesario model in two different ways. One has to do with substitution at destinations and attractivity, and the other with substitutability at an origin and emissiveness.

CONCLUSION

The comments of the last few paragraphs show that much of the apparent conceptual elegance of the "generalization" of the Cesario model is counterbalanced by problems that have been ignored or are not dealt with in this paper. Still, the authors believe that the questions raised regarding ways of understanding visitor flows represent useful progress in developing more meaningful models of behaviour.

This paper has shown how the Cesario model can be generalized so that inherent emissiveness can be extracted from the emissiveness parameters that Cesario originally showed how to calculate. This means that if the generalization is even approximately structurally appropriate to a given situation, system model parameters can be calculated which can validly be used in estimating the consequences of, for example, deleting a given destination from a system. Cesario's Stage I emissiveness values should not be used because their value is conditional on the system for which parameters are estimated. Thus a critical step in making the Cesario model useful to planners has been taken. It also seems fair to claim that a useful theoretical contribution has been made in clarifying what the emissiveness parameters in the Cesario model really measure in a particular case.

If the paper prompts some other researcher to pursue research that was suggested it will have served another useful purpose. For example, it seems very important that some research be done to clarify how much Stage I emissiveness depends on supply configuration in a variety of circumstances as opposed to aggregate socio-economic characteristics associated with an origin.