TRANSPORTATION SCIENCE, Vol. 11, No. 3, August 1977

On Stochastic Models of Traffic Assignment

CARLOS F. DAGANZO

University of California, Berkeley, California

and

YOSEF SHEFFI

Massachusetts Institute of Technology, Cambridge, Massachusetts

This paper contains a quantitative evaluation of probabilistic traffic assignment models and proposes an alternate formulation. First, the concept of stochastic-user-equilibration (S-U-E) is formalized as an extension of Wardrop's user-equilibration criterion. Then, the stochastic-network-loading (S-N-L) problem (a special case of S-U-E for networks with constant link costs) is analyzed in detail and an expression for the probability of route choice which is based on two general postulates of user behavior is derived. The paper also discusses the weaknesses of existing S-N-L techniques (with special attention paid to Dial's multipath method) and compares them to the suggested approach. The proposed model seems reasonable and does not exhibit the inherent weaknesses of the logit *model when applied to sets of routes which overlap heavily. The discussion is supported by several numerical examples on small contrived networks. The paper concludes with the discussion of two techniques that can be used to approximate the link flows resulting from the proposed model in large networks.*

Traffic assignment methods deal with the problem of allocating fixed traffic demands between origin-destination (O-D) pairs to an existing or hypothetical transportation network made up of links connecting nodes. The foundation of any traffic assignment technique is the assumed rationale behind a motorist's choice of a route between his origin and destination.

WARDROP[1] identified two criteria that can be used to allocate traffic to competing routes:

A. User equilibration (U-E):

In a user equilibrated network *no user can improve his travel time by unilaterally changing routes.* If the links of the network have unlimited capacities the trip times on all the routes actually used between an O-D pair are equal and less than those which would be experienced by a single vehicle on any unused route.

B. System optimization:

In this instance the average journey time of all motorists is a minimum. Consequently, *the overall vehicle-hours spent in the network are minimized.*

The U-E criterion is commonly accepted as the standard objective of a good traffic assignment procedure.

Heuristic techniques such as various *capacity restraint*[2, 3] and *incremental assignment*[4] methods have been suggested to achieve user equilibrium in networks exhibiting congestion effects (flow dependent costs) on the links. These techniques, however, do not converge to the desired U-E flow pattern.[5, 6]

These difficulties have recently been overcome by mathematical algorithms that can be applied to moderate sized networks and that converge to a flow pattern in accordance with Wardrop's U-E criterion.[7, 8] All these methods treat route choice deterministically; users make consistently perfect decisions because implicit in the U-E criterion is that motorists are identical and infallible individuals. This is clearly not true but seems not to matter for heavily congested networks as in this case simple minor disturbances from the U-E flow pattern result in such large route cost differences that most of the users notice them. Consequently, flow patterns very different from the U-E pattern cannot be stable in a congested network.[9]

For lightly congested networks, however, other factors (such as motorists' inaccurate and/or distorted perception of link travel costs) may play a more dominant role than congestion effects. This was recognized in the early traffic diversion studies and perhaps explains the extreme sensitivity of U-E flow patterns in uncongested networks to small changes in the network (SHEFFI[10] found, while conducting an extensive study of maintenance strategies for the Ethiopian road network, that with deterministic traffic assignment methods the flows on the network were substantially affected by small network improvements).

Some models of traffic assignment which do not necessarily allocate all the traffic to the set of routes with lowest cost have been proposed in the literature.[11-13] These models are applicable to uncongested networks and are the subject of this paper. Before reviewing them, it is useful to introduce the concept of stochastic user equilibration (S-U-E) as this will enable us to illustrate the relationship between U-E and non-U-E methods and to discuss the latter in a unified way.

We suggest modifying Wardrop's U-E criterion to read as follows:

A′. Stochastic user equilibration (S-U-E):

In a S-U-E network no user *believes* he can improve his travel time by unilaterally changing routes.

In order to restate this principle in more precise terms, we must introduce some definitions. For every set of link flows, we let T_k^{ij} be the measured travel time in the kth route joining origin i and destination j (since, even under identical traffic conditions, route travel times change from time to time, T_k^{ij} can be regarded for our purposes as the time average of such travel times). We also let t_k^{ij} denote the travel time on the kth route between origin i and destination j as believed by a user of the system at the time he makes the route choice decision.

The values of t_k^{ij} cannot be observed and will generally be different from T_k^{ij}; they will also differ from user to user and perhaps vary for the same user from time to time.

Because of this, it seems natural to model t (we use vector notation $t = (\cdots, t_k^{ij}, \cdots)$, $T = (\cdots, T_k^{ij}, \cdots)$) as a random variable distributed across the population and to regard the set of perceived travel times for any given user as an outcome of such random variable. One can then talk about a probability of route choice for a motorist selected at random from the population (at least before his choice becomes known).

The probability of selecting the kth route between origin i and destination j, p_k^{ij}, is related to the perceived route travel times by:

$$p_k^{ij} = \Pr\{t_k^{ij} < t_h^{ij};\ \forall h \neq k \mid T\}.$$

If one now assumes (not unreasonably) that motorists stimulated by the same set of measured travel times act independently and that we have a significant number of users making the same decisions, the weak law of the large numbers ensures that:

$$\Pr\{t_k^{ij} < t_h^{ij};\ \forall h \neq k \mid T\} \cong \frac{x_k^{ij}}{\sum_{\forall k} x_k^{ij}}\ ;\ \forall i, j, k \tag{1}$$

where x_k^{ij} is the flow on route k between origin i and destination j.

Since the LHS of Equation (1) is a function of the measured travel times on the network, T, which in general depend on the flow pattern $x = (\cdots, x_k^{ij}, \cdots)$, Equation (1) is an equilibrium equation which merely states principle (A').

The definition of p_k^{ij} in terms of t above is not entirely correct for deterministic models ($t \equiv T$) because, when two or more routes from i to j are tied for shortest, all routes from i to j have zero flow. Consequently, the definition should be modified for cases when there is a nonzero probability that two routes have the same perceived travel time (otherwise, $\sum_{\forall k} p_k^{ij} < 1$ and Equation (1) cannot be satisfied). We define p_k^{ij} more generally as:

$$\Pr\{t_k^{ij} < t_h^{ij}; \forall h \neq k | T\} \leq p_k^{ij} \leq \Pr\{t_k^{ij} \leq t_h^{ij}; \forall h \neq k | T\}; \forall ij$$

and the equilibrium condition as:

$$\Pr\{t_k^{ij} < t_h^{ij}; \forall h \neq k | T\} \leq \frac{x_k^{ij}}{\sum_{\forall k} x_k^{ij}} \leq \Pr\{t_k^{ij} \leq t_h^{ij}; \forall h \neq k | T\}; \forall ij. \quad \text{(1a)}$$

Note, however, that whenever t is modeled as a continuous random variable (as is the case for the model presented in this paper), the LHS and the RHS of Equation (1a) coincide and Equation (1) can be used instead of Equation (1a).

Note also that in the deterministic case in which the perceived travel times equal the measured travel times ($t \equiv T$), the variable t is discrete and Equation (1a) implies that:

$$x_k^{ij} > 0 \quad \text{only if } T_k^{ij} \leq T_h^{ij}\ ; \quad \forall h \neq k$$

$$x_k^{ij} = 0 \quad \text{only if } T_k^{ij} \geq T_h^{ij}\ ; \quad \text{for some } h \neq k.$$

Since this merely states Wardrop's U-E criterion (A), Equation (1a) can be regarded as a generalization of it.

In order to devise a procedure that will yield S-U-E flow patterns (or simply to test whether S-U-E has been achieved) it is necessary to define explicitly how users perceive travel time. In other words, we need the joint probability density function of $t = (\cdots, t_k^{ij}, \cdots)$ conditional on $T = (\cdots, T_k^{ij}, \cdots)$, $f(t|T)$. Section 1 of this paper addresses this issue.

Achieving S-U-E in a network with flow dependent link travel times is a difficult task because the perceived route travel times t_k^{ij} depend on the measured travel times which in turn depend on the route (and link) flows. Thus, Equations (1) appear very complicated except in two cases:

a) When we let t_k^{ij} be equal to T_k^{ij} with probability 1, Equation (1a) reduces to Wardrop's deterministic U-E criterion (and as mentioned above there are algorithms that solve this problem).

b) If the measured link travel times are independent of flow the LHS of Equation (1) is independent of x_k^{ij} and can consequently be computed prior to the flow levels.

This paper deals with the latter case of constant link costs. In this case, users do not have to "compete" for the use of the road and the concept of equilibration is trivial. In order to emphasize this fact, we will refer to this problem as the stochastic network loading (S-N-L) problem.

Two basic approaches have been proposed to perform S-N-L: stochastic simulation methods[12, 13] which may have been inspired by techniques used in connection with PERT networks with random activity durations and Dial's analytic method[11] which is based on a multinomial logistic diversion curve. Although the latter method is deficient,[14-16] it is getting more and more attention; it has been calibrated[17] and is regularly used in practice.[18] Since there seems to be some confusion regarding SNL models (e.g., FLORIAN AND FOX[19] have concluded that the deficiencies of Dial's method are too difficult to overcome and that in uncongested networks any reasonable traffic assignment method should produce results close to "all-or-nothing"), in the remainder of this paper, we attempt to formalize a few more concepts related to the theory of route choice and try to shed some light on the pros and cons of existing S-N-L methods.

We start by providing a generalized formulation of the S-N-L problem which is based on the economist's concept of random utility. Next, based on two reasonable postulates of user behavior, we proceed to identify a class of S-N-L models which do not exhibit the deficiencies in Dial's model.

Section 1 shows that, with the above mentioned postulates, the perceived route travel times are approximately multivariate normally distributed. This readily yields an approximate expression for the LHS of Equation (1). Such expression is complicated but enables us to solve small scale examples by hand. This is done in Section 2 where by means of small contrived networks (similar to the counterexample networks that have appeared in the literature pointing out weaknesses in Dial's method) the proposed method is compared to Dial's and is shown not to possess the latter's deficiencies. Section 2 also contains a discussion of Burrell's simulation method and of two techniques that are proposed to approximate the S-N-L flows resulting from the model presented in this paper.

A summary of the results in this paper is presented in Section 3 which concludes with a discussion of possible generalizations of the model.

1. THE CONCEPTUAL FRAMEWORK

IN THIS SECTION we develop formulas for the probability of route choice, p_k^{ij}, as a function of the set of measured route travel times, T.

Since for S-N-L problems link travel times are independent of flow, the decision of one trip-maker does not affect decisions by other trip-makers and, from a theoretical point of view, it suffices to study the problem for a single trip-maker for a unique origin-destination (O-D) pair. On that account, we drop the superscripts i, j from our notation.

In accordance with our statement of stochastic user equilibrium, we need to define how users perceive travel time since choices will be made based on this perception. We postulate that a user's perceived travel time on route k, t_k, equals the measured travel time on that route, T_k, plus a random error term with zero mean, ξ_k, which will vary from user to user. Thus:

$$t_k = T_k + \xi_k \tag{2a}$$

and

$$E(t_k) = T_k. \tag{2b}$$

Every user then evaluates the travel time on all routes $k = 1, \cdots r$ and selects the route $k_{\min}$ with the minimum perceived travel time.

$$t_{k_{\min}} < t_k \ ; \ \forall k \neq k_{\min}. \tag{2c}$$

In consumer behavior and travel demand theory these models are called *random utility* models. These models form the conceptual basis for the modern approaches to travel demand forecasting[20, 21] and it is not our purpose to evaluate or question the validity of the random utility concept here. We accept Equations (2a) (2b) and (2c) as reasonable postulates and proceed to build a consistent model of route choice based on these equations.

As mentioned in the introduction, we need the joint p.d.f. of the multivariate random variable,

$$t = (t_1 \cdots t_r),$$

as this p.d.f. will enable us to assess the probabilities of choice in conjunction with Equation (2c).

We formulate the following two postulates of motorist behavior:

(A) Nonoverlapping sections of road are perceived independently by the tripmaker.
(B) Sections of the road of equal length are perceived in identical fashion.

Assumption (A) merely states that the error terms for nonoverlapping road sections must be mutually independent random variables. Postulate (B) further requires that the p.d.f. of the error term on sections of equal length should be the same.

One possible interpretation of the assumptions (similar to the explanation of the differences in motorists' route choice provided in reference 22) consists in assuming that motorists use the actual driving times as their perceived travel times, and that before making his final route choice decision, a motorist has actually driven once on all the links of the network. Since actual driving times are subject to variation because drivers cannot adjust their vehicle's speed perfectly, the perceived travel times will differ from driver to driver.

For instance, it has been suggested[23] that the time-space trajectory of a human driven vehicle can be modeled as a Brownian motion process with positive drift equal to the desired speed of the driver. In such situation, the error term, ξ, on any given section of the road for any given driver, corresponds to a realization of a Brownian motion process, and if one assumes that all drivers possess the same desired speed and the same variance per unit length, assumptions (A) and (B) follow.

Postulates (A) and (B) can also be visualized as arising from a similar but more general scenario in which drivers' perception of their actual travel times during the preliminary stage before selecting the route is not accurate. All one has to assume is that the accumulated error term while evaluating a section of road also follows a Brownian motion process (with zero drift and variance per unit length determined by the human capability to perceive accurately travel time).

Postulate (A) does not hold if one allows drivers to have different variances per unit time or different desired speeds. Nevertheless, this deficiency is less serious than assuming that the variance is zero ("all-or-nothing" approach) or that, as in Dial's model the route error terms are independent ard identically distributed.

We now use these postulates to obtain an approximate expression for the probability of route choice.

Defining θ as the variance of the error term, ξ, on a section of the road of unit length we show that on a section of the road of length, T, the corresponding error term, $\xi(T)$, must have zero mean and variance θT.

We first note that:

$$\theta = \mathrm{var}(\xi(1)) \leq \|\Delta T^{-1} + 1\| \cdot \mathrm{var}(\xi(\Delta T)) \leq (\Delta T^{-1} + 1). \quad \mathrm{var}(\xi(\Delta T)) \tag{3a}$$

and

$$\theta = \mathrm{var}(\xi(1)) \geq \|\Delta T^{-1}\| \cdot \mathrm{var}(\xi(\Delta T)) \geq (\Delta T^{-1} - 1). \quad \mathrm{var}(\xi(\Delta T)) \tag{3b}$$

where the symbol $\|x\|$ denotes the integer part of x.

Since Equations (3) can be alternatively written,

$$\frac{\theta}{\Delta T^{-1} + 1} \leq \mathrm{var}(\xi(\Delta T)) \leq \frac{\theta}{\Delta T^{-1} - 1} \tag{4}$$

we note that, as $\Delta T \to 0$, $\mathrm{var}(\xi(\Delta T)) = \theta \cdot \Delta T + o(\Delta T)$.

Consequently, $\mathrm{var}(\xi(T)) = n \cdot \mathrm{var}(\xi(T/n)) = \theta T + n \cdot o(T/n)$, and letting $n \to \infty$ we see that $\mathrm{var}(\xi(T)) = \theta \cdot T$.

A similar argument would yield (letting α be the mean error term on a section of unit length) that $E(\xi(T)) = \alpha.T$; and since as postulated $E(\xi_k) = 0$, $\alpha = 0$ and $E(\xi(T)) \equiv 0$.

The central limit theorem further guarantees that for a given link, lm, of sufficient length, the error terms will be approximately normally distributed with zero mean and variance $\theta T'_{lm}$ (we use primes when denoting variables associated with a link).

The accuracy of this assessment depends to a great extent on the shape of the p.d.f. of the error term for very small road sections. For instance, one could argue that since the random variables t'_{lm} are non-negative, the p.d.f. of the error terms for very small sections of the road should have a long tail to the right (positive skewness) and that, consequently, a non-negative, infinitely divisible distribution (such as the gamma distribution[24] with shape parameter porportional to $\theta T'_{lm}$) would be more appropriate than the normal for short roadway sections.

However, it is not unreasonable to think that the relative magnitude of ξ'_{lm} with respect to T'_{lm} is small for most realistic situations and that consequently even if a gamma distribution were used we would be in the region where the gamma family is well approximated by the normal distribution. Thus,

$$t'_{lm} \dot{\sim} N(T'_{lm} \ ; \ \theta T'_{lm}), \tag{5}$$

and, since by definition links do not overlap, the random variables t'_{lm} are also mutually independent.

We now obtain the distribution of the perceived route travel time vector, t. Let us define the link-route incidence matrix, Δ, for the network; its terms, $\delta_{ij,k}$, are:

$$\begin{aligned} \delta_{ij,k} &= 1 \text{ if link } ij \text{ belongs to route } k \\ &= 0 \text{ otherwise.} \end{aligned}$$

Clearly,

$$t_k = \sum_{\forall ij} t'_{ij} \cdot \delta_{ij,k};$$

or using vector notation and letting t' be the vector of perceived link travel times, we have:

$$t = t' \cdot \Delta;$$

and since t' is a vector of mutually independent normal random variables, the linear transformation $t' \cdot \Delta$ yields a MVN random variable.

The mean of t is, by definition, T, and the variance-covariance matrix can be easily obtained.

Letting t_{kp} and T_{kp} denote respectively the perceived and measured travel times on the road section shared by routes k and p, it is easy to see that,

$$\mathrm{var}(t_k) = \theta T_k \tag{6a}$$

and

$$\mathrm{cov}(t_k, t_p) = \mathrm{var}(t_{kp}) = \theta T_{kp}. \tag{6b}$$

Not surprisingly, we see that the distribution of t is completely characterized by three factors:

- the measured route travel times, T
- the accuracy with which people can perceive travel time (the parameter θ), and
- *the topology of the network.*

Since the distribution of t determines the probability of choice, all three factors affect the probability of choice.

As is discussed in the next section, Dial's flow allocation formula fails to capture the last factor and this causes it to produce unreasonable results when the topology of the network is altered.

The probability of selecting route k is given by:

$$\begin{aligned} p_k &= \Pr[t_k < t_i;\ \forall i \neq k] \\ &= \Pr[t_k - t_i < 0;\ \forall i \neq k]. \end{aligned}$$

If

$$_kt = (t_k - t_1; t_k - t_2; \cdots t_k - t_{k-1}; t_k - t_{k+1}; \cdots ; t_k - t_r)$$

then

$$p_k = \Pr[_kt < (0, 0, \cdots, 0)]. \tag{7}$$

However, $_kt$ is a linear combination of t,

$$_kt = tL_k$$

with L_k:

$$L_k = \begin{array}{rc} & \begin{array}{cccccccc} 1 & 2 & & k-1 & k+1 & & r-1 & r \end{array} \\ \begin{array}{r} 1 \\ 2 \\ \vdots \\ k-1 \\ k \\ k+1 \\ \vdots \\ r-1 \\ r \end{array} & \left[\begin{array}{cccccccc} -1 & 0 & — & 0 & 0 & — & 0 & 0 \\ 0 & -1 & — & 0 & 0 & — & 0 & 0 \\ & & & & & & & \\ 0 & 0 & — & -1 & 0 & — & 0 & 0 \\ 1 & 1 & — & 1 & 1 & — & 1 & 1 \\ 0 & 0 & — & 0 & -1 & — & 0 & 0 \\ & & & & & & & \\ 0 & 0 & — & 0 & 0 & — & -1 & 0 \\ 0 & 0 & — & 0 & 0 & — & 0 & -1 \end{array}\right] \end{array},$$

and therefore ${}_kt$ is also MVN with mean TL_k and variance-covariance matrix $L_k^t\sum L_k$ (in our notation L_k^T is the transpose of L_k and $\sum$ is the variance-covariance matrix of the perceived route travel times).

It is easy to see that:

$$E(t_k - t_p) = T_k - T_p, \tag{8}$$

and that

$$\begin{aligned}\operatorname{cov}(t_k - t_p,\, t_k - t_q) &= \operatorname{cov}(t_k, t_k) + \operatorname{cov}(t_p, t_q) - \operatorname{cov}(t_p, t_k) \\ &\quad - \operatorname{cov}(t_q, t_k) = \theta[T_k + T_{pq} - T_{pk} - T_{qk}].\end{aligned} \tag{9}$$

Therefore p_k is given by the cumulative joint distribution function of a MVN variate (with mean and variance-covariance matrix given by (8) and (9) evaluated at the origin (Equation (7)).

This is difficult to do computationally for large numbers of dimensions, however, small examples (such as the contrived networks that have appeared in the literature pointing out the weaknesses in Dial's model) can be studied by hand by means of approximate formulae.[25] We now illustrate how to use the approximation to evaluate the probabilities of route choice with an example.

Example:

Take the network in Figure 1; we want to predict p_1, p_2, and p_3. We use the results in the previous pages to obtain the distribution of t:

$$t = (t_1;\, t_2;\, t_3) \sim \text{MVN}\left[(3, 2, 2);\, \theta\begin{pmatrix}3 & 1 & 1\\ 1 & 2 & 0\\ 1 & 0 & 2\end{pmatrix}\right].$$

We now obtain p_1,

$$L_1 = \begin{array}{c c} & \begin{array}{cc} 2 & 3 \end{array} \\ \begin{array}{c} 1\\ 2\\ 3 \end{array} & \begin{pmatrix} 1 & 1\\ -1 & 0\\ 0 & -1 \end{pmatrix} \end{array}$$

$$E({}_1t) = (3, 2, 2)\begin{pmatrix}1 & 1\\ -1 & 0\\ 0 & -1\end{pmatrix} = (1, 1)$$

$$\text{var-cov}({}_1t) \doteq \theta\begin{pmatrix}1 & -1 & 0\\ 1 & 0 & -1\end{pmatrix}\begin{pmatrix}3 & 1 & 1\\ 1 & 2 & 0\\ 1 & 0 & 2\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & 0\\ 0 & -1\end{pmatrix} = \theta\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}$$

$${}_1t \sim \text{BVN}\left[(1, 1);\, \theta\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}\right].$$

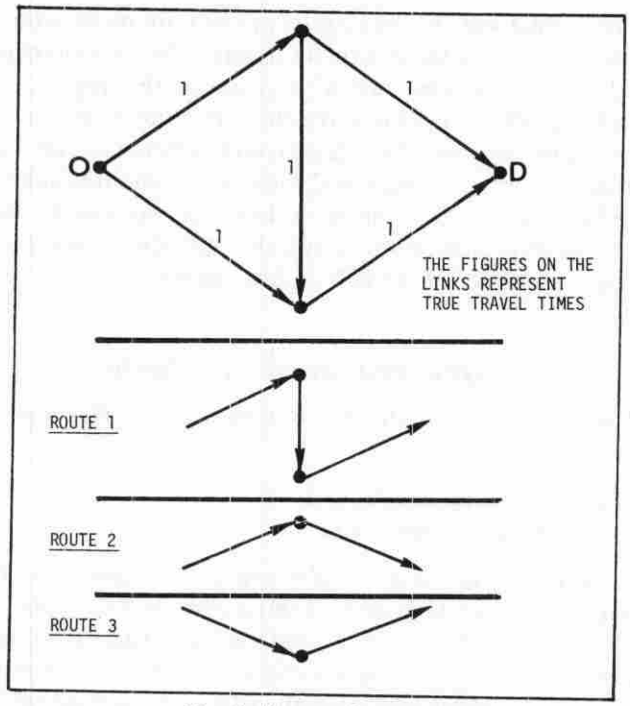

Fig. 1. A network example.

In order to find $\Pr[{}_1t < 0]$ we must find:

$$p_1 = \int_{-\infty}^{-1} \int_{-\infty}^{-1} \frac{1}{2\pi\sqrt{8\theta}} \exp\left\{-1/2\ (x, y)\, \frac{1}{8\theta} \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\} dxdy,$$

which, for $\theta = 1/9$ (a reasonable value if our unit of measurement is minutes), yields $p_1 \simeq 0.007$.

Clark's recursive formulae[25] when applied to this case yield:

$$\max({}_1t) \stackrel{\cdot}{\sim} N(1 + (2\theta/\pi)^{1/2};\, \theta(3 - 2/\pi)\,);$$

and since

$$P_r\{{}_1t < (0, 0)\} = \Pr\{\max({}_1t) < 0\}\ ,$$

$$p_1 \simeq \Phi[-(1+(2\theta/\pi)^{1/2})(\theta(3-2/\pi))^{-1/2}] = \Phi(-2.47) = 0.0068,$$

which coincides with the exact value. Reference 26 contains a detailed discussion and evaluation of Clark's approximation formulae.

As the number of alternatives becomes large (say larger than three or four), the exact numerical approach is no longer feasible. However, Clark's recursive technique can be efficiently applied to more alternatives. In the next section, we use these formula to quantify the differences between Dial's method and the formulation proposed in this paper.

At present, Clark's formulae have not been imbedded in a computer algorithm because this would require an enumeration of the n most likely paths between every O-D pair and this is a computationally unsolved problem. This is not a major problem, however, because as shown in the sequel of the paper, simulation approaches (if adequately executed) can yield unbiased estimates of the S-N-L flow pattern at a considerable lower cost.

2. APPRAISAL OF S-N-L METHODS

THIS SECTION contains a quantitative comparison of the two most widely used S-N-L techniques,

1) Dial's analytical approach, and
2) Burrell's simulation approach,

with the proposed formulation. It also contains a description of a simulation procedure that can be used to approximate S-N-L flows.

1. Dial's analytical procedure consists in allocating traffic among alternate routes by means of a diversion curve. For practical reasons, it works with a set of *reasonable* or *efficient* paths which are defined for each origin-destination pair. A path is efficient if all of its links place the user *closer* to the destination and *farther away* from the origin.

The algorithm works in such a way that the resulting traffic flows among competing efficient paths for a given O-D pair are allocated according to probabilities, p_k:

$$p_k = \frac{\exp\{\theta^*(T_o - T_k)\}}{\sum_{j\epsilon P} \exp\{\theta^*(T_o - T_j)\}} \qquad \text{if } k\epsilon P \qquad (10a)$$

$$= 0 \qquad \text{if } k\notin P \qquad (10b)$$

where:

P = set of all the efficient paths for the given O-D pair.
T_o = measured travel time along the shortest path.
T_i = measured travel time along the ith path.
θ^* = diversion intensity parameter (to be calibrated or selected by the user).

At this point, we note that Equation (10a) is not affected by the topology of the network (as was discussed in the previous section), and that it is the *logit* formulation of travel demand modeling which arises when the error terms, ξ_k, associated with the travel time along the routes between the O-D pair are independent identically distributed (i.i.d.) Gumbel variates.[20, 21]

However, as has been pointed out in the literature, its traffic flow patterns are not reasonable when routes overlap heavily[14, 15, 19] and it is sensitive to network representation.[15, 16] It is shown below that these two flaws are not exhibited by the formulation proposed in this paper.

a) *Flow allocation to overlapping routes*

A network similar to the one displayed in Figure 2 is the counter-example used by most authors[14, 15, 19] when discussing Dial's method. The argu-

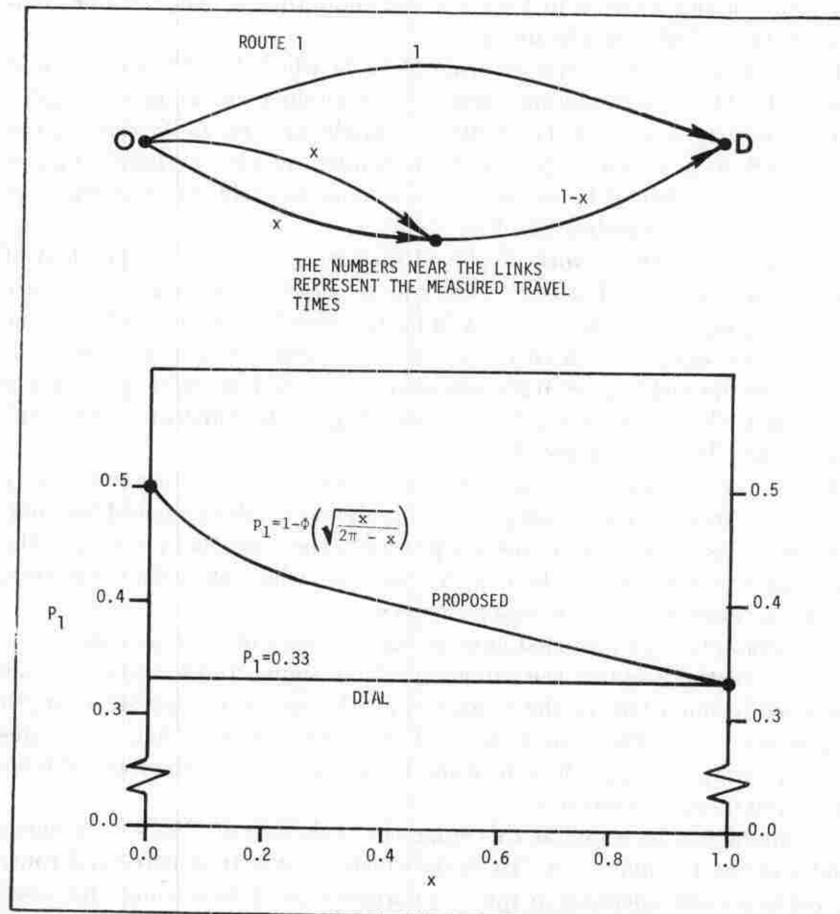

Fig. 2. Influence of route overlap on the probability of route choice.

ment is that when x is close to 0 the fraction of flow on route 1, p_1, should be close to 1/2 since the other two routes will be perceived as one by the motorists. However, since Dial's flow allocation formula is not sensitive to network topology (it is sensitive only to route length and to the motorist's ability to accurately perceive travel time), the flow assigned to route 1 is always 1/3 regardless of the amount of overlap between the other two routes.

The fraction of flow diverted to route 1 according to both the model presented in this paper and Dial's formula is plotted in Figure 2. Note in particular how the proposed model reproduces what is generally considered to be reasonable behavior.

We now look at the behavior of the proposed formulation for different variations of the network in Figure 2 and show that it behaves more reasonably than Dial's logit formula.

Figure 3a displays a network from O to D which has 17 equally long routes. In Dial's formulation then the probability of using the direct route (route 1) is $p_1 = 1/17$ (0.06). A straightforward application of the model of Section 1 results (p_1 is also independent of θ in this case) in $p_1 = 0.19$. If the number of loops, m, varies so does the relative difference between the two approaches (see Fig. 3b).

Consider now the network displayed in Figure 4a (p_1 is independent of θ in this case, as well). For $m = 1$ two routes with identical travel times are available and, clearly, $p_1 = 0.5$. Adding two short links in parallel to one of the routes results in three routes ($m = 3$) that compete with route 1. Dial's formula yields $p_1 = 0.25$ instead of the more reasonable $p_1 = 0.425$ obtained with the proposed formulation. Figure 4b illustrates these differences for different values of m.

In the last two examples, one reasonably expects p_1 to decrease slowly with the number of routes (as properly predicted by the proposed formulation) since the more new routes one provides, the more likely it is that the trip-maker *believes* one of these is the shortest. Dial's algorithm, however, is far too sensitive to these minor changes.

The examples just provided have illustrated some of the biases that arise in Dial's method and at the same time have shown the adequacy of the proposed formulation. In the examples, all the routes had the same length and consequently the error terms had the same variance; had the routes been different in length the errors and biases of Dial's method would have been even more pronounced.

It should also be noted at this stage that this flaw of Dial's assignment model is not peculiar to it. Methods which assume that perceived route travel times are independent random variables (such as the ones discussed in reference 22) also exhibit the pathological behavior depicted in Figure 2

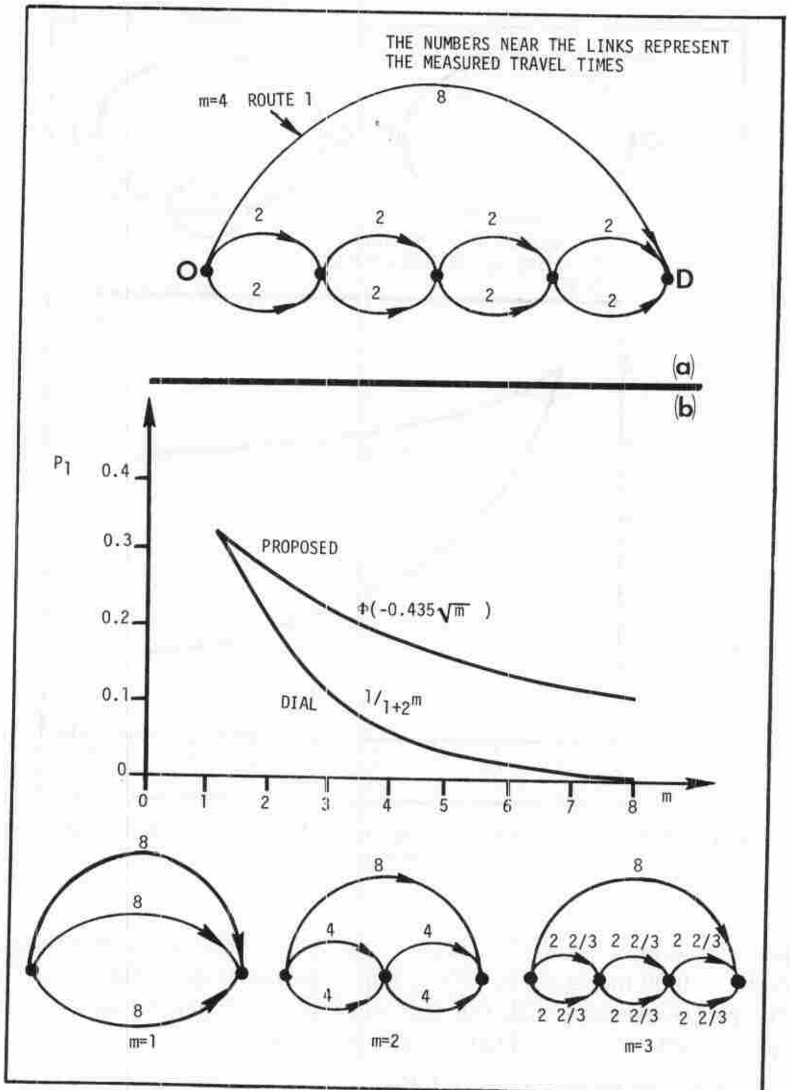

Fig. 3. Influence of number of overlapping routes on the probability of route choice. Case I.

since in this case the probability of route choice depends only on the length of the routes. Von Falkenhausen's method,[13] although a commendable first attempt at capturing the effects of route overlap, also exhibits this deficiency since (as described in reference 13), for any given route, only the overlap with the shortest is taken into account. Consequently, if in

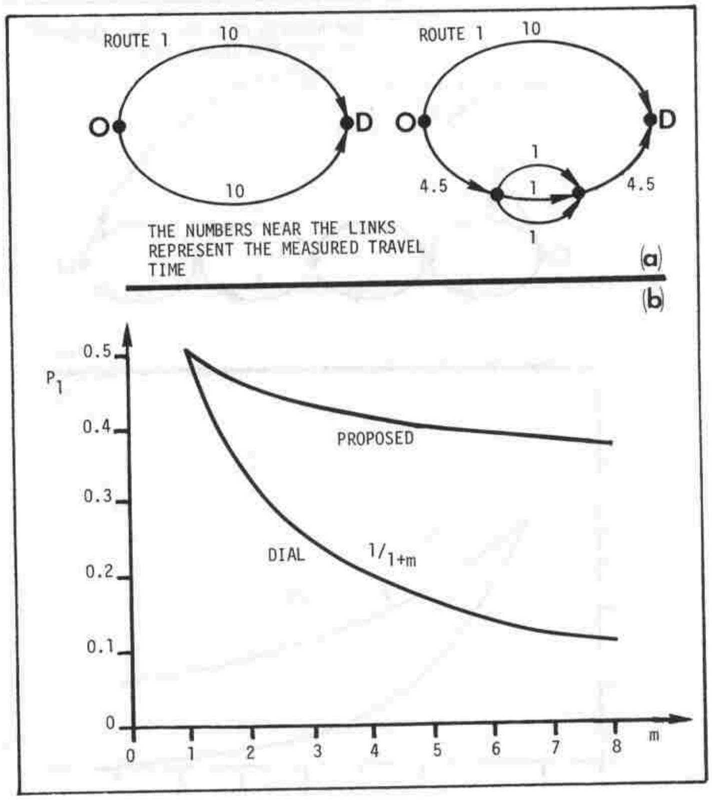

Fig. 4. Influence of number of overlapping routes on the probability of route choice. Case II.

Figures 2, 3a and 4a we let Route 1 be the shortest by a small amount (say $T_1 = 0.99$ in Figure 2, 7.99 in Figure 3a and 9.99 in Figure 4a), none of the routes overlaps with the shortest and Von Falkenhausen's strategy leads to results similar to Dial's in all three cases.

b) *Sensitivity to network representation*

Another flaw of Dial's algorithm lies in the definition of efficient paths since it is possible to transform efficient paths into inefficient paths by adding dummy nodes to the network. Although this characteristic of the model is somewhat undesirable (it is dramatized in Figure 5 where construction of a new freeway results in unreasonable loss of traffic) we need not be concerned about it since it is not directly related to the flow allocation formula.

2. Burrell's simulation method disperses flows over the network in a

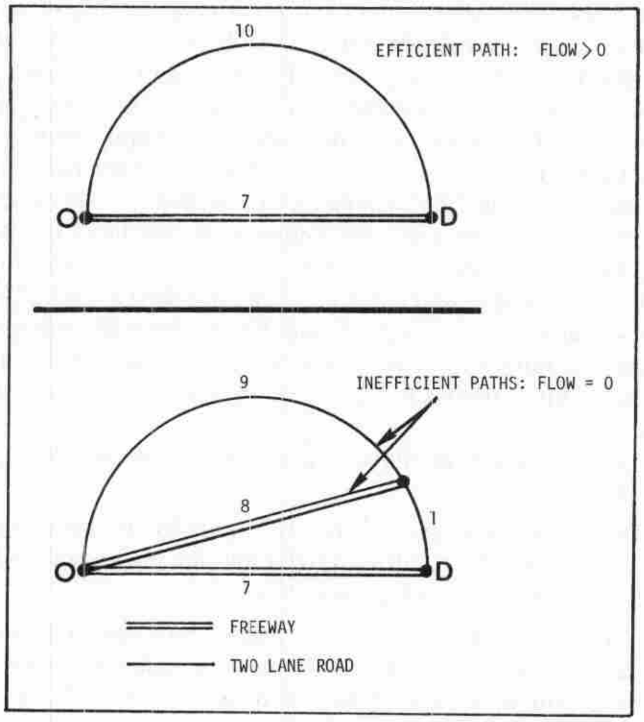

Fig. 5. Inconsistency of Dial's efficient path definition.

different way. In the method, which has a sequence of steps very similar to incremental assignment methods, one origin is considered at a time. For each origin, one obtains a set of link travel times (by sampling each link travel time from a uniform distribution with variance to mean ratio equal on all links), finds the corresponding tree, and loads all the trips emanating from that origin onto the network.

Two basic shortcomings can be identified:

a) Since the method is a stochastic simulation, it must rely on the laws of the large numbers and on repetitive sampling to converge. Burrell, however, samples link travel times only once for each origin and this may cause the results of the algorithm to vary considerably from execution to execution. Perhaps for very homogeneous and symmetric cases with many O-D pairs the errors will cancel out on the links with the highest flow levels but this, as yet, remains to be proven.

b) The second shortcoming is much less important from a practical point of view but theoretically interesting. Using a distribution that is not stable (i.e., that it does not "conserve its shape" when convoluted) is not desirable because, as is illustrated below, the resulting traffic flow patterns may depend on the way the network is represented. Consider, for example, a route consisting of only one link. In this case, the route travel time is directly sampled from a uniform distribution. If one now wishes to use a finer representation of the network, which, without changing the number, length, or general characteristics of the available routes, uses two links for the above mentioned route, Burrell's strategy yields a route travel time which is indirectly being sampled from a triangular distribution. The resulting assignments, thus, depend slightly on network representation. (In a recent paper,[15] Burrell argues that a normal distribution would have been more desirable but that the uniform is computationally easier.)

Burrell's method can be modified to attenuate the impact of these shortcomings as follows:

Step 0. Divide all the entries of the O-D table by an arbitrary integer, N. We call the resulting O-D table an increment and assume that we have N such increments.

Step 1. Exercise the algorithm as it exists (but sampling link travel times from normal distributions with same variance to mean ratio) for each one of the increments and add the resulting flows on each link.

This modified simulation algorithm is very closely related to our definition of S-U-E. (The general strategy is also reminiscent of the simulation algorithm used by Von Falkenhausen to approximate his stochastic route choice model [reference 13, p. 417[). Note that one can visualize the mean of the distribution used to sample the link travel times as the measured travel times, T'_{lm}, of Section 2, and each outcome of the sampling process as an individual's perceived travel time, t'_{lm}. The determination of the shortest paths using the results of the sampling process corresponds to Equation (2b) in this paper. As a matter of fact, it is easy to see that with the above modification, and as $N \to \infty$, the resulting flow pattern converges in probability to the link flow pattern derived from the route choice model proposed in this paper. Moreover, as indicated by tests carried out by Von Falkenhausen with his simulation model,[13] it seems that between 5 and 10 iterations are sufficient and that, consequently, the above described simulation algorithm can be used to approximate the S-N-L flow patterns suggested in this note at a cost comparable with that of existing mathematical U-E procedures.

An alternate, analytical, two-stage approach to perform S-N-L could

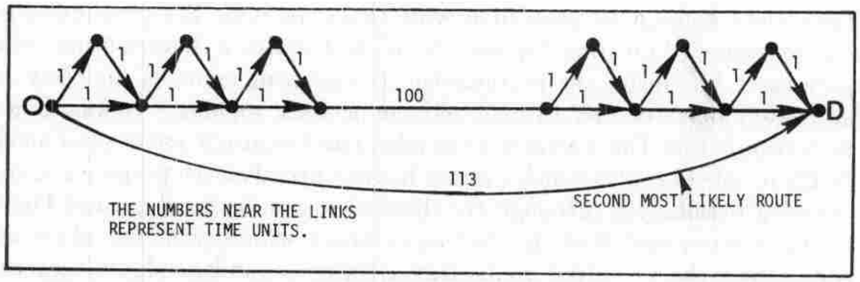

Fig. 6. Network where the longest route is actually the second most likely to be used.

consist of a route enumeration stage performed by humans and a flow allocation stage performed by the computer. The advantage of using humans for the route enumeration process is based on efficiency and accuracy considerations. Although a computer can find the n shortest routes in a graph there is no guarantee that the routes found would be adequate for our purposes since, in S-N-L, the most likely two or three routes are not always among the shortest. As a matter of fact, there is no guarantee whatsoever that the n shortest routes will contain the two or three more likely routes. Figure 6, for instance, depicts a network where the second most likely route is actually the longest (the 65th shortest) and where finding the five or six shortest paths would have been useless (even more dramatically, it is possible to construct examples—topologically similar to the networks in Figures 3 and 4—where the most likely route is also the longest!). Humans, on the other hand, can be looking at a map very quickly identify (perhaps on an interactive graphics computer terminal) the best routes and at the same time weed out the routes which although short are obviously dominated by shorter overlapping routes. Although we have not experimented with interactive graphics methods yet, they are probably competitive with the simulation approach on both the cost and accuracy aspects.

3. SUMMARY

SINCE FOR NETWORKS which are very lightly congested S-N-L seems to be a more desirable technique than deterministic U-E methods, and there seems to be some confusion in the literature regarding the relative merits of S-N-L methods, this paper has attempted to formalize the concept of stochastic user equilibrium and has evaluated alternate techniques to perform S-N-L.

Section 1 contains the development of a model of route choice which, by being based on two simple behavioral postulates, does not suffer from the serious deficiencies associated with Dial's method. The probability of choice depended on three factors: the route lengths, a diversion intensity parameter θ (which may be regarded as capturing humans' inability to accurately perceive travel time), and the network topology. It was shown in Section 2 that Dial's model is not adequate because it assigns too much traffic to sets of routes which overlap heavily and that the proposed model behaved reasonably. Although the discussion was centered around Dial's model it was noted that this deficiency arises with any model which assumes that the perceived route travel times are independent (e.g., the special cases discussed in reference 22) or with models which do not capture adequately the effects of overlapping routes (as happens in Von Falkenhausen's model[13]). The section concludes with a discussion of Burrell's simulation method and the presentation of two approaches to perform S-N-L.

The model presented in this paper can be generalized in more than one way. For instance, by letting the parameter θ either be random or take two or three different values one is allowing different motorists to have different abilities to perceive travel time. Such a model would be more realistic but one would have to calibrate from the data at least one more parameter.

If one introduces, by means of a generalized cost function, route attributes other than travel time (e.g., cost, safety, $\cdots$) in the determination of the best route, the model can be further generalized by letting different users put different weights on the elements of the generalized cost. This problem is also solvable (it is called the *taste variation* problem in econometrics) but would require calibration of up to two additional parameters for every additional attribute.

As a closing comment, we wish to state that the approximate formulas used in this paper to evaluate the probability of route choice can also be applied to other choice problems in travel demand forecasting (destination, mode, $\cdots$). Thus, prediction and estimation of a multinomial probit model has become entirely possible.[26, 27] This frees travel demand models of the nagging fallacy of the independence of the irrelevant alternatives axiom.

ACKNOWLEDGMENTS

THE AUTHORS wish to thank Professor Donald E. Cleveland of the University of Michigan for his valuable and constructive comments and Dr. Caroline Fisk of the University of British Columbia for her helpful remarks during the refereeing process.

REFERENCES

. J. G. WARDROP, "Some Theoretical Aspects of Road Traffic Research," *Proc. Inst. Civ. Eng.* (*Part II*) **1,** 325–378 (1952).

2. R. SMOCK, "An Iterative Approach to Capacity Restraint on Arterial Networks," *Highway Res. Board Bull.* *347*, 60–66 (1962).
3. N. A. IRWIN, N. DODD AND H. G. VON CUBE, "Capacity Restraint in Assignment Programs," *Highway Res. Board Bull. 297*, 109–127 (1961).
4. B. V. MARTIN AND M. L. MANHEIM, "A Research Program for Comparison of Traffic Assignment Techniques," *Highway Res. Rec. 88*, 69–84 (1965).
5. M. J. HUBER, H. B. BOUTWELL AND D. K. WHITHFORD, "Comparative Analysis of Traffic Assignment Techniques with Actual Highway Use," *N.C.H.R.P. Report 58*, 1968.
6. J. A. FERLAND, M. FLORIAN AND C. ACHIM, "On Incremental Methods for Traffic Assignment," *Trans. Res.* **9,** 237–239 (1975).
7. L. J. LEBLANC, E. K. MORLOK AND W. P. PIERSKALLA, "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem," *Trans. Res.* **9,** 309–318 (1975).
8. S. NGUYEN, "An Algorithm for the Traffic Assignment Problem," *Trans. Sci.* **8,** 203–216 (1974).
9. C. F. DAGANZO, "Some Remarks on Traffic Assignment Methodo'ogy Selection," Working Paper 7703, Stochastic Traffic Assignment Project, Institute of Transportation Studies, University of California, Berkeley, Calif., 1977.
10. Y. SHEFFI, "The Road Network Analysis Model," Vol. 4 of "Transportation Planning in Developing Countries," U.S.A.I.D. Report 2 (Contract AID/TA-C-1107), 1977.
11. R. B. DIAL, "A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration," *Trans. Res.* **5,** 83–111 (1971).
12. J. E. BURRELL, "Multipath Route Assignment and Its Application to Capacity Restraint," *Fourth International Symposium on the Theory of Traffic Flow*, Karlsruhe, Germany, 1968.
13. H. VON FALKENHAUSEN, "Traffic Assignment by a Stochastic Model," *Proceedings of the Fourth International Conference on Operational Research*, pp. 415–421, 1966.
14. M. SCHNEIDER, "Probability Maximization in Networks," *Proceedings of the International Conference on Transportation Research*, Bruges, Belgium, pp. 748–755, 1973.
15. J. E. Burrell, "Multipath Route Assignment: A Comparison of Two Methods," *Traffic Equilibrium Methods*, Lecture Notes in Economics and Mathematical Systems, Volume 118, Springer-Verlag, New York, 1976.
16. R. L. TOBIN, "Multipath Traffic Assignment Based on Tripmakers' Perceptions," *Research Publication GMR-2139* Transportation and Urban Analysis Department, Research Laboratories, General Motors Corporation, Warren, Mich., 1976.
17. P. ROBILLARD, "Calibration of Dial's Assignment Methods," *Trans. Sci.* **8,** 117–125 (1974).
18. M. H. RAPP, P. MATTENBERGER, S. PIGUET AND A. ROBERT-GRANDPIERRE, "Interactive Graphics System for Transit Route Optimization," *Trans. Res. Rec. 559,* 73–89 (1976).

19. M. FLORIAN AND B. FOX, "On the Probabilistic Origin of Dial's Multipath Traffic Assignment Model," Publication #29, Centre de Reserche sur les Transport, Université de Montréal, September 1975. *Trans. Res.*, **10,** 339–341 (1976).
20. D. MCFADDEN, "The Measurement of Urban Travel Demand," *J. Public Econ.* **3,** 303–328 (1974).
21. M. BEN-AKIVA, "Structure of Passenger Travel Demand Models," *Trans. Res. Rec. 526*, 26–42 (1974).
22. H. BEILNER AND F. JACOBS, "Probabilistic Aspects of Traffic Assignment," in G. F. Newell (Ed.) *Traffic Flow and Transportation*, pp. 183–194. Elsevier, New York, 1972.
23. G. F. NEWELL, "The Flow of Highway Traffic through a Sequence of Synchronized Traffic Lights," *Opns. Res.* **8,** 390–405 (1960).
24. W. FELLER, *An Introduction to Probability Theory and Its Application*, Vol. II. John Wiley & Sons, New York, 1971.
25. C. E. CLARK, "The Greatest of a Finite Set of Random Variables," *Opns. Res.* **9,** 145–162 (1961).
26. C. F. DAGANZO, F. BOUTHELIER AND Y. SHEFFI, "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm." (Submitted for publication to *Trans. Sci.*)
27. C. F. DAGANZO, F. BOUTHELIER AND Y. SHEFFI, "An Efficient Approach to Estimate and Predict with Multinomial Probit Models." Presented at the 56th Meeting of the Transportation Research Board, January 1977.

(Received, July 1976; revised, August 1977)