大韓交通學會誌 第二卷 第一號, pp. 69~88, 1984

□論 文□

Evaluating Value-of-Time for Travel Based on The Linear Logit Choice Model

「로-짓」模型에 의한 交通時間價値 算定에 관한 研究

俞 浣*

(延世大學校 工科大學 敎授)

目 次

要 約

通行時間價値에 대한 연구는 獨立變數를 利用한 通行時間과 費用에 의한 選擇模型을 대개 사용하여 왔다. 通行時間과 費用에 대한 媒介變數를 추정한 후에 通行時間價値를 交通時間과 費用의 「파라메타」比로써 얻고 있다. 그러나 通行時間價値에 관한 信賴區間과 統計的 有意性을 設定하기 위한 적절한 과정은 아직 研究되지 않고 있으나 「로-짓」모형 자체에 대하여는 되어있다.

本 論文은 通行時間價値를 測定하기 위한 旣存의 方法을 批評的으로 檢討하여 본것으로 通行時間價値의 統計的 有意性 檢證을 위한 標準化 過程은 아직 理論的으로 定立 可能하다고 생각되지 않지만 通行時間價値評價의 簡便式을 提示하여 보았다. 簡便式에 의하여 提示된 結果에 의하면 選擇模型에 입각한 通行時間價値의 統計的信賴區間이 매우 큰것으로 나타났는바 이에 대한 選擇模型에 입각한 通行時間價値 측정을 위한 과정의 妥當性에 대한 많은 의문이 아직도 남는다 하겠다.

* 本學會 正會員·都市計劃學博士

I. INTRODUCTION

The method currently used for evaluating the value-of-time (VOT) for travel is to construct a linear logit (modal) choice probability model, and derive the VOT within the structure of this model. Let P_i be the probability that alternative i is chosen for a trip when a choice set I is offered to an individual trip maker. Let t_i and c_i be the travel time and travel cost, respectively, if alternative i is chosen, and a_i, b_t and b_c be the statistically obtained coefficients. The linear logit choice probability model is, then, expressed as

$$P_i = \frac{\exp U_i}{\sum_j \exp U_j} \qquad (1)$$

where

$$u_i = a_i + b_t\, t_i + b_c.\, c_i$$

Hence, the VOT based on the choice model is given by

$$VOT = \frac{\partial u_i/\partial t}{\partial u_i/\partial c} \qquad (2)$$

$$= \frac{b_t}{b_c}$$

An examination of the above practice reveals that the VOT theory $U \in \mathcal{u}$ from all possible theories $\mathcal{u}$ consists of two sub-theories: the behavioral choice theory $T \in \mathcal{J}$ from its domain $\mathcal{J}$ of all possible theories, and the VOT theory $V_T \in \upsilon_T$ from its domain υ_T of consistent theories with T. Therefore, the universe set u of the overall VOT theory U is the Cartisian product of the pair (T, υ_T) over all possible $T \in \mathcal{J}$, i.e.,

$$U = \prod_{T \in \mathcal{J}} (T, \upsilon_T) \qquad (3)$$

The possibilities that can arise in testing the overall VOT theory U are:

1. T and V_T are true. $\Rightarrow$ U is ture.
2. T and/or V_T is false. $\Rightarrow$ U is false.

Therefore, if we want to employ conventional statistical hypothesis testing techniques to establish the significance for VOT estimate, we have to do so for both theories T and V_T in a consistent manner.

A number of theory testing procedures have been proposed for models explaining choice behavior, such as Stopher (1975), McFadden (1973), Boyce et al. (1974), Hansen (1975), and Goldfeld and Quandt (1972). While the theory testing procedures for choice beahvior models are standardized by the likelihood ratio test, there appears to be no procedure for testing the theory of the value of time. This paper reviews three aspects of these testing procedures: the general theory testing procedure, the theories to be tested, and the construction of a testable hypothesis. Since there is no standard testing procedure available that is satisfactory from the point of view of probability theory, an ad hoc approach for establishing the significance of the value of time estimate is proposed.

II. GENERAL THEORY TESTING

An empirical choice theory is constructed to determine some "best" estimate, e.g., maximum likelihood estimate, of unknown choice probability distribution on the basis of certain theoretical statements about the choice behavior. For example, given any random observation $Y = \{Y_{is}\}$, over a finite number of alternatives $i=1, ..., m$, $m \geq 2$, and individual choice makers, $s=1, ..., S$, with finite outcome space Y, we wish to determine an unknown probability distribution $P = \{P_{is}\}$ over Y on the basis of some theoretical statement T about the choice probability distribution. Let the set of all possible probability distributions over Y be

$$P = \{p \mid \sum_i p_{is} = 1, p_{is} \geq 0 \text{ for all } s=1, ..., S\} \cdots \quad (4)$$

and let $P_T \leq P$ denote the set of distributions in P consistent with T.

$$P_T = \{p \in P \mid T\} \cdots \quad (5)$$

Then the problem is to determine a best representative distribution $P_T \in P_T$ consistent with the theory T over Y. In this context, the maximum likelihood approach to determine P_T over Y is to find some $\hat{P}$ that maximizes a likelihood probability $L_T(P|Y)$ of p in P_T,

$$\hat{P} \in \left\{ p_T \mid L_T(P_T|Y) = \max_{p \in P_T} L_T(p|Y) \right\} \cdots \quad (6)$$

To compute $L_T(P|Y)$, observe first that

If we note $N = \{n_{is}\}$ as the number of trips observed for i by s from all possible trip observations of size $N = \sum_s \sum_i n_{is}$, and $X = \{x_{is}\} = \{t_{is}, c_{is}\}$ as the descriptions of i relevant to s from all possible descriptions $\supset\subset$ of alternatives $i=1, ..., m$, then the observation Y consists of $N = \{n_{is}\}$ and $x = \{\underline{x}_{is}\}$;

$$Y = \{N \vdots X\} \quad \cdots\cdots (7)$$

Let the theory T be specified as a choice model, f, which consists of a set of behavioral assumptions and a model specification, expressed as choice probabilities of choosing i by s ;

$$P_{is} = f(\underline{x}_{is} ; \underline{x}_{1s}, ..., \underline{x}_{ms}) \text{ for all } s=1, ..., S. \quad \cdots\cdots (8)$$

Then, the likelihood of p given observation Y under T is defined to be

$$L_T(P \mid Y) = \prod_s \prod_i f(x_{is}; \underline{x}_{1s}, ..., x_{ms})^{n_{is}} \quad \cdots\cdots (9)$$

The "best representative" among all possible distributions under any theory is then defined to be the unconstrained maximum likelihood estimate p* of the true distribution given our sample Y;

$$p^* \in \{p \mid L(P^* \mid Y) = \max_{p \in P} L_A (p \mid Y)\} \quad \cdots\cdots (10)$$

Where L_A (P|Y) is the likelihood of P given Y under any theory. For finite distributions, this unconstrained maximum always corresponds to the relative frequency distribution of the sample itself.

$$p^*_{is} = \left\{\frac{n_{is}}{\sum_j n_{js}}\right\} \quad \cdots\cdots (11)$$

Thus the likelihood of $p \in P$ is defined as

$$L_A(p \mid Y) = \prod_s \prod_i \left(\frac{n_{is}}{\sum_j n_{js}}\right)^{n_{is}}. \quad \cdots\cdots (12)$$

A simple hypothesis test, then, can be regarded as having the form

null hypothesis $H_T : p=\hat{p}$
alternative hypothesis $H_A : p=p^*$.

Hence, following the standard Neyman-Pearson argument for simple hypothesis tests, the most powerful test of H_T for any given level of significance α is defined by the critical region C_T^{α} in the sample space $\supset\subset_N$ of all possible samples of size N from $\supset\subset$ such that the likelihood ratio values λ_T are as large as possible :

$$\lambda_T = \frac{L_T(\hat{p}|Y)}{L_A(p^* Y)} = \frac{\prod_s \prod_i f(x_{is}; x_{1s}, \ldots, x_{ms})^{n_{is}}}{\prod_s \prod_i \left(\frac{n_{is}}{\sum_j n_{js}}\right)^{n_{is}}} \quad \ldots\ldots (13)$$

If the set of distributions P_T consistent with the theory T constitutes a k dimensional subspace (Number of parameters) within the S (m-1) dimensional space P of all distributions, it is well known that under the null hypothesis H_T, the statistic -2 log λ_T is asymptotically Chi-square distributed with S (m-1)-k degrees of freedom. The importance of reviewing the likelihood ratio procedure for theory testing is to emphasize its relation to the overall theory testing, which includes a set of behavioral assumptions, a model specification, and parameter estimates.

If the overall theory is accepted (cannot be rejected) the likelihood ratio test can be extended to test the significance of parameter values, commonly known as a goodness of fit test, under the given set of behavioral assumptions and model specification. Suppose we have k (or more commonly k + m-1) parameters, $\underline{b} = (a_1, \ldots, a_{m_1}, b_t, b_c)$, from the space Ω, in expressing the choice probabilities for each s=1, ..., s ;

$$P_T = \left\{ p(\underline{x}_s ; \underline{b}) ; s=1, \ldots, S. \right\} \quad \ldots\ldots (14)$$

Let ω be a subset of Ω. If $\underline{b}$ is the true value of the parameters in the popolation, we set up the statistical null hypothesis H_o that $\underline{b} \in \omega$ against the alternative hypothesis H_a that $\underline{b} \in \Omega - \omega$. In practice, the parameter values are judged significant if $\underline{b} \neq \underline{0}$ for a given level of significance $\underline{\alpha}$.

$$H_o : \underline{b} \in \omega = \{0\}$$

$$H_a : \underline{b} \in \Omega - \omega.$$

Let the new theory T_o consist of the same behavioral assumptions and a model specification, given Y and $\underline{b} \in \omega$. Assuming that the maximum likelihood estimate is $\hat{\underline{b}} \neq \underline{0}$, the two competing hypotheses can be written as

H_o : T_o is true.
H_a : T_a is true.

Then the likelihood ratio for this problem is given by

$$\lambda = \frac{L_{H_0}(\hat{P}|\underline{X})}{L_{H_a}(P^*|\underline{X})} = \frac{P(\underline{X}\,|\,\underline{b}=\underline{0})}{P(\underline{X}\,|\,\underline{b}=\hat{\underline{b}})} = \frac{\prod_s \prod_i P(\underline{X}_s;0)^{n_{is}}}{\prod_s \prod_i P(\underline{X}_s;\hat{\underline{b}})^{n_{is}}} \quad \cdots\cdots (15)$$

Consider a partitioning of $\underline{b}$ as $\underline{b}_I$ and $\underline{b}_{II}$:

$$\underline{b} = (\underline{b}_I \,\vdots\, \underline{b}_{II}) \quad \cdots\cdots (16)$$

If we want to test $\underline{b}_I = \underline{0}$ against $\underline{b}_I = \hat{\underline{b}}_I$, the unspecified parameter values $\underline{b}_{II}$ under the null hypothesis should be estimated through a constrained maximum likelihood approach

$$\hat{\hat{\underline{b}}}_{II} \in \left\{ \underline{b}_{II} \,|\, p(\underline{x}|\underline{b}_I = 0, \hat{\hat{\underline{b}}}_{II}) = \max_{\underline{b}_{II} \in \Omega_{II}} P(\underline{x}|\underline{b}_I = \underline{0}, \underline{b}_{II}) \right\} \quad \cdots\cdots (17)$$

where Ω_{II} is a parameter space for $\underline{b}_{II}$. Then the two hypotheses are specified as, following Wald's reduction for a single hypothesis;

$$H_0 : (\underline{b}_I, \underline{b}_{II}) = (\underline{0}, \hat{\hat{\underline{b}}}_{II})$$

$$H_a : (\underline{b}_I, \underline{b}_{II}) = (\hat{\underline{b}}, \hat{\underline{b}}_{II}).$$

The above testing procedure should be distinguished from the common practice (see for example, Watson, 1974) of setting the null hypothesis to

$$H_0 : (\underline{b}_I, \underline{b}_{II}) = (\underline{0}, \hat{\underline{b}}_{II})$$

against the same alternative hypothesis. This concludes the conventional hypothesis testing procedure for theory T.

Once the coefficients, Q_t, $t = 1, \ldots, m$, b_t and b_c are estimated, tests for the theory T can be conducted for the choice model of (1) and the significance of each coefficient can be tested. Although the construction of the theory V is consistent with the theory T, a test of T cannot be regarded as a test of the theory U, nor of V. To test U, we need to test V in a consistent manner after it is established that T cannot be rejected. Suppose that T cannot be rejected by tests based on the asymptotic normality of the maximum likelihood estimators of b_t and b_c. Then the underlying probability distribution of the VOT may be assumed to be a porbability density function of the ratio of two normally distributed random variables. Here, two cases are possible. First, if the sample estimator $\hat{b}_t/\hat{b}_c$ of the ratio b_t/b_c does not appear to converge (in probability) to a stable value, then the ratio of the standardized variates

$$B_t = \frac{b_t - E(b_t)}{\sigma_t^2} \sim N(0, 1)$$

$$B_c = \frac{b_c - E(b_c)}{\sigma_c^2} \sim N(0, 1) \quad \ldots\ldots\ldots (18)$$

can be assumed to be Cauchy distributed with zero median and unit scale parameter (Johnson and Kotz, 1970). Hence in this case the only possible tests of the ratio $\hat{b}_t/\hat{b}_c$ are in terms of median of the Cauchy distribution (for such tests, see Johnson and Kotz, 1970).

On the other hand, if we are able to establish that $\hat{b}_t/\hat{b}_c$ converges (in probability) to a finite mean, then $\hat{b}_t/\hat{b}_c$ can be inferred to be asymptotically normally distributed (Hansen, 1975; Kendall and Stuart, 1958) with variance;

$$\sigma^2 \frac{b_t}{b_c} = \left(\frac{\hat{b}_t}{\hat{b}_c}\right)^2 \left[\frac{\sigma_t^2}{\hat{b}_t^2} + \frac{\sigma_c^2}{\hat{b}_c^2} - \frac{2\sigma_{tc}^2}{\hat{b}_t \hat{b}_c}\right] \quad \ldots\ldots\ldots (19)$$

Hence the standard t-test can be applied in this case.

Alternatively, the choice model can be specified to include the VOT as a parameter;

$$P_i = \frac{\exp\left[b_c(VOT \cdot t_i + c_i) + a_i\right]}{\sum_j \exp\left[b_c(VOT \cdot t_j + c_j) + a_j\right]} \quad \ldots\ldots\ldots (20)$$

We would obtain the estimate $\widehat{VOT}$ for VOT directly through the maximum likelihood estimation procedure. Since $\widehat{VOT}$ is then distirubted normally, we would be able to conduct a test of the $\widehat{VOT}$. However, the maximization procedure for the maximum likelihood estimation associated with the direct estimate of $\widehat{VOT}$ tends to be computationally more difficult.

III. INFERENCE AND NULL HYPOTHESIS TENTING

Testing the null hypothesis, H_o, against alternatives, H_a, is a well established technique in traditional statistical inference. Suppose that the theory is established that the VOT is \$2.00/hr. based on the choice behavior. We wish to know whether or not the magnitude of \$2.00/hr. could have been due to random effects. We set up H_o: VOT = 0, against H_a: VOT = 2, choose a significance level, α, select an appropriate test statistic, say t or F, and proceed with the test. Rejection of H_o permits us to assert, with a precisely defined risk of being wrong at α%, that the VOT = 2 was not due to random effects. Failure to reject H_o is commonly referred to as "accepting" H_o. However, we can never conclude that the null hypothesis is true, but only that it cannot be rejected at the given level of significance.

This significance testing procedure, as Edwards (1971) pointed out, is always severely biased against H_o from a Bayesian point of view. Since H_a is to be confirmed, H_o is set up to be characteristically rejected with large samples (Lehmann, 1959). The probability of rejecting the null hypothesis is a function of five factors; whether it is a one or two tail test, the level of significance, the value of the standard deviation, the size of the deviation from H_o, and the number of samples. Once a set of samples is collected, the level of significance and the choice of H_o are left to the researcher. Yet, any failure of rejecting the null hypothesis when this goes against the researcher's intuition is too often blamed on the small sample size.

Wilson (1971), among others, proposed to identify "one's theory with the null hypothesis, and basing support on the acceptance of the null hypothesis" (p. 163).

Statistical inference is so difficult and perplexing a subject that a considerable amount of debate on the null hypothesis decision procedure has gone on, with the result that no resolution can be applied across the board. At present, the well-documented statistical inference techniques discussed are still philosophically problematic. We should not be overly willing to accept a theory by rejecting a null hypothesis which is designed to be rejected. The choice of the hypothesis, if the traditional inference method is to be taken, should be made with care rather than setting up a "straw man" hypothesis $H_o: \underline{b} = \underline{0}$ when it is known that it can be easily rejected, for reasons irrelevant to the purpose of the research.

When the maximum likelihood estimates, $\hat{\underline{b}}$, of the coefficients are obtained, the assymptotic variance-covariance is given by Kendall and Stuart (1961)

$$[\mathrm{cov}(b_h, b_k)] = [E\left\{\frac{\partial^2(-\log L)}{\partial b_h \, \partial b_k}\right\}]^{-1} \qquad \text{(21)}$$

which is an expected value of the inverse of the Hessian matrix of the negative log likelihood function. As the explanatory variables x_h and x_k associated with b_h and b_k become collinear to each other, it can be shown that the covariance, cov (b_h, b_k), between them becomes larger. In other words, the shape of the confidence region for the coefficients depends upon the properties of the data matrix. The maximum likelihood estimates we can obtain are conditional on a given finite sample.

Suppose that a linear compound of two explanatory variables x_1 and x_2 are hypothesized to explain the choice probability P:

$$P = f(a) \qquad \text{(22)}$$
$$a = b_1 x_1 + b_2 x_2$$

where b_1 and b_2 are coefficients corresponding to x_1 and x_2, respectively. Then the negative log likelihood function is

$$-\log L = g\,(b_1 x_1 + b_2 x_2). \quad \cdots\cdots \quad (23)$$

Suppose x_1 and x_2 are collinear, so that the scattered samples in the variable space form an ellipsoid. Let x_1 and x_2 be written as

$$x_2 = \alpha x_1 + e \quad \cdots\cdots \quad (24)$$

where $E(e) = 0$ and $E(e^2) \geqslant 0$ expresses the degree of independence between x_1 and x_2. Substituting (24) into (23) we obtain

$$-\log L = g\,[(b_1 + \alpha b_2)\,x_1 + e b_2 x_2]. \quad \cdots\cdots \quad (25)$$

If x_1 and x_2 are perfectly collinear, then $E(e^2) = 0$, and the coefficients b_1 and b_2 therefore cannot be determined uniquely. But it is possible to obtain an estimate $\hat{d}$ of the linear combination

$$d = b_1 + \alpha b_2 \quad \cdots\cdots \quad (26)$$

of the coefficients in (25). Consider a given confidence level corresponding to

$$C = g\,(\hat{d} x_1). \quad \cdots\cdots \quad (27)$$

If we let $\hat{b}_1$ and $\hat{b}_2$ be any estimates of b_1 and b_2 consistent with $\hat{d}$, i.e.,

$$\hat{d} = \hat{b}_1 + \alpha \hat{b}_2 \quad \cdots\cdots \quad (28)$$

then it is clear that, although $\hat{d}$ is uniquely determined for C, $\hat{b}_1$ and $\hat{b}_2$ are not.

More generally, even if x_1 and x_2 are not perfectly collinear, the equation

$$x_2 = \alpha x_1 \quad \cdots\cdots \quad (29)$$

may still represent a dominant principal axis of the ellipsoid of concentration for the sample scatter in the variable space. In this case, equation (28) may then approximate the principal axis of the corresponding confidence region in the parameter space. Hence, rewriting (28) as

$$\hat{b}_1 = \hat{d} - \alpha \hat{b}_2 \quad \cdots\cdots \quad (30)$$

it is clear that the slope of the principal axis of the $(\hat{b}_1, \hat{b}_2)$ confidence region is approximately equal to $-\alpha$. That is, if x_1 and x_2 are positively related ($\alpha > 0$), we expect the principal axis of the confindence region to be negatively sloped ($-\alpha < 0$). If samples are clustered into groups or other irregularities are presented from a unimodal distribution with decreasing density from the mean value, specific values of the slope of the principal axis of the confidence region would be different at different levels of significance. However, we expect, in general, an inverse relationship between the sample distribution and confidence region. Figure 1 shows the general relationship between the scattered samples in the variable space and the confidence region in the parameter space.

Case A is when the variables x_1 and x_2 are independent of each other. The case B and C show x_1 and x_2 which are correlated, and case D is when x_2 accounts for all of the variation in the dependent variables,such as choice probability. In the cases B and C, as the variables become strongly correlated, the confidence region becomes narrower. Notice that the travel time and travel costs are likely to be positively correlated as in the case B. If the mode choice model is developed to explain choice behavior and the composite hypothesis is set up for H_o:b = 0, the choice theory is likely to be accepted. This is acceptable when the explanation of the choice behavior is of importance, but how much is explained by each of the explanatory variables is not.

If the purpose of developing the choice model is to estimate the value of time, expressed as a ratio of choice model parameters, the acceptance of the choice theory by rejecting $H_o = \underline{b} = \underline{0}$ is apt to be misleading. Minimized Type I error for the choice theory for a fixed significance level does not minimized Type I error for the value of time. To avoid the problem, we should set up the statistical null hypotheses $H_o:\underline{b} = (0, b_2)$ and $H_o:\underline{b} = (b_1, o)$, where b_1 and b_2 are the constrained maximum likelihood estimators as shown in case F of Figure 2, which is obtained by (17).

Often each coefficient is tested separately with two simple tests, with the statistical null hypotheses, $H_o:\underline{b} = (0, b_2)$ and $H_o:\underline{b} = (b_1, 0)$, as conducted in analogue to the t-test for each regression coefficient in the multiple regression model. The results of this test can also be misleading with respect to the value of time, when the two variables are collinear. Figure 2 shows the case in which the multicollinearity is mild (case E) and severe (case F). In case F, both null hypotheses words, when the multicollinearity is mild, as in the case of E, simple hypothesis testing would properly reject the choice theory in light of the effect on each variable.

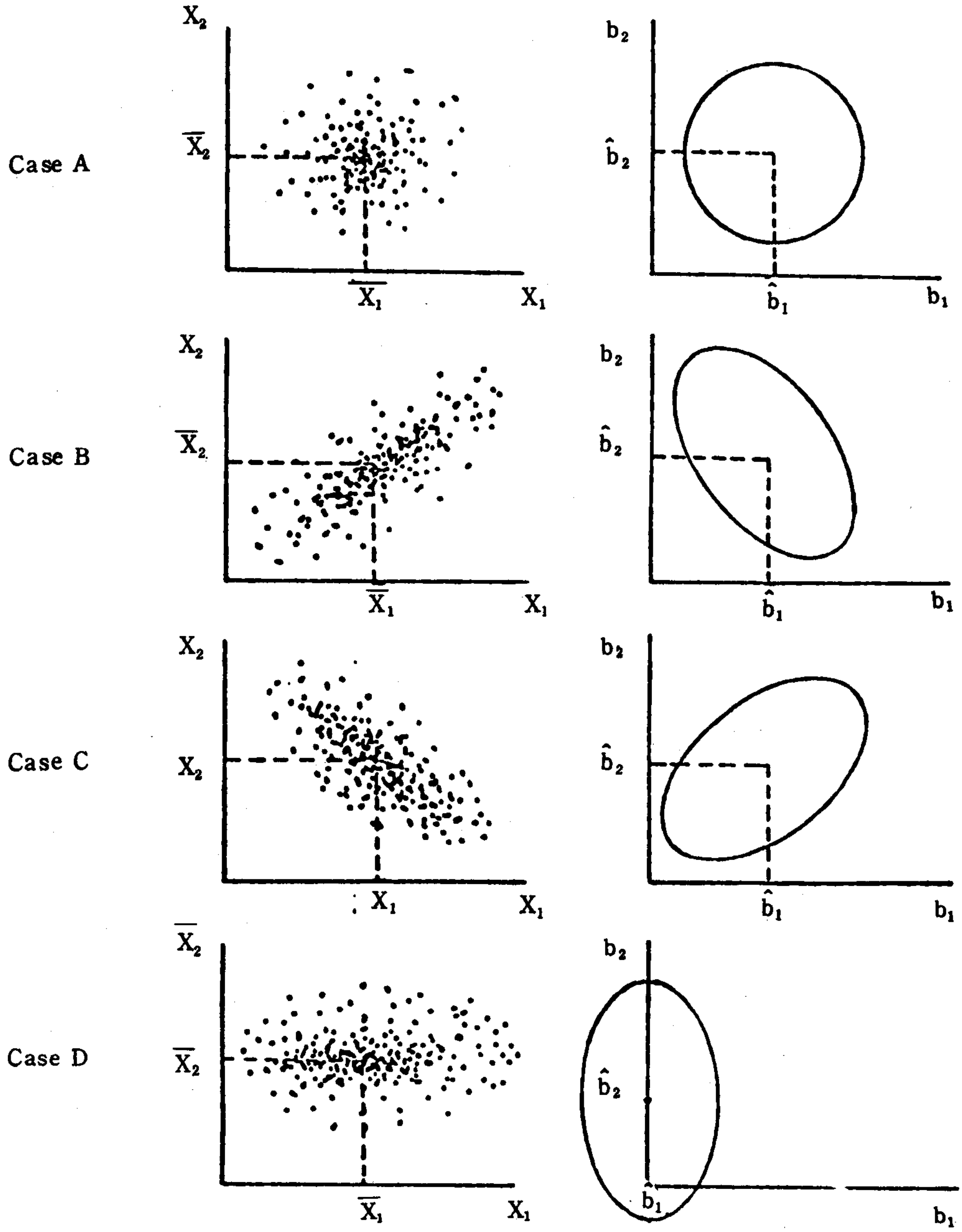
If, however, the multicollinearity is severe, the confidence region becomes a long and narrow band and the simple hypothesis testing may still be in danger of giving a wrong conclusion for the value of time.

If the average speed and unit travel costs per mile are used to explain the choice behavior instead of the travel time and travel costs, the value of time is expressed as a product of the coefficients associated with the variables, assuming that the same linear logit model of (1) is used.

$$VOT = b_t \, b_c \qquad (31)$$

where b_t is a coefficient associated with the average speed and b_c unit travel costs. The situation we should worry most about in this case is the one in which the explanatory variables are negatively correlated as in the case C in Figure 1. The conclusion drawn from the composite hypothesis with H_o: $\underline{b} = \underline{0}$ is less likely to be misleading about the value of time than in the case C, while the simple hypo-

Figure 1. Direction of Confidence Region and Sample Scatter



thesis testing would not help determine the value of time. However, if the confidence region is tilted with different angles in the coefficient space as shown in case B of Figure 3, then composite hypothesis testing for choice behavior can also be misleading for the conclusion of the value of time study.

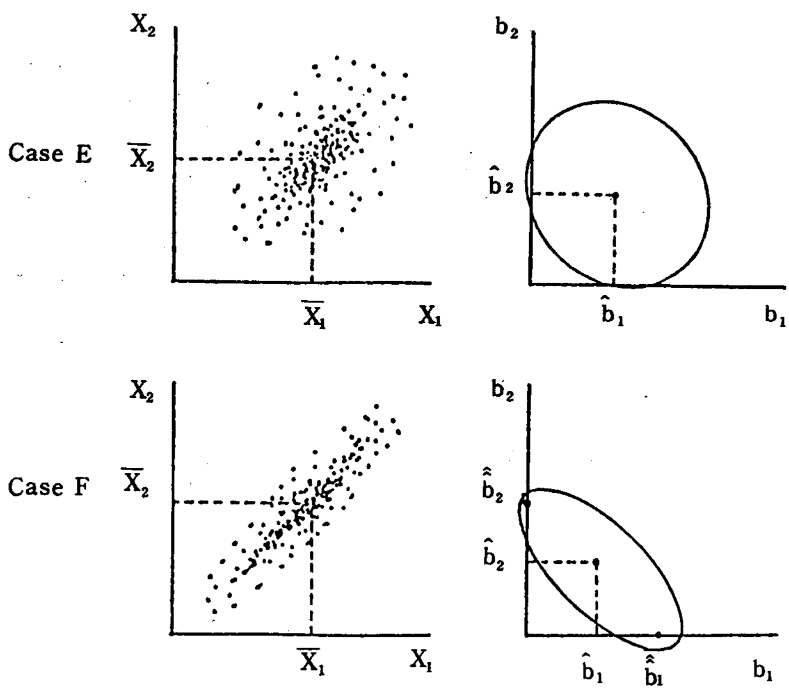
Since the confidence region is the area in the coefficient space where the true values of the parameters are located with a given level of probability or higher, it is drawn as an isolikelihood function, a contour in which the values of the likelihood function are all equal, under the assumption that the maximum likelihood estimators are asymptotically normally distributed. If the likelihood function is roughly dome-shaped, then the confidence region is approximately circular. On the other hand, if a ridge is present on the surface of the likelihood function, then the confidence region appears as an ellipsoid. Although the likelihood function for the linear logit model is everywhere concave, it is difficult to generalize the shape of the likelihood function over a large range of $\underline{b}$ values. Figure 4 shows some possible irregular shapes of the likelihood function which may be difficult to trace, and for which the possibility of committing either Type I or Type II error may be unexpectedly large.

Consider a binary mode choice problem in which the log likelihood function is expressed as

$$L = n_1 \log p_1 + n_2 \log p_2 \quad \ldots\ldots\ldots\ldots \quad (32)$$

where $p_1 + p_2 = 1$ and $n_1 + n_2 = N$. Suppose each mode is characterized by travel time and travel costs and the choice behavior is explained by a linear logit model with mode specific parameters.

Figure 2. Shape of Confidence Region and Sample Scatter



$$p_1 = f(b_{1t}\, t_1 + b_{1c} c_1 + \alpha_1) \qquad (33)$$
$$p_2 = f(b_{2t}\, t_2 + b_{2c} c_2 + \alpha_2)$$

where f (.) is a logit model. Let n_1 samples and n_2 samples be scattered and the confidence region for the mode parameters be given as shown in Figure 5.

If we impose additional constraints on the parameters that

$$b_{1t} = b_{2t} = b_t \qquad (34)$$
$$b_{1c} = b_{2c} = b_c$$

then, we have the choice model expressed as (1). The confidence region for $(\hat{b}_t, \hat{b}_c)$ is related as in Figure 6 with the confidence regions shown in Figure 5. Even if the confidence regions for each mode at different levels of significance (contours of iso-likelihood functions) may be systematically changing, confidence regions for all modes could vary much more radically than for any one mode. As the number of modes in the choice set becomes large, the "direction" of the confidence regions (the direction of the principal axis of the ellipsoid) may resist intuitive prediction for the various levels of significance (any ridge on the likelihood surface may not be constant in its direction, nor even continuous). Therefore, a priori data examination to detect a possibly unstable VOT estimate is not an easy task, although it is necessary.

Figure 3. Slope of Principal Axis of Confidence Region

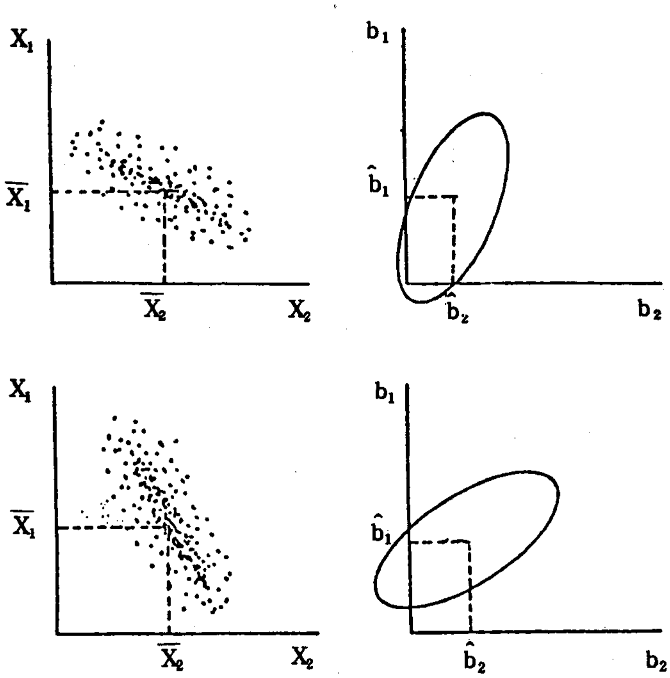
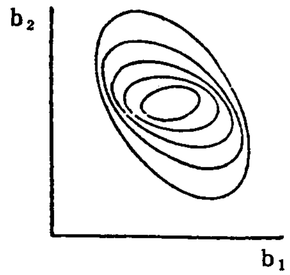
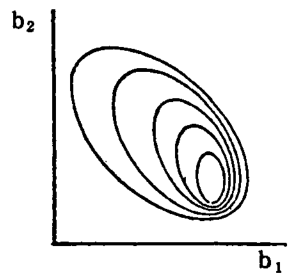


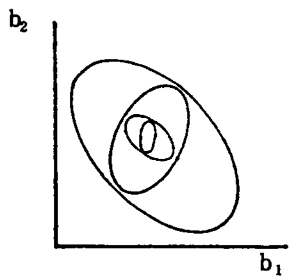
Figure 4. Irregular Shape of Likelihood Surface



The direction of cricitcal region may change drastically depending upon the significance level chosen.



Uneven tail of confidence region may result contradicting conclusions of one-tail and two-tail tests.



Derivatives may mislead the conclusion for shape of likelihood surface.

Figure 5. Confidence Region for Each Mode

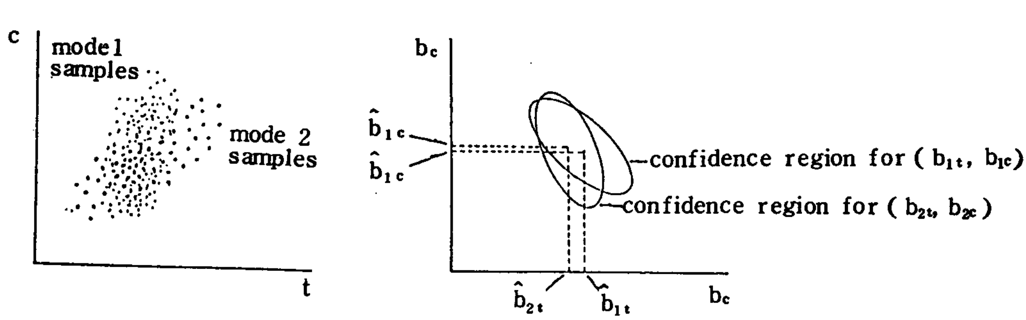


Figure 6. Confidence Region for All Modes

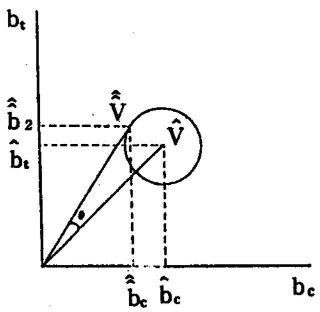
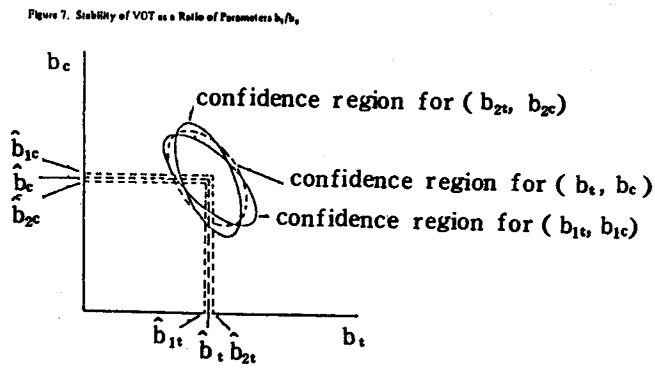


Figure 7. Stability of VOT as a Ratio of Parameters b_t/b_c



IV. TESTING THE THEORY OF VOT

Suppose the VOT is expressed as a ratio of coefficient estimates of a linear logit model, as shown in equation (2). The significance of the VOT implies that the slope of the parameters are stable for any values of the parameters within a given confidence region. Hence, the significance test of the VOT estimate is a test of the stability of the direction of the vector $\widehat{V}$ connecting the origin and coefficient estimates $(\widehat{b}_t, \widehat{b}_c)$ in the parameter space.

Suppose travel time t and travel cost c are appropriately transformed (having the same sample variance) so that $\widehat{b}_t$ and $\widehat{b}_c$ can be assumed to be independently and identically normally distributed with variance σ^2. Hence, the confidence region for a given level of significance is a perfect circle. In this case, the stability test of the vector $\widehat{V}$ direction is of the angle θ between the vector $\widehat{V}$ and the vector $\widehat{\widehat{V}}$ connecting the origin and tangent to the circular boundary of the confidence region as shown in Figure 7. Let $(\widehat{\widehat{b}}_t, \widehat{\widehat{b}}_c)$ be the point on the boundary of the confidence region tangent to $\widehat{\widehat{V}}$. The angle θ can be expressed as

$$W = \cos\theta = \frac{\widehat{b}_t \widehat{\widehat{b}}_t + \widehat{b}_c \widehat{\widehat{b}}_c}{\sqrt{(\widehat{b}_t^2 + \widehat{b}_c^2)(\widehat{\widehat{b}}_t^2 + \widehat{\widehat{b}}_c^2)}} \qquad \cdots\cdots (35)$$

Hence, W is a statistic corresponding to the maximum possible angle between vectors in the confidence region of the ratio of the coefficient estimates. If the variances of $\hat{b}_t$ and $\hat{b}_c$ are not identical and/or t and c are not truly independent, the confidence region becomes ellipsoidal if it is closed, and finding $\hat{\hat{b}}_t$ and $\hat{\hat{b}}_c$ becomes complicated. In particular, the following nonlinear program must be solved (see for example Goldfeld and Quandt, 1972):

$$\begin{aligned} &\max && (\hat{\underline{b}} - \hat{\hat{\underline{b}}}) \cdot (\hat{\underline{b}} - \hat{\hat{\underline{b}}}) \\ &\text{subject to} && L(t, c; \hat{\hat{\underline{b}}}) = L(t, c; \hat{b} + z_\alpha \hat{H}^{-1} \hat{\underline{b}}) \end{aligned} \quad \cdots\cdots (36)$$

where z_α is the area under the normal curve with a given confidence level and $\hat{H}^{-1}$ is a maximum likelihood estimate of the variance-covariance matrix. Unfortunately the solution of (36) is very difficult and no explicit solution exists.

Alternatively, an approximate method can be developed for $\hat{\hat{\underline{b}}}$ by taking $\hat{\hat{\underline{b}}}$ to be an end point of the principal axis of the confidence region (obtainable (approximately) from (29)). This approximate method gives us a reasonable estimate of the range of VOT when the principal axis of the confidence region is negatively sloped (e.g., case B). This is the situation when the VOT value is most unstable. On the other hand, if the VOT value is stable, that is, the principal axis of the confidence region is positively sloped, this approximate method would perform badly (e.g., case C). In other words, this approximation is most reliable for the worst situation (where we are most concerned with the stability of the VOT) and most unreliable for the best situation (when we need not be overly concerned with the stability of the VOT).

V. AN ILLUSTRATIVE EXAMPLE

To illustrate the stability of the VOT estimates, a hypothetical data set of bivariate attributes of time savings, t, and cost savings, c, for binary alternatives, 1, and 2, are constructed as in Table 1 with simple closed ellipsoidal iso-likelihood contours. The principal axis of the ellipsoid is toward the origin, in order to have a reasonably stable VOT estimate. First, maximum likelihood estimators for a linear logit model for choice behavior are estimated. The choice model is

$$P_i = \frac{\exp u_i}{\exp u_1 + \exp u_2} \qquad i = 1, 2$$

$$\cdots\cdots (37)$$

$$u_1 = b_t t_1 + b_c c_1$$

$$u_2 = a + b_t t_2 + b_c c_2 .$$

The parameter estimates are obtained as

Table 1. Hypothetical Data for Iso-Likelihood Evaluations

t_1	c_1	t_2	c_2	f_1	f_2
15	20	25	5	50	50
25	15	15	4	1	99
25	10	10	14.4	99	

(f_1, f_2) are frequency observations for (1, 2).
Total sample size is 300.

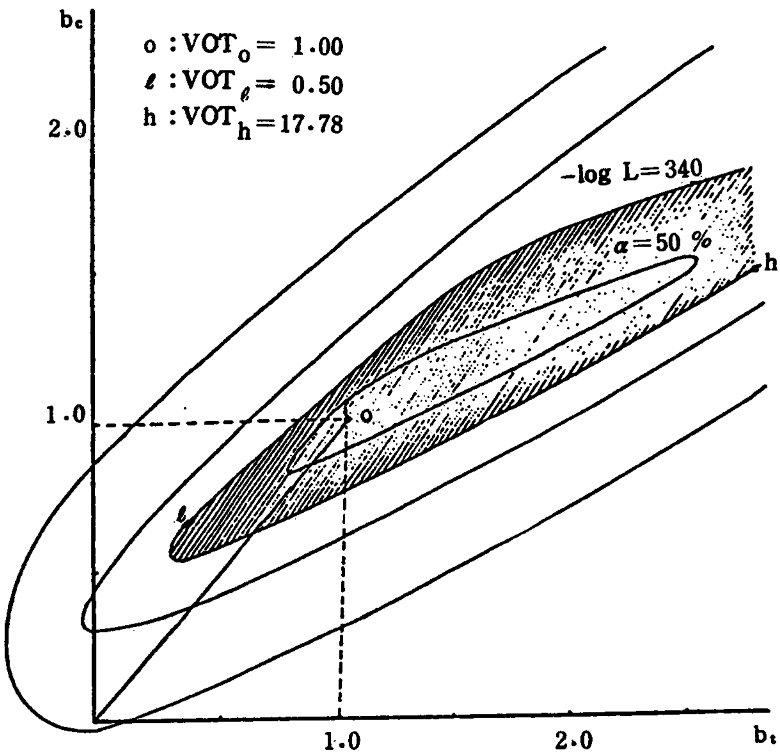
$$\hat{a} = 15.0 \quad \hat{b}_t = 1.0 \quad \hat{b}_c = 1.0. \qquad (38)$$

Therefore, the VOT estimate is given by

$$VOT = 1.0 \qquad (39)$$

To determine the iso-likelihood contours,the negative log likelihood function is evaluated in the vicinity of the maximum likelihood estimates of b_t and b_c, given a = 15, as shown in Figure 8. The negative log likelihood value at the maximum likelihood estimate point (minimum value of the negative log likelihood function) is 81.2. Assuming that a = 15 is true, the true values of b_t and b_c at the 95% confidence level lie anywhere within the negative log likelihood value of approximately 340. Among the points that are evaluated, the lowest and highest values of VOT that are within the range are 0.50 and 17.78, respectively. The highest value for the VOT estimate that can be accepted with 95% confidence level is over 35 times larger than the lowest estimate which is acceptable. Yet, the composite hypothesis tests for parameters reveal that the parameters (b_t and b_c) are significantly different from zero at the 95% confidence level. Hence, accepting the VOT on the basis of accepting the choice model can be grossly misleading.

Figure 8. Likelihood Contour Simulation



VI. CONCLUSION

Hypothesis testing of the choice model occurs on two different levels: overall theory testing and individual parameter testing. Based on the Neyman-Pearson procedure, the likelihood ratio test can be used for both overall theory testing and individual parameter testing. The likelihood ratio test is here considered to be a more valid testing procedure than the standard t-test, in that the distribution theory for the latter is far more difficult to justify in the logit model context.

However, the Neyman-Pearson argument does not provide a fault-free testing procedure. Its main drawback is that it provides only an indirect test. The classical "test-of-hypothesis" procedure always involves a "null" hypothesis, i.e., one designed to be rejected, rather than a hypothesis which is designed to be accepted. Unfortunately, there is no widely accepted testing procedure which enables one to accept a given hypothesis. Although examining the ciritical region would give us some insight, it does not allow for a method of concrete statistical inference.

Hence, testing the choice theory should be distinguished from testing the VOT theory. Although composite hypothesis testing of parameters for travel time and travel cost can potentially provide some insights, it does not provide a statistical test for the VOT, which is a ratio of two parameters within the linear logit model. Without knowing the underlying distribution of the VOT estimator, the Neyman-Pearson argument cannot be applied. An alternative procedure was developed in section IV to analyze in a direct way the "shape" of the critical region in the parameter space. In this context. the correlation-like angular measure (36) of the VOT dispersion in the parameter space was shown to provide potentially useful information concerning the stability of the VOT estimator. Finally, the example in section V estimator can be highly unstable, although the choice theory itself cannot be rejected. Further, the example raises a question of the validity of the procedure for estimating choice model base VOT.

BIBLIOGRAPHY

1. Ben-Akiva, M. (1974): "Alternative Travel Behavior Structures of Passenger Travel Demand Models, "*TRB Record* 526, TRB, Washington, D.C.
2. Boyce, D.E. et al. (1974): *New Approaches to Understanding Travel Behavior, Proposal,* mimeo, University of Pennsylvania, Philadelphia.
3. Burns, L.D. et al. (1976): "Theory of Urban Household Automobile-Ownership Decisions," *TRB Report* 569, TRB, Washington D.C.
4. Charles River Association (1972): *A Disaggregated Behavioral Model of Urban Travel Demand,* Cambridge.
5. Cox, D.R. (1970): *The Analysis of Binary Data,* Methuen Company, London, England.
6. Edwards, W. (1971): "Tactical Note on the Relation between Scientific and Statistical Hypotheses," in Lieberman, B. (ed.), *Contemporary Problems in Statistics,* Oxford University Press, New York.
7. Gehner, C.D. (1974): *Investigation of Alternative Calibration Techniques for the n-Dimensional Logit Model,* Ph. D. thesis, University of Washington, Seattle.
8. Goldfeld, S.M. and Quandt, R.E. (1972): *Nonlinear Methods in Economitrics,* North Holland Publishing Company, Amsterdam, The Netherlands.
9. Hansen, S. (1974): "Structure of Disaggregate Behavioral Choice Models," *TRB Special Report* 149, TRB, Washington, D.C.
10. ———— (1975) Private Letter on June 8.
11. Johnson, N.L. and Kotz, S. (1970): *Continuous Univariate Distributions-1,* Houghton Mifflin, New York.
12. Kendall, M.G. and Stuart, A. (1958): *The Advanced Theory of Statistics, Volume I, Distribution Theory,* Griffin Company, London, England.
13. ———— and ———— (1961): *The Advanced Theory of Statistics, Volume II, Inference and Relationship,* Griffin Company, London, England.
14. Lehmann, E.L. (1959): *Testing Statistical Hypotheses,* John Wiley, New York.

15. Lerman, S.R. (1975): *A Disaggregate Behavioral Model of Urban Mobility Decisions,* Ph. D. Thesis, MIT, Cambridge.
16. Liou, P.S. and Talvitie, A.P. (1974): "Disaggregate Access Mode and Station Choice Models for Rail Trips," *TRB Report* 526, TRB, Washington, D.C.
17. McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, P. (ed.) *Frontiers of Econometrics,* Academic Press, New York.
18. __________ (1974): *Quantal Choice Analysis,* mimeo, University of California, Berkeley.
19. __________ et al. (1977): *Demand Model Estimation and Validation,* University of California, Berkeley.
20. Rao, C.R. (1973): *Linear Statistical Inference and Its Applications,* 2nd ed., John Wiley, New York.
21. Smith, T.E. (1975a): "A Choice Theory of Spatial Interaction," *Regional Science and Urban Economics* 5, pp.137-176.
22. __________ (1975b): "An Axiomatic Theory of Spatial Discounting Behavior," *Papers of the Regional Science Association* 35.
23. Stopher, P.R. (1974): "Discussion" *TRB Record* 526, TRB, Washington, D.C.
24. __________ (1975): "Goodness-of-Fit Measures for Probabilistic Travel Demand Models," *Transportation* 4, pp.67-83.
25. Watson, P.L. (9174): *The Value of Time: Behavioral Models of Modal Choice,* Heath, Lexington.
26. Wilson, W. et al. (1971): "Much Ado about the Null-Hypothesis," in Lieberman, B., (ed.) *Contemporary Problems in Statistics,* Oxford University Press, New York.