the American Statistical Association, No. 50, 1955, pp. 130-161.
9. H. Theil. Principles of Econometrics. Wiley, New York, 1971, pp. 65-76.
10. M. Gaudry and M. G. Dagenais. The Dogit Model. Centre de Recherche sur les Transports, Univ. de Montréal, Publication 82, 1978, 14 pp.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Confidence Intervals for Choice Probabilities of the Multinomial Logit Model

Joel Horowitz, U.S. Environmental Protection Agency

This paper describes three methods for developing confidence intervals for the choice probabilities in multinomial logit models. The confidence intervals reflect the effects of sampling errors in the parameters of the models. The first method is based on the asymptotic sampling distribution of the choice probabilities and leads to a joint confidence region for these probabilities. This confidence region is not rectangular and is useful mainly for testing hypotheses about the values of the choice probabilities. The second method is based on an asymptotic linear approximation of the relation between errors in models' parameters and errors in choice probabilities. The method yields confidence intervals for individual choice probabilities as well as rectangular joint confidence regions for all of the choice probabilities. However, the linear approximation on which the method is based can yield erroneous results, thus limiting the applicability of the method. A procedure for setting an upper bound on the error caused by the linear approximation is described. The third method is based on nonlinear programming. This method also leads to rectangular joint confidence regions for the choice probabilities. The nonlinear programming method is exact and, therefore, more generally applicable than the linear approximation method. However, when the linear approximation is accurate, it tends to produce narrower confidence intervals than does the nonlinear programming method, except in cases where the number of alternatives in the choice set is either two or very large. Several numerical examples are given in which the nonlinear programming method is illustrated and compared with the linear approximation method.

The multinomial logit formulation of urban travel-demand models has a variety of theoretical and computational advantages over other demand-model formulations and is receiving widespread use both for research purposes and as a practical demand-forecasting tool (1-3). However, travel-demand forecasts derived from logit models, like forecasts derived from other types of econometric models, are subject to errors that arise from several sources, including sampling errors in the estimated values of parameters of the models, errors in the values of explanatory variables, and errors in the functional specifications of the models. Knowledge of the magnitudes of forecasting errors can be important in practice, particularly if either the errors themselves or the costs of making erroneous decisions are large. This paper deals with the problem of estimating the magnitudes of forecasting errors that result from sampling errors in the estimated values of the parameters of logit models. Specifically, the paper describes techniques for developing confidence intervals for choice probabilities and functions of choice probabilities (e.g., aggregate market shares, changes in choice probabilities caused by changes in independent variables) derived from logit models, conditional on correct functional specification of the models and use of correct values of the explanatory variables.

A model's forecasting error can be characterized in a variety of ways, including average forecasting error and root-mean-square forecasting error, in addition to confidence intervals for the forecast. Among the various error characterizations, only the confidence interval provides a range in which the true value of the forecast quantity is likely to lie. Methods for developing confidence intervals for the forecasts of linear econometric models are well known (4). However, these methods are not applicable to logit models, which are nonlinear in parameters. Koppelman (5, 6) has analyzed the forecasting errors of logit models and has described the ways in which various sources of error contribute to total error in forecasts in choice probabilities. Koppelman's error measures do not include confidence intervals for the choice probabilities although, as will be shown later in this paper, one of his error measures can be used to derive approximate confidence intervals.

Three methods for estimating confidence intervals for the choice probabilities of logit models are described in this paper. All of the methods lead to asymptotic confidence intervals in that they are based on the large-sample properties of the estimated parameters of the models. The first method is based on the exact asymptotic sampling distribution of the choice probabilities and leads to a joint confidence region for these probabilities. This region is useful mainly for testing hypotheses about the values of the choice probabilities. The region is not rectangular and, therefore, is difficult to use in practical forecasting. Moreover, the methods used to derive the confidence region cannot be readily extended to functions of the choice probabilities.

The second method is based on an asymptotic linear approximation of the relation between sampling errors in models' parameters and sampling errors in choice probabilities. The linear approximation method yields confidence intervals for individual choice probabilities as well as rectangular joint confidence regions for all of the choice probabilities. The method can easily be extended to functions of the choice probabilities. However, the linear approximation on which the method is based can yield erroneous results, thus limiting the method's applicability. A procedure for placing an upper bound on the error caused by the linear approximation is described.

The third method is based on nonlinear programming. This method yields rectangular joint confidence regions for the choice probabilities and can be extended to functions of the choice probabilities. The method does not require approximation of the relations between sampling errors in models' parameters and sampling errors in choice probabilities and, therefore, is more generally applicable than is the linear approximation method. Several numerical examples are given in which the nonlinear programming method is illustrated and compared with the linear approximation method.

PROPERTIES OF THE LOGIT MODEL

In the multinomial logit model, the probability that individual n selects alternative i from a set of J_n available alternatives is given by

$$P_{in} = \exp(V_{in}) \Big/ \sum_{j=1}^{J_n} \exp(V_{jn}) \tag{1}$$

where P_{in} is the probability that alternative i is chosen by individual n, and V_{jn} (j=1, ..., J_n) is the systematic component of the utility of alternative j to individual n.

For each alternative i, V_{in} is assumed to be a linear function of appropriate explanatory variables. Thus

$$V_{in} = \sum_{m=1}^{M} X_{imn}\alpha_m \tag{2}$$

where

M = the number of explanatory variables,
X_{imn} = the value of the m th explanatory variable for alternative i and individual n, and
α_m = the coefficient of explanatory variable m.

The values of the coefficients (or parameters) α_m ordinarily are not known a priori and are estimated from observations of individuals' choices by using the method of maximum likelihood. Details of the estimation procedure and the statistical properties of the estimated coefficients are described by McFadden (7).

Denote the estimated coefficients by $\{a_m; m=1, \ldots, M\}$. For each alternative i and individual n define $\hat{V}_{in}$ by

$$\hat{V}_{in} = \sum_{m=1}^{M} X_{imn} a_m \tag{3}$$

$\hat{V}_{in}$ is the estimated systematic utility function for alternative i and individual n. $\hat{V}_{in}$ is a random variable by virtue of its dependence on the random variables $\{a_m\}$. Define

$$\hat{P}_{in} = \exp(\hat{V}_{in}) \Big/ \sum_{j=1}^{J_n} \exp(\hat{V}_{jn}) \qquad (i = 1, \ldots, J_n; n = 1, \ldots, N) \tag{4}$$

$\hat{P}_{in}$ estimates the probability that individual n makes choice i and is the forecast of the choice probability that is used in applications of the logit model. Accordingly, the subsequent sections of this paper are concerned with the development of ranges about the $\hat{P}_{in}$ that are likely to contain the true choice probabilities P_{in}.

Assume that the coefficients $\{a_m\}$ have been estimated by the method of maximum likelihood by using a data set that consists of observations of N individuals' choices. Then for large N, the estimated coefficients $\{a_m\}$ are asymptotically jointly normally distributed with mean values $\{\alpha_m\}$ and covariance matrix A^{-1}, where

$$A_{rs} = -\sum_{n=1}^{N} \sum_{j=1}^{J_n} (X_{jrn} - X_{\cdot rn})(X_{jsn} - X_{\cdot sn}) P_{jn} \qquad (r,s = 1, \ldots, M) \tag{5}$$

and

$$X_{\cdot rn} = \sum_{j=1}^{J_n} X_{jrn} P_{jn} \tag{6}$$

In addition, the quadratic form

$$Q(\underline{a}, \underline{\alpha}) = \sum_{i=1}^{M} \sum_{j=1}^{M} (a_i - \alpha_i) A_{ij} (a_j - \alpha_j) \tag{7}$$

tends asymptotically to the chi-square distribution with M degrees of freedom.

Let one of the J_n alternatives available to individual n be considered a numeraire, and denote this alternative by t. Then the random variables $\{\hat{V}_{in} - \hat{V}_{tn}; i=1, \ldots, J_n; i \neq t\}$ are linear combinations of the asymptotically normally distributed random variables $\{a_m\}$ and are themselves asymptotically jointly normally distributed with mean values $\{V_{in} - V_{tn}; i=1, \ldots, J_n; i \neq t\}$ and covariance matrix C_n^{-1}, where

$$(C_n^{-1})_{ij} = \sum_{r=1}^{M} \sum_{s=1}^{M} (A^{-1})_{rs} (X_{irn} - X_{trn})(X_{jsn} - X_{tsn}) \tag{8}$$

and (i, j=1, ..., J_n; i, j $\neq$ t). In addition the quadratic form

$$R(\hat{\underline{V}}_n, \underline{V}_n) = \sum_{\substack{i=1 \\ i,j \neq t}}^{J_n} \sum_{j=1}^{J_n} (C_n)_{ij} [(\hat{V}_{in} - \hat{V}_{tn}) - (V_{in} - V_{tn})] \times [(\hat{V}_{jn} - \hat{V}_{tn}) - (V_{jn} - V_{tn})] \tag{9}$$

is asymptotically distributed as chi-square with $J_n - 1$ degrees of freedom.

In practical applications of logit models, the probabilities P_{in} and, therefore, the matrix A in Equation 5 are not known due to their dependence on the unknown coefficients $\{\alpha_m\}$. Therefore, P_{in} is approximated by $\hat{P}_{in}$ in Equation 5. This approximation is used without further comment in the rest of this paper.

In the following discussion the subscript n, which denotes the individual, will not be used unless needed to prevent confusion. The choice probabilities will be understood to apply to an individual. The explanatory variables X_{imn} will be assumed to have known, fixed values. All uncertainty in the choice probabilities will be due to their dependence on the unknown coefficients $\{\alpha_m\}$.

Confidence Intervals for Choice Probabilities in Binary Logit Models

If there are only two alternatives in the choice set (J=2), then C^{-1} is a scalar. Therefore, if $Z_{\epsilon/2}$ is the 100 (1 - ϵ/2) percentile of the standard normal distribution, a 100(1-ϵ) percent confidence interval for $V_1 - V_2$ is

$$(\hat{V}_1 - \hat{V}_2) - Z_{\epsilon/2} C^{-1/2} \leq V_1 - V_2 \leq (\hat{V}_1 - \hat{V}_2) + Z_{\epsilon/2} C^{-1/2} \tag{10}$$

Denote the left- and right-hand expressions of inequalities by b and B, respectively. Then, the expressions for the 100 (1-ϵ) confidence intervals for P_1 and P_2 in the binary logit model are

$$1/[1 + \exp(-b)] \leq P_1 \leq 1/[1 + \exp(-B)] \tag{11}$$

and

$$1/[1+\exp(B)] \le P_2 \le 1/[1+\exp(b)] \quad (12)$$

These simple expressions for confidence intervals exist only for binary choice models.

Joint Confidence Regions for the Choice Probabilities Based on Asymptotic Sampling Distribution

Equation 4 for the estimated choice probabilities can be rewritten as

$$\hat{P}_i = \exp(\hat{V}_i - \hat{V}_t) \Big/ \Big[1 + \sum_{j \neq t} \exp(\hat{V}_j - \hat{V}_t)\Big] \quad (i \neq t) \quad (13)$$

$$\hat{P}_t = 1 - \sum_{j \neq t} \hat{P}_j \quad (14)$$

where t denotes the numeraire alternative. Equation 13 defines a transformation from the random variables $\{\hat{V}_j - \hat{V}_t; j \neq t\}$ to the random variables $\{\hat{P}_i; i \neq t\}$. This transformation has a nonsingular Jacobian matrix. Accordingly, the joint probability-density function of the random variables $\{\hat{P}_i; i \neq t\}$, conditional on $\hat{P}_t$, can be derived by using standard procedures (8). The result is

$$f(\{\hat{P}_i \quad i \neq t\} \mid \hat{P}_t) = (2\pi)^{(1-n)/2} |C|^{1/2} \left(\prod_{j=1}^{J} \hat{P}_j\right)^{-1}$$
$$\times \exp\{-(1/2) \sum_i \sum_j C_{ij}[\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t)]$$
$$\times [\log(\hat{P}_j/\hat{P}_t) - \log(P_j/P_t)]\} \quad (15)$$

where $|C|$ denotes the determinant of the matrix C and the quantity on the left-hand side denotes the joint probability-density function of $\{\hat{P}_i; i \neq t\}$, conditional on $\hat{P}_t$.

The conditioning of density function 15 on $\hat{P}_t$ can be removed by noting from Equation 14 that $\hat{P}_t$ is completely determined by $\{\hat{P}_i; i \neq t\}$. Thus, the joint probability-density function of all of the $\hat{P}_i$ ($i=1, \ldots, J$) is

$$f(\{\hat{P}_i \quad i = 1, \ldots, J\}) = \delta\left(1 - \sum_{j=1}^{J} \hat{P}_j\right) f(\{\hat{P}_i, i \neq t\} \mid \hat{P}_t) \quad (16)$$

where δ is the Dirac delta function. Equation 16 constitutes a multivariate generalization of the univariate S_B distribution (9). The univariate distribution has been applied in a transportation context by Westin (10), who used the distribution to develop aggregate forecasts from a binary logit model.

The distribution in Equation 16 is highly intractable. To develop a confidence region for P_i it is more convenient to work with the distribution of the logarithms of the choice probabilities than with the distribution of the probabilities themselves. Specifically, Equation 4 implies that

$$\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t) = (\hat{V}_i - \hat{V}_t) - (V_i - V_t) \quad (17)$$

Equations 9 and 17 together imply that the random variable R* defined by

$$R^*(\hat{\underline{P}}, \underline{P}) = \sum_{\substack{i=1 \\ i,j \neq t}}^{J} \sum_{j=1}^{J} C_{ij}[\log(\hat{P}_i/\hat{P}_t) - \log(P_i/P_t)]$$
$$\times [\log(\hat{P}_j/\hat{P}_t) - \log(P_j/P_t)] \quad (18)$$

has the chi-square distribution with J-1 degrees of freedom. Let $\chi^2(\epsilon, K)$ denote the 100 $(1-\epsilon)$ percentile of the chi-square distribution with K degrees of freedom. Then, the inequality

$$R^*(\hat{\underline{P}}, \underline{P}) \le \chi^2(\epsilon, J-1) \quad (19)$$

together with Equation 14 define a joint 100 $(1-\epsilon)$ percent confidence region for $\{P_i; i=1, \ldots, J\}$. Specifically, given estimated values of $\{\hat{P}_i; i=1, \ldots, J\}$, the confidence region consists of the set of all P_i ($i=1, \ldots, M$) such that Equation 14 and inequality 19 are satisfied.

The confidence region defined by Equation 14 and inequality 19 is not rectangular and, therefore, is difficult to use in practical forecasting. In particular, the confidence region does not directly yield constants b_i and B_i ($i=1, \ldots, J$) such that $b_i \le P_i \le B_i$ with a specified level of confidence. However, the confidence region can be used to test hypotheses about the values of the P_i. Let the null hypothesis be $P_1 = P_1^*$, $P_2 = P_2^*$, $\ldots$, $P_J = P_J^*$, and assume that $\Sigma P_i^* = 1$. Substitute P_i^* for P_i in Equation 18 and compute R*. Then, the null hypothesis is rejected at the ϵ significance level if R* fails to satisfy inequality 19.

The method used to develop inequality 19 for individual choice probabilities cannot be extended to functions of the choice probabilities, such as aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The number of utility components $\hat{V}_i - \hat{V}_t$ in such functions exceeds the number of dependent variables (e.g., aggregate shares, changes in choice probabilities) defined by the functions. Therefore, equations such as Equation 17, which define one-to-one mappings between the utility components and the dependent variables, do not exist, and chi-square distributed quadratic forms analogous to R* cannot be developed. Moreover, the sampling distributions of aggregate shares and changes in choice probabilities contain intractable integrals that prevent these distributions from being used to form confidence regions.

Confidence Regions Based on a Linear Approximation

Equation 4 for the estimated choice probabilities can be expanded in a Taylor series about $\hat{V}_j = V_j$ ($j=1, \ldots, J$) to obtain

$$\hat{P}_i = P_i + \sum_{j=1}^{J} (\partial P_i/\partial V_j)(V_j - V_j) + \Delta \quad (i = 1, \ldots, J) \quad (20)$$

where Δ is a remainder term. As the size of the sample used in estimating the $\hat{V}_i$ approaches infinity, Δ converges in probability to zero and $\hat{P}_i$ converges in probability to (11):

$$\hat{P}_i = P_i + \sum_{j=1}^{J} (\partial P_i/\partial V_j)(\hat{V}_j - V_j) \quad (21)$$

The random variables $\{\hat{V}_j - V_j\}$ are asymptotically jointly normally distributed with mean values of zero and covariance matrix D^{-1}, where

$$(D^{-1})_{jk} = \sum_{r=1}^{M} \sum_{s=1}^{M} X_{jr} X_{ks} (A^{-1})_{rs} \quad (22)$$

and A is the matrix defined in Equation 5. Therefore, $\hat{P}_i$ is asymptotically normally distributed with mean value P_i and variance

$$\hat{\text{var}}(\hat{P}_i) = \sum_{j=1}^{J} \sum_{k=1}^{J} (\partial P_i/\partial V_j)(\partial P_i/\partial V_k)(D^{-1})_{jk} \quad (i = 1, \ldots, J) \tag{23}$$

It follows that an asymptotic 100(1-ϵ) percent confidence interval for P_i is

$$\hat{P}_i - Z_{\epsilon/2}[\text{var}(\hat{P}_i)]^{1/2} \leq P_i \leq \hat{P}_i + Z_{\epsilon/2}[\text{var}(\hat{P}_i)]^{1/2} \tag{24}$$

where $Z_{\epsilon/2}$ is the 1-ϵ/2 percentile of the standard normal distribution. The numerical value of var $(\hat{P}_i)$ can be approximated by substituting $\hat{V}$ for V and $\hat{P}$ for P in Equation 23. Equation 21, which is a well-known approximation in mathematical statistics, formed the basis of Koppelman's analysis of errors in disaggregate models (5, 6).

Equation 24 can also be used to develop rectangular joint confidence regions for the P_i. Let I_i be a 100(1-ϵ/J) confidence region for P_i as given by Equation 24. Then

$$\Pr(P_1 \epsilon I_1, P_2 \epsilon I_2, \ldots, P_J \epsilon I_J) \geq 1 - \epsilon \tag{25}$$

Thus $\{P_i;\ i=1, \ldots, J\}$ is contained in the J-dimensional rectangular region $P_1 \in I_1, \ldots, P_J \in I_J$ and has a confidence level that equals or exceeds 100(1 - ϵ) percent.

The confidence interval defined by inequalities 24 and the joint confidence region defined by inequality 25 can easily be generalized to apply to functions of choice probabilities, including aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The generalization consists of substituting the functions of interest in place of the choice probabilities in Equations 21-24. The generalization of Equation 23 to aggregate market shares is given by Koppelman (5, 6).

The advantages of the confidence regions defined by inequalities 24 and 25 are substantial: The regions are rectangular, generalizable to functions of the choice probabilities, and computationally tractable. However, because of the regions' reliance on the asymptotic approximation of Equation 21, the accuracy of the confidence levels associated with the regions can vary greatly and may be highly erroneous. This variation in accuracy is illustrated in the following examples.

Consider the univariate, binomial logit model

$$P_i = \exp(\alpha X_i)/[\exp(\alpha X_1) + \exp(\alpha X_2)] \quad (i = 1, 2) \tag{26}$$

where X_i is the explanatory variable of the model evaluated for alternative i and α is a constant. Let a be the maximum likelihood estimator of α, and let the sampling variance of a be σ^2. Assume that $X_1 = 0$, $X_2 = 0.1$, a = 3, and $\sigma = 1$. Then from inequalities 24, a 95 percent confidence interval for P_1 is $0.378 \leq P_1 \leq 0.474$. The confidence level associated with this interval also can be computed without using approximation 21 by noting that $0.378 \leq P_1 \leq 0.474$ is equivalent to $1.041 \leq \alpha \leq 4.980$. Using the asymptotic normality of the estimated coefficient a, the confidence level associated with $1.041 \leq \alpha \leq 4.980$ and, therefore, with $0.378 \leq P_1 \leq 0.474$ can be computed to be 95.12 percent. Thus, in this example, inequalities 24 yield an accurate estimate of the confidence level.

Now let $X_2 = 1.0$ while X_1, a, and σ remain unchanged. Then inequalities 24 yield $-0.041 \leq P_1 \leq 0.136$ as a 95 percent confidence interval for P_1. If the confidence level associated with this interval is computed directly from the asymptotic distribution of a without using the approximation 21, a confidence level of 87.5 percent is obtained. A true 95 percent confidence interval for P_1 is $0 \leq P_1 \leq 0.205$. Thus, in this case inequalities 24 yield erroneous results.

Nonlinear Programming Approach to Developing Confidence Regions

A method for deriving joint rectangular confidence regions for multinomial logit-choice probabilities without using approximation 21 is described in this section. Denote the vectors of true coefficients ($\alpha_1, \ldots, \alpha_M$) and estimated coefficients ($\alpha_1, \ldots, \alpha_M$) by $\underline{\alpha}$ and $\underline{a}$, respectively. Let $Q(\underline{a}, \underline{\alpha})$ be the quadratic form defined in Equation 7, and let $\chi^2(\epsilon, M)$ be the 100(1 - ϵ) percentile of the chi-square distribution with M degrees of freedom. Recall that P_i (i = 1, ..., J) is a function of $\underline{\alpha}$. Given $\underline{a}$ and ϵ, define $b_i(\epsilon)$ and $B_i(\epsilon)$ for each i by the following nonlinear programming problems:

$$b_i(\epsilon) = \min P_i(\alpha) \quad (i = 1, \ldots, J) \tag{27}$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\epsilon, M)$

$$B_i(\epsilon) = \max P_i(\underline{\alpha}) \quad (i = 1, \ldots, J) \tag{28}$$

subject to $Q(\underline{a}, \alpha) \leq \chi^2(\epsilon, M)$. The maximizations and minimizations are carried out over variations in $\underline{\alpha}$. Then the inequalities

$$b_i(\epsilon) \leq P_i \leq B_i(\epsilon) \quad (i = 1, \ldots, J) \tag{29}$$

define a rectangular joint confidence region for the P_i with confidence level equal to or greater than 100(1 - ϵ) percent (12).

Another rectangular joint confidence region for the P_i with the same confidence level can be computed by considering P_i to be a function of the utilities ($V_1, \ldots, V_J$). Let $R(\hat{\underline{V}}, \underline{V})$ be the quadratic form defined in Equation 9. Then the solutions to the nonlinear programming problems

$$b_i(\epsilon) = \min P_i(\underline{V}) \quad (i = 1, \ldots, J) \tag{30}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon, J-1)$

$$B_i(\epsilon) = \max P_i(\underline{V}) \quad (i = 1, \ldots, J) \tag{31}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\epsilon, J-1)$ define joint lower and upper confidence limits for P_i with confidence level equal to or greater than 100(1-ϵ) percent. The maximizations and minimizations are performed over variations in $\underline{V}$. The confidence limits thus defined are closer together than the 100(1-ϵ) confidence limits defined by problems 27 and 28 when J-1 < M.

The confidence limits defined by problems 27 and 28 can easily be extended to functions of the choice probabilities. The extension consists of using the relevant functions of the choice probabilities as the objective functions of problems 27 and 28. For example, if P_{in} is the probability that individual n chooses alternative i, the aggregate market share of alternative i in a population of N individuals is

$$\Pi_i = (1/N) \sum_{n=1}^{N} P_{in} \quad (i = 1, \ldots, J) \tag{32}$$

Π_i is a function of α through the P_{in}. Joint confidence limits b_i and B_i for the Π_i with confidence level equal to at least 100(1-ϵ) percent are given by

$$b_i(\epsilon) = \min \Pi_i(\underline{\alpha}) \quad (i = 1, \ldots, J) \tag{33}$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\epsilon, M)$

$$B_i(\epsilon) = \max \underset{i}{\Pi}(\underline{\alpha}) \quad (i = 1, \ldots, J) \tag{34}$$

subject to $Q(a, \alpha) \le \chi^2(\epsilon, M)$.

The joint rectangular confidence region that results from the asymptotic linear approximation (inequalities 24 and 25) can be obtained by solving the nonlinear programming problems

$$b_i^*(\epsilon) = \min P_i^*(\underline{V}) \quad (i = 1, \ldots, J) \tag{35}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \le \chi^2(\epsilon/J, 1)$

$$B_i^*(\epsilon) = \max P_i^*(\underline{V}) \quad (i = 1, \ldots, J) \tag{36}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \le \chi^2(\epsilon/J, 1)$, where $\hat{P}_i^*(\underline{V})$ is the expression obtained by exchanging P_i with $\hat{P}_i$ and $\hat{V}_j$ with V_j in Equation 21, and b_i^* and B_i^*, respectively, are the lower and upper confidence limits for P_i obtained by the linear approximation method. As the accuracy of the asymptotic linear approximation increases, problems 35 and 36 approach equivalence with the problems

$$b_i^*(\epsilon) = \min P_i(\underline{V}) \quad (i = 1, \ldots, J) \tag{37}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \le \chi^2(\epsilon/J, 1)$

$$B_i^*(\epsilon) = \max P_i(\underline{V}) \quad (i = 1, \ldots, J) \tag{38}$$

subject to $R(\hat{\underline{V}}, \underline{V}) \le \chi^2(\epsilon/J, 1)$. Problems 37 and 38 differ from problems 30 and 31 only in the right-hand sides of their constraints. Comparison of problems 37 and 38 with problems 30 and 31 and problems 27 and 28 provides a means of determining whether the asymptotic linear approximation or the nonlinear programming method yields a smaller joint confidence region for the choice probabilities when the linear approximation is accurate. If $J-1<M$, the linear approximation yields narrower confidence limits for each of the P_i whenever

$$\chi^2(\epsilon/J, 1) < \chi^2(\epsilon, J-1) \tag{39}$$

If $M \le J-1$, the linear approximation yields narrower limits whenever

$$\chi^2(\epsilon/J, 1) < \chi^2(\epsilon, M) \tag{40}$$

Conditions 39 and 40 will be satisfied at normal confidence levels unless the number of coefficients M is very small or the number of alternatives J is either two or very large. For example, if $M = 4$ and $\epsilon = 0.05$, conditions 39 and 40 will be satisfied if $3 \le J \le 24$. If $M = 5$ and $\epsilon = 0.05$, the conditions will be satisfied if $3 \le J \le 61$. Thus, when the asymptotic linear approximation is accurate it will tend to produce smaller joint confidence regions than will the nonlinear programming method unless the choice set either is large or contains only two alternatives. Numerical illustrations of the differences in the sizes of the linear approximation and nonlinear programming confidence regions are given in a later section.

A BOUND ON THE ERROR IN THE CONFIDENCE LEVEL

The errors in the linear approximation confidence levels of a binary choice model were previously computed exactly. This exact computation is not possible for models that have more than two alternatives in their choice sets. In multinomial models, nonlinear programming can be used to establish upper bounds on the errors in the confidence levels obtained from inequalities 24.

Let P_i^* be defined as in problems 35 and 36, and let σ^* be the linear approximation estimate of the standard deviation of $\hat{P}_i$ obtained from Equation 23. Note that P_i^* depends on the true coefficients α through $\underline{V}$. For arbitrary positive K and k define the following sets:

$$S_1(K) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \le K\} \tag{41}$$

$$S_2(K) = \{\underline{\alpha} \mid |\hat{P}_i - P_i| \le K\} \tag{42}$$

$$S_3(K) = \{\underline{\alpha} \mid |P_i^* - P_i| \le K\} \tag{43}$$

$$S_4(K, k) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \le K - k\} \tag{44}$$

$$S_5(K, k) = \{\underline{\alpha} \mid |P_i^* - \hat{P}_i| \le K + k\} \tag{45}$$

The sets S_1 through S_5 all depend on the estimated coefficients $\underline{a}$ and, therefore, are random events. Let $\Pr(S_j)$ be the probability of the event S_j ($j = 1, \ldots, 5$). Note that

$$S_2 \cap S_3 \subset S_5 \tag{46}$$

and

$$\Pr(S_2 \cap S_3) \ge \Pr(S_2) + \Pr(S_3) - 1 \tag{47}$$

Therefore,

$$\Pr(S_2) \le \Pr(S_5) + [1 - \Pr(S_3)] \tag{48}$$

Also,

$$S_4 \cap S_3 \subset S_2 \tag{49}$$

and

$$\Pr(S_4 \cap S_3) \ge \Pr(S_4) + \Pr(S_3) - 1 \tag{50}$$

Therefore,

$$\Pr(S_2) \ge \Pr(S_4) - [1 - \Pr(S_3)] \tag{51}$$

when probabilities 48 and 51 are combined,

$$\Pr(S_4) - [1 - \Pr(S_3)] \le \Pr(S_2) \le \Pr(S_5) + [1 - \Pr(S_3)] \tag{52}$$

$P_i^* - \hat{P}_i$ is asymptotically normally distributed with mean zero and standard deviation σ^*, by virtue of Equation 21. Let Φ denote the cumulative standard normal distribution function. Then asymptotically

$$\Pr(S_4) = 2\,\Phi\,[(K - k)/\sigma^*] - 1 \tag{53}$$

$$\Pr(S_5) = 2\,\Phi\,[(K + k)/\sigma^*] - 1 \tag{54}$$

$$\Pr(S_1) = 2\,\Phi\,(K/\sigma^*) - 1 \tag{55}$$

Inequality 52 and Equations 53-55 imply

$$\begin{aligned} 2\{\Phi[(K-k)/\sigma^*] - \Phi(K/\sigma^*)\} - [1 - \Pr(S_3)] &\le \Pr(S_2) - \Pr(S_1) \\ &\le 2\{\Phi[(K+k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - \Pr(S_3)] \end{aligned} \tag{56}$$

Given a confidence level $100(1-\epsilon)$ percent, let K be given by the solution to

$$\Pr[S_1(K)] = 1 - \epsilon \tag{57}$$

Note that in the linear approximation method for developing confidence intervals P_i^* and P_i are considered to be equal. Therefore, $100(1-\epsilon)$ is the confidence level that the linear approximation assigns to the interval $|P_i - \hat{P}_i| \le K$, whereas $100\Pr[S_2(K)]$ is the confidence level that is obtained if the linear approximation is not used. Thus, $100[\Pr(S_2) - \Pr(S_1)]$ is the error in the confidence level that is made by using the linear approximation, and inequalities 56 bound this error. Specifically, for any k

$$|\Pr(S_2) - \Pr(S_1)| \leq \max[F^+(K, k), F^-(K, k)] \quad (58)$$

where

$$F^+(K, k) = 2\{\Phi[(K + k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - \Pr(S_3)] \quad (59)$$

and

$$F^-(K, k) = -2\{\Phi[(K - k)/\sigma^*] - \Phi(K/\sigma^*)\} + [1 - \Pr(S_3)] \quad (60)$$

In practice it is usually difficult or impossible to evaluate $\Pr(S_3)$. Thus, inequality 58 is not directly useful. However, it is possible to establish a computationally tractable lower bound on $\Pr(S_3)$. Given a number δ that satisfies $0 < \delta < 1$, define $k(\delta)$ by the following nonlinear programming problem:

$$k(\delta) = \max |P_i^*(\underline{\alpha}) - P_i(\underline{\alpha})| \quad (61)$$

subject to $Q(\underline{a}, \underline{\alpha}) \leq \chi^2(\delta, M)$, if $M \leq J-1$ and $R(\hat{\underline{V}}, \underline{V}) \leq \chi^2(\delta, J-1)$ otherwise

$$\Pr[|P_i^* - P_i| \leq k(\delta)] \geq 1 - \delta \quad (62)$$

and

$$\Pr\{S_3[k(\delta)]\} \geq 1 - \delta \quad (63)$$

Given K and δ, define $G^+(K, \delta)$ and $G^-(K, \delta)$ by

$$G^+(K, \delta) = 2(\Phi\{[K + k(\delta)]/\sigma^*\} - \Phi(K/\sigma^*)) + \delta \quad (64)$$

$$G^-(K, \delta) = -2(\Phi\{[K - k(\delta)]/\sigma^*\} - \Phi(K/\sigma^*)) + \delta \quad (65)$$

Then

$$G^+(K, \delta) \geq F^+[K, k(\delta)] \quad (66)$$

$$G^-(K, \delta) \geq F^-[K, k(\delta)] \quad (67)$$

and

$$|\Pr(S_2) - \Pr(S_1)| \leq \min_{\delta}\{\max[G^+(K, \delta), G^-(K, \delta)]\} \quad (68)$$

Inequality 68 defines a computationally tractable upper bound on the error in the confidence level obtained from inequalities 24.

The degree to which the right-hand side of inequality 68 overestimates the error made by linear approximation 21 can be illustrated with the model of Equation 26. It was shown that when $X_1 = 0$, $X_2 = 0.1$, $a = 3$, and $\sigma = 1$ in Equation 26, the linear approximation assigns a confidence limit of 95 percent to a particular confidence interval for the coefficient α, whereas a confidence level of 95.12 percent is obtained for the same interval when the linear approximation is not used. In this case the linear approximation makes an error of 0.12 percent in the confidence level. When $X_2 = 1.0$ and the other parameters remain unchanged, the linear approximation assigns a confidence level of 95 percent to an interval whose confidence level is found to be 87.5 percent when the linear approximation is not used. In this case, the linear approximation makes an error of 7.5 percent in the confidence level. Inequality 68 gives an upper bound on the error in the confidence level of 1.2 percent when $X_2 = 0.1$ and 31 percent when $X_2 = 1.0$. Although inequality 68 considerably overestimates the error made by the linear approximation in both cases, the error estimates obtained from inequality 68 do distinguish between a case in which the linear approximation is useful (e.g., $X_2 = 0.1$), and a case in which the linear approximation is not useful (e.g., $X_2 = 1.0$).

Inequality 68 can be extended to functions of the choice probabilities, such as aggregate market shares. The extension is accomplished by substituting the desired functions in place of $\hat{P}_i$, P_i, and $P_i{}^*$ in equations and inequalities 41-68 and by using the Q form of the constraint in problem 61.

NUMERICAL EXAMPLES

To illustrate and compare the linear approximation and nonlinear programming methods for developing confidence regions, both methods were applied to two multinomial logit models: a 3-alternative model of work-trip mode choice (5) and a 20-alternative model of destination choice for nonwork trips (13). Typical values of the explanatory variables were used in each case. The nonlinear programming problems 27, 28, 31, 32, and 61 were solved by using the sequential unconstrained minimization technique (14).

Joint 95 percent confidence limits for the choice probabilities of the mode choice model are shown in Table 1. The upper and lower confidence limits of the choice probabilities are, respectively, approximately 17 percent above and below these probabilities. The nonlinear programming confidence intervals were obtained from problems 30 and 31 and are approximately 2 percent wider than the linear approximation intervals. Inequality 68 indicates that the errors in the confidence levels of the linear approximation confidence intervals considered individually are less than 1.14 percent. Considering the looseness of the bound provided by inequality 68, this suggests that the linear approximation achieves acceptable accuracy in this example.

Table 1. Joint 95 percent confidence intervals for the choice probabilities in a three-alternative mode choice model.

		Linear Approximation Method		Nonlinear Programming Method		
Alternative	P	b	B	b	B	R
1	0.402	0.338	0.467	0.338	0.470	1.02
2	0.312	0.262	0.362	0.263	0.366	1.02
3	0.286	0.234	0.337	0.236	0.341	1.02

Note: P = estimated choice probability; b = lower confidence limit; B = upper confidence limit; and R = width of nonlinear programming confidence interval divided by width of linear approximation interval.

Table 2. Joint 95 percent confidence intervals for the choice probabilities in a 20-alternatives destination choice model.

		Linear Approximation Method		Nonlinear Programming Method		
Alternative	P	b	B	b	B	R
1	0.022	0.017	0.027	0.017	0.028	1.10
2	0.029	0.023	0.035	0.023	0.037	1.10
3	0.017	0.011	0.022	0.012	0.023	1.11
4	0.035	0.027	0.043	0.026	0.044	1.10
5	0.024	0.013	0.035	0.015	0.039	1.13
6	0.034	0.029	0.039	0.029	0.040	1.09
7	0.056	0.039	0.073	0.040	0.078	1.11
8	0.036	0.030	0.042	0.030	0.043	1.10
9	0.025	0.020	0.031	0.020	0.032	1.10
10	0.049	0.041	0.057	0.040	0.058	1.09
11	0.111	0.075	0.147	0.077	0.157	1.11
12	0.083	0.075	0.091	0.074	0.092	1.10
13	0.089	0.066	0.112	0.066	0.117	1.10
14	0.066	0.056	0.076	0.056	0.078	1.10
15	0.080	0.069	0.090	0.069	0.091	1.10
16	0.077	0.064	0.089	0.064	0.091	1.10
17	0.018	0.010	0.025	0.011	0.028	1.13
18	0.063	0.052	0.074	0.052	0.075	1.10
19	0.044	0.033	0.056	0.033	0.058	1.10
20	0.043	0.037	0.049	0.037	0.050	1.10

Note: P = estimated choice probability; b = lower confidence limit; B = upper confidence limit; and R = width of nonlinear programming confidence interval divided by width of linear approximation interval.

Joint 95 percent confidence limits for the choice probabilities of the destination choice model are shown in Table 2. The upper and lower confidence limits of the choice probabilities are, respectively, roughly 10 to 40 percent above and below these probabilities, depending on the alternative. The nonlinear programming confidence intervals are approximately 10 percent wider than the linear approximation intervals. Inequality 68 indicates that the errors in the confidence levels of the linear approximation confidence intervals considered individually are less than 0.8 percent, again suggesting that the linear approximation is acceptably accurate.

CONCLUSIONS

This paper has described three methods for developing confidence regions for the choice probabilities of the multinomial logit model. One method involves a direct application of the asymptotic sampling distribution of the choice probabilities and yields joint confidence regions for these probabilities. The confidence regions are not rectangular and, therefore, are useful mainly for testing hypotheses about the choice probabilities.

The other two methods are based, respectively, on a linear approximation of the relation between errors in the coefficients of a model and errors in the choice probabilities, and on a nonlinear programming approach to developing confidence intervals. Both of these methods produce joint rectangular confidence regions for the choice probabilities, and both can be applied to functions of the choice probabilities, such as aggregate market shares and changes in choice probabilities caused by changes in explanatory variables. The linear approximation method also can be used to develop confidence intervals for individual choice probabilities.

The linear approximation method is computationally simpler than the nonlinear programming method. Moreover, when the linear approximation on which the method is based is accurate, the linear approximation method produces a smaller confidence region for a given confidence level than does the nonlinear programming method, unless the choice set either is very large or contains only two alternatives. However, the linear approximation method has the disadvantage that it can yield erroneous results.

A procedure for bounding the error made by the linear approximation method has been described in this paper. However, this procedure is based on nonlinear programming, and the computational effort involved in implementing it can equal or exceed the computational effort involved in developing confidence regions by the nonlinear programming method. If there are a priori reasons for believing that the linear approximation method will yield accurate results in a particular application, then the computational simplicity of this method makes it preferable to the nonlinear programming method. However, if the accuracy of the linear approximation method is questionable and resources for implementing the bounding procedure are not available, then the nonlinear programming method will yield more reliable results than will the linear approximation method.

The linear approximation and nonlinear programming methods for developing confidence regions can be applied to other utility maximizing models with linear-in-parameters utility functions (e.g., multinomial probit) by substituting the choice probabilities of the desired model in place of the logit probabilities used in this paper.

ACKNOWLEDGMENT

Thanks are due to Garth McCormick for providing a computer program that implements the sequential unconstrained minimization technique. The views expressed in this paper are mine and are not necessarily endorsed by the U.S. Environmental Protection Agency.

REFERENCES

1. T. A. Domencich and D. McFadden. Urban Travel Demand. North-Holland-Elsevier, New York, 1975.
2. B. D. Spear. Applications of New Travel Demand Forecasting Techniques to Transportation Planning. U.S. Department of Transportation, 1977.
3. M. Ben-Akiva and T. J. Atherton. Methodology for Short-Range Travel Demand Predictions. Journal of Transport Economics and Policy, Vol. 11, 1977, pp. 224-261.
4. H. Theil. Principles of Econometrics. Wiley, New York, 1971.
5. F. S. Koppelman. Travel Prediction with Models of Individual Choice Behavior. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, MA, Rept. CTS 75-7, 1975.
6. F. S. Koppelman. Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models. TRB, Transportation Research Record 592, 1976, pp. 17-23.
7. D. McFadden. Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics (P. Zarembka, ed.), Academic Press, New York, 1974, pp. 105-142.
8. M. Fisz. Probability Theory and Mathematical Statistics. Wiley, New York, 1963.
9. N. Johnson. Systems of Frequency Curves Generated by Methods of Translation. Biometrika, Vol. 36, 1949, pp. 149-176.
10. R. B. Westin. Predictions from Binary Choice Models. Journal of Econometrics, Vol. 2, 1974, pp. 1-16.
11. H. Chernoff. Large Sample Theory: Parametric Case. Annals of Mathematical Statistics, Vol. 27, 1956, pp. 1-22.
12. C. R. Rao. Linear Statistical Inference and Its Applications. 2d ed. Wiley, New York, 1973.
13. J. Horowitz. Disaggregate Model for Nonwork Travel. TRB, Transportation Research Record 673, 1978, pp. 65-71.
14. A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York, 1968.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.