MULTINOMIAL LOGIT SPECIFICATION TESTS

Cheng Hsiao
Princeton University
University of Toronto

Kenneth A. Small
Princeton University
University of California - Irvine

MULTINOMIAL LOGIT SPECIFICATION TESTS

by

Kenneth A. Small

and

Cheng Hsiao

July 1983

ABSTRACT

We investigate specification tests based on comparing multinomial logit parameter estimates formed on a full and on a restricted choice set. Using a likelihood ratio, as originally proposed by McFadden, Train and Tye (MTT), leads to a downward biased test statistic, for which we provide a very rough correction factor. Randomly splitting the sample first leads to an upward biased test statistic. We then show that by forming unrestricted-choice-set estimates separately on both parts of a split sample, a linear combination of them can be combined with the restricted-choice-set estimate into a likelihood ratio which does lead to a known (chi square) asymptotic distribution, hence provides a more rigorous test.

We compare our proposed test with one suggested by Hausman and McFadden, in which a quadratic form in the difference between the two estimates used by MTT is computed. Their test statistic and ours have the same asymptotic distribution, though ours may sometimes converge more slowly. However, ours has fewer computational problems and, based on an empiricl travel-demand application, appears much less volatile in finite samples.

MULTINOMIAL LOGIT SPECIFICATION TESTS

I. INTRODUCTION

The multinomial logit model has achieved an established place in travel demand analysis and other areas in which the choice among discrete alternatives is analyzed. At the same time, its limitations in accounting for differential degrees of substitutability among the discrete alternatives, particularly as embodied in its "independence of irrelevant alternatives" (IIA) property, have become widely known.[1] This has led to an interest in statistical tests to detect important departures from the multinomial logit specification.

Unfortunately, it is not clear how one implements the conventional likelihood ratio or Lagrangian multiplier test in this context without specifying a well defined alternative, which requires extensive computer programming. On the other hand, since the IIA assumption implies that the ratio of the probabilities of choosing any two alternatives is independent of a third choice, testing procedures have been suggested based on eliminating one or more alternatives from the choice set to see whether the coefficient estimates are affected. Such procedures can be carried out with existing computer programs and, as illustrated by our empirical application, can be designed to test against specific types of departure from logit even without formulating a formal model. The development of such tests was pioneered by McFadden, Train and Tye (1977), studied and modified by Horowitz (1981), and recently extended by Hausman and McFadden (1981), who apply the general technique of Hausman (1978) to this case. The tests in various forms have been used by Lave and Train (1979), Horowitz (1980), Abkowitz (1981), and Small (1982), among others.

---

[1] For reviews of these issues, see McFadden (1973), Domencich and McFadden (1975), Manski (1981), Amemiya (1981), or McFadden (1981).

An important class of such "diagnostic" tests which is particularly easy to implement is based on a likelihood ratio. Two parameter estimates are formed, one based on the full choice set and the other on a restricted choice set, and their likelihoods compared. The version originally proposed by McFadden, Train and Tye (1977), which we call the MTT test, computed each estimate using the largest possible sample: the first on the full sample, and the second on the subsample for which the chosen alternative is a member of the restricted choice set; the likelihoods were then computed using this subsample. By analogy with the classical likelihood ratio test for a parameter restriction, the resulting test statistic was thought to be approximately asymptotically chi square. However, we show in this paper that the MTT test is asymptotically biased toward accepting the null hypothesis; a variant which we call the "split-sample MTT test" has the opposite bias. We provide rough-and-ready correction factors which, though strictly valid only under stringent conditions, seem to work well in an empirical example. We then propose a closely related test statistic whose asymptotic distribution _is_ chi square.

The test statistic proposed by Hausman and McFadden (1981) also converges to the chi square distribution. In finite samples, however, the two tests have quite different properties. Ours has the disadvantage that parameter estimates are formed on random subsamples and therefore have larger standard errors than the estimates used by Hausman-McFadden; thus our test statistic may converge more slowly in well-behaved examples. The Hausman-McFadden test, however, requires the inversion of the difference between two closely

related matrices. Not only does this involve matrix manipulations not required by our test, but it can lead to important differences between the actual and the asymptotic distribution of the test statistic. This point is illustrated by an empirical application in which three asympotically equivalent versions of the Hausman-McFadden test yield greatly differing results, including negative test statistics; whereas our test shows much less small-sample variation.

In the next section, we define the MTT, split-sample MTT, and Hausman-McFadden tests and derive the asymptotic distribution of the test statistic for each. This makes clear the biases in the first two and the similarities among all three, while laying the foundation for our proposed modification. In Section III, we present our proposed test, and derive its asymptotic distribution. Section IV then applies the tests to an empirical model of trip timing, for which other work has already detected small but statistically significant departures from the multinomial logit specification. Concluding comments follow.

II. DISTRIBUTION OF THE LIKELIHOOD RATIO TEST STATISTICS

Suppose each member t of a sample $t = 1,\ldots,N$ is observed to choose a discrete alternative j from a choice set C. Define $S_{tj} = 1$ if j is the choice for t and 0 otherwise. The multinomial logit model explains these choices on the basis of vectors z_{tj} of observed characteristics and the (Kx1) parameter vector β. Maximum likelihood estimation chooses β to maximize

$$(1) \qquad L(\beta) = \sum_{t=1}^{N} \sum_{j \varepsilon C} S_{tj} \log P_{tj}$$

where

$$(2) \qquad P_{tj} = \exp(\beta' z_{tj}) / \sum_{\ell \varepsilon C} \exp(\beta' z_{t\ell}).$$

We will need the negative Hessian of this function which, following McFadden (1973), can be written

$$(3) \qquad H(\beta) = -\partial^2 L / \partial \beta \partial \beta' = \sum_{t=1}^{N} \sum_{j \varepsilon C} P_{tj} (z_{tj} - \bar{z}_t)(z_{tj} - \bar{z}_t)'$$

where

$$(4) \qquad \bar{z}_t = \sum_{j \varepsilon C} P_{tj} z_{tj}$$

is a function of β. Note that if we define $z_t = \sum_{j \varepsilon C} S_{tj} z_{tj}$, a random variable equal to the characteristics vector of the chosen alternative, then each term in (3) is its covariance matrix conditional on β. Note also that with a trivial extension of notation, we could generalize to the case where choice set C differs for different sample members.

The specification tests in question depend on a mathematical property peculiar to the logit model (2): The choice probability conditional on a subset $D \subset C$ of the choice set,

$$(5) \qquad P_{tj|D} = P_{tj} / \sum_{\ell \varepsilon D} P_{t\ell},$$

happens also to have the logit form (2) except with the summation just over D. Thus a consistent estimate of some components of β can be obtained by restricting the sample to those N_1 members for which the observed choice

lies in D , and estimating the model as if D were the entire choice set. However some components of β , such as the coefficient of any variable which does not vary among alternatives in D , may not be identified on this subsample. Let $\beta = (\gamma, \theta)$ be a partition of β into the $(K-k) \times 1$ vector γ which is not so identified and the $k \times 1$ vector θ which is, and let $z_t = (y_t, x_t)$ be a commensurate partition of the explanatory variables. The conditional choice probability is then

$$P_{tj}\big|_D = \exp(\theta' x_{tj}) / \sum_{\ell \in D} \exp(\theta' x_{t\ell}) \ .$$

Renumbering the sample members so that the first N_1 belong to the subsample for which the chosen alternative is in choice subset D , the conditional log likelihood is

$$(6) \qquad L_1(\theta) = \sum_{t=1}^{N_1} \sum_{j \in D} S_{tj} \log P_{tj}\big|_D$$

with negative Hessian

$$(7) \qquad H_1(\theta) = -\partial^2 L_1 / \partial \theta \partial \theta' = \sum_{t=1}^{N_1} \sum_{j \in D} P_{tj}\big|_D (x_{tj} - \bar{x}_{tD})(x_{tj} - \bar{x}_{tD})'$$

where

$$(8) \qquad \bar{x}_{tD} = \sum_{j \in D} P_{tj}\big|_A x_j \ .$$

Each term in (7) is the covariance, conditioned on choice of some alternative in D , of a random variable $x_{tD} \equiv \sum_{j \in D} S_{tj} x_{tj}$.

Let $\hat{\beta}_0 \equiv (\hat{\gamma}_0, \hat{\theta}_0)$ and $\hat{\theta}_1$ denote the parameters which maximize L and L_1, respectively, and let $\hat{q} = \hat{\theta}_1 - \hat{\theta}_0$. We seek tests of whether departures from the logit specification cause $\hat{q}$ to differ from zero.

To obtain asymptotic results, we assume that increasing the sample size involves sampling from a given underlying distribution of characteristics. Thus under the null hypothesis that the model is correctly described by (2) with true parameter vector $\beta_0 \equiv (\gamma_0, \theta_0)$, the following probability limits exist:

$$(9) \qquad S = \text{plim } H(\beta_0)/N$$

$$(10) \qquad S_1 = \text{plim } H_1(\theta_0)/N_1$$

$$(11) \qquad a = \text{plim } N_1/N.$$

<u>Hausman-McFadden</u>. Hausman and McFadden (1981) have shown that, under the null hypothesis, $\sqrt{N}\hat{q}$ is asymptotically normal with variance-covariance matrix[2]

$$(12) \quad \Sigma \equiv \text{plim Var}(\sqrt{N}\hat{q}) = \text{plim Var}(\sqrt{N}\hat{\theta}_1) - \text{plim Var}(\sqrt{N}\hat{\theta}_0) .$$

The asymptotic variance of $\sqrt{N_1}\hat{\theta}_1$ or $\sqrt{N}\hat{\beta}_0$ is the inverse of the corresponding asymptotic **negative Hessian** S_1 or S, respectively; dividing the first by $a \equiv \text{plim}(N_1/N)$ and partitioning the second gives

$$(13) \quad \Sigma = (1/a)S_1^{-1} - S_0^{-1} ,$$

where S_0^{-1} is defined to be the kxk lower right submatrix of S^{-1} .

Hence the quadratic form

$$(14) \qquad Q \equiv \hat{q}'(\hat{\Sigma}/N)^{-1}\hat{q}$$

is asymptotically distributed chi square with k degrees of freedom, provided

---

[2]The proof follows from the general result of Hausman (1978) that if, under the null hypothesis, $\hat{\theta}_0$ is efficient and $\hat{\theta}_1$ is consistent, then $(\hat{\theta}_1 - \hat{\theta}_0)$ is uncorrelated with $\hat{\theta}_0$.

$\hat{\Sigma}$ is a nonsingular[3] consistent estimate of Σ . The

"standard" estimate of Σ/N is obtained from (12) by sub-

stituting the usual estimates for the covariances of $\hat{\theta}_1$ and $\hat{\theta}_0$, namely

$H_1^{-1}(\hat{\theta}_1)$ and $H_0^{-1}(\hat{\beta}_0)$, where H_0^{-1} is the kxk lower right submatrix of H^{-1} .

An asymptotically equivalent pair of estimates is $H_1^{-1}(\hat{\theta}_1)N_1/(N_1-k)$ and

$H_0^{-1}(\hat{\theta}_0)N/(N-k)$, which are the standard estimates corrected for finite-sample

degrees of freedom in analogy to ordinary least squares. Yet a third pair is

$[E_{\hat{\theta}_0} H_1 (\hat{\theta}_0)]^{-1}$ and $H_0^{-1}(\hat{\beta}_0)$, where $E_{\hat{\theta}_0}$ is the expectation (on the full

sample) conditional on $\theta = \hat{\theta}_0$; its virtue is that $\hat{\Sigma}/N$ is then guaranteed

nonnegative definite (Hausman and McFadden, 1981, p. 15).

McFadden-Train-Tye. The original MTT test uses not $\hat{q}$ itself, but rather the

amount by which the log-likelihood calculated on the subsample is affected by

the differing estimates of θ . The test statistic used by MTT is

$$(15) \qquad \Delta_1 \equiv -2[L_1(\hat{\theta}_0) - L_1(\hat{\theta}_1)] .$$

The test is based on the hope that the distribution of Δ_1 is approximately

chi square with k degrees of freedom.

As it turns out, this in general is not true. Taking a Taylor Series

expansion of $L_1(\hat{\theta}_0)$ around $\hat{\theta}_1$ we see that

$$(16) \qquad \Delta_1 \tilde{=} \hat{q}'H_1(\hat{\theta}_1)\hat{q} .$$

---

[3] If $\hat{\Sigma}$ is singular, its inverse is replaced by a generalized inverse. This causes no problem in principle, though recognizing an exactly singular matrix can be tricky in practice due to computational inaccuracies. We looked rather carefully for singularities in the empirical work reported in Section IV, but did not find any. Hausman and McFadden examine sufficient conditions for Σ to be nonsingular, and conclude that singularity "will occur only for exceptional configurations of the . . . variables" [p. 14]. This does not preclude computational problems from near-singularity, however.

Using (10) and (13), we see that (16) converges to approximately the same thing as (14) if $a \doteq 1$ and if the second term in (13) is negligible compared to the first term. More precisely, the probability limit of (16) is a quadratic form, with symmetric matrix (aS_1), in a vector having a multivariate normal distribution with variance-covariance matrix Σ ; it therefore has a chi square distribution if and only if [4] $\Sigma = (1/a)S_1^{-1}$. But (13) shows that $(1/a)S_1^{-1}$ exceeds Σ by the positive definite matrix $S_0^{-1} = \text{plim Var}(\sqrt{N}\hat{\theta}_0)$. In the appendix we take advantage of this fact to prove that under the null hypothesis, Δ_1 is asymptotically distributed as

$$(17) \qquad \Delta_1 = \sum_{i=1}^{k} \lambda_i w_i^2$$

where λ_i are the characteristic roots of $aS_1\Sigma$, each satisfying $0 < \lambda_i < 1$, and where w_i are independently distributed standard normal random variables. If $aS_1\Sigma$ were the identity matrix, the λ_i would all be one and (17) would have a chi square distribution with k degrees of freedom; as it is, its distribution is more concentrated towards zero than a chi square statistic so that a chi-square test based on Δ_1 is asymptotically biased toward accepting the null hypothesis.

A Split-Sample MTT test. Suppose the sample is first divided randomly in two asymptotically equal parts, denoted here by superscripts A and B. One parameter estimate $\hat{\beta}_0^A \equiv (\hat{\gamma}_0^A, \hat{\theta}_0^A)$ is obtained on the N^A members of subsample A by maximizing $L^A(\beta) \equiv \sum_{t \varepsilon A} \sum_{j \varepsilon C} S_{tj} \log P_{tj}(\beta)$. Another estimate $\hat{\theta}_1^B$ is obtained on the N_1^B members of subsample B whose choices belong to the restricted choice set D, by maximizing $L_1^B(\theta) \equiv \sum_{t \varepsilon B} \sum_{j \varepsilon D} S_{tj} \log P_{tj|D}(\theta)$. Let H^A and H_1^B be the corresponding negative Hessians. It is clear from the independence of $\hat{\theta}_0^A$ and $\hat{\theta}_1^B$ that $\hat{q}^{AB} \equiv \hat{\theta}_1^B - \hat{\theta}_0^A$ has asymptotic variance-covariance matrix

$$(18) \quad \Sigma^{AB} = (1/a)S_1^{-1} + S_0^{-1} \ .$$

---

[4] This applies result (vii) of Rao (1973), p. 188, to the case where both Σ and aS_1 are nonsingular.

The split-sample MTT test statistic is

(19) $\Delta_1^{AB} \equiv -2[L_1^B(\hat{\theta}_0^A) - L_1^B(\hat{\theta}_1^B)]$.

Taking the Taylor Series expansion of $L_1^B(\hat{\theta}_0^A)$ around $\hat{\theta}_1^B$ yields the approximation

(20) $\Delta_1^{AB} \tilde{=} \hat{q}^{AB'}H_1^B(\hat{\theta}_1^B)\hat{q}^{AB}$.

An argument exactly analogous to the MTT case shows that Δ_1^{AB} cannot be asymptotically chi square unless the second term in the right-hand side of (18) is negligible compared to the first; and that

(21) $\text{plim} \ \Delta_1^{AB} = \sum_{i=1}^{k} \lambda_i^{AB} (w_i^{AB})^2$

where w_i^{AB} are independent standard normal variates and λ_i^{AB}, the characteristic roots of $aS_1\Sigma^{AB}$, satisfy $\lambda_i^{AB} > 1$. Hence the asymptotic distribution of Δ_1^{AB} is <u>less</u> concentrated towards zero than a chi square statistic, so that a chi-square test based on Δ_1^{AB} is asymptotically biased toward <u>rejecting</u> the null hypothesis. This is the exact opposite of the MTT case.

<u>Discussion.</u> It might appear that we could combine the two split samples and form a Chow-type test statistic[5]

---

[5] In fact, this test was first proposed by MTT in an earlier version of their 1977 paper, but was deleted from the published version. It has been used by Horowitz (1980, 1981) to whom we are grateful for bringing it to our attention.

(22) $\qquad 2[L_1^B(\hat{\theta}_1^B) + L^A(\hat{\beta}_o^A) - L^C(\hat{\beta}_o^C)]$,

where $L^C = L^A + L_1^B$. Using the same argument as that of the appendix, one can show that the asymptotic distribution of (22) is very complicated.[6] Moreover, the direction of this Chow-type test statistic relative to the chi-square distribution can not be worked out a priori. Hence, we are reluctant to recommend its use.

The problem with the MTT and split-sample MTT tests is that they treat $\hat{\theta}_0$ in (15) or $\hat{\theta}_0^B$ in (19) as though they were nonstochastic. Some intuition can be gained by examining what this means for the variance of the i-th component of $\hat{q}$ (or $\hat{q}^{AB}$, which is analogous):

(23) $\quad var(\hat{q}_i) = var(\hat{\theta}_{1i}) + var(\hat{\theta}_{0i}) - 2 \, cov(\hat{\theta}_{1i}, \hat{\theta}_{0i})$.

Ignoring the stochastic nature of $\hat{\theta}_{0i}$ amounts to

---

[6] The reason the classical derivation of an asymptotic chi square distribution for such a statistic does not go through in this case is that there is no sensible alternative (i.e., maintained) hypothesis which would yield the likelihood function $L^A(\beta_o) + L_1^B(\theta_1)$ with $\beta_o \equiv (\theta_o, \gamma_o)$ and $\theta_o \neq \theta_1$.

neglecting the last two terms on the right. The first of these adds to the variance of $\hat{q}_i$. The second is zero for the split-sample case, and exactly -2 times the first for the MTT case. Hence ignoring the last two terms underestimates $var(\hat{q}_i^{AB})$ in the **split-sample case and overestimates** $var(\hat{q}_i)$ **in the** MTT case. Hence the **split sample test tends to regard an observed difference** between parameter estimates as significant when it is not, and vice versa for the MTT test. Happily there is an alternative estimate of θ_{0i} for which these extra two terms exactly cancel, forming the basis for the test proposed in Section III.

<u>Correction Factors</u>. We might ask under what circumstances one could eliminate the asymptotic biases just demonstrated through some scalar correction factor. The answer is that whenever the characteristic roots of $aS_1\Sigma$ are all identical, say λ , the "corrected MTT statistic" $(1/\lambda)\Delta_1$ is asymptotically chi square with k degrees of freedom; as is the "corrected **split** **sample MTT statistic"** $(1/\lambda^{AB})\Delta_1^{AB}$ **if the characteristic roots of** $aS_1\Sigma^{AB}$ **are** all equal to λ^{AB} . But the first condition simply means that $(1/\lambda)aS_1\Sigma$ is the kxk identity matrix, or

$$\Sigma = (\lambda/a)S_1^{-1} .$$

Comparing with (13), we see this is possible if $S_1 = \alpha S_0$ in which case $\lambda = 1 - \alpha a$ Similarly, (18) shows that under precisely the same condition the roots of $aS_1\Sigma^{AB}$ are all equal to $\lambda^{AB} \equiv 1 + \alpha a$. Thus under the very restrictive condition $S_1 = \alpha S_0$, the following corrected test statistics have an asymptotic chi square distribution with k degrees of freedom:

$$(24) \quad \tilde{\Delta}_1 \equiv (\frac{1}{1-\alpha a}) \Delta_1$$

$$(25) \quad \tilde{\Delta}_1^{AB} \equiv (\frac{1}{1+\alpha a})\Delta_1^{AB} \quad .$$

Note that when $S_1 = \alpha S_0$ the positive definiteness of (13) and (18) guarantees that $-1 < \alpha a < 1$. Hence (24) and (25) are positive, and correct the MTT statistic upward and the **split-sample MTT statistic downward, as required.**

While it is unlikely that S_1 and S will ever differ by exactly a scalar multiple, except in the case of one independent variable, there may be reason to expect them to be approximately equal. In particular, suppose all components of β are identified on the restricted choice set so that $\beta = \theta$, or more generally that the variables y and x having unidentified and identified coefficients, respectively, are orthogonal. Then S_1 and S_0 are the limits of the average term in the corresponding negative Hessian H_1 or H, respectively. As noted earlier, each such term is the covariance of the vector of characteristics of the chosen alternative for a given sample member. If the dispersion of such characteristics does not greatly differ on sets C and D, then use of the statistics (24) or (25) with $\alpha = 1$ might be justified as a rough screening procedure. Most logit estimation programs can be asked to print the Hessian or its inverse, so that the reasonableness of the assumption $S_1 \overset{\sim}{=} S_0$ can at least be checked.

The way $\underline{a}$, the ratio of sample sizes, enters these correction factors is fairly intuitive. Large $\underline{a}$ means high correlation between $\hat{\theta}_1$ and $\hat{\theta}_0$, which biases the MTT test statistic toward zero, thus requiring a large correction factor in (24); whereas small $\underline{a}$ means a small sample size N_1^B and hence a large variance in $\hat{\theta}_1^B$, which biases the **split-sample test statistic (19) upward, thus requiring a large correction factor in (25).** The role of α is less obvious in the MTT case, but in the **split-sample case a small value of α again increases the variance** in $\hat{\theta}_1^B$ because it means little variation of the independent variables within the restricted choice set.

III. AN ASYMPTOTICALLY UNBIASED LIKELIHOOD RATIO TEST

Each likelihood ratio test discussed thus far employs two estimates of
the parameter vector θ : one using the complete choice set, the   other using
a restricted choice set. The likelihood function maximized by the latter is
then used to compare the two. We have seen that the resulting test statistic
is asymptotically biased downward when both estimates begin from a single sample,
and upward when they are based on independent subsamples. It turns out that
there   is   a way to combine the two approaches so as to eliminate any
asymptotic bias. The solution is to do the full-choice-set estimate twice,
once à la MTT and once à la **split-sample MTT, and take a weighted average.**

We begin just as for the **split-sample MTT test.** Divide the sample randomly into
two parts A and B of (asymptotically equal) sizes N^A and N^B. Compute estimates
$\hat{\beta}_0^A = (\hat{\gamma}_0^A, \hat{\theta}_0^A)$ and $\hat{\theta}_1^B$ as in the **split-sample test. In addition,**

compute the estimate $\hat{\beta}_0^B \equiv (\hat{\gamma}_0^B, \hat{\theta}_0^B)$ which maximizes

$$(26) \quad L^B(\beta) = \sum_{t \varepsilon B} \sum_{j \varepsilon C} S_{tj} \log P_{tj}(\beta) \; .$$

This is the estimate that would be used in the MTT test if B were the full sample. Define

$$(27) \quad \chi(\theta) = - 2[L_1^B(\theta) - L_1^B(\hat{\theta}_1^B)] \; ,$$

the likelihood ratio statistic for a test which compares θ to the value which maximizes L_1^B. The **split-sample test statistic is** $\chi(\hat{\theta}_0^A)$, whereas the MTT test statistic using B as the full sample is $\chi(\hat{\theta}_0^B)$. We know the former is less concentrated and the latter more concentrated than the chi square distribution with k degrees of freedom. Form the following average:

$$(28) \quad \hat{\theta}_0^{AB} = (1/\sqrt{2})\hat{\theta}_0^A + (1 - 1/\sqrt{2})\hat{\theta}_0^B \; .$$

We now prove that

$$(29) \quad \Delta \equiv \chi(\hat{\theta}_0^{AB}) \equiv - 2 \, [L_1^B(\hat{\theta}_0^{AB}) - L_1^B(\hat{\theta}_1^B)]$$

is asymptotically distributed chi square with k degrees of freedom.

To see this, let $b = 1 - 1/\sqrt{2}$, so that $\hat{\theta}_0^{AB} = (1-b)\hat{\theta}_0^A + b\hat{\theta}_0^B$. A Taylor Series expansion of L_1^B around $\hat{\theta}_1^B$ yields

$$(30) \quad \Delta \overset{\sim}{=} \hat{q}^{ABB'} H_1^B(\hat{\theta}_1^B)\hat{q}^{ABB}$$

where $\hat{q}^{ABB} \equiv \hat{\theta}_1^B - \hat{\theta}_0^{AB}$ and H_1^B is the Hessian of L_1^B.

Because of the random division of the sample, the following probability limits under the null hypothesis are the same as the corresponding limits for the full sample:[7]

$$(31) \quad \text{plim } N^B \text{Var}(\hat{\theta}_0^B) = \text{plim } N^B \text{Var}(\hat{\theta}_0^A) = S_0^{-1}$$

$$(32) \quad \text{plim } N^B \text{Var}(\hat{\theta}_1^B) = [\text{plim } H_1^B/N^B]^{-1} = (1/a)S_1^{-1}.$$

One can also show, by writing out equation (12) for the subsample B, that the covariance of $\hat{\theta}_1^B$ and $\hat{\theta}_0^B$ satisfies

$$(33) \quad \text{plim } N^B \text{Cov}(\hat{\theta}_1^B, \hat{\theta}_0^B) = S_0^{-1}.$$

Furthermore, $\sqrt{N^B}\hat{q}^{ABB}$ is asymptotically normal with variance-covariance matrix

$$(34) \quad \Sigma^{ABB} \equiv \text{plim } N^B \text{Var}[\hat{\theta}_1^B - b\hat{\theta}_0^B - (1-b)\hat{\theta}_0^A]$$

$$= \text{plim } N^B \text{Var}(\hat{\theta}_1^B - b\hat{\theta}_0^B) + (1-b)^2 \text{plim } N^B \text{Var}(\hat{\theta}_0^A)$$

since $\hat{\theta}_0^A$ is uncorrelated with either of the estimates formed using B. The last term is $(1-b)^2 S_0^{-1}$ from (31). The first term is

$$\text{plim } N^B \text{Var}(\hat{\theta}_1^B) + b^2 \cdot \text{plim } N^B \text{Var}(\hat{\theta}_0^B) - 2b \cdot \text{plim } N^B \text{Cov}(\hat{\theta}_1^B, \hat{\theta}_0^B)$$

$$= (1/a)S_1^{-1} + (b^2 - 2b)S_0^{-1}.$$

Combining, the terms in S_0^{-1} all cancel, since $(1-b)^2 + (b^2-2b) = 0$, leaving

$$(37) \quad \Sigma^{ABB} = (1/a)S_1^{-1}.$$

---

[7] More generally, if $\text{plim }(N^B/N^A) = h$, then equation (31) should be written as $\text{plim } N^B \text{Var}(\hat{\theta}_0^A) = aS_0^{-1}$, and b should be modified as $1-(1"h)^{-1/2}$. Thus, (28) becomes $(1+h)^{-1/2}\hat{\theta}_0^A + (1 - (1+h)^{-1/2})\hat{\theta}_0^B$.

Thus the asymptotic variance of $\sqrt{N^B}\hat{q}^{ABB}$ is just the inverse of plim H_2^B/N^B, which means the quadratic form (30) has the claimed distribution.[8]

IV. EMPIRICAL EXAMPLE

This section presents calculations of the various test statistics discussed earlier, from a model and data set previously reported by Small (1982). It illustrates the relationships between the alternate tests, and highlights the importance of applying some correction factor before using the MTT test. It also serves as a correction and extension of the results of the uncorrected MTT test reported by Small.

The model is a behavioral explanation of trip timing, i.e. the choice of time of day at which work trips are taken. Each alternative is a five-minute arrival interval, arranged so that for each commuter, alternative number 9 is the interval centered at the employer's official starting time, alternatives 10 through 12 involve arrival later than this official time, and alternatives 1 through 8 earlier. Workers who report no official work-start time on the part of their employers are excluded from the sample, as are those who choose to arrive outside the one-hour range (from 42.5 minutes early to 17.5 minutes late) covered by the twelve alternatives assumed available. The specification used here is "Model 4" of Small (1982), except that the sample size has been extended to 527 by reconstructing some previously missing data indicating whether or not the trip utilized a car pool. The 9 independent variables include travel time, several measures of the extent to which a given

---

[8] Alternatively, using an argument parallel to that of Appendix A, (35) implies that the roots λ_i^{ABB} of $aS_1 \Sigma^{ABB}$ are all equal to 1. Hence the analogue of (A2) states that plim Δ_1^{ABB} is the sum of k independent squared normal variates.

arrival time differs from the official starting time, and two variables

designed to capture tendency to round answers to multiples of 10 or 15

minutes. Several of these variables are constant on one or more of the

subsets D investigated, hence k, the degrees of freedom of the chi square

statistic, is sometimes less than 9.

Small discusses two reasons to doubt the validity of the logit speci-

fication. The first is the possibility that the alternatives are viewed

as grouped: Alternatives 1-8 (early), 9 (on-time), and 10-12 (late) might be

considered qualitatively distinct, thereby leading to an error structure

suitable for a nested logit model (McFadden, 1978). The second

is that the alternatives are ordered in such a way that com-

muters probably view pairs of alternatives whose labels are close to each

other (e.g. 2 and 3) as closer substitutes than other pairs (e.g. 2 and 8).

This leads to the "ordered logit" model developed by Small (1981). The

first departure can be tested by letting D be some combination of the likely

groupings; we have somewhat arbitrarily added the grouping {5-12}. The second

departure might be detected by letting D consist of alternatives which are separated
on the ordering by
/one or more alternatives (e.g. D could be all the even-numbered alternatives).

Small (1982) tested the first of these possibilities, and was surprised

to find only weak evidence of misspecification based on the (uncorrected)

MTT test. The reason is now apparent: As shown in Table 1, all the correct

or corrected test statistics exceed the uncorrected MTT, for six of the

seven choice subsets D tried.[9] Based on the values shown in the table,

a reasonable case could be made for rejecting the logit specification in all

four of the nested-structure tests and in two of the three ordered-structure

tests. The nested-structure results are corroborated by Small and Brownstone's

[1982] finding that logit is rejected in favor of nested logit

---

[9]These seven tests are not independent, and ideally we would like a rigorous
analysis of their joint properties. We have not found a way to do this, so can
only reiterate the usual warning against over-interpreting "significant" results
when many formulations are tried.

Table 1
Test of the Multinomial Logit Specification

	Nested-Structure Tests				Ordered-Structure Tests		
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Choice subset (D)	{1-8, 10-12}	{1-9}	{9-12}	{5-12}	{even}	{odd}	{3,6,9}
No. alternatives identified (k)	9	7	7	9	8	8	6
Ratio of Sample Sizes (N_1/N)	0.65	0.96	0.40	0.83	0.35	0.65	0.62
Hausman-McFadden Test Statistic:							
Standard[d]	8.3[g]	11.0	31.7**	9.5	18.0**	15.0	9.1
Deg.-Freedom Adjusted	5.8[g]	10.9	30.4**	9.4	20.9**	14.6	8.7
Guaranteed Pos. Def.[f]	1583.**	11.1	1436.**	31.5**	37.1**	9.2	5.8
MTT Test Statistic:							
Uncorrected[a]	5.6	0.7	14.5**	6.0	13.6	4.6	3.2
Corrected[b]	15.8	16.2**	24.1**	34.2**	20.9**	13.2	8.5
Split-Sample MTT Test Statistic:							
Uncorrected	15.6	20.7**	19.1**[h]	12.1	12.7	20.3**	12.9**
Corrected	9.5	10.4	13.5[h]	6.5	9.5	12.1	8.0
Small-Hsiao Test Statistic:	8.1	11.1	17.5**[h]	7.5	9.7	13.2	6.9
Repeat on Interchanged Subsamples A,B:							
Split-Sample Uncorrected	24.4**	16.5**	18.3**	12.5	20.1**	18.2**	17.2**
Split-Sample Corrected	14.8	8.5	13.3	7.0	14.8	11.2	10.6
Small-Hsiao	16.4	8.9	16.3**	8.3	17.5**	11.3	11.2

[a]Test statistic is $\Delta_1 \equiv 2[L_1(\hat{\theta}_1) - L_1(\hat{\theta}_0)]$.

[b]Test statistic is $[N/(N-N_1)]\Delta_1$.

[c]Test statistic is $\hat{q}'(\hat{\Sigma}/N)^{-1}\hat{q}$, with $\hat{q} = \hat{\theta}_1 - \hat{\theta}_0$ and $\hat{\Sigma}/N$ as below.

[d]Using $\hat{\Sigma}/N = H_1^{-1}(\hat{\theta}_1) - H_0^{-1}(\hat{\theta}_0)$

[e]Using $\hat{\Sigma}/N = H_1^{-1}(\hat{\theta}_1)N_1/(N_1-k) - H_0^{-1}(\hat{\theta}_0)N/N-k$.

[f]Using $\hat{\Sigma}/N = [E_{\hat{\theta}_0} H_1(\hat{\theta}_0)]^{-1} - H_0^{-1}(\hat{\theta}_0)$.

[g]$\hat{\Sigma}$ not positive definite.

[h]Estimation of θ_1 failed because of flat spot on likelihood function; however $L_1(\hat{\theta}_1)$ appears to be well defined.

**indicates that test statistic exceeds the 5% critical level of the appropriate chi-square distribution.

(though coefficient estimates of the variables of interest are only moderately affected).

Unfortunately, the results shown still leave room for doubt in five of the seven cases, due to considerable variation among even the asymptotically equivalent test statistics. The main problem is that the Hausman test statistic, as mentioned earlier, is sensitive to which of several consistent estimates $\hat{\Sigma}/N$ is used. The "standard" and the "guaranteed positive definite" estimates are those recommended in Hausman and McFadden (1981), the latter designed to guarantee that $\hat{\Sigma}$ be positive definite. In contrast to Hausman and McFadden, we find the latter estimate to give sometimes ridiculous results when $\hat{\Sigma}$ is nearly singular. Thus in four of the seven cases, these alternative estimates of the chi-square statistic (with 6 to 9 degrees of freedom) differed by 19 or more. In two cases the differences are more than 1000; applying the asymptotic theory to the results in column (a) leads to a significance level varying from 75% to less than 0.0001%, depending on which version is chosen. Also shown is the "degrees-of-freedom-corrected" estimate described earlier, using the finite-sample correction normally applied by QUAIL (the computer program used here) in its estimates of covariance matrices. In most cases this is close to the standard estimate. Note that the rough corrections to the MTT and split-sample tests give results reasonably in line with the two asymptotically unbiased tests, except in the MTT case when sample sizes N_1 and N are nearly the same (columns b and d).

Columns (a) and (c) are the two cases where the Hausman-McFadden statistic turned out to be extremely unstable. When the tests shown in column (a) were repeated on the smaller sample (N=453) used in Small (1982),

the three versions of the Hausman-McFadden test statistic were -11.9,
-21.7 and 1364.5; whereas the corrected MTT, corrected split-sample MTT, and Small-
Hsiao statistics were 11.5, 13.4, and 10.5, respectively. Computation of the
eigenvalues confirmed that the matrix $\hat{\Sigma}$ used in the standard Hausman-McFadden
test was not positive definite; the same is true for the sample (N=527)
shown in column (a), even though the test statistic happened to come out
positive.

There is a different and less severe kind of instability in the Small-
Hsiao test. As described in the previous section, our test randomly divides
the sample into two parts A and B. It is just as natural to reverse the
roles of A and B: i.e., to compute estimate $\hat{\theta}_1^A$ by maximizing the likelihood
function $L_1^A(\theta)$ of the N_1^A members of subsample A whose choices belong
to the restricted choice set D , and form the statistic

(37) $\chi^A(\hat{\theta}_0^{BA}) \equiv -2 [L_1^A(\hat{\theta}_0^{BA}) - L_1^A(\hat{\theta}_1^A)]$,

where $\hat{\theta}_0^{BA} = (1/\sqrt{2}) \hat{\theta}_0^B + (1 - 1/\sqrt{2})\hat{\theta}_0^A$. The results of the interchanged
Small-Hsiao test are reported at the bottom line of Table 1. The difference
in observed significance level between the original and interchanged tests
can pose a decision problem: Had we chosen a 5% significance level as the
size of each of these two tests, then in column (e) we would have accepted
H_0 in one case and rejected it in the other. However, the instability in
the Small-Hsiao test is of a different nature from that of the Hausman-McFadden
test. In the latter, instability is caused by different ways of computing
the asymptotically equivalent variance covariance matrix, which poses
particularly serious problems if the estimated variance covariance matrix is
near singular or not positive definite. In our test the instability is due
to sampling variability in drawing subsamples, which we suspect would disappear

more quickly with larger overall sample size. Furthermore, the possibility

of conflicting inferences in our test can be eliminated if one is willing to

accept a test whose asymptotic size is unknown but lies within a known interval.

For example, a test whose size lies in the interval $[0,\alpha]$ is defined by

the decision rule: reject H_0 if both Δ^{AB} and Δ^{BA} exceed $c(\alpha,k)$,

the critical value of the chi-square distribution corresponding to significance

level α . Alternatively, a test whose size lies within the narrower interval

$[\alpha/2 , \alpha]$ is defined by the decision rule: reject H_0 if either

Δ^{AB} or Δ^{BA} exceeds $c(\alpha/2,k)$. Since the first decision rule leads to an

interval which can always be narrowed by using the second decision rule, the

latter is preferred. We recommend it as the most natural way to take

advantage of the availability at little cost of both test statistics.

IV. CONCLUSION

We have shown that the original MTT and split-sample MTT likelihood ratio

tests are asymptotically biased toward accepting and toward rejecting,

respectively, the null hypothesis of multinomial logit.   We provide

correction factors which, under the very restrictive condition that the asymptotic

moment matrices of the independent variables on different choice sets be equal,

removes this bias. The factors are $1/(1-a)$ and $1/(1+a)$, respectively, where

a is the fraction of the full sample included in the subsample estimation.

We also provide a general test with no asymptotic bias which, like that recently

proposed by Hausman and McFadden, utilizes a test statistic which is

asymptotically chi-square distributed.

An empirical application demonstrates that in most cases the corrected MTT and split-sample tests give results much closer to each other and to the other tests than do the uncorrected versions. The application also illustrates the finite-sample variability that sometimes occurs in the Hausman-McFadden test due to inversion of a matrix which is nearly singular and/or not positive definite. The test we propose avoids this severe fluctuation when asymptotically equivalent versions are computed. Our test requires no matrix manipulation or inversion, and can be

computed using repeated applications of any logit estimation routine with the ability to compute the likelihood function at an arbitrary parameter vector. Hence, we suspect many users will find our test the easiest to use for routine applications.

REFERENCES

Abkowitz, M. D. (1981). An Analysis of the Commuter Departure Time Decision. Transpn. 10, 283-297.

Amemiya, T. (1981). Qualitative Response Models: A Survey. J. Econ. Lit. 19, 1483-1536.

Domencich, T. A. and McFadden, D. (1975). Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam.

Hausman, J. A. (1978). Specification Tests in Econometrics. Econometrica 46, 1251-1271.

Hausman, J. A. and McFadden, D. (1981). Specification Tests for the Multinomial Logit Model. Working Paper, Dept. of Econ., M.I.T., Cambridge, Mass.

Horowitz, J. (1980). A Utility Maximizing Model of the Demand for Multi-Destination Non-Work Travel. Transpn. Res. 14B, 369-386.

Horowitz, J. (1981). Identification and Diagnosis of Specification Errors in the Multinomial Logit Model. Transpn. Res. 15B, 345-360.

Lave, C. A. and Train, K. (1979). A Disagregate Model of Auto-Type Choice. Transpn. Res. 13A, 1-9.

Manski, C. F. (1981). Structural Models for Discrete Data: The Analysis of Discrete Choice. Sociological Methodology 1981, 58-109.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. Frontiers in Econometrics (Edited by P. Zarembka), 105-142. Academic Press, New York.

McFadden, D. (1978). Modelling the Choice of Residential Location. Spatial Interaction Theory and Planning Models (Edited by A. Karlqvist et al.), 75-96. North-Holland, Amsterdam.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. Structural Analysis of Discrete Data with Econometric Applications (Edited by C. F. Manski and D. McFadden), 198-272. MIT Press, Cambridge, Mass.

McFadden, D., Tye, W. and Train, K. (1977). An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. Transpn. Res. Rec. 637, 39-46.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. Wiley, New York.

Small, K. A. (1981). Ordered Logit: A Discrete Choice Model with Proximate Covariance Among Alternatives. Research Memorandum No. 292, Econometric Research Program, Princeton Univ., Princeton, NJ.

Small, K.A. and Brownstone, D. (1982). Efficient Estimation of Nested Logit
    Models: An Application to Trip Timing. Research Memorandum No. 296,
    Econometric Research Program, Princeton Univ., Princeton, NJ.

Small, K. A. (1982). The Scheduling of Consumer Activities: Work Trips.
    Amer. Econ. Rev. 72, 467-479.

APPENDIX

A. Proof of Equation (17):

We are given that $\sqrt{N}\hat{q}$ is asymptotically normally distributed with variance-covariance matrix Σ; and that $[(1/a)S_1^{-1} - \Sigma]$ is positive definite (p.d.), where S_1 is given by (10).

Since Σ is symmetric and positive definite, it can be decomposed into a product of a kxk nonsingular matrix B^{-1} and its transpose:

(A1) $\quad \Sigma = B^{-1}(B')^{-1}$

Hence $v \equiv B\sqrt{N}\hat{q}$ is a vector of independent normal variates, since $Var(v) = B\Sigma B' = I$.

Because S_1 is p.d., so is $A \equiv (B')^{-1}(aS_1)B^{-1}$. Hence there exists a kxk orthogonal matrix F which diagonalizes A:

$$FAF' = \Lambda = diag(\lambda_1,\ldots,\lambda_k)$$

where $\lambda_i > 0$ are the characteristic roots of A.

By using (10) and inserting identity matrices at various places, the probability limit of (16) can be written as

$$plim\tilde{\Delta}_1 = plim\sqrt{N}\hat{q}'\, B'(B')^{-1}(aS_1)B^{-1}B\sqrt{N}\hat{q}$$

$$= v'Av$$

$$= v'F'FAF'Fv$$

$$= w'\Lambda w$$

(A2) $$\qquad\qquad = \sum_{i=1}^{k} \lambda_i w_i^2$$

where $w \equiv Fv$ is a vector of independent standard normal variates (independent because $Var(w) = F'IF = F'F = I$).

Since λ_i are the characteristic roots of A, they are the roots of the k-th degree polynomial equation

$$0 = |A - \lambda I|$$

$$(A3) \qquad = |B'A(B')^{-1} - \lambda B'(B')^{-1}|$$

$$(A4) \qquad = |aS_1\Sigma - \lambda I|$$

$$(A5) \qquad = |\Sigma - \lambda(1/a)S_1^{-1}|$$

where (A3) and (A5) result from pre- and/or post-multiplying the matrix inside the determinant by some square matrix whose determinant is nonzero. (A4) shows that λ_i are the characteristic roots of $aS_1\Sigma$. Because $(1/a)S_1^{-1} - \Sigma$ is p.d., (A5) implies that $\lambda_i < 1$.

B. Proof of Equation (21)

The proof of (21) is exactly analogous to that of (17), with appropriate superscripts A and B; except that $(1/a)S_1^{-1} - \Sigma^{AB}$ is negative definite because of (18), so that the characteristic roots λ_1^{AB} of $aS_1\Sigma^{AB}$ are greater than 1.