Econometrica, Vol. 52, No. 5 (September, 1984)

SPECIFICATION TESTS FOR THE MULTINOMIAL LOGIT MODEL[1]

BY JERRY HAUSMAN AND DANIEL MCFADDEN

Discrete choice models are now used in a variety of situations in applied econometrics. By far the model specification which is used most often is the multinomial logit model. Yet it is widely known that a potentially important drawback of the multinomial logit model is the independence from irrelevant alternatives property. While most analysts recognize the implications of the independence of irrelevant alternatives property, it has remained basically a maintained assumption in applications.

In the paper we provide two sets of computationally convenient specification tests for the multinomial logit model. The first test is an application of the Hausman [**10**] specification test procedure. The basic idea for the test here is to test the reverse implication of the independence from irrelevant alternatives property. The test statistic is easy to compute since it only requires computation of a quadratic form which involves the difference of the parameter estimates and the differences of the estimated covariance matrices.

The second set of specification tests that we propose is based on more classical test procedures. We consider a generalization of the multinomial logit model which is called the nested logit model. Since the multinomial logit model is a special case of the more general model when a given parameter equals one, classical test procedures such as the Wald, likelihood ratio, and Lagrange multiplier tests can be used.

The two sets of specification test procedures are then compared for an example where exact and approximate comparisons are possible.

DISCRETE CHOICE MODELS are now used in a wide variety of situations in applied econometrics.[2] By far the model specification which is used most often is the multinomial logit model (McFadden [**18**]). The multinomial logit model provides a convenient closed form for the underlying choice probabilities without any requirement of multivariate integration. Therefore, choice situations characterized by many alternatives can be treated in a computationally convenient manner. Furthermore, the likelihood function for the multinomial logit specification is globally concave which also eases the computational burden. The ease of computation and the existence of a number of computer programs has led to the many applications of the logit model. Yet it is widely known that a potentially important drawback of the multinomial logit model is the independence from irrelevant alternatives property. This property states that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set.[3] Debreu [**6**] was among the first economists to discuss the implausibility of the independence from irrelevant alternatives as-

[1] J. Dubin, W. Newey, and J. Rust provided research assistance. The NSF and DOE provided research support. G. Chamberlain, T. Rothenberg, A. Zellner, the referees, and an editor made helpful comments.

[2] McFadden [**20**] provides references to many of their uses.

[3] A "universal" logit model avoids the independence from irrelevant alternatives property while maintaining the multinomial logit form by making each ratio of probabilities a function of attributes of *all* alternatives (McFadden [**20**]). It is difficult however to give an economic interpretation of this model other than as a flexible approximation to a general functional form.

sumption. Basically, no provision is made for different degrees of substitutability or complementarity among the choices. While most analysts recognize the implications of the independence of irrelevant alternatives property, it has remained basically a maintained assumption in applications.

The multinomial probit model does provide an alternative specification for discrete choice models without any need for the independence of irrelevant alternatives assumption (Hausman and Wise [14]). Furthermore, a test of the 'covariance' probit specification versus the 'independent' probit specification which is very similar to the logit specification does provide a test for the independence from irrelevant alternatives assumption. But use of the multinomial probit model has been limited due to the requirement that multivariate normal integrals must be evaluated to estimate the unknown parameters. Thus, the multinomial probit model does not provide a convenient specification test for the multinomial logit model because of its complexity.

In this paper we provide two sets of computationally convenient specification tests for the multinomial logit model. The first test is an application of the Hausman [10] specification test procedure. The basic idea for the test here is to test the reverse implication of the independence from irrelevant alternatives property. The usual implication is to note that if two choices exist, say car and bus in a transportation choice application, then addition of a third choice, subway, will not change the ratio of probabilities of the initial two choices. Our test here is based on eliminating one or more alternatives from the choice set to see if underlying choice behavior from the restricted choice set obeys the independence from irrelevant alternatives property. We estimate the unknown parameters from both the unrestricted and restricted choice sets. If the parameter estimates are here approximately the same, then we do not reject the multinomial logit specification. The test statistic is easy to compute since it only requires computation of a quadratic form which involves the difference of the parameter estimates and the differences of the estimated covariance matrices. Thus, existing logit computer programs provide all the necessary input to the test.

The second set of specification tests that we propose is based on more classical test procedures. We consider a generalization of the multinomial logit model which is called the nested logit model (McFadden [20]). Since the multinomial logit model is a special case of the more general model when a given parameter equals one, classical test procedures such as the Wald, likelihood ratio, and Lagrange multiplier tests can be used. Of course, we have added the requirement of the specification of an alternative model to test the original model specification. Maximum likelihood estimation of the nested logit model is considerably more difficult than for the multinomial logit model. However, a Wald type test can be constructed on the basis of a parameter estimated from a consistent, but inefficient, sequential logit estimation which uses standard computer packages. Alternatively, a Lagrange multiplier test can be computed from the multinomial logit estimates.

We then proceed to compare the two sets of specification test procedures for an example. We find rather unexpected results. First despite a sample size of

1000, the asymptotically equivalent classical tests differ markedly in their operating characteristics. The power of the Wald test is significantly greater than the other two classical tests, the LR test and the LM test. Perhaps, more surprising, we find the power of one Hausman test to be comparable to that of the Wald test. Thus the often quoted asymptotic power results for local departures from the null hypothesis do not provide a reliable guide to the exact performance of our specification tests in the example, despite the relatively large sample size and small departures from the multinomial logit model and despite the fact that all the tests are identical to first order, i.e. they have identical noncentrality parameters for the asymptotic noncentral χ^2 distribution. We then do a further comparison of the tests when the degrees of freedom differ for a smaller sample. We again find a marked difference in the actual performance of the tests.

The plan of the paper is as follows. In the following section we derive the Hausman-type specification test for the multinomial logit model. The distribution theory as well as computational considerations are discussed. In the following section we derive the classical tests from the nested logit model. In Section 3 we calculate the exact size and power for the two sets of specification tests for an example. We also compare the exact power with the usual asymptotic approximation to the power function. Lastly, in the conclusion we discuss some further considerations for the test procedures.

1. A HAUSMAN-TYPE TEST OF THE IIA PROPERTY

A widely used functional form for discrete probabilities is the multinomial logit (MNL) model,

$$(1.1)\qquad P(i\,|\,z, C, \beta) = e^{z_i\beta} \Big/ \sum_{j\in C} e^{z_j\beta}$$

where $C = \{1, \ldots, J\}$ is a finite choice set; i, j are alternatives in C; z_j is a K-vector of explanatory variables describing the attributes of alternatives j and/or the characteristics of the decision maker which affect the desirability of alternative j; $z = (z_1, \ldots, z_J)$ represents the attributes of C; β is a K-vector of taste parameters; $P(i\,|\,z, C, \beta)$ is the probability that a randomly selected decision maker, when faced with choice set C with attributes z, will choose i.

The MNL model has a necessary and sufficient characterization, termed independence for irrelevant alternatives (IIA), that the ratio of the probabilities of choosing any two alternatives is independent of the attributes or the availability of a third alternative, or

$$(1.2)\qquad P(i\,|\,z, C, \beta) \equiv P(i\,|\,z, A, \beta)P(A\,|\,z, C, \beta)$$

where $i \in A \subseteq C$ and

$$P(A\,|\,z, C, \beta) = \sum_{j\in A} P(j\,|\,z, C, \beta).$$

This property greatly facilitates estimation and forecasting because it implies the model can be estimated from data on binomial choices, or by restricting attention to choice within a limited subset of the full choice set. On the other hand, this property severely restricts the flexibility of the functional form, forcing equal cross-elasticities of the probabilities of choosing various alternatives with respect to an attribute of one alternative. Further discussion of the IIA property and conditions under which it is likely to be true or false is given in Domencich and McFadden [**7**], McFadden, Tye, and Train [**19**], and Hausman–Wise [**14**]. The McFadden, Tye, and Train paper suggests that the MNL specification be tested by comparing parameter estimates obtained from choice data from the full choice set with estimates obtained from conditional choice data from a restricted choice set. Here we develop an asymptotic test statistic for this comparison, using the approach to specification tests introduced by Hausman [**10**].

Consider a random sample with observations $n = 1, \ldots, N$. Let z^n be the attributes of C for case n, and define $S_{in} = 1$ if case n chooses i and $S_{in} = 0$ otherwise. The normalized log likelihood of the sample is

$$(1.4) \qquad L_C(\beta) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \in C} S_{in} \ln P(i \mid z^n, C, \beta).$$

We first review the asymptotic properties of maximum likelihood estimates of β from (1.4). We make the following regularity assumptions:

Assumption A: The vector of attributes z has a distribution μ in the population which has a bounded support.

Assumption B: The MNL specification (1.1) with a parameter vector β^* is the true model.

Assumption C: The parameter vector β^* is asymptotically identified, i.e., if $\beta \neq \beta^*$, there exists a set Z of z values and an alternative i such that

$$\int_Z P(i \mid z, C, \beta^*)\, d\mu(z) \neq \int_Z P(i \mid z, C, \beta)\, d\mu(z).$$

Under these assumptions, $ES_{in} = P(i \mid z^n, C, \beta^*)$, the normalized log likelihood converges uniformly in β to

$$(1.5) \qquad \operatorname*{plim}_{N \to \infty} L_C(\beta) = \int \sum_{i \in C} P(i \mid z, C, \beta^*) \ln P(i \mid z, C, \beta)\, d\mu(z),$$

and (1.5) has a unique maximum at $\beta = \beta^*$. Then the maximum likelihood estimator β_C is consistent, and $\sqrt{N}(\beta_C - \beta^*)$ converges in distribution to a normal random vector with zero mean and covariance matrix $\operatorname{plim}_{n\to\infty}(-\partial^2 L_C(\beta^*)/\partial\beta\,\partial\beta')^{-1}$. Discussion and proofs of these properties can be found in Manski and McFadden [**16**]. See also McFadden [**18**].

Let $A = \{1, \ldots, M\}$ be a subset of the choice set C. Consider the conditional normalized log likelihood of the subsample who make choices from A. If the MNL specification (1.1) is true, then the IIA property states that the probability of choosing i from C, given that choice is contained in A, equals the probability of choosing i from A. The conditional normalized log likelihood is then

$$(1.6)\qquad L_A(\beta) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i\in A} S_{in}\ln(P(i\,|\,z^n, C, \beta)/P(A\,|\,z^n, C, \beta))$$

$$= \frac{1}{N}\sum_{n=1}^{N}\sum_{i\in A} S_{in}\ln P(i\,|\,z^n, A, \beta).$$

Some components of β^*, such as the coefficients of alternative-specific variables for excluded alternatives, are not identified by choice from A. Let $z^n = (y^n, x^n)$ be a partition of the explanatory variables into a vector y^n which only varies outside A and a vector x^n which varies within A, and let $\beta = (\gamma, \theta)$ be a commensurate partition of the parameter vector. The conditional choice probability is then

$$(1.7)\qquad P(i\,|\,x^n, A, \theta) = e^{x_i^n\theta}\Big/\sum_{j\in A} e^{x_j^n\theta}$$

and $y_i^n = y_j^n$ for $i, j \in A$. We add to the regularity assumptions the asymptotic identification condition.

Assumption D: If $\theta \neq \theta^*$, there exists a set Z of z values and an alternative $i \in A$ such that

$$\int_Z P(i\,|\,x, A, \theta^*)\,d\mu(z) \neq \int_Z P(i\,|\,x, A, \theta)\,d\mu(z).$$

Then, as in the unconditional case, the conditional normalized log likelihood converges uniformly in θ to

$$(1.8)\qquad \operatorname*{plim}_{N\to\infty} L_A(\theta) = \int \sum_{i\in A} P(i\,|\,z, C, \beta^*)\ln P(i\,|\,x, A, \beta)\,d\mu(z)$$

$$= \int P(A\,|\,z, C, \beta^*)\sum_{i\in A} P(i\,|\,x, A, \theta^*)\ln P(i\,|\,x, A, \theta)\,d\mu(z)$$

with a unique maximum at $\theta = \theta^*$. The maximum likelihood estimator θ_A is consistent and $\sqrt{N}(\theta_A - \theta^*)$ is asymptotically normal with mean zero and covariance matrix $\operatorname{plim}_{N\to\infty}(-\partial^2 L_A(\theta^*)/\partial\theta\,\partial\theta')^{-1}$.

The specification test statistic is based on the parameter difference $\delta = \theta_A - \theta_C$, where $\beta_C = (\gamma_C, \theta_C)$. When the regularity assumptions hold and the MNL model is true, $\operatorname{plim}_{N\to\infty}\delta = 0$. Conversely, when the MNL specification (1.1) is false, then the IIA property fails, and (1.8) becomes

$$(1.9) \qquad \operatorname{plim} L_A(\theta) = \int P(A \mid z, C, \beta^*) \sum_{i \in A} \left[\frac{P(i \mid z, C, \beta^*)}{P(A \mid z, C, \beta^*)} \right] \times \ln P(i \mid x, A, \theta)\, d\mu(z)$$

with $P(i \mid z, C, \beta^*)/P(A \mid z, C, \beta^*) \neq P(i \mid x, A, \theta^*)$. In general, equation (1.8) is not maximized at $\theta = \theta^*$, implying $\operatorname{plim}_{N\to\infty} \delta \neq 0$. Thus, a test of $\delta = 0$ is a test of the MNL specification. Rejection of $\delta = 0$ indicates a failure of the restrictive structure of the MNL form embodied in the IIA property, or a misspecification of the explantory variables z in (1.1), or both. Acceptance of $\delta = 0$ implies that for the *given* specification of explanatory variables and distribution of these variables, the IIA property holds. Thus the test is consistent against this family of alternatives. However it is not necessarily consistent against all members of the family of alternatives defined by a given specification of explanatory variables and *any* distribution of these variables. To derive an asymptotic test statistic for $\delta = 0$, note that under the regularity assumptions, $\sqrt{N}(\beta_C - \beta^*, \theta_A - \theta^*)$ is asymptotically normal with mean zero and covariance matrix V calculated below; the argument is a standard application of a central limit theorem, as used in Manski and McFadden [17].

To calculate the covariance matrix for the asymptotic test statistic δ, the gradient vectors and Hessian matrices are computed for the normalized unconditional log likelihood function of equation (1.4) and the normalized conditional log likelihood function of equation (1.6). Details of the derivation are given in Hausman–McFadden [12]. We define $V_A = \operatorname{plim} H_A^{-1}$ where H_A is the minus Hessian matrix for the normalized conditional log likelihood function:

$$(1.10) \qquad H_A = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \in A} P(A \mid z^n, C, \beta^*) P(i \mid x^n, A, \theta)(x_i^n - x_A^n)(x_i^n - x_A^n)'$$

with

$$(1.11) \qquad x_A^n = \sum_{i \in A} P(i \mid x^n, A, \theta) x_i^n .$$

Likewise define $V_C = \operatorname{plim} H_C^{-1}$ where H_C is minus the Hessian matrix for the normalized unconditional log LF,

$$(1.12) \qquad H_C = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \in C} P(i \mid z^n, C, \beta)(z_i^n - z_A^n)(z_i^n - z_A^n)',$$

where z_C^n is defined analogously to x_A^n. Partition V_C so that

$$(1.13) \qquad V_C = \operatorname*{plim}_{N\to\infty} H_C^{-1} = \begin{bmatrix} V_{C\gamma\gamma} & V_{C\gamma\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} \end{bmatrix},$$

where the partition is commensurate with (γ, θ). Then

$$(1.14) \qquad V = \begin{bmatrix} V_{C\gamma\gamma} & V_{C\gamma\theta} & V_{C\gamma\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} & V_{C\theta\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} & V_A \end{bmatrix}.$$

Hausman–McFadden [12] demonstrate that the asymptotic covariance matrix of $\sqrt{N}(\theta_A - \theta_C)$ is

$$(1.15) \qquad Q = V_A - V_{C\theta\theta},$$

the difference of the asymptotic covariance matrices of θ_A and θ_C. Thus the test statistic

$$(1.16) \qquad T = N(\theta_A - \theta_C)' Q^t (\theta_A - \theta_C),$$

where Q^t is a generalized inverse of Q, is asymptotically distributed chi-square with degrees of freedom equal to the rank of Q under the null hypothesis. This statistic then coincides with the general specification test statistic developed by Hausman [10] and generalized to use of singular covariance matrices by Hausman–Taylor [13] when an efficient estimator is available under the null hypothesis.

The estimated asymptotic covariance matrices for the maximum likelihood estimators θ_C and θ_A satisfy

$$(1.17) \qquad \operatorname{cov}(\theta_C) = \left[\sum_{n=1}^{N} \sum_{i \in C} P(i \mid z^n, C, \beta_C)(z_i^n - z_C^n)(z_i^n - z_C^n)' \right]^{-1},$$

$$(1.18) \qquad \operatorname*{plim}_{N \to \infty} N \operatorname{cov}(\theta_C) = V_{C\theta\theta}$$

$$(1.19) \qquad \operatorname{cov}(\theta_A) = \left[\sum_{n=1}^{N} S_{An} \sum_{i \in A} P(i \mid x^n, A, \theta_A)(x_i^n - x_A^n)(x_i^n - x_A^n)' \right]^{-1},$$

$$(1.20) \qquad \operatorname*{plim}_{N \to \infty} N \operatorname{cov}(\theta_A) = V_A.$$

Therefore, an asymptotically equivalent computational formula for T is

$$(1.21) \qquad T = (\theta_A - \theta_C)' \left[\operatorname{cov}(\theta_A) - \operatorname{cov}(\theta_C) \right]^t (\theta_A - \theta_C),$$

which should be quite easy to calculate using existing logit programs.

The regularity assumptions do not exclude the possibility that Q is less than full rank. However, deficiencies will occur only for exceptional configurations of the x_i^n variables. In particular, a sufficient condition for Q to be nonsingular is that

$$(1.22)\qquad H^*_{C\theta\theta} - H^*_A$$
$$= \frac{1}{N}\sum_{n=1}^{N}\sum_{i\in D} P(i\,|\,z^n, C, \beta^*)(x_i^n - x_D^n)(x_i^n - x_D^n)'$$
$$+ \frac{1}{N}\sum_{n=1}^{N} P(A\,|\,z^n, C, \beta^*)P(D\,|\,z^n, C, \beta^*)(x_A^n - x_C^n)(x_A^n - x_C^n)'$$

have a nonsingular limit, where $D = C|A$ and x_D^n is defined analogously to (1.11). This is generally the case if the x variables either vary within D, or take on values within D different from their average within A.

The estimated matrix $[\mathrm{cov}(\theta_A) - \mathrm{cov}(\theta_C)]$ may fail to be definite in finite samples even when Q is nonsingular. This does not impede calculation of the statistic (1.21) or carrying out the asymptotic test. However, one can form an asymptotically equivalent estimate of $\mathrm{cov}(\theta_A)$ such that $[\mathrm{cov}(\theta_A) - \mathrm{cov}(\theta_C)]$ is always positive semidefinite by evaluating $P(i\,|\,x^n, A, \theta)$ in (1.19) at θ_C and replacing S_{An} by $P(A\,|\,z^n, C, \beta_C)$.[4]

Ruud **[25]** has provided an interesting interpretation of the test as well as an alternative computation scheme. Ruud's interpretation is based on the notion of an "out of estimation" prediction test of the ability of the parameter estimates θ_A to predict choices on the complete choice set C.[5] Ruud's computational scheme is based on a factorization of the normalized LF into two terms, one of which represents the choice between the restricted choice set and its complement while the second term represents the choice within A. Ruud then generates a class of asymptotically equivalent tests to the test statistic of equation (1.6). However, Ruud's test setup requires a separate computer program since it cannot be based on the ML estimates θ_A and θ_c.

2. AN ALTERNATIVE NESTED LOGIT SPECIFICATION AND CLASSICAL TESTS

The use of the previous specification test requires no specific alternative model.[6] In this section we consider a specific alternative model, the nested logit

[4] We have occasionally found the test statistic of equation (1.21) to be negative due to lack of positive semidefiniteness in finite sample applications. Replacement by the alternative covariance matrix always leads to a small positive number. However, in no case have we found this alternative statistic to be so large as to come close to any reasonable critical value for a χ^2 test. However, Small and Hsiao [27] do report computational difficulties when an estimate of the covariance matrix Q is near singular. They propose an alternative test of the multinomial logit specification.

[5] This interpretation is not quite so neat in the presence of parameters γ_C which are not identified in the restricted choice set. However, one can consider the use of γ_C and θ_A to make predictions.

[6] It is interesting to note that the previous test is *not* equivalent to an ANCOVA-like procedure in which β would be allowed to vary across each alternative or some subset of alternatives after a normalization. The specification test here would involve a test of equality of the β's. But it is straightforward to check that the IIA property still holds under this specification so that the most likely failing of the MNL model would not be tested.

model, on which to base test procedures. Given a parametric alternative hypothesis, we can apply classical test procedures such as the Wald test, likelihood ratio (LR) test, and Lagrange multiplier (LM) test. It is well known that for local deviations from the null hypothesis, these tests have certain optimal large sample power properties; c.f. Silvey [26] and Cox and Hinckley [5]. Of course, the optimum properties of these classical tests depend on three factors which may not be satisfied in a given application: (i) the alternative specification on which these tests are based is correct, (ii) the sample is large enough so that the asymptotic theory provides a good approximation, (iii) deviations from the null hypothesis model are of order $1/\sqrt{N}$. We investigate questions of sufficiently large samples and local deviations in the next section. The nested logit model for the three choice case has the simple hierarchical nature shown in Figure 2.1. Alternatives 1 and 2 are assumed to have more common characteristics than either alternative has with alternative 3. The idea behind the choice process is that the individual forms a weighted average of the attributes of alternatives 1 and 2, sometimes called the inclusive value, which is closely related to his consumer's surplus from these two choices considered above. The inclusive value is defined as

$$(2.1) \qquad y = \log(e^{z_1\beta/\lambda} + e^{z_2\beta/\lambda})$$

where λ is a scalar parameter of the model. The choice probabilities of the model

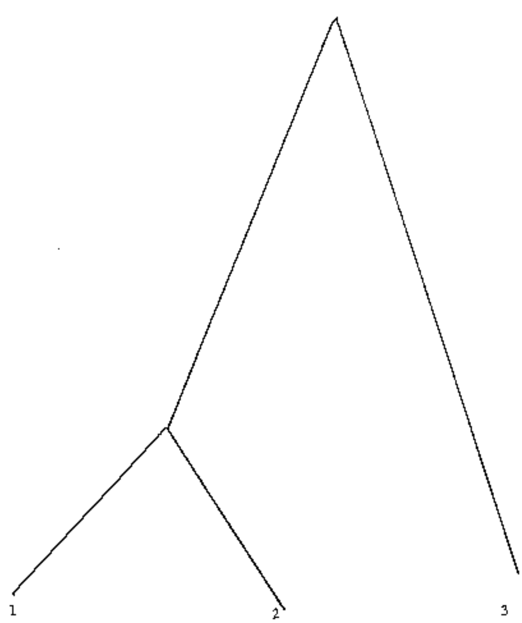

FIGURE 2.1.

are

$$(2.2a)\qquad P(1\mid z, C, \beta, \lambda) = p_1 = \frac{e^{z_1\beta/\lambda}e^{\lambda y}}{e^{y}(e^{z_3\beta}+e^{\lambda y})},$$

$$(2.2b)\qquad P(2\mid z, C, \beta, \lambda) = p_2 = \frac{e^{z_2\beta/\lambda}e^{\lambda y}}{e^{y}(e^{z_3\beta}+e^{\lambda y})},$$

$$(2.2c)\qquad P(3\mid z, C, \beta, \lambda) = p_3 = \frac{e^{z_3\beta}}{e^{z_3\beta}+e^{\lambda y}}.$$

For $\lambda = 1$, the nested logit model reduces to a MNL model. For $0 < \lambda < 1$ the model fails to satisfy the IIA property but it does satisfy the properties required for a random utility model. This proposition is proven together with a discussion of other features of the general model specification in McFadden **[20]**. For λ outside the unit interval the probabilities are still well defined. However, the interpretation of the model as a choice model is not clearcut. One property of the model which deserves mention is that on any subbranch of the tree, the IIA assumption is satisfied. However, in the three choice case, since only a binary comparison involving alternatives 1 and 2 is required, the IIA property does not lead to a testable restriction.

Given the alternative model specification the classical test procedures may be applied by basing tests on the normalized log likelihood function

$$(2.4)\qquad L_N(\beta,\lambda) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i\in C} S_{in}\log P(i\mid z^n, C, \beta, \lambda).$$

The assumptions for the MNL logit model following equation (1.4) are sufficient to prove consistency and asymptotic normality of the estimates $(\beta,\lambda) = \delta$ where we replace β by δ in the assumptions. The parameter λ need not be constrained to the unit interval for these properties to hold. The covariance matrix of the asymptotic normal distribution equals $\operatorname{plim}_{N\to\infty}(-\partial^2 L_N(\delta^*)/\partial\delta\,\partial\delta')^{-1}$. The classical tests proceed so that each test leads to a test of $\lambda = 1$, so that under the null hypothesis the test statistic is asymptotically a central χ^2 random variable with 1 degree of freedom.[7] While we have derived the alternative model and classical tests for a three alternative case, the procedure can be applied in a straightforward way for a general number of choices. McFadden **[21]** gives the formula for the LM test in the general case.

We have now specified an alternative model to the MNL specification which leads to classical tests of the MNL model. We have applied the first specification test and the three classical tests to empirical examples, some of which have led to

[7] Here we take the alternative hypothesis to be $\lambda \neq 1$. One might consider the more restricted alternative that $\lambda < 1$ given the random utility model requirements. Since only 1 parameter is under test, the critical values for the test could be easily adjusted.

strong rejections of the MNL specification.[8] We now turn to a comparison of the various tests in a situation where we can compare the exact performance of the tests.

3. EXACT AND APPROXIMATE COMPARISONS OF THE SPECIFICATION TESTS

We consider a three choice example where the specification tests have closed form solutions. Exact comparisons of the size and power of the tests can then be made. To make the comparisons, we choose a nested logit model as the correct model under the alternative hypothesis. Therefore, the Wald, LR, and LM tests are based on the *correct* alternative specification. Of course, in actual applied work it is important to remember that a particular nested logit model may not provide the correct specification. The first specification test is, of course, not based on a specific alternative model. For our first example we choose the sample size equal to 1000 so that asymptotic theory should provide a reasonable guide. Yet we find two rather unexpected results: (i) The three tests which comprise the so-called holy trinity of asymptotic tests differ markedly in both of their operating characteristics even when their nominal size is the same. (ii) The Wald test and Hausman type test have approximately equal power. Next comes the LR test and the LM test. Both results may well arise because we are considering a nonlocal alternative hypothesis.[9] That is, the expansions required for the optimal power theorems hold for parameter vectors $\theta_A = \theta_0 + \delta$ where δ must be sufficiently small, e.g., $\delta = d/\sqrt{N}$ where d is a constant vector. But, examples have been given such as Peers [**24**] where significant differences can arise. No generally accepted theory exists for the nonlocal case. Or, the samples may not be large enough for the reliable application of asymptotic theory. We see our results as a particular example and as a caution against relying too heavily on the local asymptotic theory.

For our example we consider a three choice MNL specification of equation (1.1) with only a single explanatory variable. Furthermore, we assume only a single data configuration occurs, $z_1 = 1$, $z_2 = 0$, $z_3 = 0$. We assume $N = 1000$ repetitions of the choice and cell counts n_1, n_2, n_3. We assume that the true parameter $\beta = \log 2$ generates the observations under the null hypothesis. For these data the MNL choice probabilities take the form:

$$(3.1) \qquad p_1 = e^{\beta}/(2 + e^{\beta}), \qquad p_2 = p_3 = 1/(2 + e^{\beta}),$$

which for $\beta = \log 2$ are $p_1 = .5$ and $p_2 = p_3 = .25$. The log likelihood function for

[8] See the earlier version of this paper (Hausman–McFadden [**12**]) for an application of the test of Section 1 and the tests of this section to an example of appliance choice.

[9] A discussion of local and nonlocal alternatives for an application of the Hausman type test is given in Hausman [**10**].

the unrestricted choice set is

$$L_C(\beta) = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 = n_1 \beta - N \log(2 + e^{\beta}). \tag{3.2}$$

Maximization of the likelihood function yields the estimates

$$\beta_C = \log\left(\frac{2n_1}{n_2 + n_3}\right), \qquad V_C = N/n_1(n_2 + n_3), \tag{3.3}$$

where V is the large sample estimator of the variance of β given in equation (1.13).

For the restricted choice set we eliminate choice 3. Maximization of the restricted log likelihood function $L_A(\beta)$ of equation (1.6) yields estimates

$$\beta_A = \log\left(\frac{n_1}{n_2}\right), \qquad V_A = (n_1 + n_2)(n_1 + 2n_2)/n_1 n_2 \tag{3.4}$$

where V_A is calculated from equation (1.10). The first specification test based on the deletion of alternative 3 where Q of equation (1.16) is evaluated at β_C is

$$\mathrm{H}_3 = (\beta_C - \beta_A)^2/Q = \left(\log\left(\frac{2n_2}{n_2 + n_3}\right)\right)^2 (n_2 + n_3). \tag{3.5}$$

By symmetry the statistic based on the deletion of alternative 2 is

$$\mathrm{H}_2 = \left(\log\left(\frac{2n_3}{n_2 + n_3}\right)\right)^2 (n_2 + n_3). \tag{3.6}$$

Note that the last possible test H_1 is not defined under our data configuration since β is not identified when alternative 1 is deleted.

As the model specification for the alternative hypothesis we use the nested logit model. For our uses it can be most conveniently written as

$$p_1 = \Pi_1 p, \quad p_2 = \Pi_2 p, \quad p_3 = 1 - p = e^{\beta z_3}/\left(e^{\beta z_3} + (e^{\beta z_1/\lambda} + e^{\beta z_2/\lambda})^{\lambda}\right), \tag{3.7}$$

where $\Pi_i = e^{\beta z_i/\lambda}/(e^{\beta z_1/\lambda} + e^{\beta z_2/\lambda})$ for $i = 1, 2$. The log likelihood has the form

$$L(\alpha, \lambda) = n_1 \alpha - (n_1 + n_2)(1 - \lambda)\log(e^{\alpha} + 1) - N\log\left(1 + (e^{\alpha} + 1)^{\lambda}\right) \tag{3.8}$$

where we use the parameterization $\alpha = \beta/\lambda$. The maximum likelihood estimates are

$$\tilde{\alpha} = \log\frac{n_1}{n_2}, \qquad \tilde{\lambda} = \log\left(\frac{n_3}{n_1 + n_2}\right)\Big/\log\left(\frac{n_2}{n_1 + n_2}\right). \tag{3.9}$$

Denote the partitioned large sample estimate of the information matrix as

$$-\lim_{N\to\infty} E\begin{bmatrix} L_{\alpha\alpha} & L_{\alpha\lambda} \\ L_{\alpha\lambda} & L_{\lambda\lambda}\end{bmatrix} = \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\lambda} \\ A_{\lambda\alpha} & A_{\lambda\lambda}\end{bmatrix}. \tag{3.10}$$

Then the Wald statistic for the hypothesis $H_0 : \lambda = 1$ is

$$(3.11) \qquad \tilde{W} = (\lambda - 1)^2 \tilde{V}^{-1}$$

where $\tilde{V}^{-1} = A_{\lambda\lambda} - A_{\alpha\lambda}^2 / A_{\alpha\alpha} = (n_2 n_3 (n_1 + n_2) t_1^4)/(N n_2 t_1^2 + t_2^2 n_1 n_3)$ for $t_1 = \log(n_2/(n_1 + n_2))$ and $t_2 = \log(n_3/(n_1 + n_2))$.[10] Next the LR statistic is calculated from the unrestricted log likelihood function of equation (3.2) for the MNL model which sets $\lambda = 1$ and the nested log likelihood of equation (3.8):

$$(3.12) \qquad \mathrm{LR} = 2(L(\alpha, \lambda) - L_C(\beta)) = 2n_2 \log\left(\frac{2n_2}{n_2 + n_3} \right) + 2n_3 \log\left(\frac{2n_3}{n_2 + n_3} \right).$$

Finally we derive the LM statistic. Under the null hypothesis with $\lambda = 1$, we use the MNL estimate $\beta = \log(2n_1/(n_2 + n_3))$ so that in equation (3.7) $\Pi_2 = (n_2 + n_3)/(N + n_1)$ and $p = (N + n_1)/2N$. We then evaluate the gradient of the likelihood function (3.8) at this point to find

$$(3.13) \qquad L_\lambda = \left(\frac{n_3 - n_2}{2} \right) \log(n_2 + n_3)/(N + n_1)$$

together with its large sample variance

$$(3.14) \qquad \tilde{\tilde{V}}^{-1} = -\left(EL_{\lambda\lambda} - (EL_{\lambda\alpha})^2 / EL_{\alpha\alpha}\right)$$

$$= \frac{n_2 + n_3}{4N} \left(\log(n_2 + n_3)/(N + n_1)\right)^2.$$

Therefore we calculate the LM statistics as[11]

$$(3.15) \qquad \mathrm{LM} = L_\lambda^2 / \tilde{\tilde{V}}^{-1} = \frac{(n_3 - n_2)^2}{n_2 + n_3}.$$

Asymptotically, each of the test statistics H3, H2, W, LR, LM is under the null hypothesis distributed χ_1^2. Note that (n_1, n_2, n_3) has the trinomial distribution,

$$(3.16) \qquad \Pr(n_1, n_2, n_3) = \frac{N!}{n_1! \, n_2! \, n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3},$$

where p_1, p_2, p_3 are given by equation (3.1) with $\lambda = 1$ under the null hypothesis, and $0 < \lambda < 1$ under alternatives. For the example, we calculate numerically the exact distribution of the statistics for $\lambda = 1$ and for alternative values $\lambda = (.95, .9, .85, .8, .75, .7)$. The procedure for calculation of the exact distribution is straightforward, albeit time intensive on a computer. For $\beta = \log 2$, $z_1 = 1$,

[10] Details of these derivations will be provided by the authors upon request. W. Newey helped discover and correct an error in an earlier version of the paper (Hausman–McFadden [12]).

[11] T. Rothenberg has pointed out that this test is the best unbiased test for the trinomial case with null hypothesis $p_2 = p_3$. See Lehmann [16, p. 147]. Of course, this result holds only for our particular example.

$z_2 = z_3 = 0$, and for the chosen value of λ we consider all possible realizations of n_1, n_2, and n_3 such that $n_1 + n_2 + n_3 = 1000$. We then calculate the values of the test statistics with the formulae of equations (3.5) to (3.15) for the given (n_1, n_2, n_3). Since equation (3.16) gives the probability for each realization of (n_1, n_2, n_3) it is straightforward to calculate the exact distribution of the test statistics under the null and alternative hypothesis. Note that no Monte Carlo sampling is performed nor is numerical integration used so that our distribution functions are exact. That is, we have evaluated the size and power of the tests via complete enumeration of the discrete sample space so that our results are exact to within the (double precision) accuracy of the computer. These calculations permit determination of the exact sizes of each test for various nominal sizes, and of the power functions.

Table III.1 gives the exact distribution of the five alternative test statistics under the null hypothesis: H3 is the statistic from equation (3.5) based on deletion of alternative 3, H2 the corresponding statistic for deletion of alternative 2, and WALD, LM, and LR are the Wald, Lagrange multiplier, and likelihood ratio statistics, respectively.

Table III.1 shows that the exact distributions of the test statistics are close to their asymptotic limit. All the statistics have relatively lower tails than the chi

TABLE III.1

EXACT CUMULATIVE DISTRIBUTION FUNCTIONS OF THE TEST STATISTICS, $\lambda = 1$

ARGUMENT	CHI SQUARE	H3	H2	WALD	LM	LR
.0009766	.0250000	0.0178390	0.0178390	0.0178390	0.0178390	0.0178390
.0039119	.0500000	0.0534814	0.0534814	0.0534814	0.0534814	0.0534814
.0157204	.1000000	0.0890171	0.0890171	0.0890172	0.0890171	0.0890171
.0356288	.1500000	0.1594559	0.1594559	0.1594559	0.1594646	0.1594646
.0639752	.2000000	0.1942877	0.1942877	0.1942892	0.1942862	0.1942862
.1012512	.2500000	0.2569941	0.2569941	0.2562594	0.2572682	0.2572682
.1481302	.3000000	0.2962718	0.2962718	0.2963227	0.2961984	0.2961984
.2055111	.3500000	0.3531899	0.3531899	0.3526435	0.3545815	0.3544815
.2745780	.4000000	0.3961915	0.3961915	0.3973523	0.3946576	0.3946576
.3568926	.4500000	0.4512023	0.4512023	0.4508477	0.4524609	0.4524609
.4545310	.5000000	0.4997899	0.4997800	0.4990429	0.5002996	0.5002995
.5702946	.5500000	0.5499002	0.5499002	0.5505190	0.5487446	0.5487446
.7080494	.6000000	0.6008482	0.6008402	0.6011120	0.5983838	0.5983838
.8732966	.6500000	0.6503276	0.6503276	0.6504698	0.6498035	0.6487021
1.0741902	.7000000	0.6991320	0.6991320	0.6996405	0.7004598	0.7004598
1.3235021	.7500000	0.7497382	0.7497382	0.7501819	0.7496198	0.7496150
1.6428286	.8000000	0.8000640	0.8000640	0.8003115	0.7998943	0.7998868
2.0730254	.8500000	0.8500422	0.8500422	0.8501742	0.8502551	0.8496568
2.7067207	.9000000	0.8997094	0.8997094	0.8998580	0.9000008	0.8999949
3.8431482	.9500000	0.9494465	0.9494465	0.9496468	0.9501477	0.9498332
5.0258684	.9750000	0.9743730	0.9743730	0.9743900	0.9750777	0.9749321
6.6369923	.9900000	0.9893903	0.9893903	0.9893313	0.9900406	0.9900223
10.8291046	.9990000	0.9987133	0.9987133	0.9986340	0.9990174	0.9989938
15.1371548	.9999000	0.9998161	0.9998161	0.9997878	0.9999030	0.9998992
19.5106637	.9999900	0.9999701	0.9999701	0.9999630	0.9999905	0.9999899
23.9262119	.9999990	0.9999947	0.9999947	0.9999930	0.9999991	0.9999930
28.3710278	.9999999	0.9999990	0.9999990	0.9999986	0.9999999	0.9999999

square. However, the upper tails are quite close to the central chi square distribution. Therefore, the nominal and exact test sizes agree rather well. As the calculations indicate in column 1 of Table III.1 for $\lambda = 1.0$ even at the 1 per cent test level the largest deviation between nominal and exact test size is .00067 for the Wald test which is only 6.7 per cent off. For the 5 per cent test level the maximum percentage deviation between nominal and exact test is 1 per cent while for the 10 per cent level the maximum percentage deviation is 3 per cent. Therefore, the overall agreement between the nominal and the exact test sizes is excellent.

Table III.2 gives the power of each of the tests against the nested logit model with values of λ less than one for a sample size of $N = 1000$. The exact powers of the asymptotic tests, uncorrected for size, are ranked

$$\text{WALD} > \text{H3} > \text{LR} > \text{LM} > \text{H2}$$

over most values of λ. The differences in power are of sufficient size and uniformity to suggest that the WALD and H2 tests are preferable to the other tests. The LR and LM tests have about 10 per cent less power at values of λ

TABLE III.2

EXACT POWER OF THE TESTS, $N = 1000$

	Nominal Size	$\lambda = 1.0$	$\lambda = .95$	$\lambda = .90$	$\lambda = .80$	$\lambda = .70$
1. H3:	.10	.10029	.17861	.38142	.86438	.99631
	.05	.05055	.10931	.27564	.79054	.99152
	.01	.01059	.03568	.12406	.59877	.96630
2. H2:	.10	.15318	.15318	.33613	.83543	.99466
	.05	.05055	.08363	.22158	.73703	.98646
	.01	.01059	.01869	.07302	.47905	.93588
3. WALD:	.10	.10014	.18087	.38523	.86643	.99640
	.05	.05035	.11159	.28013	.79416	.99177
	.01	.01067	.03733	.12839	.60677	.96772
4. LM:	.10	.10000	.16576	.35912	.85076	.99557
	.05	.04985	.09577	.23829	.76534	.98933
	.01	.00996	.02619	.09710	.54204	.95376
5. LR:	.10	.10001	.16577	.35915	.85080	.99558
	.05	.05017	.09622	.24902	.76580	.98934
	.01	.00998	.02622	.09717	.54211	.95377
6. Asymptotic Approximation	.10	.1000	.1632	.3404	.7914	.9793
	.05	.0500	.0941	.2328	.6901	.9577
	.01	.0100	.0256	.0889	.4523	.8662
7. Approximate Slopes	H3	.0	.00040	.00174	.00815	.02216
	H2	.0	.00038	.00155	.00639	.01500
	WALD	.0	.00041	.00175	.00832	.02277
	LM	.0	.00039	.00164	.00718	.01794
	LR	.0	.00039	.00164	.00719	.01806

TABLE III.3
EXACT POWER OF THE TESTS, $N = 100$

	Nominal Size	$\lambda = 1.0$	$\lambda = .95$	$\lambda = .90$	$\lambda = .80$	$\lambda = .70$
1. H3:	.10	.10268	.12389	.15753	.27235	.45470
	.05	.05402	.07130	.09829	.19294	.35525
	.01	.01594	.02496	.03918	.09332	.20497
2. H2:	.10	.10268	.09854	.10598	.16899	.30819
	.05	.05402	.04759	.04895	.08385	.17908
	.01	.01074	.01074	.00821	.01176	.03388
3. WALD:	.10	.10330	.12163	.15692	.27454	.45623
	.05	.05426	.07328	.10268	.20131	.36553
	.01	.01685	.02674	.04204	.09895	.21375
4. LM:	.10	.09876	.10528	.12595	.21629	.37950
	.05	.04943	.05385	.06804	.13361	.26628
	.01	.00959	.01010	.01566	.04033	.17901
5. LR:	.10	.01080	.10854	.12966	.22170	.38730
	.05	.05010	.05454	.06882	.13478	.26806
	.01	.00999	.00143	.01622	.04141	.10629
6. Asymptotic Approximation	.10	.1000	.10632	.12541	.2003	.3181
	.05	.0500	.05428	.06739	.1214	.2142
	.01	.0100	.01141	.01586	.0364	.0792

equal to .95 and .90 where power differences are most critical. However, overall power differences are not extremely large.

In Table III.3 we again do exact power calculations for a sample size of $N = 100$. Here substantial differences arise with respect to the operating characteristics of the tests. The size of the tests ($\lambda = 1.0$) is not as accurate as before with the H2 and H3 tests and Wald test too large by about 8 per cent, e.g., the size of the Wald test is .05426. The percentage difference is even larger at the 1 per cent level although the absolute difference is not large. The overall power ranking for the tests is identical to the previous case. For λ close to one, none of the tests do well. For $\lambda = .8$ the H2 and Wald tests are extremely close and are about 40 percent more powerful at the 5 per cent level than the LR and LM tests and 90 per cent more powerful than the H3 test. Comparable results occur for the $\lambda = .70$ case. Thus the Wald and H2 tests demonstrate substantially more power than the LR or LM test although their size is somewhat too large.

The difference in the test results offers a caution for relying too much on first order asymptotic expansions. The Wald test and H2 tests always do better than the other tests, sometimes substantially better. This result is surprising because all the tests have the same noncentrality parameter under local deviations. That is, the LR, LM, and Wald tests are asymptotically equivalent while the Hausman type tests and LR tests are also asymptotically equivalent for our particular example given the results of Holly **[15]**. We now calculate the noncentrality parameter for all of the tests.

The noncentrality parameter is perhaps most easily calculated from the Wald test. For a sequence of local alternatives it is $\nu = \delta^2/V(\lambda)$ where $V(\lambda)$ is the asymptotic variance of the ML estimator from equation (3.11). From the inverse of the information matrix it is calculated to be $V(\lambda) = 2(e^{\beta} + 2)/\log(e^{\beta} + 1))^2$. Therefore the noncentrality parameter is

$$\nu = \delta^2\left(\log(e^{\beta} + 1)\right)^2/(2e^{\beta} + 4). \tag{3.17}$$

Asymptotically, all the test statistics are distributed as noncentral χ^2 with one degree of freedom and noncentrality parameter ν for local deviations.

To assess the accuracy of the asymptotic approximation and to determine how local the deviation must be for asymptotic theory to provide a reliable guide, in the bottom rows of Table III.2 and III.3 we calculate the asymptotic approximation to the power of the test with a noncentral χ^2 distribution and ν from equation (3.17) where $\delta^2 = N(1-\lambda)^2$.[12] For the $N = 1000$ case the LM and LR tests are reasonably close for $\lambda = .95$ and $\lambda = .90$. The Wald and H3 tests have higher exact power than the asymptotic approximation predicts. For λ less than .9 all the tests have greater exact power than the prediction of the asymptotic approximation. Similar results are found for the case $N = 100$ although the underprediction of the exact power by the asymptotic approximation becomes more serious as λ departs from unity. For departures of size $\delta = .2$ or $\delta = .3$ the deviations between exact power and the asymptotic prediction are substantial although a sample size as small as $N = 100$ may not be large enough for the approximation to be accurate. Our conclusion here is that, for our example, deviations must be quite local for the first order asymptotic approximation to provide accurate guidance for power considerations. Presumably, higher order Edgeworth type expansions would lead to improved accuracy, although the tests are no longer equivalent to higher orders of approximation.

An alternative asymptotic approximation criterion is given by the approximate slope of the test statistics (Bahadur [**2**, **3**] and Geweke [**9**]). An interpretation of the approximate slope statistic can be given either as the rate that the size of a test approaches zero for fixed power or that the ratio of the approximate slopes gives the inverse ratio of sample sizes needed to attain equal power as the size of the nonrejection region increases. The approximate slopes of our test statistics are easily calculated by inserting the theoretical probabilities for the number of realizations. This calculation is equivalent to division by the sample size followed by taking the limit as the sample size goes to infinity. The approximate slopes are given at the bottom of Table III.2. It is interesting to note that the ranking of the approximate slopes is *identical* to the ranking of the exact power of the test statistics. Furthermore, the approximate slopes of the WALD and H3 statistics are very close as are the approximate slopes for the LM and LR tests which corresponds to our results. Therefore, in our example, the approximate slopes

[12] A formula for the infinite series representation of the noncentral χ^2 is given in Abramowitz and Stegun [1, formula (26.4.25)].

provide a valuable tool to predict the relative performance of the various tests although they do not provide a guide to the absolute power characteristics.

Now a special feature of our example was that all the tests had the same degrees of freedom. In more general situations the classical tests would remain one degree of freedom tests of the hypothesis $\lambda = 1$. The Hausman-type tests would have the degrees of freedom grow with the number of parameters in the restricted model. Holly [15] has discussed situations where Hausman type tests and classical tests differ in their degrees of freedom. While for the general case no ranking can be made because the noncentrality parameters will not be equal, here we consider a two parameter case to see if important differences emerge. We again have a three alternative case, but now we have two right-hand-side variables. We set $x_{1n} = 1$ for $n = 1, \ldots, N_a$ and $x_{1n} = 0$ for $n = N_a + 1, \ldots, N$ where $N_a + N_b = N$. Then we set $x_{2n} = 1 - x_{1n}$. We therefore have 2 slope parameters α and β and the parameter for the nested logit model λ. The probabilities take the form

$$P_{1a} = \Pi_{1a} p_a, \quad P_{2a} = \Pi_{2a} p_a, \quad \text{and} \quad P_{3a} = 1 - p_a = 1/\left(1 + (1 + e^{\alpha/\lambda})^{\lambda}\right) \tag{3.21}$$

where $\Pi_{1a} = e^{\alpha/\lambda}/(1 + e^{\alpha/\lambda})$ and $\Pi_{2a} = 1 - \Pi_{1a}$. Similar formulas apply for the P_{ib} probabilities with β replacing α. Let $a = \alpha/\lambda$, $b = \beta/\lambda$, and let n_i^a be the number choosing alternative i in the first N_a observations and n_i^b be the number choosing alternative i in the last N_b observations. The likelihood function is

$$L(\alpha, \beta, \lambda) = n_1^a \log P_{1a} + n_2^a \log P_{2a} + n_3^a \log P_{3a} + n_1^b \log P_{1b} + n_2^b \log P_{2b} + n_3^b \log P_{3b}. \tag{3.22}$$

A closed form solution for the MLE no longer exists.

We compare the LM tests and H2 and H3 tests. Both tests are based on the MNL estimates which allow us to compute the exact power characteristics of the tests. Presumably, these tests would be most often used as model specification given the difficulty of estimation of the nested logit model.[13] The LM test is calculated to be

$$\mathrm{LM} = \frac{\left[\log \Pi_{2a}(n_3^a - n_2^a) + \log \Pi_{2b}(n_3^b - n_2^b)\right]^2}{(n_2^a + n_3^a)(\log \Pi_{2a})^2 + (n_2^b + n_3^b)(\log \Pi_{2b})^2} \tag{3.23}$$

where $\Pi_{2a} = (n_2^a + n_3^a)/(N_a + n_1^a)$ and likewise for Π_{2b}. The H3 test is computed to be

$$\mathrm{H3} = \left(\log\left(\frac{2n_2^a}{n_2^a + n_3^a}\right)\right)^2 (n_2^a + n_3^a) + \left(\log\left(\frac{2n_2^b}{n_2^b + n_3^b}\right)\right)^2 (n_2^b + n_3^b) \tag{3.24}$$

while the H2 test is computed similarly with n_3^a or n_3^b in the numerator.

[13] A Wald type test could be based on the consistent, but asymptotically inefficient estimate of λ from the sequential model. The correct asymptotic standard error for this estimate would need to be computed.

TABLE III.4

EXACT POWER OF THE PARAMETER TESTS, $N = 100$

	Nominal Size	$\lambda = 1.0$	$\lambda = .95$	$\lambda = .90$	$\lambda = .80$
1. H2:	.10	.11914	.14460	.17998	.35798
(2 df)	.05	.07314	.09437	.12351	.27730
	.01	.01933	.04112	.08801	.15931
2. H3:	.10	.11914	.10200	.11638	.22938
(2 df)	.05	.07314	.05800	.05787	.13876
	.01	.01933	.02095	.01483	.03866
3. LM:	.10	.10003	.10636	.12629	.28223
(1 df)	.05	.04947	.05370	.06723	.18277
	.01	.01112	.02693	.03549	.06806

In Table III.4 we give exact results for the case of $N = 100$ where $N_a = N_b = 50$. We choose $\alpha = \log 2.5$, $\beta = \log 1.5$, and consider values of λ of (1.0, .95, .90, .75). Note that the H2 and H3 tests are now 2 degrees of freedom χ^2 tests while the LM test is based on a one degree of freedom χ^2 distribution. The results demonstrate that the H2 test again has more power than does the LM test which is better than the H3 test. In fact, the power results are very similar to Table III.3 where we considered a sample size of $N = 100$, but with only one degree of freedom situation. However, here the actual size of the H2 and H3 test is too large by a greater magnitude than in Table III.3 Therefore, judgements on the relative merits of the tests must not only consider the power characteristics but also the size where the H2 and H3 tests are not as good as the LM test. Our results from this 2 parameter case demonstrate that the power of the H2 and H3 tests do not decline in relation to the classical tests even here where the alternative model of the classical tests corresponds to the true model. In actual testing situations, where the true model is often unlikely to be a nested logit model, the performance of the Hausman type test has appeared to be comparable to the classical tests.[14]

For comparison purposes, we recalculated the exact classical tests based on a misspecified nested logit model. That is, in Equation (3.7) we interchanged p_2 and p_3 so that choices 1 and 3 now lie on the same branch. However, we continue to conduct the test procedures based on the original model specification. The H2 now has more power than either the H3 or the Wald test while the power of the LM and LR tests remain the same. The LM and LR tests have greater power than the Wald test here. Thus, it seems useful to investigate different tree structures when using the classical tests because of their sensitivity to model specifications. In the last line of Table III.5 we also present the maximum likelihood estimates of λ which all exceed the theoretical maximum of

[14] See Hausman–McFadden [12] for an example of the performance of the tests in an applied situation.

TABLE III.5

EXACT TEST RESULTS FOR THE MISSPECIFIED MODEL AT TEST SIZE .05 FOR $N = 1000$ AND $N = 100$

	$N = 1000$			$N = 100$		
	$\lambda = .95$	$\lambda = .9$	$\lambda = .75$	$\lambda = .95$	$\lambda = .90$	$\lambda = .75$
Nominal size 0.05						
H2	.10931	.27564	.94241	.07067	.09797	.26494
H3	.08363	.22158	.91877	.04607	.04665	.11518
WALD	.08042	.21460	.91541	.04359	.04096	.09534
LM	.09577	.23829	.93153	.05385	.06804	.19024
LR	.09622	.24902	.93166	.05454	.06882	.19168
MLE estimate of λ	1.051	1.0107	1.301			

1.0 for a random utility model which again indicates misspecification.[15] Likewise, it seems useful to base the Hausman type specification test on different restricted choice sets since its power also depends on the model specification. In an actual applied situation where the correct specification is unknown both different restricted choice sets and different tree structures should be investigated for the respective tests since our examples indicate the sensitivity of the operating characteristics of the tests to the choice of alternative specifications.

4. CONCLUSIONS

In terms of applying the Hausman type tests or the trinity of tests, the non-uniqueness of application can arise. For instance in the 3 choice case any of the 3 alternatives can be dropped or 3 different tree structures can be defined. As the number of choices grows, the different possible combinations grow factorially. If more than one test is performed, the problem of controlling for the size arises because the tests will not be independent. In the case of the H2 and H3 tests the different estimates could be combined to form a test after their joint asymptotic covariance matrix is calculated which would not be difficult.[16] Alternatively, the distribution of the maximum of say j test statistics under the null hypothesis of the MNL model might be possible to derive. Another approach is to use the IIA property of the null hypothesis. The restricted choice set A could be successively decreased when more than 3 choices exist and the size controlled for as we do when independent F tests in linear models are used. Close substitutes should be eliminated from the choice set C when tests of the IIA property are conducted. However, a general theory of the optimal procedure to design the restricted choice set A is a topic for further research.

We conclude that the Hausman type test and Wald test are the best choices to

[15] We emphasize that a ML estimate of λ which exceeds unity is a 'regular' outcome of the specification of the nested logit model from a statistical viewpoint. However, the choice model cannot be interpreted as a random utility model if the true λ is greater than one.

[16] McFadden [**21**] considers the combination of tests when different restricted choice sets are used to estimate test statistics.

test the IIA assumption in MNL models. The Wald test does require maximum likelihood, or at least asymptotically efficient estimates, of the nested logit model and correct specification of the tree structure in the nested logit model. The Wald test has higher power in all our examples than the LR or LM test and the H2 or H3 test for the correct specification. However, for an incorrect specification the LR and LM tests are superior to the Wald test. But the Wald test has significantly greater computational requirements than does the first type test. An alternative test based on a consistent estimate of λ from the sequential logit estimator is possible by the use of Neyman's [**23**] C_α procedure. However, our experience with sequential logit estimator is that it gave unreliable estimates of λ.[17] Since the Hausman type test gave results in general close to that of the Wald test across the range of our examples, we recommend it as a general purpose specification test for the MNL model. The Wald test should also be considered when the analyst feels that the nested logit model provides the correct specification for the choice problem under consideration. The Wald test requires more sophisticated computer software; presumably, computer software will become available which permits its convenient estimation. But certainly some test of the IIA property should be made when the MNL model is used. Our experience is that the Hausman type test and classical tests have rejected the MNL specification in a number of applications.

Massachusetts Institute of Technology

Manuscript received October, 1981; final revision received November, 1983.

[17] An indication of the unreliability is that in the applied example of the Hausman–McFadden [**12**] version, one Newton type step beginning at the sequential logit estimates led to a *decrease* in the value of the likelihood function in the *majority* of the cases considered (for step size equal to one). Since the C_α test is based on the one step methodology, we decided against its use. Furthermore, in the misspecified model case, we were often unable to find an increase in the likelihood function even when up to three Newton steps were made.

REFERENCES

[**1**] Abramowitz, M., and I. Stegun: *Handbook of Mathematical Functions.* New York: Dover Publications, 1965.
[**2**] Bahadur, R. R.: "Stochastic Comparison of Tests," *Annals of Mathematical Statistics*, 31(1960), 845–863.
[**3**] ———: *Some Limit Theorems in Statistics.* Philadelphia: SIAM, 1971.
[**4**] Berndt, E., B. Hall, and J. Hausman: "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 4(1974), 653–666.
[**5**] Cox, D. R., and D. Hinckley: *Theoretical Statistics.* London: Chapman and Hall, 1974.
[**6**] Debreu, G.: "Review of R. Luce, Individual Choice Behavior," *American Economic Review*, 50(1960), 186–188.
[**7**] Domencich, T., and D. McFadden: *Urban Travel Demand.* Amsterdam: North Holland, 1975.
[**8**] Fisher, R. A.: "Theory of Statistical Estimation, *Proceedings of the Cambridge Philosophical Society*, 22(1925), 700–725.
[**9**] Geweke, J.: "The Approximate Slopes of Economic Tests," *Econometrica*, 49(1981), 1427–1442.
[**10**] Hausman, J.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251–1272.

[11] ———: "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables," *Bell Journal of Economics*, 10(1979), 33–54.

[12] HAUSMAN, J., AND D. McFADDEN: "Specification Tests for the Multinomial Logit Model," Working Paper #292, Department of Economics, M.I.T., 1981.

[13] HAUSMAN, J., AND W. TAYLOR: "A Generalized Specification Test," *Economics Letters*, 6(1981), 239–245.

[14] HAUSMAN, J., AND D. WISE: "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(1978), 403–426.

[15] HOLLY, A.: "A Remark on the Hausman Specification Test," *Econometrica*, 50(1982), 749–760.

[16] LEHMANN, E.: *Testing Statistical Hypothesis*. New York: Wiley, 1959.

[17] MANSKI, C., AND D. McFADDEN: "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

[18] McFADDEN, D.: "Conditional Logit Analysis of Quantitative Choice Behavior," in *Frontiers in Econometrics*, ed. By P. Zarambka. New York: Academic Press, 1973.

[19] McFADDEN, D., W. TYE, AND K. TRAIN: "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," *Transportation Research Record*, 637(1976), 39–46.

[20] ———: "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

[21] ———: "Econometric Analysis of Qualitative Choice," *Handbook of Econometrics*, Vol. 2. Amsterdam: North Holland, 1984.

[22] NEWMAN, D., AND D. DAY: *The American Energy Consumer*. Cambridge: Ballinger, 1975.

[23] NEYMAN, J.: "Optimal Asymptotic Tests of Composite Statistical Hypothesis," in *Probability and Statistics*, ed. by U. Grenander. Stockholm: Almquist and Wiksell, 1959.

[24] PEERS, H.: "Likelihood Rates and Associated Test Criteria," *Biometrika*, 58(1971), 577–587.

[25] RUUD, R.: "A Score Test for Consistency," mimeo, 1982.

[26] SILVEY, S.: *Statistical Inference*. Harmondsworth: Penguin, 1980.

[27] SMALL, K., AND C. HAIAO: "Multinomial Logit Specification Tests," mimeo, 1982.