NON-NESTED TESTING IN LOGIT MODELS:
AN ALTERNATIVE APPROACH

by

James S. Eales
Assistant Professor
Agricultural Economics
University of Illinois
Urbana, IL 61820

February, 1985
Abstract

An asymptotic non-nested hypothesis testing framework is developed, which may be applied to logit models with three or more alternatives. Its advantage over the other currently available frameworks is that its application does not require the estimation of a third logit model. This can be important in economic applications.
Introduction

The increasing availability of data sets on microlevel decision behavior has presented researchers with the opportunity to test a number of interesting behavioral hypotheses. At the same time it has presented econometricians with a new set of challenges. The dependent variable of such problems often takes the form of a dummy variable, representing a qualitative or discrete result. One theoretical construct which allows the consistent aggregation of a collection of such discrete decisions into an estimable model was developed by McFadden. He began with the decision makers' utility functions and derived the multinomial or conditional logit model. The conditional logit model has gained in popularity, because the implied probabilities exist in a closed form and the likelihood function is globally concave (in theory). However, in economic applications the collinearity inherent in many data sets makes estimation of logit models difficult or impossible. This is particularly true if there are more than two alternatives (see Davidson and MacKinnon, 1984).

As logit models become more popular in applications, situations will arise in which researchers wish to test non-nested alternative hypotheses. This might happen in a number of ways. For instance, Capps and Kramer test a logit against a probit formulation in a binary choice problem. Other researchers have been interested in testing the "Independence of Irrelevant Alternatives" axiom underlying the logit development. Some of the tests which they have developed to do this are non-nested, e.g. Gourieroux, Monfort and Trognon or Hausman and McFadden. A
third aspect of model selection which leads to tests of non-nested hypotheses occurs in logit models where the returns to the decision maker are assumed to be stochastic. In such instances, expectations play a key role in the choice problem. However, the specification of the information sets upon which these expectations are based can lead to testing situations where the hypotheses are non-nested. An application to the testing of non-nested alternative information sets in a logit framework is presented below.

One test is currently available with which to do non-nested hypothesis testing in a logit model with three or more alternatives. Gourieroux, Monfort and Trognon (GMT) have developed a very general test which is applicable whether the hypotheses are nested or not. The test statistic is similar to a Wald statistic and has an asymptotic chi-square distribution. The difficulty with the GMT test is that in order to conduct it a third logit model must be estimated. In economic applications this may be a significant problem because of the collinearity problems alluded to above.

The purpose of this paper is to provide an asymptotically equivalent test which can be applied to the testing of logit models without the requirement of a third estimation. The alternative test is a modification of that developed by MacKinnon, White and Davidson (MWD). The extension of the MWD test to conditional logit problems with three or more alternatives, is made in the next section. In the third section an application
to the information sharing arrangements among fishermen in the California Pink Shrimp fishery is presented. The final section is devoted to summarizing and a few conclusions.

**The MWD Test in Logit Models**

Davidson and MacKinnon (1981) developed a technique for conducting a non-nested test which is easy to use and asymptotically equivalent to other available tests. Their work was extended by MacKinnon, White and Davidson (MWD). Briefly, they utilize a notion originally suggested by Cox of specifying a comprehensive model, in this case a convex combination of the two non-nested alternatives. They then linearize this in a Taylor series around the maximum likelihood values of the null model. MacKinnon, et. al. then prove that this formulation yields the same results in the limit as estimating the nonlinear compound model directly.

The intuition behind the MWD test is that if the null hypothesis is the true model then any alternative should have very little to add to the explanation of the residual of the null model and the coefficient, \( \hat{a} \), of the alternative's prediction should not be significantly different from zero.

To apply their test to non-nested testing in logit models, it is first necessary to modify the logit specification as follows:

\[(1) \quad H_0: \quad Y_{jt} = P_0jt + \varepsilon_0jt\]

\[(2) \quad H_1: \quad Y_{jt} = P_1jt + \varepsilon_1jt\]
where $j = 1, \ldots, J$ indexes the alternatives, $t = 1, \ldots, T$ indexes the observations, $Y_{jt} = 1$ if alternative $j$ was chosen for observation $t$ and $Y_{jt} = 0$ otherwise, and

\begin{align*}
(3) \quad P_{0jt} &= \frac{\exp(X_{0jt} \beta)}{\sum_{i=1}^{J} \exp(X_{0jt} \beta)} \\
(4) \quad P_{1jt} &= \frac{\exp(X_{1jt} \gamma)}{\sum_{i=1}^{J} \exp(X_{1jt} \gamma)}.
\end{align*}

Equations 1 and 2 are similar to the linear probability model, but differ in the following way: The linear probability model typically relates the sample proportions to a linear combination of the explanatory variables plus an error term. Equations 1 and 2 relate individual observations on the discrete choice among alternatives to the logit form of the probabilities plus an error term as a system, that is equation 1 (and 2) should be viewed as $T$ observations on $J$ equations. The advantage of this specification over the linear probability model is that the linear probability model can produce inconsistent results, i.e. predicted probabilities outside the interval $(0,1)$. In equation 1 or 2 the predicted probabilities have been restricted to this interval. The disadvantage of this specification is that it is a nonlinear system of equations with a singular covariance matrix. This makes the direct estimation of these systems difficult and expensive for most problems. However, in the context of adapting the MWD test applications in logit models, equations 1 and 2 can be exploited advantageously.

The composite model corresponding to equations 5 and 6 is:
\( Y_{jt} - \hat{P}_{jt} = \alpha (\hat{P}_{jt} - \hat{P}_{0jt}) + (\hat{\alpha}_{P_{jt}/\alpha_b^k})' b + \varepsilon_{jt} \)

where

\[ \hat{\alpha}_{P_{jt}/\alpha_b^k} = (X_0^k_{jt} - \bar{X}_0^k) \hat{P}_{0jt} \]

\[ \bar{X}_0^k = \frac{1}{J} \sum_{j=1}^{J} X_0^k_{jt} \hat{P}_{0jt} \]

\[ \varepsilon_{jt} = (1 - \hat{\alpha}) \varepsilon_{0jt} + \varepsilon_{1jt} \]

\( k = 1, \ldots, K \) indexes the number of coefficients, \( j \) the alternatives, \( t \) the observations, and the hats indicate that \( P_0 \) and \( P_1 \) are evaluated using the maximum likelihood estimates of \( B \) and \( Y \) respectively. The coefficient of interest is again \( \alpha \), while the \( b \) coefficients are estimated to improve the properties of the estimate of \( \alpha \) and its variance.

Note that equation 5 is a linear regression, so the recasting of the hypotheses into the form of equations 1 and 2 allows the formulation of the non-nested hypothesis test in the MWD framework. Assuming that consistent estimates of \( \alpha \) and its variance can be derived then an asymptotic \( t \) test can be used to compare the non-nested alternatives, \( H_0 \) and \( H_1 \). To simplify notation in the examination of the properties of equation 5, let

\[ Z_{jt} = [\hat{P}_{jt} - \hat{P}_{0jt}, (\hat{\alpha}_{P_{jt}/\alpha_b^k})'] \]

\[ d' = (\alpha, b') \]

\[ \hat{Y}_{jt} = \hat{Y}_{jt} - \hat{P}_{0jt} \]

Then equation 5 may be rewritten:

\[ \hat{Y}_{jt} = Z_{jt} d + \varepsilon_{jt} \]
or if the \( Y \)'s, \( Z \)'s, and \( \varepsilon \)'s are stacked by alternative by observation, i.e. \( Y' = [Y_{11}, \ldots, Y_{J1}, \ldots, Y_{1T}, \ldots, Y_{JT}] \), etc., then equation 5 may be written:

\[
(13) \quad \hat{Y} = Zd + \varepsilon.
\]

The expected value of the error term in equation 12, is:

\[
(14) \quad E_0(\varepsilon_{jt}) = E_0(1 - \alpha) \varepsilon_{0jt} + \alpha \varepsilon_{1jt}
\]

\[
(15) \quad = (1 - \alpha) E_0(\varepsilon_{0jt}) + \alpha E_0(\varepsilon_{1jt})
\]

\[
(16) \quad = 0
\]

where \( E_0 \) indicates that the expectation is taken under the null hypothesis, in which case \( \alpha = 0 \).

Note that \( \varepsilon_{jt} \) will take on either the value \( 1 - E_0(P_0jt) - E_0(Z_{jt}) \) or \( -E_0(P_0jt) - E_0(Z_{jt}) \). Using this it is easy to show:

\[
(17) \quad \text{Var}(\varepsilon_{jt}) = E_0(\varepsilon_{jt}) = P_0jt(1 - P_0jt)
\]

\[
(18) \quad \text{cov}(\varepsilon_{jt}, \varepsilon_{is}) = \begin{cases} -P_0jtP_0it & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}
\]

Then the covariance matrix, \( V \), of the error terms in equation 12 is block diagonal:

\[
(19) \quad V = \begin{bmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_T \end{bmatrix}
\]

and

\[
(20) \quad V_t = \begin{bmatrix} P_01t(1-P_01t) & \cdots & P_01tP_01t \\ \vdots & \ddots & \vdots \\ -P_0jtP_01t & \cdots & P_0jt(1-P_0jt) \end{bmatrix}
\]
Once the initial estimation of the logit models corresponding to the null and alternative hypotheses has been done, maximum likelihood estimates of $\beta$ will be available to use in the calculation of $P_{0j_t}$ and so a consistent estimate of $V_t$ can be calculated as well. It would then seem that equation 13 could be estimated by generalized least squares. Unfortunately, each of the $V_t$ is singular by one, i.e. the rank of $V_t$ is $J-1$. This is due to the restriction that for any $t$ the sum of the probabilities for all $J$ alternatives is one -- an alternative is chosen for every observation. It is similar to the problem encountered when estimating share equations in production or demand analysis.

One remedy to this problem is to delete an alternative and estimate the reduced system. This is an obviously inefficient loss of information.

Another approach is to use Rao’s Unified Theory of Linear Estimation (1973: pp 294-302) or Theil’s Generalized Aitken’s theorem (1971: pp 274-281). Each derives best linear unbiased estimators for the case where the covariance matrix of the estimation problem is singular. While Rao’s formulation is more general, both produce the same result in the present problem. This is due to the somewhat pathological formulation of the logit model to apply the MWD test. The reason is that in equation 13 the range space of $Z$, $\mu(Z)$, is contained in the range space of $V$, $\mu(V)$, in Rao’s terminology or that $G'Z = 0$, where $G$ is the matrix of eigenvectors corresponding to the zero eigenvalues of $V$, in
Theil's notation. In such a case the minimum semi-norm solution, i.e. the best linear unbiased estimator when the covariance matrix is singular, is unique.

Since the two solutions, Rao's and Theil's, are equivalent in the current problem, it will simplify the notation to concentrate on one of them. Therefore, in what follows Theil's solution is presented. His condition is that $G'Z = 0$. If this condition holds then:

$$\hat{d}_{Th} = (Z'V^+Z)^{-1}Z'V^+Y$$

where $V^+$ is the Moore-Penrose g-inverse of $V$, is best in the sense that any other linear unbiased estimator has a covariance matrix which exceeds that of $\hat{d}_{Th}$, i.e. $(Z'V^+Z)^{-1}$, by a positive semidefinite matrix (Theil, pp 278-9).

In summary, the non-nested logit models as specified in equations 1 and 2 can be compared statistically by first estimating the parameters of each by maximum likelihood in the normal fashion and then using the results to generate the variables, $Y_{jt} - \hat{P}_{0j} \hat{Y}_{jt}$, $\hat{P}_{1j} - \hat{P}_{0j}$, and $\hat{\gamma}_{P0j} / \hat{\gamma}_{K}$, estimating the coefficients of equation 13 according to the formulas given in equation 21 and finally conducting a t test on the coefficient of the variable, $\hat{P}_{1j} - \hat{P}_{0j}$. If the coefficient is significantly different from zero then the alternative hypothesis $H_1$, rejects the null hypothesis, $H_0$.

An Application in a Conditional Logit Model

The extension of the MWD test was developed to apply non-nested searching models developed for the California Pink
Shrimp fleet in Eales. Each search model was calibrated as a logit model relating the areas chosen for an observation to an assumed probability distribution of catch. The distribution incorporated all of the information the decision maker was assumed to have available at the time the choice was made, that is the catch history, ex-vessel prices, and the cost of moving among the areas. The reason the models are non-nested is because of different assumptions about the information sets. From personal communication with fishermen in the Pink Shrimp fleet, it was known that some of the fishermen participated in information sharing arrangements with one or two other vessels. A reasonable specification of the information set of a fisherman who did not cooperate would be that he had access to all of his past values of catch in each of the areas. If the fisherman were sharing information with another vessel then it would seem appropriate to include not only his own past value of catch but the other vessel's as well. The difficulty was that when the data was collected confidentiality restrictions were imposed and so it was not known which of the fishermen were cooperating. So, for instance, suppose fisherman A is thought to be cooperating with fisherman B. The proposed method of testing for this cooperation is to calibrate a logit model relating the location choices made by fisherman A to past catches made by fisherman A in all the areas. The alternative model relates the same choices to past catches by both fisherman A and fisherman B. Note that these two models are nonnested in that fisherman A may have
cooperated with fisherman C, in which case both hypotheses are false. The two models are then compared via the MWD test.

In Tables 1 and 2 some representative results for a potential, two-vessel information sharing group are presented. Each test matrix contains the results for one vessel, Table 1 for vessel A and Table 2 for vessel B. The null hypothesis models are listed down the left hand side of the matrices and the alternative hypothesis models across the top. The statistic reported in the off diagonal entries of the tables are the t statistics for the MWD test. As described above, it is significantly different from zero at the specified level, then the null model is rejected by the alternative model. For example, in Table 1 the value 3.63 in row A, column AB indicates that the cooperative model, AB, rejects the noncooperative model, A, at the .01 level. Similarly, the AB model is rejected by the A model at the .05 level. In Table 2 the results are similar, however, in the test of BA against B the t statistic is 4.44. There has been some disagreement as to how to interpret this result (see Dastoor). The interpretation employed here is that it merely indicates that the null hypothesis is rejected without any implication as to the validity of the alternative hypothesis.

For comparison, the hypotheses were also tested using the framework developed by Gourieroux, et. al. (GMT). As mentioned above, this required the estimation of a third logit model for each test. The results of the GMT test are presented in Tables 3 and 4. The statistics are similar to a Wald statistic and have
an asymptotic chi-square distribution with three degrees of freedom under the null hypothesis. Note the similarity in the results. Again in all cases the cooperative model rejects the non-cooperative model and vice versa. Both tests were performed on 84 different model comparisons and they agreed, in rejecting or failing to reject the null model at the .05 level of significance, in 53.7% of the cases. While there is no small sample evidence for the GMT test, Pesaran (1982) and Godfrey and Pesaran have done some Monte Carlo studies of the Davidson-MacKinnon J-test, Fisher-McAleer test and both an unadjusted and adjusted Cox-type test. Their results indicate that the unadjusted Cox-type test, which is similar to the GMT test, tends to reject too frequently in small samples. The GMT test rejected the null hypothesis in 71 out of 82 cases or 86.6% of the time. The MWD test rejects 39 out of 82 cases or 47.6% of the time.

Summary and Conclusions

In this paper a framework is developed within which it is possible to apply the non-nested test of MacKinnon, et.al. to logit models. To adapt the MWD test to the logit framework, the logit model is recast into the form of a nonlinear probability model with a system of J equations corresponding to the J alternatives which constitute the set from which the choice is made. In this form the specification avoids the inconsistency of the linear probability model in that its predicted probabilities are constrained to fall in the interval (0,1) by the logit form of the probabilities.
This system of $T$ observations on $J$ nonlinear equations is further hampered by the fact that it possesses a singular covariance matrix. This makes this specification unwieldy for estimation purposes, however, it can be exploited in adapting the MWD test successfully, because it never has to be estimated directly. Instead it is linearized with a Taylor series expansion and the previously derived maximum likelihood estimates are used to produce the variables in the linear regression given in equation 5. Due to the singularity of the covariance matrix of the system, an estimation technique developed independently by Rao and Theil is used to derive best linear unbiased estimates of the coefficients and their standard errors in the regression. The test is then performed as an asymptotic t test on the estimate of the coefficient $\hat{a}$ in equation 5.

The MWD and GMT tests are then applied to non-nested search models of the California Pink Shrimp fishery and the results compared. The MWD test tends to reject much less frequently than the GMT test, which is in agreement with the Monte Carlo results currently available.

The advantage of the MWD test over the GMT test is in situations where estimation of a third conditional logit model, as required by the GMT test is expensive or impossible. Such may be the case in many economic applications where the collinearity of the data makes the estimation of logit models difficult.
Matrices of Results for Testing Cooperation Amongst Two Fishermen in the California Pink Shrimp Fishery - Comparing the MWD and GMT Tests

**TABLE 1**

<table>
<thead>
<tr>
<th>H₀</th>
<th>MWD</th>
<th>A</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>3.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2**

<table>
<thead>
<tr>
<th>H₀</th>
<th>MWD</th>
<th>B</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>6.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BA</td>
<td>-4.44</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3**

<table>
<thead>
<tr>
<th>H₀</th>
<th>GMT</th>
<th>A</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>9.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>6.71</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4**

<table>
<thead>
<tr>
<th>H₀</th>
<th>GMT</th>
<th>B</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>79.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BA</td>
<td>16.56</td>
<td></td>
</tr>
</tbody>
</table>

1. Off diagonal entries are asymptotic values (if significant the Alternative Hypothesis, H₁, rejects the Null Hypothesis, H₀).

2. Off diagonal entries are chi square statistics with three degrees of freedom (if significant H₁ rejects H₀).
REFERENCES


