MONASH 8/93

M O N A S H
U N I V E R S I T Y

AUSTRALIA

A MONTE CARLO STUDY OF TESTS FOR THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY

Tim R.L. Fry and Mark N. Harris

Working Paper No. 8/93

September 1993

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0376 5

A MONTE CARLO STUDY OF TESTS FOR THE INDEPENDENCE OF IRRELEVANT ALTERNATIVES PROPERTY

Tim R.L. Fry and Mark N. Harris

Working Paper No. 8/93

September 1993

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

A Monte Carlo Study of Tests for the Independence of Irrelevant Alternatives Property.

Tim R.L. Fry and Mark N. Harris[1]

Department of Econometrics, Monash University, Clayton, Vic. 3168, Australia.

Abstract: A plethora of tests for the Independence of Irrelevant Alternatives (IIA) property of Logit models of discrete choice behavior has been proposed in the literature. These tests are based upon asymptotic arguments and little is known about their size and power properties in finite samples. This paper uses a Monte Carlo simulation study to investigate the size and power properties of six tests for IIA in the multinomial Logit (MNL) model. Our results show that tests based upon partitioning the choice set appear to have very poor size and power properties in small samples.

J.E.L. Classifications: C25, C12.

This paper is preliminary. The usual caveats apply.

[1]This work was supported by an Australian Research Council small grant.

1. Introduction.

The econometric literature provides a number of models for the analysis of discrete choice behavior (for a survey see *inter alia* Fry *et al* (1993), McFadden (1985)). This paper concerns the testing of a behavioral assumption implied by the most popular model for polychotomous (or multinomial) choice situations, the Logit model. The Logit model embodies the independence of irrelevant alternatives (IIA) property, which implies that the choice between two alternatives in a given choice set depends only upon the attributes of the alternatives being compared and not on the attributes, or the existence, of any other alternatives in the choice set. Whilst this property may be an advantage, it does imply strong restrictions on choice behavior. In particular, it implies a uniform pattern of response to changes in the attributes of one alternative, which is inconsistent with the heterogeneous patterns often encountered in economic data. Thus it is important to see if the data are consistent with this property. That is to test for the existence of the IIA property.

Several tests have been proposed in the literature. They fall into two categories. Those based upon partitioning the choice set and those based upon comparing the Logit model with a particular alternative model specification which does not embody the IIA property. There appears to be little agreement in the literature as to which procedure is the best. Furthermore, the existing tests all rely on asymptotic results and as a result not much is known about their performance in finite samples. This paper examines the finite sample properties of six (four choice set partitioning and two alternative model) tests of the IIA property using a Monte Carlo simulation study.

The plan of the rest of the paper is as follows. In the next section we discuss the specification of Logit models for discrete choice behavior and the associated

IIA property. In section 3 we describe the six tests for the IIA property that we consider in this paper. Section 4 contains the details of our Monte Carlo simulation study and a discussion of our results. Finally section 5 contains some concluding remarks.

2. Logit Models and Independence of Irrelevant Alternatives.

In this section we discuss the specification of Logit models of discrete choice behavior and the associated IIA property. As is common in the literature (see, for example, Fry *et al* (1993)) we will consider a random utility maximization model with (indirect) utility function given by:

$$U_{ij} = V_{ij}(\mathbf{Z}_{ij}, \mathbf{X}_i) + \varepsilon_{ij}, \qquad i = 1, ..., n; \; j = 1, ..., J, \tag{1}$$

where U_{ij} is the utility individual i derives from choosing alternative j which comprises of two components V_{ij} and ε_{ij}. V_{ij} is a deterministic component which depends upon characteristics of the individual $\mathbf{X}_i$ and variables which vary across both individuals and alternatives, $\mathbf{Z}_{ij}$. ε_{ij} is a random component which represents unobservable factors. Typically the V_{ij} function is assumed to be linear. Thus,

$$U_{ij} = \mathbf{Z}'_{ij}\alpha + \mathbf{X}'_i\beta_j + \varepsilon_{ij}, \qquad i = 1, ..., n; \; j = 1, ..., J, \tag{2}$$

and often the $\mathbf{Z}_{ij}$ and J $\mathbf{X}_i$ vectors are combined into $\mathbf{W}_{ij}$ to yield:

$$U_{ij} = \mathbf{W}'_{ij}\delta + \varepsilon_{ij}, \qquad i = 1, ..., n; \; j = 1, ..., J, \tag{3}$$

with $\delta = (\alpha', \beta'_1, ..., \beta'_J)'$.

In the random utility maximization model individuals are assumed to make selections that maximize their utility. Hence the probability that individual i chooses alternative 1 is given by:

$$P_{i1} = P(U_{i1}>U_{i2} \text{ and } U_{i1}>U_{i3} \dots \text{ and } U_{i1}>U_{iJ})$$

with similar expressions for the other selection probabilities. If we assume that the ε_{ij} are independent and identically distributed as Extreme Value then we obtain a Logit model for the selection probabilities with:

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})} \qquad i = 1, ..., n; \; j = 1, ..., \; J. \tag{4}$$

Under the usual assumption of a linear form for V_{ij} (equation (2)) we should distinguish between three forms of the Logit model. If $\alpha = \mathbf{0}$ the model is termed a multinomial Logit (MNL), when $\beta_j = \mathbf{0} \; \forall \; j$ the model is the conditional Logit and when $\alpha \neq \mathbf{0}$ and $\beta_j \neq \mathbf{0} \; \forall \; j$ the model is a mixed Logit. Typically the use of one variant of the Logit model rather than another variant is a matter of data availability. In this paper we concentrate upon the MNL variant of the Logit model which is consistent with modeling with survey data on choices and individual characteristics $\mathbf{X}_i$ (e.g. Crichton and Fry (1992)).

A particularly important, and potentially restrictive, property of the Logit model is that of independence of irrelevant alternatives (for further details see Fry *et al* (1993)). This property states that the odds of choosing alternative *j* over alternative *k* ($k \neq j$) P_{ij}/P_{ik} are independent of all other alternatives and the number of alternatives in the choice set. For a Logit model,

$$\frac{P_{ij}}{P_{ik}} = \frac{\exp(V_{ij})}{\exp(V_{ik})},$$

which using the form for the V_{ij}'s given in equation (2) above will satisfy the definition of IIA. If a model incorporating IIA provides an appropriate representation of discrete choice behavior then considerable advantages are gained in model specification, estimation and in forecasting. Thus it is important to test for the IIA property and the next section discusses six tests for IIA.

3. Tests for Independence of Irrelevant Alternatives.

Tests for the IIA property fall into two categories. Those based upon choice set partitioning and those based upon comparing the Logit model with an alternative model specification which does not embody the IIA property. Although there is little agreement in the literature as to which procedure is best, it does appear that, particularly in applied work, choice set partitioning tests dominate. In this section we describe four choice set partitioning tests and two tests based upon an alternative model specification (the DOGIT model of Gaudry and Dagenais (1979)) for the IIA property.

3.1 Choice Set Partitioning Tests.

The idea behind choice set partitioning tests is simple: if the IIA property is valid then the model structure and parameters are unchanged when choice is analyzed conditional on a restricted subset of the full choice set. Therefore, such tests are based upon seeing whether the estimated parameters or maximized log-likelihoods from the full choice set C (= {1, ..., J}) and a proper subset D of the full choice set are significantly different from each other. The choice set partitioning tests we consider are the Hausman-McFadden (HM) test (Hausman and McFadden (1984)), the McFadden-Train-Tye (MTT) test (McFadden, Train and Tye (1981)), the Horowitz (H) test (Horowitz (1981)) and the Small-Hsiao (SH) test (Small and Hsiao (1985)). The first of these, HM, is a Hausman specification test and the other three, MTT, H and SH, are variants of a likelihood ratio test.

Probably the most widely applied test for IIA is the HM test. The HM test statistic is based upon the fact that if IIA holds the model structure will be invariant to whether the parameter estimates are obtained from the full choice set C or from a restricted subset, D, of this choice set. That is consistent estimates of δ in (3) are obtained by maximizing the log-likelihood of the Logit model either over the full choice set C, yielding $\hat{\delta}_C$, or over the subset D, yielding $\hat{\delta}_D$. The test is a Hausman test (see Hausman (1978)) with two estimators being employed. One of these $\hat{\delta}_D$ is both consistent and efficient under the null hypothesis of IIA but inconsistent under the alternative that IIA does not hold. The other estimator $\hat{\delta}_C$ is consistent under *both* the null and alternative but inefficient under the alternative. The HM test statistic is given by:

$$HM = (\hat{\delta}_D - \hat{\delta}_C)'\Omega^{-}(\hat{\delta}_D - \hat{\delta}_C) = \hat{q}'\Omega^{-}\hat{q}, \tag{5}$$

where Ω^{-} is the generalized inverse of the asymptotic covariance matrix of $\hat{q}$. Asymptotically the test follows a chi-squared distribution with degrees of freedom equal to the rank of Ω^{-}.

Hausman and McFadden show that Ω^{-} is equivalent to $(\Omega_D - \Omega_C)^{-}$, the generalized inverse of the difference of the asymptotic covariance matrices of $\hat{\delta}_D$ and $\hat{\delta}_C$ respectively. Three problems arise with this test statistic. Firstly, that HM is not bounded to be positive in finite samples (Hausman and McFadden (1984) footnote 4 suggests negative values support the null). Secondly, not all elements of δ are identified over the subset D - thus the test statistic must be calculated for the identifiable component of δ and thirdly it is not clear *how* we are to choose the alternatives in C to include in the restricted subset D . We should note that the second and third problems are not unique to the HM test but will arise with all choice set partitioning tests.

Our other three choice set partitioning tests (MTT, H and SH) are variants of a likelihood ratio test. Again the basic idea is that, if IIA holds, the structure of the model and hence the maximized log-likelihood does not change if we analyze data on choice set C or data conditional on the restricted subset D. If this is true then we can base a test upon the difference between the maximized log-likelihood of the model for the restricted subset D evaluated at the maximum likelihood estimates obtained over the full choice set C and those obtained over the subset D.

The MTT test consists of estimating the Logit model by maximum likelihood on the full choice set C to obtain $\hat{\delta}_C$. The Logit model is then estimated using data over the subset D to obtain $\hat{\delta}_D$. The log-likelihood for this 'restricted' estimation is labelled $log\ L_1$ and its maximized value is $log\ L_1(\hat{\delta}_D)$. The MTT test then compares $log\ L_1(\hat{\delta}_D)$ with with $log\ L_1$ evaluated at the full choice set estimates $\hat{\delta}_C$ (i.e. with $log\ L_1(\hat{\delta}_C)$). Thus the MTT test statistic is given by:

$$MTT = -2(log\ L_1(\hat{\delta}_C) - log\ L_1(\hat{\delta}_D)). \tag{6}$$

McFadden, Tye and Train (1981) show that this test statistic has an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of $\hat{\delta}_D$ (i.e. the identifiable component of δ).

Horowitz (1981) noted that as $\hat{\delta}_C$ and $\hat{\delta}_D$ are estimated from overlapping samples the MTT test statistic is asymptotically biased towards accepting the null hypothesis of IIA. His suggestion is to randomly divide the sample into two asymptotically equal parts A and B with sample sizes n_A and n_B and to use these two samples to avoid the overlap problem. The construction of the Horowitz (H) test statistic is as follows. Firstly, the Logit model is estimated by maximum likelihood for data over the full choice set C in the sub-sample of n_A observations. This yields the estimates $\hat{\delta}_C^A$. The sub-sample A is now discarded and the Logit model estimated by maximum likelihood using data over the subset D in the sub-sample of n_B observations This estimation yields $\hat{\delta}_D^B$ and the maximized likelihood $log\ L_1(\hat{\delta}_D^B)$. The H test statistic is then calculated as:

$$H = -2(log\ L_1(\hat{\delta}_C^A) - log\ L_1(\hat{\delta}_D^B)). \tag{7}$$

This test statistic follows an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of $\hat{\delta}_D^A$ (i.e. the identifiable component of δ).

Since it is based upon independent samples the H test avoids the overlapping samples problem of the MTT test. Small and Hsiao (1985) prove the aforementioned bias of the MTT test but their result also shows that by using strictly independent samples the H test will be asymptotically too large. Hence the H test is biased towards rejecting the null hypothesis of IIA. In their paper they propose a test (SH) that combines the MTT and H test procedures and is free of any asymptotic bias. The SH test procedure involves randomly dividing the sample into two asymptotically equal parts A and B with sample sizes n_A and n_B. The Logit model is then estimated by maximum likelihood in each sub-sample over the data for the full choice set C. These estimations yield estimates $\hat{\delta}_C^A$ and $\hat{\delta}_C^B$ which are then combined in a weighted average:

$$\hat{\delta}_C^{AB} = (1/\surd 2)\hat{\delta}_C^A + (1 - 1/\surd 2)\hat{\delta}_C^B.$$

The first sub-sample A is then discarded and the Logit model is estimated by maximum likelihood in sub-sample B for data over the subset D. This yields $\hat{\delta}_D^B$ and the maximized likelihood $log\ L_1(\hat{\delta}_D^B)$. The SH test statistic is calculated as:

$$SH = -2(log\ L_1(\hat{\delta}_C^{AB}) - log\ L_1(\hat{\delta}_D^B)). \tag{8}$$

Again this test statistic follows an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of $\hat{\delta}_D^A$ (i.e. the identifiable component of δ).

3.2 Tests Against the DOGIT Model.

Another class of tests for the IIA property involves the specification of an alternative model which does not embody IIA. Typically such models are generalizations of the Logit model and IIA is tested using conventional tests for parameter restrictions. Such a generalization is the DOGIT model of Gaudry and Dagenais (1979). The selection probabilities for the DOGIT model are given by:

$$P_{ij} = \frac{\exp(V_{ij}) + \theta_j \sum_{k=1}^{J} \exp(V_{ik})}{\left(1 + \sum_{k=1}^{J} \theta_k\right) \sum_{k=1}^{J} \exp(V_{ik})} \qquad i = 1, ..., n; \; j = 1, ..., J,$$

with $\theta_j \geq 0 \; \forall \; j$.

A key feature of the DOGIT model is that it does not *a priori* impose the IIA property on all pairs of alternatives. The odds ratio P_{ij}/P_{ik} for $j \neq k = 1, ..., J$ in the DOGIT model is:

$$\frac{\exp(V_{ij}) + \theta_j \sum_{s=1}^{J} \exp(V_{is})}{\exp(V_{ik}) + \theta_k \sum_{s=1}^{J} \exp(V_{is})},$$

and since this ratio clearly depends upon all of the alternatives in the choice set the DOGIT model does not exhibit the IIA property.

The DOGIT model is convenient in that if $\theta_j = 0 \; \forall \; j = 1, ..., J$ the model collapses the the Logit model, which embodies the IIA property. Furthermore, if some, but not all, of the θ's are non-zero then IIA will hold between certain pairs of alternatives but not between other pairs of alternatives. As the Logit model is nested within the DOGIT model then a test for the Logit and hence for IIA can be carried out by testing the appropriate parameter restriction using a Wald, likelihood ratio or Lagrange multiplier (score) test procedure.

The most appealing test procedure to use is the score test as it only involves estimation under the null hypothesis of the Logit model (or equivalently the null hypothesis that IIA holds). Tse (1987) derives the score (LM) test of the IIA property using the DOGIT model as his alternative (non-IIA) specification. Using the parameterization of V_{ij} given in (3) above Tse partitions the parameter vector ϕ for the DOGIT model into two parts (i.e. $\phi = (\delta', \theta')'$). The score vector and information matrix are also partitioned in this way. The score test statistic tests the null of $\theta = \mathbf{0}$ against the alternative hypothesis that $\theta \neq \mathbf{0}$. The quadratic form for this test statistic requires the component of score vector with respect the the θ parameters and the inverse of its asymptotic covariance matrix and is given by:

$$LM = s(\theta)'V(\theta)^{-1}s(\theta). \tag{9}$$

Detailed expressions for $s(\theta)$ and $V(\theta)$ can be found in Tse (1987, p.284). This expression is evaluated at the parameter estimates obtained under the null hypothesis (i.e. under the Logit model estimates $\hat{\phi} = (\hat{\delta}', \mathbf{0}')'$) and the statistic LM has an asymptotic chi-squared distribution with J degrees of freedom under the null hypothesis.

The LM statistic ignores the inherent one-sided nature of the hypothesis test in this case. In other words, under the alternative DOGIT model $\theta > \mathbf{0}$. King and Wu (1993) extend the locally most mean powerful (LMMP) test procedure of SenGupta and Vermeire (1986) to such one-sided testing situations. The LMMP test statistic turns out to be nothing more than a standardized sum of scores (see King and Wu (1993)). Thus the LMMP test statistic for testing the null of IIA ($\theta = \mathbf{0}$) against the DOGIT model is given by:

$$LMMP = \frac{\iota' s(\theta)}{\sqrt{\iota' V(\theta) \iota}}, \tag{10}$$

where ι is a $J \times 1$ vector of ones. Again this statistic is evaluated at the estimates obtained under the null Logit model. The LMMP statistic has an asymptotic standard normal distribution under the null and tests are one-tailed tests.

In the following section of this paper we will consider the performance of the six tests (HM, MTT, H, SH, LM and LMMP) for IIA discussed above using Monte Carlo simulation methods.

4. A Monte Carlo Study.

An integral part of any Monte Carlo study is the design of the experiment(s) as the results obtained are conditional upon the design used. In all our experiments the number of alternatives J is, for simplicity, set to three. The variant of the Logit model considered is the MNL and the vector of explanatory variables $\mathbf{X}_i$ comprises of a constant and two independent drawings from the standard normal distribution. Once generated, however, the $\mathbf{X}_i$ vector was then fixed for the duration of the simulation experiments. In all experiments three sample sizes n are considered (250, 500, 1000), which range for small to moderate for cross section data, and N = 1000 replications are performed. All computations are undertaken using GAUSS and the add-in QUANTAL RESPONSE unit.

Our first set of simulations are designed to investigate the size properties of the six IIA tests. In these experiments the underlying generating process for the observed data is assumed to be the random utility maximization model given in (2) above with $\alpha = \mathbf{0}$ and the errors, ε_{ij}, being drawings from independent, identically distributed, Extreme Value distributions. The choice of the β_j (j = 1,2,3) vectors is as follows: $\beta_1 = (\gamma\ 1\ -1)'$, $\beta_2 = (\gamma\ -1\ 1)'$ and $\beta_3 = (0\ 0\ 0)'$, where γ is Euler's constant ($\cong$ 0.577216). These coefficients yield an expected split across the three alternatives of 45:40:15. That is the third alternative is chosen relatively infrequently. Furthermore, not only did these coefficients yield an acceptable split but they are attractive in their simplicity. Note also that β_3 is normalized to zero to identify the MNL model.

For each replication in our size experiment drawings from independent identically distributed Extreme Value distributions (the ε_{ij}'s) are added to the known V_{ij}'s to obtain U_{ij}'s. The observable variable, y_i, is then coded as 1,2, or 3 depending upon which value of U_{ij} ($j = 1,2,3$) is the maximum. A MNL model is then estimated and the six IIA tests are carried out. The tests are conducted at three significance levels 1%, 5% and 10%. The results of the size experiment can be found in Table 1. We should note that for the choice set partitioning tests there are three possible choices for the subset D. These are obtained by deleting alternative 1, 2 and 3 respectively. The corresponding test statistics for the Hausman-McFadden test are labelled HM1, HM2 and HM3. The same nomenclature is adopted for the other tests (MTT, SH and H).

Table 1 shows that the HM test has erratic size properties but tends to be oversized at the 1% and 5% levels and for HM3, where the less frequently chosen alternative is deleted. The results for MTT, H and SH are consistent with the underlying asymptotic theory. Namely that for all sample sizes MTT is biased towards the null and the H tests over-rejects the null. The SH test is correctly sized (lying between MTT and H) especially as the sample size n increases. The results for the LM and LMMP tests against a DOGIT specification suggest that for all sample sizes these tests are severely undersized.

Since most of our tests are incorrectly sized when we evaluate their power properties we 'size correct' by using empirical critical values from our size experiment. That is we use as a critical value in our power Monte Carlo experiments the value of the test statistic above which 5% of calculated values in our 1000 replications lie. This ensures that all of the tests start from a level playing field in that their differing size properties do not shroud conclusions about their power properties.

To ascertain empirical power properties it is necessary to generate the observable data y_i under the alternative hypothesis that IIA does not hold. We conduct two sets of power Monte Carlo experiments and in both use size corrected critical values at the 5% size. The two alternative non-IIA specifications which we consider are the DOGIT model and a multinomial probit (MNP) model. The β_j vectors remain at the values fixed in the size simulations.

A problem which we face when attempting to generate data consistent with a DOGIT model is the choice of suitable values for the θ_j ($j = 1,2,3$) parameters. The DOGIT model has not found much use in empirical work - indeed to our knowledge only two applications have appeared (Gaudry and Wills (1979) and Gaudry (1980)). These applications are for two alternatives and suggest that for cross section data values close to 0 (e.g. 0.04) might be expected for the θ's. In order to obtain further information on the possible range of *plausible* θ values it is useful to consider the generation of the DOGIT model from a sequential choice process (see Hensher and Johnson (1981) and Bordley (1990)).

Suppose that an individual is either 'captive' to one of the J alternatives or chooses freely from the full choice set C. In the first stage of the choice process we determine whether an individual is captive or exercises free choice. This can be represented as a random utility maximization model with J+1 alternatives {1, ..., J+1} (={1, ..., J, C}) and utilities given by:

$$U_{ij}^{(1)} = V_{ij}^{(1)} + \varepsilon_{ij}^{(1)}, \qquad i = 1, ..., n; \; j = 1, ..., J+1.$$

Under the assumptions that $\varepsilon_{ij}^{(1)}$ are independent identically distributed Extreme Value, that $V_{ij}^{(1)} = \log(\theta_j)$, $j = 1, ..., J$ and $V_{ij+1}^{(1)} = 0$ the probability of individual i being captive to alternative j is equal to:

$$P_{ij}^{(1)} = \frac{\theta_j}{1 + \sum_{k=1}^{J} \theta_k} .$$

For individual i exercising free choice the probability $P_{ij+1}^{(1)} = P_{ic}^{(1)}$ is given by:

$$P_{ic}^{(1)} = \frac{1}{1 + \sum_{k=1}^{J} \theta_k} .$$

If an individual exercises free choice then in the second stage of the choice process they will make selections according the the random utility maximization model in (2) with independent and identically distributed Extreme Value ε_{ij}'s (i.e according to a Logit model). If, on the other hand, they are captive to an alternative then then selection probability is determined in the first stage. The final probabilities will be given by:

$$P_{ij} = P_{ij}^{(1)} + P_{ic}^{(1)} \times \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})},$$

which is the DOGIT form given in section 3 above.

In our empirical power study we choose values of $\sum_k \theta_k$ which yield approximately 10% 'captive' choosers ($\sum_k \theta_k = 0.1$) and 50% 'captive' choosers ($\sum_k \theta_k = 1$). The individual θ_j's are chosen to move the split across alternatives away from the null of 45:40:15. Tables 2, 3 and 4 contain the results of these simulations[2]. From these we can see that typically the choice set partitioning tests have no

[2]The split across alternatives for one configuration of θ values and n=250 was such that for the choice set partitioning tests the sample sizes in sample B over choice set D was too small to estimate the model and hence the test statistics could not be computed.

power (power $\cong$ size) against the DOGIT alternative. The exception is the MTT3 test against a DOGIT model with $\theta_1 = \theta_2 = 0,\ \theta_3 \neq 0$ which has good power properties for $n = 500$ and $n = 1000$. This DOGIT model exhibits IIA between alternatives 1 and 2 but not between 1 and 3 or 1 and 2. By comparing estimates from $C = \{1,2,3\}$ and $D = \{1,2\}$ the MTT does seem to pick up the departure from the null hypothesis. We should also note that HM3, SH3 and H3 also appear to be able to detect the departure when $\theta_1 = \theta_2 = 0,\ \theta_3 = 1$ and $n = 1000$.

The tests designed against the DOGIT model (LM, LMMP) do best when $\sum_k \theta_k = 0.1$ and $\theta_j \neq 0,\ \forall\ j = 1,2,3$. They do not seem to have good power when only one θ_j is non-zero. These results also suggest that the one-sided LMMP test improves on the two-sided LM test in the cases in which the DOGIT based tests have power. Furthermore, in such cases the DOGIT based tests perform better than the choice set partitioning tests. It is interesting to note that these cases when $\sum_k \theta_k = 0.1$ correspond to the range of values of θ's found in the limited empirical work with cross section data quoted earlier.

One final point to note about the empirical powers for the DOGIT alternatives is that the two stage choice process described above also shows us that the θ_j's are not unconstrained as the proportion choosing alternative j in a sample must be greater than or equal to the proportion given by the 'captive' probability $P_{ij}^{(1)}$. This places an upper bound on the admissible θ_j values. The final column in Tables 2, 3 and 4 corresponds to a situation in which we are close to the upper bound and here the LM and LMMP tests have poor power properties.

Our second choice of alternative, non-IIA, specification is the multinomial probit (MNP) model. In this model the ε_{ij}'s in the random utility model given in (2) follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Ω. In our study we choose:

$$\Omega = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which is consistent with the MNP model used to find empirical rejection probabilities in Tse (1987). The results of this Monte Carlo experiment are found in Table 5 and we see that only MTT3 in the case of positive correlation has any power against the MNP alternative. These MNP results suggest that the six tests for IIA we have considered are unlikely to pick up the departure from the null suggested by the MNP model.

5. Conclusions.

In this paper we have taken six tests for the independence of irrelevant alternatives (IIA) property of Logit models for polychotomous choice situations and compared their size and power properties in small to moderate sample sizes using a set of Monte Carlo experiments. The six tests fall into two categories: choice set partitioning and tests against an alternative, non-IIA, model - the DOGIT model.

We find that the most popular test, the Hausman-McFadden test has poor size and power properties and is sensitive to the alternative dropped to form the subset D[3]. Of the other choice set partitioning tests the Small-Hsiao test has good size, but poor power, properties. The McFadden-Train-Tye test, when size corrected and the least popular alternative is dropped to form D (i.e. MTT3), has reasonable power properties against some variants of the DOGIT model and the multinomial probit model with positive correlation. The Horowitz test performs poorly in terms of both size and power.

---

[3]The poor performance of the HM test is partly explained by the high proportion of negative values calculated in our experiments. Further details are available on request from the authors.

Two tests against an ad-hoc generalization of the Logit model, the DOGIT, are considered. The DOGIT model is an attractive generalization as it potentially allows for IIA to hold for some, but not all, pairs of alternatives in the choice set. As the Logit model is nested within the DOGIT it also allows a test for IIA to be carried out using a score (LM) test and a one-sided variant of the score test (the LMMP test). We find that both these tests are severely undersized but have reasonable, size corrected, power, against certain parameterizations of the DOGIT model. As the DOGIT models against which the LM and LMMP tests have power are those which may be encountered in practice we recommend that researchers consider their use.

Finally, as with any set of Monte Carlo experiments, the results we obtained and the conclusions we draw are dependent upon the design of the experiments. It is possible that with a different experimental design different results might arise. Areas for future work should include the use of mixed Logit models and the consideration of the performance of a likelihood ratio test against a DOGIT specification. It would also be of interest to see the DOGIT model applied to economic data, which would give a better view of the likely values of θ to use in future simulation studies.

References.

Bordley, R.F. (1990), "The Dogit Model is Applicable Even Without Perfectly Captive Buyers", *Transportation Research*, **24B**, 315-323.

Crichton, N.J. and T.R.L. Fry (1992), "An Analysis of the Effect of an Offender's Employment Status on the Type of Sentence Chosen by the Magistrate", in P.G.M. van der Heijden *et al* (eds), *Statistical Modelling*, Elsevier Science, Amsterdam.

Fry, T.R.L., Brooks, R.D., Comley, B.R. and J. Zhang (1993), "Economic Motivations for Limited Dependent and Qualitative Variable Models", *Economic Record*, **69**, 193-205.

Gaudry, M.J.I. (1980), "Dogit and Logit Models of Travel Mode Choice in Montreal", *Canadian Journal of Economics*, **13**, 268-279.

Gaudry, M.J.I. and M.G. Dagenais (1979), "The Dogit Model", *Transportation Research*, **13B**, 105-112.

Gaudry, M.J.I. and M.J. Wills (1979), "Testing the Dogit Model with Aggregate Time-Series and Cross-Sectional Travel Data", *Transportation Research*, **13B**, 155-166.

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

Hausman, J.A. and D. McFadden (1984), "Specification Tests for the Multinomial Logit Model", *Econometrica*, **52**, 1219-1240.

Hensher, D.A. and L.W. Johnson (1981), *Applied Discrete Choice Modelling*, Croom-Helm, London.

Horowitz, J. (1981), "Identification and Diagnosis of Specification Errors in the Multinomial Logit Model", *Transportation Research*, **15B**, 345-360.

King, M.L. and P.X. Wu (1993), "Locally Optimal One-Sided Tests for Multiparameter Hypotheses", mimeo, Monash University.

McFadden, D. (1985), "Econometric Analysis of Qualitative Response Models", in Z. Griliches and M.D. Intrilligator (eds), *Handbook of Econometrics, Volume II,* North Holland, Amsterdam.

McFadden, D., Train, K. and W. Tye (1981), "An Application of Diagnostic Tests for the Independence of Irrelevant Alternatives Property of the Multinomial Logit Model", *Transportation Research Record*, **637**, 39-46.

SenGupta, A. and L. Vermeire (1986), "Locally Optimal Tests for Multiparameter Hypotheses", *Journal of the American Statistical Association*, **81**, 819-825.

Small, K.A. and C. Hsiao (1985), "Multinomial Logit Specification Tests", *International Economic Review*, **16**, 471-486.

Tse, Y.K. (1987), "A Diagnostic Test for the Multinomial Logit Model", *Journal of Business & Economic Statistics*, **5**, 283-286.

Table 1: Empirical Test Sizes for a Nominal Size of α.

	α=0.01			α=0.05			α=0.10		
Test	n=250	n=500	n=1000	n=250	n=500	n=1000	n=250	n=500	n=1000
HM1	0.022	0.038	0.053	0.030	0.055*	0.077	0.040	0.066	0.096*
HM2	0.016*	0.044	0.055	0.038	0.059*	0.081	0.051	0.079	0.103*
HM3	0.073	0.073	0.052	0.101	0.102	0.090	0.119	0.134	0.114*
MTT1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MTT2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MTT3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SH1	0.027	0.017*	0.013*	0.075	0.063*	0.049*	0.141	0.122	0.100*
SH2	0.023	0.023	0.012*	0.076	0.053*	0.059*	0.129	0.110*	0.106*
SH3	0.029	0.016*	0.016*	0.068	0.063*	0.055*	0.115*	0.119*	0.102*
H1	0.156	0.150	0.124	0.303	0.279	0.265	0.378	0.375	0.363
H2	0.145	0.142	0.126	0.284	0.285	0.262	0.398	0.375	0.367
H3	0.132	0.123	0.114	0.258	0.251	0.267	0.354	0.360	0.372
LM	0.004	0.006*	0.003	0.006	0.007	0.006	0.010	0.008	0.010
LMMP	0.003	0.003	0.002	0.005	0.006	0.005	0.006	0.013	0.008

* indicates that 95% confidence interval for estimated size contains nominal size.

Table 2: Size Corrected Power for n = 250 against the DOGIT Model.

Test	$\theta_1=0.06$ $\theta_2=0.03$ $\theta_3=0.01$	$\theta_1=0.0\dot{3}$ $\theta_2=0.0\dot{3}$ $\theta_3=0.0\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=0.10$	$\theta_1=0.60$ $\theta_2=0.30$ $\theta_3=0.10$	$\theta_1=0.3\dot{3}$ $\theta_2=0.3\dot{3}$ $\theta_3=0.3\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=1.00$
HM1	0.043	0.064	0.093	0.014	0.014	cannot compute
HM2	0.056	0.055	0.072	0.010	0.010	
HM3	0.049	0.051	0.012	0.007	0.019	
MTT1	0.178	0.072	0.051	0.088	0.019	
MTT2	0.039	0.059	0.038	0.000	0.003	
MTT3	0.015	0.038	0.387	0.000	0.000	
SH1	0.051	0.057	0.053	0.040	0.047	
SH2	0.048	0.043	0.057	0.033	0.043	
SH3	0.071	0.060	0.038	0.055	0.060	
H1	0.045	0.059	0.048	0.038	0.049	
H2	0.041	0.039	0.048	0.036	0.045	
H3	0.075	0.063	0.038	0.067	0.062	
LM	0.303	0.245	0.166	0.002	0.006	
LMMP	0.425	0.363	0.075	0.025	0.082	

Table 3: Size Corrected Power for n = 500 against the DOGIT Model.

Test	$\theta_1=0.06$ $\theta_2=0.03$ $\theta_3=0.01$	$\theta_1=0.0\dot{3}$ $\theta_2=0.0\dot{3}$ $\theta_3=0.0\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=0.10$	$\theta_1=0.60$ $\theta_2=0.30$ $\theta_3=0.10$	$\theta_1=0.3\dot{3}$ $\theta_2=0.3\dot{3}$ $\theta_3=0.3\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=1.00$
HM1	0.032	0.048	0.111	0.013	0.010	0.039
HM2	0.059	0.038	0.108	0.012	0.017	0.040
HM3	0.112	0.059	0.001	0.011	0.035	0.002
MTT1	0.229	0.087	0.122	0.092	0.015	0.000
MTT2	0.092	0.094	0.166	0.001	0.010	0.002
MTT3	0.031	0.036	0.617	0.000	0.000	0.999
SH1	0.047	0.056	0.061	0.050	0.046	0.049
SH2	0.073	0.067	0.069	0.046	0.052	0.066
SH3	0.062	0.061	0.045	0.057	0.055	0.352
H1	0.046	0.056	0.056	0.046	0.051	0.049
H2	0.070	0.064	0.070	0.053	0.054	0.070
H3	0.061	0.057	0.032	0.061	0.066	0.130
LM	0.510	0.456	0.430	0.000	0.011	0.032
LMMP	0.603	0.578	0.079	0.021	0.244	0.000

Table 4: Size Corrected Power for n = 1000 against the DOGIT Model.

Test	$\theta_1=0.06$ $\theta_2=0.03$ $\theta_3=0.01$	$\theta_1=0.0\dot{3}$ $\theta_2=0.0\dot{3}$ $\theta_3=0.0\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=0.10$	$\theta_1=0.60$ $\theta_2=0.30$ $\theta_3=0.10$	$\theta_1=0.3\dot{3}$ $\theta_2=0.3\dot{3}$ $\theta_3=0.3\dot{3}$	$\theta_1=0.00$ $\theta_2=0.00$ $\theta_3=1.00$
HM1	0.021	0.032	0.176	0.002	0.010	0.050
HM2	0.037	0.041	0.153	0.025	0.018	0.089
HM3	0.218	0.103	0.042	0.044	0.102	0.939
MTT1	0.268	0.090	0.266	0.092	0.007	0.018
MTT2	0.099	0.084	0.258	0.001	0.011	0.028
MTT3	0.074	0.039	0.833	0.000	0.004	1.000
SH1	0.058	0.063	0.081	0.057	0.045	0.070
SH2	0.059	0.054	0.060	0.039	0.050	0.053
SH3	0.084	0.066	0.094	0.059	0.063	0.640
H1	0.054	0.066	0.079	0.058	0.047	0.070
H2	0.055	0.052	0.051	0.041	0.052	0.053
H3	0.087	0.069	0.058	0.070	0.075	0.327
LM	0.631	0.505	0.546	0.000	0.002	0.025
LMMP	0.748	0.695	0.061	0.002	0.160	0.000

Table 5: Size Corrected Rejection Probabilities against Multinomial Probit.

	$\rho=0.50$			$\rho=-0.50$		
Test	n=250	n=500	n=1000	n=250	n=500	n=1000
HM1	0.066	0.084	0.125	0.054	0.038	0.027
HM2	0.058	0.067	0.121	0.061	0.039	0.044
HM3	0.020	0.008	0.004	0.041	0.072	0.123
MTT1	0.049	0.097	0.180	0.081	0.086	0.090
MTT2	0.039	0.101	0.197	0.065	0.090	0.090
MTT3	0.318	0.515	0.714	0.020	0.016	0.024
SH1	0.071	0.050	0.066	0.043	0.040	0.065
SH2	0.049	0.067	0.063	0.039	0.052	0.047
SH3	0.043	0.045	0.058	0.041	0.038	0.050
H1	0.059	0.051	0.061	0.041	0.038	0.061
H2	0.048	0.070	0.055	0.040	0.051	0.046
H3	0.038	0.031	0.044	0.047	0.040	0.057
LM	0.032	0.039	0.057	0.006	0.008	0.008
LMMP	0.000	0.000	0.000	0.008	0.006	0.006