Tests of IIA: Details

1. Hausman-McFadden Test on a subset of alternatives.

- Estimate logit model twice:

  a. on full set of alternatives
  b. on a specified subset of alternatives (and the subsample with choices from this subset)

- If IIA holds, the two sets of estimates should not be statistically different: Let β_b denote the estimates obtained from setup b. above, and Ω_b denote their estimated covariance matrix. Let β_a denote the estimates of the same parameters obtained from setup a. above, and Ω_a denote their estimated covariance matrix. (Some parameters that can be estimated in setup a. may not be identified in setup b, in which case β_a refers to estimates under setup a. of the subvector of parameters that are identified in both setups.) Then, the quadratic form

$$(\beta_a - \beta_b)'(\Omega_b - \Omega_a)^{-1}(\beta_a - \beta_b)$$

  has a chi-square distribution when IIA is true. In calculating this test, one must be careful to restrict the comparison of parameters, dropping components as necessary, to get $\Omega_b - \Omega_a$ non-singular. When this is done, the degrees of freedom of the chi-square test equals the rank of $\Omega_b - \Omega_a$.

Reference: Hausman-McFadden, <u>Econometrica</u>, 1984.

2. McFadden omitted variables test.

- Estimate the basic logit model, using all the observations.

- Suppose A is a specified subset of alternatives. Create new variables in one of the following three forms:

a. If x_{in} are the variables in the basic logit model, define new variables

$$z_{in} = \begin{cases} x_{in} - \left(\sum_{j\in A} P_{jn}x_{jn}\right) / \left(\sum_{j\in A} P_{jn}\right) & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

where P_{jn} is calculated from the basic estimated model. The variables z_{in} can be written in abbreviated form as $z_{in} = \delta_{iA}(x_{in} - x_{An})$, where $\delta_{iA} = 1$ iff $i \in A$ and $x_{An} = \sum_{j\in A}P_{jn|A}x_j$, and $P_{jn|A}$ is the conditional probability of choice of j given choice from A, calculated from the base model.

b. If $V_{in} = x_{in}\beta$ is the representative utility from the basic model, calculated at the basic model estimated parameters, define the new variable

$$z_{in} = \begin{cases} V_{in} - \left(\sum_{j\in A} P_{jn}V_{jn}\right) / \left(\sum_{j\in A} P_{jn}\right) & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

or more compactly, $z_{in} = \delta_{iA}(V_{in} - V_{An})$.

c. Define the new variable

$$z_{in} = \begin{cases} -\log\left(P_{in} / \sum_{j\in A} P_{jn}\right) & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

where P_{in} is calculated using the basic model estimates.

- Estimate an expanded model that contains the basic model variables plus the new variables z_{in} , and carry out a LR test that the coefficients of z_{in} are zero:

  LR = 2[(Log likelihood with z's) - (Log likelihood without z's)]

  If IIA holds, then this statistic has a chi-square distribution with one degree of freedom.

Properties:

- The test using variables of type a. is equivalent to the Hausman-McFadden test for the subset of alternatives A.

- The test using variables of type b. is equivalent to a one-degree-of-freedom Hausman-McFadden test focused in the direction determined by the parameters β. It is likely to have greater power than the previous test if there is substantial variation in the V's across A.

- The test of type c. is equivalent to a test of the basic MNL model against a nested MNL model in which subjects discriminate more sharply between alternatives within A than they do between alternatives that are not both in A. One plus the coefficient of the variable can be interpreted as a preliminary estimate of the inclusive value coefficient for the nest A.

- The tests b. and c. are closely related. The expression $-\log(P_{in} / \sum_{j\in A} P_{jn})$ that appears in the type c. variable can be rewritten

$$-\log\left(P_{in} \;/\; \sum_{j\in A} P_{jn}\right) = -V_{in} + \log\left(\sum_{j\in A} e^{V_{jn}}\right) = -\left(V_{in} - V_{An}\right) + \log\left(\sum_{j\in A} e^{V_{jn}-V_{An}}\right)$$

  The first term on the right-hand-side of this expression coincides with the type b. variable, except for a change of sign. The last term is a constant if the V's are uniform in A. Making a Taylor's expansion in powers of $V_{in} - V_{An}$, one finds that the first-order term vanishes, and variation appears only in the second-order term. Thus, type b. and type c. tests will tend to give similar results except in the case that there is large variation across the V's within A.

- The tests described above are for a single specified subset A. However, it is trivial to test the MNL model against several nests at once, simply by introducing an omitted variable for each suspected nest, and testing jointly that the coefficients of these omitted variables are zero. Alternative nests in the test be overlapping. The coefficients on the omitted variables provide some guide to choice of nesting structure if the IIA hypothesis fails.

- If there are subset-A-specific dummy variables in the basic model, then some of the omitted type a. variables duplicate these variables, and cannot be used in the testing procedure.

- One may get a rejection of the null hypothesis either if IIA is false, or if there is some other problem with the model specification, such as omitted variables or a failure of the logit form due, say, to asymmetry or to fat tails in the disturbances.

- Rejection of the IIA test will often occur when IIA is false, even if the nest A does not correctly represent the pattern of nesting. However, the test will typically have greatest power when A is a nest for which an IIA failure occurs.

Reference: D. McFadden, "Regression based specification tests for the multinomial logit model" Journal of Econometrics, 1987.

References for Tests of IIA

McFadden, D., K. Train, and W. Tye, 1976, "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," *Transportation Research Record*, No. 637, pp. 39-45.

Hausman, J. and D. McFadden, 1984, "Specification Tests for the Multinomial Logit Model," *Econometrica*, Vol. 52, No. 5, pp. 1219-1240.

Small, K. and C. Hsiao, 1985, "Multinomial Logit Specification Tests," *International Economic Review*, Vol. 26, No. 3, pp. 619-627.

McFadden, D., 1987, "Regression-Based Specification Tests for the Multinomial Logit Model," *Journal of Econometrics*, Vol. 34, No. 1/2, pp. 63-82.

Train, K., M. Ben-Akiva, and T. Atherton, 1989, "Consumption Patterns and Self-Selecting Tariffs," *Review of Economics and Statistics*, Vol. 71, No. 1, pp. 62-73.