Econ. 240B, Second Half   Daniel McFadden, 1999

EXERCISE 0. HYPOTHESIS TESTING IN A MNL MODEL FITTED BY MLE

(Not to be handed in, for discussion the week of Oct. 12 and 19)

This exercise estimates a multinomial logit model by maximum likelihood, and tests several hypotheses using Wald, LR, and LM tests. The application is discrete choice of travel mode for shopping, with bus and auto alternatives, or the alternative of staying at home. The data are artificially generated. The SST program ex0.cmd is in the class data directory, and can be run using the unix command sst ex0. Output from one run of this program, which describes the hypotheses and test procedures, is reproduced below. Using SST, or transcribing this program into TSP, carry out the tests requested below. This exercise may be helpful in preparing for the midterm and for practical understanding of the lectures on discrete response models.

```
SST Spool File:  ex0.out
Wed Sep 29 07:05:06 1999

AN EXAMPLE OF MAXIMUM LIKELIHOOD ESTIMATION
AND HYPOTHESIS TESTING FOR A MULTINOMIAL LOGIT MODEL


GENERATE SOME MADE-UP DATA, SHOPPING CHOICE OF (1) BUS,
(2) AUTO, OR (3) HOME (NO TRIP)

     range obs[1-400]
     set u = nrnd
     set tt2= 60 + 10*u + 5*nrnd
     set tt1= 90 + 5*u + 5*nrnd
     set tc2=2+0.04*tt2-.5*nrnd
     set tc1=1.25+0.01*tt1-.5*nrnd
     set u = log(-log(urnd))
     set y1=-0.1*tt1-tc1+10+ log(-log(urnd)) - u
     set y2=-0.1*tt2-tc2+10+ log(-log(urnd)) - u
     set one=1
     set z=0
     set i=1*(y1>vmax(y2,0))+2*(y2>=vmax(y1,0))+3*(0>vmax(y1,y2))
     label var[tt1] lab[travel time bus]
     label var[tt2] lab[travel time auto]
     label var[tc1] lab[travel cost bus]
     label var[tc2] lab[travel cost auto]
     label var[i] lab[choice] val[1 bus 2 auto 3 home]
     cova var[tt1 tt2 tc1 tc2] cov
```

```
Variable:      tt1     travel time bus

Mean                   90.23568      Standard deviation      7.07601
Minimum                70.83353      Skewness               -1.56729e-002
Maximum                 1.13171e+002 Kurtosis                2.96962
Valid observations       400

Variable:      tt2     travel time auto

Mean                   60.29345      Standard deviation     11.58801
Minimum                19.05803      Skewness               -0.12238
Maximum                 1.02718e+002 Kurtosis                3.41181
Valid observations       400

Variable:      tc1     travel cost bus

Mean                    2.14241      Standard deviation      0.51380
Minimum                 0.61633      Skewness               -0.14033
Maximum                 3.72038      Kurtosis                2.98541
Valid observations       400

Variable:      tc2     travel cost auto

Mean                    4.42621      Standard deviation      0.70304
Minimum                 2.48229      Skewness                2.39518e-002
Maximum                 6.53560      Kurtosis                2.80964
Valid observations       400

Correlation and Covariance matrix

         tt1              tt2              tc1
tt1     49.94471         59.03648         0.51180
tt2      0.72179          1.33946e+002    0.18747
tc1      0.14113          3.15654e-002    0.26333
tc2      0.52834          0.72451        -1.57086e-002

                   tc2
          tt1      2.62176
          tt2      5.88772
          tc1     -5.66009e-003
          tc2      0.49303

freq var[i]
```

```
i    choice
400 valid observations

                   bus        auto        home
                   1           2           3
                   ----------  ----------  ----------
Count               57         143         200
Percent           14.25       35.75       50.00


THIS COMPLETES GENERATION OF MADE-UP DATA


ESTIMATION OF BASE MNL MODEL
  mnl dep[i] covmat[cvmat] coef[beta] \
  ivalt[tt: tt1 tt2 z tc: tc1 tc2 z d1: one z z d2: z one z]


  ********** MULTINOMIAL LOGIT **********
  Dependent variable: i

  Value        Label        Count        Percent
     1          bus          57          14.25
     2         auto         143          35.75
     3         home         200          50.00

ITERATION  1:   OLD LLF =      -4.39445e+002     STEP =            1.29693
NEW LLF =      -3.11731e+002     GRAD*DIREC =      2.30453e+002

ITERATION  2:   OLD LLF =      -3.11731e+002     STEP =            0.99416
NEW LLF =      -3.10431e+002     GRAD*DIREC =     2.61166

ITERATION  3:   OLD LLF =      -3.10431e+002     STEP =            1.00326
NEW LLF =      -3.10429e+002     GRAD*DIREC =      3.90115e-003

  At convergence grad * dir =       1.48498e-009


  Independent        Estimated             Standard              t-
   Variable         Coefficient             Error              Statistic

      tt          -8.57271e-002          1.50024e-002         -5.71422
      tc          -1.25794               0.20111              -6.25483
      d1           9.07420               1.28205               7.07790
      d2          10.17796               0.99883              10.18987
```

```
auxiliary statistics              at convergence      initial
log likelihood                       -310.43          -439.44
number of observations                   400
percent correctly predicted               64

calc ll0=_llk
matrix cvmat
           [  1]             [  2]             [  3]
[  1]      2.25072e-004     -1.05793e-003     -1.82733e-002
[  2]     -1.05793e-003      4.04472e-002      1.66843e-002
[  3]     -1.82733e-002      1.66843e-002      1.64365
[  4]     -8.68450e-003     -0.11046           1.00648

                [  4]
     [  1]     -8.68450e-003
     [  2]     -0.11046
     [  3]      1.00648
     [  4]      0.99766
```

CONSIDER THE FOLLOWING HYPOTHESES

H1: beta_tt = 0.1*beta_tc

H2: beta_d1 = beta_d2

h3: beta_d1 = beta_d2 = 10

HYPTHESIS H1 TESTS WHETHER TRAVEL TIME IS VALUED AT 10 CENTS/MINUTE OR $6/HOUR. IT IS A ONE-DIMENSIONAL HYPOTHESIS.

HYPOTHESIS H2 TESTS WHETHER THE ALTERNATIVE-SPECIFIC EFFECTS FOR AUTO AND BUS ARE THE SAME. THIS WILL BE TRUE IF THESE REFLECT THE COMMON PAYOFF TO SHOPPING, BUT FALSE IF THERE IS ANY SYSTEMATIC DIFFERENCE IN THE ATTRACTIVENESS OF AUTO AND BUS OTHER THAN TRAVEL TIME AND COST DIFFERENCES.

HYPOTHESIS H3 TESTS WHETHER THE ALTERNATIVE-SPECIFIC EFFECTS FOR AUTO AND BUS ARE BOTH EQUAL TO 8. THIS IS A 2-DIMENSIONAL HYPOTHESIS

TESTING H1. FIRST, BY REPARAMETERIZATION, THIS CAN BE STATED AS A HYPOTHESIS THAT A PARAMETER IS ZERO. DEFINE A NEW VARIABLE, GENERALIZED COST gc = tc + tt/10 , AND ESTIMATE THE MODEL BELOW. IF H1 IS TRUE, THEN THE TRUE COEFFICIENT ON tt IS ZERO, AND A T-TEST ON THIS COEFFICIENT IS A TEST OF H1. FOR COMPARISON WITH OTHER TEST STATISTICS, THE SQUARE OF THIS T-STATISTIC IS ALSO COMPUTED.

```
     set gc1 = tc1+tt1/10
     set gc2 = tc2+tt2/10
     mnl dep[i] ivalt[tt: tt1 tt2 z gc: gc1 gc2 z d1: one z z d2: z one z] \
            coef[beta1] covmat[cvmat1]


********** MULTINOMIAL LOGIT **********
Dependent variable: i

Value        Label        Count        Percent
  1           bus          57           14.25
  2          auto         143           35.75
  3          home         200           50.00

ITERATION  1:   OLD LLF =      -4.39445e+002     STEP =            1.29693
NEW LLF =      -3.11731e+002     GRAD*DIREC =      2.30453e+002

ITERATION  2:   OLD LLF =      -3.11731e+002     STEP =            0.99416
NEW LLF =      -3.10431e+002     GRAD*DIREC =     2.61166

ITERATION  3:   OLD LLF =      -3.10431e+002     STEP =            1.00326
NEW LLF =      -3.10429e+002     GRAD*DIREC =      3.90115e-003

  At convergence grad * dir =       1.48498e-009


  Independent        Estimated            Standard                t-
   Variable         Coefficient            Error              Statistic

     tt           4.00669e-002         2.90022e-002          1.38151
     gc          -1.25794              0.20111              -6.25483
     d1           9.07420              1.28205               7.07790
     d2          10.17796              0.99883              10.18986

  auxiliary statistics                at convergence      initial
  log likelihood                        -310.43          -439.44
  number of observations                    400
  percent correctly predicted                64

     calc beta1[1]^2/cvmat1[1,1]
           1.90857
```

SECOND, A WALD TEST CAN BE CALCULATED FROM THE BASE MODEL.
THE TEST STATISTIC IS ASYMP. CHI-SQUARED WITH 1 D.F. UNDER H1.

```
matrix A = {1; -0.1; 0; 0}
matrix A
             [  1]
  [  1]      1.00000
  [  2]     -0.10000
  [  3]      0.00000
  [  4]      0.00000

matrix beta'*A*inv(A'*cvmat*A)*A'*beta
             [  1]
  [  1]      1.90857
```

THIRD, A LR TEST CAN BE CALCULATED BY RUNNING THE MODEL UNDER H1. THE TEST STATISTIC IS AGAIN ASYMP. CHI-SQUARED WITH 1 D.F.

```
mnl dep[i] ivalt[gc: gc1 gc2 z d1: one z z d2: z one z] \
      prob[pb2 pb3]
```

```
********** MULTINOMIAL LOGIT **********
Dependent variable: i

Value        Label        Count        Percent
      1          bus          57           14.25
      2         auto         143           35.75
      3         home         200           50.00

ITERATION  1:   OLD LLF =      -4.39445e+002     STEP =            1.29878
NEW LLF =      -3.11608e+002     GRAD*DIREC =       2.30446e+002

ITERATION  2:   OLD LLF =      -3.11608e+002     STEP =            0.99555
NEW LLF =      -3.11390e+002     GRAD*DIREC =      0.43751

At convergence grad * dir =       1.73179e-004
```

Independent Variable	Estimated Coefficient	Standard Error	t-Statistic
gc	-1.01759	9.78282e-002	-10.40186
d1	10.04989	1.08799	9.23710
d2	10.09161	0.99991	10.09254

auxiliary statistics	at convergence	initial
log likelihood	-311.39	-439.44
number of observations	400	
percent correctly predicted	63	

```
calc lla=_llk
calc lr = 2*(ll0 - lla)
calc lr
      1.92212
```

FOURTH, A LM TEST CAN BE CALCULATED BY AUXILIARY REGRESSION CONSTRUCT THE SCORE FOR THE UNRESTRICTED MODEL AT THE RESTRICTED ESTIMATES. REGRESS 1 ON THIS SCORE. THEN LM IS THE SUM OF SQUARED FITTED VALUES.

```
set pb1 = 1 - pb2 - pb3      # fitted probabilities
set r1 = (i==1)-pb1          # generalized residuals
set r2 = (i==2)-pb2
set st = tt1*r1+tt2*r2       # score, tt
set sc = tc1*r1+tc2*r2       # score, tc
set sd1 = r1                 # score, d1
set sd2 = r2                 # score, d2
reg dep[(1)] ind[st sc sd1 sd2] pred[onehat]
```

```
********* ORDINARY LEAST SQUARES ESTIMATION *********

Dependent Variable:       (1)
```

Independent Variable	Estimated Coefficient	Standard Error	t-Statistic
st	1.54258e-002	1.50540e-002	1.02469
sc	-0.22701	0.19346	-1.17338
sd1	-0.93538	1.27653	-0.73275
sd2	8.49096e-002	0.99540	8.53022e-002

Number of Observations	400
R-squared	1.00000
Corrected R-squared	0.00000
Sum of Squared Residuals	3.98165e+002
Standard Error of the Regression	1.00273
Durbin-Watson Statistic	9.42493e-003
Mean of Dependent Variable	1.00000

```
calc sum(onehat^2)
      1.83473
```

THE HYPOTHESIS H1 IS TRUE, AND THE RESULTING TEST STATISTICS ARE SIMILAR. TEST ANOTHER H1 HYPOTHESIS THAT IS FALSE, SAY beta_tt = 0.05*beta_tc. THE STATISTICS IN THIS CASE ARE LIKELY TO ALL REJECT THE NULL, BUT WILL BE MORE SPREAD OUT IN VALUES.

FOR THE EXERCISE, FORM THE HYPOTHESES H2 AND H3 AND CALCULATE THE TEST STATISTICS FOR THEM USING THE PROCEDURES GIVEN ABOVE, MODIFIED AS NEEDED.