Department of Economics

ECONOMETRICS I

Professor William Greene
Phone: 212.998.0876
Office: KMC 7-90
Home page: people.stern.nyu.edu/wgreene
Email: wgreene@stern.nyu.edu
Course web page: people.stern.nyu.edu/wgreene/Econometrics/Econometrics.htm

Assignment 5

Estimation

I. Econometric Theory

There are 38 cells on American roulette wheel, denoted 00, 0, and 1-36. The probability that any one particular cell will be hit on a spin of the wheel is 1/38, unless the wheel is rigged. A scientist visiting Amarillo Slim's casino in Reno believes that the wheel is rigged against the positive numbers. Let θ be the true proportion of hitting 1-36 on a wheel at Slim's casino. In theory, $\theta = 36/38$. Suppose our scientist decides to conduct an experiment. He is going to watch N gamblers, i=1,…,N, and count the number of spins, K_i that it takes until the first hit occurs. For each gambler, the probability that they lose K_i times before they win is

$$\text{Prob}[\text{Spins} = K_i] = \theta (1 - \theta)^{Ki}.$$

The sample data will consist of $y = (K_1,\ldots,K_N)$. Formulate the log likelihood for estimation of θ and derive the MLE. Derive the likelihood equation and show how to solve it to estimat θ. Derive the asymptotic variance of the MLE. [Hint: $E[K_i] = (1-\theta)/\theta$.]

II. Maximum Likelihood Estimation

(**Hierarchical distribution**) An exponential regression model might be formulated as follows (this is called a "loglinear model"): Let y_i be the time until failure of some electronic component. A model that is often used for this phenomenon is the exponential model that we have used in class:

$$f(y_i) = \theta_i \exp(-\theta_i y_i),\ \theta_i > 0,\ y_i \geq 0.$$

We believe that the lifetime of parts depends on a certain other variable, x_i, such that

$$\theta_i = \exp(\beta_1 + \beta_2 x_i).$$

We are interested in estimation of the parameters β_1 and β_2 and in manipulation of the model after estimation.

a. Write out the conditional (on x) log likelihood function. (Note, the density does involve the exponential of an exponential function, so the log of the density will still involve an exponential.)

b. Show the likelihood equations (first order conditions) for estimation of β_1 and β_2. Define the vector $\mathbf{x}_i = [1,x_i]'$ and $\boldsymbol{\beta} = [\beta_1,\beta_2]'$. Then, show that this first derivative vector can be written in the form

$$\partial \log L/\partial\boldsymbol{\beta} \;=\; \Sigma_i\, d_i\mathbf{x}_i \;\text{ where }\; d_i \;=\; (1 - \theta_i y_i).$$

It will also be convenient to write the gradient as $\partial \log L/\partial\boldsymbol{\beta} = \Sigma_i\, \mathbf{g}_i = \mathbf{g}$, where $\mathbf{g}_i = d_i\mathbf{x}_i$. It is now possible to show that the expected value of the first derivative vector is zero, as the theory requires. Explain, then do the proof. (It's trivial.)

c. We will also need the second derivatives. Show that

$$\partial^2 \log L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' \;=\; -\Sigma_i\, h_i\mathbf{x}_i\mathbf{x}_i' \;=\; -\Sigma_i\, \mathbf{H}_i = -\mathbf{H}, \;\text{ where }\; h_i \;=\; d_i - 1.$$

(Note that these values are all negative. It follows that the Hessian is a negative definite matrix.)

d. What is $-E[h_i]$? What is the asymptotic covariance matrix of the maximum likelihood estimator in this model?

An algorithm for estimation (that is, for finding the maximum likelihood estimator) in this model is Newton's method:

$$\mathbf{b}(k+1) \;=\; \mathbf{b}(k) \;-\; \mathbf{H}(k)^{-1}\mathbf{g}(k).$$

where "k" indicates the iteration, **b** is the estimator of $\boldsymbol{\beta}$ and **g** and **H** are the first derivative vector (the sum of terms) and Hessian (also sum of terms) of the log likelihood. This shows how one could locate the solution to the likelihood equations. Where should one begin the process? There are actually two natural candidates here. The first is (β_1=0, β_2=0). The second is a little more creative. Suppose $\beta_2 = 0$. Then, as we saw in class, the MLE of θ would be $1/\bar{y}$. In the model, if $\beta_2 = 0$, then $\beta_1 = \log\theta$, so an initial estimator would be $\log(1/\bar{y})$. You will be doing the estimation in the next part of the problem set. You might want to try both starting points. (Final observation. This is what is known as a 'globally concave log likelihood.' Because the Hessian is always negative definite, no matter what β_1, β_2, and x_i are, it makes no difference where you start the iterations, you will always end up at the same point (estimate).

e. The data listed below are generated by the model assumed above. You will use NLOGIT or any other program you wish to estimate the parameters using Newton's method. Here is a routine that will do the computations for you if you are using NLOGIT. As always, you may use a different computer program if you prefer to. Report your results.

```
READ the data using procedures that are familiar to you.
(Or, enter them in the data editor.)
Sample   ; 1 - 50 $
Namelist ; X = one,xi $
? Use zero and zero for starting values
Matrix ; c = [0/0] $
? Define a procedure that we can use over and over again.
Procedure
?  First derivative and second derivative, scalar part
Create ; gi = 1  - yi*exp(c(1) + c(2)*xi) ; hi = gi - 1 $
? This computes the (negative of) the second derivatives matrix
? and the first derivative vector then the next coefficient vector.
Matrix ; H = -X'[hi]X ; g = X'gi ; c = c + <H>*g ; List ; gt = g' $
EndProc
?  How many iterations will it take to converge?  Watch the first
?  derivatives, g.  When they get very small, we are done.
Execute ; n = 10 $
?  Display the results
Matrix  ; Stat (c,<H>,X) $
```

f. The mean value of x is .529271. The expected value of y_i is $1/\theta_i$. Use the delta method and your maximum likelihood results to obtain a confidence interval for this conditional mean function evaluated at the mean of x_i.

g. We are always interested in regression slopes. Since $E[y_i|x_i] = 1/\theta_i$, what is $\partial E[y_i|x_i]/\partial x_i$? Compute this value using the mean of the x's and your estimates and, once again, use the delta method to obtain an asymptotic standard error. Some researchers suggest that there is a better way to compute these marginal effects. The alternative is to compute the marginal effect for each observation separately, then average these separate observations. Note, in the first case, you are computing $\delta(\overline{x})$ while in the second, you are computing $(1/n)\Sigma_i\delta(x_i)$. Do this second computation and compare the two results.

h. Linearly regress y on x and report the least squares regression slope. How does this compare to the value you obtained for the marginal effect in *g*? Do you have an intuition for this result? What do you think is the explanation for your finding.

i. As in all regular maximum likelihood problems, there are three alternative estimators for the asymptotic variance of the MLE, the negative inverse of the actual Hessian that you used in part c, the negative inverse of the expected Hessian in part d. and the sum of outer products of the first derivatives, which in this case would be $\text{BHHH} = [\Sigma_i\, g_i^2\mathbf{x}_i\mathbf{x}_i']^{-1}$. (The BHHH stands for Berndt, Hall, Hall, and Hausman, the four econometricians who first suggested this estimator to the econometrics literature in 1971.) Compute all three estimates and compare the results.

j. We are now interested in testing the hypothesis that $\beta_2 = 0$. We will apply the 'trinity' of tests.
(1) Use a Wald test based on your earlier results. What do you conclude?
(2) The likelihood ratio test. You can compute the log likelihood function in part a. by plugging in your maximum likelihood estimates. Call this function $\log L_1$. As we showed in class, the maximum likelihood estimator of θ when there is no x in the model is $1/\bar{y}$, so the maximum likelihood estimator of β_1 will be $\log(1/\bar{y})$. You can compute the log likelihood function in part a. again by plugging in this value and 0.0 for γ. Call this $\log L_0$. Compare the two values. Which is larger. Explain. The likelihood ratio statistic is $\chi^2[1] = 2\times(\log L_1 - \log L_0)$. Compute the statistic and carry out the test.
(3) The Lagrange Multiplier Test. The logic of the LM test is to test whether the first derivatives of the log likelihood function are zero when they are evaluated at the restricted estimator (which we discussed in (2) just above). To compute the LM statistic, do the following: (a) Evaluate the first derivative vector (only two elements) in part b. and the second derivative in part c. at the restricted MLE, $[\log(1/\bar{y}),0]$ for $[\alpha,\gamma]$. (b) Use a Wald test to test the hypothesis that the gradient is zero. The statistic is

$$LM = \mathbf{g}_0' (-\mathbf{H}_0)\, \mathbf{g}_0.$$

This is also a chi-squared statistic with 1 degree of freedom. Carry out this test, and report your results.

I	Yi	Xi
1.	.066160	.946800
2.	1.120200	.264050
3.	.488590	.370610
4.	.092126	.939910
5.	2.139800	.019091
6.	1.126200	.302710
7.	1.015000	.904730
8.	.154100	.690400
9.	1.004700	.697850
10.	.154030	.312350
11.	.067202	.288070
12.	.310240	.925220
13.	.157290	.599210
14.	1.352700	.714170
15.	.055464	.653310
16.	.503090	.633900
17.	.128890	.629960
18.	.405520	.683460
19.	.285340	.465400
20.	1.444700	.569690
21.	.088804	.653300
22.	1.267100	.378620
23.	.284470	.366850
24.	.071322	.348690
25.	.312380	.037744
26.	.282310	.731110
27.	.407930	.598430
28.	.277310	.653960
29.	.134230	.336180
30.	.225810	.053202
31.	.150300	.323130
32.	1.144900	.501200
33.	.112230	.704040
34.	1.018200	.042694
35.	.230820	.774310
36.	.095829	.537350
37.	.414960	.966150
38.	.917050	.763620
39.	.134340	.967000
40.	2.274800	.154660
41.	.284190	.325830
42.	.183810	.244440
43.	.271540	.557660
44.	.065958	.125680
45.	.289010	.675920
46.	2.536700	.711490
47.	.143330	.677970
48.	1.082500	.177430
49.	1.329800	.846460
50.	.483260	.617530

```
? Maximum Likelihood Estimator. Parts a-e
SAMPLE   ; 1 - 50 $
NAMELIST ; X = one,xi $
? Use zero and zero for starting values
Matrix ; c = [0/0] $
? Define a procedure that we can use over and over again.
Procedure
?  First derivative and second derivative, scalar part
Create ; gi = 1  - yi*exp(c(1) + c(2)*xi) ; hi = gi - 1 $
? This computes the (negative of) the second derivatives matrix
? and the first derivative vector then the next coefficient vector.
Matrix ; H = -X'[hi]X ; g = X'gi ; c = c + <H>*g ; List ; gt = g'$
EndProc
?  How many iterations will it take to converge?  Watch the first
?  derivatives, g.  When they get very small, we are done.
Execute ; n = 10 $
?  Display the results
Matrix  ; Stat (c,<H>,X) $

? Part f.  Compute Mean at mean of x.
? For the delta method, we need the covariance matrix
Matrix ; V = <H> $
? Delta method for E[y|x=xbar].  We also list mean of y
Calc ; xbar = xbr(xi) $
Calc ; list ; thetabar = 1/exp(c(1)+c(2)*xbar) ; xbr(yi) $
Calc ; d1 = -1/thetabar^2 * thetabar ; d2 = d1*xbar $
Matrix ; d = [d1/d2] $
Calc ; List ; se = sqr(qfr(d,V))
     ; lower=thetabar-1.96*se ; upper = thetabar+1.96*se $
? Part g. Delta method for partial effect at mean.
Calc ; list ; mebar = c(2)*d1 $
Calc ; dm1 = -mebar ; dm2=(-1+c(2)*xbar)/thetabar $
Matrix ; dm = [dm1/dm2] $
Calc ; list ; se=sqr(qfr(dm,V)) $
? Partial effect at each observation, then averaged.
Create ; mei = -c(2)/exp(c(1)+c(2)*xi) $
Calc ; list ; xbr(mei)$
? Part h.  Compare linear regression slope to partial effects.
Regress;lhs=yi;rhs=one,xi$
? Part i.  Three variance estimators.
? So far, H is based on actual derivatives.
? But, hi = -yi*thetai and E[yi]=1/thetai, so E[hi]=-1.
? Therefore, covariance matrix based on E[H] is just <X'X>
? BHHH estimator uses squares of first derivatives.
Create ; gi2 = gi*gi $
Matrix ; List ; V ; EH = <X'X>  ; BHHH = <X'[gi2]X> $
? Part j.  Testing Hypothesis that beta(2) = 0.
Calc Wald test    ; List ; Wald = c(2)/sqr(v(2,2)) $
Calc LR statistic ; b01 = log(1/xbr(yi)) ; b02=0 $
Create            ; theta0=exp(b01 + b02*xi)$
Create     ; fi0 = theta0*exp(theta0*yi) ; logli0=log(fi0)$
Calc       ; logl0 = sum(logli0)$
Calc       ; b1 = c(1) ; b2=c(2) $
Create     ; thetai=exp(b1 + b2*xi)$
Create     ; fi = thetai*exp(thetai*yi) ; logli=log(fi)$
Calc       ; logl1 = sum(logli)$
calc       ; list ; LRstat = 2*(logl1 - logl0) $
Create     ; gi0 = 1  - yi*exp(b01 + b02*xi) ; hi0 = gi0 - 1 $
Matrix     ; H0 = -X'[hi0]X ; V0 = <H0> $
Matrix     ; list ; LM = gi0'X * V0 * X'gi0 $
```