UC Berkeley **Fall Semester 2004**

Economics 241A **Guido Imbens**

PROBLEM SET I, Part 3

Due: Wednesday February 11$^{\text{th}}$

1. (a) Suppose that $Y_1, \ldots, Y_N$ are independent and identically distributed random variables, with a distribution with hazard rate $h(t) = \lambda_0 + \lambda_1 \cdot 1\{t > 3\}$, where $1\{\cdot\}$ is the indicator function. Do a likelihood ratio test of the hypothesis that $\lambda_1 = 0$.

   (b) Check the second result in the Appendix to lecture notes 3.

2. Use the same data as in Part 1 of the problem set.

   (a) Discard all observations with durations less than 80. This should leave you with 3338 observations. What is the likelihood function, taking into account the (right) censoring and the (left) truncation, based on an exponential distribution with hazard rate λ?

   (b) Suppose the hazard function is, as in the last problem set, $\exp(-\beta_0 - \beta_1 \text{AGE} - \beta_2 \text{ED} - \beta_3 \text{WHITE} - \beta_4 \text{LOCRATE})$. Show that at $(1, 0, 0, 0, 0)$ the log likelihood function is -1.9839e+005.

   (c) Calculate the analytic first derivatives at this point.

   (d) Compare the analytic first derivatives to numerical first derivatives, using $c =$ 0.000001.

   (e) Compute the maximum likelihood estimates using the DFP algorithm with starting values $(1, 0, 0, 0, 0)$, and the identity matrix for A_0. At each step report β_k, p_k, d_k, q_k, λ_k and A_k. Make sure you do the normalization of the direction vector. Use the golden section method to calculate the optimal step length. In the step length calculate stop when the upper and lower bound are less than 0.00001 apart. Stop the main algorithm if the sum of the absolute values of the derivatives is less than 0.001.

   (f) Estimate the standard errors for the ml estimates using the average of the second derivatives to estimate the information matrix.

(g) Now we will consider testing the hypothesis H_0: $\beta_3 = \beta_4 = 0$, against the alternative hypothesis that either β_3 or β_4 (or both) differs from zero, at the 5% level. First estimate the parameters under the null hypothesis. using the DFP algorithm with starting values $(1, 0, 0)$, and the identity matrix for A_0. At each step report β_k, p_k, d_k, q_k, λ_k and A_k. Make sure you do the normalization of the direction vector. Use the golden section method to calculate the optimal step length. In the step length calculate stop when the upper and lower bound are less than 0.00001 apart. Stop the main algorithm if the sum of the absolute values of the derivatives is less than 0.001.

(h) Estimate the standard errors for the restricted ml estimates using the average of the second derivatives to estimate the information matrix.

(i) Next do the likelihood ratio test.

(j) Do the Wald test, using the second derivatives and estimates under the alternative to estimate the information matrix.

(k) Do the LM or score test, using the N times R^2 representation.