3. The SAS Macro for the Additive Hazards Regression Model

A SAS macro is stored text that performs one or more functions and is identified by a specified name. When a macro is invoked, or executed, values are passed from the user's program to the macro through variables called macro parameters.

The macro presented here is called "additive." The additive macro fits the additive hazards model for continuous or binary covariates. The test statistic for the global test of no effects is calculated, and an analysis of variance table is printed. Options such as testing linear contrasts, plotting parameter estimates over time, and creating output data sets are available. This macro is written in IML (SAS/IML 1989) which is a matrix language built into SAS.

Even though the additive hazards model can have time-dependent covariates, this macro is not equipped to handle them. Only fixed covariates should be used. Before any estimation begins, the macro will check for missing data. If an observation has a missing value for the event time, censoring indicator, or any of the covariates then that observation will be deleted and will not be used in any calculations. The output will contain a message stating which cases, if any, were deleted due to missing values.

This macro uses a data set specified by the user. In order for the program to run correctly, the data set must meet certain criteria. First, the data set must have the following ordering of its variables: Time, Censoring Indicator, Covariates. Second, the data set must be sorted by ascending time. Third, the censoring indicator needs to specified as follows:

0 - censored; 1 - event.

Finally, all categorical variables should be defined as binary variables. The user now needs to be in IML, which is done by issuing the statement "proc iml" in the program. The additive macro is included in the program via the "%include 'filename'" statement where 'filename' is the name of the additive macro file. Next the eight parameters are defined:

1. Data set parameter: name of user's data set.

2. Confidence level parameter: defines   for (1-) 100% confidence intervals.

.e. 0(574 3(V5433 02ed b)-22((f)38( )]TJ933 0 Tlist0.25 0 TDtive)]TJ10 -2.2 (3. Time uTw(the event)Tj -1.1 TD80.001 Tlis

parameters need to be included in the invocation statement in these positions because the positions in name-style invocation are positional.

The macro calculates the parameter estimates for each unique time. The parameters are only estimable in the range where the Y(t) matrix is of full rank. The macro checks if this matrix is of full rank at each time and stops estimating once the matrix is singular. The output will display a message that defines the range of estimability. Using the given data set, the macro will automatically calculate the chi-square statistic for the global test of no effects. The hypothesis is

$$H_{\ j}: \ _j(t) = 0 \quad \text{for all } j = 1, ..., p \quad \text{and all } t \quad .$$

The heading "Global Test" and the test statistic, degrees of freedom, and the corresponding p-value are printed. An analysis of variance table is printed that will list the individual effects and each effects' chi-square statistic, degrees of freedom, and corresponding p-value. The global test and analysis of variance table are the only items calculated and printed automatically. Whatever options specified by the user will then be processed.

4. Examples of The SAS Additive Hazards Regression Macro

Several examples will be given to illustrate the additive macro. The data used is from Kardaun (1983) who reported data on 90 males diagnosed with cancer of the larynx during the period 1970-1978 at a Dutch hospital. Times recorded are the intervals (in years) between first treatment and either death or the end of the study (January 1, 1983). Also recorded are the patients age in years at the time of diagnosis and the stage of the patient's cancer. The four stages of disease in the study were based on the T.N.M. classification used by the American Joint Committee for Cancer Staging (1972). The four groups are labeled Stage 1 through Stage 4, which is ordering the stages from least serious to most serious. Stage 1 is the baseline. The variables Stage2 through Stage4 are binary variables created to define stage. They are defined as follows:

Stage2= 1 if Stage 2 disease
0 if Stage 1, 3, or 4 disease

Stage3= 1 if Stage 3 disease
0 if Stage 1, 2, or 4 disease

Stage4= 1 if Stage 4 disease
0 if Stage 1, 2, or 3 disease.

The age variable, age, has been centered at its mean:

Age = age at diagnosis - 64.11.

The first example is the simplest possible program, one that defines "no" for all five options in the option parameter. Output follows the SAS program.

Example 1: Basic program

```
options nocenter pagesize=59 linesize=80;
 data cancer;
  infile 'larynx.dat';
  input stage time age year censor;
  stage2=0; if stage=2 then stage2=1;
  stage3=0; if stage=3 then stage3=1;
  stage4=0; if stage=4 then stage4=1;
  age=age-64.11;
 proc sort; by time;
* The following routine in proc iml creates the input SAS data set for this problem;
proc iml;
  use cancer;
  read all var _num_ into temdat;
   dummy=j(90,6,0);
  dummy[,1]=temdat[,2];   * time;
  dummy[,2]=temdat[,5];   * censor;
  dummy[,3]=temdat[,6];   * stage 2;
  dummy[,4]=temdat[,7];   * stage 3;
  dummy[,5]=temdat[,8];   * stage 4;
  dummy[,6]=temdat[,9];   * age-64.11;
  create mydata from dummy;
  append from dummy;
   quit;
proc iml;
        %include 'addmacro';
         level=0.05;
         unit={"Years"};
         varlist={"Stage 2", "Stage 3", "Stage 4" ,"Age"};
          option={n,n,n,n,n};
%additive(mydata,level,unit,varlist,option,dummy1,
dummy2,dummy3);
         quit;
```

The following output is the result from the first example:

```
The SAS System
             Additive Hazards Model

    No missing data: all observations were used in
analysis.
             90 observations used.

Estimates are restricted to the time interval 0  to   4.30

             Global Test
Chi-Square   d.f   p-value
10.9613        4    0.0270
```

2. Aalen, O.O. (1989) 0̂A linear regression model for the analysis of life time0̃,