Chapter 6

Conclusion

In this thesis a flexible parametric model, log hazard regression model for paired censored failure times, is proposed. In this model B-splines are used to estimate the log hazards of marginal failure times and the log hazards ratio of the two failure times. Consistent estimates of the standard errors for the estimated marginal log hazards and the log hazards ratio are presented as well. Based on this model we propose test procedures for the four hypotheses that the marginals follow an exponential or Weibull distributions and that the two failure times have the same distribution or have proportional hazards.

A simulation study indicates that when the censoring rate is not too high, the estimates of the log hazards and the log hazards ratio perform well within the range of the knots used and the estimates of standard errors for the estimated log hazards ratio are good, but a little bit small. This underestimation is bigger at time zero.

The simulation study also shows that the tests that the marginal failure times follow an exponential or Weibull distribution perform very well. But the test for the hypotheses that the two failure times have proportional hazards or have the same distribution tend to over-reject the null hypothesis when the null hypothesis is true. Also, the test is not powerful.

The model was applied to the Diabetic Retinopathy Study data to assess the effectiveness of laser photocoagulation in delaying the onset of blindness for diabetic retinopathy patients. The conclusion of the analysis with the log hazard regression model agrees with those from the standard parametric models fitted by Huster et. al (1989), that is, there is a significant laser photocoagulation treatment effect. We use the log hazard model to test the hypothesis that the failure times of the treated eye and untreated eye have an exponential or a Weibull distribution. The test results indicate that there is significant evidence to reject the hypotheses that the failure time of the treated eye follows either exponential or Weibull distribution. The test results also show that there is significant evidence to reject that the failure time of the untreated eye follows an exponential distribution. But the test that the failure time of the untreated eye follows a Weibull yields a marginal result. The p-value is 0.053. The p-value for testing the hypothesis that the failure times of the treated eye and the untreated are proportional hazards is very high, 0.25. This, along with the tendency of the test to over-reject (see Section 5.3), indicates that it is reasonable to assume that the failure times of the treated eye and the untreated eye have proportional hazards. This result supports the analysis of Huster et. al. (1989).

We note that knot selection is a basic step to use the log hazard regression model. The simulation study shows that our Quantile Knots procedure performs well when the censoring rate is not too high. Even when we restrict the range of the knots, the estimates are fairly reliable within the range of the knots used.

We can not always choose knots using Kooperberg et. al's stepwise addition and stepwise deletion method, since the chosen knots often cause numerical problems (see Section 5.2). However, the knots they selected will give us some indication of the number and the location of the knots that should be used. From the experience of our simulation, we give the following suggestions for using the log hazard regression models to analyze paired data:

1. Use the *heft* algorithm to choose a set of knots for each failure time as we mentioned in Section 3.4. If there is no warning message from the *heft* code and the calculation of estimated standard errors can be performed, then use the Different Knots method to estimate the log hazards ratio with the two sets of knots. If there is any warning message or the calculation of the estimated standard errors is impossible for one or both failure times, choose the knots for the one or both failure times by the quantiles method as we discussed in Section 3.4. Use the number of knots used by *heft*.

2. Test for exponential or Weibull. Fit the data with the log hazards model decided by the selected knots above. Then use the methods we mentioned in Section 3.5, to test if the marginals have exponential or Weibull distributions. If they do have those standard parametric distributions, use those distributions for modeling the marginal failure times.

3. To test proportional hazards, choose one set of knots for the two failure times. Use the quantiles of the non-censored failure times as discussed in Section 3.4. The number of knots should be between $\max\{K_1, K_2\}$ and $K_1 + K_2$, where K_1 and K_2 are the numbers of of knots selected by *heft* for T_1 and T_2, respectively. Then use the method given in Section 3.5 to test whether the two failure times have proportional hazards.

From the definition of our log hazard regression model, we can generalize the model to multivariate censored survival data easily. We also can include covariates in this model. If we assume that the effects of covariates are independent of time, then the calculation of the estimates is straightforward based on the *heft* algorithm. We can model the log hazard as a function of time plus a function of the covariate. If the effects of the covariates are dependent on time, then the log hazard would be modeled as a bivariate function of time and the covariate. However, the calculations of the estimates would be more complicated. To simplify calculations somewhat, Kooperberg et. al. modeled this bivariate function using linear B-spline and their tensor products.

There are some unanswered questions. First, why is the variability of the estimate of log hazard or log hazards ratio big at time zero? We might reduce the problem if we restrict the B-spline functions to be constant between zero and the smallest knot. Second, how do we choose the range of knots? As we see from our simulations, the estimates perform better within the range of the knots than outside of the range. We need more simulations to find a better way to determine the range of knots. Third, why do the test procedures for testing that the failure times have proportional hazard or have the same distribution work poorly, while the test procedures for testing that marginal failure times follows an exponential or Weibull distribution work well? We expect that a big sample size might improve the behavior of the test procedure. Finally, why does a high correlation level yield a better test for testing that failure times have proportional hazards? We need more simulations to understand the effect of the correlation level.