Chapter 3

Regression Space

In the previous chapter we discussed the regression model for log hazard functions for a general regression space G. Here, we will use a family of extremely flexible functions, cubic splines, as the regression space. This family was used by Kooperberg, Stone, and Truong in 1995 for univariate log hazard regression. In this chapter we first give a brief introduction to cubic splines. Then we give the definition of a restricted cubic spline regression space $\mathcal{B}$ and a method to construct a cubic B-spline basis of $\mathcal{B}$. In Section 3.3 we introduce the numerical computation method that we use to calculate the estimates of the log hazard regression model for paired failure data. In Section 3.4 we explain our methods for choosing knots for the log regression model.

Finally, in Section 3.5, we show that the log hazard regression model with the regression space $\mathcal{B}$ can be used for hypothesis testing. In the univariate case, we can test the hypotheses that the failure times have an exponential or a Weibull distribution. In the bivariate case, we can test the hypotheses that the two failure times have the same distribution or that the two failures have proportional hazards.

3.1 Cubic Splines

Definition 3.1.1 *The function ϕ is a* cubic spline on $[a, b]$ with knots $t_1, \cdots, t_K$ *($a < t_1 < \cdots < t_K < b$) if ϕ is a cubic polynomial on the subintervals $[a, t_1]$, $[t_1, t_2]$, $\cdots$, $[t_K, b]$ and ϕ has 2 continuous derivatives on $[a, b]$. Denote the collection of these splines by $Sp(t_1, \cdots, t_K)$.*

By De Boor (1978), $Sp(t_1, \cdots, t_K)$ is a linear space with dimension $K+4$. The *power basis* $\{p_k, k = -3, \cdots, K\}$ of $Sp(t_1, \cdots, t_K)$ is defined as:

$$p_0(t) = 1, \quad p_{-1}(t) = t, \quad p_{-2} = t^2, \quad p_{-3} = t^3,$$

$$p_1(t) = (t - t_1)_+^3, \quad \cdots, \quad p_K(t) = (t - t_K)_+^3,$$

where

$$(t - t')_+^3 = \begin{cases} (t - t')^3 & \text{for } t \geq t' \\ 0 & \text{for } t < t'. \end{cases}$$

Therefore, a power basis representation of a cubic spline $\phi \in Sp(t_1, \cdots, t_K)$ is

$$\phi(t) = \sum_{k=-3}^{K} \theta_k p_k(t) = \sum_{k=0}^{3} \theta_{-k} t^k + \sum_{k=1}^{K} \theta_k (t - t_k)_+^3. \tag{3.1}$$

The power basis is very easy to understand, but isn't used in computation since it has bad numerical properties (see De Boor, 1978). A numerically much better basis is the *B-spline basis*, which consists of functions that are zero except on some sequential subintervals. For more details about the B-spline basis, see De Boor (1978) and Shikin (1995). We will use a restricted B-spline basis, defined in Section 3.2, for our regression.

3.2 The Regression Space $\mathcal{B}$

In this section we will introduce the cubic spline regression spaces defined by Kooperberg et al. (1995) and we will give an algorithm to construct restricted B-spline bases in those spaces.

First, we assume the log hazard function $\alpha(t|\boldsymbol{\beta})$ of a failure time to be a cubic spline from $Sp(t_1, \cdots, t_K)$ satisfying

(**) α is linear on $[0, t_1]$ and is constant on $[t_K, \infty)$.

If we use the power basis representation $\alpha = \sum_{k=-3}^{K} \theta_k p_k$, then (**) implies θ_{-3} and θ_{-2} equal 0, and it places three constraints on $\theta_{-1}, \theta_1, \cdots, \theta_K$. This leaves a $K+4-2-3 = K-1$ dimensional space. More precisely, define

$$\mathcal{P} = \operatorname{span}\{p_{-1}, p_0, p_1, \cdots, p_K\}$$

and

$$\mathcal{B} = \{\phi \in \mathcal{P} : \phi \text{ is linear on } [0, t_1] \text{ and constant on } [t_k, \infty)\}.$$

Then $\dim(\mathcal{P}) = K + 2$ and $\dim(\mathcal{B}) = K - 1$.

Next, we give the definition of a restricted B-spline basis of $\mathcal{B}$. We assume that $K \geq 3$, and place restrictions on b_1, b_{K-1} and, if $K > 3$, on $b_2, \cdots, b_{K-2}$.

Definition 3.2.1 *The set of functions $\{b_1, \cdots, b_{K-1}\}$ in $\mathcal{B}$ is called a restricted B-spline basis of $\mathcal{B}$ if it has the following properties:*

1. *b_1 is linear but not constant on $[0, t_1]$ and is zero on $[t_3, \infty)$;*

2. *if $K > 3$, for $1 < j < K-2$, b_j is zero on $[0, t_{j-1})$ and $[t_{j+3}, \infty)$ but not zero on (t_{j-1}, t_j);*

3. *if $K > 3$, b_{K-2} is zero on $[0, t_{K-3})$, not zero on (t_{K-3}, t_{K-2}), and a non-zero constant on $[t_K, \infty)$;*

4. *b_{K-1} is a non-zero constant on $[0, \infty)$.*

Clearly, any b_j's satisfying 1 - 4 are bounded, a condition required in Section 2.2.1. Now we need to show that the restricted B-spline basis is well defined. That means that we need to prove that b_j's with the above properties exist and they are a basis of $\mathcal{B}$. We will give an algorithm for constructing b_j's (this implies the existence), but first we will show that any b_j's satisfying 1 - 4 are a basis.

Theorem 3.2.2 *If $b_1, \cdots, b_{K-1}, K \geq 3$, in $\mathcal{B}$ have properties 1 - 4 of Definition 3.2.1, then they are a basis of $\mathcal{B}$.*

Proof. Since $\mathcal{B}$ is a $K-1$ dimensional space, to show that $\{b_1, \cdots, b_{K-1}\}$ is a basis of $\mathcal{B}$, we need only prove that $b_1, \cdots, b_{K-1}$ are linearly independent. Let

$$\phi(t) = \sum_{j=1}^{K-1} \alpha_j b_j(t).$$

Suppose that $\phi(t) \equiv 0$. We are going to show that this implies that $\alpha_j = 0, j = 1, \cdots, K-1$. Consider $t \in [0, t_1)$. If $K > 3$, by properties 2 and 3, $b_j(t) = 0$, for $j = 2, \cdots, K-2$. If $K = 3$, then $\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t)$. Therefore, for $K \geq 3$ and $t \in [0, t_1)$

$$\phi(t) = \alpha_1 b_1(t) + \alpha_{K-1} b_{K-1}(t) = 0.$$

By property 4, b_{K-1} is non-zero constant, and by property 1, b_1 is linear but not a constant. Hence we have

$$\alpha_1 = \alpha_{K-1} = 0. \tag{3.2}$$

Hence, if $K = 3$, then b_1 and b_2 are a basis of $\mathcal{B}$. Now we assume that $K > 3$. Consider $t \in (t_1, t_2)$. If $K > 4$, by properties 2 and 3, $b_j(t) = 0, j = 3, \cdots, K-2$. If $K = 4$, then

$$\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t) + \alpha_3 b_3(t).$$

Hence, for $K \geq 4$ and $t \in (t_1, t_2)$, using (3.2), we have

$$\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t) + \alpha_{K-1} b_{K-1}(t) = \alpha_2 b_2(t) = 0.$$

So $\alpha_2 = 0$ since $b_2(t) \neq 0$ by property 2. By induction on j, using a similar explanation, for $t \in (t_{j-1}, t_j)$,

$$\phi(t) = \alpha_j b_j(t) = 0,$$

which implies that $\alpha_j = 0$ for $j = 3, \cdots, K-2$. Hence we have that $\alpha_j = 0, j = 1, \cdots, K-1$, which shows that $\{b_1, \cdots, b_{K-1}\}$ is a basis of $\mathcal{B}$. $\square$

For the convenience of our constructing a restricted B-spline basis, we give the following result as a lemma.

Lemma 3.2.3 *Fix $J = 2, \cdots, K-2$, and let*

$$b(t) = \sum_{k=-3}^{0} \theta_k p_k(t) + \sum_{k=J-1}^{J+2} \theta_k p_k(t).$$

Then $b \equiv c$ *on* $[t_{J+2}, \infty)$ *if and only if* $\theta_{-3}, \theta_{-2},\ \theta_{-1}, \theta_0, \theta_{J-1},\ \theta_J,\ \theta_{J+1},\ \theta_{J+2}$ *solve the linear system*

$$\begin{aligned}
\theta_{-3} + \theta_{J-1} + \theta_J + \theta_{J+1} + \theta_{J+2} &= 0 \\
\theta_{-2} + 3t_{J-1}\theta_{J-1} + 3t_J\theta_J + 3t_{J+1}\theta_{J+1} + 3t_{J+2}\theta_{J+2} &= 0 \\
\theta_{-1} + 3t_{J-1}^2\theta_{J-1} + 3t_J^2\theta_J + 3t_{J+1}^2\theta_{J+1} + 3t_{J+2}^2\theta_{J+2} &= 0 \\
\theta_0 + t_{J-1}^3\theta_{J-1} + t_J^3\theta_J + t_{J+1}^3\theta_{J+1} + t_{J+2}^3\theta_{J+2} &= c.
\end{aligned}$$

Proof. First regroup terms of $b(t)$ by powers of t, so that

$$b(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3,$$

with α's depending on indicator functions involving the θ_k's and the t_j's. We see that $b \equiv c$ on $[t_{J+2}, \infty)$ is equivalent to $\alpha_0 = c, \alpha_1 = \alpha_2 = \alpha_3 = 0$, which is equivalent to the above equations. □

Now we start our construction of a restricted B-spline basis of $\mathcal{B}$. Let $b_j = \sum_{k=-1}^{K} \theta_{kj} p_k$, $j = 1, \cdots, K-1$. We will find θ_{kj}'s so that the b_j's satisfy properties 1 - 4 of Definition 3.2.1.

Step 1: Let $b_{K-1} = p_0$, i.e., $\theta_{0(K-1)} = 1$ and $\theta_{k(K-1)} = 0$, for $k = -1, 1, \cdots, K$.

Step 2: To define $b_1 = \sum_{k=-1}^{K} \theta_{k1} p_k$, let $\theta_{k1} = 0$ if $k > 3$. As in the proof of Lemma 3.2.3, we regroup terms of b_1 by powers of t. Then $b_1 \equiv 0$ on $[t_3, \infty)$ implies that $\theta_{01}, \theta_{(-1)1}, \theta_{11}, \theta_{21}, \theta_{31}$ solve the following linear system

$$\begin{aligned}
\theta_{11} + \theta_{21} + \theta_{31} &= 0 \\
t_1\theta_{11} + t_2\theta_{21} + t_3\theta_{31} &= 0
\end{aligned}$$

$$\begin{aligned}\theta_{(-1)1} + 3t_1^2\theta_{11} + 3t_2^2\theta_{21} + 3t_3^2\theta_{31} &= 0 \\ \theta_{01} + t_1^3\theta_{11} + t_2^3\theta_{21} + t_2^3\theta_{31} &= 0.\end{aligned}$$

Since the t_k's are not equal to each other, there is a unique solution for the linear system once one of the $\theta_{k1}, k = -1, \cdots, 3$, is fixed. We let $\theta_{11} = 1$ and solve the linear system. Note that $\theta_{(-1)1} \neq 0$ otherwise the linear system has only zero solution which would contradict $\theta_{11} = 1$. Thus, b_1 is linear but not constant on $[0, t_1)$ and $b_1 \equiv 0$ on (t_3, ∞).

To construct b_j, $1 < j < K-1$, we first construct b_j^* which is zero on $[0, t_{j-1})$ and a non-zero constant on $[t_{j+2}, \infty)$. We will define the b_j^*'s as linear combinations of the b_j's such that b_j is zero on $[t_{j+3}, \infty)$, $1 < j < K-2$.

Step 3 For $1 < j < K-1$, let

$$b_j^* = \theta'_{(j-1)j}p_{j-1} + \theta'_{jj}p_j + \theta'_{(j+1)j}p_{j+1} + \theta'_{(j+2)j}p_{j+2}.$$

Then for any constant c, by Lemma 3.2.3, $b_j^* \equiv c$ on $[t_{j+2}, \infty)$ and $\theta_{sj} = 0$, $s = -3, -2, -1, 0$ imply that $\theta'_{(j-1)j}, \theta'_{jj}, \theta'_{(j+1)j}, \theta'_{(j+2)j}$ satisfy the linear system

$$\begin{aligned}\theta'_{(j-1)j} + \theta'_{jj} + \theta'_{(j+1)j} + \theta'_{(j+2)j} &= 0 \\ t_{j-1}\theta'_{(j-1)j} + t_j\theta'_{jj} + t_{j+1}\theta'_{(j+1)j} + t_{j+2}\theta'_{(j+2)j} &= 0 \\ t_{j-1}^2\theta'_{(j-1)j} + t_j^2\theta'_{jj} + t_{j+1}^2\theta'_{(j+1)j} + t_{j+2}^2\theta'_{(j+2)j} &= 0 \\ t_{j-1}^3\theta_{(j-1)j} + t_j^3\theta_{jj} + t_{j+1}^3\theta_{(j+1)j} + t_{j+2}^3\theta_{(j+2)j} &= c.\end{aligned}$$

Let $\theta'_{(j-1)j} = 1$. Then we can solve the first three equations for $\theta'_{jj}, \theta'_{(j+1)j}, \theta'_{(j+2)j}$ and then set c to satisfy the fourth equation. We must show that $c \neq 0$. It is easy to see that, if $c = 0$, then the linear system in Lemma 3.2.3 has only the zero solution. Hence here $\theta_{(j-1)j} = 1$ implies that $c \neq 0$. Then b_j^* is zero on $[0, t_{j-1}]$, non-zero on (t_{j-1}, t_j), and a non-zero constant on (t_{j+2}, ∞).

Step 4: Let $b_{K-2} = b_{K-2}^*$. Then b_{K-2} satisfies property 3.

Step 5: Recall that b_j^* is a non-zero constant on $[t_{j+2}, \infty)$. For $1 < j < K-1$, let a_j be that constant on $[t_{j+2}, \infty)$. For $1 < j < K-2$, if we let

$$b_j = b_j^* - \frac{a_j}{a_{j+1}} b_{j+1}^*$$

then b_j is zero on $[t_{j+3}, \infty)$ and also on $[0, t_{j-1})$, since b_j^* and b_{j+1}^* are zero on $[0, t_{j-1})$. Also b_j is not zero on (t_{j-1}, t_j) since b_{j+1}^* is zero but b_j^* is not zero on (t_{j-1}, t_j).

Thus $b_1, \cdots, b_{K-1}$ satisfy properties 1 - 4 and are a B-spline basis of $\mathcal{B}$.

From now on, for given knots, we use $\mathcal{B}$ as our regression space with the restricted B-spline basis $b_1, \cdots, b_{K-1}$ defined as above. By a bounded basis plus log tail, we mean the basis is $\{b_1, \cdots, b_{K-1}, b_K\}$ and the regression space is $\text{span}\{b_1, \cdots, b_K\}$, where $b_K(t) \equiv \log(t + c)$, $c \geq 0$ is a constant.

Since the basis $\boldsymbol{b}$ is bounded or bounded plus log tail, all relevant results discussed in Chapter 2 can be used. To simplify notation, we will denote both regression spaces by $\mathcal{B}$.

3.3 Numerical Implementation

In this section we introduce the algorithm used to calculate the estimates $\hat{\underline{\boldsymbol{\beta}}}$ and $\hat{Q}$ in the log hazard regression model of paired failure times.

By the definition of $\hat{\underset{\sim}{\boldsymbol{\beta}}}$ from (2.33), the estimates $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are calculated separately. So we need only discuss the calculation of $\hat{\boldsymbol{\beta}}_1$. We can use Kooperberg's *heft* algorithm (see Kooperberg, Stone, Truong, 1995) to find $\hat{\boldsymbol{\beta}}_1$. They use the Newton-Raphson method for computing $\hat{\boldsymbol{\beta}}_1$ since $\boldsymbol{\beta}_1$ is the unique solution of the equations (2.20). Specifically, they start with an initial guess $\hat{\boldsymbol{\beta}}_1^{(0)}$, then iteratively determine $\hat{\boldsymbol{\beta}}_1^{(k+1)}$ from $\hat{\boldsymbol{\beta}}_1^{(k)}$ according to the formula

$$\hat{\boldsymbol{\beta}}_1^{(k+1)} = \hat{\boldsymbol{\beta}}_1^{(k)} + [I^{(n)}(\hat{\boldsymbol{\beta}}_1^{(k)}|\boldsymbol{X}_1)]^{-1}\boldsymbol{U}^{(n)}(\hat{\boldsymbol{\beta}}_1^{(k)}|\boldsymbol{X}_1),$$

and stop the iterations when $\log L_q(\hat{\boldsymbol{\beta}}_1^{(k+1)}|\boldsymbol{X}_1) - \log L_q(\hat{\boldsymbol{\beta}}_1^{(k)}|\boldsymbol{X}_1) \leq \epsilon$, where ϵ is a given positive number chosen so that estimates with the desired accuracy can be obtained.

Thus the main numerical task in calculating $\hat{\underset{\sim}{\boldsymbol{\beta}}}$ is the computation of the log likelihood $\log L_q(\boldsymbol{\beta}_1|\boldsymbol{X}_1)$, score function $\boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1)$, and observed information matrix $I_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1)$ for various values of $\boldsymbol{\beta}$. By the definitions of $\log L_q(\boldsymbol{\beta}_1|\boldsymbol{X}_1), \boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1)$ and $I_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1)$, see (2.17), (2.18), and (2.19), this computation involves the numerical approximation of

$$\sum_{i=1}^{n} \int_0^{X_{1i}} \psi(t)dt, \tag{3.3}$$

for ψ of the form

$$\psi(t) = b_{1m}(t)b_{1l}(t)\exp(\sum_{j=1}^{p_1} \beta_{1j}b_{1j}(t)),$$

$l, m \in \{1, \cdots, p_1\}$.

Kooperberg et. al. do not calculate $\int_0^{X_{1i}} \psi(t)dt$ for each i. Rather, they rewrite (3.3) as

$$\int_0^{\infty} N(t)\psi(t)dt, \tag{3.4}$$

where $N(t)$ is the number of i satisfying $X_{ki} \geq t$. The function $N(\cdot)$ is piecewise constant, has jumps at the observations $X_{k1}, \cdots, X_{kn}$, and equals zero to the right of the maximum observation. Then they rewrite (3.4) as

$$\sum_{\nu} \int_{a_{\nu-1}}^{a_\nu} N(t)\psi(t)dt,$$

where $\{a_\nu\}$ is a finite grid of points containing all knots. Then they calculate $\int_{a_{\nu-1}}^{a_\nu} N(t)\psi(t)dt$ numerically.

Now consider calculation of $\hat{Q}$. By the definition of $\hat{Q}$ from (2.36), the calculation of $\hat{Q}$ involves the calculation of $\hat{\boldsymbol{\beta}}_k, I_k^{(n)}(\hat{\boldsymbol{\beta}}_k|\boldsymbol{X}_k)$, and the summands $\boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k)$, $k = 1, 2$, $i = 1, \cdots, n$. As we mentioned above, we can get $\hat{\boldsymbol{\beta}}_k$ from the *heft* algorithm directly. By slightly modifying the *heft* code, we can obtain $I_k^{(n)}(\hat{\boldsymbol{\beta}}_k|\boldsymbol{X}_k)$ from *heft* as well. But to evaluate $\boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k)$, we need to calculate $\int_0^{X_{1i}} \psi(t)dt$ separately. We can not get these integrals from *heft* without rewriting its entire integration program. Instead, we use Gaussian Quadrature (see Abramowitz and Stegun 1964, p. 916) for the integration $\int_0^{X_{ki}} \psi(t)dt$ to calculate $\boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k), k = 1, 2$.

3.4 Knot Selection

In Section 3.2, we give the method to define the regression space $\mathcal{B}$ for given knots. In this section we introduce the methods that we use to choose the knots.

For the univariate case, we use the following two methods to select the knots.

- Choose knots by Kooperberg' *heft* algorithm.

For a data set, Kooperberg's *heft* algorithm can choose the knots for the model fitting. The algorithm chooses knots by a stepwise addition and stepwise deletion procedure. See Kooperberg et. al. 1995 for details. We hoped that *heft* would select knots well. But from our simulation study we find that there are some numerical calculation problems with the knots chosen by *heft*. If the ratio of the maximum of the knots to the minimum of the knots is too big, then the resulting log hazard estimate and the calculation of the estimated standard errors may be impossible, see Section 5.2.2. So we should not use the knots from *heft*, if we receive a warning message from the *heft* code, or the calculation of the estimated standard error is not possible.

- Choose the knots by the quantiles.

We choose the quantiles of the non-censored observations as the knots to define the regression space $\mathcal{B}$. This procedure is based on ideas in Kooperberg et. al's knot selection for their stepwise addition method and Abrahamowicz et. al's knot selection method, Abrahamowicz, Ciampi, and Ramsay (1992). Choosing knots equal to quantiles may also result in a large ratio of maximum of the knots to the minimum of the knots. As noted above, this causes numerical problems. We solve this problem by truncating the data, that is, if there are any warnings, we use the quantiles of a truncated data set. We truncate the observations which are greater than 80 in our simulations.

For paired data we use the following methods:

- Use different knots for modeling the two log hazard functions.

Choose knots for each failure time by the above two methods. We then use the two sets of knots to define the two regression spaces and marginally fit the log hazard regression models for the paired data.

- Use the the same knots for modeling the two log hazard functions.

In this method we need to choose one set of knots which defines one regression space for both log hazard functions. If we denote the ranges of non-censored observations $\{X_{ki} : \delta_{ki} = 1\}$ by R_k, $k = 1, 2$, then our knots must lie in $R_1 \cap R_2$. There are two way to choose this set of knots.

a) Use the quantiles of the non-censored observations which lie in $R_1 \cap R_2$.

b) Use the union of the knots selected for each marginal log hazard estimate. Denote the sets of knots selected for the two log hazards by $\mathcal{K}_1$ and $\mathcal{K}_2$. Use $(\mathcal{K}_1 \cup \mathcal{K}_2) \cap (R_1 \cap R_2)$ as the set of knots for the regression space.

3.5 Hypothesis Testing

In this section we show how to use the restricted B-spline basis and the regression space $\mathcal{B}$ and given data to test hypotheses.

3.5.1 Univariate Case

In the univariate case, we can test:

- H_0^e: the failure time has an exponential distribution;

- H_0^w: the failure time has a Weibull distribution.

By the definition of the log hazard regression model in (2.5), the log hazard function is

$$\alpha(t|\boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j b_j(t).$$

If the failure time is exponentially distributed, then $\alpha(t|\boldsymbol{\beta}) \equiv$ constant. Therefore, for the restricted B-spline basis defined in Section 3.2, ($\{b_j, j = 1, \cdots, K-1\}$ if $\boldsymbol{b}$ is bounded or $\{b_j, j = 1, \cdots, K\}$ if $\boldsymbol{b}$ is bounded plus log tail, where K is the number of knots), we can rewrite H_0^e as

H_0^e: $\beta_j = 0$ for $j \neq K-1$.

Then we can write H_0^e as

H_0^e: $X_e^T \boldsymbol{\beta} = 0$

for an appropriate matrix X_e. By Corollary 2.3.3,

$$[nI^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow N(0, I).$$

Hence, under H_0^e,

$$n(X_e^T\hat{\boldsymbol{\beta}})^T(X_e^T[I^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{-1}X_e)^{-1}(X_e^T\hat{\boldsymbol{\beta}}) \Rightarrow \begin{cases} \chi^2_{K-2} & \text{if } \boldsymbol{b} \text{ is bounded;} \\ \chi^2_{K-1} & \text{if } \boldsymbol{b} \text{ is bounded plus log tail.} \end{cases}$$

Therefore, $n(X_e^T\hat{\boldsymbol{\beta}})^T(X_e^T[I^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{-1}X_e)^{-1}(X_e^T\hat{\boldsymbol{\beta}})$ can be used as a test statistic.

To test H_0^w, we need to use the restricted B-spline basis plus log tail $b_K(t) \equiv \log(t)$. Thus, if the failure time has a Weibull distribution, then

$$\alpha(t|\boldsymbol{\beta}) = \beta_{K-1}b_{K-1}(t) + \beta_K b_K(t).$$

Hence, testing H_0^w is equivalent to testing $\beta_j = 0$ for $j < K-1$. By a similar procedure as for testing H_0^e, we can find a matrix X_w such that

$$n(X_w^T\hat{\boldsymbol{\beta}})^T(X_w^T[I^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{-1}X_w)^{-1}(X_w^T\hat{\boldsymbol{\beta}}) \Rightarrow \chi^2_{K-2}$$

and we can use $n(X_w^T\hat{\boldsymbol{\beta}})^T(X_w^T[I^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{-1}X_w)^{-1}(X_w^T\hat{\boldsymbol{\beta}})$ as our test statistic.

3.5.2 Bivariate Case

In bivariate case, we can test:

- H_0: the two failure times have the same distribution;
- H_0^p: the two failure times have proportional hazards.

For given paired data $(X_{1i}, \delta_{1i}, X_{2i}, \delta_{2i}), i = 1, \cdots, n$, to test H_0 and H_0^p, we model the log hazard functions of the two failure times with the same regression space $\mathcal{B}$. Then the log hazard regressions are

$$\begin{aligned} \alpha_1(t|\boldsymbol{\beta}_1) &= \sum_{j=1}^{p} \beta_{1j} b_j(t) \\ \alpha_2(t|\boldsymbol{\beta}_2) &= \sum_{j=1}^{p} \beta_{2j} b_j(t). \end{aligned}$$

If the two failure times have the same distribution, then

$$\alpha_1(t|\boldsymbol{\beta}_1) - \alpha_2(t|\boldsymbol{\beta}_2) = \sum_{j=1}^{p} (\beta_{1j} - \beta_{2j}) b_j(t) \equiv 0.$$

Therefore, testing H_0 is equivalent to testing

H_0^*: $\beta_{1j} - \beta_{2j} = 0$ for $j = 1, \cdots, p$.

Then we can rewrite H_0^* as

$$H_0^* : X^T \underset{\sim}{\boldsymbol{\beta}} = \mathbf{0}$$

for an appropriate matrix X. For bases described in Section 3.2, by Corollary 2.5.4,

$$\sqrt{n}\, \hat{Q}^{-\frac{1}{2}}(\hat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}) \Rightarrow N(0, I).$$

Then under H_0,

$$n(X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underset{\sim}{\boldsymbol{\beta}}}) \Rightarrow \begin{cases} \chi^2_{K-1} & \text{if } \boldsymbol{b} \text{ is bounded;} \\ \chi^2_K & \text{if } \boldsymbol{b} \text{ is bounded plus log tail.} \end{cases}$$

So, we use $n(X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})$ as test statistic.

If the two hazards are proportional, then

$$\alpha_1(t|\boldsymbol{\beta}_1) - \alpha_2(t|\boldsymbol{\beta}_2) = \sum_{j=1}^{p} (\beta_{1j} - \beta_{2j}) b_j(t) \equiv \text{constant}.$$

Hence, for the regression spaces and b_j's defined in Section 3.2, the test for H_0^p is equivalent to the test for $\beta_{1j} - \beta_{2j} = 0$ for $j \neq K - 1$. Then we can rewrite H_0^p as

$$H_0^p : X^T \underset{\sim}{\boldsymbol{\beta}} = \mathbf{0}$$

for an appropriate matrix X. By Corollary 2.5.4,

$$\sqrt{n}\, \hat{Q}^{-\frac{1}{2}}(\hat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}) \Rightarrow N(0, I).$$

Then under H_0^p,

$$n(X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underset{\sim}{\boldsymbol{\beta}}}) \Rightarrow \begin{cases} \chi^2_{K-2} & \text{if } \boldsymbol{b} \text{ is bounded;} \\ \chi^2_{K-1} & \text{if } \boldsymbol{b} \text{ is bounded plus log tail.} \end{cases}$$

Therefore, $n(X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underset{\sim}{\boldsymbol{\beta}}})$ can be used as test statistic.

Note: When T_{1i}'s and T_{2i}'s are independent, the Cox proportional hazards model

$$h_C(t) = h_T(t)\exp(\gamma) \tag{3.5}$$

provides an estimate of the relative risk γ. Thus under independence assumption, we can also test H_0 by testing $H_0^c : \gamma = 0$. To our knowledge, there is no non-parametric test for proportional hazards for dependent paired data as we defined in Chapter 2. We might use the usual Cox proportional model, assuming the two failure times are independent. However, while the usual estimate of γ may be good, the standard errors are probably biased.