BIOMETRICS 000, 1–2 DOI: 000
000 2014

Web suplementary materials for "Hypothesis Testing for an Extended Cox Model with Time-Varying Coefficients"

Takumi Saegusa, Chongzhi Di *, Ying Qing Chen

Fred Hutchinson Cancer Research Center and University of Washington

**email:* cdi@fhcrc.org

Appendix

A. *Asymptotic distributions of test statistics*

The two proposed test statistics (T_1 and T_2), as well as the log-rank (T_{LR}) and modified PH (T_{mPH}) test statistics, are all asymptotically equivalent to quadratic forms of $S(\widehat{\beta},0,0)$. In the following, we describe asymptotic distributions of such quadratic forms, as well as approximation methods to calculate p-values. The following derivations are similar to those of Lin et al. (2006).

We consider a general quadratic form $Q = S^TUS$, where $S := S(\widehat{\beta},0,0)$ is a vector of length r and U is a positive semi-definite matrix of size $r \times r$. Since each element of S is a realization of the score function, S has mean 0 and its variance-variance matrix is the Fisher information V. One can rewrite $Q = S^TUS = (V^{-1/2}S)^T(V^{1/2}UV^{1/2})(V^{-1/2}S)$, where $V^{-1/2}S$ are standardized S with identity matrix as its covariance matrix. Using quadratic form theory and the central limit theorem, one obtains the following result.

Proposition. Asymptotically, the distribution of the quadratic form $Q = S^TUS$ is approximately a weighted average of χ_1^2, more specifically,

$$Q \to \sum_{k=1}^{r} \lambda_k \chi_1^2,$$

where λ_k's are eigenvalues of the matrix UV. The mean and variance of the limiting distribution are $tr(UV)$ and $tr(UVUV)$, respectively.

In practice, it is often the case that the first few eigenvalues capture the most variations and the remaining ones are negligible. To calculate p-values, one can use another approximation $c\chi_v^2$, i.e., a scaled χ^2 distribution with degree of freedom v. By matching the mean and variance of the two distributions, one can obtain the choice of parameters $c = tr(UVUV)/tr(UV)$ and $v = \{tr(UV)\}^2/tr(UVUV)$. In simulations, we found that both approximations work reasonably well in finite samples.

B. *Connection with weighted log-rank tests via spectral decomposition*

In this section, we will apply the spectral decomposition to understand the connection between the proposed tests and the weighted log-rank test. Consider the general quadratic form $Q = S^TUS$, where U is a non-negative semi-definite matrix. One has spectral decomposition $U = \sum_{k=1}^{r} \lambda_k P_k P_k^T$, where λ_k's and P_k's are eigenvalues and eigenvectors of U, respectively. Using such decomposition, the quadratic form can be written as

$$Q = S^TUS = \sum_{k=1}^{r} \lambda_k \, S^T P_k P_k^T S = \sum_{k=1}^{r} \lambda_k (P_k^T S)^T (P_k^T S). \tag{A.1}$$

Note that the k^{th} term is equivalent to the weighted log-rank statistic with weight P_k. Thus, the test statistic Q is equivalent to a linear combination of several weighted log-rank statistics, with weights determined by the eigenvectors of the matrix U. The relative importance of each weighted log-rank statistic in the linear combination is determined by the eigenvalues λ_k.

If the matrix U has rank 1 and thus only one eigenvector P_1, the test statistic Q is actually weighted log-rank test with weight P_1 (unweighted log-rank test if and only if $U \propto \mathbf{1}\mathbf{1}^T$, or equivalently, $P_1 \propto \mathbf{1}$). If $rank(U) > 1$, the quadratic form Q is equivalent to a linear combination of several weighted log-rank statistics, different from any weighted log-rank tests. The resulting test statistics incorporate information from deviation from the null in several different directions, and thus are expected to be omnibus when the shape of true hazard ratio function is unknown. In Lin et al. (2006), they chose $U = \Sigma$, which was derived from the differential operator, and their test statistic would summarize information from possible non-proportionality. For the proposed tests T_1 and T_2, we choose the matrix U to be a linear combination of $\mathbf{1}\mathbf{1}^T$ and Σ, and thus our test statistics combine information from both the magnitude and shape of the hazard ratio function.

C. *A sketch of proof of the properties of* T_2

We provide a sketch of proof to show that $\mathbf{1}^T S(\widehat{\beta}, 0, 0)$ and $W(\widehat{\beta})S(\widehat{\beta}, 0, 0)$ are approximately uncorrelated under the null. Therefore, T_2 is expected to combine information from $\mathbf{1}^T S(\widehat{\beta}, 0, 0)$ and $S(\widehat{\beta}, 0, 0)$ effectively. Because the profile likelihood is an approximately least favorable submodel of the Cox model (Murphy and van der Vaart, 2000), we treat the partial likelihood as a legitimate likelihood from a parametric model without a nuisance parameter. Denote the partial likelihood as $\ell(\beta, \theta_0, \underline{\theta})$. Note that the projection of a random vector X onto a random vector Y is $Z = \{\mathrm{cov}(X, Y)/\mathrm{var}(Y)\}Y$. We consider $X = S(\beta_0, 0, 0)$ and $Y = \mathbf{1}^T S(\beta_0, 0, 0)$ where β_0 is a true parameter. Then the projection of X onto Y is given by

$$\begin{aligned} Z &= \frac{\mathrm{cov}\{\ S(\beta_0, 0, 0), \mathbf{1}^T S(\beta_0, 0, 0)\ \}}{var(\mathbf{1}^T S(\beta_0, 0, 0))} \mathbf{1}^T S(\beta_0, 0, 0) \\ &= \frac{\mathrm{E}\{S(\beta_0, 0, 0) S(\beta_0, 0, 0)^T\}\,\mathbf{1}}{\mathbf{1}^T \mathrm{var}\{S(\beta_0, 0, 0)\}\mathbf{1}} \mathbf{1}^T S(\beta_0, 0, 0), \end{aligned}$$

Let

$$V = \mathrm{E}\{\ S(\beta_0, 0, 0) S(\beta_0, 0, 0)^T\ \} = \mathrm{E}\left[\frac{\partial \ell(\beta_0, 0, 0)}{\partial \underline{\theta}}\ \frac{\partial \ell(\beta_0, 0, 0)}{\partial \underline{\theta}^T}\right] = -\mathrm{E}\left[\frac{\partial^2 \ell(\beta_0, 0, 0)}{\partial \underline{\theta} \partial \underline{\theta}^T}\right],$$

we obtain

$$Z = (\mathbf{1}^T V \mathbf{1})^{-1} V \mathbf{1}\mathbf{1}^T S(\beta_0, 0, 0).$$

Since $\mathrm{E}\{\dot{S}(\beta_0, 0, 0)\}$ and β_0 are unknown, we plug in their empirical estimates, $n^{-1}\dot{S}(\widehat{\beta}, 0, 0)$ and $\widehat{\beta}$, to obtain

$$\widehat{Z} = (\mathbf{1}^T \widehat{V} \mathbf{1})^{-1} \widehat{V} \mathbf{1}\mathbf{1}^T S(\widehat{\beta}, 0, 0).$$

Thus, $WS(\widehat{\beta}, 0, 0) = S(\widehat{\beta}, 0, 0) - \widehat{Z}$ and $\mathbf{1}^T S(\widehat{\beta}, 0, 0)$ are asymptotically uncorrelated.

REFERENCES

Lin, J., Zhang, D., and Davidian, M. (2006). Smoothing spline-based score tests for proportional hazards models. *Biometrics* **62,** 803–812.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95,** 449–485.