Improving the Accuracy of Parameter Estimation of Proportional Hazards Regression with Kernel Resampling

Haiyan Bai
University of Central Florida

The accuracy of parameter estimation of proportional hazards regression (PHR) has been a concern. To improve the accuracy of the estimation, the bootstrap has been used; unfortunately, prior research revealed inconsistent findings. The current study applies a new resampling method, the kernel resampling technique (KRT), to PHR. Two empirical datasets were employed to cross-validate and compare the accuracy and stability of the estimation results through multiple replications from KRT with those from the naïve bootstrap as well as the maximum likelihood method. The study results revealed that KRT outperformed the bootstrap and maximum likelihood method in estimating parameters of PHR. The application of KRT to PHR improved the accuracy of the parameter estimation.

Proportional hazards regression (PHR) (or Cox model) is a method for investigating the effect of several variables upon the time-specified outcome for an event to occur. PHR is most commonly applied in time-to-event studies (Cox, 1972). It assumes that the effects of the predictor variables upon survival are constant over time and are additive in one scale. If the assumptions are met, the PHR model can provide better estimates of survival probabilities and cumulative hazard than those provided by the Kaplan-Meier function; a log-rank test method for comparing survival curves in two or more groups (Cox). The PHR model has been used widely in medical studies and increasingly employed in a variety of disciplines under various rubrics, for example, "event-history analysis" in sociology (Allison, 1984), or "teacher survivals" and "student retention" in Education (cf. Adams, 1996; Adams & Dial, 1993; Plank, DeLuca, & Estacion, 2008). However, the accuracy of the estimation of the PHR model parameters has been a concern because estimating density functions or hazard rate functions is complicated (Burr, 1994). To improve the estimation accuracy from PHR models, the bootstrap method was implemented. Unfortunately, the effectiveness of this method is questionable due to the inconsistent findings of the performance of the bootstrap in the PHR model in prior research (Burr; Hjort, 1985; Singh, 1981).

Studies on the PHR model using the bootstrap are classified into two types: one for PHR model selections and the other for parameter estimation. The following is a brief review of these studies. Chen and George (1985) conducted a primary study using the bootstrap to investigate the variable selection in PHR, but they neither considered the prognostic implications for individuals nor discussed the accuracy of the parameter estimation. Extending Chen and George's study, Sauerbrei and Schumacher (1992) proposed a bootstrap-model selection procedure, but this study still focused on the model selection without considering the use of the bootstrap procedures directly in the parameter estimation. Altman and Andersen (1989) explored the confidence interval estimation of hazard ratios while conducting a bootstrap investigation of the stability of the PHR model, and the results revealed that the bootstrap intervals were graphically wider than those obtained from the original model. Hjort (1985) discussed using the bootstrap in the PHR model and found that the bootstrap procedure was first-order equivalent to the standard procedure. This was consistent with later research findings (e.g., Burr, 1994).

Burr (1994) presented a comprehensive study focusing on the methodological discussion about using the bootstrap procedures in PHR parameter estimation. This study compared bootstrap confidence intervals for the following three types of parameters in PHR: the regression parameters, the survival function at fixed time points, and the median survival time at fixed values of a covariate. The study revealed that the bootstrap-t intervals consistently outperformed both bootstrap percentile and hybrid interval estimations. The results also showed that the bootstrap did not improve the quality of regression parameter estimation on the asymptotic method, but it did improve the estimation of the survival function. Burr provided useful information to employ the bootstrap for parameter estimation in PHR; however, as Burr (p. 1301) stated, "We would like to be able to recommend a single method appropriate for all parameters, but currently this is not possible." Therefore, further research in this area is desirable. The current study aims at exploring the potential improvement of parameter estimation in PHR using kernel resampling procedures.

PHR Model

PHR for hazard rate was first introduced by Cox (1972) and it is often expressed as:

$$\lambda(t; \mathbf{X}) = \lambda_0(t)\exp(\mathbf{X}\beta),$$

where $\lambda(t; \mathbf{X})$ is the hazard (risk of event) at time t with respect to covariate matrix $\boldsymbol{X}$. The parameter β is a log relative risk and $\exp(\beta)$ is a relative risk of response. PHR is sometimes called relative risk regression, Cox regression, or Cox model. $\lambda_0(t)$ represents a reference point that depends on time, which is the "baseline" hazard (when covariates $\mathbf{X}$ are zero) just as β_0 denotes an arbitrary reference point in other types of regression analysis. PHR is a useful tool for studying patient survival time in medical studies, historical event in social science, company bankruptcy in economic investigations, and students' departure and teachers' survival in educational research.

The Bootstrap and KRT

As a modern statistical technique, the bootstrap has been used in many procedures to improve the validity of studies through estimating more accurate standard errors (Efron & Tibshirani, 1993). The basic concept of the bootstrap is to construct empirical distributions of parameter estimates to assess the standard errors or confidence intervals to obtain improved statistical estimates. The bootstrap empirical distribution is usually constructed from bootstrap resamples, which are obtained through resampling from the original data with replacement. Existing studies have revealed the usefulness of the bootstrap in PHR (Gonzalez, Pena, & Delicado, 2010)

Kernel resampling technique (KRT) is an alternative resampling method which extends the bootstrap by sampling with random errors from Gaussian Kernels using a fixed bandwidth (Bai & Pan, 2009). KRT is a product of integrating the distribution theory into the smoothing technique. By design, KRT is fundamentally different from the bootstrap and its variant, the smoothed bootstrap, which requires researchers to find the optimal bandwidth to smooth the bootstrap distribution. KRT uses the Gaussian kernel technique to capture the covariance structure of multivariate data (Silverman, 1986; Simonoff, 1996).

The multivariate Gaussian kernel is defined as

$$K(\mathbf{x}) \sim N_d(\mathbf{X}_i, \mathbf{H}^2),$$

where d is the number of variables, $\mathbf{X}_i$ ($i = 1, \ldots, n$) are multivariate data or a vector from a d-dimensional space R_d, n is the number of cases, and $\mathbf{H}$ is the bandwidth matrix that can be chosen as an optimal one to minimize the *mean integrated square error* (MISE) (Silverman, 1986; Simonoff, 1996):

$$\mathbf{H}_{\mathrm{o}} = \left(\frac{4}{d+2}\right)^{1/(d+4)} \mathbf{S}^{1/2} n^{-1/(d+4)}$$

KRT has been successfully used in multiple regression models for increasing the accuracy of parameter estimation (Bai & Pan, 2009).

Purpose of the Study

Considering the usefulness of the bootstrap in PHR (Gonzalez et al., 2010), the current study was proposed to use the KRT, an alternative to the bootstrap, to improve the accuracy of parameter estimation of PHR. The application of KRT to a multiple regression model has successfully provided more accurate parameter estimation than both naïve bootstrap and smoothed bootstrap (Bai, 2008; Bai & Pan, 2009); therefore, the purpose of the current study is to examine the performance of the application of KRT to PHR. Empirical data from education were employed for the methodological comparison through resampling at multiple numbers of replications to study the accuracy and stability of the estimation, while the medical data set was used to cross-validate the results. The findings from the applications of KRT to the PHR model using both data sets are compared with those from the bootstrap and the classical maximum likelihood (ML) method to determine which method is the most effective.

Studies with Empirical Data

Study 1: Educational Data

The study data were collected from an urban public school district in the Southeast region of the United States after obtaining Institutional Review Board approval at the author's university. This data set documented the departure records of 8462 students who departed from public schools between 2006 and 2010 including regular graduations. There were conceivably seven ways of departure in the data set that are listed in Table 1. In the current study, the PHR model was used to study the hazard rate for students' departure from public schools versus regular graduations.

For this study, two variables, student age and accumulated GPA, were used in the PHR as the covariates for the purpose of methodological evaluation of the performance of KRT application to PHR. The two variables were utilized as covariates (i.e., predictors in the PHR) because of their influential impacts on high school student departure based on the extant literature. Hauser, Simmons, and Pager (2000) stated that the likelihood of student departure increased with age in general; therefore, high school students tend to have a higher dropout rate than elementary and middle school students. The association between academic performance and dropout rates has been well studied (cf. Fagan & Pabon, 1990; Krohn, Thornberry, Collins-Hall, & Lizotte, 1995; Rumberger, 1987). Student academic performance is a major predictor of graduation rates and departure rates (Battin-Pearson et al., 2000). Prior studies examined and identified many influential factors or predictors for high school student departure including a variety of demographic, individual, family, and school characteristics (Neild, Stoner-Eby, & Furstenberg, 2008). However, for the focus of the current study on methodological discussions, only two major factors were included, student age and cumulative GPA, in the model to compare the accuracy of the statistics from different statistical procedures with no intention of providing any statistical inferences from the empirical example. The variables used in the model:

Table 1. The Numbers of Public High School Students' Departure and Regular Graduation

Type of Departure	Students	Male/Female
Non-Public	277	156/121
Nowhere	157	87/70
Home School	255	128/127
Adult Program	1770	1007/763
Another District	1548	847/701
Out of State	1257	638/619
Regular Diploma	3398	1662/1736

- Departure and Graduation: Move to non-public schools, go nowhere, home school, adult program, move to other in-state public schools, or move to other states versus obtain a regular diploma.
- Age: Student age was recorded at the time of departure.
- Cumulative GPA: The student GPA measure was the accumulated GPA since the semester a student entered the public high school.
- Survival Months: Months of staying in the public schools.

PHR on Student Departure Data

A PHR model for the current study was defined as:

$$\text{Log}[\lambda(t;\mathbf{X})] = \log[\lambda_0(t)] + \beta\mathbf{X},$$

where **X** represents the predictors, *age* and *Weighted Cumulative GPA,* and β is *the logarithm of the ratio* of the hazard rate for students belonging to departure versus regular graduation in the hazard function.

The PHR model was fitted with *age* and *Weighted Cumulative GPA* to estimate the hazard ratio. No evidence was found that students' departure in general depends on age (while adjusting only for *Weighted Cumulative GPA*) with $\chi^2 = 0.13$ ($p = 0.98$) (see Table 2); therefore, age was eliminated in the final model.

Table 2. Estimates for Predictors

Variable	*df*	β	*SE*	χ^2	$Pr > \chi^2$	HazardRatio	95% CI	
AGE	1	0.02	0.06	0.13	0.72	0.980	0.88	1.09
GPA	1	0.58	0.06	83.61	<0.001	0.562	0.50	0.63

Table 3. *Comparisons of Estimates, CIs, and Bias of PH Model with Asymptotic, Bootstrap, and KRT.*

	Estimates	Replicates	Estimate	*SE*	CI(2.5%)	CI(97.5%)	Bias
Hazard Ratio	ML		0.5660	0.0333	0.5030	0.6360	
	Bootstrap	200	0.5694	0.0404	0.4983	0.6541	0.0034
		500	0.5685	0.0421	0.4966	0.6604	0.0025
		1000	0.5678	0.0421	0.4925	0.6593	0.0018
	KRT	200	0.5700	0.0167	0.5384	0.6048	0.0040
		500	0.5678	0.0155	0.5353	0.5993	0.0018
		1000	0.5672	0.0130	0.5426	0.5918	0.0012
GPA	ML		-0.5693	0.0598	-0.6889	-0.4498	
	Bootstrap	200	-0.5656	0.0703	-0.6966	-0.4245	0.0038
		500	-0.5675	0.0736	-0.6999	-0.4149	0.0018
		1000	-0.5687	0.0737	-0.7082	-0.4166	0.0006
	KRT	200	-0.5663	0.0289	-0.6210	-0.5073	0.0030
		500	-0.5686	0.0254	-0.6209	-0.5165	0.0007
		1000	-0.5680	0.0235	-0.6127	-0.5218	0.0014

Results of Study 1

In order to conduct the methodological study, *Weighted Cumulative GPA* was selected to estimate the hazard ratio to examine the performance of KRT in PHR. Both KRT and the bootstrap procedures were used to obtain parameter estimates of *Weighted Cumulative GPA* and *the estimate of hazard ratio* for comparing the results. Two hundred, 500, and 1000 replications of both the bootstrap and KRT were conducted based on the original student departure data using the SAS macro (SAS Institute Inc., 2008) for parameter estimation and hazard ratio estimation of the PHR model.

From Table 3 we can see that the KRT estimates were comparable to the estimates for both *hazard ratio* and *β* for *Weighted Cumulative GPA* from the bootstrap and ML estimates; however, the standard errors from the KRT estimates for the hazard ratio and *β for Weighted Cumulative GPA* were systematically smaller than those from the bootstrap procedure and the Maximum Likelihood estimates across various numbers of replications with less biases in most cases. The confidence intervals (percentiles) for the estimates using the KRT procedure were narrower than those from both the bootstrap procedure and the Maximum Likelihood method.

Study 2: Cross-Validating Data

To cross-validate the results of Study 1 for further evaluation on the performance of the application of KRT to PHR, a study was conducted using a large national medical data set, *Localized colon carcinoma 1975–1994*, as the original input data collected by the Institute for Statistical and Epidemiological Cancer. *Localized colon carcinoma 1975–1994* contains individual-level data of 6,274 patients diagnosed with localized tumors among 15,564 patients diagnosed with colon carcinoma in Finland 1975-1994 with follow-up to the end of 1995.

For the purpose of the methodological research focusing on comparison of the accuracy of the PHR model parameter estimations, the model selection is not discussed in the current study. With regard to the focus of the current study, the hazard ratio of mortality from colon cancer versus mortality due to other reasons was studied using the PHR model (i.e., mortality among the 6,274 patients diagnosed with localized tumors).

Table 4. Localized Stage

Status	Patient *N*
0: Alive	2979
1: Dead: colon cancer	1734
2: Dead: other	1557
3: Lost to follow-up	4

Study Variables:

In the current study, four variables of interest were used:

- Gender: Gender is defined as male or female.
- Year of Diagnosis: The year diagnosed as having localized tumors.
- Survival Months: Months survived since the time of diagnosed localized tumors.
- Status: Vital status at last date of contact.

Table 5. Estimates for Predictors

Variable	*df*	*β*	*SE*	χ^2	$Pr > \chi^2$	Hazard Ratio	95% CI	
Gender	1	-0.002	0.049	0.020	0.966	0.998	0.907	1.098
Year85--94	1	-0.232	0.049	22.258	<0.001	0.793	0.720	0.873

PHR Model on Localized Colon Carcinoma Data

A PHR model for the current study was defined as:

$$\text{Log}[\lambda(t;\mathbf{X})] = \log[\lambda_0(t)] + \beta\mathbf{X}$$

where **X** represents the predictors, *gender* and *Year of Diagnosis,* and *β* is *the logarithm of the ratio* of the hazard rate for patients belonging to the *mortality from colon cancer* group versus the *mortality group because of other reasons* in the hazard function. The PHR model was fitted with *gender* and *Year of Diagnosis* as predictors just for the purpose of the methodological discussion focus of this study. No evidence was found that mortality depends on gender while adjusting for year of diagnosis with $\chi^2 = .02$ ($p = .966$) (see Table 5). Therefore, *Year of Diagnosis* was selected to estimate the parameters and the hazard ratio for examining the performance of KRT in the Cox model with respect to the preliminary model fitting information. The KRT, the bootstrap, and the Maximum Likelihood method were used to obtain parameter estimates of *Year of Diagnosis* and *the estimate of hazard ratio* for comparing the results for examining the performance of KRT in the PHR model.

Cross-Validating Results from Study 2

Table 6 presents the parameter and hazard ratio estimation from the PHR model with 200, 500, and 1000 replications of both the bootstrap and KRT and the results from the Maximum Likelihood applied to the original Localized Colon Carcinoma data. From Table 6 we can see that the KRT estimates were comparable to the estimates for both *hazard ratio* and *β for Year of Diagnosis* from the bootstrap and asymptotic estimates. With this in mind, it is evident that the standard errors from the KRT resamples were systematically smaller. The estimation biases were consistently less in most cases than those from both the bootstrap procedure and the conventional maximum likelihood method across of 200, 500, and 1000 replications. The 95% confidence intervals (percentiles) for the estimates using the KRT procedure were narrower than those from both the bootstrap procedure and the conventional maximum likelihood estimates. Methodologically, the evaluation results from the cross-validating sample were consistent with the results from Study 1 from the educational data; therefore, the findings of the KRT application to PHR model were cross-validated and proved to be replicable.

Table 6. Comparisons of Estimates, CIs, and Bias of Cox Model with the Conventional Asymptotic, Bootstrap, and KRT Methods

	Estimates	Replicates	Estimate	*SE*	CI (2.5%)	CI (97.5%)	Bias
Hazard Ratio	ML		0.7930	0.0383	0.7200	0.8730	
	Bootstrap	200	0.7947	0.0401	0.7246	0.8748	0.0017
		500	0.7936	0.0389	0.7194	0.8704	-0.0012
		1000	0.7932	0.0386	0.7201	0.8728	-0.0004
	KRT	200	0.7938	0.0165	0.7651	0.8278	0.0007
		500	0.7930	0.0154	0.7653	0.8240	-0.0008
		1000	0.7945	0.0153	0.7688	0.8243	0.0015
Year of Diagnosis	ML		-0.2310	0.0503	-0.3222	-0.1338	
	Bootstrap	200	-0.2321	0.0492	-0.3311	0.1334	-0.0011
		500	-0.2324	0.0490	-0.3293	-0.1388	-0.0003
		1000	-0.2329	0.0487	-0.3284	-0.1360	-0.0005
	KRT	200	-0.2296	0.0205	-0.2683	-0.1920	0.0033
		500	-0.2322	0.0180	-0.2653	-0.1961	-0.0027
		1000	-0.2329	0.0202	-0.2718	-0.1919	-0.0007

Discussion and Further Study

Using data from different research areas, the findings from two studies provide strong evidence that the KRT outperformed both the bootstrap and the Maximum Likelihood method in the PHR parameter estimation. The application of KRT in PHR provided more accurate confidence interval estimation with narrower bands, smaller standard errors with less or comparable biases, and equivalent accurate point estimates. The KRT procedure produced stable estimation results across various replications. KRT application to PHR provides a solution for "a single method appropriate for all parameters" (Burr, 1994, p. 1301). This study produced preliminary results of the KRT application in PHR models for parameter estimation. The findings suggest that applications of KRT to PHR models improve the accuracy of parameter estimation for more valid statistical inference in survival research.

Future studies are desired to compare the results of other types of confidence interval estimation. In the current study, only empirical datasets were used to study the performance of the application of the KRT in a PHR model. Even though the cross-validating study provided strong evidence of the current study findings, a simulation study is expected to provide more information and further confirmation of the study results in terms of the stability of the findings under other conditions. Future studies should engage in (1) comparison of the results of other types of confidence interval estimation and (2) simulation studies with different data conditions (e.g., sample sizes or distributions) to explore the stability of the application results.

Significance of the Study

In education, teachers' survival, students' dropout, and on-time graduation are all important factors influencing the quality of education. Understanding these factors is crucial for educators and educational administrators to work on effective solutions. PHR is an appropriate and effective statistical analytical tool for studies in such areas, and applications of KRT to PHR will improve the accuracy of parameter estimation to provide more valid statistical inference in educational research.

Acknowledgement

Sincere thanks to Drs. Paul Dickman and Timo Hakulinen for providing the medical data.

References

Adams, G. J. (1996). Using a PHR regression model to examine voluntary teacher turnover. *Journal of Experimental Education*, *643,* 267-285.

Adams, G. J., & Dial, M. (1993). Teacher survival a PHR regression model. *Education and Urban Society, 26*, 90-99.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage Publications.

Altman, D. G., & Andersen, P. K. (1989). Bootstrap investigation of the stability of a PHR regression model. *Statistics in Medicine, 8*, 771-783.

Bai, H. (2008). Kernel resampling to improve the performance of multiple regression with small samples. *Journal of Applied Statistical Science, 17*, 573-582.

Bai, H., & Pan, W. (2009). An application of a new multivariate resampling method to multiple regression. *Multiple Linear Regression Viewpoints*, *35*, 1-5.

Battin-Pearson, S., Newcomb, M.D., Abbott, R. D., Hill K. G., Catalano R. F., & Hawkins J. D. (2000). Predictors of early high school drop-out: a test of five theories. *Journal of Educational Psychology*. *92*, 568–582.

Burr, D. A. (1994). A comparison of certain bootstrap confidence intervals in the PHR model. *Journal of the American Statistical Association*, *89*, 1290-1302.

Chen, S. H., & George, S. L. (1985). The bootstrap and identification of prognostic factors via PHR's proportional hazards regression model. *Statistics in Medicine, 4*, 39–46,

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B, 34,* 187–220.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall, Inc.

Fagan, J., & Pabon, E. (1990). Contributions of delinquency and substance use to school dropout among inner-city youth. *Youth & Society, 21,* 306-354.

Gonzalez, J. R., Pena, A. E., & Delicado, P. (2010). Confidence intervals for median survival time with recurrent event data. *Computational Statistics & Data Analysis*, 54, 78-89

Hauser, R. M., Simmons, S., & Pager, D. (2000). *High school dropout, race-ethnicity and social background from the 1970s to the 1990s*. Retrieved from http://www.ssc.wisc.edu/ ~hauser /Trends2001_03.pdf.

Hjort, H. (1985). *Boostrapping PHR's regression model*. Technical Report. Stanford University.

Krohn, M. D., Thornberry, T. P., Collins-Hall, L., & Lizotte, A. J. (1995). School dropout, delinquent behavior, and drug use: An examination of the causes and consequences of dropping out of school. In H. B. Kaplan (Ed.), *Drugs, crime, and other deviant adaptations: Longitudinal studies* (pp. 163-183). New York: Plenum Press.

Neild, R. C., Stoner -Eby, S., & Furstenberg, F. (2008). Connecting entrance and departure: The transition to ninth grade and high school dropout. *Education and Urban, 40*, 543-569.

Plank, S., DeLuca, S., & Estacion, A. (2008). High school dropout and the role of career and technical education: A survival analysis of surviving high school. *Sociology of Education*, *81*, 345-370.

Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research*, *57*, 101-121.

SAS Institute Inc. (2008). *SAS/STAT user's guide, Version 9.2.* Cary, NC: SAS Institute Inc.

Sauerbrei, W., & Schumacher M. (1992). A bootstrap resampling procedure for model building: Application to the PHR regression model. *Statistics in Medicine*, *11*, 2093–2109.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Simonoff, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics, 9*, 1187-1195.

Send correspondence to: Haiyan Bai
University of Central Florida
Email: haiyan.bai@ucf.edu