⊠연구논문

A Study on Estimating Mean Lifetime After Modifying Censored Observations

Jinheum Kim
Dept. of Applied Statistics, Suwon University
Jeehoon Kim
Statistical Research Institute, Seoul National University

Abstract

Kim and Kim (1997) developed a method of estimating the mean lifetime based on the augmented data after imputing censored observations. Assuming the linear relationship between lifetime and covariates, and then introducing the procedure of Buckley and James (1979) to estimate the mean lifetimes of censored observations, they proposed a mean lifetime estimator and its consistency under the regularity conditions. In this article, the Kim and Kim's estimator is compared with the estimator introduced by Gill (1983) through simulations under the various configurations. Also, their estimator is illustrated with two real data sets.

1. Introduction

Under censorship, the researchers in reliability has been interested in estimating the mean lifetime of a component or a system. The various methods of estimating it have been proposed by Yang (1977), Susarla and Van Ryzin (1980), Gill (1983), Kumazawa (1987), and so forth. Those methods are not to modify censored incomplete observations. In many lifetime testing of reliability field, however, the covariate information related to components or systems is often observed together with those times to failure or censoring. In these situations we may consider an estimator different from the previous estimators by using the covariate information in modifying the censored observations.

Kim and Kim (1997) developed a method of estimating the mean lifetime based on the augmented data after imputing censored observations. Assuming the linear relationship between lifetime and covariates, and then introducing the procedure of Buckley and James (1979) to estimate the mean lifetimes of censored observations, they proposed a mean lifetime estimator and showed its consistency under the regularity conditions.

In the next section we briefly review the Kim and Kim's estimator. Section 3 performs simulation studies under the various configurations to examine the relative performance of Kim and Kim's estimator to the estimator introduced by Gill (1983). Finally, the Kim and Kim's estimation procedure is illustrated with two real data sets in Section 4.

2. Kim and Kim's Estimator

Suppose that the observable data consist of independent and identical (iid) random vectors (X_i, Z_i, Δ_i) $(i=1, \ldots, n)$, where $X_i = Y_i \wedge C_i$, the Z_i is a covariate, and $\Delta_i = I(Y_i \le C_i)$. Here, the Y_i is a transformation of true lifetime and the C_i is a censoring time. $\wedge$ denotes the minimum and I is the indicator function. Suppose that the transformed lifetime Y_i is linearly related to the covariate Z_i such as

$$Y_i = \alpha_0 + \beta_0 Z_i + \varepsilon_i, \quad i=1, \ldots, n,$$

where the ε_i are iid with mean 0, independent of Z_i and C_i.

For any real b, let $\eta_i^b = Y_i - bZ_i$ and $\xi_i^b = C_i - bZ_i$. Define the product-limit estimator $\hat{F}_b$ based on $\{(\eta_i^b \wedge \xi_i^b, \Delta_i), i=1, \ldots, n\}$ as

$$\hat{F}_b(t) = 1 - \prod_{s \le t} \left\{ 1 - \frac{\Delta N_b(s)}{Y_b(s)} \right\},$$

where $N_b(t) = \sum_{i=1}^{n} I(\eta_i^b \wedge \xi_i^b \le t, \Delta_i = 1)$, $\Delta N_b(s) = N_b(s) - N_b(s-)$, and

$Y_b(t) = \sum_{i=1}^{n} I(\eta_i^b \wedge \xi_i^b \ge t)$.

Kim and Kim (1997) proposed to estimate the mean lifetime $E(Y_i \mid Y_i > C_i, Z_i)$ as

$$\hat{E}_b(Y_i|Y_i > C_i, Z_i) = C_i + \int_{s > \xi_i^b} \{1 - \hat{F}_b(s)\} ds / \{1 - \hat{F}_b(\xi_i^b)\}.$$

Also, they estimated β_0 by iterative solution of the equation

$$b = \left\{ \sum_{i=1}^{n} \hat{Y}_i(b)(Z_i - \bar{Z}) \right\} / \sum_{i=1}^{n} (Z_i - \bar{Z})^2,$$

where $\hat{Y}_i(b) = Y_i\Delta_i + \hat{E}_b(Y_i|Y_i > C_i, Z_i)(1-\Delta_i)$ and $\bar{Z} = n^{-1}\sum_{i=1}^{n} Z_i$. Since for the limiting value $\hat{\beta}$ of b, $\hat{Y}_i(\hat{\beta})$ does not depend on censorship, Kim and Kim (1997) proposed an estimator of mean lifetime $E(Y)$ as

$$\hat{E}(Y) = n^{-1} \sum_{i=1}^{n} \hat{Y}_i(\hat{\beta}),$$

and showed its consistency under the regularity conditions as follows. In the proofs of the following theorem and lemmas in Appendix, we assume, without loss of generality, that $\beta_0 = 0$ and Z has support on $[0, 1]$.

Theorem. Let $\hat{\beta}$ be any $\sqrt{n}$-consistent estimator of β_0, and assume $P(\sup_n|\hat{\beta}| < \infty) = 1$ and that C has the finite variance. Then,

$$\hat{E}(Y) \rightarrow_p E(Y), \quad \text{as } n \rightarrow \infty,$$

where $\rightarrow_p$ denotes convergence in probability.

Proof. First, note that

$$\hat{Y}_i(\hat{\beta}) = \Delta_i Y_i + (1-\Delta_i)C_i + (1-\Delta_i)\int_{\xi_i\hat{\beta}}^{\infty} \{1 - \hat{F}_{\hat{\beta}}(u)\} du / \{1 - \hat{F}_{\hat{\beta}}(\xi_i^{\hat{\beta}})\}, \quad (1)$$

where the last term is defined as 0 whenever its denominator is equal to 0. As $n \rightarrow \infty$, by the weak law of large numbers,

$$n^{-1}\sum_{i=1}^{n} \Delta_i Y_i \rightarrow_p E(\Delta Y),$$

$$n^{-1}\sum_{i=1}^{n}(1-\Delta_i)C_i \rightarrow_p E\{(1-\Delta)C\}.$$

For the last term in (1),

$$\left| n^{-1}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i\hat{\beta}}^{\tau_{F_0}+|\hat{\beta}|}\{1-\hat{F}_{\hat{\beta}}(u)\}du}{1-\hat{F}_{\hat{\beta}}(\xi_i^{\hat{\beta}})} - E\left[(1-\Delta)\frac{\int_{\xi^0}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi^0)}\right]\right|$$

$$\leq I+J+K,$$

where

$$I=\left| n^{-1}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i\hat{\beta}}^{\tau_{F_0}+|\hat{\beta}|}\{1-\hat{F}_{\hat{\beta}}(u)\}du}{1-\hat{F}_{\hat{\beta}}(\xi_i^{\hat{\beta}})} - n^{-1}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i\hat{\beta}}^{\tau_{F_0}+|\hat{\beta}|}\{1-F_0(u)\}du}{1-F_0(\xi_i^{\hat{\beta}})}\right|,$$

$$J=\left| n^{-1}\sum_{i=1}^{n}(1-\Delta_i)\frac{\int_{\xi_i\hat{\beta}}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi_i^{\hat{\beta}})} - E\left[(1-\Delta)\frac{\int_{\xi\hat{\beta}}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi^{\hat{\beta}})}\right]\right|,$$

$$K=\left| E\left[(1-\Delta)\frac{\int_{\xi\hat{\beta}}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi^{\hat{\beta}})}\right] - E\left[(1-\Delta)\frac{\int_{\xi^0}^{\tau_{F_0}}\{1-F_0(u)\}du}{1-F_0(\xi^0)}\right]\right|.$$

In particular,

$$I\leq sup_{t<\tau_{F_0}+|\hat{\beta}|}\left|\frac{\int_{t}^{\tau_{F_0}+|\hat{\beta}|}\{1-\hat{F}_{\hat{\beta}}(u)\}du}{1-\hat{F}_{\hat{\beta}}(t)} - \frac{\int_{t}^{\tau_{F_0}+|\hat{\beta}|}\{1-F_0(u)\}du}{1-F_0(t)}\right|$$

$$\leq sup_{\tau_{F_0}\leq t<\tau_{F_0}+|\hat{\beta}|}\left|\frac{\int_{t}^{\tau_{F_0}+|\hat{\beta}|}\{1-\hat{F}_{\hat{\beta}}(u)\}du}{1-\hat{F}_{\hat{\beta}}(t)}\right|$$

$$+ sup_{t<\tau_{F_0}} \left| \frac{\int_{\tau_{F_0}}^{\tau_{F_0}+|\widehat{\beta}|} \{1-\widehat{F}_{\widehat{\beta}}(u)\}du}{1-\widehat{F}_{\widehat{\beta}}(t)} \right|$$

$$+ sup_{t<\tau_{F_0}} \left| \frac{\int_{t}^{\tau_{F_0}} \{1-\widehat{F}_{\widehat{\beta}}(u)\}du}{1-\widehat{F}_{\widehat{\beta}}(t)} - \frac{\int_{t}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(t)} \right|$$

$$= I_1 + I_2 + I_3 .$$

Here, $I_1 \le |\widehat{\beta}|$, $I_2 \le |\widehat{\beta}|$, and using Lemma A.3 in Appendix, $I_3 \rightarrow_p 0$ as $n \rightarrow \infty$. On the other hand, by Chebychev's inequality, for any $\varepsilon > 0$,

$$P(|J| > \varepsilon) \le \frac{1}{n^2\varepsilon^2}\{nJ_1 + n(n-1)J_2\},$$

where

$$J_1 = Var\left[(1-\Delta_1)\frac{\int_{\xi_1^{\widehat{\beta}}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi_1^{\widehat{\beta}})}\right],$$

$$J_2 = Cov\left[(1-\Delta_1)\frac{\int_{\xi_1^{\widehat{\beta}}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi_1^{\widehat{\beta}})}, (1-\Delta_2)\frac{\int_{\xi_2^{\widehat{\beta}}}^{\tau_{F_0}} \{1-F_0(u)\}du}{1-F_0(\xi_2^{\widehat{\beta}})}\right].$$

Here, $\lim_{n\to\infty} J_1 < \infty$, $\lim_{n\to\infty} J_2 = 0$ by the boundedness of $\widehat{\beta}$, the finite variance of C, and Theorem 5.4 (Billingsley (1968), p. 32). Similarly, $K \rightarrow 0$ as $n \rightarrow \infty$. ■

Remark. In the multiple regression $\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{Z}_{(n\times p)}{}^T \boldsymbol{\beta}_{0(p\times 1)} + \boldsymbol{\varepsilon}_{(n\times 1)}$, the estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ can be obtained by iteratively solving

$$\boldsymbol{b} = (\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \widehat{\boldsymbol{Y}}(\boldsymbol{b}),$$

where $\widehat{\boldsymbol{Y}}(\boldsymbol{b}) = (\widehat{Y}_i(\boldsymbol{b}))_{(n\times 1)}$, $\widehat{Y}_i(\boldsymbol{b}) = Y_i\Delta_i + \widehat{E}_{\boldsymbol{b}}(Y_i | Y_i > C_i, Z_i)(1-\Delta_i)$, and

$\hat{E}_b(Y_i | Y_i > C_i, Z_i)$ is the multiple version of the simple regression [James and Smith, 1984]. The procedure of estimating $E(Y)$ in the simple regression is similarly extended to the multiple regression, in which the multiple version estimator of $E(Y)$ is given as follows;

$$\hat{E}(Y) = n^{-1} \sum_{i=1}^{n} \hat{Y}_i(\hat{\beta}).$$

3. Small Sample Monte Carlo Studies

Monte Carlo studies were carried out to assess the relative performance of Kim and Kim's estimator to Gill's for practical sample sizes. Here, the Gill's estimator is defined by $\int_0^{\max_i X_i} \{1 - \hat{F}(s)\} ds$, where $\hat{F}(\cdot)$ is the product-limit estimator based on $\{(X_i, \Delta_i),\ i = 1, \ldots, n\}$ [Kaplan and Meier, 1958]. The following tables include the values of bias, means squared error (mse), and ratio of mse of Gill's estimate to mse of Kim and Kim's (ratio). All values of bias and mse are the means ($\times 10^4$) of 1,000 replicates.

In simulation studies we consider a linear regression model with a single covariate. The covariate z_i was generated from the standard normal distribution truncated at ± 5 and the error term ε_i from a normal distribution with mean 0 and variance 0.25. Let the true values for α_0 and β_0 be 0 and 1. The censoring variables was considered as $U(c_1, c_2)$, where c_1 and c_2 were chosen to yield the prespecified proportion of censoring. As the sample size is larger and the proportion of censoring is smaller, both bias and mse of two estimators gradually decrease. Specially, the entries in the last column of <Table 1> show that the Kim and Kim's estimator is always preferred to Gill's.

We also conduct simulations to examine how effective the Kim and Kim's procedure is with respect to Type I censorship usually performed in lifetime testing of reliability field. The covariate z_i was generated from the $U(-2, 2)$ and the error term ε_i from a standard normal distribution. Let α_0 and β_0 be equal to 0 and 1 respectively. The censoring distribution was assumed to be

$U[c, c]$ having constant value of c, where c was chosen to yield the prespecified proportion of censoring. <Table 2> implies that the Kim and Kim's estimator is substantially preferred; both its bias and mse are consistently smaller than those of Gill's estimator.

< Table 1 > Comparison of Gill and Kim and Kim Estimators of Mean Lifetime Based on 1,000 Replications in Each Configuration Under Random Censorship

sample size	censoring proportion	Gill		Kim and Kim		ratio
		bias	mse	bias	mse	
50	[a]25%	-40	296	-19	258	1.148
	[b]50%	-10	413	-198	290	1.423
	[c]70%	558	696	-694	393	1.772
100	[a]25%	-7	152	-2	132	1.149
	[b]50%	12	220	-82	157	1.400
	[c]70%	875	456	-566	214	2.128

Note: a. $U(-2.5, 7.5)$ censoring distribution.
b. $U(-5, 5)$ censoring distribution.
c. $U(-7.5, 2.5)$ censoring distribution.

In conclusion the Kim and Kim's estimator has better performance than Gill's with no concern for modifications of censored observations. This trend is remarkably found out under the constant censoring distribution like Type I censorship. As the censoring proportion is heavier, the Kim and Kim procedure is quite effective but negatively biased. On the contrary, in the additional simulation studies not reported here, the proposed estimator has trend to be positively biased under the light' censorship.

< Table 2 > Comparison of Gill and Kim and Kim Estimators of Mean Lifetime Based on 1,000 Replications in Each Configuration Under Type I Censorship

sample size	censoring proportion	Gill		Kim and Kim		ratio
		bias	mse	bias	mse	
50	[a]25%	-2362	859	42	566	1.516
	[b]50%	-6158	3927	-114	582	6.749
	[c]70%	-12547	15797	-3985	2203	7.172
100	[a]25%	-2376	711	-13	255	2.786
	[b]50%	-6238	3969	-107	335	11.858
	[c]70%	-12512	15682	-3363	1410	11.125

Note : a. $U[1, 1]$ censoring distribution.
b. $U[0, 0]$ censoring distribution.
c. $U[-1, -1]$ censoring distribution.

4. Real Examples

In this section we illustrate the Kim and Kim's estimator with two real data sets. The first one is taken from Gertsbakh (1989, p. 206). During the development of a new type of motor, the Manoa company had decided to carry out a lifetime test by combining two levels of two factors, the temperature and the combined load index (CLI). For this purpose, four samples of 10 motors were randomly selected from the assembly line and then each subsample was tested under one of four combinations of rescaled covariates; $(z_1, z_2)=(-1,-1),(-1,+1),(+1,-1),(+1,+1)$, where z_1 and z_2 are the covariates corresponding to CLI and temperature respectively. The censoring was of Type I for each of four subsamples. We fit the multiple regression $\beta_1(\text{ CLI})+\beta_2(\text{ temperature})$ of the lifetime of motor against CLI and temperature. The Buckley-James estimates of regression parameters are obtained as $\widehat{\beta}_1$=0.2361(0.1092) and $\widehat{\beta}_2$=0.4340(0.1083), where the values in parenthesis are the estimated standard errors of estimates. The significance of effects of two covariates indicates the appropriateness of assumed model. In the meantime, when this resulting model is used to modify the censored observations, the Kim and Kim estimate of mean lifetime of motor is 5.995 and its estimated standard error based on 100,000 Bootstrap samples, 0.1271.

As another example, we consider the updated Stanford heart transplant data as of February 1980 [Miller and Halpern, 1982]. This data set contains the lifetimes of 184 heart-transplanted patients along with their ages at the time of the first transplant and T5 mismatch scores. The 27 patients who did not have T5 mismatch scores are excluded from this analysis. Out of the remaining 157 patients, 55 were censored as of February 1980. This data set was analyzed by many authors including Miller and Halpern (1982), Wei, Ying and Lin (1990), and so forth. Miller and Halpern (1982) first regressed the base 10 logarithm of the lifetime on age and T5 mismatch score. They founded out that T5 mismatch score is not significant and also the goodness-of-fit tests provide strong evidence to discredit this model. In an attempt to achieve a better fit, a quadratic age model without T5 mismatch score was tried by Miller and Halpern. For their analysis, Miller and Halpern (1982) deleted the 5 patients with survival times less than 10 days to symmetrize the data. However, since there exists the strong multicollinearity between age and age^2, in order to avoid this, Wei et al. (1990) proposed to fit the quadratic age model $\beta_1(\text{ age}-42)+\beta_2(\text{ age}-42)^2$ with age being centered around 42, the approximate sample mean of the patients' ages. For the resulting model, the Buckley-James estimates of regression parameters are $\hat{\beta}_1$ =-0.0333(0.0074) and $\hat{\beta}_2$=-0.0017(0.0005). Also, $\hat{E}(Y)$ is equal to 2.716 with the Bootstrap estimate of standard error, 0.0788.

References

[1] Billingsley, P.(1968), *Convergence of Probability Measures*, Wiley, New York.

[2] Buckley, J. and James, I.R.(1979), "Linear Regression with Censored Data," *Biometrika*, Vol. 66, pp. 429-436.

[3] Gertsbakh, I.B.(1989), *Statistical Reliability Theory*, Marcel Dekker, New York.

[4] Gill, R.D.(1983), "Large Sample Behavior of the Product-Limit Estimator on the Whole line," *The Annals of Statistics*, Vol. 11, pp. 49-58.

[5] James, I.R. and Smith, P.J.(1984), "Consistency Results for Linear Regression With Censored Data," *The Annals of Statistics*, Vol. 12, pp. 590-600.

[6] Kaplan, E.L. and Meier, P.(1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, Vol. 53, pp. 457-481.

[7] Kim, J.H. and Kim, J.H.(1997), "Mean Lifetime Estimation with Censored Observations," *Journal of the Korean Statistical Society*, Vol. 26, pp. 299-308.

[8] Kumazawa, Y.(1987), "A Note on an Estimator of Life Expectancy with Random Censorship," *Biometrika*, Vol. 74, pp. 655-658.

[9] Miller, R. and Halpern, J.(1982), "Regression with Censored Data," *Biometrika*, Vol. 69, pp. 521-531.

[10] Susarla, V. and Van Ryzin, J.(1980), "Large Sample Theory for an Estimator of the Mean Survival Time from Censored Samples," *The Annals of Statistics*, Vol. 8, pp. 1002-1016.

[11] Wei, L.J., Ying, Z. and Lin, D.Y.(1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests," *Biometrika*, Vol. 77, pp. 845-851.

[12] Yang, G.(1977), "Life Expectancy Under Random Censorship," *Stochastic Processes and Their Applications*, Vol. 6, pp. 33-39.

Appendix

Lemma A.1 Let $\hat{\beta}$ be any $\sqrt{n}$-consistent estimator of β_0.

(i) For any τ such that $F_0(\tau)<1$,

$$sup_{t\leq\tau} | \hat{F}_{\hat{\beta}}(t)-F_0(t) | \to_p 0, \text{ as } n\to\infty.$$

(ii) $sup_{t\leq\tau_{F_0}} | \hat{F}_{\hat{\beta}}(t)-F_0(t) | \to_p 0$, as $n\to\infty$.

Proof. (i) Since $F_0(\tau)<1$, Z has support on $[0,1]$, and F_0 is continuous, $P(Y-\beta Z<\tau)=E\{F_0(\beta Z+\tau)\}\leq F_0(|\beta|+\tau)<1$ for any β in some neighborhood N_0 of 0. Since $\tau_{F_0}\leq\tau_{G_0}$, $P(Y\wedge C-\beta Z<\tau)<1$ for any $\beta\in N_0$. Let $\{\beta_n\}$ be any nonrandom sequence converging to 0. Then, using the Lemma 7.1 (i) and 7.2 (i) of Ritov (1990),

$$sup_{t\leq\tau} | \hat{F}_{\beta_n}(t)-F_0(t) | \to_p 0, \text{ as } n\to\infty.$$

Since $\hat{F}_\beta(t)$ is monotone in β, for any sequence $\{b_n\}$ with $b_n\geq 0$, $\lim_{n\to\infty} b_n=0$, and $\lim_{n\to\infty}\sqrt{n}b_n=\infty$,

$$sup_{t\leq\tau,\ |\beta|\leq b_n} | \hat{F}_\beta(t)-F_0(t) | \to_p 0, \text{ as } n\to\infty,$$

and so (i) holds. ■

(ii) Our proof is similar to the Kim and Kim's (1997) arguments. For given $\varepsilon>0$, choose $\tau<\tau_{F_0}$ such that $1>F_0(\tau)>1-\varepsilon$. Now, for $t\in(\tau,\tau_{F_0}]$,

$$\hat{F}_{\hat{\beta}}(\tau)\leq \hat{F}_{\hat{\beta}}(t)\leq 1,\ 1-\varepsilon<F_0(\tau)\leq F_0(t)\leq 1.$$

Therefore,

$$sup_{\tau\leq t\leq\tau_{F_0}} | \hat{F}_{\hat{\beta}}(t)-F_0(t) | \leq\max\{\varepsilon, 1-\hat{F}_{\hat{\beta}}(\tau)\}.$$

Combining this, (i) and the arbitrariness of $\varepsilon>0$, (ii) holds. ■

Lemma A.2 Let $\hat{\beta}$ be any $\sqrt{n}$-consistent estimator of β_0. For any $\tau < 0$,

$$\int_{-\infty}^{\tau} \hat{F}_{\hat{\beta}}(u)du \rightarrow_p \int_{-\infty}^{\tau} F_0(u)du, \quad \text{as } n \rightarrow \infty.$$

Proof. Let $\{\beta_n\}$ be any nonrandom sequence converging to 0. Then

$$\int_{-\infty}^{\tau} \hat{F}_{\beta_n}(u)du = \tau \hat{F}_{\beta_n}(\tau) - \int_{-\infty}^{\tau} ud\,\hat{F}_{\beta_n}(u) \rightarrow_p \tau F_0(\tau) - \int_{-\infty}^{\tau} ud\,F_0(u)$$

by Lemma 7.1 (ii) and 7.2 (ii) of Ritov (1990), and Lemma A.1. Since $\hat{F}_{\beta}(t)$ is monotone in β, this lemma holds by the same arguments as used in Lemma A.1 (i). ■

Lemma A.3 Let $\hat{\beta}$ be any $\sqrt{n}$-consistent estimator of β_0. Then,

$$sup_{t < \tau_{F_0}} \left| \frac{\int_t^{\tau_{F_0}} \{1 - \hat{F}_{\hat{\beta}}(u)\}du}{1 - \hat{F}_{\hat{\beta}}(t)} - \frac{\int_t^{\tau_{F_0}} \{1 - F_0(u)\}du}{1 - F_0(t)} \right| \rightarrow_p 0, \quad \text{as } n \rightarrow \infty.$$

Proof. James and Smith (1984) proved the above result when $\hat{\beta} = \beta_0$. But their procedures can be exactly applied to this lemma by using Lemma A.1 and A.2. We omit the details for this reason. ■