*The Annals of Statistics*
1996, Vol. 24, No. 6, 2705–2719

# EMPIRICAL PROCESS APPROACH IN A TWO-SAMPLE LOCATION–SCALE MODEL WITH CENSORED DATA[1]

BY FUSHING HSIEH

*National Taiwan University*

To compare two samples of possibly right-censored failure times, a location–scale model, without assuming the distribution form, is considered for the log-transformed data. This new accelerated failure time model is introduced to accommodate the possible heterogeneity between within-treatment variations of the two groups considered. One distinct feature of this model is that, with the presence of heterogeneity, statistical inferences on both location and scale parameters based on log-rank tests or linear rank tests will become inappropriate. In this paper we propose the empirical process approach to construct a regression setup derived from the theory of strong approximation of the Kaplan–Meier product–limit empirical quantile process. A generalized least squares (GLS) estimator is obtained and shown to be semiparametric efficient. Also it is shown that this estimation is adaptive for a special case. At the end, results on the two-sample setting are applied to the $K$-sample problem.

**1. Introduction.** To evaluate treatment effects on the hazard function in comparative survival and life testing studies, the Cox proportional hazard model has been the most successful and widely used model for the past two decades. On the other hand, as a useful alternative, the accelerated failure time model, which regresses the logarithm of the failure time over the covariates in the same studies, bears the treatment effects directly related to the time scale of the response variables. [See Wei (1992) for a nice review on this model.] For the latter model, recently theoretical justifications for the linear rank tests employed as estimating equations are intensively and successfully studied in Tsiatis (1990), Ritov (1990) and Ying (1993), among many others. Given further advances in computational related issues, more applications of this linear regression model with censored data could be expected in failure time related research.

However, most of the models considered in the literature related to this topic are homogeneous models in the sense that the error variables in the linear regression model are assumed to be identically distributed with unknown distribution form. This assumption could be very unrealistic. From the parsimonious point of view, the unequal within-treatment variations corresponding to different covariates are frequently encountered in medical research. This has been emphasized in O'Brien (1988). Further, in contrast to

Received August 1994; revised October 1995.
[1]This research was supported in part by the National Science Council of Taiwan.
*AMS* 1991 *subject classifications*. 62G05, 62G20.
*Key words and phrases.* Generalized Kiefer process, generalized least squares estimator, product–limit empirical quantile process, semiparametric ANOVA.

heteroscedastic normal regression studies, the homogeneity assumption could become a limitation for its potential applications, especially when a prediction of survival probability is needed given a specific covariate.

The models considered here will allow for unequal within-treatment variations between different groups. A treatment variable having this type of inequality is said to have heterogeneous effects. In this paper, we will focus on the two-sample and $K$-sample problems, while the heterogeneous type of censored regression model is studied and reported in Hsieh (1996) in which a totally different way of constructing empirical processes adapting to the regression setup is required.

More specifically, for the two-sample problem, the accelerated failure time models discussed in Louis (1981) and Wei and Gail (1983) are homogeneous models in the sense that each becomes a two-sample location shift model after having log-transformation on the data. This can be seen from the following generalized model, which will be kept in mind throughout this paper,

$$T_1 = (T_0/\lambda)^{\gamma}, \tag{1}$$

where $T_1$ and $T_0$ are random variables of failure times in two treatment groups, respectively. After log-transformation, model (1) is equivalent to the following two-sample location–scale model:

$$\log T_1 = \gamma(\log T_0 - \log \lambda). \tag{2}$$

In terms of distributions, the above model is also equivalent to assuming

$$F_1(x) = F_0\left(\frac{x - \mu}{\sigma}\right), \tag{3}$$

where $F_i$ is the distribution function of random variable $\log T_i$, $i = 0, 1$, and $\mu = \log \lambda$ and $\sigma = 1/\gamma$. If $\gamma = 1$, it becomes the location–shift model and it is called the scale–change model in Wei and Gail (1983). Hence a treatment variable having heterogeneous effects means that $\sigma \neq 1$. Parsimoniously and practically, it is necessary to test the homogeneity hypothesis $\sigma = 1$ before the location–shift model can be further used.

One significant feature of this paper is that the inference techniques employed in the aforementioned references based on log-rank test or linear rank test statistics turn out to be inappropriate for the location–scale model considered here. The reason is that the pooled rank vector or order statistics of the two samples generated from two distributions with different location and scale parameters will not carry useful information of either location or scale parameters. [See Hájek and Šidák (1967), page 94, or Moses (1963).] More precisely speaking, the linear rank test statistics used as estimating equations in Tsiatis (1990) and Ying (1993), in general, will not be a martingale with the presence of the heterogeneity. Therefore, different estimation approaches are called for the inference in this location–scale model.

The estimating approach proposed here is based on the theory of empirical processes and is called the empirical process approach (EPA). The first step of this approach is to construct a regression setup based on the strong approxi-

mations of the Kaplan–Meier product–limit (PL) empirical quantile process pertaining to the two samples of log-transformed life times. [For a comprehensive account of this theory, see Chapter 8 of Csörgő (1983).] The second step is estimate the parameters of interest $\mu$ and $\sigma$ by using the generalized least squares (GLS) estimator, a type of weighted least squares estimator. For more applications of the EPA approach, see Hsieh (1995), Hsieh and Turnbull (1996) and Hsieh and Turnbull (1996).

From the aspects of computation, the EPA approach needs only the estimate of a density quantile function in the construction of a covariance matrix for the weighting purpose and for the asymptotic precision of the GLS estimator derived from this approach. This relative simplicity certainly gives GLS estimation advantages over the $M$-estimator approach proposed in Ritov (1990) if an efficient estimator is considered. The reason is that, for obtaining an efficient estimator, estimated score functions are involved with the estimation of the density function as well as its derivative. Hsieh and Manski (1987), in their Monte Carlo studies for complete data, showed that the adaptive estimator constructed by using the score estimate is severely sensitive to the smoothing parameters. Here we expect that even more severe sensitivity will be encountered with the presence of censored data.

In regard to the asymptotical properties, this estimation problem is shown to be adaptive in the sense of Bickel (1982) when condition (22) holds, in which the two censoring mechanisms are assumed to give the same censoring schemes on the two samples. Further, in general, the GLS estimator is shown to be semiparametric efficient in the sense of achieving the Fisher information bound.

Further the results of the two-sample problem are easily extended to the $K$-sample problem. The key point for this extension being possible is that independent Kaplan–Meier PL empirical quantile processes will result in independent generalized Kiefer processes in their strong approximations. So that, by taking one treatment group as a baseline, there are $K - 1$ pairs of GLS estimators and their covariance matrix can be derived without much effort. This type of problem is important in its own right. In particular, the semiparametric version of analysis of variance (ANOVA) with censored data should be of interest from a statistical application point of view.

The paper is organized as follows. In Section 2, the empirical process approach is described and the GLS estimator is derived and its asymptotic properties are also given. A small simulation study is reported for comparing the performance of the GLS estimator and the log-rank estimator in both homogeneous and heterogeneous cases. In Section 3, the asymptotical variance and the semiparametric efficiency of the GLS estimator are discussed. In Section 4, the EPA approach is applied to $K$-sample problems. All the proofs of theorems and corollaries are given in the Appendix.

**2. The empirical process approach.** There are two sets of random samples $(X_{ij}, C_{ij}) \in R \times R$, $i = 0, 1$, with sample sizes $n_0$ and $n_1$, respectively, and we observe $(X_{ij} \wedge C_{ij}, \Delta_{ij})$, where $\Delta_{ij} = 1_{(X_{ij} \le C_{ij})}$. Here the ran-

dom variable $X_{ij}$, $i = 0, 1$, can be $\log T_i$, $i = 0, 1$, or any other random variables even not related to life time at all. We assume that the joint density of $(X_{ij}, C_{ij})$ relative to the Lebesgue measure on $R \times R$ is given by $f_i(X_{ij})h_i(C_{ij})$. Let $F_i$ and $H_i$, $i = 0, 1$, be distribution functions with density functions $f_i$ and $h_i$, respectively. Denote the Kaplan–Meier product–limit (PL) estimators $\tilde{F}_{i,n_i}$ of $F_i$ pertaining to the set of observable samples $\{(X_{ij} \wedge C_{ij}, \Delta_{ij}),\ 1 \le j \le n_i\}$. And the PL empirical quantile functions $\tilde{F}_{i,n_i}^{-1}$ are defined by

$$\tilde{F}_{i,n_i}^{-1}(u) = \inf\left\{t\colon \tilde{F}_{i,n_i}(t) \ge u\right\}, \qquad 0 < u < 1.$$

Furthermore, the notion of a generalized Kiefer process $\{K(s,t),\ 0 \le s \le 1,\ t \ge 0\}$ is recalled here from Csörgő (1983), page 118. The latter is a separable Gaussian process such that:

$$\begin{aligned} &\text{(i)} \qquad K(0,t) = K(1,t) = K(s,0) = 0 \quad \text{for } s \in [0,1],\ t \ge 0, \\ &\text{(ii)} \qquad EK(s,t) = 0, \qquad EK(s,t)K(s',t') = (t \wedge t')\Gamma(s,s'), \end{aligned} \tag{4}$$

where $\Gamma(s,s')$ is the covariance function of a separable Gaussian process $\{B(s),\ 0 \le t \le 1\}$, with $B(0) = B(1) = 0$. The $\{K(s,t),\ 0 \le s \le 1,\ t \ge 0\}$ is a Kiefer process if $\{B(s),\ 0 \le t \le 1\}$ is a Brownian bridge, that is, $\Gamma(s,s') = s \wedge s' - ss'$.

We assume that the following smoothness and tail assumptions are satisfied by distributions $F_i$, $i = 0, 1$, throughout this paper.

A1. The distribution function $F$ is twice differentiable on $(t_F, T_F)$, where $-\infty \le t_F = \sup\{x\colon\ F(x) = 0\}$ and $\infty \ge T_F = \inf\{x\colon\ F(x) = 1\}$ and $F' = f > 0$ on $(t_F, T_F)$.

A2. For some $\gamma > 0$,

$$\sup_{0<u<1} u(1-u)\frac{\left|f'(F^{-1}(u))\right|}{f^2(F^{-1}(u))} < \gamma. \tag{5}$$

By applying Theorem 8.4.2 in Csörgő (1983) and assuming the location–scale model (3), we have the strong approximation for the PL empirical quantile functions $\tilde{F}_{i,n_i}^{-1}$, $i = 0, 1$, in the following form:

$$\begin{aligned} \tilde{F}_{1,n_1}^{-1}(u) &= \mu + \sigma F_0^{-1}(u) \\ &\quad + \left[\sigma/\left(n_1 f_0\left(F_0^{-1}(u)\right)\right)\right]K_1(u, n_1) + \varepsilon_{1,n_1} \quad \text{a.s.} \end{aligned} \tag{6}$$

and

$$\tilde{F}_{0,n_0}^{-1}(u) = F_0^{-1}(u) + \left[1/\left(n_0 f_0\left(F_0^{-1}(u)\right)\right)\right]K_0(u, n_0) + \varepsilon_{0,n_0} \quad \text{a.s.} \tag{7}$$

uniformly on $(\delta_n, T)$ with $n = n_0 \wedge n_1$, $T < \min\{1, T_{H_0(F_0^{-1})}, T_{H_1(F_1^{-1})}\}$ and $\delta_n = n^{-2/3}(\log n)^5$, where the two generalized Kiefer processes $K_i$, $i = 0, 1$, are independent and each has covariance function $EK_i(s,t)K_i(s',t') = (t \wedge t')\Gamma_i(s,s')$ with

$$\Gamma_i(s,s') = (1-s)(1-s')\int_0^{s\wedge s'} \frac{du}{(1-u)^2\left(1 - H_i\left(F_i^{-1}(u)\right)\right)}, \tag{8}$$

and the two error terms $\varepsilon_{1,n_1}$ and $\varepsilon_{0,n_0}$ are of order

$$O\big(n^{-3/4}(\log\log n)^{1/4}(\log n)^{3/4}\big)$$

almost surely.

Let $\mathbf{u}^{(m)} = (u_1, \ldots, u_m)^{\mathsf{T}}$ with $\delta_n < u_i < T$, $1 \le i \le m$. For notational simplicity, we suppress the dependence of vector $\mathbf{u}^{(m)}$ on $m$, that is, $\mathbf{u}^{(m)} = \mathbf{u}$. Also let $G(\mathbf{u})$ be the $m \times 1$ vector $(G(u_1), \ldots, G(u_m))^{\mathsf{T}}$ for $G$ being any function on $u$ or stochastic process having time scale $u$. We denote as $D_{(G(\mathbf{u}))}$ the diagonal matrix with main diagonal vector $G(\mathbf{u})$. However, we further denote the diagonal matrix $D_q = D(f_0(F_0^{-1}(\mathbf{u})))$ since it is frequently needed from here on.

With (6) and (7) at $\mathbf{u}$, we then have a regression setup with measurement error on the covariate as follows:

$$\tilde{F}_{1,n_1}^{-1}(\mathbf{u}) = \mu + \sigma F_0^{-1}(\mathbf{u}) + \sigma/n_1 D_q^{-1} K_1(\mathbf{u}, n_1) + \varepsilon_{1,n_1} \quad \text{a.s.}, \tag{9}$$

$$\tilde{F}_{0,n_0}^{-1}(\mathbf{u}) = F_0^{-1}(\mathbf{u}) + 1/n_0 D_q^{-1} K_0(\mathbf{u}, n_0) + \varepsilon_{0,n_0} \quad \text{a.s.} \tag{10}$$

The following lemma gives a decomposition for the covariance matrices of $K_1(\mathbf{u}, n_1)$ and $K_0(\mathbf{u}, n_0)$. These decompositions will enable us to calculate the finite sample approximated precision of the GLS estimator of $\mu$ and $\sigma$ derived from the above regression setup and its asymptotical covariance matrix in the next section.

LEMMA 2.1. *Let the covariance matrices of Gaussian random vectors* $K_i(\mathbf{u}, n_i)$ *be* $\Lambda_i$, $i = 0, 1$. *Then, for* $i = 0, 1$,

$$\Lambda_i = D_{(1-\mathbf{u})} C^{-1} D_{(C\mathbf{t}^{(i)})} C^{\mathsf{T}-1} D_{(1-\mathbf{u})}, \tag{11}$$

*where the matrix* $C$ *is the linear operator such that* $C\mathbf{u} = (u_1, u_2 - u_1, \ldots, u_m - u_{m-1})^{\mathsf{T}}$ *and* $\mathbf{t}^{(i)} = (t_1^{(i)}, \ldots, t_m^{(i)})^{\mathsf{T}}$, *with*

$$t_j^{(i)} = \int_0^{u_j} \frac{du}{(1-u)^2\big(1 - H_i\big(F_i^{-1}(u)\big)\big)}, \qquad 1 \le j \le m. \tag{12}$$

It is noted that, for fixed $u$, $(1-u)(1 - H_i(F_i^{-1}(u)))$ can be $\sqrt{n}$-consistently estimated by the empirical probability of $((X_{ij} \wedge C_{ij}) > \tilde{F}_{i,n_i}^{-1}(u))$, $i = 0, 1$. Accordingly, $t_j^{(i)}$ can be $\sqrt{n}$-consistently estimated by plugging the empirical probability into (12). We denote this estimator of $t_j^{(i)}$ as $\hat{t}_j^{(i)}$, $i = 0, 1$, $1 \le j \le l$, and we denote the vector as $\hat{\mathbf{t}}^{(i)} = (\hat{t}_1^{(i)}, \ldots, \hat{t}_m^{(i)})^{\mathsf{T}}$. Alternatively, kernel-type smoothed estimates can be used here as well.

Also in the regression setup of (9) and (10), the regression error is involved with unknown matrix $D_q = D_{(f_0(F_0^{-1}(\mathbf{u})))}$ since $f_0$ is unknown. We estimate the $f_0$ by a kernel-smoothed density estimate, $\hat{f}_0$ say, based on the Kaplan–Meier PL estimator $\tilde{F}_{0,n_0}$ with bandwidths, or smoothing parameters, chosen according to the results given in Marron and Padgett (1987). Then we plug the

PL empirical quantile into $\hat{f}$ to estimate the density quantile function $f_0(F_0^{-1})$. An alternative estimate of this density quantile function is obtained through the first derivative of a smoothed quantile function estimate.

With the consistency of vector estimator $\hat{\mathbf{t}}^{(i)}$, $i = 0, 1$, and the kernel estimator of the density quantile function, we then are able to construct an estimate for the covariance structure of the regression setup of (9) and (10). Hence the parameters of interest $\mu$ and $\sigma$ can be simply estimated by a type of weighted least squares estimator, which turns out to be an approximated maximum likelihood estimator (MLE) in the functional model with measurement error on the covariate. Furthermore, this functional can be thought of as having replicates tending to $\infty$ as $n \to \infty$. Therefore, the consistency of this MLE estimator is ensured.

Next we derive the weighted least squares estimator. From here on, we call this type of weighted least squares estimator the generalized least squares (GLS) estimator, and we denote it by $\hat{\theta}$, where $\theta = (\mu, \sigma)$. That is,

$$\hat{\theta} = \left(\hat{\mathbf{X}}^{\top}\hat{\Sigma}_1^{-1}\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^{\top}\hat{\Sigma}_1^{-1}\tilde{F}_{1,n_1}^{-1}(\mathbf{u}), \tag{13}$$

where

$$\hat{\mathbf{X}}^{\top} = \begin{pmatrix} 1 & \cdots & 1 \\ \tilde{F}_{0,n_0}^{-1}(u_1) & \cdots & \tilde{F}_{0,n_0}^{-1}(u_l) \end{pmatrix}$$

and

$$\hat{\Sigma}_1 = \hat{D}_q^{-1}\left[1/n_1\hat{\Lambda}_1 + 1/n_0\hat{\Lambda}_0\right]\hat{D}_q^{-1}, \tag{14}$$

with $\Sigma_1 = D_q^{-1}[1/n_1\Lambda_1 + 1/n_0\Lambda_0]D_q^{-1}$ denoting the covariance matrix of random vector $D_q^{-1}[n_1^{-1}K_1(\mathbf{u}, n_1) - n_0^{-1}K_0(\mathbf{u}, n_0)]$, $\hat{D}_q = D^{-1}_{(\hat{f}_0(\tilde{F}^{-1}_{0,n_0}(\mathbf{u})))}$ and $\hat{\Lambda}_i$ obtained by plugging $\hat{t}_j^{(i)}$, $1 \le j \le m$, into $\Lambda_i$, $i = 0, 1$, defined in (11).

Further, from the regression setup of (9) and (10), we have the approximated error of the GLS estimator

$$\begin{aligned}\hat{\theta} - \theta = {} & \sigma\left(\hat{\mathbf{X}}^{\top}\hat{\Sigma}_1^{-1}\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^{\top}\hat{\Sigma}_1^{-1}\hat{D}_q^{-1}\left[n_1^{-1}K_1(\mathbf{u}, n_1) - n_0^{-1}K_0(\mathbf{u}, n_0)\right] \\ & + o_p(n^{-1/2}).\end{aligned} \tag{15}$$

Asymptotic properties of the GLS estimator $\hat{\theta}$ based on (15) will be discussed in Theorem 3.2 in the next section.

At the end of this section, a small simulation study is reported for comparing the performance of the GLS estimator with the estimator derived from the original log-rank estimating function in the homogeneous model as well as in the heterogeneous model. Since the estimation of the weighting needed in the GLS estimator is involved with kernel function and bandwidth choices, it seems desirable to bring the performance of the following simple least squares estimator into comparison:

$$\hat{\theta}_{\mathrm{LS}} = \left(\hat{\mathbf{X}}^{\top}\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^{\top}\tilde{F}_{1,n_1}^{-1}(\mathbf{u}). \tag{16}$$

Asymptotic properties of $\hat{\theta}_{\mathrm{LS}}$ will also be discussed in the next section.

This simulation is done with 500 replications having equal sample size $n = 100$. The distribution of $X_{0j}$ is taken as Weibull(1, 1), and the two censoring random variables $C_{ij}$, $i = 0, 1$, are simulated from Weibull($\frac{1}{4}$, 1) and Weibull($\frac{1}{8}$, 1), respectively. For constructing the GLS and LS estimators, the vector of regression points $\mathbf{u}$ consists of eight points with equal spacing between 0.1 and 0.8. And, for simulation simplicity, the normal kernel function with bandwidth $n^{-1/5}$ is used in estimating the density quantile function in the covariance matrix of the regression setup. Simulation averages and mean square errors (MSEs) are reported for estimating $\mu$ and $\sigma$, both taking values 0.5, 1.0 and 1.5. The case with $\sigma = 1$ corresponds to the homogeneous case. The MSE is given in the parentheses right below its average value. In the last column of Table 1, the average percentage of uncensored data in the two groups are reported with the percentage of the second group in the parentheses.

We can see from Table 1 that the GLS estimator is very competitive with, even better than, the estimator derived from the log-rank estimating function in the homogeneous case, $\sigma = 1$. Further, in estimating $\mu$, the latter estimator is biased upward or downward according to whether the value of $\sigma$ is greater than or smaller than 1. Similar results are also found in several simulation studies with different choice of $\mathbf{u}$ and with different distribution forms, such as the log-normal distribution. For estimating $\sigma$, both the GLS and LS estimators are rather unbiased with not much difference in the mean square error.

TABLE 1
*Comparison of the log-rank*, GLS *and* LS *estimators of* $\mu$ *and* $\sigma$

| | | Estimating μ | | | Estimating σ | | |
|---|---|---|---|---|---|---|---|
| **σ** | **μ** | **Log-rank** | **GLS** | **LS** | **GLS** | **LS** | **(%)** |
| 0.5 | 0.5 | 0.46149 | 0.49513 | 0.49999 | 0.50028 | 0.50013 | 80 |
| | | (0.00475) | (0.00184) | (0.00181) | (0.00666) | (0.00713) | (94) |
| | 1.0 | 0.93430 | 0.98909 | 0.99799 | 0.49649 | 0.49684 | 80 |
| | | (0.01800) | (0.00743) | (0.00788) | (0.00570) | (0.00611) | (89) |
| | 1.5 | 1.40393 | 1.47074 | 1.48879 | 0.49273 | 0.49177 | 79 |
| | | (0.04173) | (0.01753) | (0.01675) | (0.00682) | (0.00691) | (84) |
| 1.0 | 0.5 | 0.53886 | 0.49206 | 0.50227 | 1.01605 | 1.01795 | 79 |
| | | (0.00816) | (0.00782) | (0.00835) | (0.02673) | (0.02961) | (94) |
| | 1.0 | 1.07500 | 0.98544 | 1.01095 | 1.00883 | 1.01541 | 79 |
| | | (0.03541) | (0.03229) | (0.03397) | (0.02525) | (0.02749) | (88) |
| | 1.5 | 1.59530 | 1.46306 | 1.49856 | 0.99686 | 1.00130 | 79 |
| | | (0.06775) | (0.06114) | (0.06453) | (0.02768) | (0.02901) | (84) |
| 1.5 | 0.5 | 0.62105 | 0.48648 | 0.50380 | 1.51685 | 1.52154 | 80 |
| | | (0.02880) | (0.01477) | (0.01657) | (0.05035) | (0.05607) | (92) |
| | 1.0 | 1.22839 | 0.99396 | 1.02889 | 1.50812 | 1.50602 | 80 |
| | | (0.11287) | (0.07738) | (0.08438) | (0.06263) | (0.06605) | (86) |
| | 1.5 | 1.78302 | 1.48756 | 1.54155 | 1.51365 | 1.51346 | 80 |
| | | (0.19143) | (0.15395) | (0.16720) | (0.06322) | (0.06718) | (81) |

**3. Asymptotic properties of the GLS estimator.** In order to study the asymptotical properties of the GLS estimator $\hat{\theta}$, we first extend the theory of strong approximation of the PL empirical quantile process on a fixed interval, which has been exclusively used in Section 2, to that on an interval tending to $(0, T_{H_0(F_0^{-1})} \wedge T_{H_1(F_1^{-1})})$. Then we calculate the asymptotic variance of the GLS estimator. If we further assume that the equality $H_0(F_0^{-1}(t)) = H_1(F_1^{-1}(t))$, $\forall\ t \in (0,1)$, holds, then this two-sample problem is adaptive. At the end of this section, the GLS estimator $\hat{\theta}$ is shown to be semiparametric efficient in general in the sense of achieving the semiparametric Fisher information bound.

To start, we denote $T^0 = T_{H_0(F_0^{-1})} \wedge T_{H_1(F_1^{-1})}$, and recall that $n = n_0 \wedge n_1$. The following theorem gives the strong approximation of the PL empirical quantile process to hold uniformly on $(\delta_n, T_n)$ with $T_n \to T^0$ in a rather slow rate as $n \to \infty$. This theorem is an extension of Theorem 8.4.2 of Csörgő (1983) which has been described in the Remark 8.4.4 of the cited reference as the concluding remark. And its proof can be done mainly by following the scheme already described at the end of Chapter 8 of Csörgő (1983) together with the improved PL empirical process results in Burke, Csörgő and Horváth (1988).

Before stating the theorem we need a polynomial varying tail condition:

A3. For some positive integer $k_0$, we have $1 - H_i(F_i^{-1}(u)) = O((T_{H_i(F_i^{-1})} - u)^{k_0})$ as $u \to T_{H_i(F_i^{-1})}$, $i = 0, 1$.

It should be noted that the rate of convergence given below will only be taken as a lower bound. The exact rate is not attempted here. And it could be significantly faster than that given here.

THEOREM 3.1. *Assume that the distribution function $F$ satisfies conditions* A1, A2 *and* A3. *Then, with $\delta_n = n^{-2/3}(\log n)^5$ and $T_n \to T_{H(F^{-1})}$ at a rate of $O(1/\log n)$, we have*

$$\sup_{\delta_n \le u \le T_n} \left| \sqrt{n}\, f\big(F^{-1}(y)\big)\big(\tilde{F}_n^{-1}(u) - F^{-1}(u)\big) - n^{-1/2}K(u,n) \right| = O\Big(n^{-1/4}(\log n)^{k_0+2}\Big) \quad a.s., \tag{17}$$

*where $\tilde{F}_n^{-1}$ is the Kaplan–Meier PL empirical quantile function pertaining to $n$ i.i.d. samples possibly right censored by independent variables with distribution function $H$.*

The next theorem gives the asymptotic covariance matrix of the GLS estimator $\hat{\theta}$ with $n_0/n_1 \to \lambda$, a positive number, as $n \to \infty$.

THEOREM 3.2. *Given the assumptions in Theorem* 3.1 *and letting* $\mathbf{u} = (u_1, \ldots, u_m)$ *be the chosen vector of time points, we have the following*:

1. *If* $\mathbf{u}$ *is chosen to be fixed and independent of* $n$, *then*

$$\hat{\theta} - \theta =_d N\Big(0, \sigma^2\big[\tilde{\mathbf{X}}^\top \Sigma_1^{-1} \tilde{\mathbf{X}}\big]^{-1}\Big) + o_p(n^{-1/2}), \tag{18}$$

*where* $\Sigma_1$ *is the true value of its estimate* $\hat{\Sigma}_1$ *defined in* (14), *and*

$$\tilde{\mathbf{X}}^\top = \begin{pmatrix} 1 & \cdots & 1 \\ F_0^{-1}(u_1) & \cdots & F_0^{-1}(u_m) \end{pmatrix}.$$

2. *If* $m = O(n^{1/4-\varepsilon})$ *for some small positive value* $\varepsilon < \frac{1}{4}$ *and* $\delta_n \le u_1 < \cdots < u_m \le T_n$ *and tends to be dense on* $(0, T^0)$ *in the sense that the mesh* $\max_i(u_i - u_{i-1}) \to 0$ *as* $n \to \infty$, *then the GLS estimator* $\hat{\theta}$ *has its asymptotic covariance matrix being equal to the limiting form of* $\sigma^2[\tilde{X}^\top \hat{\Sigma}^{-1} \tilde{X}]^{-1}$ *and*

$$\lim_{n\to\infty} n_0\Big[\tilde{X}^\top \hat{\Sigma}^{-1} \tilde{X}\Big] = \begin{pmatrix} a_{11} & a_{12} \\ * & a_{22} \end{pmatrix}, \tag{19}$$

*where these entries are as follows and all the integrals are assumed to exist*:

$$a_{11} = \int_0^{T^0} c(u) l(u)^2 \, du,$$

$$a_{12} = \int_0^{T^0} c(u) l(u) s(u) \, du,$$

$$a_{22} = \int_0^{T^0} c(u) s(u)^2 \, du,$$

*with*

$$c(u) = \frac{\big(1 - H_0\big(F_0^{-1}(u)\big)\big)\big(1 - H_1\big(F_1^{-1}(u)\big)\big)}{\lambda\big(1 - H_0\big(F_0^{-1}(u)\big)\big) + \big(1 - H_1\big(F_1^{-1}(u)\big)\big)},$$

$$l(u) = \left[\frac{f_0'\big(F_0^{-1}(u)\big)}{f_0\big(F_0^{-1}(u)\big)} + \frac{f_0\big(F_0^{-1}(u)\big)}{1-u}\right],$$

$$s(u) = \left[\frac{F_0^{-1}(u) f_0'\big(F_0^{-1}(u)\big)}{f_0\big(F_0^{-1}(u)\big)} + 1 + \frac{F_0^{-1}(u) f_0\big(F_0^{-1}(u)\big)}{1-u}\right].$$

Using the same argument for result (1), we can derive the following asymptotic property of the simple least squares estimator defined in (20):

$$\hat{\theta}_{\mathrm{LS}} = \big(\hat{\mathbf{X}}^\top \hat{\mathbf{X}}\big)^{-1} \hat{\mathbf{X}}^\top \tilde{F}_{1,n_1}^{-1}(\mathbf{u}), \tag{20}$$

$$\hat{\theta}_{\mathrm{LS}} - \theta =_d N\Big(0, \sigma^2\big[\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\big]^{-1}\big[\tilde{\mathbf{X}}^\top \Sigma_1^{-1} \tilde{\mathbf{X}}\big]\big[\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\big]^{-1}\Big) + o_p(n^{-1/2}). \tag{21}$$

Next, we give a sufficient condition under which this estimator is adaptive in the sense of Bickel (1982). Consider a parametric setup in which we suppose that $G((x-\mu_0)/\sigma_0) = F_0(x)$ and $G((x-\mu_1)/\sigma_1) = F_1(x)$ with known distribution form $G$. Under the location–scale model assumption (3),

these two pairs of location and scale parameters satisfy

$$\sigma = \sigma_1/\sigma_0,$$
$$\mu = \mu_1 - \sigma\mu_0.$$

Suppose that censoring mechanisms $C_{ij}$, $i = 0, 1$, the same as in Theorem 3.2, are applied on the two samples, respectively. Then the parametric Fisher information matrix for $\theta_i = (\mu_i, \sigma_i)^{\mathrm{T}}$ is $\sigma_i^2 I_i$, $i = 0, 1$, where

$$I_i = \begin{pmatrix} a_{11}^{(i)} & a_{12}^{(i)} \\ * & a_{22}^{(i)} \end{pmatrix},$$

where

$$a_{11}^{(i)} = \int_0^{T^0} c^{(i)}(u) l(u)^2 \, du,$$
$$a_{12}^{(i)} = \int_0^{T^0} c^{(i)}(u) l(u) s(u) \, du,$$
$$a_{22}^{(i)} = \int_0^{T^0} c^{(i)}(u) s(u)^2 \, du.$$

with $c^{(i)}(u) = (1 - H_i(F_i^{-1}(u)))$.

From the expression of $I_i$, $i = 0, 1$, and $c(u)$ in Theorem 3.2, it is clear that if

$$c^{(0)}(u) = c^{(1)}(u), \qquad u \in (0, T^0), \tag{22}$$

then the Fisher information matrix for $(\mu, \sigma)$ in this two-sample parametric problem is equal to the matrix in (19). Therefore, when the strict condition (22) holds, the semiparametric two-sample location–scale model with right-censored data is adaptive in the same sense as that in the case of complete data. [See Bickel, Klaassen, Ritov and Wellner (1993) or Hsieh (1995).] Furthermore, the GLS estimator should be semiparametric efficient given that condition (22) holds.

In fact, in general the GLS estimator $\hat{\theta}$ defined in (13) is semiparametric efficient in the sense of achieving the Fisher information bound. This result is stated in the next theorem. It is noted that the homogeneous regression model considered in Ritov and Wellner (1988), Ritov (1990) and Bickel, Klaassen, Ritov and Wellner (1993) will cover the homogeneous case, $\sigma = 1$, of this two-sample problem. And the Fisher information bound for the parameter $\mu$ given in these references is exactly equal to $a_{11}$ in (19). Furthermore, by using the $R$ and $L$ operators and the martingale techniques developed in Ritov and Wellner (1988), the Fisher information bound for the location and scale parameters in our two-sample location–scale model can also be calculated based on the same line as their arguments and found to be exactly equal to $\sigma^2$ multiplying the matrix given in (19). To avoid replication, the proof of the following theorem is therefore omitted here.

THEOREM 3.3. *Given the assumptions made in Theorem* 3.2, *the GLS estimator* $\hat{\theta}$ *defined in* (13) *is a regular and semiparametric efficient estimator*.

As an alternative way of estimating $(\mu, \sigma)$, a semiparametric efficient estimator can be obtained by using Ritov's (1990) $M$-type estimators with a system of two estimating equations with estimated score functions corresponding to the location and the scale parameters, respectively. Nevertheless, if finite sample properties, like robustness, are concerned, then the GLS estimator might be more desirable, since Ritov's $M$-type estimator derived from estimated score equations can be very sensitive to the smoothing parameters used. See Hsieh and Manski (1987). In this aspect, the GLS estimator is better because only the estimate of the density function is needed, not its derivative. As an alternative to remedy this type of sensitivity problem, the $B$-spline method used in Faraway (1992) and Jin (1992) could be a useful technique here as well.

REMARK 1. In regard to the convergence rate given in Theorem 3.1, the rate $n^{-1/4}$ is not essential for our first-order asymptotic result, since we need only the rate $n^{-\varepsilon}$ for some small positive number $\varepsilon$. Further if the rate in assumption A3 is exponential, instead of polynomial, then we have to make $T_n$ tend to $T_{H(F^{-1})}$ at a rate of $O(1/\log\log n)$. And then the convergence rate in Theorem 3.1 is changed accordingly.

REMARK 2. So far, in our two-sample setting, we require censoring random variables $\{C_{ij}\}$, $i = 0, 1$, defined in Section 2, to be independent and identically distributed in each of the two groups. Based on the recent results of Zhou (1991, 1992), the "identically distributed" assumption can be taken off. That is, random variables $C_{ij}$ can be nonidentically distributed according to $H_{ij}$, $j = 1, \ldots, n_i$. Correspondingly, the strong approximations of the PL empirical quantile functions given in (6) and (7) are still valid with generalized Kiefer processes whose covariance functions, given in (8), will have $H_i$ being replaced by $n_i^{-1}\sum_1^{n_i} H_{ij}$. And similar results as in Theorem 3.2 can also be derived.

**4. *K*-sample problem and semiparametric ANOVA.** In this section, we discuss the $K$-sample problem. The semiparametric version of analysis of variance (ANOVA) can be regarded as a special case in which the assumption of homogeneity on within-treatment variations is assumed. This type of homogeneity corresponds to the valid condition of equal normal variances in the classic ANOVA setting.

Throughout the discussion in this section, we extend the notation used for the two-sample setting in the previous sections to a $K(= k + 1)$-sample problem without repeating their definitions explicitly. For example, the $k + 1$ samples of random variables are denoted as $(X_{ij}, C_{ij})$, $i = 0, \ldots, k$, $j = 1, \ldots, n_i$, and so forth. Also the within-group and between-group independence of $X_{ij}$ and $C_{ij}$ are assumed as well. For this $K$-sample problem, we

similarly assume the location–scale model as

$$F_i(x) = F_0\left(\frac{x-\mu_i}{\sigma_i}\right), \qquad i = 1, \ldots, k. \tag{23}$$

By applying the empirical approach used in Section 2, we can estimate $\theta_i = (\mu_i, \sigma_i)$, $i = 1, \ldots, k$, by the corresponding GLS estimator, $\hat{\theta}_i$ say.

In order to test for the equality among $\{\mu_i, i = 1, \ldots, k\}$ or $\{\sigma_i, i = 1, \ldots, k\}$, we need to construct the covariance matrix of the vector of $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$. For this purpose, it is essential to observe from (15) that, for each $i = 0, \ldots, k$, the generalized Kiefer process $K_i(n_i, t)$ is exclusively generated from the PL empirical quantile processes pertaining to the Kaplan–Meier PL empirical quantile function $\tilde{F}^{-1}_{i,n_i}$, so that they are mutually stochastically independent. Finally, with the results in Theorem 3.1 and (15) and the independency between these generalized Kiefer processes $K_i(n_i, \cdot)$, $i = 0, \ldots, k$, we can derive the asymptotic covariance matrix of $\hat{\theta}_i - \theta$ and $\hat{\theta}_j - \theta$, $i \neq j$, in the following corollary. Its proof will be also omitted.

COROLLARY 4.1. *Assume that the distribution functions* $F_i$, $i = 0, \ldots, k$, *satisfy conditions* A1 *and* A2, *and right-censoring mechanisms* $C_{ij}$ *are independent of* $X_{ij}$ *for each* $1 \le j \le n_i$. *Let the location–scale model* (23) *hold for* $F_i$, $i = 0, \ldots, k$. *Then the GLS estimator* $\hat{\theta}_i$ *defined in* (13) *has the following asymptotic covariance structure*:

$$\begin{aligned} &\operatorname{cov}\left(\hat{\theta}_i - \theta, \hat{\theta}_j - \theta\right) \\ &\quad = \frac{1}{n_0}\sigma_i\sigma_j\left(\tilde{\mathbf{X}}^{\top}\Sigma_i^{-1}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{\top}\Sigma_i^{-1}D_q^{-1}\Lambda_0 D_q^{-1}\Sigma_j^{-1}\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}^{\top}\Sigma_j^{-1}\tilde{\mathbf{X}}\right)^{-1}, \end{aligned} \tag{24}$$

*where* $\Sigma_i$, $1 < i \le k$, *are likewise defined as* $\Sigma_1$ *and recall that* $D_q = D_{(f_0(F_0^{-1}(\mathbf{u})))}$.

With the above corollary we are able to do an analysis of variance (ANOVA) in the semiparametric setting. From the ANOVA point of view, two types of preliminary tests of the homogeneity of $\sigma_i$, $i = 0, \ldots, k$, are of interest. The first type is $H_0^{(1)}$: $\sigma_i = 1$, $i = 0, \ldots, k$, when the equality of the $K$ treatments effect is concerned. If the $F_0$ is corresponding to the standard treatment, while the others are innovative ones, then the second type of hypothesis $H_0^{(2)}$: $\sigma_i = \sigma$, $i = 1, \ldots, k$, for some unknown $\sigma$ is needed for testing the equality of these $k$ innovative treatments.

To test $H_0^{(1)}$ and $H_0^{(2)}$, a corresponding $\chi^2$ testing statistic can be constructed by applying the above corollary. After not rejecting the homogeneity assumption as a valid condition for the ANOVA, the corresponding $\chi^2$ testing statistic can also be constructed concerning the equality among $\mu_i$, $i = 0, \ldots, k$.

If the hypothesis of equality of $F_i$, $i = 0, \ldots, k$, is of interest, then the corresponding $\chi^2$ testing statistic can be regarded as a type of generalized $t$-test. For the two-sample problem with complete data, O'Brien (1988) pro-

posed another type of generalization of $t$-test based on the regression of the group membership on the pooled order statistics and its square value. The idea of using the square rank is also to accommodate the possible heterogeneity between the treatment effects. However, our $\chi^2$ testing statistic not only works for the censored data, but also, from a parsimonious point of view, it will indicate the direction of deviation of the data from the homogeneity hypothesis.

We remark that a $\chi^2$ testing statistic for the two-sample location–scale model assumptions can be easily constructed, as a by-product, based on the regression setup (9) in Section 2. On the other hand, the corresponding $\chi^2$ testing statistic for testing the $K$-sample model assumption seems quite involved and is not known to the author.

## APPENDIX

In this Appendix, the proofs are only sketched to show the key points and ideas involved.

PROOF OF LEMMA 2.1. Let $W(\mathbf{u})$ be the standard Wiener process with positive time scale $z$. Consider the random vector $W(\mathbf{u})$, where $0 < u_1 < \cdots < u_m$. Then the covariance matrix of $W(\mathbf{u})$, $\Lambda$ say, satisfies

$$\Lambda = C^{-1} D_{(C\mathbf{u}))} C^{\mathsf{T}-1}.$$

With the definition of the covariance structure (8) of the generalized Kiefer process and the above result, we prove the lemma. □

PROOF OF THEOREM 3.1. With assumption A3 and Theorem 8.2.1 of Csörgő (1983), page 122, we have the result related to small increments of a generalized Kiefer process

$$\sup_{0 \le s \le T_n - h_n} \sup_{0 \le t \le h_n} |K(t+s,n) - K(s,n)| = O\left(n^{1/2} (\log n)^{k_0 + 3/4} h_n^{1/2}\right) \quad \text{a.s.}$$

By applying Theorem 5.5 of Burke, Csörgő and Horváth (1981), taking $h_n = O(n^{1/2}(\log n)^{2k_0+5/2})$ and then following the development of Theorems 8.3.1, 8.4.1 and 8.4.2, we prove our result. □

PROOF OF THEOREM 3.2. Based on (15) with the vector $\mathbf{u}$ being fixed, the strong consistency of estimators $\tilde{F}_{0,n_0}^{-1}(\mathbf{u})$, $\hat{f}_0(\tilde{F}_{0,n_0}^{-1}(\mathbf{u}))$ and $\hat{t}_j^{(i)}$, $1 \le j \le m$, will ensure result (1) of this theorem.

For result (2), we consider the strong approximation of the PL empirical quantile process on $(\delta_n, T_n)$ ensured by Theorem 3.1. Therefore, the error terms in (15) are still valid since the two error terms in (10) and (9) are of order $O(n^{-3/4}(\log n)^{k_0+2})$ and $m$ is chosen to have order smaller than $O(n^{1/4})$. Then, by applying result (1), we can have the asymptotical variance of the GLS estimator as $\sigma^2[\tilde{X}^{\mathsf{T}} \hat{\Sigma}^{-1} \tilde{X}]^{-1}$.

By Lemma 2.1, we further have

$$\lim_{n\to\infty} n_0\left[\tilde{X}^{\top}\hat{\Sigma}^{-1}\tilde{X}\right] = \lim_{n\to\infty}\begin{pmatrix}\hat{a}_{11} & \hat{a}_{12}\\ * & \hat{a}_{22}\end{pmatrix},$$

where, letting $u_j \le u_j^* \le u_{j+1}$, $j = 1,\dots,m-1$,

$$\hat{a}_{11} = \sum_{j=1}^{m-1}\frac{1}{u_{j+1}-u_j}\left\{(1-u_j^*)^2\frac{\left(1-H_0\left(F_0^{-1}(u_j^*)\right)\right)\left(1-H_1\left(F_1^{-1}(u_j^*)\right)\right)}{\lambda\left(1-H_0\left(F_0^{-1}(u_j^*)\right)\right)+\left(1-H_1\left(F_1^{-1}(u_j^*)\right)\right)}\right.$$
$$\left.\times\left[\frac{f_0\left(F_0^{-1}(u_{j+1}^*)\right)}{1-u_{j+1}^*}-\frac{f_0\left(F_0^{-1}(u_j^*)\right)}{1-u_j^*}\right]^2\right\}$$
$$\doteq \int_0^{T^0}(1-u)^2\frac{\left(1-H_0\left(F_0^{-1}(u)\right)\right)\left(1-H_1\left(F_1^{-1}(u)\right)\right)}{\lambda\left(1-H_1\left(F_1^{-1}(u)\right)\right)+\left(1-H_0\left(F_0^{-1}(u)\right)\right)}$$
$$\times\left[\frac{\partial}{\partial u}\left(\frac{f_0\left(F_0^{-1}(u)\right)}{1-u}\right)\right]du$$
$$= a_{11}.$$

$a_{12}$ and $a_{22}$ are calculated likewise. □

## REFERENCES


BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–761.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins Univ. Press.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. Verw. Gebiete* **56** 87–112.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1988). A correction to and improvement of Strong approximations of some biometric estimates under random censorship. *Probab. Theory Related Fields* **79** 51–57.

CSÖRGŐ, M. (1983). *Quantile Processes with Statistical Applications*. SIAM, Philadelphia.

FARAWAY, J. J. (1992). Smoothing in adaptive estimation. *Ann. Statist.* **20** 414–427.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

HSIEH, D. and MANSKI, C. F. (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Ann. Statist.* **15** 541–551.

HSIEH, F. (1995). The empirical process approach for semiparametric two-sample models with heterogeneous treatment effect. *J. Roy. Statist. Soc. Ser. B* **57** 735–748.

HSIEH, F. (1996). Estimations in homoscedastic and heteroscedastic linear regression models with censored data: an empirical process approach. Unpublished manuscript.

HSIEH, F. and TURNBULL, B. W. (1996). Non- and semi-parametric estimation of the receiver operating characteristic (ROC) curve. *Ann. Statist.* **24** 25–40.

HSIEH, F. and TURNBULL, B. W. (1996). The empirical process approach for general semiparametric regression models: theory and applications. Unpublished lecture notes.

JIN, K. (1992). Empirical smoothing parameter selection in adaptive estimation. *Ann. Statist.* **20** 1844–1874.

LOUIS, T. A. (1981). Non-parametric analysis of an accelerated failure time model. *Biometrika* **68** 381–390.

MARRON, J. S. and PADGETT, W. J. (1987). Asymptotically optimal bandwidth selection for kernel for density estimators from randomly right-censored samples. *Ann. Statist.* **15** 1520–1535.

MOSES, L. E. (1963). Rank tests of dispersion. *Ann. Math. Statist.* **34** 973–983.

O'BRIEN, P. C. (1988). Comparing two samples: extensions of the $t$, rank sum, and log-rank tests. *J. Amer. Statist. Assoc.* **83** 52–61.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. In *Contemporary Mathematics, Statistical Inference for Stochastic Processes* (N. U. Prabhu, ed.) **80** 191–220. Amer. Math. Soc., Providence, RI.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372.

WEI, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statist. in Medicine* **11** 1871–1879.

WEI, L. J. and GAIL, M. H. (1983). Nonparametric estimation for a scale-change with censored observations. *J. Amer. Statist. Assoc.* **78** 382–388.

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21** 76–99.

ZHOU, M. (1991). Some properties of the Kaplan–Meier estimator for independent non-identically distributed random variables. *Ann. Statist.* **19** 2266–2274.

ZHOU, M. (1992). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20** 1002–1021.

DEPARTMENT OF MATHEMATICS
NATIONAL TAIWAN UNIVERSITY
NUMBER 1, SECTION 4, ROOSEVELT ROAD
TAIPEI
TAIWAN