Statistica Sinica **29** (2019), 1963-1975
doi:https://doi.org/10.5705/ss.202016.0345

ON ESTIMATION OF PARTIALLY LINEAR VARYING-COEFFICIENT TRANSFORMATION MODELS WITH CENSORED DATA

Bo Li, Baosheng Liang, Xingwei Tong and Jianguo Sun

Tsinghua University, Peking University, Beijing Normal University and University of Missouri

Abstract: Failure time data occur in many fields and in various forms and, as a result, have been studied extensively by researchers. To handle such data, many semiparametric regression models have been proposed and investigated, including the proportional hazards model and the additive hazards model. In this study, we consider a class of partially linear varying-coefficient transformation models. Given the limitations of existing inference procedures, we propose a more general rank estimation procedure. Furthermore, we establish the finite and asymptotic properties of the resulting estimators of the regression parameters and illustrate the proposed procedure by means of an example.

Key words and phrases: B-splines, smoothing partial rank, transform model, varying coefficients, weighted bootstrap.

1. Introduction

In this study, we conduct a regression analysis of right-censored failure time data, for which many regression models have been proposed and investigated. Of these, the proportional hazards or Cox model is perhaps the most widely used (Cox (1972); Kalbfleisch and Prentice (2002)). Other popular models include the accelerated failure time model (Jin et al. (2003)) and the additive hazards model (Lin and Ying (1994)). However, these models share a common limitation in the sense that they can only be applied to specific situations. To address this problem, several general models have been proposed, including various types of transformation models, which have the advantage of flexibility and include many commonly used specific models as special cases (Chen, Jin and Ying (2002); Chen and Tong (2010); Jin, Ying and Wei (2001); Li and Zhang (2012)).

We begin by introducing the transformation models. Let T denote the failure time of interest and suppose that the covariates of interest can be written in three parts, X, Z, and W, for notational simplicity. Here, X is a vector, and Z and

W are scalar values, as defined below. To model the effect of X on T, one of the first transformation models to be used was the linear transformation model

$$H_0(T) = X^T\beta_0 + \varepsilon, \tag{1.1}$$

where H_0 denotes an unknown, strictly increasing function, β_0 is a vector of regression parameters, and ε is an error term. Several studies have developed inference approaches for estimating β_0 under the assumption that the distribution of ε is known (Chen, Jin and Ying (2002); Khan and Tamer (2007)). However, this assumption does not always hold in practice. To address this, Khan and Tamer (2007) and Song et al. (2007) examined the situation where the distribution of ε is unknown. The former proposed a partial rank (PR) estimation procedure that was a generalization of the method proposed in Han (1987) and Sherman (1993) for uncensored data, and the latter proposed a smooth version of the PR estimation procedure.

Although model (1.1) includes some commonly used models as special cases, it can be restrictive in certain applications (Chen and Tong (2010); Lu and Zhang (2010); Li and Zhang (2012)). For example, it only allows covariates that have linear effects, and thus excludes those that have nonlinear effects. To address this limitation, several studies have examined the following class of partially linear varying-coefficient transformation models:

$$H_0(T) = X^T\beta_0^* + Z\phi_0(W) + \varepsilon, \tag{1.2}$$

where $\beta_0^* = (1, \beta_0^T)^T$, H_0, and ε are defined as in model (1.1), and Z represents a covariate that has a possible nonlinear effect on the response variable through covariate W and an unknown function ϕ_0. Note that X represents the covariates that have linear effects on the response variable, and W is usually a time-related (but fixed) covariate over which the effect of Z on the response variable may vary. For example, W might be the age of onset of a disease in a prevalent cohort. Furthermore, in model (1.2), we assume that the first component of β_0^* is equal to one in order to avoid the identifiability problem.

Among others, Chen and Tong (2010) and Lu and Zhang (2010) have examined inference problems using model (1.2) when the distribution of ε can be specified by a parametric model. Li and Zhang (2012) considered model (1.2) under the assumption that T can have a known special conditional cumulative distribution function. However, their method relies heavily on the assumption of a standard normal distribution function, and also the proposed algorithm is complicated. Because these assumptions may not hold in practice, we propose an inference procedure that does not require such assumptions. In other words, we will consider the situation in which the distribution of ε is completely unknown.

The remainder of the paper is organized as follows. Section 2 presents a smooth estimation procedure for model (1.2), where following Chen and Tong (2010), we employ a linear combination of B-splines to approximate the nonparametric function ϕ_0. In addition, we establish the convergence rate and asymptotic normality of the proposed estimators and present a bootstrapping method for inferences. Section 3 discusses the results of an extensive simulation study conducted to evaluate the proposed inference procedure, showing that it works well in practical situations. In Section 4, we provide an illustrative example of the proposed procedure, and Section 5 concludes the paper.

2. Inference Procedure

Consider a failure time study that consists of n independent subjects. Let T_i, X_i, Z_i, and W_i be defined as above, but associated with subject i. Furthermore, let C_i denote the censoring time on subject i, which is assumed to be independent of T_i. Assume that the observed data have the form

$$\{V_i = \min(T_i, C_i), \Delta_i = I(T_i \le C_i), X_i, Z_i, W_i; i = 1, \ldots, n\}.$$

That is, the data are right-censored. Here, we aim to estimate the regression parameters.

For an estimation or inference using model (1.2), denote $B_n(\cdot) = (b_1(\cdot), \ldots, b_{q_n}(\cdot))^T$ and consider a B-spline basis of order $\ell + 1$, where $q_n = K_n + \ell$ and K_n denotes the number of knots and is the integer part of n^ν, with $0 < \nu < 0.5$. Then, following Song et al. (2007), we propose estimating β_0 and $\phi_0(\cdot)$ by $(\hat{\beta}_n, \hat{\phi}_n(\cdot))$, defined as $\hat{\phi}_n(\cdot) = B_n(\cdot)^T \hat{\alpha}_n$ and

$$(\hat{\beta}_n, \hat{\alpha}_n) = \text{argmax}_{(\beta,\alpha)} O_n(\beta, \alpha), \tag{2.1}$$

where

$$\begin{aligned} O_n(\beta, \alpha) = &\frac{1}{n(n-1)} \sum_{i \ne j} \Delta_j I(V_i \ge V_j) \\ &\times s_n(X_i^T\beta + Z_i B_n(W_i)^T \alpha - X_j^T \beta - Z_j B_n(W_j)^T \alpha). \end{aligned} \tag{2.2}$$

Here, $s_n(u) = s(u/\eta_n)$ is a smooth function, where $s(u)$ is typically set as the sigmoid function $s(u) = 1/\{1+\exp(-u)\}$ and η_n is a sequence of strictly positive numbers converging to zero.

Clearly, if $\phi_0(\cdot) = 0$, model (1.2) reduces to model (1.1) and the proposed estimation procedure reduces to that of Song et al. (2007). Furthermore, if we employ $s(u) = I(u \ge 0)$ as the indicator function, the estimation procedure reduces to that of Khan and Tamer (2007). In practice, we may omit the indicator function to avoid using a computationally intensive grid search for the maximization.

Now, we establish the asymptotic properties of the estimators $\hat{\beta}_n$ and $\hat{\phi}_n(\cdot)$ defined above. Let $\theta = (\beta^T, \phi(\cdot))^T$ and $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\phi}_n(\cdot))^T$, and without loss of generality, assume that W has support on $[0,1]$. Define $y = (x, z, w, \delta, v)$ and

$$\tau_n(y, \theta) = E\{\Delta I(v \geq V) s_n(x^T\beta + z\phi(w) - X^T\beta - Z\phi(W))\}.$$

We require the following regularity conditions:

Condition A1. The true value $\beta_0 \in \mathcal{B}$, a compact subset of R^p.

Condition A2. The true function $\phi_0 \in \mathcal{F}_r$, with $r = l + \gamma > 0.5$, where

$$\mathcal{F}_r = \{\phi(\cdot) : |\phi^{(l)}(w_1) - \phi^{(l)}(w_2)| \leq A_0 |w_1 - w_2|^\gamma \text{ for all } 0 \leq w_1 \leq w_2 \leq 1\},$$

$l+1$ is the order of the B-spline functions, and $\phi^{(l)}$ is the lth derivative function of $\phi(\cdot)$.

Condition A3. The random variable ε_i is independent of the random vector (C_i, X_i, Z_i, W_i) and is independent and identically distributed. In addition, there exists a positive constant c_0 such that the censoring time C satisfies $\inf_{x,z,w} P(C > \tau | X = x, Z = z, W = w) > c_0$, where τ is the largest followup study period.

Condition A4. The first component of X has a density with respect to the Lebesgue measure that is positive everywhere, conditional on the other components of X and on Z and W.

Condition A5. For each y, $\tau_n(y, \theta)$ is twice differentiable with respect to θ (in the Gâteux sense) in a neighborhood of θ_0 with the kth derivative $\nabla_k \tau_n(y, \theta)$, for $k = 1, 2$. The second derivative $\nabla_2 \tau_n(y, \theta)$ satisfies the Lipschitz condition.

Condition A6. $E\| \nabla_1 \tau_n(Y, \theta)\|^2$ and $E\| \nabla_2 \tau_n(Y, \theta)\|$ are finite, and the eigenvalues of $E\{\nabla_2 \tau_n(Y, \theta)\}$ are bounded away from zero.

Note that Conditions A1 and A2 are standard regularity conditions (Chen and Tong (2010)), and Conditions A3 and A4 ensure the identification of θ_0 (Khan and Tamer (2007)). Conditions A5 and A6 are inherited from Song et al. (2007) and are needed to establish the Lipschitz conditions required for the Taylor expansion arguments. Let θ_0 denote the true value of θ and $\| \cdot \|_2$ be the usual L_2 norm. In addition, define the metric

$$\rho(\theta_1, \theta_2) = |\beta_1 - \beta_2| + \|\phi_1 - \phi_2\|_2,$$

for $\theta_1,\ \theta_2 \in \mathcal{B} \times \mathcal{F}_r$. In the following, we first give the convergence rate of $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\phi}_n(\cdot))^T$ and then the asymptotic normality of $\hat{\beta}_n$. The proofs are provided in the Supplementary Material.

Theorem 1 (Convergence Rate)**.** *Assume that Conditions A1–A6 hold and that* $\eta_n \to 0$ *as* $n \to \infty$*. Then, we have that*

$$\rho(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-v)/2} + n^{-rv}).$$

Theorem 2 (Asymptotic Normality)**.** *Assume that Conditions* A1–A6 *hold and that* $\eta_n \to 0$ *as* $n \to \infty$*. Furthermore, assume that* $q_n = O(n^v)$ *with* $1/4r < v < 0.5$*. Then, we have that*

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma)$$

in distribution, where Σ *is defined in the Supplementary Material.*

To make an inference about β_0 based on the above results, it is apparent that we need to estimate the covariance matrix Σ. A natural way to do so is to derive a consistent estimate of Σ, which is possible, but complicated. As such, we instead suggest using the weighted bootstrap strategies proposed by Jin, Ying and Wei (2001) and Cai, Tian and Wei (2005). Specifically, consider the following perturbed objective function:

$$\begin{aligned} O_n^w(\beta, \alpha) =& \frac{1}{n(n-1)} \sum_{i \neq j} \psi(R_i, R_j) \Delta_j I(V_i \geq V_j) \\ & \times s_n(X_i^T \beta + Z_i B_n(W_i)^T \alpha - X_j^T \beta - Z_j B_n(W_j)^T \alpha), \end{aligned} \tag{2.3}$$

where $\psi(R_i, R_j)$ satisfies one of the following two conditions:

(I) R has a known mean $\mu > 0$ and variance $4\mu^2$, and $\psi(R_i, R_j) = R_i + R_j$;

(II) R has mean 1 and variance 1, and $\psi(R_i, R_j) = R_i R_j$.

For a given integer B and each $1 \leq b \leq B$, let $(R_1^{(b)}, \ldots, R_n^{(b)})$ denote a set of random variables generated from one of the two scenarios above, and let $\hat{\theta}_n^{*(b)}$ be the minimizer of (2.3) corresponding to $(R_1^{(b)}, \ldots, R_n^{(b)})$. Then, we can approximate the distribution of $\hat{\theta}_n - \theta_0$ using the empirical distribution of the sample $\{\hat{\theta}_n^{*(b)} - \hat{\theta}_n; b = 1, \ldots, B\}$, and make an inference about θ_0. The simulation study below indicates that this approach works well in practical situations.

To implement the aforementioned weighted bootstrap inference procedure, we need to choose B, some distribution for the generation of $R_i^{(b)}$, the degree of the B-splines, the number and location of the knots, and the smoothing parameter η_n. Here, a larger B yields better results, and many distributions can be used for $R_i^{(b)}$. In the numerical studies below, we use $B = 400$, generate $R_i^{(b)}/10$ from the beta distribution Beta(0.125, 1.125) for the type-I bootstrap method, and generate $(\sqrt{2}-1)R_i^{(b)}/\sqrt{2}$ from the beta distribution Beta($\sqrt{2}-1, 1$) for the type-II bootstrap method. For the B-spline approximation, a common choice is to use cubic B-splines with $1.5n^{1/3}$ knots located according to the quantiles of the observed event times. The simulation study below suggests that these choices work well and that the estimation results are not too sensitive to them. With respect to η_n, one choice is to set $\eta_n = cn^{-1/2}$ and to compare the results given by different c; see the simulation study described below. An alternative, used in the example below, is to apply the approach suggested by Gammerman (1996) and Song et al. (2007) for choosing an optimal c. More specifically, we first choose an initial value, say, $c_0 = 1$, and then estimate $\hat{\beta}_{c_0}$ based on c_0. Next, we determine the largest constant c_1 such that 95% of the pairs $\{(Z_i, Z_j)\}_{i \neq j}$ satisfy $|\hat{\beta}_{c_0}^T (Z_i - Z_j)/(c_1 n^{-1/2})| > 5$. Finally, we choose $c_{opt} = \min(c_0, c_1)$.

3. A Simulation Study

In this section, we present the results of an extensive simulation study conducted to evaluate the finite-sample performance of the proposed estimation procedure. We first generated the covariates $X = (X_1, X_2)^T$ from the bivariate normal distribution with mean $(0.2, 0.2)^T$, and the variance for both covariates, with correlations of 0.5 and 0.1, respectively. We also generated the covariate Z from the binary distribution with $\Pr(Z = 0) = \Pr(Z = 1) = 0.5$, and the covariate W from the uniform distribution over $(0, 1)$. Given the covariates, the failure times are generated from model (1.2), with $H_0(t) = \log(t)$ $(t > 0)$ and $\phi_0(w) = \sin(2\pi w)$. For the distribution of ε, we considered three situations: the normal distribution with mean 0 and standard error σ, the Gumbel (or type-I extreme value) distribution with location 0 and scale $\sqrt{6}\sigma/\pi$, and the logistic distribution with location 0 and scale $\sqrt{3}\sigma/\pi$. Note that under the above transformation function, the Gumbel distribution corresponds to the proportional hazards model, whereas the logistic distribution corresponds to the proportional odds model. The censoring times were generated from an exponential distribution to give the required percentage of right-censoring. The simulation results given below are based on 500 replications with sample sizes of $n = 200$ or 400.

Tables 1 and 2 present the results for the estimation of the regression parameter β_0, with true values of $\beta_0 = 1$ or -1, $\sigma^2 = 0.5$ or 1, and $c = 0.5$ or 1. In Table 1, the percentage of right-censored observations is set to 20%, and the corresponding percentage in Table 2 is 40%. The results include the estimated bias (Bias) given by the average of the estimates minus the true value, empirical standard error (SE), normalized median absolute deviation (MAD) of the

Table 1. Simulation results under different scenarios with a 20% censoring rate.

				$n = 200$								$n = 400$							
						Bootstrap I			Bootstrap II					Bootstrap I			Bootstrap II		
β	σ^2	Dis.	c	Bias	SE	MAD	SEE	CP	MAD	SEE	CP	Bias	SE	MAD	SEE	CP	MAD	SEE	CP
1	0.5	N	0.5	−0.037	0.132	0.133	0.142	94	0.139	0.146	94	−0.016	0.079	0.079	0.082	95	0.080	0.083	95
			1.0	0.022	0.128	0.131	0.136	93	0.133	0.138	95	0.011	0.077	0.077	0.079	94	0.078	0.079	94
		G	0.5	−0.061	0.203	0.201	0.218	94	0.203	0.214	95	−0.033	0.114	0.124	0.130	94	0.123	0.129	95
			1.0	0.048	0.201	0.204	0.216	94	0.202	0.213	95	0.027	0.113	0.122	0.127	95	0.119	0.124	95
		L	0.5	−0.033	0.108	0.118	0.126	95	0.122	0.128	95	−0.012	0.070	0.069	0.072	95	0.071	0.073	93
			1.0	−0.019	0.106	0.114	0.120	95	0.118	0.122	95	−0.006	0.069	0.068	0.069	94	0.069	0.070	95
	1.0	N	0.5	0.050	0.166	0.179	0.189	94	0.180	0.188	95	0.027	0.107	0.108	0.111	94	0.107	0.110	94
			1.0	−0.035	0.162	0.178	0.186	95	0.176	0.184	94	−0.021	0.104	0.105	0.108	94	0.104	0.107	94
		G	0.5	0.085	0.250	0.257	0.280	91	0.253	0.266	91	0.054	0.150	0.164	0.173	96	0.162	0.171	95
			1.0	−0.074	0.251	0.264	0.284	92	0.260	0.275	91	−0.051	0.148	0.163	0.172	95	0.160	0.168	95
		L	0.5	−0.043	0.154	0.158	0.169	93	0.160	0.170	93	−0.018	0.092	0.093	0.097	95	0.093	0.097	95
			1.0	−0.028	0.151	0.156	0.165	94	0.154	0.162	93	−0.012	0.090	0.092	0.095	95	0.090	0.093	96
−1	0.5	N	0.5	−0.035	0.134	0.132	0.142	92	0.138	0.146	93	−0.010	0.079	0.079	0.083	95	0.082	0.085	96
			1	−0.019	0.129	0.131	0.138	93	0.135	0.140	94	−0.005	0.078	0.078	0.080	94	0.080	0.081	94
		G	0.5	−0.059	0.205	0.202	0.216	93	0.202	0.213	93	−0.029	0.121	0.124	0.130	94	0.125	0.131	95
			1	−0.045	0.198	0.204	0.215	93	0.202	0.212	93	−0.024	0.119	0.124	0.128	95	0.122	0.127	95
		L	0.5	0.023	0.116	0.120	0.128	95	0.125	0.132	95	0.010	0.064	0.070	0.072	96	0.071	0.074	96
			1	−0.006	0.114	0.119	0.123	94	0.122	0.126	95	−0.004	0.063	0.067	0.068	95	0.070	0.071	96
	1.0	N	0.5	−0.038	0.176	0.177	0.189	94	0.182	0.190	93	−0.028	0.104	0.106	0.110	93	0.107	0.112	94
			1	−0.022	0.171	0.175	0.186	95	0.177	0.186	96	−0.022	0.103	0.106	0.109	94	0.105	0.108	93
		G	0.5	0.075	0.279	0.261	0.283	91	0.256	0.271	90	0.035	0.160	0.165	0.176	94	0.163	0.172	94
			1	0.064	0.275	0.267	0.289	92	0.261	0.279	92	0.030	0.157	0.163	0.173	94	0.162	0.170	94
		L	0.5	0.028	0.160	0.155	0.168	93	0.162	0.171	95	0.013	0.096	0.094	0.098	94	0.096	0.099	94
			1	−0.012	0.154	0.158	0.166	95	0.159	0.166	94	−0.008	0.095	0.094	0.096	95	0.094	0.096	95

estimates, average of the estimated standard errors (SEE), and 95% empirical coverage probability (CP). Note that in the tables, we use N, G, and L to denote the normal, Gumbel, and logistic distribution error terms, respectively. In general, the reults suggest that the proposed estimator is unbiased and that the weighted bootstrap method works well in terms of both the variance estimation and the approximation to the distribution of the proposed estimator. Moreover, as expected, the results improve as the sample size increases.

In addition, the results given in Tables 1 and 2 indicate that the tuning parameter c and the constant σ^2 may have some effect on the estimation. More specifically, the proposed estimators with $c = 1$ show a relatively smaller bias and are more efficient than those with $c = 0.5$. Furthermore, as expected, the estimators with a smaller censoring percentage and σ^2 perform relatively better. However, when the sample size is sufficiently large, the proposed estimator does not seem sensitive to the choice of c. In addition, the type-I bootstrap method performs slightly better than the type-II bootstrap method does under most simulation settings. However, for large sample sizes, the difference tends to

Table 2. Simulation results under different scenarios with a 40% censoring rate.

				$n = 200$								$n = 400$							
						Bootstrap I			Bootstrap II					Bootstrap I			Bootstrap II		
β	σ^2	Dis.	c	Bias	SE	MAD	SEE	CP	MAD	SEE	CP	Bias	SE	MAD	SEE	CP	MAD	SEE	CP
1	0.5	N	0.5	−0.065	0.134	0.142	0.148	91	0.145	0.152	92	−0.032	0.076	0.085	0.089	95	0.086	0.089	95
			1	0.053	0.127	0.136	0.143	93	0.139	0.145	94	0.027	0.075	0.083	0.084	95	0.083	0.084	94
		G	0.5	−0.078	0.199	0.206	0.222	92	0.209	0.219	92	−0.049	0.126	0.130	0.138	92	0.131	0.137	94
			1	−0.068	0.198	0.210	0.225	92	0.208	0.219	93	−0.045	0.123	0.129	0.135	93	0.129	0.134	93
		L	0.5	0.040	0.117	0.123	0.132	95	0.131	0.137	96	0.018	0.070	0.072	0.075	94	0.073	0.076	93
			1	−0.026	0.110	0.121	0.127	95	0.125	0.130	97	−0.013	0.069	0.070	0.072	94	0.071	0.072	94
	1.0	N	0.5	0.053	0.177	0.181	0.196	94	0.183	0.194	94	0.031	0.112	0.111	0.118	95	0.113	0.118	95
			1	−0.040	0.174	0.181	0.193	94	0.181	0.192	94	−0.027	0.110	0.110	0.115	94	0.109	0.113	95
		G	0.5	−0.102	0.267	0.262	0.286	90	0.257	0.273	90	−0.052	0.167	0.172	0.181	94	0.171	0.178	94
			1	0.087	0.271	0.273	0.295	91	0.267	0.282	90	0.050	0.165	0.171	0.180	94	0.168	0.176	94
		L	0.5	0.065	0.156	0.156	0.168	92	0.161	0.169	92	0.031	0.092	0.095	0.099	93	0.095	0.099	94
			1	−0.051	0.151	0.156	0.163	93	0.158	0.165	95	−0.026	0.091	0.093	0.096	93	0.092	0.095	94
−1	0.5	N	0.5	−0.045	0.137	0.146	0.152	93	0.150	0.156	94	−0.012	0.086	0.086	0.089	95	0.088	0.091	95
			1	0.031	0.132	0.142	0.147	95	0.145	0.149	94	0.006	0.084	0.084	0.086	96	0.085	0.087	95
		G	0.5	−0.066	0.214	0.209	0.227	92	0.210	0.223	91	−0.035	0.126	0.133	0.139	95	0.132	0.137	95
			1	−0.054	0.208	0.213	0.228	92	0.213	0.225	93	−0.031	0.124	0.130	0.136	94	0.129	0.134	94
		L	0.5	0.028	0.123	0.128	0.136	93	0.133	0.139	94	0.012	0.070	0.073	0.076	95	0.075	0.078	96
			1	−0.013	0.118	0.124	0.130	93	0.127	0.131	94	−0.007	0.068	0.072	0.073	96	0.072	0.074	96
	1.0	N	0.5	−0.081	0.182	0.182	0.196	91	0.186	0.196	91	−0.025	0.110	0.114	0.119	94	0.114	0.119	95
			1	0.071	0.177	0.181	0.193	91	0.184	0.192	91	0.019	0.109	0.114	0.117	94	0.112	0.115	95
		G	0.5	0.102	0.258	0.266	0.290	92	0.256	0.274	91	0.060	0.172	0.173	0.184	93	0.171	0.180	93
			1	0.089	0.258	0.270	0.293	93	0.270	0.286	93	0.058	0.169	0.171	0.182	93	0.166	0.178	93
		L	0.5	−0.045	0.158	0.158	0.169	93	0.161	0.171	93	−0.024	0.091	0.094	0.099	94	0.097	0.101	95
			1	−0.030	0.154	0.154	0.163	94	0.158	0.166	94	−0.017	0.090	0.093	0.096	95	0.094	0.096	95

disappear. Note that the weighted bootstrap distribution could be skewed for a small sample size and when outlier estimates are unavoidable. That is why we calculate the MAD of the proposed estimates, which may give better results for the standard error of the estimate when sample size is small.

Tables 1 and 2 also show that the shape of the error distribution may affect the proposed estimate for small sample sizes. More specifically, the proposed method seems to perform slightly better under symmetric error distributions, such as normal and logistic distributions, than it does under skewed error distributions, such as the Gumbel distribution. Once again, when the sample size increases, the performance difference and the shape effect become smaller or can be ignored. To further illustrate this point and to compare the proposed method with an existing method, Figure 1 shows the averages of the estimates of the nonlinear effect function ϕ_0 given by the method of Chen and Tong (2010) and the proposed method under the three error distributions. The graphs are based on the simulated data giving the results in Table 1 with $\beta_0 = -1$, $c = 1$, $\sigma = 0.5$, and $n = 200$. Note that the former method requires a known error distribution,

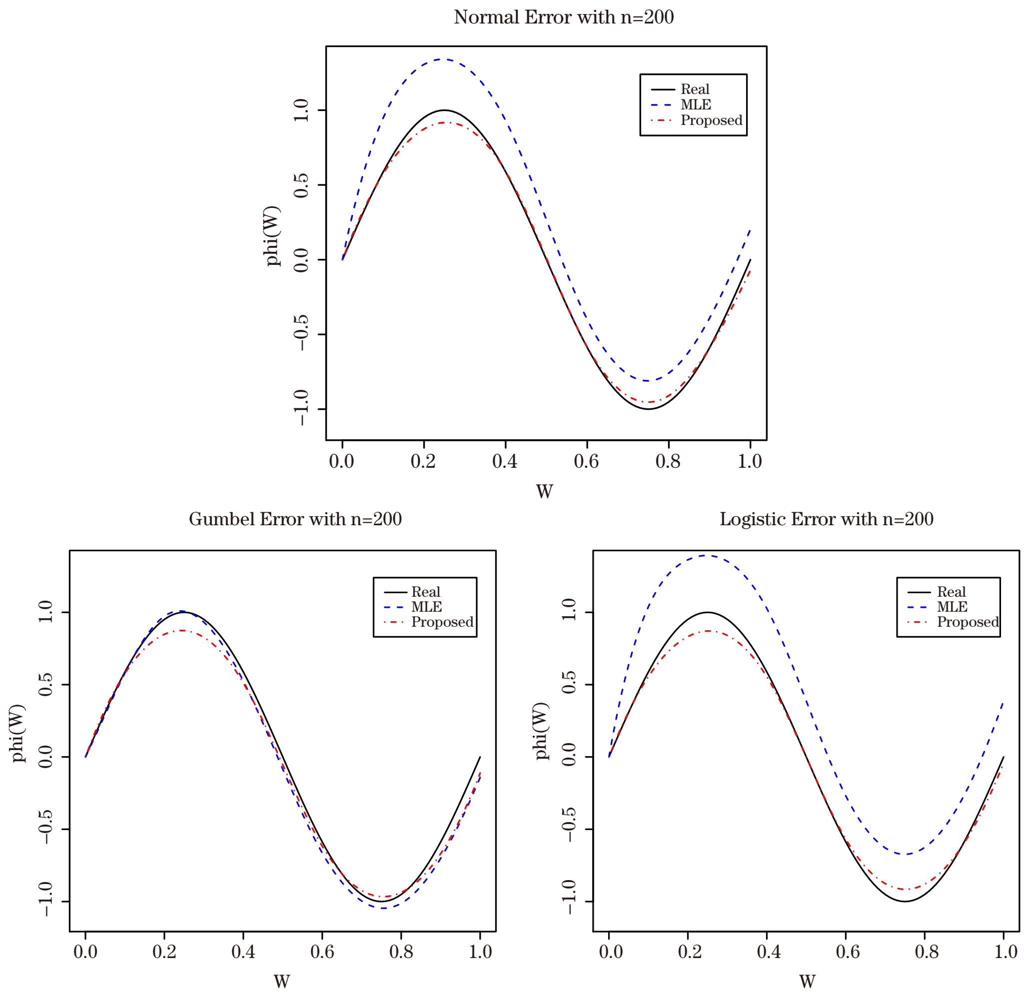

Figure 1. The estimated nonlinear effect function ϕ_0 using Chen and Tong's (2010) method and the proposed method.

which is assumed to be the Gumbel distribution in the figure. Figure 1 clearly shows that the proposed method gives better estimates and is more robust than the method of Chen and Tong (2010).

We also considered other setups and obtained similar conclusions. In the Supplementary Material, we provide estimation results obtained from error distributions that are more heavy-tailed than that considered above. There, we also present the results of comparisons between the proposed menthod and the methods of Khan and Tamer (2007) and Lu and Zhang (2010). In all cases, the proposed estimator tends to be more stable and robust than the other methods.

4. An Illustrative Example

In this section, we apply the proposed methodology to the Veterans' Administration lung cancer data on patients with advanced inoperable lung cancer, discussed by Kalbfleisch and Prentice (2002) and Li and Zhang (2012), among others. The data set consists of 137 patients who were randomized to receive either standard or a test form of chemotherapy. In the study, one of the primary endpoints for the therapy comparison is the time to death. Of the patients in the sample, 128 were followed to death. In addition to the treatment, several covariates were observed, including the Karnofsky score, the time in months from the diagnosis to randomization (diagtime), prior therapy (yes or no), the patient's age in years, and the type of lung cancer cell (small, squamous, or large). Note that to fit model (1.2), we first need to choose a benchmark variable or covariate, the coefficient of which is set to one. A common method of doing so is to choose the most interesting or important covariate as the benchmark. To determine this, we calculated the Kendall-τ between the failure time and each covariate. As a result, we selected the Karnofsky score, which yielded the largest Kendall-τ of 0.387, as the benchmark variable. Note that both Kalbfleisch and Prentice (2002) and Li and Zhang (2012) also concluded that the Karnofsky score is the most important covariate.

In addition to the treatment and covariate effects on the time to death, identifing the optimal age for chemotherapy treatment is also of interest. Thus, we consider the following varying-coefficient transformation model:

$$H(T) = X_1 + \sum_{i=1}^{5} \beta_i X_{i+1} + \phi_1(W) + Z\,\phi_2(W) + \varepsilon\,, \tag{4.1}$$

where the covariates are defined as follows: $X_1 = \text{Karnofsky}/10$, $X_2 = \text{diagtime}/100$, $X_3 = \text{prior}/10$, $X_4 = X_5 = X_6 = 1$ if the cell type is small, squamous, or large, respectively, and 0 otherwise; $Z = 1$ for the patients given the test chemotherapy, and 0 otherwise; and $W = \text{age}/100$, where prior $= 0$ if there has been no prior therapy, and 1 otherwise. Note that in model (4.1), $\phi_1(W)$ characterizes the possible nonlinear effect of the patient's age, and $\phi_2(W)$ represents the possible effect of the chemotherapy treatment at different age points or the interaction effect between the treatment and age on the death time.

Table 3 presents the estimation results for the covariates $X_2, \ldots, X_6$ given by the proposed inference procedure. The estimated nonlinear functions $\hat{\phi}_1(w)$ and $\hat{\phi}_2(w)$ are given in Figures 2, with pointwise 95% confidence bands. Note that the estimated standard errors were obtained using the type-I bootstrap method,

Table 3. Estimation results for the Veterans' Administration data.

			Cell type		
	β_1	β_2	β_3	β_4	β_5
Estimate	0.626	−0.154	−0.324	1.387	1.703
Stand Error	1.017	0.514	0.505	0.774	0.711
P-values	0.538	0.765	0.521	0.073	0.017

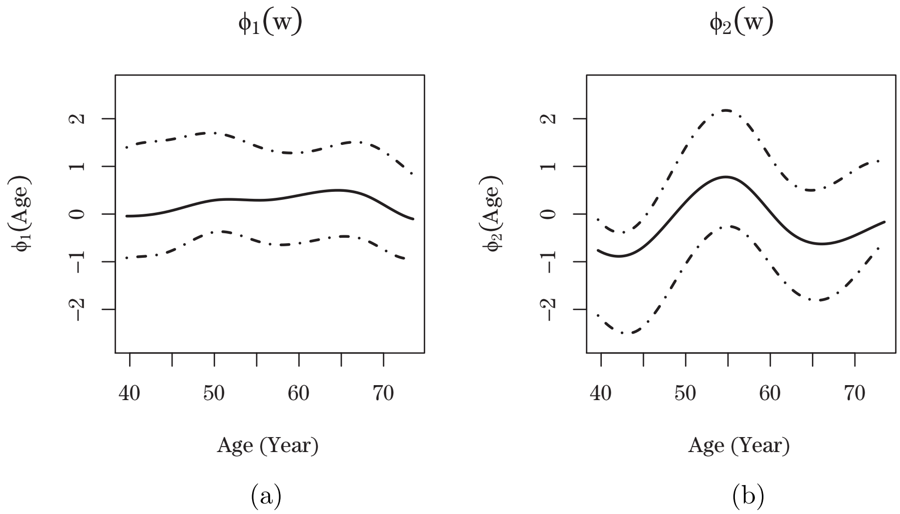

Figure 2. The estimates of curves $\phi_1(\cdot)$ and $\phi_2(\cdot)$ for the Veterans' Administration data.

and the confidence bands were determined using the 0.025 and 0.975 quantiles of 1,000 resampling estimates. Table 3 shows that, given the Karnofsky score, the death time seems to be significantly related to the tumor type, but not to the other covariates. Figure 2 (b) shows that the chemotherapy treatment may benefit patients between the ages of 48 and 60. The optimal age for the chemotherapy treatment is approximately 54. The treatment becomes less effective in younger and older patients. Furthermore, for patients younger than 48 years and older than 60 years, the treatment effect on survival may even be negative. These conclusions support those of Li and Zhang (2012), who used the normal distribution assumption, as discussed above.

5. Conclusion

In this study, we investigated a class of partly linear varying-coefficient transformation models for regression analyses of right-censored failure time data. To estimate the regression parameters, we presented a rank-based objective function, showing that it yields valid estimators. In addition, the asymptotic consistency and normality of the resulting estimators were established. The results of an extensive simulation study suggeste that the proposed methodology works well in practice. Although other researchers have discussed the same model, their inference methods require restrictive assumptions or apply only to limited situations.

We have focused on estimating linear or nonlinear covariate effects. However, it may be necessary at times to estimate the transformation function $H_0(\cdot)$ and the error distribution function, in which case new estimation procedures are required. In the proposed methodology, we assume that we know which covariates have linear effects and which have nonlinear effects on the failure time of interest. However, this information might not always be available. Thus, it would be useful to develop procedures to identify these covariates. Another limitation of the proposed method is its lack of efficiency, the improvement of which is left to future research.

Supplementary Materials

The Supplementary Material provides proofs of Theorems 1 and 2, as well as additional simulation results obtained under heavy-tailed error distributions for comparative purposes.

Acknowledgment

The authors wish to thank the editor, Dr. Zhiliang Ying, the associate editor, and two reviewers for their helpful comments and suggestions that greatly improved the paper. Liang's work was supported by the Fundamental Research Funds for the Central Universities. Tong's work was partly supported by the National Key Research and Development Program of China (2017YFA0604903) and NSFC 11671338.

References

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

Cai, T., Tian, L. and Wei, L. J. (2005). Semi-parametric Box-Cox power transformation models for censored survival observations. *Biometrika* **92**, 619-632.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.

Chen, K. and Tong, X. (2010). Varying coefficient transformation models with censored data. *Biometrika* **97**, 969-976.

Gammerman, A. (1996). *Computational Learning and Probabilistic Reasoning*. Wiley, New York.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics* **35**, 303-316.

Jin, Z., Ying, Z. and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.

Jin, Z., Lin, D. Y., Wei, L. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.

Khan, S. and Tamer, E. (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics* **136**, 251-280.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Li, J. and Zhang, R. (2012). General partially linear varying-coefficient transformation model with right censored data. *Journal of Statistical Planning and Inference* **142**, 1285-1293.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

Lu, W. and Zhang, H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association* **105**, 683-691.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123-137.

Song, X., Ma S., Huang, J. and Zhou, X. H. (2007). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics* **8**, 197-211.

School of Economics and Management, Tsinghua University, Beijing, 100084, China.

E-mail: libo@sem.tsinghua.edu.cn

Department of Natural Sciences in Medicine, Health Science Center, Peking University, Beijing 100191, P.R. China.

E-mail: liangbs@hsc.pku.edu.cn

School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China.

E-mail: xweitong@bnu.edu.cn

Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.

E-mail: sunj@missouri.edu

(Received July 2016; accepted February 2018)