Statistica Sinica Preprint No: SS-2016-0345

Title	ON ESTIMATION OF PARTIALLY LINEAR VARYING-COEFFICIENT TRANSFORMATION MODELS WITH CENSORED DATA
Manuscript ID	SS-2016-0345
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0345
Complete List of Authors	Bo Li Baosheng Liang Xingwei Tong and Jianguo Sun
Corresponding Author	Jianguo Sun
E-mail	sunj@missouri.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

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

Bo Li[1], Baosheng Liang[2], Xingwei Tong[2] and Jianguo Sun[3]

[1] *Tsinghua University,* [2] *Beijing Normal University,*

and [3] *University of Missouri*

Abstract: Failure time data occur in many fields and in various forms and their analysis has been attracting a great deal of attention (Cox (1972); Kalbfleisch and Prentice (2002); Sun (2006)). In particular, many semiparametric regression models such as the proportional hazards model and the additive hazards model have been proposed and investigated in the literature for various situations. In this paper, we consider a class of partially linear varying coefficient transformation models that have been studied by many authors (Chen, Jin and Ying (2002); Chen and Tong (2010); Jin, Ying and Wei (2001); Li and Zhang (2012)). For the situation considered here, however, all of the existing inference procedures have some limitations and corresponding to these, we propose a rank estimation procedure that applies to more general situations. Both the finite and asymptotic properties of the resulting estimators of regression parameters are established and an illustrative example is provided.

Key words and phrases: B-splines; Smoothing partial rank; Transform model; Varying coefficients; Weighted bootstrap.

1. Introduction

This paper discusses regression analysis of right-censored failure time data and for this, many regression models have been proposed and investigated. Among them, the proportional hazards or Cox model is perhaps the most commonly used one (Cox (1972); Kalbfleisch and Prentice (2002)). Some other commonly used models include the accelerated failure time model (Jin et al. (2003)) and the additive hazards model (Lin and Ying (1994)). On the other hand, each of these models has some shortcomings and one common one is that all of them are specific models that may only apply to some specific situations. To address these, many other models or generalizations of them have been proposed and in particular, various types of transformation models, which have the advantage of flexibility and include many commonly used specific models as special cases, have drawn a great deal of attention (Chen, Jin and Ying (2002); Chen and Tong (2010); Jin, Ying and Wei (2001); Li and Zhang (2012)).

To introduce the transformation model, 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 notation simplicity. Here X is a vector and Z and W

are scalar with their meanings defined below more specifically. To model the effect of X on T, one of the early transformation models is the linear transformation model

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

In the above, H_0 denotes an unknown strictly increasing function, β_0 is a vector of regression parameters and ε is an error term. Several authors have studied this type of models and developed inference approaches for estimation of β_0 under the assumption that the distribution of ε is known (Chen, Jin and Ying (2002); Khan and Tamer (2007)). It is easy to see that sometimes this assumption may not hold and to address this, Khan and Tamer (2007) and Song et al. (2007) discussed the situation where the distribution of ε is unknown. The former gave a partial rank (PR) estimation procedure, a generalization of the method proposed in Han (1987) and Sherman (1993) for uncensored data, and the latter discussed 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 some applications (Chen and Tong (2010); Lu and Zhang (2010); Li and Zhang (2012)). For example, it only allows the covariates that have linear effects and sometimes there may exist some covariates that have nonlinear effects. To address this, several authors studied the following class of partially linear varying coefficient transformation models

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

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

Among others, Chen and Tong (2010) and Lu and Zhang (2010) discussed the inference problem about model (1.2) when the distribution of ε can be specified by some parametric models. Li and Zhang (2012) also considered model (1.2) for the situation where T can be assumed to have a known special conditional cumulative distribution function. In addition, their method heavily relies on the standard normal distribution function assumption and the given algorithm is complicated. It is apparent that in reality, these assumptions may not hold and to address these, in the following, we will present an inference procedure that does not require these assumptions. In other words, we will consider the situation where the distribution of ε is completely unknown.

The remainder of the paper is organized as follows. Section 2 will present a smooth estimation procedure for model (1.2) and in the procedure, following Chen and Tong (2010), we will employ the linear combination of B-splines to approximate the nonparametric function ϕ_0. In addition, the convergence rate and asymptotic normality of the proposed estimators are established and a bootstrapping method is presented for inference. Section 3 gives some results obtained from an extensive simulation study conducted for the evaluation of the proposed inference procedure, which show that it works well for practical situations. In Section 4, an illustrative example is provided and Section 5 contains some discussion and concluding remarks.

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, and 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

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

That is, one observes right-censored data. Suppose that the main interest is to estimate regression parameters.

For estimation or inference about model (1.2), denote $B_n(\cdot) = (b_1(\cdot), ..., b_{q_n}(\cdot))^T$, and the 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 to estimate 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

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

In the above, $s_n(u) = s(u/\eta_n)$ is a smooth function with $s(u)$ typically setting to be the sigmoid function $s(u) = 1/\{1 + \exp(-u)\}$ and η_n being a sequence of strictly positive numbers converging to 0.

It is easy to see that if $\phi_0(\cdot) = 0$ or $\alpha = 0$, model (1.2) reduces to model (1.1) and in this case, the estimation procedure above reduces to that given in Song et al. (2007). Furthermore, if we take $s(u) = I(u \geq 0)$, the indicator function, the estimation procedure reduces to that discussed by Khan and Tamer (2007). In practice, one may want to avoid the use of the indicator function since then one has to employ the 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. For this, 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 the support on $[0, 1]$. Define $y = (x, z, w, \delta, v)$ and

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

Also we need 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 the ε_i's are independent and identically distributed. Also 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 time.

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

Condition A5. For each y, $\tau_n(y, \theta)$ is twice differentiable with respect to θ (in Gâteux sense) in a neighborhood of θ_0 with the kth derivative $\bigtriangledown_k \tau_n(y, \theta), k = 1, 2$. The second derivative $\bigtriangledown_2 \tau_n(y, \theta)$ satisfies the Lipschitz condition.

Condition A6. $E\| \bigtriangledown_1 \tau_n(Y, \theta)\|^2$ and $E\| \bigtriangledown_2 \tau_n(Y, \theta)\|$ are finite and the eigenvalues of $E\{\bigtriangledown_2 \tau_n(Y, \theta)\}$ is bounded away from 0.

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 needed to establish the Lipschitz conditions needed for Taylor expansion arguments. Let θ_0 denote the true value of θ and $\|\cdot\|_2$ be the usual L_2 norm. Also 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 $\hat{\beta}_n$ with their proofs given in the Supplementary Material.

Theorem 1. *[Convergence Rate] Assume that Conditions A1 - A6 hold and $\eta_n \to 0$ 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* $\eta_n \to 0$ *as* $n \to \infty$*. Also 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.*

For inference about β_0 based on the results above, it is apparent that we need to estimate the covariance matrix Σ, and for this, one natural way would be to derive a consistent estimate of Σ, which is possible but the estimate would be very complicated. Corresponding to this, for the inference, we instead suggest to use the weighted bootstrap strategies proposed in 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 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 $(R_1^{(b)}, \cdots, R_n^{(b)})$ denote a set of random variables generated from one of the two scenarios above and

$\hat{\theta}_n^{*(b)}$ the minimizer of (2.3) corresponding to $(R_1^{(b)}, \cdots, R_n^{(b)})$. Then one can approximate the distribution of $\hat{\theta}_n - \theta_0$ by the empirical distribution of the sample $\{\, \hat{\theta}_n^{*(b)} - \hat{\theta}_n;\, b = 1, ..., B \,\}$ given the data and make inference about θ_0. The simulation study below indicates that this approach works well for practical situations.

To implement the weighted bootstrap inference procedure above, it is apparent that we need to choose B, some distribution for the generation of the $R_i^{(b)}$'s, the degree of B-splines and the number and location of the knots as well as the smoothing parameter η_n. It is obvious that the larger B, the better the results, and the many distributions can be used for the $R_i^{(b)}$'s. In the numerical studies below, we used $B = 400$, and generated $R_i^{(b)}/10$ from the beta distribution Beta(0.125, 1.125) for the type (I) bootstrap method and $(\sqrt{2}-1)R_i^{(b)}/\sqrt{2}$ from the beta distribution Beta$(\sqrt{2}-1, 1)$ for the type (II) bootstrap method. For the B-spline approximation, a common choice is cubic B-splines with the knot number equal to $1.5n^{1/3}$ and the knots located according to the quantiles of the observed event times. The simulation study below suggests that these choices seem to work well and 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 as used in the simulation study below. An alternative, which was used in the example below, is to apply the approach suggested by Gammerman (1996) and Song et al. (2007) for choosing an optimal c. In this, one first chooses an initial value, say, $c_0 = 1$ and obtains an estimate $\hat{\beta}_{c_0}$ based on c_0. Then we determine the largest constant c_1 such that 95% of the pairs$\{(Z_i, Z_j)\}_{i \neq j}$ satisfy $|\hat{\beta}_{c_0}^T (Z_i - Z_j)/(c_1 n^{-1/2})| > 5$, and finally choose $c_{opt} = \min(c_0, c_1)$.

3. A Simulation Study

In this section, we present some results obtained from an extensive simulation study conducted to evaluate the finite sample performance of the estimation procedure proposed in the previous section. In the study, we first generated the covariates $X = (X_1, X_2)'$ from the bivariate normal distribution with mean $(0.2, 0.2)'$, the variance for both covariates and the correlation being 0.5 and 0.1, respectively, 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 were 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 transformation function above, the Gumbel distribution gives to the proportional hazards model, while the logistic distribution corresponds to the proportional odds model. The censoring times were generated from an exponential distribution to given the required percentage of right censoring, and the simulation results given below are based on 500 replications with the sample size $n = 200$ or 400.

Tables 1 and 2 present the results on estimation of regression parameter β_0 with the true value $\beta_0 = 1$ or -1, $\sigma^2 = 0.5$ or 1, and $c = 0.5$ or 1. In Table 1, the percentage of right-censored observations was set to be 20%, while 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, the empirical standard error (SE), the normalized median absolute deviation (MAD) of the estimates, the average of the estimated standard errors (SEE), and the 95% empirical coverage probability (CP). Note that in the tables, we used N, G, and L to denote the normal, Gumbel and logistic distribution error terms, respectively. It is easy to see from the two tables that in general, the proposed estimator seems to be unbiased and the weighted bootstrap method also seems to work well in terms of both variance estimation and the approximation to the distribution of the proposed estimator. In addition, as expected, the results became better as the sample size increased.

The results given in Tables 1 and 2 also indicate that the tuning parameter c and the constant σ^2 may have some effects on the estimation. More specifically, the proposed estimators with $c = 1$ seems to show relatively smaller biases and more efficient than these with $c = 0.5$, and as expected, the estimators with smaller censoring percentage and σ^2 had relatively better performances. However, when the sample size is enough, the proposed estimator does not seem to be too sensitive to the choice of c. In addition, one can see from the tables that the type (I) bootstrap method seems to give a little better performance than the type (II) bootstrap method under most simulation settings, but again for large sample sizes, the difference tends to disappear. Note that the weighted bootstrap distribution could be skewed when the sample size is not large and the occurrence of outlier estimates are unavoidable. That is why the MAD of the proposed estimates was calculated, which may give better estimates of the standard error of the estimate when sample size is not large.

Another point that one may see from Tables 1 and 2 is that the shape of the error distribution may also have some effects on the proposed estimate for small sample size situations. More specifically, the proposed method seems to perform a little bit better under the symmetric error distributions like the normal and logistic distribution than under the skewed error distri-

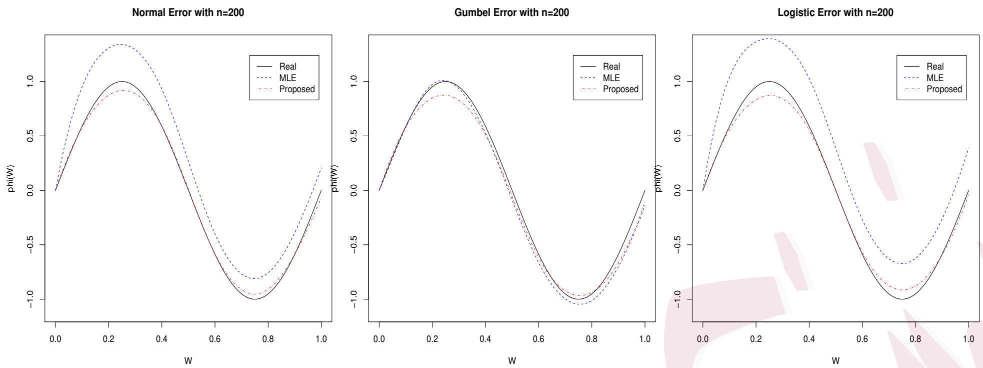


Figure 1: The estimated nonlinear effect function ϕ_0 by Chen and Tong (2010)'s method and the proposed method.

bution like the Gumbel distribution. Again as before, when the sample size increased, the performance difference or the shape effect became smaller or ignorable. To further see this and also to compare the presented method to the existing method, Figure 1 shows the averages of the estimates of the nonlinear effect function ϕ_0 given by both the method proposed in Chen and Tong (2010) and the method presented above under the three error distributions based on the simulated data yielding Table 1 with $\beta_0 = -1$, $c = 1$, $\sigma = 0.5$, and $n = 200$. Note that the former method needs to know the error distribution, which was assumed to be the Gumble distribution in the figure. One can see from Figure 1 that the proposed method clearly gave much better estimates or showed a significant robustness advantage than that in Chen and Tong (2010).

We also considered some other set-ups in the study here and obtained similar conclusions. In particular, in the Supplementary Material, we provide some estimation results obtained with more heavy-tailed error distributions than the error distribution considered above. Also we present some results obtained for comparing the method proposed here and those given in Khan and Tamer (2007) and Lu and Zhang (2010), and they indicated that as expected, the proposed estimator tends to be more stable or robust than the others.

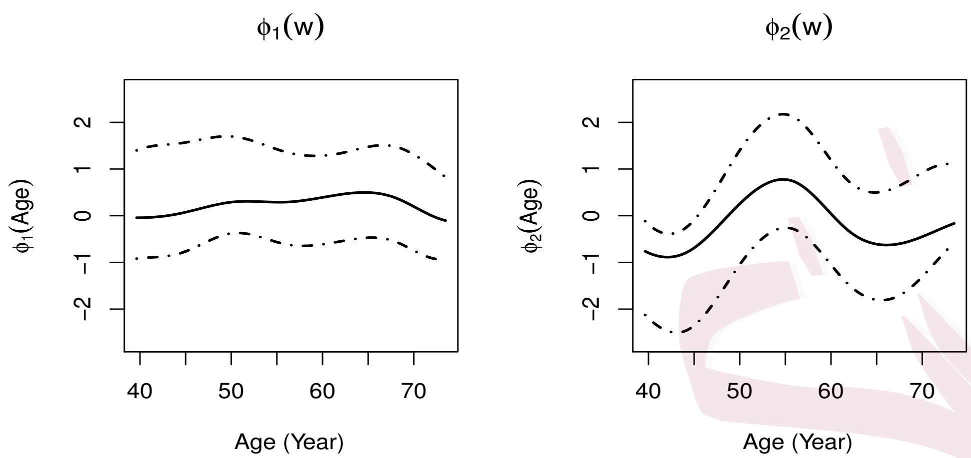
4. An Illustrative Example

Now we apply the methodology proposed in the previous sections to the Veterans' Administration lung cancer data on the patients with advanced inoperable lung cancer discussed in Kalbfleisch and Prentice (2002) and Li and Zhang (2012) among others. The data set consists of 137 patients who were randomized to either a standard or test chemotherapy and in the study, one primary endpoint for the therapy comparison is the time to death. Among them, 128 were followed to death. In addition to the treatment, several covariates were also observed and include 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 lung cancer cell type (small, squamous and large). Note that to fit model (1.2), first we need to choose a benchmark variable or covariate whose coefficient will be set to be one and for this, it is common to choose the most interesting or important covariate as the benchmark. To determine it, we calculated the Kendall-τ between the failure time and each covariate and selected the karnofsky score, which gave 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, the identification of the optimal age for chemotherapy treatment is also of interest. Based on these and the discussion above, for the analysis, we considered the following varying coefficient transformation model

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

In the above, the covariates are defined as follows: X_1 = karnofsky/10, X_2 = diagtime/100, X_3 = prior/10, $X_4 = X_5 = X_6 = 1$ if the cell type was small, squamous or large, respectively, and 0 otherwise, $Z = 1$ for the patients given test chemotherapy and 0 otherwise, W = age/100, where prior = 0 if 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



(a) (b)

Figure 2: The estimates of curves $\phi_1(\cdot)$ and $\phi_2(\cdot)$ for the Veterans Administration Data

death time.

Table 3 presents the estimation results on the covariates $X_2, \cdots, X_6$ given by the proposed inference procedure and the estimated nonlinear functions $\hat{\phi}_1(w)$ and $\hat{\phi}_2(w)$ are given in Figures 2 with the point-wise 95% confidence bands. Note that here the estimated standard errors were obtained by using the type (I) bootstrap method and the confidence bands were determined by the 0.025 and 0.975 quantiles of 1000 resampling estimates. One can see from Table 3 that given the karnofsky score, the death time seems to be significantly related to the tumor type, but not to other covariates. Figure 2 (b) shows that the chemotherapy treatment could benefit patients of 48 – 60 years' old. The optimal age for chemotherapy treatment is around 54. The treatment becomes less effective for both younger and older patients than the age 54, what is more, for the patients younger than 48 years and older than 60 years the treatment effect on survival may even be negative. The conclusions above are basically similar to or confirm those given in Li and Zhang (2012), which required the normal distribution assumption as discussed above.

5. Discussion and Concluding Remarks

In this paper, we investigated a class of partly linear varying coefficient transformation models for regression analysis of right-censored failure time data and for the estimation of regression parameters, a rank-based objective function was presented and shown to give valid estimators. In addition, the asymptotic consistency and normality of the resulting estimators were established, and an extensive simulation study conducted suggested that the proposed methodology seems to work well in practice. As discussed above, some authors discussed the same model but their inference methods require some restrictive assumptions or apply only to limited situations compared to the proposed approach.

In the preceding sections, we have mainly focused on estimation of linear or nonlinear covariate effects and it is apparent that sometimes one may be interested in estimation of the transformation function $H_0(\cdot)$ and the error distribution function too. For these, it is clear that one needs to develop some new estimation procedures. In the proposed methodology, it has been assumed that one knows the covariates that have linear effects and the covariates that may have nonlinear effects on the failure time of interest. Sometimes this may not be true and thus it will be useful to develop some procedures to identify these covariates. Another issue about the proposed method that was not discussed above is the efficiency and one future research direction related to this is to develop some more efficient estimation procedures.

Supplemental Material

In the Supplementary Material given at the journal website, we will sketch the proofs of Theorems 1 and 2 and also provide some additional simulation results obtained under some heavy-tail error distributions and for the comparison of the method proposed here and those given by the others.

Acknowledgements

The authors wish to thank the Editor, Dr. Zhiliang Ying, the Associate Editor and two reviewers for their many helpful comments and suggestions that greatly improved the paper. Tong's work was partly supported by the

National Key Research and Development Program of China2017YFA0604903 and NSFC 11671338.

References

Arcones, M. A. and Gin, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21**, 1494-1542.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B* **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.

Chen, S. (2002). Rank estimation of transformation model. *Econometrica* **70**, 1683-97.

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.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* **64**, 103-37.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics* **24**, 540-568.

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.

Schumaker, L. (2007). *Spline Functions: Basic Theory.* Cambridge University Press, Cambridge.

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

Sherman, R. (1994). Maximal inequalities for degenerate, U-processes with applications to optimization estimators. *Annals of Statistics* **22**, 439-459.

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.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* Springer-Verlag, New York.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Tsinghua University, Beijing, P. R. China.

Email: libo@sem.tsinghua.edu.cn

Beijing Normal University, Beijing, P. R. China.

E-mail: liangbs@mail.bnu.edu.cn

Beijing Normal University, Beijing, P. R. China.

E-mail: xweitong@bnu.edu.cn

University of Missouri, Columbia, Missouri, U.S.A.

E-mail: sunj@missouri.edu

Table 1: Simulation results under different scenarios with 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

Table 2: Simulation results under different scenarios with 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

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