The Annals of Statistics
2019, Vol. 47, No. 4, 1864–1892
https://doi.org/10.1214/18-AOS1726

MAXIMUM LIKELIHOOD ESTIMATION IN TRANSFORMED LINEAR REGRESSION WITH NONNORMAL ERRORS

BY XINGWEI TONG[*,1], FUQING GAO[†,2], KANI CHEN[‡], DINGJIAO CAI[§] AND JIANGUO SUN[¶]

Beijing Normal University[*], *Wuhan University*[†], *Hong Kong University of Science and Technology*[‡], *Henan University of Economics and Law*[§] *and University of Missouri* [¶]

This paper discusses the transformed linear regression with non-normal error distributions, a problem that often occurs in many areas such as economics and social sciences as well as medical studies. The linear transformation model is an important tool in survival analysis partly due to its flexibility. In particular, it includes the Cox model and the proportional odds model as special cases when the error follows the extreme value distribution and the logistic distribution, respectively. Despite the popularity and generality of linear transformation models, however, there is no general theory on the maximum likelihood estimation of the regression parameter and the transformation function. One main difficulty for this is that the transformation function near the tails diverges to infinity and can be quite unstable. It affects the accuracy of the estimation of the transformation function and regression parameters. In this paper, we develop the maximum likelihood estimation approach and provide the near optimal conditions on the error distribution under which the consistency and asymptotic normality of the resulting estimators can be established. Extensive numerical studies suggest that the methodology works well, and an application to the data on a typhoon forecast is provided.

1. Introduction.

Linear regression is a traditional and commonly used technique for characterizing the relationship between a response variable, say Y, and a group of predictor variables, say X, and a great deal of literature has been established about inference on various linear models and their generalizations. In particular, various transformations such as log and Box–Cox transformations have been proposed to make the transformed response variable more close to a normal variable among other reasons (Box and Cox [1]; Sherman [18]). However, it is well known that sometimes there may not exist such transformation and, furthermore, one may prefer to leave the transformation arbitrary for the flexibility. That

Received August 2016; revised May 2018.
[1]Supported by the National Key Research and Development Program of China Grant 2017YFA0604903 and National Natural Science Foundation of China Grant 11671338.
[2]Supported by the National Natural Science Foundation of China under grants 11571262 and 11731012.
MSC2010 subject classifications. 62F12.
Key words and phrases. Linear transformation model, maximum likelihood estimation.

is, we face a transformed linear regression model with unknown transformation or an inference problem with some distributions other than the normal distribution.

To be specific, in this paper, we will consider the following linear transformation model:

$$H_0(Y) = -X'\beta_0 + \varepsilon, \tag{1.1}$$

where H_0 is a strictly increasing function on $\mathbb{R}$, β_0 a p-dimensional vector of regression parameters, and ε the error term with a known density function f. Note that the model above is invariant to any increasing transformation function. That is, for any strictly increasing function G, the variable $Y^* = G(Y)$ still follows the same model. On the density function f, among others, one popular class of functions is given by

$$f(x) = \frac{e^x}{(1 + re^x)^{1+1/r}},$$

and it gives the Cox model and the proportional odds model with $r = 0$ and 1, respectively. Among others, two areas where model (1.1) has been commonly used and studied are econometrics and survival analysis (Chen [4]; Doksum [8]; Han [10]; Horowitz [11]; Ma [14]; Sherman[18]). In the following, we will discuss inference about model (1.1) with f belonging to a broad family, to be denoted by $\mathcal{F}$, of density functions defined in the next section over the entire real line.

Several inference procedures have been developed for model (1.1) under different situations (Cheng [5]; Diao [7]; Zeng [23–25]; Fine [9]; Huang [12]). However, most of these methods only considered the function H_0 defined over a bounded interval mainly because this can render the technical ease when applying the empirical process theory to establish the asymptotical properties of the resulting estimators. For example, Zeng [25] investigated the problem with the normal error distribution over a bounded region. In practice, as the typhoon example discussed below, one may often face the error distribution over the entire line and/or being a nonnormal distribution. Although it may seem to be easy to generalize the idea used in Zeng [25] and others to the current situation, as will be seen below, their techniques do not work anymore for the derivation or establishment of the asymptotic properties of the derived estimators here. Correspondingly, for the proposed nonparametric maximum likelihood estimation (NPMLE) approach with the focus on the nonnormal distribution, we will employ a philosophically different approach along with some novel techniques and observations.

In particular, a key contribution or an innovation of the approach used here is the inequalities given in the Lemma A.1 in the Appendix, which allow one to be able to handle the density that is of polynomial decay. By citing the maximum likelihood criterion, we will prove that the NPMLE of the transformation function shall not have large jumps, an essential step for the establishment of the consistency of the NPMLE. Also, we will derive the minimal conditions on $\mathcal{F}$ that allow the establishment of the consistency and asymptotic normality of the NPMLE. A key step in proving the asymptotic normality will be to show that the score function of the regression parameter is a P-Donsker class (van der Vaart [20–22]), which will be achieved under stronger conditions on the tail of the transformation function. Note that as mentioned above, the focus here will be on the situation with the nonnormal error distribution. In other words, although the class $\mathcal{F}$ to be defined below include many commonly used distributions or models such as the Cox model and the proportional odds model as special case, it does not include the normal distribution. More comments on this are given below.

The rest of this paper is organized as follows. In Section 2, after introducing some notation and assumptions, we will develop the maximum likelihood estimation approach for model (1.1). Asymptotic properties of the resulting estimators are also established in the section with the proofs sketched in the Appendix. Section 3 presents some results from an extensive simulation study conducted for assessing the finite sample properties of the proposed approach, which suggest that it works well for practical situations. An application which is provided in Section 4 and Section 5 contains some discussion and concluding remarks.

2. Parameter estimation and inference procedure. In this section, we will first discuss estimation of model (1.1) and then establish the asymptotic properties of the resulting estimators along with the description of the class of the density distributions $\mathcal{F}$.

2.1. *Maximum likelihood estimation.* Suppose that the observed data consist of n i.i.d. random samples (Y_i, X_i)'s of (Y, X). To derive the likelihood function, note that the conditional distribution function of Y given X has the form

$$F_{Y|x}(y) = P(Y \le y|X = x) = P\bigl(\varepsilon \le H_0(y) + X'\beta_0\bigr)|_{X=x}.$$

This yields the conditional density function of Y

$$f_{Y|X}(y) = h_0(y) f\bigl(H_0(y) + X'\beta_0\bigr)$$

given X, where $h_0(y) = dH_0(y)/dy$. It thus follows that the log-likelihood function of $\theta = (\beta, H)$ is given by

$$l^*(\theta) = \sum_{i=1}^{n} \bigl[\log f\bigl\{H(Y_i) + X_i'\beta\bigr\} + \log h(Y_i)\bigr],$$

where $h(y) = dH(y)/dy$.

For estimation of θ, as with usual nonparametric settings, we will restrict the function H to be right continuous and have jumps only at the points Y_i's and consider the following log-likelihood function:

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \bigl[\log f\bigl\{H(Y_i) + X_i'\beta\bigr\} + \log H\{Y_i\} + \log n\bigr]. \tag{2.1}$$

In the above, $H\{Y_i\} = H(Y_i) - H(Y_i-)$ denotes the jump size of H at the point Y_i. More specifically, define the estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{H}_n)$ to be the value of θ that maximizes the log-likelihood function $l_n(\theta)$ over the parameter space $\Theta = \mathcal{B} \times \mathcal{H}$, where $\mathcal{B}$ is a subset of R^p and

$$\mathcal{H} = \{H : H(\cdot) \text{ is a nondecreasing right continuous function}\}.$$

Define $g(t) = -d \log f(t)/dt$ and $\mathbb{I}(\cdot)$ be the indicator function. For the determination of $\hat{\theta}_n$, we propose to employ the following algorithm.

Step 1. Choose the initial values $\beta^{(0)}$ and $H\{Y_k\}^{(0)}$.

Step 2. At the $t+1$th iteration, obtain the updated estimator $H\{Y_k\}^{(t+1)}$ by: for $Y_k > 0$,

$$H\{Y_k\}^{(t+1)} = \frac{\sum_{i=1}^n \mathbb{I}(Y_i = Y_k)}{\sum_{i=1}^n \mathbb{I}(Y_i \geq Y_k) g(H^{(t)}(Y_i) + X_i' \beta^{(t)})},$$

and for $Y_k < 0$,

$$H\{-Y_k\}^{(t+1)} = \frac{\sum_{i=1}^n \mathbb{I}(Y_i = Y_k)}{\sum_{i=1}^n \mathbb{I}(-Y_i \geq -Y_k) g(H^{(t)}(-Y_i) + X_i' \beta^{(t)})},$$

where $H^{(t)}(Y_i) = \sum_{j=1}^n H\{-Y_j\}^{(t)} \mathbb{I}(Y_i \geq Y_j)$.

Step 3. Obtain the updated estimator $\beta^{(t+1)}$ as the root to the following score equation:

$$U(\beta) = \sum_{i=1}^n X_i g\{H^{(t+1)}(Y_i) + X_i' \beta\} = 0.$$

Step 4. Repeat Steps 2 and 3 until the desired convergence.

For the selection of the initial estimators $\beta^{(0)}$ and $H\{Y_k\}^{(0)}$ in Step 1, one has to be careful since we allow the error distribution f to be over the entire real line. In practice and also in the numerical studies below, we suggest to create some censored data and then employ some existing methods to determine the initial estimators. More specifically, choose a large $M > 0$ and for any i such that if $|Y_i| > M$, set $Y_i^* = M$ and $\delta_i = 0$. Otherwise, set $Y_i^* = Y_i$ and $\delta_i = 1$, which gives the censored data $\{Y_i^*, \delta_i, X_i\}_{i=1}^n$. Then one can use the resulting estimators of β and H given by some existing method, say that proposed by Chen and Tong [3], based on the censored data as the initial estimators. The numerical study below indicates that this seems to work well and relatively be robust with respect to M. For the convergence of the algorithm above, we suggest to focus on estimation of β, which is usually the focus of the inference, and stop the algorithm if $\beta^{(t+1)}$ and $\beta^{(t)}$ are close enough such that their distance is smaller than a prespecified positive number. In all of the simulation studies conducted, the number of iterations needed for the convergence was less than 100 times.

2.2. *Asymptotical properties and inference.* Now we will establish the asymptotic properties of the estimator $\hat{\theta}_n$ and for this, we will first describe the regularity conditions needed. For a function g, let $\dot{g}$ and $\ddot{g}$ denote the first and second derivatives of g. We will need the following regularity conditions:

(C1) The true value β_0 is an interior point of a known compact set $\mathcal{B}$ in R^p.

(C2) The true function H_0 is strictly increasing and $\dot{H}_0(s)$ is continuous.

(C3) (i) The covariate X is uniformly bounded. That is, X takes its values in a known compact set $\mathcal{X} \subset \mathbb{R}^p$; (ii). Assume that $E\{XX'\} > 0$.

(C4) The density function f is positive and there exists a positive constant $\nu > 0$ such that

$$\limsup_{|t|\to\infty} |t|^{1+\nu} f(t) < \infty.$$

(C5) The function $g(x) = -\dot{f}(x)/f(x)$ is strictly increasing with $g(0) = 0$ and

$$\lim_{x\to\infty} g(x) = g_{+\infty}, \qquad \lim_{x\to-\infty} g(x) = g_{-\infty}.$$

(C6) The function g defined in (C5) has a bounded and continuous second derivative.

Note that Conditions (C1)–(C3) are very common in the statistical literature and Condition (C4) ensures that the density f cannot be too heavy tailed. The latter condition is generally very weak as if there exists no such ν, then both the expectation and variance of the error do not exist. In addition, many functions f satisfy Conditions (C4)–(C6) and one example is when the hazard function of ε takes the form $\exp(t)/(1 + r\exp(t))$ with $r \geq 0$. In the following, we will first describe the existence and consistency of $\hat{\theta}_n$.

THEOREM 2.1. *Assume that Conditions* (C1)–(C4) *given above hold. Then* $\hat{\theta}_n = (\hat{\beta}_n, \hat{H}_n)$ *exists and, furthermore, for any* $\tau \in (-\infty, \infty)$, *we have*

$$P\Big(-\infty < \inf_{n\geq 1} \hat{H}_n(\tau) \leq \sup_{n\geq 1} \hat{H}_n(\tau) < \infty\Big) = 1.$$

THEOREM 2.2. *Assume that Conditions* (C1)–(C4) *given above hold. Then for any* $0 < \tau < \infty$, *we have*

$$|\widehat{\beta}_n - \beta_0| \to 0 \quad \textit{and} \quad \sup_{t\in[-\tau,\tau]} \big|\widehat{H}_n(t) - H_0(t)\big| \to 0 \qquad \textit{almost surely}.$$

Note that the results above state that the estimator $\hat{H}_n$ is uniformly bounded from both above and below in any compact set of $\mathbb{R}$ and $\hat{\theta}_n$ is asymptotically consistent. To establish the asymptotic normality, without loss of generality, assume that $H(0) = 0$.

THEOREM 2.3. *Assume that Conditions* (C1)–(C6) *given above hold. Then we have that as* $n \to \infty$, $n^{1/2}(\hat{\beta}_n - \beta_0)$ *converges in distribution to a normal random variable with zero mean and the covariance matrix* $\Sigma = A_\beta^{-1} \Sigma_\beta A_\beta^{-1'}$, *where* A_β *and* Σ_β *are given in* (A.14) *and* (A.15) *of the Appendix below, respectively.*

The proofs of all results above are sketched in the Appendix. It is worth pointing out that if f is the standard normal density function, we have that $|t|^{1+\nu} f(t) \to 0$ as $t \to \pm\infty$ for any $\nu > 0$ and, therefore, f satisfies Condition (C4). In other words, although the focus here is on nonnormal distributions, the results given in Theorems 2.1 and 2.2 also hold and the proposed maximum likelihood estimator is still asymptotically consistent for the normal error distribution. In contrast, the result given in Theorem 2.3 only applies to nonnormal error distributions since the normal density does not satisfy Conditions (C5) and (C6).

For inference about β_0 and the use of the results above, it is apparent that we need to estimate the covariance matrix of $\hat{\beta}_n$, and a natural approach would be to develop some consistent estimators of the covariance matrix given in Theorem 2.3. However, this is quite difficult or not straightforward due to the large number of parameters involved. To address this, we instead propose to employ the following profile likelihood approach (Murphy [15, 16]; Murphy and van der Vaart [17]).

Specifically, for given $\beta \in \mathcal{B}$, let $\hat{H}_n(\cdot; \beta)$ denote the estimator of H_0 obtained by maximizing the log-likelihood function given in (2.1) and define the profile log-likelihood function $pl_n(\beta) = l_n\{\beta, \hat{H}_n(\cdot; \beta)\}$. For $k = 1, \dots, p$, let e_k be the p-dimensional vector of zeros except its kth component being one. Then it follows from Corollary 3 of Murphy and van der Vaart [17] that for σ_k^2, the asymptotic variance of the kth component of $\hat{\beta}_n$, we have that

$$\frac{2pl_n(\hat{\beta}_n) - pl_n(\hat{\beta}_n + s_n e_k) - pl_n(\hat{\beta}_n - s_n e_k)}{n s_n^2} \to \sigma_k^{-2}$$

in probability, where s_n is any sequence that converges to 0 in probability. This suggests that one can estimate σ_k^2 by the quantity above for finite samples and for all results below, we will use $s_n = n^{-1/2} \max\{1, |\hat{\beta}_{nk}|\}$, where $\hat{\beta}_{nk}$ denotes the kth component of $\hat{\beta}_n$. Note that in general, the profile approach for variance estimation may be sensitive to the choice of s_n sometimes. To assess this, we performed a simulation study and the results, given in part 3 of the Supplementary Material [19], indicate that the variance estimator above seems to be robust with respect to s_n.

3. A simulation study. In this section, we present some results obtained from an extensive simulation study conducted to assess the finite sample performance of the maximum likelihood estimation approach proposed in the previous section. In the study, it was assumed that the response variable Y follows the linear transformation model:

$$H_0(Y) = -X_1 \beta_{01} - X_2 \beta_{02} + \varepsilon.$$

TABLE 1
Simulation results on estimation of β_{01} with $\beta_{02} = 1$

		$n = 100$				$n = 200$			
r	β_{01}	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
		With $H_0(Y) = \log(Y)$							
0	1	−0.0293	0.2287	0.2271	0.949	−0.0105	0.1453	0.1575	0.962
	−1	0.0127	0.2197	0.2263	0.950	0.0098	0.1524	0.1576	0.960
0.5	1	0.02511	0.2796	0.3006	0.964	−0.0016	0.2005	0.2104	0.953
	−1	−0.0024	0.2875	0.3002	0.960	0.0114	0.2064	0.2108	0.957
1	1	0.0008	0.3627	0.3624	0.957	−0.0021	0.2441	0.2545	0.963
	−1	−0.0091	0.3515	0.362	0.960	0.0053	0.2481	0.2545	0.953
2	1	−0.0193	0.4697	0.464	0.954	−0.0173	0.299	0.3254	0.967
	−1	−0.0066	0.4535	0.4638	0.949	−0.0148	0.3065	0.3256	0.960
		With $H_0(Y) = Y^3$							
0	1	0.0245	0.2306	0.2368	0.956	0.0054	0.1624	0.1619	0.948
	−1	0.0169	0.2277	0.2359	0.966	0.0069	0.1629	0.1614	0.951
0.5	1	0.0123	0.3116	0.3049	0.95	0.005	0.2093	0.2118	0.956
	−1	−0.0075	0.3114	0.304	0.947	0	0.2121	0.2116	0.948
1	1	0.0043	0.3703	0.3671	0.952	−0.0063	0.2623	0.256	0.941
	−1	0.0186	0.3727	0.3679	0.942	−0.0063	0.2536	0.2557	0.941
2	1	−0.0047	0.4564	0.4751	0.955	−0.0104	0.3154	0.3296	0.963
	−1	−0.0071	0.4788	0.4756	0.955	−0.0057	0.3314	0.3297	0.947

In the above, we assumed that $H_0(Y) = \log(Y)$ for the logarithm function or $H_0(Y) = Y^3$ for the power function that is equivalent to the famous Box–Cox transformation function, the two covariates X_1 and X_2 followed the Bernoulli distribution with the success probability of 0.5 and the uniform distribution over $[-1, 1]$, respectively, and ε had the distribution with the hazard function $\exp(t)/(1 + r\exp(t))$ (Chen [2, 3]), where r is a constant. As mentioned above, when $r = 0$, we have the Cox model (Cox [6]) and $r = 1$ gives the proportional odds model. The results given below are based on $n = 100$ or 200 with 1000 replications.

Table 1 presents the results on estimation of the regression parameter β_{01} with $\beta_{01} = -1$ or 1, $\beta_{02} = 1$ and $r = 0, 0.5, 1$ and 2. The results include the estimated bias (Bias) of the proposed estimates given by the average of the estimates minus the true value, the sample standard deviations (SSE) of the estimates, the average of the estimated standard errors (ESE) and the empirical 95% coverage probability (CP). They suggest that the proposed estimator seems to be unbiased and the variance estimation also seems to be appropriate. In addition, they indicate that the normal distribution approximation appears to be reasonable. Also as expected, the estimator became more accurate when the sample size increased or when r decreased. To further see the performance of the proposed method, Figure 1 gives

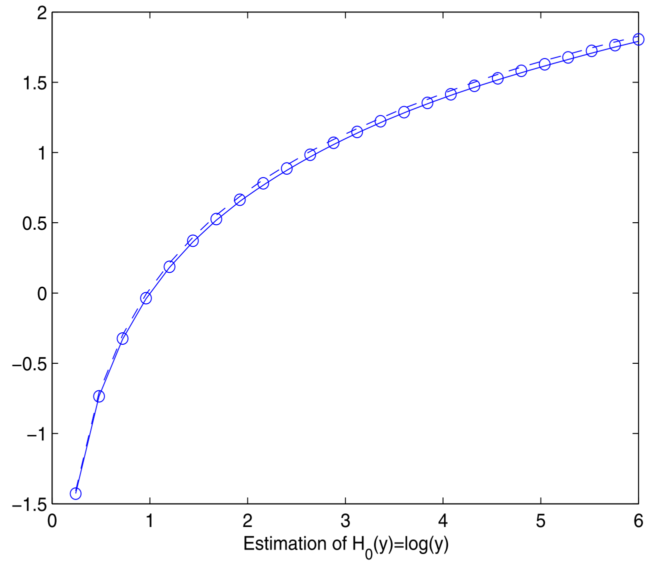

FIG. 1. *The proposed estimates of H_0 with the solid line being the true curve, the dashed line for $n = 100$ and the circle line for $n = 200$.*

the average of the proposed estimates of H_0 based on the simulated data giving Table 1 with $\beta_{01} = 1$, $r = 1$ and $H_0(Y) = \log(Y)$ and again suggests that the method seems to perform well. The results on the estimation of β_{02} are similar and some of them are given in Table 2 along with some other results.

As mentioned above, among others, Khan and Tamer [13] and Sherman [18] discussed the similar problems and developed some rank-based estimation procedures, and thus it would be interesting to compare them to the proposed estimation approach. For this, Table 2 displays the results on estimation of β_{02} given by the three methods with $\beta_{01} = 1$, $\beta_{02} = -1$ or 1, $H_0(Y) = \log(Y)$, and the other set-ups being the same as in Table 1. One can see from the table that the proposed estimator seems to have better performance than the ones given in Khan and Tamer [13] and Sherman [18]. In particular, the proposed estimator seems to be more efficient than the other two estimators.

4. Analysis of typhoon data in China. To illustrate the inference procedure proposed in the previous sections, we apply it to a set of the data consisting of 280 typhoons that occurred on the mainland of China since 1949. Among others, one objective of the study is to predict the duration time of a typhoon or investigate the relationship between the duration time and some predictive factors or covariates. It is apparent that the longer the duration time is, the more loss or damage a typhoon causes. In the analysis below, we will focus on the following five covariates, the maximum wind speed (MWS), the longitude of landing site, the central pressure,

TABLE 2
Simulation results on estimation of β_{02} with $\beta_{01} = 1$

			Proposed method				KT's method		Sherman's method	
n	r	β_{02}	Bias	SSE	ESE	CP	Bias	SSE	Bias	SSE
100	0	1	0.02188	0.2028	0.2031	0.954	0.0260	0.4088	0.0380	0.4110
		−1	−0.0157	0.2084	0.2025	0.949	−0.0585	0.3689	−0.0468	0.3678
	0.5	1	0.0079	0.2687	0.2643	0.942	0.0381	0.4768	0.0449	0.4821
		−1	−0.023	0.259	0.2643	0.967	−0.0842	0.4136	−0.0710	0.4159
	1	1	0.0205	0.3273	0.3176	0.946	0.0692	0.5114	0.0685	0.5121
		−1	−0.0174	0.3202	0.3175	0.952	−0.09616	0.4541	−0.08678	0.4506
	2	1	0.0125	0.4051	0.4066	0.949	0.0744	0.5908	0.0822	0.5934
		−1	0.0122	0.409	0.4058	0.962	−0.1394	0.5160	−0.1385	0.5161
200	0	1	0.0071	0.1404	0.1405	0.954	0.01274	0.2898	0.02592	0.2991
		−1	−0.0102	0.1432	0.1409	0.947	−0.04192	0.2715	−0.03348	0.2756
	0.5	1	0.0064	0.1862	0.1852	0.953	0.01738	0.3675	0.02488	0.3694
		−1	0.0048	0.1792	0.1854	0.951	−0.05048	0.3337	−0.04774	0.3356
	1	1	0.0022	0.2170	0.2229	0.958	0.0291	0.4389	0.02886	0.432
		−1	−0.0074	0.2217	0.2229	0.949	−0.0738	0.3792	−0.06742	0.3836
	2	1	0.0081	0.2919	0.2849	0.948	0.0266	0.4919	0.0302	0.4903
		−1	−0.0066	0.2878	0.2849	0.945	−0.0648	0.4466	−0.0623	0.4498

the moving speed (MS) and the rotation angle (RA). As a preliminary analysis, we first calculated the Kendall τ and Spearman's rank correlation coefficients between the duration time and the five covariates and present them in Table 3. One can see that it seems that the duration time is positively related to the MWS and longitude, but negatively correlated with the central pressure, MS and RA.

To apply the proposed maximum likelihood approach, let Y denote the duration time and $X_1, \dots, X_5$ the five covariates described above, respectively. Furthermore, as in the simulation study, we assume that Y follows the linear transformation model $\log\{H(Y)\} = X'\beta + \varepsilon$ with ε following the distribution whose hazard function has the form $\exp(t)/(1 + r\exp(t))$. Table 4 gives the estimated effects of the five covariates on the duration time along with the estimated standard errors for $r = 0$, 1 and 2. Here, we also tried some other values for r, too, and obtained similar results. One can see that the estimation results are quite consistent with respect

TABLE 3
Correlation coefficients between the duration time and the five covariates

	MWS	Longitude	Central Pressure	MS	RA
Spearman	0.4721	0.0464	−0.4339	−0.1888	−0.2781
Kendall	0.3562	0.0315	−0.3227	−0.1342	−0.2005

TABLE 4
Estimated effects of the five covariates on the duration time

Error distribution		MWS	Longitude	Central pressure	MS	RA
$r = 0$	Estimator	0.0285	0.0023	−0.4442	−0.0244	−0.0022
	ESE	0.0037	0.0004	0.0183	0.0043	0.0008
$r = 1$	Estimator	0.0650	0.0167	−0.4316	−0.0317	−0.0047
	ESE	0.0043	0.0045	0.0057	0.0045	0.0015
$r = 2$	Estimator	0.0851	0.0187	−0.5363	−0.0349	−0.0052
	ESE	0.0036	0.0034	0.0039	0.0036	0.0016
Normal	Estimator	0.1015	−0.0008	−0.4458	−0.0361	−0.0063

to r and also the signs of all the estimated effects are consistent with those given in Table 3. They indicate that all five factors seem to be significantly related to the duration time of a typhoon and can be useful predictors for the typhoon duration.

As mentioned above, the focus of this paper is on model (1.1) with non-normal error distributions, for which there was no established method available before. To check the appropriateness of the normality assumption here, we obtained the quantile plot of the estimated errors $\hat{\varepsilon}_i$'s against the standard normal distribution and present it in Figure 2. It is apparent that the normality assumption seems to be questionable or it does not seem to be reasonable for the typhoon data considered here. To further investigate this, we applied the proposed method to the data by as-

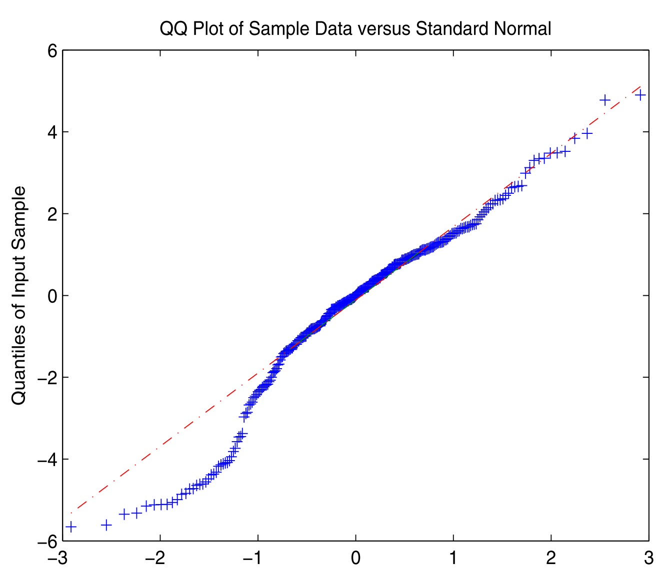

FIG. 2. *Quantile plot of the estimated errors against the standard normal distribution.*

suming that ε follows the normal distribution and include the estimated covariate effects in Table 4 too. Although the estimated effects of the other four covariates are similar to other estimates given in the table, the estimated effect of the longitude of landing site under the normal error assumption is quite different or has the opposite sign from other estimates. This again suggests that the normal assumption may not be appropriate here and one needs to apply the proposed method.

5. Discussion and concluding remarks. This paper discussed the linear transformation regression with unknown transformations with the focus on the situation where the error term follows a general, nonnormal distribution. For the problem, the maximum likelihood estimation approach was developed and the resulting estimators have been shown to be consistent. Furthermore, the estimators of regression parameters were shown to asymptotically follow the normal distribution. The simulation study suggested that the approach works well for practical situations and can be more efficient than the commonly used rank estimation procedures.

As seen from the example in Section 4, the nonnormal error distribution can occur in practice and for the situation, the analysis could yield biased or even misleading results or conclusions if the normal error distribution was used. Also as discussed above, although the similar problem has been discussed for censored data in the literature, the case considered here is actually more difficult due to the unboundedness of $\hat{H}_n$. To deal with this, some fundamentally different ideas or methods had to be used to establish the asymptotic properties of the proposed estimators. Unfortunately, in this paper, the normal case is not included in the class $\mathcal{F}$. The main reason, with the normal distribution, we cannot write the score function for β evaluated at the maximum likelihood estimator as a positive matrix times $(\hat{\beta}_n - \beta_0)$ plus $o_p(n^{-1/2})$. More specifically, $P_n X[g(\hat{H}_n(Y) + X'\hat{\beta}_n) - g(H_0(Y) + X'\beta_0)]$ may not be $O_p(n^{-1/2})$ since $\hat{H}_n(Y)$ largely deviates from $H_0(Y)$ for large $|Y|$. For the normal case, it may need some other techniques to prove the asymptotical normality of the proposed estimator.

Note that in the proposed approach, as other authors, we have assumed that the error distribution is known and it is apparent that this may not be true in reality. In other words, it would be useful to generalize the proposed method to the situation where the error distribution is completely unknown, which may not be easy. A relatively simpler problem is to assume that the error distribution belongs to some class such as the class used in the numerical studies above and in this situation, one can treat r as another parameter and estimate it by using the maximum likelihood estimation along with other parameters together. For the situation, in part 2 of the Supplementary Material [19], we show that model (1.1) is still identifiable and, furthermore, one can also generalize the profile likelihood approach described at the end of Section 2 for variance estimation. For the assessment of this approach on estimation of regression parameters, we carried out a limited simulation study and the results, given in part 4 of the Supplementary Material [19], indicate that it performed well. But more research is clearly needed. Another possible direction for the generalization of the proposed method can be to fit model (1.1) to correlated or clustered data with either normal or nonnormal error distributions.

APPENDIX A: PROOFS OF THE ASYMPTOTIC PROPERTIES OF $\hat{\theta}_n$

In this appendix, we will sketch the proof of the asymptotic properties of $\hat{\theta}$ described in Theorems 2.1–2.3. For this, we will first present two lemmas, whose proofs are shown in Appendix B.

LEMMA A.1. *For any* $c \in \mathbb{R}$, $\nu > 0$, $d_n \geq 0$, *set*

$$S_m = \max_{\delta_1,\dots,\delta_m>0}\left\{-(1+\nu)\sum_{j=1}^{m}\log(1+d_n+\Delta_j)+\sum_{j=1}^{m}\log\delta_j+mc+m\log n\right\},$$

where $\Delta_j = \sum_{k=1}^{j}\delta_k$ *and* $\Delta_0 = 0$. *Then*

$$S_m = -m\nu\log(1+d_n) - \sum_{j=1}^{m}(1+\nu j)\log\frac{1+\nu j}{\nu j}$$

$$- \sum_{j=1}^{m}\log\nu j + mc + m\log n. \tag{A.1}$$

As a result, if $\lim_{n\to\infty}\frac{m}{n} = 0$ *and* $\limsup_{n\to\infty} d_n < \infty$,

$$\limsup_{n\to\infty}\frac{S_m}{n} \leq 0. \tag{A.2}$$

LEMMA A.2. *Assume that Conditions* (C1)–(C4) *hold. Then there exists a random variable* $-\infty < \tau_0 < \infty$ *such that*

$$-\infty < \inf_{n\geq 1}\widehat{H}_n(\tau_0) \leq \sup_{n\geq 1}\widehat{H}_n(\tau_0) < \infty \qquad a.s.$$

PROOF OF THEOREM 2.1. To show the existence, it suffices to show that the jump size of $\widehat{H}_n$ at Y_i is finite and away from 0. By the compactness of $\mathcal{B}$, X_i, $i = 1,\dots,n$, and

$$l_n(\beta, H) = \frac{1}{n}\sum_{i=1}^{n}\bigl(\log f\bigl(H(Y_i)+X_i'\beta\bigr)+\log H\{Y_i\}+\log n\bigr),$$

if for some i such that $H\{Y_i\}\to\infty$, or $H\{Y_i\}\to 0$, then $l_n(\beta, H)\to -\infty$ by Condition (C4). We conclude that the jump sizes of $\widehat{H}_n$ must be finite and away from 0.

Next, we will show the boundedness. For this, suppose that $p := P(\inf_{n\geq 1} \widehat{H}_n(\tau) = +\infty) > 0$, and without loss of generality, assume $P(\widehat{H}_n(\tau) \to +\infty) > 0$. Also for convenience, assume that $Y_1 \leq Y_2 \leq \cdots \leq Y_n$, and suppose there exists a subsequence (for convenience of notation, let the subsequence be the sequence itself) such that for some $d_n \uparrow \infty$,

$$\tau_n = \sup\{t : \hat{H}_n(t) \leq d_n\} < \infty.$$

Then $P(\sup_{n\geq 1} \tau_n \leq \tau) \geq p > 0$. The proof will follow from the following four steps.

Step 1. Show

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i:\tau_0<Y_i\leq\tau_n} \{\log f(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n) + \log \widehat{H}_n\{Y_i\} + \log n\} \leq c \tag{A.3}$$

for some $c > 0$.

Let $(\tilde{\beta}_n, \tilde{H}_n)$ be the maximizer of

$$\sum_{i:\tau_0<Y_i\leq\tau_n} \{\log f(\tilde{H}(Y_i) + X_i'\tilde{\beta}) + \log \tilde{H}\{Y_i\} + \log n\}$$

subject to

$$\tilde{H}(\tau_0) = \widehat{H}(\tau_0).$$

For fixed M large enough, set $\tau_n^* = \inf\{t; \tilde{H}_n(t) > M\}$. Then

$$\frac{1}{n} \sum_{i:\tau_0<Y_i\leq\tau_n^*} \{\log f(\tilde{H}_n(Y_i) + X_i'\tilde{\beta}_n) + \log \tilde{H}_n\{Y_i\} + \log n\} \leq c_1$$

for some constant $c_1 > 0$.

By Lemma A.1,

$$\begin{aligned}
&\frac{1}{n} \sum_{i:\tau_n^*<Y_i\leq\tau_n} \{\log f(\tilde{H}_n(Y_i) + X_i'\tilde{\beta}_n) + \log \tilde{H}_n\{Y_i\} + \log n\} \\
&\quad \leq \frac{1}{n} \sum_{i:\tau_n^*<Y_i\leq\tau_n} \bigl(-(1+\nu)\log(1+\tilde{H}_n\{Y_i\}) + \log \tilde{H}_n\{Y_i\} + \log C + \log n\bigr) \\
&\quad \leq \frac{-k_n \log k_n + k_n \log n}{n} + O(1) = O(1),
\end{aligned}$$

where $k_n =$ the number of the set $\{i : \tau_n^* < Y_i \leq \tau_n\}$. Therefore,

$$\frac{1}{n} \sum_{i:\tau_0<Y_i\leq\tau_n} \{\log f(\tilde{H}_n(Y_i) + X_i'\tilde{\beta}_n) + \log \tilde{H}_n\{Y_i\} + \log n\} = O(1),$$

and so (A.3) holds.

Step 2. Show that on $\{\sup_{n\geq 1}\tau_n \leq \tau\}$,

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i:Y_i>\tau_n}\{\log f(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n)+\log\widehat{H}_n\{Y_i\}+\log n\}=-\infty. \tag{A.4}$$

Denote by $k_n^*=$ the number of the set $\{i: Y_i>\tau_n\}$. Set $J=\sup\{i; Y_i\leq\tau_n\}$ and

$$\delta_j=\widehat{H}_n\{Y_{j+J}\},\qquad \Delta_j=\sum_{i=1}^{j}\delta_i$$

Then $\widehat{H}_n(Y_{j+J})=\Delta_j+d_n$, and on $\{\sup_{n\geq 1}\tau_n\leq\tau\}$,

$$\liminf_{n\to\infty}\frac{k_n^*}{n}>0.$$

By Lemma A.1, on $\{\sup_{n\geq 1}\tau_n\leq\tau\}$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i:Y_i>\tau_n}\{\log f(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n)+\log\widehat{H}_n\{Y_i\}+\log n\}\\
&\quad\leq\frac{1}{n}\sum_{i:Y_i>\tau_n}\bigl(-(1+\nu)\log(1+d_n+\Delta_i)+\log\tilde{H}_n\{Y_i\}+\log C+\log n\bigr)\\
&\quad\leq\frac{k_n^*}{n}\biggl((-\nu\log d_n+\log C+\log n)-\frac{1}{k_n^*}\sum_{i:Y_i>\tau_n}(1+\nu j)\log\frac{1+\nu j}{\nu j}\\
&\qquad-\frac{1}{k_n^*}\sum_{i:Y_i>\tau_n}\log\nu j\biggr)\\
&\quad\leq-\frac{k_n^*}{n}\nu\log d_n-\frac{k_n^*}{n}\log\frac{k_n^*}{n}+O(1)\to-\infty.
\end{aligned}$$

Step 3. Combining Step 2 and Step 3, we have that on $\{\sup_{n\geq 1}\tau_n\leq\tau\}$,

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i:Y_i>\tau_0}\{\log f(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n)+\log\widehat{H}_n\{Y_i\}+\log n\}=-\infty. \tag{A.5}$$

Step 4. We define a nondecreasing function $H_n^*(t)=\sum_{i=1}^n H_n^*\{Y_i\}\mathbb{I}(t\geq Y_i)$, where $H_n^*(t)=\widehat{H}_n(\tau_0)$ for $t\leq\tau_0$ and $1/n$ for $Y_i>\tau_0$.

Then

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i:Y_i>\tau_0}\{\log f(H_n^*(Y_i)+X_i'\widehat{\beta}_n)+\log H_n^*\{Y_i\}+\log n\}>-\infty. \tag{A.6}$$

Since $(\widehat{\beta}_n,\widehat{H}_n)$ maximizes the log-likelihood function, one obtains that on $\{\sup_{n\geq 1}\tau_n\leq\tau\}$,

$$0\leq\limsup_{n\to\infty}l_n(\widehat{\beta}_n,\widehat{H}_n)-l_n(\widehat{\beta}_n,H_n^*)$$

$$
\begin{aligned}
&= \limsup_{n\to\infty} \frac{1}{n} \sum_{i:Y_i>\tau_0} \{\log f(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n) \\
&\quad + \log \widehat{H}_n\{Y_i\} + \log n\} \\
&\quad - \liminf_{n\to\infty} \frac{1}{n} \sum_{i:Y_i>\tau_0} \{\log f(H_n^*(Y_i) + X_i'\widehat{\beta}_n) \\
&\quad + \log H_n^*\{Y_i\} + \log n\} \\
&= -\infty.
\end{aligned}
$$

It is a contradiction. This proves $P(\sup_{n\geq 1} \widehat{H}_n(\tau) < \infty) = 1$ for all finite τ. Thus symmetrically, we can prove that $P(\sup_{n\geq 1} \widehat{H}_n(\tau) > -\infty) = 1$ for all finite τ and we complete the proof. $\square$

Now we will sketch the proof of Theorem 2.2. For this, first note that it suffices to show that for any convergent subsequence of $(\widehat{\beta}_n, \widehat{H}_n)$ such as $(\widehat{\beta}_{n_k}, \widehat{H}_{n_k}) \to (\beta^*, H^*)$, we have $(\beta^*, H^*) = (\beta_0, H_0)$. For the simplicity of notation, let the subsequence $\{n_k, k \geq 1\}$ be the original sequence. Since $(\widehat{\beta}_n, \widehat{H}_n)$ is a symmetric function of the sample $X_1, \dots, X_n$, (β^*, H^*) is measurable with respect to the exchangeable σ-field of $\{X_n, n \geq 1\}$. It then follows from the Hewitt–Savage's $0-1$ law that (β^*, H^*) is a constant. To provide the proof, we need one more lemma, whose proof is given in the Appendix B.

LEMMA A.3. *Assume that Conditions* (C1)–(C4) *hold. Then the function* $H^*(t)$ *is continuous in* $(-\infty, \infty)$.

PROOF OF THEOREM 2.2. We write

$$
\begin{aligned}
&l_n(\widehat{\beta}_n, \widehat{H}_n) \\
&\quad = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{Y_i \leq t_0\}\big(\log f(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n) + \log \widehat{H}_n\{Y_i\} + \log n\big) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{t_0 < Y_i \leq t_N\}\big(\log f(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n) + \log \widehat{H}_n\{Y_i\} + \log n\big) \\
&\qquad \times \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{Y_i > t_N\}\big(\log f(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n) + \log \widehat{H}_n\{Y_i\} + \log n\big) \\
&\quad =: \eta_1(\widehat{\beta}_n, \widehat{H}_n) + \eta_2(\widehat{\beta}_n, \widehat{H}_n) + \eta_3(\widehat{\beta}_n, \widehat{H}_n),
\end{aligned}
$$

where $-M = t_0 < t_1 < t_2 < \cdots < t_N = M$ are fixed and $M > 0$ is large enough. Since $\frac{1}{n}\sum_{i=1}^n \mathbb{I}\{Y_i > M\} \to P(Y > M)$ which goes to 0 when $M \to \infty$, by

Lemma A.1,

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\eta_1(\widehat{\beta}_n,\widehat{H}_n)\le 0. \tag{A.7}$$

Likewise,

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\eta_3(\widehat{\beta}_n,\widehat{H}_n)\le 0. \tag{A.8}$$

Now let us consider $t_k < Y_i \le t_{k+1}$ for $k = 0, 1, \dots, N-1$. Let $n_k =$ the number of $\{i; Y_i \in (t_k, t_{k+1}]\}$. Then when $n\to\infty$,

$$\begin{aligned}
&\frac{1}{n}\sum_{t_k<Y_i\le t_{k+1}}\bigl(\log f\bigl(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n\bigr)+\log\widehat{H}_n\{Y_i\}+\log n\bigr)\\
&\quad\le\frac{1}{n}\sum_{t_k<Y_i\le t_{k+1}}\Bigl(\log f\bigl(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n\bigr)+\log\frac{\widehat{H}_n(t_{k+1})-\widehat{H}_n(t_k)}{n_k}+\log n\Bigr)\\
&\quad=E\bigl(\log f\bigl(H^*(Y)+X'\beta^*\bigr)+\log\bigl(H^*(t_{k+1})-H^*(t_k)\bigr)\bigr)\mathbb{I}\{t_k<Y_i\le t_{k+1}\}\\
&\qquad+p_k\log p_k+o(1),
\end{aligned}$$

where $p_k = P(t_k < Y \le t_{k+1})$.

Let $\tilde{H}_n(t_k) = H_0(t_k)$ for $k = 0, 1, \dots, N$, and

$$\Delta\tilde{H}_n(Y_i)=\frac{\widehat{H}_n(t_{k+1})-\widehat{H}_n(t_k)}{n_k}\qquad\text{for } Y_i\in(t_k,t_{k+1}), k=0,1,\dots,N-1,$$

$$\Delta\tilde{H}_n(Y_i)=\frac{1}{n}\qquad\text{for } Y_i<-M \text{ or } Y_i>M.$$

Then by Lemma A.1,

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\eta_1(\beta_0,\tilde{H}_n)\le 0 \tag{A.9}$$

and

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\eta_3(\beta_0,\tilde{H}_n)\le 0. \tag{A.10}$$

For $k = 0, 1, \dots, N-1$,

$$\begin{aligned}
&\frac{1}{n}\sum_{t_k<Y_i\le t_{k+1}}\bigl(\log f\bigl(\tilde{H}_n(Y_i)+X_i'\beta_0\bigr)+\log\tilde{H}_n\{Y_i\}+\log n\bigr)\\
&\quad=E\bigl(\log f\bigl(H_0(Y)+X'\beta_0\bigr)\mathbb{I}\{t_k<Y_i\le t_{k+1}\}\bigr)+p_k\log\bigl(H_0(t_{k+1})-H_0(t_k)\bigr)\\
&\qquad+p_k\log p_k+r_{n,k},
\end{aligned}$$

where $r_{n,k}$ is such that

$$\lim_{\max_{-\infty\le k\le N-1}|t_{k+1}-t_k|\to 0}\ \limsup_{n\to\infty}\sum_{k=0}^{N-1}|r_{n,k}|=0.$$

Then

$$\begin{aligned}
&\limsup_{n\to\infty}\bigl(\eta_2(\widehat{\beta}_n,\widehat{H}_n)-\eta_2(\beta_0,\tilde{H}_n)\bigr)\\
&\qquad\le \sum_{k=0}^{N-1} E\bigl(\bigl(\log f\bigl(H^*(Y)+X'\beta^*\bigr)+\log\bigl(H^*(t_{k+1})-H^*(t_k)\bigr)\bigr)\\
&\qquad\quad\times \mathbb{I}\{t_k<Y\le t_{k+1}\}\bigr)\\
&\qquad\quad-\sum_{k=0}^{N-1}\bigl(E\bigl(\bigl(\log f\bigl(H_0(Y)+X'\beta_0\bigr)+\log\bigl(H_0(t_{k+1})-H_0(t_k)\bigr)\bigr)\\
&\qquad\quad\times \mathbb{I}\{t_k<Y\le t_{k+1}\}\bigr)\bigr)\\
&\qquad\quad+\limsup_{n\to\infty}\sum_{k=0}^{N-1}|r_{n,k}|
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k=0}^{N-1}\bigl(E\bigl(\bigl(\log f\bigl(H^*(Y)+X'\beta^*\bigr)+\log\bigl(H^*(t_{k+1})-H^*(t_k)\bigr)\bigr)\mathbb{I}\{t_k<Y\le t_{k+1}\}\bigr)\bigr)\\
&\quad-\sum_{k=0}^{N-1}\bigl(E\bigl(\bigl(\log f\bigl(H_0(Y)+X'\beta_0\bigr)+\log\bigl(H_0(t_{k+1})-H_0(t_k)\bigr)\bigr)\mathbb{I}\{t_k<Y\le t_{k+1}\}\bigr)\bigr)\\
&\quad=E\Biggl(\mathbb{I}\{t_0<Y\le t_N\}\log\sum_{k=0}^{N-1}\frac{f(H^*(Y)+X'\beta^*)(H^*(t_{k+1})-H^*(t_k))}{f(H_0(Y)+X'\beta_0)(H_0(t_{k+1})-H_0(t_k))}\\
&\qquad\times\mathbb{I}\{t_k<Y\le t_{k+1}\}\Biggr)\\
&\quad\le\alpha_N\log E\Biggl(\sum_{k=0}^{N-1}\frac{f(H^*(Y)+X'\beta^*)(H^*(t_{k+1})-H^*(t_k))}{f(H_0(Y)+X'\beta_0)(H_0(t_{k+1})-H_0(t_k))}\mathbb{I}\{t_k<Y\le t_{k+1}\}\Biggr)\\
&\qquad-\alpha_N\log\alpha_N,
\end{aligned}$$

where $\alpha_N=P(t_0<Y\le t_N)$.

Let $-M=t_0<t_1<\cdots<t_N=M$ such that

$$\lim_{N\to\infty}\max_{-\infty\le k\le N-1}|t_{k+1}-t_k|=0.$$

Then using Lemma A.3, as $N\to\infty$,

$$\log E\Biggl(\sum_{k=0}^{N-1}\frac{f(H^*(Y)+X'\beta^*)(H^*(t_{k+1})-H^*(t_k))}{f(H_0(Y)+X'\beta_0)(H_0(t_{k+1})-H_0(t_k))}\mathbb{I}\{t_k<Y\le t_{k+1}\}\Biggr)$$

$$= \log \int_{\mathcal{A}} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \frac{f(H^*(t) + x'\beta^*)(H^*(t_{k+1}) - H^*(t_k))}{f(H_0(t) + x'\beta_0)(H_0(t_{k+1}) - H_0(t_k))} \times f\big(H_0(t) + x'\beta_0\big) h_0(t)\, dt\, dF_X(x)$$

$$\to \log \int_{\mathcal{A}} \int_{-M}^{M} f\big(H^*(t) + x'\beta^*\big)\, dH^*(t)\, dF_X(x) \to 0 \qquad \text{as } M \to \infty.$$

Thus

$$\limsup_{M\to\infty} \limsup_{N\to\infty} \limsup_{n\to\infty} \big(l_n(\widehat{\beta}_n, \widehat{H}_n) - l_n(\beta_0, \tilde{H}_n)\big) = 0. \tag{A.11}$$

Noting that

$$E\left(\mathbb{I}t_0 < Y \le t_N \log \sum_{k=0}^{N-1} \frac{f(H^*(Y) + X'\beta^*)(H^*(t_{k+1}) - H^*(t_k))}{f(H_0(Y) + X'\beta_0)(H_0(t_{k+1}) - H_0(t_k))} \times \mathbb{I}\{t_k < Y \le t_{k+1}\}\right)$$

$$= \int_{\mathcal{A}} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \log \frac{f(H^*(t) + x'\beta^*)(H^*(t_{k+1}) - H^*(t_k))}{f(H_0(t) + x'\beta_0)(H_0(t_{k+1}) - H_0(t_k))} \times f\big(H_0(t) + x'\beta_0\big) h_0(t)\, dt\, dF_X(x)$$

and by (A.11),

$$\limsup_{M\to\infty} \limsup_{N\to\infty} \left| \int_{\mathcal{A}} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \log \frac{f(H^*(t) + x'\beta^*)(H^*(t_{k+1}) - H^*(t_k))}{f(H_0(t) + x'\beta_0)(H_0(t_{k+1}) - H_0(t_k))} \times f\big(H_0(t) + x'\beta_0\big) h_0(t)\, dt\, dF_X(x) \right| = 0.$$

Since f and h_0 have a positive lower bound and a finite upper bound on each finite interval, the above equation implies that for any $M \in (0, \infty)$,

$$\sup_{N\ge 1} \left| \sum_{k=0}^{N-1} \big(H_0(t_{k+1}) - H_0(t_k)\big) \log \frac{H_0(t_{k+1}) - H_0(t_k)}{H^*(t_{k+1}) - H^*(t_k)} \right| < \infty.$$

Now, by $x \log x = \sup_{\alpha\in\mathbb{R}}\{\alpha x - e^{\alpha-1}\}$, we have that for any $r > 0$,

$$\left| \sum_{k=0}^{N-1} \big(H_0(t_{k+1}) - H_0(t_k)\big) \log \frac{(H_0(t_{k+1}) - H_0(t_k))}{H^*(t_{k+1}) - H^*(t_k)} \right|$$

$$= \left| \sum_{k=0}^{N-1} \sup_{\alpha\in\mathbb{R}} \big\{\alpha\big(H_0(t_{k+1}) - H_0(t_k)\big) - \big(H^*(t_{k+1}) - H^*(t_k)\big) e^{\alpha-1}\big\} \right|$$

$$\geq r \sum_{k=0}^{N-1} \big|H_0(t_{k+1}) - H_0(t_k)\big| - e^{r-1} \sum_{k=0}^{N-1} \big|H^*(t_{k+1}) - H^*(t_k)\big|.$$

Thus if $\lim_{N\to\infty} \sum_{k=0}^{N-1} |H^*(t_{k+1}) - H^*(t_k)| = 0$, then

$$\begin{aligned}
&\limsup_{N\to\infty} \sum_{k=0}^{N-1} \big|H_0(t_{k+1}) - H_0(t_k)\big| \\
&\qquad \leq \frac{1}{r} \sup_{N\geq 1} \left| \sum_{k=0}^{N-1} \big(H_0(t_{k+1}) - H_0(t_k)\big) \log \frac{H_0(t_{k+1}) - H_0(t_k)}{H^*(t_{k+1}) - H^*(t_k)} \right| \\
&\qquad \to 0 \qquad \text{as } r \to \infty.
\end{aligned}$$

Thus, H_0 is absolutely continuous with respect to H^*, and for any $M > 0$,

$$\begin{aligned}
&\limsup_{N\to\infty} \int_{\mathcal{A}} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} \log \frac{f(H^*(t) + x'\beta^*)(H^*(t_{k+1}) - H^*(t_k))}{f(H_0(t) + x'\beta_0)(H_0(t_{k+1}) - H_0(t_k))} \\
&\quad \times f\big(H_0(t) + x'\beta_0\big) h_0(t)\, dt\, dF_X(x) \\
&\qquad = \int_{\mathcal{A}} \int_{-M}^{M} \log \frac{dG_x(y)}{dF_{Y|x}(y)}\, dF_{Y|x}(y)\, dF_X(x) \\
&\qquad = -E\left(\int_{-M}^{M} \left(\log \frac{dF_{Y|X}(y)}{dG_X(y)} \right) dF_{Y|X}(y) \right),
\end{aligned}$$

where

$$dG_x(y) = f\big(H^*(y) + x'\beta^*\big)\, dH^*(y).$$

Note that $|\log x| \leq 1/x$ for all $0 < x \leq 1$. We have that $(\log \frac{dF_{Y|X}(y)}{dG_X(y)})^- \leq (\frac{dF_{Y|X}(y)}{dG_X(y)})^{-1}$. Thus

$$\int_{-M}^{M} \left(\log \frac{dF_{Y|X}(y)}{dG_X(y)} \right)^- dF_{Y|X}(y) \leq G_X(M) \leq 1.$$

By the monotone convergent theorem,

$$\begin{aligned}
&\lim_{M\to\infty} E\left(\int_{-M}^{M} \left(\log \frac{dF_{Y|X}(y)}{dG_X(y)} \right)^+ dF_{Y|X}(y) \right) \\
&\qquad = E\left(\int_{-\infty}^{\infty} \left(\log \frac{dF_{Y|X}(y)}{dG_X(y)} \right)^+ dF_{Y|X}(y) \right)
\end{aligned}$$

and

$$\lim_{M\to\infty} E\left(\int_{-M}^{M} \left(\log \frac{dF_{Y|X}(y)}{dG_X(y)} \right)^- dF_{Y|X}(y) \right)$$

$$= E\left(\int_{-\infty}^{\infty}\left(\log \frac{dF_{Y|X}(y)}{dG_X(y)}\right)^{-} dF_{Y|X}(y)\right) < \infty.$$

Therefore,

$$\begin{aligned} 0 &\leq -\lim_{M\to\infty} E\left(\int_{-M}^{M}\left(\log \frac{dF_{Y|X}(y)}{dG_X(y)}\right) dF_{Y|X}(y)\right) \\ &= -E\left(\int_{-\infty}^{\infty}\left(\log \frac{dF_{Y|X}(y)}{dG_X(y)}\right) dF_{Y|X}(y)\right) \leq 0, \end{aligned}$$

which implies that

$$F_{Y|X}(y) = G_X(y).$$

That is, for any $t \in \mathbb{R}$,

$$\int_{-\infty}^{t} f\big(H^*(y) + X'\beta^*\big)\, dH^*(y) = \int_{-\infty}^{t} f\big(H_0(y) + X'\beta_0\big)\, dH_0(y).$$

This yields

$$\int_{-\infty}^{H^*(t)+X'\beta^*} f(s)\, ds = \int_{-\infty}^{H_0(t)+X'\beta_0} f(s)\, ds.$$

Thus,

$$H^*(t) - H_0(t) = -X'\big(\beta^* - \beta_0\big) \qquad \text{for all } t \in \mathbb{R}.$$

By letting $t \to -\infty$, we obtain $\beta^* = \beta_0$, and so $H^* = H_0$. This completes the proof. □

Now we consider the proof of the asymptotic normality or Theorem 2.3. For this, define $\{T_n\}$ as the values such that $P(|Y| \geq T_n) = o(n^{-3/4})$ and $E\{\dot{g}(H_0(Y) + X'\beta_0)H_0(Y)\mathbb{I}(|Y| \geq T_n)\} = o_p(n^{-3/4})$. Also for any $T > 0$, define the pseudo-distance of (θ_1, θ_2) as

$$d_T(\theta_1, \theta_2) = \sup_{|t|\leq T} \big|H_1(t) - H_2(t)\big| + \|\beta_1 - \beta_2\|.$$

For convenience, we write $\varepsilon(t) = H_0(t) + X'\beta_0$ and $\hat{\varepsilon} = \hat{H}_n(Y) + X'\hat{\beta}_n$. Again we need the following lemma, whose proof will be shown in the Appendix B.

LEMMA A.4. *Assume Conditions* (C1)–(C6) *hold. Then* $d_{T_n}(\hat{\theta}_n, \theta_0) = o_p(1)$.

PROOF OF THEOREM 2.3. For the convenience of the proof, we define the function

$$\widetilde{H}_n(t) = \begin{cases} \hat{H}_n(t) & \text{if } |t| \leq T_n, \\ \hat{H}_n(T_n) & \text{if } t \geq T_n, \\ \hat{H}_n(-T_n) & \text{if } t \leq -T_n, \end{cases}$$

which can be viewed as the truncated function of $\hat{H}_n$ at T_n. Also it can be approximately viewed as the maximizer of the likelihood function for the censored data which Y are censored by $\pm T_n$. Also for notational convenience, we write $\widetilde{\beta}_n = \hat{\beta}_n$ and $\widetilde{\theta}_n = (\widetilde{\beta}_n, \widetilde{H}_n)$.

Since the score function for β evaluated at $\hat{\theta}_n$ is zero, one obtains that $P_n g(\hat{H}_n(Y) + X'\hat{\beta}_n)X = 0$. Then it follows from the proof of Lemma A.4 that

$$
\begin{aligned}
&o_p(n^{-1/2})\\
&\quad = P_n \mathbb{I}(|Y| \le T_n) g(\hat{H}_n(Y) + X'\hat{\beta}_n)X\\
&\quad = P_n g(\varepsilon)X + P_n\{\mathbb{I}(|Y| \le T_n)\dot{g}(H_0(Y) + X'\beta_0)X[\hat{H}_n(Y) - H_0(Y)]\}\\
&\qquad + P_n\{I(|Y| \le T_n)\dot{g}(H_0(Y) + X'\beta_0)XX'\}(\hat{\beta}_n - \beta_0) + o_p(d_{T_n}(\hat{\theta}_n, \theta_0)).
\end{aligned}
$$

It follows from the definition of $\widetilde{H}_n$ that the above equation still holds by replacing $\hat{H}_n$ by $\widetilde{H}_n$ and in addition,

$$
\begin{aligned}
&E[\dot{g}(\varepsilon)XX'](\hat{\beta}_n - \beta_0) + \mathcal{A}[\widetilde{H}_n - H_0]\\
&\quad = -P_n g(\varepsilon)X + o_p(d_{T_n}(\widetilde{\theta}_n, \theta_0)) + o_p(n^{-1/2}),
\end{aligned}
\tag{A.12}
$$

where $\mathcal{A}$ is the functional operator such that

$$
\mathcal{A}[H] = \int_{-\infty}^{\infty} E_X[\dot{g}(H_0(t) + X'\beta_0) f(H_0(t) + X'\beta_0)X] H(t)\, dH_0(t),
$$

where $H \in \{H : |H(t)| \le 3H_0(t)\}$.

Now we consider the asymptotical representation of $\widetilde{H}_n(t)$. First,

$$
\widetilde{H}_n(t) = \int_0^t \frac{P_n dN_+(s)}{P_n \mathbb{I}(Y \ge s) g(\widetilde{H}_n(Y) + X'\hat{\beta}_n)},
$$

where $N_+(s) = \mathbb{I}(0 < Y \le s)$. Define the martingale process $M_+(t)$ as

$$
dM_+(s) = dN_+(s) - \mathbb{I}(Y \ge s)\lambda(H_0(s) + X'\beta_0)\, dH_0(s),
$$

where $\lambda(s)$ is the hazards function of ε.

Then it follows from the proof of Lemma A.4 that for any $s \in [0, T_n]$,

$$
\begin{aligned}
&\widetilde{H}_n(t) - H_0(t)\\
&\quad = \int_0^t \frac{P_n dN_+(s)}{P_n \mathbb{I}(Y \ge s) g(\widetilde{H}_n(Y) + X'\hat{\beta}_n)} - H_0(t)\\
&\quad = \int_0^t \frac{P_n dM_+(s)}{P \mathbb{I}(Y \ge s) g(\widetilde{H}_n(Y) + X'\hat{\beta}_n)}\\
&\qquad + \int_0^t \frac{P_n \mathbb{I}(Y \ge s)[\lambda(\varepsilon(s)) - g(\varepsilon)]\, dH_0(s)}{P_n \mathbb{I}(Y \ge s) g(\widetilde{H}_n(Y) + X'\hat{\beta}_n)}
\end{aligned}
$$

$$
\begin{aligned}
&- \int_0^t \frac{P_n \mathbb{I}(Y \geq s)[g(\widetilde{H}_n(Y) + X'\hat{\beta}_n) - g(H_0(Y) + X'\beta_0)]\, dH_0(s)}{P_n \mathbb{I}(Y \geq s) g(\widetilde{H}_n(Y) + X'\hat{\beta}_n)} \\
&= \int_0^t \frac{P_n dM_+(s)}{P\mathbb{I}(Y \geq s) g(\varepsilon)} + \int_0^t \frac{P_n \mathbb{I}(Y \geq s)[\lambda(\varepsilon(s)) - g(\varepsilon)]\, dH_0(s)}{E\mathbb{I}(Y \geq s) g(\varepsilon)} \\
&- \int_0^t \frac{P_n \mathbb{I}(Y \geq s) \dot{g}(\varepsilon)[(\widetilde{H}_n(Y) - H_0(Y)) + X'(\hat{\beta}_n - \beta_0)]}{E\mathbb{I}(Y \geq s) g(\varepsilon)}\, dH_0(s) \\
&+ o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0)\big) + o_p\big(n^{-1/2}\big).
\end{aligned}
$$

Define $U_1^+(t) = \int_0^t \frac{dM_+(s)}{E\mathbb{I}(Y \geq s) g(\varepsilon)} + \int_0^t \frac{\mathbb{I}(Y \geq s)[\lambda(\varepsilon(s)) - g(\varepsilon)]}{E\mathbb{I}(Y \geq s) g(\varepsilon)}\, dH_0(s)$, $U_2^+(t) = \int_0^t \frac{E\mathbb{I}(Y \geq s) \dot{g}(\varepsilon) X}{E\mathbb{I}(Y \geq s) g(\varepsilon)}\, dH_0(s)$ and the functional operator

$$
\mathcal{G}_+[H](t) = \int_0^\infty H(s) E_X\big[\dot{g}\big(\varepsilon(s)\big) f\big(\varepsilon(s)\big)\big]\, dH_0(s) \int_0^s \frac{\mathbb{I}(u \leq t)}{E\mathbb{I}(Y \geq s) g(\varepsilon)}\, dH_0(s),
$$

where the notation E_X is the expectation with respect to the distribution of X. This immediately yields that

$$
\begin{aligned}
\widetilde{H}_n(t) - H_0(t) &= P_n U_1^+(t) - U_2^+(t)'(\hat{\beta}_n - \beta_n) \\
&\quad - \mathcal{G}_+[\widetilde{H}_n - H_0](t) + o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0) + n^{-1/2}\big).
\end{aligned}
$$

Similar to the above arguments, we can obtain that for $t \in [-T_n, 0]$,

$$
\begin{aligned}
\widetilde{H}_n(t) - H_0(t) &= P_n U_1^-(t) - U_2^-(t)'(\hat{\beta}_n - \beta_n) \\
&\quad - \mathcal{G}_-[\widetilde{H}_n - H_0](t) + o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0) + n^{-1/2}\big),
\end{aligned}
$$

where $U_1^-(t), U_2^-(t), \mathcal{G}_-$ can be similarly defined as above. Combing with the above two equalities yields that for $t \in [-T_n, T_n]$,

$$
\begin{aligned}
\widetilde{H}_n(t) - H_0(t) &= P_n U_1(t) - U_2(t)'(\hat{\beta}_n - \beta_n) \\
&\quad - \mathcal{G}[\widetilde{H}_n - H_0](t) + o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0) + n^{-1/2}\big),
\end{aligned}
$$

where $U_1(t) = U_1^+(t) + U_1^-(t)$ and so are $U_2(t)$ and $\mathcal{G}$, which gives that

$$
\begin{aligned}
\widetilde{H}_n(\cdot) - H_0(\cdot) &= P_n (I + \mathcal{G})^{-1}[U_1](\cdot) - (I + \mathcal{G})^{-1}[U_2](\cdot)'(\hat{\beta}_n - \beta_n) \\
&\quad + o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0) + n^{-1/2}\big).
\end{aligned}
$$

This, together with (A.12), gives that

$$
\begin{aligned}
&\big[E\{\dot{g}(\varepsilon) X X'\} - \mathcal{A}\big[(I + \mathcal{G})^{-1}[U_2](Y)\big]'\big](\hat{\beta}_n - \beta_0) \\
&\quad = -P_n\big[g(\varepsilon) X + (I + \mathcal{G})^{-1}[U_1](Y)\big] + o_p\big(d_{T_n}(\widetilde{\theta}_n, \theta_0)\big) + o_p\big(n^{-1/2}\big),
\end{aligned}
$$

and thus that $\sqrt{n}(\hat{\beta}_n - \beta_0) = -[E\{\dot{g}(\varepsilon)XX'\} - \mathcal{A}[(I+\mathcal{G})^{-1}[U_2](Y)]']^{-1}\sqrt{n}P_n \times [g(\varepsilon)X + (I+\mathcal{G})^{-1}[U_1](Y)] + o_p(\sqrt{n}\, d_{T_n}(\widetilde{\theta}_n, \theta_0)) + o_p(1)$. Therefore, we obtain that

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

where

$$\Sigma = A_\beta^{-1}\Sigma_\beta A_\beta^{-1'}, \tag{A.13}$$

and

$$A_\beta = E\{\dot{g}(\varepsilon)XX'\} - \mathcal{A}\big[(I+\mathcal{G})^{-1}[U_2](Y)\big]', \tag{A.14}$$

$$\Sigma_\beta = \operatorname{cov}\big[g(\varepsilon)X + (I+\mathcal{G})^{-1}[U_1](Y)\big]. \tag{A.15}$$

The proof is complete. □

APPENDIX B: PROOFS OF THE LEMMAS IN APPENDIX A

PROOF OF LEMMA A.1. Set $\nu_j = j\nu$. For any $a > 0$, consider the function the function

$$g(\delta) = -(1+\nu)\log(a+\delta) + \log\delta.$$

Then $g'(\delta) = \frac{-\nu\delta + a}{\delta(a+\delta)}$, $g'(0+) > 0$ and $g'(+\infty) = -\infty$. The function $\tilde{g}(\delta) = -\nu\delta + a$ is strictly decreasing. Thus $g(\delta)$ attains its maximum at a/ν:

$$-(1+\nu)\log(1+\nu) - \nu\log\frac{a}{\nu} = -(1+\nu)\log\frac{1+\nu}{\nu} - \log\nu - \nu\log a.$$

For $d_n > 0$, set

$$g_k(\nu) = \max_{\delta_k > 0} -(1+\nu)\log(1 + d_n + \Delta_{k-1} + \delta_k) + \log\delta_k$$

for $k = 1, \dots, m$. Then it follows that

$$g_m(\nu) = -\nu\log(1 + d_n + \Delta_{m-1}) - (1+\nu)\log\frac{1+\nu}{\nu} - \log\nu,$$

and thus

$$S_m = \max_{\delta_1,\dots,\delta_{m-1}>0}\left\{g_{m-1}(\nu_2) + \sum_{j=1}^{m-2} g_j(\nu_1)\right\} - (1+\nu)\log\frac{1+\nu}{\nu} - \log\nu + mc + m\log n.$$

This yields (A.1) by induction. □

PROOF OF LEMMA A.2. Set $\vartheta = 2\max_{x\in\mathcal{X},\beta\in\mathcal{B}}|x'\beta|$. By Condition (C4), there exists a positive constant C such that $\sup_{|s-t|\le\vartheta} f(s) \le \frac{C}{(1+|t|)^{1+\nu}}$, $t \in \mathbb{R}$, For a given $M > \vartheta + 2$, assume that there exists a subsequence $\{n_k, k \ge 1\}$ such that

$$|\widehat{H}_{n_k}(Y_i)| > M \qquad \text{for all } i = 1, 2, \dots, n_k$$

For simplicity of notation, let the sequence $\{n_k, k \ge 1\}$ be the original sequence $\{n, n \ge 1\}$ and assume $Y_1 < \cdots < Y_n$. Set

$$m = \sup\{k \le n; \widehat{H}_n(Y_k) < -M\},$$

here $\sup\varnothing = 0$. Then $\widehat{H}_n(Y_{m+1}) > M$, and

$$\begin{aligned}
l_n(\widehat{\beta}_n, \widehat{H}_n) &= \frac{1}{n}\sum_{i=1}^{n}\bigl(\log f\bigl(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n\bigr) + \widehat{H}_n\{Y_i\} + \log n\bigr) \\
&= \frac{1}{n}\sum_{i=1}^{m}\bigl(\log f\bigl(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n\bigr) + \widehat{H}_n\{Y_i\} + \log n\bigr) \\
&\quad + \frac{1}{n}\sum_{i=m+1}^{n}\bigl(\log f\bigl(\widehat{H}_n(Y_i) + X_i'\widehat{\beta}_n\bigr) + \widehat{H}_n\{Y_i\} + \log n\bigr) \\
&=: \xi_1 + \xi_2.
\end{aligned}$$

It follows from $\sup_{|s-t|\le\vartheta} f(s) \le \frac{C}{(1+|t|)^{1+\nu}}$, $t \in \mathbb{R}$, that

$$n\xi_1 \le m\log C - (1+\nu)\sum_{i=1}^{m}\bigl(\log\bigl(1 + \widehat{H}_n(Y_i)\bigr) + \log\widehat{H}_n\{Y_i\} + m\log n\bigr).$$

Applying Lemma A.1 to $d_n = -\widehat{H}_n(Y_m)$, $\delta_j = \widehat{H}_n\{Y_{n-j+1}\}$, $j = 1, \dots, n$, we obtain that

$$n\xi_1 \le -m\nu\log(1+M) - \sum_{j=1}^{m}(1+\nu j)\log\frac{1+\nu j}{\nu j} - \sum_{j=1}^{m}\log\nu j + m\log C + m\log n.$$

Similarly,

$$\begin{aligned}
n\xi_2 \le{}& -(n-m)\nu\log(1+M) - \sum_{j=1}^{n-m}(1+\nu j)\log\frac{1+\nu j}{\nu j} - \sum_{j=1}^{n-m}\log\nu j \\
&+ (n-m)\log C + (n-m)\log n.
\end{aligned}$$

Thus

$$\begin{aligned}
&\limsup_{n\to\infty} l_n(\widehat{\beta}_n, \widehat{H}_n) \\
&\quad \le -\mu\log(1+M) + \limsup_{n\to\infty}\frac{(n-m)(\log C + \log\nu) + m(\log C + \log\nu)}{n}
\end{aligned}$$

$$+\limsup_{n\to\infty}\frac{(n-m)\log\frac{n}{n-m}+m\log\frac{n}{m}}{n}$$

$$\le -\mu\log(1+M)+(\log C+\log\nu)+2\sup_{x\ge 1}\frac{\log x}{x}\to -\infty$$

as $M\to\infty$. $\square$

PROOF OF LEMMA A.3. Suppose H^* is discontinuous at some point $\tau\in\mathbb{R}$. Let $\tau_1<\tau_2<\tau<\tau_3$. Let $\bar{H}_n$ be defined by

$$\bar{H}_n(t)=\widehat{H}_n(t)\qquad\text{for } t<\tau_1 \text{ or } t>\tau_3,$$

and

$$\Delta\bar{H}_n(Y_i)=\frac{\widehat{H}_n(\tau_3)-\widehat{H}_n(\tau_1)}{m}\qquad\text{for } Y_i\in(\tau_1,\tau_3),$$

where $m=$ the number of $\{i;Y_i\in(\tau_1,\tau_3)\}$. Then

$$\begin{aligned}
&l_n(\widehat{\beta}_n,\widehat{H}_n)-l_n(\widehat{\beta}_n,\bar{H}_n)\\
&\quad=\sum_{Y_i\in(\tau_1,\tau_3)}\bigl(\log f\bigl(\widehat{H}_n(Y_i)+X_i'\widehat{\beta}_n\bigr)-\log f\bigl(\bar{H}_n(Y_i)+X_i'\widehat{\beta}_n\bigr)\bigr)\\
&\qquad+\sum_{Y_i\in(\tau_1,\tau_3)}\bigl(\log\widehat{H}_n\{Y_i\}-\log\bar{H}_n\{Y_i\}\bigr)\\
&\quad\le m\tilde{C}+\sum_{Y_i\in(\tau_1,\tau_2)}\log\frac{\widehat{H}_n(\tau_2)-\widehat{H}_n(\tau_1)}{m_1}+\sum_{Y_i\in(\tau_2,\tau_3)}\log\frac{\widehat{H}_n(\tau_3)-\widehat{H}_n(\tau_2)}{m_2}\\
&\qquad+m\log\frac{\widehat{H}_n(\tau_3)-\widehat{H}_n(\tau_1)}{m}\\
&\quad\le m\tilde{C}+m_1\log\frac{\widehat{H}_n(\tau_2)-\widehat{H}_n(\tau_1)}{m_1}+m_2\log\frac{\widehat{H}_n(\tau_3)-\widehat{H}_n(\tau_2)}{m_2}\\
&\qquad+m\log\frac{\widehat{H}_n(\tau_3)-\widehat{H}_n(\tau_1)}{m},
\end{aligned}$$

where $m_1=$ the number of $\{i;Y_i\in(\tau_1,\tau_2)\}$, $m_2=$ the number of $\{i;Y_i\in(\tau_2,\tau_3)\}$ and

$$\tilde{C}=\sup_{|x-y|\le\tau_3-\tau_1,|x|\le M,|y|\le M}\bigl|\log f(x)-\log f(y)\bigr|.$$

Let $\tau_2-\tau_1=\tau_3-\tau_2$. Then

$$\limsup_{n\to\infty}\frac{1}{m}\bigl(l_n(\widehat{\beta}_n,\widehat{H}_n)-l_n(\widehat{\beta}_n,\bar{H}_n)\bigr)$$

$$\le \tilde{C} + p_1 \log \frac{\widehat{H}_n(\tau_2) - \widehat{H}_n(\tau_1)}{\widehat{H}_n(\tau_3) - \widehat{H}_n(\tau_1)} + p_2 \log \frac{\widehat{H}_n(\tau_3) - \widehat{H}_n(\tau_2)}{\widehat{H}_n(\tau_3) - \widehat{H}_n(\tau_1)} + p_1 \log p_1 + p_2 \log p_2,$$

where

$$p_1 = \limsup_{n\to\infty} \frac{m_1}{m} = \frac{P(Y \in (\tau_1, \tau_2))}{P(Y \in (\tau_1, \tau_3))}, \qquad p_2 = 1 - p_1.$$

Let $|\tau_2 - \tau_1| = |\tau_3 - \tau_2| \to 0$, then $p_1 \to 1/2$ and $p_2 \to 1/2$, and

$$H^*(\tau_2) - H^*(\tau_1) \to 0, H^*(\tau_3) - H^*(\tau_2) \to \Delta H^*(\tau).$$

As a result,

$$\limsup_{|\tau_2-\tau_1|=|\tau_3-\tau_2|\to 0} \limsup_{n\to\infty} \frac{1}{m}\big(l_n(\widehat{\beta}_n, \widehat{H}_n) - l_n(\widehat{\beta}_n, \bar{H}_n)\big) = -\infty.$$

This is a contradiction. □

PROOF OF LEMMA A.4. Since Theorem 2.2 shows that $|\hat{\beta}_n - \beta_0| = o_p(1)$, it suffices to prove that $\sup_{|t|\le T_n} |\hat{H}_n(t) - H_0(t)| = o_p(1)$.

First, we show that $\sup_{0<t\le T_n} |\hat{H}_n(t) - H_0(t)| = o_p(1)$. This is equivalent to proving that for any sufficiently small $\eta > 0$, the probability of the event $A_n = \{\sup_{0<t\le T_n} |\hat{H}_n(t) - H_0(t)| < C\eta\}$ for some constant $C > 0$ tends to one.

From Conditions (C5)–(C6), there exists a $T_\eta > 0$, such that

$$g_{+\infty} \ge \sup_{t>T_\eta, x\in\mathcal{X}, \beta\in\mathcal{B}} g\big(H_0(t) + x'\beta\big) \ge \inf_{t>T_\eta, x\in\mathcal{X}, \beta\in\mathcal{B}} g\big(H_0(t) + x'\beta\big) \ge g_{+\infty} - \eta.$$

By differentiating the log-likelihood function w.r.t. $H\{Y_i\}$, we obtain that

$$\hat{H}_n(t) = \int_0^t \frac{P_n dI(Y \le s)}{P_n\{\mathbb{I}(Y \ge s) g(\hat{H}_n(Y) + X'\hat{\beta}_n)\}}.$$

It follows from Theorem 2.2 that $|\hat{H}_n(t) - H_0(t)| = o_p(1)$ uniformly for $t \in [0, T_\eta]$. Observe that uniformly for $s \in [0, T_n]$,

$$\begin{aligned}
&\big|P_n \mathbb{I}(Y \ge s)\big(g(\hat{\varepsilon}) - g(\varepsilon)\big)\big| \\
&\quad \le P_n\big|\mathbb{I}(Y > T_n) g(\varepsilon)\big| + \big|P_n \mathbb{I}(Y > T_n) g(\hat{\varepsilon})\big| \\
&\qquad + \big|P_n \mathbb{I}(Y \ge s, Y < T_\eta)\big[g(\hat{\varepsilon}) - g(\varepsilon)\big]\big| \\
&\qquad + P_n \mathbb{I}(Y \ge s, T_\eta < Y < T_n)\big|g(\hat{\varepsilon}) - g(\varepsilon)\big| \\
&\quad \le 2g_{+\infty} P(Y > T_n) + \text{constant} * d_{T_\eta}(\hat{\theta}_n, \theta_0) + 2\eta P_n \mathbb{I}(Y > s).
\end{aligned}$$

Using the equality: $a^{-1} = b^{-1} - (a-b)/[ab]$ for any two scalars $a, b \neq 0$, we have that

$$\begin{aligned}\hat{H}_n(t) &= \int_0^t \frac{P_n dI(Y \leq s)}{P_n\{\mathbb{I}(Y \geq s)g(\hat{H}_n(Y) + X'\hat{\beta}_n)\}} \\ &= \int_0^t \frac{P_n dI(Y \leq s)}{P_n\{\mathbb{I}(Y \geq s)g(H_0(Y) + X'\beta_0)\}} \\ &\quad - \int_0^t \frac{P_n\{\mathbb{I}(Y \geq s)(g(\hat{\varepsilon}) - g(\varepsilon))\}}{P_n\{\mathbb{I}(Y \geq s)g(\varepsilon)\}P_n\{I(Y \geq s)g(\hat{\varepsilon})\}} P_n dI(Y \leq s).\end{aligned}$$

Since the function class of the $\{\mathbb{I}(y \geq s) : s \geq 0\}$ is the Glinvenko–Cantelli class, the first term of the right-hand side uniformly converges to $H_0(t)$ as n tends to infinity.

This yields that, for sufficiently large n,

$$|\hat{H}_n(t) - H_0(t)| \leq 3\eta \int_0^t \frac{P\mathbb{I}(Y \geq s)}{P\{g(H_0(Y) + X'\beta_0)\mathbb{I}(Y \geq s)\}} dH_0(s) + o_p(1),$$

which completes the proof of this lemma. □

Acknowledgments. The authors are grateful to the former Editor, Dr. George, the Associate Editor and two reviewers for their many comments and suggestions that greatly improved the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Maximum likelihood estimation in transformed linear regression with nonnormal errors" (DOI: 10.1214/18-AOS1726SUPP; .pdf). Due to space constraints, the proofs of the consistency of the proposed covariance matrix estimator and the identifiability of model (1.1) along with some additional simulation results are provided in the Supplementary Material.

REFERENCES

[1] BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252. MR0192611

[2] CHEN, K., JIN, Z. and YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89** 659–668. MR1929170

[3] CHEN, K. and TONG, X. (2010). Varying coefficient transformation models with censored data. *Biometrika* **97** 969–976. MR2746165

[4] CHEN, S. (2002). Rank estimation of transformation models. *Econometrica* **70** 1683–1697. MR1929984

[5] CHENG, S. C., WEI, L. J. and YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82** 835–845. MR1380818

[6] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

[7] DIAO, G., ZENG, D. and YANG, S. (2013). Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics* **69** 840–849. MR3146780

[8] DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15** 325–345. MR0885740

[9] FINE, J. P., YING, Z. and WEI, L. J. (1998). On the linear transformation model with censored observations. *Biometrika* **85** 980–986.

[10] HAN, A. K. (1987). A nonparametric analysis of transformations. *J. Econometrics* **35** 191–209. MR0903182

[11] HOROWITZ, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* **64** 103–137. MR1366143

[12] HUANG, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563. MR1742499

[13] KHAN, S. and TAMER, E. (2007). Partial rank estimation of duration models with general forms of censoring. *J. Econometrics* **136** 251–280. MR2328593

[14] MA, S. and KOSOROK, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Ann. Statist.* **33** 2256–2290. MR2211086

[15] MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731. MR1292537

[16] MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23** 182–198. MR1331663

[17] MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168

[18] SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705

[19] TONG, X., GAO, F., CHEN, K., CAI, D. and SUN, J. (2019). Supplement to "Maximum likelihood estimation in transformed linear regression with nonnormal errors." DOI:10.1214/18-AOS1726SUPP.

[20] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[21] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.

[22] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics. Springer Series in Statistics*. Springer, New York. MR1385671

[23] ZENG, D. and LIN, D. Y. (2006). Maximum likelihood estimation in semiparametric transformation models for counting processes. *Biometrika* **93** 627–640.

[24] ZENG, D. and LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 507–564. MR2370068

[25] ZENG, D. L., MICHAEL, R. and KOSOROK, L. D. Y. (2014). Nonparametric maximum likelihood estimation in linear regression with unknown transformation. Unpublished manuscript.

X. TONG
SCHOOL OF STATISTICS
BEIJING NORMAL UNIVERSITY
BEIJING
CHINA
E-MAIL: xweitong@bnu.edu.cn

F. GAO
SCHOOL OF MATHEMATICS AND STATISTICS
WUHAN UNIVERSITY
WUHAN
CHINA
E-MAIL: fqgao@whu.edu

K. CHEN
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY
OF SCIENCE AND TECHNOLOGY
KOWLOON
HONG KONG
E-MAIL: makchen@hkust.edu

D. CAI
SCHOOL OF MATHEMATICS
AND INFORMATION SCIENCE
HENAN UNIVERSITY OF ECONOMICS AND LAW
HENAN
CHINA
E-MAIL: djcai2002@163.com

J. SUN
DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65201
USA
E-MAIL: sunj@missouri.edu