SUPPLEMENTAL MATERIAL FOR "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*

S.1. Proofs of the Asymptotic Properties of the Proposed Estimator

In this section, we will sketch the proofs of Theorems 1 and 2 given in the paper.

Proof of Theorem 1: We will prove the rate of convergence by using empirical process theory. Denote $\Theta = \{(\beta, \phi) : \beta \in \mathcal{B}, \phi \in \mathcal{F}_r\}$, $\Theta_n = \{\theta = (\beta, \phi) : \beta \in \mathcal{B}, \phi(\cdot) = B_n(\cdot)^T \alpha\}$, where $q_n = [n^v] + l$. And let $y_i = (x_i, z_i, w_i, \delta_i, v_i)$ and $f_n(y_1, y_2, \theta) = f_n^*(y_1, y_2, \theta) - f_n^*(y_1, y_2, \theta_0)$, where

$$f_n^*(y_1, y_2, \theta) = \delta_2 I(v_1 \geq v_2) s_n(x_1^T \beta + z_1 B_n(w_1)^T \alpha - x_2^T \beta - z_2 B_n(w_2)^T \alpha)$$

and we further denote $\Gamma_n(\theta) = O_n(\theta) - O_n(\theta_{n0})$. Then by following Song et al. (2007), we can write

$$\Gamma_n(\theta) = \Gamma_{n0}(\theta) + P_n g_n(\cdot, \theta) + U_n h_n(\cdot, \cdot, \theta), \tag{1}$$

where

$$\Gamma_{n0}(\theta) = Ef_n(\cdot, \cdot, \theta),$$

$$g_n(y, \theta) = Ef_n(y, \cdot, \theta) + Ef_n(\cdot, y, \theta) - 2\Gamma_{n0}(\theta),$$

$$h_n(y_1, y_2, \theta) = f_n(y_1, y_2, \theta) - Ef_n(y_1, \cdot, \theta) - Ef_n(\cdot, y_2, \theta) + \Gamma_{n0}(\theta).$$

Define $\mathcal{F}_n = \{f_n(y_1, y_2, \theta), \theta \in \Theta_n\}$, $\mathcal{G}_n = \{g_n(y, \theta), \theta \in \Theta_n\}$, $\mathcal{H}_n = \{h_n(y_1, y_2, \theta), \theta \in \Theta_n\}$. In the following, we use C to denote a generic constant not depending on n which may vary at different places. Note that the first order derivative of s_n is bounded by 1, we can easily verify the following Lipschitz conditions:

$$\|f_n(x_1, x_2, \theta_1) - f_n(x_1, x_2, \theta_2)\|_2 \le C\rho(\theta_1, \theta_2),$$

$$\|g_n(x, \theta_1) - g_n(x, \theta_2)\|_2 \le C\rho(\theta_1, \theta_2),$$

and Corollary 2.7.2 in van der Vaart and Wellner (1996) implies that

$$\log N_{[]}(\epsilon, \Theta_n, \rho) \le Cq_n \log(1/\epsilon).$$

We can combine the preceding 3 inequalities to derive

$$\int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{G}_n, \|\cdot\|_2)} \le Cq_n^{1/2}\delta. \tag{2}$$

Then by appling Lemma 3.4.3 in van der Vaart and Wellner (1996), we obtain

$$E\big\{\sup_{\rho(\theta,\theta_{n0})\leq\delta}|P_n g_n(\theta))|\big\} \leq C q_n^{1/2}\delta\,. \tag{3}$$

Similarly, we can obtain the entropy bound for the class of functions $\mathcal{H}_n$. Because $\{U_n h_n(\cdot,\cdot,\theta),\ \theta\in\Theta_n\}$ is a degenerated U-process. We can apply the maximum inequality for a degenerated U-process (Section 5, Sherman (1994)) to derive

$$E\big\{\sup_{\rho(\theta,\theta_{n0})\leq\delta}|U_n h_n(\theta))|\big\} \leq C q_n^{1/2}\delta. \tag{4}$$

Also by using a Taylor expansion argument, we can show that

$$E\big\{\sup_{\rho(\theta,\theta_{n0})\leq\delta}|\Gamma_{n0}(\theta)|\big\} \leq C\delta. \tag{5}$$

In light of (1), we can combine (3), (4) and (5) to obtain that

$$E\big\{\sup_{\rho(\theta,\theta_{n0})\leq\delta}|\Gamma_n(\theta)|\big\} \leq C q_n^{1/2}\delta. \tag{6}$$

It can be seen from the proof (A.19) of Theorem 2.1 in Khan and Tamer (2007) and the proof of Theorem 4.1 in Song et al. (2007) that θ_{n0} is the maximizer of $E[O_n(X,\theta)]$ and $\hat{\theta}$ approaches to θ_{n0} in probability. We can use a Taylor expansion argument to show $0 \leq E\Gamma_n(\theta) = E[O_n(X,\theta)] - E[O_n(X,\theta_{n0})] \leq -C\rho^2(\theta,\theta_{n0})$ for $\theta\in\{\theta:\rho(\theta,\theta_{n0})\leq\delta\}$. Furthermore by applying the Theorem 3.2.5 in van der Vaart and Wellner (1996) with

$\phi(w) = q_n^{1/2} w$, we obtain $\rho(\hat{\theta}, \theta_{n0}) = O_p((n/q_n)^{1/2}) = O_p(n^{-(1-v)/2})$. This plus the fact that the approximation error is well known to be $\rho(\theta_{n0}, \theta_0) = O(n^{-rv})$ (Schumaker (2007)) yields that $\rho(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-v)/2} + n^{-rv})$.

□

Proof of Theorem 2: We will extend the general arguments in the Theorem 6.1 of Huang (1996) to U-type Z-estimation to show asymptotic normality. First define $\Theta_{n0} = \{\theta : \rho(\theta, \theta_0) < \delta\}$, and denote the following (Gâteau) derivatives $f_1(y_1, y_2, \theta) = \frac{\partial}{\partial \beta} f(y_1, y_2, \theta)$, $f_2(y_1, y_2, \theta)[\phi] = \frac{\partial}{\partial \phi} f(y_1, y_2, \theta)[\phi]$, $F_{11}(\theta) = \frac{\partial^2}{\partial \beta^2} E[f(Y_1, Y_2, \theta)]$, $F_{12}(\theta)[\phi_1] = \frac{\partial^2}{\partial \beta \partial \phi} E[f(Y_1, Y_2, \theta)][\phi_1]$, $F_{22}(\theta)[\phi_1, \phi_2] = \frac{\partial^2}{\partial \phi^2} E[f(Y_1, Y_2, \theta)][\phi_1, \phi_2]$, and the U-operator

$$U_n f = \frac{2}{n(n-1)} \sum_{i \neq j} f(Y_i, Y_j) \,.$$

By the definition of $\hat{\beta}$ and $\hat{\phi}$, we have

$$U_n f_1(\hat{\beta}, \hat{\phi}) = 0, \; U_n f_2(\hat{\beta}, \hat{\phi})[\phi] = 0 \text{ for all } \phi. \tag{7}$$

It then follows from the identification condition that

$$P f_1(\beta_0, \phi_0) = 0, P f_2(\beta_0, \phi_0)[\phi] = 0 \text{ for all } \phi. \tag{8}$$

By the geometry of Hilbert spaces, there exists a ϕ^* satisfying $F_{12}[\phi] - F_{22}[\phi^*, \phi] = 0$ for all ϕ. Using the calculation of the bracketing number and similar decomposition in (1), we can obtain the following asymptotic equicontinuties (Arcones and Gin (1993))

$$\sup_{\theta\in\Theta_{0n}}|(U_n-P)f_1(\beta,\phi)-(U_n-P)f_1(\beta_0,\phi_0)|=o_p(n^{-1/2}),$$

$$\sup_{\theta\in\Theta_{0n}}|(U_n-P)f_2(\beta,\phi)[\phi^*]-(U_n-P)f_2(\beta_0,\phi_0)[\phi^*]|=o_p(n^{-1/2}).$$

The combination of (7) and (8) with the foregoing uniform bounds leads to

$$U_nf_1(\beta_0,\phi_0)-Pf_1(\hat\beta,\hat\phi)=o_p(n^{-1/2}),$$

$$U_nf_2(\beta_0,\phi_0)[\phi^*]-Pf_2(\hat\beta,\hat\phi)[\phi^*]=o_p(n^{-1/2}).$$

Furthermore, by Taylor expansion and noting that $\rho(\hat\theta_n,\theta_0)=o_p(n^{-1/4})$, we have

$$U_nf_1(\beta_0,\phi_0)-(F_{11}(\theta_0)(\hat\beta-\beta_0)+F_{12}(\theta_0)[\hat\phi-\phi_0])=o_p(n^{-1/2}),$$

$$U_nf_2(\beta_0,\phi_0)[\phi^*]-(F_{21}(\theta_0)[\phi^*](\hat\beta-\beta_0)+F_{22}(\theta_0)[\phi^*,\hat\phi-\phi_0])=o_p(n^{-1/2}).$$

Since $F_{12}[\phi]-F_{22}[\phi^*,\phi]=0$ for all ϕ, it follows from the preceding two equations that

$$\begin{aligned}&n^{1/2}(\hat\beta-\beta_0)\\ &=n^{1/2}(F_{11}(\theta_0)-F_{21}(\theta_0)[\phi^*])^{-1}U_n(f_1(\beta_0,\phi_0)-f_2(\beta_0,\phi_0)[\phi^*])+o_p(1).\end{aligned}$$

Denote $\Sigma_1=F_{11}(\theta_0)-F_{21}(\theta_0)[\phi^*]$ and $f^*=f_1(\beta_0,\phi_0)-f_2(\beta_0,\phi_0)[\phi^*]$. It follows from the classic degenerated U-statistic theory (see, e.g., Theorem

12.3 in van der Vaart (2000) that

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

in distribution, which completes the proof. In the above,

$$\Sigma = 4\Sigma_1^{-1} Cov(f^*(Y_1, Y_2), f^*(Y_1, Y_2'))\Sigma_1^{-1} \tag{9}$$

with Y_1, Y_2, Y_2' being i.i.d copies of Y. □

S.2. Additional Simulation Results

In this part, we present some additional simulation results obtained similarly as those given in the paper but for different purposes. One is that for all of the error distributions considered in the paper, they have relatively light-tails and it is apparent that it would be useful to assess the performance of the proposed method with heavy-tail error distributions. Corresponding to this, we considered the t-distribution with the degree of freedoms being 1 or 5, which will be denoted by $t(1)$ or $t(5)$, respectively. Note that both have heavy-tails than the error distribution considered in the paper and $t(1)$ corresponds to the Cauchy distribution, well-known for its fat tails. Table S.1 presents some results on estimation of regression parameter β_0 given by the proposed method under the two error distributions with $\beta_0 = 1$, $\sigma^2 = 0.5$ and the other set-ups being the same as with Table

1 of the paper. In particular, the censoring times were generated from the exponential distribution that yielded around 20% censoring rates.

The results in Table S.1 suggest that as with the situations considered in the paper, the proposed method seems to perform well for the cases considered here. On the other hand, as expected, we can see some differences as the empirical bias and variance seem to be slightly larger than those obtained under the light-tail error distributions. However, it is clear that both bias and variance became smaller when the error distribution changed from $t(1)$ to $t(5)$ or became less flat. To further see the performance of the proposed method, Figure S.1 displays the estimated varying-coefficient curve obtained with $n = 200$ and $c = 1$ and again indicates that the proposed method can well identify the varying-coefficient function and perform well.

As discussed in the paper, Chen and Tong (2010), Li and Zhang (2012) and Lu and Zhang (2010) discussed the same problem as that considered in the paper but their estimation procedures need some restrictive assumptions or only apply to some limited situations in comparison to that given here. Also Khan and Tamer (2007) investigated a special case of the model discussed in the paper and gave a complex implementation algorithm. Suggested by a reviewer, for the comparison of the proposed method to these methods, we repeated the study that gave the results in Table S.1 but with the error distribution being the t-distribution with the degree of freedoms of 3 and $c = 1$ and presented the results in Table S.2. Note that here we only calculated the estimated bias (Bias) and the sample standard error (SE) for each case and compared the proposed method (Prop) to those given by Khan and Tamer (2007) and Lu and Zhang (2010), which are referred to as PRE and GLE, respectively. This is because some discussion on the comparison between the proposed method and that given in Chen and Tong (2010) has been given in Section 3 of the paper and the method given in Li and Zhang (2012) is similar to that given in Lu and Zhang (2010). In addition to Table S.2, we also obtained the the estimated varying- coefficient curves given by the three methods for the case of $n = 400$ and present them in Figure S.2. One can see from Table S.2 and Figure S.2 that all three methods gave similar performance in general. On the other hand, it is clear that as expected, the proposed method is more robust than the other two and also the proposed method was about four times faster than that given Khan and Tamer (2007) on average.

References

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

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

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

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

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.

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. (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.

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.

S.1 Simulation results with heavy-tail error distributions

Error					Bootstrap I			Bootstrap II		
Dis.	n	c	Bias	SE	MAD	SEE	CP	MAD	SEE	CP
$t(1)$	200	0.5	0.078	0.262	0.271	0.288	93.4	0.270	0.280	93.2
		1	-0.067	0.265	0.269	0.307	94.2	0.269	0.289	93.6
	400	0.5	0.052	0.196	0.202	0.218	95.2	0.200	0.207	94.5
		1	-0.053	0.198	0.203	0.217	95.6	0.201	0.207	95.2
$t(5)$	200	0.5	-0.063	0.197	0.190	0.208	94.3	0.194	0.206	92.9
		1	-0.056	0.190	0.189	0.204	93.6	0.192	0.204	93.0
	400	0.5	-0.047	0.138	0.142	0.145	94.5	0.135	0.141	94.2
		1	-0.035	0.136	0.138	0.144	94.3	0.135	0.137	93.9

S.2 Simulation results on the comparison of three estimation procedures.

	$n = 200$		$n = 400$	
Method	Bias	SE	Bias	SE
Prop	-0.055	0.196	-0.041	0.138
PRE	0.039	0.303	0.006	0.187
GLE	0.031	0.285	0.022	0.205

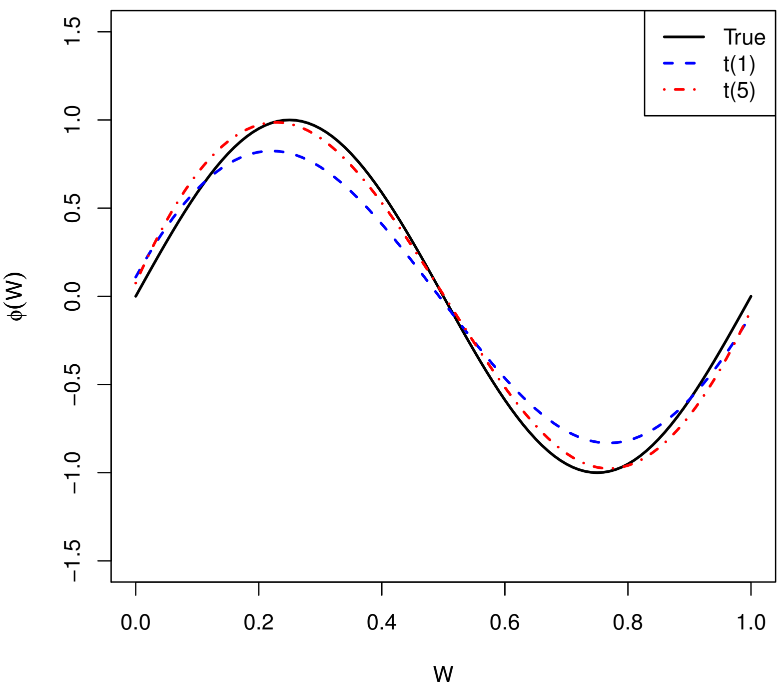

Figure S.1 The estimated varying-coefficient function $\phi(\cdot)$ with heavy-tail error distributions $t(1)$ and $t(5)$.

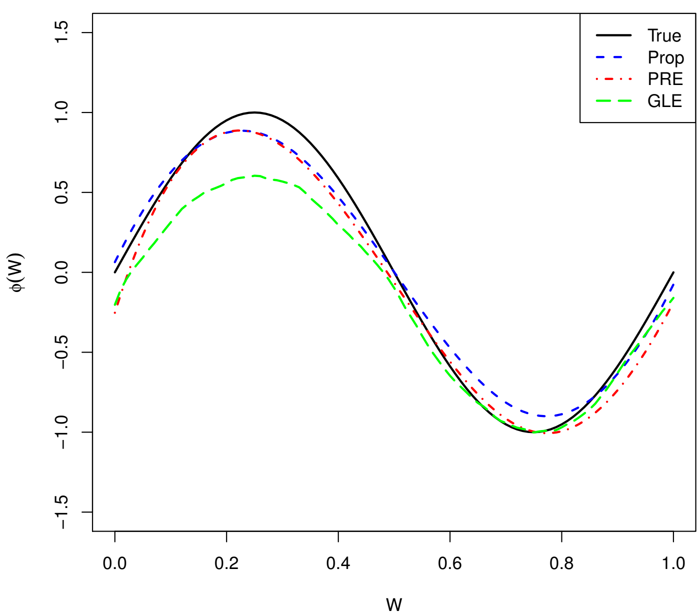

Figure S.2 The estimated varying-coefficient function $\phi(\cdot)$ given by three different methods.